Master's Thesis

Modelling and analysis of a financial market with slow and fast trading agents acting on time-delayed market information

Halfdan Rump

Department of Computer Science and Engineerring
School of Fundamental Science and Engineering
Waseda University

Student ID	5115BG02-1
Date of Submission	February 5, 2014
Advisor	Prof. Toshiharu Sugawara

Contents

I Experiments 5

1 Overview of experiments 5
1.1 Correlation between fitness measures 6

2 Fitness and parameter evolution 8
2.1 Variable number of market makers 8
2.2 Fixed number of chartists and market makers 14
2.3 Variable number of chartists 14

3 Population-wide parameter/fitness correlations 14
3.1 Number of market makers 17
3.2 Market maker latency 18
3.2.1 Fixing the number of market makers 20
3.3 Number of chartists 20
3.4 Chartist latency 24
3.5 Chartist to market maker ratio 30

4 Grouping models by behavior 30
4.1 Manually grouping simulations by behavior 32
4.1.1 Fast simulations 35
4.2 Using clustering algorithms to group markets by behavior 36
4.3 Clustering with mixture of Gaussians 36

5 Summary of results 39

II Discussion 43

6 Benefits of fast traders 43
6.1 Fast market makers reduce price flickering 43

7 Agent strategies market crashes 43
7.1 Frequency of crashes 45
7.2 Agent speed and market crashes 46

8 Market makers causing the stock to be over-valued 46

9 title 46

10 Co-location 46

11 Strategy crowding 46

12 Future work **46**

Abstract

This work proposes a model in which multiple heterogeneous agents use time delayed price information to trade an imaginary financial instrument in a market with a continuous double auction. The main innovation of the model is that, just as in the real world, trading agents do not have access to the market information at the same point in time, which means that the agents generally trade on different information. The model contains agent models of slow human traders using a noisy fundamentalist strategy, and high speed software traders using market maker and chartist strategies. Slow traders base their trading decisions on market information which has been delayed several seconds, while the fast traders observe the market with much smaller delays ranging between a few milliseconds to a few hundred milliseconds. By containing agents of such different speeds, the model is relevant to the ongoing discussion of the pros and cons of high frequency trading in financial markets.

The model simulates the market events in the minutes after the advent of bad news represented by sudden negative shock to the fundamental stock price, and does so with a time resolution high enough to register events that unfold from millisecond to millisecond. A genetic algorithm is used to search for model parameters that cause the market to be stable, and parameters that cause the market to crash. Analysis of the results shows that a moderate level of high speed trading activity is not in itself problematic. In fact, high speed market makers are found to reduce price flickering, while high speed chartists are found to decrease the time required for the market to respond to changes in the fundamental price. However, the market is found to respond unfavorably to a large presence of fast traders. First of all, a large presence of fast market makers causes the market to respond sluggishly to the shock, leading to a prolonged disparity between the traded price and the true fundamental price. Secondly, a large presence of fast chartists causes increased flickering of the traded price. Finally it is found that flash-crashes can occur in markets in which ratio of the number of chartists to the number of market makers is high, while at the same time the market makers are faster than the chartists. These results are interesting, as they show both benefits and dangers of having markets where fast computer algorithms trade side by side with human traders.

Contents

List of Figures

1 Correlation matrix of the four fitness measures in the first generation of dataset $\mathcal{D}_{10}$ . . . 7
2 Evolution of the four fitness measures in experiment $\mathcal{D}_{10}$ . . . 8
3 Evolution of the model parameters in experiment $\mathcal{D}_{10}$ . . . 10
4 Evolution of the four fitness measures in experiment $\mathcal{D}_{9}$ . . . 12
5 Evolution of the model parameters in experiment $\mathcal{D}_{10}$ . . . 13
6 Evolution of the four fitness measures in experiment $\mathcal{D}_{11}$ . . . 15
7 Evolution of the model parameters in experiment $\mathcal{D}_{11}$ . . . 16
8 Correlation between $N_m$and the four fitness measures in experiment $\mathcal{D}_{10}$ . . . 17
9 Correlation between $\lambda_{m,\mu}$ and fitness values (fixed N_c, variable N_m) . . . 19
10 Relation between N_m, $\lambda_{m,\mu}$, and the model fitness when the number of chartists was fixed to $N_c = 150 agents$. Due to missing data, some of the curves are not complete. . . . 20
11 Correlation between $\lambda_{m,\mu}$ and fitness values (fixed N_m, variable N_c) . . . 21
12 Relation between N_m, $\lambda_{m,\mu}$, and the model fitness when the number of market makers was fixed to $N_m = 52$ agents. . . . 22
13 Correlation between $N_c$and the four fitness measures when $N_m = 52$ (experiment $\mathcal{D}_{11}$) . . . 23
14 Correlation between chartist latency and fitness values (fixed N_c, variable N_m) . . . 25
15 Relation between N_m, $\lambda_{c,\mu}$, and the model fitness when the number of chartists was fixed to $N_c = 150 agents$. Due to missing data, some of the curves are not complete. . . . 26
16 Correlation between chartist latency and fitness values (fixed N_m, variable N_c) . . . 27
17 Relation between N_c, $\lambda_{c,\mu}$, and the model fitness when the number of market makers was fixed to $N_m = 52$ agents. . . . 28
18 Correlations between $\rho_{\mathrm{A}}$and the fitness values when $N_c = 150$ . . . 29
19 Correlations between $\rho_{\mathrm{A}}$and the fitness values when $N_m = 52$ . . . 30
20 Scatter plots of fitness measures in experiment $\mathcal{D}_{9}$. . . . 31
21 Scatter plot of f_{s}against $f_{\mathrm{t}}$with coloring showing $\log f_\sigma$ and f_{o} . . . 33
22 Scatter plot of $\log f_\sigma$ against $f_{\mathrm{t}}$with coloring showing f_{s} and f_{o} . . . 34
23 Jaccard index between the sets of data points extracted by each filter . . . 36
24 Scatter plots in fitness space showing the grouping of data points when using a GMM with 12 clusters. . . . 40
25 Two examples of market crashes . . . 45

List of Tables

1 Overview of datasets . . . 6
2 Average fitness values for the market with the top 10% highest and 10% lowest number of market makers . . . 18
3 Average fitness values for the market with the top 10% highest and 10% lowest number of market makers . . . 19
4 Filter IDs and fitness-regions . . . 37
5 XXX . . . 37
6 XXX . . . 38
7 XXX . . . 38
8 Cluster means ($\mathcal{D}_{10}$) . . . 41
9 Cluster means ($\mathcal{D}_{11}$) . . . 41
10 Cluster standard deviations ($\mathcal{D}_{11}$) . . . 42

Part I
Experiments

As mentioned in the previous chapter, the process of finding was an iterative one of running an experiment, analyzing the generated data, draw conclusions and then repeat the steps with a new experiments designed to amend the mistakes of the previous experiment. This chapter will go through the results that were obtained from each of the data sets. A summary and discussion of the results is found in chapter II.

1 Overview of experiments

The model has many parameters, and doing an exhaustive search over the entire parameter space is not possible. Instead, a genetic algorithm (GA) was used to do targeted searches of the parameters. For the details of the GA, please see chapter **??**. Depending on how the GA was set up, different areas of the search space was searched. Even when using a GA, some parameters had to be fixed. However, fixing parameters means that the effect of the fixed parameter on the behavior of the model remains unknown, since only a subspace of the entire parameter space is searched. For this reason the genetic algorithm was executed several times, each creating a data set containing fitness values for different parts of the parameter space. Each of these data set can be analyzed, providing information which can be corroborated in order to form an understanding of the overall market behavior.

Table 1 contains an overview of the different data sets, showing which parameters were fixed, and which were included as genes in the genetic algorithm.

In the following four experiments, the genetic algorithm was set to minimize all four fitness-measures.

$\mathcal{D}_3$: **Varying the number of HFT agents, and all latency related parameters** This data set was generated by including all the model parameters concerning time latency as well as the number of agents into the individuals in the genetic algorithm. Due to the high number of variables, the data turned out to be difficult to analyze, as too many factors pertaining to the simultaneous change of several parameters influenced the fitness values.

$\mathcal{D}_9$: **Fixing the number of agents while varying latency parameters** The analysis of $\mathcal{D}_3$ showed that when minimizing the four fitness-measures, the genetic algorithm tended to select model containing few or no HFT agents. The case of a market with no market makers and no chartists can safely be said to be trivial. Hence, in experiment $\mathcal{D}_9$, the number of HFT agents were fixed to $N_m = 30$ and $N_c = 100$.

$\mathcal{D}_{10}$: **Fixing the number of HFT chartists** Since $\mathcal{D}_9$kept $N_m$and $N_c$constant, the experiment did not reveal anything on how the market behavior changes when the number of agents changes. In order to investigate the impact of having many or few HFT market makers, $N_m$was varied in this experiment. Although it is also of

ID	Description	Fixed parameters	As genes
$\mathcal{D}_3$	All parameters varied	$V_{c,\mu} = 10$,$V_{c,\sigma} = 3$,$V_{m,\mu} = 50$,$V_{m,\sigma} = 20$,$\gamma_{c,\mu} = 2$,$\gamma_{c,\sigma} = 5$,$\alpha_{c,\mu} = 3$,$\alpha_{c,\sigma} = 2$	$\lambda_{c,\mu}$, $\lambda_{c,\sigma}$, N_c, $\tau_{c,\mu}$, $\tau_{c,\sigma}$, $H_{c,\mu}$, $H_{c,\sigma}$, $W_{c,\mu}$, $W_{c,\sigma}$, $\lambda_{m,\mu}$, $\lambda_{m,\sigma}$, N_m, $\tau_{m,\mu}$, $\tau_{m,\sigma}$
$\mathcal{D}_9$	Fixed number of HFT agents	$N_m = 30$, $N_c = 100$,$V_{c,\mu} = 10$,$V_{c,\sigma} = 3$,$V_{m,\mu} = 50$,$V_{m,\sigma} = 20$,$\gamma_{c,\mu} = 2$,$\gamma_{c,\sigma} = 5$,$\alpha_{c,\mu} = 3$, $\alpha_{c,\sigma} = 2$	$\tau_{c,\mu}$, $\tau_{c,\sigma}$, $H_{c,\mu}$, $H_{c,\sigma}$, $W_{c,\mu}$, $W_{c,\sigma}$, $\lambda_{m,\mu}$, $\lambda_{m,\sigma}$, N_m, $\tau_{m,\mu}$, $\tau_{m,\sigma}$
$\mathcal{D}_{10}$	Fixed number of HFT chartists and fixed strategy parameters	$N_c = 150$, $\tau_{m,\mu} = \tau_{c,\mu} = 50$, $\tau_{m,\sigma} = \tau_{c,\sigma} = 20$, $H_{c,\mu} = 5000$, $H_{c,\sigma} = 2000$, $W_{c,\mu} = 50$, $W_{c,\sigma} = 20$,$V_{c,\mu} = 10$,$V_{c,\sigma} = 3$,$V_{m,\mu} = 50$,$V_{m,\sigma} = 20$,$\gamma_{c,\mu} = 2$,$\gamma_{c,\sigma} = 5$, $\alpha_{c,\mu} = 3$,$\alpha_{c,\sigma} = 2$	N_m, $\lambda_{c,\mu}$, $\lambda_{c,\sigma}$, $\lambda_{m,\mu}$, $\lambda_{m,\sigma}$
$\mathcal{D}_{11}$	Fixed number of HFT market makers and fixed strategy parameters	$N_m = 52$, $\tau_{m,\mu} = \tau_{c,\mu} = 50$, $\tau_{m,\sigma} = \tau_{c,\sigma} = 20$, $H_{c,\mu} = 5000$, $H_{c,\sigma} = 2000$, $W_{c,\mu} = 50$, $W_{c,\sigma} = 20$,$V_{c,\mu} = 10$,$V_{c,\sigma} = 3$,$V_{m,\mu} = 50$,$V_{m,\sigma} = 20$,$\gamma_{c,\mu} = 2$,$\gamma_{c,\sigma} = 5$, $\alpha_{c,\mu} = 3$,$\alpha_{c,\sigma} = 2$	N_m, $\lambda_{c,\mu}$, $\lambda_{c,\sigma}$, $\lambda_{m,\mu}$, $\lambda_{m,\sigma}$

Table 1: Overview of datasets

interest how the market behavior depends on the number of HFT chartists, including $N_c$as a gene would yield results similar to those obtained in $\mathcal{D}_3$. For this reason the number of HFT chartists was fixed to $N_c = 150$.

$\mathcal{D}_{11}$: **Fixing the number of HFT market makers** This experiment was carried out in order to investigate the impact of the number of HFT chartists on the market behavior, and is supplementary to $\mathcal{D}_{10}$.

1.1 Correlation between fitness measures

A factor which influences the evolution of parameters is correlation between the fitness-measures. If two or more fitness measures have non-negative correlation coefficients, individuals will be statistically more likely to get good scores in the correlated fitness measures at the same time. Since all fitness measures are given equal weight in the selection process, individuals scoring well in the correlated fitness-measures will win over individuals which score well on another, statistically independent fitness measure. It is therefore important to compare the selection tendencies with the correlation between

Figure 1: Correlation matrix of the four fitness measures in the first generation of dataset $\mathcal{D}_{10}$

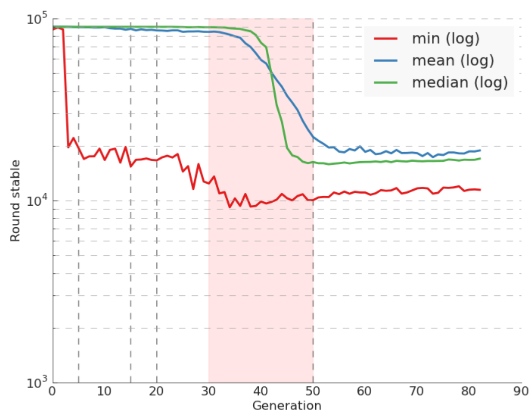
fitness-measures. Figure 1 shows a plot of the correlation matrix for $\mathcal{D}_{10}$. Since later generations will be affected by the biased selection and therefore contain more individuals which did well on the correlated fitness measures, the correlation coefficients in the figure were calculated over individuals in the first generation only.

For instance, the correlation between $f_{\mathrm{o}}$and f_{σ}means that an individual which scores a good f_{o}-fitness will be statistically likely to also score a good f_{σ}-fitness. Since all four fitness measures are weighed evenly in the selection, models with behavior which is assigned good values for $f_{\mathrm{o}}$and f_{σ}will score a better overall fitness than a simulation with a good fitness

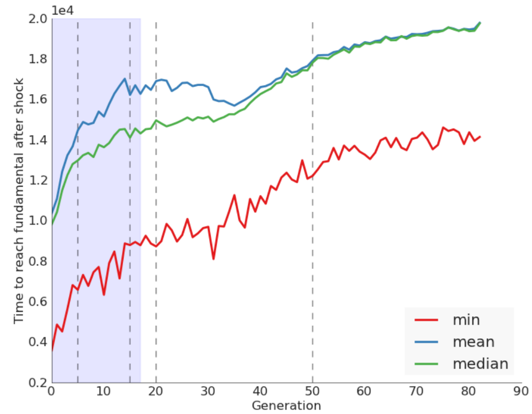
In other words, the correlation between $f_{\sigma}$and f_{o}means that stable individuals will outlive fast individuals as they are selected for breeding more often. This is not a property of a model itself, but rather a problem with the definition of the fitness measures. This problem can be circumvented by not using of the correlated fitness values. XXX

Both $f_{\mathrm{o}}$and f_{σ}were used for the GA selection, and although these two fitness measures do reflect different properties of the simulations, they were found to be somewhat correlated. That is, a simulation which tends to have a small overshoot also tends to have stable traded prices.

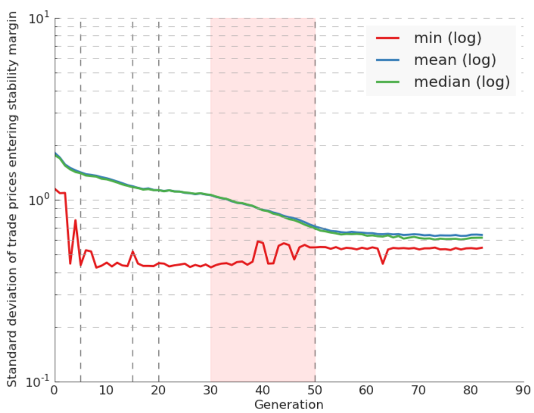
(simulations with a small overshoot also tend to have more stable trade prices), and there work together towards selecting the same type of simulations. $f_{\mathrm{t}}$and f_{s}both The same is not the case for $f_{\mathrm{t}}$and f_{s}, as it possible that a simulation responds quickly to the shock, but does not stay within the stability margin.



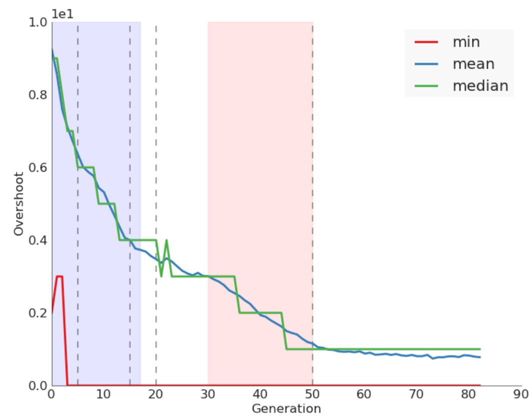
(a) Evolution of f_s



(b) Evolution of f_t



(c) Evolution of f_σ



(d) Evolution of f_o

Figure 2: Evolution of the four fitness measures in experiment $\mathcal{D}_{10}$

2 Fitness and parameter evolution

2.1 Variable number of market makers

Figure 2 shows the evolution of the four fitness measures. The population wide mean is plotted along the median and minimum statistics. Since all four fitness measures were minimized, the curve for the minimum value shows the best individual alive during each generation, with respect to each fitness measure. While the mean reflects how the overall population is evolving, the median is useful as it gives an insight into how skewed the population wide distribution of parameters is.

