12-2018

Commonality: A longitudinal study

Raja VELU
Syracuse University

Zhaoque ZHOU
Syracuse University

Chyng Wen TEE
Singapore Management University, cwtee@smu.edu.sg

Commonality: A Longitudinal Study

Raja Velu* Zhaoque Zhou* Chyng Wen Tee†‡

ABSTRACT

Commonality in asset characteristics such as returns, order flows, liquidity, and other non-trade parameters has attracted intensive research interest in the literature. In this paper, we investigate the trend in commonality by performing a longitudinal study for the duration covering 2000-2016, a period that includes periods of boom and bust in the financial market. We develop a unified methodology to accommodate systematically all factors that may better explain the commonality. The relationship between market microstructure models and the statistical representation framework of commonality is also explored in great detail. Finally, we demonstrate that commonality has increased over time, though exchange-wide variation has reduced. This points to increased information efficiency and high-frequency trading as the cause.

Keywords: *Market microstructure; liquidity measures; commonality; structural models.*

*Department of Finance, Whitman School of Management, Syracuse University, Syracuse, New York 13244.

†Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road #05-01, Singapore 178899.

‡Corresponding Author: Tel: +65 6828 0819. Email: `cwtee@smu.edu.sg`.

1. Introduction

While the idea of investment through portfolios has become so ingrained around the world, microstructure research has focused mainly on the study of single security setting. Since the early 2000s, both academics and practitioners have paid more attention to the magnitudes of cross-sectional interactions between stocks at the microstructure level. However, the study of commonality in short-horizon returns, order flows, and liquidity is still of interest in the microstructure analysis of equity markets. Following Hasbrouck and Seppi (2001b), who find that both short-horizon returns and order flows are characterized by common factors, two new research questions have emerged.

First, are the commonalities in short-horizon returns and order flows stable over time? Otherwise, do the results still hold in recent years? Although Kamara, Lou, and Sadka (2008) provide the evidence that the cross-sectional variation of liquidity commonality has increased over the period 1963-2005 using daily data, the time-variant commonality in liquidity does not necessarily imply that the commonalities in returns and order flows are not stable. Liquidity commonality can easily arise when trading activity runs in different directions for different stocks, since both heavy buyer-motivated trading and heavy sell-motivated trading can strain liquidity. But commonality in returns can arise because of less firm-specific and more market-wide, public information flows and also because of correlated order imbalances with the same sign across stocks. Furthermore, commonality in order flows may be influenced by the differential liquidity of individual stocks as well as by other factors such as asymmetric information, idiosyncratic risks, transaction costs and other forms of market imperfections. In this paper, we document that the commonality in short-horizon returns slightly increases from 2000 to 2016, but the order flows commonality decreases from 2000 to 2016.

Second, does commonality in stocks' order flows account for the covariance structure of short-term returns? How to characterize relationships involving returns and order flows? Microstructure research focuses on how price adjusts to new information. If the market is efficient, new information would be immediately disseminated and interpreted by all market participants, thus prices would immediately adjust to a new equilibrium value determined by the content of the information. But in practice, the price adjustment is not processed at the same speed for all stocks. Therefore, the price discovery and order flow dynamics have more complex relationship when we consider multiple assets at the same time.

The remainder of the paper is organized as follows. Section 2 outlines the literature review and research hypotheses that have been studied before. We describe the data and empirical methodologies in Section 3. Section 4 present the unified model we developed and the empirical results. We extend MRR structural model to multiple stocks case that is a special case of our unified model in Section 5. Section 6 provides several possible explanations for the change of commonality. The final section concludes with an outline of possible future work.

2. Literature Review

2.1. Commonality in financial markets

The degree of returns commonality determines how market participants can reduce systematic risk via effective diversification. Moreover, investors are also concerned about liquidity systematic variation, referred to as "commonality in liquidity". Therefore, a better understanding of what causes commonality can help investors to constitute their portfolio more efficiently.

In recent years, a stream of research studies commonality in liquidity starting with Chordia, Roll, and Subrahmanyam (2000),Huberman and Halka (2001),Hasbrouck and Seppi (2001b).They find that the liquidity of individual securities co-moves with each other, which constitutes an undiversifiable risk factor in financial markets. However,despite the solid evidence on liquidity commonality, the source of commonality in liquidity is still an open question. Both demand-side and supply-side explanations for commonality in liquidity have been proposed. The demand-side explanations suggest that the liquidity commonality could be driven by correlated trading demands and/or correlated sentiment. On the other hand, the supply-side explanations suggest that shocks to the funding liquidity cause the liquidity commonality.

The determinants of commonality in returns have also been widely studied. Researchers provide evidence that the degree of returns commonality is influenced by various market frictions,such as investors reallocate funds based on the performance of different style assets, correlated trading with different groups of traders, the information diffusion in different stocks with different speed.

2.2. *Modeling for commonality*

Many papers study the empirical evidence on whether liquidity commonality exists and what factors drive commonality in liquidity over time. The focus on a single asset may lead to inventory risk due to lack of diversification. Even after controlling for individual asset's liquidity determinants, there is some commonality among the assets. Chordia et al. (2000) are among the first to study the common factors and correlated movements of liquidity based on 1169 NYSE stocks on the 254 trading days during 1992. Applying a simple and intuitive market model regression to each stock, they compute various liquidity measures (quoted spreads, effective spreads, and quoted depths) defined below in Table 1 and regress changes in individual stock liquidity on changes in market liquidity. Defining $DL_t = (L_t - L_{t-1})/L_{t-1}$, where L_t is a liquidity measure, the following regression model is constructed.

$$DL_{j,t} = \alpha_j + \beta_j DL_{M,t} + \varepsilon_{j,t} \tag{1}$$

Here $DL_{M,t}$ is simply the percentage change in the cross-sectional average excluding the j-th stock of the same liquidity variable. The additional variables included in (1) are $DL_{M,t-1}$, $DL_{M,t+1}$, the market return variable $r_{M,t}$, $r_{M,t-1}$, $r_{M,t+1}$, and squared return $r^2_{j,t}$. The lead, lag variables are meant to account for dependence between returns and spread measures. The squared return is a proxy for volatility that may influence liquidity. The commonality is measured mainly through the significance of 'β_j' coefficients. Chordia et al. (2000) report that around 84% of the contemporaneous slope coefficients are positive and statistically significant for approximately 33% of the sample stocks. The economic interpretation is that individual stock liquidity, in terms of both spreads and depth, co-moves with market-wide liquidity.

Table 1
Liquidity variables in Chordia et al. (2000)

Liquidity measure	Acronym	Definition	Units
Quoted spread	QSPR	$P_A - P_B$	$
Proportional quoted spread	PQSPR	$(P_A - P_B)/P_M$	None
Depth	DEP	$\frac{1}{2}(Q_A + Q_B)$	Shares
Effective spread	ESPR	$2\,\lvert P_t - P_M\rvert$	$
Proportional effective spread	PESPR	$2\,\lvert P_t - P_M\rvert\,/P_t$	None

P denotes price and subscripts indicate: t=actual transaction, A=ask, B=bid, M=bid-ask midpoint. Q denotes the quantity guaranteed available for trade at the quotes, (with subscripts: A=ask, B=bid).

Huberman and Halka (2001) study the presence of common factor among liquidity proxies of different stocks. The proxies-spread, spread/price ratio, depth in shares and depth in dollars-are adjusted for their time series dependence and discreteness. The time series are modeled via autoregressive process for mutually exclusive sets of stocks and if the residuals from different sets are correlated, it is taken to indicate the presence of commonality. As the determinants of common movements, two sources, the cost of holding inventory and adverse selection, are examined. The inventory cost is captured by interest rates, by market-wide shocks such as shifts in the yield curve and the adverse selection component is reflected by the depth related variables.

Hasbrouck and Seppi (2001b) utilize the factor models framework to describe the cross-section of returns, orders and public non-trade information for the 30 Dow stocks. Using principal components analysis that relies solely on the variance-covariance matrix to extract commonality, they find that both returns and order flows are characterized by common factors. Furthermore, they introduce canonical correlation analysis to analyze the relations between returns commonality and order flows commonality, and provide the evidence that commonality in order flows explains roughly two-thirds of the commonality in returns. More precisely, their model can be stated as follows:

Let r_t be the 30-dimensional vector of return and x_t be the vector of an order flow measure. Then the following factor models are constructed:

$$\begin{aligned} r_t &= \varphi G_t + \eta_t \\ x_t &= \theta F_t + \varepsilon_t \end{aligned} \tag{2}$$

The estimated factors G_t and F_t are then related through a linear model. It is shown that the residuals that result from this model (ω_t) can have additional factor structure. This can be more completely summarized via multivariate regression:

$$\begin{aligned} r_t &= \Lambda x_t + u_t \\ u_t &= \xi H_t + \omega_t \end{aligned} \tag{3}$$

Here H_t are the factors of the residuals. The correspondence is clear, if (2) are substituted in the firt equation in (3) as

$$\varphi G_t = \Lambda(\theta F_t + \varepsilon_t) + u_t - \eta_t$$
$$= \Lambda\theta F_t + \omega_t \tag{4}$$

We want to make some observations here: It can be taken that the residuals, u_t, capture public non-trade information. The contribution of other stocks order flows beyond a stock's own order flow must be isolated to separate the commonality factor.

2.3. Commonality of Asset Characteristics

2.3.1. Return commonality: Results

Many factors can drive commonality in returns. The traditional view is that return commonality is driven by comovement in fundamentals. However, in economies with frictions, return commonality is delinked from comovement in fundamentals. Barberis, Shleifer, and Wurgler (2005) classify the explanations for return commonality into three groups. First, commonality could be driven by style investing. Barberis and Shleifer (2003) argue that some investors categorize assets into different styles and allocate funds between style categories based on relative performance. This type of style investing leads to common factors in returns of assets within the same style.

