RedFame

Applied Economics and Finance
Vol. 9, No. 1; February 2022
ISSN 2332-7294 E-ISSN 2332-7308
Published by Redfame Publishing
URL: http://aef.redfame.com

Clustering Structure of Microstructure Measures

Liao Zhu[1], Ningning Sun[2], & Martin T. Wells[1]

[1] Department of Statistics and Data Science, Cornell University, Ithaca, New York, 14853, USA

[2] Department of Computer Science, Cornell University, Ithaca, New York,14853, USA

Correspondence: Liao Zhu, Department of Statistics and Data Science, Cornell University, Ithaca, New York, 14853, USA.

Received: December 27, 2021 Accepted: January 31, 2022 Available online: February 8, 2022

doi:10.11114/aef.v9i1.5431 URL: https://doi.org/10.11114/aef.v9i1.5431

Abstract

This paper investigates popular market microstructure measures for stock returns prediction and builds a clustering model for them to study their correlation and the best measures to use as representatives. Using high-dimensional statistical methods, we build the clustering dendrogram and select 20 representatives from all measures. Furthermore, we provide several interesting insights of the market microstructure measures from our clustering results. We found that the time-weighting technique is only useful for Herfindahl-Hirschman Index (HHI) related measures. HHI measures on the number of trades are always redundant. However, the HHI measures on quotes are very important. Also, we find a strong relationship between quote prices and quote shares.

Keywords: market microstructure, interpretable machine learning, artificial intelligence in finance, prototype clustering, high-dimensional statistics, dimension reduction

1. Introduction

Finding the features to explain and predict future stock returns has always been intensely studied in the financial economics literature. For example, Jegadeesh and Titman (1993) show that a stock's past returns predict its future returns, Jegadeesh et al. (1995) showed that short-horizon return reversals are related to the liquidity, which was measured by the bid-ask spread. Glosten et al. (1993) showed that the expected value and the volatility of the nominal excess return on stocks are related. Chan and Fong (2000) mentioned the role of the imbalance of trades and Bessembinder and Seguin (1993) discussed the relationship among price volatility, trading volume, market depth, and stock returns. Apart from that, other market microstructure features can be potentially related, such as the Herfindahl-Hirschman index in Herfindahl (1950).

Furthermore, there are different measures created for particular market features. Take liquidity as an example, the most commonly used measure is the bid-ask spread. However, the most commonly used measure may not always be the best one for all purposes. Some other versions of measure of liquidity may be more insightful, such as the effective bid-ask spread, see Roll (1984). Even for the same measure, taking the last-prevailing value and the time-weighted value may still be different.

With the increase of trading frequency, volume, and the boom of technology, information is more likely to be captured faster. In the paper by Chinco et al. (2017), a simple LASSO model was applied to choose significant factors from about 6000 stocks' 1-minute ahead returns as the sparse and short-lived signals. This suggests the need to use short-term measures instead of long-term ones to predict short-term future returns. Instead of using only the returns from stocks as candidate features to predict short-term future returns, many other features can be taken into account, as stated above. For example, in a 10-second horizon, liquidity may play an important role in predicting future returns.

Considering the above, instead of the traditional approach of studying the impact of one feature at a time, this paper investigates various measures of a larger number of features together and tries to find the clustering structure of those measures. In this way, we can find out the best representatives of the measures and find which measures are ``redundant'' in case of the presence of the others. Considering that there are many related features there are several different market measures. Consequently, several kinds of ways to calculate each measure, we are facing a large number of candidate measures to choose from. Since some of these measures are related to the same feature and some features may be strongly related, the measures are sometimes highly correlated and will cause dimension issues and redundancy if we use them all for stock return predicting. For further prediction of one stock return, we need to consider the impact from a large number of stocks, which will be in a high-dimensional regime. Any redundant measures included for each company can cause large unforeseen difficulties for downstream analysis.

Considering these challenges, it is worthwhile to use high-dimensional statistical methods to reduce the dimension and find out the good representatives of the measures. In the meantime, by high-dimensional statistical methods, the relationship between different features can also be studied. The most traditional method for dimension reduction is the Principal Component Analysis (PCA) method. However, in our case, the covariance decomposition type of methods (including the PCA) has some intrinsic weakness since it can mix the underlying features or separate them into several principal components. In addition, PCA may ignore some features that matter since this type of method is based on variance. A feature with low variance and low loadings in the principal components (and thus ignored by the PCA) can still have a high correlation with the future stock returns. This issue can be well solved by the prototype-clustering due to its interpretability. Using the prototype clustering method to form the clustering structure of the measure, we can have a clear insight and interpretation of the relationship among different measures. These methods were also used in fitting the Adaptive Multi-Factor Models, see Zhu et al. (2020, 2021a); and Jarrow et al. (2021).