Model stability Figure 2a: shows that the GA quickly manages to find some parameters which cause the simulation to stabilize quickly. However, these individuals do not manage to dominate the population evident by the mean and median curves remaining almost the same until generation 30 or so. In the next 20 generations the population undergoes a rapid change, as the population wide average of $f_s$drop from close to 10^5 to around $2 \cdot 10^4$ rounds on average. The disparity between the mean and the median indicates that the population undergoes a rapid change in the same period, from mostly containing unstable individuals to mostly containing stable individuals. In generation 42, the median curve crosses the mean curve, which means that the the population contain as many stable simulations as it contain unstable simulations. From that point on the unstable simulations are quickly replaced by stable individuals.

Price fluctuations and overshoot During the same period, the population average $f_\sigma$also decreases fairly rapidly, but the drop is less pronounced than the drop in f_σ. As figures 3a and 3b show, the number of market makers rapidly increased during this period, as did the average latency of the market makers. Since the mean and median are close in both figure[1], the mean is representative of the evolution of the entire population.

Responsiveness $f_t$measures the time it takes for the model to react to the shock in the fundamental, and the evolution of the population wide statistics is shown on figure 2b. Although the GA is instructed to minimize $f_t$in order to look for more faster models, it clearly fails to do this. Indeed, the most responsive simulation took only about 4000 rounds to reach the new fundamental, but this individual died out in favor of slower individuals. In the last generation the most responsive simulation took around 14000 rounds to reach the new fundamental. The reasons for this failure to locate responsive models is discussed in section 1.1. In the A large change of the average of $f_t$happens in the rounds five to 15. In this period, the median is lower than the mean, which means that the growth in the mean can be attributed to a minority of individuals.

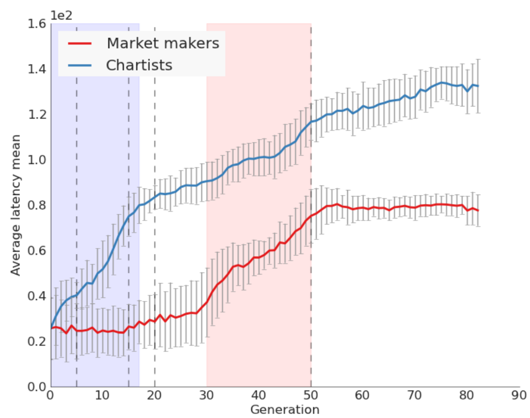
On figures 2 and 3, the two areas shaded in a light blue and light red respectively show the two periods during which there was a drastic change in parameters and fitness-values. By comparing the time at which parameters and fitness-values change, it is possible to get an idea of how parameters influence the fitness-values. To that end, figure 3 shows the evolution of each of the parameters that were varied by the GA[2].

The two periods indicated by the shaded squares seem to reflect some sudden changes in the parameters.

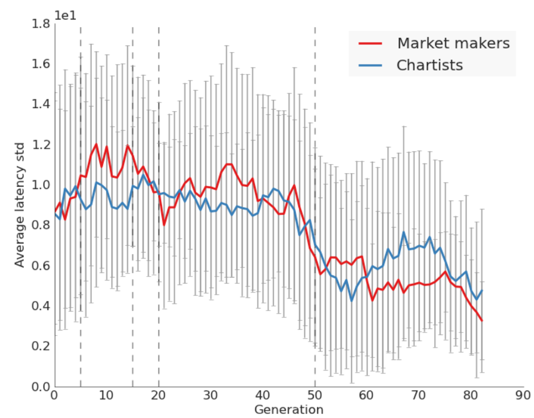
Average agent latency As is shown on figure 3a, individuals containing large latency parameters are selected for both HFT market makers and HFT chartists. $\mathrm{E}_{\mathcal{P}}[\lambda_{c,\mu}]$ grows grows quickly during the first 20 rounds (blue shade). Referring back to figure 2, it is seen that $\mathrm{E}_{\mathcal{P}}[f_t]$ and $\mathrm{E}_{\mathcal{P}}[f_o]$ grows and shrinks respectively. As for $\mathrm{E}_{\mathcal{P}}[\lambda_{m,\mu}]$, it grows from rounds 20 through 50 (red shade), and this seems to be strongly reflected in the growth of $\mathrm{E}_{\mathcal{P}}[f_s]$, and to a lesser degree a decline in $\mathrm{E}_{\mathcal{P}}[f_o]$

[1]Since $f_o$is discrete, the median and min statistics are also discrete

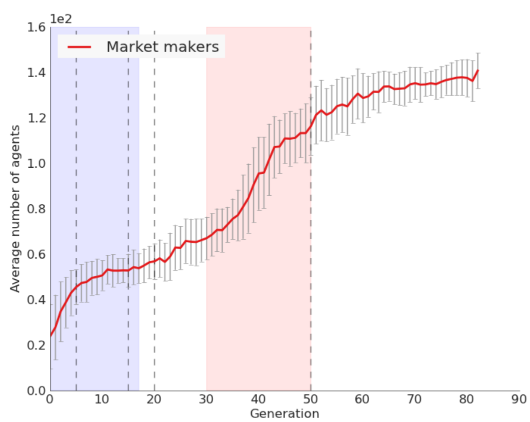
[2]Since the median was found to follow the mean nicely for all the parameters, the medians are not displayed. Also, the gray error bars show the population wide variance



(a) Evolution of $\lambda_{m,\mu}$ and $\lambda_{c,\mu}$



(b) Evolution of $\lambda_{m,\sigma}$ and $\lambda_{c,\sigma}$



(c) Evolution of N_m

Figure 3: Evolution of the model parameters in experiment $\mathcal{D}_{10}$

and $\mathrm{E}_{\mathcal{P}}[f_\sigma]$. Furthermore, the small size of the error bars on both curves show that the population consistently moves towards containing more individuals with larger latency parameters for both HFT agent types. While initially $\mathrm{E}_{\mathcal{P}}[\lambda_{m,\mu}] \approx \mathrm{E}_{\mathcal{P}}[\lambda_{c,\mu}]$, the population wide mean $\mathrm{E}_{\mathcal{P}}[\lambda_{c,\mu}]$ ends up being roughly 1.5 times larger than $\mathrm{E}_{\mathcal{P}}[\lambda_{c,\mu}]$. Finally, note also that the growth of $\mathrm{E}_{\mathcal{P}}[\lambda_{c,\mu}]$ and $\mathrm{E}_{\mathcal{P}}[\lambda_{m,\mu}]$ seem to be somewhat independent, as they sometimes grow together, sometimes not.

Number of market makers The number of market makers increases almost every generation, but grows especially quickly through rounds 20 to 50 (red shade)

Agent latency variance Figure 3b: The trends for $\mathrm{E}_{\mathcal{P}}[\lambda_{c,\sigma}]$ and $\mathrm{E}_{\mathcal{P}}[\lambda_{m,\sigma}]$ are less clear, as the population-wide variances $\mathrm{Var}_{\mathcal{P}}[\lambda_{c,\mu}]$ and $\lambda_{m,\mu}$ illustrated by the large error bars are high compared to the change in $\mathrm{E}_{\mathcal{P}}[\lambda_{c,\mu}]$ and $\mathrm{E}_{\mathcal{P}}[\lambda_{m,\mu}]$. While this could mean that the simulation behaves more nicely when the difference between the latency parameters of the trading agents is smaller, further experiments would have to be carried out to confirm this fact. XXX

In summary, the genetic algorithm prefers simulations with many, but relatively slow market makers. Apparently simulations with slow chartists also outperformed those with fast chartists, but since the number of HFT chartists was fixed at $N_c =$, this experiment does not reveal how the simulation would perform with more (or less) chartists. It is possible to imagine that the market would perform just as well with a few and fast chartists. Section **??** contains the analysis of an experiment in which the number of chartists were varied. The discussion above can be summarized as follows:

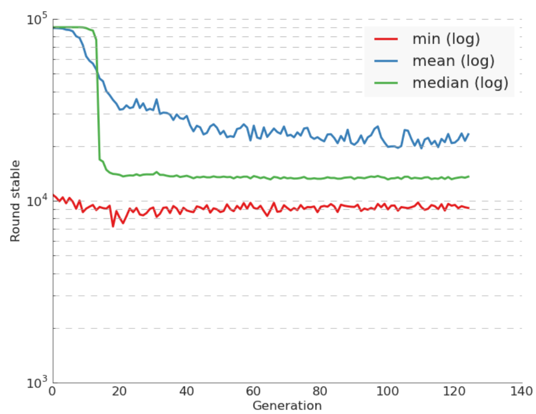
1. The responsiveness of the market is influenced by latency of the chartists. Slower chartists made the market require more time to respond to the fundamental shock.

2. The time it takes for the market to become in influenced by the number of market makers and on the latency of the market makers. More but slower market makers seems to make the market settle within the stability margin faster.

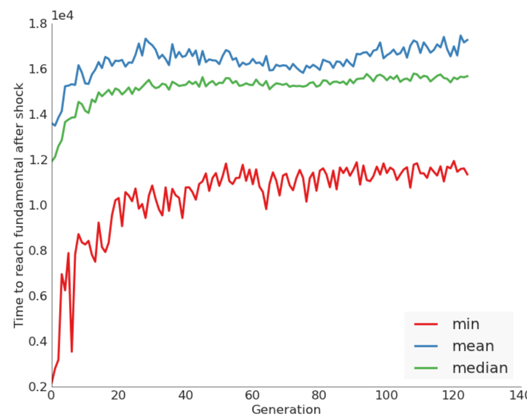
3. The overshoot, as well as the average size of the price fluctuations of the market, are both influenced by the latency of both agent types, as well as the number of market makers.

4. The market was more stable but reacted slowly when the chartists were slower than the market makers.

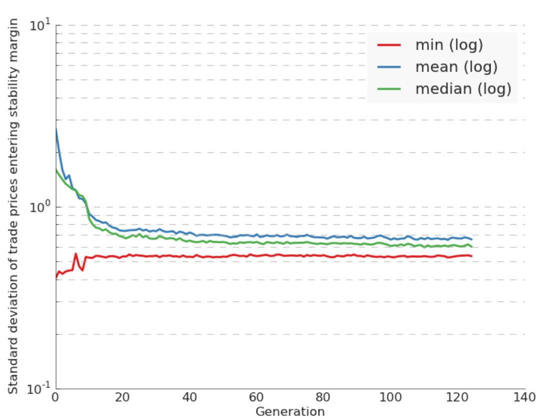
The accuracy of the above analysis is limited as it only looks at population wide statistics at a given point in the duration of the GA. The following section contain an analysis in which the generation to which each data point belongs is considered irrelevant. The analysis will try to confirm each of the four statements above.



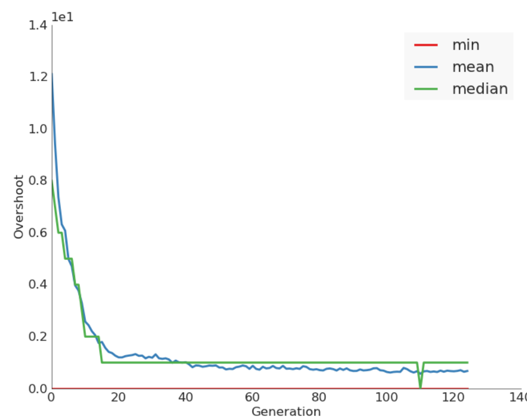
(a) Evolution of f_s



(b) Evolution of f_t

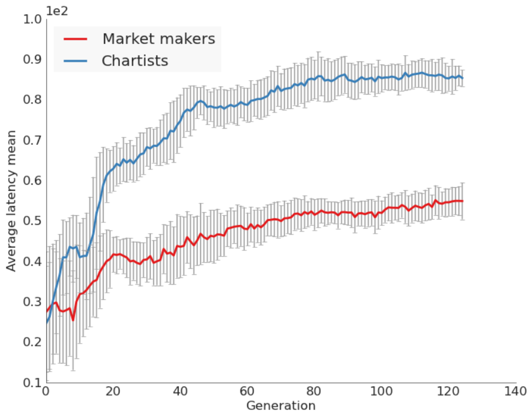


(c) Evolution of f_σ

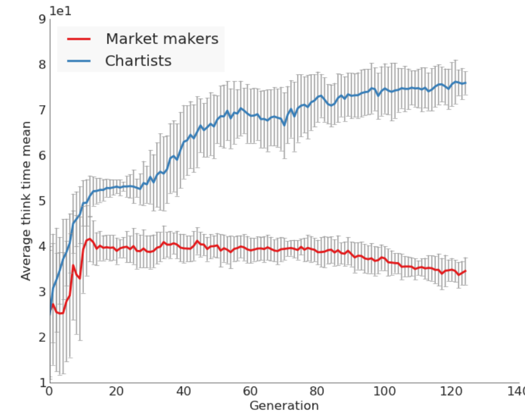


(d) Evolution of f_o

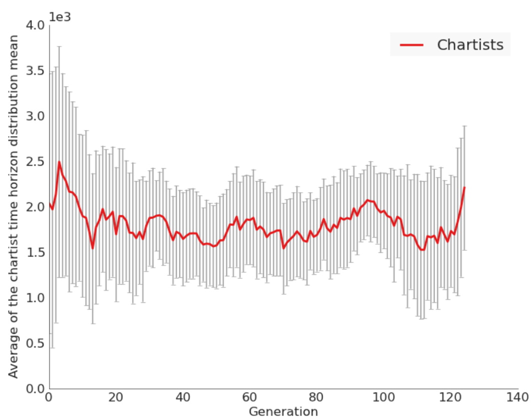
Figure 4: Evolution of the four fitness measures in experiment $\mathcal{D}_9$



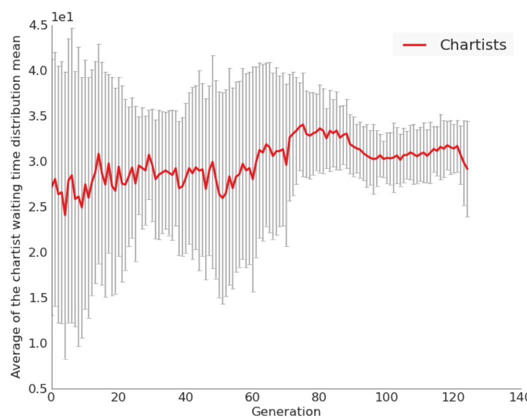
(a) Evolution of $\lambda_{m,\mu}$ and $\lambda_{c,\mu}$



(b) Evolution of $\lambda_{m,\sigma}$ and $\lambda_{c,\sigma}$



(c) Evolution of N_m



(d) Evolution of N_m

Figure 5: Evolution of the model parameters in experiment $\mathcal{D}_{10}$

2.2 Fixed number of chartists and market makers

When the GA cannot change the number of chartists and market makers, it has to find better fitness values by selecting the right latency parameters. As shown on figure 4, the GA managed to find models with little or no overshoot, non-flickering prices, and which become stable. The price of having these nice qualities seems to be a slower response time to the shock. The GA find these well-behaving models by selecting latency parameters such that the chartists are slower than the market makers. $\mathrm{E}_{\mathcal{P}}[H_{c,\mu}]$ and $\mathrm{E}_{\mathcal{P}}[W_{c,\mu}]$ change little over the are more or less unchanged, which seems to indicate that they have little effect of the fitness values, at least compared to other time related parameters such as $\lambda_{c,\mu}$, $\lambda_{m,\mu}$, $\tau_{c,\mu}$and $\tau_{m,\mu}$. This can either mean that these parameters does

2.3 Variable number of chartists

The evolution of the fitness values in $\mathcal{D}_{11}$, shown in figure 6 fluctuate significantly more than in $\mathcal{D}_9$ and $\mathcal{D}_{10}$. In contrary two the two other experiments, the GA here manages to decrease f_{t}, and $\mathrm{E}_{\mathcal{P}}[f_{\mathrm{t}}]$ drops almost 10000 rounds around generation 200. At the same time $\mathrm{E}_{\mathcal{P}}[f_{\mathrm{o}}]$, $\mathrm{E}_{\mathcal{P}}[f_{\sigma}]$ and $\mathrm{E}_{\mathcal{P}}[f_{\mathrm{s}}]$ all rise, again indicating the there exists a trade-off between speed and stability in the model. At the point in the evolution where $\mathrm{E}_{\mathcal{P}}[f_{\mathrm{t}}]$ drops, several interesting things happen with the model parameters that live in the population. First of all the number of chartists increase dramatically, again pointing towards more chartists making the markets fast and unstable. Secondly, the market maker latency also drops to around the same level as the chartist latency. This is interesting because it could mean that the faster market makers help drive the market towards a larger overshoot.

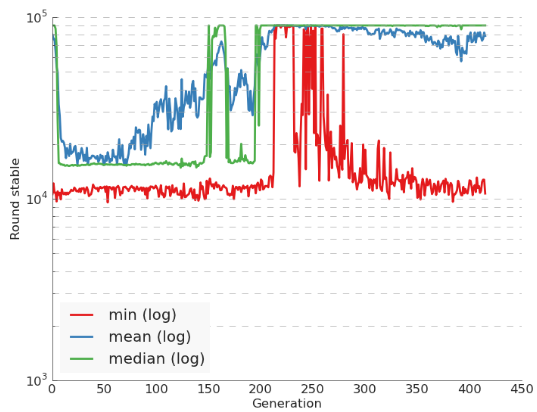
Market makers become slower and the chartists become faster. At the same time, the number of chartists rise rapidly

- A high number of market makers enable the market to respond quickly to the shock, but also cause the traded price to flicker more, and for the model to have a larger overshoot.

