Predictability of Order Imbalance, Market Quality and Equity Cost of Capital*

Daniele Bianchi† Matthias Büchner‡ Roman Kozhan§

First draft: August 2018. This draft: March 29, 2019

Abstract

We study the effect of the predictability of order imbalance on market quality. We measure the degree of predictability by using the predictive likelihood from a dynamic linear model where the dependent variable is the day-ahead order imbalance. Empirically, we show that increasing order imbalance predictability corresponds to significantly higher market liquidity and efficiency. This positive relationship is economically significant: a long-short portfolio based on past predictability generates significant risk-adjusted returns. Predictability of order imbalance measures a cost of asymmetric information that is not captured by traditional measures of adverse selection. The risk factor that is associated with asymmetric information is priced in the cross-section of stock returns, controlling for a variety of conventional sources of systematic risk. These results suggest the existence of a tight link between market microstructure features affecting order imbalance predictability and both market quality and the cost of capital of firms.

Keywords: Order imbalance predictability, Market efficiency, Liquidity provision, Asymmetric information.

JEL classification: G10, G11, G14, G17.

*We are grateful to Constantinos Antoniou, Vimal Balasubramaniam, Björn Hagströmer, Kolja Johannsen, Olga Klein, Marc Lipson, Andrew Patton, Richard Payne, Vikas Raman, Elvira Sojli, as well as participants at the 2018 London Empirical Asset Pricing workshop and research seminars at Stockholm Business School and Warwick Business School for providing helpful comments and suggestions. We also thank the Centre for Scientific Computing at the University of Warwick for support with the supercomputing clusters.

†Warwick Business School, University of Warwick, Scarman Road, Coventry CV4 7AL, UK. E-mail: Daniele.Bianchi@wbs.ac.uk Web: whitesphd.com

‡Warwick Business School, University of Warwick, Scarman Road, Coventry CV4 7AL, UK. E-mail: matthias.buechner.16@mail.wbs.ac.uk Web: https://sites.google.com/view/matthiasbuechner

§Warwick Business School, University of Warwick, Scarman Road, Coventry CV4 7AL, UK. E-mail: roman.kozhan@wbs.ac.uk Web: https://homepages.warwick.ac.uk/staff/R.Kozhan/

1 Introduction

Most market micro-structure models recognize order imbalance as the main driver of the dynamics of asset prices. On the one hand, marker makers use information contained in observed order flow to efficiently set up prices and provide liquidity. Therefore, the ability to accurately predict the uninformed part of order imbalance is essential for both liquidity provision and efficient price setting as it allows market makers to mitigate both adverse selection and inventory risks. On the other hand, a high degree of order imbalance predictability can be detrimental for market quality as it facilitates predatory trading and front-running by informed traders. While academic literature studies the effect of uninformed order imbalance on trading costs and price efficiency, little attention has been devoted to the effect of predictability of order imbalance on market quality and expected returns. In this paper we fill this gap and empirically address this question.

To measure the degree of order imbalance predictability we estimate a dynamic linear regression model to predict the day-ahead order imbalance for each stock in the sample. We use a goodness-of-fit measure – the predictive likelihood – as a measure of predictability. In order to contrast the effect of predictability of total versus uninformed order flow we use two types of order imbalances: total order imbalances of a large cross-section of U.S. equities for the period from 1995 to 2013 and order imbalances from pension funds on a subset of U.S. stocks with active trading for the period from 2006 to 2010. We document substantial time-series and cross-sectional variation in the degree of predictability of the order imbalance for both types of order imbalances. In addition, we show that order imbalance predictability is cross-sectionally correlated with several stock characteristics, such as size, bid-ask spreads, share turnover, trade-to-quote ratios, and the volatility of order imbalance.

We provide evidence of a negative and statistically significant relation between the predictability of order imbalances and illiquidity, while controlling for variables which have been widely shown to affect the bid-ask spread in addition to time and individual fixed effects. Given the predictive nature of these regressions, the results strongly suggest that the predictability of order imbalance can be thought of as an additional channel in determining market liquidity. The effect is economically large: a one standard deviation change in predictive likelihoods leads to a 2.99 bps change in effective spreads; this corresponds to about 35.7% change relative to the sample average. The main empirical results also show that the predictability of order imbalance has a significant impact on price efficiency. By using the variance ratios and the autocorrelations of daily stock returns as proxies for price efficiency, we show that a one standard deviation increase in predictive likelihoods leads to 1.43% decrease in variance ratio and 1.02% decrease in absolute autocorrelation as compared to their sample averages. We obtain qualitatively similar effects when the predictability of the uninformed component of the order imbalance is proxied by the predictive likelihood of order imbalances from active pension funds. Furthermore, the results remain strong when splitting our sample into quintiles based on firms' characteristics, such as market capitalization, trade-to-quote ratio, idiosyncratic volatility, passive institutional ownership. Finally, we show that both effects for liquidity and price efficiency are more pronounced during periods of high market uncertainty, as proxied by the VIX index.

If predictability of order imbalance captures costs of information asymmetry, it poses a systematic risk for investors (see, e.g., Easley and O'Hara 2004). To test this intuition we examine if the predictability of order imbalance proxies for a risk factor that is priced in the cross section of stock returns. We provide evidence that the degree of order imbalance predictability correlates with monthly future stock returns: the stocks with the highest degree of predictability in order imbalance underperform stocks with lowest predictive likelihoods by 17.16% annually. Based on this evidence, we construct a low-minus-high predictability portfolio, denoted POF, and show that such strategy yields an annualized Sharpe ratio of 1.00. For small stocks, the portfolio with highest predictive likelihood underperforms stocks with lowest predictive likelihood by 23.96% annually. Such portfolio evidence is robust to a double-sorting procedure in which we control for other measures of adverse selection, such as bid-ask spread, share turnover, trade-to-quote ratio, and both the level and the volatility of the order imbalance.

Delving further into the asset pricing implications of the predictability of the order imbalance, we run a battery of regression-based asset pricing tests. First, we document the existence of a positive relation between future monthly return and the exposure to our POF factor. That is, for a given market cap quintile, stocks with lower predictability of order imbalance exhibit higher beta to the POF factor. By implementing a conventional Fama-MacBeth two-step regression we show that the risk proxied by our POF factor is priced in the cross-section of stock returns. More precisely, the price of risk associated with our POF factor is positive and statistically significant after controlling for Fama and French (1993) factors, the Jegadeesh and Titman (1993) momentum factor, and a number of factors that proxy asymmetric information, such as the Pastor and Stambaugh (2003) liquidity factor, returns on stocks sorted on trade-to-quote ratio, and both the level and the volatility of order imbalance (see, e.g., Chordia, Hu, Subrahmanyam, and Tong 2018). These results suggest that the main mechanism behind the effect of order imbalance predictability on market quality is via changes in adverse selection rather than reduction of inventory risk.

Our paper contributes to an existing debate on the role of order imbalance predictability in price discovery and the consequential effect on market quality. A number of papers document an improvement of market quality during events where order imbalance of uninformed parties can be predicted by market participants.[1] The main argument in favor of this intuition is that an announced, or anticipated, uninformed trade can attract additional liquidity suppliers and hence might improve liquidity and market resiliency.[2] On the contrary, there are a number of theoretical and empirical results advocating for a negative effect of order imbalance predictability on liquidity and efficiency due to predatory trading and front-running, especially around events with predictable order flow.[3] We directly measure predictability of order imbalances using a dynamic regression model that allows for time-varying coefficients and show that the positive effect significantly prevails in the data.

This paper connects to a substantial literature that investigates the asset pricing implications of asymmetric information and inventory risk. Several theory papers argued that adverse selection costs generate a non-diversifiable risk for which investors require compensation (see Admati 1985, 1993; Dow and Gorton 1953; Easley and O'Hara 2004). Madhavan and Smidt (1993), Brunnermeier and Pedersen (2005), Duffie (2010), Hendershott and Menkveld (2014) demonstrate theoretically how inventories of intermediaries generate substantial pricing errors in asset returns. There is also a large body of empirical research testing for the pricing effects of adverse selection and inventory risks, although the evidence is mixed. For instance, a number of papers using different proxies of adverse selection show significant pricing effects.[4] On the opposite, existing research claims that

[1] See Admati and Pfleiderer (1991); Degryse, De Jong, and van Kervel (2013); Bessembinder, Carrion, Tuttle, and Venkataraman (2015); Skjeltorp, Sojli, and Tham (2017) among others.

[2] An emerging literature on endogenous liquidity provision also makes an implicit assumption that predictability of order imbalance is a main factor that attracts high-frequency liquidity provision (see Raman, Robe, and Yadav 2016; Aït-Sahalia and Saglam 2017; Saglam 2018; Brogaard, Carrion, Moyaert, Riordan, Shkilko, and Sokolov 2018).

[3] See Brunnermeier and Pedersen (2005); Chen, Noronha, and Singal (2006); Carlin, Lobo, and Viswanathan (2007); Coval and Stafford (2007); Cheng and Madhavan (2009); Petajisto (2011); Mou (2011); Tuzan (2013) among others.

[4] Papers include, but are not limited to, Easley, Hvidkjaer, and O'Hara (2002) who estimate the probability of

some of these effects are not robust, possibly due to the fact that adverse selection costs are hard to measure empirically (see, e.g. Mohanram and Rajgopal 2009). Hendershott and Menkveld (2014) show that specialists inventory contributes to pricing errors in stock prices and Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010) show that inventories affect market liquidity. We contribute to this literature by documenting the asset pricing effects of predictability of order imbalance, which is related to both adverse selection costs and inventory risks.

Our paper also speaks to a recent literature that investigates the role of order imbalances in price discovery. Chordia, Roll, and Subrahmanyam (2002) demonstrate that market-wide returns are driven by lagged order imbalances after controlling for volumes and liquidity measures. Chordia and Subrahmanyam (2004) focus on order imbalances as predictors for individual stocks returns. In particular, they present evidence for order imbalance as a significant force linking market makers' inventories and stock price movements. Andrade, Chang, and Seasholes (2008) show that order imbalance in one stock affects prices in other stocks. More recent literature focuses on the effect of non-informational order imbalances. Griffin, Harris, and Topaloglu (2003) classify order imbalances for NASDAQ stocks as institutional- vs. retail-originated and document that institutional-driven order imbalances are persistent over several days and have a positive contemporaneous relation to returns. Boehmer and Wu (2008) document that institutional program trades provide liquidity to the market. Finally, Hendershott and Seasholes (2009) find that non-informational market-wide order imbalances affect individual stock returns and their cross-sectional co-movement.

The paper proceeds as follows. Section 2 develops testable hypotheses. Section 3 summarizes the data employed in our empirical analysis. Section 4 contains the findings from the main empirical analysis linking order imbalance predictability and market quality. Finally, Section 5 reviews our results and concludes.

informed trading (PIN), Pastor and Stambaugh (2003) who estimate the price impact, Akbas, Petkova, and Armstrong (2011) who use volatility of liquidity, Hwang and Qian (2011) who estimate an information risk measure based on the price discovery of large trades, Johnson and So (2015) who use the option-to-stock volume ratio as a measure of information asymmetry, Choi, Jin, and Yan (2016) who consider institutional ownership volatility, Yang, Zhang, and Zhang (2015) who use abnormal idiosyncratic volatility as a measure of information asymmetry and Chordia et al. (2018) who consider volatility of order imbalance.

2 Hypotheses Development

In well-functioning and competitive markets the predictability of order imbalance can affect both market quality and prices via several channels. By observing a signal about uninformed order imbalance market makers are able to filter out the informed demand as the residual component of the total contemporaneous order imbalance. The ability to separate informed and uninformed order flow allows market makers to adjust the level of liquidity in the market as a consequence of a reduced degree of asymmetric information. Market efficiency also improves through new information about future order imbalance being incorporated into the price. In addition, better knowledge of future order imbalance helps market makers to better handle their inventories and further improve liquidity and minimize price pressures due to a reduction of inventory risk.

On the other hand, predictable uninformed order imbalance can also have a detrimental effect on market quality. Sophisticated investors can incorporate knowledge of uninformed order imbalance into their trading strategies and better mask their demand. Moreover, informed traders, in addition to exploiting knowledge of fundamental information, can also front run the uninformed traders to earn additional returns. This reduces profits that market makers can earn to compensate their potential losses against informed traders and forces them to widen the bid-ask spread.

In order to guide the empirical analysis and demonstrate the competing effect of these two forces, we extend the the framework of Subrahmanyam (1991) and build an equilibrium model where market makers face both adverse selection and inventory concerns. We show that the net effect of predictability on market quality and prices depends on the model parameters. The model is described in detail in Appendix A. For most of the parameter values the positive effects of predictability of order imbalance on market liquidity tend to dominate, whereas for some other values of the parameters the opposite holds. As a result, understanding which one of the two effects described above is dominating in the data becomes ultimately an empirical question. In the following we outline the main hypotheses that come from the model, which we empirically test later on. The first two hypotheses concern the effect of predictability of uninformed order imbalance on market quality:

Hypothesis 1 (H1)**.** *The predictability of* uninformed *order imbalance in a stock has no effect on* *the liquidity of that stock.*

Our second testable hypothesis relates price efficiency.

Hypothesis 2 (H2). *The predictability of* uninformed *order imbalance in a stock has no effect on the price efficiency of that stock.*

Our main goal is to establish the effect of predictability of *total* order imbalance on market quality. It is not clear however if predictability of uninformed order imbalance is automatically translated into predictability of total order imbalance. In fact, although the uninformed order imbalance is a part of the total order imbalance, the part of the informed traders' demand is endogenously adjusted and might potentially eliminate or reduce the predictability of total order imbalance. Yet, our model establishes that in equilibrium the predictability of uninformed order imbalance generates predictability of total order imbalance (see Appendix A). To test this empirically, we re-formulate H1 and H2 by focusing on total order imbalance. Our re-formulated first hypothesis is again related to market liquidity:

Hypothesis 1b (H1b). *The predictability of* total *order imbalance in a stock has no effect on the liquidity of that stock.*

Similarly to H2 we are interested in the effect of the predictability of total order imbalance on price efficiency. Therefore our re-formulated second hypothesis is defined as

Hypothesis 2b (H2b). *The predictability of* total *order imbalance in a stock has no effect on the price efficiency of that stock.*

Finally, our third hypothesis concerns the relationship between the predictability of order imbalance and the cross-section of stock returns. There are two potential channels through which the predictability of order imbalance proxies some source of priced systematic risk: adverse selection and inventory management. Easley and O'hara (2004) provide a theoretical argument that adverse selection costs are significant determinant of the equity cost of capital. In particular, they show that investors demand a higher return to hold stocks with greater private information. This higher return arises because informed investors are able to better shift their portfolio to incorporate new information, and thus putting uninformed investors at a disadvantage. On the other hand, Hendershott and Menkveld (2014) demonstrate that shocks to inventories affect the pricing errors of stocks. In both cases, changes in order imbalance predictability are expected to partly explain the cross-sectional variation of stock returns. Therefore, our third hypothesis reads as follows:

Hypothesis 3 (H3). *The predictability of* total *order imbalance has no effect on the equity cost of capital.*

By testing H3 we can investigate whether the predictability of order imbalance proxies a source of systematic risk and hence contributes to the understanding of the cross-sectional variation of stock returns alongside typical risk factors such as size, value, liquidity as well as widely acknowledged anomalies such as momentum.

3 Data and Variables Definitions

We construct total order imbalance using data from the Trades and Quotes tape (TAQ) spanning all market places with reporting obligations. The main sample (we refer to this sample as *Full sample of stocks*) spans the period from January 1, 1995 to December 31, 2013. The TAQ data is pre-processed following Holden and Jacobsen (2014) and then matched with CRSP data using CUSIP / NCUSIP pairs. We exclude all securities with CRSP share code other than 10 or 11, as well as those stocks with average closing price smaller than \$5 and greater than \$1,000. We further require stocks to have more than 5 trades per day on average and a history of quotes longer than three years within our sample period. All variables used in the analysis are winsorized at the 0.1% and 99.9% quantiles so that values smaller / larger than these thresholds are set equal to the quantile value. After accounting for all filters, our sample includes 6,121 firms in total.

We use the algorithm proposed by Lee and Ready (1991) to infer the trade direction as "buy" or "sell" in the TAQ sample and use contemporaneous quotes to sign trades and calculate effective spreads (see, e.g., Holden and Jacobsen 2014). We define order imbalance oib_{it}^{tot} of stock i on day t as the daily number of buyer-initiated minus seller-initiated dollar volume of transactions scaled by total dollar volume.

We use another sample of stocks in order to proxy for the predictability of uninformed order imbalance. The uninformed order imbalance is approximated by using pension fund trading from the ANcerno database. We obtain institutional daily trading data for the period from January

1, 2006 to December 31, 2010.[5] Each institution type identifier distinguishes between clients that are pension plan sponsors vs. money managers. For each execution, the database reports the date of the trade, the execution price of the trade, the stock traded, the number of shares traded, whether the trade was a buy or a sell, and identity codes for the institution making the trade. The granularity of the data makes it particularly useful to compute stock level order imbalance of pension funds (denoted by oib_{it}^{pf}).

In using the pension funds as a proxy for the "uninformed" part of order imbalance we acknowledge that pension funds cover a fraction of the aggregate uninformed order imbalance (that also includes, among others, order imbalance from retail investors and ETFs). Moreover, there are many stocks where pension fund trading is not intense. Sparse trading can lead to spurious evidence of low predictability of order imbalance. To address this issue, we delete stocks with low level of pension fund trading activity. That means, a stock is included in our ANcerno sample if a trade has been executed for at least 50% of the days in the sample.

Insert Figure 1 about here

Our initial ANcerno sample includes 1,560 stocks that record at least one trade by a pension fund. Figure 1 presents the empirical cumulative distribution of the fraction of those firms as a function of the number of days during which we observe at least one trade by a pension fund. It shows that about 50% of the firms in the sample experienced a trade by pension funds for at least half their sample length, i.e. on average 791 days of trades. By applying this filter, we retain 789 firms in our *Active pension funds* sample.

We measure liquidity using quoted and effective spreads for both samples. The quoted spread $qspread_{it}$ is the difference between the bid and ask quotes of stock i scaled by the prevailing midpoint. The effective spread $espread_{it}$ is the difference between the midpoint of the bid and ask quotes and the actual transaction price of stock i, expressed as a proportion of the prevailing midpoint. We use all trades and quotes to calculate quoted and effective spreads for each reported transaction and calculate a share-weighted average across all trades in a given day. The realized spread $rspread_{it}$ is the difference between the transaction price and the midpoint five minutes prior

[5] The data stops in 2010 since ANcerno database stops providing the client type identifier after 2010.

to the trade scaled by the midpoint prevailing immediately before the transaction. The price impact $prcimpact_{it}$ is obtained from the difference of midpoint five minutes prior the current trade and the current midpoint scaled by the current midpoint. Finally, the intraday measures of $rspread$ and $prcimpact$ are averaged within a day.

We use two measures of market efficiency for both samples in our analysis: return autocorrelations and variance ratios. Return autocorrelations have been used to assess price efficiency in a number of papers (see,.e.g. Chordia, Roll, and Subrahmanyam 2008; Saffi and Sigurdsson 2011). We construct stock's i daily return autocorrelation ac_{it} based on a past 20 days rolling window basis. We use the past 20 observation days to construct our measure to capture almost a full month of trading activity and is motivated by earlier literature on long-memory processes in financial markets (see, e.g. Corsi 2009). Given that we are interested in the magnitude of the autocorrelation, we define an absolute value of a transformed autocorrelation measure as[6]

$$abs\_ac_{it} = \left| \frac{1 + ac_{it}}{1 - ac_{it}} \right|.$$

The variance ratio represents our second measure of price efficiency. The literature shows that the variance of longer horizon returns should equal the variance of shorter horizon returns times the frequency of the short horizon returns in the absence of autocorrelations (see Lo and MacKinlay 1988). This finding is a property of any random walk process since variance increases linearly with time. In the market micro-structure literature variance ratios have been used by Chordia et al. (2008). Similar to the autocorrelation measure we compute variance ratio using a 20-day rolling window. In the case of a perfect random walk, the variance ratio, defined as $2\, var[r_{it}^{(1)}]/var[r_{it}^{(2)}]$ should be equal to one in expectation, where $var[r_{it}^{(1)}]$ and $var[r_{it}^{(2)}]$ are the variance of one-day and two-day returns, respectively. We use the absolute value of the deviation of this ratio from one as a measure of market efficiency:

$$vratio_{it} = \left| 1 - \frac{2\, var[ret_{it}^{(1)}]}{var[ret_{it}^{(2)}]} \right|.$$

We use a set of order book and market characteristics to forecast order imbalances for both

[6]We transform this response variable to increase its variation from $-\infty$ to ∞.

samples. The choice of the predictors is motivated by taking the perspective of a market maker who is interested in forecasting order imbalance. In a typical setting, the market maker has access to past observed order imbalance and further order book measures such as volumes, intraday volatility, etc. Thus, we limit our set of predictors to this class of variables. The daily realized variance $rvar_{it}$ is defined as a sum of squared 5-minute midpoint price returns in stock i during day t. The trade-to-quote ratio ttq_{it} is constructed as a ratio of daily number of quotes (as measured by the changes in bid and ask quotes in the TAQ dataset) with respect to the number of trades in stock i on day t.[7] In addition, we also construct a measure of order book depth imbalance, i.e. the overhang of buy / sell orders in the order book.[8] It is defined a difference between depth on the ask side minus the depth on the bid side of the market divided by the total depth:

$$dib_{it} = \frac{ask\_depth_{it} - bid\_depth_{it}}{ask\_depth_{it} + bid\_depth_{it}}.$$

3.1 Predicting order imbalance

A standard approach for predicting order imbalance is a simple linear regression of the form

$$y_t = \boldsymbol{x}'_{t-1}\boldsymbol{\beta} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \tag{1}$$

where y_t represents the one-day ahead order imbalance, $\boldsymbol{x}_{t-1}$ a $(k+1) \times 1$ vector of predictors (see description below) in addition to the intercept, and ϵ_t an orthogonal error term with constant variance. A dynamic relationship between order imbalances can be imposed by recursively estimating Eq.(1) in a rolling window fashion for a constant window of size n. Although simple to implement such an approach is both ad-hoc and inefficient, as the lack of a more specific parametric form makes testing for time-variation highly dependent on hard-to-justify choices of the window size. As a matter of fact rolling window estimates $\hat{\beta}_t$ exploit a limited amount of information on order imbalances implicitly assigning equal weight to each observation in the estimation sample, that is, betas are assumed to change (remain constant) across (within) sub-samples with probability one.