In this paper, we first list all the calculations of the measures in Section 2. Then a brief discussion of the statistical methods used is included in Section 3. Section 4 describes our analysis procedure and results and Section 5 concludes.

2. Related Features and Measures

Here is the list of all features and the measures we use in the article.

2.1 Returns

The first is the return. We use mid-quote prices instead of trade prices since trade prices are only effective at the time spot when they occur and are not effective before nor after. However, quotes prices are almost continuous since they are effective for a time period before a trade ends it. Therefore, the equation we use is

$$r_t = \frac{q_t}{q_{t-\Delta t}} - 1.$$

Where r_t and q_t denote the return and mid-quote at time *t*, respectively. In this paper, we use frequency of 10-seconds as our Δt. The mid-quotes are the last prevailing values.

2.2 Bid-Ask Spreads

For the liquidity, we include different measures of the spreads. For each measure, we consider both prevailing and time-weighted values. For each 10-second time interval, the measures for spreads are:

1. Dollar bid-ask (quoted) spread = ask price - bid price.
2. Proportional bid-ask (quoted) spread = (ask price - bid price) / (mid quote).
3. Dollar effective spread: 2 |trade price-mid quote|.
4. Proportional effective spread: $2\left|\frac{trade\ price - mid\ quote}{mid\ quote}\right|$.

2.3 Volatility of Prices

For volatilities, since we are on a short horizon, the standard deviation of returns does not seem reasonable since there may not be enough records for accurate estimation. Consequently, we use the (high - low) for bid prices, ask prices, mid-quote, or trade prices within the 10-second time interval, normalized by the Average Daily Realized Volatility (ADRV) defined by

$$\text{ADRV} = \text{daily ask high - daily bid low}$$

of the prevailing month as measures of volatility. The daily data are from the CRSP database.

2.4 Measures Related to Trades

To measure related the imbalance of trades, we should first classify the trades into buyer-initiated trades and seller-initiated trades. The paper Chakrabarty et al. (2006) gives a nice brief review of these methods and proposed their own improved algorithm. The popular algorithms are: the LR algorithm by Lee and Ready (1991), the EMO algorithm by Ellis et al. (2000) and Chakrabarty et al. (2006). Here we use the CLNV algorithm Chakrabarty et al. (2006) since it is slightly more accurate.

The measures of trade frequency include the count number of all trades within each 10-second time period, and the average time between trades.

The measures of the trade volume are the last prevailing / time-weighted average value of: dollar amount / number of shares / number of shares normalized by ADRV of the prevailing month.

For the imbalance of trades, for each 10-second time period, the equations are

$$\frac{buyer - seller}{buyer + seller} \text{ (directional) or } \left|\frac{buyer - seller}{buyer + seller}\right| \text{ (non-directional)}$$

We can use the previous measures for trade frequencies and trade volumes for the buyer and seller in the equations above.

2.5 Measures Related to Quotes

The measures of the quote frequencies can be the count number of: all quote records / bid changes / ask changes, and the average time between the quote records/ bid changes/ ask changes within the 10-second intervals. The “quote change” measures only count for quotes with different quote prices, as a measure of new information. Note that for the count number of bid changes, we will count the consecutive same bid prices as only 1 count. Similarly, for asks, we include average time between quotes / quote changes here. It is almost reciprocal (multiplied by a constant) of the count number. But since it is not a linear function of count number, it is totally different from the count number in a regression.

The depths of the market are based on the quotes for each stock, as a measure of liquidity and an indicator of price movement direction. The measures related to the depth of the market can be:

1. The last prevailing ask / bid / ask - bid / |ask - bid|.
2. The time weighted values, specifically, $\int ask_t dt, \int bid_t dt, \int (ask_t - bid_t)dt, \int |ask_t - bid_t|dt.$

For each ask (or bid) in the equation above, we can use dollar volume / number of shares / number of shares normalized by Average Daily Trading Volume (ADTV) of the prevailing month. We can consider the measure for quotes in each exchange or the best quote nation-wide.

We also include the imbalance of quotes using a similar expression with that of imbalance of trades. For each 10-second time interval, the equations are:

$$\frac{ask - bid}{ask + bid} \text{ (directional) or } \left|\frac{ask - bid}{ask + bid}\right| \text{ (non-directional)}$$

We use the previous measures for quote frequencies and depth for the Ask and Bid in the equations above. In addition to calculating the time weighted value and then plug in the fraction as before and compute the time-weighted averages of the fractions:

$$\int \frac{(ask\ size)_t - (bid\ size)_t}{(ask\ size)_t + (bid\ size)_t} dt \text{ and } \int \left|\frac{(ask\ size)_t - (bid\ size)_t}{(ask\ size)_t + (bid\ size)_t}\right| dt.$$

Note that since we have them in both nominator and denominator, normalizing by ADTV will not make any difference now, in other words, these measures are already normalized.

