Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis - Quantitative Finance Programme

Order Book Dynamics - Two-Dimensional Exit Problems on a Cryptocurrency Exchange

Student	ir. J.M. de Jong
Student I.D.	450700jj
Supervisor	drs. S.H.L.C.G. Vermeulen
Second assessor	dr. R-J. Lange
Date final version	January 21, 2020

I would like to express my gratitude to my thesis supervisor, drs. Sebastiaan Vermeulen, and coordinator, dr. Rutger-Jan Lange, for their helpful comments, feedback and patience. I would also like to thank prof. dr. ir. Johan Grasman for our discussion on asymptotic solutions of exit problems, and the people at Blocktraders for pointing me in the right direction for automated data collection. Last but not least, I would like to thank my parents for their unconditional support during the writing of this thesis.

ERASMUS UNIVERSITEIT ROTTERDAM

Erasmus School of Economics

Abstract

This thesis reports on the applicability of drift-diffusion and diffusion-only models to estimate future price directions from Level-1 limit order book data. We fit and evaluate the probability models on a data set of 78M order book updates (1 month) of the Bitcoin – U.S. Dollar market on *BitMEX*, a large cryptocurrency exchange.

We show that diffusion-only models outperform drift-diffusion models, which is likely a result of the limited Level-1 description of the order book. Both types of models offer significant improvements over baseline models that do not incorporate order book information. On selected evaluation points leading up to price changes, the best-performing model obtains an RMS direction prediction error of $9.18 - 30.52\%$. A model-based price direction classifier obtains an average classification precision on the same points in the range of $93.83 - 99.64\%$.

We finally demonstrate effectiveness of the classifier in a simple trading strategy, resulting in an average return of 0.062% per trade and win rate of 69.12% over 38K trades. A realistic implementation of the strategy will require further analysis in price move sizes and/or limit order execution.

Keywords: Drift-diffusion, cryptocurrency, exit problems, high-frequency trading, limit order books, market microstructure, option valuation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

At the heart of modern trading exchanges is an electronic matching system to track and execute orders given by market participants. This system continuously updates a list of open orders, which express buying or selling interest, with their associated price level and order volume. This double list is generally called the Limit Order Book (LOB), and the process of executing these orders the Continuous Double Auction (CDA). As the market participants send their instructions, liquidity is either created by accumulation of open interest (limit order), or removed through a trade (market order) or a cancellation of a limit order.

This interplay of existing and arriving orders results in the price discovery of the asset trading on the exchange. The arrivals of orders and cancellations can therefore be viewed as the microscopic process driving the macroscopic price development and is hence of major interest to a multitude of parties. Market participants wish to inform themselves of the behaviour of other participants and the potential impact of a large-volume trade. Market regulators want to monitor trading behaviour and evaluate the performance of the order matching system. Scientific and quantitative researchers may use the LOB to develop new dynamical models or to back-test trading strategies.

Arguing that the main order flow is directed at the best bid and ask prices, Cont and De Larrard (2013) proposed a reduced-form model, which takes into account only the volume at these two levels. Inspired by Cont, Stoikov, and Talreja (2010), Avellaneda, Reed, and Stoikov (2011) define a similarly reduced, but symmetrical model which includes the negative correlation of the best bid and ask queue sizes, and derive the price increase probability.

It is known from Biais, Hillion, and Spatt (1995) that limit order arrivals show clustering behaviour. It is therefore a serious limitation of both the Cont-Stoikov-Talreja (CST) model and the Avellaneda-Reed-Stoikov (ARS) model that the arrival rates are modeled as constants.

In this thesis, we investigate whether the analytical predictions of the price process, provided by the aforementioned CST and ARS models, can be improved by allowing the Poisson process intensities to vary between arrival events. Additionally, we introduce and compare two new models for predictions of the price increase probability, based on convection-diffusion results of López and Sinusía (2004). These correspond to heavy-traffic approximations of more general imbalanced (drifting) arrival processes with and without correlation.

The application of these reduced-form models on cryptocurrency markets is motivated by findings of Schnaubelt, Rende, and Krauss (2019), who investigated the Bitcoin (XBT) − United States Dollar (USD) market. At the time of writing, Bitcoin is the world's largest cryptocurrency by market capacity. The authors point out that average order book volume has a global maximum at the best bid and ask levels, indicating a high information content at the first LOB level. Secondly, real-time LOB data is readily available for such markets for free, facilitating the ready application of our models. The LOB under investigation in this thesis is a three-month-long data set of the XBT − USD market on *BitMEX*, the largest cryptocurrency exchange by reported trading volume (CoinMarketCap, 2019).

In a rolling window approach, we estimate the order flow parameters on the bid and ask sides of the LOB using exponentially filtered cumulative order flow processes. We assume that the fitted parameters at a specific point in time remain constant until the next price move. This way, the available analytical results for computing transition probabilities still apply. We investigate the effectiveness of the rolling window method by varying the weighting parameter of both the exponential and rolling window filters.

For the chosen set of evaluation points, all order flow models show significant improvement over baseline models that do not take the order flow into account. This indicates that the incorporation of Level-1 order volume information can provide a superior estimate of future price direction. The inclusion of drift effects does not provide better direction estimates, which is likely a result of the limited Level-1 data set. Inclusion of correlation, and parameter estimation from short rolling windows do not improve prediction performance either.

When the models are used as binary threshold classifiers, we observe that the time-averaged models generate (nearly) identical classifications. Classifier precision on the evaluation points ranges between $87.69 - 99.00\%$. If we exclude certain order volume combinations for prediction, classifier precision of the best-performing model can be increased to $93.83 - 99.64\%$, at the cost of reduced recall.

Finally, we show the effectiveness of the classifier for generating trading signals. Over the course of 1