[7] Rosu, Sojli, and Tham (2018) demonstrate theoretically and empirically an association between order imbalance and trade-to-quote ratio.

[8] Hagströmer (2018) shows that a micro-price effective spread measure, which is closely related to the order book depth imbalance, can be a valuable source of fundamental information for market makers.

In this paper, we follow existing research and assume a stochastic dynamics for the model parameters (see Hendershott and Menkveld 2014; Pascual and Veredas 2010; Brogaard, Hendershott, and Riordan 2014). In particular, we assume that the dynamics of the slope parameters follow a random walk, resembling the information acquisition process of market makers and informed traders who form beliefs based on the available information, i.e.,

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(0, \boldsymbol{\Omega}), \tag{2}$$

where $\boldsymbol{\Omega}$ is a $(k+1) \times (k+1)$ full covariance matrix. Such random walk dynamics imply that the weight assigned to past information now decrease exponentially rather than being constant, depending on how much the predictors are informative about future order imbalances, i.e., the signal-to-noise ratio (see West and Harrison 1997; Koopman and Durbin 2012 for details). The missing parameters $\boldsymbol{\Omega}$ and σ^2 are estimated by maximum likelihood, whereas the dynamic betas $\boldsymbol{\beta}_t$ are assumed latent and are extracted through a standard Expectation Maximization (EM) algorithm. In particular, we can obtain the forecast of the time t observation using data up to time $t-1$ as $y_{t|t-1}$ and the associated forecast variance. The model parameters are estimated by using the entire sample for each stock i.

3.1.1 Predictability Measure. Our objective is to forecast the one-day ahead order imbalance for stock i based on some conditioning information available at time $t-1$, $\boldsymbol{x}_{t-1}$. The predictor variables are the order imbalance of stock i on day $t-1$, $oib_{i,t-1}$, the individual stock return $ret_{i,t-1}$, the return on the S&P 500 index $spret_{t-1}$, the realized variance $rvar_{i,t-1}$, the trade-to-quote ratio $ttq_{i,t-1}$ and the depth imbalance $dib_{i,t-1}$.

The literature has entertained a number of out-of-sample forecast evaluation measures. For instance, Campbell and Thompson (2008) uses out-of-sample R^2 that is evaluating model forecast performance against historical mean forecasts in the context of return predictability. We choose a measure of out-of-sample forecast performance more related to the dynamic linear model. In particular, the model introduced above provides a forecast of the order imbalance $oib_{t|t-1}$ and the associated forecast variance $Var_t^2(oib_{t|t-1})$. The normality assumptions in Equation (1) ensure that the forecasted observations are distributed as $N\big(oib_{t|t-1}, Var_t(oib_{t|t-1})\big)$. Thus, we can obtain a measure of predictability that is akin to a predictive likelihood as

$$\tilde{pl}_t = p\big(oib_t \mid oib_{t|t-1}, \sigma^2(oib_{t|t-1})\big), \tag{3}$$

where $p(\mu, \sigma^2)$ is the density function of the Normal distribution with parameters μ and σ^2. This is the value of the normal probability density function evaluated at oib_t with mean $oib_{t|t-1}$ and variance $\sigma^2(oib_{t|t-1})$. One comment is in order. The predictive likelihood evaluated in (3) is not bounded from above such as for example empirical R^2. Notwithstanding, as we are interested in the degree of order imbalance predictability in relative terms (across stocks and time) rather than its absolute value, this does not pose any particular challenge for our analysis. Since a daily measure of predictability derived from the predictive likelihood is inherently noisy we opt to smooth the measure by taking its rolling 20-day average. Thus, henceforth pl_t is understood as the 20-day rolling average over $\tilde{pl}_t$ constructed from past observations from $t-19$ to t. We denote the predictability of total order imbalance in firm i on day t by pl^{tot}_{it} and the predictability of pension funds order imbalance in firm i on day t by pl^{pf}_{it}.

3.1.2 Summary Statistics. Table 1 reports the summary statistics for both samples: *Active pension funds* (Panel A) and *Full sample of stocks* (Panel B). For both samples we sort stocks based on their market capitalization using the monthly NYSE breakpoints and present summary statistics for each size quintile.[9] We refer to the smallest value quintile, e.g. small-cap, as Quintile 1 and to the highest value quintile as Quintile 5.

Insert Table 1 about here

Consistent with prior evidence, smaller stocks tend to be less liquid than large stocks. As a matter of fact, the average values of the four illiquidity variables are monotonically decreasing with size. Stocks traded by active pension funds seem to be more liquid, on average, than stocks in the aggregate sample. This is consistent with the idea that pension funds trade more actively in large stocks. Such observation is also confirmed by the number of observations in the Q5 quintile. Both

[9] The monthly NYSE breakpoints are retrieved from Kenneth French's data library `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_me_breakpoints.html`

market efficiency measures indicate that large stocks' prices are more efficient than the small stock ones. While for active pension funds this evidence is somewhat weak, average price inefficiency monotonically decreases from Q1 to Q5 in the *Full sample of stocks*.

The average daily total order imbalance is -1.87% and monotonically increases with the firm size, i.e., from -6.55% for small stocks to 4.32% for large stocks. The order imbalances from the *Active pension funds* sample are smaller in magnitude (see Panel A). For both samples, the order imbalance is autocorrelated with the first order autocorrelation coefficients being about 15% for the order imbalance in the *Full sample of stocks* and 33% for the order imbalance of *Active pension funds*, respectively.

The *Active pension funds* sample average of the predictive likelihood is 36.62%. This means that the probability of observing the order imbalance value of oib_t^{pf} within an ϵ-neighborhood of $oib_{t|t-1}$ is approximately $2\epsilon \times 0.3662$ for small $\epsilon > 0$. The sample average of predictive likelihood for the total order imbalance sample is 34.76%. On average, the order imbalance of large stocks is more predictable than the order imbalance of small stocks. This result holds for both samples. The descriptive statistics of the additional predictors, such as trading volume, trade-to-quote ratio, depth imbalance, realized variance and stock returns are summarized in Table A1 in the Appendix.

Figure 2 plots the cross-sectional mean and 95% quantiles of the predictive likelihoods as well as the time-series average of the predictive likelihoods across five market capitalization quintiles, both for the *Active pension funds* sample (Panel A) and the *Full sample of stocks* (Panel B).

Insert Figure 2 about here

Figure 2 documents a substantial variation of predictive likelihood in both the cross-section and time series. Furthermore, we observe that the median predictive likelihood (see Panel A of Figure 2) drops noticeably during the global financial crisis of 2008. The degree of predictability is the lowest for small cap stocks and is monotonically increasing with size. Panel B of Figure 2) also reveals that the slump in predictive likelihoods around the crisis is particularly pronounced for large cap stocks. Unsurprisingly, we find a trending behavior in the time series of predictive likelihoods. This can originate from the convergence characteristics of the dynamic linear model. Through the accumulation of information the model estimates improve and converge to their long-run levels. Another reason for the existence of trends can come from the sample characteristics due to increasing number of institutional investors and high-frequency trading in the market. In the following analysis, we will deal with both of those issues separately.

Insert Table 2 about here

Table 2 reports the sample pairwise correlations of the predictors that are used in the main empirical analysis (see above for a detailed description). With the only exception of the quoted and effective spreads, correlations tend to be low. The absence of contemporaneous correlations across predictors should rule out multi-collinearity issues in the battery of panel regressions estimated in the main empirical analysis. There is a potential concern that the degree of order imbalance predictability is related to the direction of order imbalance. It it is worth stressing a low correlation between predictive likelihood of order imbalance and the level of order imbalance itself which mitigates this concern.

4 Empirical Results

In this section we test the hypotheses outlined in Section 2. Hypothesis 1 and 2 concern the effect of predictability of uninformed order imbalance on market quality. We test both hypotheses by using the order imbalance from active pension funds. Hypothesis 1b and 2b are based on the conventional wisdom that posits that the predictability of uninformed order imbalance is translated in the predictability of the total order imbalance. It is evident that this assumption does not hold automatically due to the fact that the uninformed order imbalance is only a part of the total order imbalance. Informed traders can endogenously adjust their trading activity in response to the signal about the uninformed orders and which might reduce or completely eliminate the predictability of total order imbalance.

The theoretical model developed in the Appendix A shows that, in equilibrium, there is a clear association between the levels of predictability of the two types of order imbalance. While there is a demand adjustment by informed traders, its extent is insufficient to completely eliminate the predictability of the total order imbalance. The empirical evidence supports the positive association between the predictability of uninformed and total order imbalance. In particular, we merge our two samples and estimate the following regression

$$pl_{it}^{tot} = \alpha_i + \omega_t + \beta pl_{it}^{pf} + u_{it}, \tag{4}$$

where α_i contains firm fixed effects, and u_{it} is an idiosyncratic error term. Table 3 presents the estimation results for a simple pooled OLS, i.e., without firm and time fixed effects, as well as two panel regressions with firm fixed effects as well as time fixed effects.

Insert Table 3 about here

The estimates of β show that regardless of the regression specification there is a positive and statistically significant association between the predictability of total versus uninformed order imbalance. Interestingly, there is a substantial unobservable heterogeneity captured by firm-level fixed effects as suggested by a much higher R^2 that is obtained with respect to a simple pooled OLS regression (column 1).

4.1 Order imbalance predictability and liquidity

Following, our hypotheses 1 and 1b we directly test the relationship between the predictability of order imbalance and market quality by using a variety of measures. Specifically, we estimate the following regression

$$illiq_{it} = \alpha_i + \beta pl_{i,t-1} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it}, \tag{5}$$

where $illiq_{it}$ corresponds to one of several illiquidity variables $qspread_{it}$, $espread_{it}$, $rspread_{it}$ and $prcimpact_{it}$ and $pl_{i,t-1}$ stands for either $pl_{i,t-1}^{pf}$ or $pl_{i,t-1}^{tot}$. The set of control variables $\boldsymbol{X}_{it}$ include the log market cap of the firm i ($lsize_{it}$), the inverse of the price ($1/prc_{it}$), the trade-to-quote ratio (ttq_{it}), the share turnover ($turn_{it}$), the value of the VIX index (vix_t), the lagged value of the dependent variable ($illiq_{i,t-1}$) and firm fixed effects dummies. We also control for contemporaneous order imbalance oib_{it} to mitigate a potential concern that predictability of order imbalance is related to the level of order imbalance itself. While the regressions are predictive in nature (with respect to the predictability measure $pl_{i,t-1}$) we still control for a number of contemporaneous day t market variables to ensure that our results are not driven by other factors known to impact liquidity (see, e.g., Hendershott, Jones, and Menkveld 2011). Given that illiquidity levels are persistent on the daily timescale, we include lagged spread as a control to proxy for any omitted but potentially relevant variables. We report least-squares estimates of $\hat{\beta}$ and $\hat{\boldsymbol{\gamma}}$ with the corresponding t-statistics (in brackets) based on double-clustered standard errors.

Panel A of Table 4 reports the estimation results using the *Active pension funds* sample and pl^{pf} as the proxy for the predictability of the uninformed component of order imbalance. The empirical evidence supports a significant and negative relation between the predictability of uninformed order imbalance and illiquidity. Such negative relationship holds across all of the liquidity measures used. In addition, the economic size of the effect is quite substantial; a one standard deviation increase in predictive likelihoods leads to an average decrease in the quoted spreads by 0.56 basis points and a decrease in the effective spreads by 0.27 basis points.

Insert Table 4 about here

The coefficients for pl^{pf} are statistically significant at the 1% level. This rejects our Hypothesis 1 in favor of the alternative that an increase in the predictability of uninformed order imbalance improves market liquidity, on average. The estimation results based on the *Full sample of stocks* and pl^{tot} as the proxy for predictability of total order imbalance (see Panel B of Table 4) also rejects Hypothesis 1b. In fact, the effect of total order imbalance predictability on liquidity is an order of magnitude larger than the predictability of pension funds order imbalance. A one standard deviation increase in total order imbalance predictive likelihood leads to an average decrease in the quoted and effective spreads by 2.99 bps. The coefficients for pl^{tot} are also statistically significant at 1% level. It is worth noting that the effect of predictability of order imbalance on the adverse selection component of the effective spread (*prcimpact*) is about three times higher than on the realized spreads. This emphasizes the claim that predictability of order imbalance is closely related to the cost of adverse selection.

4.2 Order imbalance predictability and market efficiency

Next, we test the effect of predictability of order imbalance on market efficiency (Hypotheses 2 and 2b). To do so we estimate the following regression

$$ineff_{it} = \alpha_i + \beta pl_{i,t-1} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it}, \tag{6}$$

where $ineff_{it}$ corresponds to one of the two illiquidity variables $vratio_{it}$ and abs_ac_{it}, whereas $pl_{i,t-1}$ stands for either $pl^{pf}_{i,t-1}$ or $pl^{tot}_{i,t-1}$. Similar to Eq.(5), the set of control variables include the log market cap of the firm i ($lsize_{it}$), the inverse of the price ($1/prc_{it}$), the trade-to-quote ratio (ttq_{it}), the share turnover ($turn_{it}$), the value of the VIX index (vix_t). The parameters α_i capture a firm fixed effect. The t-statistics are presented in brackets and are based on double-clustered standard errors.

Insert Table 5 about here

Panel A of Table 5 provides the estimation results for the *Active pension funds* sample and $pl^{pf}_{i,t-1}$ as the independent variable. We find a significant negative relation between the predictability of order imbalance and the degree of market inefficiency. Again, such negative relationship holds for both measures of market inefficiency. That is, an increase by one standard deviation in the predictive likelihood corresponds to a decrease in the variance ratio by 0.196% (which is about 1.43% of its sample average value) and absolute autocorrelation of daily returns by 0.362% (which is about 1.02% of its sample average). Both of these effects are statistically significant at the 1% level. This results rejects Hypothesis 2 that a higher predictability of uninformed order imbalance does not improve market efficiency, on average.

We obtain similar results for the predictability of order imbalance from the *Full sample of stocks*, i.e., $pl^{tot}_{i,t-1}$. Indeed, the coefficient β is still negative and significant at the 1% level (see Panel B of Table 5). An increase by one standard deviation in predictive likelihood of total order imbalance decreases the variance ratio by 0.547% (which is about 3.44% of its sample average value) and the absolute autocorrelation of daily returns by 0.969% (which is about 2.45% of its sample average). This results rejects Hypothesis 2b that a higher predictability of order imbalance does not improve market efficiency, on average.

4.3 Order imbalance predictability, market conditions and stock characteristics

Next, we explore how the relation between order imbalance predictability and market quality depends on aggregate market conditions. We interact the predictive likelihoods with a time dummy corresponding to different levels of the VIX index, which is a widely used proxy for aggregate market uncertainty (see, e.g., Bloom 2009). We define a dummy variable D_t^{high} that takes value one whenever vix_t falls within the upper tercile of its historical distribution drawn from our sample and zero otherwise. Similarly, dummy variables D_t^{med} and D_t^{low} take the value one when vix_t is in the middle and bottom terciles, respectively. Each of these dummy variables are then interacted with the predictive likelihood to study the differential impact of predictability on spreads during periods of high vs. low market uncertainty. Specifically, we estimate a panel regression of the following form:

$$y_{it} = \sum_{j \in \{\text{high,med,low}\}} \beta_j \cdot pl_{i,t-1} \cdot D_t^{(j)} + \alpha_i + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it}, \tag{7}$$

where y_{it} represents either one of the illiquidity measures $illiq_{it}$ or one on the proxies used for market inefficiency $ineff_{it}$, and α_i represents a firm fixed effect. Here, $pl_{i,t-1}$ stands for either $pl_{i,t-1}^{pf}$ or $pl_{i,t-1}^{tot}$. The set of control variables is the same as in the previous regressions. Standard errors are double clustered by firm and time.

Insert Table 6 about here

Table 6 reports the estimation results. For the ease of exposition we present the results for a measure of illiquidity and a measure of inefficiency, namely the quoted spread $qspread_{it}$ and the variance ration $vratio_{it}$. Panel A displays the estimation results for the *Active pension funds* sample. For all three levels of the VIX index we confirm the positive effect of the predictive likelihoods on liquidity and market efficiency. The marginal effect of the predictability of uninformed order imbalances on market liquidity tends to increase as aggregate market uncertainty increases. That is, bid-ask spreads decrease by a greater extent during periods of high investors' uncertainty. On the other hand, the effect of predictability of uninformed order imbalance on market efficiency tends to be higher during calmer times.

Panel B reports the results for the *Full sample of stocks*. The magnitude is stronger than for the sub-sample of pension funds. Nevertheless the monotonic relationship of the interaction terms is preserved. Again, the effect of predictability of total order imbalance on market liquidity is much more pronounced as aggregate uncertainty increases. On the opposite, the effect of predictability on price efficiency decreases as the level of the VIX index increases. Table A2 in the Appendix shows further results for additional measures of market liquidity such as the effective spread, $espread_{it}$, the realized spread, $rspread_{it}$, and price impact, $prcimpact_{it}$, as well as returns autocorrelation, abs_ac_{it}, as additional measure of market inefficiency. The results confirm the evidence provided in Table 6, namely the higher the market uncertainty the bigger (smaller) is the impact of order imbalance predictability on liquidity (inefficiency).

A potential concern is that the documented effects exist only for a specific group of stocks and is attributed to size and volatility of a firm due to limits to arbitrage, activity of institutional investors or high-frequency traders. We address this question by sorting the stocks in quintiles for a given characteristic and re-estimating Equations (5) and (6) separately for each quintile.

Insert Table 7 about here

Table 7 reports the estimates of the slope parameter $\hat{\beta}$ on the measure of predictability. Panel A reports the results for the sample of stocks traded by *Active pension funds*, whereas Panel B reports the results for the *Full sample of stocks*. For the ease of exposition we report the results for quoted spreads as a measure of liquidity. Few interesting aspects emerge; when stocks are sorted by market cap, the coefficient on pl_{t-1}^{pf} (see Panel A) is negative and statistically significant for all five quintiles. The strongest effect is for small stocks and the weakest effect is for large stocks. A one standard deviation increase in predictive likelihood pl_{t-1}^{pf} leads to about 0.91 basis points decline in quoted spreads of small stocks and only to about 0.08 basis points decline in quoted spreads of large stocks. We obtain qualitatively similar results when stocks are sorted by trade-to-quote ratio, by idiosyncratic volatility and passive institutional ownership. The highest magnitude of coefficient on pl_{t-1}^{pf} is for stocks with the lowest trade-to-quote ratio, with the highest idiosyncratic volatility and with lowest passive institutional ownership. Our results are also robust to the choice of liquidity measure.

Panel B confirms that the coefficient on the predictability measure pl_{t-1}^{tot} is negative and statistically significant at 1% level for all specifications. The pattern of the estimate across stock characteristics remains the same as for Panel A. These results confirm that an increase in the predictability of order imbalance reduces the extent of adverse selection in the market and overall improves liquidity, conditional on a wide range of stock characteristics.