2.6 Concentration among Exchange Places

The concentration among exchange places can be measured by the Herfindahl-Hirschman Index (HHI) of concentration, which is defined as

$$HHI_v = \sum_{i=1}^{N} v_i^2$$

where v_i denotes the fraction of the value of the *i-th* exchange.

The original measure didn’t use the fraction, but directly uses the shares of each part. However, in our case, using fractions is better because we don’t want the difference of shares, volume, etc. across different stocks to be taken into account, since they are taken care of in other measures. And since we use the HHI to measure the concentration of trades and quotes for each single company, normalizing by average daily value is no longer needed. For each 10-second time interval, the choices of values (vi) can be the measures of trade frequency, trade volume, quote frequency, or the depth.

Considering the ensemble of measures above, we have more than 100 measures in total, and many of them are highly correlated. Therefore, we use a statistical approach to remove redundancies and find informative representatives.

3. Statistical Methods

This section describes the prototype clustering to be used to efficiently deal with the problem of high correlation among the measure. To remove unnecessary independent variables, using clustering methods, we classify them into homogeneous groups and then choose representatives from each group with small pairwise correlations. First, we define a distance metric to measure the similarity between points (in our case, the returns of the independent variables). Here, the distance metric is related to the correlation of the two points, i.e.

$$d(r_1, r_2) = 1 - |corr(r_1, r_2)| \tag{1}$$

where $r_i = \left(r_{i,t}, r_{i,t+1}, \dots, r_{i,T}\right)'$ is the time series vector for independent variable $i = 1, 2$ and $corr(r_1, r_2)$ is their correlation. Second, the distance between two clusters needs to be defined. Once a cluster distance is defined, hierarchical clustering methods (see Kaufman and Rousseeuw, 2009) can be used to organize the data into trees.

In these trees, each leaf corresponds to one of the original data points. Agglomerative hierarchical clustering algorithms build trees in a bottom-up approach, initializing each cluster as a single point, then merging the two closest clusters at each successive stage. This merging is repeated until only one cluster remains. Traditionally, the distance between two clusters is defined as either a complete distance, single distance, average distance, or centroid distance. However, all of these approaches suffer from interpretation difficulties and inversions (which means parent nodes can sometimes have a lower distance than their children). To avoid these difficulties, Bien and Tibshirani (2011) introduced hierarchical clustering with prototypes via a minimax linkage measure, defined as follows. For any point *x* and cluster *C*, let

$$d_{max}(x, C) = \max_{x' \in C} (x, x')$$

be the distance to the farthest point in *C* to *x*. Define the minimax radius of the cluster *C* as

$$r(C) = \min_{x \in C} d_{max}(x, C)$$

that is, this measures the distance from the farthest point $x \in C$ which is as close as possible to all the other elements in *C*. We call the minimizing point the prototype for *C*. Intuitively, it is the point at the center of this cluster. The minimax linkage between two clusters *G* and *H* is then defined as

$$d(G, H) = r(G \cup H)$$

Using this approach, we can easily find a good representative for each cluster, which is the prototype defined above. It is important to note that minimax linkage trees do not have inversions. Also, in our application as described below, to guarantee interpretability and tractability, using a single representative independent variable is better than using other approaches (for example, PCA) which employ linear combinations of the independent variables.

Similar approaches can also be found in the literature related to the Adaptive Multi-Factor Model, see Zhu et al. (2020, 2021b, a); Jarrow et al. (2021); and Zhu (2020). This proposed method can be potentially used with time series analysis, see Zhao et al. (2020, 2021), and some other machine learning models, such as Huang et al. (2021, 2020); Li et al. (2021); Jie et al. (2018); Jie (2018); Zhang et al. (2021); Mao et al. (2020b, a); Stein et al. (2021b, 2020, 2021a); Du et al. (2021); Li (2021); and Bo et al. (2021).

4. Analysis Procedure and Results

We use the Daily TAQ database to form all the measures for each 10-second period. The average daily values (average daily trading volume, average daily realized volatility, etc.) are calculated for the prevailing month from the CRSP database. In this paper, we use the day April 3rd, 2018 as an example, and all the daily average value is calculated in the prevailing month, which is from March 2nd, 2018 to April 2nd, 2018.

We calculate the distance between measures by Equation (1) for each company among all 7263 companies in the database. Then we take the average of distances of each pair of measures over all the companies. The first prototype clustering dendrogram result is in Figure 1. There are 91 measures in total, but some of them are nearly perfectly correlated. The following are the reasons for the strong correlation between some measures.

From the dendrogram, it is clear that the measures related to dollar volume always have a strong correlation with that related to the number of shares, since the price within a day does not change very much. Since we already have measures related to prices, here we should always use the measures related to the number of shares and remove redundant ones based on the dollar volume.

