What Drives Trading in Financial Markets? A Big Data Perspective *

Shikun Ke[1] and Anton Lines[2]

[1] *Yale University*
[2] *Columbia University*

January 30, 2023

Abstract

We use deep Bayesian neural networks to investigate the determinants of trading activity in a large sample of institutional equity portfolios. Our methodology allows us to evaluate hundreds of potentially relevant explanatory variables, estimate arbitrary nonlinear interactions among them, and aggregate them into interpretable categories. Deep learning models predict trading decisions with up to 86% accuracy out-of-sample, with macroeconomic conditions and market liquidity together accounting for most $(66-91\%)$ of the explained variance. Stock fundamentals, firm-specific corporate news, and analyst forecasts have comparatively low explanatory power. Our results suggest that macroeconomic risk and market microstructure considerations are the most crucial factors in understanding institutional trading patterns.

Keywords: Institutional trading; machine learning; big data; public information

*Correspondence: Anton Lines, e-mail: anton.lines@columbia.edu. We thank seminar participants at Columbia Business School, INSEAD, HEC Paris, Copenhagen Business School, and BI Norwegian Business School for helpful comments.

1 Introduction

Despite receiving considerable attention in recent decades, asset market trading activity remains one of the great unsolved mysteries in financial economics. Illustrating the scale of the puzzle, Daniel and Hirshleifer (2015) report that annual turnover in the five hundred largest US stocks was as high as 223% from 1980 to 2014. The "no trade" theorems of Milgrom and Stokey (1982) and Tirole (1982) have spurred a large body of theoretical work exploring the conditions under which high trading volume can occur in equilibrium, with many calibrations able to match broad stylized facts about volume and returns.[1] However, with many potential explanations, it is not known which mechanisms are of primary importance in real-world markets and which are secondary (or trivial). To date, empirical studies have examined the proposed explanations largely in isolation,[2] while a large-scale, data-driven comparison between potential motivations for trade is missing from the literature.

This paper aims to provide such a large-scale approach. Specifically, we estimate the relationship between public information signals and trading activity using multi-layered ("deep") Bayesian neural networks. This methodology allows us to overcome the econometric challenges that typically plague regression models with many variables, such as multicollinearity and over-fitting, and therefore to include an expansive set of potential explanatory signals—over 500 in total. We are also able to estimate complex empirical relationships in a flexible, non-linear, semi-parametric way, ensuring that all potential drivers of trade (at least, those that are measurable in our dataset) are fully accounted for. This completeness is crucial for evaluating the *relative* importance of different explanations, which is the primary goal of the paper.[3]

[1] A non-exhaustive list: Kyle (1985); Easley and O'hara (1987); Harris and Raviv (1993); Wang (1993, 1994); He and Wang (1995); Daniel et al. (1998); Scheinkman and Xiong (2003); Lo et al. (2004); Banerjee and Kremer (2010); Kondor (2012); Du and Zhu (2017); and Kyle et al. (2018), among many others.

[2] For instance, empirical papers have examined trading responses to corporate news (Barber and Odean (2008); Tetlock (2010); Fang et al. (2014); Hendershott et al. (2015); Huang et al. (2020)), earnings announcements (Bhattacharya (2001); Kaniel et al. (2012)); publication of anomalies in the academic literature (Calluzzo et al. (2019)), behavioural biases (Statman et al. (2006); Grinblatt and Keloharju (2009)), analyst forecasts (Irvine et al. (2007); Goldstein et al. (2009)), market variables and holding-period returns (Grinblatt and Keloharju (2001); Griffin et al. (2007)), and private information (Collin-Dufresne and Fos (2015); Di Mascio et al. (2021)).

[3] Without prior knowledge of the information structure of the market, simpler methods such as LASSO or ridge regression will not suffice. Even if a linear framework were able to capture most of the estimated variation, we would not know this without first estimating the more complex model.

Our analysis focuses on the trades of financial institutions, for which representative portfolio-level data are available. This granularity turns out to be necessary for capturing heterogeneous trade motivations, which are otherwise obscured in aggregate market statistics. Daily institutional transactions are provided by Abel Noser Solutions (also known as ANcerno). The dataset accounts for as much as 12% of daily US equity trading volume (Hu et al. (2018)), covering over a thousand institutions and over thirty thousand individual portfolios from January 2000 to December 2010.

For potential explanatory variables, we include the entirety of CRSP (for all stocks in the investment universe of the ANcerno institutions), the entirety of Compustat, all available news coverage and sentiment statistics in the Ravenpack news database, all sell-side analyst forecasts in the IBES database, and the major macroeconomic time-series from the St Louis Federal Reserve Economic Database (FRED). We allow for various timings of each variable, up to the four most recent quarters for monthly/quarterly data, and up to the most recent five days for daily variables. For daily variables, such as returns, market volume, news, and analyst forecasts, we also construct long-run past averages. These data sources provide a comprehensive picture of the public information environment facing institutional traders in the United States.

Before training the models, we first group the ANcerno portfolios into clusters based on similarities in trading styles. Using a hierarchical clustering algorithm, we endogenously extract five interpretable sub-groups: (i) small-cap traders; (ii) quantitative traders; (iii) block traders; (iv) frequent buyers; and (v) concentrated anomaly traders. These groups show significant differences in trading patterns and stock characteristics, and allow us to capture heterogeneous trade motivations. For example, at the extensive margin of trade decisions, different investment styles cause most effects except market liquidity to offset each other in the aggregated sample, while these effects shine through in the disaggregated groups.

We model both the extensive and the intensive margins of trading activity: respectively, the probability that a particular stock in the investment universe is traded on a particular day, and the log dollar volume conditional on a transaction taking place. For the combined universe of all portfolios, the model correctly identifies whether a trade will take place 71% of the time, and $77-86\%$ of the time in the separate portfolio groups. These accuracies should be judged against a baseline probability of 50% due to down-sampling of "no trade" observations. The corresponding pseudo-R^2s range from 0.24 to 0.60, comparable to values of

0.01 to 0.16 in prior studies.[4] For the intensive margin, the model achieves an out-of-sample adjusted R^2 of 0.53 in the combined sample, and from 0.19 to 0.49 in the separate portfolio groups, comparable to R^2s of around 0.04 in prior work using transaction-level data.[5] All of the above statistics are computed out-of-sample, as are all results presented below.

The most important contribution of our approach is to estimate the relative importance of the various information sources. We find that for the extensive margin of trading activity, the variables that account for the highest share of explained variance are non-ANcerno market dollar volume (primarily on and close to the day of the trade, but also throughout the previous quarter) and the collective set of macroeconomic variables, both prevailing conditions and innovations. Market volume contributes 71.8% to the explained variance in the aggregated sample of all portfolios, and between 22.6% (for the concentrated anomaly traders) and 47.5% (for the block traders) in the separate groups.[6] Macroeconomic conditions explain 13.0% of the variance in the aggregated sample, and between 22.1% (block traders) and 42.1% (anomaly traders) in the separate groups. For the intensive margin, market volume accounts for 30.1% of the variance in the full-sample model and between 6.5% (small cap traders) and 40.5% (block traders) in the separate portfolio groups. Macroeconomic variables explain 56.1% in the aggregated sample and from 50.5% (block traders) to as high as 82.5% (anomaly traders) in the separate groups.

Firm fundamentals, corporate news, and analyst forecasts contribute much less to the explained variance. In the aggregated sample of all portfolios, fundamentals—including market capitalization, dividend yield, and the major balance sheet, income statement, and cash flow statement items—account for only 8.2% of variance at the extensive margin, and 10.6% at the intensive margin. However, the aggregate results mask significant inter-group variation. At the extensive margin, for example, the contribution of fundamentals rises to a minimum of 16.8% (block traders) and a maximum of 23.5% (anomaly traders). The contribution of corporate news, on the other hand, remains low throughout the full sample (3.1% at the extensive margin and 2.3% at the intensive margin) as well as in the separate groups (2.3% $-$ 12.11% at the intensive margin, and 6.7% $-$ 9.04% at the extensive margin). Moreover, most of the explanatory power of news comes from earnings announcements rather

[4]See section 4.4.1.

[5]See section 4.4.2.

[6]We first subtract ANcerno volume from market volume to avoid estimating a purely mechanical relationship.

than articles in the financial press. The least important variable category turns out to be analyst forecasts, which explain only $1.2\% - 3.1\%$ across all models at the extensive margin, and $0.1\% - 0.4\%$ at the intensive margin.

Another useful way to decompose the explained variance is by the timing of information arrival relative to the day of the trade. This decomposition tells us how quickly institutional traders react to new information, and whether they react to stale information. For both the intensive and extensive margin, we find that most types of traders react immediately to news—that is, on the day the news is first released—with the exceptions of the quantitative traders (group 2) at the extensive margin, and the small cap traders (group 1) at the intensive margin. In these two cases, the trade reaction occurs more gradually over five business days from the information release date.

To provide deeper insight into our model preditions, we derive the equivalent of regression coefficients within the networks: i.e., the responses of the predicted model output to changes in individual or composite variables. Unlike with standard feed-forward neural networks, in Bayesian models it is straightforward to construct confidence intervals for these responses by repeatedly sampling from the estimated posterior network weights. Given the high contribution of market volume to the models' explanatory power at the extensive margin, we start with predicted responses to changes in the Amihud illiquidity measure (a nonlinear function of daily volume and returns; see Amihud (2002)). For most portfolio groups, the probability of trading one particular stock as opposed to another decreases sharply in the day-of-trade illiquidity of the former. This finding is striking, as it implies that market microstructure considerations are highly relevant—not only for trade execution, but also for the stock selection decision itself. We also find limited evidence consistent with exploitation of illiquidity premia by some groups of traders (quants and block traders), and that the general preference for liquidity is specifically a preference for *temporarily* high liquidity. These effects are statistically and economically large.

One concern regarding the interpretation of the Amihud measure is that volume itself may act as an information aggregator. If this is the case, much of our analysis may be redundant. To address this concern, we train a separate neural network to predict log market dollar volume, using the same suite of explanatory variables as we use for our ANcerno trade models. Although this network achieves a very high R^2 (81.6% out of sample), its explanatory power is driven mostly by stocks' market capitalization (55.5% of the explained variance).

While we do find that corporate news and analyst forecasts play a non-negligible role (7.8% and 8.5%, respectively), these effects are sufficiently small that most of the effects of market volume can still be attributed to liquidity. Overall, the findings in this paper imply that liquidity motivations for trade are first-order.

Lastly, we examine model responses to the macroeconomic variables in more detail. The most important single macro time-series turns out to be inflation, and we document significant heterogeneity in traders' responses to this variable. Portfolio groups 2 and 3 (quants and block traders, respectively) tend to trade more aggressively when inflation is high, while groups 4 and 5 (frequent buyers and concentrated anomaly traders, respectively) tend to show the opposite behavior. Traders in group 1 (small cap traders) tend to trade more aggressively when current price levels are more extreme in either direction. There is also significant variation in traders' response to new information releases versus prevailing economic conditions. This heterogeneity suggests that differences of opinion about aggregate conditions—or, alternatively, aggregate risk hedging based on differing underlying exposures—is also a first-order determinant of trading activity.

Our results indicate that macroeconomic risk and market microstructure considerations are the main drivers of institutional trading, and that theoretical models relying on these mechanisms will be closer to empirical reality. In particular, our results support the quantitative importance of models with differences of opinion, such as Harris and Raviv (1993) and Scheinkman and Xiong (2003), and for models where liquidity dynamics are centre-stage, such as Kyle (1985) and its many extensions. Although mechanisms involving firm-specific characteristics, news, or pre-processed information (analyst forecasts) are all statistically significant, their effects are comparatively small.

In the section following this introduction, we provide additional details regarding the data used in the paper. Section 3 then describes how we group portfolios together according to similarities in trading styles, and examines the features of each group. Section 4 explains the methodology for and presents results from the Bayesian neural network estimation. Section 5 concludes.

2 Data

Equity transaction data are obtained from Abel Noser Solutions (also known as ANcerno), an execution-cost consultancy serving many large US financial institutions. Although these institutions self-select the trades they submit for analysis, Puckett and Yan (2011) and Anand et al. (2012) find that the ANcerno dataset is representative of the broader universe of Form-13F-filing institutions. The data are indexed by date, institution ID (*clientcode*), portfolio ID (*clientmgrcode*), and stock CUSIP; they include trade quantities, execution prices in US dollars (as well as other prices such as daily highs, lows, VWAPs, etc), intraday timestamps, and brokerage commissions. However, importantly, the dataset does not include information on portfolio holdings. The sample contains approximately 230 million transactions by 1014 institutions from January 2000 to December 2010, after which ANcerno ceases to provide institution identifiers and trades can no longer be associated with the same entity over time.

Our analysis is conducted primarily at the portfolio level (i.e., *clientcode-clientmgrcode*). We aggregate all intraday trades in the same stock into daily parent orders so that our unit of analysis corresponds to portfolio manager decision-making rather than trading desk execution. This is also due to concerns about the reliability of the intraday timestamps in the ANcerno data (see Hu et al. (2018)).[7] To minimize sample selection and data integrity concerns, we drop portfolios with fewer than 100 total trades, fewer than 10 trades per month, fewer than 10 unique stocks in their trading history, average dollar volume per trade less than 1,000 USD, average daily market dollar volume of traded stocks less than 100,000 USD, more than 95% of trades in the same industry (see appendix A.2.2), and more than 80% of trades in the same direction (buy or sell).[8] After applying these filters, we retain 34,828 portfolios.

We then merge the ANcerno data with potential information signals from five standard datasets, containing stock characteristics (CRSP and Compustat), news coverage and sentiment (Ravenpack), analyst forecasts (IBES), and macroeconomic time series (FRED).

From CRSP, we obtain prices, market dollar volume, dividends and shares outstanding at a daily frequency. From Compustat we obtain the major aggregated balance sheet, income

[7] We do not observe parent order IDs directly and so may mistakenly classify some multi-day orders as separate decisions; however earlier work using Plexus trading data (e.g. Keim and Madhavan (1995)) finds that most parent orders are fully executed within a single day.