Insert Table 8 about here

Table 8 confirms the relationship between the predictability of order imbalance and market efficiency that we showed at the aggregate level (see Table 5). Panel A reports the results for the sample of stocks traded by *Active pension funds*; in the vast majority of cases, price efficiency – measured by the variance ratio – substantially improves with an increase of order imbalance predictability for each quintile sorted by market cap, trade-to-quote ratio, idiosyncratic volatility and passive institutional ownership. The exceptions are mainly for stocks in the top and bottom market cap quintiles, stocks in top trade-to-quote ratio quintile, stocks with the highest idiosyncratic volatility and lowest passive institutional ownership. In these cases the coefficient on pl_{t-1}^{pf} is not statistically significant. Similar to Table 7, the results for the *Full sample of stocks* are much stronger (see Panel B). The coefficient on pl_{t-1}^{tot} is negative and statistically significant at the 1% level for all characteristic quintiles, with the only exception of the largest stocks.

Table A3 in the Appendix reports further results on alternative measures of liquidity and market efficiency. Panel A displays the results for the stocks traded by *Active pension funds*. Except few nuances, the effect of the predictability of order imbalance on liquidity tends to be almost monotonically decreasing with the market capitalization, trade-to-quote ratio, and the amount of passive institutional ownership. Panel B confirms the results for the *Full sample of stocks*; in fact, the magnitude of the coefficients is much higher and the monotonic behavior of the beta on predictability for the liquidity measures. As far as *abs_ac* is concerned, both Panel A and panel B confirm the results on market efficiency outlined in Table 8.

4.4 De-Trending the Predictive Likelihood

In addition to the main empirical analysis we investigate the robustness of the results to the concern of a spurious regression problem due to the trending behavior of our predictability measure pl_t. To address this issue, we perform a linear de-trend of the predictive likelihood and repeat the regression exercises using the de-trended series. For each given stock, we define the de-trended predictive likelihood pl_t^* as residuals from the regression of pl_t on the corresponding time index t, i.e., $pl_t = \alpha + \beta\, t + pl_t^*$.

Table A4 in the Appendix contains the results for regressions of quoted spreads onto the de-trended predictive likelihood. The coefficient on the de-trended predictive likelihood remains negative and statistically significant. Moreover its magnitude increases in absolute value. This largely confirms the main results outlined above, both in relation to the effect of predictability of order imbalance on market quality and on efficiency. Notice the additional results for the de-trended predictive measure hold for both the sample of stocks traded by *Active pension funds* (Panel A) and the *Full sample of stocks* available in our dataset (Panel B).

4.5 The asset pricing implications

In this section we test Hypothesis 3 which lays out the relation between the predictability of order imbalance and the cross-sectional variation of stock returns. In Sections 4.1 - 4.3 we have shown that the predictability of pension funds and total order imbalances has similar effects on market quality. For the investigation of asset pricing implications of the predictability of order imbalance, we concentrate on the *Full sample of stocks* in order to benefit from the wider cross-section of stocks and longer time period.

We start by presenting the portfolio characteristics across quintiles sorted by the predictability of total order imbalance. To do so, we first aggregate our predictability measure to monthly frequency by averaging the daily predictive likelihood within a given month. Table 9 reveals some clear path in the characteristics of stocks conditional on the amount of order imbalance predictability.

Insert Table 9 about here

For instance, stocks with the highest level of predictability in the trading imbalance tend to be growth stocks, larger in terms of market capitalization, and experience past returns which are both lower and less volatile. Although not exhaustive, this evidence may imply that our predictability measure simply captures systematic risk features such as size, value, and momentum. In addition, the fact that the order imbalance of less liquid stocks tend to be more predictable may suggests that our predictability may be subsumed by a liquidity risk factor.

4.6 Portfolio analysis

We test the implications of our predictability measure to capture the cross-sectional variation of stock returns against a variety of competing explanations both based on a portfolio sorting procedure and a regression analysis. In particular, we first present univariate portfolio sort results for the predictability of order imbalance. The dark-blue bars in Figure 3 show the average future returns for each of quintiles sorted on previous month's value of $pl_{i,t-1}^{tot}$. We also present the average return of the trading strategy that buys an equally weighted portfolio of stocks which experienced a low level of predictability of order imbalance during past month and sells short the portfolio with high predictability instead.

Insert Figure 3 about here

The average returns are monotonically decreasing with respect to the predictability of order imbalance. Low predictable stocks experience a return of 2.14% per month, whereas stocks with more predictable order imbalance generate a much lower 0.86% monthly return over the period 01:1997-12:2013. The second red bar from the right shows that our low-minus-high predictability portfolio generates a statistically significant returns at a conventional 1% credibility level. Such average return is also economically large: the strategy yields on average 1.29% per month with the annualized Sharpe ratio of 1.00. In addition, the first red bar from the right shows that our strategy also generates a substantial risk-adjusted return calculated conditional on several sources of systematic risks, such as market, size, value, momentum and the liquidity factor introduced by Pastor and Stambaugh (2003).

Insert Figure 4 about here

Figure 4 plots the cumulative returns of our low minus high predictability portfolio, which we denote by POF. The returns are pro-cyclical and experience positive performance for most of the sample except 1999 and 2008 crises. The value-weighted strategy generates lower returns as compared of its equally-weighted counterpart. This is consistent with the evidence that the bulk of the economic magnitude is within small stocks, which are mechanically underweighted in a value-weighted portfolio.

As discussed above, Table 9 reveals some monotonic relation of the predictability of order imbalance with other characteristics that are shown to be priced in the empirical finance literature, e.g., size and value. To investigate the pricing implications of POF we perform a set of double sorts based on $pl_{i,t-1}^{tot}$ and other variables, like size, bid-ask spread, turnover, trade-to-quote ratio and volatility of order imbalance. Table 10 reports the results.

Insert Table 10 about here

Panel A reports the returns of 5×5 portfolios sorted on past predictability of order imbalance and the lagged value of market capitalization. The average return on the low-minus-high predictability portfolio, conditional on small stocks, is positive and statistically significant. Conditional on the smallest stocks, the portfolio yields 1.64% per month. For large stocks the return on the low minus high portfolio is still positive but not statistically significant. Interestingly, when controlling for market, size, value, momentum and liquidity factors, the Jensen's alphas are large and statistically significant at 1% level for all quintiles with the only exception of Q5, i.e., largest stocks.

Panel B shows the results for a double-sort of 5×5 portfolios based on lagged predictability of order imbalance and past value of the bid-ask spread. Interestingly, a conditional strategy based on low-minus-high predictability and high bid-ask spreads generate the highest performance, with a 2.18% on a monthly basis and a risk-adjusted return, once controlling for the aforementioned risk factors, equal to 2.4% per month. The first column shows that there is no significant premium captured by our POF portfolio for high liquid stocks. The results for a double sort based on past order imbalance predictability and trade-to-quote ratios are significant across most portfolios. All low-minus-high portfolios generate substantial risk-adjusted returns in the range 1.12% for Q2 to 3.00% for Q5.

Panel D and E shows the performance of our strategy conditional on different levels of idiosyncratic volatility and shares turnover, respectively. Our strategy tends to perform better for stocks with high levels of idiosyncratic volatility and high turnover. As a matter of fact, when idiosyncratic volatility is low, i.e. Q1, our low-minus-high predictability strategy does not generate statistically significant returns. On the opposite, for highly volatile stocks our POF strategy generates up to 1.84% risk-adjusted returns on a monthly basis.

Finally, Panel F reports the returns of portfolios sorted on past order imbalance predictability and the order imbalance. The average return on the low-minus-high predictability portfolio, conditional on low order imbalance stocks, is 1.398% per month. For high order imbalance stocks the average return on the high-low portfolio drops to 1.036% per month.

As a whole, the results are consistent across all double-sort characteristics. Moreover, the average returns for low-minus-high predictability portfolios typically increase when the sorts are based on other characteristics instead of size. These results suggest that the predictability of order imbalance proxies some aggregate and non-diversifiable source of risk that is priced in the cross-section of stock returns. Yet, this pricing effect is robust when controlling for other relevant characteristics which are commonly assumed to proxy for systematic risks.

4.7 Cross-sectional regressions

If low predictability of order imbalance poses a risk to marginal investors due to increased asymmetric information, we should observe a risk-return trade-off. To investigate this we construct 25 portfolios based on size and the predictability of order imbalance, and test the exposure of these test assets to a set of commonly used systematic risk factors as well as the returns on our POF.[10] More specifically, we first estimate the following time series regression,

$$ret_{i,t} = \alpha_i + \beta_i POF_t + \boldsymbol{\gamma}_i' \boldsymbol{F}_t + u_{i,t}, \qquad i = 1, \ldots, 25 \tag{8}$$

[10] We sort by predictability of order imbalance to ensure that the POF betas have high variation across portfolios and we choose size as the second sorting characteristics because it is a hard benchmark to beat given results in Table 10.

where $\boldsymbol{F}_t = (Mkt_t, SMB_t, HML_t, MOM_t, LIQ_t, REV_t, TTQ_t, VOIB_t)'$ contains the excess returns on the market, the size and value factors of Fama and French (1993), the momentum factor calculated as in Jegadeesh and Titman (1993), the liquidity risk factor from Pastor and Stambaugh (2003), as well as the short-term reversals factor and the returns on high-minus low portfolios sorted by past month trade-to-quote ratio and volatility of order imbalance.

Insert Table 11 about here

Table 11 reports the estimates of the slope parameter $\hat{\beta}_i$ for each of the 25 test portfolios. Notice that the nature of the regression (8) implies that higher estimates $\hat{\beta}_i$ are associated with higher expected returns controlling for other risk factors. Panel A shows the results for equally-weighted test assets. Conditional on the level of predictability, the exposure to POF tends to decrease with the size of a firm. Similarly, conditional on size, the exposure of stock returns to the POF returns decrease with the level of predictability. That is, small size-low predictability stocks tend to be the most correlated to the POF factor, whereas the opposite is true for the large size-high predictability stocks. Interestingly, the vast majority of betas are highly statistical significant.

Panel B shows the results for value-weighted test assets. The results are largely similar to the equally-weighted portfolio both in terms of sign, magnitude and significance of the slope parameter estimates. As a whole, the results for the value-weighted test portfolios reinforce the evidence of Panel A whereby large size-high predictability stocks tend to have the most negative correlation with the POF factor, while small size-low predictability stocks tend to have the highest correlation with the POF factor.

Although instructive, the results reported in Table 11 do not necessarily imply that the risk associated with the time series variation of our POF factor is priced in the cross-section of stock returns. To address this issue, and further investigate the asset pricing implications of the predictability of order imbalance, we perform a set of standard Fama-MacBeth regressions. This allows to estimate the equilibrium price of risk implied by order imbalance predictability while controlling for alternative, and possibly competing, sources of systematic risk (see Fama and MacBeth

1973). Specifically, we estimate the following cross-sectional regression

$$returns_{i,t} = \lambda_{POF,t}\hat{\beta}_i + \boldsymbol{\lambda}_t'\hat{\gamma}_i + e_{i,t}, \qquad t = 1, \ldots, T \tag{9}$$

where the parameters $\hat{\beta}_i, \hat{\gamma}_i$ are estimated for each test portfolio by using the time series regression in Eq.(8). The estimates $\hat{\lambda}_{POF}, \hat{\boldsymbol{\lambda}}'$ are obtained as the sample averages of $\hat{\lambda}_{POF,t}, \hat{\boldsymbol{\lambda}}_t'$ (see Ch.12 Cochrane 2009).[11] As additional control variables, we use the returns on a set of alternative strategies such as a buy-and-hold exposure on the market, long-short portfolios based on size, value, momentum, liquidity as well as short-term reversals. In addition, we include the returns on two alternative high-minus-low portfolio strategies constructed based on the past month trade-to-quote ratio and the volatility of order imbalances (see Chordia et al. 2018).[12] Panel A of Table 12 reports the estimation results for 5×5 portfolios sorted on past predictability of order imbalance and the lagged value of market capitalization, and each portfolio is equally weighted. The baseline specification is a standard three-factor Fama-French model with the addition of our *POF* factor. Then each of the competing risk factors is added one by one to investigate the marginal effect of *POF* conditioning for alternative risk factors which have been investigated in the market microstructure literature. Last column shows an implementation in which all factors have been included.

Insert Table 12 about here

Two interesting facts emerge: first, the price of risk for our *POF* is positive and significant at the 1% confidence level for all regression specifications. This means that the risk associated with the time series variation of our *POF* factor is priced in the cross-section of stock returns. Consistent with the fact that $\hat{\beta}_i$ load on a tradable factor, the price of risk is statistically indistinguishable from the average return on the *POF* portfolio. Second, the prices of risk for both the size factor and a long-short portfolio based on the volatility of order imbalances – which is labeled *VOIB* in the table – remain positive and highly significant across model specifications. The results for *VOIB* are consistent with recent evidence provided by Chordia et al. (2018).

[11]To address the time-series correlation of the lambdas estimates, the standard errors are calculated based on a Newey-West estimator with 12 lags (see Newey and West 1987).

[12]Notice the two additional portfolios constructed based on the past month trade-to-quote ratio and the volatility of order imbalances are built by sorting stocks into the corresponding quintiles on our sample.

Panel B reports the results whereby the 5×5 test portfolios are still sorted on past predictability of order imbalance and the lagged value of market capitalization, but now are value-weighted. The value weighting scheme naturally dilutes the effect of small stocks in the time-series variation of the test portfolios. Despite the different weighting scheme, the sign, magnitude and significance of the price of risk for the *POF* factor remain intact. Again, both the *SMB* and the *VOIB* risk proxies turn out to be significant and positive across model specifications.

As a whole, Table 12 provide evidence that the risk associated with our *POF* factor is neither subsumed nor substantially affected by other competing (or alternative) risk factors, such as liquidity, short-term reversals, size and value. This result, coupled with the portfolio evidence in Table 10, suggests that we can reject Hypothesis 3 and therefore provide significant evidence that the predictability of order imbalance affects the equity cost of capital.

5 Conclusion

In this paper, we provide evidence that the predictability of order imbalance is an important determinant of market quality. We take the perspective of a market maker that forecasts order imbalances through a dynamic linear model and show that there is a significant and positive relationship between the predictability of order imbalances and both measures of market liquidity and market efficiency. Such positive relationship is robust to controlling for periods of high and low market uncertainty and across different groups of stocks sorted by size, idiosyncratic volatility, trade-to-quote ratio and institutional ownership.

These findings are consistent with the notion that increase in predictability of order imbalance reduces adverse selection costs and inventory risks. As a result, order imbalance predictability significantly affects stocks prices. We document that a strategy that buys stocks with low order imbalance predictability and sell short stocks with high order imbalance predictability produces significant and economically large returns. Thieve returns are not explained by a set of traditional risk factors as well as other characteristics that capture different aspects of adverse selection costs. The exposure of stock returns to the returns of this strategy commands for a systematic risk premium.

Our results demonstrate that many features of market microstructure that potentially influencing predictability of order imbalance can affect firms' equilibrium return. This dictates that a firm's cost of capital is influenced by not only by information, but also by market anonymity, organization of exchanges, proliferation of high-frequency trading and increase in ETF trading. This provides another linkage between market microstructure, asset pricing and corporate finance.

References

Admati, A., 1985, A noisy rational expectations equilibrium for multi-asset securities markets, *Econometrica* 53, 629–658.

Admati, A., 1993, A model of intertemporal asset prices under asymmetric information, *Review of Economic Studies* 60, 249–282.

Admati, A., and P. Pfleiderer, 1991, Sunshine trading and financial market equilibrium, *Review of Financial Studies* 4, 443–481.

Aït-Sahalia, Y., and M. Saglam, 2017, High frequency market making: Optimal quoting, *Working paper* .

Akbas, F., R. Petkova, and W. Armstrong, 2011, The volatility of liquidity and expected stock returns, *Working paper* .

Andrade, Sandro C, Charles Chang, and Mark S Seasholes, 2008, Trading imbalances, predictable reversals, and cross-stock price pressure, *Journal of Financial Economics* 88, 406–423.

Bessembinder, H., A. Carrion, L. Tuttle, and K. Venkataraman, 2015, Liquidity, resiliency, and market quality around predictable trades: Theory and evidence, *Journal of Financial Economics* 7, 1–20.

Bloom, N., 2009, The impact of uncertainty shocks, *Econometrica* 77, 623–685.

Boehmer, Ekkehart, and Juan Julie Wu, 2008, Order Flow and Prices, *Working Paper* .

Brogaard, J., A. Carrion, Th. Moyaert, R. Riordan, A. Shkilko, and K.n Sokolov, 2018, High frequency trading and extreme price movements, *Journal of Financial Economics* 128, 253–265.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2014, High-frequency trading and price discovery, *Review of Financial Studies* 27, 2267–2306.

Brunnermeier, M., and L. Pedersen, 2005, Predatory trading, *Journal of Finance* 60, 1825–1863.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Carlin, B., M. Lobo, and S. Viswanathan, 2007, Episodic liquidity crises: cooperative and predatory trading, *Journal of Finance* 62, 2235–2274.

Chen, H., G. Noronha, and V. Singal, 2006, Index changes and losses to index fund investors, *Financial Analysts Journal* 62, 31–47.

Cheng, M., and A. Madhavan, 2009, The dynamics of leveraged and inverse exchange-traded-funds, *Journal of Investment Management* 7, 1–20.

Choi, J., L. Jin, and H. Yan, 2016, Informed trading and expected returns, *Working paper* .

Chordia, T., J. Hu, A. Subrahmanyam, and Q. Tong, 2018, Order flow volatility and equity costs of capital, *Management Science* forthcoming.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2002, Order imbalance, liquidity, and market returns, *Journal of Financial Economics* 65, 111–130.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2008, Liquidity and market efficiency, *Journal of Financial Economics* 87, 249–268.

Chordia, Tarun, and Avanidhar Subrahmanyam, 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics* 72, 485–518.

Cochrane, John H, 2009, *Asset pricing: Revised edition* (Princeton university press).

Comerton-Forde, C., T. Hendershott, C. Jones, P. Moulton, and M. Seasholes, 2010, Time variation in liquidity: The role of market maker inventories and revenues, *Journal of Finance* 65, 295–331.

Corsi, Fulvio, 2009, A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics* 7, 174–196.

Coval, J., and E. Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

Degryse, H., F. De Jong, and V. van Kervel, 2013, Does order splitting signal uninformed order flow?, *Working paper* .

Dow, J., and G. Gorton, 1953, Profitable informed trading in a simple general equilibrium model of asset pricing, *Journal of Economic Theory* 67, 327–369.

Duffie, D., 2010, Presidential address: asset price dynamics with slow-moving capital, *Journal of Finance* 65, 1237–1267.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, Is information risk a determinant of asset returns?, *Journal of Finance* 57, 2185–2221.

Easley, D., and M. O'Hara, 2004, Information and the cost of capital, *Journal of Finance* 59, 1553–1583.

Easley, David, and Maureen O'hara, 2004, Information and the cost of capital, *The journal of finance* 59, 1553–1583.

Fama, E., and K. French, 1993, Common risk factors in the returns on stocks and bonds., *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Griffin, John M, Jeffrey H Harris, and Selim Topaloglu, 2003, The Dynamics of Institutional and Individual Trading, *Journal of Finance* 58, 2285–2320.

Hagströmer, Björn, 2018, Overestimated effective spreads: Implications for investors, *Working paper* .

Hendershott, T., and A. Menkveld, 2014, Price pressures, *Journal of Financial Economics* 114, 405–423.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does Algorithmic Trading Improve Liquidity?, *Journal of Finance* 66, 1–33.

Hendershott, Terrence, and Mark S. Seasholes, 2009, Market Predictability and Non-Informational Trading, *Working Paper* .

Holden, Craig W, and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *Journal of Finance* 69, 1747–1785.

Hwang, C., and X. Qian, 2011, Is information risk priced? evidence from the price discovery of large trades, *Working paper* .

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of finance* 48, 65–91.

Johnson, T., and E. So, 2015, A simple multimarketmeasure of information asymmetry, *Working paper* .

Koopman, S.J., and J. Durbin, 2012, *Time Series Analysis by State Space Methods*, second edition (Oxford University Press, Oxford).

Lee, Charles M C, and Mark J. Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46, 733–746.

Lo, Andrew W., and A. Craig MacKinlay, 1988, Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test, *Review of Financial Studies* 1, 41–66.

Madhavan, A., and S. Smidt, 1993, An analysis of changes in specialist inventories and quotations, *Journal of Finance* 48, 1595–1628.

Mohanram, P., and S. Rajgopal, 2009, Is PIN priced risk?, *Journal of Accounting and Economics* 47, 226–243.

Mou, Y., 2011, Limits to arbitrage and commodity index investment: front-running the Goldman Roll, *Working paper* .