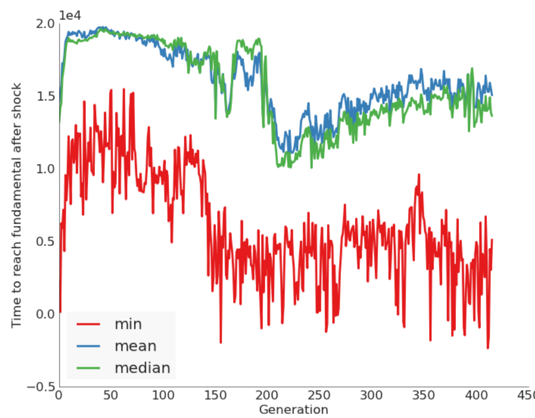
- Slower market makers also cause the market to respond faster to the shock

3 Population-wide parameter/fitness correlations

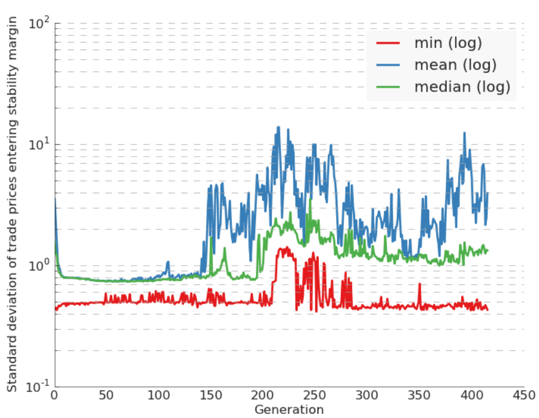
This section contains several figures which illustrate how the model fitness varies with the model parameters. In the figures showing the data from experiment $\mathcal{D}_{11}$, simulations with $f_{\mathrm{o}} > 10$ are removed in order to make the figures easier to interpret. In the following, the notation $\mathrm{E}_{\mathrm{p}}[\cdot]$ is used for the population wide average of a model parameter of fitness measure. For instance, $\mathrm{E}_{\mathrm{p}}[f_{\mathrm{o}}]$ is the average market overshoot, where the average is calculated over the total population of individuals that ever lived in the genetic algorithm.



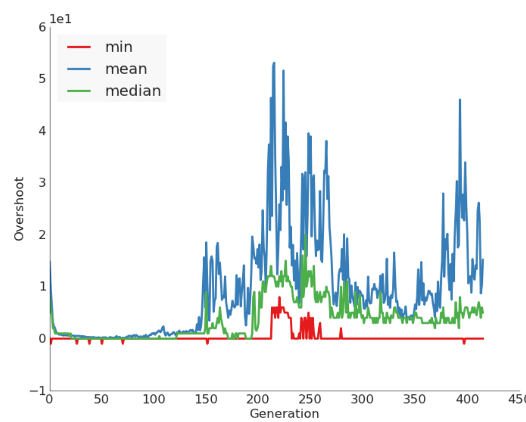
(a) Evolution of f_s



(b) Evolution of f_t

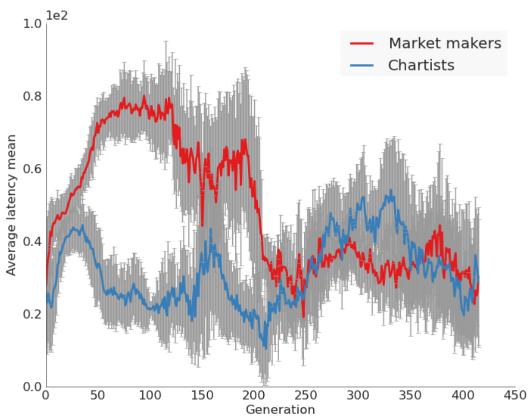


(c) Evolution of f_σ

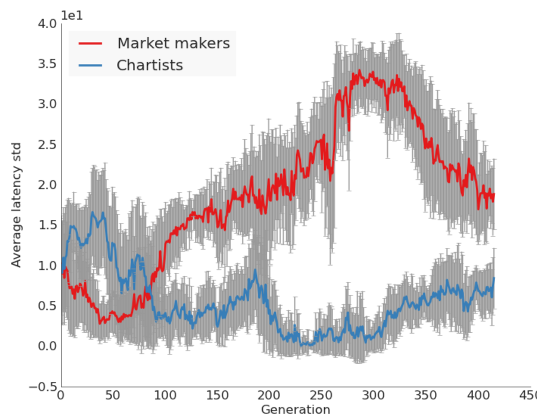


(d) Evolution of f_o

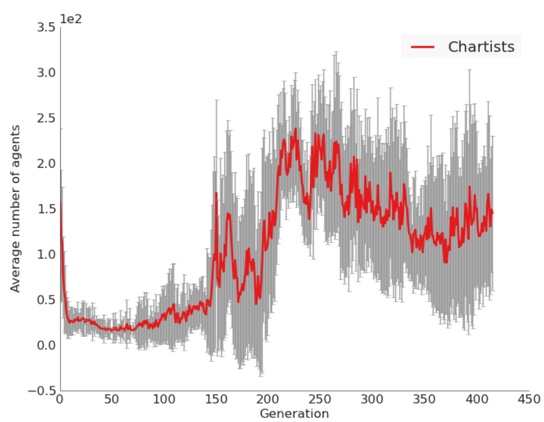
Figure 6: Evolution of the four fitness measures in experiment $\mathcal{D}_{11}$



(a) Evolution of $\lambda_{m,\mu}$ and $\lambda_{c,\mu}$

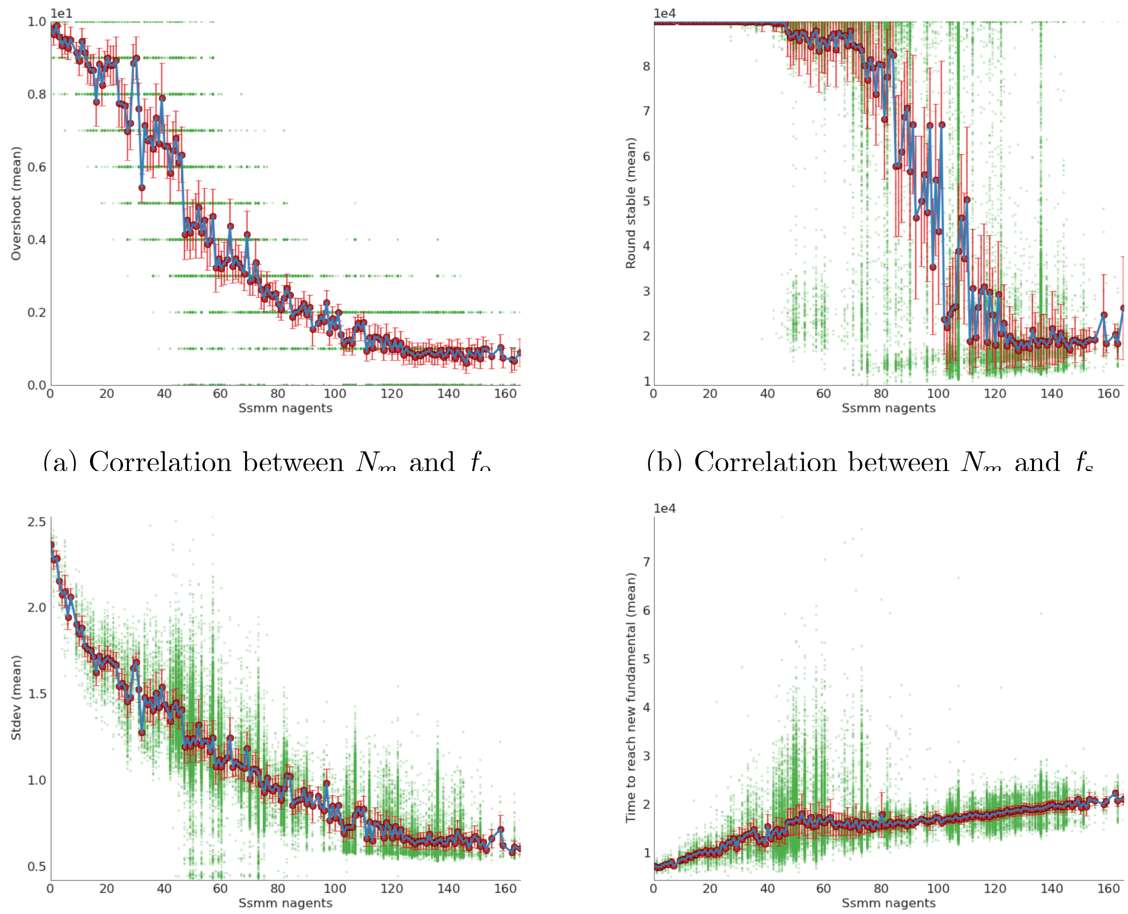


(b) Evolution of $\lambda_{m,\sigma}$ and $\lambda_{c,\sigma}$



(c) Evolution of N_c

Figure 7: Evolution of the model parameters in experiment $\mathcal{D}_{11}$



(a) Correlation between N_m and f_o

(b) Correlation between N_m and f_s

(c) Correlation between N_m and f_σ

(d) Correlation between N_m and f_t

Figure 8: Correlation between $N_m$and the four fitness measures in experiment $\mathcal{D}_{10}$

3.1 Number of market makers

The number of market makers was kept fixed in experiments $\mathcal{D}_9$ and $\mathcal{D}_{11}$, but was varied in experiment $\mathcal{D}_{10}$. Figure 8 shows how the number of market makers correlates with the model fitness in experiment $\mathcal{D}_{10}$. Evidently a large number of market makers reduces the overshoot, whereas the market virtually always have an overshoot when there are few or no market makers. The same is true for the trade prices flicker: few market makers always means flickering prices. A higher number of market makers cause the market to be less responsive, while fewer market makers have the opposite effect. Finally, the number of market makers also influences how quickly the market settles within the stability margin, as markets with more market makers become stable faster than markets with few agents. Table 2 shows the average fitness values for models where the number of market maker agents was respectively below and above the first (q_1) and ninth (q_9) 10-quantiles in the dataset. The two quantiles were at $q_1 = 46$ and $q_9 = 136$. It is seen that the market containing a large number of market makers (more than 136) had a much

$\mathcal{D}_{10}$	$N_m < q_1$	$N_m > q_9$
f_o	7.5	0.8
f_s	89876.2	18546.9
f_σ	1.6	0.6
f_t	12590.5	19832.1

Table 2: Average fitness values for the market with the top 10% highest and 10% lowest number of market makers

smaller overshoot, less flickering prices, a slower response time, and stayed within the stability margin faster, compared to the markets with a low (less than 46) number of market makers.

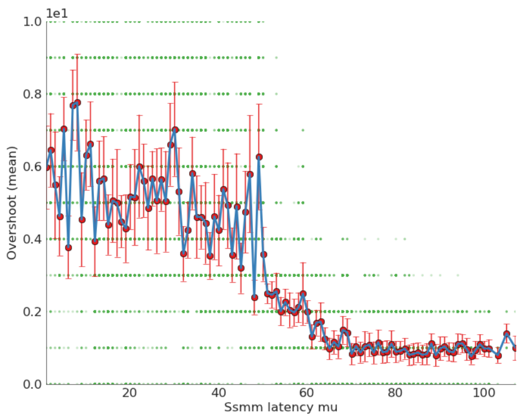
3.2 Market maker latency

The parameters controlling the market maker latency was varied in experiments $\mathcal{D}_9$, $\mathcal{D}_{10}$ and $\mathcal{D}_{11}$. However, since the data from $\mathcal{D}_9$ turned out to be too noise due to the large number of parameters included in the search, only data from $\mathcal{D}_{10}$ and $\mathcal{D}_{11}$ was used.

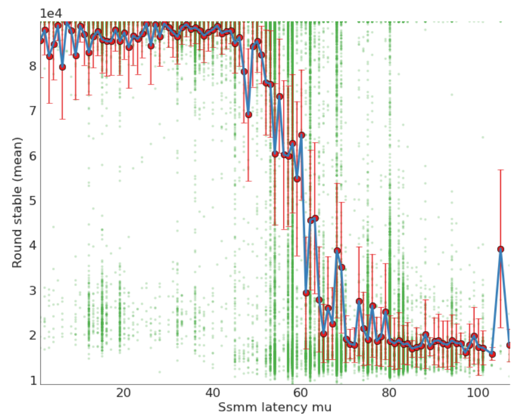
Fixing the number of chartists

In experiment $\mathcal{D}_{10}$, the number of chartists was kept fixed at $N_c = 150$, while $N_m$was varied by the GA. In this case, the $\lambda_{m,\mu}$is found to be somewhat correlated with the fitness measures as illustrated on figure 9. Especially for $\lambda_{m,\mu} > 50$, the data seems to be consistent, as the error bars showing the standard deviation of the data are small in this region. However, for $\lambda_{m,\mu} < 50$, the model behavior is no longer predictable by using $\lambda_{m,\mu}$alone. Figure 10 shows line plots of the four fitness measures plotted against $\lambda_{m,\mu}$. Each line shows the average fitness of markets in which the number of market makers is in a limited range as shown in the legend. The figure shows that even though the market maker latency does influence the market, the effect is secondary to that of the number of agents. For instance, when the market contains less than 25 market makers, all four fitness measures are more or less unchanged, as is evident by the nearly flat red curves. As the number of market makers grow, so does the importance of how fast they are. The average overshoot and the average time to catch up to the new fundamental only change slightly, even with over 100 market makers (yellow line). On the other hand, the average price flickering and the average number of rounds it takes for the market to settle within the stability margin both change significantly with the market maker speed for $N_m > 50$. In summary, figure 10 shows that in a market with only a few market makers, these agents have little influence no matter how fast they are. As the number of market makers grow, so does the collective force of all the market makers, and so does the importance of how slow or fast these agents are. Although the influence of $\lambda_{m,\mu}$ depends on N_m, . q_1 and q_1

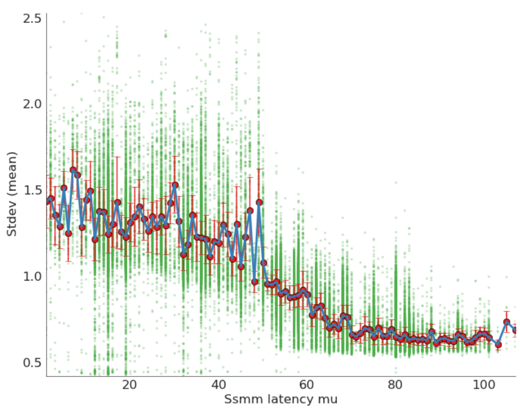
The next section will examine how the market behaves with respect to how many



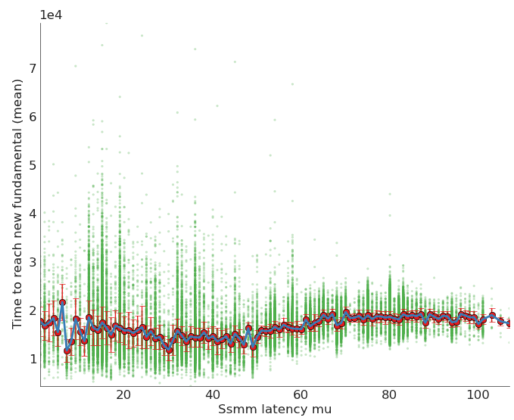
(a) Correlation between $\lambda_{m,\mu}$and f_{o}



(b) Correlation between $\lambda_{m,\mu}$and f_{s}



(c) Correlation between $\lambda_{m,\mu}$and f_σ



(d) Correlation between $\lambda_{m,\mu}$and f_{t}

Figure 9: Correlation between $\lambda_{m,\mu}$ and fitness values (fixed N_c, variable N_m)

$\mathcal{D}_{10}$	$\lambda_{m,\mu} < q_1$	$\lambda_{m,\mu} > q_9$
f_{o}	5.6	0.9
f_{s}	85927.9	18329.3
f_σ	1.4	0.6
f_{t}	16649.4	18795.3

$\mathcal{D}_{11}$	$\lambda_{m,\mu} < q_1$	$\lambda_{m,\mu} > q_9$
f_{o}	14.2	1.7
f_{s}	81379.2	26153.0
f_σ	3.6	1.0
f_{t}	15856.2	18107.4

Table 3: Average fitness values for the market with the top 10% highest and 10% lowest number of market makers

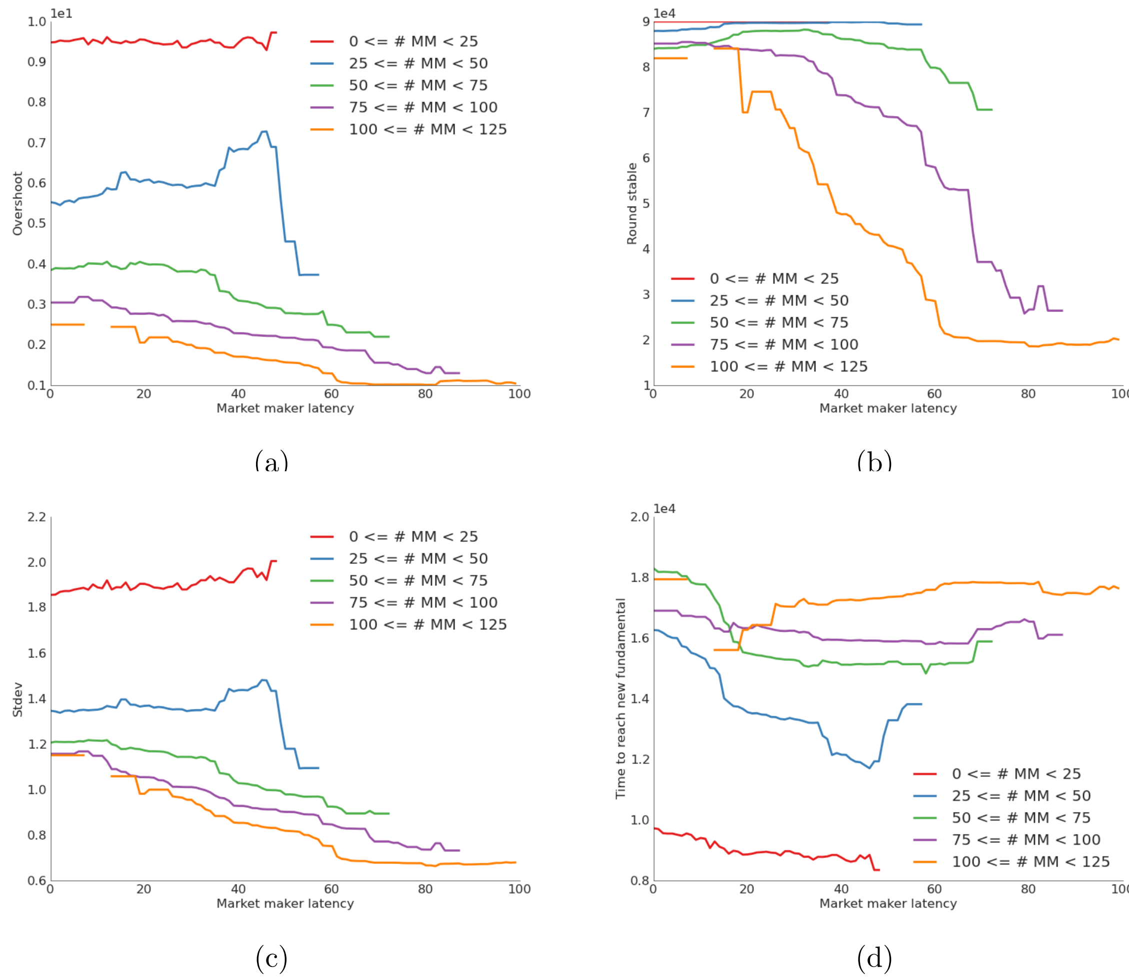


Figure 10: Relation between N_m, $\lambda_{m,\mu}$, and the model fitness when the number of chartists was fixed to $N_c = 150 agents$. Due to missing data, some of the curves are not complete.