[8] Results are not sensitive to the particular cutoffs.

statement, and cash flow statement items at a quarterly frequency. From Ravenpack, we obtain the number of articles published per day referencing each company in our sample, along with Ravenpack's relevance score, "news impact projection" (*NIP*), and sentiment scores for each article.[9] NIP and sentiment are based on Ravenpack's proprietary algorithms: sentiment identifies textual characteristics associated with directional stock price movements, providing a measure of positive vs negative news, and NIP captures the predicted magnitude of the stock price response regardless of direction.

From IBES, we obtain individual security analysts' one-quarter, one-year, and two-year earnings forecasts, as well as earnings announcement dates. And from FRED, we obtain the major macroeconomic time series: real GDP (quarterly), consumption (monthly), unemployment rate (monthly), employment-to-population ratio (monthly), industrial production index (monthly), consumer price index (monthly), composite dollar exchange rate (daily), S&P500 Index level (daily), VIX Index level (daily), federal funds rate (daily), and various other government and corporate bond yields (daily).

3 Portfolio Grouping

The goal of this paper is to understand the broad informational drivers of trading activity in equity markets. The level of aggregation of the data is therefore an important consideration. Ancerno's clients include a variety of different institution types—investment managers, pension plan sponsors, and brokers—and within each type, there are many different investment objectives, strategies, and information processing capabilities. On the one hand, estimating trade determinants separately for each portfolio would deliver highly idiosyncratic results, and would also be prone to over-fitting due to limited sample sizes. On the other hand, aggregating trades across *all* portfolios would conflate the information sets of the various groups, potentially biasing estimated coefficients towards zero.

Ideally we would like to group together investment managers with similar information sets and portfolio characteristics, and study each group separately. Unfortunately, information sets are not directly observable. Moreover, most Ancerno portfolios cannot be matched to

[9] Original news sources include Dow Jones Newswires, the Wall Street Journal, Barron's, Reuters, Associated Press, and other major news organizations. Other recent papers using Ravenpack data include Kolasinski et al. (2013), Shroff et al. (2013), Shroff et al. (2014), and Dang et al. (2015). Shroff et al. (2013) compare news coverage in the Ravenpack and Factiva databases and find a 95% overlap. Following Huang et al. (2020), we drop articles from magazines, blogs, and smaller local news outlets.

public mutual fund databases such as CRSP, Thomson Reuters, or Morningstar, so data on the portfolio characteristics are limited.[10] Our solution is to compute available trade- and stock-level summary statistics for each portfolio, and to run an unsupervised clustering algorithm using these statistics. To the extent that differences in information sets result in systematic differences in measured trading behaviour, they will be captured by the algorithm.

3.1 Hierarchical Agglomerative Clustering

Specifically, for each *clientcode-clientmgrcode* combination (which identifies a unique portfolio), we compute time-series averages for three sets of features:

1. *Trade features*: the number of separate transactions per month (*trade frequency*); the log dollar volume per transaction (*trade size*); the Herfindahl index of trade sizes (*trade concentration*); the number of different stocks traded (*universe size*); and the fraction of transactions that are purchases (*buy fraction*).

2. *Stock features*: market capitalization of traded stocks (*market cap*); book-to-market ratio of traded stocks (*B/M*); the operating profitability of traded firms (*profitability*); the annual growth in total assets of traded firms (*investment*);[11] the annual dividend yield of traded stocks (*div. yield*); various past returns of traded stocks (*1D ret; 1M ret; 3M ret; 1Y ret; 3Y ret*); annual volatility of traded stocks, estimated from daily returns (*volatility*); and log daily market dollar volume of traded stocks (*mkt volume*).

3. *Industry composition*: the fraction of traded stocks (weighted by dollar volume) in each of the 10 one-digit SIC industry codes (*agriculture; mining; construction; manufacturing; communications; wholesale trade; retail trade; finance; services; public admin*).

After filtering the sample according to the procedure described in section 2, we apply the *hierarchical agglomerative clustering algorithm*. Intuitively, the algorithm is an iterative process that joins the two most similar portfolios (according to the cosine distance between vectors of their average characteristics) into one cluster, then treats the newly-formed cluster as if it were a single portfolio. This process is repeated until all portfolios have been

[10] Mention how many portfolios are matched in the Agarwal paper.

[11] Profitability and investment are constructed as per Fama and French (2015).

assigned to a cluster; it is hierarchical because observations are added sequentially in order of similarity. See appendix A.2 for further details and a formal description of the algorithm.[12]

Figure 1 plots the resulting dendrogram of hierarchical linkages between portfolios, as well as the corresponding heatmap of pairwise cosine distances. One advantage of this method over other clustering algorithms (e.g. *k-means*) is that we are not required to pre-specify the number of clusters. Rather, the dendrogram encodes all possible numbers of clusters, depending on the desired degree of similarity between them. The hierarchical method also allows us to view the structure within each broad group, and to assess at a glance how different the groups are from each other. After examining the dendrogram, we obtain our final groups by applying a cutoff equal to 75% of the maximum cosine distance. Our aim is to obtain the finest possible grouping while ensuring that most groups contain similar numbers of portfolios.

The chosen cutoff results in four large groups and one small group, represented by different colours in the dendrogram (the small group is barely visible as red linkages). The larger groups are also clearly visible as darker regions along the diagonal in the distance heatmap. The structure of the data turns out to be such that decreasing the distance cutoff further would form many smaller groups at the periphery of the large groups instead of splitting them evenly. Therefore, we stop at 5 groups.

3.2 Descriptive Statistics

To characterize the five portfolio groups, we run a series of portfolio-level panel regressions of trade/stock features on dummy variables for each group k:

$$X_{i,t} = \alpha + \beta \mathbb{1}_{\{i \in G_k\}} + \varepsilon_{i,t},$$

where $X_{i,t}$ refers to a particular feature of interest for portfolio i in month t, and $\mathbb{1}_{\{i \in G_k\}}$ is an indicator that takes the value 1 if portfolio i is in group k, and 0 if not. The coefficient β therefore measures the difference between the feature's average value in group k and its average in the other groups. Standard errors are double-clustered by portfolio and month.

[12]The clustering algorithm is actually applied twice. The first stage, discussed in appendix A.2.2, results in one cluster with many duplicate portfolios and 95% of trades in just one industry. Thus, we drop this cluster as part of our data filters, and re-run the algorithm on the remaining portfolios to obtain the final results presented in this section.

The $\hat{\beta}$ coefficient estimates, alongside full-panel averages for each feature, are reported in table 1. Panel A reports coefficients for trade characteristics, panel B for stock characteristics, and panel C for industry composition. The many significant coefficients are evidence that the clustering algorithm identifies groups of portfolios with meaningfully different trading patterns. In what follows below, we discuss the most salient distinctions among the five groups and assign each a descriptive label.

Group 1: Small-Cap Traders (131 portfolios). Fund managers in this group tend to transact in much smaller stocks with lower market volume than those in other groups (average market cap only $e^{-1.58} = 20.62\%$ of the average market cap among the others, and average volume only $e^{-0.73} = 48.19\%$ of that for the others). The traded stocks also tend to have higher volatility (by 10.97%), lower profitability (by 0.12) and lower dividend yields (about half of that in other groups). Lastly, they appear to target stocks with negative long-term (3-year) past returns (25.68% lower than for the other groups).

Group 2: Quantitative Traders ($10,833$ portfolios). The most obvious distinguishing feature of this group is that they trade far more frequently than the other groups (about 832 more transactions per month). These portfolios also trade in a much wider range of firms (average investment universe larger by ~80 stocks) and their trades tend to be less concentrated (0.11 lower trade-size Herfindahl Index). These features align broadly with the characteristics of quantitative funds identified by Abis (2020).

Group 3: Block Traders ($8,457$ portfolios). Managers of portfolios in this group tend to initiate trades that are about eight times larger ($e^{2.08}$) than those initiated by other groups. These trades also tend to be in stocks with greater market capitalization (by $e^{0.35} - 1 = 42\%$) and higher liquidity (market volume greater by $e^{0.23} - 1 = 26\%$).

Group 4: Frequent Buyers ($3,975$ portfolios). This group of portfolios is distinguished by a significant majority of their trades being purchases (above 60%, where the other groups are all close to 50%). Mechanically, by the time of eventual liquidation, the total dollar volume of buys must equal the total dollar volume of sells; thus, a high fraction of buys indicates that fund managers are making many small purchases (possibly to conceal private information) and relatively fewer, larger, sales. These managers also tend to trade in stocks that have significantly higher dividend yields (0.76% higher than for other groups), a 14% higher concentration in the finance industry, and a 7% lower concentration in the manufacturing industry. This is the only group with meaningful industry tilts.

Group 5: Concentrated Anomaly Traders (11,382 portfolios). In terms of trading behaviour, the last group is characterized by a low frequency of trading (about 443 fewer trades per month) and small trade sizes (only $e^{-2.53} = 7.97\%$ of the size of the average trade in the other groups). These trades are highly concentrated (Herfindahl index greater by 0.17) and occur in a small universe of stocks (about 48 fewer than average). Most importantly, the firms traded by this group are, on average, on the "correct" side of the major asset pricing anomalies: they have by far the highest book-to-market ratio (higher than other groups by 1.04), the highest profitability (higher by 0.07), the lowest investment (lower by 0.12), and the lowest volatility (lower by 6.47%).

Overall, the descriptive results in this section provide a rich and novel characterization of the different types of institutional traders operating in the US equity market during our sample period (2000-2010). As we show in subsequent sections, the groups also allow us to develop a finer understanding of trade motivations.

4 Trade Determinants

In this section we describe our main empirical analysis. We start with the construction of candidate information signals that could plausibly motivate institutional trading, then outline our interpretable neural network estimation procedure. Finally, we present the results of the estimation and discuss implications for theories of trading activity.

4.1 Variable Construction

In this paper we are concerned primarily with publicly-available information. We capture private information to the extent that differing interpretations of public signals could constitute an information advantage for some traders (e.g. Chuprinin et al. (2019), and that trading in advance of scheduled information release dates may signal advance knowledge.

Our goal is to estimate traders' information sets without imposing any particular structure on the data; thus we apply only limited transformations to the raw variables. Note that, because multi-layer feed-forward neural networks are able to estimate arbitrary nonlinear transformations of variables (and interactions among them), the results of our analysis are not sensitive to the initial data transformations, or lack thereof (unlike with OLS, for example; see further discussion in section 4.2). We do not manually compute composite characteristics such as book-to-market ratios or apply scalings such as dividing by total assets, as the network is able to learn such configurations if they contribute to the model's predictive power.

We log-transform all positive variables, and apply the log-modulus transformation to variables with negative and positive values.[13] This is done primarily for convenience, as it compresses the distributions of variables with high kurtosis, making the results easier to display. We also winsorize all continuously-distributed data at the 0.1% level to reduce the impact of outliers and potential measurement error. Lastly, we normalize each variable by subtracting its sample mean and dividing by its sample standard deviation.

The timing of each observation is defined relative to the date of each trade (or batch of trades) in the ANcerno dataset. Since the ANcerno timestamps are not generally reliable, or missing entirely, we assume that traders have access to any information released on the day of the trade prior to market close. Institutional traders are likely to be aware of high-frequency intraday information in real time, and certain managers may even have advance access to some sources. For low-frequency variables, we include dummy variables for their (scheduled) announcement dates, as well as five leads and lags. The interaction of these dummies with the newly-released data allows us to capture whether traders react immediately on the day of the trade or with a delay (up to five business days). The leading dummies capture any trading in advance of scheduled announcements.

The following paragraphs provide an overview of the variables included in our models and their construction from the raw data. Appendix A.1 contains a full list of variable names and descriptions.

CRSP variables. For prices and shares outstanding, we record the most recent observations as of the close of business on the day before the trade. For stock returns, we record the

[13] The log-modulus transformation of x is given by $sign(x) \times ln(1 + |x|)$.

contemporaneous return on the same day as the trade, as well as the past daily returns on the five prior business days. We separately construct one-month, one-year, and three-year compounded past returns as of the close of business on the day before the trade (to capture the standard momentum and reversal patterns). For the market dollar volume in the traded stock, we record the day-of-trade and the five previous days' volume, as well as the average volume over the quarter before the trade. Since ANcerno institutions constitute up to 12% of daily trading volume, we always subtract ANcerno volume from market volume to avoid estimating mechanical effects. For dividends, we construct the trailing one-year dividend yield relative to the stock price on the day before the trade.

Compustat variables. For all major items (from the balance sheet, income statement, and cash flow statement), we record observations from the firm's four most recent quarterly financial statements, up to and including the day of the trade. We also construct dummy variables indicating whether the trade took place on the day of a statement release (earnings announcement date) or, separately, on any of the five business days before or after the release date.

Ravenpack variables. For each trade, we record the number of articles published about the firm from all news sources in the Ravenpack database: (i) on the day of the trade; (ii) on each of the five days prior to the trade; (iii) over the month before the trade (exclusive of the trade date); and (iv) over the quarter before the trade (exclusive). In addition to the number of articles, we also record the average news impact projection (NIP) and sentiment scores. We follow Ravenpack's suggestion and exclude articles with a relevance score below 90%, which eliminates cases where the firm is mentioned but is not the primary subject of the story.

IBES variables. For the one-quarter, one-year, and two-year forecast horizons, we record the number of individual analyst forecasts released on the day of the trade and on each of the five business days prior to the trade. We also aggregate all forecasts over the past four quarters (separate aggregations for each quarter). In addition to the number of forecasts, we record the mean forecast value and, for the quarterly aggregations, the forecast standard deviation.[14] The IBES "actuals" file also contains headline earnings-per-share (EPS) and the associated announcement dates. We record the four most recent quarterly EPS observations, as well as announcement date dummies (also with five leads and lags). We merge these dates with the announcement dates reported in Compustat, using the IBES dates if there is a conflict between the two.

FRED variables. For each of the monthly and quarterly macroeconomic time series (real GDP, consumption, unemployment, employment-to-population, industrial production, and CPI), we match the values reported in FRED to their various release dates from the Bureau of Economic Analysis (BEA), Bureau of Labour Statistics (BLS), and the Federal Reserve System. We then record the four most recent quarterly observations up to and including the day of the trade. Furthermore, we create release date dummies (as well as dummies for the five days before and after the release date) for each group of variables with separate schedules: GDP data, income/consumption data, employment data, production data, and inflation data. Lastly, we record daily market index levels (for the S&P 500, the VIX, and a composite dollar index) and various interest rates: 3-month, 1-year, 10-year, 20-year, and 30-year Treasury yields, and Moody's AAA and BAA corporate bond indices.