Newey, Whitney K, and Kenneth D West, 1987, Hypothesis testing with efficient method of moments estimation, *International Economic Review* 777–787.

Pascual, Roberto, and David Veredas, 2010, Does the Open Limit Order Book Matter in Explaining Informational Volatility?, *Journal of Financial Econometrics* 8, 57–87.

Pastor, L., and R. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 113, 642–685.

Petajisto, A., 2011, The index premium and its hidden cost for index funds, *Journal of Empirical Finance* 18, 271–288.

Raman, V., M. Robe, and P. Yadav, 2016, Liquidity provision under stress: The fast, the slow, and the dead, *Working paper* .

Rosu, I., E. Sojli, and W.W. Tham, 2018, Quoting activity and the cost of capital, *Working paper* .

Saffi, Pedro A.C., and Kari Sigurdsson, 2011, Price efficiency and short selling, *Review of Financial Studies* 24, 821–852.

Saglam, M., 2018, Order anticipation around predictable trades, *Working paper* .

Skjeltorp, J., E. Sojli, and W.W. Tham, 2017, Flashes of trading intent at NASDAQ, *Journal of Financial and Quantitative Analysis* 51, 165–196.

Subrahmanyam, Avanidhar, 1991, Risk Aversion, Market Liquidity, and Price Efficiency, *Review of Financial Studies* 4, 417–441.

Tuzan, T., 2013, Are leveraged and inverse ETFs the new portfolio insurers?, *Working paper* .

West, Mike, and Jeff Harrison, 1997, *Bayesian Forecasting and Dynamic Models*, second edition (Springer-Verlag New York, New York).

Yang, C., B. Zhang, and C. Zhang, 2015, Is information risk priced? evidence from abnormal idiosyncratic volatility, *Working paper* .

Figures and Tables

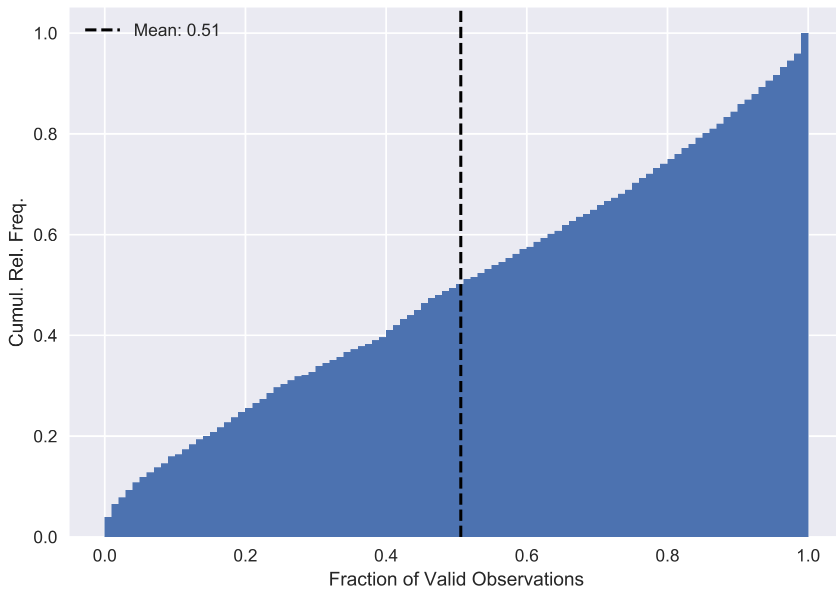
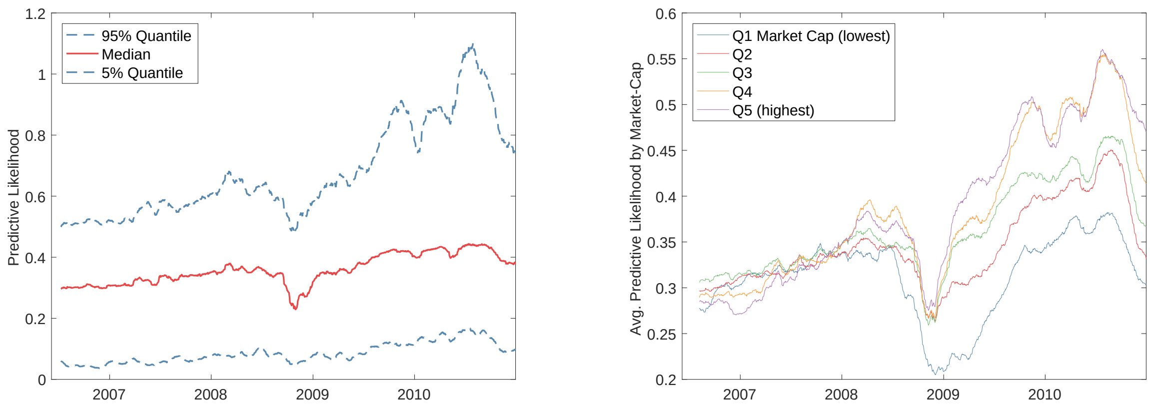


Figure 1. The number of days with a trade by a pension fund.

The figure presents a histogram for the number of observations with a trade from a pension fund in the ANcerno dataset. It shows a fraction of firms in the original dataset (vertical axis) that has a given proportion of trading days (horizontal axis) with a trade in the stock from a pension fund. The sample is from January 1, 2006 to December 31, 2010 (1,144 trading days in total) and 1,560 firms.

Panel A: Predictive Likelihood of Order Imbalances from Pension Funds



Panel B: Predictive Likelihood of Total Order Imbalances

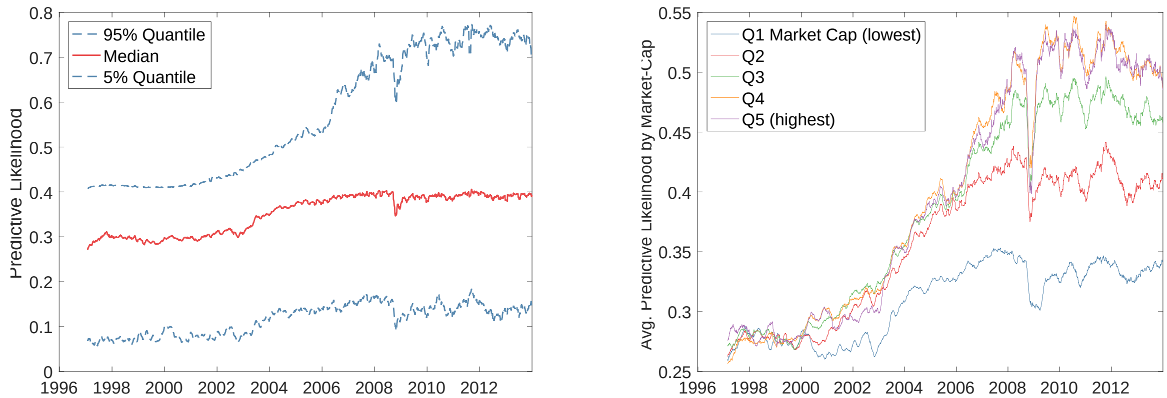


Figure 2. Predictive likelihoods

This figure shows the daily estimates of the smoothed predictive likelihood as outlined in Section (3.1.1). Top panel shows the results for the order imbalance of those stocks traded by active pension funds, i.e., the ANcerno sample, whereas the bottom panel shows the results for the total order imbalances. The left column of both panels report the 5%, 50% (median) and 95% quintiles across the daily estimates at each time t. The right column shows the average predictive likelihood for stocks sorted into quintiles based on the market capitalisation at each time t. The Ancerno sample is from January 1, 2006 to December 31, 2010 (Panel A) and the full sample of stocks is from January 1, 1997 to December 31, 2013 (Panel B).

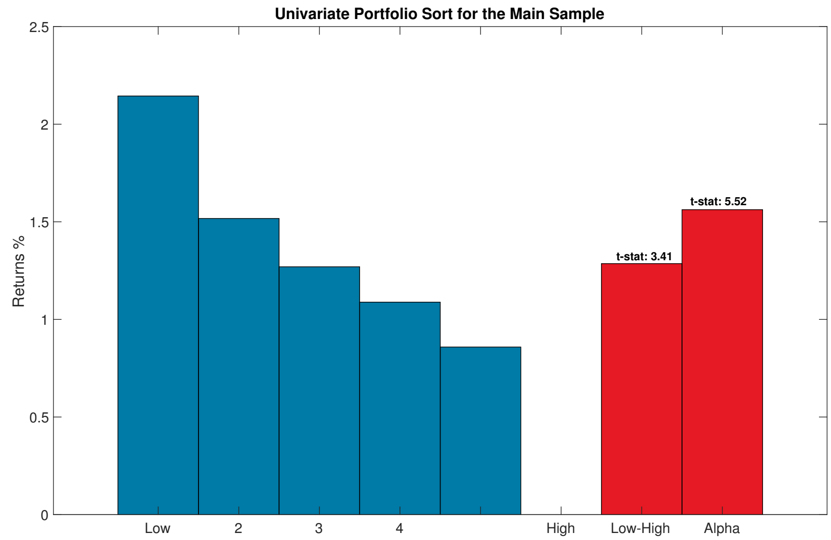


Figure 3. Average returns on portfolios of stocks sorted on pl_t^{tot}

This figure shows the average returns on portfolios of stocks sorted based on the lagged predictive likelihood of total order imbalance pl_t^{tot}. Every month we sort stocks into quintiles, form five portfolios and rebalance next month. Portfolio 1 corresponds to the quintile of stocks with the lowest predictability of order imbalance pl_t^{tot} and portfolio 5 is for stocks with the highest predictability of order imbalance. The sample period is from January 1, 1997 to December 31, 2013. The red bars report the returns on the low-minus-high portfolio and the Jensen's alpha obtained once controlling for the three Fama-French factors, momentum and liquidity as in Pastor and Stambaugh (2003).

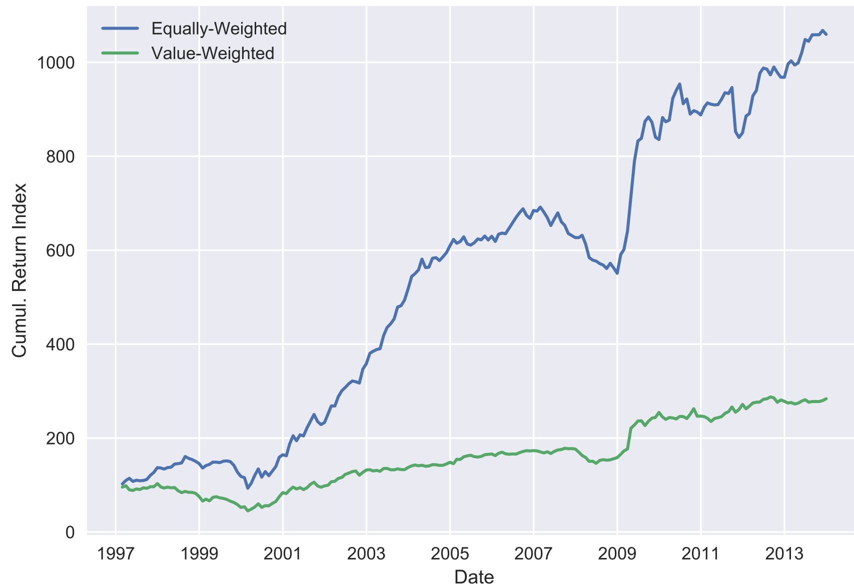


Figure 4. Cumulative returns on the POF portfolio.

This figure presents time-series of cumulative returns on equally and value weighted portfolios of stocks sorted based on the lagged predictive likelihood of total order imbalance pl_t^{tot}. Every month we sort stocks into quintiles, form five portfolios and rebalance next month. The returns are generated long-short strategy which goes long on the portfolio of stocks with the lowest level of predictability and short the portfolio of stocks with the highest level of predictability. The blue line shows the cumulative returns when portfolios are equally weighted whereas the green line shows the cumulative returns when portfolios are value weighted. The sample period is from January 1, 1997 to December 31, 2013.

Table 1. Summary statistics

This table presents means and standard deviations (in squared brackets) for the variables used in this study. We report the summary statistics both for the full sample and across different quintiles of market capitalization. **Panel A:** summarizes the variables for the sample of active pension funds for the period from January 1, 2006 to December 31, 2010. **Panel B:** reports the summary statistics for the full sample of stocks for the period from January 1, 1997 to December 31, 2013. The labels *espread*, *qspread* and *rspread* denote effective, quoted and realized spreads respectively. *prcimpact* is the 5 minutes price impact, *vratio* is the absolute value of the variance ratio minus one, $abs\_ac = |(1+ac)/(1-ac)|$, where ac is the autocorrelation of daily returns, oib^{pf} is the order imbalance of pension funds, and oib^{pf} autocorr is the first order autocorrelation in the pension funds order imbalance. oib^{tot} and oib^{tot} autocorr are total order imbalance and its autocorrelation coefficient. The variables pl^{pf} and pl^{tot} stand for the cross-sectional average of the predictive likelihoods of pension and total order imbalances, respectively. The predictive likelihoods pl^{pf}_{it} is defined as the 20-day rolling average over $\tilde{pl^{pf}}_{it} = \phi\big(oib^{pf}_{i,t} \mid oib^{pf}_{i,t|t-1}, \sigma^2(oib^{pf}_{i,t|t-1})\big)$, where $\phi(\mu, \sigma^2)$ denotes the Normal distribution density function. Here $oib^{pf}_{i,t|t-1}$ denote the out-of-sample forecast for stock i on day t, which is traded by an active pension fund, using data up to day $t-1$ and $\sigma^2(oib^{pf}_{i,t|t-1})$ is the associated forecast variance. pl^{tot}_{it} is defined analogously using oib^{tot}_{it} instead of oib^{pf}_{it}.

Panel A: Active Pension Funds

	Full Sample	Q1 (small)	Q2	Q3	Q4	Q5 (large)
espread (*bp*)	8.37 [4.36]	21.96 [9.72]	11.65 [4.28]	7.82 [2.87]	5.92 [2.22]	4.04 [1.68]
qspread (*bp*)	13.91 [7.63]	39.85 [18.56]	21.36 [8.93]	13.11 [5.19]	8.64 [3.36]	4.93 [1.93]
rspread (*bp*)	2.61 [3.87]	7.55 [9.93]	3.52 [4.98]	2.34 [3.28]	1.80 [2.45]	1.31 [1.81]
prcimpact (*bp*)	5.64 [4.55]	14.24 [10.62]	8.02 [5.75]	5.37 [3.70]	4.01 [2.87]	2.63 [1.79]
vratio (%)	13.70 [9.18]	15.19 [9.69]	13.76 [9.69]	13.42 [9.39]	13.57 [9.33]	13.61 [9.44]
abs_ac (%)	35.39 [24.01]	38.70 [25.29]	35.57 [25.29]	34.59 [24.73]	35.13 [24.70]	35.32 [25.47]
oib^{pf} (%)	0.01 [6.84]	0.01 [11.12]	0.10 [7.90]	0.07 [6.04]	-0.10 [4.79]	-0.05 [3.70]
oib^{pf} acorr (%)	32.56 [12.71]	31.90 [34.17]	32.71 [22.29]	26.75 [23.80]	23.28 [23.46]	20.69 [21.06]
pl^{pf} (%)	36.62 [9.79]	31.32 [6.09]	34.41 [6.70]	36.26 [6.97]	38.58 [8.55]	38.78 [10.05]
Nr. obs.	780,605	59,141	161,021	201,805	174,132	184,506
Nr. firms	789					

Panel B: Full Sample of Stocks

	Full Sample	Q1 (small)	Q2	Q3	Q4	Q5 (large)
espread (*bp*)	77.11 [82.47]	140.4 [112.1]	38.74 [28.08]	23.62 [16.53]	15.82 [11.57]	9.10 [7.37]
qspread (*bp*)	109.4 [101.6]	199.5 [137.9]	55.52 [33.71]	33.22 [18.91]	21.85 [12.31]	11.86 [7.46]
rspread (*bp*)	45.75 [328.9]	89.42 [392.9]	16.11 [238.3]	8.03 [154.8]	5.24 [97.21]	3.62 [42.18]
prcimpact (*bp*)	32.05 [233.5]	52.29 [280.1]	22.49 [170.0]	15.94 [113.7]	10.78 [71.31]	5.89 [26.46]
vratio (%)	15.89 [11.86]	17.88 [12.09]	15.06 [10.48]	14.19 [9.88]	13.68 [9.74]	13.40 [9.86]
abs_ac (%)	39.58 [30.01]	42.98 [30.64]	38.29 [27.00]	36.61 [25.68]	35.67 [25.18]	35.28 [25.33]
oib^{tot} (%)	-1.87 [34.83]	-6.56 [41.02]	-0.24 [26.58]	2.17 [21.61]	3.54 [18.63]	4.34 [14.76]
oib^{tot} acorr (%)	15.41 [6.73]	13.71 [9.75]	11.22 [15.44]	11.81 [15.95]	13.17 [16.76]	16.76 [17.22]
pl^{tot} (%)	34.73 [11.40]	31.15 [10.46]	35.46 [10.12]	37.87 [10.63]	39.55 [11.35]	39.54 [12.34]
Nr. obs.	13,029,155	5,917,088	2,478,330	1,772,525	1,510,898	1,350,314
Nr. firms	6,122					

Table 2. Pairwise Correlations of the Variables of Interest

This table presents the pairwise sample correlation for each predictors used in the main empirical analysis. We report the correlations both for the full sample and across different quintiles of market capitalization. **Panel A:** summarizes the variables for the sample of active pension funds for the period from January 1, 2006 to December 31, 2010. **Panel B:** reports the summary statistics for the full sample of stocks for the period from January 1, 1997 to December 31, 2013. The labels *espread*, *qspread* and *rspread* denote effective, quoted and realized spreads respectively. *prcimpact* is the 5 minutes price impact, *vratio* is the absolute value of the variance ratio minus one, $abs\_ac = |(1+ac)/(1-ac)|$, where ac is the autocorrelation of daily returns, oib^{tot} is total order imbalance, ttq_{it} is the trade-to-quote ratio, $rvar_t$ the daily realized variance, ret_t the realized returns, and vol_t the daily realized volume. The variables pl^{pf} and pl^{tot} stand for the cross-sectional average of the predictive likelihoods of pension and total order imbalances, respectively. The predictive likelihoods pl^{pf}_{it} is defined as the 20-day rolling average over $\widetilde{pl^{pf}}_{it} = \phi\big(oib^{pf}_{i,t} \mid oib^{pf}_{i,t|t-1}, \sigma^2(oib^{pf}_{i,t|t-1})\big)$, where $\phi(\mu,\sigma^2)$ denotes the Normal distribution density function. Here $oib^{pf}_{i,t|t-1}$ denote the out-of-sample forecast for stock i on day t, which is traded by an active pension fund, using data up to day $t-1$ and $\sigma^2(oib^{pf}_{i,t|t-1})$ is the associated forecast variance. pl^{tot}_{it} is defined analogously using oib^{tot}_{it} instead of oib^{pf}_{it}.

