COMPARING PREDICTION MARKET MECHANISMS USING AN EXPERIMENT-BASED MULTI-AGENT SIMULATION

Frank M. A. Klingert and Matthias Meyer
Institute of Management Control and Accounting
Hamburg University of Technology
Hamburg, Germany
Email: frank.klingert@tu-harburg.de and matthias.meyer@tu-harburg.de

KEYWORDS
Prediction markets, mechanisms, continuous double auction, logarithmic market scoring rule, multi-agent simulation

ABSTRACT

Prediction markets are an interesting instrument to draw on the "wisdom of the crowds", e.g., to forecast sales or project risks. So far, mainly two market mechanisms have been implemented in prediction markets, the continuous double auction and logarithmic market scoring rule. However, the effects of the choice between these two market mechanisms on relevant variables such as prediction market accuracy are not fully understood. These effects are relevant as faulty prediction market outcomes might cause wrong decisions. This work contributes via an experiment-based simulation model to understand the mechanism-related effects and to direct further laboratory experiments. Our results show, that the mechanism decision does matter. Due to the higher amount of trades and the lower standard deviation of the price, the logarithmic market scoring rule seems to have a clear advantage on a first view. Taking the accuracy error as an independent variable, the effects are not as straightforward and depend on the environment and actors.

INTRODUCTION

Prediction markets can be described as markets which are "designed specifically for information aggregation and revelation" (Wolfers and Zitzewitz 2004, p. 108). They are an important example of the use of the "wisdom of the crowds" (Surowiecki 2010). The basic idea is to provide individuals with the possibility to trade their expectations concerning the relevant variable (e.g., the expected number of sales of a product in the next year) on a virtual market and to aggregate this local knowledge of individual actors in form of the market price, which is again visible to all market participants. Thus, these markets can be used to reveal and aggregate the diverse knowledge of even large groups at different locations and they differ considerably from traditional forecasting methods like expert forecasting or statistical methods. They have been successfully applied in research (e.g., Iowa Electronic Markets, e.g., Forsythe et al. 1992) as well as practice. A prominent example is Hewlett Packard's prediction market, which has shown an increased accuracy compared to traditional sales forecasts (Chen and Plott 2002).

A major design question when setting up a prediction market is the choice of a market mechanism (Spann and Skiera 2003, p. 1314), i.e. how trades by individuals on this market can be conducted. Several different market mechanisms have been applied to prediction markets to coordinate the trading interactions between individual actors. This paper focuses on the most commonly used, the continuous double auction (CDA, e.g. used in the Iowa Electronic Market by Forsythe et al. 1992, p. 1144) and the logarithmic market scoring rule (LMSR, invented by Hanson 2003; e.g. used in the Gates Hillman Prediction Market by Othman and Sandholm 2010, p. 368). On average, the performance of prediction markets has been "pretty good" (Wolfers and Zitzewitz 2004, p. 119), while there are cases, where they failed to aggregate information well (e.g., Hansen et al. 2004).

The effects of different market mechanisms on the results are often not intensively discussed in existing research publications, but research does document their importance. In field prediction markets (e.g., Forsythe et al. 1992; Othman and Sandholm 2010; Hansen et al. 2004), the mechanisms are regularly not varied as this would double the effort. Furthermore, they are often chosen without an intensive discussion or at least without documenting it. Some software providers in the area of prediction markets even do not offer alternative mechanisms, e.g., Crowdworx does focus on the LMSR (Ivanov 2008). Nevertheless, some differences are known. From the technical perspective, the LMSR offers constant liquidity and only needs one trader to execute a transaction while the CDA demands at least two.

The possible importance of the choice of an appropriate market mechanism stands, however, in contrast to the current understanding of the effects of different mechanisms on prediction market outcomes (Healy et al. 2010, p. 1995). The existing laboratory results focus on a small selection of aspects, e.g., the influence of knowledge distribution (Healy et al. 2010; Ledyard et al. 2009). Beyond them, especially the influence of the context together with a certain mechanism is unclear (Healy et al. 2010, p. 1995). For instance, the mentioned studies do not vary important aspects, such as the initial money endowment and the trading strategies of their actors. Furthermore, some existing results contradict each other. According to Healy et al. (2010) the

Proceedings 26th European Conference on Modelling and Simulation ©ECMS Klaus G. Troitzsch, Michael Möhring, Ulf Lotzmann (Editors)
ISBN: 978-0-9564944-4-3 / ISBN: 978-0-9564944-5-0 (CD)

accuracy of the LMSR is much worse than the accuracy of the CDA in a simple environment with few traders. This differs to the outcome of an experiment mentioned in a talk of J. O. Ledyard referenced in the same paper (Healy et al. 2010, p. 1995) and another experiment (Ledyard et al. 2009). Finally, the available laboratory experiments (Healy et al. 2010; Ledyard et al. 2009) share a problem as they all have only a very small number of traders starting from 3 up to 6. This number is lower than an average prediction market. It is even lower compared to the average prediction market in a corporate context which generally involves fewer traders than other prediction markets.

