Order Flow and Cryptocurrency Returns*

Alexia Anastasopoulos
U of Guelph

Nikola Gradojevic
U of Guelph

Fred Liu
U of Guelph

Alex Maynard
U of Guelph

Ilias Tsiakas
U of Guelph

November 2024

Abstract

We assess the information content of order flow for the cross-section of cryptocurrency returns. Our analysis is based on a set of international order flows denominated in 11 major currencies that reflect world order flow. We find that world order flow has strong explanatory and predictive power for cryptocurrency returns. Order flow tends to dominate economic fundamentals for out-of-sample prediction, especially in the context of non-linear machine learning models, and its performance cannot be explained by limits to arbitrage. Overall, our findings indicate that order flow has a permanent effect for cryptocurrency returns.

JEL classification: C58, G11, G15.

Keywords: Order Flow, Cryptocurrencies; Machine Learning; Forecasting; Portfolio Sorts.

***Acknowledgements:** The authors are grateful for useful comments to Adam Bozman, Zhenzhen Fan, Haibo Jiang, Vikas Kakkar, Hong Li, Jiacheng Liu, Steven Riddiough, Razvan Sufana, Yiguo Sun and participants at the Generative AI conference in Montreal, the CEA 2024 conference in Toronto and seminars at the University of Guelph. This research was supported by the Social Sciences and Humanities Research Council of Canada (grant number 430-2024-00947 for Fred Liu, 435-2017-0582 for Alex Maynard and 430-2024-00258 for Ilias Tsiakas) and the Digital Research Alliance of Canada Resources for Research Groups 2022 competition (ID 4286). Declarations of interest: none. *Corresponding author*: Ilias Tsiakas, Department of Economics and Finance, Lang School of Business and Economics, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Email: itsiakas@uoguelph.ca.

1 Introduction

The emergence of the cryptocurrency market over the past decade has been accompanied by remarkable growth. Market capitalization has increased from $11 billion US dollars (USD) in 2014 to a peak of $2.9 trillion USD in 2021. Similarly, daily trading volume has risen from $40 million USD in 2014 to a peak of $308 billion USD in 2021. Unlike traditional foreign exchange markets, where trading volume is concentrated among a few major dealers (Menkhoff et al., 2016), cryptocurrency markets consist of independently owned, non-integrated exchanges that operate in parallel across countries and multiple currencies (Makarov and Schoar, 2020). Furthermore, cryptocurrency markets are relatively less liquid and face more pervasive asymmetric information compared to traditional markets (Bianchi, Babiak and Dickerson, 2022). These unique features make cryptocurrency markets a novel setting for studying the role of order flow in the price formation process.[1]

In this context, this paper contributes to the emerging cryptocurrency literature by assessing the information content of order flow for explaining and predicting cryptocurrency returns. Specifically, we address four related questions. First, can order flow explain the cross-section of cryptocurrency returns? Second, does order flow convey predictive information for the cross-section of cryptocurrency returns? Third, do machine learning techniques (linear and non-linear) improve the out-of-sample predictive power of order flow relative to standard predictive approaches? Finally, fourth, can the predictive power of order flow be explained by economic restrictions such as limits to arbitrage? Addressing these questions is relevant to the study of market microstructure and paves the way for a deeper investigation into

[1] According to coinmarketcap.com, the market capitalization of the cryptocurrency market on January 1, 2014 was $10.8 billion USD and subsequently peaked on November 8, 2021 at $2.9 trillion USD. Daily volume was $40.2 million USD on January 1, 2014 and peaked on April 16, 2021 at $308 billion USD. This trading volume was spread across more than 700 exchanges operating in various countries and transacting in over 50 fiat currencies.

the economic mechanisms driving order flow and price discovery in cryptocurrency markets. Although for traditional financial markets these questions have been addressed by a long line of research, for cryptocurrencies they remain open. In this paper, we bridge this gap in the literature by focusing on the information conveyed by order flow in the cryptocurrency market.[2]

Order flow is a measure of the net demand for a particular cryptocurrency defined as the value of buyer-initiated transactions minus the value of seller-initiated transactions. In general, there are three economic explanations on the extent to which order flow may be related to cryptocurrency returns. First, it is possible that order flow is in fact unrelated to cryptocurrency returns. In this case, it is economic fundamentals, not order flow, that may explain and predict cryptocurrency returns. We refer to this view as the "no-predictability view" and it is essentially the null hypothesis tested in our empirical analysis.

Second, it is possible that order flow has a transitory effect on cryptocurrency returns. In this case, order flow has a temporary but not a permanent effect. The market microstructure literature associates transitory effects with changes in liquidity, price pressure and temporary preference shocks (see, e.g., Froot and Ramadorai, 2005 and Menkhoff et al., 2016). We refer to this view as the "transitory view."

Finally, third, it is possible that order flow has a permanent effect on cryptocurrency returns. The market microstructure literature associates permanent effects with asymmetric information among market participants so that trades convey information with a persistent effect on security prices (see, e.g., Glosten and Milgrom, 1985, Kyle, 1985, and Hasbrouck,

[2] For the theory and empirical evidence on the effect of order flow on fiat exchange rates, see among others, Evans and Lyons (2002, 2005), Berger et al. (2008), Evans (2010), Rime, Sarno and Sojli (2010), Evans and Rime (2012) and Menkhoff et al. (2016). For US treasury markets see Brandt and Kavajecz (2004), for the S&P 500 futures market see Deuskar and Johnson (2011), and for NYSE stocks see Chordia, Roll and Subrahmanyam (2002), Goyenko, Holden and Trzcinka (2009) and Hendershott and Menkveld (2014).

1988, 1991). According to this view, if trades convey information about future economic fundamentals that is not currently known by all market participants, then order flow acts as the key vehicle that impounds this information into cryptocurrency returns through the process of price discovery (see, e.g., Evans and Lyons, 2002, 2005 and Menkhoff et al., 2016). We refer to this view as the "permanent view."

Our empirical analysis employs daily data from a cross-section of 82 cryptocurrencies for the sample period of January 1, 2018 to June 30, 2022. We use a rich dataset of order flows denominated in 11 major currencies, including the G10 currencies plus South Korea, which is a prominent centre for cryptocurrency trading. We then form a measure of world order flow by aggregating the 11 international order flows. Accordingly, our analysis assesses the information content of world order flow and disaggregated international order flows, including US order flow. In what follows, we discuss our empirical approach and main findings step by step.

We begin by estimating panel regressions to assess whether order flow can explain and predict cryptocurrency returns in sample. The panel regressions condition on combinations of world order flow and the 11 international order flows together with a set of coin-specific control variables and economic fundamentals. We estimate contemporaneous and predictive panel regressions at the daily and weekly frequency.

With respect to the in-sample evidence, our main finding is that world order flow has significant explanatory and predictive power for the cross-section of cryptocurrency returns. The contemporaneous relation of world order flow and cryptocurrency returns is positive and highly significant. World order flow together with control variables can explain about 11% of daily returns and 20% of weekly returns.

For predictive regressions, it is crucial to separate two opposing components of order flow: a transitory component that reverses in the short term and a permanent component that persists over the long term. Following Bianchi, Babiak and Dickerson (2022), we use lagged returns as a proxy for short-term reversal. In other words, the predictive regressions condition on lagged order flow, while controlling for lagged returns. We find that the portion of lagged order flow that is uncorrelated with lagged returns has a positive and significant predictive relation with future returns. The positive coefficient of world order flow for both daily and weekly returns provides empirical support for the permanent view.

In considering all international order flows, world order flow consistently exhibits the highest and most significant coefficient compared to all international order flows. In contemporaneous regressions, the US and South Korean order flows are also significant. In predictive regressions, however, the significance of US and South Korean order flows disappears in the presence of world order flow. Overall, these findings indicate that world order flow is a more powerful and significant predictor than any of the individual international order flows in the context of in-sample panel regressions.

Next, we assess the out-of-sample predictive ability of order flow for daily cryptocurrency returns using machine learning (ML) techniques (see, e.g., Gu, Kelly and Xiu, 2020, Cakici et al., 2024, Fieberg et al., 2024 and Filippou, Rapach and Thimsen, 2024). These methods are ideally suited for out-of-sample forecasting because they emphasize techniques for variable selection and dimension reduction, which can accommodate a large set of predictors as well as a richer specification of functional forms. Therefore, ML forecasting allows us to condition on the 11 international order flows and additional economic fundamentals for prediction of the one-day ahead cryptocurrency returns. We estimate a set of linear and non-linear ML models, which include the following specifications: ridge regression (RR), lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), stochastic gradient boosted regression trees (SGB), and neural networks with 1-4 hidden layers (NN1-NN4). In addition, we compute forecast combinations across the linear models (L-Mean), and the non-linear models (NL-Mean).[3]

In terms of the out-of-sample statistical analysis, we find that non-linear ML forecasts which condition on all order flows consistently outperform: (1) linear models that condition on all order flows, (2) linear and non-linear models that condition on economic fundamentals, and (3) the zero-forecast benchmark. The best performer is the SGB model that conditions on all order flows with an out-of-sample R^2 of 0.66%. In short, non-linear ML models conditioning on order flow have significant out-of-sample predictive ability for cryptocurrency returns.

In addition to the statistical analysis, we also assess the out-of-sample economic value of conditioning on order flow using portfolio sorts. Our approach is based on generating cryptocurrency portfolios sorted on lagged values of either world order flow or the ML return forecast of models conditioning on all international order flows. The ML approach enables us to exploit the predictive information of all international order flows. Specifically, on each day, we allocate the 82 coins into quintile portfolios and assess the one-day ahead performance of the five quintile portfolios as well as the long-short (top-minus-bottom) portfolio.

In assessing portfolio performance, we first demonstrate that portfolios sorted on world order flow exhibit strong performance at the weekly frequency. For example, the long-short mean return spread when conditioning on world order flow is equal to 1.61% per week, it is significant (t-stat=2.13) and exhibits an annualized Sharpe ratio of 1.68. Furthermore, the long-short portfolio based on world order flow delivers a weekly alpha of 1.44% (t-stat=2.02) with respect to the cryptocurrency three-factor model (Liu, Tsyvinski and Wu, 2022). For daily portfolio sorts, we first orthogonalize lagged world order flow with respect to lagged returns to remove the short-term reversal component. The long-short portfolio based on orthogonalized world order flow delivers an alpha of 0.30% per day (t-stat=2.07) and an

[3] Although most of our analysis is performed for both daily and weekly returns, the out-of-sample ML forecasting is only performed for daily returns. Due to the short sample period, which is specific to the cryptocurrency market, we do not have enough data to perform weekly ML forecasting.

annualized Sharpe ratio of 1.34.

Importantly, portfolio performance improves substantially when the portfolio sorts are based on machine learning models that generate forecasts which condition on all order flows. Although all ML models perform well, the non-linear models conditioning on order flow perform the best. For example, the best model (SGB) delivers a daily long-short alpha of 0.79% (t-stat=5.67) and an annualized Sharpe ratio of 3.63. Moreover, non-linear ML models conditioning on order flow outperform all ML models conditioning on economic fundamentals. Furthermore, non-linear ML models conditioning on order flow outperform leading ML-based benchmarks in the literature for forecasting cryptocurrency returns. For example, the NL-Mean model in Filippou, Rapach and Thimsen (2024), which conditions on 54 characteristics but excludes order flow, generates a Sharpe ratio of 2.57, while the same model conditioning on order flow in this paper achieves a Sharpe ratio of 3.45. Taken together, these findings indicate that there is significant additional economic value in conditioning on all international order flows in the context of out-of-sample machine learning forecasts.