A second explanation is based on habitat investing. Some securities may be held and traded by only particular subsets of investors. For example, individual investors tend to hold small stocks and closed-end funds. As the sentiment or liquidity needs or risk preferences of these investors change, they alter their holding in their habitat, thereby leading to a common factor in the returns of the securities they hold. Notably, style and habitat-based explanations suggest that common factors reflect correlated trading decisions within specific groups of traders.

A third potential explanation for commonality is related to the speed of information diffusion in different stocks. Barberis et al. (2005) introduce a bivariate regression to show that betas relative to the S&P portfolio increase, while betas relative to non-S&P stocks decrease following S&P additions. The opposite results are found following S&P deletions. These findings provide the evidence to support the friction- and sentiment-based explanations. Furthermore, when they decompose the friction, sentiment effects and include five lead and lag S&P and non-S&P returns in bivariate regression to identify the effect of information diffusion, Barberis et al. (2005) find that at least a portion of the effect is driven by differences in information diffusion across stocks.

2.3.2. Liquidity commonality: Results

Researchers offer various hypotheses of why liquidities co-move. These hypotheses can be broadly grouped into supply or demand based. The supply side hypotheses focus on the role of funding constraints of financial intermediaries. Significant market declines or high volatility increase the demand for liquidity as agents liquidate their positions across many assets and reduce the supply of liquidity as liquidity suppliers hit their capital constraints. So commonality in liquidity arises and is intensified during periods of large market declines or high market volatility. Coughenour and Saad (2004) identify specialist portfolios, use the market model method employed by Chordia et al. (2000) and find commonality in liquidity among NYSE stocks handled by the same specialist firm. This key argument is a result of shared capital and information among specialists within a firm. Hameed, Kang, and Viswanathan (2010) add the downside dummy variable to show that the asymmetric effect of market returns on liquidity exists. Furthermore, they consider the capital constraints and design three proxies to capture tightness of capital in the market. The first proxy is the excess returns on the portfolio of financial intermediaries (SIC code 6211). A negative aggregate return in the firms operating in investment banking and securities brokerage services implies a weak aggregate balance sheet of the funding sector and high capital constraints. The second proxy is the weekly changes in aggregate repurchase agreements(repos). When financial intermediaries have weak balance sheets, their leverage is too high. These intermediaries will contract their balance sheets through repos. Hence, a decline in aggregate repos means the funding market is capital constrained. The third proxy relies on the weekly spread in commercial paper (CP), measured as the difference in the weekly returns on the 3-month CP rate and 3-month Treasury bill rate. This CP spread reflects a liquidity premium that are related to the willingness of the financial intermediaries to provide liquidity. Finally, Hameed et al. (2010) conclude that commonality in liquidity on the NYSE increases during market declines, especially when funding liquidity is tight. The findings suggest that spillover effects among securities during market declines are important and provide strong support that the contagion in illiquidity is due to supply effects.

The demand side theory postulates that liquidity commonality arises mainly due to the correlated trading behavior of institutional investors. Using quarterly institutional ownership data from CDA/Spectrum database, Kamara et al. (2008) provide evidence that the increase in commonality in liquidity among U.S. large-cap stocks in particular over the past 25 years can be attributed to the increased institutional and index-related trading for these stocks. Koch, Ruenzi, and Starks (2016) use the same database and design a turnover-weighted measure of mutual fund ownership as a proxy for correlated trading. They show that stocks with higher mutual fund ownership and stocks owned by mutual funds with high turnover or funds that experience liquidity shocks exhibit greater commonality in liquidity. The intuition is that growing institutional ownership may give rise to correlated trading across stocks, which, in turn, creates common buying or selling pressure, and thus higher levels of common variation in liquidity.

Some relent work toward commonality include price-based return comovement.Green and Hwang (2009) argue that similarly priced stocks move together. The patterns are confirmed even after stock splits. Although this phenomenon cannot be explained by economic theory, it is empirically confirmed that investors categorize stocks based on price. The evidence is captured through the differences in regression coefficients before and after the split. Chen, Singal, and Whitelaw (2016) argue that the bivariate regression of previous studies Barberis et al. (2005) and Green and Hwang (2009) provide little information about the comovement and hence suggest matched case control study using robust univariate regressions. It appears that the excess comovement does not seem to hold, but the increase in betas is mainly due to the cross-sectional momentum effect.

2.3.3. Market Quality Commonality: Results

Not only returns and liquidity, but also some other asset characteristics such as transaction costs and realized volatility display a tendency to move together in aggregate. Marsh and Mazza (2018) investigate a set of market quality proxies that encompasses all the liquidity dimensions, market efficiency, transaction costs and realized volatility measures. They show that the presence of market quality comovements in diversified aggregations of Euro-zone stocks and suggest that these comovements are mainly driven by country-specific effects.

3. Longitudinal study

3.1. Data

The data for this study are from the TAQ database of the New York Stock Exchange, which contains trade-by-trade data of all listed stocks. TAQ records transactions prices and quantities of all trades, as well as a record of all stock price quotes that were made. Our sample begins in January 2000 and ends in December 2016. For our sample of equities, we choose the firms in the Dow Jones Industrial Average index constituent.

We choose the Dow stocks as our sample because first, the rapid pace of trading provides frequently updated prices and allows us to construct some high-frequency grading measures. Second, these 29 stocks, considered as the large cap stocks, are normally categorized in the same style and traded mainly by institutional traders, that is, there should be more correlated trading on these stocks such as index arbitrage, dynamic hedging strategies and naive momentum trading.

We establish a standard time frame for the data series using 15-minute intervals covering 9:30-9:45, 9:45-10:00, ..., 15:45-16:00 for a total of 26 intervals per trading day. The 15-minute time resolution represents a compromise between, on the one hand, needing to look at correlations in contemporaneous order flows across stocks and, on the other hand, requiring enough time for feedback effects from prices into subsequent order submissions.

We calculate the log quote midpoint return as

$$r_{i,t} = \log(m_{i,t,Last}/m_{i,t,First}) \tag{5}$$

where $m_{i,t,Last}$ is the midpoint of the National Best bid and offer quotes for firm i prevailing at the end of interval t; $m_{i,t,Last}$ is the midpoint of the first quote in interval t.

We also develop both unsigned order flow and signed order flow measures. Let $n_{i,t}$ denote the number of trades for firm i in interval t. For the jth trade,$j = 1, 2, ..., n_{i,t}$,let $P_{i,j}$ and $v_{i,j}$ be the price per share and share volume. Four unsigned order flow measures are derived from the consolidated trade data. 1) The total number of trades in the interval is $n_{i,t}$; 2) the total share volume is $\sum_{j=1}^{n_{i,t}} v_{i,j}$; 3) the total dollar volume is $\sum_{j=1}^{n_{i,t}} \log(P_{i,j})v_{i,j}$ and 4) the square root of the dollar volume is $\sum_{j=1}^{n_{i,t}} \sqrt{P_{i,j}v_{i,j}}$.

TAQ does not classify transactions as either buyer-initiated or seller-initiated. To classify the direction of each trade, we use a matching algorithm suggested by Lee and Ready (1991). We define $sign(v_{i,j})$ equals 1 when the jth trade is a buy, and -1 when it is a sell. Therefore, the four corresponding signed order flow measures are 1) the signed trades $\sum_{j=1}^{n_{i,t}} \text{sign}(v_{i,j})$; 2) the signed share volume $\sum_{j=1}^{n_{i,t}} \text{sign}(v_{i,j})v_{i,j}$; 3) the signed dollar volume $\sum_{j=1}^{n_{i,t}} \text{sign}(v_{i,j}) \log(P_{i,j})v_{i,j}$ and 4) the signed square root of dollar volume $\sum_{j=1}^{n_{i,t}} \text{sign}(v_{i,j})\sqrt{P_{i,j}v_{i,j}}$.

Table 7 shows summary statistics for market activity in the sample. If we compare the statistics in Hasbrouck and Seppi (2001b), we find that the number of average daily trades dramatically increases in recent years, and the mean volatility which can be considered as a proxy of market volatility remains the same (1.5% in 1994 and 1.5% on average from 2008 to 2016).

3.2. Relation between returns and order flows

To investigate whether the commonality exist, we rely on the principal component analysis (PCA). The PCA generally sensitive to the unit of the underlying variable. Therefore, we standardize variables to have unit variance and to remove the time-of-day effects documented in Wood, McInish, and Ord (1985). For a representative variable "z", let $z_{i,d,k}$ denote the observation from firm i on the k-th 15-miniute subperiod of day d. Then the standardized variable becomes $z^*_{i,d,k} = (z_{i,d,k} - \mu_{i,k})/\sigma_{i,k}$,where $\mu_{i,k}$ and $\sigma_{i,k}$ are the mean and standard deviation for firm i and subperiod k, estimated across days.

Panel A in Table 2 contains the results for returns. We show that the first component captures 25% to 54% of the total cross-sectional variation in returns, that is, a single common factor can explain one fourth to one half of the total variation. The second and third components are lower than 6%, however, indicating that additional common factors are negligible. Furthermore, the results present that the commonality in returns vary over time. it reached the peak in 2008 (49.49%) and in 2011 (51.04%) ,then decreased in the recent year. This coincides with the financial crisis in 2008 and European debt crisis in 2011.

