Simulating a Limit Order Book: Metrics for High-Fidelity Simulations and Impact of Agent Behaviour

Meinte Ringia, Misho Yanakiev, Diogo Franquinho

May 28, 2021

Korteweg-de Vries Instituut voor Wiskunde
Faculteit der Natuurwetenschappen, Wiskunde en Informatica
University of Amsterdam

Abstract

Using real market data to make inference about the effects of specific strategies on the market and its participants is a challenging task, often simply infeasible. Various methods have been developed over the years to tackle this issue. In our paper, we make use of agent-based modelling as an alternative way of studying this type of market interaction and develop two simulation configurations for that purpose. The simulations are tested for realism using well-known stylized facts and yield the emergence of a subset of them clearly related to the types of strategies used in the markets. Building on that, we then study the impact of large order placements on the market and compare the performance of instantaneous order placement against interval (Percent of Volume) execution strategies. We find that in a semi-realistic market, interval orders perform better both for the party that places the order and for the overall health of the market.

Title: Simulating a Limit Order Book: Metrics for High-Fidelity Simulations and Impact of Agent Behaviour

Authors:

Meinte Ringia, meinte.ringia@student.uva.nl, 10626840

Misho Yanakiev, misho.yanakiev@student.uva.nl, 13150855

Diogo Franquinho, diogo.sousa.franquinho@student.uva.nl, 13470701

Date: May 28, 2021

Korteweg-de Vries Instituut voor Wiskunde

University of Amsterdam
Science Park 904, 1098 XH Amsterdam
`http://www.science.uva.nl/math`

Contents

1. Introduction

Having a thorough understanding of market dynamics has been a long standing goal (Friedman 2018) of any market participant, be it an investor or researcher, but its achievement has been inhibited by a myriad of issues of varying complexity. The most natural one is that of human behavioral psychology. While one may argue that the current state of technology makes it less relevant, it is still an overwhelmingly prevalent problem as 35% and 50% of the volume traded across EU and US equity markets, respectively, up to 2014 (Kaya 2016) has been done algorithmically. The complexity of this issue can be boiled down to the degree of unpredictability of how an individual market participant (be it a retail trader or an institutional participant such as a hedge/pension fund etc.) makes a decision (Friedman 2018). It may be argued that an investor will follow a thorough mathematical analysis of the current market trends along with proper predictive modelling to make the optimal decision possible at that point in time, but it is no secret that humans are emotional beings and often times succumb to less than perfectly objective reasoning. The manner in which a strategy universe is developed by such a participant and the manner in which a strategy is chosen and executed at a given time remains unknown and impossible to study from public domain market data. This touches upon the other major issue we identify - there is simply no 'equation of the universe' for market dynamics. The absence of a generalized mathematical model which encapsulates all fundamental market characteristics prompts researchers into using data-based analysis to derive insights under stringent conditions. This data-based approach is itself inhibited by either the basic availability of the needed data or the granularity of the data itself Vyetrenko et al. 2019. There is also the additional historical data-related issue which is simply unavoidable - replaying public domain historical data cannot reveal much about the actual sets of trading strategies used by market participants, nor can it give much insight as to how decisions were made.

There have been three major approaches employed to circumvent these problems (Friedman 2018), namely field studies of real markets; laboratory studies of small, synthetically created market structures and finally, computer simulated markets. Field studies are successful in gaining insight but suffer from lacking access to complete market information. Laboratory studies on the other hand have a more controlled approach and full access to relevant information, but are small and expensive. Finally, computer simulation would seem to be an efficient solution. Insofar, the biggest fundamental issue with computer simulation has been the fact that trading strategies are externally specified by the researcher(s) rather than the market participant and thus cannot represent an actual market environment, but this is no longer the case (Vyetrenko et al. 2019). In the paradigm of agent-based modelling, this issue can be overcome by allowing for basic background agents that have fixed strategies to comprise the bulk of market participants and then introducing reinforcement learning agents who adjust strategies on their own. In that vein, computer based simulations tick all boxes for having the capacity to generate sufficiently realistic market conditions suitable for relevant research.

In our paper, we make use of agent-based modelling via the open source ABIDES (Agent-Based Interactive Discrete Event Simulation) framework available on GitHub to explore two main questions: How realistic can we make computer simulations be and then, given a certain degree of achieved realism, we compare the effects of specific agent behaviour on the market. The specific behaviour we are interested in are the effects of large orders and how can an agent diminish this effect. Our interest in large orders stems from the idea that any large orders should move the market due the to orderbook micro-structure, namely that they can impact the price through swinging liquidity heavily in one direction and introducing a large spread. This in turn would cause a noticeable, at least short term, change in the price in comparison to a market where no anomalously sized orders exist. To explore these effects, the large orders come in two shapes, as a single high impact order and as an execution strategy order that is spread out over a given interval in smaller amounts using a percent of volume strategy. We follow the work done by Wellman and Wah 2017, as well as Vyetrenko et al. 2019. From Wellman and Wah 2017 we take the game-like market setting and the two fundamental type of market participants, namely the so-called Zero Intelligence Agent and the Heuristic Belief Learning Agent, both of which are described in detail in subsection 3.1.1. From Vyetrenko et al. 2019 we make use of the market stylized facts as measures of market realism to judge the simulation configurations we run. Based on these results, we then follow with the research done in Byrd, Hybinette, and Balch 2019 to explore the effects of a single high impact order on the execution price. Finally, we then compare the efficiency of that large single order against a same sized order, this time executed using a percent of volume strategy over a given time interval.

We outline our approach as follows. In Chapter 2 we present the stylized facts we make use of in this paper along with some brief theory supporting them. In Chapter 3, we outline the full market model in use with the agent types and simulation settings within which we use said agents. In Chapter 4, the reader can find all results obtained based on the previously stated models and hypotheses, and finally, Chapter 5 and 6 present our discussion and suggestions for future research.

2. Stylized Facts

To understand what stylized facts are, we first need to be familiar with some concepts and mechanisms from real-world public exchanges. These exchanges maintain a limit order book (LOB) from each asset traded at the exchange. The LOB represents a snapshot of the supply and demand for the asset at each point in time. It is a record of different types of events such as submission of a new limit order, cancellation, deletion, execution of a visible limit order and execution of a hidden limit order. Stylized facts are properties of the LOB behaviour that are repeated across a wide range of instruments, markets, and time periods. By comparing the stylized facts of data generated from real markets with data generated by our market simulations we can analyse the degree of fidelity by which our simulations are able to reproduce real market dynamics. There is not a consensus on the literature on whether such stylized facts are created by traders behaviour or are a consequence of the market's mechanisms and in our paper we will further explore this. Furthermore, we will only present the stylized facts that we analyse later, more information on other stylized facts can be found in (Vyetrenko et al. 2019).

2.1. Asset Return Stylized Facts

The asset return stylized facts are statistically descriptive regarding the asset price dynamics through time. In the literature (Cont 2001) (Ballocchi et al. 1999), there is consensus on the dynamics of the asset price that we will discuss in this section. At time t, the best bid b_t and best ask a_t are used to derive the mid-price $m_t = \frac{a_t - b_t}{2}$. For each increment Δt these are the log returns are $r_t = \log\left(\frac{m_t}{m_{t-\Delta t}}\right)$.