Also, the dollar effective spread always has a strong correlation with proportional effective spread, since mid-quotes do not change much during one day. Therefore, in the short horizon, we can abandon the dollar effective spread and only use the proportional effective spreads since they have normalization and will be comparable across the stocks.

Apart from the normalizing issues above, the measures normalizing by the average daily trading volume (or daily volatility) has a correlation equal to one with their original measures. Consequently, we only need one of each normalized / non-normalized pair. Since we want to make measures stable across all stocks, it is better to use the normalized ones. Therefore, we remove all non-normalized measures if there is an appropriate normalized one. The dendrogram after removing these redundant measures looks neater in Figure 2. Note there are no longer perfect correlated measures in Figure 2.

Finally, we use prototype clustering to find good representatives (prototypes) of measures within each cluster. We require the absolute correlation between prototypes to be no higher than 0.3 (in other words, distance less than 0.7). The dendrogram of the prototypes selected is shown in Figure 3. There are 20 measures (out of 91 original measures) selected. The list of descriptions of the selected measures is in Table 1.

Comparing the clustering results shown in Figure 3 and all non-redundant measures in Figure 2, we found some interesting patterns.

The time-weighting technique is only useful for HHI-related measures. For non-HHI measures, time-weighted measures are mostly redundant with the presence of last-prevailing values. All the 12 non-HHI related time-weighted measures in Figure 2 are redundant. The explanation is that the non-HHI-related measures do not change too much over the short horizon, so the time-weighted measures are not too different from the corresponding last-prevailing measures which are easier to compute.

HHI measures on the number of trades/buys/sells are always redundant. However, the HHI measures on quotes are very useful. This indicates that the market is quite efficient in terms of the actual trading counts across exchange places, although institutions may try to behave differently in terms of quotes aiming to search for arbitrage opportunities. It follows that the HHI measures on the number of trades are redundant, but HHI measures on quotes provide crucial extra information.

The clustering structure also reveals many interesting relationships between the measures. For example, the last prevailing bid-ask spread is strongly correlated with the last prevailing (ask shares - bid shares) / (average daily trading shares). This suggests a strong relationship between quote prices and quote shares. In the clustering result, with the presence of the last-prevailing bid-ask spread, the last prevailing (ask shares - bid shares) / (average daily trading shares) does not provide much extra information. Further studies can be done by using the selected prototype measure to predict stock returns. Another fruitful research direction is to investigate why some measures (e.g. the 2 measures mentioned above) are so similar that they are clustered into one group.