This lack of knowledge and the resulting problems are relevant as faulty prediction market outcomes might cause wrong decisions. Due to the increased application of prediction markets in the corporate context (Bray et al. 2008, p. 6), the relevance of differences between mechanism has further increased. Beyond, corporate prediction markets might face different conditions. For example, the amount of traders might be limited in corporate internal markets. As a result, traders who do not have any knowledge on how to trade optimally on a stock market might matter more than in large prediction markets. Therefore, the trading of some traders should rather be represented by a random strategy than by a strategy based on the expected value. A further aspect in the corporate context is that even marginal differences can have a high influence on important corporate decisions, e.g., the acquisition of a company. If there is an effect of the mechanism choice, an intentional decision is therefore necessary. This fact has already been recognized by Spann and Skiera (2003, p. 1314), who include the "choice of [a] trading mechanism" as a crucial step in the design process of a prediction market. But what are the relevant aspects for this mechanism decision?

Against this backdrop, this paper provides a comparative analysis of the CDA and LMSR concerning relevant prediction market output variables such as amount of trades, standard deviation of the price and accuracy error. It contributes by introducing and analyzing an agent-based simulation model to understand the mechanism-related effects and the dynamics of the collective knowledge aggregation of the participants. It aims at strong links with economic laboratory experiments following an iterative process (Klingert and Meyer 2012), as it is based on a laboratory experiment, i.e. the simulation model is constructed, micro and macro validated based on the experimental data and results of Hanson et al. (2006). It also wants to provide input for future laboratory experiments by identifying factors which should further be investigated. Our results show, that the mechanism decision does matter. Due to the higher amount of trades and the lower standard deviation of the price, the LMSR seems to have a clear advantage on a first view. Taking the accuracy error as an independent variable, the effects are less simple and depend on the environment and actors.

The paper is structured as follows. First, the research questions and hypothesis are derived. Second, the simulation model is introduced. Third, this model is validated from several perspectives. Fourth, experiments are conducted, analyzed and tested on robustness. The paper concludes with a discussion of the results.

LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

In this paper the continuous double auction and the logarithmic market scoring rule (Hanson 2003) are compared. The continuous double auction is the most widely used market mechanism. It allows traders to place asks and bids in an order book and executes a trade when a bid or ask is accepted or an ask or bid are crossing. Contrary in LMSR, the market maker is always the trading partner. This continuously offers to buy and sell for a certain price according to a logarithmic function (see Hanson 2003 for details).

What is the most relevant criterion to measure the quality of prediction market results? Which mechanism is superior to the other one? In this section, these questions will be split up to testable research questions and hypotheses in two steps. First, the dimensions of the evaluation are chosen. These dimensions support the judgment of when a certain prediction market mechanism can be seen as "superior" to another one. Second, along these dimensions, the research questions and hypotheses will be derived based on prior research and the documented differences of both mechanisms.

There are three evaluation criteria used subsequently. The first evaluation criterion is the quantity of trades. For the success of a market, the amount of trades is important. With an increased amount of trades, potentially more pieces of information can be added to the market price. The case of "no trade" does even end without any result and should therefore be avoided. The second evaluation criterion is the accuracy error. This checks the quality of the prediction market consistently with Hanson et al. (2006) and is taken as the central figure. The accuracy error is even more important than the amount of trades as the amount does not always lead to a high quality. The accuracy error is defined as variance of the price from the correct value (Hanson et al. 2006, p. 456). The third evaluation criterion is the standard deviation of the price and ensures the reliability as a good average accuracy might still not be sufficient. The existence of extreme faults might be more important than a slight reduction of the average accuracy as extreme predictions might influence decisions negatively. All three figures are measured at the end of the prediction market.

The accuracy error (hypothesis 2) is consequently used as the most important measure of prediction market quality. As the accuracy seems not to depend on the mechanism only, interaction effects with other factors are considered. Besides, the amount of trades and the standard deviation of the price are explored although the results are expected to be more straightforward.