Panel B in Table 2 contains the results for order flows. We use the signed trades measures as the proxy of order flows. We choose this measure because among all of eight order flow measures, it is generally the most highly correlated with returns at the individual firm level. First, the first principal component in 2007 only explains 10.78% of total variation in order flows. It may be related to the implementation of Regulation National Market System(Reg NMS) in 2007. In 2005, Reg NMS, a set of rules, was passed by the Securities and Exchange Commission (SEC) to bind the fragmented markets into a unified national market and was implemented market-wide in 2007. This implementation changed the mechanisms for achieving queue position in a price-time priority market. This fundamentally changed trading strategies and exchange matching practices.

If removed the outlier data in 2007, the first principal component explains 31.10% of total cross-sectional variation in order flows in 2001, monotonically decreases to 11.52% in 2013 and reverse the decreasing trend to 15.38% in 2016. Combined the results in Panel A, we have the hypothesis that the high commonality in order flows causes the high commonality in returns in 2008.

Table 2
Principal component analysis(PCA) and canonical correlation analysis(CCA) for returns and order flows variables

Loading across firms	First	Second	Third	First	Second	Third	First	Second	Third
	A. PCA for returns			B. PCA for signed trades			C. CCA between returns and signed trades		
2000	17.89%	4.65%	4.46%	14.94%	5.82%	5.39%	78.47%	59.66%	57.19%
2001	24.62%	4.57%	4.10%	23.74%	5.89%	5.03%	76.84%	45.98%	44.02%
2002	37.93%	3.82%	3.43%	31.20%	6.23%	3.77%	81.05%	43.94%	40.88%
2003	40.76%	3.96%	3.20%	29.02%	5.46%	4.49%	80.60%	38.85%	36.45%
2004	36.13%	4.13%	3.60%	26.05%	4.39%	3.79%	71.83%	36.70%	35.05%
2005	34.25%	5.16%	3.50%	26.44%	4.01%	3.90%	72.29%	37.12%	35.13%
2006	29.87%	5.27%	3.96%	22.71%	4.64%	3.82%	76.38%	48.93%	46.36%
2007	39.56%	4.21%	3.79%	10.78%	5.79%	4.04%	72.06%	36.97%	30.49%
2008	54.11%	4.18%	3.75%	23.69%	4.35%	4.13%	71.06%	34.63%	32.36%
2009	48.05%	5.03%	3.47%	23.44%	4.85%	3.82%	79.22%	53.87%	43.62%
2010	47.82%	4.46%	3.19%	18.84%	4.34%	3.75%	81.43%	62.47%	56.84%
2011	51.70%	4.21%	3.04%	15.64%	4.27%	4.08%	77.88%	58.66%	56.58%
2012	39.38%	5.29%	3.38%	12.91%	4.41%	4.33%	80.75%	62.33%	58.38%
2013	35.07%	4.84%	3.45%	11.52%	4.79%	4.07%	76.46%	56.57%	54.31%
2014	37.94%	4.95%	3.94%	14.58%	4.88%	4.36%	73.40%	53.65%	51.26%
2015	45.95%	5.18%	3.26%	17.11%	4.57%	3.90%	71.29%	51.86%	48.37%
2016	39.97%	6.39%	3.96%	15.38%	4.75%	4.27%	70.80%	49.91%	46.26%

Panel A and B list the proportion of total variance explained by the first three principal components. All variables are calculated at 15-minute intervals from 9:30 AM through 4:00 PM and standardized using firm and time-of-day specific means and standard deviations. Returns are defined based on log quote midpoints at the beginning and end of each period. Trades are labeled as buy or sell using the methodology of Lee and Ready (1991). Panel C reports the correlations between the return and order flow (signed trade) canonical variates.

Given the presence of common factors in returns and order flows, we want to explore whether these commonalities are statistically correlated with each other. In this section, we limit our analysis to the signed trades measure. Panel C in Table 2 reports the canonical correlation analyses. For each firm i, if the return $r_{i,t}$ were only correlated with its own order flow $x_{i,t}$, but were uncorrelated across other firms, that is, the cross-covariance matrix $\Sigma_{r,x} = diag(Corr(r_i, x_i))$,then the canonical correlation which is the largest eigenvalue of matrix $\Sigma_{r,x}$ would be $max(Corr(r_i, x_i))$. In unreported results, the $max(Corr(r_i, x_i))$ reaches a peak at 0.5207 in 2009 then monotonically decreases to 0.2967 in 2013, thus Table 2 provides the evidence that the commonalities in returns and order flows are statistically interrelated.

3.3. Public non-trade information

Table 3
Principal component analysis(PCA) for non-trade information

	PCA for residuals			Autocorrelation analysis	
	First PC	Second PC	Third PC	Mean ACF(1)	# of significant stocks
2000	10.04%	5.77%	4.98%	-0.053*	16
2001	13.44%	5.21%	4.81%	-0.041*	15
2002	19.31%	5.45%	4.55%	-0.033*	11
2003	20.40%	13.37%	4.78%	-0.004	15
2004	22.81%	4.85%	4.29%	0.006	8
2005	21.32%	5.73%	4.14%	-0.004	4
2006	18.00%	5.48%	4.49%	0.016	10
2007	26.46%	5.17%	4.64%	-0.019	14
2008	38.57%	5.19%	4.74%	0.007	7
2009	28.84%	5.88%	4.28%	0.011	9
2010	29.56%	5.04%	4.04%	0.041*	21
2011	36.46%	4.82%	3.82%	0.031*	17
2012	24.71%	5.45%	4.04%	0.035*	20
2013	23.20%	5.23%	3.97%	0.036*	20
2014	25.98%	5.25%	4.59%	0.047*	26
2015	33.40%	5.87%	3.89%	0.063*	29
2016	28.35%	6.76%	4.54%	0.053*	27

* denotes significant at 5% level.

In this section, we focus on the residuals part of the first equation in (3) that captures public non-trade information. The second to fourth columns in Table 3 contains the results for non-trade information. We find that the peaked commonality happened in 2008 and 2011. Perhaps it is due to the relation information of financial crisis of 2008 and European debt crisis may have passed through the returns.

The last two columns in Table3 describes the property of residuals time series. Before 2010, less than one-half stocks follow MA(1) process with small first order auto-correlation. Therefore, we can conclude that market has become fairly efficient that the stock price reflects the public non-trade information in a short time (less than 15 minutes). But after 2010, more and more stocks display the positive auto-correlation property. Since order flows are also positively auto-correlated, we suppose that as the market structure has been dramatically changed in recent year, only one order flows measure can not capture all information from the trades. Therefore, the model (3) can be improved by adding some valuable variables.

4. Critical Evaluation of the Models

4.1. PCA models drawbacks

Table 4
Average Adjusted Rsquared for different models

	Adjusted Rsquared		
	Simple Model	PCA Model	Full Model
2000	0.334	0.099	0.356
2001	0.200	0.149	0.260
2002	0.273	0.310	0.382
2003	0.233	0.323	0.357
2004	0.156	0.173	0.219
2005	0.116	0.086	0.161
2006	0.150	0.092	0.231
2007	0.020	0.118	0.241
2008	0.101	0.197	0.257
2009	0.178	0.326	0.376
2010	0.215	0.355	0.410
2011	0.205	0.329	0.393
2012	0.265	0.300	0.401
2013	0.171	0.193	0.292
2014	0.129	0.165	0.255
2015	0.085	0.155	0.231
2016	0.086	0.143	0.207

Simple Model is to regress individual stocks return on its own signed trades measure. PCA Model is to regress the stocks return on the first principle component of signed trades measure. Full Model is to regress the individual stocks returns on all stocks' signed trades measure.

PCA defines a set of linearly uncorrelated principal components that optimally describes variance in a single dataset, while CCA defines a coordinate system that optimally describe the cross-covariance between two datasets. In this sense, the first principal component of order flows may only capture limited information that has passed through the returns. Table 4 presents the average Adjusted R^2 for three models: Simple model ($r_{i,t} = \alpha + \beta_i x_{i,t} + u_{i,t}$), PCA model ($r_{i,t} = \alpha + \beta_i \lambda_t + u_{i,t}$, where λ_t is the first principle component of signed trades measure) and full model ($r_{i,t} = \alpha + \sum_{i=1}^{n} \beta_i x_{i,t} + u_{i,t}$).

We can make some observations here. First, PCA model does not always perform as well as simple model. It is in line with what we discuss before. Second, the Adjusted R^2 of PCA model decreases since 2008. This coincides with our hypothesis that the signed trades measure captures less information in recent year because of the change of market structure. Third, full model with all stocks' order flow information is always the best model and significantly better than PCA model. It means that order flow information from other stocks is very important and can help to explain its own return.