4.2 Basic Methodology

To estimate the mapping from information signals $X_{i,t}$ to investors' trading decisions $y_{i,t}$, we use feed-forward neural networks (FFNNs) with multiple hidden layers (i.e., "deep" networks). This network structure is extremely popular in the modern machine learning literature as a result of its ability to approximate arbitrary nonlinear functions to any desired precision—formalized as the *universal approximation theorem* (see Hornik et al. (1989) for the proof). This approximation includes complex interactions among explanatory variables, which are often highly important in large models such as ours.

FFNNs have the simplest network architecture, with each layer formed as an affine function of the previous layer's output, before applying a vector-valued nonlinear "activation function". Formally, we have input layer $X^{(0)} = X \in R^{K_0}$, a K_0-dimensional vector of

[14] We do not record the standard deviation for the daily aggregations, as there are typically only one or two forecasts published on a given day.

explanatory variables. Then, for layers $l = 1, ..., L$, we have:

$$X^{(l)} = \text{Activ}^{(l)}\left(W^{(l)}X^{(l-1)} + b^{(l)}\right), \tag{1}$$

with outputs $X^{(l)} \in R^{K_l}$, weight parameters $W^{(l)} \in R^{K_l \times K_{l-1}}$ and "bias" parameters $b^{(l)} \in R^{K_l}$, where K_l is the number of outputs ("nodes") for layer l. Intermediate layers are referred to as "hidden layers", and their outputs as "hidden nodes". To allow the network to learn nonlinear functions, $\text{Activ}^{(l)}(x)$ can be any convex function of x. In practice, for the hidden layers of the network, we choose an activation function with efficient numerical properties: the leaky rectified linear unit ("leaky ReLU"):

$$\text{Leaky-ReLU}(x) = \begin{cases} \alpha x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

where $\alpha \in (0, 1)$ is the Leaky-ReLU weight. Intuitively, networks with this activation function learn piecewise-linear approximations to the true underlying continuous function. We use leaky ReLU instead of the standard $\text{ReLU}(x) = \max(x, 0)$, which can suffer from information loss as negative network weights are transformed to zeros. The final layer of the network ("output layer") has as many nodes as the dimension of the target function output—in our case, one—and its activation depends on the range of the target function.

We consider two different outcome variables, representing the extensive and intensive margins of the trade decision:

1. (Extensive Margin) $y_{i,t} = \mathbb{1}_{\{Volume_{i,t} \neq 0\}}$
   $Volume_{i,t}$ is the total dollar volume of stock i traded by ANcerno institutions on date t. Thus, $y_{i,t}$ is an indicator variable that takes the value 1 if stock $i \in N$ was traded by at least one institution on date t, and 0 otherwise. The universe of stocks N is defined as all stocks traded by ANcerno institutions at some point in the sample. In other words, the control group is formed by stocks held in the ANcerno portfolios but not traded on date t. To construct balanced treatment and control groups, we randomly down-sample observations where $y_{i,t} = 0$, such that the baseline probability of a trade occurring is set to 0.5. For the extensive margin, the final-layer activation function of the network

is the logistic function: $y = 1/\left(1 + \exp\left(W^{(L)}X^{(L)} + b^{(L)}\right)\right)$*, which can be interpreted as the predicted probability of a trade occurring.*

2. (Intensive Margin) $y_{i,t} = |ln(Volume_{i,t})| \times \mathbb{1}_{\{Volume_{i,t}>0\}}$
   In this case, $y_{i,t}$ *is equal to the unsigned log dollar volume of stock i traded on date t (i.e., conditional on a trade taking place). For the intensive margin, the network has no final-layer activation, such that the final prediction for* $y_{i,t}$ *is an affine function of the penultimate layer output and thus can take any value in* R^1.

As discussed in section 3, we divide the ANcerno portfolios into five groups based on similarities in their trading behavior. We then estimate the model in equation 1 separately for each portfolio group, as well as for the full sample. Within each group, ANcerno trades are summed for each stock-day pair. The separate groups allow us to study disaggregated trade motivations that may net out when combining trades from all portfolios.

4.3 Bayesian Neural Networks

One drawback of standard FFNNs is that they do not allow for estimation of sampling variance, and thus cannot be used natively to construct confidence intervals for parameters or predictions. These intervals, and other forms of hypothesis testing, can be constructed by bootstrapping the network: re-estimating it multiple times on randomly-chosen subsamples of the data. However, this procedure is computationally expensive for large models such as the ones we estimate. A second drawback is that FFNNs, with their many parameters and highly flexible functional form, are prone to overfitting in-sample, leading to poor out-of-sample forecasting.

Both of these issues can be solved parsimonously by estimating a Bayesian neural network (BNN), with a *posterior distribution* over the weights W and biases b. Thus, instead of yielding a single prediction, a forward pass through the network will yield a draw from the estimated posterior distribution over model parameters. Confidence intervals can then be constructed by repeatedly sampling from the posterior and taking percentiles of the sample. BNNs also have the added advantage of mitigating the over-fitting problem. Firstly, the posterior distribution over parameters is formed as a weighted average of a prior—which can be set to a tight distribution around zero values for the weights and biases—and the likelihood function of the data. Effectively, the posterior weights and biases are shrunk towards the prior, regularizing the network. Secondly, estimating a distribution over parameters instead of point estimates reduces the scope for overfitting, as the model cannot converge too strongly to particular weights because it must also match the uncertainty in the underlying data.

In practice, we estimate Gaussian posteriors over the model parameters. Thus, the weights matrix W is replaced by separate matrices of means W_μ and variances W_{σ^2}, and the bias vector b is replaced by vectors of means b_μ and variances b_{σ^2}. The parameters W_μ, W_{σ^2}, b_μ, and b_{σ^2} are chosen to minimize a loss function that depends on the form of the output variable $y_{i,t}$ and the model's prediction $\hat{y}_{i,t}$. For the extensive margin, we use binary cross entropy, and for the intensive margin, we use the the mean-squared error. The parameters are then estimated using backpropagation, specifically the "Bayes by Backprop" variational Bayesian algorithm of Blundell et al. (2015) and the *Adam* stochastic optimizer (Kingma and Ba (2014)).

For each model, we randomly divide the dataset into 90% training (in-sample) data and 10% validation (out-of-sample) data. The models are estimated using the training data, but predictive accuracy/goodness-of-fit statistics are also computed for the validation data in order to verify that the models have not been over-fit.

Further details on the estimation methodology and network structure can be found in appendix A.3.

4.4 Results

4.4.1 Extensive Margin

Model Accuracy. Figure 2 plots out-of-sample accuracy statistics for the six models—all portfolios, plus the five separate portfolio groups, as derived by hierarchical agglomerative clustering (see section 3)—after training until convergence. Recall that the baseline probability of a trade occurring in stock i at date t has been normalized to 0.5. Each panel of the figure contains a confusion matrix, which shows the fraction of observations correctly and incorrectly classified, for each of the true labels: "trade" and "no trade". Correct classifications are shown along the diagonals, while incorrect classifications are shown on the off-diagonals.

From panel (a), when aggregating trades across all portfolios, we find that the model has an overall classification accuracy of 71% out-of-sample, but exhibits a bias towards predicting the occurrence of a trade (65% of observations). We hypothesized in section 4.2 that aggregating all trades may conflate the trade motivations of different groups of investors. This hypothesis is confirmed by the greater prediction accuracy obtained when the model is estimated separately on the trades of the five sub-groups. From panels (b) to (f), we see that overall out-of-sample accuracy ranges from 77-86%, and aside from the group with the fewest number of portfolios (group 1; small-cap traders), the models do not show the same strong bias towards predicting the occurrence of trades. The group with the highest accuracy is group 2 (quantitative traders), which is consistent with the notion that quants rely most on public signals in choosing which stocks to buy and sell. The small differences between in-sample and out-of-sample accuracy confirm that the models do not suffer from over-fitting.

For comparison with existing studies that use linear or logistic regression—particularly Grinblatt and Keloharju (2001), the spiritual ancestor to this paper—we also report Efron pseudo-R^2s in panel C of table 2. Efron's pseudo R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \widehat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \overline{y})^2}$$

where $\widehat{y}_n$ is the predicted probability that the n^{th} observation in the dataset will be a trade, while y_n is the actual outcome (1 for a trade, 0 for no trade), and $\overline{y}$ is the sample average probability of a trade (normalized again to 0.5). We also apply the standard "adjusted R^2" penalty,[15] and compute the statistic both in- and out-of-sample.

We find out-of-sample pseudo-R^2s ranging from 0.24 to 0.60, which are considerably higher than those found in existing studies. For instance, Grinblatt and Keloharju (2001) report pseudo-R^2s of 0.014 to 0.157 for the trades of financial institutions in their Finnish data. Although these numbers are not directly comparable to ours—their data includes information on portfolio holdings that ours does not, while our dataset includes many public signals missing in theirs—they help to contextualize the accuracy that our models are able to achieve.[16]

[15] Adj. $R^2 = 1 - \frac{(1-R^2)(N-1)}{N-K_0-1}$

[16] Collin-Dufresne and Fos (2015) also regress trade dummies on various market-wide and stock-specific characteristics; however, they do not report the regression R^2s.

Variance Decomposition. To answer the main question posed by the title of this paper, we turn to estimates of the importance of explanatory variables in our model predictions. Specifically, we compute the share of outcome variance explained by each variable in the input layer of the network. Following Dimopoulos et al. (1995), for input variable k, we compute the importance

$$I(k) = \sum_{n=1}^{N} \left(\frac{\partial \widehat{y}_n(x)}{\partial x_k} \bigg| x = X_n \right)^2 ,$$

where x_k is the k^{th} variable in the input layer. The importance is equal to the average of squared partial derivatives of the model prediction with respect to the k^{th} input variable, evaluated at each observation n in the sample. The derivatives are approximated numerically by ε-perturbations of variable k, while holding all other variables at their sample means.

Table 2 reports the results of this variance decomposition. As the input dataset contains over five hundred variables, we cannot show the importance individually for each one. Thus, we group together variables with specific informational themes and report the sum of their constituent importances. We identify five main themes: (i) market variables; (ii) firm fundamentals; (iii) corporate news; (iv) analyst forecasts; and (v) macroeconomic news.

The most striking result is that market variables, specifically those associated with liquidity (volume and returns), explain by far the largest share of trade decision variance for the model estimated on the full sample of ANcerno portfolios (74.60%). The next most important variable category is macroeconomic news (12.96%), followed by firm fundamentals (8.20%), corporate news (3.07%), and then analyst forecasts (1.17%). Perhaps the most surprising result is the low relative importance of corporate news, especially given the recent findings of Huang et al. (2020), who examine institutional trading around corporate news in the same (ANcerno) dataset and find significant contemporaneous trading. Part of the difference may be that their measures of news are more carefully constructed than ours, as we simply use the statistics available in the Ravenpack News database; however, we note that the R^2s in their full-sample regressions are quite low (< 0.04), which is consistent with our results.

As noted in the previous section, the lack of explanatory power of the non-market variables may be because these variables matter in offsetting ways for portfolios with different investment styles, while liquidity matters in the same way for all of them. Indeed, when we disaggregate the trade data into the five groups produced by hierarchical agglomerative clustering, the non-market variables take on greater importance. Corporate news rises to explain between 6.71% and 9.04% of the model variance, although most of the increase comes from earnings announcements rather than qualitative news articles. Firm fundamentals now explain between 16.84% and 23.51%, consistent with the differences in investment styles discussed in section 3.2.

The biggest winner of disaggregation is macroeconomic news, which now explains between 22.11% and 42.14% of the total predictive variance. All four macro sub-categories are meaningful, although the largest share comes from economic activity (such as GDP, consumption, inflation, and industrial production) as well as market interest rates (Treasury yields, corporate bond yields, etc) and indices such as the S&P500 and the VIX. Even in the disaggregated data, analyst forecasts are the least important (maximum 3.09% of total variance) for predicting the extensive margin of institutional trading. This result is also perhaps surprising given the prior literature, which has found evidence of institutional trading responses to analyst recommendations (e.g. Irvine et al. (2007); Goldstein et al. (2009); Busse et al. (2012)). It is worth noting that our results do not suggest that these factors have no relationship with trading activity; only that their importance is low *relative* to other factors.

Panel B of table 2 presents an alternative summary of the variable importance. Here, we aggregate information by the timing of its release relative to the date of the trade (or non-trade). Our input dataset can be broadly divided into prevailing characteristics (for example, financial statements from 1-4 quarters prior to the trade date, last quarter's GDP, the stock's market capitalization, etc) and information innovations (for example, newly-published news articles, earnings announcements, the latest jobs report, etc). In the latter case, we always include five business days' lags of the new information release, in order to measure the speed of investors' response. With one major exception, day-of-trade innovations explain the vast majority of the predictive variance—up to 39.20%—while the lagged innovations explain only 1-3% each. These results indicate that most institutional traders are quick to respond to new information, while their response to even day-old information is an order of magnitude smaller.

The exception to the above finding is the result for quantitative traders (group 2), who do not react strongly to information innovations at all. In total, only 12% of the explained variance is accounted for by the innovation variables, with the remainder being explained by prevailing characteristics. This finding is consistent with quantitative traders relying on trend-following or reversal patterns in stock returns, but is inconsistent with high-frequency trading.

Model Responses. The results in the previous section indicate which data are most important for explaining the extensive margin of institutional trading activity. However, they do not indicate the direction or shape of the effects. To see these effects, we would like to obtain the equivalent of linear regression coefficients from our models. Although the functions estimated by neural networks are complex, we can obtain a flavor of the responses by fixing all input variables at their sample means, then varying one or a few of them at a time (from two standard deviations below to two standard deviations above the mean), and plotting the corresponding model predictions. The use of Bayesian neural networks also allows us to estimate 95% confidence intervals for these responses by repeatedly sampling from the learned posterior distribution and computing appropriate sample percentiles.