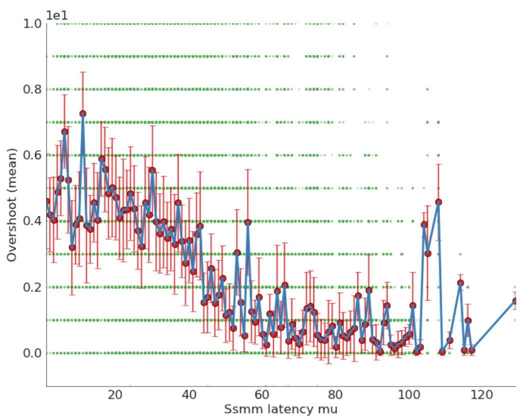
chartists are active in the market, and with respect to the latency of the chartists.

3.2.1 Fixing the number of market makers

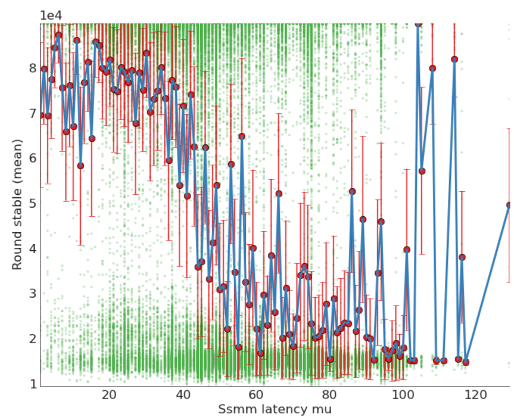
XXXNOT FINISHEDXXX In experiment $\mathcal{D}_{11}$, the number of chartists was varied, while the number of market makers were kept at a constant $N_m = 52$ agents. While it was fairly obvious that the market would not be impacted by changing $\lambda_{m,\mu}$when only a few market makers were active, it is less obvious that the same is true for the number of chartists. Yet figure

3.3 Number of chartists

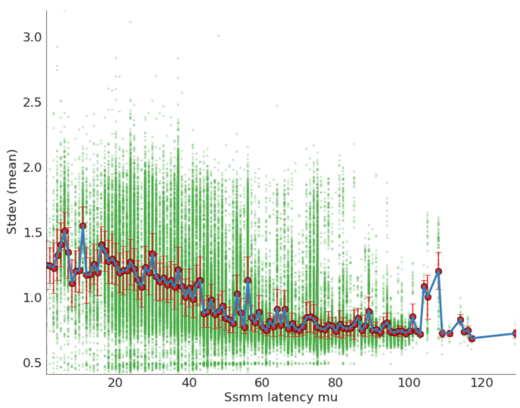
Figure 13 shows the average population wide number of agents $\mathrm{E}_{\mathcal{P}}[N_c]$ plotted against each of the four fitness measures, and the figures are summarized below.



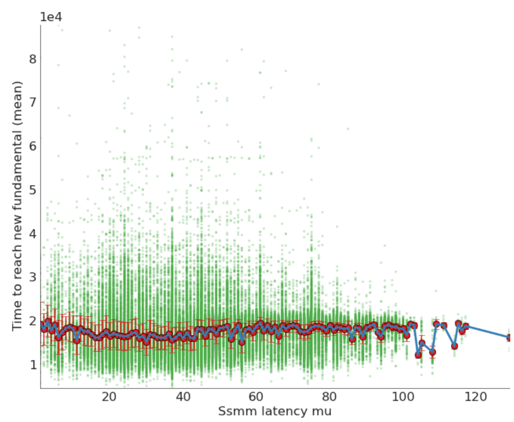
(a) Correlation between $\lambda_{m,\mu}$and f_o



(b) Correlation between $\lambda_{m,\mu}$and f_s



(c) Correlation between $\lambda_{m,\mu}$and f_σ



(d) Correlation between $\lambda_{m,\mu}$and f_{t}

Figure 11: Correlation between $\lambda_{m,\mu}$ and fitness values (fixed N_m, variable N_c)

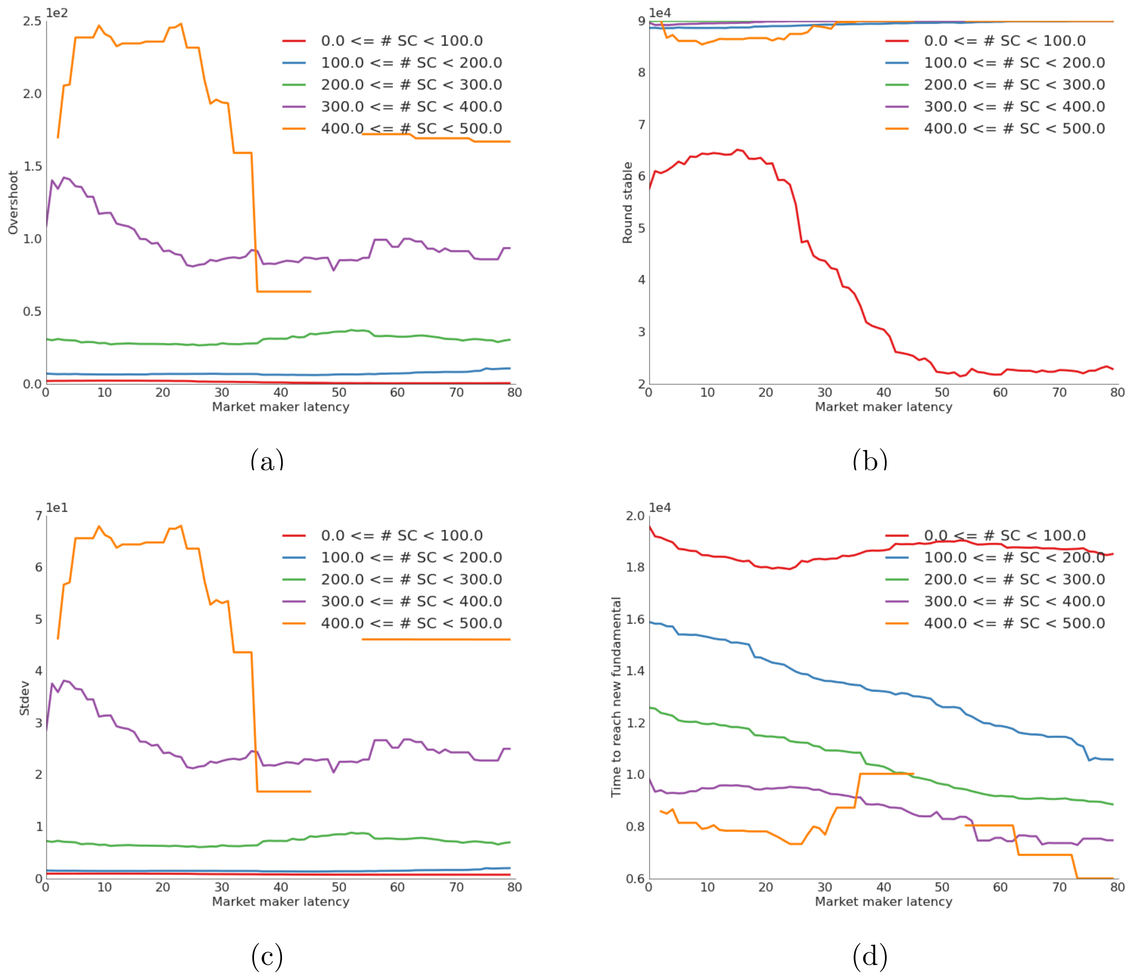
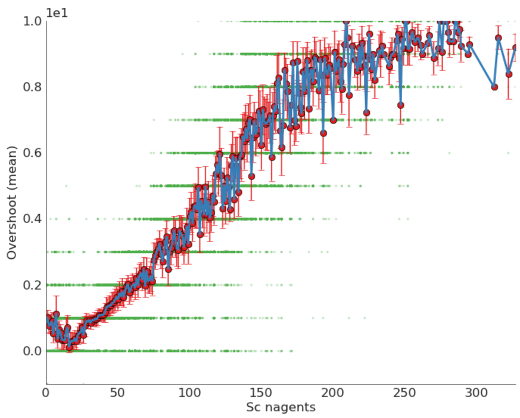
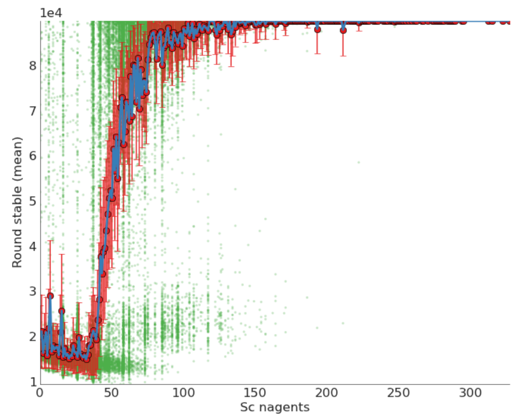


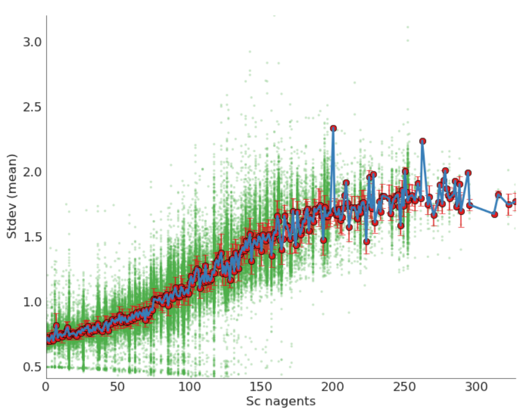
Figure 12: Relation between N_m, $\lambda_{m,\mu}$, and the model fitness when the number of market makers was fixed to $N_m = 52$ agents.



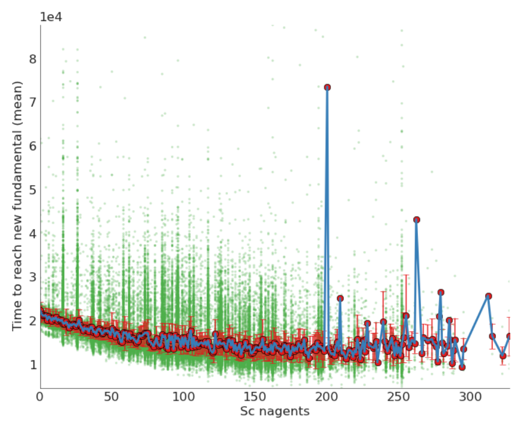
(a) Correlation between $N_c$and f_o



(b) Correlation between $N_c$and f_s



(c) Correlation between $N_c$and f_σ



(d) Correlation between $N_c$and f_t

Figure 13: Correlation between $N_c$and the four fitness measures when $N_m = 52$ (experiment $\mathcal{D}_{11}$)

- The more chartists a market has, the faster it responds to the fundamental. This is especially true when comparing markets with less than 100 chartists, and less pronounced when comparing markets with over 100 chartists.
- The model overshoot is also correlated with the number of chartists in such a way that markets with more chartists have a larger overshoot on average. Whereas the market only seemed to benefit from a decreased response time when the number of chartists were kept below 100, the overshoot continues to grow steadily larger even as the number of chartists is increased beyond 100 agents.
- $f_\sigma$is correlated with the number of chartists in the same way as f_o, such that more chartists make the traded prices flicker more.
- Finally, the graph for $f_\mathrm{s}$show that the market rarely becomes stable when it contains more than 50 chartists or so.

The large errorbars around the points with a large value of $N_c$is caused by data sparsity in this region. The GA was set to search for stable markets, and since markets with a large number of chartists tend to be unstable, such markets were rarely selected for creating offspring.

3.4 Chartist latency

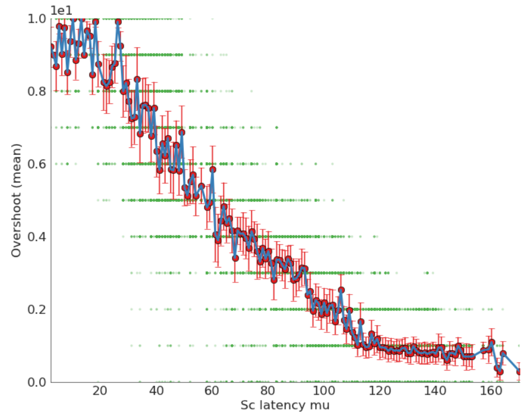
Fixed number of chartists

Figure 14a shows that $\lambda_{c,\mu}$is negatively correlated with f_o, such that markets with faster chartists are more likely to have a larger overshoot. Next, figure 14 shows that $\lambda_{c,\mu}$is negatively correlated with f_σ, such that markets with faster chartists are more likely to have flickering trade prices.

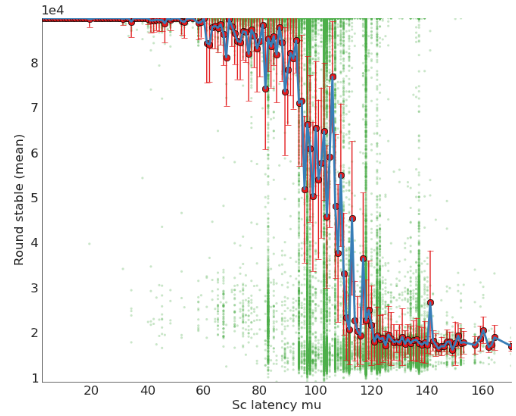
As for the market responsiveness, it is seen that $\lambda_{c,\mu}$is positively correlated with f_t, such that markets with faster agents is more likely to have a shorter response time to the market. Figure 14d confirms that markets with fast chartists did actually manage to reach the new fundamental price faster than those markets having slow chartists. The average response time of markets in which the chartists had a latency of less than 30 rounds was around 18000 rounds, whereas it was around 25000 rounds with chartists with more than 100 rounds of latency. The market response time is most sensitive in the range $20 < \lambda_{c,\mu} < 60$, and does not change much for larger latencies.

The plots of f_o, $f_\sigma$and $f_\mathrm{t}$show that predicting the three fitness measures in markets with slow chartists would be more accurate than for markets with fast chartists, as the correlation of f_o, $f_\sigma$and $f_\mathrm{t}$with $\lambda_{c,\mu}$is stronger for larger values of $\lambda_{c,\mu}$.

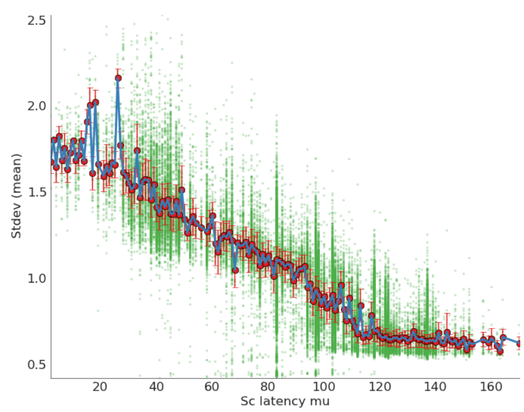
Figure 14b shows that $\lambda_{c,\mu}$is positively correlated with f_s, but also that the relationship between $\lambda_{c,\mu}$and $f_\mathrm{s}$seems highly non-linear. The figure illustrates the binary nature of the stability criteria, that is, that a simulation is either stable or not stable. This causes $f_\mathrm{s}$to have a high conditional variance of $f_\mathrm{s}$given $\lambda_{c,\mu}$in the region $50 < \lambda_{c,\mu} < 120$, meaning that that prediction of $f_\mathrm{s}$from $\lambda_{c,\mu}$in this region would not be very accurate. What this means is that the stability of a simulation is highly dependent on factors other than $\lambda_{c,\mu}$,



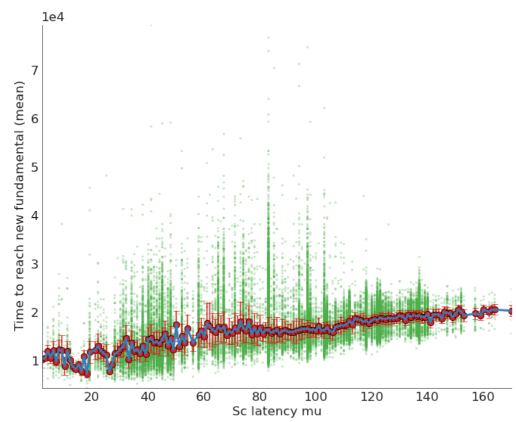
(a) Correlation between $\lambda_{c,\mu}$and f_{o}



(b) Correlation between $\lambda_{c,\mu}$and f_{s}



(c) Correlation between $\lambda_{c,\mu}$and f_σ



(d) Correlation between $\lambda_{c,\mu}$and f_{t}

Figure 14: Correlation between chartist latency and fitness values (fixed N_c, variable N_m)

when the parameter is within 50 to 120 rounds. When the chartists are faster than 50 rounds, the market is almost always unstable, and when the chartists are slower than 120 rounds the market is almost always stable.

Fixed number of market makers

Figures 16 seems to indicate that no correlations exist between the speed of the chartist agents, and the model fitness measures. However, since figure 14 does point towards the existence of such correlations, something else must be obscuring the scatter plots in 16. The reason is found to be that the number of chartists was not kept constant in experiment $\mathcal{D}_{11}$. It turns out that $\lambda_{c,\mu}$ is in fact correlated with f_{o}, $f_\sigma$and f_{t}, but that the correlation depends heavily on the number of chartists in the market.