Panel A: Active Pension Funds

	$espread$	oib^{tot}	$prcimpact$	$rspread$	pl^{pf}	$qspread$	ret_t	ttq	vol_t	abs_ac	$vratio$	$rvar_t$
$espread$	1.00	-0.05	0.36	0.37	-0.37	0.84	-0.02	0.21	-0.13	0.03	0.05	0.23
oib^{tot}	-0.05	1.00	0.01	-0.04	0.01	-0.04	0.25	-0.01	0.01	-0.01	-0.01	0.01
$prcimpact$	0.36	0.01	1.00	-0.51	-0.14	0.31	0.00	0.07	-0.02	0.00	0.01	0.14
$rspread$	0.37	-0.04	-0.51	1.00	-0.15	0.31	-0.01	0.09	-0.07	0.02	0.02	0.05
pl^{pf}	-0.37	0.01	-0.14	-0.15	1.00	-0.38	-0.01	-0.24	0.18	-0.05	-0.07	-0.04
$qspread$	0.84	-0.04	0.31	0.31	-0.38	1.00	-0.01	0.17	-0.17	0.03	0.05	0.27
ret_t	-0.02	0.25	0.00	-0.01	-0.01	-0.01	1.00	0.02	0.07	-0.01	0.01	0.01
ttq	0.21	-0.01	0.07	0.09	-0.24	0.17	0.02	1.00	0.10	0.00	0.02	-0.04
vol_t	-0.13	0.01	-0.02	-0.07	0.18	-0.17	0.07	0.10	1.00	-0.04	0.00	0.10
abs_ac	0.03	-0.01	0.00	0.02	-0.05	0.03	-0.01	0.00	-0.04	1.00	0.41	0.00
$vratio$	0.05	-0.01	0.01	0.02	-0.07	0.05	0.01	0.02	0.00	0.41	1.00	0.02
$rvar_t$	0.23	0.01	0.14	0.05	-0.04	0.27	0.01	-0.04	0.10	0.00	0.02	1.00

Panel B: Full Sample of Stocks

	$espread$	oib^{tot}	$prcimpact$	$rspread$	pl^{tot}	$qspread$	ret_t	ttq	vol_t	abs_ac	$vratio$	$rvar_t$
$espread$	1.00	-0.04	0.56	0.35	-0.21	0.78	-0.03	0.01	0.16	0.01	0.02	0.41
oib^{tot}	-0.04	1.00	-0.01	-0.03	-0.05	-0.05	0.14	0.04	-0.05	-0.01	-0.01	-0.02
$prcimpact$	0.56	-0.01	1.00	-0.53	-0.11	0.51	0.00	0.01	0.14	-0.01	0.01	0.29
$rspread$	0.35	-0.03	-0.53	1.00	-0.09	0.21	-0.03	0.01	0.00	0.02	0.01	0.11
pl^{tot}	-0.21	-0.05	-0.11	-0.09	1.00	-0.19	0.00	-0.28	0.00	-0.01	-0.02	-0.12
$qspread$	0.78	-0.05	0.51	0.21	-0.19	1.00	-0.03	-0.07	0.07	0.01	0.03	0.46
ret_t	-0.03	0.14	0.00	-0.03	0.00	-0.03	1.00	0.03	0.04	-0.01	0.01	-0.01
ttq	0.01	0.04	0.01	0.01	-0.28	-0.07	0.03	1.00	0.40	-0.01	0.01	0.04
vol_t	0.16	-0.05	0.14	0.00	0.00	0.07	0.04	0.40	1.00	-0.02	0.04	0.20
abs_ac	0.01	-0.01	-0.01	0.02	-0.01	0.01	-0.01	-0.01	-0.02	1.00	0.39	0.00
$vratio$	0.02	-0.01	0.01	0.01	-0.02	0.03	0.01	0.01	0.04	0.39	1.00	0.03
$rvar_t$	0.41	-0.02	0.29	0.11	-0.12	0.46	-0.01	0.04	0.20	0.00	0.03	1.00

Table 3. Relation between total and uninformed order imbalance predictability

This table reports the estimation results from the following regression:

$$pl_{it}^{tot} = \alpha_i + \omega_t + \beta pl_{it}^{pf} + u_{it},$$

where pl_{it}^{tot} and pl_{it}^{pf} are predictive likelihoods of total and pension funds order imbalance respectively, and α_i, ω_t contain firm and time fixed effects, respectively. The t-statistics are given in parentheses are based on double-clustered standard errors. The sample is from period from January 1, 2006 to December 31, 2010.

	pl_t^{tot}	pl_t^{tot}	pl_t^{tot}	pl_t^{tot}
pl_t^{pf}	0.196 (9.67)	0.219 (18.5)	0.150 (6.13)	0.1339(11.431)
const	0.398 (48.3)			
fixed effects	none	firm	time	firm, time
R^2	3.59%	29.30%	20.71%	18.62%
Nr. obs.	710,633	710,633	710,633	710,633

Table 4. Predictability of order imbalance and market liquidity.

This table reports the results from the regression:

$$illiq_{it} = \alpha_i + \beta pl_{i,t-1} + \boldsymbol{\gamma}'\boldsymbol{X}_{it} + \epsilon_{it},$$

where $illiq_{it}$ takes one of following illiquidity variables (expressed in basis points): quoted spread *qspread*, effective spread *espread*, realized spread *rspread*, and price impact *prcimpact*. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(3). **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($obi_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	qspread	*espread*	*rspread*	*prcimpact*
pl_{t-1}	-0.057 (-11.5)	-0.028 (-7.41)	-0.022 (-11.3)	-0.023 (-9.91)
$illiq_{t-1}$	0.574 (23.7)	0.556 (7.62)	0.210 (8.18)	0.271 (11.7)
oib_t	0.001 (0.927)	-0.001 (-0.608)	-0.005 (-3.81)	0.006 (5.361)
lsize	-0.017 (-6.01)	-0.011 (-4.68)	0.003 (1.32)	-0.020 (-10.8)
turn	-0.069 (-2.20)	0.106 (4.19)	-0.057 (-2.42)	0.162 (4.53)
$1/prc$	0.229 (4.32)	0.242 (3.86)	0.232 (6.53)	0.169 (4.19)
ttq	-14.72 (-11.2)	-3.702 (-5.06)	-2.013 (-3.46)	-1.114 (-1.64)
vix	0.094 (12.0)	0.039 (5.96)	-0.007 (-2.98)	0.072 (22.3)
R^2	54.96%	65.71%	16.67%	32.37%
No. obs.	764,825	764,825	764,825	764,825

Panel B: Full Sample of Stocks

	qspread	*espread*	*rspread*	*prcimpact*
pl_{t-1}	-0.263 (-11.7)	-0.263 (-11.7)	-0.163 (-3.56)	-0.485 (-17.1)
$illiq_{t-1}$	0.688 (171.4)	0.579 (125.5)	0.047 (22.9)	0.036 (21.3)
oib_t	-0.052 (-22.12)	-0.073 (-25.77)	-0.086 (-5.99)	-0.006 (-0.607)
lsize	-0.168 (-31.9)	-0.159 (-31.4)	-0.262 (-24.9)	-0.097 (-19.5)
turn	-1.542 (-9.87)	-1.182 (-9.75)	-2.474 (-9.01)	-0.014 (-0.24)
$1/prc$	0.559 (21.2)	0.611 (22.9)	1.174 (20.7)	0.371 (14.6)
ttq	1.722 (5.67)	7.342 (18.9)	16.46 (9.51)	1.433 (1.24)
vix	0.543 (17.49)	0.485 (19.34)	0.464 (11.61)	0.649 (28.07)
R^2	59.84%	49.50%	1.04%	0.42%
No. obs.	12,906,715	12,906,715	12,906,715	12,906,715

Table 5. Predictability of order imbalance and market efficiency.

This table reports the results from the regression:

$$ineff_{it} = \alpha_i + \beta pl_{i,t-1} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it},$$

where $ineff_{it}$ corresponds to one of the two market efficiency variables $vratio_{it}$ and abs_ac_{it}. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(3). **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), and the contemporaneous order imbalance ($oib_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	vratio	*abs_ac*
pl_{t-1}	-0.020 (-3.92)	-0.037 (-3.08)
oib_t	-0.002 (-0.97)	-0.002 (-0.47)
lsize	0.009 (3.80)	0.010 (2.17)
turn	0.245 (7.02)	-0.456 (-6.38)
$1/prc$	0.066 (3.88)	0.133 (5.36)
ttq	-0.410 (-0.40)	-4.222 (-1.76)
vix	0.024 (3.28)	0.040 (2.13)
R^2	0.32%	0.18%
No. obs.	764,825	764,825

Panel B: Full Sample of Stocks

	vratio	*abs_ac*
pl_{t-1}	-0.048 (-14.4)	-0.086 (-11.4)
oib_t	-0.002 (-8.60)	-0.006 (-11.6)
lsize	-0.008 (-18.9)	-0.014 (-16.2)
turn	0.054 (6.45)	-0.440 (-8.35)
$1/prc$	0.013 (8.48)	0.028 (9.98)
ttq	-0.032 (-0.89)	-0.095 (-1.25)
vix	0.035 (7.33)	0.050 (3.83)
R^2	0.85%	0.49%
No. obs.	12,906,715	12,906,715

Table 6. Aggregate market uncertainty and the effect of order imbalance predictability.

This table reports the results of the following regression:

$$y_{it} = \alpha_i + \beta_{high} \cdot pl_{i,t-1} \cdot D_t^{high} + \beta_{med} \cdot pl_{i,t-1} \cdot D_t^{med} + \beta_{low} \cdot pl_{i,t-1} \cdot D_t^{low} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it},$$

where y_{it} represents either one of the illiquidity measures $illiq_{it}$ or one on the proxies used for market inefficiency $ineff_{it}$, and α_i represents a firm fixed effect. For the ease of exposition we present the results for a measure of illiquidity and a measure of inefficiency, namely the quoted spread $qspread_{it}$ and the variance ration $vratio_{it}$. We define a dummy variable D_t^{high} that takes value one whenever vix_t falls within the upper tercile of its historical distribution drawn from our 1997-2013 sample and zero otherwise. Similarly, dummy variables D_t^{med} and D_t^{low} take the value one when vix_t is in the middle and bottom terciles, respectively. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(3). **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($obi_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	qspread	*vratio*
$pl_{t-1}^{pf} \times D_t^{high}$	-0.065 (-10.2)	-0.018 (-2.72)
$pl_{t-1}^{pf} \times D_t^{med}$	-0.054 (-11.2)	-0.022 (-4.26)
$pl_{t-1}^{pf} \times D_t^{low}$	-0.050 (-7.59)	-0.032 (-3.72)
oib_t	0.001 (0.79)	-0.001 (-0.53)
$illiq_{t-1}$	0.573 (23.8)	
$lsize$	-0.018 (-5.85)	0.009 (3.82)
$turn$	-0.064 (-1.87)	0.237 (6.82)
$1/prc$	0.228 (4.29)	0.066 (3.85)
ttq	-14.80 (-10.7)	-0.190 (-0.18)
vix	0.105 (8.87)	0.017 (1.82)
No. obs.	764,825	764,825
R^2	54.97%	0.33%

Panel B: Full Sample of Stocks

	qspread	*vratio*
$pl_{t-1}^{pf} \times D_t^{high}$	-0.427 (-15.3)	-0.032 (-7.11)
$pl_{t-1}^{pf} \times D_t^{med}$	-0.223 (-9.79)	-0.052 (-16.1)
$pl_{t-1}^{pf} \times D_t^{low}$	-0.179 (-6.74)	-0.053 (-14.6)
oib_t	-0.052 (-21.9)	-0.002 (-8.79)
$illiq_{t-1}$	0.680 (171.3)	
$lsize$	-0.170 (-32.2)	-0.008 (-18.0)
$turn$	-1.534 (-9.87)	0.052 (6.40)
$1/prc$	0.555 (21.0)	0.014 (8.75)
ttq	1.577 (5.17)	-0.010 (-0.28)
vix	0.836 (14.1)	0.010 (1.561)
No. obs.	12,906,715	12,906,715
R^2	59.85%	0.88%

Table 7. The effect of order imbalance predictability on liquidity by stock characteristics.

This table presents the results of the following regression:

$$qspread_{it} = \alpha_i + \beta pl_{i,t-1} + \boldsymbol{\gamma}'\boldsymbol{X}_{it} + \epsilon_{it},$$

where $qspread_{it}$ is the quoted bid-ask spread and pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(3). We report the estimates $\hat{\beta}$ by sorting stocks in quintiles based on the following characteristics: market cap, trade-to-quote ratio, idiosyncratic volatility, and passive institutional ownership. **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), and the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($obi_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors. Results for analogous regression using the alternative liquidity measures *espread*, *rspread* and *prcimpact* are reported in Table A3.

Panel A: Active Pension Funds

Sorting variable	Q1	Q2	Q3	Q4	Q5
Market cap	-0.149 (-3.98)	-0.069 (-7.39)	-0.047 (-9.81)	-0.023 (-9.51)	-0.008 (-6.48)
Trade-to-quote ratio	-0.072 (-7.20)	-0.055 (-7.38)	-0.046 (-9.80)	-0.047 (-11.4)	-0.051 (-5.47)
Idiosyncratic vol.	-0.022 (-8.57)	-0.041 (-9.95)	-0.063 (-9.94)	-0.073 (-10.0)	-0.079 (-6.26)
Passive inst. ownership	-0.090 (-7.28)	-0.044 (-7.83)	-0.048 (-6.95)	-0.039 (-8.86)	-0.046 (-7.53)

Panel B: Full Sample of Stocks

Sorting variable	Q1	Q2	Q3	Q4	Q5
Market cap	-1.044 (-13.6)	-0.354 (-13.9)	-0.119 (-12.8)	-0.084 (-7.34)	-0.028 (-14.3)
Trade-to-quote ratio	-0.702 (-14.6)	-0.383 (-14.6)	-0.354 (-14.3)	-0.406 (-13.7)	-0.417 (-8.41)
Idiosyncratic vol.	-0.201 (-18.1)	-0.300 (-18.1)	-0.429 (-15.3)	-0.712 (-16.1)	-1.563 (-16.0)
Passive inst. ownership	-1.069 (-8.93)	-0.575 (-11.0)	-0.289 (-11.3)	-0.257 (-13.7)	-0.273 (-12.9)

Table 8. The effect of order imbalance predictability on efficiency by stock characteristics.

This table presents the results of the following regression:

$$vratio_{it} = \alpha_i + \beta pl_{i,t-1} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it},$$

where the variance ratio $vratio_{it}$ is calculated as shown in Section (3) and pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(3). We report the estimates $\hat{\beta}$ by sorting stocks in quintiles based on the following characteristics: market cap, trade-to-quote ratio, idiosyncratic volatility, and passive institutional ownership. **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), and the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($obi_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors. Results using the alternative measure of price efficiency abs_ac are detailed in Table A3.

Panel A: Active Pension Funds

Sorting variable:	Q1	Q2	Q3	Q4	Q5
Market cap	-0.016 (-0.66)	-0.028 (-1.88)	0.015 (0.13)	-0.026 (-3.01)	-0.011 (-1.19)
Trade-to-quote ratio	-0.044 (-3.62)	-0.017 (-2.03)	-0.017 (-1.98)	-0.019 (-1.91)	-0.009 (-0.72)
Idiosyncratic vol.	-0.033 (-3.31)	-0.021 (-2.27)	-0.029 (-3.54)	-0.020 (-2.09)	-0.007 (-0.75)
Passive inst. ownership	0.001 (0.08)	-0.026 (-2.33)	-0.039 (-2.63)	-0.014 (-1.21)	-0.018 (-1.27)

Panel B: Full Sample of Stocks

Sorting variable:	Q1	Q2	Q3	Q4	Q5
Market cap	-0.170 (-19.1)	-0.060 (-8.31)	-0.040 (-7.30)	-0.026 (-5.11)	-0.004 (-0.79)
Trade-to-quote ratio	-0.066 (-8.16)	-0.038 (-8.40)	-0.045 (-9.63)	-0.059 (-11.8)	-0.072 (-10.8)
Idiosyncratic vol.	-0.031 (-6.31)	-0.043 (-9.63)	-0.047 (-10.3)	-0.067 (-13.2)	-0.133 (-16.1)
Passive inst. ownership	-0.191 (-13.7)	-0.076 (-8.88)	-0.039 (-6.44)	-0.028 (-5.07)	-0.029 (-4.98)

Table 9. Characteristics of stocks sorted by the predictability of order imbalance.

This table reports the average value of the characteristics of stocks sorted on quintiles based on the predictability of order imbalances as proxied by pl_{it}^{tot}. The predictive likelihood for the full sample of stocks is defined as the 20-day rolling average over $\tilde{pl}_{it}^{tot}$ which is calculated as in Eq.(3). Here, MCap is the market capitalisation of the firm i (in million of dollars), BE/ME is the book-to-market ratio, ret_t and $\sigma(ret_t)$ are the monthly return and its standard deviation, $qspread$ is the quoted spread, ttq denotes the trade-to-quote ratio, $voib$ is the volatility of order imbalance, and $turn_{it}$ is the share turnover (in percent). Sample period is from January 1, 1997 to December 31, 2013.

Quint. of pl_t^{tot}	MCap	BE/ME	ret_t	$\sigma(ret_t)$	$qspread$	ttq	$voib$	$turn$
Q1 (low)	1,756	0.93	1.87	3.07	212.38	0.32	0.43	0.38
Q2	2,565	0.79	1.65	2.90	129.06	0.34	0.32	0.54
Q3	3,373	0.70	1.59	2.89	89.98	0.36	0.26	0.67
Q4	5,643	0.62	1.55	2.80	62.79	0.40	0.21	0.81
Q5 (high)	6,310	0.54	1.58	2.92	41.28	0.47	0.15	1.13

Table 10. Double sorted portfolio returns.

This table reports the returns on portfolios sorted on the lagged value of a given stock characteristic and the past predictability of order imbalance for the full sample of stocks. The latter is defined as the 20-day rolling average over $\tilde{pl}_{it}^{tot}$ which is calculated as in Eq.(3). The characteristics used for double sorting are the market capitalisation (Panel A), the quoted bid-ask spread (Panel B), share turnover (Panel C), the trade-to-quote ratio (Panel D), the volatility of order imbalance (Panel E), and order imbalance (Panel F). We first sort stocks into quintile portfolios based on a characteristic, and then sort on pl_{it}^{tot} into quintile portfolios in each group at month t. Portfolio returns and return differences (Low-High) in month $t+1$ are reported. All returns are reported in percent. The sample period is from January 1, 1997 to December 31, 2013. The t-statistics for the low-minus-high predictability portfolio are based on Newey-West adjusted standard errors. The Jensen's alphas are calculated conditioning on the three Fama and French (1993) factors along with the momentum factor of Jegadeesh and Titman (1993) and the Pastor and Stambaugh (2003) traded liquidity factor.

Panel A: Double sort by order imbalance predictability and market cap

	Quintile of $lsize$				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	2.785	1.735	1.394	1.175	0.949
Q2	1.956	1.196	1.220	0.993	0.903
Q3	1.720	1.174	1.170	0.969	0.828
Q4	1.363	1.098	0.765	0.888	0.913
Q5 (high)	1.146	1.043	0.721	0.725	0.691
Low-High	1.639 (4.47)	0.691 (1.84)	0.673 (1.97)	0.450 (1.35)	0.259 (0.71)
Alpha	1.922 (4.64)	1.025 (2.48)	1.023 (2.82)	0.730 (2.05)	0.501 (1.45)

Panel B: Double sort by order imbalance predictability and quoted bid-ask spread

	Quintile of $qspread$				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	0.983	1.418	1.886	2.873	3.216
Q2	1.006	1.178	1.324	1.812	2.009
Q3	0.916	1.020	1.234	1.518	1.810
Q4	0.896	1.078	1.220	1.229	1.485
Q5 (high)	0.832	0.804	1.048	1.196	1.033
Low-High	0.151 (0.46)	0.614 (1.56)	0.838 (2.04)	1.677 (3.58)	2.183 (5.50)
Alpha	0.392 (1.28)	0.975 (2.42)	1.139 (2.48)	2.076 (4.24)	2.369 (5.53)

Panel C: Double sort by order imbalance predictability and trade-to-quote ratio

	Quintile of ttq				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	1.641	1.951	2.028	2.253	3.327
Q2	1.040	1.280	1.487	1.695	2.280
Q3	0.926	1.039	1.128	1.658	1.674
Q4	0.943	0.967	1.193	1.178	1.190
Q5 (high)	0.796	1.025	0.818	0.900	0.589
Low-High	0.845 (3.30)	0.926 (3.26)	1.210 (3.39)	1.353 (3.55)	2.74 (5.50)
Alpha	1.922 (3.82)	1.215 (4.01)	1.366 (3.94)	1.641 (4.29)	3.00 (5.85)

Table 10 continued.

Panel D: Double sort by order imbalance predictability and volatility of order imbalance

	Quintile of *voib*				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	1.142	1.753	2.402	2.802	2.621
Q2	1.122	1.466	1.699	1.799	1.650
Q3	1.131	1.160	1.397	1.623	1.291
Q4	0.890	1.061	1.153	1.048	1.032
Q5 (high)	0.650	0.874	1.144	1.136	0.880
Low-High	0.492 (1.19)	0.879 (2.58)	1.257 (3.35)	1.667 (3.80)	1.741 (5.00)
Alpha	0.834 (1.93)	1.188 (3.07)	1.407 (3.88)	1.856 (4.25)	1.835 (5.67)

Panel E: Double sort by order imbalance predictability and share turnover

	Quintile of *turn*				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	1.831	2.719	2.485	2.469	2.631
Q2	1.049	1.705	1.661	1.606	1.832
Q3	0.903	1.157	1.442	1.349	1.249
Q4	0.557	0.915	0.971	1.247	1.098
Q5 (high)	0.759	0.903	1.049	0.663	0.641
Low-High	1.027 (3.706)	1.816 (3.57)	1.436 (3.67)	1.806 (4.46)	1.990 (4.65)
Alpha	1.166 (4.35)	1.968 (3.97)	1.559 (4.11)	1.984 (5.04)	2.176 (5.29)

Panel F: Double sort by order imbalance predictability and order imbalance

	Quintile of *oib*				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	2.439	2.462	2.078	1.872	1.787
Q2	1.813	1.672	1.422	1.441	1.324
Q3	1.438	1.664	1.137	1.137	1.076
Q4	1.346	1.379	1.015	1.037	0.927
Q5 (high)	1.041	1.145	0.838	0.770	0.750
Low-High	1.398 (4.56)	1.317 (2.92)	1.240 (2.96)	1.102 (2.48)	1.036 (3.70)
Alpha	1.165 (4.85)	1.648 (3.52)	1.615 (3.91)	1.366 (3.28)	1.265 (4.81)

Table 11. Betas on the order imbalance predictability factor.

