How Smart Is Institutional Trading?

JinGi Ha and Jianfeng Hu *

January 19, 2017

ABSTRACT

We estimate daily aggregate order flow at the stock level from all institutional investors as well as for hedge funds and the other institutions separately. We achieve this by extrapolating the relation between quarterly institutional ownership in 13F filings, aggregate market order imbalance in TAQ, and a representative group of institutional investors' transaction data. We find that the estimated institutional order imbalance has positive price impact in the short term, which reverses in the long term. The "smart" order flow from hedge funds generates greater and more persistent price impact than the "dumb" order flow from all the other institutions. We also find that hedge funds trade on well known anomalies around month ends while the other institutions do not.

*JinGi Ha and Jianfeng Hu are at Singapore Management University. We would like to thank Ekkehart Boehmer, Jarred Harford, Dashan Huang, Roger Loh, Avanidhar Subrahmanyam, Yuehua Tang, Joe Zhang, and the seminar participants at Singapore Management University for comments. All remaining errors are ours. Please address correspondence to JinGi Ha (jingiha.2014@pbs.smu.edu.sg, +65 9189 0853) and Jianfeng Hu (jianfenghu@smu.edu.sg) at Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore, 178899.

I. Introduction

Empirical research on the effect of institutional trading on financial markets has been largely constrained by the availability of institutional transaction data because the Securities and Exchange Commission (SEC) in the US only requires institutional investors to report their equity positions in quarterly 13-F filings. Therefore, most researchers rely on quarterly changes in reported institutional positions to identify institutional trading intension.[1] While traditional analysis usually investigates institutional investors as a whole group, several recent studies such as Frazzini and Lamont (2008), Akbas, Armstrong, Sorescu, and Subrahmanyam (2014) and Caglayman and Celiker (2016) find hedge funds and the other institutional investors differ significantly in their flow impact on well-known market anomalies, indicating that the hedge fund flow is smart and the mutual fund flow is dumb. Without detailed trading records, however, the direct evidence from portfolio rebalancing at different types of institutions is largely absent in the literature. Recently, several studies such as Lipson and Puckett (2010), Hendershott, Livdan, and Schurhoff (2015), and Kadan, Michaely, and Moulton (2016) examine institutional trading around corporate events at finer granularities using a unique data set from the NYSE's Consolidated Equity Audit Trial Data. Irvine, Lipson, and Puckett (2007) obtain a limited sample of institutional transactions from the Plexus Group to study analysts' tipping before stock recommendation initiations. Pucket and Yan (2011) find institutions profit from intra-quarter trading using data from ANcerno Ltd. A general concern regarding results from these studies is the representativeness of the samples due to coverage limitations. Alternatively, algorithms have been proposed to identify institutional orders from publicly available tick data. For example, Lee and Radhakrishna (2000, LR hereafter) use transaction sizes to differentiate retail and institution orders. Campbell, Ramadorai, and Schwartz (2009, CRS hereafter) regress the quarterly change of institutional holdings on the order imbalances of different trade size bins in TAQ and fit the relation to daily TAQ order imbalance to retrieve daily institutional order flow. These methods can be applied to all stocks at the cost of being noisy identification of institutional trades. Therefore, reducing measurement error is the key to success along this type of methodology.

In this article, we propose a new method to estimate daily aggregate institutional order flow using publicly available data. The idea is similar to CRS by extrapolating cleanly identified quarterly relation between institutional position changes and microstructure level trading data to a higher frequency. While CRS use only order imbalance of different size bins from TAQ to achieve the goal, we use both TAQ order imbalance and ANcerno's institutional order imbalance. The benefit of including the ANcerno data is not trivial. The underlying assumption of the CRS method is that institutional investors are more likely to submit large orders given their trading needs. While this assumption makes sense in perfectly liquid markets, it is less appealing when institutions take transactions cost and price impact into consideration. Indeed, Cready, Kumas, and Subasi (2014) show that institutions actively use small-size orders to manage the market impact and aggressively increase the order size during announcement periods. Therefore, the aggregate order imbalance of a given size bin can contain substantial amount of noise to represent either institutional or retail investors. Although Puckett and Yan (2011) conclude that ANcerno covers only about 10% of total institutional trading, the ANcerno data provide additional identification of institutional trades at different sizes and at the same granularity as the TAQ data.

We apply the method to all common stocks listed on NYSE, AMEX, and NASDAQ between January 1999 and March 2012. In addition to total institutional order flow, we also apply the estimation method to hedge fund and the other institutions separately. Our estimated total institutional order flow (HH hereafter) has strong serial correlations of 0.363 at the first lag and 0.164 at the fifth lag. The estimated hedge fund order flow ($SMART$) always has smaller autocorrelations than the estimated non-hedge fund ($DUMB$) order flow at all lags. HH as well as $SMART$ and $DUMB$ also has a large and positive contemporaneous price impact as the correlation with returns is around 20%. While other institutional order imbalance estimates such as LR and CRS also exhibit similar serial correlations and contemporaneous price impact, our estimated institutional order flow behaves differently in return prediction from the other methods. Given the positive serial correlations and positive contemporaneous price impact, it is expected that the institutional order flow also positively predicts future returns. This prediction is also consistent with the theoretical result from order splitting as in Chordia and Subrahmanyam (2004). In our empirical tests, however, only HH shows robust and positive predictive ability for the subsequent day's return. The LR and CRS estimates of institutional order flow positively predict subsequent returns in univarite regressions but the predictive relation turns negative and significant once aggregate order imbalance in the TAQ data is added or when the mid-quote return is used as the dependent variable instead of raw returns. In an investment analysis, we sort all stocks on the estimated institutional order imbalance every day and buy the stocks in the highest imbalance decile and sell the stocks in the lowest imbalance decile. This strategy generates an abnormal daily return of 5.7 basis points (bp) with a t-statistic of 5.59 when we use HH as the imbalance measure. The alpha with respect to the Fama-French (1993) three factors is 6.8 bp per day with a t-statistic of 7.78. When we use LR or CRS imbalances, however, the long-short strategy generates negative returns that is statistically significant for CRS and insignificant for LR. The robustness of our method is favorable over traditional methods and we believe the benefit comes from additional and cleaner identification from using the Ancerno data.

We also find that both $SMART$ and $DUMB$ order imbalances have positive and significant price impact on the following day. However, $SMART$ imbalance has greater statistical and economic significance than $DUMB$. In our benchmark multivariate regression, the coefficient estimate of $SMART$ is 1.178 with a t-statistic of 15.51. A one-standard deviation increase in $SMART$ is expected to increase the next day's return by 3 bp. The coefficient estimate of $DUMB$ is 0.139 with a t-statistic of 9.19. A one-standard deviation increase in $DUMB$ is expected to increase the next day's stock return by 2 bp. The investment analysis confirms the positive price impact from both $SMART$ and $DUMB$. The long-short portfolio generates an average daily abnormal return (alpha) of 6.4 (7.5) bp for $SMART$ imbalance, and an average daily abnormal return (alpha) of 3.5 (4.4) bp for $DUMB$ imbalance. All of the abnormal returns and alphas are statistically significant at the 1% level. The larger price impact and profitability of $SMART$ indicates that hedge funds can have better trading skills than the other institutional investors on average.

We turn to the return predictability in the subsamples next. We find that institutional trading as a whole, has larger price impact for small, illiquid, and Nasdaq stocks and the price impact becomes weaker in the recent period. The results are consistent with both an information based story and a liquidity based story because the level of information asymmetry as well as illiquidity is higher for small and Nasdaq stocks, and in the early period. Moreover, we explore the difference between hedge fund and non-hedge fund trading in the subsamples too. We find that hedge fund order imbalance has positive and significant price impact on the next day for both large and small firms although the predictive ability is stronger for small firms. However, the non-hedge fund order imbalance is able to predict returns only for small firms. We find similar results when we use stock illiquidity as a conditioning variable. Hedge fund order imbalance predicts future returns in both liquid and illiquid stock groups but the non-hedge fund order imbalance positively predicts returns only in the illiquid stock group. When we split the sample by exchanges, we find the hedge fund order imbalance has similar predictive ability in NYSE, AMEX, and Nasdaq stocks but the predictive ability of non-hedge fund imbalance is stronger for Nasdaq stocks than stocks on the other two exchanges. Finally, the predictive ability of $SMART$ slightly weakens over time while the predictive ability of $DUMB$ seems unaffected by time. These results suggest that hedge funds can have better trading skills on average than non-hedge fund institutions but that advantage is shrinking over time, possibly due to competition among hedge fund managers.

After documenting positive and significant price impact of institutional trading in the short term, we then investigate if the predictive ability of institutional order flow is a result of informed trading by institutional investors. If institutions, either as a whole group or the smart component of it, bring fundamental information to the market, we expect the price impact to be at least partially permanent. We examine the relation between long-term cumulative returns and estimated institutional order imbalances as well as the behavior of institutional imbalance around significant corporate events to answer the question. In the cross section, we find that the positive price impact of institutional trading completely reverses in a week. When we compare the long-term price impact of $SMART$ and $DUMB$ order imbalances, we find that both types of institutional trading has only transitory price impact. However, the hedge fund imbalance, $SMART$, has more persistent price impact than non-hedge fund imbalance, $DUMB$. Therefore, the taking-profit window could be longer for hedge funds engaging in short-term oriented trading. In the event study using both scheduled and unscheduled corporate events including earnings announcements, analyst recommendation changes, price jumps, 8k filings, and 13D filings, we do not observe abnormal behavior of any institutional order flow estimate. The results combined seem to suggest that the return predictability of institutional trading is unlikely to come from informed trading by institutions.

Finally, we use the estimated institutional order flow to examine the relation between well-known anomalies and institutional trading. We use Stambaugh, Yu, and Yuan's (2015) mispricing index to measure the aggregate anomaly effect on individual stocks. The mispricing index is constructed at the end of each month. With daily order imbalance estimates, we can examine how institutions trade on the anomaly signals around month ends. We uncover significant difference in the response of hedges and non-hedge funds. The hedge fund order imbalance, $SMART$, on the last trading day of a month, as well as the cumulative imbalance over the next one to five days, is significantly and positively correlated with the expected stock return due to mispricing. On the other hand, the non-hedge fund imbalance, $DUMB$, has positive but insignificant relations with the mispricing index. Our findings suggest that hedge funds rebalance portfolios around month ends to capture the mispricing signals while non-hedge funds largely ignore those signals.

We make several contributes to the finance literature. First, we introduce a new method of estimating institutional order flow for individual stocks at the daily level. Empirical analysis shows that this new method has more robust performance in terms of return predictability than prior methods by Lee and Radhakrishna (2000) and Campbell, Ramadorai, and Schwartz (2009). Our method can be applied in many empirical studies that examine institutional trading behavior at the daily frequency. Second, our findings provide new evidence of the effect of institutional trading in financial markets. While many studies find that institutions as a whole group, or some types of institutions trade on advanced information ahead of corporate events, we find that on an average day, institutional trading presents only transitory price pressure on the stock. Third, we find that hedge funds appear smarter than the other institutional investors because their trading generates greater and more persistent price impact, hence a longer profit-taking window, and hedge funds trade on well-known stock return anomalies at the end of the month. These findings complement the studies using longer-horizon observations such as Frazzini and Lamont (2008), Akbas and Armstrong, Sorescu, and Subrahmanyam (2014) by providing direct evidence at a finer granularity.

The rest of the paper is organized as follows. Section II describes how to construct our sample data and estimate institutional order flows. Section III reports the return predictive power of our estimated institutional order flow, comparing with other estimated institutional order flows. Section V documents whether institutional order flow can capture fundmantal information flow around corporate events such as earnings announcement, extreme price movement, analyst recommendation update, value-related 8-K filing, scheduled 13-D filing. Lastly, we conlude in Section VI.

II. Data and variable description

A. Sample selection

We employ four data sets, Trades and Automated Quotes (hereafter TAQ), ANcerno Ltd institutional trading data (hereafter AN), Thomson Reuters Legacy Institutional Holdings Data (hereafter 13F), and Center of Research in Security Prices (hereafter CRSP) in the study. Our sample includes all of common stocks in three exchange markets, e.g., NYSE, AMEX, and NASDAQ, from January 1999 to March 2012 where AN covers. From TAQ, we extract trade and quote messages between 9:30 AM to 4 PM EST with positive trading price and trading volume. After that, we estimate stock-day order imbalance from Lee and Ready (1993) algorithm in nineteen size bins whose lower cutoffs are \$0, \$2,000, \$3,000, \$5,000, \$7,000, \$9,000, \$10,000, \$20,000, \$30,000, \$50,000, \$70,000, \$90,000, \$100,000, \$200,000, \$300,000, \$500,000, \$700,000, \$900,000, and \$1 million. From AN, we obtain stock-day institutional order flow of ANcerno Ltd. in the above nineteen trade-size bins. From 13F, we have quarterly institutional holdings as well as its quarterly change. From CRSP, we extract information on common stock characteristics including daily stock return, daily stock price, close bid and ask prices, shares outstanding, and daily trading volume. We exclude observations from our sample data if they have a price lower than five dollars and if their relative bid-ask spread, defined as bid-ask spread scaled by the average of bid and ask prices, is outside the interval between zero and one half.

B. Estimation of institutional order flow

We estimate institutional order flow (hereafter IOF) in three ways, e.g., a cut-off rule (LR hereafter) following Lee and Radhakrishna (2000, LR hereafter), a quarterly regression model ($CAMPBELL$ hereafter) following Campbell, Ramadorai, and Schwartz (2009, CRS hereafter), and our proposed regression model (HH hereafter).

LR is based on a \$5,000 cut-off rule. The cut-off rule is under an assumption where individual investors are more likely to submit small trade-size orders than institutional investors due to the limitation of investment budget. Trades with its trade-size above \$5,000 are classified as those organized by institutional investors but trades below \$5,000 are classified as those established by individual investors. Although LR recommend \$20,000 cut-off rule to distinguish institution-initiated order flow from individual-initiated order flow, we choose \$5,000 as our lower cut-off because IOF based on the \$5,000 lower cut-off shows the strongest predictive power for daily future return among \$5,000, \$10,000, \$20,000, \$50,000, and \$100,000 cut-off rules.

CRS propose a regression methodology extrapolating daily IOF from quarterly relation between 13F institutional holding change and TAQ order imbalance. LR is likely to mis-estimate IOF because it, by definition, ignores small trade-size trades initiated by institutional investors. However, institutional investors submit small-size orders to circumvent concerns about price impact of large-size orders from illiquidity and information leakage caused by large-size orders. $CAMPBELL$ is constructed in two stages of calculation to exploit the information in diverse trade-size order flows. In an estimation stage, CRS regress change of quarterly institutional holding on order imbalance in nineteen trade-size bins as the following regression model.

$$\Delta Y_{i,q} = \alpha_q + \rho \Delta Y_{i,q-1} + \phi Y_{i,q-1} + \beta^U U_{i,q} + \beta^{UY} Y_{i,q} \times U_{i,q} + \sum_{Z=1}^{19} \beta_F^Z F_{Z,i,q} + \epsilon_{i,q} \qquad (1)$$

, where α is four quarter dummies, $Y_{i,q}$ is quarterly institutional holdings of a stock i at quarter q, $U_{i,q}$ is undefined order flow of a stock i at quarter q, and $F_{Z,i,q}$ is quarterly aggregated order imbalance of a trade-size bin Z from TAQ of a stock i at quarter q. Also CRS report that the distribution of trade intensities in different trade size depends on market capitalization of stocks. Therefore, they run the above regression model in each quintile size portfolio which is constructed based on NYSE breakpoints of market capitalization. In addition, CRS estimate β_F^Z in a non-linear form suggested by Nelson and Siegel (1987) to model yield curves.

$$\beta_F^Z = b_{01} + b_{02}Y + (b_{11} + b_{12}Y + b_{21} + b_{22}Y)[1 - e^{-Z/\tau}]\frac{\tau}{Z} - (b_{21} + b_{22}Y)e^{-Z/\tau} \qquad (2)$$

In retrieval stage, they recover daily estimated IOF from the following equation by using the estimated coefficients in the above regression model.

$$\Delta Y_{i,d} = \widehat{\beta^U} U_{i,d} + \widehat{\beta^{UY}} Y_{i,d} \times U_{i,d} + \sum_{Z=1}^{19} \widehat{\beta_F^Z} F_{Z,i,d} \qquad (3)$$

, where d indexes daily observations.

In the same spirit of CRS, we propose a new regression methodology to estimate IOF, extrapolating from quarterly relation of 13F institutional holding change not only with TAQ order imbalance but also with AN order imbalance. TAQ is noisy information resource for estimated IOF since it includes order flow from individual investors as well as institutional investors. In contrast, AN deals with daily institutional trading only which covers about ten percent of total institutional trading, according to Puckett and Yan (2011). Adding information on actual IOF from AN in Equation (1) and (3) may reduce estimation errors of β_F^Z and therefore estimated IOF may become more accurate. This paper provides evidence that the addition of actual IOF improves the predictive power of estimated IOF for future returns.

We constructed our estimated IOF, HH, in two stage of calculation, following CRS. In an estimation stage, we run the following regression model within each quintile size portfolio.

$$\Delta Y_{i,q} = \alpha_q + \rho \Delta Y_{i,q-1} + \phi Y_{i,q-1} + \beta^U U_{i,q} + \beta^{UY} Y_{i,q} \times U_{i,q} + \sum_{Z=1}^{19} \beta_F^Z F_{Z,i,q} + \sum_{Z=1}^{19} \beta_D^Z D_{Z,i,q} + \epsilon_{i,q} \qquad (4)$$

, where α is four quarter dummies, $Y_{i,q}$ is quarterly institutional holdings of a stock i at quarter q, $U_{i,q}$ is undefined order flow of a stock i at quarter q, and $F_{Z,i,q}$ and $D_{Z,i,q}$ are quarterly aggregated order imbalance of a trade-size bin Z from TAQ and AN of a stock i

at quarter q, respectively.