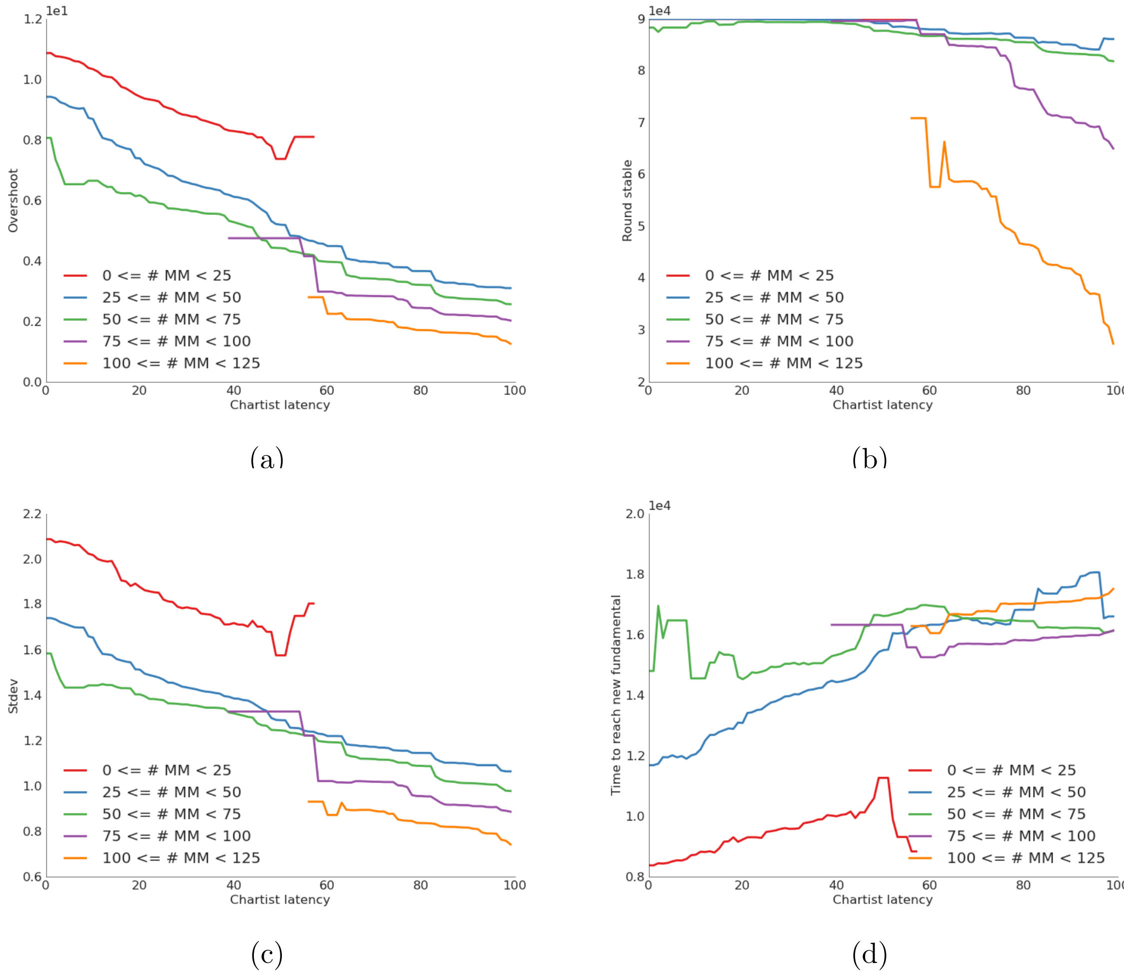
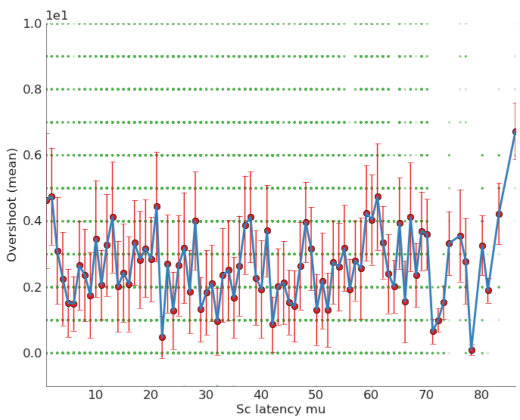
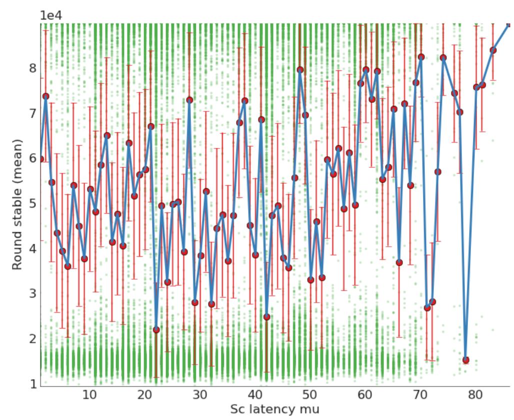


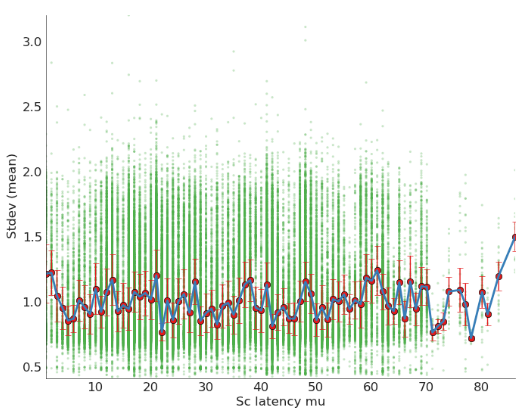
Figure 15: Relation between N_m, $\lambda_{c,\mu}$, and the model fitness when the number of chartists was fixed to $N_c = 150agents$. Due to missing data, some of the curves are not complete.



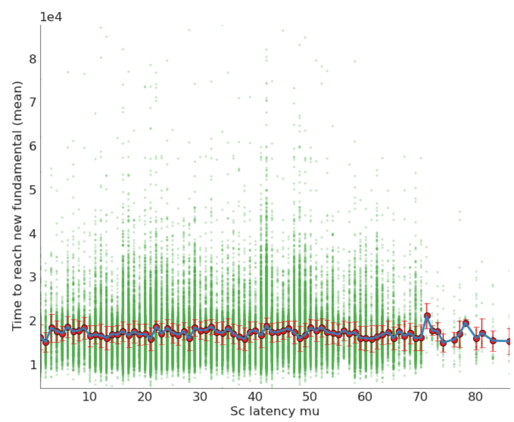
(a) Correlation between $\lambda_{c,\mu}$and f_o



(b) Correlation between $\lambda_{c,\mu}$and f_s



(c) Correlation between $\lambda_{c,\mu}$and f_σ



(d) Correlation between $\lambda_{c,\mu}$and f_{t}

Figure 16: Correlation between chartist latency and fitness values (fixed N_m, variable N_c)

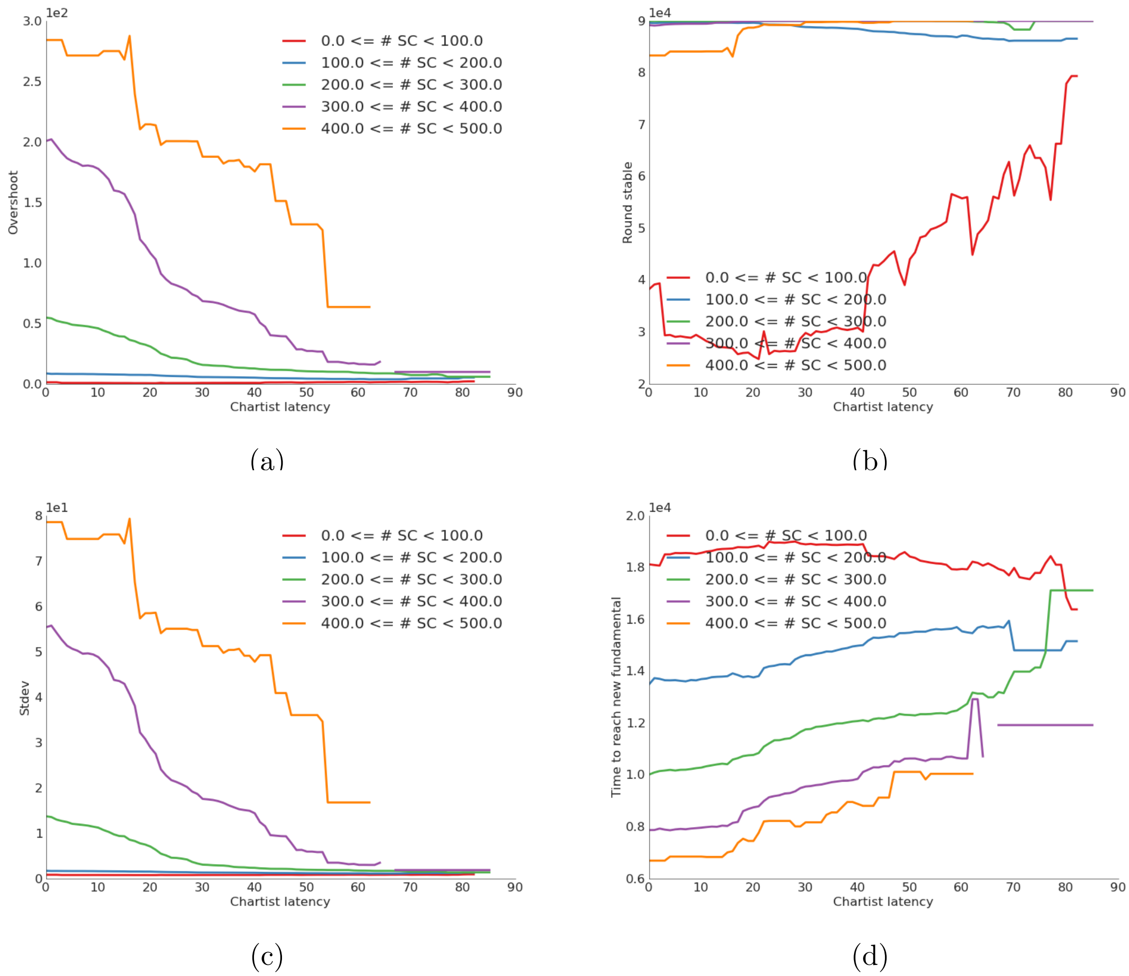
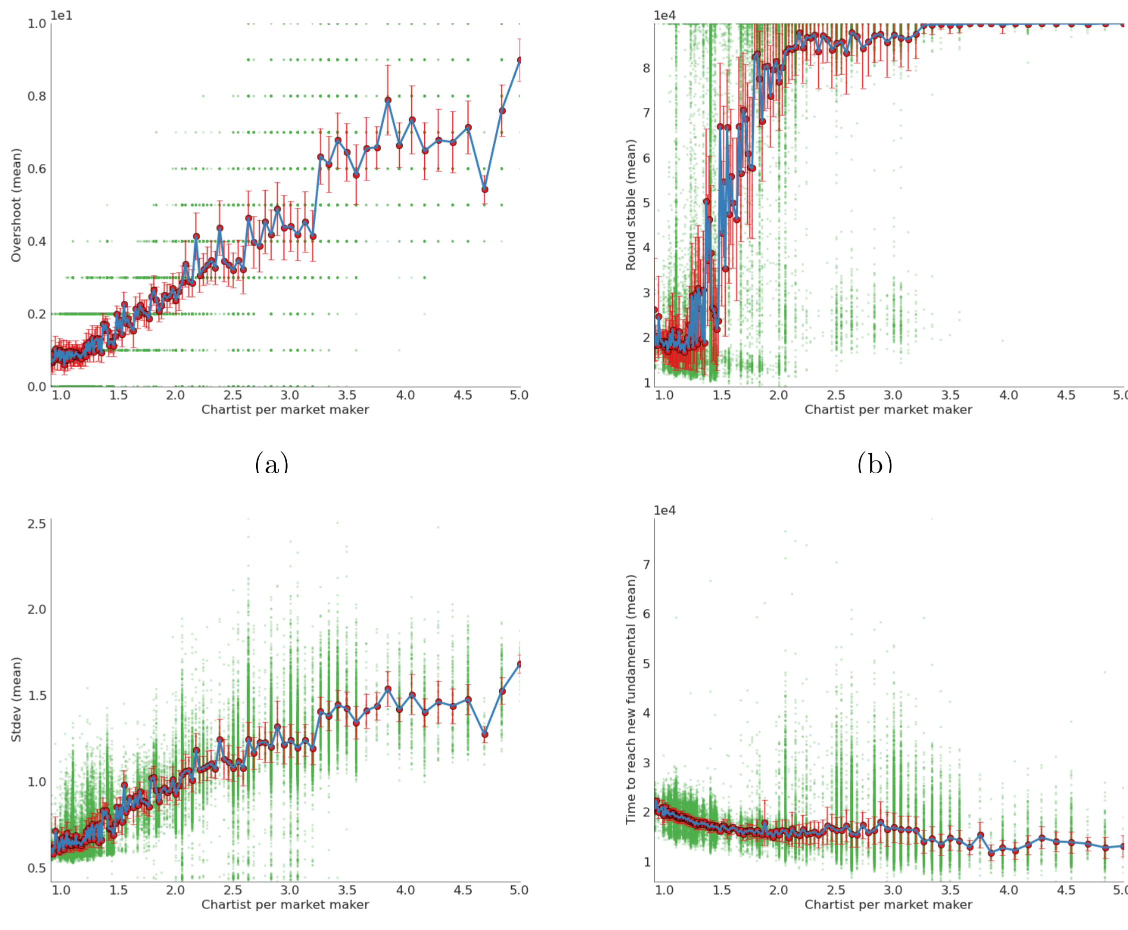


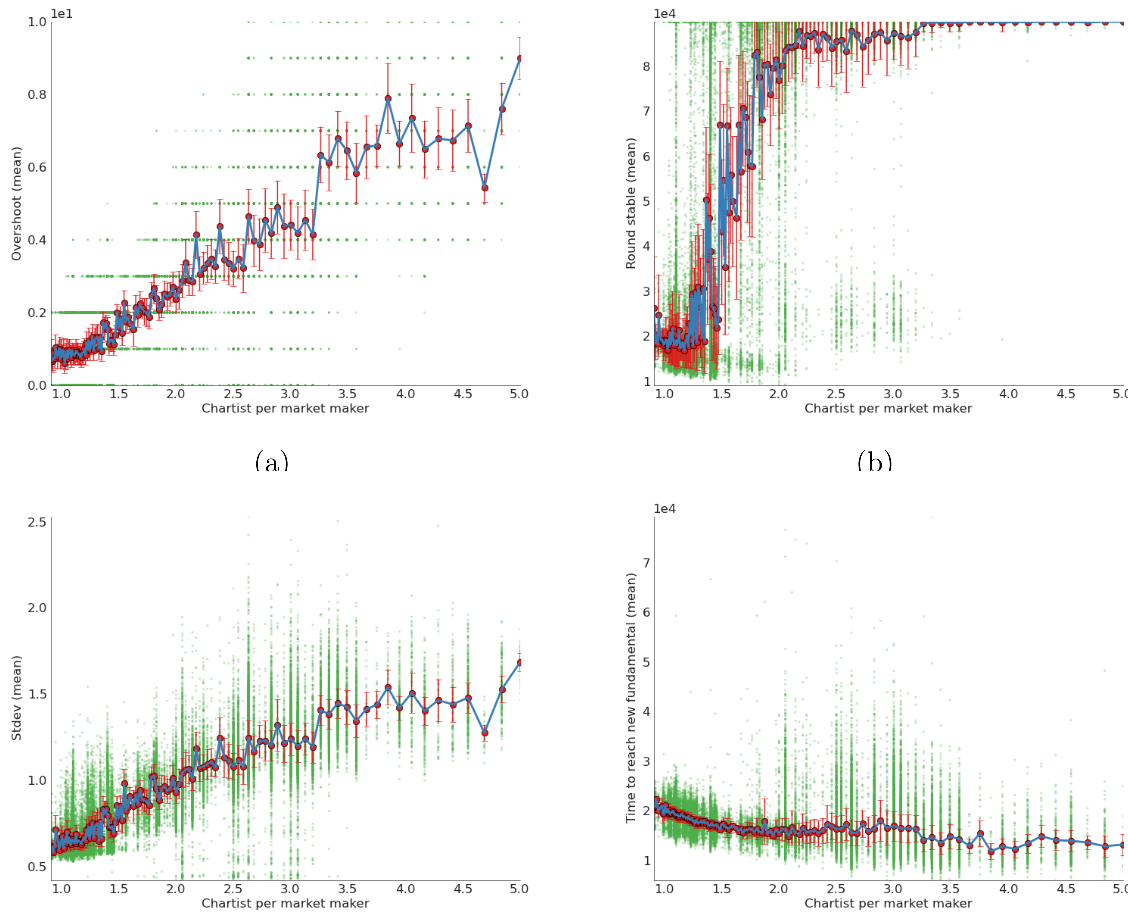
Figure 17: Relation between N_c, $\lambda_{c,\mu}$, and the model fitness when the number of market makers was fixed to $N_m = 52$ agents.