Here, we examine the role of market liquidity in more detail. From the variance decomposition, we know that market volume and returns are important drivers of the decision of which stocks to buy or sell on a given day—but are these variables measuring liquidity, or something else? A major benefit of our flexible estimation methodology is that we can also examine responses to changes in composite variables constructed from the raw data. We therefore plot the model's response specifically to changes in the Amihud illiquidity measure (Amihud (2002)), constructed as the ratio of absolute daily returns to daily market dollar volume.[17] While finer microstructure-based liquidity measures exist, the Amihud measure is a convenient combination of two basic variables in our dataset, and has been shown to successfully capture price impact (e.g. Goyenko et al. (2009)).

Figures 3 and 4 plot, respectively, the model's responses to changes in day-of-trade and past-quarter Amihud illiquidity for the five portfolio groups. Note that the x-axes in all of the plots run from -2 to $+4$ (rather than -2 to $+2$). Axis values of -2 represent the case with an absolute return equal to zero and a market liquidity 2 standard deviations below

[17] To calculate the model's response to the Amihud measure, we first compute its responses to market dollar volume and returns, separately, for grids from -2 to $+2$, then compute the Amihud measure using each combination of grid points. This procedure results in multiple model outputs for each value of the Amihud measure, which we then average to obtain the final response plot.

its mean. Similiarly, axis values of +4 represent cases where absolute return is 2 standard deviations above or below zero and market liquidity is 2 standard deviations below its mean. The y-axes show predicted trade probabilities. Due to the size of the dataset, we estimate the model responses very precisely for most of the domain of the output function, with three exceptions where the confidence intervals are wider: (i) day-of-trade illiquidity above 2 for group 3 (block traders), (ii) quarterly average illiquidity for group 1 (small cap traders), and (iii) quarterly average illiquidity below 1 for group 3 (block traders).

The patterns are similar across portfolio groups for day-of-trade illiquidity but quite different for past-quarter average illiquidity. Figure 3 shows that trade probabilities are generally decreasing as day-of-trade illiquidity rises, suggesting that traders in all five groups tend to select highly liquid stocks and largely avoid stocks with low liquidity on the day of the trade decision. It is surprising that liquidity plays such a large role for the actual choice of stock, not just the implementation of the trade. In group 2 (quantitative traders) and, to a lesser extent, group 3 (block traders), very high levels of illiquidity are associated with a change in the slope of the relationship, and the probability of trading more illiquid stocks begins to increase again. For the block traders, the change of slope may be a chance effect of high sampling variation in the region of high illiquidity, but the change is clear for the quantitative traders. In both cases, however, illiquidity values above +2 represent only a small fraction of the empirical distribution. This finding may indicate that these stocks carry a high illiquidity premium that attracts certain sophisticated traders.

Model responses to past-quarter average illiquidity are shown in figure 4 and display starkly different relationships. For three of the five groups—quantitative traders (group 2), block traders (group 3), and frequent buyers (group 4)—the relationship between illiquidity and predicted trade probability is hump-shaped. This suggests that certain traders tend to avoid not only illiquid stocks, but also those stocks whose liquidity is *typically* high (the high liquidity may indicate that it is more difficult to collect private information about these stocks), while they favour stocks whose liquidity is *temporarily* high. Group 5 (concentrated anomaly traders) display the same decreasing trade probabilities as we saw for day-of-trade illiquidity, and group 1 (small-cap traders) shows the opposite result. This latter finding is consistent with the preferred universe of small-cap traders having naturally higher average illiquidity.

In all cases discussed above, the effects are economically as well as statistically significant. Predicted trade probabilities vary by at least 17.5% (panel (a) of figure 4) and by as much as 80% in some cases (panel (e) of figure 3).

4.4.2 Intensive Margin

Model Accuracy. Panel C of table 3 reports goodness-of-fit statistics for the models estimated on log dollar trade volume. Unlike the extensive margin, the total share of variance explained by the model is higher for the aggregated sample of all ANcerno portfolios, and lower for the disaggregated groups estimated in section 3. In our view, the most likely reason for this difference is that dollar trading volume is not only a function of traders' information sets, but also of the size of their portfolios. Unfortunately, we do not observe portfolio sizes in the ANcerno dataset—large trades may indicate large portfolios, but they may also indicate high turnover in smaller portfolios. As such, aggregating the data across a greater number of portfolios should cause size differences to average out and improve the fit of the model. We should also not expect to obtain as high R^2s here as we do for the extensive margin, since a major determinant of trade size (i.e., portfolio size) is missing from the input dataset.

Nonetheless, we are still able to obtain impressively accuracy predictions for the intensive margin of trade. The out-of-sample adjusted R^2 is 0.531 for the combined sample of all ANcerno portfolios, and ranges from 0.189 (group 4; frequent buyers) to 0.492 (group 2; quantitative traders) in the separate portfolio groups. Similar values for the in-sample statistics indicate, again, that the models have not been over-fit.

These numbers are larger than comparable R^2s in prior empirical studies, most of which are in-sample. For instance, Campbell et al. (2009) infer aggregate institutional trades from the TAQ database and report R^2s of 0.072 to 0.123 in a vector autoregression model of trades and returns. Huang et al. (2020) examine institutional trading around corporate news in the ANcerno dataset and report R^2s ranging from 0.038 to 0.253; however, the higher numbers in their analysis are for intensive news days and are not representative of the average trade size predictability. Di Mascio et al. (2021) study informed institutional trading patterns and report R^2s ranging from 0.096 to 0.116.

Variance Decomposition. Panels A and B of table 3 report the share of variance explained by the model's input variables, aggregated by information type (panel A) and information release timing (panel B). Section 4.4.1 describes the computation of this variance decomposition.

Simliar to the extensive margin, in the combined sample of all portfolios, the most important drivers of trade size are the market liquidity of the stock (30.22% of variance explained) and the state of the macroeconomy (56.05%). The other factors are again less important, with firm fundamentals explaining the next largest share of the variance (10.63%), and corporate news (2.30%) and analyst forecasts (0.19%) having the smallest contribution.

In the separate portfolio groups, the disaggregation allows us to capture the different ways that different types of traders make use of the available public information. As such, corporate news explains a larger share of the predictive variance (between 2.27% and 12.11%), though firm fundamentals explain about the same as in the combined sample. Most importantly, the state of the macroeconomy rises to become overwhelmingly the dominant explanatory factor, with explained variance ranging from a minimum of 50.52% (for group 3; block traders) to a maximum of 82.52% (group 5; concentrated anomaly traders). Although the state of the macroeconomy is the same for all stocks, the fact that its effects are uniformly stronger in the disaggregated groups (compared to the combined sample) suggests that it is not simply the case that overall trading intensity is higher in certain macroeconomic conditions. Rather, stocks' characteristics must be interacting with the macroeconomy in different ways in order to predict trade size; for instance, certain stocks may provide a better hedge against the different aggregate risk exposures that matter for different investor groups.

As for information release timing, the overall message from the decomposition at the intensive margin is similar to that at the extensive margin: most ANcerno institutions react an order-of-magnitude more strongly to day-of-trade information than they do to lagged information, even from the previous day. The major exception is group 1 (small cap traders), who continue to react to lagged information for at least five business days, with similar magnitudes to the day-of-trade response (though gradually declining).

Model Responses. To understand the direction and shape of the relationships between input variables and predicted trade sizes, we use the same methodology described in section 4.4.1. Here, we further explore the role played by the most important determinant of trading at the intensive margin: macroeconomic variables. Since it is not feasible to display the responses to all of the many macroeconomic indicators in our dataset, we focus, for illustration, on the single variable with the highest explanatory power. This variable turns out to be the consumer price index (CPI). However, note that the other macro indicators are of comparable importance, and the network forms a composite picture of the overall state of the macroeconomy rather than placing all weight on one or two variables.

Figure 6 displays the model responses to the standardized CPI level (from -2 to $+2$ standard deviations from its mean) for the two highest-importance quarterly observations (relative to the trade date), which are not always the most recent. Only for groups 1 and 5 is the most recent observation (indicated by the variable name suffix "_q0") among the two most important. This is also generally true for the responses to other macroeconomic variables (not shown here for brevity). Thus the main driver of trade decisions at the intensive margin appears to be more the prevailing state of the macroeconomy than innovations to that state (i.e., news), although the innovations do still play a significant role in all portfolio groups.

The main take-away from figure 6 is that the different portfolio groups have very different responses to the state of the macroeconomy. Groups 2 (quantitative traders) and 3 (block traders) tend to trade larger amounts when inflation has been high in the past six months (i.e., when the difference between the second and third quarterly CPI observations is greater). Groups 4 (frequent buyers) and 5 (concentrated anomaly traders) tend to trade larger amounts when inflation has been low in the recent past. And group 1 (small cap traders) makes larger trades when current price levels are more extreme, but when past price levels were low.

We estimate these relationships with a high degree of statistical precision, with the 95% confidence intervals barely even visible in some cases. The results are again also economically meaningful, with predicted log dollar trade volume varying by at least $e^{0.6} - 1 = 80.2\%$ (panel (a); small cap traders) and as much as $e^{2.5} - 1 = 1,118\%$ (panel (b); quantitative traders).

4.4.3 Market Volume

In this final section of results, we investigate further the most important single explanatory variable: market dollar volume. Recall that we subtract the ANcerno trading volume from the market volume to avoid estimating mechanical effects. However, while we interpret the effect of market volume primarily as stock liquidity—backed up by results for the Amihud illiquidity measure—trading volume could still aggregate other aspects of the information environment. If such information is incorporated into volume, this may explain why our model does not load strongly on corporate news or analyst forecasts. To examine this possibility, we estimate a new Bayesian neural network with market dollar volume (less ANcerno volume) as the dependent variable.

Table 4 reports the resulting variance decomposition and goodness-of-fit statistics. Note that the explanatory variables are the same as for the ANcerno trade models with the exception of market dollar volume.

Our predictive model for non-ANcerno volume is highly accurate both in- and out-of-sample, with adjusted R^2s of 0.866 and 0.816, respectively. However, most of this explanatory power (55.52%) comes just from the market capitalization of the stock. This is unsurprising, as it is well known that larger stocks tend to be more liquid. Our concern that market volume may a better aggregator of information, driving out responses to corporate news and analyst forecasts, turns out to have been valid: both of those information types explain a higher share of variance of market volume than they do of ANcerno volume: 7.83% and 8.45%, respectively. However, in absolute terms these percentages are fairly small, and do not change our overall conclusions about the importance of market liquidity or the state of the macroeconomy.

Finally, we note that market volume reacts more slowly to newly-released information than do the ANcerno institutions, with the highest share of explained variance coming from information the day *before* the trade as opposed to the day of, and the overall share of variance explained by information innovations being much lower (12.16%).

4.4.4 Discussion

The prominence of market liquidity and macroeconomic variables in our analysis leads to several provocative conclusions. In seeking to understand trading patterns, our results imply that more attention should be paid to market microstructure considerations even at the macro scale of stock selection. They also suggest that macroeconomic hedging motives or, alternatively, disagreement among traders about the state of the macroeconomy, is a key driver of trading decisions, particularly at the intensive margin. Models that prioritize these aspects are likely to be closer to empirical reality.

5 Conclusion

Using a large dataset of institutional equity transactions (ANcerno), we study the informational drivers of trading activity, drawing on data from diverse sources: stock fundamentals, corporate news, market data, analyst forecasts, and macroeconomic news. We estimate deep Bayesian neural networks to overcome the typical problems of multicollinearity and over-fitting, leading to models with up to 86% accuracy out-of-sample.

Decomposing the variance explained by the model into data categories, we find that market microstructure considerations (specifically market liquidity) as well as macroeconomic risk (whether hedging or disagreement about the state of the aggregate economy) are the major drivers of trading activity in the US equity market. We also find that institutional traders are quick to react to newly-released information, with same-day data releases having an order of magnitude higher explanatory power than signals with up to five days' lag.

References

Abis, S. (2020). Man vs. machine: Quantitative and discretionary equity management. *Machine: Quantitative and Discretionary Equity Management (October 23, 2020).*

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets 5*(1), 31–56.

Anand, A., P. Irvine, A. Puckett, and K. Venkataraman (2012). Performance of institutional trading desks: An analysis of persistence in trading costs. *The Review of Financial Studies 25*(2), 557–598.

Banerjee, S. and I. Kremer (2010). Disagreement and learning: Dynamic patterns of trade. *Journal of Finance 65*(4), 1269–1302.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*(2), 785–818.

Bhattacharya, N. (2001). Investors' trade size and trading responses around earnings announcements: An empirical investigation. *The Accounting Review 76*(2), 221–244.

Blundell, C., J. Cornebise, K. Kavukcuoglu, and D. Wierstra (2015). Weight uncertainty in neural network. In *International conference on machine learning*, pp. 1613–1622. PMLR.

Busse, J. A., T. C. Green, and N. Jegadeesh (2012). Buy-side trades and sell-side recommendations: Interactions and information content. *Journal of Financial Markets 15*(2), 207–232.

Calluzzo, P., F. Moneta, and S. Topaloglu (2019). When anomalies are publicized broadly, do institutions trade accordingly? *Management Science 65*(10), 4555–4574.

Campbell, J. Y., T. Ramadorai, and A. Schwartz (2009). Caught on tape: Institutional trading, stock returns, and earnings announcements. *Journal of Financial Economics 92*(1), 66–91.

Chuprinin, O., S. Gaspar, and M. Massa (2019). Adjusting to the information environment: News tangibility and mutual fund performance. *Management Science 65*(3), 1430–1453.

Collin-Dufresne, P. and V. Fos (2015). Do prices reveal the presence of informed trading? *The Journal of Finance 70*(4), 1555–1582.

Dang, T. L., F. Moshirian, and B. Zhang (2015). Commonality in news around the world. *Journal of Financial Economics 116*(1), 82–110.

Daniel, K. and D. Hirshleifer (2015). Overconfident investors, predictable returns, and excessive trading. *Journal of Economic Perspectives 29*(4), 61–88.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under-and overreactions. *Journal of Finance 53*(6), 1839–1885.

Di Mascio, R., A. Lines, and N. Y. Naik (2021). Alpha decay and institutional trading. *Working Paper*.

Dimopoulos, Y., P. Bourret, and S. Lek (1995). Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters* (2), 1–4.

Du, S. and H. Zhu (2017). What is the optimal trading frequency in financial markets? *Review of Economic Studies 84*(4), 1606–1651.