We estimate β_F^Z and β_D^Z in a non-linear form of the equation (2), following CRS. Since non-linear regression relies on iterative numerical analysis based on non-linear least-squares estimation, it sometimes encounters a convergence problem. The convergence problem mainly comes from how to set the value of τ in the equation (2). Also the estimated coefficients are sensitive to the initial value of τ. To quench the convergence concern and τ sensitivity concern, we set three sequential strategies. First of all, we iterate running the non-linear regression model with the different value of τ. From the iteration of estimation, we obtain the sum of squared errors of prediction (SSE hereafter) and then we consider the value of τ with global minimum value of SSE to be appropriate in terms of least-squares estimation. The second strategy is necessary in the case we cannot find global minimum value of SSE. We cannot solve the convergence problem itself, but we are able to circumvent τ sensitivity concern by checking the sensitivity of $\widehat{\beta_F^Z}$ and $\widehat{\beta_D^Z}$ to the value of τ. We compute $\widehat{\beta_F^Z}$ and $\widehat{\beta_D^Z}$ from non-linear estimation with different value of τ, and then we check whether they converge at certain points as the value of τ increases. We consider the value of τ to be appropriate in terms of τ insensitivity when $\widehat{\beta_F^Z}$ and $\widehat{\beta_D^Z}$ converge to some degree. Lastly we need to take the last strategy when the first and second strategies do not work. We set the value of τ to our arbitrary maximum value like 100,000.

In retrieval stage, we recover daily estimated IOF from the following equation by using the estimated coefficients in the above regression model.

$$\Delta Y_{i,d} = \widehat{\beta^U} U_{i,d} + \widehat{\beta^{UY}} Y_{i,d} \times U_{i,d} + \sum_{Z=1}^{19} \widehat{\beta_F^Z} F_{Z,i,d} + \sum_{Z=1}^{19} \widehat{\beta_D^Z} D_{Z,i,d} \tag{5}$$

, where d indexes daily observations.

In addition, We create two more IOFs for hedge fund (hereafter, $SMART$) and non-hedge fund (hereafter, $DUMB$), following the estimation methodology of HH. The only difference is that we use quarterly hedge and non-hedge fund holdings, respectively, instead of all institutional holdings for $Y_{i,q}$ in Equation (4). We follow Agarwal, Jiang, Tang, and Yang (2013) in order to correctly classify 13F holdings into hedge funds and non-hedge funds categories.

C. *Other variables and outlier control*

We calculate total order imbalance in TAQ (TAQOI) and AN (ANOI), share volume turnover ratio (TURN), and relative bid-ask spread (BASPRD) for each stock-day in order to obtain control variables for return prediction models. The detailed definitions are following.

- TAQOI: the number of buyer-initiated shares less the number of seller-initiated shares in TAQ from Lee and Ready algorithm, scaled by the number of shares outstanding for each stock-day.
- ANOI: the number of buyer-initiated shares less the number of seller-initiated shares in AN, scaled by the number of shares outstanding for each stock-day.
- TURN: daily trading volume over the number of shares outstanding.
- BASPRD: the difference of bid and ask prices scaled by the average of bid and ask prices for each stock-day.

After variable construction we control outliers in three ways. Firstly, we remove odd observations with negative market capitalization, relative bid-ask spread below zero and above one half, and negative turnover ratio. Secondly, we get rid of penny stocks below $5 of a stock price in order to circumvent any concern related with unexpected market microstructure effects. In addition, we conduct time-series winsorization on every independent variable at 1 and 99 percent to mitigate the effect of outliers in our sample.

D. *Summary statistics*

[Place Table I about here]

Table I documents summary statistics of our sample. Panel A is for descriptive statistics. The number of dates in our study is 3,517 from January 1999 to March 2012, and the average number of stocks per day is about 3,230. Our estimated IOF, HH, has positive mean similar to $CAMPBELL$ with 3.7% of daily order imbalance. That is, institutional investors are likely to buy, rather than to sell on average. In addition, the standard deviation of HH is 0.212 higher than other estimated IOFs. LR is very similar to $TAQOI$ in terms of mean, standard deviation, minimum, median, and maximum. This is because, by construction, LR simply set observations below \$5,000 trade size to zero. $SMART$ has smaller mean, standard deviation, minimum, median, and maximum than $DUMB$ has.

Panel B is for autocorrelation of estimated IOFs. HH shows high and positive autocorrelation of 0.363, which is similar to $ANOI$. The characteristic of HH is consistent with Chordia and Subrahmanyam (2004) where they argue that institutional investors are likely to split their order inter-day to avoid price impact and information leakage from their trading. Comparing with HH and $ANOI$, other IOFs such as $CAMPBELL$, LR, and $TAQOI$ have lower but positive autocorrelation. Again LR is statistically similar to $TAQOI$ even in Panel B because LR simply classify order flows beyond \$5,000 trade size in institutional order flow. SMART is less serially correlated than DUMB.

Panel C is for correlation between our control variables. HH has reasonably high correlation with other IOFs including $ANOI$ from 0.405 to 0.941. $CAMPBELL$ is also highly correlated with other IOFs except $ANOI$. In particular, its correlation with $TAQOI$ is 0.769 while it shows weak correlation with $ANOI$. LR is similar to $CAMPBELL$; it is highly correlated with $TAQOI$ by construction but weakly with $ANOI$. The correlation coefficients of $CAMPBELL$ and LR implies that two estimated IOFs, $CAMPBELL$ and LR, may not include enough information on institutional trading. When it comes to $SMART$ and $DUMB$, they are reasonably correlated with one another in the correlation of 0.641. Also $SMART$ is less correlated with HH and $ANOI$, comparing with $DUMB$. All the estimated IOFs are more likely to buy liquid stocks in terms of turnover ratio and bid-ask spread. In addition, all the IOFs except $ANOI$ are well associated with contemporaneous stock returns, which means that they have an impact on contemporaneous stock price.

E. Seasonality

[Place Figure 1 about here]

Figure 1 describes the seasonality in estimated IOFs; weekday seasonality in Panel A, week seasonality in Panel B, and month seasonality in Panel C. Panels have two figures. We analyze estimated IOFs on the top figure and their absolute values on the bottom figure. Both the top and bottom figures show similar pattern to one another. In Panel A, we take an average of weekday IOFs and scale it by IOF on Monday. All the IOFs monotonically increase from Monday to Friday. Particularly, all of them except LR show a clear rise on Friday. We do not have a clear answer to why estimated IOFs have weekday seasonality. In Panel B, we calculate weekly aggregated IOFs for the first, last, or the other weeks in a given month, scaling it by IOF in the first week. All the IOFs except LR have the lowest value in the last week, implying that institutions are more likely to submit sell orders in the last week. In Panel C, we take an average of monthly IOFs and scale it on January. According to the top figure, estimated IOFs, $SMART$ in particular, have quarter seasonality; it increases on March, June, September, and December.

III. Cross-sectional return prediction

A. Daily return prediction

[Place Table II about here]

This table presents estimated coefficients from Fama-MacBeth (1973) regression to measure return predictability of four different estimated institutional order flows (estimated

IOFs),

$$\begin{aligned}\mathrm{RET}_{i,t} &= \alpha_t + \sum_{k=1}^{5} \beta_{t,k} \mathrm{IOF}_{i,t-k} + \sum_{k=1}^{5} \gamma_{t,k} \mathrm{TAQ}_{i,t-k} + \gamma_t^B \mathrm{BASPRD}_{i,t-1} + \gamma_t^T \mathrm{TURN}_{i,t-1} \\ &\quad + \sum_{k=1}^{5} \gamma_{t,k}^R \mathrm{RET}_{i,t-k} + \sum_{k=1}^{5} \gamma_{t,k}^{R2} \mathrm{RET}_{i,t-k}^2 + \epsilon_{i,t}\end{aligned}$$

, where for stock i on day t, α is a weekday dummy, RET is (mid-quote) stock return adjusted by Fama-French three factors, and IOF is HH, $CAMPBELL$, LR, $SMART$, or $DUMB$. All the IOFs except $CAMPBELL$ and LR have predictive power for one-day-ahead stock returns. We also add control variables, relative bid-ask spread ($BASPRD$), turnover ratio ($TURN$), lagged returns (RET), and lagged squared returns (RET^2). We put those control variables to isolate the effect of lagged IOFs on current stock returns. $BASPRD$ has a positive sign with is consistent with Amihud and Mendelson (1986, 1989). This is because, according to the model in Amihud and Mendelson (1986), market participants expect higher returns when they put their money into stocks with wider bid-ask spread. $TURN$ also have desirable signes in all the regression models. Gervais, Kaniel, and Mingelgrin (2001) prove that there is the high-volumn return preminum resulted from stock's visibility. The positive sign of estimated coefficients on $TURN$ indicates the high-volume return premium. In addition, all the lagged returns are negative because of stock return reversal. The lagged squared returns represent volatility of returns, so it is natural that higher lagged squared returns lead higher current returns according to the expectation of high return from high risk.

Table II indicates that HH outperform $CAMPBELL$ and LR in terms of statistical significance and economical significance. The t-statistics of HH is 9.68 higher than those of $CAMPBELL$ with -9.97 and LR with -14.05. Also the economical significance of HH is 4.77% from the multiplication of estimated coefficient, 0.214, with one standard deviation, 0.223. That is, the increase of one standard deviation of HH makes price impact of 4.77% per day. On the other hands, the economic significane of $CAMPBELL$ and LR is 1.99%

and 2.59%, respectively, which is about half of HH's. IOFs are considered as informed order flow because institutional investors have better environment to gether fundamental information on a particular securities than individual investors and therefore they are more likley to make profit from their investment than individual investors. From this point of view, HH seems to better capture IOFs than the IOFs in prior studies, $CAMPBELL$ and LR.

B. *Investment strategy*

[Place Table III about here]

Table III documents the profitability of investment strategy based on one-trading-day lagged IOFs. We rank all the stocks in our sample by one-trading-day lagged IOFs for each day, and classify them into decile portfolios. Stocks with the lowest (highest) IOF belong to *Low (High)* portfolio. We take short positions for stocks in the *Low* portfolio and long position for stocks in the *High* portfolio at day t. HH is the only IOF with positive and significant investment profit. Its profit of investment strategy is 5.60% per day with annual Sharpe Ratio of 146.20%. Other IOFs make negative or insignificant investment profit. However, the performance of decile portfolios based on one-trading-day lagged HH is not monotonically increasing from the 'Low' portfolio to the 'High' portfolio. The daily performance of 'Low' portfolio is 10.90% per day, but the '5' portfolio has lower daily performance of 2.10%. Other IOFs such as $CAMPBELL$, LR, and $TAQOI$ also show similar patterns over the decile portfolios.

Table III is consistent with the previous tables; HH is more informative for future stock returns than $CAMPBELL$, LR, and $TAQOI$. From the perspective that institutional investors are informed, HH is a better proxy for IOF. This is because investment strategy based on HH is more profitable with statistical significance. However, the non-monotonical increase in decile portfoios may mean that HH is not informative and its profitability is random coincidence. Since institutional investors are likely to have low preference on investing in illiquid stocks for the reason of transaction costs, the non-linear of decile portfolio performance may come from liquidity premium. Hence Table III have the same implication as the previous tables provide. The following tables also indicates not only that the performance of HH investment strategy is profitable but also that it is persistent over our sample period.

[**Place Figure 2 about here**]

This figure shows time evolution of investment performance based on one-trading-day lagged IOFs. We cumulate daily performance of investment strategies in Table III during our whole sample period from January 1999 to March 2012. As Table III shown, the investment strategy based on HH is profitable only while other strategies is not profitable at all except before 2001. All the investment strategies are profitable before 2001, but they experience sudden fall in their performance in 2001. After that, strategies based on $CAMPBELL$, LR, and $TAQOI$ keep losing money.

Consistent with Table III, Figure 2 also proves that HH investment strategy is lucrative. Its profitability lasts for the whole sample period except sudden drop in 2001. The sudden drop can be explained in two ways; 2000 Regulation Fair Disclosure (Reg FD hereafter) and 2001 IT bubble. The Reg FD mandates companies in exchange markets disclose material information to the public, so it may harm the information advantage of institutional investors against individual investors. Before 2001 IT bubble, an equity market was in a bull market and investors are likely to get profit from their trading. During IT bubble, however, the equity market suddenly shifted from the bull market to a bear market. The sudden drop may indicates the shift of market condition.

C. Subsample tests

This table presents estimated coefficients from the regression model in Panel A of Table II in order to examine predictive power of estimated IOFs in subsamples based on firm characteristics.

[Place Table IV about here]

Table IV presents the predictive power of IOFs in size subsample. We separate whole sample dataset into five subsamples based on market capitalization. In this table, we report Fama-MacBeth coefficients in three subsample regression. Panel A is for the samllest-size stocks, Panel B is for middle-sized stocks, and Panel C is for the largest-size stocks. The one-day lagged of HH is positive and significant regardless with the market capitalization of stocks while $CAMPBELL$, LR, and $TAQOI$ are negative or insignificance in Panel B and C for middle-size and large-size stocks, repectively.

Panel A of Table IV also proves that HH seems a better measure for IOF. Institutional investors tend to trade stocks with large market capitalization for liquidity reason, and therefore IOF is supposed to predict future stock returns of large-size stocks. According to middle-sized and large-sized stock columns, $CAMPBELL$ and LR is not informative while HH shows predictive power for future stock returns. The predictive power of $CAMPBELL$ and LR in Panel A could be explained by illiquidity of small-size stocks even though $CAMPBELL$ and LR have positive and significance estimated coefficients in Panel A.

Panel B of Table IV reports return predictability of IOFs in liquidity subsamples. We separate whole sample dataset into five subsamples based on relative bid-ask spread ($BASPRD$). In this table, we report Fama-MacBeth coefficients in three subsample regression. Panel A is for stocks with the narrowest $BASPRD$, Panel B is for stocks with midium $BASPRD$, and Panel C is for stocks with the widest $BASPRD$. In Panel A, no IOFs are informative for one-day-ahead stock returns. However, in Panel B, HH only can predict future stock return at 1% significance with t-statistics of 10.58, and in Panel C, all the IOFs have positive and significant estimated coefficients in their first lagged IOFs.

Panel B of Table IV also shows the same implication which HH is a better proxy for

IOF than $CAMPBELL$ and LR. Institutional investors need liquidity to minimize price impact and information leakage. That is, they prefer trading liquid stocks to trading illiquidity stocks. Therefore correctly estimated IOFs should have predictive power even within liquid subsamples. Although HH loses its predictive power in Panel A, it shows significant predictive power in Panel B while $CAMPBELL$ and LR does not. The predictive power of $CAMPBELL$ and LR in Panel C may come from illiquidity because wide bid-ask spread is likely to cause stock price to sensitively react to unbalanced order imbalance.

Panel C of Table IV reports return predictability of IOFs in each subperiod. We separate whole sample dataset into three subperiods. Panel A is for early subperiod from 1999 to 2002, Panel B is for middle subperiod from 2003 to 2007, and Panel C is for late subperiod from 2008 to 2012. HH is the only IOF which shows predictive power for future stock returns over the whole sample period from 1999 to 2012. Other IOFs such as $CAMPBELL$, LR, and $TAQOI$ is predictive for stock retrun after 2003.

In Panel C of Table IV, the economic significance and statistical significance of HH always dominates those of other IOFs. The economic significance of one-day lagged HH is 8.07%, 2.99%, and 3.77% while LR is 3.69%, 1.29%, 3.09% in Panel A, B, and C, respectively. T-statistics of one-day lagged HH is 2.29, 13.54, and 11.16 while LR is 1.42, 5.37, and 6.28 in Panel A, B, and C, respectively. $CAMPBELL$ is even lower than LR in terms of economic and statistical significance.

Panel D of Table IV documents the prediction power of IOFs in different exchange markets, NYSE and AMEX versus Nasdaq. We separate whole sample dataset into two subsamples based on an exchange market. Panel A is for NYSE and AMEX, and Panel B is for Nasdaq. Regardless with the exchange markets, all the IOFs have strong predictive power for future stock returns. The value of estimated coefficients in NYSE and AMEX in Panel A is half of their value in Nasdaq in Panel B. This is because Nasdaq holds smaller size stocks than NYSE and AMEX do and therefore the effect of IOF is stronger on stocks in Nasdaq than stocks in NYSE and AMEX due to illiquidity of smaller stocks.

D. Price impact

[Place Figure 3 about here]

This figure describes k estimated coefficients of the first lagged IOFs from the following Fama-MacBeth regression model in order to gauge long-term return predictability of four different IOFs,

$$\mathrm{CR}_{i,t,t+k} = \alpha_t + \beta_t \mathrm{IOF}_{i,t-1} + \epsilon_{i,t}$$

, where $CR_{i,t,t+k}$ is raw cumulative return of stock i from day t to $t+k$, and $IOF_{i,t}$ is HH, $CAMPBELL$, LR, or $TAQOI$ of stock i on day t. Figure 3 visually shows evidence that estimated IOFs contain transient price impact only, rather than permanent price impact. The predictive power of HH lasts for about three days, but it is quickly diminishing. $CAMPBELL$ does not even show any predictive power for future stock returns consistent with Table I. LR and $TAQOI$ can predict stock returns on the right next day. Figure **??** implies that IOFs may not contain fundamental information but make a temporaneous price impact.

IV. Information flow around corporate events

[Place Figure 4 to Figure 8 about here]

We study five corporate events; earnings announcements in Figure 4, extreme price movement in Figure 5, recommendation updates in Figure 6, value related 8K filings in Figure 7, and scheduled 13D filings in Figure 8. We gether the earnings announcement date and recommendation update date from I/B/E/S. We define extreme price movement as daily unreversed abnormal return above two standard deviations for abnormal return in the last twenty trading days. Abnormal return is a residual term of Fama-French three factor regression model. Moreover, we collect the 8-K and 13-D filing dates from WRDS SEC Analytics

Suites. All the IOFs cannot capture fundamental information flow on corporate events.

The dynamic of IOFs is consistent with Figure 3 where IOFs does not have any permanent price impact and therefore it does not include fundamental information on a given stock. Our event studies also indicate that IOFs do not contain fundamental information on corporate events. Together with Figure 3, estiamted IOFs including HH have only transitory price impact but not permanent price impact which comes from fundamental information flow.