- **Heavy tails and aggregational normality:** It is commonly accepted that the distribution of financial asset returns possesses heavy tails. When the period of time Δt over when we calculate the returns is increased, the distribution of the asset returns will start to resemble more the normal distribution. A method to quantify deviation from normal distribution is to calculate its kurtosis.

- **Volatility clustering:** Periods of high market volatility events tend to cluster in time(Cont 2001). A measure for volatility clustering at time lag τ is though autocorrelation $corr(r_t^2, r_{t-\tau\Delta t}^2)$. Empirical studies using returns from various equities indicate that this autocorrelation function remains significantly positive over several days, which suggests periods of high volatility clustering.

- **Absence of autocorrelations:** Linear autocorrelations $corr(r_t, r_{t-\tau\Delta t})$ of asset returns over periods τ longer than 20 minutes are insignificant.

2.2. Stylized Facts about volume and order flow

- **Number of orders in a fixed time window:** Number of orders in a fixed time window can be approximated by gamma or lognormal distributions.

- **Order inter-arrival times:** In the literature, LOB order inter-arrival times are suggested to be fit into exponential, lognormal, and Weibull distributions (Abergel et al. 2016).

- **Intraday volume patterns:** LOB volumes are known to exhibit strong intraday patterns. For instance, historical exchange trading volumes can be approximated by a "U-shaped" quadratic-fit. Similarly, in most equity markets, volumes are highest in the beginning of trading day, followed by a period of lower activity, and then spike again at the end of the trading day (*Trades, Quotes and Prices: Financial Markets Under the Microscope, 2018*).

3. Market Model

The market model follows a Continuous Double Auction model, similar to the NASDAQ and NYSE exchanges as discussed in (Friedman 2018). In our simulations, we have the same basic exchange setup for each market type which is as follows. The exchange is facilitated by a single exchange agent who controls all communication between the market participants, the orderbook and the relaying of information about the fundamental price process and orderbook status to the participants. No communication outside the exchange agent is allowed between market participants and all messages between the exchange and the trading agents are delayed according to a latency model.

We have made use of two latency models. In the stylized facts simulations, we use a deterministic latency model which samples uniformly a large range of values and then creates the same number of points as agents in the simulation. It then computes the pairwise differences between these points and assigns the minimum of these values as the latency for the agents. In the impact and POV Agent simulations, we use a model with zero latency which means that agents trade at the nanosecond scale and have instant access to orderbook and price process information.

In every simulated market setting, the fundamental price follows a mean-reverting stochastic process of the form:

$$r_t = \max[0, \kappa\bar{r} + (1-\kappa)r_{t-1} + u_t], \tag{3.1}$$

where $\kappa \in [0,1]$ specifies the degree to which the fundamental reverts back to the mean and the parameter $u_t \sim \mathcal{N}(0, \sigma_s^2)$ is a random shock at time t. The fundamental price process represents the intrinsic value of the stock which varies over time according to a stochastic process. Finally, the last thing to note is that all prices observed in our paper are given in cents.

3.1. Market Participants

All agents in our simulation have some common constraints. With respect to trading, the Zero Intelligence and Heuristic Belief Learning agents have private valuation vectors $\Theta = (\theta_i^{-q_{max}+1}, ..., \theta^{q_{max}})$ where q_{max} specifies the maximum number of units a trader can be long or short, set to $q_{max} = 10$ throughout all simulations, and Θ is to be understood as the private valuation over all possible positions in terms of marginal private gain by either buying or selling one share, i.e. $\theta^i \rightarrow \theta^{i\pm1}$. The vector Θ is randomly sampled from an $\mathcal{N}(0, \sigma_{PV}^2)$ distribution for each agent at the start of the simulation and kept fixed. Furthermore, each agent arrives at the market according to a Poisson process with arrival rate λ_a which is adjusted depending on the time frame of the simulation whether we want minute or nanosecond scales.

3.1.1. Agent Types

Exchange Agent: In addition to the price function mentioned in Equation 3.1, the exchange agent provides each market participant with a noisy observation $o_t = r_t + n_t$ of the fundamental price where $n_t \sim \mathcal{N}(0, \sigma_n^2)$ and also maintains an available orderbook history up to a certain number of trades, adjustable by a memory parameter. The latter function is of high importance when dealing with agents who make use of the history of the orderbook in their strategies.

Zero Intelligence Agent (ZI): The ZI agent is the most basic type of trader since its trading strategy is totally independent of the orderbook and is solely based on a noisy observation of the fundamental price. At the start of a simulation, the agent is provided with a strategy represented by a triplet (R_{min}, R_{max}, η) where R_{min}, R_{max} define a range around the agent's valuation of the price of the asset within which to place an order at a uniformly random chosen price, and $\eta \in [0,1]$ gives the probability that the agent will actually place that order in terms of fractional surplus which can be achieved given the current market orders. The agent's asset price valuation is a function both of the history of the asset and the current observed price o_t, which is achieved through Bayesian updating via:

$$\tilde{r}_t = \frac{\sigma_n^2}{\sigma_n^2 + \tilde{\sigma}_t^2}\tilde{r}_{t-1} + \frac{\tilde{\sigma}_{t-1}^2}{\sigma_n^2 + \tilde{\sigma}_{t-1}^2} o_t,$$

$$\tilde{\sigma}_t^2 = \frac{\sigma_n^2 \tilde{\sigma}_{t-1}^2}{\sigma_n^2 + \tilde{\sigma}_{t-1}^2}.$$

The price at which the ZI places an order is a combination of the fundamental price estimate, the range from the strategy and its private valuation, all randomized over a uniform interval:

$$p_i(t) = \begin{cases} U[\tilde{r}_t + \theta_i^{q+1} - R_{max}, \tilde{r}_t + \theta_i^{q+1} - R_{min}] & \text{if buying,} \\ U[\tilde{r}_t - \theta_i^{q} + R_{min}, \tilde{r}_t - \theta_i^{q} + R_{max}] & \text{if selling.} \end{cases}$$

Heuristic Belief Learning Agent (HBL): The HBL agent goes a step further in strategic complexity. Its main purpose is to estimate the probability that orders at various prices will be viable in the market and thus pick the price which optimizes its surplus given its private valuation vector Θ. This is achieved through adding an extra memory parameter which allows the agent to store information about a given number of past market states in addition to a two step process of first constructing a personal belief function:

$$f_t(P) = \begin{cases} \frac{TBL_t(P)+AL_t(P)}{TBL_t(P)+AL_t(P)+RBG_t(P)} & \text{if buying,} \\ \frac{TAG_t(P)+BG_t(P)}{TAG_t(P)+BG_t(P)+RBG_t(P)} & \text{if selling.} \end{cases}$$

which assesses the probability that an order at price P will result in a transaction and a price optimization problem for the actual order placement of the form:

$$P_i^{\star}(t) = \begin{cases} \text{argmax}_p(\hat{r}_t + \theta_i^{q+1} - p)f_t(p) & \text{if buying,} \\ \text{argmax}_p(p - \hat{r}_t - \theta_i^{q})f_t(p) & \text{if selling.} \end{cases}$$

In the belief function A and B represent asks and bids; T and R transacted and rejected orders; L and G describe orders with prices less than or equal to or greater than or equal to price P. Note that for this strategy to work, the HBL agent needs an orderbook history. When the length of the orderbook history is not enough to compute the prices, the HBL will behave exactly like a ZI agent. This also implies that when the exchange agent does not keep an order book, the HBL agent inherits the ZI agent behaviour.