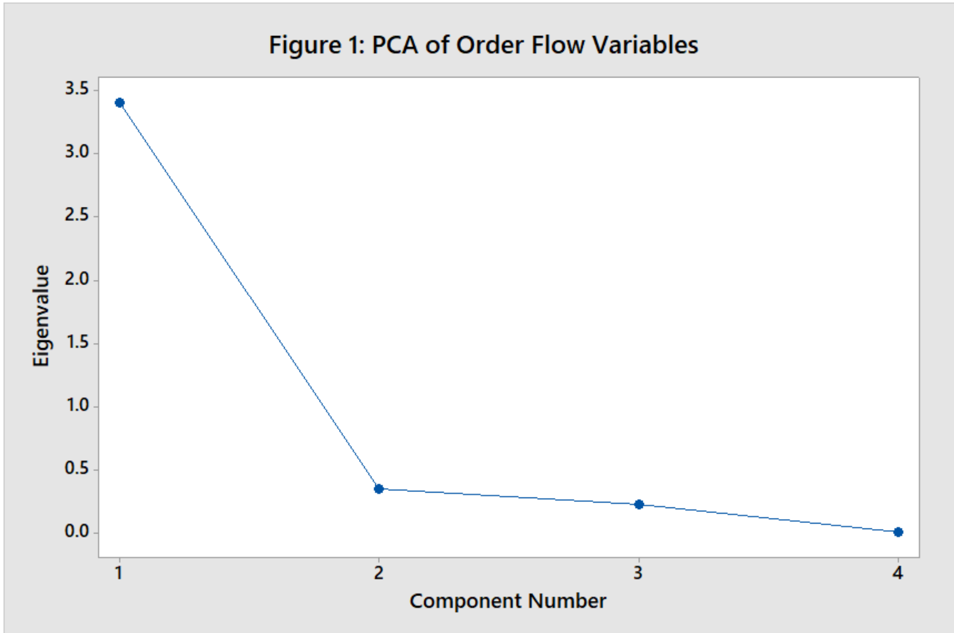
4.2. Unifying methodology

In the multivariate regression model proposed in (3), the need for additional PCA analysis on the residuals is required possibly due to the fact that the order flows have more than one component and therefore simply using one variable via 'x_t' may not be sufficient. Among the four order flow variables only signed trades are used as 'x_t' variable. The PCA of the four signed order flow measures indicate that there may be two dimensions to these measures (refer to Figure 1). The first two components explain 94% of the correlations. Thus the multivariate regression part in Eq.(3) can be improved by adding an additional flow variable.

It can also be empirically verified that in high frequency setting, there is some stickiness in the order flows and therefore the data will exhibit some autocorrelations in the errors. These correlations due to efficiency in the market do not last for too long. Considering these aspects, the following model is proposed:

$$\begin{aligned} r_t &= \Lambda_1 x_{1,t} + \Lambda_2 x_{2,t} + u_t \\ u_t &= a_t + \theta a_{t-1} \end{aligned} \tag{6}$$

Figure 1. PCA of order flow variables



This model will capture both the cross sectional variation as well as time series effects.

The dynamic vector regression model in (6) can be used to capture contagion, comovement and commonality, the concepts used for studying the dependence among the components of the series. Assume that returns 'r_t' are driven by a lower dimensional common features, 'f_t', we postulate the model

$$r_t = A \cdot f_t + \varepsilon_t \tag{7}$$

If the factors are unknown, then the 'f_t' vector is constructed through the PCA of r_t and its past. But if 'f_t' is determined by exogenous variables, $x_t^{'} = (x_{1,t}, x_{2,t})$', then

$$f_t = B \cdot x_t + \varepsilon_t^* \tag{8}$$

Combining (7) and (8), we have

$$r_t = AB \cdot x_t + u_t \tag{9}$$

Thus leading the regression coefficient matrix that is of lower rank. The comovement in the returns 'r_t can be derived from the orthogonal vectors (l's) to 'A' matrix, so that, $l'r_t \sim l'u_t$, indicating that the movements are independent of order flow variables. The dimension of the comovement is captured by the rank of the regression coefficient matrix. More specifically the comovement concept as studied in vector autoregressive (VAR) models where x_t's in (6) are past values of 'r_t', the feature is captured by the rank of VAR coefficients. If the dependent series are non-stationary, such as stock prices, with error correction form,

$$r_t = \Lambda p_t + u_t = A \cdot Bp_t + u_t \tag{10}$$

where Bp_t is taken as cointegrated series.

The above models can also be used to study contagion. One definition used in Forbes and Rigobon (2002) is that it is "contagion if cross-market comovement increases significantly after the shock". This can be tested from the rank of the VAR coefficient matrix. For testing commonality which is the focus of this paper, it is clear that it depends on the dimension of 'f_t' and therefore can also be inferred from the rank of the coefficient matrix.

Computational steps for the paper:

——Estimate the regression model in (6) with the judicious choice of order flow variables.

——Examine the rank of the coefficient matrices via canonical or partial canonical correlations.

——Check to see if the structure has changed over time. Interpret the change in structural coefficients and relate them to demand/supply side variables.

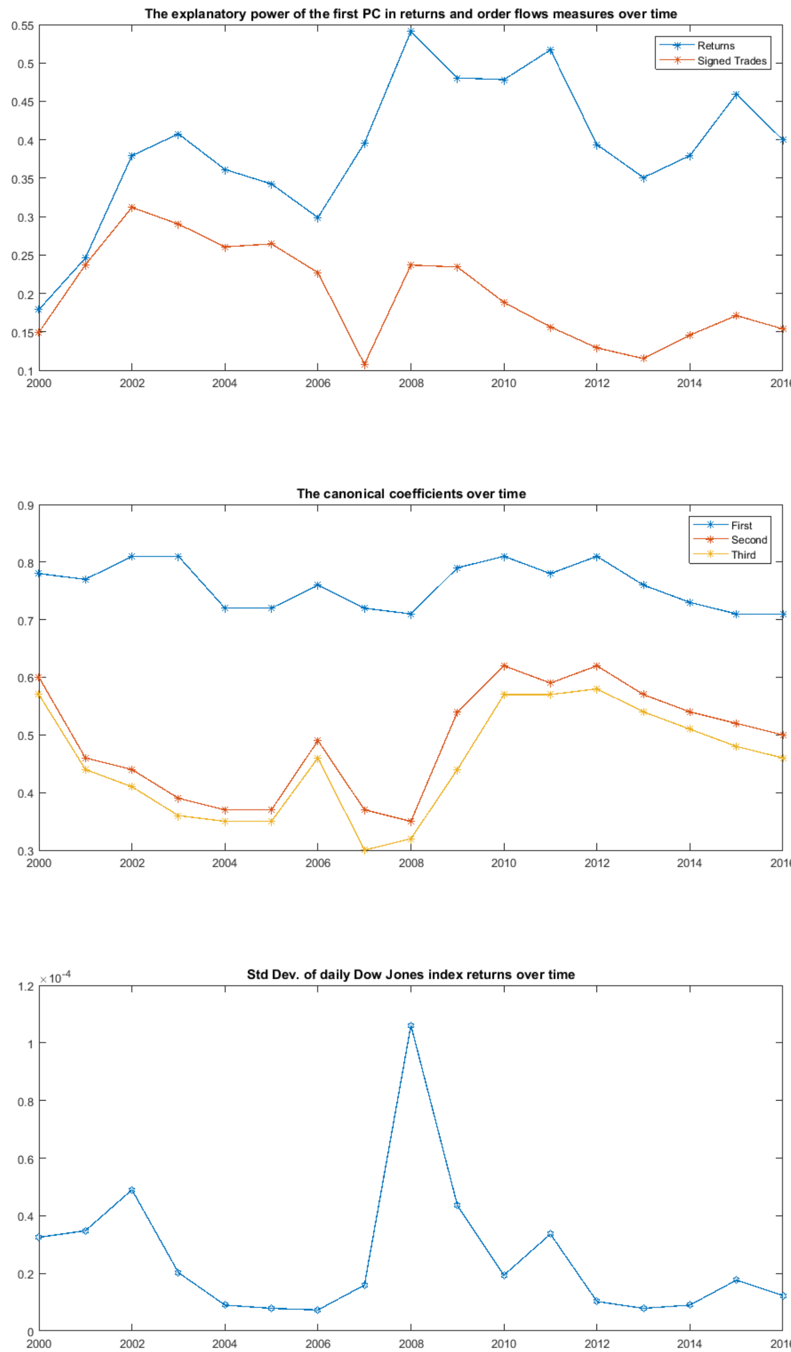
4.3. Results of the General Model

4.3.1. Commonality over time

Table 2 shows that the first principal components of returns and signed trades measure remain dominant in the whole sample period. Therefore, in this section, we focus on the first principal component to explore how the commonality changes over time.

First, Figure 2 describes the explanatory power of the first principal component in returns and order flow measure from 2000 to 2016. we can observe that the return commonality increases over time, but the order flows commonality decreases. The internet bubble occurred roughly over the period 1997 to 2001, but the index slid steadily starting in March 2002, with dramatic declines in July and September. This may be related to the high return commonality in 2002. Moreover, because of 2008 financial crisis and European debt crisis, the returns commonality remains at a

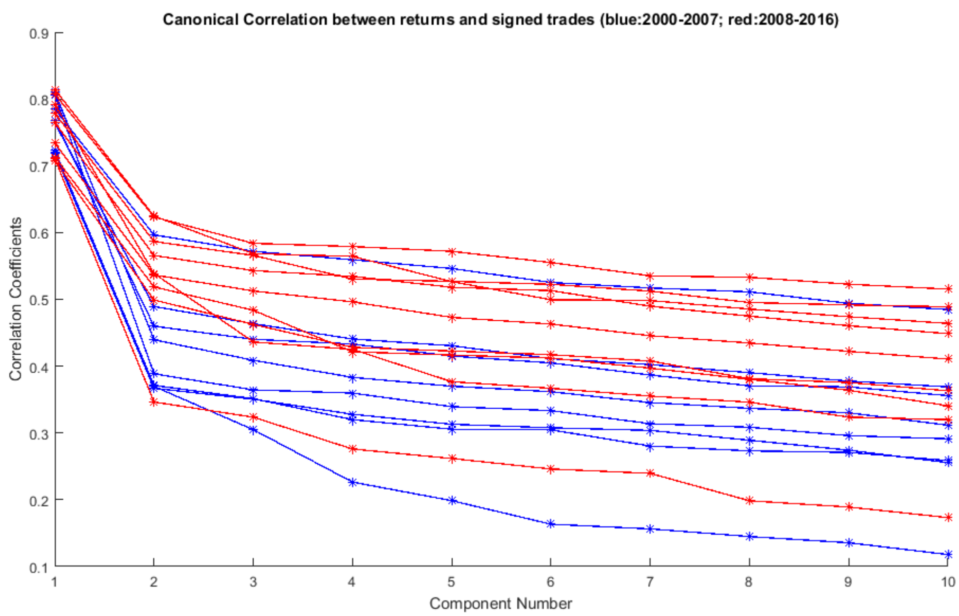
Figure 2. PCA and CCA over time



high level from 2008 to 2011, then slowly decrease since 2012. In 2015, investors' concern over the impending end to the quantitative easing policy in US then choke off investment in emerging markets, causing negative global financial effects. Thus, the return commonality reaches another peak in 2015.