V. Relation between anomalies and institutional trading

[Place Table V about here]

Table V presents Fama-MacBeth (1973) regression results for the following equation between 1999 and 2012,

$$\text{MISPRICING}_{i,m} = \alpha_m + \beta_m \text{IOF}_{i,t,t+k,m} + \epsilon_m$$

, where for stocks i on month m, MISPRICING is mispricing index suggested by Stambaugh, Yu, and Yuan (2012, 2015), and $\text{IOF}_{t,t+k}$ is an cumulative IOF from the end of month t to $t+k$. HH and $SMART$ trade in the direction where mispricing is mitigiated.

VI. Conclusion

In this article, we propose a new methodology to estimate institutional order flow (IOF hereafter) that improves the method of Campbell, Ramadorai, and Schwatz (2009, CRS hereafter). CRS extrapolates daily IOF from the relation between quarter change in institutional holdings and quarterly aggregate TAQ order imbalances in several trade-size bins. Our estimation method is similar to CRS but we add actual institutional order flow from

ANcerno trade data along with TAQ order flow. We expect our estimated IOF (HH) to be a better measure than CRS's because we utilize actual invitational order flow to minimize estimation errors caused by TAQ mixed order flow with invitational and individual order flow.

Our empirical analysis shows that HH has more robust predictive ability about future stock returns than benchmark IOFs such as CRS's IOF ($CAMBPELL$) and a $5,000 cut-off rule ($LR$). The return predictability of HH is robust in four subsample tests based on firm size, liquidity, subperiod, and exchange market. Moreover, investment strategy based on HH is profitable while the other TAQ-based IOFs, $CAMPBELL$ and LR, are not. Applying our estimation methods to hedge funds and non-hedge funds separately, we find that both hedge fund ($SMART$) and non-hedge fund ($DUMB$) order imbalances have positive and significant impact on the next day's return. The hedge fund price impact is statistically and economically larger, more robust in subsamples, and more persistent than the non-hedge fund price impact. Nevertheless, we do not find that any of the estimated institutional order flow including HH and $SMART$ can capture permanent information flow in the cross section. The return predictive power of all the IOFs disappears within three days in long-term return prediction models. Also the IOFs do not seem to show abnormal behavior prior to significant corporate events. Those suggest that institutional investors do not trade on information generally. If institutional investors, including hedge funds, do not trade on information, their profitability must depend on something else. We find that hedge funds actively trade on well-known return anomalies around month ends while non-hedge funds largely ignore the anomalies. Given well documented anomaly returns, it is possible that hedge funds outperform their institutional peers by capturing such predictive return patterns.

The proposed estimation method for institutional order flow can be applied in other empirical studies that require institutional order flow estimates at the daily frequency. We apply the method to hedge funds and non-hedge funds in this study. The method can also be applied to other types of institutional trading such as long-term and short-term investors, and active and passive investors. Our analysis is based on stock level institutional order imbalance, it would be interesting to combine it with fund level analysis to gain a more complete picture of the effect from institutional trading. We leave this question to future studies.

REFERENCES

[1] Amihud, Y., and H. Mendelson, 1986, Asset pricing and the bid ask spread, Journal of Financial Economics 17, 223-249.

[2] Amihud, Y., and H. Mendelson, 1989, The effects of beta, bid-ask spread, residual risk, and size on stock returns, Journal of Finance 44, 479- 486.

[3] Andy, Lipson Marc; Pucket, 2010, Institutional trading during extreme market movements, Working paper.

[4] Balkrishna, Lee Charles M. C.; Radhakrishna, 2000, Inferring investor behavior - evidence from torq data, Journal of Financial Market 3, 29.

[5] Bennett, J. A., R. W. Sias, and L. T. Starks, 2003, Greener pastures and the impact of dynamic institutional preferences, Review of Financial Studies 16, 1203-1238.

[6] Bhaskaran, Burch R. Timothy; Swaminathan, 2002, Earnings news and institutional trading, Working paper.

[7] C., Kadan Ohad; Michaely Roni; Moulton Pamela, 2016, Trading in the presence of short-lived private information: Evidence from analyst recommendation changes, Working paper.

[8] Campbell, Ramadorai, Schwatz, 2009, Caught on tape, Journal of Financial Economics 92, 26.

[9] Chen, H. L., N. Jegadeesh, and R. Wermers, 2000, The value of active mutual fund management: An examination of the stockholdings and trades of fund managers, Journal of Financial and Quantitative Analysis 35, 343-368.

[10] Chordia, T., and A. Subrahmanyam, 2004, Order imbalance and individual stock returns: Theory and evidence, Journal of Financial Economics 72, 485-518.

[11] Cready, W., A. Kumas, and M. Subasi, 2014, Are trade size-based inferences about traders reliable? Evidence from institutional earnings-related trading, Journal of Accounting Research 52, 877-909.

[12] Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, Journal of Finance 52, 1035-1058.

[13] Fama, E. F., and J. D. Macbeth, 1973, Risk, return, and equilibirum: Empirical tests, Journal of Political Economy 81, 607-636.

[14] Gervais, S., R. Kaniel, and D. H. Mingelgrin, 2001, The high-volume return premium, Journal of Finance 56, 877-919.

[15] Gompers, P. A., and A. Metrick, 2001, Institutional investors and equity prices, Quarterly Journal of Economics 116, 229-259.

[16] Grinblatt, M., S. Titman, and R. Wermers, 1995, Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, American Economic Review 85, 1088-1105.

[17] Hendershott, T., D. Livdan, and N. Schurhoff, 2015, Are institutions informed about news?, Journal of Financial Economics 117, 249-287.

[18] Irvine, P., M. Lipson, and A. Puckett, 2007, Tipping, Review of Financial Studies 20, 741-768.

[19] Lee, C. M. C., and M. J. Ready, 1991, Inferring trade direction from intraday data, Journal of Finance 46, 733-746.

[20] Lu, Cai Feng; Zheng, 2004, Institutional trading and stock returns, Finance Research Letters 1, 178.

[21] M., Cready. W.; Kumas A.; Subasi, 2014, Are trade size-based inferences about traders reliable evidence from institutional earnings-related trading, Journal of Accounting Research 52, 33.

[22] Newey, W. K., and K. D. West, 1987, A simple, positive semidefinite, heteroskedasticity and autocorrelation cosistent covariance-matrix, Econometrica 55, 703-708.

[23] Nofsinger, J. R., and R. W. Sias, 1999, Herding and feedback trading by institutional and individual investors, Journal of Finance 54, 2263-2295.

[24] Puckett, Andy, and Xuemin Yan, 2011, The interim trading skills of institutional investors, Journal of Finance 66, 601-633.

[25] Wermers, R., 1999, Mutual fund herding and the impact on stock prices, Journal of Finance 54, 581-622.

Table I
Summary statistics

This table shows the time-series averages of the cross-sectional statistics for the sample during January 1999 to March 2012 sample period. We combine daily Center for Research in Security Prices (daily CRSP) with Trades and Quotes (TAQ) and ANcerno Ltd. institutional trading (AN) database. We obtain estimated institutional order flow (HH) from two stages of calculation; estimation stage and retrieval stage. In estimation stage, we regress the following regression model,

$$\Delta Y_{i,q} = \alpha_q + \rho \Delta Y_{i,q-1} + \phi Y_{i,q-1} + \beta^U U_{i,q} + \beta^{UY} Y_{i,q} \times U_{i,q} + \sum_{Z=1}^{19} \beta_F^U F_{Z,i,q} + \sum_{Z=1}^{19} \beta_D^U D_{Z,i,q} + \epsilon_{i,q}$$

, where α is four quarter dummies, $Y_{i,q}$ is quarterly institutional holdings of a stock i at quarter q, $U_{i,q}$ is undefined order flow of a stock i at quarter q, $F_{Z,i,q}$ is quarterly aggregated order imbalance of a trade-size bin Z from TAQ of a stock i at quarter q, and $D_{Z,i,q}$ is quarterly aggregated order imbalance of a trade size bin Z from AN of a stock i at quarter q. We estimate β_F^U and β_D^U in a non-linear form following Campbell, Ramadoral, and Schwatz (2009). In retrieval stage, we recover daily estimated institutional order flow from the following equation by using the estimated coefficients in the above regression model.

$$\Delta Y_{i,d} = \widehat{\beta^U} U_{i,d} + \widehat{\beta^{UY}} Y_{i,d} \times U_{i,d} + \sum_{Z=1}^{19} \widehat{\beta_F^U} F_{Z,i,d} + \sum_{Z=1}^{19} \widehat{\beta_D^U} D_{Z,i,d}$$

, where d indexes daily observations. To estimate hedge fund order flow ($SMART$) at daily level, we employ quarterly hedge fund holding for $Y_{i,q}$. We also put quarterly non-hedge fund holding in $Y_{i,q}$ to measure non-hedge fund order flow ($DUMB$). $CAMPBELL$ is estimated institutional order flow proposed by Campbell, Ramadoral, and Schwatz (2009). It utilizes similar estimation methodology to HH except not including $D_{Z,i,q}$'s. LR is estimated institutional order flow from Lee and Radhakrishna (2000) cutoff rules of \$ 5,000. $ANOI$ is daily institutional order flow in ANcerno Ltd.. $TAQOI$ is daily total order flow in TAQ. $TURN$ is daily turnover ratio defined as trading volume over the number of shares outstanding. $BASPRD$ is daily relative spreads measured as twice the distance between daily close offer and bid prices scaled by the quote midpoint. RET is daily mid-quote stock return, adjusted by Fama-French three factors. Panel A is for discriptive statistics. *Number of Dates* stands for the number of working days during sample period. *Avg Number of Stocks* is the average number of stocks at a day. Moreover, this table reports mean, standard deviation, minimum, median and maximum of each variable. Panal B is for autocorrelation of seven institutional order flows (IOFs). Panel C is for correlation of seven IOFs with each other and other variables.

(Continued)

Table I – *Continued*

Panel A. Descriptive Statistics

	Number of Dates	Avg Number of Stocks	Mean	Standard Deviation	Min	Med	Max
HH	3517	3230.970	0.037	0.212	-0.713	0.008	0.917
$CAMPBELL$	3517	3230.970	0.034	0.118	-0.300	0.009	0.561
LR	3517	3230.970	0.015	0.156	-0.570	0.002	0.667
$SMART$	3517	3230.970	0.006	0.025	-0.084	0.002	0.119
$DUMB$	3517	3230.970	0.028	0.142	-0.469	0.007	0.624
$ANOI$	3517	3230.970	0.003	0.099	-0.405	0.000	0.399
$TAQOI$	3517	3230.970	0.015	0.185	-0.708	0.003	0.801
$TURN$	3517	3230.970	0.692	0.836	0.003	0.440	5.146
$BASPRD$	3517	3230.970	0.009	0.013	0.000	0.004	0.126
RET	3517	3226.718	0.000	0.026	-0.121	-0.001	0.151

Panel B. Autocorrelation

	HH_t	$CAMPBELL_t$	LR_t	$SMART_t$	$DUMB_t$	$ANOI_t$	$TAQOI_t$
IOF_{t-1}	0.363	0.282	0.192	0.288	0.365	0.389	0.207
IOF_{t-2}	0.255	0.228	0.143	0.212	0.258	0.259	0.146
IOF_{t-3}	0.210	0.211	0.128	0.183	0.213	0.200	0.130
IOF_{t-4}	0.183	0.204	0.120	0.166	0.187	0.165	0.122
IOF_{t-5}	0.164	0.197	0.116	0.154	0.170	0.142	0.117

Panel C. Correlation

	HH	$CAMPBELL$	LR	$SMART$	$DUMB$	$ANOI$	$TAQOI$
HH	1.000						
$CAMPBELL$	0.556	1.000					
LR	0.405	0.734	1.000				
$SMART$	0.629	0.528	0.444	1.000			
$DUMB$	0.941	0.492	0.345	0.641	1.000		
$ANOI$	0.785	0.052	0.052	0.425	0.818	1.000	
$TAQOI$	0.430	0.769	0.945	0.475	0.361	0.057	1.000
$TURN$	0.202	0.348	0.151	0.250	0.222	0.006	0.156
$BASPRD$	-0.067	-0.108	-0.052	-0.068	-0.072	-0.008	-0.053
RET	0.190	0.229	0.232	0.174	0.174	0.087	0.295

Table II
Return predictability

This table presents estimated coefficients from Fama-MacBeth (1973) regression to measure return predictability of five different estimated institutional order flows (estimated IOFs),

$$\begin{aligned}\mathrm{RET}_{i,t} &= \alpha_t + \sum_{k=1}^{5} \beta_{t,k}\mathrm{IOF}_{i,t-k} + \sum_{k=1}^{5} \gamma_{t,k}\mathrm{TAQ}_{i,t-k} + \gamma_t^B \mathrm{BASPRD}_{i,t-1} + \gamma_t^T \mathrm{TURN}_{i,t-1} \\ &+ \sum_{k=1}^{5} \gamma_{t,k}^R \mathrm{RET}_{i,t-k} + \sum_{k=1}^{5} \gamma_{t,k}^{R2} \mathrm{RET}_{i,t-k}^2 + \epsilon_{i,t}\end{aligned}$$

, where for stock i on day t, α is a weekday dummy, RET is (mid-quote) stock return adjusted by Fama-French three factors, IOF is HH, $CAMPBELL$, LR, $SMART$, or $DUMB$. The sample period is from January 1999 to March 2012. The first row shows which IOF is utilized as independent variable in the regression model. Panel A reports regression results using daily risk-adjusted mid-quote return as a dependent variable. Panel B takes daily risk-adjusted return as a dependent variable. Panel C is for weekly regression results using risk-adjusted return as a dependent variable. In parentheses, we report t-statistics of the average coefficient over sample period based on New-West (1987) standard errors. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent lavel, respectively.

(Continued)

Table II – *Continued*

Panel A. Daily prediction model using risk-adjusted mid-quote return	HH	$CAMPBELL$	LR	$SMART$	$DUMB$
Mon	-0.008***	-0.007***	-0.008***	-0.008***	-0.008***
	(-3.90)	(-3.16)	(-3.83)	(-3.98)	(-3.79)
Tue	-0.000	0.001	-0.000	-0.000	-0.000
	(-0.16)	(0.44)	(-0.16)	(-0.22)	(-0.09)
Wed	-0.004**	-0.003*	-0.004**	-0.004**	-0.004*
	(-2.06)	(-1.68)	(-2.13)	(-2.04)	(-1.95)
Thu	-0.004*	-0.003	-0.004**	-0.004*	-0.004*
	(-1.90)	(-1.41)	(-2.01)	(-1.92)	(-1.84)
Fri	-0.010***	-0.009***	-0.010***	-0.009***	-0.009***
	(-5.42)	(-4.85)	(-5.65)	(-5.29)	(-5.28)
IOF_{t-1}	0.088***	-0.198***	-1.168***	1.392***	0.100***
	(9.68)	(-9.97)	(-14.05)	(21.62)	(7.89)
IOF_{t-2}	-0.031***	-0.146***	0.133***	-0.070	-0.050***
	(-3.99)	(-8.84)	(3.36)	(-1.34)	(-4.44)
IOF_{t-3}	-0.019**	-0.068***	0.048	-0.176***	-0.034***
	(-2.49)	(-4.06)	(1.20)	(-3.45)	(-2.87)
IOF_{t-4}	-0.032***	-0.052***	0.006	-0.277***	-0.050***
	(-4.42)	(-3.21)	(0.14)	(-5.06)	(-4.74)
IOF_{t-5}	-0.037***	-0.067***	0.058	-0.288***	-0.057***
	(-5.08)	(-4.00)	(1.46)	(-5.68)	(-5.46)
$TAQOI_{t-1}$	0.078***	0.210***	1.161***	0.027***	0.094***
	(7.90)	(15.67)	(14.99)	(2.67)	(9.77)
$TAQOI_{t-2}$	-0.071***	-0.009	-0.196***	-0.075***	-0.072***
	(-8.92)	(-0.90)	(-5.31)	(-9.41)	(-9.60)
$TAQOI_{t-3}$	-0.028***	0.000	-0.080**	-0.023***	-0.030***
	(-3.45)	(0.01)	(-2.13)	(-2.71)	(-3.71)
$TAQOI_{t-4}$	-0.005	0.010	-0.028	-0.001	-0.006
	(-0.61)	(0.86)	(-0.73)	(-0.09)	(-0.82)
$TAQOI_{t-5}$	-0.003	0.018*	-0.062*	0.001	-0.005
	(-0.34)	(1.70)	(-1.69)	(0.17)	(-0.58)
$BASPRD_{t-1}$	1.517***	1.436***	1.565***	1.503***	1.502***
	(7.46)	(7.11)	(7.70)	(7.40)	(7.39)
$TURN_{t-1}$	0.024***	0.041***	0.025***	0.023***	0.026***
	(4.84)	(7.60)	(5.03)	(4.55)	(5.11)
RET_{t-1}	-1.411***	-1.350***	-1.674***	-1.381***	-1.399***
	(-9.38)	(-9.04)	(-11.59)	(-9.21)	(-9.30)
RET_{t-2}	-1.180***	-1.161***	-1.130***	-1.184***	-1.171***
	(-10.20)	(-10.09)	(-9.75)	(-10.30)	(-10.11)
RET_{t-3}	-0.852***	-0.864***	-0.835***	-0.864***	-0.842***
	(-8.72)	(-8.85)	(-8.51)	(-8.82)	(-8.60)
RET_{t-4}	-0.649***	-0.678***	-0.667***	-0.670***	-0.643***
	(-6.29)	(-6.63)	(-6.50)	(-6.54)	(-6.23)
RET_{t-5}	-0.360***	-0.396***	-0.370***	-0.383***	-0.354***
	(-3.61)	(-3.99)	(-3.70)	(-3.84)	(-3.54)
RET^2_{t-1}	29.725***	28.266***	29.044***	29.218***	29.564***
	(18.58)	(18.15)	(18.40)	(18.42)	(18.49)
RET^2_{t-2}	4.611***	4.217***	4.438***	4.421***	4.555***
	(4.03)	(3.69)	(3.89)	(3.85)	(3.98)
RET^2_{t-3}	8.164***	8.068***	8.082***	8.360***	8.193***
	(7.45)	(7.37)	(7.46)	(7.64)	(7.47)
RET^2_{t-4}	7.438***	7.322***	7.345***	7.689***	7.529***
	(6.55)	(6.42)	(6.46)	(6.78)	(6.63)
RET^2_{t-5}	7.786***	7.529***	7.834***	8.171***	7.863***
	(6.65)	(6.42)	(6.71)	(7.00)	(6.71)
Adjusted R^2	0.034***	0.035***	0.036***	0.035***	0.034***
	(39.70)	(40.21)	(41.51)	(40.15)	(39.72)
Observation	11,143,554	11,143,554	11,143,554	11,143,554	11,143,554