Therewith, the analysis of these two variables should further contribute to model validation.
Regarding the first criterion, the following question will be analyzed: “Which mechanism shows a higher amount of trades?” This question is relevant as only trades can improve the prediction. It is straightforward, that the LMSR has an advantage in achieving a higher amount of trades as the CDA needs at least two traders to execute a trade. Contrary, the LMSR constantly provides liquidity and therewith is able to act as the second trader. This advantage has also been observed empirically by Ken Kittlitz, a software engineer of Consensus Point: “The number of trades in a market using the market maker is at least an order of magnitude higher than in one not using it.” (cited in Hanson 2009, p. 62) Therefore, hypothesis 1 is formulated:
Hypothesis 1: The LMSR leads to a higher amount of trades than the CDA.
The second criterion is analyzed based on the following question: “Which mechanism achieves a better accuracy?” This question is central to prediction markets as the accuracy error measures the quality of the prediction market result. Unlike the first research question, finding a hypothesis for or against one of the mechanisms is more complicated in this case. As the LMSR was introduced after the CDA and specifically tackles its problems, the LMSR might be seen as favorable. However, most of the properties of the mechanisms have two sides. Having a quote-driven market maker (LMSR) might be beneficial, if no trades are expected with very few traders. However, in the absence of no trades and assuming an early end of the market, the traded liquidity in LMSR might be too low to achieve an accurate value. The market maker needs a certain minimum liquidity to move the price. Contrary, the CDA is able to directly change the price with one trade of a single stock. Summarizing, finding a hypothesis is less straightforward compared to the first one. However, considering the increased application due to the enhanced liquidity, the LMSR is supposed to have a slight advantage, leading to our second hypothesis
Hypothesis 2: *The LMSR achieves a better accuracy than the CDA.*
The question linked to the third and last criterion is: “Which mechanism shows a lower standard deviation of the price?” This question is relevant, as it reflects the constancy of a prediction market which might be an important measure for applications. For example, a high probability of a small derivation might be more acceptable than a low probability of a high derivation which results in a disastrous decision. Here, the LMSR seems to have a lower standard deviation as it restricts the action space of traders. Furthermore, the traders constantly only can choose out of exactly three actions, they can accept to buy or sell from or to the market maker or do nothing. Furthermore, the prices of the market maker direct the trading and the liquidity hinders fast price changes. Therefore, extreme trades do not directly result in extreme derivations from the correct value. Summarizing, this leads to our third hypothesis:
Hypothesis 3: *The LMSR has a lower standard deviation of the price than the CDA.*

SIMULATION MODEL

The purpose of the agent-based simulation model is to analyze the effect of the two mechanisms on the number of trades, the accuracy of prediction markets and the standard deviation of the prices. The motivation is to go beyond existing laboratory experiments. In this case, simulation can utilize at least two advantages compared to laboratory experiments (for a more detailed discussion on complementarities of simulation and laboratory experiments see Klingert and Meyer 2012). First, actor strategies can be controlled and therefore varied intentionally. Second, simulation experiments can be executed more efficiently which enables a much broader experimental design including up to 7 factors and 100 simulation runs for each factor combination. Beside its advantages, simulation itself benefits from the strong link to laboratory experiments as outcomes of the default model are validated and strategies are chosen by classification based on experimental data. In this section only an excerpt of the model description and reasoning concerning model design can be given due to space limitations. A more detailed model documentation based on the ODD protocol (Grimm et al. 2010) as well as a comprehensive validation and analysis can be found in Klingert (2012).
As a starting point, the default setting of the simulation model is mostly identical to the laboratory experiment of Hanson et al. (2006, p. 451). Twelve agents are initially endowed with 200 monetary units and 2 stocks. Both stocks give the right to receive an unknown payoff of 0, 40 or 100 with equal probability at the end of the trading period. As private knowledge, the agents know that one value can be excluded with certainty from the possible outcomes. For example, if the true value is 40, half of the agents know it is not 0 and half of the agents know it is not 100. Therefore, an agent which can exclude 0 as true value, knows that the true value is either 40 or 100 and that the stock has an expected value of 70.
Regarding the procedure, there are only few differences to the laboratory experiment of Hanson et al. (2006), e.g., the simulation model is executed in steps instead of a continuous execution. In each step based on CDA, the agents are allowed to place a bid or ask with a limit within the natural numbers between [0, 100]. Alternatively, they can accept an order from the order book. If an offer is accepted, the trade is executed and money and stocks are exchanged. The agent order is determined randomly to align the simulation with the laboratory experiment. The simulation ends after step 60 and then shows a comparable amount of trades (0-41 trades instead of 8-37) as in the laboratory experiment which lasts 5 minutes (Hanson et al. 2006, p. 451). Consistently with other market simulations (Gode and Sunder 1993, p. 122), the agents are limited to trade one