Easley, D. and M. O'hara (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics 19*(1), 69–90.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fang, L. H., J. Peress, and L. Zheng (2014). Does media coverage of stocks affect mutual funds' trading and performance? *The Review of Financial Studies 27*(12), 3441–3466.

Goldstein, M. A., P. Irvine, E. Kandel, and Z. Wiener (2009). Brokerage commissions and institutional trading patterns. *The Review of Financial Studies 22*(12), 5175–5212.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009). Do liquidity measures measure liquidity? *Journal of financial Economics 92*(2), 153–181.

Griffin, J. M., F. Nardari, and R. M. Stulz (2007). Do investors trade more when stocks have performed well? evidence from 46 countries. *Review of Financial Studies 20*(3), 905–951.

Grinblatt, M. and M. Keloharju (2001). What makes investors trade? *Journal of Finance 56*(2), 589–616.

Grinblatt, M. and M. Keloharju (2009). Sensation seeking, overconfidence, and trading activity. *Journal of Finance 64*(2), 549–578.

Harris, M. and A. Raviv (1993). Differences of opinion make a horse race. *Review of Financial Studies 6*(3), 473–506.

He, H. and J. Wang (1995). Differential information and dynamic behavior of stock trading volume. *Review of Financial Studies 8*(4), 919–972.

Hendershott, T., D. Livdan, and N. Schürhoff (2015). Are institutions informed about news? *Journal of Financial Economics 117*(2), 249–287.

Hornik, K., M. Stinchcombe, and H. White (1989). Multilayer feedforward networks are universal approximators. *Neural Networks 2*(5), 359–366.

Hu, G., K. M. Jo, Y. A. Wang, and J. Xie (2018). Institutional trading and abel noser data. *Journal of Corporate Finance 52*, 143–167.

Huang, A. G., H. Tan, and R. Wermers (2020). Institutional trading around corporate news: Evidence from textual analysis. *Review of Financial Studies 33*(10), 4627–4675.

Irvine, P., M. Lipson, and A. Puckett (2007). Tipping. *The Review of Financial Studies 20*(3), 741–768.

Kaniel, R., S. Liu, G. Saar, and S. Titman (2012). Individual investor trading and return patterns around earnings announcements. *The Journal of Finance 67*(2), 639–680.

Keim, D. B. and A. Madhavan (1995). Anatomy of the trading process empirical evidence on the behavior of institutional traders. *Journal of Financial Economics 37*(3), 371–398.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kolasinski, A. C., A. V. Reed, and M. C. Ringgenberg (2013). A multiple lender approach to understanding supply and search in the equity lending market. *The Journal of Finance 68*(2), 559–595.

Kondor, P. (2012). The more we know about the fundamental, the less we agree on the price. *Review of Economic Studies 79*(3), 1175–1207.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.

Kyle, A. S., A. A. Obizhaeva, and Y. Wang (2018). Smooth trading with overconfidence and market power. *Review of Economic Studies 85*(1), 611–662.

Llorente, G., R. Michaely, G. Saar, and J. Wang (2002). Dynamic volume-return relation of individual stocks. *Review of Financial Studies 15*(4), 1005–1047.

Lo, A. W., H. Mamaysky, and J. Wang (2004). Asset prices and trading volume under fixed transactions costs. *Journal of Political Economy 112*(5), 1054–1090.

Milgrom, P. and N. Stokey (1982). Information, trade and common knowledge. *Journal of Economic Theory 26*(1), 17–27.

Puckett, A. and X. Yan (2011). The interim trading skills of institutional investors. *The Journal of Finance 66*(2), 601–633.

Scheinkman, J. A. and W. Xiong (2003). Overconfidence and speculative bubbles. *Journal of Political Economy 111*(6), 1183–1220.

Shroff, N., A. X. Sun, H. D. White, and W. Zhang (2013). Voluntary disclosure and information asymmetry: Evidence from the 2005 securities offering reform. *Journal of Accounting Research 51*(5), 1299–1345.

Shroff, N., R. S. Verdi, and G. Yu (2014). Information environment and the investment decisions of multinational corporations. *The Accounting Review 89*(2), 759–790.

Statman, M., S. Thorley, and K. Vorkink (2006). Investor overconfidence and trading volume. *Review of Financial Studies 19*(4), 1531–1565.

Tetlock, P. C. (2010). Does public financial news resolve asymmetric information? *The Review of Financial Studies 23*(9), 3520–3557.

Tirole, J. (1982). On the possibility of speculation under rational expectations. *Econometrica: Journal of the Econometric Society*, 1163–1181.

Wang, J. (1993). A model of intertemporal asset prices under asymmetric information. *The Review of Economic Studies 60*(2), 249–282.

Wang, J. (1994). A model of competitive stock trading volume. *Journal of Political Economy 102*(1), 127–168.

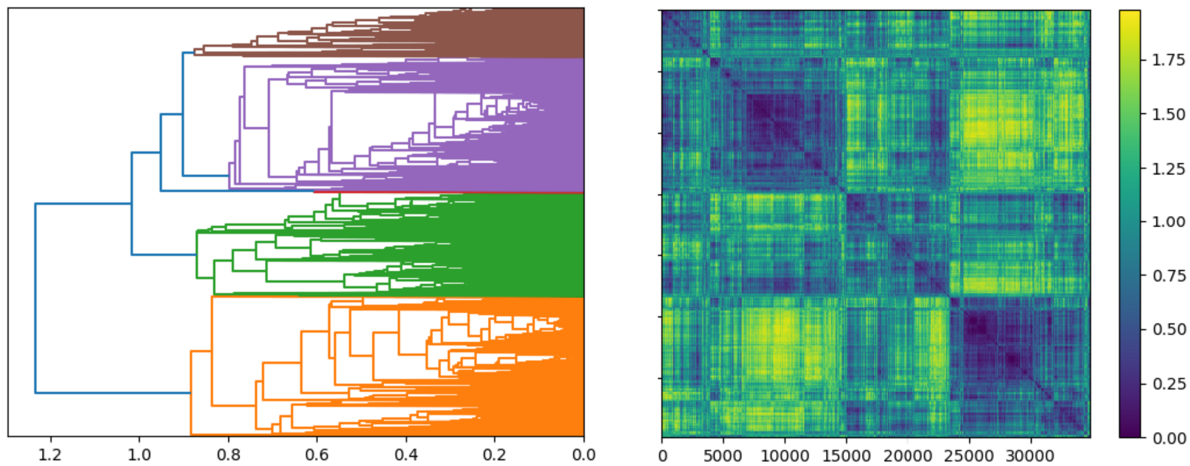
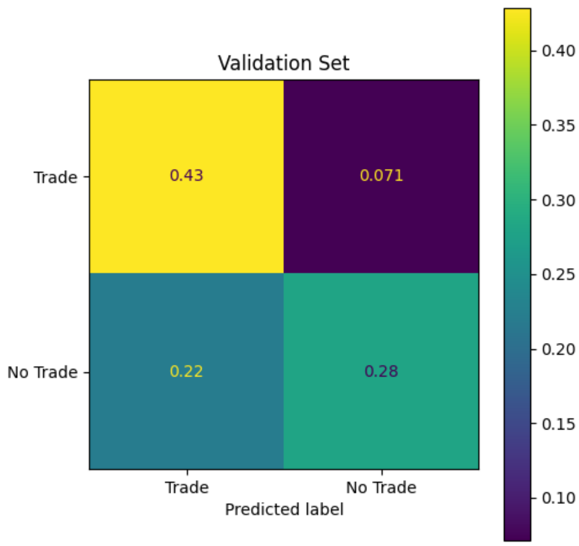
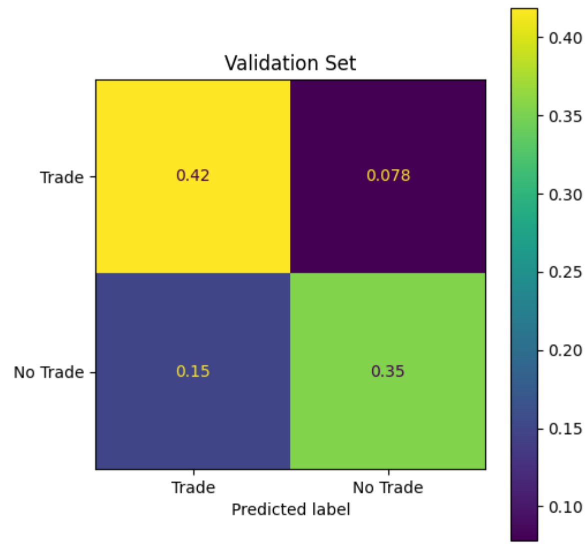


Figure 1: Dendrogram showing linkages between portfolios, as determined by the hierarchical agglomerative clustering algorithm (left panel) and heatmap showing corresponding pairwise cosine distances (right panel). See section 3.

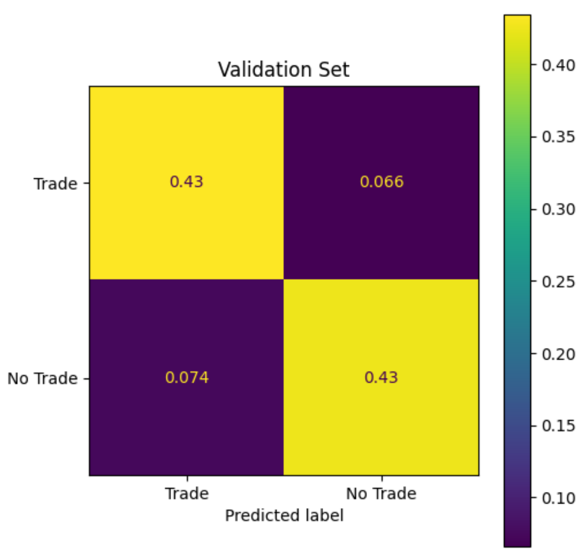
(a) All Portfolios



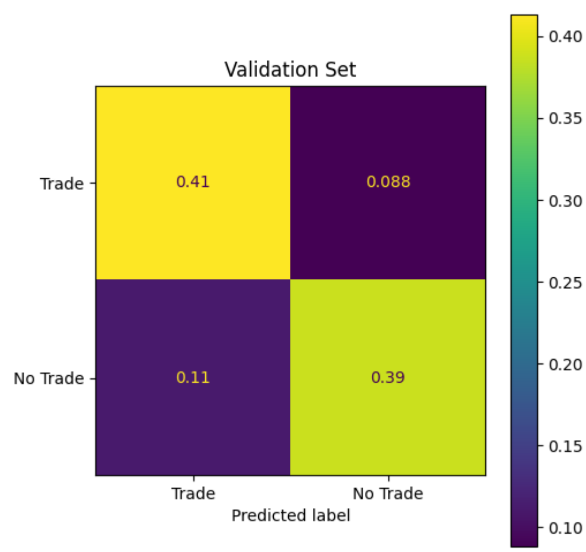
(b) Group 1: Small Cap Traders



(c) Group 2: Quantitative Traders



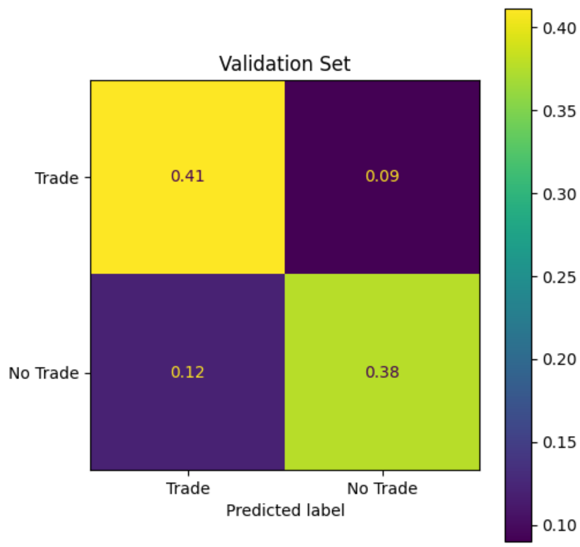
(d) Group 3: Block Traders



Continued on following page ...

... continued from previous page.

(e) Group 4: Frequent Buyers



(f) Group 5: Anomaly Traders

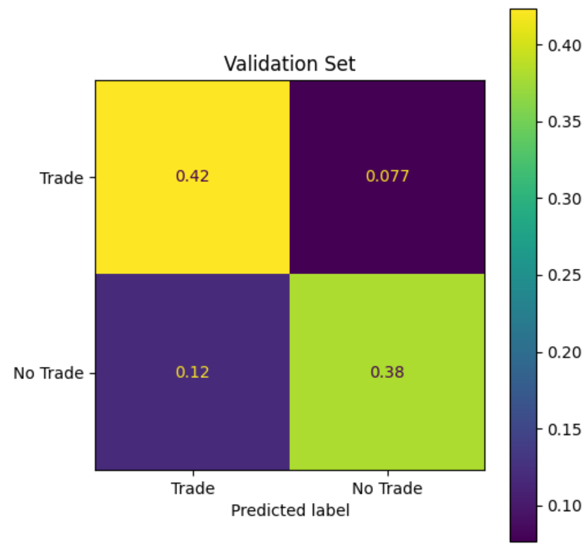
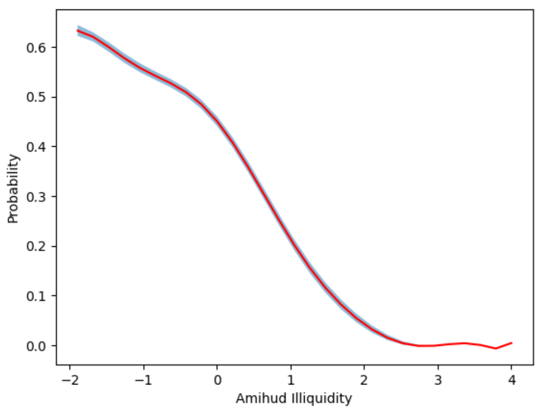
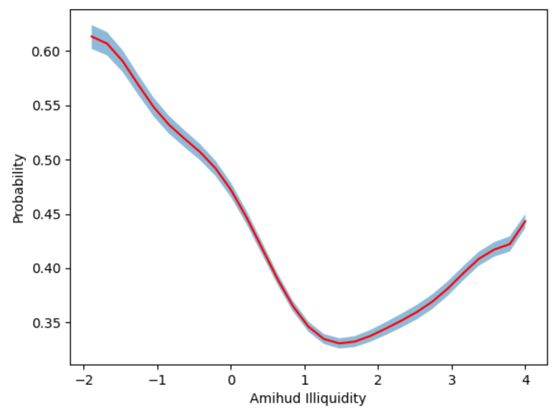


Figure 2: Confusion matrices for the extensive-margin trade model. Each cell shows the fraction of observations classified by the model as trades or non-trades, either correctly or incorrectly. Correct classifications appear along the diagonal. See section 4.4.1.