(a) (b)

(c) (d)

Figure 18: Correlations between $\rho_{\mathrm{A}}$and the fitness values when $N_c = 150$



(a) (b) (c) (d)

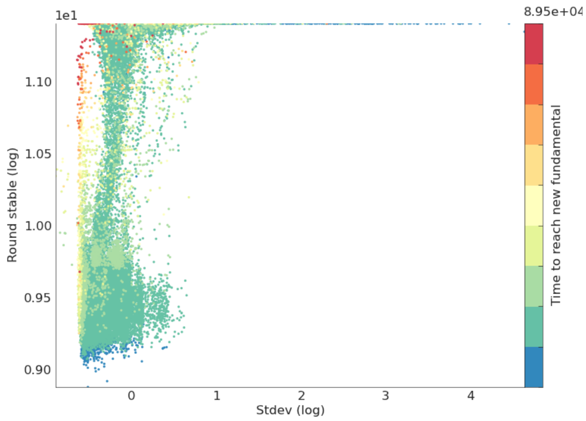
Figure 19: Correlations between $\rho_{\mathrm{A}}$and the fitness values when $N_m = 52$

3.5 Chartist to market maker ratio

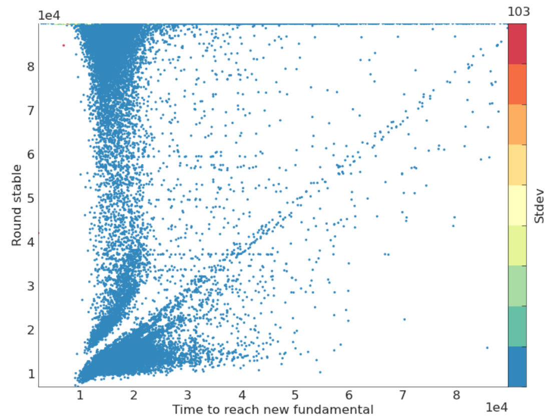
The above observations about how the number of agents influence the stability and speed of the market pointed out that more market makers made the market slow but stable, while more chartists made the market fast, but unstable. By merging $\boldsymbol{P}\mathcal{D}_{10}$ and $\boldsymbol{P}\mathcal{D}_{11}$, we can calculate the ratio, ρ_{A}, between the number of chartists and the number of market makers, and see how the fitness values correlate. Figure **??** shows the resulting scatter plots.

4 Grouping models by behavior

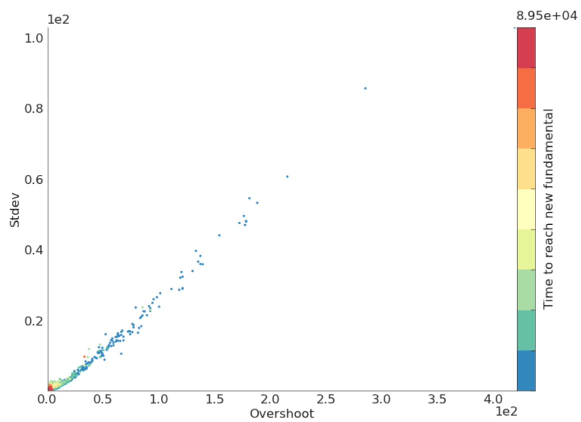
This section is be concerned with trying to tie various patterns of model behavior to different regions in the parameter space. The quickest way to get an idea of how the data generated by the simulations is distributed is to make scatter plots. Scatter plots



(a) $\log f_\sigma$ vs. $\log f_s$ vs. f_t



(b) $f_s$vs. $f_t$vs. f_σ



(c) $\log f_o$ vs . $\log f_\sigma$ vs. f_t

Figure 20: Scatter plots of fitness measures in experiment $\mathcal{D}_9$.

are probably among the most rudimentary of techniques for data analysis, yet they can be incredibly informative, especially when the data that is visualized is low-dimensional. The data of the model fitness is four-dimensional, requiring twelve plots to visualize all combinations. However, since $f_o$is discrete with a small range of values, it is not suitable for a scatter plot. Furthermore, some scatter plots are not useful for interpretation if they do not show any structure in the data. Figure 20 shows three scatter plots which were found to best illustrate the structure of the dataset from $\mathcal{D}_9$. Note also that coloring each point corresponding to its value in one dimension makes it possible to show how the data is distributed in three dimensions. The scatter plots do seem to reveal some structure, the presence of large values in the $f_\sigma$feature obscures the nature of this structure, in spite of the logarithmic scaling. The plot showing $\log f_\sigma$ vs. $\log f_s$ is squeezed to the left, and the color grading on the scatter plot for $\log f_o$ vs . $\log f_\sigma$ reveals no variety in the $f_\sigma$feature. In an attempt to get some more information out of the scatter plot, data points with an overshoot of over 100 % of the shock to the fundamental (corresponding to $f_o > 10$) are removed. The resulting scatter plots for the reduced data set are shown on figures 21 and **??**.

First of all, it is seen that while the data is distributed similarly in the three data sets, there are some differences. The data from $\mathcal{D}_9$ seems to have many "lonely" data points, which are not part of any cluster, whereas the data from $\mathcal{D}_{10}$ somehow seems to be the cleanest of the three. In all three data sets, there are clusters of data. The clusters do not necessarily mean anything in themselves. They might simply be due to the way that data points are mutated and crossed by the GA. However, by considering which regions of the fitness space that each cluster covers, it is possible to add meaning to the clusters in terms of model behavior.

4.1 Manually grouping simulations by behavior

Table 4 contains an overview of the named criteria used for roughly grouping simulations into different types of behavior. The following text contains the reasoning for why each of these groups are interesting.

In figure 21, the black dashed lines at $f_\mathrm{t} = f_\mathrm{s}$ divide each plot into region A, (upper left triangle) and region B (lower right triangle). Region A contains the fitness-points of the simulations which are counted as stable *after* they reach the new fundamental, and region B contain those that become stable before.

The description below provides a brief summary of which simulations belong in the two regions.

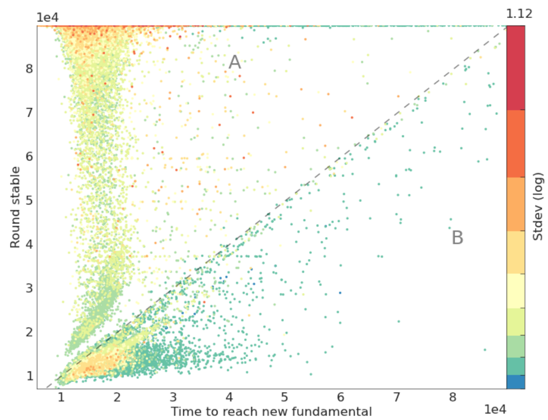
$f_\mathbf{s} < f_\mathbf{t}$ This happens when the traded price never leaves the stability margin after reaching the new fundamental price. Note however that this case does not necessarily mean that the prices do not flicker.

$f_\mathbf{s} > f_\mathbf{t}$ This happens when the traded price leaves the stability margin once or more after reaching the new fundamental. The traded price can be close to the fundamental, but flickers in and out of the stability margin as on Figure **??** shows an example where the trade price fairly stable and with no overshoot, leading to good (low) $f_\sigma$and $f_\mathrm{o}$fitness values to be assigned to the parameters. However, even though the traded prices are mostly within the stability margin, occasional flickers out of the margin causes the simulation to score a bad (high) $f_\mathrm{s}$fitness. Note also that $f_\mathrm{t}$is undefined in this case.

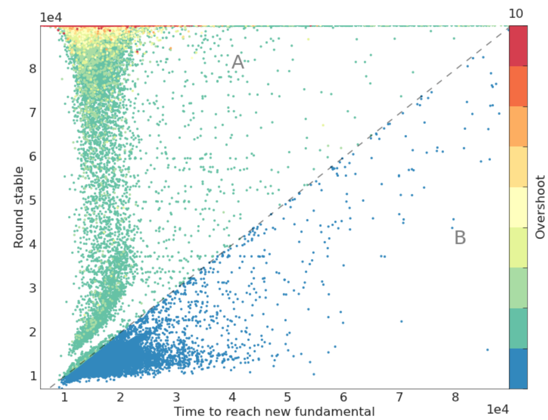
$f_\mathbf{s} = f_\mathbf{t}$ This happens if a trade is executed at price $m_\mathrm{stable} - p_\mathrm{fas} < p_\mathrm{match} < m_\mathrm{stable} + p_\mathrm{fas}$, and another trade is executed at price $p_\mathrm{match} = p_\mathrm{fas}$ in the same round.

Fast and stable simulations with flickering prices

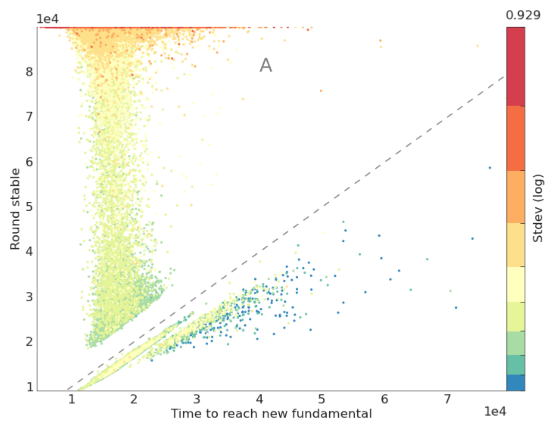
The points in the lower left corner in are those which quickly reached the new fundamental price, and quickly because stable, leading those simulations to be assigned low $f_\mathrm{t}$and $f_\mathrm{s}$fitness-values. These points are extracted by using filter F1 (see table 4)



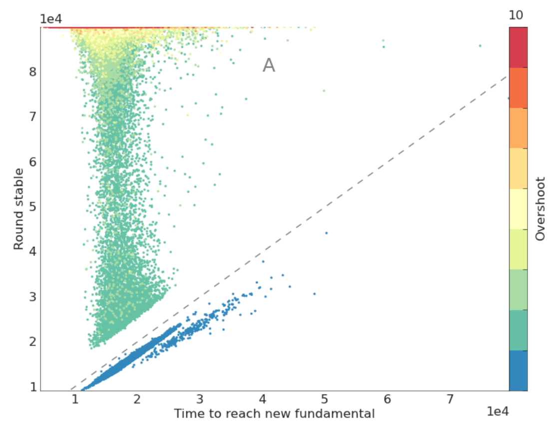
(a) Experiment $\mathcal{D}_9$



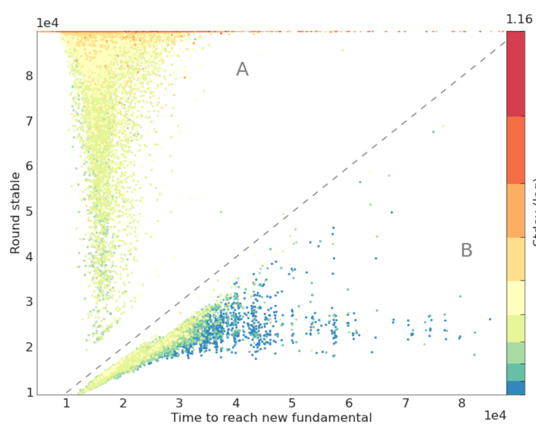
(b) Experiment $\mathcal{D}_9$



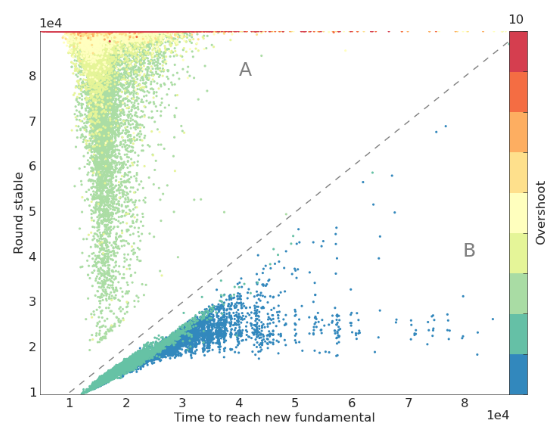
(c) Experiment $\mathcal{D}_{10}$



(d) Experiment $\mathcal{D}_{10}$

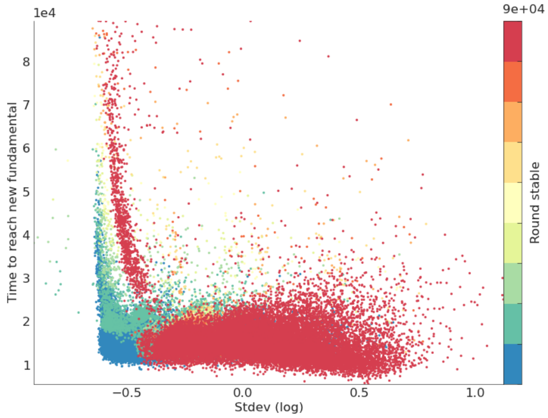


(e) Experiment $\mathcal{D}_{11}$

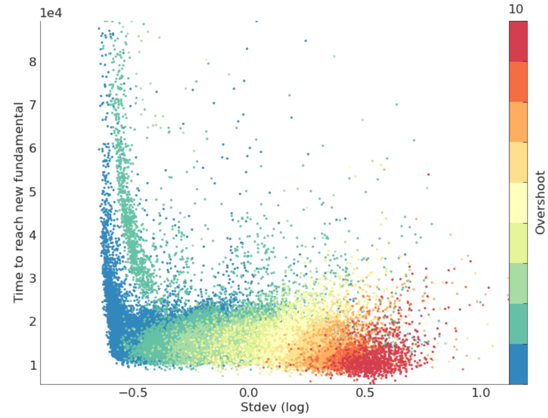


(f) Experiment $\mathcal{D}_{11}$

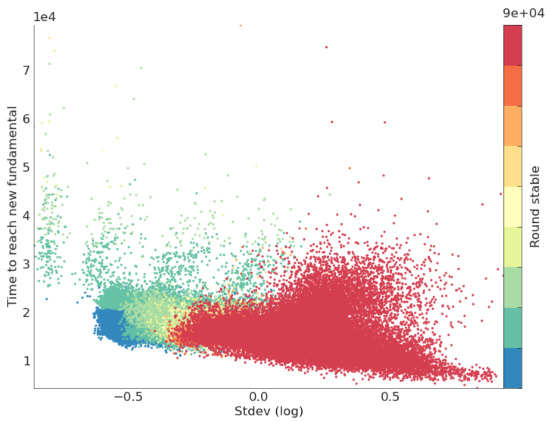
Figure 21: Scatter plot of $f_s$against $f_t$with coloring showing $\log f_\sigma$ and f_o



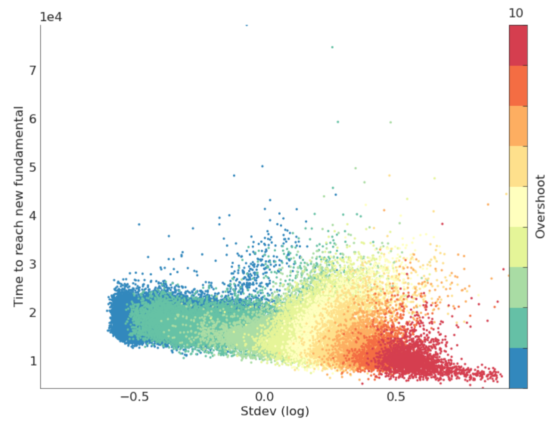
(a) Experiment $\mathcal{D}_9$



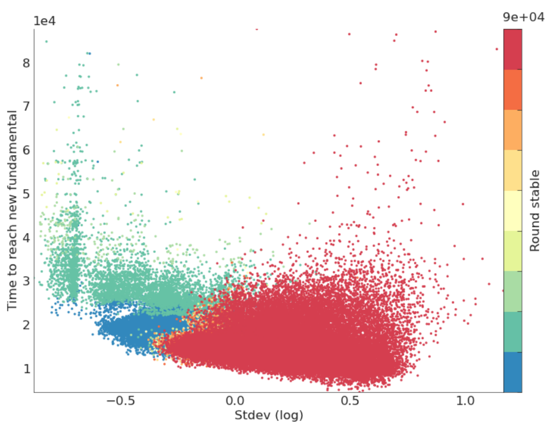
(b) Experiment $\mathcal{D}_9$



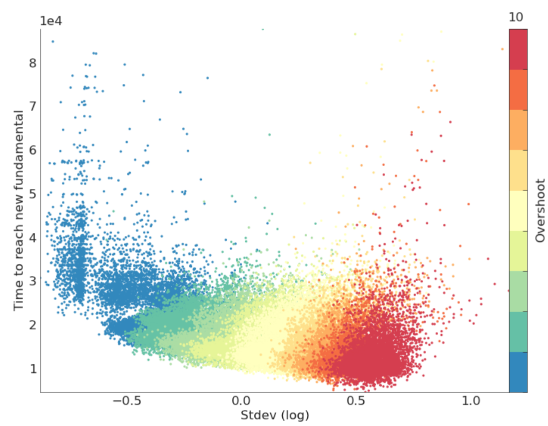
(c) Experiment $\mathcal{D}_{10}$



(d) Experiment $\mathcal{D}_{10}$



(e) Experiment $\mathcal{D}_{11}$



(f) Experiment $\mathcal{D}_{11}$

Figure 22: Scatter plot of $\log f_\sigma$ against $f_\text{t}$with coloring showing f_s and f_o

Slow or fast and stable simulations with non-flickering prices

All three data sets have data points which are close to the diagonal. However, only $\mathcal{D}_9$ has data points which are close to the diagonal in the upper right corner of the figure. These points are interesting because they belong to simulations which became stable as soon as they reached the new fundamental price. Hence, these simulations should have prices that do not flicker, and therefore yield a small f_σ-fitness. This is confirmed by looking at the left scatter plot of $\mathcal{D}_9$, as all the points close to the dotted line has a green/blue color. The right plot of $\mathcal{D}_9$ shows that these simulations did not have any overshoot. It is interesting that both slow simulations which take a long time to reach the new fundamental, as well as simulations who manage to be fast, have no overshoot, and this observation begs the question of whether or not these simulations have some common parameters that make them behave in such a way. These points are extracted by applying filters F2 and F3 (see table 4) to the data matrix $\boldsymbol{F}\mathcal{D}_9$ which selects points that lie within a distance of 400 rounds of the diagonal.

Stable before reaching the new fundamental

Most of the simulations falling in region B, meaning that they became stable before reaching the new fundamental price, had no overshoot. However, when the model was allowed to have a large number of chartists, but in $\mathcal{D}_{11}$, a group of simulations did have a small overshoot. These two groups of points were extracted by applying filters F4 and F5 (see table 4).