stock per step and the unmatched offers are deleted after a trade to simplify their decision and as the agents are allowed to place the same offers in the next step.
To provide additional contributions, the simulation model goes beyond the laboratory experiment and is varied along its elements, the market institution, the environment and the agents. The market institution is defined by its rules, i.e. the market mechanism. The CDA and the LMSR are used in the simulation experiments. The implementation of the LMSR demands a concrete b-value, which determines the maximum loss of the market maker, for the LMSR. To achieve a comparable result for CDA and LMSR, the b-value of the LMSR should be linked to the CDA setup. It is chosen such as the maximum loss of the market maker and the initial money endowment in LMSR is equal to the maximum value of the initial stocks and the money endowment in CDA.
Finally, the agents trade based on simple rules. The "family" of zero intelligence (ZI) traders is used because these strategies fulfill four criteria. First, the simplicity of the zero intelligence traders allows separating the influence of the market mechanism and the agent strategy. The separation is needed as the analysis should focus on the mechanisms. Second, zero intelligence agents allow cumulative research as they have been used in a variety of simulations. Third, the zero intelligence agents are already validated on the macro level as prior research has recognized similar efficiency of markets with ZI traders compared to markets with human traders (Gode and Sunder 1993, p. 133). Therefore, zero intelligence agents seem to be appropriate to direct subsequent experiments. Fourth, strategies which can be validated on the micro level would be valuable. The zero intelligence strategies have not been micro-validated statistically in their original papers (e.g., Gode and Sunder 1993). However, the simplicity of the zero intelligence strategies allows the statistical micro validation.
As ZI-strategies, ZI.EV, N-ZI, ZIP and two models mixing different strategies are chosen. Fundamental trading is represented by ZI.EV traders. The ZI.EV traders (derivation from ZIC (Gode and Sunder 1993) and N-ZI (Duffy and Ünver 2006)) are selling above and buying below the expected value (exact price is chosen randomly with a uniform distribution as in Gode and Sunder 1993). N-ZI traders (Duffy and Ünver 2006) weight the expected value and the last price to select when to buy and to sell. ZIP traders (Cliff and Bruten 1997) are learning a profit margin and only have a small price span in which they trade. Finally, the Mix 1 (4*ZI.EV, 2*N-ZI) and Mix 2 (2*ZI.EV, 2*N-ZI, 2*ZIU (fully random traders, for details see Gode and Sunder 1993)) are a model with several strategies for different agents.
The strategies (all beside ZIP) have been derived by micro validation similar to Boero et al. (2010). Instead of clustering as in Boero et al. (2010), a classification of experimental actors to the pre-defined ZI-strategies is done based on the data of the laboratory experiment of Hanson et al. (2006). Regarding strategies which are based on one information source only, almost all laboratory actors have been categorized to ZI.EV instead of trading on the trend or last price. After strategies which use more than one information source are added to the classification, the ZI.EV and N-ZI strategy equally fit to the laboratory data. Finally, mixed models have been introduced according to the distribution of these strategies to cover the minimum as well as the maximum amount of traders which are classified as fully random traders (ZIU).
The verification (see Gilbert and Troitzsch 2005, p. 19) of the model relies on three basic procedures: The model has been built based on (semi-) formalized models, it has been tested several times and the code has been inspected with a step-by-step debugging.

EXPERIMENTAL DESIGN, RESULTS AND ROBUSTNESS

In this section, the experimental design and results are described, including tests for robustness. After introducing the simulation results with an exemplary simulation, the results are evaluated along the three evaluation criteria in several steps (cf. Kleijnen et al. 2005; Lorscheid et al. 2012). First, each research hypothesis is evaluated comparing the averages of 100 runs which are based on the default setting. Second, several robustness tests are performed. These include the comparison of the averages over all 3^k-combinations and for each of the 3^k-combinations. Third, an overview of the effect sizes of main and 2-way-interaction effects is given based on a 3^k-experimental design. As effect size measure, partial eta squared is chosen to weight the influence of the factors against the size of the error. Finally, the results are tested for robustness from two further perspectives. They are analyzed in different environments assuming extreme factor levels at the border of all possibilities or assuming different knowledge distributions.

Table 1: Factor levels in experimental design

Scale	Factors	Low	Default	High
Ratio	Steps	30	**60**	90
	Agents	6	**12**	18
	Initial stocks	1	**2**	3
	Initial money	100	**200**	300
Nominal	Mechanism	{**CDA**, LMSR}		
	Strategies	{**ZI.EV**, N-ZI, Mix 1, Mix 2, ZIP}		
	Knowledge distributions	correct value of {0, **40**, 100}		

Table 1 shows the factors which are varied in a 3^k-experimental design. The 3^k- is chosen instead of a 2^k-experimental design. It allows choosing the low and high values linearly around the default value and therewith systemizes the analysis of the results. As the model is mainly validated based on the experiment of Hanson et al. (2006), the default factor levels are chosen accordingly to the experiment as well. For example, the default number of agents is 12 and is able to reflect the corporate context with few active traders.