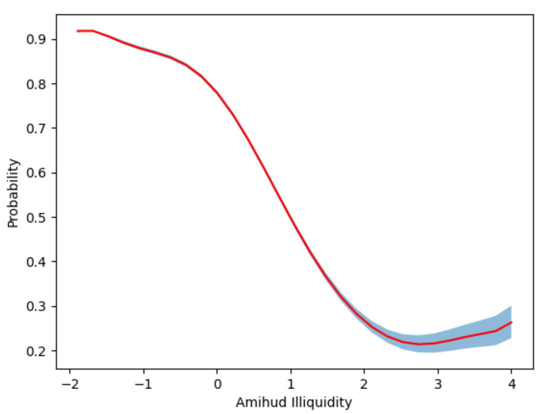
(a) Group 1: Small Cap Traders



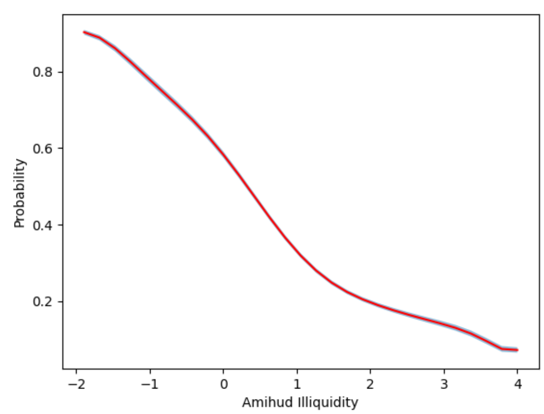
(b) Group 2: Quantitative Traders



(c) Group 3: Block Traders



(d) Group 4: Frequent Buyers



(e) Group 5: Anomaly Traders

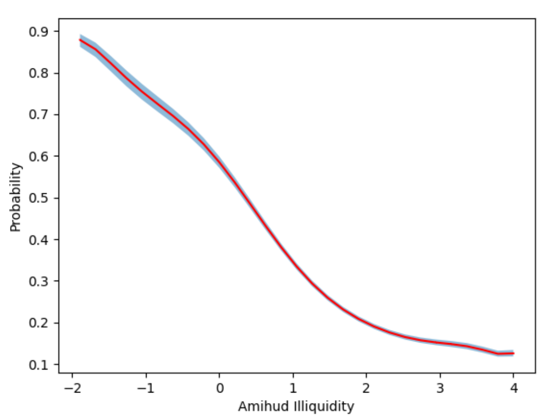
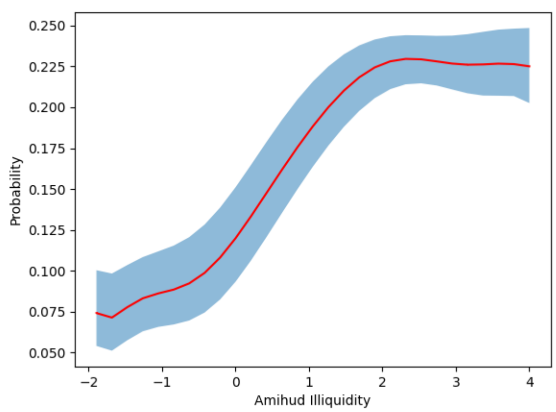
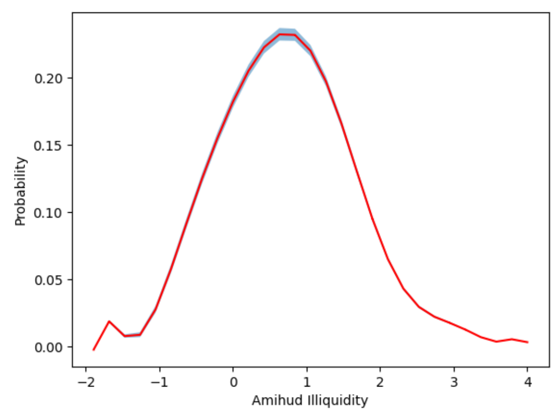


Figure 3: Response in predicted trade probability (extensive margin) to variations in day-of-trade Amihud illiquidity (mean ± 2 standard deviations). 95% confidence intervals shown in blue. See section 4.4.1.

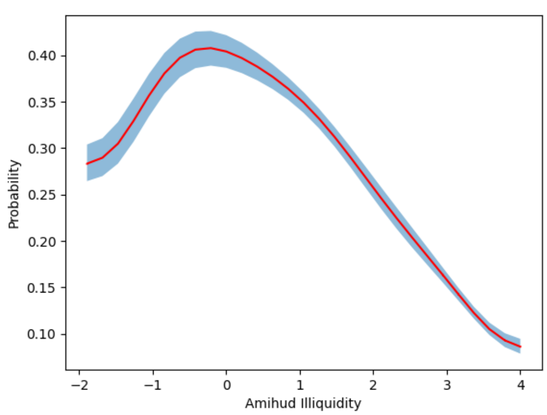
(a) Group 1: Small Cap Traders



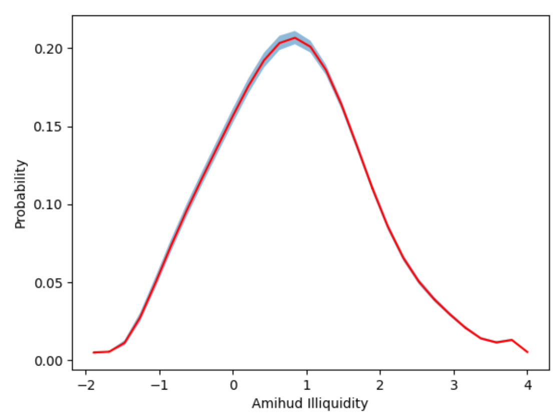
(b) Group 2: Quantitative Traders



(c) Group 3: Block Traders



(d) Group 4: Frequent Buyers



(e) Group 5: Anomaly Traders

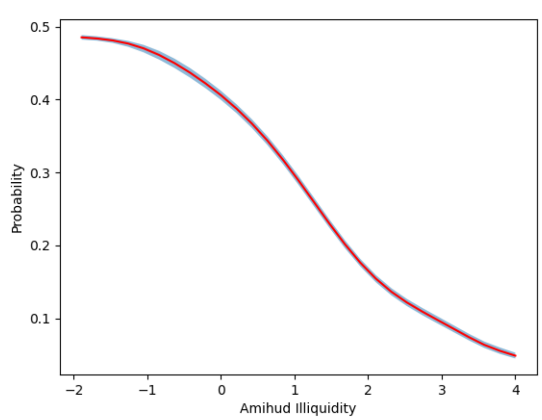
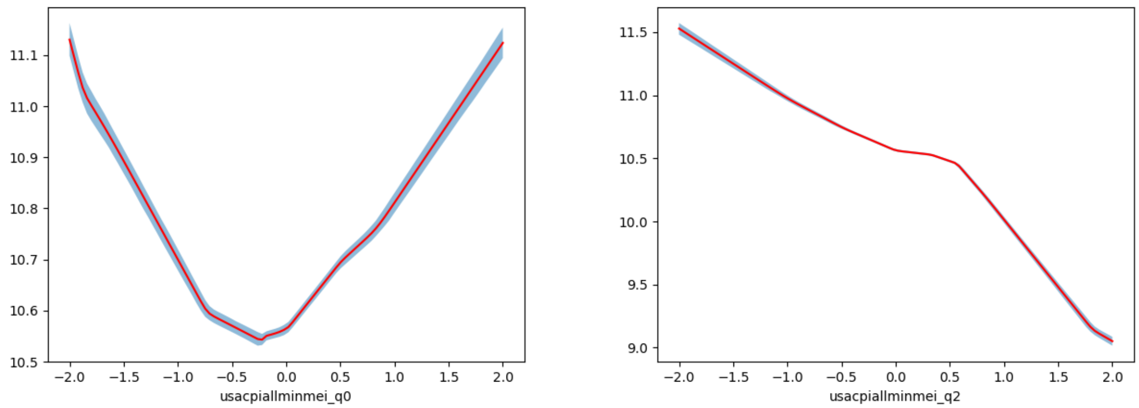
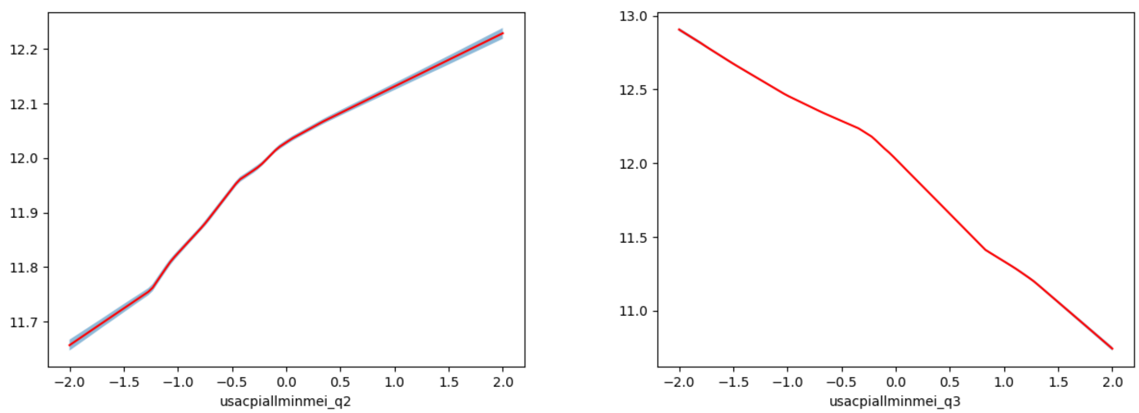


Figure 4: Response in predicted trade probability (extensive margin) to variations in past-quarter average Amihud illiquidity (mean $\pm$ 2 standard deviations). 95% confidence intervals shown in blue. See section 4.4.1.

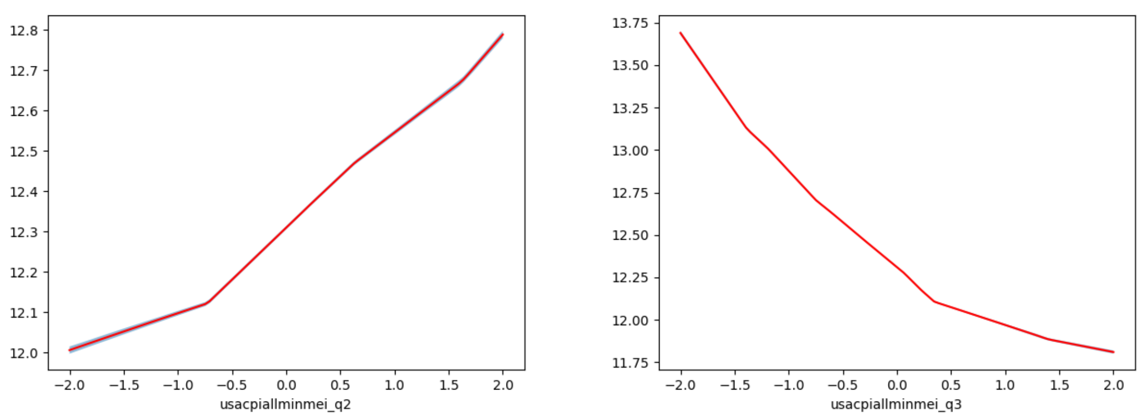
(a) Group 1: Small Cap Traders



(b) Group 2: Quantitative Traders



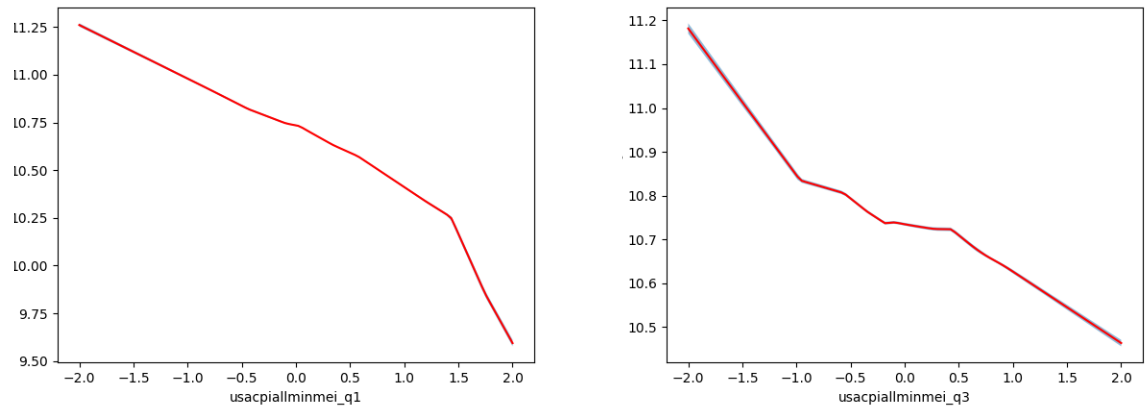
(c) Group 3: Block Traders



Continued on following page ...

... continued from previous page.