Simulations with overshoot

XXX NOT FINISHED

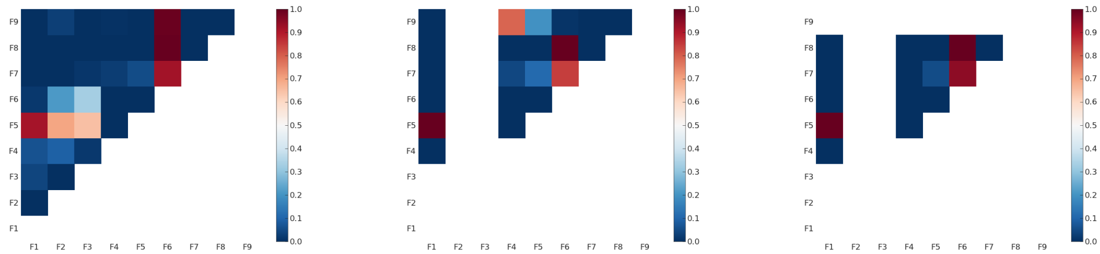
4.1.1 Fast simulations

All three experiments produces a group of simulations which had a quick response to the shock, but took longer to become stable. The simulations are in the column-shaped cluster in figure 21 and the all have relatively low f_{t}-fitness of less than 25000 rounds or so. These data points were extracted using filters F7 and F8 (see table 4).

Unstable simulations with non-flickering prices

The final group of simulation that are singles out in this section are those that had very smooth price curves (that is, a small value for f_σ), yet did not manage to become stable. These simulations were selected by filter F9.

The criteria in table 4 do not prevent a simulation to be selected by different filters. The groups of points are therefore likely to have a non-empty intersection. Using the filters is an the attempt to separate the simulations by their behavior. Hence, the filters should in general not select the same data points. The Jaccard index $J(A, B)$ is calculated between sets A and B and used to determine the overlap between the sets. Figure 23 shows the Jaccard index between the sets created by applying the filters to each of the



(a) Experiment $\mathcal{D}_9$ (b) Experiment $\mathcal{D}_9$ (c) Experiment $\mathcal{D}_{10}$

Figure 23: Jaccard index between the sets of data points extracted by each filter

three data sets. Since the distance matrix is symmetrical, only the upper left part has been plotted. In case that the filter produced an empty set, the Jaccard index is undefined, resulting in white squares.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

$$J(A, B) = \frac{|A \cap B|}{\min|A|, |B|} \tag{2}$$

Tables **??**, **??** and **??** show the fitness and parameter arithmetic means for each of the nine groups selected by the filters

When the number of chartists were fixed as in experiment $\mathcal{D}_{10}$, the simulations with overshoot (picked out by filter F6) has an average overshot of $\mathrm{E}_{\mathcal{F}f_o}[6] = 4.1$ ticks. These markets had comparatively few, but fast market makers ($\mathrm{E}_{\mathcal{F}N_m}[6] = 67.1$ and $\mathrm{E}_{\mathcal{F}\lambda_{m,\mu}}[6]$), and fast chartists ($\mathrm{E}_{\mathcal{F}\lambda_{c,\mu}}[6] = 78.3$). Figure **??** shows that F7 and F8 mostly contain points that are also contained in F6. Both F7 and F8 a lower average overshoot, and this is caused by the markets to have slower chartists, and fewer, slower market makers. When the number of chartists were varied as in experiment $\mathcal{D}_{11}$ (and with a constant of $N_m = 52$ market makers), the average latency of the chartists $\mathrm{E}_{\mathcal{F}\lambda_{c,\mu}}[6]$ did not differ from the other groups. Instead, the biggest difference was that markets with overshoot on average had a high number of chartists. As for the market maker latency, it was smaller than any of the other groups. On the the other hand, markets with no overshoot on average had a large number of marker makers when $N_c$is fixed to $N_c = 150$, although the market makers were not particularly fast. Markets with no overshoot also had the lowest average number of chartists among the nine groups.

4.2 Using clustering algorithms to group markets by behavior

4.3 Clustering with mixture of Gaussians

In this section, the focus is shifted from looking at population wide statistics to analysis sub-groups within each population. Whereas the previous sections showed that there do

ID	Target simulations	Filter criteria
F1	Fast and stable (but maybe flickering)	$f_\mathrm{t} < 12000$, $f_\mathrm{s} < 12000$, $\log f_\sigma > 0$
F2	Slow, stable and not flickering (diagonal)	$\lvert f_\mathrm{t} - f_\mathrm{s}\rvert < 400$
F3	Fast and stable and not flickering (diagonal)	$\lvert f_\mathrm{t} - f_\mathrm{s}\rvert < 400$
F4	Stable before reaching fundamental, no overshoot	$f_\mathrm{s} < f_\mathrm{t}$, $f_\mathrm{o} = 0$
F5	Stable before reaching fundamental, with overshoot	$f_\mathrm{s} < f_\mathrm{t}$, $f_\mathrm{o} =>$
F6	Has overshoot	$f_\mathrm{s} > f_\mathrm{t}$, $f_\mathrm{o} > 0$
F7	Fast response, quick to stabilize	$1000 < f_\mathrm{t} < 25000$, $20000 < f_\mathrm{s} < 40000$
F8	Fast response, slow to stabilize	$1000 < f_\mathrm{t} < 25000$, $40000 < f_\mathrm{s} < 75000$
F9	Smooth prices with a small overshoot, yet unstable	$e^{-0.5} - 0.1 < \log f_\sigma < e^{-0.5} + 0.1$

Table 4: Filter IDs and fitness-regions

	F1	F2	F3	F4	F5	F6	F7	F8	F9
$\lambda_{c,\mu}$	70.1	65.5	49.5	81.5	76.7	58.5	75.8	75.6	76.0
$\lambda_{c,\sigma}$	5.4	5.5	7.8	4.5	4.9	6.8	5.0	4.9	5.0
$\tau_{c,\mu}$	59.7	58.8	51.8	70.1	65.8	53.5	65.7	64.8	65.9
$\tau_{c,\sigma}$	12.0	12.0	12.8	11.2	11.6	11.2	11.4	11.6	11.6
$H_{c,\mu}$	1744.0	1676.1	1758.1	1766.8	1773.6	1906.9	1795.3	1757.2	1777.9
$H_{c,\sigma}$	1483.2	1472.7	1573.4	1415.7	1444.2	1346.7	1437.3	1442.4	1465.4
$W_{c,\mu}$	29.5	27.4	29.3	30.5	30.0	28.6	30.0	29.7	29.9
$W_{c,\sigma}$	3.2	4.1	4.2	3.0	3.1	5.1	3.1	3.0	3.1
$\lambda_{m,\mu}$	45.2	38.5	37.6	49.8	47.3	39.2	46.1	46.3	44.5
$\lambda_{m,\sigma}$	4.2	4.9	4.6	4.3	4.4	5.4	4.1	4.2	5.1
$\tau_{m,\mu}$	39.8	38.2	40.1	37.9	38.6	35.5	39.0	39.1	35.5
$\tau_{m,\sigma}$	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.3	0.2
f_o	1.0	1.1	1.3	0.0	1.0	3.9	1.9	2.1	1.4
f_s	11712.3	44993.4	13454.5	13121.9	13468.1	80410.6	26840.4	61259.6	54143.9
f_σ	1.2	0.7	0.7	0.6	0.7	1.1	0.8	0.9	0.6
f_t	13767.0	45089.4	13579.4	15564.3	16743.6	16513.3	17502.8	16812.7	40564.2
Count	498	33	1031	33928	41982	30468	2016	1733	2160

Table 5: XXX

	F1	F2	F3	F4	F5	F6	F7	F8	F9
$\lambda_{c,\mu}$	N/A	N/A	N/A	123.8	121.5	78.3	116.9	104.2	95.0
$\lambda_{c,\sigma}$	N/A	N/A	N/A	6.5	6.5	9.0	7.1	8.7	8.7
$\lambda_{m,\mu}$	N/A	N/A	N/A	73.9	75.8	39.2	73.3	59.6	35.9
$\lambda_{m,\sigma}$	N/A	N/A	N/A	5.6	6.0	9.6	6.6	8.7	9.2
N_m	N/A	N/A	N/A	126.5	125.6	67.1	121.2	98.0	81.2
f_{o}	N/A	N/A	N/A	0.0	1.0	4.1	1.8	2.1	0.2
f_{s}	N/A	N/A	N/A	17025.9	15920.1	81539.2	27387.5	59444.5	24392.3
f_{σ}	N/A	N/A	N/A	0.6	0.7	1.2	0.7	0.9	0.6
f_{t}	N/A	N/A	N/A	20405.2	18724.9	15504.3	18725.1	16840.5	31017.6
Count	0	0	0	8486	25574	47493	5379	2528	399

Table 6: XXX

	F1	F2	F3	F4	F5	F6	F7	F8	F9
$\lambda_{c,\mu}$	N/A	N/A	N/A	30.4	32.3	33.3	34.3	35.1	35.8
$\lambda_{c,\sigma}$	N/A	N/A	N/A	9.3	9.9	4.6	7.2	6.6	9.0
N_c	N/A	N/A	N/A	19.0	26.9	154.0	46.3	54.6	41.1
$\lambda_{m,\mu}$	N/A	N/A	N/A	67.2	55.0	38.9	50.0	47.3	46.0
$\lambda_{m,\sigma}$	N/A	N/A	N/A	8.4	11.6	22.9	17.5	19.5	14.7
f_{o}	N/A	N/A	N/A	0.0	1.0	14.6	2.0	2.6	0.0
f_{s}	N/A	N/A	N/A	15678.2	15075.3	87971.4	31954.9	63613.9	21851.4
f_{σ}	N/A	N/A	N/A	0.7	0.8	3.6	0.9	1.0	0.6
f_{t}	N/A	N/A	N/A	19965.3	18765.4	13685.0	17265.5	16449.5	31048.9
Count	0	0	0	71329	31377	84597	349	3014	1706

Table 7: XXX

indeed exist statistical relationships between the latency of the agents and the behavior of the model, each discovered correlation was calculated over the entire population. Although the correlations reveal overall tendencies of the model behavior when changing a single parameter, little can be said about how the various parameters interact to determine model behavior. For instance, even though prediction of, say a negative correlation between $f_{\mathrm{t}}$was $\lambda_{c,\mu}$was found, there might be configurations of the model in which faster chartists were actually beneficial to the market.

One way to approach this problem is to see whether or not there exists relationships between partitions in the parameter space to partitions in the fitness space. As in section **??**, the first step is to partition space, since each partition can be interpreted in terms of the model behavior. For instance, a partition covering the lower left half of the 2-dimensional space in figure 21 would encompass all the simulations which had a fast response time and became stable quickly (no matter if they had prices that flickered within the stability margin or not).

In order to investigate this, a Gaussian mixture model (GMM) was used to find clusters in the fitness space. All four fitness measures were used for the clustering. After discarding simulations with undefined fitness values and removing outliers, the data set contained 80813 data points. The large number of data points and the low number of dimensions made it possible to allow a each Gaussian component to have a full covariance matrix, giving the model a high level of flexibility. Figure 24 shows scatter plots of the data after it has been grouped. Tables **??** and **??** respectively show the mean and standard deviation calculated over each cluster. The tables are sorted by the average value of f_{o}.

XXX NOT FINISHED. WRITE ABOUT THE CLUSTERS THAT HAVE SOME INTERPRETABLE VALUE. IT DOES NOT HAVE TO BE ALL OF THEM. $\mathrm{Var}_{\mathcal{C}8}[f_{\mathrm{s}}]$ and $\mathrm{Var}_{\mathcal{C}11}[f_{\mathrm{s}}]$ and $\mathrm{Var}_{\mathcal{C}5}[f_{\mathrm{s}}]$ are large. The points in this cluster have parameters which

Chartist latency and market response time

As was noted earlier, the evolution of $\lambda_{c,\mu}$and f_{t}indicates that slow chartists made the market slow, and fast chartists made the market fast. C1 is the cluster with the $\mathrm{E}_{\mathcal{C}1}[f_{\mathrm{t}}]$

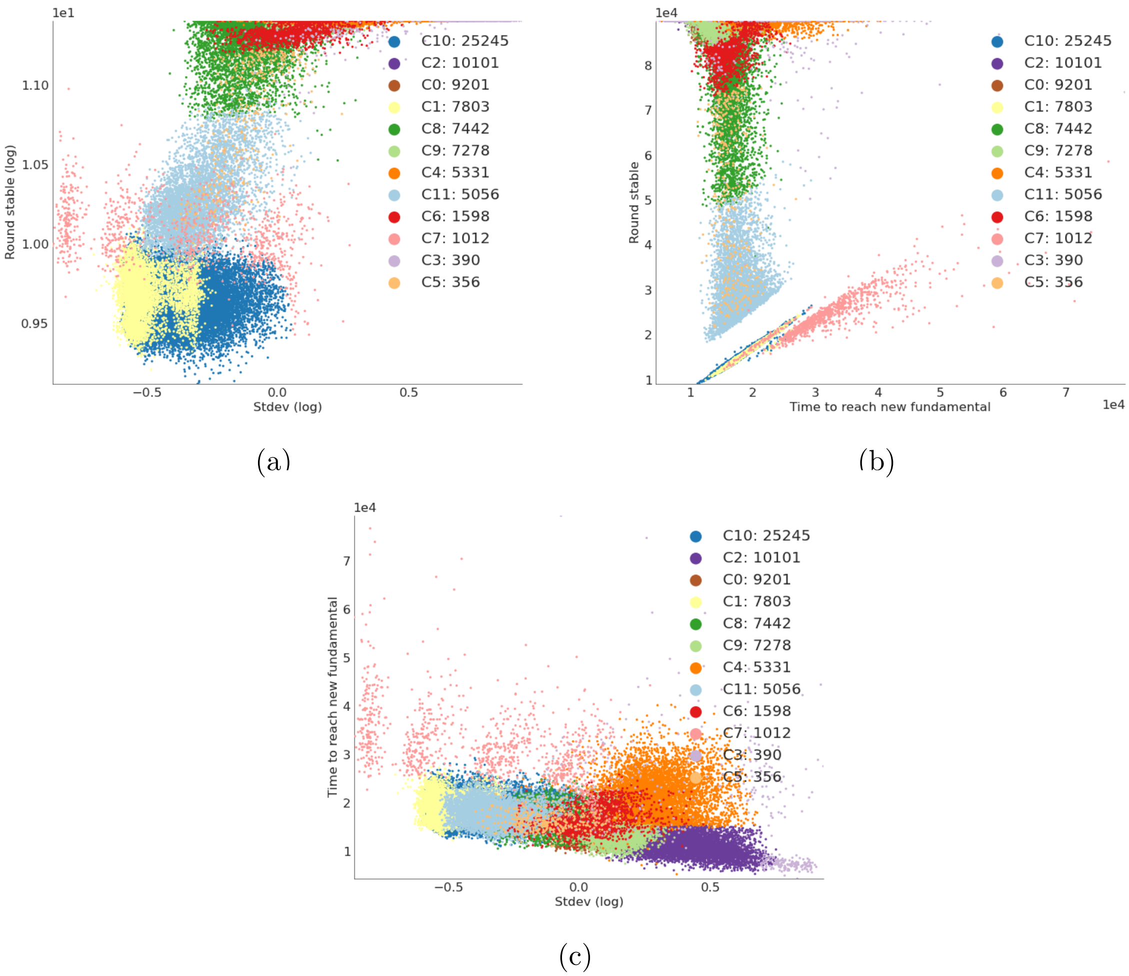
XXX CONSIDER MOVING SOME OF THE TABLES TO THE APPENDIX?

5 Summary of results

The results showed that both types of HFT agents influence the market in positive and negative ways. This section contains a summary of the results.

Analysis of the results showed that a moderate amount of high speed trading activity was not in itself problematic.

- Market makers have a stabilizing effect on the markets and make the market more robust
- High speed chartists influence the market so as to decrease the duration of the period of time in which the stock price stock price is overvalued by reducing the



(a) (b) (c)

Figure 24: Scatter plots in fitness space showing the grouping of data points when using a GMM with 12 clusters.

	f_o	f_s	f_σ	f_t	$\lambda_{c,\mu}$	$\lambda_{c,\sigma}$	$\lambda_{m,\mu}$	$\lambda_{m,\sigma}$	N_m	Count
C1	0.0	16301.2	0.6	19217.4	127.2	6.2	78.4	5.2	132.2	7803
C7	0.3	24383.1	0.7	32005.5	87.0	9.5	25.8	10.5	66.7	1012
C10	1.0	15832.8	0.7	18602.8	121.9	6.5	76.3	6.0	126.3	25245
C8	2.0	81599.7	0.9	16333.6	101.8	8.7	56.4	9.2	91.3	7442
C11	2.0	30515.0	0.7	17822.9	114.2	7.3	71.7	7.0	117.5	5056
C0	3.0	89015.0	1.1	14687.5	89.0	9.2	39.5	10.0	66.8	9201
C5	3.0	51306.4	0.9	16508.7	102.2	9.0	59.2	8.8	96.0	356
C6	3.0	83914.2	1.1	16820.9	92.3	9.2	40.6	10.0	72.4	1598
C9	4.0	89496.0	1.2	14459.8	77.7	9.6	29.9	10.4	56.7	7278
C4	5.0	89896.6	1.4	22034.2	62.4	9.4	21.5	10.3	51.6	5331
C3	6.7	86440.9	1.8	18692.7	54.9	9.2	26.2	10.5	38.0	390
C2	7.9	89996.1	1.6	11574.7	36.4	9.2	26.4	9.6	36.4	10101
Outliers	11.5	89998.8	2.2	8835.1	17.8	9.3	24.3	8.5	19.4	740

Table 8: Cluster means ($\mathcal{D}_{10}$)

	f_o	f_s	f_σ	f_t	$\lambda_{c,\mu}$	$\lambda_{c,\sigma}$	N_c	$\lambda_{m,\mu}$	$\lambda_{m,\sigma}$	Count
C5	0.0	15270.3	0.7	19209.4	30.0	9.3	17.5	68.7	7.9	66665
C11	0.0	21182	0.6	29334.9	35.5	8.9	40.2	46.6	14.5	4369
C0	1.0	14897	0.8	18523.0	32.2	10.0	25.9	55.6	11.3	30258
C4	1.0	19666	0.9	25087.1	35.3	7.3	55.5	39.0	18.9	1108
C9	1.0	30399	0.7	35789.7	34.4	7.8	45.6	43.6	17.9	591
C6	2.0	79198	0.9	15620.1	37.5	6.1	63.5	47.3	20.6	11141
C7	2.6	78771	1.0	20524.6	37.7	5.6	63.8	38.5	22.6	967
C1	3.0	88210	1.0	14225.8	37.9	4.6	88.0	44.2	23.3	10577
C8	4.0	89633	1.1	13452.7	33.6	4.4	103.9	44.4	22.7	8099
C3	5.7	89842	1.4	20087.3	34.6	4.5	136.1	29.6	23.9	9613
C10	7.0	89935	1.8	27587.6	32.8	4.2	165.8	25.9	24.3	1401
C2	7.0	89982	1.5	11691.4	32.3	4.4	154.4	37.6	22.7	19067
Outliers	40.5	89951	9.9	10434.7	29.2	4.0	255.5	36.4	23.5	23454

Table 9: Cluster means ($\mathcal{D}_{11}$)

	f_o	f_s	f_σ	f_t	$\lambda_{c,\mu}$	$\lambda_{c,\sigma}$	N_c	$\lambda_{m,\mu}$	$\lambda_{m,\sigma}$	Count
C0	0.0	1349.9	0.1	1678.1	13.7	6.5	12.1	16.3	8.1	30258
C1	0.0	2448	0.1	1383.3	16.8	4.6	15.5	18.6	10.1	10577
C4	0.0	3347	0.1	4304.5	16.8	5.7	33.3	17.4	10.8	1108
C5	0.0	1102	0.0	1155.7	12.0	6.8	4.2	13.0	6.9	66665
C6	0.0	11795	0.1	1640.9	17.7	5.1	18.1	19.4	10.1	11141
C8	0.0	893	0.1	1446.5	16.7	4.4	19.1	20.5	9.7	8099
C11	0.0	2801	0.1	4505.8	15.0	6.4	31.7	18.6	11.2	4369
C7	0.6	6266	0.2	4695.2	17.4	4.9	29.9	16.7	9.7	967
C9	1.1	9640	0.2	18435.6	16.5	5.8	33.1	18.6	10.2	591
C2	1.7	143	0.2	1566.7	16.3	4.7	39.8	17.6	8.9	19067
C3	2.2	522	0.3	3765.1	16.7	4.5	48.7	13.0	8.9	9613
C10	2.3	407	0.3	10643.3	16.8	4.4	52.4	12.4	8.8	1401
Outliers	53.1	1107	15.0	3790.6	15.5	4.5	50.7	15.9	8.3	23454

Table 10: Cluster standard deviations ($\mathcal{D}_{11}$)

time required for the price to return to the new fundamental price after the negative shock.