Momentum Agent: A trader that assumes that the market oscillates and places orders accordingly. The agent wakes up each minute during the day and queries the price of the last trade. The agent calculates the price average of the past 20 and 50 time frames. If the 20-step average is larger than the 50-step average, the agent buys the asset, whereas it sells when the opposite happens.

Market Maker Agent: This agent is designed to provide liquidity to the market. The idea of a market maker is to place both buy and sell orders around the mid price. The agent queries market prices every second and sets a window around the mid price. The market maker places orders at multiple levels in the order book around this window.

Impact Agent: The impact agent places a single large order at a specified time. Often institutional participants have to spontaneously buy or sell large orders of an asset due to for example expiry of contracts or a new hedge. The impact agent places one large buy order at a predetermined time of the day. The agent has a greed parameter which represents the proportion of available order book liquidity near the spread it consumes at the time of trade. An impact trader that for example buys with greed $g = 0.25$ will place a market buy order for 25% of the shares on offer.

Percentage Of Volume (POV) Execution Agent: There is a consensus in the financial market that directly placing large orders negatively influences the direction of the price for the party placing the order. This agent aims to fill a large order without excessive impact on the market price. The POV agent accepts small quotes depending on a small percentage of the previously traded volume and does so until the desired large order is filled. The POV agent has a time window in which it places an order after a certain delay according to

$$(\text{order volume})_{i+1} = \frac{p}{1-p} V(T_i),$$

where p is the percentage of the volume and $V(T_i)$ is the volume that was traded in the previous time period.

3.2. Simulation Settings

3.2.1. Homogeneous Zero Intelligence

We introduce the ZI setting with the purpose of exploring whether or not some stylized facts emerge simply from fundamental market dynamics, disregarding any role of strategy-based order placing and price response. In that sense, we only populate the market with zero intelligence traders so that the state of the orderbook is not a part of the strategies available. We run a simulation comprised of 100 ZI agents with strategy profiles $(15, 0, 250, 1)$, $(15, 0, 500, 1)$, $(14, 0, 1000, 0.8)$, $(14, 0, 1000, 1)$, $(14, 0, 2000, 0.8)$, $(14, 250, 500, 0.8)$,$(14, 250, 500, 1)$ where the first entry in the tuple represents the number of agents with the given strategy (R_{min}, R_{max}, η). Since the default time scale is in nanoseconds, we adjust the arrival rate for the ZI agents to be $\lambda_a = 1 \times 10^{-12}$ which is equivalent to 1 arrival every minute on average. The remainder of the parameters are given in Table 3.1 below.

Parameter	$\overline{r}$	κ	σ_s^2	σ_n^2	σ_{PV}^2	$\$_{initial}$
Value	1e5	1.67×10^{-12}	0	10^3	5×10^6	10^6

Table 3.1: General parameter values

3.2.2. Complex Market (RMSC01)

In the RMSC01 simulation, we create a market setting with non-trivial strategic complexity. The purpose of this simulation is to fully explore existence of stylized facts and test the realism of the setting. The simulation is comprised of an exchange agent, a market maker, 50 ZI agents with a common strategy, 25 HBL agents and 24 momentum agents. All ZI agents share the same strategy profile of $(0, 100, 1)$. The HBL agent in turn has a memory length of 2 and inherits the same strategy profile as the ZI agent in case of no current market history. Our Market Maker agent is equipped with an order size range of $[500, 1000]$ and places orders at the current mid-price so as to decrease the bid-ask spread. Last but not least, the momentum agent has an order range set to $[1, 10]$. The simulation is again adjusted to a minute scale with arrival rate for all agents of $\lambda_a = 1 \times 10^{-12}$. The remainder of the general parameters are found in Table 3.1.

3.2.3. Impact and Execution

For the impact and execution simulations, we use the nanosecond scale with 1000 time steps with no latency on the same set of agents as the RMSC01 over which we then test the effects of impact and execution agents independently. The impact agent is set to trade at exactly the 200th time step for a buy order with greed parameter $g \in \{0.3, 0.7, 1.3, 1.9\}$ while the execution trader is set to trade over the interval between the 200th and 600th time steps with $p = 0.412$ chosen so that the order sizes of the impact and execution agents are the same. Finally, we run a comparison of mid-range greed of 0.7 against the pov of 0.412 and analyze how the agents perform. The remainder of the general parameters are the same as in Table 3.1 with the exception of $\kappa = 0.05, \sigma_s^2 = 5 \times 10^5$ and the arrival time, with $\lambda_a = 5 \times 10^{-3}$.

4. Results and Analysis

As outlined in subsection 3.2.1 and subsection 3.2.2, we ran these setups to observe the emergence or absence of stylized facts. We test this against historical data collected by picking 25 random stocks, provided in the appendix Appendix A, then taking a random 5-tuple per day, normalizing and aggregating the data and then repeating the process for 22 trading days, from 15th of April to 15th of May, 2021. The motivation behind this approach is to obtain a comprehensive view of the market so that the simulation comparison is not confined to a single asset but can be tested against a broader range.

4.1. Stylized Facts on the Asset Returns

Heavy tails and aggregation normality: From the graphs presented in Figure 4.1 we conclude that both simulations seem to produce heavy tail distributions and aggregation normality, with the RMSC01 simulation being more consistent at reproducing heavy tail distributions and aggregation normality observed in historical data. The sharp contrast between the ZI and RMSC01 tails would suggest that heavy tails are primarily a result of strategic trading.

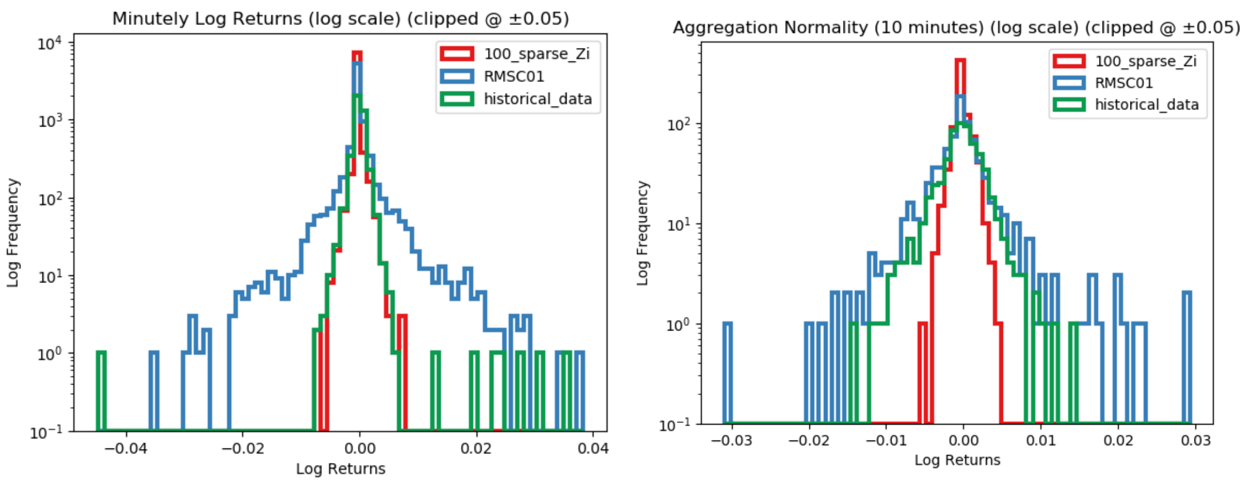


Figure 4.1: *(left)* Distribution of minutely log returns and *(right)* Distribution of log returns over 10 minute intervals