Motivated by Avramov, Cheng and Metzker (2023), we conduct an analysis on the profitability of our models in the presence of economic restrictions such as short-selling constraints, transaction costs and limits-to-arbitrage. To be more specific, in the cryptocurrency market, it may be difficult or even impossible to short certain coins. Therefore, to generate a realistic trading strategy that avoids shorting, we may focus on the long portfolio rather than the long-short portfolio spread. In light of this, we show that compared to the long-short portfolios, the long portfolios exhibit a similar risk-adjusted return, lower Sharpe ratio, lower turnover and hence a higher break-even transaction cost. For example, the long portfolio of the NL-Mean combination exhibits a daily alpha of 0.81% (t-stat=2.99), an annualized Sharpe ratio of 1.88 and a break-even transaction cost of 1.02% per day. In contrast, the same model for the long-short portfolio delivers an alpha of 0.76% (t-stat=5.57), an annualized Sharpe ratio of 3.52 and a break-even transaction cost of 0.48% per day. We conclude, therefore, that ML models that condition on order flow generate high economic value for long-only investors who pursue a realistic trading strategy investing in cryptocurrencies.

Finally, we study whether the profitability of ML models conditioning on order flow can be explained by limits to arbitrage. To test this, we create an arbitrage cost index based on multiple indicators known to capture limits to arbitrage. We find that orthogonalized order flow and non-linear ML models conditioning on order flow produce significant economic value for coins with low arbitrage costs. In contrast, models conditioning on economic fundamentals perform well only for coins with higher arbitrage costs, indicating that their economic value can be explained by limits to arbitrage. Overall, the inability of limits to arbitrage to explain the positive predictive ability of order flow further supports the permanent view of the effect of order flow on cryptocurrency returns.

In summary, our analysis provides compelling empirical evidence to show that order flow has strong and economically valuable out-of-sample predictive power for cryptocurrency returns. Our findings indicate that order flow has a permanent effect for cryptocurrency returns since this effect is consistently positive and stronger for weekly compared to daily returns. Non-linear ML models conditioning on order flow outperform ML models conditioning on economic fundamentals as well as leading ML benchmarks that do not include order flow. Forecast combinations of non-linear ML models also exhibit strong performance. Finally, our empirical findings are robust to economic restrictions such as short-selling constraints and high transaction costs that characterize the cryptocurrency market. In short, order flow matters for explaining and predicting cryptocurrency returns.

Our empirical analysis is related to the study of Makarov and Schoar (2020), who examine the relation between bitcoin returns and order flow. However, the focus of Makarov and Schoar (2020) is on assessing arbitrage opportunities across cryptocurrency exchanges. Therefore, it substantially differs from our study since it only focuses on one coin and the analysis is exclusively in-sample. Our portfolio sort analysis is related to Bianchi, Babiak and Dickerson

(2022), who study cryptocurrency portfolio sorts on lagged returns and trading volume shocks (denominated in US dollars), which differ from order flow. Our approach is distinct from the extant literature as indicated by the use of a large cross-section of coins, a large cross-section of order flows denominated in different currencies, the out-of-sample focus of assessing predictability, the use of machine learning methods and the economic evaluation based on portfolio sorts.

The remainder of the paper is organized as follows. In the next section, we describe the cryptocurrency data and define order flow. In Section 3, we describe the in-sample panel regression framework for explaining and predicting cryptocurrency returns. The framework for out-of-sample prediction based on machine learning models is examined in Section 4. In Section 5, we assess the economic value of order flow based on portfolio sorts and, in Section 6, we discuss the effect of economic restrictions. Finally, we conclude in Section 7. An Online Appendix provides further details on the machine learning models and additional results.

2 Data

2.1 Cryptocurrency Returns

Our empirical analysis employs data from a cross-section of 82 cryptocurrencies for the sample period of January 1, 2018 to June 30, 2022. For each cryptocurrency (crypto or coin), we collect daily US dollar (USD) prices from coinmarketcap.com (henceforth CMC) at 00:00am GMT time. The CMC data aggregates prices from over 700 exchanges and is a reliable source of data that is used extensively in the cryptocurrency literature (for a detailed description of the CMC data, see, e.g., Liu and Tsyvinski, 2021 and Liu, Tsyvinski and Wu, 2022). We exclude weekends and US holidays from the daily sample.

The CMC database contains data on thousands of coins, though most have low market capitalizations and/or limited historical data. To obtain the final cross-section of 82 coins, we impose three criteria to ensure the liquidity of our crypto sample. First, we require that coins must have a market capitalization of greater than 1 million USD on the last day of the sample period. Second, coins must be continuously traded with a non-zero price and non-zero volume in each time period of our sample. Finally, third, we exclude all stable coins from the cross-section. As a result, we are left with a balanced panel of 82 coins.[4]

We use daily prices to compute returns:

$$r_{i,t} = \frac{P_{i,t}}{P_{i,t-1}} - 1, \tag{1}$$

where $P_{i,t}$ is the price of coin i at time t. Weekly crypto returns are defined over the interval beginning on Saturday at 00:00am GMT time and ending the following Friday at 11:59pm GMT for a total of seven days including weekends.

2.2 Order Flow

Our primary variable of interest is order flow, which is defined as the log difference between buyer-initiated and seller-initiated transaction volume denominated in a particular currency over a period of time. We compute order flow using data on signed volume (buy-volume and sell-volume) obtained from cryptocompare.com (henceforth CC). CC is a reliable source for cryptocurrency volume data (see, e.g., Bianchi, Babiak and Dickerson, 2022). CC aggregates

[4] We set January 2018 as the beginning of the sample to allow for a reasonable number of observations across time, while maintaining a reasonably large cross-section of coins. Starting earlier than 2018 severely limits the breadth of the balanced cross-section. Notably, the number of coins in our sample is similar to other studies using ML to predict crypto returns. For example, Filippou, Rapach and Thimsen (2024) include 41 coins, while the number of coins in Cakici et al. (2024) dynamically ranges from approximately 30 to 300.

information from over 300 exchanges and provides data on signed volume that is not available on CMC. Throughout our analysis, we use crypto data from CMC except for signed volume (i.e., order flow) that is only available from CC.

A critical advantage of the CC data is that they provide signed volume data denominated in several currencies. In our analysis, we compute the order flow denominated in each of the G10 currencies plus South Korea. We add South Korea to the sample because it is a prominent centre for trading cryptocurrencies outside of the G10. We designate this group as the G11 currencies, which include the following: US dollar (USD), Euro (EUR), British pound (GBP), Japanese yen (JPY), Swiss frank (CHF), Canadian dollar (CAD), Australian dollar (AUD), New Zealand dollar (NZD), Norwegian krone (NOK), Swedish krona (SEK) and the Korean won (KRW).

We compute a measure of order flow that is comparable across cryptocurrencies. Following Menkhoff et al. (2016), we standardize each series as follows:

$$OF_{i,t} = \frac{of_{i,t}}{\sigma(of_{i,t-29:t})}, \tag{2}$$

where $OF_{i,t}$ denotes the standardized order flow of coin i at time t, $of_{i,t}$ is the original order flow, and $\sigma(of_{i,t-29:t})$ is the order flow volatility over the last 30 days.

In addition to the 11 international order flows, we also define an aggregate measure that we term world order flow (OF^W). World order flow is computed by aggregating all 11 buy-volumes and all 11 sell-volumes, taking the log difference, and performing the same standardization as applied to the disaggregated order flows.

To assess the contribution of each cryptocurrency to world order flow, we conduct a variance decomposition as follows:

$$VC_i^W = \frac{\beta_i^2 Var(OF_i)}{\sum_{i=1}^{11} \beta_i^2 Var(OF_i)}, \tag{3}$$

where VC_i^W is the variance contribution of coin i to world order flow, β_i is the estimated coefficient of each order flow when regressing world order flow on its 11 components, and $Var(OF_i)$ is the variance of each order flow.

The variance decomposition results are reported in Table 1. As expected, for daily data, the US order flow has the highest variance contribution at 39%. The second contribution is from South Korea at 24%. This motivates our decision to add South Korea to the G10 order flows. The euro is third at 10% followed by the British pound at about 6%. Overall, world order flow is dominated by few countries since the top two contribute 63% to the total variance, whereas the top three contribute 73%. The weekly VCs are similar to the daily VCs but now the top three order flows contribute 81% to world order flow.

2.3 Control Variables

We use a combination of coin-specific variables and economic fundamentals as controls in the empirical analysis. The coin-specific variables include market capitalization (henceforth market cap), total volume and volatility. Total volume is the sum of a coin's 24-hour trading volume across all exchanges. Volume and market cap are expressed in logs. Volatility is defined as the log difference between a coin's high price and low price on a given day. As mentioned earlier, all coin-specific data other than order flow are obtained from CMC.

The economic fundamentals used in our analysis include the following. The short rate is defined as the 3-month US T-bill rate. The term spread is defined as the difference between the 10-year and 3-month US treasury rates. The default spread is defined as the difference between the BAA- and AAA-rated corporate bond yields. All interest rate data are obtained from the Federal Reserve Bank of St. Louis. The TED spread is defined as the difference between the 3-month T-Bill rate and the 3-month LIBOR rate, and is obtained from the LSEG Eikon database. The S&P 500 index returns (including dividends) are from the

CRSP database. The MSCI global index returns are taken from msci.com. The VIX index is obtained from CBOE.com.

2.4 Summary Statistics

In Tables A2 and A3 of the Online Appendix, we report summary statistics for all coin-specific variables: returns, US and world order flow, market cap, volume and volatility. The tables report results for all 82 coins as well as for the top 10 coins by market cap. In brief, our main findings are as follows: (1) daily cryptocurrency returns are extremely volatile with high positive skewness and very high kurtosis; (2) daily standardized order flow is well behaved relative to returns since it exhibits low skewness and moderate kurtosis; (3) the coins exhibit higher volatility, skewness and kurtosis in the time-series than the cross-section; and (4) the average full-sample cryptocurrency return is equal to 0.21% per day or 52.9% per year.

3 In-Sample Analysis

Our empirical approach is based on panel regressions for in-sample analysis and machine learning techniques for out-of-sample analysis. This section presents the in-sample panel regression results on the effect of order flow in explaining and predicting cryptocurrency returns.