- Crashes never occur in markets with only fast market makers or fast chartists, no matter how many agents are active and how fast they are. Even markets with

  In fact, the markets that had the fastest and most stable response to the shock had both several fast chartists and fast market makers. However, it was found that flash-crashes could occur in markets in which the number of chartists was four times or greater that of the number of market makers. Furthermore, crashes occurred four times more often when the chartists were faster than the market makers.

The results were analysed while focussing on the market dynamics caused by the aggregate actions of all the traders. High speed market makers were found to reduce price flickering, confirming the results of [Wang].

On the negative side, the market was found to respond unfavorably to a large presence of fast traders. First of all, a large presence of very fast market makers pushed the market towards responding sluggishly to the shock, leading to the stock being traded at a price higher than the fundamental price. Secondly, a large presence of fast chartists caused increased flickering of the traded price.

Lastly, the analysis discovered indications that strategy crowding could occur when the fast traders had narrowly distributed latencies causing them to act simultaneously on the same information, but a verfification of this would require further experiments. These results add to the current debate on computerized high frequency trading as they show both benefits and dangers of having markets where fast computer algorithms trade side by side with human traders.

Market makers were found to

Part II
Discussion

6 Benefits of fast traders

6.1 Fast market makers reduce price flickering

In

7 Agent strategies market crashes

It is somewhat intuitive that HFT chartists should be suspected of having an influence on the market such that the market become more likely to crash. The results did indeed confirm this, as it was shown that a negative correlation exists between the number of chartists active in the market, and the size of the overshoot (see figure **??**).

It is conceivable that the market makers also contribute to the market crashing, since the market makers also ignore the true fundamental price. Indeed, the results showed that the market does not crash from having fast chartists alone. The results also showed that the market will not crash from having only fast market makers either. Instead, both types of fast traders were required to make the market crash. The following text will provide an explanation as to why this is so.

It was found that markets containing no market makers will almost never crash. Without the presence of market makers, even a market saturated with chartists will eventually return to the fundamental price, as the force of the initial downtrend created by the fundamentalists dissipates.

Case with no fast traders

The shock to the fundamental creates a drive in the market for falling prices, due to the presence of the fundamentalists. The fundamentalists have a large delay, and the downwards drive is therefore initially small, as most of the fundamentalists fail to observe that the shock has happened. As the fundamentalists begin to observe the shock, they start submitting sell orders at lower prices, as they believe that the stock is no longer worth the price at which they were previously willing to sell. Hence, the number of sell orders starts to increase.

As for the buy side of the order book, the number of new buy orders starts to fall, as the fundamentalists start to register the shock. The buy orders that were previously submitted by slow traders at prices slightly below the old fundamental are not canceled, as the model assumes that the fundamentalists are too slow to register the change in the fundamental. Furthermore, in order to simulate an order book with a long trade history, the order book was initialized with a large number of market orders with a normal price distribution centered around the initial fundamental. These buy orders provide matches for the increasing number of sell orders, and the traded price begins to drop. If the market has no fast traders, the traded price will eventually reach the new fundamental, and stay within the stability margin. Thus, in the rather simple case where the market only contains fundamentalists, crashes do not occur.

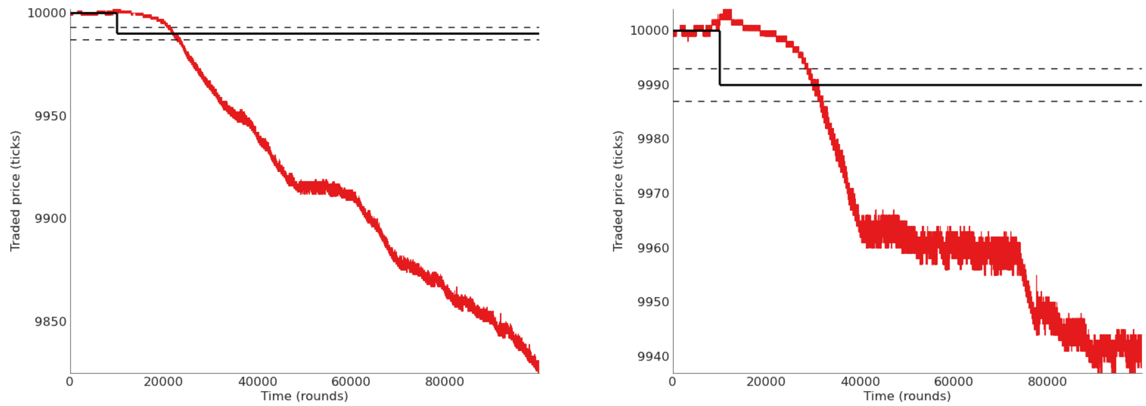
Case with chartists but no market makers

When adding chartists, the market starts to behave in a different manner. The chartists do not use any information about the true true fundamental price, but are instead only concerned with the actual traded price. After the shock, the fundamentalists start submitting bids to sell at lower prices, Depending on the parameters of the chartists, some chartists will interpret this as a downtrend, while others will not. The chartists that detect a downtrend will start submitting bids to sell at a lower price, as they believe the price will continue to drop. The chartists that did not detect a trend will remain inactive. The sell orders submitted by the chartists are matched by previously existing buy market orders at lower prices. Hence, the chartists add to the force that drives the traded price down by submitting sell orders at lower prices. However, since the only active traders in the market are fundamentalists and chartists, the supply of buy orders at prices lower than the new fundamental are limited. The chartists that detected a downtrend will exclusively place sell orders, and the fundamentalists will rarely submit buy orders at prices much lower than the fundamental. When the supply of buy orders at prices below the new fundamental dries out, the execution price will not drop further. The chartists that detected a downtrend will continue to submit sell orders for as long as they believe that the trend continues, but the only new buy orders are submitted by the fundamentalists. As some of the fundamentalist buy orders are placed a few ticks below the fundamental price, the execution price will flicker, but always in a region close the true fundamental price.

Case with chartists and market makers

When the market also contains market makers, the situation is quite different. Like the chartists, the market makers ignore the fundamental price. Instead they submit buy and sell orders just above and below the best buy and sell prices existing in the order book at the time that the market maker requested the market information. The market maker strategy is such that it will always try to follow a narrowing spread, in order to stay competitive. On the other hand, if the market maker discovers that the spread is widening, the agent will attempt to avoid buying/selling at a higher/lower price than necessary. The agent therefore tries to follow the widening spread by submitting buy/sell orders at lower/higher prices.

When the sell price starts to drop after the shock due to the activity of the fundamentalists and the chartists, the market maker will try to stay competitive on the sell side be decreasing its own sell price. If the decrease in the sell price is large enough to make the spread smaller than what the agent is prepared to risk, the market maker submits a new sell order with as low a price than its strategy allows.



(a) (b)

Figure 25: Two examples of market crashes

On the other side of the order book, the best buy price starts to drop as the sell orders submitted by the chartists start to eat away at the existing buy orders. If the market maker orders are among the orders that match the chartist orders, the market makers request the latest market information and use it to submit new buy orders. If the market maker orders were not matched by chartist orders the market makers will cancel their existing buy order and submit a new one at a lower price in order to stay a competitive buyer. In any case, the market maker will eventually start submitting buy orders at a lower price than before. Hence, the market makers provide the market with a new supply of buy orders, the prices of which can be arbitrarily low. As these buy orders are filled by sell orders, initially by both fundamentalists and chartists but eventually solely by chartists, the traded price will drop, and the chartists will continue detecting a trend and continue to drive the market down into a crash.

7.1 Frequency of crashes

Crashes did not occur often particularly. Even during experiment $\mathcal{D}_{11}$, in which the genetic algorithm ended up prioritizing the market response times, and generate a large number of genes with many chartists and very fast market makers, the market had an overshoot of over 25 tick in just less than 0.2% of the cases[3]. In experiment $\mathcal{D}_{10}$, not a single case of markets with an overshoot of over 17 ticks was generated.

XXX ADD A SMALL DISCUSSION OF HOW OFTEN CRASHES OCCUR IN REAL MARKETS

[3]The simulation was run around $4 \cdot 10^5$ times, and 7989 of these had $f_o > 25$

7.2 Agent speed and market crashes

XXX INSERT DATA THAT SHOWS THAT CRASHING MARKETS HAD FASTER AGENTS THAN NON CRASHING MARKETS

8 Market makers causing the stock to be over-valued

XXX NOT FINISHED. COLLECT EVIDENCE This temporary over-evaluation of the assert is on the expense of Hence, the

in [], the authors discussed the scenario in which a disparity between the this scenario and

9 title

XXX discuss the variability of market behavior for markets simulated with the same parameters. is it reasonable that the same set of parameters can cause various types of behavior? XXX

10 Co-location

Stable markets had an average market maker latency of $\lambda_{m,\mu} = 60ms$, while crashing markets had an average of $\lambda_{m,\mu} = 30ms$. Is it realistic that just a factor of two can have such a dramatic effect on the markets?

The recent construction of the transatlantic fiber line reduced the latency from European markets to American markets

Note also that latency is nt just a result of physical distance in the market. Network crowding can cause an increase in the latency as well.

11 Strategy crowding

Strategy crowding is a commonly observed phenomenon in the field of multi-agent systems. XXX FIND REFERENCE XXX

12 Future work

A commonly used strategy among high frequency traders is that of arbitrage,

References

[1] Franklin Allen and Risto Karjalainen. Using genetic algorithms to find technical trading rules. *Journal of financial Economics*, 51(2):245–271, 1999.

[2] Yakov Amihud and Haim Mendelson. Dealership market: Market-making with inventory. *Journal of Financial Economics*, 8(1):31–53, 1980.

[3] Volkan Arslan, Patrick Eugster, Piotr Nienaltowski, and Sebastien Vaucouleur. Scoop–concurrency made easy. In *Dependable Systems: Software, Computing, Networks*, pages 82–102. Springer, 2006.

[4] Bruno Biais and Paul Woolley. High frequency trading. *Manuscript, Toulouse University, IDEI*, 2011.

[5] Jeff A Bilmes et al. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126, 1998.

[6] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.

[7] Tobias Blickle and Lothar Thiele. A comparison of selection schemes used in genetic algorithms, 1995.

[8] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[9] Jonathan Brogaard, Björn Hagströmer, Lars L Norden, and Ryan Riordan. Trading fast and slow: Colocation and market quality. *Manuscript*, 8:25, 2013.

[10] Alain Chaboud, Erik Hjalmarsson, Clara Vega, and Ben Chiquoine. Rise of the machines: Algorithmic trading in the foreign exchange market. *FRB International Finance Discussion Paper*, (980), 2009.

[11] Carl Chiarella, Giulia Iori, and Josep Perelló. The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics and Control*, 33(3):525–537, 2009.

[12] Silvano Cincotti, Sergio M Focardi, Linda Ponta, Marco Raberto, and Enrico Scalas. The waiting-time distribution of trading activity in a double auction artificial financial market. In *The Complex Networks of Economic Interactions*, pages 239–247. Springer, 2006.

[13] Paul Davidsson. *Multi agent based simulation: beyond social simulation*. Springer, 2001.

[14] Lawrence Davis. Handbook of genetic algorithms. 1991.

[15] Ann Devitt and Khurshid Ahmad. Sentiment polarity identification in financial news: A cohesion-based approach. In *ACL*, 2007.

[16] Alexis Drogoul, Diane Vanbergue, and Thomas Meurisse. Multi-agent based simulation: Where are the agents? In *Multi-agent-based simulation II*, pages 1–15. Springer, 2003.

[17] David Easley and Jon Kleinberg. *Networks, crowds, and markets*, volume 8. Cambridge Univ Press, 2010.

[18] J Doyne Farmer and Spyros Skouras. An ecological perspective on the future of computer trading. *Quantitative Finance*, 13(3):325–346, 2013.

[19] Thierry Foucault, Johan Hombert, and Ioanid Rosu. News trading and speed. *Available at SSRN 2188822*, 2012.

[20] Brendan J Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.

[21] Nigel Gilbert. Agent-based social simulation: dealing with complexity. *The Complex Systems Network of Excellence*, 9(25):1–14, 2004.

[22] Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. Large-scale sentiment analysis for news and blogs. *ICWSM*, 7, 2007.

[23] David E Goldberg and Kalyanmoy Deb. A comparative analysis of selection schemes used in genetic algorithms. *Urbana*, 51:61801–2996, 1991.

[24] Markus Gsell. Assessing the impact of algorithmic trading on markets: A simulation approach. Technical report, CFS working paper, 2008.

[25] Peter Hoffmann. A dynamic limit order market with fast and slow traders. *Available at SSRN 1969392*, 2012.

[26] Cars H Hommes. Heterogeneous agent models in economics and finance. *Handbook of computational economics*, 2:1109–1186, 2006.

[27] Alan J Izenman. *Modern multivariate statistical techniques: regression, classification, and manifold learning.* Springer, 2008.

[28] Kiyoshi Izumi, Fujio Toriumi, and Hiroki Matsui. Evaluation of automated-trading strategies using an artificial market. *Neurocomputing*, 72(16):3469–3476, 2009.

[29] Neil Johnson. Financial black swans driven by ultrafast machine ecology. *Available at SSRN*, 2012.

[30] Eric Jones, Travis Oliphant, and Pearu Peterson. Scipy: Open source scientific tools for python. *http://www. scipy. org/*, 2001.

[31] Moshe Koppel and Itai Shtrimberg. Good news or bad news? let the market decide. In *Computing attitude and affect in text: Theory and applications*, pages 297–301. Springer, 2006.

[32] Blake LeBaron. Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control*, 24(5):679–702, 2000.

[33] Blake LeBaron. A builders guide to agent-based financial markets. *Quantitative Finance*, 1(2):254–261, 2001.

[34] Charles M Macal and Michael J North. Tutorial on agent-based modeling and simulation. In *Proceedings of the 37th conference on Winter simulation*, pages 2–15. Winter Simulation Conference, 2005.

[35] Michael J McGowan. Rise of computerized high frequency trading: Use and controversy, the. *Duke L. & Tech. Rev.*, page i, 2010.

[36] Thomas McInish and James Upson. Strategic liquidity supply in a market with fast and slow traders. *Available at SSRN 1924991*, 2012.

[37] Travis E Oliphant. *A Guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[38] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[39] J. Pitman. *Probability*. Springer Texts in Statistics. Springer, 1993.

[40] Jean-Yves Potvin, Patrick Soriano, and Maxime Vallée. Generating trading rules on the stock markets with genetic programming. *Computers & Operations Research*, 31(7):1033–1047, 2004.

[41] Marco Raberto, Silvano Cincotti, Sergio M Focardi, and Michele Marchesi. Agent-based simulation of a financial market. *Physica A: Statistical Mechanics and its Applications*, 299(1):319–327, 2001.

[42] De Rainville, Félix-Antoine Fortin, Marc-André Gardner, Marc Parizeau, Christian Gagné, et al. Deap: A python framework for evolutionary algorithms. In *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference companion*, pages 85–92. ACM, 2012.

[43] Neil J Salkind. *Encyclopedia of research design*, volume 1. Sage Publications, Incorporated, 2010.

[44] D Sculley. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, pages 1177–1178. ACM, 2010.

[45] Jonathon Shlens. A tutorial on principal component analysis. *Systems Neurobiology Laboratory, University of California at San Diego*, 2005.

[46] Steven Skiena. *The Algorithm Design Manual: Text*, volume 1. Springer, 1998.

[47] Leigh Tesfatsion. Agent-based computational economics: Growing economies from the bottom up. *Artificial life*, 8(1):55–82, 2002.

[48] Leigh Tesfatsion. Agent-based computational economics: A constructive approach to economic theory. *Handbook of computational economics*, 2:831–880, 2006.

[49] Stefan Van Der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.

[50] Chi Wang, Kiyoshi Izumi, Takanobu Mizuta, and Shinobu Yoshimura. Investigating the impact of trading frequencies of market makers: a multi-agent simulation approach. *SICE Journal of Control Measurement, and System Integration*, 4(1):001–005, 2013.

[51] Darrell Whitley. A genetic algorithm tutorial. *Statistics and computing*, 4(2):65–85, 1994.