(Continued)

Table II – *Continued*

Panel B. Daily prediction model using risk-adjusted return					
	HH	$CAMPBELL$	LR	$SMART$	$DUMB$
Mon	-0.013***	-0.011***	-0.013***	-0.013***	-0.013***
	(-5.75)	(-5.00)	(-5.64)	(-5.78)	(-5.63)
Tue	-0.006***	-0.005**	-0.006***	-0.006***	-0.006***
	(-2.80)	(-2.25)	(-2.77)	(-2.81)	(-2.72)
Wed	-0.009***	-0.008***	-0.009***	-0.009***	-0.009***
	(-4.52)	(-4.23)	(-4.55)	(-4.52)	(-4.40)
Thu	-0.009***	-0.008***	-0.009***	-0.009***	-0.008***
	(-3.97)	(-3.54)	(-4.08)	(-3.93)	(-3.90)
Fri	-0.013***	-0.012***	-0.013***	-0.012***	-0.012***
	(-6.27)	(-5.73)	(-6.46)	(-6.12)	(-6.15)
IOF_{t-1}	0.133***	-0.134***	-1.319***	1.178***	0.139***
	(12.67)	(-6.08)	(-12.73)	(15.51)	(9.19)
IOF_{t-2}	-0.047***	-0.166***	0.130**	-0.053	-0.068***
	(-5.39)	(-9.09)	(2.46)	(-0.87)	(-5.32)
IOF_{t-3}	-0.023***	-0.076***	-0.005	-0.157***	-0.035***
	(-2.69)	(-4.10)	(-0.09)	(-2.68)	(-2.67)
IOF_{t-4}	-0.037***	-0.053***	0.073	-0.328***	-0.061***
	(-4.54)	(-2.82)	(1.39)	(-5.21)	(-5.08)
IOF_{t-5}	-0.044***	-0.070***	0.088*	-0.318***	-0.069***
	(-5.21)	(-3.75)	(1.68)	(-5.39)	(-5.70)
$TAQOI_{t-1}$	0.098***	0.218***	1.339***	0.078***	0.123***
	(8.09)	(13.45)	(13.61)	(6.37)	(10.43)
$TAQOI_{t-2}$	-0.059***	0.005	-0.185***	-0.071***	-0.063***
	(-6.14)	(0.40)	(-3.81)	(-7.45)	(-7.04)
$TAQOI_{t-3}$	-0.027***	0.005	-0.032	-0.025**	-0.029***
	(-2.86)	(0.36)	(-0.66)	(-2.57)	(-3.15)
$TAQOI_{t-4}$	-0.017*	-0.004	-0.099**	-0.012	-0.019**
	(-1.79)	(-0.29)	(-2.04)	(-1.26)	(-2.05)
$TAQOI_{t-5}$	0.001	0.020	-0.091*	0.003	-0.001
	(0.08)	(1.55)	(-1.87)	(0.35)	(-0.14)
$BASPRD_{t-1}$	1.703***	1.635***	1.753***	1.693***	1.690***
	(7.30)	(7.03)	(7.53)	(7.27)	(7.26)
$TURN_{t-1}$	0.029***	0.045***	0.029***	0.030***	0.031***
	(5.01)	(7.28)	(5.27)	(5.20)	(5.33)
RET_{t-1}	-4.358***	-4.283***	-4.626***	-4.313***	-4.345***
	(-28.05)	(-27.80)	(-30.04)	(-27.93)	(-27.98)
RET_{t-2}	-1.504***	-1.479***	-1.469***	-1.500***	-1.497***
	(-12.56)	(-12.42)	(-12.27)	(-12.61)	(-12.48)
RET_{t-3}	-0.950***	-0.962***	-0.948***	-0.958***	-0.941***
	(-9.35)	(-9.48)	(-9.29)	(-9.43)	(-9.25)
RET_{t-4}	-0.617***	-0.647***	-0.615***	-0.635***	-0.611***
	(-5.86)	(-6.20)	(-5.87)	(-6.06)	(-5.80)
RET_{t-5}	-0.379***	-0.415***	-0.379***	-0.397***	-0.371***
	(-3.55)	(-3.92)	(-3.56)	(-3.74)	(-3.48)
RET^2_{t-1}	33.947***	32.814***	33.071***	33.505***	33.810***
	(19.20)	(18.81)	(19.16)	(19.05)	(19.13)
RET^2_{t-2}	5.737***	5.443***	5.552***	5.545***	5.671***
	(4.92)	(4.67)	(4.79)	(4.76)	(4.87)
RET^2_{t-3}	7.738***	7.695***	7.497***	7.856***	7.732***
	(7.15)	(7.14)	(6.99)	(7.30)	(7.15)
RET^2_{t-4}	7.348***	7.248***	7.157***	7.580***	7.418***
	(6.15)	(6.04)	(6.01)	(6.33)	(6.21)
RET^2_{t-5}	9.441***	9.299***	9.508***	9.754***	9.501***
	(7.33)	(7.20)	(7.45)	(7.55)	(7.36)
Adjusted R^2	0.033***	0.033***	0.035***	0.033***	0.033***
	(40.98)	(41.42)	(44.23)	(41.53)	(41.00)
Observation	11,362,415	11,362,415	11,362,415	11,362,415	11,362,415

(Continued)

Table II – *Continued*

Panel C. Weekly prediction model using risk-adjusted return

	HH	$CAMPBELL$	LR	$SMART$	$DUMB$
Intercept	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.70)	(-0.16)	(-1.01)	(-0.63)	(-0.58)
IOF_{t-1}	0.000	-0.001***	-0.003***	0.003***	-0.000
	(0.49)	(-3.64)	(-2.68)	(3.62)	(-1.02)
IOF_{t-2}	-0.000***	-0.001***	-0.000	-0.004***	-0.001***
	(-4.54)	(-3.91)	(-0.03)	(-5.00)	(-4.84)
IOF_{t-3}	-0.000***	-0.001***	-0.000	-0.004***	-0.001***
	(-4.17)	(-3.59)	(-0.10)	(-4.20)	(-4.98)
IOF_{t-4}	-0.000***	-0.001***	0.001	-0.002***	-0.000***
	(-3.12)	(-4.58)	(0.70)	(-2.66)	(-3.29)
$TAQOI_{t-1}$	-0.000**	0.000	0.002**	-0.001***	-0.000**
	(-2.32)	(1.08)	(2.25)	(-3.74)	(-1.98)
$TAQOI_{t-2}$	0.000	0.000**	-0.000	0.000	0.000
	(0.30)	(2.20)	(-0.21)	(0.97)	(0.17)
$TAQOI_{t-3}$	-0.000**	0.000	-0.000	-0.000	-0.000**
	(-2.29)	(0.23)	(-0.60)	(-1.57)	(-2.39)
$TAQOI_{t-4}$	-0.000**	0.000	-0.001	-0.000**	-0.000**
	(-2.40)	(0.78)	(-1.42)	(-2.35)	(-2.54)
$BASPRD_{t-1}$	0.042***	0.040***	0.043***	0.041***	0.041***
	(4.53)	(4.37)	(4.70)	(4.49)	(4.47)
$TURN_{t-1}$	-0.000	0.000	-0.000	0.000	0.000
	(-0.19)	(1.26)	(-1.21)	(0.08)	(0.10)
RET_{t-1}	-0.031***	-0.031***	-0.032***	-0.031***	-0.031***
	(-10.70)	(-10.81)	(-10.72)	(-10.80)	(-10.63)
RET_{t-2}	-0.016***	-0.017***	-0.017***	-0.017***	-0.016***
	(-6.01)	(-6.30)	(-6.19)	(-6.14)	(-5.97)
RET_{t-3}	-0.012***	-0.013***	-0.013***	-0.013***	-0.012***
	(-5.05)	(-5.46)	(-5.43)	(-5.16)	(-5.00)
RET_{t-4}	-0.007***	-0.007***	-0.006***	-0.007***	-0.006***
	(-2.64)	(-2.92)	(-2.66)	(-2.70)	(-2.60)
RET^2_{t-1}	0.235***	0.227***	0.235***	0.233***	0.234***
	(15.35)	(14.85)	(15.37)	(15.24)	(15.30)
RET^2_{t-2}	0.146***	0.143***	0.145***	0.149***	0.145***
	(9.25)	(9.16)	(9.26)	(9.42)	(9.25)
RET^2_{t-3}	0.100***	0.098***	0.098***	0.103***	0.101***
	(6.58)	(6.44)	(6.56)	(6.78)	(6.62)
RET^2_{t-4}	0.100***	0.101***	0.098***	0.104***	0.101***
	(5.88)	(5.97)	(5.79)	(6.05)	(5.91)
Adjusted R^2	0.032***	0.032***	0.033***	0.032***	0.032***
	(20.75)	(21.20)	(21.02)	(21.05)	(20.84)
Observation	2,236,814	2,236,814	2,236,814	2,236,814	2,236,814

Table III
Investment simulation

This table documents the performance of investment strategies based on estimated institutional order flows (IOFs). We rank all the stocks in our sample by one-trading-day lagged LFOIs for each day, and classify them into decile portfolios. Stocks with the lowest (highest) IOF belong to *Low (High)* portfolio. We take short positions for stocks in the *Low* portfolio and long position for stocks in the *High* portfolio at day t. The sample period is from January 1999 to March 2012. The first row indicates the type of IOF which we utilize in our investment strategy. The first to tenth rows (*Low* to *High*) report portfolio return in each decile IOF portfolio. The eleventh row, HML, reports the performance and t-statistics of our investment strategies, short *Low* and long *High*. The thirteenth row, *FF3 Alpha*, reports Fama-French three factor alpha of our investment strategy. In parentheses, we report t-statistics of *HML* and *FF3 Alpha* based on New-West (1987) standard errors.

	HH	$CAMPBELL$	LR	$SMART$	$DUMB$
Low (%)	0.116	0.182	0.182	0.153	0.126
2	0.092	0.142	0.122	0.124	0.097
3	0.126	0.146	0.105	0.117	0.132
4	0.121	0.089	0.095	0.044	0.094
5	0.018	0.018	0.070	0.004	0.004
6	0.030	0.020	0.040	0.035	0.042
7	0.070	0.051	0.034	0.060	0.080
8	0.096	0.073	0.062	0.088	0.104
9	0.126	0.106	0.093	0.126	0.129
High	0.173	0.144	0.166	0.218	0.161
HML (%)	0.057	-0.038	-0.015	0.064	0.035
	(5.59)	(-2.93)	(-1.12)	(7.62)	(3.69)
FF3 Alpha (x1000)	0.068	-0.032	-0.011	0.075	0.044
	(7.78)	(-2.95)	(-1.04)	(10.06)	(5.45)

Table IV
Return predictability in subsamples

This table presents estimated coefficients from the regression model in Panel A of Table II in order to examine predictive power of estimated IOFs in subsamples based on firm characteristics. For brevity, we only report the coefficient estimates of the estimated IOF. The sample period is from January 1999 to March 2012. In Panel A, we divide the sample into small, middle, and large stock groups based on market capitalization everyday. In Panel B, we divide the sample into liquid, less liquid, and illiquid stock groups based on relative bid-ask spread everyday. In Panel C, we compare stocks listed on NYSE and AMEX versus Nasdaq. In Panel D, we divide the sample into an early (1999-2002), a middle (2003-2007), and a late (2008-2012) subperiods everyday. Corresponding t-statistics based on New-West (1987) standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent lavel, respectively.

Panel A. Size

	Small-sized stocks			Middle-sized stocks			Large-sized stocks		
	HH	$SMART$	$DUMB$	HH	$SMART$	$DUMB$	HH	$SMART$	$DUMB$
IOF_{t-1}	0.172***	1.784***	0.175***	0.081***	0.758***	0.031	0.084***	3.030***	0.013
	(10.79)	(13.45)	(6.81)	(4.70)	(7.38)	(1.22)	(2.99)	(9.22)	(0.48)
IOF_{t-2}	-0.073***	-0.324***	-0.119***	-0.012	0.066	-0.029	0.035	0.380	0.020
	(-4.87)	(-2.85)	(-5.08)	(-0.82)	(0.72)	(-1.24)	(1.27)	(1.22)	(0.73)
IOF_{t-3}	-0.053***	-0.438***	-0.076***	-0.016	0.048	-0.025	-0.030	0.255	-0.037
	(-3.66)	(-3.95)	(-3.31)	(-1.05)	(0.53)	(-1.09)	(-1.04)	(0.79)	(-1.37)
IOF_{t-4}	-0.065***	-0.609***	-0.111***	-0.011	-0.003	-0.008	-0.026	-0.128	-0.006
	(-4.48)	(-5.75)	(-4.82)	(-0.77)	(-0.04)	(-0.36)	(-1.02)	(-0.39)	(-0.26)
IOF_{t-5}	-0.057***	-0.605***	-0.107***	-0.033**	0.008	-0.037*	-0.016	0.165	-0.021
	(-4.22)	(-5.96)	(-5.04)	(-2.44)	(0.09)	(-1.77)	(-0.63)	(0.50)	(-0.87)
Adjusted R^2	0.038***	0.038***	0.038***	0.081***	0.081***	0.081***	0.134***	0.134***	0.133***
	(47.61)	(47.47)	(47.67)	(50.77)	(51.17)	(50.99)	(48.71)	(48.44)	(48.51)
Observation	4,937,678	4,937,678	4,937,678	1,553,935	1,553,935	1,553,935	1,100,451	1,100,451	1,100,451

(Continued)

Table IV – *Continued*

Panel B. Liquidity

	Liquid stocks			Less liquid stocks			Illiquid stocks		
	HH	*SMART*	*DUMB*	*HH*	*SMART*	*DUMB*	*HH*	*SMART*	*DUMB*
IOF_{t-1}	0.052***	0.849***	0.019	0.118***	1.541***	0.130***	0.144***	0.921***	0.114**
	(3.72)	(7.29)	(1.11)	(7.85)	(13.93)	(6.16)	(4.74)	(3.99)	(2.21)
IOF_{t-2}	-0.015	0.119	-0.025	-0.046***	0.021	-0.061***	-0.021	0.268	-0.022
	(-1.14)	(1.17)	(-1.43)	(-3.03)	(0.19)	(-2.66)	(-0.77)	(1.36)	(-0.54)
IOF_{t-3}	-0.009	-0.032	-0.011	-0.017	-0.281***	-0.033	-0.073***	-0.011	-0.078*
	(-0.66)	(-0.32)	(-0.62)	(-1.14)	(-2.74)	(-1.53)	(-2.60)	(-0.05)	(-1.75)
IOF_{t-4}	-0.015	-0.138	-0.033*	-0.049***	-0.325***	-0.065***	-0.022	-0.030	-0.043
	(-1.09)	(-1.44)	(-1.92)	(-3.52)	(-3.21)	(-3.27)	(-0.70)	(-0.14)	(-0.87)
IOF_{t-5}	-0.034**	-0.148	-0.036**	-0.024*	-0.248**	-0.048**	-0.069***	-0.531***	-0.122***
	(-2.55)	(-1.49)	(-2.06)	(-1.80)	(-2.55)	(-2.39)	(-2.75)	(-2.76)	(-3.17)
Adjusted R^2	0.077***	0.078***	0.077***	0.057***	0.058***	0.057***	0.053***	0.054***	0.053***
	(48.93)	(49.74)	(48.93)	(53.72)	(54.40)	(53.78)	(60.45)	(60.60)	(60.43)
Observation	2,205,849	2,205,849	2,205,849	2,227,989	2,227,989	2,227,989	2,241,882	2,241,882	2,241,882

Panel C. Subperiod

	Early period from 1999 to 2002			Middle period from 2003 to 2007			Late period from 2008 to 2012		
	HH	*SMART*	*DUMB*	*HH*	*SMART*	*DUMB*	*HH*	*SMART*	*DUMB*
IOF_{t-1}	0.108***	1.615***	0.084***	0.106***	1.668***	0.137***	0.055***	0.941***	0.075***
	(5.63)	(12.25)	(3.01)	(13.33)	(20.55)	(11.91)	(2.99)	(8.42)	(3.05)
IOF_{t-2}	-0.018	0.302***	-0.026	-0.015**	-0.102	-0.031***	-0.057***	-0.331***	-0.089***
	(-1.03)	(2.82)	(-1.11)	(-2.00)	(-1.44)	(-2.70)	(-3.87)	(-3.87)	(-3.98)
IOF_{t-3}	0.004	0.098	0.013	-0.010	-0.158**	-0.015	-0.047***	-0.413***	-0.089***
	(0.21)	(0.86)	(0.54)	(-1.37)	(-2.46)	(-1.46)	(-3.02)	(-4.83)	(-3.65)
IOF_{t-4}	-0.033*	-0.326***	-0.050**	-0.030***	-0.201**	-0.046***	-0.031**	-0.313***	-0.054***
	(-1.95)	(-2.64)	(-2.13)	(-3.87)	(-2.56)	(-3.91)	(-2.52)	(-3.63)	(-2.75)
IOF_{t-5}	-0.060***	-0.319***	-0.087***	-0.018**	-0.119*	-0.031***	-0.037***	-0.432***	-0.060***
	(-4.04)	(-3.11)	(-4.07)	(-2.25)	(-1.65)	(-2.68)	(-2.63)	(-4.90)	(-2.92)
Adjusted R^2	0.036***	0.036***	0.036***	0.028***	0.028***	0.028***	0.040***	0.040***	0.040***
	(23.36)	(23.62)	(23.36)	(25.04)	(25.62)	(25.11)	(24.75)	(24.87)	(24.80)
Observation	3,643,928	3,643,928	3,643,928	3,954,754	3,954,754	3,954,754	3,544,872	3,544,872	3,544,872

