NBER - Big Data and High-Performance Computing for Financial Economics

Using Big Data to Understand Cryptocurrency Derivatives Markets and Regulation

April 5, 2021

Nicolas Christin,[1] **Bryan R. Routledge**,[2] **Kyle Soska**,[3] and **Ariel Zetlin-Jones**,[4]

Abstract

Cryptocurrencies and virtual assets have gained considerable popularity, if not widespread acceptance, in the past decade. In the past few years, trading instruments for these virtual assets have become both increasingly popular and complex. Besides spot trading cryptocurrencies for traditional currencies, many online platforms now offer the ability to trade derivatives. As part of an ongoing data collection effort, we have access to detailed account-level data from one of the largest cryptocurrency derivatives exchanges, Binance Futures. Specifically, between August 2020 and April 2021 (on-going), we have gathered close to three billion observations of account balances, roughly hourly, for over four million accounts. Using additional public data (e.g., minute-level price data for many cryptotcurrencies), we propose to solve the inference problem to obtain estimates of each accounts' positions and, then, trades. This fine-grained user detail allows us to better understand how users trade cryptocurrency derivatives. Our evidence and inferences provides insights into the role of these new platforms and on the role that existing or future regulations might play. This project builds on the work by Soska et al. (2021), which uses data from the BitMEX derivative exchange to explore the evolution of exchange-traded products.

[1]School of Computer Science and Department of Engineering and Public Policy, Carnegie Mellon University
[2]Tepper School of Business, Carnegie Mellon University
[3]Department of Electrical and Computer Engineering, Carnegie Mellon University
[4]Tepper School of Business, Carnegie Mellon University

Using Big Data to Understand Cryptocurrency Derivatives Markets and Regulation

Introduction

Markets for cryptocurrency derivatives are sizable and growing. In the past year, the average daily volume of cryptocurrency futures trades has grown from $25 billion to over $100 billion and is now comparable to total daily volume of spot trading on the New York Stock Exchange. Some traditional exchanges, such as the CME Group, offer cryptocurrency forward contracts. However, these exchanges are only small part in this emerging marketplace. Most volume in this market takes place at start-up exchanges, such as Binance, Bybit, Okex, BitMEX and others who offer complex, non-traditional derivatives products. Specifically, many of these derivatives incorporate large non-recourse leverage. Regulators have taken notice. In the United Kingdom, the Financial Conduct Authority banned the sale of crypto-derivatives to retail customers.[5] In the United States, the Commodity Futures Modernization Act of 2000 (CFMA) and 7 U.S.C. §1 makes it illegal to offer foreign currency derivative products to retail investors. While this Act predates modern cryptocurrencies, the current suit filed by the SEC against BitMEX[6] on alleged violations of the Bank Secrecy Act (31 U.S.C. §5311), seems to indicate that regulators are following the space closely.

Data

In previous research (Soska et al., 2021), we documented new evidence on BitMEX, a central cryptocurrency derivatives exchange. For the period January 2018 and April 2021, BitMEX trading volume was on average 3 billion dollars per day. Much of the growth of BitMEX was attributed to its innovative structure that allows traders to take levered positions, up to 100x, long or short Bitcoin. We analyzed the evolution of BitMEX products—both settled forward contracts and their innovative "perpetual forward." We used on-chain Bitcoin blockchain forensics, public liquidation events, and a site-wide chat room to describe the diverse ensemble of amateur and professional traders that forms this community. Our findings include evidence of around the clock ("potentially obsessive") use of the exchange by some traders and evidence of large liquidations of smaller accounts seemingly benefiting larger players resulting in increased wealth concentration over time.