This table reports the betas on the returns on the POF factor calculated from the following regression

$$returns_{i,t} = \alpha_i + \beta_i POF_t + \boldsymbol{\gamma}_i' \boldsymbol{F}_t + u_{i,t}, \qquad i = 1, \ldots, 25$$

where $\boldsymbol{F}_t = (Mkt_t, SMB_t, HML_t, MOM_t, LIQ_t)'$ contains the excess returns on the market, the size and value factors of Fama and French (1993), the momentum factor calculated as in Jegadeesh and Titman (1993), the liquidity risk factor from Pastor and Stambaugh (2003), as well as the short-term reversals factor and the returns on high-minus low portfolios sorted by past month trade-to-quote ratio and volatility of order imbalance. The returns on the 5×5 test assets are constructed by sorting on both size and the predictability of order imbalance. The POF risk factor is defined as monthly returns on the low minus high portfolio sorted by predictive likelihood of total order imbalance pl_{it}^{tot}. We report the estimates $\hat{\beta}_i$ for each of these test portfolios. **Panel A:** reports the results for test assets which are constructed as equally-weighted portfolios. **Panel B:** reports the results for value-weighted test portfolios. The sample period is from January 1, 1997 to December 31, 2013. The t-statistics are based on Newey-West adjusted standard errors. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Equally-Weighted Portfolios

Decile of pl_t^{tot}	Q1 (small)	Q2	Q3	Q4	Q5 (large)
Q1 (low)	0.662***	0.310***	0.460***	0.556***	0.236***
Q2	0.320***	0.212***	0.208***	0.271***	0.236***
Q3	0.147**	0.073	0.072	0.139**	0.142***
Q4	-0.047	-0.137***	-0.022	0.024	0.021
Q5 (high)	-0.345***	-0.437***	-0.349***	-0.209***	-0.240***

Panel B: Value-Weighted Portfolios

Decile of pl_t^{tot}	Q1 (small)	Q2	Q3	Q4	Q5 (large)
Q1 (low)	0.507***	0.305***	0.343***	0.493***	0.159**
Q2	0.240***	0.204***	0.180***	0.222***	0.189***
Q3	0.110**	0.085*	0.085	0.126**	0.124***
Q4	-0.095*	-0.125**	-0.015	0.027	-0.002
Q5 (high)	-0.376***	-0.433***	-0.325***	-0.208***	-0.239***

Table 12. Fama-MacBeth Regressions.

This table reports the estimates of the following cross-sectional regression

$$returns_{i,t} = \lambda_{POF,t}\hat{\beta}_i + \boldsymbol{\lambda}_t'\hat{\boldsymbol{\gamma}}_i + e_{i,t}, \qquad t = 1, \ldots, T$$

where the parameters $\hat{\beta}_i, \hat{\boldsymbol{\gamma}}_i$ are estimated for each test portfolio by using the time series regression in Eq.(8). The estimates $\hat{\lambda}_{POF}, \hat{\boldsymbol{\lambda}}'$ are obtained as the sample averages of $\hat{\lambda}_{POF,t}, \hat{\boldsymbol{\lambda}}_t'$. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(3). **Panel A:** reports the results for test assets which are constructed as equally-weighted portfolios. **Panel B:** reports the results for value-weighted test portfolios. The sample period is from January 1, 1997 to December 31, 2013. The sample period is from January 1, 1997 to December 31, 2013. The t-statistics (in parentheses) are based on Newey-West adjusted standard errors.

Panel A: Equally-Weighted Portfolios

POF	1.206 (2.97)	1.185 (3.48)	1.145 (2.88)	1.182 (3.30)	1.194 (3.29)	1.181 (2.958)	1.299 (2.96)
Mkt	0.603 (1.56)	0.602 (1.64)	0.565 (1.46)	0.620 (1.63)	0.591 (1.58)	0.595 (1.531)	0.579 (1.41)
SMB	0.695 (2.14)	0.540 (1.78)	0.647 (2.15)	0.534 (1.83)	0.662 (2.20)	0.560 (2.166)	0.768 (2.85)
HML	-0.425 (-0.91)	-0.396 (-0.92)	-0.417 (-0.88)	-0.330 (-0.80)	-0.534 (-1.23)	-0.381 (-0.916)	-0.474 (-1.162)
MOM	-0.640 (-0.82)						-0.852 (-0.98)
REV		0.273 (0.40)					-1.410 (-1.38)
LIQ			0.960 (0.89)				0.091 (0.14)
$VOIB$				0.711 (2.09)			0.548 (1.37)
TTQ					-0.905 (-1.70)		-0.803 (-1.62)
OIB						-0.718 (-2.634)	-1.009 (-3.08)

Panel B: Value-Weighted Portfolios

POF	1.429 (3.27)	1.436 (3.41)	1.418 (3.37)	1.437 (3.48)	1.430 (3.30)	1.440 (3.53)	1.564 (3.66)
Mkt	0.611 (1.62)	0.606 (1.62)	0.607 (1.61)	0.618 (1.61)	0.609 (1.62)	0.611 (1.62)	0.599 (1.52)
SMB	0.601 (2.09)	0.611 (2.11)	0.602 (2.21)	0.578 (2.12)	0.599 (2.25)	0.566 (2.08)	0.700 (2.48)
HML	-0.485 (-1.03)	-0.49 (-1.12)	-0.497 (-1.14)	-0.459 (-1.16)	-0.497 (-1.34)	-0.457 (-1.14)	-0.561 (-1.20)
MOM	-1.194 (-1.41)						-1.191 (-1.26)
REV		-0.033(-0.04)					-0.141 (-1.09)
LIQ			0.512 (0.62)				-0.244 (-0.32)
$VOIB$				0.824 (2.24)			0.684 (1.52)
TTQ					-1.026 (-1.64)		-0.897 (-1.50)
OIB						-0.860 (-2.50)	-1.072 (-2.48)

Appendix to

"Predictability of Order Imbalance, Market Quality and Equity Cost of Capital"

A A Simple Microstructure Model

We present a simple equilibrium model that guides our empirical analysis. Before describing it formally, we outline its main ingredients and discuss why these are required for our analysis. In our model, one risky security is traded by three types of traders: informed traders, liquidity traders, and a competitive, risk-averse market maker. The model is in spirit of Subrahmanyam (1991) with one substantial difference: order imbalance of liquidity traders is predictable to both informed traders and market makers. Our main goal it to establish the effect of order imbalance predictability on market liquidity and efficiency. In the model, the market maker absorbs the net demand of the other traders and sets the price such that he expects to earn zero utility conditional upon observing the net total market order imbalance. Predictability of uninformed order imbalance has two effects on liquidity. On one hand predictable order imbalance helps the market maker to better estimate the total order imbalance and reduce costly inventory. This effect improves market liquidity. On the other hand it might exacerbate the adverse selection problem leading to deterioration of liquidity. Our model aims to answer which effect and when dominates in the market and helps us to formulate testable hypotheses.

There are three types of agents: an informed trader, uninformed / liquidity traders, and a competitive risk-averse market maker.[13] The informed trader is assumed to be risk-neutral. Market maker has risk aversion coefficient A_m and CARA utility. The agents trade in one risky security and market makers absorb the residual demand of the other types of traders, setting prices such that her expected profit conditional on her signals is zero. Trading taking place at time 0 and liquidation of the security taking place at time 1. The security is liquidated at time 1 for value δ, where δ is a random variable with $\delta \sim \mathcal{N}\left(0, \sigma_\delta^2\right)$. The informed trader observes the fundamental

[13] Introducing a number of competing informed traders as in Subrahmanyam (1991) does not change the results qualitatively so we consider only a single informed trader for simplicity of exposition.

value δ. Uninformed traders submit an order z, $z \sim \mathcal{N}\left(0, \sigma_z^2\right)$, and both the informed trader and the market maker observe a noisy signal of the uninformed order imbalance in form of $z + y$, $y \sim \mathcal{N}\left(0, \sigma_y^2\right)$. The total order imbalance is denoted by ω.

We focus on characterising the unique linear Nash equilibrium. The pricing rule of the market maker has the form

$$p = \lambda_1 \omega + \lambda_2 \left(z + y\right), \tag{A1}$$

where λ_1 and λ_2 are measures of price impact (inverse market depth / liquidity) for the total order imbalance ω and the predictable portion of the order imbalance, respectively. The term $\lambda_2 \left(z + y\right)$ captures the partial price impact originating from the predictability of uninformed order imbalance z.

Let us denote the order of the informed trader as x. The informed trader's profit is

$$x\left(\delta - p\right) = x\left(\delta - \lambda_1 \omega - \lambda_2 \left(z + y\right)\right). \tag{A2}$$

We conjecture the linear demand function of the informed trader in the form

$$x = \beta_1 \delta + \beta_2 \left(z + y\right). \tag{A3}$$

This assumption is the linear extension of the original response function from Subrahmanyam (1991) to account for the impact of order imbalance predictability. Given x, the total order imbalance ω can be written as $\omega = x + z = \beta_1 \delta + \beta_2 \left(z + y\right) + z$.

Given the linear pricing rule and publicly observed signal about the final dividend, the following lemma characterizes the equilibrium price at $t = 2$ and the informed trader's orders.

Proposition 1. The price impact coefficients λ_1 and λ_2 and informed traders' demand response coefficients β_1 and β_2 are the solutions to the following system of equations:

$$\beta_1 = \frac{1}{2\lambda_1}, \tag{A4}$$

$$\beta_2 = -\frac{\sigma_z^2}{2(\sigma_z^2+\sigma_y^2)} + \frac{\lambda_2}{2\lambda_1}, \tag{A5}$$

$$\lambda_1 = -\frac{\lambda_2}{\beta_2 + \sigma_z^2/(\sigma_z^2+\sigma_y^2)} + \frac{A_m}{2} Var[\delta \mid \omega, z+y] \tag{A6}$$

$$\lambda_2 = -\frac{\beta_1 \sigma_\delta^2 (\beta_2 + \sigma_z^2/(\sigma_z^2+\sigma_y^2))}{\beta_1^2\sigma_\delta^2 + \sigma_z^2\sigma_y^2/(\sigma_z^2+\sigma_y^2)}, \tag{A7}$$

where

$$Var[\delta \mid \omega, z+y] = Var[\delta \mid p] = \sigma_\delta^2 \left(1 - 0.5\sqrt{\sigma_\delta^2/Var[p]}\right), \tag{A8}$$

$$Var[p] = \lambda_1^2 \left[\beta_1^2\sigma_\delta^2 + \beta_2^2\left(\sigma_z^2+\sigma_y^2\right) + \sigma_z^2 + 2\beta_2\sigma_z^2\right] + \lambda_2^2\left(\sigma_z^2+\sigma_y^2\right) + 2\lambda_1\lambda_2\beta_2\left(\sigma_z^2+\sigma_y^2\right) \tag{A9}$$

Proof of Proposition 1. The informed trader derives his utility from the profit as

$$\begin{aligned} E\big[U\left(x\left(\delta - p\right)\right) \mid \delta, z+y\big] = E\big[x\left(\delta - \lambda_1\omega - \lambda_2\left(z+y\right)\right) \mid \delta, z+y\big] \\ - \left(\frac{A}{2}\right) Var\big[x\left(\delta - \lambda_1\omega - \lambda_2\left(z+y\right)\right) \mid \delta, z+y\big]. \end{aligned} \tag{A10}$$

Maximizing the expected utility with respect to x yields

$$x = \frac{\delta - \lambda_1 \frac{\sigma_z^2(z+y)}{\sigma_z^2+\sigma_y^2} - \lambda_2(z+y)}{2\lambda_1 + A\lambda_1^2\frac{\sigma_z^2\sigma_y^2}{\sigma_z^2+\sigma_y^2}}. \tag{A11}$$

The second order condition for the maximisation is

$$\lambda_1 > \lambda_1^2 \frac{\sigma_z^2\sigma_y^2}{\sigma_z^2+\sigma_y^2}. \tag{A12}$$

We obtain the unique linear Nash equilibrium by setting $x = \beta_1\delta + \beta_2(z+y)$ in Equation (A11) and solving for β_1 and β_2. We obtain

$$\beta_1 = \frac{1}{\lambda_1 + A\lambda_1^2 \frac{\sigma_z^2\sigma_y^2}{\sigma_z^2+\sigma_y^2}}, \tag{A13}$$

$$\beta_2 = -\frac{\lambda_1 \frac{\sigma_z^2}{\sigma_z^2+\sigma_y^2} + \lambda_2}{\lambda_1 + A\lambda_1^2 \frac{\sigma_z^2\sigma_y^2}{\sigma_z^2+\sigma_y^2}}. \tag{A14}$$

The market maker's objective function is

$$E\big[U_m\left(\omega(p-\delta)\right) \mid \omega, z+y\big] = E\big[\omega(p-\delta) \mid \omega, z+y\big] - \left(\frac{A_m}{2}\right) Var\big[\omega(p-\delta) \mid \omega, z+y\big]. \tag{A15}$$

Following standard procedure we assume Bertrand competition between market makers and, thus, market makers set the prices so that their expected utility equals zero. Rearranging the previous equation we obtain an equation of bivariate regression layout

$$E\left[\delta \mid \omega, z+y\right] = \underbrace{\left(\lambda_1 - \left(\frac{A_m}{2}\right) Var\left[\delta \mid \omega, z+y\right]\right)}_{=:\gamma_1} \omega + \lambda_2(z+y). \tag{A16}$$

Comparing this equation with the general bivariate regression design equations we obtain the coefficients as

$$\gamma_1 = \frac{Var[z+y]\, Cov[\delta,\omega] - Cov[\omega, z+y]\, Cov[z+y,\delta]}{Var[\omega]\, Var[z+y] - Cov^2[\omega, z+y]} \tag{A17}$$

$$\lambda_2 = \frac{Var[\omega]\, Cov[\delta, z+y] - Cov[\omega, z+y]\, Cov[\omega,\delta]}{Var[\omega]\, Var[z+y] - Cov^2[\omega, z+y]}, \tag{A18}$$

where

$$Var[z+y] = \sigma_z^2 + \sigma_y^2, \tag{A19}$$

$$Var[\omega] = \beta_1^2\sigma_\delta^2 + \beta_2^2\left(\sigma_z^2 + \sigma_y^2\right) + \sigma_z^2 + 2\beta_2\sigma_z^2, \tag{A20}$$

$$Cov[\delta,\omega] = \beta_1\sigma_\delta^2, \tag{A21}$$

$$Cov[\omega, z+y] = \beta_2\left(\sigma_z^2 + \sigma_y^2\right) + \sigma_z^2 \tag{A22}$$

$$Cov[z+y,\delta] = 0. \tag{A23}$$

Finally,

$$
\begin{aligned}
Var[\delta \mid \omega, z+y] &= Var[\delta - \gamma_1\omega - \lambda_2(z+y)] \\
&= Var[\delta - \gamma_1\omega - \lambda_2(z+y)] = \sigma_d^2 + \gamma_1^2\sigma_\omega^2 + \lambda_2^2(\sigma_z^2 + \sigma_y^2) \\
&\quad - 2\gamma_1 Cov[\delta, \omega] + 2\gamma_1\lambda_2 Cov[\omega, z+y]
\end{aligned}
$$

Q.E.D.

Price Efficiency We follow Subrahmanyam (1991) and compute a measure of price efficiency as

$$
Q = \left(Var[\delta \mid P]\right)^{-1}, \tag{A24}
$$

where we can use the result from Equation (A9). This definition of price efficiency can be interpreted as the uncertainty about the fundamental value conditional on observing the price. The smaller the uncertainty about the fundamental value given all available information, the greater price efficiency.

Total Price Impact One of our main variables of interest is the market quality as measured by the price impact of the total order imbalance unconditional of the predicted part of uninformed order imbalance. Writing $p = \alpha + \lambda_{uncond}\,\omega$, we get

$$
\lambda_{uncond} = \frac{\mathrm{cov}[p, \omega]}{Var[\omega]} = \frac{Cov[\lambda_1\omega + \lambda_2(z+y), \omega]}{Var[\omega]} = \lambda_1 + \frac{\lambda_2\left(\beta_2\left(\sigma_z^2 + \sigma_y^2\right) + \sigma_z^2\right)}{\beta_1^2\sigma_\delta^2 + \beta_2^2\left(\sigma_z^2 + \sigma_y^2\right) + \sigma_z^2(1 + 2\beta_2)}. \tag{A25}
$$

Order Imbalance Predictability Predictability of the order imbalance of uninformed traders affects the optimal strategy of the informed trader who adjusts his optimal demand to extract maximum revenue. Despite this, the total order imbalance remains predictable.

We define the predictability of order imbalance in two ways. Firstly, we consider how the market quality (as measured by total price impact and market efficiency) is affected by the predictability in the uninformed order imbalance. We define the degree of uninformed order imbalance predictability as the ratio of the variance of predicted uninformed order imbalance to the total variance of the uninformed order imbalance:

$$
R_u^2 = \frac{Var[\hat{z}]}{Var[z]} = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_y^2}. \tag{A26}
$$

Secondly, we consider how the market quality is affected by the predictability in the total order imbalance (resulted from the predictability of the uninformed orders). We define the degree of total order imbalance predictability as the ratio of the variance of predicted total order imbalance to the variance of the total order imbalance:

$$R^2 = \frac{Var[\hat{\omega}]}{Var[\omega]} = \frac{\left(\beta_2 + \frac{\sigma_z^2}{\sigma_z^2+\sigma_y^2}\right)^2 (\sigma_z^2 + \sigma_y^2)}{\beta_1^2\sigma_\delta^2 + \beta_2(\sigma_z^2 + \sigma_y^2) + \sigma_z^2(1+2\beta_2)}, \tag{A27}$$

where we have used that $Var(\hat{\omega}) = \gamma^2(\sigma_z^2 + \sigma_y^2)$ with γ given as $\gamma = Cov[\omega, z+y]/Var[z+y]$.

Numerical solution and comparative statics The system of four four Equations (A13), (A14), (A17), (A18) with respect to variables $\lambda_1, \lambda_2, \beta_1, \beta_2$ cannot be solved analytically. Instead, we present numerical solutions to the system and perform comparative statistics for a set of parameter values. The focus of this exercise is to gain an intuition for the model reaction to changes in the predictability of total and uninformed order imbalance R^2 and R_u^2 respectively.

We set our baseline scenario set of parameters as $\sigma_z^2 = 1$, $\sigma_\delta^2 = 1$ and let the noise of the uninformed order imbalance component vary according to $\sigma_y^2 = (1, \ldots, 100)$. The predictability variables R^2 and R_u^2 are then computed from the specific combination of parameters in each of the scenarios analysed. We perform our analysis for two different values of market maker's risk aversion: $A_m = 0$ and $A_m = 1$.[14]

Insert Figure A1 about here

Figure A1 plots the relation between the two measures of predictability: R^2 and R_u^2. It reveals that for a wide range of parameter values, R^2 is monotonically increasing in R_u^2. This goes in line with a common belief that predictability of uninformed order imbalance is translated into predictability of total order imbalance despite endogenous nature of informed traders' demand.

Insert Figures A2 and A3 about here

The optimal demand of informed traders increases with fundamental value of the asset (see Figure A2). Moreover, the sensitivity of informed orders to fundamental value monotonically

[14]We keep the risk-aversion parameter A_m rather small since earlier empirical work by Hendershott and Seasholes (2009) concludes that there is little evidence for overly limited risk-bearing capacity of market makers.

declines with predictability of order imbalance. Higher degree of order imbalance predictability makes it harder for informed traders to hide behind uninformed orders. This leads to optimal reduction of trading aggressiveness as R^2 increases. This is true regardless of the degree of market-makers' risk aversion. The response of informed demand to the predictive part of uninformed order imbalance is zero when market maker is risk neutral (Panel A of Figure A2). Price impact of total order imbalance increases (Panel A of Figure A3) while price response to the predicted part of uninformed order imbalance decreases with R^2 (Panel B of Figure A3).

Figure A4 presents numerical comparative statics of λ_{uncond} – price impact of total order imbalance ω unconditional of predicted uninformed order imbalance. In general the relationship between the liquidity and order imbalance predictability is non-monotonic and has an inverse U-shape. When the degree of order imbalance predictability is low, marginal increase in informed trader's profit from front-running the uninformed orders is higher than marginal decrease in profit due to revelation of private information (price impact). For higher degrees of order imbalance predictability marginal profit from front-running drops and eventually a positive effect on market quality prevails.