(Continued)

Table IV – *Continued*

Panel D. Exchange market

	NYSE and AMEX			NASDAQ		
	HH	$SMART$	$DUMB$	HH	$SMART$	$DUMB$
IOF_{t-1}	0.056***	1.351***	0.053***	0.122***	1.569***	0.152***
	(5.48)	(16.01)	(3.97)	(10.05)	(19.43)	(9.05)
IOF_{t-2}	-0.008	0.041	-0.022	-0.046***	-0.120*	-0.070***
	(-0.77)	(0.55)	(-1.55)	(-4.13)	(-1.66)	(-4.42)
IOF_{t-3}	-0.020**	-0.064	-0.031**	-0.020*	-0.234***	-0.035**
	(-2.07)	(-0.94)	(-2.23)	(-1.79)	(-3.25)	(-2.16)
IOF_{t-4}	-0.027***	-0.245***	-0.037***	-0.029***	-0.279***	-0.055***
	(-2.86)	(-3.40)	(-2.68)	(-2.93)	(-3.91)	(-3.70)
IOF_{t-5}	-0.011	-0.117*	-0.025*	-0.052***	-0.354***	-0.076***
	(-1.15)	(-1.72)	(-1.87)	(-5.68)	(-5.29)	(-5.49)
Adjusted R^2	0.056***	0.057***	0.056***	0.036***	0.036***	0.036***
	(40.20)	(40.67)	(40.17)	(47.05)	(47.27)	(47.07)
Observation	4,826,579	4,826,579	4,826,579	6,309,105	6,309,105	6,309,105

Table V
Reaction of IOF to mispricing

This table presents Fama-MacBeth (1973) regression results for the following equation between 1999 and 2012,

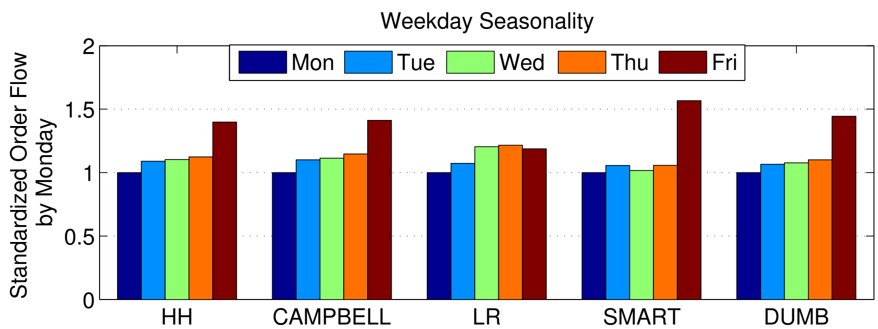
$$\text{MISPRICING}_{i,m} = \alpha_m + \beta_m \text{IOF}_{i,t,t+k,m} + \epsilon_m$$

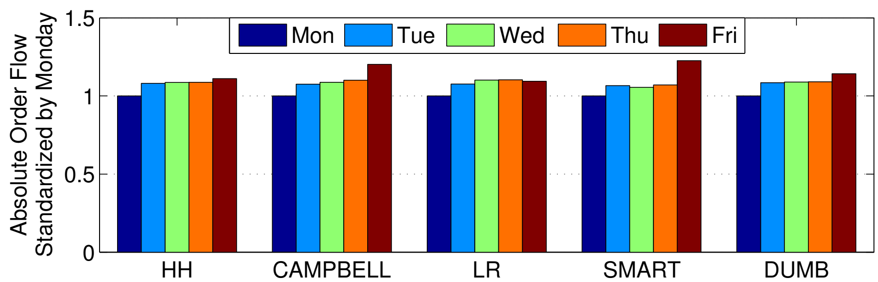
, where for stocks i on month m, MISPRICING is mispricing index suggested by Stambaugh, Yu, and Yuan (2012, 2015), and $\text{IOF}_{t,t+k}$ is an cumulative IOF from the end of month t to $t+k$. Corresponding t-statistics based on New-West (1987) standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent lavel, respectively.

	HH	$CAMPBELL$	LR	$SMART$	$DUMB$
$IOF_{t,t}$	2.075**	0.765	2.472***	0.220***	0.833
	(2.51)	(1.10)	(3.11)	(3.65)	(1.61)
$IOF_{t,t+1}$	3.077	0.874	5.561**	0.337**	0.637
	(1.49)	(0.45)	(2.51)	(2.24)	(0.53)
$IOF_{t,t+5}$	4.967	2.044	11.310**	0.585*	0.321
	(1.16)	(0.49)	(2.40)	(1.87)	(0.13)

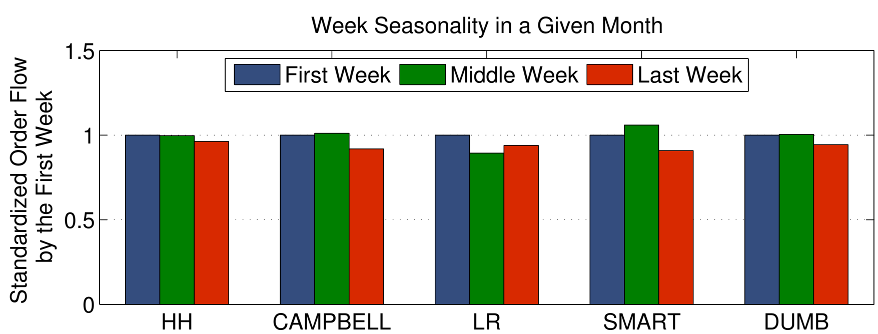
Figure 1. Seasonality This figure describes the seasonality of estimated institutional order flows. In Panel A, we take an average of weekday IOFs and scale it by IOF on Monday. In Panel B, we calculate weekly aggregated IOFs for the first, last, or other week in a given month, scaling it by IOF in the first week. In Panel C, we take an average of monthly IOFs and scale it by IOF on January.

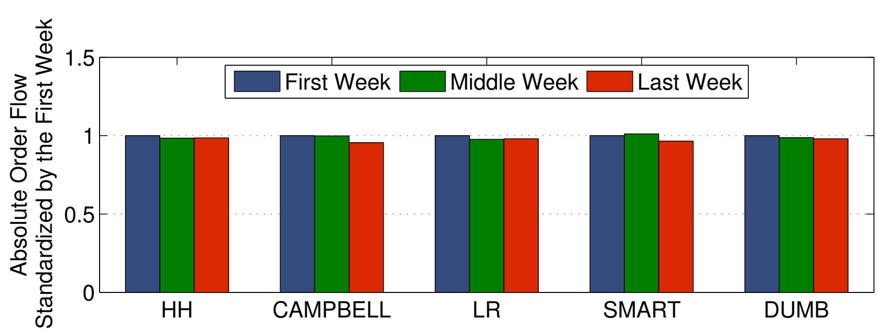
Panel A. Weekday seasonality





Panel B. Week seasonality





(Continued)

Figure 1 – *Continued*

Panel C. Month seasonality

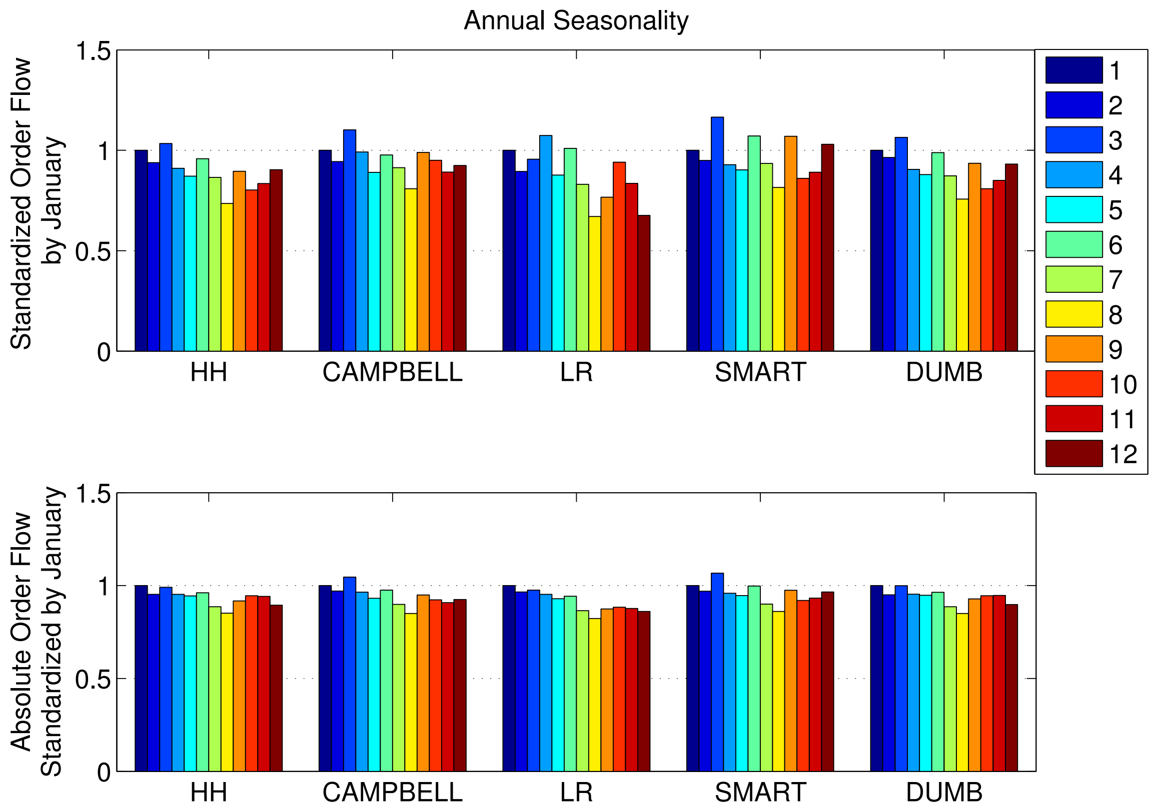


Figure 2. Performance of investment strategy based on estimated order flow This figure shows the log cumulative returns of investment strategies based on lagged IOFs. We sort all stocks in our sample into decile portfolios based on different IOFs each day, and calculate the return of a portfolio long stocks in the highest IOF decile and shrot stocks in the lowest IOF decile. The sample period is from January 1999 to March 2012.

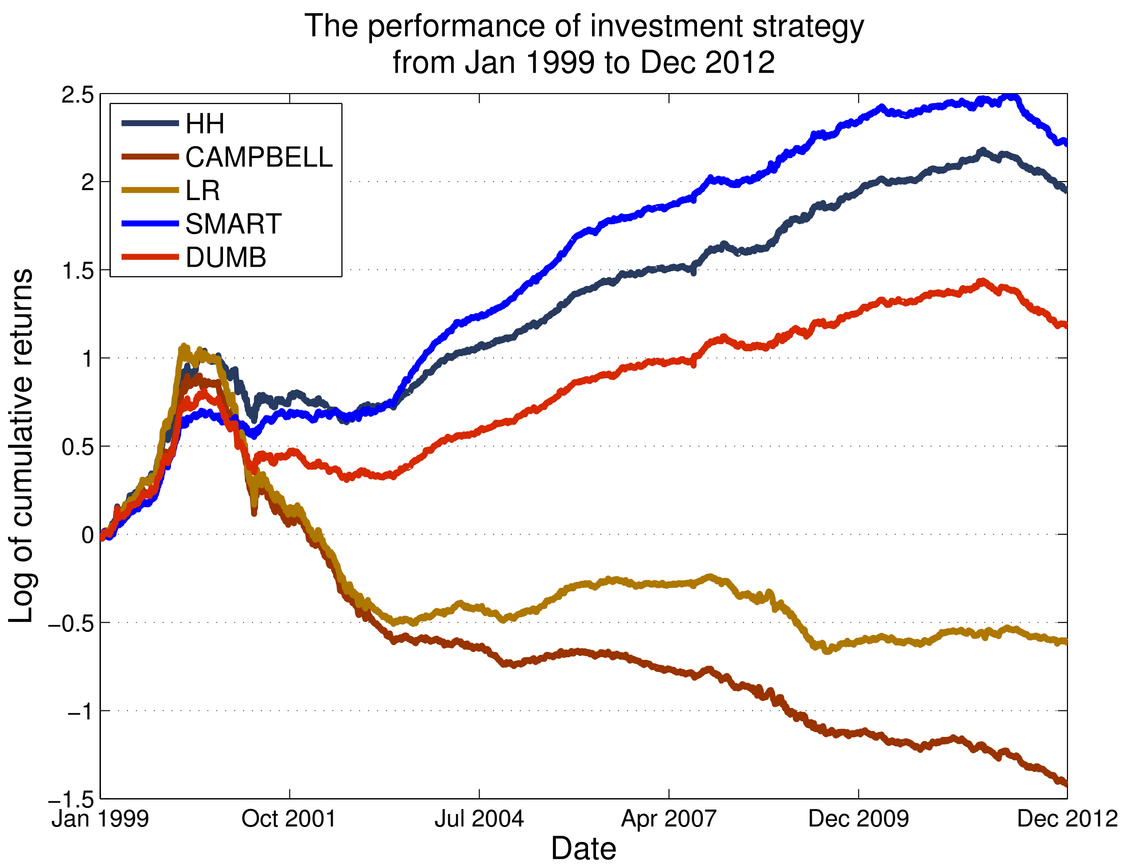
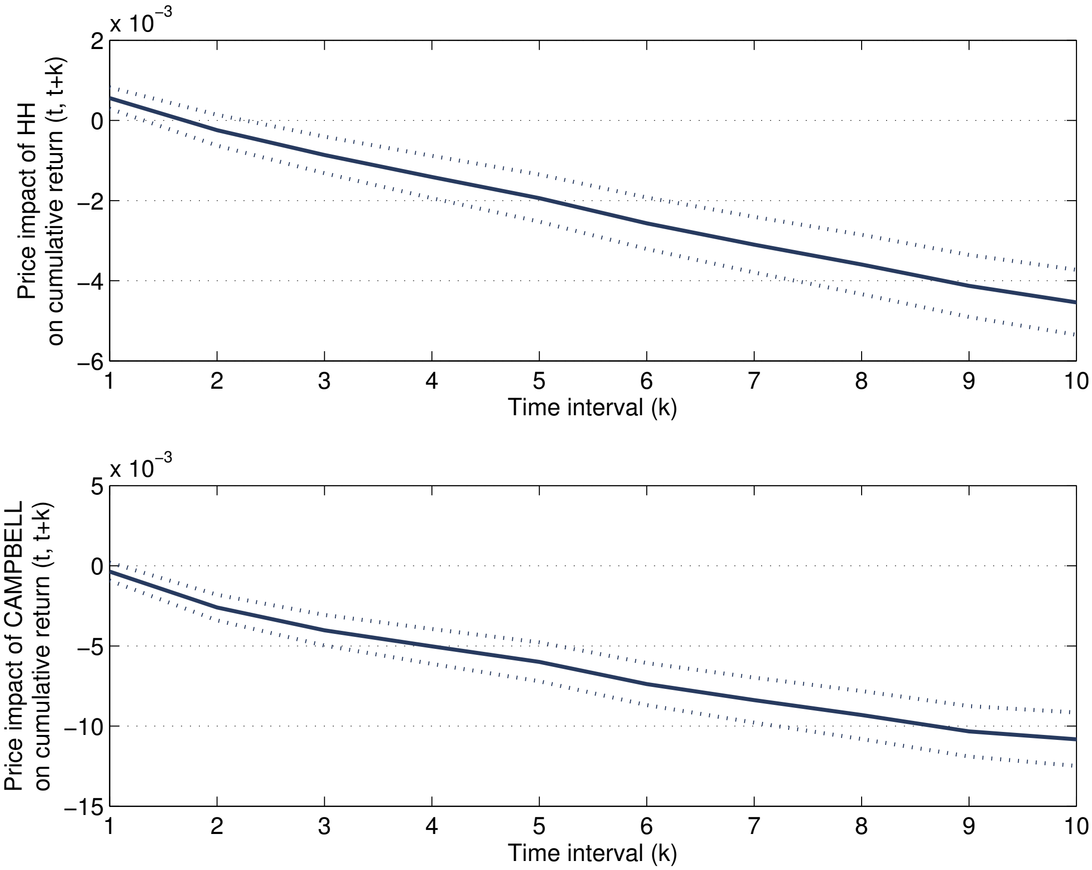


Figure 3. Price impact of institutional order flow at different horizons This figure plots the Fama-MacBeth (1973) coefficient estimates and 95% confidence intervals of lagged IOF from the following equation between 1999 and 2012,

$$\mathrm{CR}_{i,t,t+k} = \alpha_t + \beta_t \mathrm{IOF}_{i,t-1} + \epsilon_{i,t}$$

, where for stock i on day t, $\mathrm{CR}_{t,t+k}$ is the cumulative return from day t to $t+k$.