Second, even though the return and order flows commonalities follow a trend, the first canonical coefficient between returns and order flows swings in a narrow range (0.70 to 0.81) from 2000 to 2016. Moreover, the second and third canonical coefficients increase dramatically in 2009 and decrease slowly after 2012. It provides the evidence that using one dimension to describe the returns or order flows of multiple stocks can not properly investigate the relationship between returns and order flows.

Figure 3. CCA over time

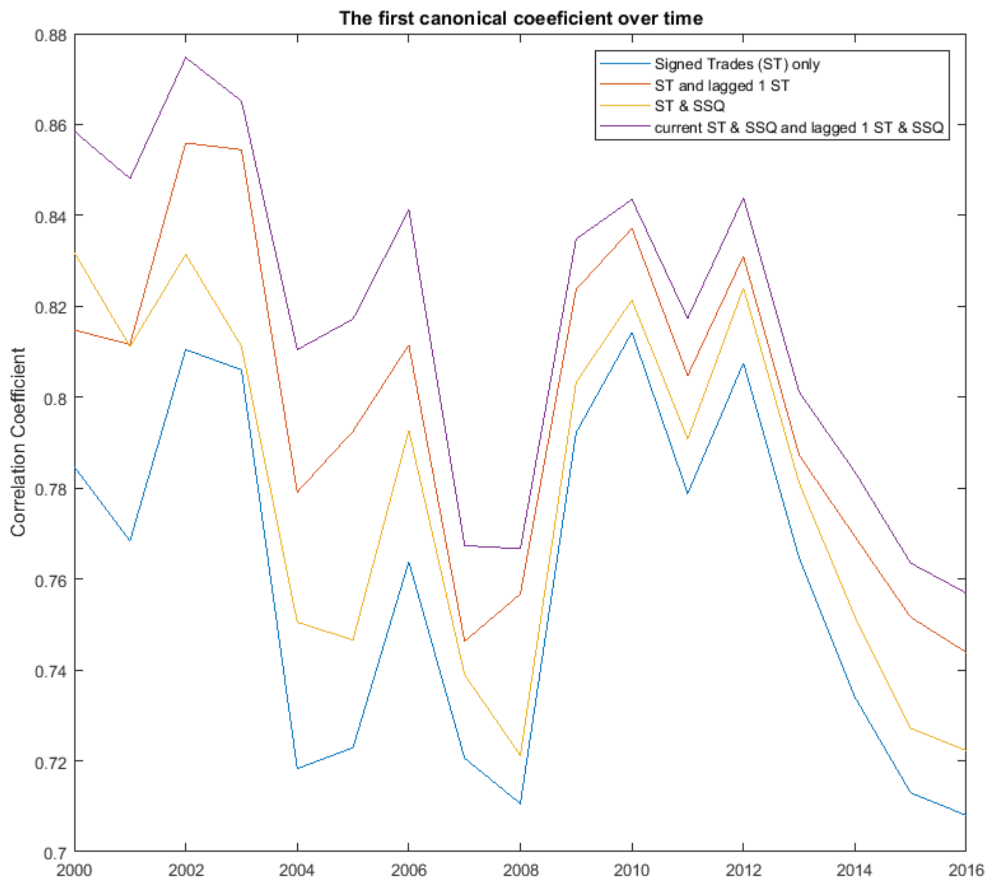


Third, Figure 3 shows the results of canonical correlation between returns and signed trades over time. The blue lines are generated by the data from 2000 to 2007. The red lines are generated by the data from 2008 to 2016. The lowest blue line presents the result in 2007, and the lowest red line presents the result in 2008. The highest blue line presents the result in 2000. It may relate to the decimal pricing rule. The implementation period of this rule began on August 28,2000 and ended with full implementation of decimal pricing for all equities and options by April 9,2001.

Furthermore, Figure 3 also shows that after 2007, not only the first three ,but also all the canonical coefficients between returns and signed trades increase dramatically.

4.3.2. Empirical results for the unified model

Figure 4. The first canonical coefficient for four different models



In this section, we consider four different models. The first one is the basic model that describes the relationship between returns and signed trades. The second one includes lagged one signed trades. The third one is to regress returns on signed trades and the signed square root of dollar volume. The fourth one is added two lagged one variables into the third model. Figure 4 describes the first canonical coefficient over time. It shows that adding signed square root of dollar volume only marginally increase the canonical coefficient. Maybe most of the information has been captured in the signed trades measure. However, the canonical coefficient distinctly increases when we add the lagged one signed trades measure to the basic model. Therefore, we can conclude that the lagged one signed trades variable captures more information than the signed square root of dollar volume.

5. Relation to Market Microstructure Models

This section explores the relationship between market microstructure models and the statistical representation of the framework used in commonality studies. Microstructure models of intraday price formation are widely used to analyze patterns in price discovery and transaction cost. For instance, Green (2004) uses microstructure models to measure the informational role of trading and to infer the asymmetric information component of the effective bid-ask spread. On the other hand, Sadka (2006) uses microstructure models to investigate momentum, a closely related concept to commonality, and whether returns can be related to the time variation of liquidity. Korajczyk and Sadka (2008) also investigate commonality in alternative measures of liquidity, notably price impact, using microstructure models. More recently, Riordan, Storkenmaier, Wagener, and Zhang (2013) employ microstructure models to study the impact of information arrival on intraday price discovery, liquidity, and trading intensity in an electronic limit order market.

Hasbrouck and Seppi (2001a) has pointed out that the statistical representation of the framework used for commonality studies can be readily expressed in a market microstructure model setup by linking returns of assets to order flow and other non-trade components. As elaborated in earlier sections, a simple model of return can be written as $r_{it} = \beta_i x_{it} + u_{it}$, which can be extended to account for commonality using the multivariate relationship $r = \Lambda x_t + u_t$. However, a microstructure model that merely relates returns to order flow (signed trades in this case) inevitably lacks sufficient explanatory power for most analyses. Based on the unifying methodology presented in the previous section, we show how more sophisticated microstructure models can be related to the commonality framework.

In the microstructure model proposed by Roll (1984), the unobservable true price process of an asset is assumed to follow a random walk

$$\pi_t = \pi_{t-1} + \epsilon_t,$$

where $\epsilon_t \sim N(0, \sigma^2)$ and is independent and identically distributed. The observed transaction price p_t as a result of bid-offer spread is written as

$$p_t = \pi_t + \frac{s}{2} x_t,$$

where s is the bid-offer spread. The change in price, or return, can therefore be written as

$$\Delta p_t = r_t = \frac{s}{2}\Big(x_t - x_{t-1}\Big) + \epsilon_t.$$

Note that Roll's model can be extended to account for commonality via the multi-variate relationship:

$$r_t = \Lambda\Big(x_t - x_{t-1}\Big) + u_t.$$

This is a special case of the unified methodology in Eq (6) with $\Lambda_1 = -\Lambda_2$ and $x_{2,t} = x_{1,t-1}$, i.e. the coefficients of the second variable are identical to that of the first variable, albeit with the opposite sign, and the second variable is simply the first variable with lag 1.

Separately, Madhavan, Richardson, and Roomans (1997) (MRR model) postulate that the market belief about a security at time t is π_t:

$$\pi_t = \pi_{t-1} + \theta\Big(x_t - \mathbb{E}[x_t | x_{t-1}]\Big) + \varepsilon_t.$$

This unobservable market belief is conditional on public information, and the change in belief is dependent on the surprise element in the trade direction, which is given by $x_t - \mathbb{E}[x_t|x_{t-1}]$, as well as other public information denoted by ϵ_t. The change in belief is modeled as $\theta\left(x_t - \mathbb{E}[x_t|x_{t-1}]\right)$, where $\theta \geq 0$ measures the degree of information asymmetry. Market makers' quotations also reflect their compensation for their service in providing liquidity on demand. Let $\phi \geq 0$ represent market makers' cost for supplying liquidity, which captures the temporary (or transitory) effect of order flow on prices. The transaction price is expressed as a noisy reflection of the market belief plus the term ϕx_t, i.e.

$$p_t = \pi_t + \phi x_t + \xi_t,$$

here ξ_t is an iid random variable with zero mean. Substituting, the transaction price is

$$p_t = \pi_{t-1} + \theta(x_t - \mathbb{E}[x_t|x_{t-1}]) + \phi x_t + \varepsilon_t + \xi_t$$

Assume that $\mathbb{E}[x_t|x_{t-1}] = \rho x_{t-1}$, the price change can be expressed as

$$\Delta p_t = r_t = (\phi + \theta)x_t - (\phi + \rho\theta)x_{t-1} + \epsilon_t,$$

where $\epsilon_t = \varepsilon_t + \Delta\xi_t$. Under our unifying methodology, the MRR microstructure model can also be extended to account for commonality under our unifying framework as follows:

$$r_t = \Lambda_1 x_t + \Lambda_2 x_{t-1} + u_t.$$

We note in passing that a related microstructure model formulated by Huang and Stoll (1997) (HS model) can also be written in the same representation.