Insert Figure A4 about here

Figure A4 reveals that the sensitivity of the total price impact to predictability of order imbalance increases in magnitude with σ_δ^2. This suggests that order imbalance predictability reduces the degree of asymmetric information the most where such asymmetries are the most pronounced: during periods of high uncertainty and in stocks with high volatility of the fundamental values. In Figure A4, we also show that price impact sensitivity of the total price impact to order imbalance predictability exhibit similar patters for different values of uninformed order imbalance volatility σ_z^2.

Insert Figure A5 about here

Similarly, Figure A5 presents numerical comparative statics of the level and the gradient market efficiency measure Q with respect to predictability of order imbalance R_t^2. In general efficiency increases with order imbalance predictability. The higher is the predictability, the easier it is for the market maker to infer the information from the total order imbalance. Although the informed trader tends to reduce the amount of trading as predictability of order imbalance increases, the positive effect of market efficiency still dominates.

Finally, Figures A6 and A7 present comparative statics of λ_{uncond} and Q with respect to R_u^2. The results are similar qualitatively with those presented in Figures A4 and A5.

Insert Figures A6 and A7 about here

B Figures and Tables

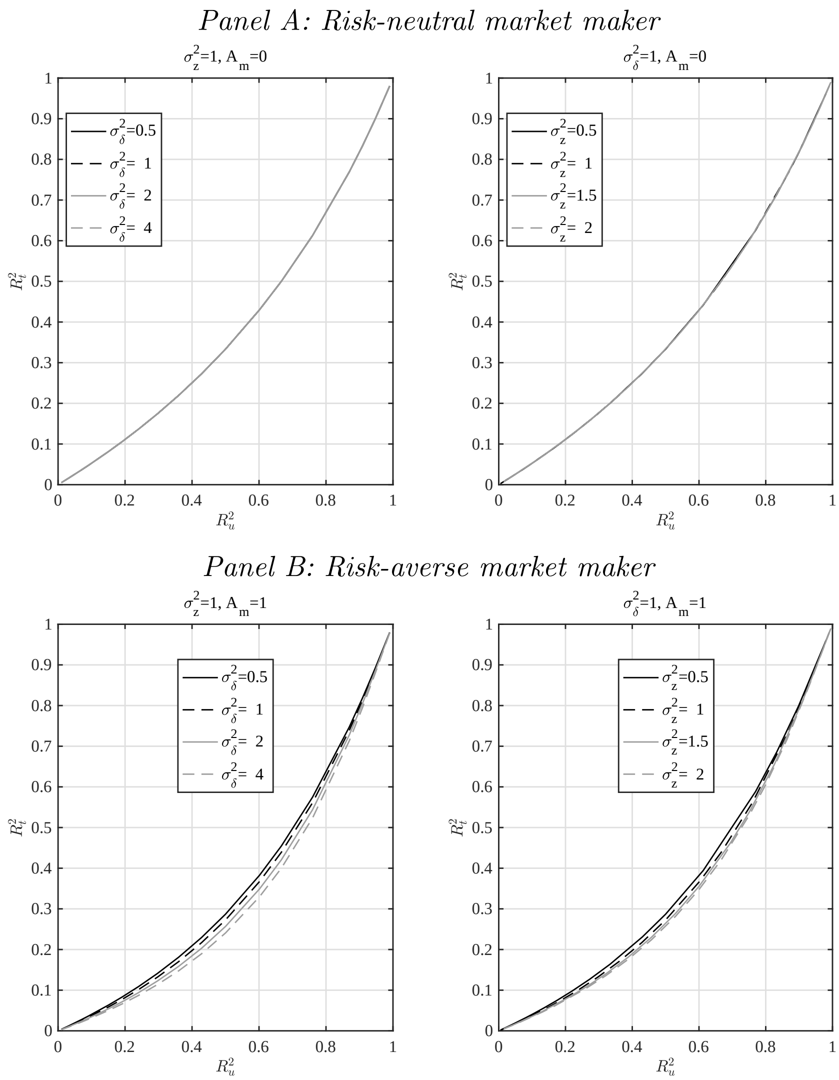
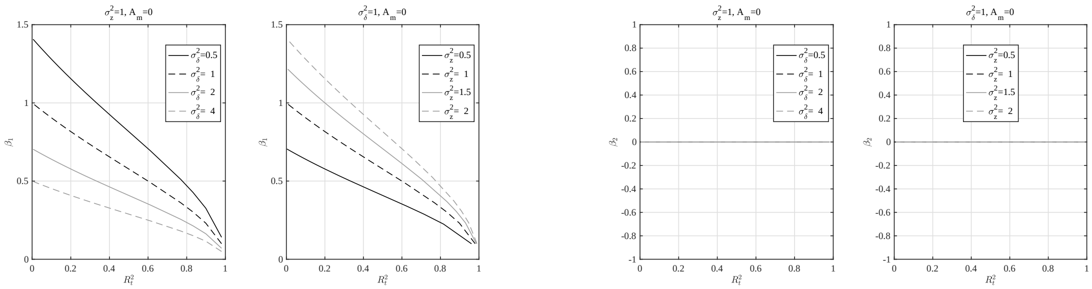


Figure A1. Comparative statics: R^2_{tot} versus R^2_u This figure plots the values of R^2 coefficient of total order imbalance predictability as a function of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the nonlinear system (A13)-(A18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R_u^2 derived from the chosen parameter set. Panel A corresponds to the case on risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

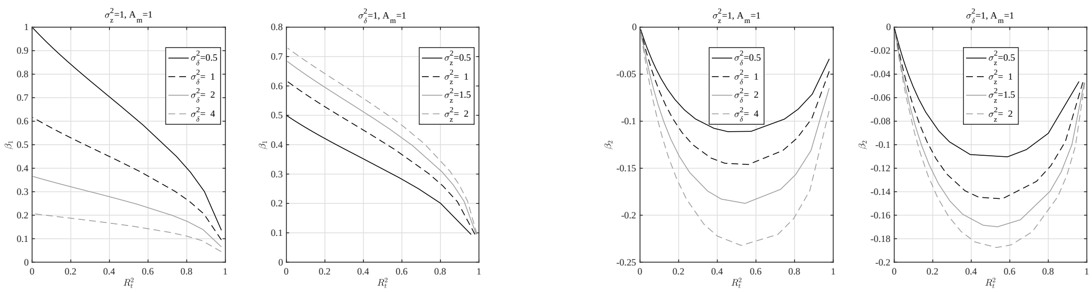
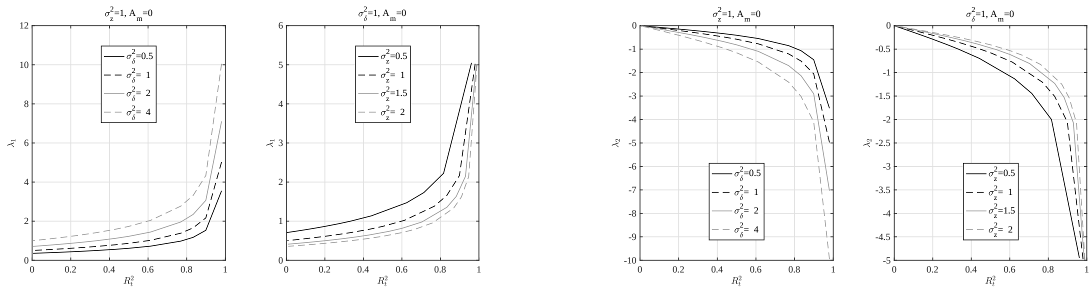


Figure A2. Comparative statics: β_1 and β_2 coefficients This figure plots the values of parameters β_1 and β_2 determining the demand function of informed traders as functions of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the nonlinear system (A13)-(A18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. Panel A corresponds to the case on risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

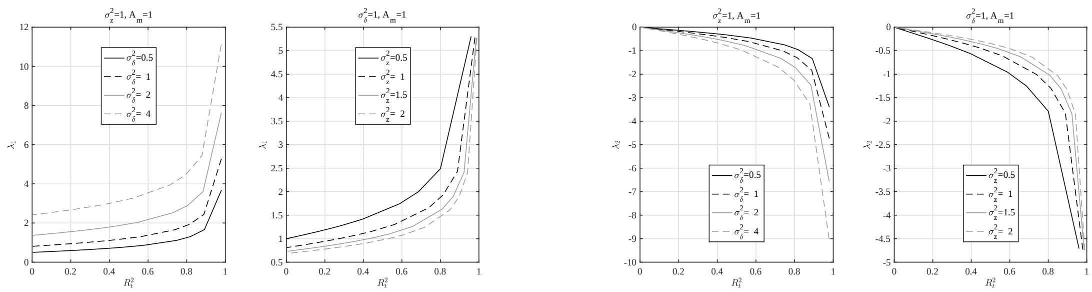
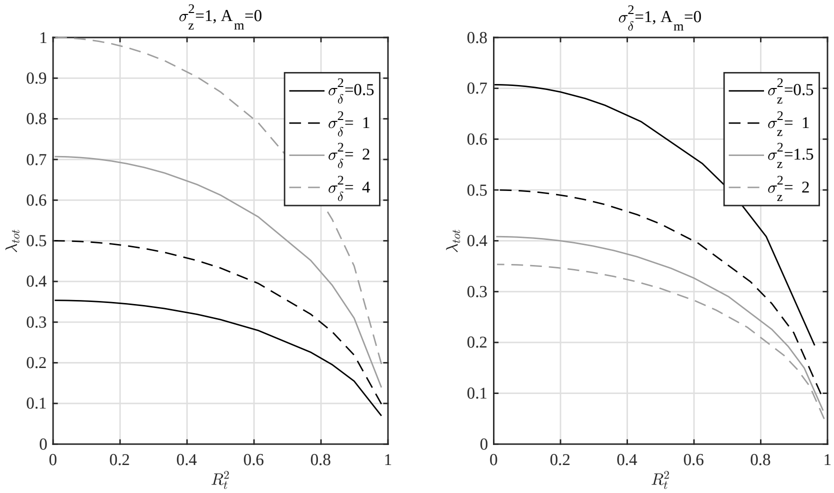


Figure A3. Comparative statics: λ_1 and λ_2 coefficients. This figure plots the values of price impact parameters λ_1 and λ_2 as functions of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the nonlinear system (A13)-(A18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. Panel A corresponds to the case on risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

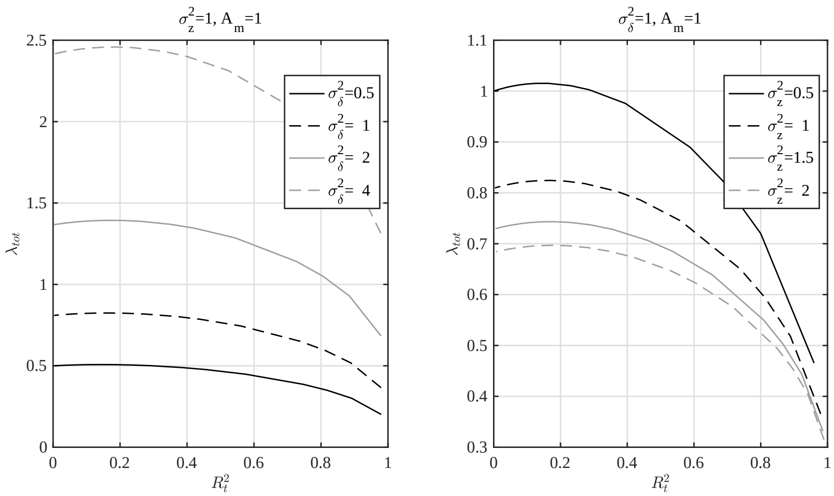


Figure A4. Comparative statics: Price impact versus predictability of total order imbalance. This figure plots the value of total price impact λ_{uncond} as a function of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the nonlinear system (A13)-(A18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \dots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. The gradient is approximated by means of a finite difference. Panel A corresponds to the case on risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

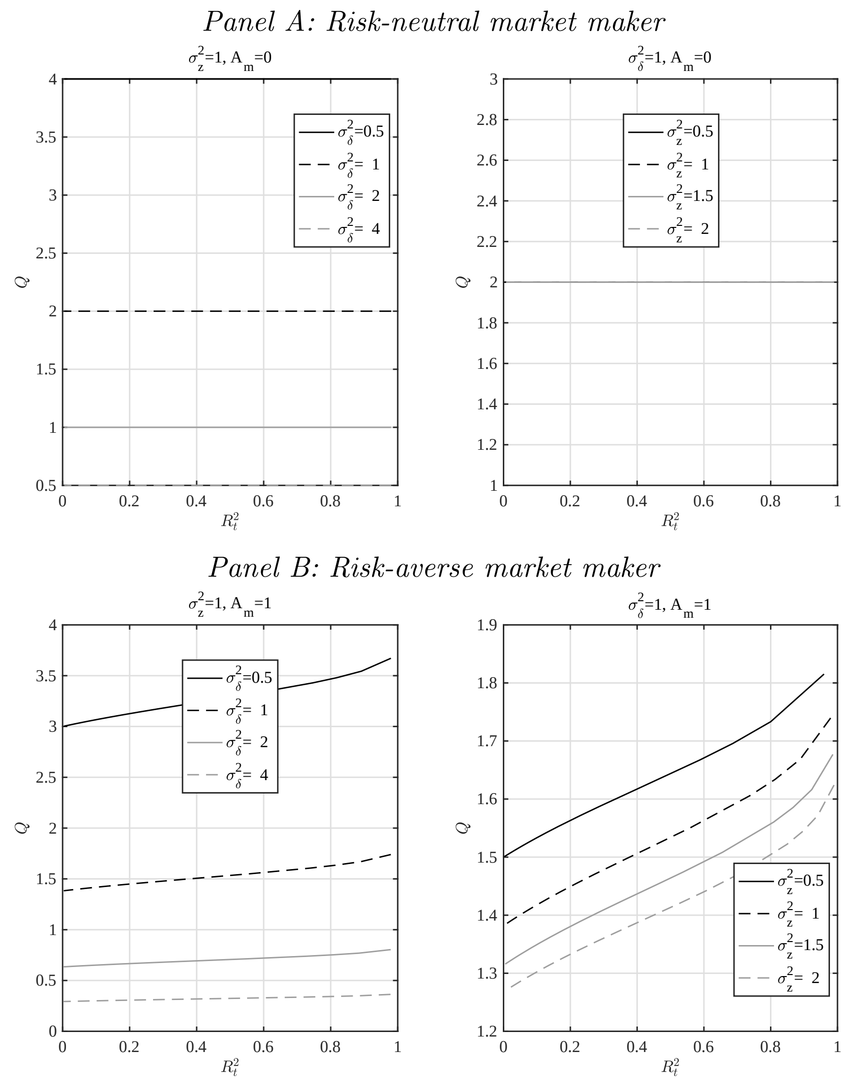


Figure A5. Comparative statics: Efficiency versus predictability of total order imbalance. This figure plots the value of price efficiency Q as a function of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the nonlinear system (A13)-(A18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. The gradient is approximated by means of a finite difference. Panel A corresponds to the case on risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

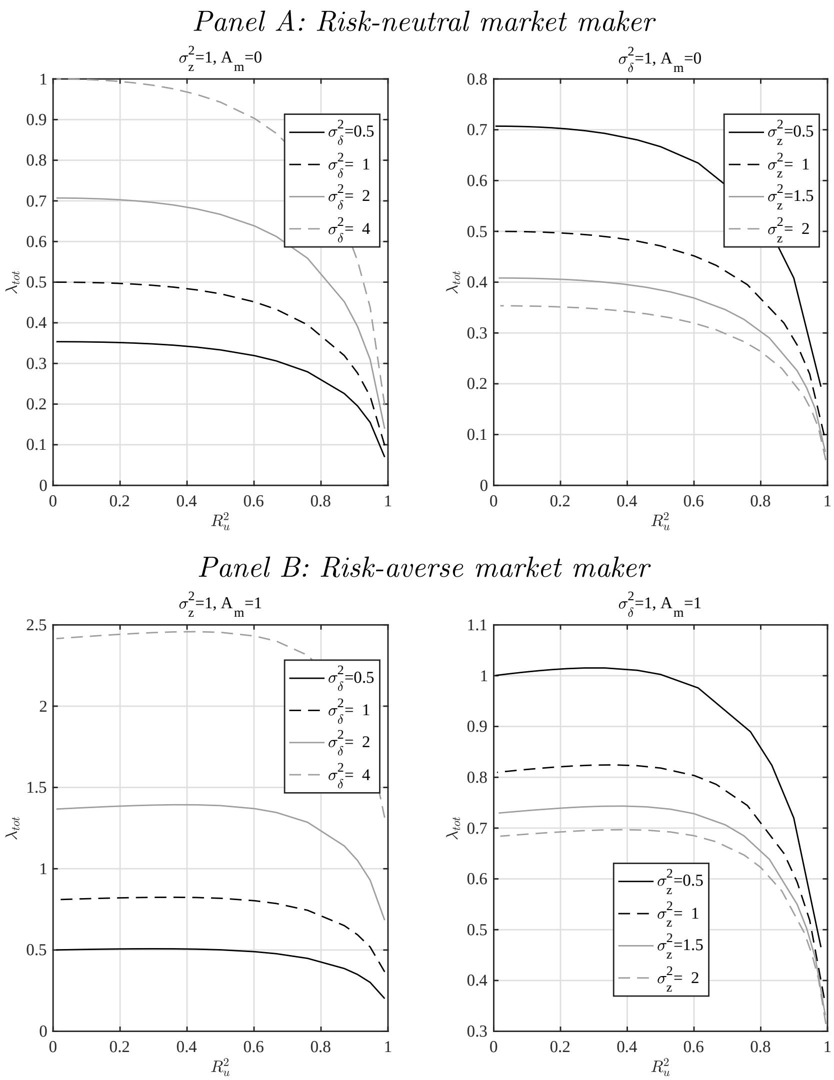


Figure A6. Comparative statics: Price impact versus predictability of uninformed order imbalance. This figure plots the value of total price impact λ_{uncond} as a function of the predictability of total order imbalance R_u^2. The data points correspond to individual numerical solutions of the nonlinear system (A13)-(A18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R_u^2 derived from the chosen parameter set. The gradient is approximated by means of a finite difference. Panel A corresponds to the case on risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

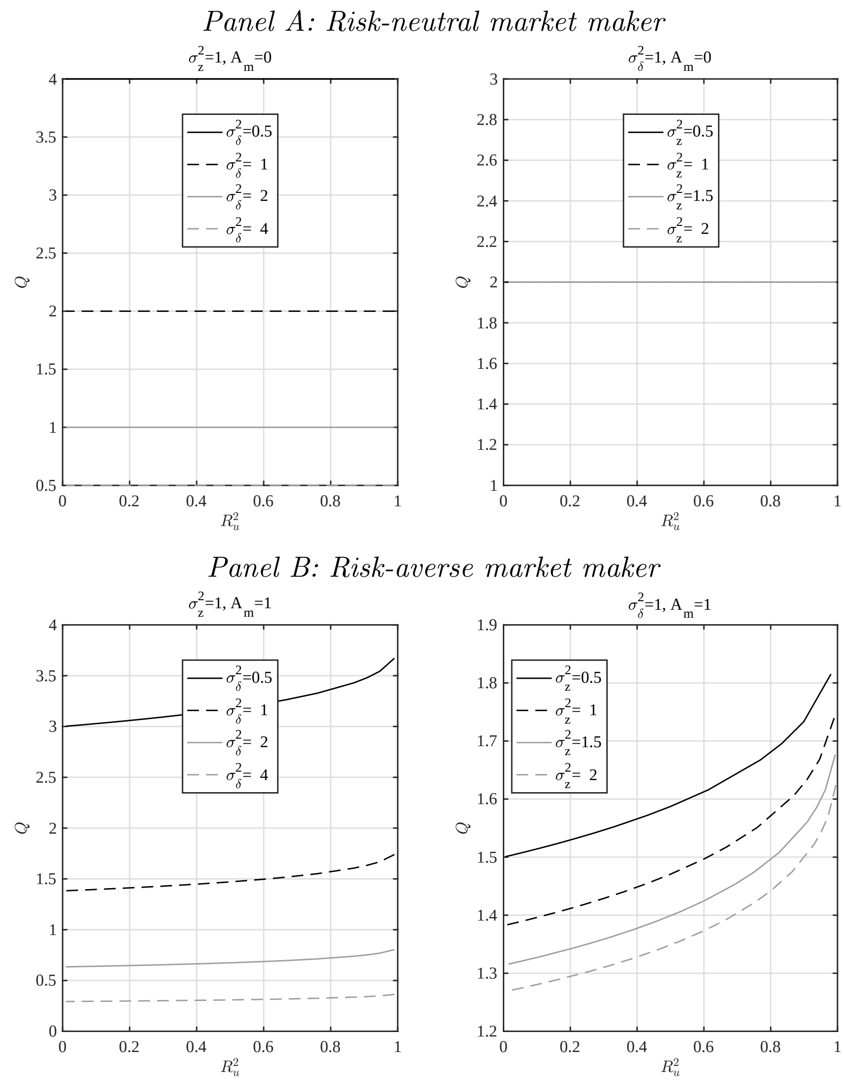
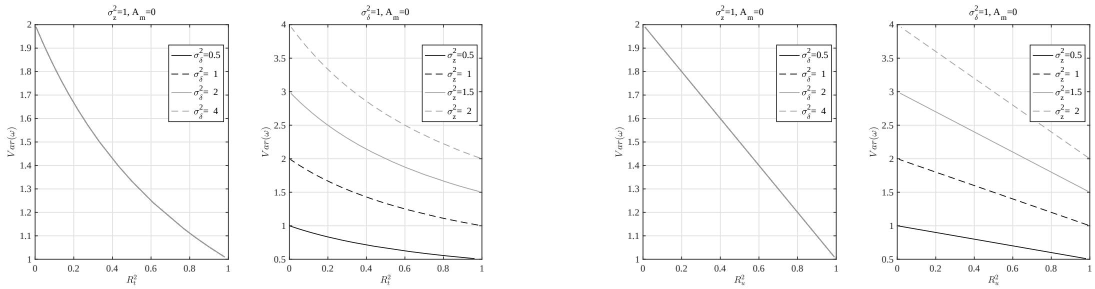