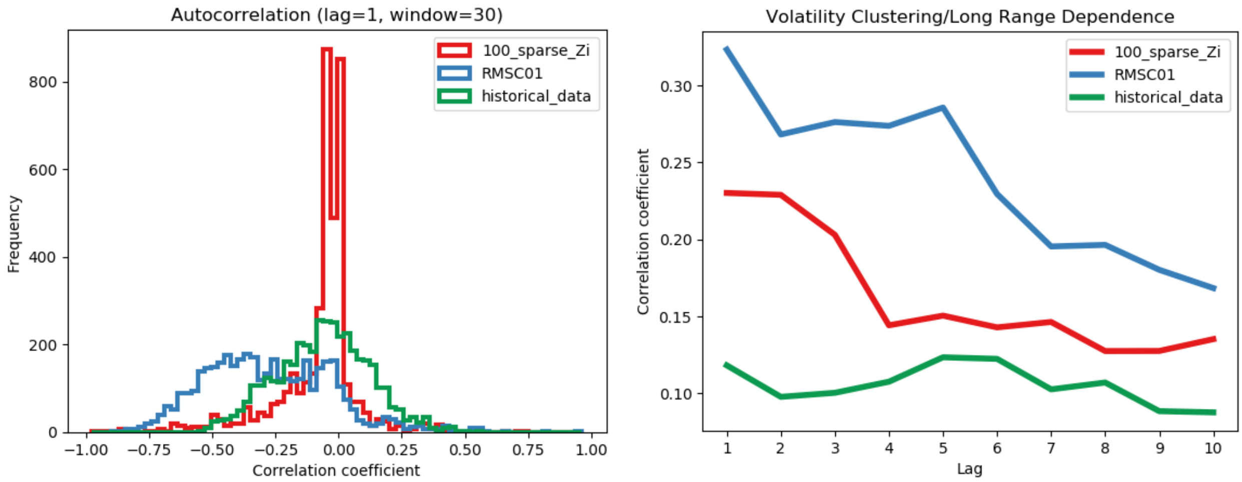


Figure 4.2: *(left)* Correlation coefficient distributions for lag 1 autocorrelation of log returns over 30 minute intervals and *(right)* Volatility Clustering

Autocorrelation: From the graphs presented in Figure 4.2 *(left)* we observe that the historical data seems to be centered around zero but with a larger than expected spread. This can be attributed to the fact that we have a relatively small sample size and also a relatively short time interval (only one trading month). As for the simulated data, we observed that the ZI simulation produces a clear spike in correlation around 0 but the RMSC01 simulation appears to be highly inaccurate in reproducing the zero correlation expected. The comparison between the three results would suggest that absence of autocorrelation could be both, a result of a strategy-independent market (ZI), and also a result of an intelligent market (historical). The subtlety to be pointed out here is that while in the historical market all participants are free to change their strategies, in the RMSC01 simulation we do not have a strategy-adjusting agent and as such there is no participant that can capitalize on the existing autocorrelation and affect the market in a way that corrects itself.

Volatility Clustering: From Figure 4.2 *(right)* we can see a common trend in all correlation coefficients, namely they all decay as time lag increases. In addition, we can observe that the correlation coefficient for the ZI simulation follows the historical data much closer, but making an inference on the nature of this stylized fact is difficult as the overall behavior is rather homogeneous with the only distinctive factor being the magnitude of correlation. The magnitude discrepancy can be attributed to the fact that none of the agents have the capacity to change their strategies and thus are susceptible to trading in a way that reinforces current market trends.

4.2. Stylized Facts on the Order Flow and Volume

To analyse order flow and volume stylized facts we run the ZI and RMSC01 simulations for a 1 day trading period. Moreover, to derive historical distributions, we consider order book historical data for MSFT stock traded on the NASDAQ exchange for 1 trading day on January 6th 2009, the data source can be found in Appendix A.

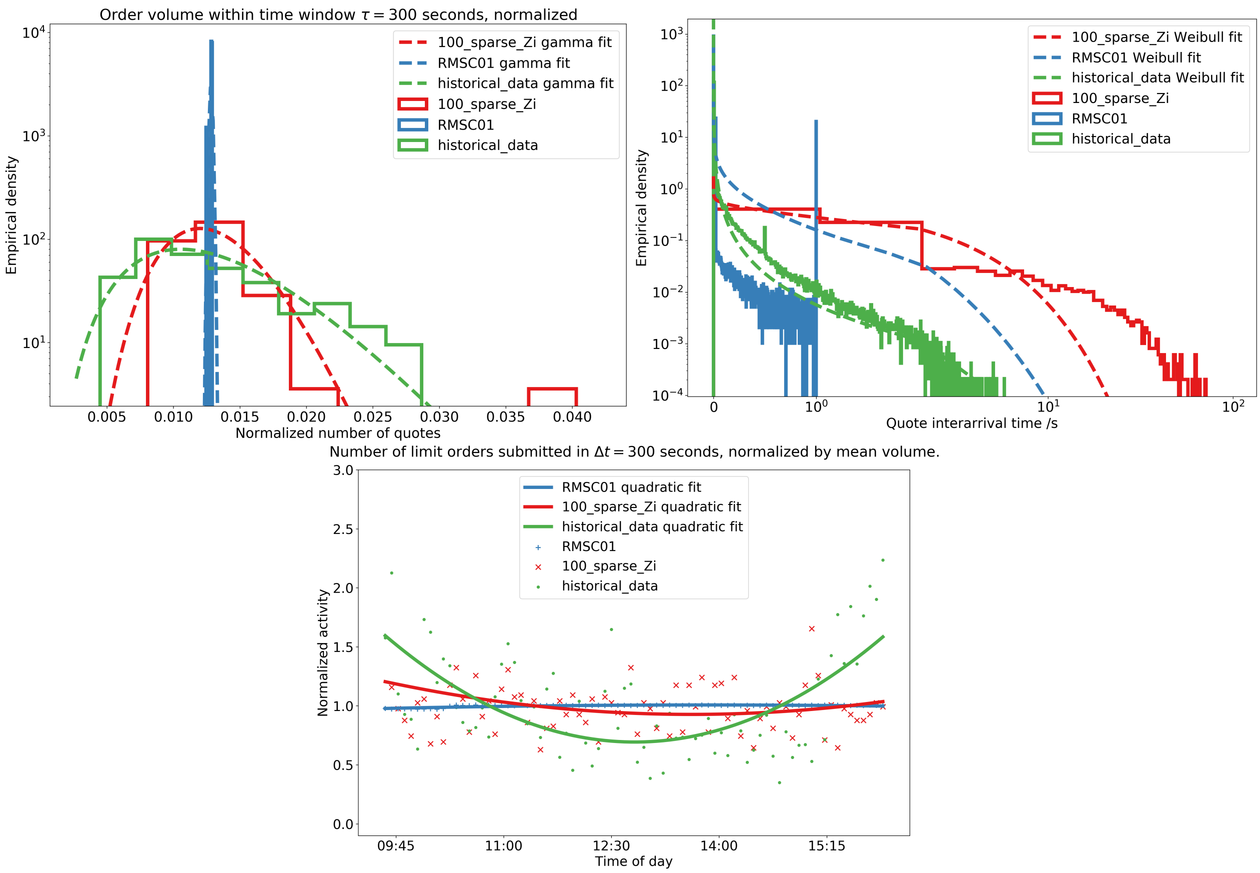


Figure 4.3: *(top left)* Number of orders in a 5 minute interval, *(top right)* returns as a function of time lag and *(bottom)* Volume/volatility correlation distributions.