In addition to order flow, it is conceivable that volume information should also play an important role in asset returns. In the microstructure model proposed by Glosten and Harris (1988) (GH model), the unobservable true price process is postulated to follow

$$\pi_t = \pi_{t-1} + \alpha_0 x_t + \alpha_1 v_t + \epsilon_t,$$

while the observable transaction price is modeled as

$$p_t = \pi_t + \beta_0 x_t + \beta_1 v_t.$$

The return can therefore be written as

$$r_t = (\alpha_0 + \beta_0)x_t - \beta_0 x_{t-1} + (\alpha_1 + \beta_1)v_t - \beta_1 v_{t-1} + \epsilon_t,$$

which can again be related to our unifying methodology as follows:

$$r_t = \Lambda_0^x x_t + \Lambda_1^x x_{t-1} + \Lambda_0^v v_t + \Lambda_1^v v_{t-1} + u_t.$$

As a summary, Table 5 compares the microstructure models with their corresponding extended multi-variate version under our unified commonality methodology:

Table 5 Comparison of Microstructure Models to Unifying Commonality Models

Standard market microstructure models relate asset's returns to the asset's own order flow and other non-trade liquidity measures, including signed trade volume and lagged variables. Under the commonality framework, returns of individual asset can also be related to the order flow and liquidity measures of other assets. The unifying methodology proposed in this paper can be seen as an extension to standard microstructure models to account for commonality between these assets.

	Microstructure Model	Extended Unified Model
Simple Model	$r_t = \lambda x_t + \epsilon_t$	$r_t = \Lambda x_t + u_t$
Roll Model	$r_t = \lambda\Big(x_t - x_{t-1}\Big) + \epsilon_t$	$r_t = \Lambda\Big(x_t - x_{t-1}\Big) + u_t$
MRR/HS Model	$r_t = \lambda_1 x_t + \lambda_2 x_{t-1} + \epsilon_t$	$r_t = \Lambda_1 x_t + \Lambda_2 x_{t-1} + u_t$
GH Model	$r_t = \lambda_1^x x_t + \lambda_2^x x_{t-1} + \lambda_1^v v_t + \lambda_2^v v_{t-1} + \epsilon_t$	$r_t = \Lambda_1^x x_t + \Lambda_2^x x_{t-1} + \Lambda_1^v v_t + \Lambda_2^v v_{t-1} + u_t$

Figure 5 plots the eigenvalues of the first principle component of the residuals of the extended unified model (top panel) and standard microstruture model (bottom panel). From the figure, it is clear that the residual variance for both family of models can be reduced by adding additional non-trade liquidity variables. The simple model relating returns only to signed trades has the highest residual variance. Adding volume information (signed square root dollar volume in this case) and lagged version of these variables reduces the residual variance. Comparing between the two family of models, the extended unified model proposed in this paper has noticeably lower residual variance, highlighting the importance of accounting for commonality when using microstructure models to analyze asset returns.

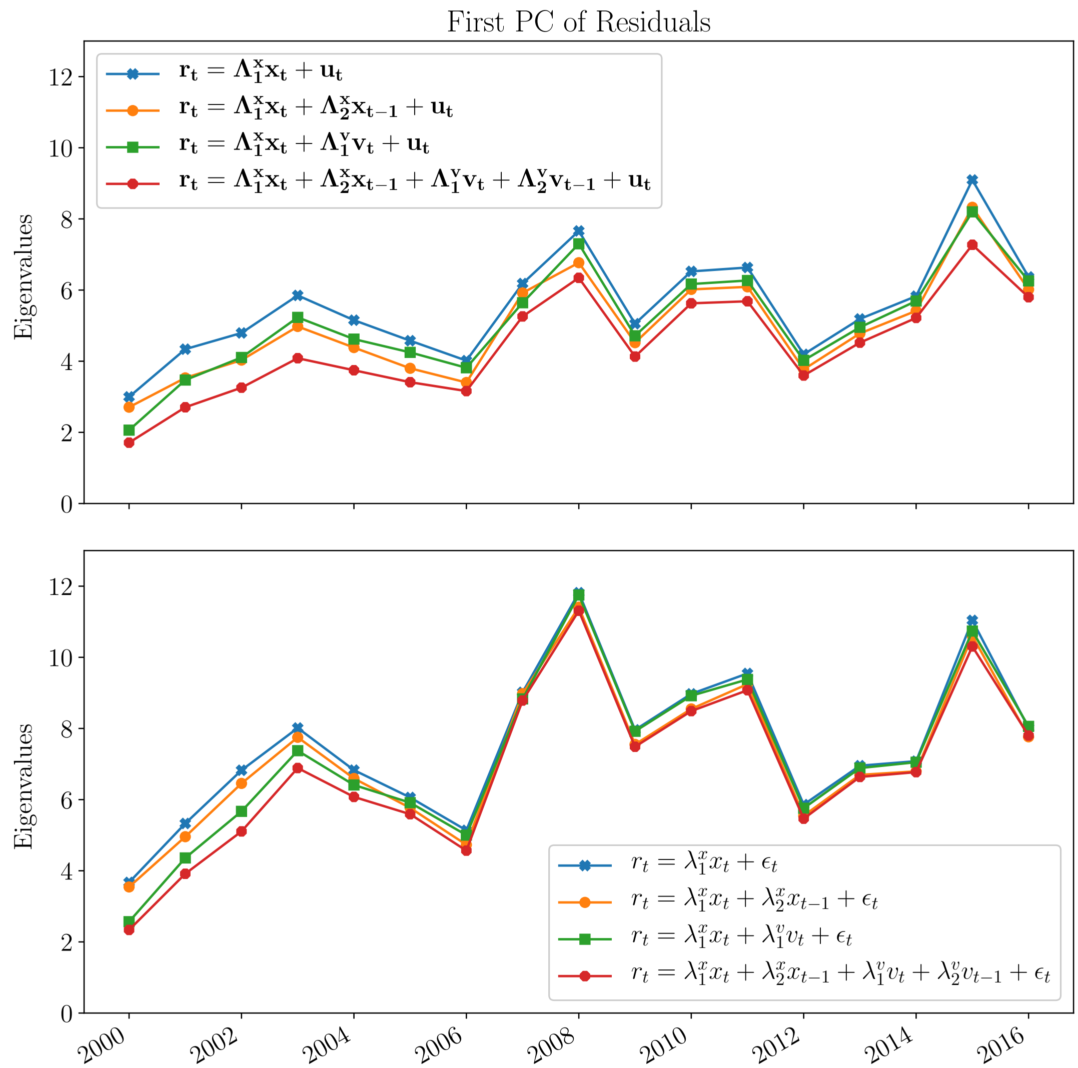


Figure 5. Comparison between the first principle component's eigenvalues of the extended unified commonality model (top panel) *vs* standard microstructure models (bottom panel).

6. Why commonality changes?

6.1. Demand and supply side reasons

Barberis and Shleifer (2003) classify the explanations for return commonality into three groups: style investing, habitat investing and information diffusion with different speed. Style investing and habitat investing would increase the correlated trading behavior, therefore, they are in line with demand side hypotheses. Suppose the increase of return commonality mainly comes from the "demand side hypotheses", we should expect that the order flows commonality also increases due to the correlated trading. But our results do not support this hypotheses.

Suppose the returns commonality mainly comes from information diffusion with different speed. We expect the stocks returns should present the lead-lag effect, therefore we perform a lagged-correlation analysis between all possible stock pairs. The lead-lag correlation coefficients define as $\rho_{r_{i,j}} = \frac{Cov(r_{i,t}, r_{j,t-1})}{\sqrt{Var(r_{i,t})}\sqrt{Var(r_{j,t-1})}}$ where $i \neq j$.

Table 6 shows that the percentage of pairs stocks with significant coefficients decreases from 2000 to 2016. It is coincide with the fact that market has become more and more efficient in recent year, or at lease the Dow Jones stocks in our sample reflect the market-wide information within 15 minutes. Therefore, the hypotheses that information diffusion with different speed can not fully explain why the return commonality increases from 2000 to 2016.

Table 6

Lead-lag correlation analysis

	Lead-lag effects on returns		Lead-lag effects on signed trades	
	% of pairs stocks with significant coefficients	Average of the significant coefficients	% of pairs stocks with significant coefficients	Average of the significant coefficients
2000	21.50%	0.017	60.50%	0.054
2001	34.31%	0.020	93.23%	0.088
2002	34.62%	0.010	97.72%	0.097
2003	25.74%	0.023	95.94%	0.077
2004	19.97%	-0.002	96.96%	0.086
2005	12.43%	0.007	99.74%	0.102
2006	11.64%	-0.019	98.28%	0.080
2007	16.26%	-0.020	55.54%	0.040
2008	20.57%	-0.017	93.60%	0.067
2009	21.67%	-0.010	91.50%	0.073
2010	18.10%	-0.027	65.02%	0.042
2011	8.37%	-0.015	61.95%	0.044
2012	7.64%	-0.010	44.70%	0.039
2013	7.64%	0.000	45.57%	0.042
2014	8.99%	0.027	71.06%	0.054
2015	10.59%	-0.027	73.77%	0.063
2016	12.81%	-0.020	67.86%	0.054

6.2. Exchange-wide Variation and Fragmentation

As O'Hara (2015) points out, the US equity market is highly fragmented, with different trading venues appealing to different clienteles. Although the fragmented US markets can be viewed as a single virtual market with multiple points of entry (see, for instance, O'Hara and Ye (2011)), there exists a strong competition between exchanges and trading venues.