(a) Group 4: Frequent Buyers



(b) Group 5: Anomaly Traders

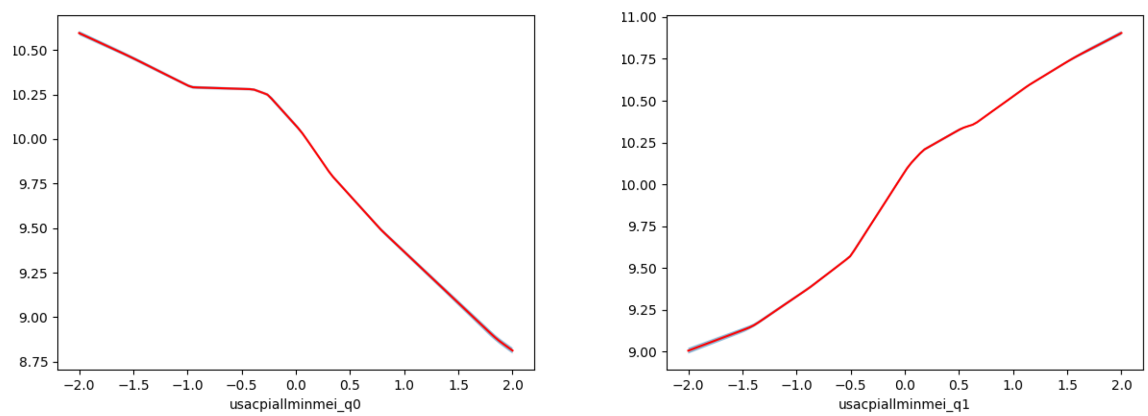


Figure 6: Response in predicted log dollar trade volume (intensive margin) to variations in different lags of the Consumer Price Index (mean $\pm$ 2 standard deviations). 95% confidence intervals shown in blue. See section 4.4.2.

Table 1: **Portfolio Group Characteristics**

This table reports estimated coefficients from a series of portfolio-level panel regressions of trade/stock characteristics on dummies for each portfolio group (see section 3). The first column shows the full-sample average of each feature as a baseline. Standard errors are two-way clustered at the month and portfolio levels. t-statistics are reported in brackets, and *, **, *** denote significance at the 1%, 5% and 10% levels, respectively.

	All Portfolios	Group 1 (SmallCap)	Group 2 (Quants)	Group 3 (Block)	Group 4 (Buyers)	Group 5 (Anomalies)
Num. portfolios	34,828	131	10,883	8,457	3,975	11,382
Panel A: Trade Characteristics						
Trades per month	318.09	-198.72***	832.38***	-326.45***	-172.01***	-443.54***
		(-5.03)	(6.66)	(-6.55)	(-4.17)	(-7.56)
Trade size (log)	10.49	0.31	0.65***	2.08***	0.31***	-2.53***
		(1.64)	(8.84)	(28.12)	(4.02)	(-38.70)
Trade concentration	0.11	-0.05***	-0.11***	-0.05***	-0.06***	0.17***
		(-7.71)	(-22.66)	(-9.61)	(-15.03)	(21.59)
Universe size	48.20	-13.04***	80.68***	-30.65***	-4.99***	-48.32***
		(-6.23)	(38.48)	(-22.65)	(-3.92)	(-40.17)
Buy fraction	0.51	-0.00	-0.03***	0.01	0.10***	-0.02**
		(-0.04)	(-6.73)	(1.59)	(27.49)	(-2.41)
Panel B: Stock Characteristics						
Market cap (log)	23.26	-1.58***	0.04	0.35***	-0.49***	-0.13***
		(-10.93)	(1.18)	(9.85)	(-10.79)	(-2.75)
B/M	1.05	-0.55***	-0.67***	-0.24	-0.50***	1.04***
		(-7.93)	(-7.82)	(-1.65)	(-7.24)	(6.99)
Profitability	0.25	-0.12***	-0.04***	-0.02	-0.04*	0.07**
		(-3.71)	(-2.65)	(-1.55)	(-1.91)	(2.53)
Investment	0.41	0.02	0.24***	-0.09***	-0.10**	-0.12**
		(0.13)	(3.96)	(-2.88)	(-2.36)	(-2.26)
1D ret (%)	0.03	0.28**	0.01	-0.06**	0.02	0.03
		(2.30)	(0.48)	(-2.28)	(0.76)	(0.48)
1M ret (%)	0.21	1.54*	-0.07	-0.51***	0.26	0.38
		(1.87)	(-0.37)	(-2.93)	(1.62)	(1.21)
1Y ret (%)	9.14	5.64	0.99	-1.12**	-0.45	0.08
		(1.44)	(1.49)	(-1.99)	(-0.65)	(0.09)
3Y ret (%)	22.12	-25.68***	-1.77*	-1.16	-3.14***	4.33**
		(-5.24)	(-1.76)	(-0.99)	(-3.45)	(2.25)
Volatility (%)	39.16	10.97***	4.86***	2.25***	-0.76	-6.47***
		(5.23)	(10.97)	(5.63)	(-1.53)	(-9.87)
Div. yield (%)	1.53	-0.81***	-0.24***	-0.12***	0.76***	0.06***
		(-10.44)	(-5.56)	(-2.63)	(10.13)	(0.71)
Market volume (log)	18.75	-0.73***	0.14***	0.23***	-0.54***	-0.12**
		(-5.52)	(4.42)	(6.80)	(-13.13)	(-2.15)

Continued on following page ...

... Continued from previous page

	All Portfolios	Group 1 (SmallCap)	Group 2 (Quants)	Group 3 (Block)	Group 4 (Buyers)	Group 5 (Anomalies)
Panel C: Industry Composition						
Agriculture	0.00	0.00	0.00*	0.00***	-0.00	-0.00***
		(0.16)	(1.69)	(4.92)	(-1.18)	(-3.26)
Mining	0.06	-0.00	-0.01***	-0.00	-0.02***	0.02***
		(-0.33)	(-2.68)	(-0.59)	(-6.81)	(2.66)
Construction	0.01	0.01***	-0.00	-0.00***	0.00***	0.00
		(3.81)	(-0.62)	(-3.38)	(3.88)	(0.90)
Manufacturing	0.39	-0.00	0.01**	0.01**	-0.07***	0.00
		(-0.08)	(2.05)	(2.33)	(-9.92)	(0.22)
Communications	0.09	-0.03***	0.01**	0.03***	-0.02***	-0.02***
		(-7.38)	(2.09)	(9.09)	(-6.83)	(-4.84)
Wholesale trade	0.02	0.00*	0.00	0.00***	-0.00	-0.00**
		(1.66)	(0.81)	(3.50)	(-0.01)	(-2.20)
Retail trade	0.07	-0.01*	0.00	0.01***	-0.02***	-0.00
		(-1.80)	(0.71)	(4.09)	(-8.93)	(-0.33)
Finance	0.18	-0.07***	-0.01***	-0.03***	0.14***	-0.01*
		(-5.19)	(-2.92)	(-8.44)	(14.86)	(-1.88)
Service	0.16	0.08***	0.01	-0.01***	-0.02***	0.01
		(8.00)	(1.02)	(-2.69)	(-3.41)	(1.17)
Public administration	0.00	0.02***	-0.00	-0.00***	-0.00**	-0.00**
		(5.79)	(-0.46)	(-3.38)	(-2.06)	(-2.06)

Table 2: **Extensive Margin: Variance Decomposition**

This table reports the percentage of explained variance attributable to different explanatory variables, grouped according to the type of information (panel A) or the timing of the information release (panel B). Explained variance is derived from a Bayesian feed-forward neural network trained on the extensive margin of trading decisions—that is, whether a particular stock drawn from the universe of ANcerno portfolios was traded on a particular day, or not (see section 4.2). Each column shows the results of training the network on a different subset of portfolios; the subsets derived from the hierarchical clustering algorithm shown in figure 1 and table 1. Panel C reports goodness-of-fit statistics for each model.

	All Portfolios	Group 1 (SmallCap)	Group 2 (Quants)	Group 3 (Block)	Group 4 (Buyers)	Group 5 (Anomalies)
			Panel A: Information Type			
Market Variables	74.60%	38.62%	31.29%	51.52%	34.66%	25.90%
Volume	71.84%	34.63%	28.95%	47.50%	32.16%	22.60%
Returns	2.32%	3.22%	1.10%	1.37%	0.86%	1.09%
Volatility	0.44%	0.77%	1.23%	2.65%	1.64%	2.21%
Firm Fundamentals	8.20%	18.90%	22.16%	16.84%	20.37%	23.51%
Market Capitalization	4.06%	2.94%	9.62%	7.58%	8.30%	12.55%
Dividends	0.17%	2.71%	0.39%	1.18%	0.69%	0.61%
Balance Sheet	2.43%	6.14%	6.66%	4.50%	6.31%	5.22%
Income Statement	0.93%	4.08%	3.17%	2.03%	2.93%	2.58%
Cash Flow Statement	0.61%	3.04%	2.31%	1.56%	2.14%	2.56%
Corporate News	3.07%	9.04%	8.30%	7.67%	8.05%	6.71%
Earnings Announcements	2.21%	5.77%	6.26%	4.37%	6.16%	4.58%
Num. Articles	0.47%	0.94%	1.25%	2.07%	0.76%	0.73%
News Impact Projection	0.12%	0.31%	0.21%	0.32%	0.28%	0.28%
Sentiment	0.26%	2.01%	0.58%	0.91%	0.85%	1.11%
Analyst Forecasts	1.17%	2.86%	3.09%	1.85%	2.06%	1.74%
Num. Analysts	0.44%	0.94%	1.82%	0.44%	0.83%	0.51%
Forecasts FQ1	0.25%	0.53%	0.44%	0.45%	0.47%	0.38%
Forecasts FY1	0.29%	0.72%	0.44%	0.46%	0.40%	0.49%
Forecasts FY2	0.20%	0.67%	0.38%	0.51%	0.36%	0.36%
Macroeconomy	12.96%	30.58%	35.18%	22.11%	34.85%	42.14%
Economic Activity	5.87%	11.56%	12.89%	9.45%	14.99%	18.10%
Employment	2.08%	4.99%	4.56%	3.67%	4.94%	6.93%
FOMC Announcements	1.00%	2.51%	2.54%	1.79%	3.53%	6.06%
Interest Rates & Indices	4.07%	11.85%	15.35%	7.33%	11.67%	11.38%

Continued on following page ...

... Continued from previous page

	All Portfolios	Group 1 (SmallCap)	Group 2 (Quants)	Group 3 (Block)	Group 4 (Buyers)	Group 5 (Anomalies)
Panel B: Information Release Timing						
Day of Trade	39.20%	34.63%	3.84%	35.45%	25.06%	21.26%
1 Day Before Trade	1.43%	2.23%	2.00%	1.53%	1.38%	3.48%
2 Days Before Trade	0.85%	2.48%	2.36%	1.15%	3.20%	2.74%
3 Days Before Trade	1.18%	1.53%	1.42%	1.27%	1.22%	1.62%
4 Days Before Trade	0.76%	0.85%	1.30%	1.02%	1.21%	1.93%
5 Days Before Trade	0.91%	1.48%	1.27%	1.15%	1.12%	3.10%
Panel C: Model Fit						
In-Sample Accuracy	0.709	0.772	0.860	0.803	0.792	0.804
In-Sample Adj. R^2	0.244	0.380	0.604	0.455	0.434	0.465
In-Sample Obs.	14,760,650	231,978	13,764,112	4,582,528	4,311,674	2,362,896
Out-of-Sample Accuracy	0.709	0.770	0.860	0.802	0.790	0.803
Out-of-Sample Adj. R^2	0.242	0.374	0.601	0.451	0.427	0.461
Out-of-Sample Obs.	1,640,074	25,776	1,529,346	509,170	479,076	262,544

Table 3: **Intensive Margin: Variance Decomposition**

This table reports the percentage of explained variance attributable to different explanatory variables, grouped according to the type of information (panel A) or the timing of the information release (panel B). Explained variance is derived from a Bayesian feed-forward neural network trained on the intensive margin of trading decisions—that is, the unsigned dollar volume conditional on a trade taking place (see section 4.2). Each column shows the results of training the network on a different subset of portfolios; the subsets derived from the hierarchical clustering algorithm shown in figure 1 and table 1. Panel C reports goodness-of-fit statistics for each model.

	All Portfolios	Group 1 (SmallCap)	Group 2 (Quants)	Group 3 (Block)	Group 4 (Buyers)	Group 5 (Anomalies)
			Panel A: Information Type			
Market Variables	30.55%	6.52%	24.29%	40.45%	21.57%	10.11%
Volume	29.80%	6.18%	23.64%	40.23%	21.36%	9.81%
Returns	0.42%	0.12%	0.38%	0.17%	0.13%	0.07%
Volatility	0.33%	0.22%	0.27%	0.05%	0.08%	0.22%
Firm Fundamentals	10.63%	8.65%	8.84%	5.43%	8.07%	3.88%
Market Capitalization	5.63%	0.50%	4.19%	2.08%	1.47%	0.65%
Dividends	0.17%	0.02%	0.24%	0.09%	0.07%	0.01%
Balance Sheet	3.51%	5.75%	3.29%	2.21%	5.13%	2.09%
Income Statement	1.16%	1.85%	0.95%	0.90%	1.21%	0.96%
Cash Flow Statement	0.16%	0.54%	0.17%	0.14%	0.19%	0.17%
Corporate News	2.30%	12.11%	2.27%	3.23%	2.69%	3.12%
Earnings Announcements	1.80%	11.73%	1.83%	2.88%	2.41%	2.93%
Num. Articles	0.31%	0.09%	0.27%	0.12%	0.11%	0.04%
News Impact Projection	0.04%	0.01%	0.03%	0.02%	0.01%	0.02%
Sentiment	0.16%	0.28%	0.13%	0.21%	0.17%	0.13%
Analyst Forecasts	0.19%	0.43%	0.19%	0.17%	0.17%	0.13%
Num. Analysts	0.07%	0.20%	0.07%	0.08%	0.08%	0.05%
Forecasts FQ1	0.03%	0.08%	0.04%	0.03%	0.03%	0.02%
Forecasts FY1	0.05%	0.08%	0.04%	0.03%	0.03%	0.03%
Forecasts FY2	0.04%	0.07%	0.04%	0.03%	0.03%	0.03%
Macroeconomy	56.05%	71.62%	64.14%	50.52%	67.19%	82.52%
Economic Activity	22.10%	45.32%	20.91%	23.91%	31.93%	36.29%
Employment	5.38%	4.98%	4.91%	4.90%	7.28%	7.56%
FOMC Announcements	2.65%	6.06%	3.32%	1.36%	5.04%	5.30%
Interest Rates & Indices	25.92%	15.25%	35.01%	20.36%	22.94%	33.38%

Continued on following page ...