(Continued)

Figure 3 – *Continued*

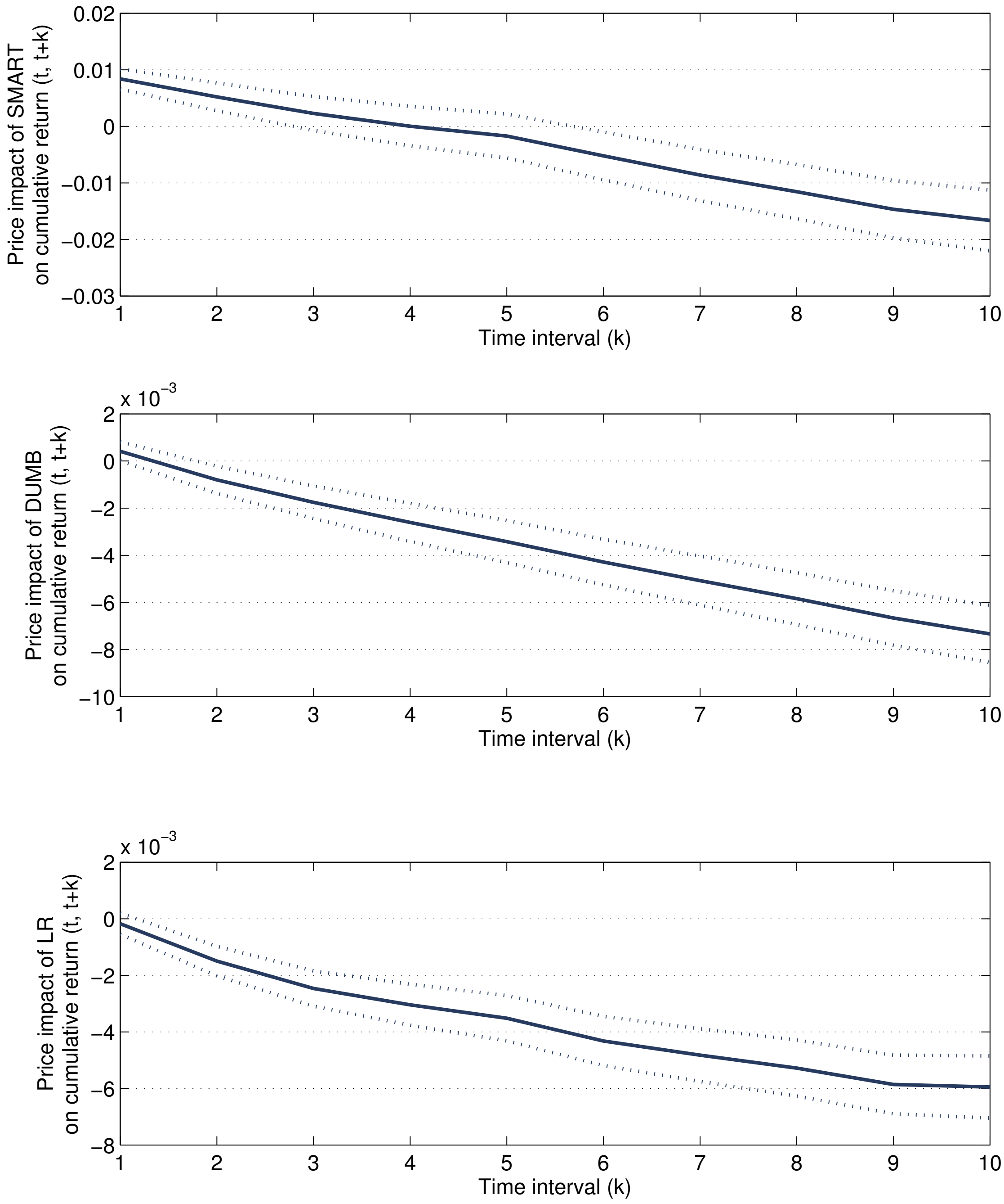
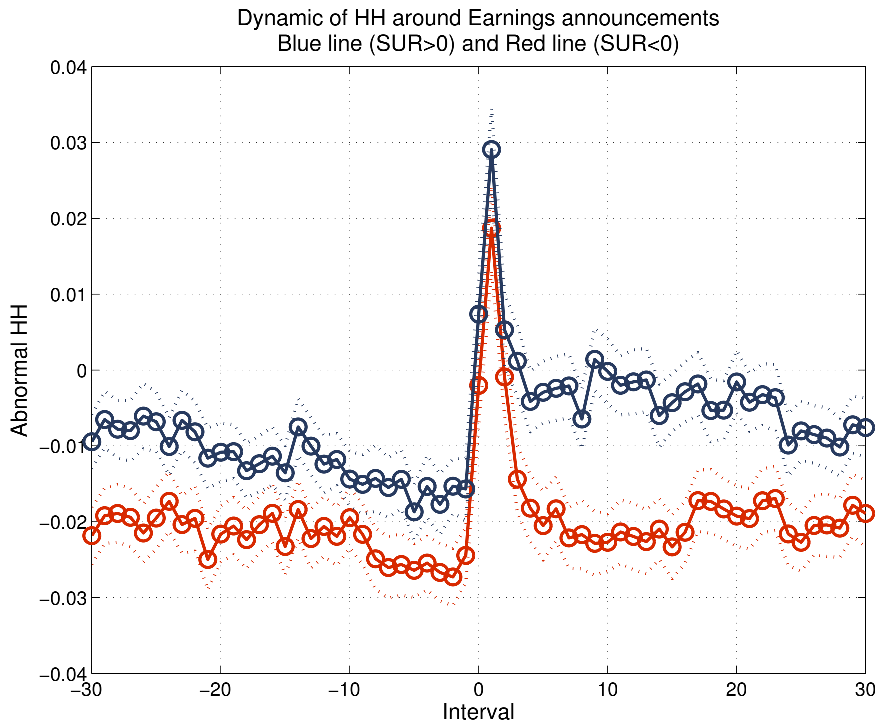
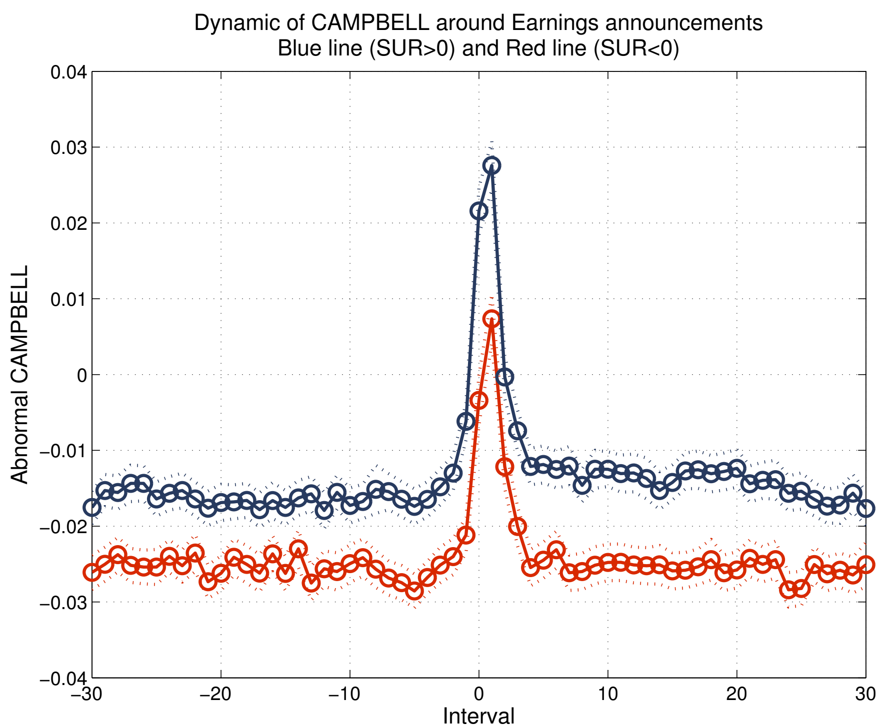


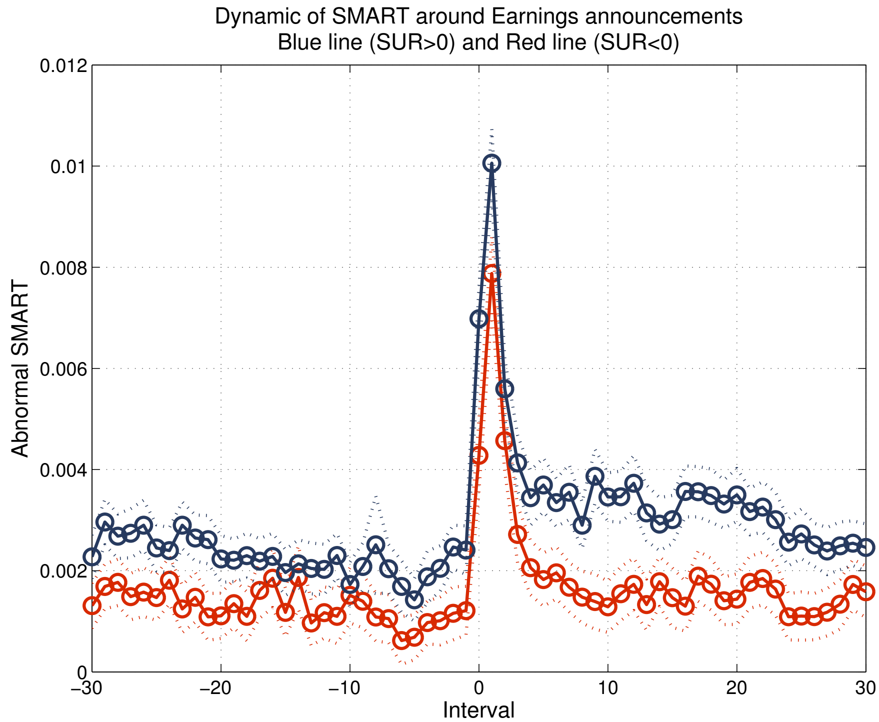
Figure 4. Dynamic of institutional order flow around earnings announcements This figure presents the pattern of institutional order flow (IOF) around quarterly earnings announcement. We plot abnormal IOF in excess of market average IOF on the same day from thirty days before to thirty days after an announcement. The earnings announcements are classified as positive (negative) if the scaled earnings surprise (SUR) is positive (negative). We define SUR as the difference between actual earnings and the average of earnings forecasts in analysts from the Institutional Brokers' Estimate System (IBES), scaled by stock price.

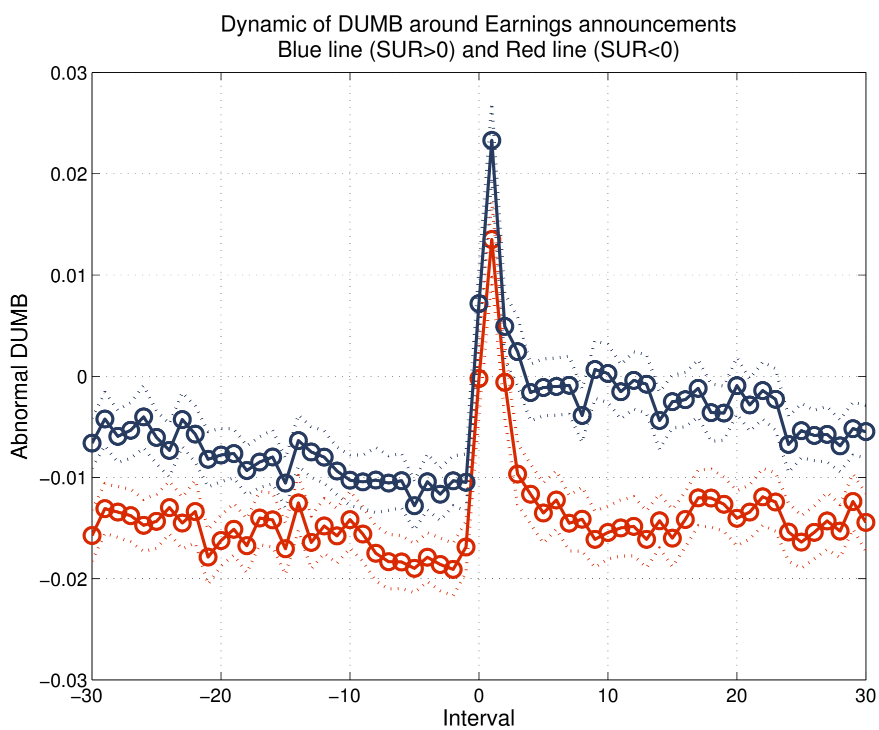




(Continued)

Figure 4 – *Continued*





(Continued)

Figure 4 – *Continued*

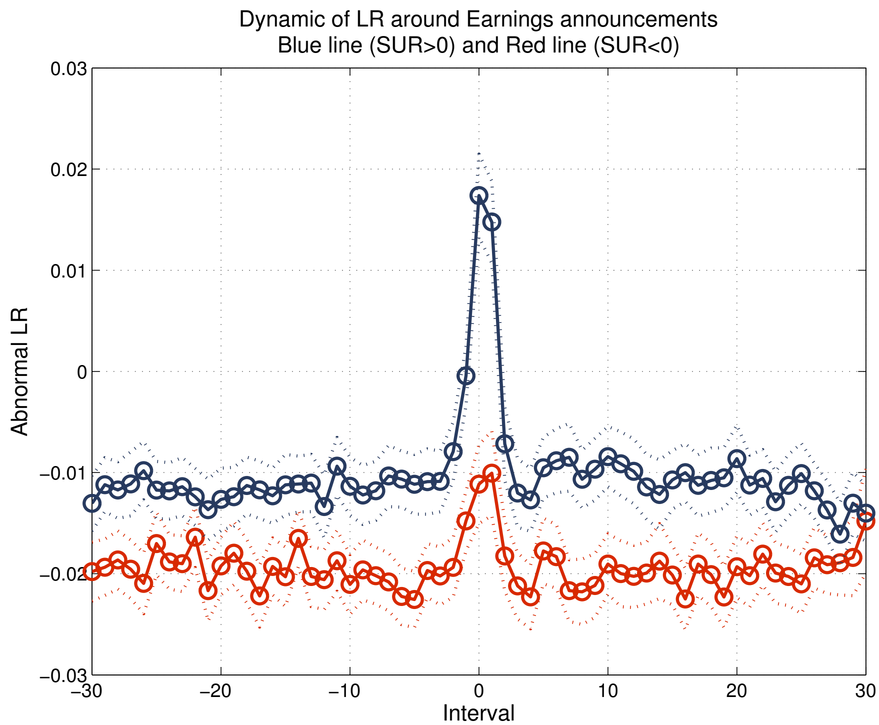
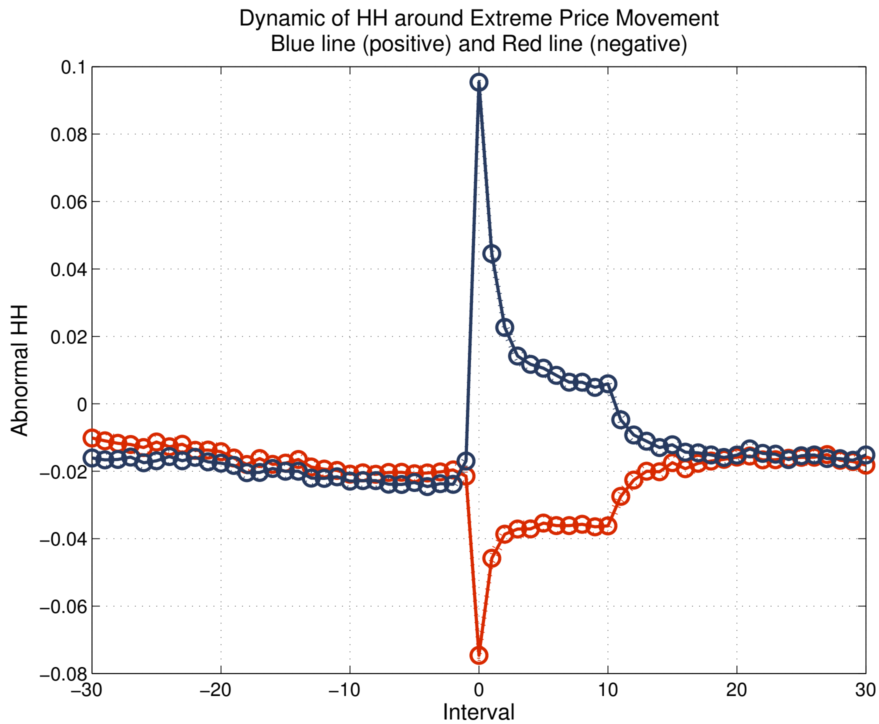
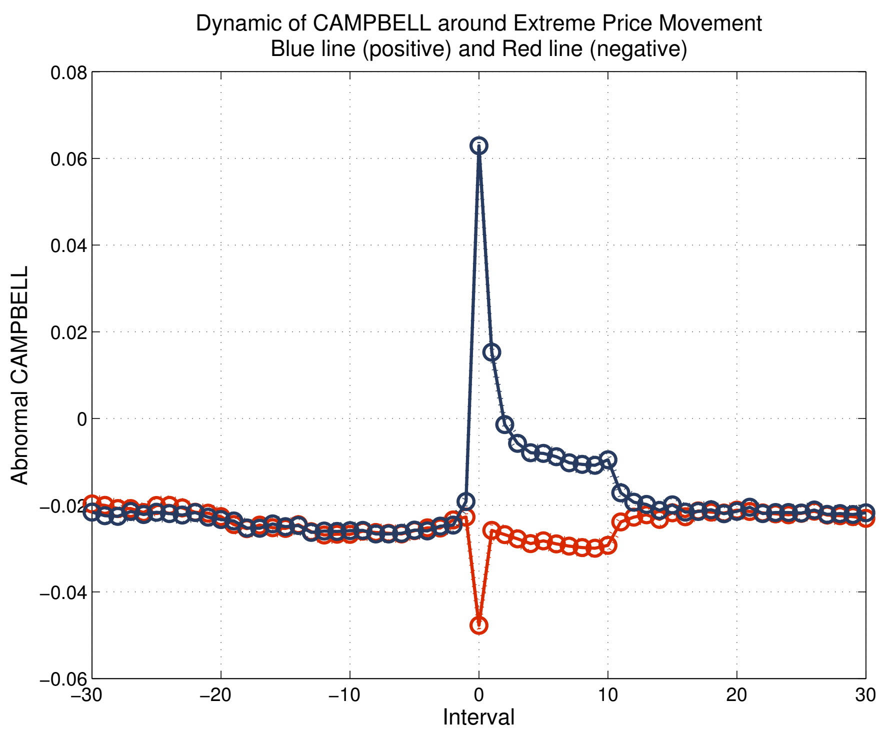


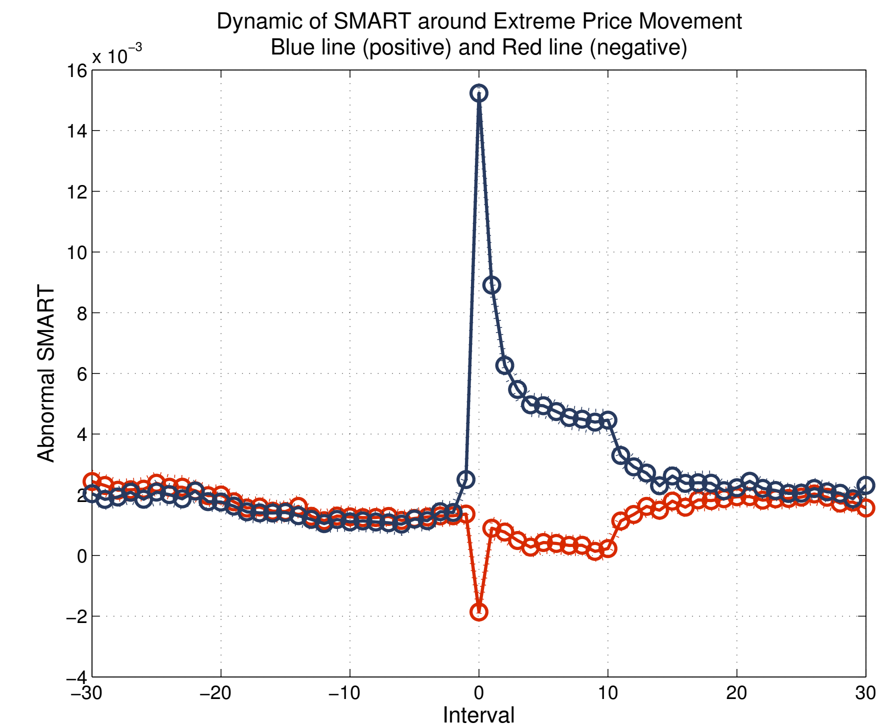
Figure 5. Dynamic of institutional order flow around extreme price movement This figure presents the pattern of institutional order flow (IOF) around extreme price movement. We plot abnormal IOF in excess of market average IOF on the same day from thirty days before to thirty days after an event. The extreme price movement is defined as those exceeding two standard deviation and not fully reversed during ten days afterward. We break the sample into two groups based on the direction of the price movement.