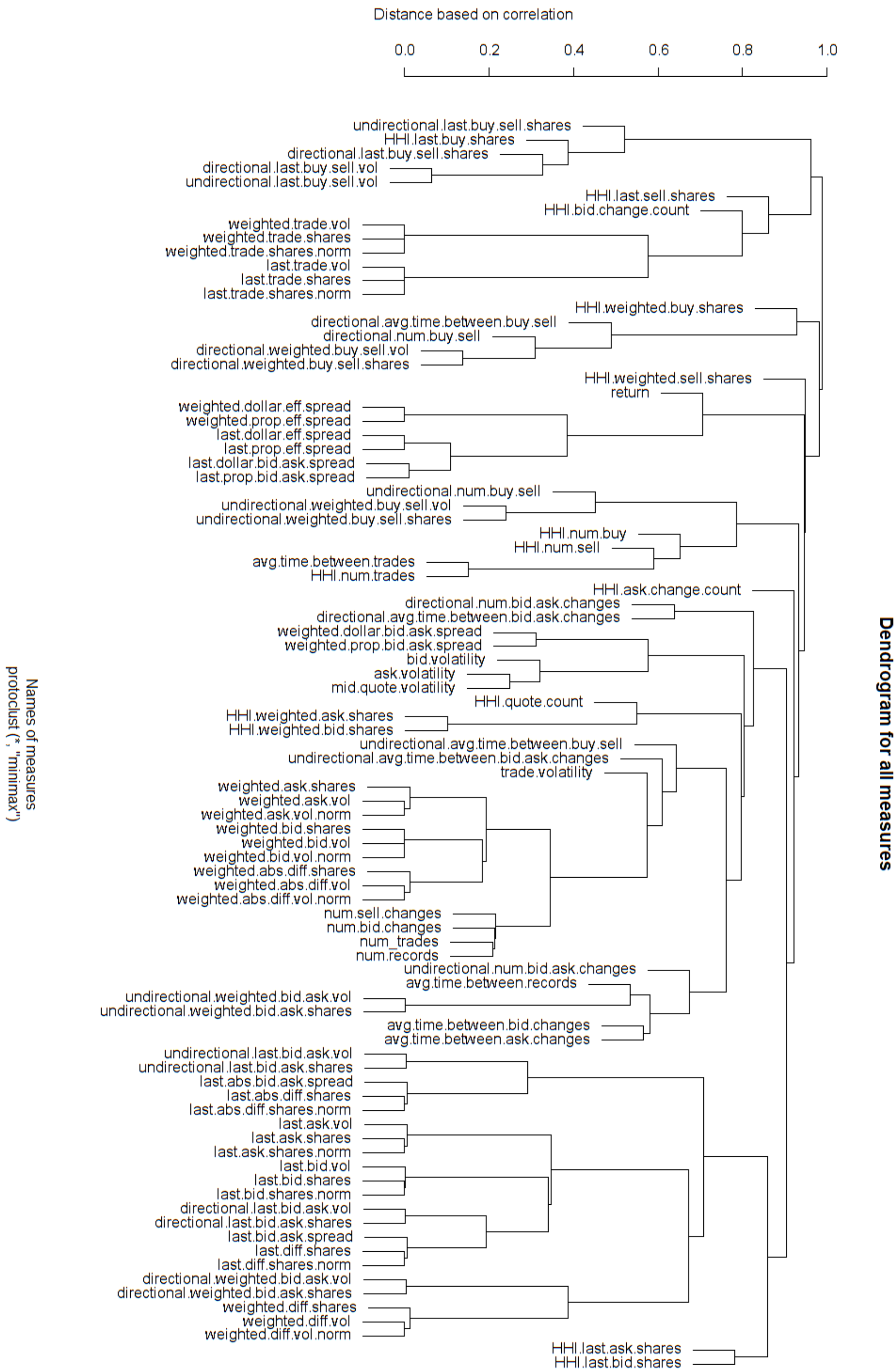

Figure 1. Dendrogram of all measures

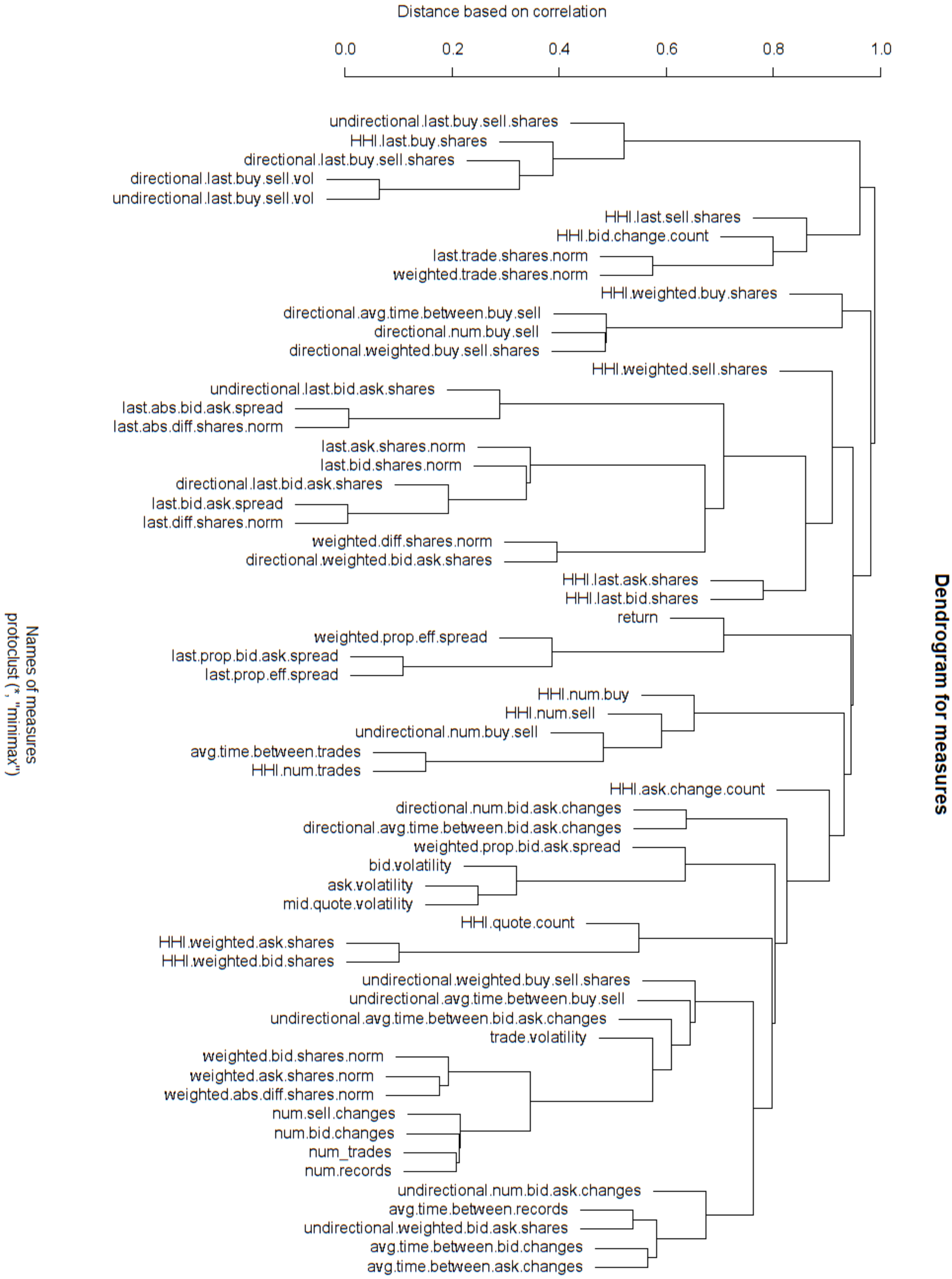

Figure 2. Dendrogram of measures after removing redundant ones

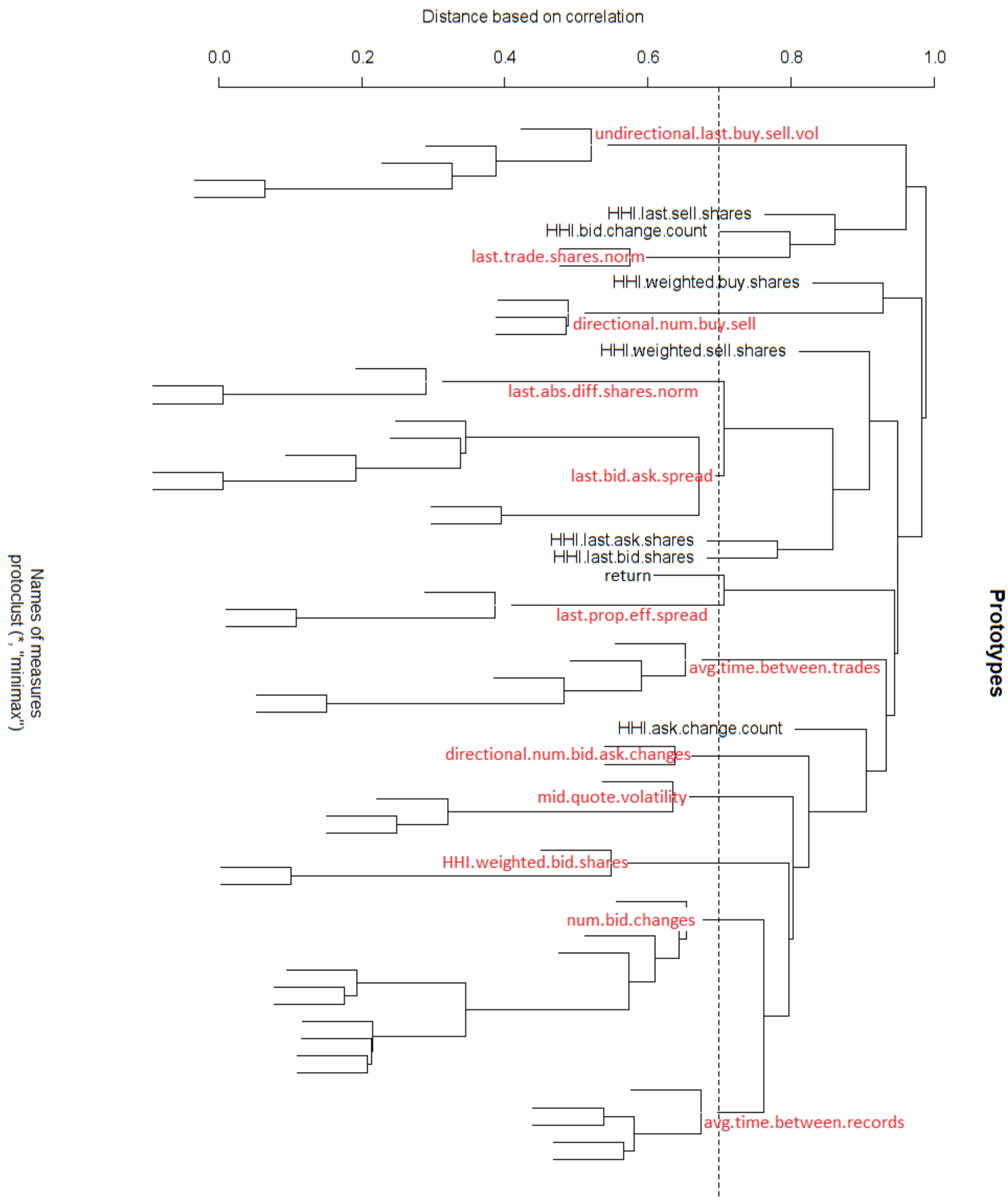

Figure 3. Prototypes selected

Table 1. List of all selected measures

No.	Names	Descriptions
1	avg.time.between.records	The average time between quote records
2	avg.time.between.trades	The average time between trades
3	directional.num.bid.ask.changes	(ask - bid) / (bid - ask) where ask (bid) in the equation means the number of ask (bid) changes
4	directional.num.buy.sell	(buy - sell) / (buy + sell) where buy (sell) in the equation means the number of buy (sell) trades
5	HHI.ask.change.count	HHI of count of ask changes
6	HHI.bid.change.count	HHI of count of bid changes
7	HHI.last.ask.shares	HHI of time-last ask shares
8	HHI.last.bid.shares	HHI of last prevailing bid shares
9	HHI.last.sell.shares	HHI of last sell prevailing shares
10	HHI.weighted.bid.shares	HHI of time-weighted bid shares
11	HHI.weighted.buy.shares	HHI of time-weighted buy shares
12	HHI.weighted.sell.shares	HHI of time-weighted sell shares
13	last.abs.diff.shares.norm	The last prevailing (ask shares - bid shares) normalized by the Average daily trading shares over the last month
14	last.bid.ask.spread	The last prevailing bid-ask spread
15	last.prop.eff.spread	The last prevailing proportional effective spread
16	last.trade.shares.norm	The last prevailing trade shares normalized by The average daily trading shares
17	mid.quote.volatility	Volatility based on mid-quote
18	num.bid.changes	The number of bid changes
19	return	Return