3.1 Panel Regressions with World Order Flow

3.1.1 Contemporaneous Regressions

Can world order flow explain cryptocurrency returns? We address this question by estimating panel regressions that condition on contemporaneous world order flow and the control variables. The panel regressions are estimated using the full sample ranging from February 14, 2018 to June 30, 2022, and include coin fixed effects. The results for both daily and weekly data are reported in Table 2.[5]

Our main empirical finding is that world order flow has a positive contemporaneous relation to returns, which is highly significant in all cases. Therefore, world order flow does explain cryptocurrency returns. For example, in the presence of control variables, a one standard deviation increase in world order flow is associated with a 2.2% increase in returns at the daily frequency and a 4.0% increase at the weekly frequency. These results are highly significant since the Newey and West (1987) t-statistic is equal to 37.62 for daily regressions and 16.97 for weekly regressions. Our results align with Makarov and Schoar (2020), who find a positive contemporaneous relation between order flow and Bitcoin returns in daily regressions. We show that this positive relation holds across a broad cross-section of cryptocurrencies and demonstrate that it also persists in weekly regressions.

It is interesting to note that the weekly coefficient is larger than the daily coefficient and this is also true for the $\bar{R}^2$. Specifically, the $\bar{R}^2$ is equal to 11.0% for daily regressions and 19.9% for weekly regressions. A possible explanation for the stronger weekly results is that order flow at the daily horizon may include uninformed traders (also referred to as liquidity or noise traders). Since the trades of uninformed traders are driven primarily by liquidity preferences than information, the effect of these traders on the price tends to be transitory.

[5] In the Online Appendix, we also report detailed results for US order flow.

Aggregating order flow over one week mitigates the effect of market microstructure noise. Consequently, order flow may explain a greater variation of price movements at the weekly than the daily horizon.[6]

In addition to order flow, the control variables are also highly significant. For daily regressions, all control variables but the market cap are significant. For weekly regressions, all control variables are significant with no exceptions. In conclusion, world order flow has strong and highly significant explanatory power for the cross-section of cryptocurrency returns in the presence of crypto-specific control variables and standard economic fundamentals.

3.1.2 Predictive Regressions

Can world order flow predict cryptocurrency returns? We address this question by estimating the same panel regressions as previously but with lagged values for all independent variables. Specifically, we condition on the same control variables plus lagged cryptocurrency returns. Short-term returns exhibit a negative autocorrelation due to transitory effects such as price pressure and other short-term liquidity effects, resulting in reversal, i.e., a negative relation between lagged and future returns. Table 2 shows that lagged order flow is positively contemporaneously related to lagged returns, suggesting that order flow has two opposing components: a transitory component that reverses in the short term and a permanent component that persists in the long term. Following Bianchi, Babiak and Dickerson (2022), we use lagged returns as a proxy for short-term liquidity effects, and assess the predictive relation of lagged order flow on future returns, while controlling for lagged returns.

The results reported in Table 3 indicate that, across all specifications, world order flow is

[6] This is in line with a long literature on market microstructure, which distinguishes between informed and uninformed traders and assesses the effect of information-based trading on asset prices. See, for example, Glosten and Milgrom (1985), Easley and O'Hara (1987) and Easley et al. (1996).

a highly significant predictor of both one-day ahead and one-week ahead cryptocurrency returns. The daily predictive regressions exhibit an interesting pattern. Lagged order flow on its own is not significant (t-stat=1.56). However, after controlling for lagged returns, it is highly significant (t-stat=5.54). In contrast, lagged returns have a negative and significant effect (t-stat=-10.63). These findings confirm that lagged order flow has two opposing components: the transitory component correlated with lagged returns that exhibits short-term reversal, and a permanent component, uncorrelated to lagged returns, that is positively related to future returns. By conditioning on both lagged order flow and lagged returns, we uncover the positive predictive ability of daily world order flow.

For the weekly predictive regressions, world order flow has a consistent positive and significant predictive power for future returns whether we condition on lagged returns or not. Weekly lagged returns still exhibit a significant negative coefficient, though of much lower magnitude and significance compared to daily regressions. This aligns with transitory effects exerting a stronger impact on prices at the daily horizon, while at the weekly horizon, permanent effects such as adverse selection become more dominant. Accordingly, the component of lagged weekly order flow that is uncorrelated with lagged weekly returns appears to be the dominant component of weekly order flow leading to consistently positive and significant predictions.

The strong positive predictive effect of world order flow for both daily and weekly returns is consistent with the permanent view. Our findings provide evidence against both the non-predictability and the transitory view. Specifically, in the full specification that includes the control variables, an increase in lagged world order flow by one standard deviation is associated with a 0.2% increase in daily returns and a 0.9% increase in weekly returns. The increasing strength of this predictive relation for longer horizons provides empirical support for the permanent view of the effect of order flow on cryptocurrency returns.

Relevant to our analysis, Makarov and Schoar (2020) study the relation between order flow and future returns specifically for Bitcoin. They find a significant negative relation between order flow and one-day-ahead returns, and a positive but statistically insignificant relationship with one-week-ahead returns. In this context, by using a broad cross-section of cryptocurrencies, our approach reveals the full predictive power of order flow.

In addition to order flow, most control variables are significant in the predictive regressions. The exceptions are as follows: the TED spread for daily returns and the term and default spreads for weekly returns. Finally, note that the daily predictive regression delivers an $\bar{R}^2 = 1.2\%$ and the weekly predictive regressions an $\bar{R}^2 = 3.1\%$.

3.2 Panel Regressions with All International Order Flows

In addition to panel regressions conditioning on just world order flow, we also estimate panel regressions that condition on all 11 international order flows. The panel regressions are set up in the same way as previously and condition on the same control variables. We estimate two types regressions: one with just the 11 international order flows and one with the 11 order flows plus the aggregate world order flow. We use daily and weekly data and repeat for both contemporaneous and predictive regressions.[7]

3.2.1 Contemporaneous Regressions

The results from contemporaneous regressions are shown in Table 4. Our main finding is that, for both daily and weekly returns, world order flow has the highest and most significant coefficient. Specifically, a one standard deviation increase in world order flow is associated with a 1.9% increase in returns at the daily horizon (t-stat=29.33) and a 2.7% increase at

[7] Note that regressions which condition on the 11 international order flows plus world order flow do not suffer from multicollinearity due to the standardization of all order flows. Specifically, if we regress the standardized world order flow on its 11 standardized components, the R^2 is equal to 0.50.

the weekly horizon (t-stat=10.91). Other than world order flow, US and Korean order flows remain significant in all specifications. The full set of order flows together with the control variables explain 12% of daily crypto returns and 21.2% of weekly returns.

These results justify the use of disaggregated international order flows (including US order flow) over and above world order flow. They also motivate the inclusion of South Korea to the sample since it is consistently highly significant. In short, the empirical evidence indicates that order flows on a global scale have explanatory power for contemporaneous cryptocurrency returns.[8]

3.2.2 Predictive Regressions

Turning to the predictive regressions, the results reported in Table 5 indicate that world order flow still has the highest and most significant coefficient. Now, a one standard deviation increase in world order flow is associated with a 0.2% increase in future daily returns (t-stat=5.09) and a 0.9% increase in future weekly returns (t-stat=4.35). For the predictive regressions, very few of the international order flows are significant: South Korea and Canada for daily returns and Sweden for weekly returns. Notably, US order flow is not significant in the presence of world order flow, which motivates the focus of our main analysis on world rather than US order flow. The $\bar{R}^2$ is equal to 1.3% for daily predictive regressions and 3.1% for weekly predictive regressions. This indicates that there is higher predictive power associated with weekly than daily returns. In conclusion, our main finding is that world order flow is a more powerful and significant predictor than any of the individual order flows in the context of in-sample panel regressions.[9]

[8] In the foreign exchange literature, several studies use disaggregated order flows by counterparty type (e.g., customer types). See, for example, Evans and Lyons (2005) and Menkhoff et al. (2016).

[9] For brevity, we do not report the coefficient estimates for the control variables in the panel regressions with international order flows. These coefficients are largely the same as with just world order flow.

4 Out-of-Sample Analysis

We perform an out-of-sample analysis of the predictive ability of order flow for cryptocurrency returns using popular machine learning (ML) methods (see, e.g., Gu, Kelly and Xiu, 2020). Out-of-sample forecasting of asset returns is notoriously difficult (see, e.g., Welch and Goyal, 2008), which is likely exacerbated in the case of cryptos due to their high volatility. Traditional prediction methods based on ordinary least squares (OLS) regressions are likely to underperform when conditioning on a high number of predictors such as the 11 international order flows and the additional control variables that we use to forecast one-day ahead cryptocurrency returns. ML methods are more suitable for out-of-sample forecasting because they emphasize techniques for variable selection and dimension reduction, which can accommodate a large set of predictors as well as a richer specification of functional forms.

4.1 ML Models

As our benchmark model, we use linear regression estimated with pooled OLS. The benchmark model is motivated by the predictive panel regression approach for in-sample analysis. In addition to the OLS benchmark, we consider a variety of linear and non-linear models from the rapidly expanding ML literature on forecasting asset returns (Gu, Kelly and Xiu, 2020). For penalized linear models, we include Ridge regression (RR), Lasso (LAS) and the Elastic Net (EN). We also employ principal component regression (PCR) as a linear dimension reduction technique. To incorporate non-linearities and predictor interactions, we consider tree-based models, which include the random forest (RF) and stochastic gradient boosted regression trees (SGB) as well as feed-forward neural networks with 1 to 4 hidden layers (NN1-NN4). A comprehensive description of these models can be found in Section A1 of the Online Appendix.

As previously discussed, cryptocurrency returns exhibit extremely heavy tails, characterized by high positive skewness and very high kurtosis. To mitigate the impact of heavy-tailed data, we adopt the robust Huber objective function, which combines the ℓ_2 loss for small errors and ℓ_1 loss for large errors. We estimate all models, except for RF, using the Huber objective function. For RF only, we follow Gu, Kelly and Xiu (2020) in using the default ℓ_2 loss function due to RF's inherent robustness to outliers. In unreported results, we find that models estimated using the Huber objective function generally outperform those estimated with the mean squared error (MSE) objective function for cryptocurrency data. See Section A1.7 of the Online Appendix for a detailed description of the Huber objective function.

We also form forecast combinations, which are designed to combine the forecasts of a set of models (see, e.g., Timmermann, 2006). Following Rapach, Strauss and Zhou (2010), we compute the equally-weighted average of all forecasts across a set of models at each point in time, referred to as the "mean" combination. We compute combined forecasts for two cases: linear models (L-Mean) and non-linear models (NL-Mean). Comparing the non-linear forecast combination to the linear combination allows us to assess the improved predictive ability gained from incorporating non-linearities and predictor interactions. For a detailed description of forecast combinations, see Section A1.6 of the Online Appendix.