Exemplary Simulations

To introduce the simulation results, Figure 1 contains one exemplary simulation run for each mechanism. It depicts the prices achieved in both methods from the step 0 up to 100 within the default setup.

Figure 1 shows clear differences between the price development based on CDA and LMSR. While the price changes are small with LMSR (maximum of 3), the price changes with CDA range up to 42. Furthermore, the number of price changes is higher in LMSR (33) than in CDA (10). Therefore, the two exemplary simulations already give a first indication concerning H1 and H3.

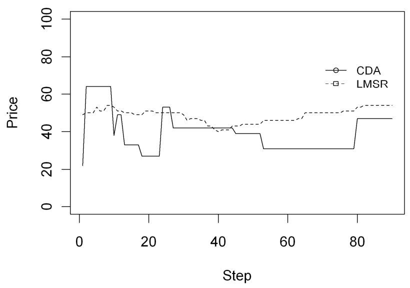

Figure 1: Exemplary simulation based on default model, e.g. a correct value of 40

Amount of trades

Strong support for hypothesis 1 can be found comparing the averages of the default setting, the averages over all 3^k settings and the averages for each 3^k setting in Table 2. Within the default setting, the LMSR achieves on average more than three times the trades than the CDA. Over all 3^k settings, CDA achieves more trades, but still using the LMSR results in nearly twice the amount of trades. It is obvious, that this result is also reflected in the winning settings, i.e. comparing in the 1215 settings the means of 100 runs for each mechanism (1215 times the mean of 100 runs with CDA and 100 runs with LMSR are compared). In 88.07% of the cases the LMSR has more trades. The majority of the cases against the hypothesis are driven by the ZIP-strategy (93 of 243 ZIP settings). Agents applying the ZIP strategy learn a price range. This might be too small to move the given prices of the LMSR, but it can be sufficient to exchange stocks at a constant price over a longer time period. As a reason for the higher amount of trades in LMSR one could argue the fact, that the LMSR steadily provides liquidity and therefore an execution of a trade is always possible. Furthermore, assuming there is just one ask and one bid, the CDA results in a maximum of one trade while the maximum in LMSR is two.

Table 2: Mean and winning settings with independent variable "amount of trades"

	default setting	3^k settings (robustness)	
	Mean trades	Mean trades	Win. settings
CDA	10.70	16.76	107
LMSR	34.95***	32.62***	1108***

Note. Two sided t-test for means and p-value assuming a probability of 50% for winning settings (p < .05. ** p < .01. *** p < .001)*

Summarizing, the effects on the amount of trades are high and the LMSR is superior in achieving a high amount of trades.

To understand the importance of the influence of the market mechanism, an ANOVA is conducted over all 3^k settings. As a result, the main effect "market mechanism" is the third largest effect in an ANOVA including all main and interaction effects. This shows that changing the market mechanism has an important influence on the amount of trades. The agent strategies have an even higher effect size due to the increased amount of trades in the presence of certain traders, e.g., random traders. However, the higher amount of trades due to random traders might not necessarily result in a better accuracy error. Finally, the R^2 of .917 is relatively high. This outlines that the amount of trades can be influenced easily by intentionally adapting the factors, e.g., by choosing LMSR instead of CDA.

Accuracy error

Hypothesis 2 has to be declined, because the expected advantage of LMSR cannot consistently be found in Table 3.

Table 3: Mean and winning settings with fixed factor levels and the independent variable "accuracy error"

		default setting	3^k setting (robustness)	
Fixed values		Mean acc. err.	Mean acc. err.	Win. settings
Default (none)	CDA	224.17	1326.82***	630
	LMSR	110.13***	1361.29	585
Strategy: Mix2	CDA	639.95	1920.93	32
	LMSR	113.90***	1440.95***	211***
Knowl.: CV0	CDA	1321.75***	1732.83***	290***
	LMSR	1792.71	1941.43	115
Knowl.: CV100	CDA	1691.58**	1880.51***	267***
	LMSR	1849.50	2023.47	138
Money: 300	CDA	227.45	1339.36	159
	LMSR	83.35***	1239.65***	246***

Note. Two sided t-test for means and p-value assuming a probability of 50% for winning settings (p < .05. ** p < .01. *** p < .001)*