20	undirectional.last.buy.sell.vol	|buy - sell| / |buy + sell| where buy (sell) in the equation means shares of buy (sell) trades

5. Conclusion

This paper investigates the popular market microstructure measures for stock returns predicting and builds the clustering model for them to study their correlation and the best measures to use as representatives. Using high-dimensional statistical methods, we build the clustering dendrogram and select 20 representatives from all measures, and meanwhile study the relationships between measures. A clear clustering structure is given by the dendrogram in Figure 3.

From the clustering results, we provide several interesting insights of the market microstructure measures. We found that the time-weighting technique is only useful for HHI-related measures. For non-HHI measures, time-weighted measures are mostly redundant with the presence of last-prevailing values. HHI measures on the number of trades/buys/sells are always redundant. However, the HHI measures on quotes are very important. We find a strong relationship between quote prices and quote shares.

Future work can be done to use the selected measures to fit more exotic statistical models, such as time series models, machine learning models, etc., to predict the stock returns. On the other hand, another fruitful research direction is to investigate the reasons on intuitively less correlated measures appear to be close in the cluster structure

References

Bessembinder, H., & Seguin, P. J. (1993). Price volatility, trading volume, and market depth: Evidence from futures markets. *Journal of financial and Quantitative Analysis, 28*(1), 21-39. https://doi.org/10.2307/2331149