4.2 Estimation Procedure and Performance Evaluation

We recursively generate out-of-sample daily forecasts for each model by dividing our sample into three disjoint sets: the training, validation and test sets. First, we estimate the model parameters on a training sample $\mathcal{T}_1$ spanning one year (February 14, 2018 to February 14, 2019). Second, we conduct an extensive hyperparameter optimization on a one-year validation sample $\mathcal{T}_2$ (February 15, 2019 to February 14, 2020). Third, we evaluate the out-of-sample performance of each model using an initial one-month test sample (February 18, 2020 to March 13, 2020). We keep the model parameters fixed for one month and repeat this process by rolling forward the validation and test sets by one month, while expanding the training sample by one month in each iteration. We continue this process to the end of the sample so that the total test sample ($\mathcal{T}_3$) ranges from February 18, 2020 to June 30, 2022. Section A2 in the Online Appendix provides detailed information on the hyperparameter optimization setup for each model.[10]

Our primary performance metric is the modified Welch and Goyal (2008) out-of-sample R^2 (R^2_{oos}) statistic used by Gu, Kelly and Xiu (2020):

$$R^2_{oos} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1})^2}, \tag{4}$$

where $r_{i,t+1}$ and $\hat{r}_{i,t+1}$ denote the realized $t+1$ return and one-day ahead $t+1$ forecast for coin i, respectively, and $\mathcal{T}_3$ indicates that the metric is only assessed on the test sample. The R^2_{oos} statistic pools forecast errors across coins and over time into a comprehensive panel-level assessment of each model's predictive ability. It measures the reduction in MSE achieved by each model relative to a naive forecast, which predicts zero returns for all coins (i.e., a random walk for prices with no drift). In unreported results, we explore several alternative benchmarks, including the use of historical mean returns typically employed in equity market forecasts. However, we find that the naive forecast of zero returns consistently produces the lowest R^2_{oos}, making it the most conservative benchmark.[11]

10 Note that our out-of-sample analysis focuses on daily forecasts since the short sample period prohibits performing weekly ML forecasting.

11 For the R^2_{oos} statistic, there is a standard way of assessing statistical significance based on the Clark and West (2006, 2007) testing procedure. However, we find that the R^2_{oos} values are significant at the 1% confidence level for nearly every model and specification. Hence, we omit these results for presentation purposes, but they are available upon request.

4.3 Out-of-Sample Performance

In Table 6, we report results on the out-of-sample statistical performance of the models. Our analysis compares models conditioning on three distinct information sets: (1) ML models that condition on just order flow (OF); (2) ML models that condition on just economic fundamentals (EF) (i.e., models that condition on just the control variables, not order flow); and (3) ML models that condition on both OF and EF. This grouping of models allows us to assess whether order flow is essential in generating reliable out-of-sample forecasts for cryptocurrency returns over and above economic fundamentals. All models condition on lagged returns to capture short-term reversals. Additionally, for the training, validation and test sets, we standardize each variable using its mean and variance from the training set as is standard practice in the ML literature. Our main findings can be summarized as follows.[12]

The first column of Table 6 shows that OF models generate higher R^2_{oos} values than both EF and OF+EF models, presented in the second and third columns, across 13 of the 14 ML specifications. The better performance of OF over EF indicates that international order flows convey more predictive information than economic fundamentals. Similarly, the better performance of OF over OF+EF suggests that economic fundamentals contribute limited additional predictive value, as their information is largely spanned by order flow. Hence, adding economic fundamentals to the order flow variables primarily increases variance without enhancing signal, causing most models to overfit and yield negative R^2_{oos} values in the OF+EF specification. Overall, this comparison across columns provides compelling evidence that order flow is the best information set in our analysis for predicting future crypto returns.[13]

[12] For standard exchange rates, machine learning techniques have been used by Gradojevic and Yang (2006) in assessing the predictive ability of order flow and Li, Tsiakas and Wang (2015) in assessing the predictive ability of economic fundamentals.

[13] The R^2_{oos} statistic is based on the MSE, which places high emphasis on large errors thus enhancing the effect of outliers. To address this issue, we also compute the R^1_{oos} statistic, which is based on the mean absolute error (MAE) and is, therefore, less sensitive to outliers. In unreported results, we find that the

Next, comparing across linear OF models, the first column of Table 6 indicates that all linear models yield negative R^2_{oos} values, suggesting that a linear specification lacks the complexity needed to outperform the naive benchmark. OLS has the lowest R^2_{oos} value of any individual model (-1.24%), which implies that OLS forecasts are outperformed by a naive forecast of zero.[14] This result is unsurprising, as OLS is prone to overfitting due to its lack of regularization. Adding shrinkage with a ridge penalty slightly improves the R^2_{oos} value to -1.18% for RR. Restricting OLS to a sparse parameterization with the lasso or elastic net penalty improves the R^2_{oos} value to -0.38% for both LAS and EN, highlighting the importance of sparsity in predicting crypto returns. In contrast, regularization via dimension reduction performs poorly; PCR achieves an R^2_{oos} value of only -1.01%, suggesting that the PCA-derived latent factors are not highly predictive of crypto returns. L-Mean has the lowest R^2_{oos} value at -1.33%, indicating that even combining forecasts cannot enhance the performance of linear models.

Notably, all non-linear OF models generate positive R^2_{oos} values, which demonstrates the importance of non-linearities and variable interactions in the predictive relation between order flow and future crypto returns. RF raises the R^2_{oos} to 0.36% by capturing non-linearities and interactions through decision trees. SGB is the best performer, further increasing the R^2_{oos} to 0.66%, reinforcing the role of sparsity since SGB is often considered a sparse model. Among neural networks, NN3 achieves the highest R^2_{oos} at 0.39%, which aligns with the findings of Gu, Kelly and Xiu (2020). NL-Mean achieves an R^2_{oos} of 0.52%, suggesting that combining forecasts generally enhances the performance of non-linear models.

Finally, we compare our ML OF models to other studies on crypto return prediction using ML techniques, which do not condition on order flow. For example, Cakici et al. (2024)

R^1_{oos} produces the same ranking across models as the R^2_{oos}.

[14] These results do not contradict the strong in-sample predictive ability of order flow shown in Table 3 since out-of-sample predictability is inherently more challenging than in-sample performance (Welch and Goyal, 2008).

predict crypto returns using ML models conditioning on 40 characteristics, not including order flow. Their best-performing model (OLS) achieves an R^2_{oos} value of -0.21%, which outperforms our OLS model's R^2_{oos} value of -1.24% but substantially underperforms all our non-linear models conditioning on order flow.[15]

In light of these results, we conclude the following: (1) conditioning on order flow provides superior forecasts relative to conditioning on economic fundamentals, and (2) non-linear ML models for cryptocurrency returns consistently outperform both linear models and the naive benchmark. In summary, order flow offers significant out-of-sample predictive power for cryptocurrency returns beyond that of economic fundamentals.

5 The Economic Value of Order Flow

5.1 Portfolio Sorts

In this section, we assess the out-of-sample economic value of conditioning on order flow. Our approach is based on generating portfolios by sorting the cross-section of cryptocurrencies on one of two criteria: (1) the world order flow of each coin; (2) the ML return forecast for each coin generated by models conditioning on one of the three information sets discussed in Section 4.3 (OF, EF, OF+EF). This approach provides a comprehensive assessment of the economic gains associated with the out-of-sample predictive ability of order flow, and its relative economic gain over other information sets.

The portfolio sorting procedure is implemented as follows. At each time t, we rank all coins using one of the two criteria listed above: world order flow or the ML forecasts. Based on this

15 Filippou, Rapach and Thimsen (2024) report R^2_{oos} values up to 2.08% for SGB, but their R^2_{oos} benchmark is the historical mean rather than zero, making direct comparison with our results challenging.

ranking, we allocate the 82 coins into five quintile portfolios. Allocating assets to quintile portfolios is a standard approach in the asset pricing literature and ensures that there is a sufficient number of cryptos in each portfolio. Then, we compute the $t + 1$ equally-weighted portfolio returns and the Newey and West (1987) t-statistic.[16] When sorting on world order flow, we report results for both daily and weekly rebalancing. When sorting based on ML forecasts, we only report results for daily rebalancing since the short sample period is not well-suited for weekly ML forecasts. All portfolio sorts are performed purely out of sample with no look-ahead bias. In short, the portfolio sorts allow us to assess whether order flow has predictive information for portfolio returns over the next period.[17]

We report results on the mean daily (or weekly) percent return for the five quintile portfolios (P_1 to P_5). We primarily focus on the zero-cost investment portfolio that goes long on the top portfolio and short on the bottom portfolio ($P_5 - P_1$). For this portfolio, we report the mean return, alpha and Sharpe ratio. Following Liu, Tsyvinski and Wu (2022), the alpha is estimated by regressing the portfolio return on three factors: (1) the market return defined as the value-weighted return of all 82 coins; (2) the size factor defined as the return difference between the small size portfolio (bottom 30% in market cap) and the big size portfolio (top 30% in market cap), and (3) the momentum factor based on 2×3 sorts on size and three-week returns. To construct these common factors at the daily (weekly) interval, we replicate their methodology using our sample of 82 coins and reconstitute the portfolios daily (weekly).

[16] In computing the Newey and West (1987) t-statistic, we use the Bartlett kernel with the data-driven bandwidth parameter selected by the AR(1) model (see, e.g., Andrews, 1991).

[17] For cryptocurrencies, it is sensible to use equal weights as opposed to value weights because value weights would be dominated by the top 5 cryptos, which on a given day comprise approximately 90% of the total market cap. Consequently, using equal weights allows us to take full advantage of the breadth of information available in our data set.

5.2 Portfolio Performance

5.2.1 Sorting on Order Flow

We begin our discussion of portfolio performance by reporting results in Table 7 for daily and weekly portfolio sorts on world order flow (OF^W). In addition to sorting on OF^W, we also sort on an orthogonalized OF^W variable (ortho-OF^W), which is orthogonalized relative to same-period returns.[18] As shown in Table 3, lagged order flow is comprised of one component that is correlated with lagged returns, reflecting transitory effects, and another component that is uncorrelated with returns, reflecting permanent effects. Using an orthogonalized OF^W ensures that portfolio sorts rely on lagged order flow information exclusive of any information in lagged returns. The sample period spans from February 18, 2020, to June 30, 2022.[19]

Panel A in Table 7 indicates that the performance of the daily long-short portfolio ($P_5 - P_1$) tends to be strong for the orthogonalized OF^W but not for the original non-orthogonalized OF^W. For example, the daily mean return of $P_5 - P_1$ when sorting on OF^W is equal to -0.03% with a t-stat=-0.19, which however jumps to 0.29% with a t-stat=2.11 for sorting on ortho-OF^W. Furthermore, ortho-OF^W exhibits an annualized Sharpe ratio of 1.34.[20]

These findings are not surprising. As we have seen in the daily predictive panel regressions, order flow alone is insignificant but when controlling for lagged returns it is highly significant. This indicates that the part of lagged order flow that is uncorrelated with lagged returns has strong positive predictive power. In portfolio sorts, we capture the same effect by

18 Daily (weekly) ortho-OF^W is defined as the last residual from a recursive regression of lagged order flow on lagged returns, which is estimated using an expanding window that is updated daily (weekly). Notably, this approach avoids any forward-looking bias as only information available up to time t is used to construct ortho-OF^W.

19 The portfolio sorts on world order flow use the same sample period as the sorts on ML forecasts so that all portfolio results are comparable.

20 For the results on portfolio sorts conditioning on just US order flow, please see Table A6.

orthogonalizing OF^W, which is essential in generating strong positive portfolio performance. Importantly, the positive relation between ortho-OF^W and future coin returns suggests a permanent effect of order flow on crypto prices.