To measure the change in market fragmentation over time, we follow Madhavan (2012) by calculating the volume Herfindahl index for the Dow Jones Index constituent stocks over the period included in this study. Herfindahl index is a metric commonly used to measure market fragmentation. The volume Herfindahl index on day t is defined as

$$H_t^v = \sum_{k=1}^{K} \left(s_t^k\right)^2,$$

where s_t^k is the volume share of venue k on day t, and K is the number of exchanges. The Herfindahl index ranges from 0 to 1, with higher figures indicating less fragmentation. As shown in Figure 6, our results reveal that over the 17 calendar year studied (2000 through to 2016), the Herfindahl index reduces progressively, indicating that the market is becoming more fragmented over time.

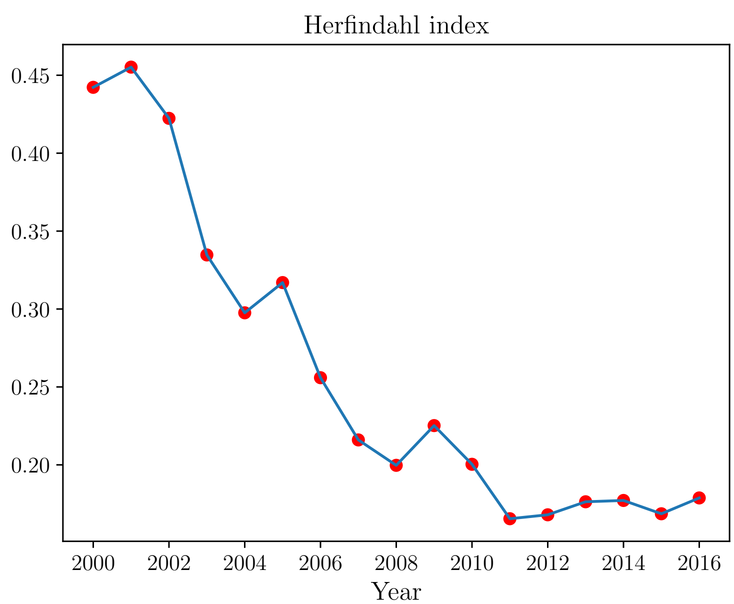


Figure 6. Annual volume Herfindahl index of the Dow Jones Index constituent stocks over the 2000-2016 period.

Kwan, Masulis, and McInish (2015) argue that competition between traditional exchanges and trading venues for order flow is changing the structure of financial market. Consequently, a thorough assessment of commonality and the informational role of trading across different exchanges and trading venues is important for investors and regulators alike. O'Hara and Ye (2011) find that market fragmentation generally reduces transactions costs and increases execution speeds.

Fragmentation does increase short-term volatility, but prices are more efficient.

To access this assertion from the commonality perspective, we split our dataset by reporting exchanges and aggregate them at the exchange level. We then apply our commonality analysis on each of the exchanges. Figure 7 plots the eigenvalues of the first principal component across the exchanges for returns (top panel) and signed trades (bottom panel). As one would expect, the commonality in returns and signed trades follow closely the general trend observed in the overall aggregated study presented in previous sections of the paper. Of particular interest in the exchange-wide results is the convergence of eigenvalues in both returns and signed trades over the years. This observation is consistent with the hypothesis in the market microstructure literature that liquidity is shared across all trading venues, and that market fragmentation does not lead to a deterioration in participants' access to liquidity. On the contrary, competition across different trading venues lead to a reduction in transaction costs and improved price efficiency, manifesting in the observed trend of convergence in eigenvalues over time.

6.3. High frequency explanations

Malceniece, Malcenieks, and Putniņš (2018) use 2 years European equity data to prove that high frequency trading (HFT) causes significant increases in co-movement in returns and in liquidity. They show that HFT impacts co-movement via three possible mechanisms. The first channel is through correlated trading across stocks. A second channel is by increasing the speed with which prices reflect public market-wide information. HFT shorten the time which market-wide information is transmitted from large stocks to small stocks, therefore increase co-movement. The third channel is that the increase in liquidity due to HFT activity makes medium and small stocks more attractive to other non-HFT participants. The increase in stock liquidity makes it more likely that the medium and small stocks become part of the "habitat" of institutional traders and therefore the co-movement in returns increases due to the habitat trading by institutional traders.

Because our sample is limited on the Dow Jones stocks, the effect from the second and third channel is negligible. At first glance, although the increased correlated trading results in the increase of commonality in returns, it cannot explain why the commonality in order flows decreases. However, high frequency traders move in and out of short-term positions at high speeds. The correlated trading mentioned in Malceniece et al. (2018),considered as "short-term correlated trading",

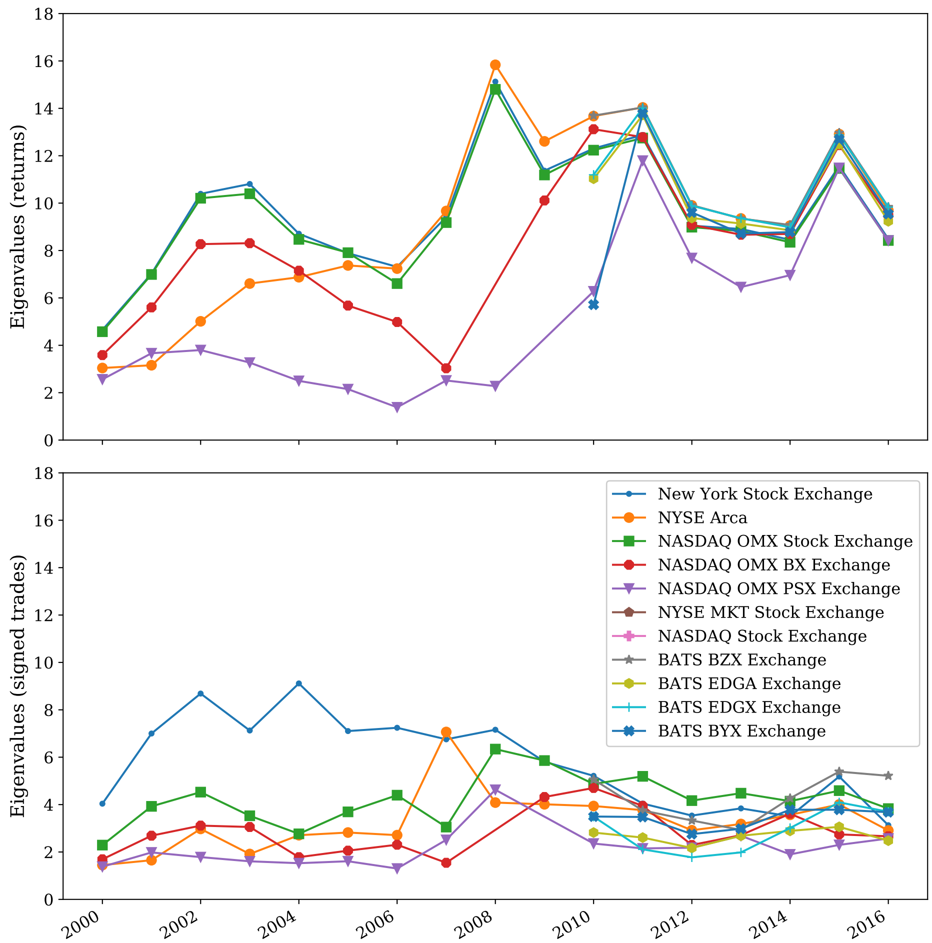


Figure 7. Exchange-wide first principle component's eigenvalues of returns and signed trades.

is not the same term mentioned in Koch et al. (2016) that describes the correlated trading comes from different mutual funds. Both kinds of correlated trading lead to the increase of return commonality, but their trades may be uncorrelated. Therefore, these two kinds of correlated trading decrease the order flow commonality and increase the return commonality at the same time.

In order to explore this hypothesis, we develop the proxies for HFT activity based on electronic message traffic, similar to Hendershott, Jones, and Menkveld (2011) and Boehmer, Fong, and Wu (2015). The HFT measure is the number of messages divided by the number of trades in each 15-minutes time interval.

$$HFT_{i,t} = \frac{messages_{i,t}}{trades_{i,t}}$$

Because high-frequency traders hold their positions in a short period (less than 15 minute), the signed high frequency trades of individual stock would be closed to 0. Therefore, the signed trades measure we discussed before cannot capture too much information related to the high frequency trading. Adding the HFT measure into our model can improve the explanatory power. Furthermore, suppose HFT measure is positive correlated with the short-term correlated trading, we can identify how the HFT correlated trading affect the returns commonality and order flow commonality.

We regress daily returns on daily signed trades measures and daily HFT measures. The unreported result provides the evidence that the HFT measures capture additional significant information besides the signed trades measures.

6.4. Herding

Hirschey (2017) uses data that classifies market participants as either an HFT or a non-HFT to show that liquidity demand by HFTs can predict subsequent liquidity demand by non-HFTs. Given that liquidity demand by non-HFTs has information about subsequent returns, such predictability shows that HFT measures may capture similar information as the lagged-one signed trade measure.

Furthermore, Table 6 shows that most of pairwise stocks present the significant lead-lag correlation. Therefore, some investors such as liquidity providers can predict the subsequent order flow via other stocks' order flow. They can finish more trades among different stocks base on this predictability. In this case, the order flow commonality may decrease, but the returns commonality increases.