Number of orders in a fixed time window: In Figure 4.3 *(top left)* we can observe the limit order flow distribution over a five minute window. We conclude that the all types of data can be fitted to a gamma distribution, although the shape of the historical data distribution is much better approximated by the ZI simulation.

Order interarrival times: In Figure 4.3 *(top right)* we can see that the weibull distribution is a good fit for the historical data. However, this distribution do not seem to be a good fit for either ZI and RMSC01 simulation data, especially the RMSC01 as we can observe that this data takes mostly values around 0. The lack of fitting for both the ZI and RMSC01 would suggest that the Weibull distribution of interarrival times is strictly dependent on self-adjusting agent behavior as this is the one fundamental difference between the historical and simulated datasets.

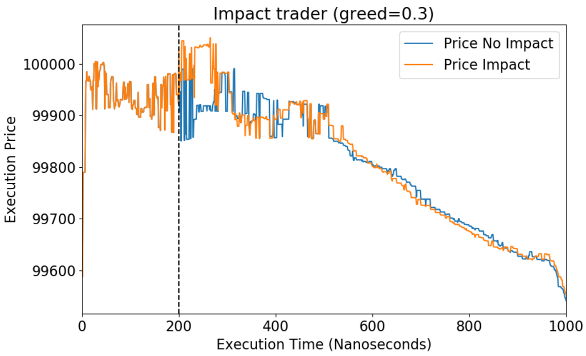
Intraday volume patterns: In Figure 4.3 *(bottom)* we have fitted quadratic-curves to the data, we can see that the historical data clearly demonstrate a "U-shaped" pattern in accordance to the theory. We can see that the ZI simulation reproduces this stylized fact to some degree, while the RMSC01 simulation produces slightly opposite results to the ones expected. This is probably due to the fact that in none of this simulations do the agents take market opening and close times into account, to better reproduce this stylized facts we would probably need to introducing liquidity during the market open and close better approximating real market behaviour.

4.3. Large Order Price Impact

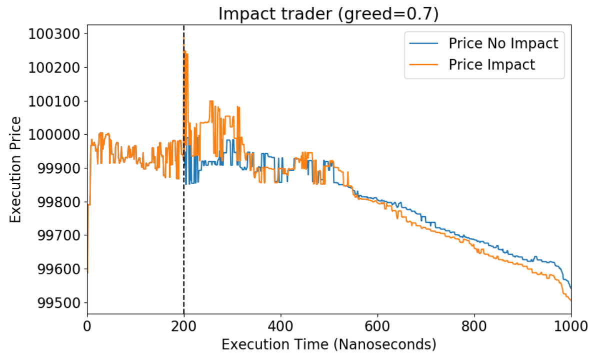
To analyse the effects of placing a large order we run the Impact and Execution simulations described in subsection 3.2.3, for a 1000 nanoseconds trading period, while making the Execution Agent inactive. We recall that the impact agent is set to trade at exactly the 200th time step.

4.3.1. Instantaneous Order Placement

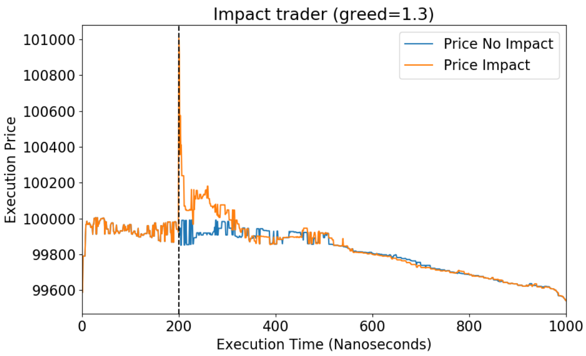
To analyse the effects of an Impact trader on price we plotted the execution price of the asset against the time, Figure 4.4, in comparison to the price in the same market with no impact agent. For this setting we that the greed parameters of 0.3, 0.7, 1.3 and 1.9 correspond to orders of sizes 5108, 11920, 22136 and 32353 shares respectively.



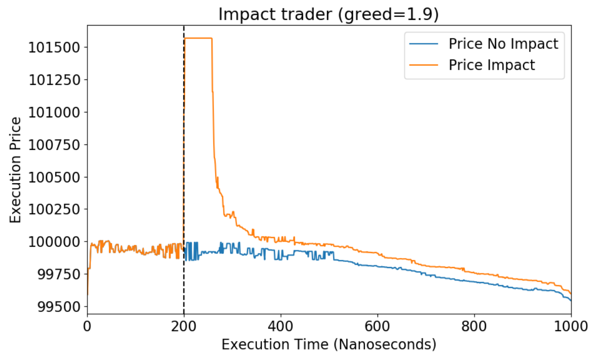
Order of 5108 shares



Order of 11920 shares



Order of 22136 shares



Order of 32353 shares

Figure 4.4: Impact agent execution price sensitivity

We can see that the initial impact on price varies from around 0.1% to 1.5% in the agents that have greed to by 0.3 and 1.9. In all settings we can see that all prices suffer a big impact at around the time the order is placed but after that the prices quickly recovers to a price close to the original, although never fully the same. We can observe for orders of large size there is a lack of liquidity to satisfy that order. This would account for the prolonged interval of higher prices in the first few steps following the impact trades of larger order size. Moreover, as one would expect, larger trades have a more prolonged effect on price.

As mentioned in the introduction, basic trading theory dictates that placing instantaneous big orders can lead to large losses for the party placing the orders. The losses of the impact agents during these simulations are provided in Table 4.1, which illustrates exactly how unwise it is to instantaneously place a large order.

Greed	0.3	0.7	1.3	1.9
Percent change	−33%	−74%	−120%	−312%

Table 4.1: Percentual change in final holdings of the impact agents

4.3.2. Partitioned Order Placement

As previously illustrated in Figure 4.3.1, instantaneous large orders negatively influence the party that places them and as such are not a good strategy when one has to either buy or sell large volumes. To investigate whether big orders can be filled without incurring major losses due to large price fluctuations, we compare the impact and execution strategies as seen in Figure 4.5. After running the simulations we compared the final valuations of the agents' portfolios and found that for the same volume of an order, the impact agent had incurred a loss of roughly 74% whereas the execution agent stood at a position 13% lower relative to their starting starting capital.