In this project, we obtain complementary, finer grained data from one of the largest exchanges in the marketplace today, Binance. Binance mimics many of the features of BitMEX. Binance offers two major, segregated services: a "spot" market where customers can exchange various "fiat" currencies for Bitcoin and other cryptocurrencies or virtual assets, and a "futures" market, on which Binance offers trading on over 120 derivative products with 125x long/short leverage available. By querying Binance's interface we have gathered close to three billion observations of account balances, roughly hourly,[7] for over four million accounts – nearly every trader on this exchange. The raw data include profit (loss) and return on investment. Both of these are for various time intervals. Effectively, this lets us infer each account's balance at the fine-grained hourly time scale. We also have 75 million observations of trading positions that were voluntarily disclosed by traders as well as meta-data on minute-level cryptocurrency prices

[5] See `https://www.fca.org.uk/news/press-releases/fca-bans-sale-crypto-derivatives-retail-consumers`.

[6] United States of America vs. Arthur Hayes, Benjamin Delo, Samuel Reed, and Gregory Dwyer. United States District Court, Southern District of New York. Indictment 20Cr.500. September 2020.

[7] We over-sample more active accounts.

on Binance and other exchanges. Our data covers the period from August 2020 to April 2021 (and ongoing). For reference, over this period Bitcoin traded in a range from $10,123 to $61,130.

Proposal - Infer Trader Positions

Our research proposal seeks to estimate the latent trading positions of each trader from the observed hourly account balances. For exposition, denote time as discrete $\mathcal{T} = 0, 1, ..$ (say, minutes). We observe trader i's wealth, $W_{i,t}$ for $t \in T$ for some points in time (say hourly), where $T \subset \mathcal{T}$. A trader's position at date t is the N-vector $\theta_{i,t}$ that represents a position in each contract on the exchange $N \approx 120$. We observe prices on the N assets, $P_{n,t}$. Putting this together, we wish to infer $\theta_{i,t}$ imposing for $t \in T$, for each i, $W_{i,t} = \sum_n \theta_{i,n,t-1} P_{n,t}$.[8] The high dimensionality of the portfolios makes this a challenging inference problem. The problem is also "under-identified" since there will typically be many possible portfolios that are consistent with an observed change in wealth. However, this problem is solvable with machine learning tools similar to those used for inferring a latent representation for observed data (e.g., "collaborative filtering," "matrix completion," "word embeddings"). We also plan to incorporate ancillary data, like aggregate market volume data to refine the estimate (e.g., Bitcoin perpetual future is the most heavily traded of the Binance contracts).

Impact

We propose to estimate detailed trade and position data for almost all of the four million traders on one of the largest cryptocurrency exchanges. This is of immediate interest to the design and regulation of cryptocurrency exchanges. For example, our prior work on BitMEX (Soska et al., 2021), suggested that distinct trader types were present on the exchange. From news reports and regulation filings exchanges have customers that are retail, traditional institutional investors, and funds specializing only in cryptocurrency. Can we identify these classes of traders from the position estimates? Does performance differ across these groups? Second, leverage on Binance, like other cryptoexchanges, is typically higher than other derivative markets (especially given the the high level of price volatility in the underlying asset like Bitcoin). The result is a large number of "liquidations" where a trader's position is closed out. Finally, cryptocurrency exchanges are a laboratory for security design. The most popular is the "perpetual futures contracts" (first introduced on BitMEX). The contract has characteristics that are similar to a swap contract in that there are interim cash-flows based on the contract's price relative to an index and has characteristics that are similar to an option contract in that price volatility can cause a liquidation. Why is this contract predominant? How is it used in a trader's portfolio. Does the structure of the contract contribute to the concentration of wealth on the exchange? Finally, the data is of interest to financial economics more generally. Detailed and high frequency portfolio and trading data is rare and allows research into many aspects of portfolio choice and asset pricing.

References

Soska, K., Dong, J.-D., Khodaverdian, A., Zetlin-Jones, A., Routledge, B., and Christin, N. (2021). Towards understanding cryptocurrency derivatives: A case study of BitMEX. In *Proceedings of the 30th Web Conference (WWW'21)*, Ljubljana, Slovenia (online).

[8] If we ignore transaction costs and interim cash-flows, the budget constraint at date t implies that $W_{i,t} = \sum_n \theta_{i,n,t} P_{n,t}$.