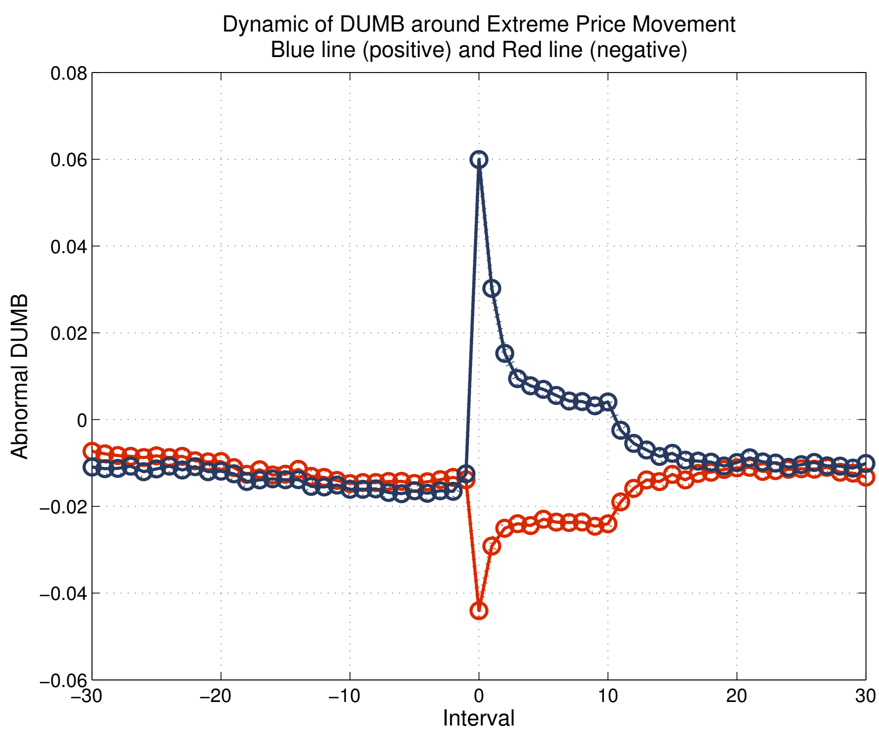




(Continued)

Figure 5 – *Continued*





(Continued)

Figure 5 – *Continued*

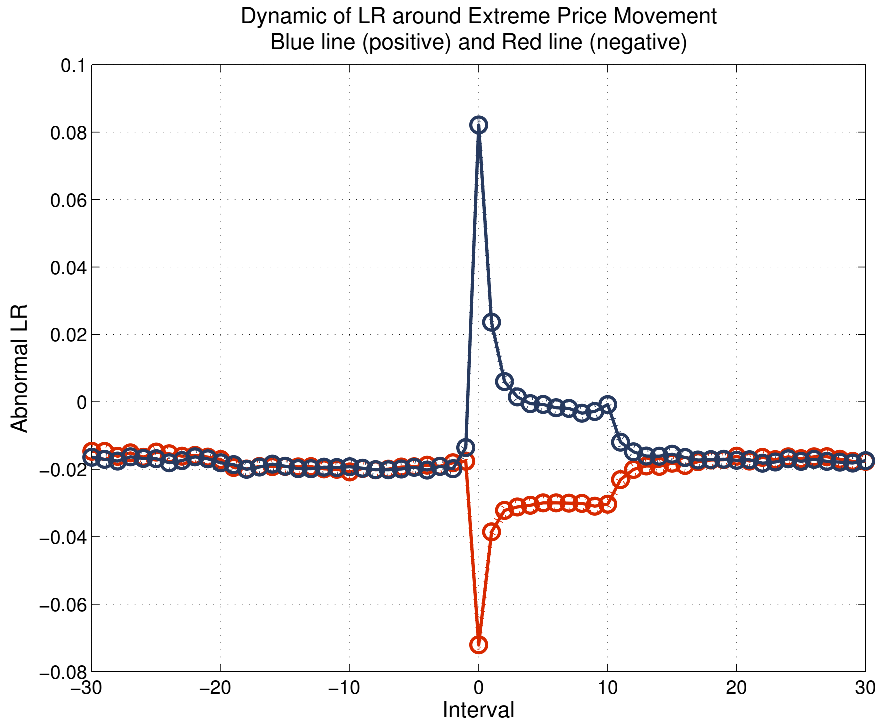
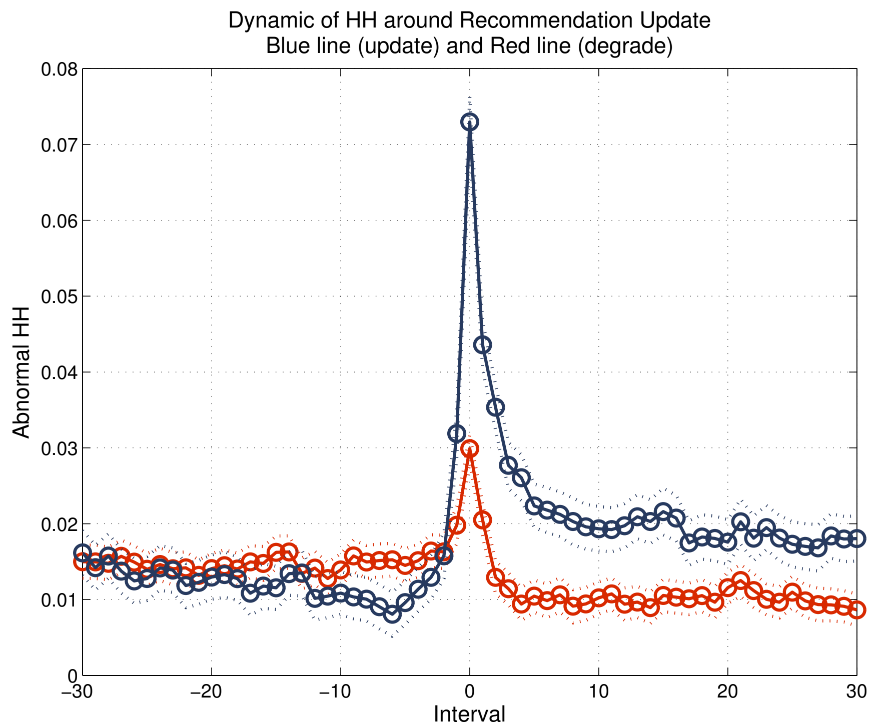
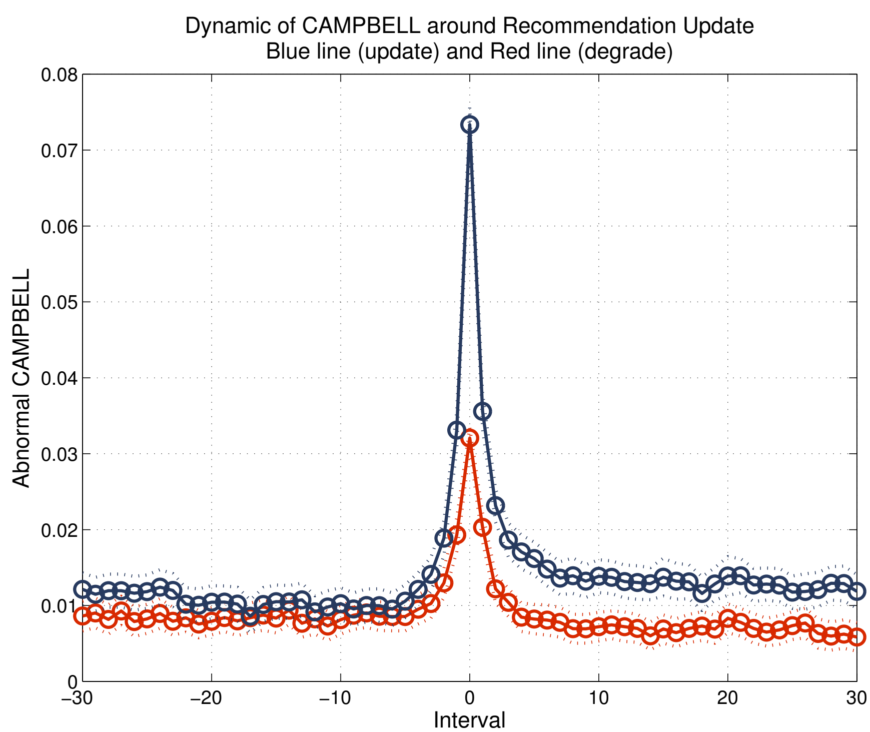


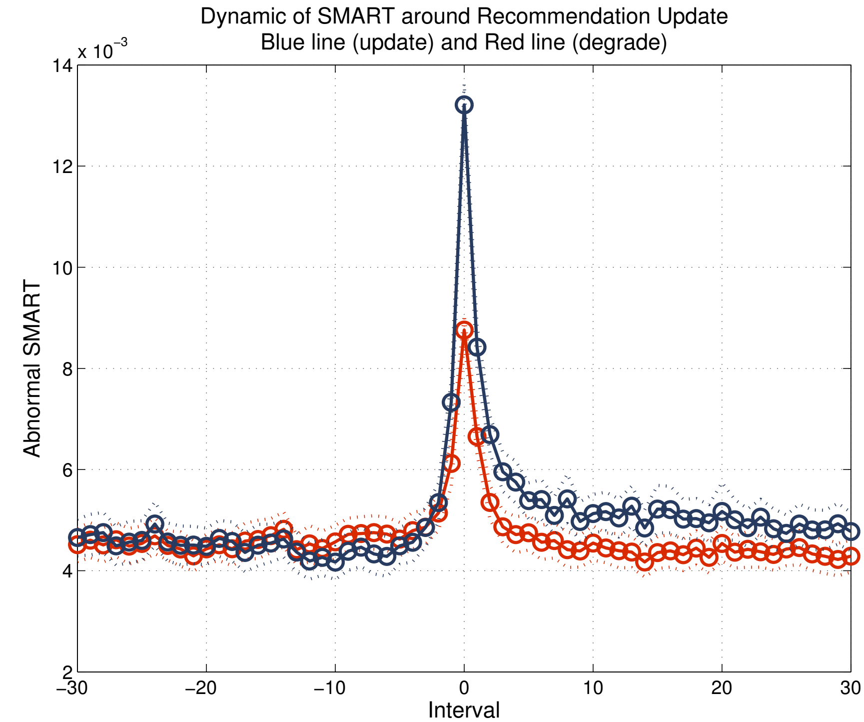
Figure 6. Dynamic of institutional order flow around recommendation updates This figure presents the pattern of institutional order flow (IOF) around analyst recommendation updates. We plot abnormal IOF in excess of market average IOF on the same day from thirty days before to thirty days after an announcement. We break the sample into two groups based on recommendation upgrade and downgrade.

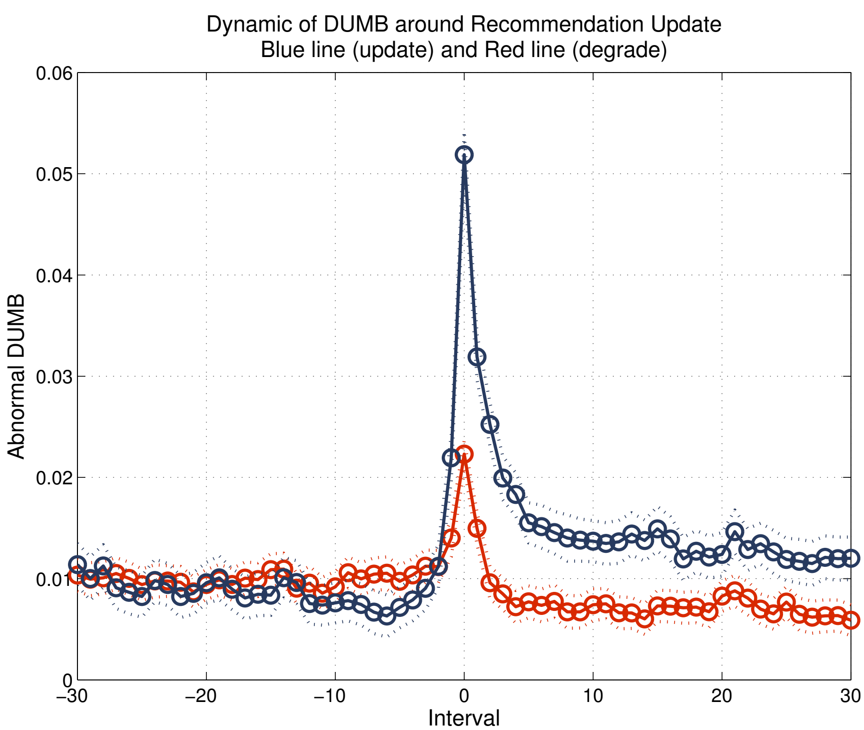




(Continued)

Figure 6 – *Continued*





(Continued)