While the default setting is advantageous for the LMSR, the mean over all settings is slightly documenting an advantage for the CDA. The comparison of the mean accuracy error of LMSR and CDA in all 1215 different settings further underlines the missing clear direction. The difference is not significant assuming each of the mechanisms to be superior with a probability of 50%. This result indicates the need for a more detailed analysis considering the interaction effects.
A selection of interaction effects including the "market mechanism" range upon the most important effects as documented in Table 4. This fact documents the importance of analyzing the interaction effects of the mechanism on the accuracy error as well. However, the knowledge distribution is the most important effect and more than ten times as large as the highest effect size including the "market mechanism". Beyond, the R^2 is lower than regarding the amount of trades. This is caused by two reasons. First, due to the lower R^2 the influence on the accuracy seems to be less straightforward than on the amount of trades. Second, the error is higher compared to the analysis in the prior subsection. As a result, one could argue that the knowledge distribution of Hanson et al. (2006) which has been applied in the simulation is too extreme. Therefore, this main effect can be seen as dominant compared to the other effects, which will be further analyzed in the robustness subsection.

Table 4: The main (on diagonal) and interaction effects with independent variable "accuracy error"

Effect size (accuracy error)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Steps	.003						
(2) Knowledge distributions	.003	.552					
(3) Mechanisms	.003	.020	.001				
(4) Agents	.000	.002	.002	.000			
(5) Initial stocks	.000	.004	.001	.000	.000		
(6) Strategies	.009	.015	.033	.003	.000	.068	
(7) Initial money	.000	.013	.008	.000	.000	.002	.005

Note. Partial η^2 used as effect size; R^2= .589; 100 runs per 2430 combinations; For all effects >.000: $p < .001$.

Considering the three biggest interaction effects to analyze the results in more detail, the mechanisms show differing results. With the existence of some fully random traders (Mix 2 which includes ZIU), the LMSR has an advantage, because it lowers the maximum influence of each trader per trade. The LMSR can profit even more with a higher initial endowment of money (300), as the traders in LMSR need certain liquidity to move the price. Contrary, the CDA is advantageous assuming a knowledge distribution at the border of the possible values (correct value of 0 or 100). One reason is the market maker in LMSR which is partly holding the money while in CDA the cumulative money of all traders remains constant. Therefore, the range of possible actions is more restricted to some individual traders in LMSR than in CDA. This especially fits to extreme values, where more money is needed to move the price to the correct value.
Summarizing, the effects of the mechanisms on the accuracy error are less strong and more diverse than the effects on the amount of trades.

Standard deviation of the price

Strong support can be found for hypothesis 3. The standard deviation of the price in the default setting and over all 3^k settings is lower in LMSR as documented in Table 5. Comparing all settings, this holds for 93.09% of all cases which is the highest value of all comparisons. Here, 84 of 84 cases with an advantage for CDA are out of the 243 ZIP cases. Therefore, the learning capabilities of the ZIP traders seem to lower the standard deviation within the CDA. However, even in the ZIP case 65.43% of the cases have a lower standard deviation with the LMSR. The reason is the ability of the LMSR to reduce high price changes (maximum only 5 instead of 47) by determining the price based on all past trades instead of only the last one.

Table 5: Mean and winning settings with independent variable "standard deviation of price"

	default setting	3^k setting (robustness)	
	Std. dev. price	Std. dev. price	Superior settings
CDA	13.79	16.12	84
LMSR	2.17	6.74	***1131

Note. P-value assuming a probability of 50% for winning settings ($p < .05$. ** $p < .01$. *** $p < .001$)*

Robustness of results with different knowledge and extreme factor levels

Finally, at least two further robustness tests seem desirable. First, a robustness test with different knowledge distributions, i.e. less extreme distributions. Second, a robustness test with extreme factor levels as an alternative to the linearly chosen ratio-scaled factor levels. Here, very low factors levels are chosen, e.g., the amount of traders (2 instead of 6 agents) and the end of trading (5 instead of 30 steps) is lower.
Four alternative knowledge distributions are chosen. The first considers a higher diversity of knowledge among the traders and is adapted from a knowledge distribution used by Smith (1962). In this case, each agent has an expected value which differs to all other agents. The expected value of agent 1 is: 25, 2: 35, 3: 20, 4: 40, 5: 15, 6: 45. Here, 30 is assumed as correct value. The second knowledge distribution is a mirrored one with an expected value of 70. The third (fourth) is

adapted from Oprea et al. (2007) and provides 50% of the agents with an expected value of 20 (80) and 40 (60) with a correct value of 30 (70).

The results in Table 6 provide further support for the findings in the prior subsections. While hypothesis 1 and 3 are again supported, the results regarding hypothesis 2 are again ambiguous. The results regarding the interaction effects find further support as well.