Panel B in Table 7 shows that the weekly results are strong regardless of orthogonalization. For portfolio sorts on the weekly OF^W, the mean of $P_5 - P_1$ is equal to 1.61% per week and significant (t-stat=2.13). The alpha is similar (1.44% per week) and also significant (t-stat=2.02), demonstrating that the economic value cannot be explained by the crypto market, size, or momentum factors. The annualized Sharpe ratio is equal to 1.68. The results for ortho-OF^W are only slightly stronger: the long-short mean is equal to 1.74% (t-stat=2.28), the alpha is 1.52% (t-stat=2.11) and a Sharpe ratio of 1.79. Therefore, the orthogonalization of world order flow leads to a large improvement for daily returns but a small improvement for weekly returns, which is consistent with lagged returns having weaker predictive power for future weekly returns in the in-sample panel regressions. Taken together, these results indicate that there is strong out-of-sample economic value for portfolios sorted on world order flow.

5.2.2 Sorting on ML Forecasts

In Table 8, we report results on portfolio sorts based on the daily forecasts generated by ML models, which condition on the 11 international order flows. This approach enables an assessment of whether there are economic gains with conditioning on all order flows as opposed to just world order flow as we did previously. Given that the training and validation sets require two years of data, the test period begins on February 18, 2020, which is the same period as the sorts on world order flow.

We emphasize three key findings regarding ML model performance. First, non-linear models conditioning on OF achieve higher Sharpe ratios than those conditioning on EF or OF+EF across all non-linear models, except NN1. For example, the NL-Mean model generates an annualized Sharpe ratio of 3.45 for OF, compared to 1.88 for EF and 2.04 for OF+EF. This means the Sharpe ratio for NL-Mean with OF is nearly double that with EF. The top-performing model overall is SGB conditioning on OF, with a Sharpe ratio of 3.63, aligning with the highest R^2_{oos} value delivered by the same model shown in Table 6. RF and NN3 also deliver high Sharpe ratios of 3.19 and 3.04, respectively, aligning with their relatively high R^2_{oos} values.

Second, while linear models perform similarly across the three information sets, they are outperformed by non-linear models conditioning on OF. For example, the best linear model, OLS conditioning on EF, delivers a Sharpe ratio of 2.54, which is lower than that of RF, SGB, NN2, NN3, and NL-Mean conditioning on OF. These results underscore the importance of model complexity in generating economic value with ML OF models, aligning with the "virtue of complexity" in return prediction (Kelly, Malamud and Zhou, 2024).[21]

Third, our models conditioning on OF outperform leading benchmarks for crypto return prediction using ML techniques that do not condition on OF. For example, Filippou, Rapach and Thimsen (2024) construct daily equally-weighted long-short portfolios using various ML models conditioning on 54 characteristics representing network value, activity, momentum, technical signals and online activity, but excluding order flow. Their models yield annualized Sharpe ratios of 1.44, 2.68, 1.14, and 2.57 for RF, SGB, NN3, and NL-Mean, respectively. In comparison, Table 8 shows that the same models conditioning on OF achieve Sharpe ratios of 3.19, 3.63, 3.04, and 3.45, respectively, demonstrating the additional economic value of order flow. Additionally, the best-performing model (OLS) in Cakici et al. (2024) achieves a weekly three-factor alpha of 1.98%, which is considerably lower than SGB conditioning on

[21] OLS is considered to have higher complexity, i.e., more parameterization, than regularized linear models such as RR, LAS and EN. Therefore, OLS outperforming these regularized linear models aligns with the "virtue of complexity" (Kelly, Malamud and Zhou, 2024).

OF, yielding a weekly alpha of $5 \times 0.79 = 3.95\%$.

To illustrate these results, in Figure 1 we report the cumulative return of the long-short portfolio of select ML OF models against the return of the market portfolio, which is the value-weighted return of all coins in the sample. The figure confirms that the non-linear models consistently outperform the linear models throughout the sample period. It is also interesting to note that the non-linear models continue to perform well even when the market portfolio exhibits a negative return in the last year of the sample.

Overall, we can summarize the main results as follows: (1) the non-linear ML OF models perform the best and their performance is excellent by any standard. A daily long-short mean return of 0.78% for the best model (SGB) corresponds to an annualized return of 196.56% and an annualized $SR = 3.63$; (2) the linear ML models consistently underperform the non-linear models, demonstrating that model complexity is important for generating economic value; (3) the OLS benchmark performs well relative to regularized linear models; (4) the combined forecasts are among the best performers (5) the portfolios sorted on the daily ML models perform substantially better than portfolios sorted directly on world order flow; and, finally, (6) ML models conditioning on OF outperform leading benchmarks that use ML models without conditioning on OF. In short, there is high economic value in conditioning on all 11 international order flows in the context of non-linear ML models.

6 Economic Restrictions

6.1 Short-sale Constraints, Drawdowns, and Transaction Costs

Our portfolio analysis has so far focused on the long-short portfolio because it is well suited for illustrating the economic value of conditioning on order flow. In other words, if order flow is a good predictor, it should be able to identify which portfolio performs well and which one does not. Then, investors would benefit from a zero-cost trading strategy that goes long on the top portfolio and short on the bottom portfolio. However, in the cryptocurrency market, it may be difficult, even impossible, to short some cryptocurrencies. When possible, additional shorting fees will apply (see, e.g., Liu, Tsyvinski and Wu, 2022). Since the objective of a typical investor is to generate a realistic trading strategy that avoids shorting and is subject to lower transaction costs, we focus on just the long portfolio. For this reason, in Table 9, we report detailed results on the performance of the daily long OF portfolio (P_5) and compare them to the corresponding daily long-short OF portfolio ($P_5 - P_1$).

For a comprehensive assessment of performance of the most realistic trading strategies, we report the mean return, alpha, and Sharpe ratio, Additionally, in assessing economic restrictions, we report the maximum drawdown (MDD), turnover (TO) and break-even transaction cost (BE-TC) as advocated by Avramov, Cheng and Metzker (2023). The maximum drawdown is defined as the maximum cumulative loss from the portfolio's price peak to the following trough. A reasonably low MDD is indicative of the success of a trading strategy because large drawdowns often lead to fund redemptions.

Following Gu, Kelly and Xiu (2020), we define the turnover of the long-short portfolio at time t as follows:

$$TO_t = \frac{1}{2}\sum_{i\in L}\left| w_{i,t} - \frac{w_{i,t-1}(1+r_{i,t})}{\sum_{k\in L} w_{k,t-1}(1+r_{k,t})}\right| + \frac{1}{2}\sum_{j\in S}\left| w_{j,t} - \frac{w_{j,t-1}(1+r_{j,t})}{\sum_{n\in S} w_{n,t-1}(1+r_{n,t})}\right|, \qquad (5)$$

where $i \in L$ ($j \in S$) indicates that coin i (j) belongs to the long (short) portfolio, $w_{i,t}$ ($w_{j,t}$) refers to the weight of coin i (j) at time t, and $r_{i,t}$ ($r_{j,t}$) refers to the return of coin i (j) at time t. By design, the turnover of long and short positions ranges between 0 and 1. Accordingly, the turnover of the long-short portfolio defined above ranges between 0 and 2.

Finally, we compute the break-even transaction cost (BE-TC) as the portfolio return divided by turnover (see, e.g., Avramov, Cheng and Metzker, 2023). The BE-TC is equal to the daily/weekly proportional transaction cost required to eliminate the gains from the trading strategy. For example, if the portfolio return is equal to 1% per day and the turnover is 0.8, then the BE-TC is equal to $1\% \div 0.8 = 1.25\%$ per day, and the proportional transaction cost would have to be 1.25% per day to eliminate all the gains of the portfolio.

Overall, we find that the performance of the long portfolio is strong. Compared to the long-short portfolio, the long portfolio tends to have a higher mean return and alpha, lower Sharpe ratio in addition to lower turnover and higher break-even transaction cost. Consider, for example, the NL-Mean OF model. The long portfolio delivers a daily mean return of 0.79% (t-stat=2.96), an alpha of 0.81% (t-stat=2.99) and $SR = 1.88$. The maximum drawdown is equal to 0.66, the turnover is equal to 0.77 and the break-even transaction cost is equal to 1.02. For reference, the long-short portfolio for the same model delivers a similar mean return (0.75%) and alpha (0.76%) but a higher Sharpe ratio (3.52) at the expense of much higher turnover (1.55) and much lower break-even transaction cost (0.48%).

In terms of evaluating the level of the break-even transaction cost for the long portfolio, it is helpful to note that realistic transaction costs for the cryptocurrency market are estimated to be in the range of 0.3%-0.5%. For example, Bianchi, Babiak and Dickerson (2022) apply a fixed transaction cost of 0.3% for their long strategy, whereas Liu, Tsyvinski and Wu (2022) estimate the effective bid-ask spread to be around 0.5% since 2018. In this context, a break-even transaction cost of about 1% for the NL-Mean (or other non-linear models) comfortably clears the bar, thus leading to a profitable strategy net of transaction costs. Based on these results, we conclude that conditioning on order flow, especially in the context of non-linear machine learning forecasts, generates high economic value for long-only investors who pursue a realistic trading strategy investing in cryptocurrencies.

6.2 Limits to Arbitrage

Our main result is that order flow itself as well as ML models that condition on order flow positively predict future crypto returns. It might be expected that arbitragers would identify this opportunity and drive prices toward their fundamental values. However, Shleifer and Vishny (1997) and Pontiff (2006) argue that there are limits to arbitrage, which in our context could potentially explain the strong positive relation between order flow and future returns. A testable implication of this argument is that the relation between order flow and future returns should be more pronounced in coins that are more difficult to arbitrage.

Rather than relying on a single proxy for arbitrage costs, we follow Atilgan et al. (2020) and Liu, Tsyvinski and Wu (2022) in constructing an arbitrage index (AI) using a number of indicators known to capture important dimensions of limits to arbitrage. First, we build an index at the coin-level, for which we include size, idiosyncratic volatility (Ang et al., 2006), and illiquidity (Amihud, 2002).[22] To construct the AI, we sort coins in increasing order based on their idiosyncratic volatility and illiquidity, and decreasing order based on their size. Each coin is given the corresponding score of its quintile rank for each variable. Then, the arbitrage cost index is the sum of the three scores, which ranges from 3 to 15. A higher value implies greater limits to arbitrage.

Based on the AI of each coin at time t, we form tercile splits in order to compare the predictability of machine learning forecasts with different levels of the arbitrage index. In the last column of Table 10, we report the R^2_{oos} values for each AI tercile. The results indicate that the L-Mean model exhibits a negative R^2_{oos} for the low and medium AI tercile and a barely positive R^2_{oos} for the high AI tercile. In contrast, the NL-Mean model exhibits a positive R^2_{oos} for all three AI terciles with the highest R^2_{oos} being associated with the low

[22] Idiosyncratic volatility is defined as the volatility of residuals from the three factor model (Liu, Tsyvinski and Wu, 2022) over the previous 21 days. Illiqudity is defined as the ratio of the absolute value of returns over daily total volume, averaged over the past 21 days.

AI tercile. This is strong statistical evidence that the NL-Mean model performs best for the crypto portfolio that is the easiest to arbitrage.