Bien, J., & Tibshirani, R. (2011). Hierarchical clustering with prototypes via minimax linkage. *Journal of the American Statistical Association, 106*(495), 1075-1084. https://doi.org/10.1198/jasa.2011.tm10183

Bo, D., Hwangbo, H., Sharma, V., Arndt, C., & TerMaath, S. C. (2021). *A subspace-based approach for dimensionality reduction and important variable selection.*

Chakrabarty, B., Li, B., Nguyen, V., & Van Ness, R. (2006). Trade classification algorithms for electronic communication networks. *Journal of Banking and Finance, 31*, 3806-3821. https://doi.org/10.1016/j.jbankfin.2007.03.003

Chan, K., & Fong, W. M. (2000). Trade size, order imbalance, and the volatility–volume relation. *Journal of Financial*

Economics, 57(2), 247-273. https://doi.org/10.1016/S0304-405X(00)00057-X

Chinco, A. M., Clark-Joseph, A. D., & Ye, M. (2017). *Sparse signals in the cross-section of returns.* Technical report, National Bureau of Economic Research. https://doi.org/10.3386/w23933

Du, W., Hyeon, E., Pei, H., Huang, Z., Cheong, Y., Zheng, S., & Galvanauskas, A. (2021). *Improved machine learning algorithms for optimizing coherent pulse stacking amplification.* In CLEO: Science and Innovations, pages JTh3A–1. Optical Society of America. https://doi.org/10.1364/CLEO_AT.2021.JTh3A.1

Ellis, K., Michaely, R., & O'Hara, M. (2000). The accuracy of trade classification rules: Evidence from Nasdaq. *Journal of Financial and Quantitative Analysis, 35*(4), 529-551. https://doi.org/10.2307/2676254

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance, 48*(5), 1779-1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

Herfindahl, O. C. (1950). *Concentration in the steel industry.* PhD thesis, Columbia University New York.