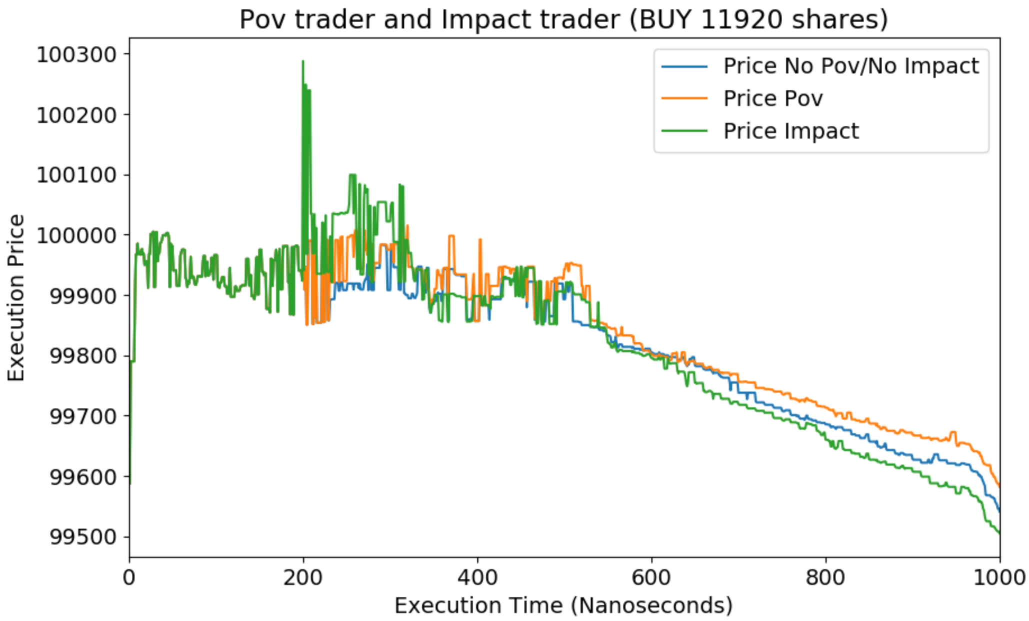


Figure 4.5: Comparison of price sensitivity between Impact and Execution orders

Figure 4.5 illustrates how the POV agent changes the execution price compared to a market without POV agent and a market without any party placing big orders. The POV agent places orders of 41.2% of the volume traded in the previous time interval every 10 nanoseconds. The orders take place between 200 and 600 nanoseconds and stop when the desired big order is filled. As we can see, the price of the asset is lower when the big order is completed by the POV agent compared to when the impact trader fills the order. This can be explained by the orderbook that gets time to recover from the placed trades. Overall, the POV order placement has a better outcome for the trader placing the order, with a percentual change of roughly -13%. Figure 4.5 also shows a lower effect on the execution price in the market. This suggests that placing big orders with respect to the Percentage Of Volume is preferable to placing the order all at once.

Note that in the Appendix, we also provide the stylized asset facts on the asset return distribution in Figure B.3. Despite that there are parties in the real market that place big orders in relatively short amounts of time, we are doubtful that a POV or impact agent has a significant influence on the stylized facts of the market.

5. Discussion

In our analysis we observed that some stylized facts emerge in only one of the two settings (ZI and RMSC01) and some in both. When it comes to asset return facts, we observe that autocorrelation and volatility clustering emerge in a setting where agents make decisions without taking the orderbook into account, which would imply that these stylized facts are a consequence of the market mechanism and not agent behaviour. On the other hand, heavy tails and aggregation normality seem to be stylized facts where agent behaviour appears necessary for their existence as we can see that the more complex simulation (RMSC01) is better at reproducing them. As for limit order book facts, we see quite an unusual result. While one would expect that the stylized facts of the ZI simulations would not resemble the stylized facts of historical data, as they do not apart from the Γ fit of the number of quotes, it could be expected that the more complex agents who take the orderbook into account, such as the HBL and market maker agents, would be able to at least partially reproduce some of the stylized facts. Contrary to expectation, the RMSC01 simulation does not yield any realism in that regard. While it does follow a Γ fit in the order distribution, the fit is too far from the historical data and in that sense can hardly be considered a realism marker. Apart from this one instance, the RMSC01 simulation fails at the remainder of the limit orderbook stylized facts and hints at the importance of having higher levels of strategic complexity.

On that note, one could argue that limit order book facts appear as a consequence of agents with the capacity to change their strategies based on the current market state. An agent of this sort will have the ability to exploit emerging market trends and strategize in a way that would yield higher returns until the trends have disappeared. In a sense, it would not be too far from the mind to hypothesize that self-adjusting agents are healthy for a trading market as they can exploit emerging trends (such as autocorrelation between returns, for example) and by doing that would create a trend of methods of exploitation, thus a meta-trend, which could in turn be exploited to beat the first wave of strategies, thus pushing the market in a way that negates the initial trend. At this level of strategizing, some of the orders in the orderbook will only implicitly follow the market and as such generate patterns that cannot be observed by fixed strategies that consider simply the orderbook status and the observation of the market price. As such, we believe that considerations of this sort make good motivation for introducing multiple learning agents and observing whether the realism of the simulated market improves in reference to the stylized facts.

As for the practical aspects of the accuracy of our simulations, we have to point out that we used a rather small historical data sample size and for a relatively short time period of time. A more comprehensive market representation would include a greater variety in stocks with larger daily sample sizes over longer trading periods. Furthermore, the simulated data also follows the same date ranges as the historical (1 trading month in asset returns and 1 trading day in limit order book facts) and as such can definitely be improved by running the simulation over a longer time period. The discrepancy in length of the trading periods between the asset return and limit orderbook facts, and the overall small dataset sizes, are due to the lack of freely available market data. As such, we had to make compromise and knowingly carry out a simulation which carries the risk of being refuted on grounds of statistical significance.

Turning now to the second stage of our analysis, the results appear to be in accordance with the theoretical expectation. We observed that the single placement of a large order consistently achieves negative returns with varied magnitudes and furthermore observed how it affects liquidity by the period of time at which the price remained significantly elevated after the order placement. The comparison between the impact and execution agents clearly shows that the POV execution strategy is favorable to a straightforward large order placement, despite the POV agent incurring losses as well. Building on that, we could infer that a more sophisticated interval trading strategy would be even better suited at handling large orders and result in lower losses.

Similarly to the stylized facts simulations, we also faced limitations in the analysis of the differences in the two trading strategies. On that note we believe that testing the strategies in a higher simulation range would produce more statistically consistent results, also testing the strategy in a more complex market structure involving learning agents could produce a more reasonable real world representation of the outcome of such orders. However, this type of analysis would come at a greater computational cost.

6. Future Research

As mentioned in the discussion, next to improvement on the methods and statistics, there is also a lot more to be done to simulate a better representation of a market. Our paper discusses mostly how to generally reproduce the outlines of a limit orderbook. However, there is a lot of oversight in choosing market participants, their strategy ranges and types and also the omittance of reinforcement learning agents. Future work can explore the interactions of these three levels of strategic behavior and go into more depth on the generating mechanisms for the real world stylized facts.

Latency is also a big variable where high frequency firms have invested large capital to ensure they achieve the lowest possible latencies. Researching the influence of the minimal and maximal viable latency for certain traders would be a topic of great interest within the financial world (Zook and Grote 2017).