We also conduct a bivariate sort analysis. First, at time t, we split the coins in our sample into terciles based on the AI. Second, within each tercile, we further sort coins into terciles by either the orthogonalized order flow or the ML forecast conditioning on order flows. Table 10 reports the double sort results for ortho-OF^W, L-Mean, and NL-Mean. The first panel shows that for the low AI tercile, the long-short portfolio based on ortho-OF^W achieves an alpha of 0.33% (t-stat=2.86) and a Sharpe ratio of 1.96. In contrast, for the high AI tercile, the long-short portfolio based on ortho-OF^W achieves an insignificant alpha of 0.13% (t-stat=0.66) and a Sharpe ratio of only 0.36. These results demonstrate that the economic value of order flow cannot be explained by limits to arbitrage, as the long-short performance is strongest in coins that are the easiest to arbitrage, i.e., the largest, most liquid, and lowest volatility coins.

The second panel shows that for the low and medium AI terciles, the long-short portfolio based on L-Mean has insignificant alphas. In contrast, for the high AI tercile, the portfolio achieves an alpha of 0.71 (t-stat=3.46) and a Sharpe ratio of 2.19. This indicates that the economic value of linear ML models is concentrated in coins with the highest arbitrage costs.

Importantly, the third panel reveals that for NL-Mean, all AI terciles display significant alphas. Additionally, the Sharpe ratios for the low, medium, and high AI terciles are 2.15, 1.25 and 2.65, respectively, suggesting that the performance of non-linear ML models is robust across coins with varying arbitrage costs. These findings are remarkable considering that the economic value of ML models not conditioning on order flow can mostly be explained by limits to arbitrage in stocks (Avramov, Cheng and Metzker, 2023), options (Bali et al., 2023), and cryptos (Cakici et al., 2024, and Filippou, Rapach and Thimsen, 2024).

Table A7 presents a bivariate sort analysis for L-Mean and NL-Mean conditioning purely on economic fundamentals (EF). For both models, the Sharpe ratios are the lowest in the bottom AI tercile and increase monotonically as we move to the top AI tercile. This clearly indicates that the long-short portfolio returns for models conditioning on economic fundamentals have economic value only for small, illiquid, and volatile coins. This finding further illustrates the economic value of ML predictions conditioning on OF compared to EF.

Overall, these results show that the economic value of orthogonalized order flow and non-linear ML models cannot be explained by limits to arbitrage. In contrast, arbitrage costs can explain the economic value of linear ML models conditioning on order flow as well as (linear and non-linear) ML models conditioning on economic fundamentals. In light of these results, we conclude that the permanent effect of order flow on crypto returns is not limited to small, illiquid and volatile coins but rather extends to coins which are large, liquid and less volatile.

7 Conclusion

The meteoric rise of cryptocurrencies as a new investment asset class has triggered an emerging literature in financial economics. A central theme of this literature is the out-of-sample prediction of cryptocurrency returns and the design of trading strategies, which are consistently profitable. The foreign exchange literature has established order flow as a prominent predictor of fiat currency returns representing the microstructure approach to exchange rate prediction as an alternative to models conditioning on standard economic fundamentals. For cryptocurrencies, however, there is a gap in the literature as there is little work on the ability of order flow to explain and predict cryptocurrency returns. Our analysis contributes to this literature by highlighting the role of order flow for out-of-sample return prediction over and above standard economic fundamentals.

We find that order flow matters both for explaining and for predicting the cross-section of cryptocurrency returns. World order flow has a strong contemporaneous and predictive relation to cryptocurrency returns in the context of in-sample panel regressions. The combination of non-linear machine learning techniques with international order flow information provides an especially powerful predictive toolkit for this market and is associated with high economic value. For example, the forecast combination of non-linear machine learning models conditioning on daily order flow delivers an annualized Sharpe ratio of 1.88 for the long portfolio and 3.52 for the long-short portfolio. Overall, our findings are robust to economic restrictions, including short-selling constraints, transaction costs, and limits to arbitrage, providing compelling empirical support for the permanent view of the effect of order flow on returns. In conclusion, our empirical evidence strongly indicates that information conveyed by order flow matters for cryptocurrency returns.

References

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

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-858.

Ang, A., R.J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance* 61, 259-299.

Atilgan, Y., T.G. Bali, K.O. Demirtas, and A.D. Gunaydin (2020). Left-tail momentum: Underreaction to bad news, costly arbitrage and equity returns. *Journal of Financial Economics* 135, 725-753.

Avramov, D., S. Cheng, and L. Metzker (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science* 69, 2587–2619.

Bali, T.G., H. Beckmeyer, M. Moerke, and F. Weigert (2023). Option return predictability with machine learning and big data. *Review of Financial Studies* 36, 3548-3602.

Berger, D., A. Chaboud, S. Chernenko, E. Howorka, and J. Wright (2008). Order flow and exchange rate dynamics in electronic brokerage system data. *Journal of International Economics* 75, 93–109.

Bianchi, D., M. Babiak, and A. Dickerson (2022). Trading volume and liquidity provision in cryptocurrency markets. *Journal of Banking and Finance* 142, 106547.

Brandt, M.W. and K.A. Kavajecz (2004). Price discovery in the US treasury market: The impact of order flow and liquidity on the yield curve. *Journal of Finance* 59, 2623-2654.

Cakici, N., S.J.H. Shahzad, B. Bedowska-Sojka, and A. Zaremba (2024). Machine learning and the cross-section of cryptocurrency returns. *International Review of Financial Analysis*, 103244.

Chordia, T., R. Roll, and A. Subrahmanyam (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics* 65, 111–130.

Clark, T.E. and K.D. West (2006). Using out-of-sample mean squared prediction errors to test the Martingale difference hypothesis. *Journal of Econometrics* 135, 155-186.

Clark, T.E. and K.D. West (2007). Approximately normal tests for equal predictive accuracy. *Journal of Econometrics* 138, 291-311.

Deuskar, P. and T.C. Johnson (2011). Market liquidity and flow-driven risk. *Review of Financial Studies* 24, 721–753.

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

Easley, D., N.M. Kiefer, M. O'Hara, and J.B. Paperman (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance* 51, 1405-1436.

Evans, M.D.D. (2010). Order flows and the exchange rate disconnect puzzle. *Journal of International Economics* 80, 58-71.

Evans, M.D.D. and R. K. Lyons (2002). Order flow and exchange rate dynamics. *Journal of Political Economy* 110, 170-180.

Evans, M.D.D. and R.K. Lyons (2005). Meese-Rogoff redux: Micro-based exchange-rate forecasting. *American Economic Review Papers and Proceedings* 95, 405-414.

Evans, M.D.D. and D. Rime (2012). Micro approaches to foreign exchange determination, in James, J., L. Sarno and I. W. Marsh (eds.) *Handbook of Exchange Rates*. Hoboken, NJ: Wiley.

Fieberg, C., G. Liedtke, T. Poddig, T. Walker, and A. Zaremba (2023). A trend factor for the cross-section of cryptocurrency returns. *Journal of Financial and Quantitative Analysis* (forthcoming).

Filippou, I., D. Rapach, and C. Thimsen (2024). Cryptocurrency return predictability: A machine-learning analysis. Available at SSRN: https://ssrn.com/abstract=3914414.

Froot, K.A. and T. Ramadorai (2005). Currency returns, intrinsic value, and institutional-investor flows. *Journal of Finance* 60, 1535–1566.

Glosten, L.R. and P.R. Milgrom (1985). Bid, ask and transaction prices in a market-maker market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71-100.

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

Gradojevic N. and J. Yang (2006). Non-linear, non-parametric, non-fundamental exchange rate forecasting. *Journal of Forecasting* 25, 227-245.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies* 33, 2223–2273.

Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of Financial Economics* 22, 229–252.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance* 46, 179–207.

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

Kyle, A.S. (1985). Continuous auctions and insider trading. *Econometrica* 53, 1335–1355.

Kelly, B., S. Malamud, and K. Zhou (2024). The virtue of complexity in return prediction. *Journal of Finance* 79, 459-503.

Li, J., I. Tsiakas, and W. Wang (2015). Predicting exchange rates out of sample: Can economic fundamentals beat the random walk?" *Journal of Financial Econometrics* 13, 293-341.

Liu, Y., and A. Tsyvinski (2021). Risks and returns of cryptocurrency. *Review of Financial Studies* 34, 2689–2727.

Liu, Y., A. Tsyvinski, and X. Wu (2022). Common risk factors in cryptocurrency. *Journal of Finance* 77, 1133-1177.

Makarov, I. and A. Schoar (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics* 135, 293–319.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2016). Information flow in foreign exchange markets: Dissecting customer currency trades. *Journal of Finance* 71, 601–633.

Newey, W.K. and K.D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics* 42, 35-52.

Rapach, D.E., J.K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies* 23, 821-862.

Rime, D., L. Sarno, and E. Sojli (2010). Exchange rate forecasting, order flow and macroeconomic information. *Journal of International Economics* 80, 72-88.

Shleifer, A., and R. W. Vishny (1997). The limits of arbitrage. *Journal of Finance* 52, 35-55.

Timmermann, A. (2006). Forecast combinations, in Elliott, G., C. W. J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 1, Amsterdam: North-Holland.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455-1508.

Table 1: Variance Decomposition of World Order Flow

This table reports the variance decomposition of world order flow into its 11 country components. The values represent the percent variance contribution of each country to world order flow at the daily and weekly frequency. The sample period ranges from February 14, 2018 to June 30, 2022.

	Daily %	Weekly %
OF^{USD}	39.11	45.56
OF^{KRW}	23.70	26.84
OF^{EUR}	10.35	8.47
OF^{GBP}	5.59	3.52
OF^{NZD}	3.63	3.60
OF^{CHF}	3.61	1.67
OF^{AUD}	3.37	2.22
OF^{CAD}	3.24	1.94
OF^{NOK}	3.04	2.39
OF^{SEK}	2.75	2.31
OF^{JPY}	1.62	1.49
OF^{W}	100	100

Table 2: Contemporaneous Panel Regressions with World Order Flow

This table displays panel regression results of cryptocurrency returns on contemporaneous world order flow and control variables. Order flow is in logs and standardized as described in the data section. Market cap and volume are in logs. Volatility is the log difference between a coin's high and low price. The VIX, S&P 500 and MSCI are log returns. The TED spread is the difference between the three-month Treasury bill and the three-month LIBOR. The short rate is the three-month T-bill rate. The term spread is the difference between the ten-year and three-month T-bill rates. The default spread is the difference between the AAA and BAA corporate bonds. All panel regressions use balanced panels with fixed coin effects. Columns (1)-(2) are for daily returns. Columns (3)-(4) are for weekly returns. The t-statistics shown in parentheses are computed using Newey and West (1987) standard errors. The sample period ranges from February 14, 2018 to June 30, 2022.