7. Conclusions

Financial time series are known to exhibit common characteristics in returns, order flows, and other non-trade parameters. In this work, we have performed an in-depth longitudinal study of the Dow Jones constituent stocks over 17 calendar years (2000-2016) to identify key trends in commonality over a long period. A longitudinal study is crucial to understand the variation in commonality characteristics over periods of economic growth and recession.

There are three main contributions in this paper to existing literature. First, our extensive empirical analyses show that commonality in returns has increased over time, while commonality in order flow has decreased marginally. Spikes and drops in either returns or order flow commonality can be related to specific financial or economic events in the market. More importantly, we performed a critical assessment on the main drawbacks of existing commonality framework in terms of its explanatory power, and formulated a more general unifying framework to systematically accommodate additional important factors related to commonality.

Second, building on the insight of Hasbrouck and Seppi (2001), we explore the relationship between the statistical representation of commonality framework and well-known market microstructure models. We show that different microstructure models can be expressed as special cases of the general unifying methodology formulated in this paper, taking commonality into consideration. Comparing standard market microstructure models against the our unifying model, the residuals of our model has significantly lower eigenvalues. This is an important insight, highlighting that instead of modelling individual assets in isolation, market microstructure models can improve their explanatory power by taking commonality into account.

Third, we investigate the cause of commonality variation over time. Our results do not directly support the demand- or supply-side hypothesis. By breaking down our dataset with respect to reporting exchanges, we demonstrate that while commonality in returns and order flow vary over time, there is a clear trend of convergence over time. In other words, exchange-wide commonality variation is decreasing. This points to high frequency traders (HFTs) as a possible explanations. Using electronic messages as a proxy for high frequency traders, we are able to estimate the participation of HFTs via the ratio of messages over the number of trades executed in a given time interval. We show that this HFT proxy measure captures significant amount of information, indicating HFTs activities as a viable explanation of commonality variation over time.

REFERENCES

Barberis, N., Shleifer, A., 2003. Style investing. Journal of financial Economics 68, 161–199.

Barberis, N., Shleifer, A., Wurgler, J., 2005. Comovement. Journal of Financial Economics 75, 283–317.

Boehmer, E., Fong, K., Wu, J., 2015. International evidence on algorithmic trading .

Chen, H., Singal, V., Whitelaw, R. F., 2016. Comovement revisited. Journal of Financial Economics 121, 624–644.

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. Journal of financial economics 56, 3–28.

Corwin, S. A., Schultz, P., 2012. A simple way to estimate bid-ask spreads from daily high and low prices. The Journal of Finance 67, 719–760.

Coughenour, J. F., Saad, M. M., 2004. Common market makers and commonality in liquidity. Journal of Financial Economics 73, 37–69.

Forbes, K. J., Rigobon, R., 2002. No contagion, only interdependence: measuring stock market comovements. The Journal of Finance 57, 2223–2261.

Glosten, L., Harris, L., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Green, T. C., 2004. Economic news and the impact of trading on bond prices. Journal of Finance 59, 1201–1233.

Green, T. C., Hwang, B.-H., 2009. Price-based return comovement. Journal of Financial Economics 93, 37–50.

Hameed, A., Kang, W., Viswanathan, S., 2010. Stock market declines and liquidity. The Journal of Finance 65, 257–293.

Hasbrouck, H., Seppi, D. J., 2001a. Common factors in prices, order flows, and liquidity. Journal of Financial Economics 59, 383–411.

Hasbrouck, J., Seppi, D. J., 2001b. Common factors in prices, order flows, and liquidity. Journal of financial Economics 59, 383–411.

Hendershott, T., Jones, C. M., Menkveld, A. J., 2011. Does algorithmic trading improve liquidity? The Journal of Finance 66, 1–33.

Hirschey, N., 2017. Do high-frequency traders anticipate buying and selling pressure? .

Huang, R., Stoll, H., 1997. The components of the bid-ask spread: A general approach. Review of Financial Studies 10, 995–1034.

Huberman, G., Halka, D., 2001. Systematic liquidity. Journal of Financial Research 24, 161–178.

Kamara, A., Lou, X., Sadka, R., 2008. The divergence of liquidity commonality in the cross-section of stocks. Journal of Financial Economics 89, 444–466.

Koch, A., Ruenzi, S., Starks, L., 2016. Commonality in liquidity: a demand-side explanation. The Review of Financial Studies 29, 1943–1974.

Korajczyk, R. A., Sadka, R., 2008. Pricing the commonality across alternative measures of liquidity. Journal of Financial Economics 87, 45–72.

Kwan, A., Masulis, R., McInish, T. H., 2015. Trading rules, competition for order flow and market fragmentation. Journal of Financial Economics 115, 330–348.

Lee, C., Ready, M. J., 1991. Inferring trade direction from intraday data. The Journal of Finance 46, 733–746.

Madhavan, A., 2012. Exchange-traded funds, market structure, and the flash crash. Financial Analysts Journal 68, 20–35.

Madhavan, A., Richardson, M., Roomans, M., 1997. Why do security prices change? a transaction-level analysis of nyse stocks. Review of Financial Studies 10, 1035–1064.

Malceniece, L., Malcenieks, K., Putniņš, T. J., 2018. High frequency trading and co-movement in financial markets .

Marsh, Mazza, 2018. Common factors in market quality: Evidence from european stocks .

O'Hara, M., 2015. High frequency market microstructure. Journal of Financial Economics 116, 257–270.

O'Hara, M., Ye, M., 2011. Is market fragmentation harming market quality? Journal of Financial Economics 100, 459–474.

Riordan, R., Storkenmaier, A., Wagener, M., Zhang, S. S., 2013. Public information arrival: price discovery and liquidity in electronic limit order markets. Journal of Banking and Finance 37, 1148–1159.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Sadka, R., 2006. Momentum and post-earnings-announcement drift anomalies: the role of liquidity risk. Journal of Financial Economics 80, 309–349.

Wood, R. A., McInish, T. H., Ord, J. K., 1985. An investigation of transactions data for nyse stocks. The Journal of Finance 40, 723–739.

Table 7
Descriptive statistics

Symbol	Name	Mean price ($/price)	Mean bid-ask spread ($/share)	Average daily trades	Average daily turnover *100(%)	Average daily Amihud illiquidity measure (10^{-10})	Average daily return (basis point)	Std Dev. of daily return *100(%)	Beta
	DJ index	13669	0.003			0.480	2.3	1.2	0.89
	Cross-stock mean:	59	0.014	65,564	0.72	0.212	1.4	1.5	0.99
AA	Alcoa	14	0.010	58,269	2.39	0.720	-14.4	2.5	1.62
AXP	American Express	57	0.014	43,272	0.78	0.374	4.9	2.1	1.46
BA	Boeing	92	0.023	30,695	0.74	0.311	1.7	1.5	1.01
BAC	Bank of America	15	0.010	183,796	1.79	0.108	-20.5	3.1	2.02
CAT	Caterpillar	79	0.018	40,924	1.22	0.270	-1.4	1.7	1.22
CSCO	Cisco	23	0.010	100,143	0.82	0.127	1.8	1.5	1.04
CVX	Chevron	98	0.016	52,255	0.47	0.130	2.8	1.5	1.06
DD	Du Pont	51	0.014	31,086	0.67	0.455	2.5	1.6	1.12
DIS	Disney	58	0.012	46,387	0.57	0.264	6.4	1.4	1.07
GE	General Electric	23	0.010	116,577	0.58	0.101	-7.0	1.7	1.15
HD	Home Depot	64	0.015	49,302	0.70	0.231	6.9	1.6	0.96
HPQ	Hewlett-Packard	32	0.011	62,982	0.86	0.334	8.7	1.7	0.97
IBM	IBM	156	0.033	33,896	0.48	0.117	9.5	1.1	0.76
INTC	Intel	25	0.010	111,270	0.91	0.119	4.4	1.5	1.02
JNJ	Johnson & Johnson	79	0.012	50,533	0.39	0.085	1.6	0.9	0.56
JPM	JPMorgan Chase	48	0.011	123,744	0.87	0.111	1.7	2.4	1.61
KO	Coca-Cola	50	0.011	49,001	0.39	0.134	3.0	1.0	0.57
MCD	McDonald	86	0.014	36,057	0.66	0.147	2.6	1.0	0.57
MMM	3M	108	0.026	22,286	0.51	0.282	5.6	1.2	0.85
MRK	Merck	44	0.011	55,858	0.54	0.189	-0.9	1.4	0.79
MSFT	Microsoft	34	0.010	132,953	0.61	0.073	3.8	1.5	0.95
PFE	Pfizer	24	0.010	89,731	0.58	0.112	-0.5	1.2	0.77
PG	Procter & Gamble	71	0.012	49,975	0.40	0.100	3.2	1.0	0.57
T	AT&T	32	0.010	78,895	0.47	0.105	-3.0	1.2	0.75
TRV	Travelers	73	0.020	19,329	0.75	0.530	7.8	1.6	1.04
UTX	United Technologies	85	0.019	28,306	0.51	0.285	1.9	1.3	0.96
VZ	Verizon	41	0.011	61,463	0.53	0.153	-1.4	1.2	0.70
WMT	Wal-Mart	64	0.012	55,842	0.34	0.111	4.0	1.1	0.51
XOM	Exxon Mobil	82	0.012	86,537	0.42	0.064	5.6	1.3	0.94

The sample is the 29 Dow stocks for all trading days from 2008 to 2016. The daily return is computed as the first difference between the log end-of-day quote midpoint and the log begin-of-day quote midpoint. The mean bid-ask spread is calculated as the average bid-ask spread of National Best Quotes when the trade happened. The mean bid-ask spread of Dow Jones Index derived by Corwin and Schultz (2012). The beta uses the S&P 500 as a benchmark.