Figure 6 – *Continued*

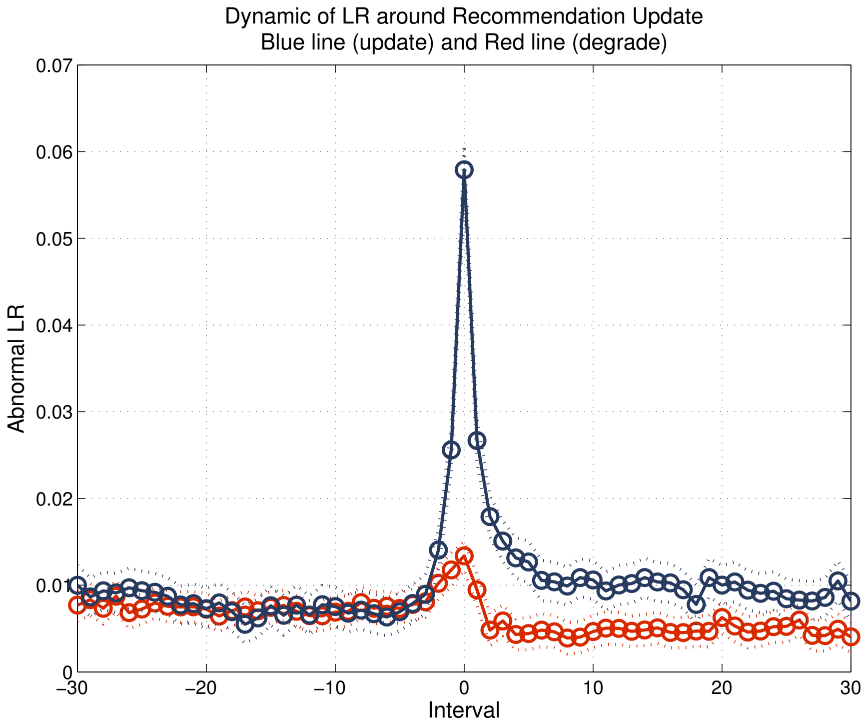
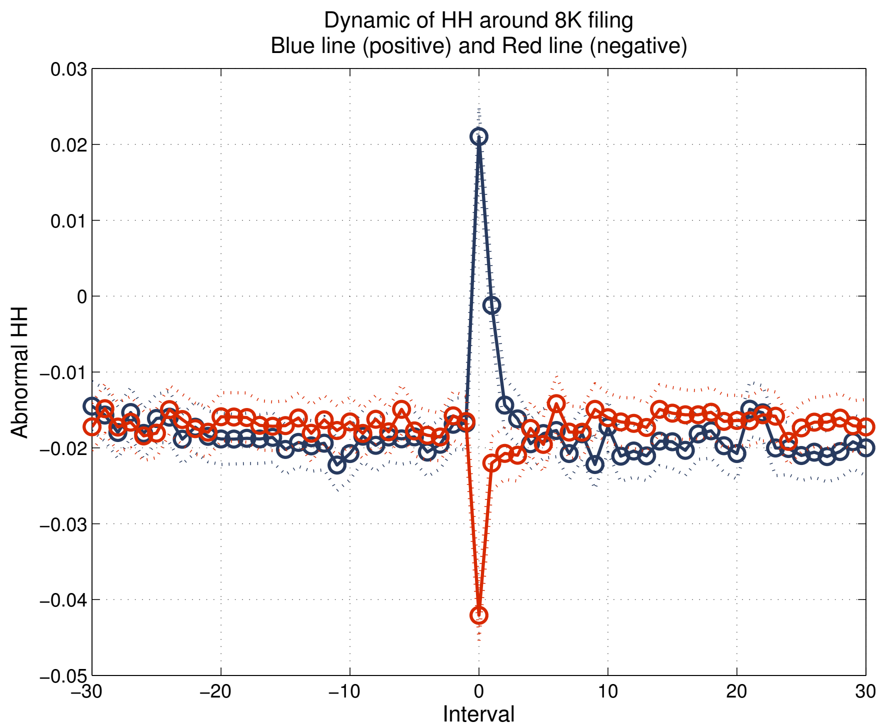
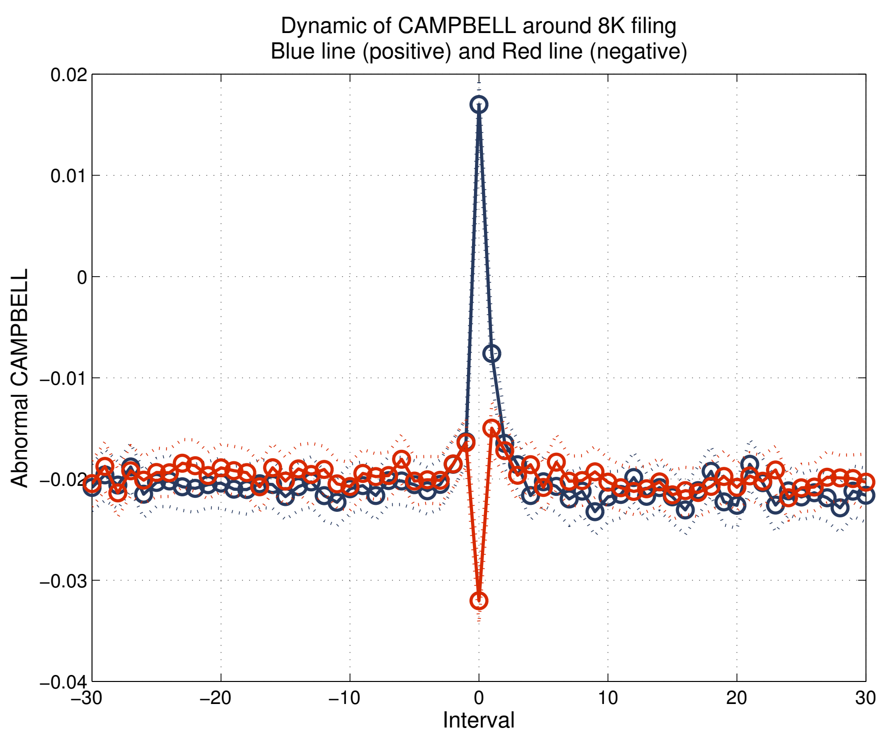


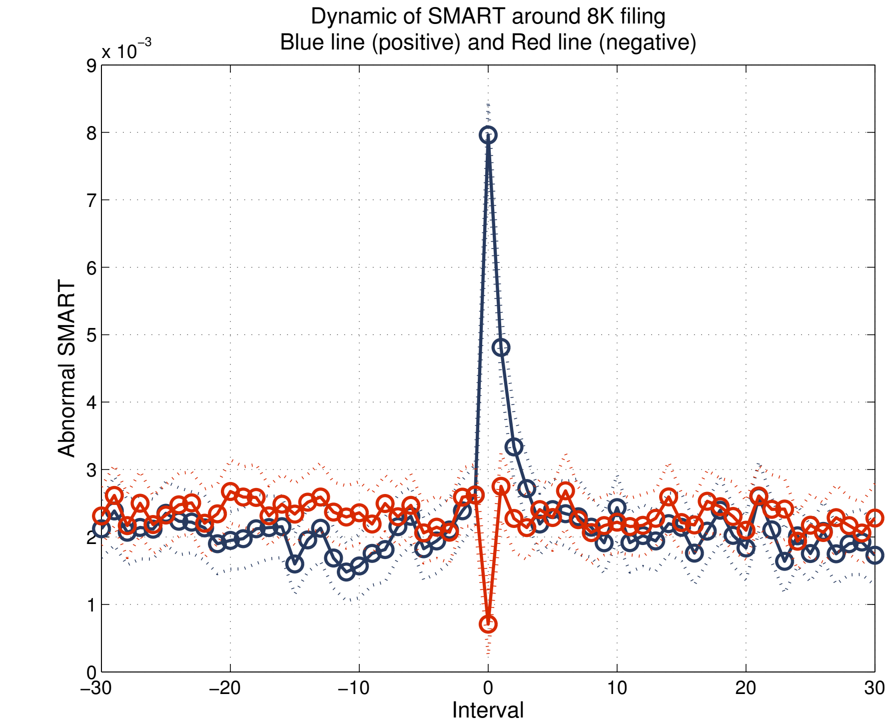
Figure 7. Dynamic of institutional order flow around value related 8K filing dates This figure presents the pattern of institutional order flow (IOF) around corporate 8K filings. We plot abnormal IOF in excess of market average IOF on the same day from thirty days before to thirty days after a filing. We classify 8K filings into positive (negative) ones based on abnormal stock return around the filing date.

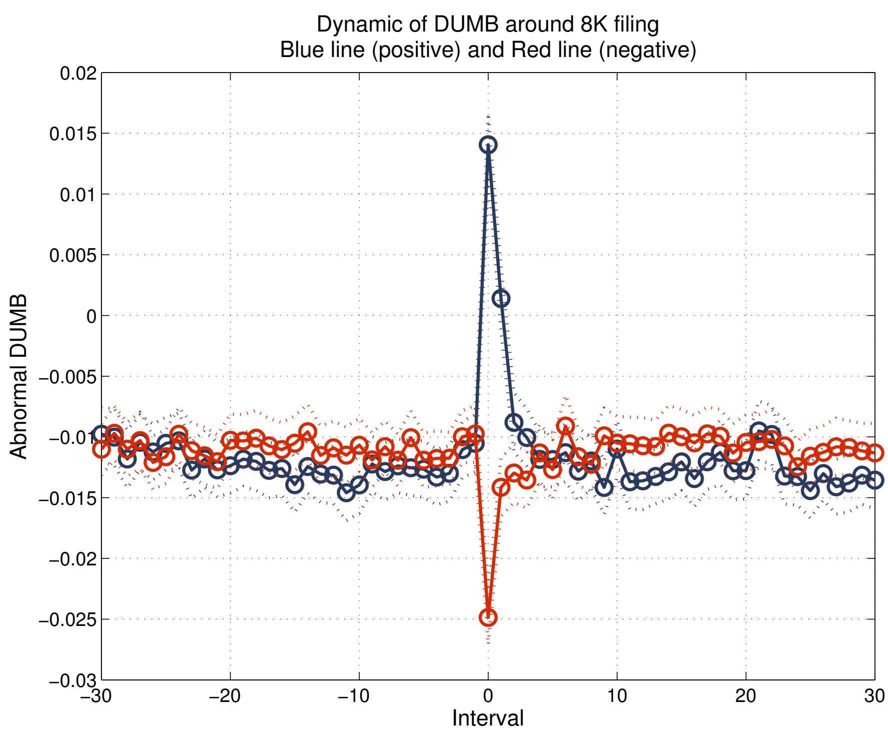




(Continued)

Figure 7 – *Continued*





(Continued)

Figure 7 – *Continued*

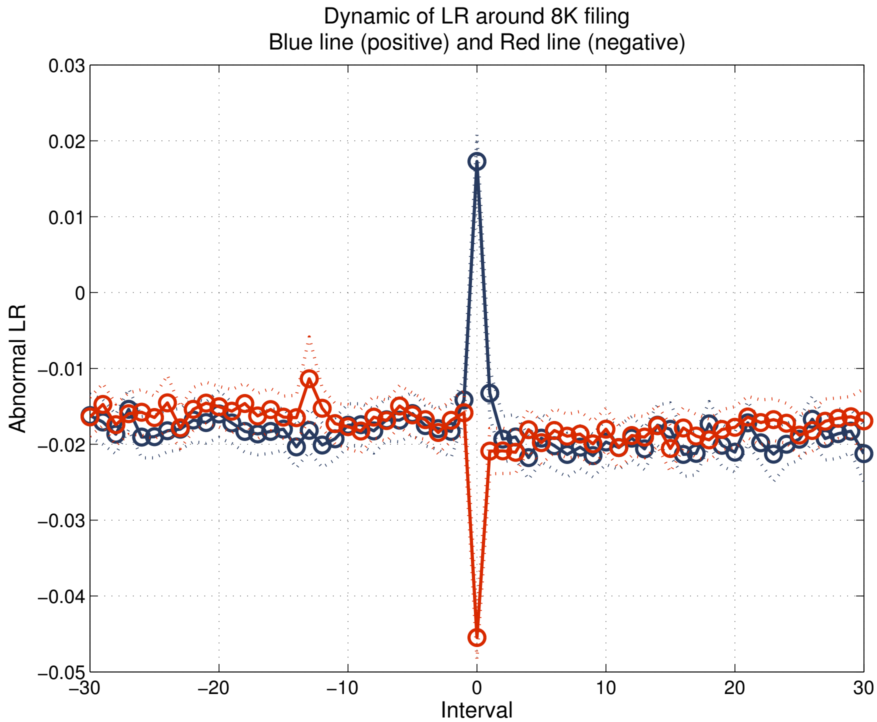
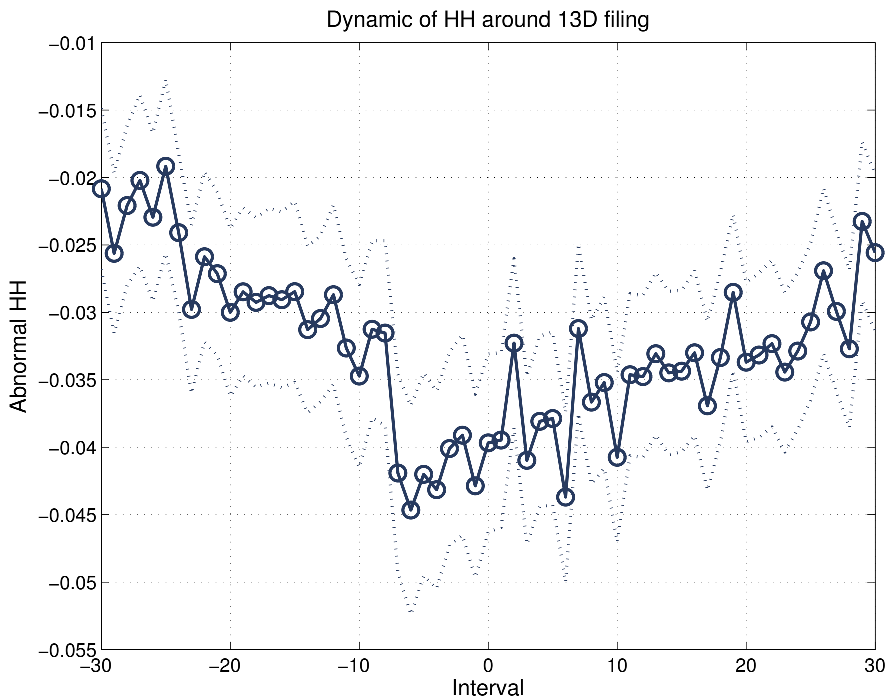
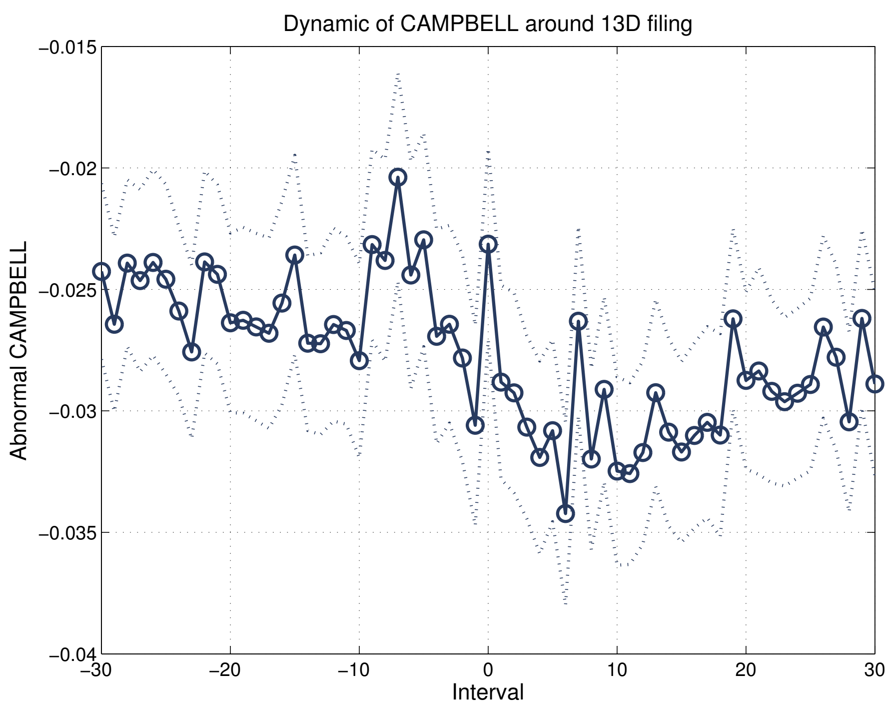


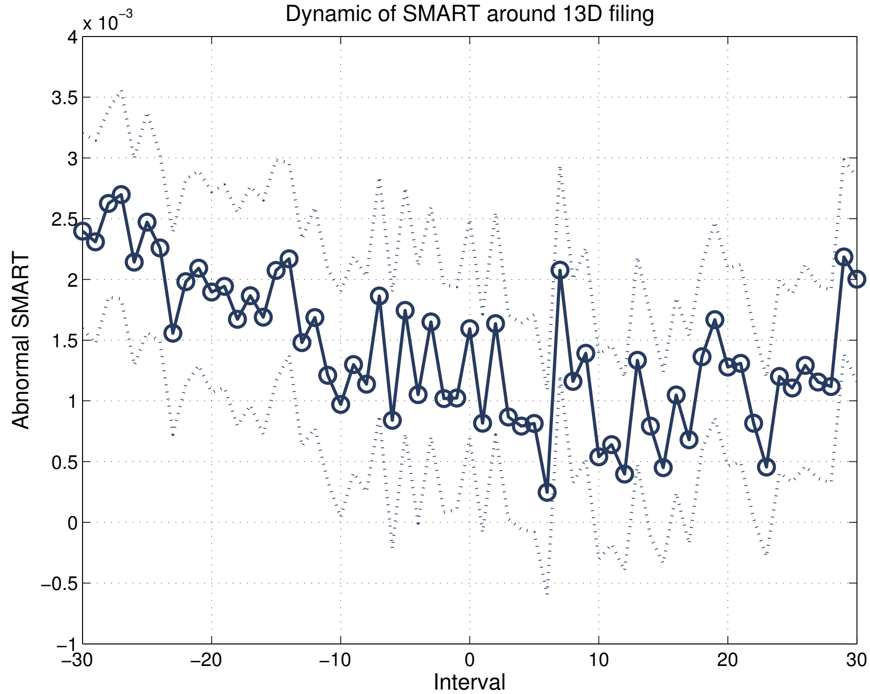
Figure 8. Dynamic of institutional order flow around scheduled 13D filing dates This figure presents the pattern of institutional order flow (IOF) around 13D filings. We plot abnormal IOF in excess of market average IOF on the same day from thirty days before to thirty days after a filing.

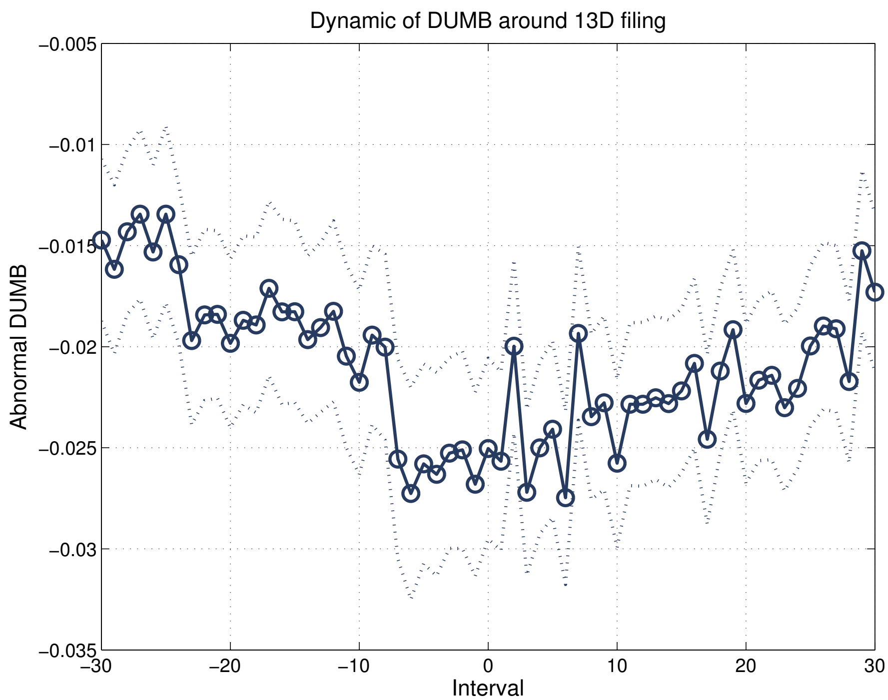




(Continued)

Figure 8 – *Continued*





(Continued)

Figure 8 – *Continued*