	Daily		Weekly	
	(1)	(2)	(3)	(4)
constant	0.002***	0.001	0.011***	0.058***
	(6.32)	(0.55)	(7.38)	(6.08)
$OF_{i,t}^{W}$	0.022***	0.022***	0.045***	0.040***
	(36.83)	(37.62)	(16.43)	(16.97)
$MCAP_{i,t}$		-0.001		0.019***
		(-1.35)		(6.16)
$Volume_{i,t}$		0.007***		0.010***
		(9.81)		(4.33)
$Volatility_{i,t}$		0.145***		0.246***
		(4.78)		(5.86)
VIX_t		-0.044***		-0.089***
		(-6.49)		(-6.44)
TED $Spread_t$		0.001***		0.001***
		(3.89)		(5.86)
S&P 500_t		-0.491***		-4.065***
		(-6.49)		(-13.12)
$MSCI_t^{Global}$		0.767***		5.675***
		(8.15)		(16.97)
Short $Rate_t$		-0.002***		-0.016***
		(-3.61)		(-7.03)
Term $Spread_t$		-0.009***		-0.056***
		(-7.12)		(-10.90)
Default $Spread_t$		0.010***		0.012*
		(6.34)		(1.69)
$\bar{R}^2$ (%)	7.1	11.0	6.5	19.9
# obs.	90364	90364	16728	16728

Table 3: Predictive Panel Regressions with World Order Flow

This table displays results of predictive panel regressions of cryptocurrency returns on lagged world order flow and lagged control variables. The table contains the same information as Table 2 but all variables are lagged and $r_{i,t-1}$ is the lagged return of each cryptocurrency. The t-statistics shown in brackets are computed using Newey and West (1987) standard errors. The sample period ranges from February 14, 2018 to June 30, 2022.

	Daily			Weekly		
	(1)	(2)	(3)	(4)	(5)	(6)
constant	0.002***	0.002***	0.014***	0.013***	0.013***	0.048***
	(5.87)	(5.96)	(5.99)	(7.85)	(7.96)	(4.36)
$\mathrm{OF}^{W}_{i,t-1}$	0.001	0.002***	0.002***	0.008***	0.009***	0.009***
	(1.56)	(5.54)	(6.43)	(5.03)	(5.39)	(5.57)
$\mathrm{r}_{i,t-1}$		-0.060***	-0.072***		-0.020**	-0.037***
		(-10.63)	(-13.43)		(-1.96)	(-3.74)
$\mathrm{MCAP}_{i,t-1}$			-0.005***			-0.026***
			(-9.44)			(-8.62)
$\mathrm{Volume}_{i,t-1}$			0.001**			0.005**
			(2.34)			(2.45)
$\mathrm{Volatility}_{i,t-1}$			0.048***			0.058***
			(7.59)			(3.33)
VIX_{t-1}			-0.012**			-0.109***
			(-2.12)			(-8.55)
TED Spread_{t-1}			0.000			-0.001*
			(0.34)			(-1.95)
S&P 500_{t-1}			-0.187**			1.564***
			(-2.03)			(6.57)
$\mathrm{MSCI}^{Global}_{t-1}$			0.546***			-1.976***
			(5.68)			(-6.51)
Short Rate_{t-1}			-0.006***			-0.024***
			(-9.86)			(-8.66)
Term Spread_{t-1}			-0.004***			-0.009
			(-2.87)			(-1.46)
Default Spread_{t-1}			-0.004**			-0.004
			(-2.41)			(-0.46)
$\bar{R}^2$ (%)	0.1	0.3	1.2	0.2	0.2	3.1
# obs.	90282	90282	90282	16646	16646	16646

Table 4: Contemporaneous Panel Regressions with All International Order Flows

This table displays panel regression results of contemporaneous cryptocurrency returns on world order flow, international order flows and control variables. The table contains the same information as Table 2 but for all international order flows plus the world order flow. The t-statistics shown in parentheses are computed using Newey and West (1987) standard errors. The sample period ranges from February 14, 2018 to June 30, 2022. The panel regressions also condition on the control variables reported in the previous tables but these results are not reported to save space.

	Daily		Weekly	
	(1)	(2)	(3)	(4)
constant	0.002	0.001	0.059***	0.056***
	(0.88)	(0.23)	(6.17)	(5.86)
$\text{OF}_{i,t}^{W}$		0.019***		0.027***
		(29.33)		(10.91)
$\text{OF}_{i,t}^{USD}$	0.012***	0.005***	0.025***	0.015***
	(24.93)	(10.68)	(12.48)	(7.28)
$\text{OF}_{i,t}^{KRW}$	0.012***	0.006***	0.021***	0.014***
	(23.53)	(13.77)	(10.88)	(7.24)
$\text{OF}_{i,t}^{EUR}$	0.004***	0.001	0.011***	0.007***
	(10.77)	(1.48)	(5.81)	(3.46)
$\text{OF}_{i,t}^{GBP}$	0.002***	-0.001	0.004**	0.001
	(5.23)	(-3.32)	(2.30)	(0.42)
$\text{OF}_{i,t}^{JPY}$	0.001***	-0.001	0.001	-0.001
	(3.89)	(-0.90)	(0.58)	(-0.92)
$\text{OF}_{i,t}^{CAD}$	0.001***	-0.001***	0.004***	0.002
	(3.40)	(-3.46)	(3.14)	(1.49)
$\text{OF}_{i,t}^{AUD}$	0.001**	-0.002***	0.003*	0.001
	(2.12)	(-4.72)	(1.74)	(0.07)
$\text{OF}_{i,t}^{CHF}$	0.000	-0.002***	-0.001	-0.003*
	(0.68)	(-6.90)	(-0.27)	(-1.82)
$\text{OF}_{i,t}^{NZD}$	0.001**	-0.002***	0.001	-0.003*
	(1.96)	(-5.45)	(0.10)	(-1.95)
$\text{OF}_{i,t}^{NOK}$	0.000	-0.002***	0.002	-0.001
	(0.88)	(-5.96)	(1.52)	(-0.23)
$\text{OF}_{i,t}^{SEK}$	0.000	-0.002***	0.001	-0.001
	(0.64)	(5.61)	(0.82)	(-0.81)
$\bar{R}^2$ (%)	9.4	12.0	19.8	21.1
# obs.	90364	90364	16728	16728

Table 5: Predictive Panel Regressions with All International Order Flows

This table displays results of predictive panel regressions of cryptocurrency returns on lagged world order flow, international order flows and control variables. The table contains the same information as Table 3 but for all international order flows plus the world order flow. The t-statistics shown in parentheses are computed using Newey and West (1987) standard errors. The sample period ranges from February 14, 2018 to June 30, 2022. The panel regressions also condition on the control variables reported in the previous tables but these results are not reported to save space.

	Daily		Weekly	
	(1)	(2)	(3)	(4)
constant	0.014***	0.014***	0.049***	0.048***
	(6.03)	(5.98)	(4.43)	(4.36)
$\text{OF}^{W}_{i,t-1}$		0.002***		0.009***
		(5.09)		(4.35)
$\text{OF}^{USD}_{i,t-1}$	-0.001*	-0.001	0.004***	0.001
	(1.70)	(-0.77)	(2.94)	(0.55)
$\text{OF}^{KRW}_{i,t-1}$	0.002***	0.001***	0.001	-0.001
	(5.55)	(3.29)	(0.91)	(-0.86)
$\text{OF}^{EUR}_{i,t-1}$	0.001	-0.001	0.003*	0.001
	(1.02)	(-0.38)	(1.78)	(0.71)
$\text{OF}^{GBP}_{i,t-1}$	0.000	-0.001	-0.001	-0.001
	(0.12)	(-0.86)	(-0.27)	(-1.01)
$\text{OF}^{JPY}_{i,t-1}$	0.001*	0.001	-0.001	-0.001
	(1.69)	(1.09)	(-0.13)	(-0.66)
$\text{OF}^{CAD}_{i,t-1}$	-0.001	-0.001**	0.002	0.001
	(-1.57)	(-2.38)	(1.55)	(1.00)
$\text{OF}^{AUD}_{i,t-1}$	-0.001	-0.001*	-0.001	-0.002
	(-0.85)	(-1.68)	(-0.79)	(-1.35)
$\text{OF}^{CHF}_{i,t-1}$	0.001	-0.000	0.001	0.001
	(0.57)	(-0.30)	(1.03)	(0.47)
$\text{OF}^{NZD}_{i,t-1}$	0.000	-0.001	0.003**	0.002
	(0.11)	(-0.76)	(2.38)	(1.58)
$\text{OF}^{NOK}_{i,t-1}$	0.001	0.000	0.001	0.001
	(0.96)	(0.13)	(0.97)	(0.34)
$\text{OF}^{SEK}_{i,t-1}$	-0.001	-0.001	-0.002*	-0.003**
	(-0.74)	(-1.51)	(-1.78)	(-2.40)
$\text{r}_{i,t-1}$	-0.070***	-0.073***	-0.032***	-0.037***
	(-13.16)	(-13.51)	(-3.22)	(-3.74)
$\bar{R}^2$ (%)	1.2	1.3	3.0	3.1
# obs.	90282	90282	16646	16646

Table 6: Out-of-Sample Statistical Performance

This table reports the out-sample statistical performance of the models using the R^2_{oos} statistic in percent based on the Mean Squared Error (MSE) of the forecasts. The MSE is computed using an expanding window. Daily results are reported for three information sets: OF only conditions on order flows; EF only conditions on economic fundamentals; and OF+EF conditions on both. All models condition on lagged returns. Bold entries indicate the highest value in each row.

	Daily R^2_{oos} (%)		
	OF	EF	OF+EF
OLS	**-1.24**	-19.80	-20.11
RR	**-1.18**	-16.83	-17.23
LAS	**-0.38**	-0.55	-0.51
EN	-0.38	-0.30	**-0.30**
PCR	**-1.01**	-1.50	-1.72
RF	**0.36**	-0.15	-0.30
SGB	**0.66**	0.61	0.50
NN1	**0.20**	-0.12	-0.23
NN2	**0.28**	0.04	-0.01
NN3	**0.39**	-1.64	-1.14
NN4	**0.13**	-2.05	-0.02
L-Mean	**-1.33**	-2.88	-2.96
NL-Mean	**0.52**	-0.12	0.23

Table 7: Portfolios Sorted on World Order Flow

This table displays the performance of cryptocurrency portfolios sorted on world order flow (OF^W). ortho-OF^W is world order flow orthogonalized relative to same-period returns. P_1 represents the portfolio with the lowest lagged order flow and P_5 the portfolio with the highest lagged order flow. The portfolios are equally-weighted and rebalanced daily (Panel A) or weekly (Panel B). The alpha is based on a three-factor model, which includes the cryptocurrency market, size and momentum factors (Liu, Tsyvinski and Wu, 2022). The returns and alphas are in daily/weekly percent. SR is the annualized Sharpe ratio. The Newey and West (1987) t-statistics are shown in parenthesis. The full sample period ranges from February 18, 2020 to June 30, 2022.