Bibliography

[1] Frédéric Abergel et al. *Limit order books*. Cambridge University Press, 2016.

[2] Giuseppe Ballocchi et al. "The intraday multivariate structure of the Eurofutures markets". In: *Journal of Empirical Finance* 6.5 (1999), pp. 479–513.

[3] Jerry Banks. *Handbook of simulation: principles, methodology, advances, applications, and practice*. John Wiley & Sons, 1998.

[4] JP Bouchaud et al. *Trades, Quotes and Prices: Financial Markets Under the Microscope, 2018*.

[5] David Byrd, Maria Hybinette, and Tucker Hybinette Balch. "Abides: Towards high-fidelity market simulation for ai research". In: *arXiv preprint arXiv:1904.12066* (2019).

[6] Rama Cont. "Empirical properties of asset returns: stylized facts and statistical issues". In: (2001).

[7] Daniel Friedman. "The double auction market institution: A survey". In: *The Double Auction Market Institutions, Theories, and Evidence*. Routledge, 2018, pp. 3–26.

[8] Orçun Kaya. "High-frequency trading: Reaching the limits". In: *Deutsche Bank Research* (2016).

[9] Svitlana Vyetrenko et al. "Get Real: Realism Metrics for Robust Limit Order Book Market Simulations". In: *arXiv preprint arXiv:1912.04941* (2019).

[10] Elaine Wah and Michael P Wellman. "Latency arbitrage, market fragmentation, and efficiency: A two-market model". In: *Proceedings of the fourteenth ACM conference on Electronic commerce*. 2013, pp. 855–872.

[11] Xintong Wang and Michael Wellman. "Spoofing the limit order book: An agent-based model". In: (2017).

[12] Michael P Wellman. "Methods for empirical game-theoretic analysis". In: *AAAI*. 2006, pp. 1552–1556.

[13] Michael P Wellman and Elaine Wah. "Strategic agent-based modeling of financial markets". In: *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3.1 (2017), pp. 104–119.

[14] Matthew Zook and Michael H Grote. "The microgeographies of global finance: High-frequency trading and the construction of information inequality". In: *Environment and Planning A: Economy and Space* 49.1 (2017), pp. 121–140.

Appendices

A. Historical data

List of tickers for 22-day simulation: "AAPL", "GME", "TSLA", "LEGN", "CRCT", "GEVO", "VNT", "WISH", "UPST", "DASH", "PCT", "NKTX", "SKLZ", "PUBM", "JPM", "KO", "MCD", "MMM", "NKE", "PG", "AMGN", "AXP", "BA", "CAT", "CSCO".

MSFT (10 level) LOB data, January 6th 2009, source:
`https://lobsterdata.com/info/DataSamples.php`