Table 6: Summary of all results including the mean of all settings with alternative knowledge and extreme values

Hypothesis	Default	3^k	3^k (settings)	Alt. Knowl.	Extr. values
LMSR: Amount of trades	+	+	++	+	+
LMSR: Accuracy	+	-	o	+	-
LMSR: Mix 2 strategy	+	+	++	+	NA
CDA: CV 0 / CV 100	+	+	++	NA	NA
LMSR: 300 money	+	+	+	+	NA
LMSR: Std. dev.	+	+	+++	+	+

Note. - contrary to hypothesis; o neutral; + support for hypothesis

DISCUSSION AND CONCLUSION

As expected, the CDA and the LMSR result in different outcomes, e.g., the amount of trades is higher and the standard deviation of the price is lower in LMSR. This is due to the fact that the existence of a market maker increases the amount of trades. Two offers can result in a maximum of one trade in CDA, while 2 trades are possible in LMSR. Therefore, the LMSR shows fewer trades only in some exceptional cases. For example, the price movements allowed by LMSR may be too big to cover all possible offers. This might be the case with ZIP learning agents, which negotiate a very small price corridor. Second, the lower volatility is rooted in the fact that the LMSR considers past trades in the price calculation. High price changes in LMSR are not immediately possible and therefore less probable due to noise trading. The maximum price change in LMSR standard setting is 5 compared to 47 in CDA.

Interestingly, the accuracy error which can be seen as the most important evaluation criteria is influenced in a much less straightforward way. Here, the interaction effects with several factors are more important than the main effect. These interaction effects affect the accuracy error in different directions. Here, an important effect is the interaction effect including the strategy of the agent. Adding random traders to a prediction market can decrease the accuracy and decreases it less with LMSR. Another important interaction effect includes the knowledge distribution. While the LMSR has advantages with the default distribution and a correct value in the middle of the possible outcomes, it has problems with the two extreme outcomes. Liquidity issues of traders are one reason as they can prevent them from further participating in trading. Regarding the accuracy error, the mechanism is much less important than the knowledge distribution. This is straightforward as even a prediction market with the best mechanism still will suffer from bad knowledge. However, in a prediction market it is often much harder to influence the knowledge distribution among the participants than to choose a different mechanism.

Overall, the choice of a certain mechanism does clearly matter when drawing on the "wisdom of the crowds" (Surowiecki 2010) in prediction market and a tradeoff has to be considered when setting up such a market. Based on all simulations, the LMSR seems to be advantageous in most of the cases. Limiting the choices of the traders with LMSR seems to enhance the results. The advantage of the flexibility by CDA is only advantageous in the exceptional case of an extreme knowledge distribution. Therefore, the prediction market owner has the option to choose the CDA which gives the full freedom to the traders risking "no trades". In most of the cases this full flexibility seems to be a disadvantage compared to the steady liquidity and cumulative price building process of the LMSR.

Beyond, further factors can be influenced to enhance the accuracy of prediction markets and might direct further valuable laboratory experiments. With LMSR, the initial amount of money has to be decided as well. On the one hand, the cumulative initial money endowment over all participants should be higher than the maximum loss of the market maker. On the other hand, unlimited money endowment would not allow noise or manipulative traders to be sorted out and therefore should be avoided as well. With CDA, random traders can have a major influence. Therefore, an appropriate training of participants in a prediction market with CDA should be ensured.

Finally, limitations of our research have to be considered. A major limitation is caused by the limits of simulation. Therefore, the simulation results should be further validated by subsequent laboratory and field experiments. Furthermore, the mechanisms are compared under the consideration of constant agent strategies. Choosing the best mechanism should also be based on their understandability. If one of the mechanisms is not fully understood by human traders, the quality of the strategies might be lower. The simplicity as a major learning from market design research is also recognized by Roth (2008, p. 306). The LMSR has a smaller freedom in how to trade as only the market maker is allowed to place offers. Therefore, it is the more simple mechanism. Simplifying the options and strategies possibly leads to a simplified strategy choice. This effect might lead to better strategies and should further be considered in subsequent laboratory and simulation experiments.

ACKNOWLEDGEMENTS

We thank Robin Hanson, Ryan Oprea and David Porter for giving us access to the data of their experiment as well as participants of workshops in Groningen and Koblenz for valuable feedback.

REFERENCES

Boero R, Bravo G, Castellani M, Squazzoni F (2010) Why bother with what others tell you? An experimental data-driven agent-based model. *Journal of Artificial Societies and Social Simulation* 13 (3).

Bray D, Croxson K, Dutton W (2008) Information markets: feasibility and performance. Oxford Internet Institute DPSN Working Paper Series No. 10.