Panel A: Daily Portfolios Sorted on World Order Flow

	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$		
	Mean return (% daily)					Mean	alpha	SR
$\mathbf{OF}^W$	0.46	0.30	0.38	0.50	0.44	-0.03	-0.06	-0.12
(t-stat)	(1.84)	(1.22)	(1.61)	(1.98)	(1.73)	(-0.19)	(-0.43)	
ortho-$\mathbf{OF}^W$	0.27	0.40	0.46	0.41	0.56	0.29	0.30	1.34
(t-stat)	(1.09)	(1.57)	(1.84)	(1.70)	(2.19)	(2.11)	(2.07)	

Panel B: Weekly Portfolios Sorted on World Order Flow

	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$		
	Mean return (% weekly)					Mean	alpha	SR
$\mathbf{OF}^W$	1.08	1.82	1.86	1.65	2.69	1.61	1.44	1.68
(t-stat)	(0.71)	(1.12)	(1.16)	(1.03)	(1.69)	(2.13)	(2.02)	
ortho-$\mathbf{OF}^W$	1.14	1.33	2.06	1.67	2.88	1.74	1.52	1.79
(t-stat)	(0.73)	(0.85)	(1.23)	(1.10)	(1.80)	(2.28)	(2.11)	

Table 8: Daily Portfolios Sorted on ML Forecasts

This table displays the performance of cryptocurrency portfolios that are sorted on ML forecasts. The portfolios are equally-weighted and rebalanced daily. Results are reported for three ML models: OF only conditions on order flows; EF only conditions on economic fundamentals; and OF+EF conditions on both. All models condition on lagged returns. The alpha is based on a three-factor model, which includes the cryptocurrency market, size and momentum factors (Liu, Tsyvinski and Wu, 2022). The returns and alphas are in daily percent. SR is the annualized Sharpe ratio. The Newey and West (1987) t-statistics are shown in parenthesis. Bold entries indicate the highest Sharpe ratio for each ML model across the three information sets (OF, EF, OF+EF). The sample period ranges from February 18, 2020 to June 30, 2022.

	OF Model $P_5 - P_1$			**EF Model** $P_5 - P_1$			**OF+EF Model** $P_5 - P_1$		
	Mean	alpha	SR	Mean	alpha	SR	Mean	alpha	SR
OLS	0.45	0.44	2.26	0.62	0.63	**2.54**	0.55	0.57	2.27
(t-stat)	(3.31)	(3.42)		(3.75)	(4.20)		(3.40)	(3.75)	
RR	0.44	0.42	2.22	0.61	0.63	**2.51**	0.49	0.50	2.09
(t-stat)	(3.24)	(3.33)		(3.71)	(4.16)		(3.13)	(3.42)	
LAS	0.32	0.33	1.57	0.31	0.32	1.60	0.32	0.33	**1.67**
(t-stat)	(2.30)	(2.48)		(2.53)	(2.55)		(2.61)	(2.66)	
EN	0.29	0.33	1.46	0.34	0.39	1.78	0.34	0.40	**1.81**
(t-stat)	(2.14)	(2.65)		(2.67)	(3.22)		(2.69)	(3.27)	
PCR	0.17	0.17	**0.84**	-0.05	-0.11	-0.24	0.10	0.05	0.46
(t-stat)	(1.24)	(1.26)		(-0.37)	(-0.76)		(0.74)	(0.36)	
RF	0.66	0.67	**3.19**	0.22	0.23	1.03	0.21	0.17	0.97
(t-stat)	(5.00)	(4.94)		(1.62)	(1.72)		(1.49)	(1.32)	
SGB	0.78	0.79	**3.63**	0.56	0.65	2.54	0.44	0.51	2.15
(t-stat)	(5.62)	(5.67)		(3.82)	(4.66)		(3.20)	(3.92)	
NN1	0.45	0.49	2.13	0.65	0.64	**2.70**	0.55	0.60	2.46
(t-stat)	(3.21)	(3.64)		(4.14)	(4.41)		(3.67)	(4.15)	
NN2	0.55	0.59	**2.77**	0.42	0.42	1.72	0.67	0.67	2.76
(t-stat)	(4.07)	(4.61)		(2.79)	(2.71)		(4.20)	(4.40)	
NN3	0.64	0.63	**3.04**	0.57	0.58	2.60	0.38	0.36	1.69
(t-stat)	(4.47)	(4.63)		(3.85)	(4.24)		(2.48)	(2.59)	
NN4	0.42	0.43	**2.16**	0.28	0.33	1.19	0.46	0.46	2.06
(t-stat)	(3.23)	(3.41)		(1.85)	(2.19)		(3.10)	(3.31)	
L-Mean	0.32	0.32	1.63	0.57	0.59	**2.35**	0.51	0.53	2.20
(t-stat)	(2.37)	(2.49)		(3.46)	(3.87)		(3.26)	(3.61)	
NL-Mean	0.74	0.75	**3.45**	0.47	0.52	1.88	0.48	0.51	2.04
(t-stat)	(5.21)	(5.44)		(2.82)	(3.41)		(2.99)	(3.47)	

Table 9: The Performance of Long vs Long-Short Portfolios

This table compares the performance of the long to the long-short portfolio for daily returns. The portfolios are sorted on the ML forecasts from the OF model that conditions on all lagged international order flows and lagged returns. The long portfolio is the top quintile portfolio (P_5) and the long-short portfolio is P_5-P_1. The portfolios are equally-weighted and rebalanced daily. The alpha is based on a three-factor model, which includes the cryptocurrency market, size and momentum factors (Liu, Tsyvinski and Wu, 2022). The returns and alphas are in daily percent. SR is the annualized Sharpe ratio. MDD is the maximum drawdown. TO is the daily turnover. BE-TC is the break-even transaction cost in daily percent. The Newey and West (1987) t-statistics are shown in parenthesis. The sample period ranges from February 18, 2020 to June 30, 2022.

	Long Portfolio (P_5) - OF Model						Long-Short Portfolio ($P_5 - P_1$) - OF Model					
	Mean	alpha	SR	MDD	TO	BE TC	Mean	alpha	SR	MDD	TO	BE TC
OLS	0.56	0.56	1.41	0.75	0.80	0.68	0.45	0.44	2.26	0.34	1.59	0.28
(t-stat)	(2.24)	(2.18)					(3.31)	(3.42)				
RR	0.55	0.55	1.41	0.77	0.81	0.67	0.44	0.42	2.22	0.35	1.59	0.27
(t-stat)	(2.22)	(2.16)					(3.24)	(3.33)				
LAS	0.54	0.54	1.32	0.75	0.53	1.04	0.32	0.33	1.57	0.40	1.03	0.31
(t-stat)	(2.11)	(2.02)					(2.30)	(2.48)				
EN	0.50	0.52	1.27	0.74	0.53	0.97	0.29	0.33	1.46	0.42	1.03	0.27
(t-stat)	(2.04)	(2.01)					(2.14)	(2.65)				
PCR	0.47	0.48	1.21	0.75	0.80	0.58	0.17	0.17	0.84	0.48	1.59	0.11
(t-stat)	(1.92)	(1.87)					(1.24)	(1.26)				
RF	0.71	0.73	1.71	0.68	0.77	0.92	0.66	0.67	3.19	0.23	1.55	0.43
(t-stat)	(2.72)	(2.70)					(5.00)	(4.94)				
SGB	0.77	0.80	1.85	0.68	0.76	1.01	0.78	0.79	3.63	0.29	1.54	0.51
(t-stat)	(2.94)	(2.92)					(5.62)	(5.67)				
NN1	0.61	0.64	1.54	0.64	0.81	0.76	0.50	0.53	2.34	0.35	1.59	0.32
(t-stat)	(2.48)	(2.46)					(3.56)	(3.82)				
NN2	0.72	0.72	1.73	0.65	0.80	0.89	0.61	0.63	2.63	0.32	1.59	0.38
(t-stat)	(2.74)	(2.67)					(3.89)	(4.23)				
NN3	0.68	0.69	1.69	0.63	0.80	0.84	0.70	0.71	3.37	0.28	1.59	0.44
(t-stat)	(2.69)	(2.64)					(4.93)	(5.31)				
NN4	0.65	0.64	1.56	0.66	0.71	0.91	0.56	0.55	2.49	0.43	1.42	0.39
(t-stat)	(2.48)	(2.39)					(3.77)	(3.80)				
L-Mean	0.48	0.49	1.21	0.78	0.80	0.58	0.32	0.32	1.63	0.36	1.59	0.20
(t-stat)	(1.93)	(1.89)					(2.37)	(2.49)				
NL-Mean	0.79	0.81	1.88	0.66	0.77	1.02	0.75	0.76	3.52	0.28	1.55	0.48
(t-stat)	(2.96)	(2.99)					(5.28)	(5.57)				

Table 10: Double-Sorted Portfolios

This table displays the performance of double-sorted cryptocurrency portfolios. We first sort on the arbitrage index, which is an index based on quintiles of size, idiosyncratic volatility and illiquidity. Then, we double sort on one of the following: orthogonalized world order flow (ortho-OF^W), the linear ML forecast combination (L-Mean) that conditions on all order flows (OF model), and the non-linear ML forecast combination (NL-Mean) for the OF model. The portfolios are equally-weighted and rebalanced daily. The alpha is based on a three-factor model, which includes the cryptocurrency market, size and momentum factors (Liu, Tsyvinski and Wu, 2022). The mean returns and alphas are in daily percent. SR is the annualized Sharpe ratio. The Newey and West (1987) t-statistics are shown in parenthesis. The table also reports the R^2_{oos} in percent for the two ML models for each level of the arbitrage index. The full sample period ranges from February 18, 2020 to June 30, 2022.

ortho-OF^W

		P_1	P_2	P_3	$P_3 - P_1$		
		Mean return (% daily)			Mean	alpha	SR
Arbitrage Index	1	0.32	0.53	0.66	0.34	0.33	1.96
		(1.34)	(2.14)	(2.69)	(3.10)	(2.86)	
	2	0.28	0.74	0.59	0.31	0.29	1.21
		(1.09)	(2.76)	(2.29)	(1.90)	(1.69)	
	3	0.49	0.55	0.60	0.11	0.13	0.36
		(1.83)	(2.09)	(2.29)	(0.55)	(0.66)	

L-Mean OF

		P_1	P_2	P_3	$P_3 - P_1$			
		Mean return (% daily)			Mean	alpha	SR	R^2_{oos}
Arbitrage Index	1	0.55	0.50	0.43	-0.13	-0.12	-0.80	-0.80
		(2.20)	(2.11)	(1.78)	(-1.12)	(-1.19)		
	2	0.41	0.72	0.49	0.09	0.13	0.37	-0.74
		(1.49)	(2.69)	(1.98)	(0.52)	(0.84)		
	3	0.15	0.67	0.82	0.67	0.71	2.19	0.02
		(0.55)	(2.68)	(2.90)	(3.27)	(3.46)		

NL-Mean OF

		P_1	P_2	P_3	$P_3 - P_1$			
		Mean return (% daily)			Mean	alpha	SR	R^2_{oos}
Arbitrage Index	1	0.33	0.50	0.68	0.35	0.36	2.15	0.83
		(1.39)	(2.07)	(2.71)	(3.28)	(3.32)		
	2	0.32	0.66	0.64	0.33	0.38	1.25	0.45
		(1.20)	(2.61)	(2.35)	(2.01)	(2.22)		
	3	0.12	0.50	1.02	0.90	0.86	2.65	0.62
		(0.48)	(2.06)	(3.30)	(3.94)	(3.95)		

Figure 1: The Cumulative Return of Long-Short Portfolios

The figure plots the cumulative return for the long-short portfolio of select machine learning models that condition on the international order flows (ML OF models). The figure also reports the return of the market portfolio, which is the value-weighted return of all the coins in the sample. The sample period ranges from February 18, 2020 to June 30, 2022.