Huang, X., Li, Z., Lu, J., Wang, S., Wei, H., & Chen, B. (2020). Time-series clustering for home dwell time during COVID-19: what can we learn from it? *ISPRS International Journal of Geo-Information, 9*(11), 675. https://doi.org/10.3390/ijgi9110675

Huang, X., Lu, J., Gao, S., Wang, S., Liu, Z., & Wei, H. (2021). Staying at home is a privilege: Evidence from fine-grained mobile phone location data in the united states during the covid-19 pandemic. *Annals of the American Association of Geographers, 0*(0), 1- 20. https://doi.org/10.1080/24694452.2021.1943301

Jarrow, R. A., Murataj, R., Wells, M. T., & Zhu, L. (2021). *The low-volatility anomaly and the adaptive multi-factor model.* https://doi.org/10.2139/ssrn.3834026

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance, 48*(1), 65-91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Jegadeesh, N., & Titman, S. et al. (1995). Short-horizon return reversals and the bid-ask spread. *Journal of Financial Intermediation, 4*(2), 116-132. https://doi.org/10.1006/jfin.1995.1006

Jie, C. (2018). *Decision Making Under Uncertainty: New Models and Applications.* PhD thesis, University of Maryland, College Park.

Jie, C., Prashanth, L., Fu, M., Marcus, S., & Szepesv ́ari, C. (2018). Stochastic optimization in a cumulative prospect theory framework. *IEEE Transactions on Automatic Control, 63*(9), 2867-2882. https://doi.org/10.1109/TAC.2018.2822658

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis, 344.* John Wiley & Sons.

Lee, C., & Ready, M. J. (1991). Inferring trade direction from intraday data. *The Journal of Finance, 46*(2), 733-746. https://doi.org/10.1111/j.1540-6261.1991.tb02683.x

Li, J. (2021). *Two ways towards combining sequential neural network and statistical methods to improve the prediction of time series.*

Li, Y., Song, K., Sun, Y., & Zhu, L. (2021). *Frequentnet: A novel interpretable deep learning model for image classification.* https://doi.org/10.2139/ssrn.3895462

Mao, Y., Fu, Y., Zheng, W., Cheng, L., Liu, Q., & Tao, D. (2020a). *Speculative container scheduling for deep learning applications in a kubernetes cluster.* https://doi.org/10.20944/preprints202010.0534.v1

Mao, Y., Yan, W., Song, Y., Zeng, Y., Chen, M., Cheng, L., & Liu, Q. (2020b). *Differentiate quality of experience scheduling for deep learning applications with docker containers in the cloud.*

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance, 39*(4), 1127-1139. https://doi.org/10.1111/j.1540-6261.1984.tb03897.x

Stein, S. A., Baheri, B., Chen, D., Mao, Y., Guan, Q., Li, A., Fang, B., & Xu, S. (2021a). *Qugan: A quantum state fidelity based generative adversarial network.* In 2021 IEEE International Conference on Quantum Computing and Engineering (QCE), pages 71-81. IEEE. https://doi.org/10.1109/QCE52317.2021.00023

Stein, S. A., Baheri, B., Chen, D., Mao, Y., Guan, Q., Li, A., Xu, S., & Ding, C. (2021b). *Quclassi: A hybrid deep neural network architecture based on quantum state fidelity.*

Stein, S. A., Baheri, B., Tischio, R. M., Chen, Y., Mao, Y., Guan, Q., Li, A., & Fang, B. (2020). *A hybrid system for learning classical data in quantum states.* https://doi.org/10.1109/IPCCC51483.2021.9679430

Zhang, Y., Du, T., Sun, Y., Donohue, L., & Dai, R. (2021). *Form 10-q itemization.* https://doi.org/10.1145/3459637.3481989

Zhao, Y., Landgrebe, E., Shekhtman, E., & Udell, M. (2020). *Online missing value imputation and correlation change detection for mixed-type data via Gaussian copula.* https://doi.org/10.1145/3394486.3403106

Zhao, Y., Matteson, D. S., Nebel, M. B., Mostofsky, S. H., & Risk, B. (2021). *Group linear non-gaussian component analysis with applications to neuroimaging.*

Zhu, L. (2020). *The Adaptive Multi-Factor Model and the Financial Market. eCommons.*

Zhu, L., Basu, S., Jarrow, R. A., & Wells, M. T. (2020). High-dimensional estimation, basis assets, and the adaptive multi-factor model. *The Quarterly Journal of Finance, 10*(04), 2050017. https://doi.org/10.1142/S2010139220500172

Zhu, L., Jarrow, R. A., & Wells, M. T. (2021a). Time-invariance coefficients tests with the adaptive multi-factor model. *The Quarterly Journal of Finance*, *11*(04), 2150019. https://doi.org/10.1142/S2010139221500191

Zhu, L., Wu, H., & Wells, M. T. (2021b). *A news-based machine learning model for adaptive asset pricing.*