Figure A7. Comparative statics: Efficiency versus predictability of uninformed order imbalance. This figure plots the value of price efficiency Q as a function of the predictability of total order imbalance R_u^2. The data points correspond to individual numerical solutions of the nonlinear system (A13)-(A18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R_u^2 derived from the chosen parameter set. The gradient is approximated by means of a finite difference. Panel A corresponds to the case on risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

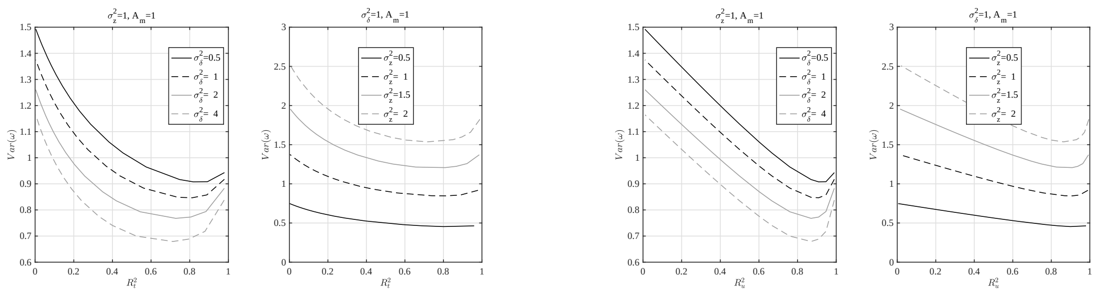


Figure A8. Comparative statics: $Var[w]$ **versus** R^2 **and** R_u^2**.** This figure plots the variance of total order imbalance as a function of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the nonlinear system (A13)-(A18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. Panel A corresponds to the case on risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Table A1. Summary statistics for the predictors of order imbalances

This table presents means and standard deviations (in squared brackets) for the predictors used to forecast order imbalances. We report the summary statistics both for the full sample and across different quintiles of market capitalization. **Panel A:** summarize the variables for the sample of active pension funds for the period from January 1, 2006 to December 31, 2010. **Panel B:** reports the summary statistics for the full sample of stocks for the period from January 1, 1997 to December 31, 2013. Here *ttq* is the trade-to-quote ratio, *ret* is the daily stock return, *rvar* is the realized variance based on 5 minutes individual stock returns, *size* is the market capitalization in billions of dollars, *prc* denotes the price, and *trvol* is the daily trading volume in millions of dollars. Finally, *dib* denotes the depth imbalance defined as $dib_{it} = (ask\_depth - bid\_depth)/(ask\_depth + bid\_depth)$, where *ask_depth* and *bid_depth* are daily average depth levels at the best bid and ask quotes respectively.

Panel A: Active Pension Funds

	Full Sample	Q1 (small)	Q2	Q3	Q4	Q5 (large)
ttq	0.12 [0.05]	0.12 [0.05]	0.11 [0.05]	0.12 [0.04]	0.12 [0.04]	0.14 [0.05]
ret (%)	0.06 [3.09]	-0.02 [5.11]	0.05 [4.08]	0.08 [3.55]	0.08 [3.09]	0.07 [2.58]
dib (%)	1.12 [14.52]	0.11 [19.65]	0.79 [15.65]	1.31 [13.74]	1.63 [12.67]	1.05 [11.19]
rvar (%)	0.15 [0.56]	0.38 [1.12]	0.21 [0.60]	0.15 [0.51]	0.11 [0.37]	0.06 [0.20]
size (B$)	9.70 [2.06]	0.37 [0.08]	0.90 [0.18]	1.92 [0.34]	4.40 [0.78]	33.87 [5.06]
prc ($)	36.82 [9.69]	14.99 [2.57]	23.80 [4.34]	32.54 [5.62]	43.53 [7.33]	53.55 [10.82]
trvol (M$)	78.24 [46.78]	3.73 [5.76]	10.92 [10.99]	25.00 [21.99]	55.31 [43.86]	240.7 [107.7]

Panel B: Full Sample of Stocks

	Full Sample	Q1 (small)	Q2	Q3	Q4	Q5 (large)
ttq	0.38 [0.41]	0.38 [0.41]	0.38 [0.32]	0.38 [0.27]	0.37 [0.25]	0.41 [0.26]
ret (%)	0.08 [3.74]	0.05 [4.61]	0.10 [4.03]	0.10 [3.68]	0.10 [3.37]	0.09 [2.90]
dib (%)	2.32 [29.51]	0.14 [33.70]	2.42 [24.45]	4.10 [21.12]	5.31 [18.96]	5.97 [16.41]
rvar (%)	0.36 [1.36]	0.65 [1.87]	0.17 [0.55]	0.11 [0.37]	0.09 [0.38]	0.05 [0.39]
size (B$)	3.83 [1.10]	0.17 [0.07]	0.67 [0.15]	1.53 [0.30]	3.75 [0.76]	28.80 [5.34]
prc ($)	24.81 [9.38]	12.50 [4.35]	23.57 [5.17]	31.98 [6.81]	41.37 [9.43]	53.06 [15.63]
trvol (M$)	24.80 [17.37]	1.08 [2.44]	6.23 [7.52]	15.22 [16.06]	36.51 [31.82]	162.3 [80.24]

Table A2. Aggregate market uncertainty and the effect of order imbalance predictability: Additional results.

This table reports the results of the following regression:

$$y_{it} = \alpha_i + \beta_{high} \cdot pl_{i,t-1} \cdot D_t^{high} + \beta_{med} \cdot pl_{i,t-1} \cdot D_t^{med} + \beta_{low} \cdot pl_{i,t-1} \cdot D_t^{low} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it},$$

where y_{it} represents either some of the additional liquidity measures, such as the effective spread *espread*, the realized spread (*rspread*) and price impact (*prcimpact*), or the additional measure of market inefficiency defined as $abs\_ac = |(1+ac)/(1-ac)|$, where ac is the autocorrelation of daily returns. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(3). **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($obi_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	espread	*rspread*	*prcimpact*	*abs_ac*
$pl_{t-1}^{pf} \times D_t^{high}$	-0.031 (-6.57)	-0.018 (-6.55)	-0.032 (-9.92)	-0.013 (-0.74)
$pl_{t-1}^{pf} \times D_t^{med}$	-0.025 (-7.19)	-0.022 (-11.7)	-0.019 (-8.31)	-0.047 (-3.88)
$pl_{t-1}^{pf} \times D_t^{low}$	-0.010 (-1.97)	-0.015 (-3.23)	-0.004 (-1.05)	-0.076 (-4.02)
oib_t	-0.002 (-1.47)	-0.006 (-4.05)	0.005 (4.39)	0.0001 (0.01)
$illiq_{t-1}$	0.554 (7.58)	0.209 (8.14)	0.268 (11.7)	
$lsize$	-0.011 (-4.22)	0.003 (1.33)	-0.021 (-10.5)	0.011 (2.37)
$turn$	0.118 (4.33)	-0.053 (-2.09)	0.176 (4.67)	-0.484 (-6.79)
$1/prc$	0.244 (3.78)	0.233 (6.40)	0.169 (4.17)	0.135 (5.49)
ttq	-4.032 (-4.82)	-2.156 (-3.04)	-1.475 (-2.07)	-3.596 (-1.52)
vix	0.051 (4.87)	-0.008 (-1.63)	0.091 (15.3)	-0.003 (-0.11)
No. obs.	764,825	764,825	764,825	764,825
R^2	65.76%	16.67%	32.49%	0.21%

Panel B: Full Sample of Stocks

	espread	*rspread*	*prcimpact*	*abs_ac*
$pl_{t-1}^{pf} \times D_t^{high}$	-0.410 (-15.9)	-0.429 (-8.41)	-0.580 (-18.8)	-0.054 (-4.39)
$pl_{t-1}^{pf} \times D_t^{med}$	-0.229 (-10.1)	-0.098 (-2.13)	-0.464 (-16.4)	-0.096 (-13.0)
$pl_{t-1}^{pf} \times D_t^{low}$	-0.179 (-7.19)	-0.021 (-0.42)	-0.424 (-13.9)	-0.096 (-11.6)
oib_t	-0.073 (-25.7)	-0.086 (-5.95)	-0.006 (-0.59)	-0.006 (-11.8)
$illiq_{t-1}$	0.578 (125.1)	0.047 (22.8)	0.036 (21.2)	
$lsize$	-0.161 (-31.5)	-0.264 (-24.9)	-0.099 (-19.6)	-0.014 (-16.1)
$turn$	-1.177 (-9.74)	-2.462 (-9.00)	-0.010 (-0.17)	-0.442 (-8.35)
$1/prc$	0.606 (22.7)	1.166 (20.5)	0.367 (14.5)	0.029 (10.2)
ttq	7.241 (18.3)	16.22 (9.32)	1.378 (1.19)	-0.041 (-0.55)
vix	0.755 (16.3)	0.956 (12.8)	0.828 (21.1)	-0.001 (-0.11)
No. obs.	12,906,715	12,906,715	12,906,715	12,906,715
R^2	49.58%	1.05%	0.43%	0.51%

Table A3. The effect of order imbalance predictability on liquidity and market efficiency by stock characteristics: Additional results.

This table presents the results of the following regression:

$$y_{it} = \alpha_i + \beta pl_{i,t-1} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it},$$

where y_{it} represents either some of the additional liquidity measures, such as the effective spread *espread*, the realized spread (*rspread*) and price impact (*prcimpact*), or the additional measure of market inefficiency defined as $abs\_ac = |(1+ac)/(1-ac)|$, where ac is the autocorrelation of daily returns. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(3). For the ease of exposition, only the loading and t-statistics on pl_t are reported. The sample stocks are sorted into quintiles Q1 to Q5 based on NYSE market capitalization breakpoints, trade-to-quote ratio, idiosyncratic volatility and passive institutional ownership. **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($obi_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	Q1	Q2	Q3	Q4	Q5
			Sorting by Market Cap		
estpread	-0.05 (-3.96)	-0.03 (-8.06)	-0.04 (-4.47)	-0.02 (-8.28)	-0.01 (-5.16)
rspread	-0.03 (-1.69)	-0.02 (-7.33)	-0.02 (-9.42)	-0.02 (-11.42)	-0.01 (-6.10)
prcimpact	-0.07 (-4.11)	-0.02 (-5.61)	-0.02(-7.77)	-0.01 (-5.57)	-0.01 (-4.13)
abs_ac	-0.11 (-2.07)	-0.02 (-0.97)	0.00 (0.17)	-0.05 (-2.56)	-0.03 (-1.35)
			Sorting by Trade-to-Quote ratio		
estpread	-0.04 (-6.09)	-0.02 (-10.30)	-0.03 (-4.80)	-0.03 (-7.58)	-0.02 (-4.71)
rspread	-0.02 (-6.44)	-0.02 (-10.39)	-0.02 (-9.34)	-0.02 (-10.11)	-0.03 (-5.21)
prcimpact	-0.03 (-6.81)	-0.02 (-7.92)	-0.02 (-7.87)	-0.02 (-8.67)	-0.02 (-3.65)
abs_ac	-0.06 (-2.43)	-0.04 (-1.98)	-0.02 (-1.32)	-0.04 (-2.02)	-0.01 (-0.24)
			Sorting by Idiosyncratic Volatility		
estpread	-0.02 (-7.33)	-0.03 (-5.94)	-0.04 (-5.43)	-0.04 (-5.95)	-0.03 (-7.40)
rspread	-0.01 (-10.07)	-0.01 (-11.47)	-0.02 (-9.10)	-0.03 (-9.99)	-0.03 (-5.50)
prcimpact	-0.01 (-5.80)	-0.02 (-8.68)	-0.03 (-9.11)	-0.03 (-8.77)	-0.03 (-6.48)
abs_ac	-0.05 (-2.22)	-0.06 (-2.53)	-0.03 (-1.35)	-0.04 (-1.69)	-0.01 (-0.013)
			Sorting by Passive Institutional Ownership		
estpread	-0.03 (-7.80)	-0.02 (-6.71)	-0.02 (-6.61)	-0.03 (-5.87)	-0.02 (-4.19)
rspread	-0.03 (-7.63)	-0.01 (-5.79)	-0.01 (-6.34)	-0.01 (-6.51)	-0.01 (-6.00)
prcimpact	-0.03 (-6.47)	-0.02 (-6.26)	-0.02 (-5.91)	-0.02 (-4.77)	-0.02 (-4.58)
abs_ac	-0.00(-0.05)	-0.04 (-1.51)	-0.05 (-1.54)	-0.05 (-2.12)	-0.03 (-1.56)

Table A3 continued.

Panel B: Full Sample of Stocks

	Q1	Q2	Q3	Q4	Q5
			Sorting by Market Cap		
estpread	-1.25 (-16.81)	-0.40 (-9.98)	-0.15 (-5.81)	-0.14 (-5.49)	-0.14 (-5.99)
rspread	-1.61 (-10.82)	-0.40 (-5.79)	0.02 (0.35)	0.01 (0.24)	-0.05 (-1.82)
prcimpact	-1.15 (-15.07)	-0.44 (-9.47)	-0.29 (-6.19)	-0.23 (-6.25)	-0.13 (-6.24)
abs_ac	-0.27 (-16.30)	-0.12 (-7.73)	-0.07 (-5.63)	-0.05 (-4.09)	-0.010 (-0.85)
			Sorting by Trade-to-Quote ratio		
estpread	-0.80 (-16.10)	-0.30 (-12.43)	-0.28 (-11.41)	-0.38 (-13.09)	-0.41 (-9.34)
rspread	-0.78 (-10.92)	-0.15 (-3.74)	-0.34 (-4.94)	-0.39 (-5.19)	-0.47 (-4.54)
prcimpact	-0.59 (-16.52)	-0.37 (-12.68)	-0.30 (-6.35)	-0.48 (-10.08)	-0.47 (-8.19)
abs_ac	-0.09 (-5.65)	-0.07 (-7.10)	-0.09 (-9.34)	-0.10 (-8.90)	-0.12 (-8.86)
			Sorting by Idiosyncratic Volatility		
estpread	-0.15 (-12.56)	-0.26 (-15.31)	-0.39 (-14.23)	-0.70 (-16.06)	-1.54 (-16.50)
rspread	-0.26 (-4.10)	-0.16 (-2.99)	-0.36 (-6.17)	-0.77 (-9.04)	-2.16 (-12.86)
prcimpact	-0.08 (-1.75)	-0.36 (-9.01)	-0.46 (-11.86)	-0.67 (-12.82)	-1.25 (-14.98)
abs_ac	-0.06 (-5.17)	-0.05 (-5.38)	-0.06 (-6.34)	-0.09 (-8.92)	-0.23 (-14.19)
			Sorting by Passive Institutional Ownership		
estpread	-1.21 (-10.02)	-0.61 (-11.64)	-0.28 (-11.05)	-0.23 (-12.91)	-0.23 (-11.57)
rspread	-1.67 (-7.26)	-0.58 (-4.19)	-0.22 (-2.73)	-0.16 (-3.00)	-0.18 (-2.92)
prcimpact	-0.94 (-8.60)	-0.83 (-8.91)	-0.44 (-7.77)	-0.37 (-9.74)	-0.32 (-7.98)
abs_ac	-0.32 (-12.93)	-0.12 (-7.32)	-0.08 (-6.44)	-0.05 (-4.40)	-0.065 (-5.18)

Table A4. De-trended predictive likelihood

This table reports the results for the main regression analysis but with the predictive likelihood replaced by its de-trended transformation. For each given stock, we define the de-trended predictive likelihood pl^*_t as residuals from the regression of pl_t on the corresponding time index t, i.e., $pl_t = \alpha + \beta\, t + pl^*_t$. As a result, for instance, the effect of order imbalance (de-trended) predictability is investigate by the following regression:

$$illiq_{it} = \alpha_i + \beta pl^*_{i,t-1} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \epsilon_{it},$$

where $illiq_{it}$ takes one of following illiquidity variables (expressed in basis points): quoted spread *qspread*, effective spread *espread*, realized spread *rspread*, and price impact *prcimpact*. We run the same regression for market inefficiency measures, such as the variance ratio $vratio_{it}$ and a measure of absolute correlation abs_ac_{it}. **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), and the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($obi_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	qspread	*espread*	*rspread*	*prcimpact*	*vratio*	*abs_ac*
pl^*_{t-1}	-0.02 (-6.00)	-0.01 (-4.91)	-0.01 (-3.77)	-0.01 (-4.60)	-0.02 (-2.94)	-0.03 (-2.21)
$illiq_{t-1}$	0.58 (23.88)	0.56 (7.75)	0.21 (8.24)	0.27 (11.81)		
oib_t	0.003 (2.08)	0.001 (0.06)	-0.001 (-3.16)	0.07 (5.99)	-0.001 (-0.63)	-0.001 (-0.23)
lsize	-0.02 (-5.79)	-0.01 (-4.22)	0.003 (1.11)	-0.02 (-10.36)	0.01 (3.72)	0.01 (2.09)
turn	-0.16 (-4.31)	0.06 (2.22)	-0.09 (-3.70)	0.12 (3.49)	0.22 (6.27)	-0.50 (-6.80)
$1/prc$	0.23 (4.28)	0.24 (3.76)	0.23 (6.37)	0.17 (4.17)	0.07 (3.94)	0.13 (5.45)
ttq	-10.50 (-8.35)	-1.60 (-1.70)	-0.22 (-0.30)	0.55 (0.78)	0.75 (0.70)	-1.99 (-0.76)
vix	0.09 (9.45)	0.04 (4.67)	-0.01 (-1.79)	0.07 (15.24)	0.02 (3.00)	0.04 (1.99)
No. obs.	764,825	764,825	764,825	764,825	764,825	764,825
R^2	54.67%	65.52%	16.44%	32.21%	0.30%	0.16%

Panel B: Full Sample of Stocks

	qspread	*espread*	*rspread*	*prcimpact*	*vratio*	*abs_ac*
pl^*_{t-1}	-0.43 (-15.27)	-0.42 (-15.52)	-0.65 (-10.80)	-0.43 (-12.42)	-0.09 (-18.55)	-0.17 (-16.99)
$illiq_{t-1}$	0.68 (170.8)	0.58 (125.21)	0.05 (22.87)	0.04 (21.34)		
oib_t	-0.05 (21.44)	-0.07 (-25.16)	-0.08 (-5.90)	-0.01 (-0.28)	-0.001 (7.13)	-0.001 (-10.51)
lsize	-0.18 (-36.23)	-0.17 (-35.37)	-0.27 (-26.61)	-0.13 (-25.92)	-0.01 (-23.29)	-0.02 (-19.82)
turn	-1.61 (-9.86)	-1.27 (-9.75)	-2.45 (-9.11)	-0.21 (-3.60)	0.04 (5.61)	-0.46 (-8.38)
$1/prc$	0.53 (20.54)	0.59 (22.24)	1.15 (20.39)	0.33 (13.10)	0.01 (5.43)	0.02 (6.90)
ttq	2.27 (7.46)	7.89 (20.50)	16.61 (9.76)	2.63 (2.31)	0.07 (1.62)	0.08 (0.89)
vix	0.53 (17.11)	0.48 (18.69)	0.49 (11.46)	0.63 (26.41)	0.03 (6.92)	0.05 (3.57)
No. obs.	12,906,715	12,906,715	12,906,715	12,906,715	12,906,715	12,906,715
R^2	59.84%	49.50%	1.05%	0.41%	0.88%	0.51%