B. Figures

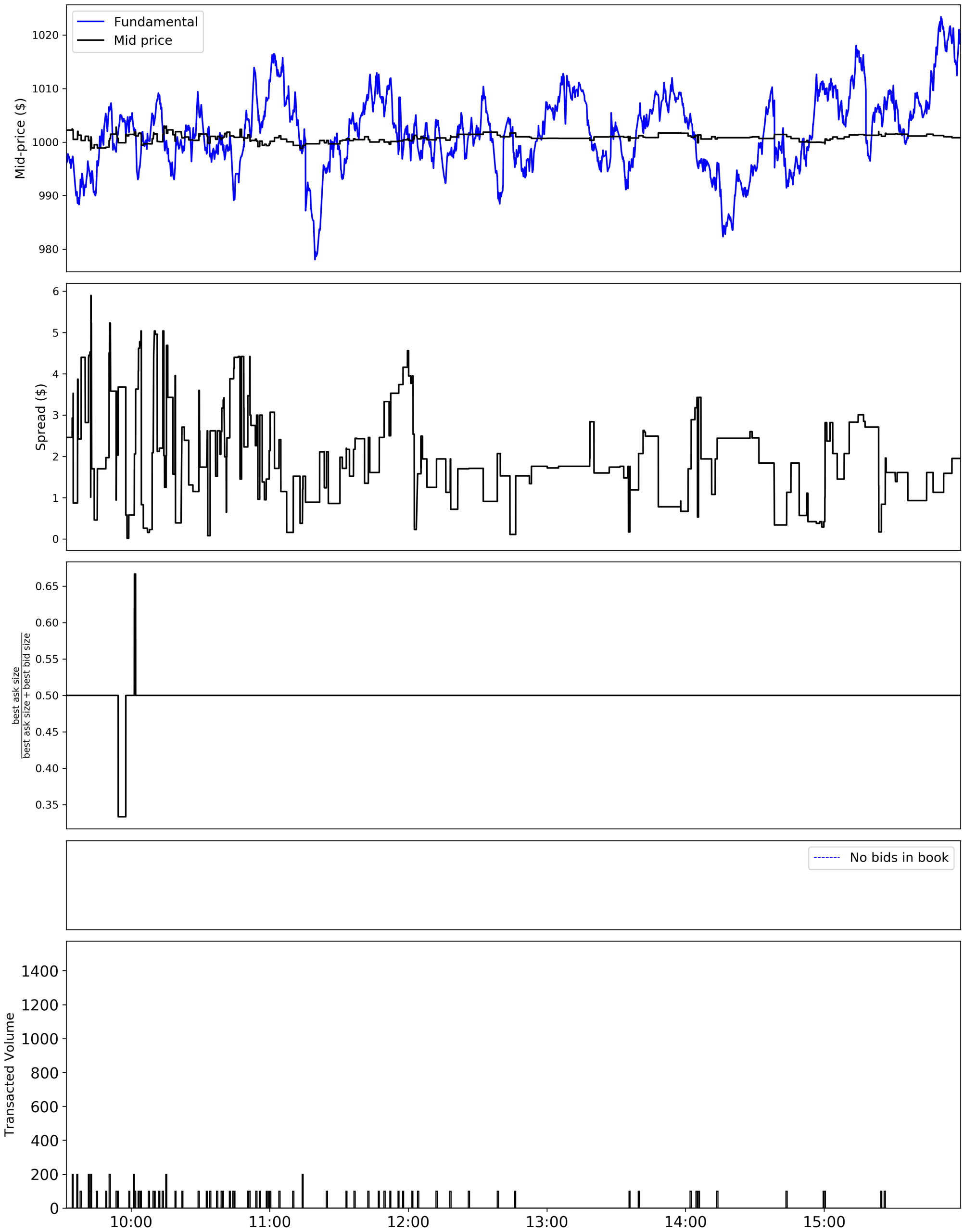


Figure B.1: Overall features of the Homogeneous ZI simulation

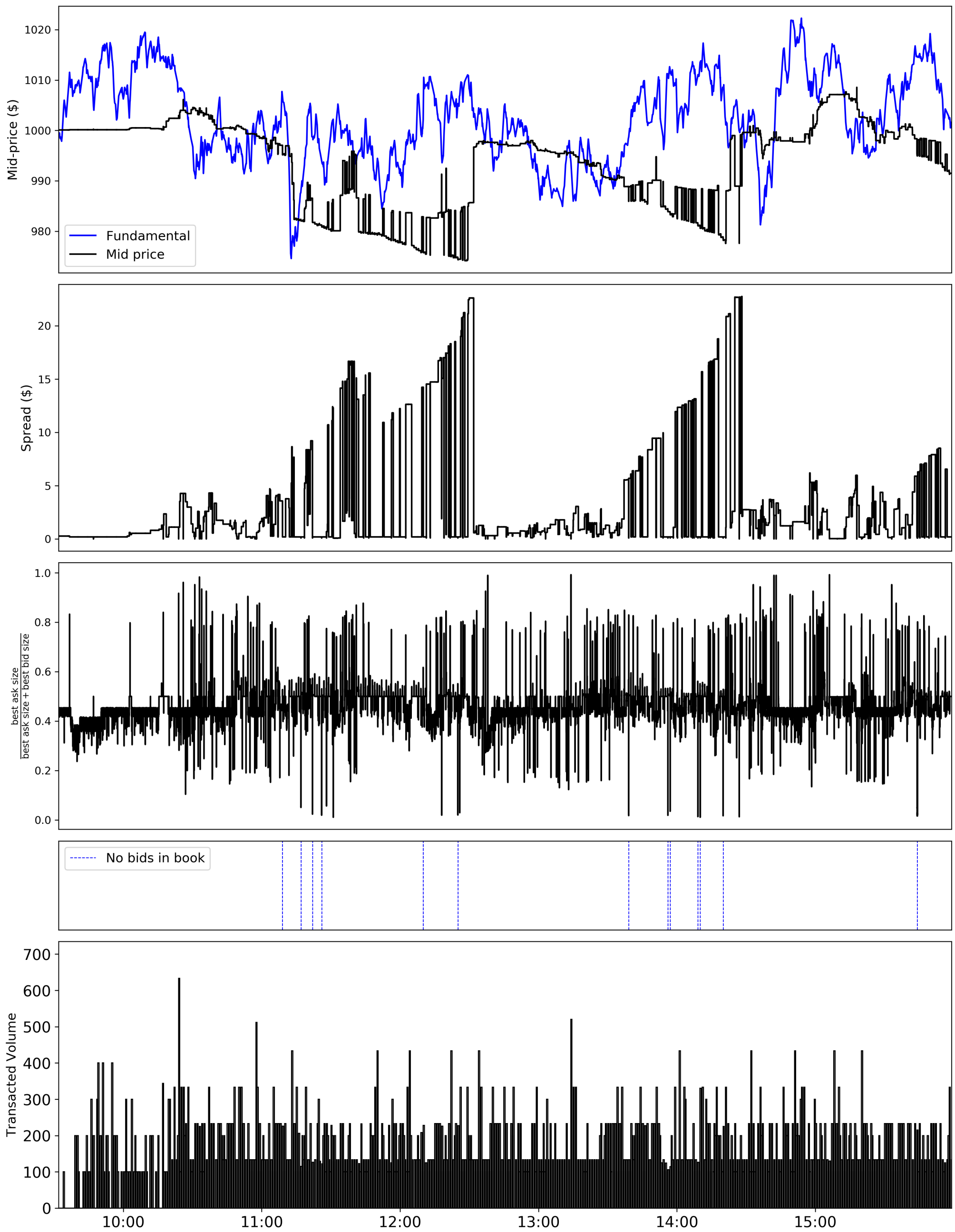


Figure B.2: Overall features of the RMSC01 simulation

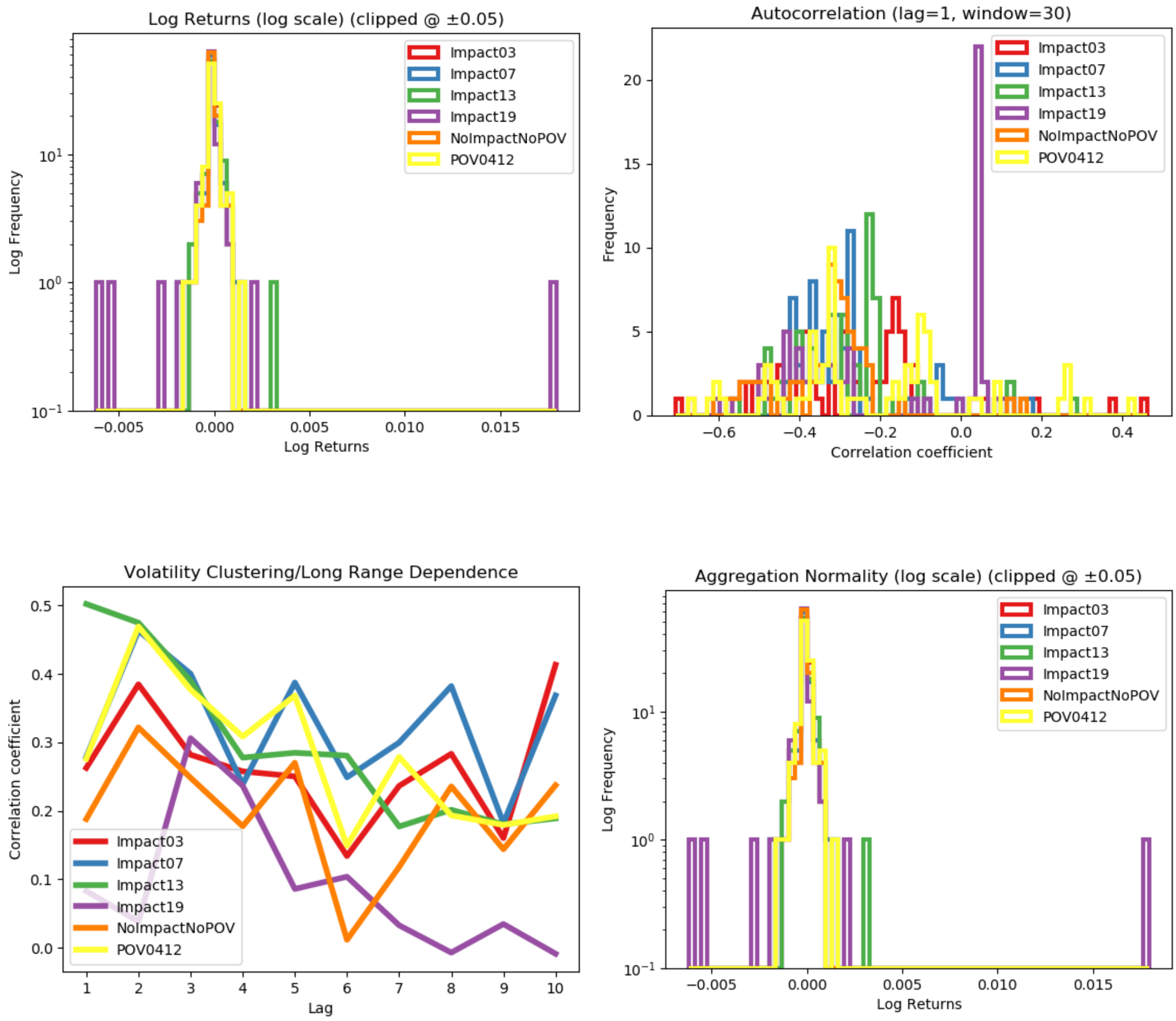


Figure B.3: Asset Return Stylized Facts of the Market configurations from Chapter 4