Chen K-Y, Plott C (2002) Information aggregation mechanisms: concept, design and implementation for a sales forecasting problem. Social Science Working Paper 1131. California Institute of Technology, Pasadena.

Cliff D, Bruten J (1997) More than zero intelligence needed for continuous double-auction trading. HP Technical Report. HP Laboratories, Bristol.

Duffy J, Ünver MU (2006) Asset price bubbles and crashes with near-zero-intelligence traders. *Economic Theory* 27 (3):537-563.

Forsythe R, Forrest N, George RN, Jack W (1992) Anatomy of an experimental political stock market. *American Economic Review* 82 (5):1142-1161.

Gilbert N, Troitzsch KG (2005) *Simulation for the social scientist*. 2nd edn. Open University Press, Maidenhead.

Gode DK, Sunder S (1993) Allocative efficiency of markets with zero-intelligence traders: market as a partial substitute for individual rationality. *Journal of Political Economy* 101 (1):119-137.

Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF (2010) The ODD protocol for describing individual-based and agent-based models: a first update. *Ecological Modelling* 221:2760-2768.

Hansen J, Schmidt C, Strobel M (2004) Manipulation in political stock markets - preconditions and evidence. *Applied Economics Letters* 11 (7):459-463.

Hanson R (2003) Combinatorial information market design. *Information Systems Frontiers* 5 (1):107-119.

Hanson R (2009) On market maker functions. *The Journal of Prediction Markets* 3 (1):61-63.

Hanson R, Oprea R, Porter D (2006) Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior & Organization* 60 (4):449-459.

Healy PJ, Linardi S, Lowery JR, Ledyard JO (2010) Prediction markets: alternative mechanisms for complex environments with few traders. *Management Science* 56 (11):1977-1996.

Ivanov A (2008) Electronic prediction markets, Analyx Reference Information.

Kleijnen J, Sanchez S, Lucas T, Cioppa T (2005) A user's guide to the brave new world of designing simulation experiments. *INFORMS Journal on Computing* 17 (3):263-289.

Klingert FMA (2012) *Prediction market mechanisms and manipulation: a multi-agent simulation based on experimental economics*. PhD thesis, Hamburg University of Technology, Hamburg, forthcoming.

Klingert FMA, Meyer M (2012) Effectively combining experimental economics and multi-agent simulation: suggestions for a procedural integration with an example from prediction markets research. *Computational & Mathematical Organization Theory* 18 (1):63-90.

Ledyard J, Hanson R, Ishikida T (2009) An experimental test of combinatorial information markets. *Journal of Economic Behavior & Organization* 69 (2):182-189.

Lorscheid I, Heine B-O, Meyer M (2012) Opening the 'black box' of simulations: increased transparency and effective communication through the systematic design of experiments. *Computational & Mathematical Organization Theory* 18 (1):22-62.

Oprea R, Porter D, Hibbert C, Hanson R, Tila D (2007) Can manipulators mislead market observers? Working Paper.

Othman A, Sandholm T Automated market-making in the large: the gates hillman prediction market. In: *Cambridge, MA*, 2010. ACM, pp 367-376.

Roth AE (2008) What have we learned from market design? *The Economic Journal* 118 (527):285-310.

Smith VL (1962) An experimental study of competitive market behavior. *The Journal of Political Economy* 70 (2):111-137.

Spann M, Skiera B (2003) Internet-based virtual stock markets for business forecasting. *Management Science* 49 (10):1310-1326.

Surowiecki J (2010) *The wisdom of crowds : why the many are smarter than the few*. Abacus, London.

Wolfers J, Zitzewitz E (2004) Prediction markets. *Journal of Economic Perspectives* 18 (2):107-126.

AUTHOR BIOGRAPHIES

FRANK M. A. KLINGERT is a Ph.D. candidate at the Institute of Management Control and Accounting at Hamburg University of Technology. Previously, he was employed in the enterprise performance management area at SAP headquarters in Walldorf. He studied computer science and information systems at the University of Koblenz-Landau in Germany and at the Leiden University in the Netherlands. His research interests focus on the linkages between computer and business sciences, e.g. multi-agent simulation and prediction markets.

MATTHIAS MEYER is professor of management control and accounting and director of the Institute of Management Control and Accounting at Hamburg University of Technology. He has a habilitation degree from WHU-Otto Beisheim School of Management in Koblenz and received 2003 his Ph.D. from the Ludwig-Maximilians-University in Munich/Germany with a dissertation on principal agent theory and methodology.
He studied business administration, economics, philosophy and philosophy of science. His research interests include computer simulation, management control and accounting, institutional economics and methodology.