... Continued from previous page

	All Portfolios	Group 1 (SmallCap)	Group 2 (Quants)	Group 3 (Block)	Group 4 (Buyers)	Group 5 (Anomalies)
Panel B: Information Release Timing						
Day of Trade	20.08%	9.31%	16.34%	34.83%	19.48%	10.31%
1 Day Before Trade	1.89%	7.76%	1.47%	1.05%	2.01%	2.84%
2 Days Before Trade	1.33%	7.60%	1.52%	0.57%	1.10%	3.74%
3 Days Before Trade	1.18%	5.73%	1.26%	0.88%	1.78%	2.64%
4 Days Before Trade	0.85%	3.08%	1.15%	0.93%	1.40%	2.08%
5 Days Before Trade	0.62%	3.70%	0.86%	0.56%	1.19%	2.74%
Panel C: Model Fit						
In-Sample Adj. R^2	0.533	0.252	0.493	0.309	0.194	0.200
In-Sample Obs.	7,380,325	115,989	6,882,056	2,291,264	2,155,837	1,181,448
Out-of-Sample Adj. R^2	0.531	0.193	0.492	0.306	0.189	0.189
Out-of-Sample Obs.	820,037	12,888	764,673	254,585	239,538	131,272

Table 4: **Market Volume: Variance Decomposition**

This table reports the percentage of explained variance attributable to different explanatory variables, grouped according to the type of information or the timing of the information release. Explained variance is derived from a Bayesian feed-forward neural network trained on log market dollar volume (see section 4.4.3). Out-of-sample predictive (adjusted) R-squared is reported in the final row.

Firm Fundamentals	63.38%	Market Variables	5.10%
Market Capitalization	55.52%	Returns	3.25%
Dividends	0.83%	Volatility	1.85%
Balance Sheet	4.12%	Macroeconomy	14.93%
Income Statement	1.99%	Economic Activity	5.86%
Cash Flow Statement	0.93%	Employment	1.49%
News	7.83%	FOMC Announcements	1.30%
Earnings Announcements	4.92%	Rates & Indices	6.27%
Num. Articles	2.69%		
News Impact Projection	0.11%	Information Release	
Sentiment	0.11%	Day of Trade	1.54%
Analyst Forecasts	8.45%	1 Day Before Trade	4.38%
Num. Forecasts	6.56%	2 Days Before Trade	1.48%
Forecasts FQ1	0.57%	3 Days Before Trade	1.38%
Forecasts FY1	0.64%	4 Days Before Trade	1.77%
Forecasts FY2	0.68%	5 Days Before Trade	1.60%
In-Sample Obs.	16,426,667		
In-Sample Adj R^2	0.866		
Out-of-Sample Obs.	1,825,186		
Out-of-Sample Adj R^2	0.816		

A Appendix

A.1 Variable Descriptions

Variable Name	Variable Description
	CRSP
	To be completed.

A.2 Clustering Methodology

A.2.1 Details on hierarchical agglomerative clustering algorithm

We perform the following clustering algorithm: let

$$X_i = \left[\frac{1}{2}X'_{i,Trading}; \frac{1}{4}X'_{i,Stock}; \frac{1}{4}X'_{i,Industry}\right]'$$

be the average features associated with portfolio i, where $x_{i,G}$ is the scaled features of set G so each feature has zero mean and unit standard deviation[18]. We apply $\frac{1}{2}$ weights on Trading features, $\frac{1}{4}$ on Stock features and $\frac{1}{4}$ on Industry compositions because we think the trading patterns are more important defining features for an institution than the stock characteristics of the stocks it trades, and thus we want the clustering algorithm to focus more on these trading features.

We use agglomerative hierarchical clustering algorithm to build hierarchy of clusters. A detailed description of the algorithm can be found in A.2.1. The algorithm starts from individual institutions and in each step groups the most similar portfolios together to form a cluster. We compute the pair-wise distance between X_i of portfolio i and X_j of portfolio j using cosine distance measure:

$$D(i,j) = 1 - \frac{X_i \cdot X_j}{||X_i||_2||X_j||_2}$$

where $||\cdot||_2$ is the L^2 norm. A lower value of $D(i,j)$ means portfolio i and j are more similar. The algorithm proceeds by merging most similar clusters together to form bigger clusters, where the distance between clusters A and B is the computed as the unweighted average

[18] We also winsorize Stock features at 0.1% and 99.9% percentile to remove outliers.

linkage:

$$D(A,B) = \frac{1}{|A| \cdot |B|} \sum_{i \in A} \sum_{j \in B} D(i,j)$$

where $|\cdot|$ is the cardinality of the set.

We use the `scikit-learn.cluster.AgglomerativeClustering` class in python to implement the algorithm. Let X_i be the average features associated with portfolio i in full sample I, $D(i,j) = 1 - \frac{X_i \cdot X_j}{||X_i||_2 ||X_j||_2}$ for two portfolios i and j (cosine distance), and $D(A,B) = \frac{1}{|A| \cdot |B|} \sum_{i \in A} \sum_{j \in B} D(i,j)$ (group average linkage), the clustering algorithm is described in Algorithm 1:

Algorithm 1 Hierarchical Agglomerative Clustering

Initialize for each X_i its cluster singleton G_i
Define a set of all clusters $H = \{G_i\}_{i \in I}$
while $|H| > 1$ **do**
  Choose G_i and G_j such that $D(G_i, G_j) = \min_{G_{i'}, G_{j'} \in H} D(G_{i'}, G_{j'})$
  Merge $G_{i,j} = G_i \cup G_j$
  Add $G_{i,j}$ to set H
  Remove G_i and G_j from set H
end while

It is obvious that for N data points, we will perform $N-1$ merges to reach the root. The dendrogram keeps track of those merges, and we use the dendrogram to decide where should we stop merging clusters. We stop merging once the minimum distance between two clusters G_i and G_j in set H is less than 75% of the max distance between any two data points, i.e. $\min_{G_i, G_j \in H} \leq 0.75 \max_{i,j} D(X_i, X_j)$.

A.2.2 Preliminary clustering results and removal of repeated institutions

«To be completed.»

We perform a preliminary clustering algorithm on the sample after applying the filters described in section 2. Figure A.1 shows the dendrogram from hierarchical clustering algorithm on the heatmap of pairwise cosine distances. We see that there is a set of portfolios gets clustered together (red cluster) that has very pairwise distances. After inspecting those institutions, we find that they all have more than 95% of the trading volume in the finance industry (SIC code H) and mostly comes from the same clientcode. This observation makes us suspect that these were duplicated records in the Ancerno data set. Due to the potential duplication issue and the atypical industry composition, we do not think these records describe the trading behavior of a typical institution, and thus remove those institutions in this cluster and re-estimate the clustering result.

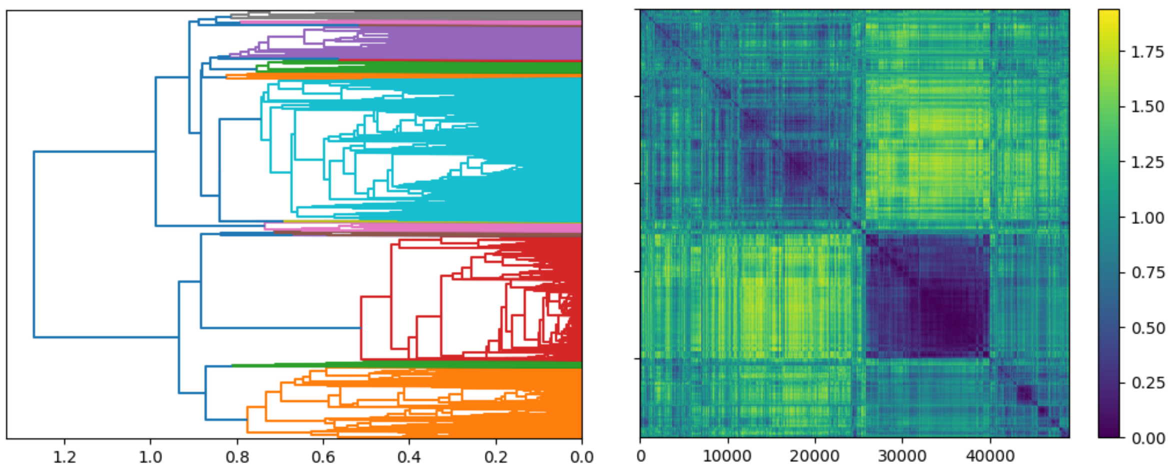


Figure A.1: Heatmap of pairwise distances ordered by the clustering results and the dendrogram, before removing the duplicated records.

A.3 Deep Learning Methodology

«To be completed.»

We use a Bayesian Neural Network (BNN) to model institutional trader's decisions. A BNN is similar to a standard feed forward neural network, except instead of learning the parameters of weights and biases that connect the hidden nodes, it learns a distribution over the weights and biases. A stylized comparison between a standard neural network and a Bayesian neural network is depicted in Figure A.2. A BNN has potential advantages over a regular feed forward network because by specifying a prior distribution and shrinking the posterior distribution of the parameters towards the prior, it prevents overfitting. Also we can construct confidence intervals of the predictions by sampling parameters from the posterior distribution.

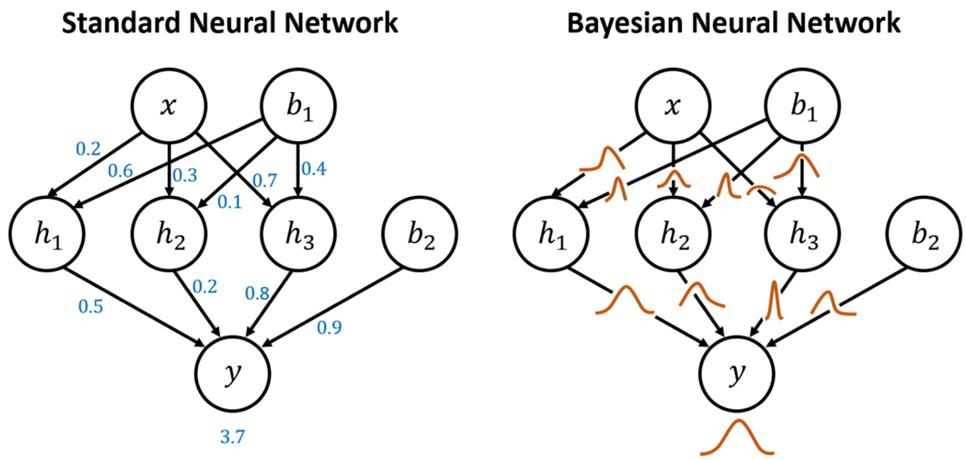


Figure A.2: A stylized comparison between a regular feed forward neural network and a Bayesian neural network. Source: Blundell et al. (2015).

We use a two-hidden-layer Bayesian neural network as our main model structure. The first hidden layer has 100 hidden nodes, and the second hidden layer has 10 hidden nodes. All connections are dense. We choose a two-hidden-layer structure such that with a wide first hidden layer, we can capture a wide range of interactions and nonlinearities among the input variables, and with a parsimonious second hidden layer we want the model to condense predictability into a number of high-level aggregations. We specify the prior for all weights and biases to be independent standard normal distributions and we use variational inference to learn the posterior distribution by approximating it using independent normal distributions, with trainable parameters being their means and variances.

We use `binary cross-entropy` as our loss function for categorical predictions (extensive margin), and `mean squared errors` for continuous variable predictions (intensive margin). We use a batch size of 10000, and we use `Adam` training optimizer with learning rate set to 0.001 as the default in `keras`. We use early-stopping based on the 10-epoch moving average changes by less than 0.0001. When such criteria is reached, we set a patience of 10 so that it stops training 10 epochs later, if the loss doesn't increase within those 10 epochs. The reason for this is to prevent non-monotonic behavior in training losses. We also impose a L2 regularization on the parameter learning, which shrinks the trainable parameters (mean and variances in the independent normal distributions for the posterior) towards 0.

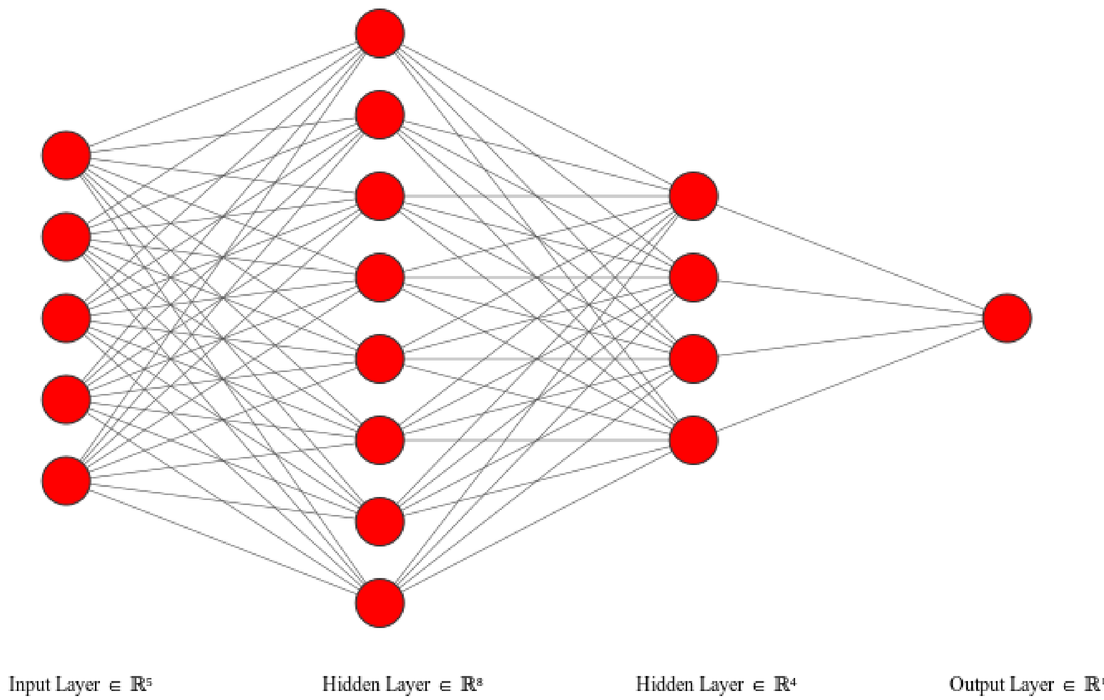


Figure A.3: Illustration of feed-forward neural network