

Stockholm School of Economics  
Department of Economics  
Bachelor's Thesis  
July 06, 2012

# Using Prediction Markets for Group Probability Elicitation in Risk and Vulnerability Analysis

## **Abstract**

This thesis evaluates an implementation of a novel method to handle uncertainty and foresight in risk and vulnerability analysis. Prediction markets were introduced to address the problems regarding forecasting, and a smaller experiment was set up to test the suitability of prediction markets in this area. The pilot experiment was insufficient in extent to make a conclusive evaluation, still some tendencies are discussed. The theoretical evidence is promising, however questions about the experimental setting remain and in which stage prediction markets are best implemented. The conclusion drawn is that there are no serious obstacles to prediction markets as a promising approach to strengthen the reliability in risk and vulnerability analysis. Nevertheless, how to implement prediction markets is still unclear and a lot of future research on the subject is necessary.

## **Key Words:**

Prediction markets, Foresight, Risk and Vulnerability Analyses

**Authors:** Karin Ehlin Kolk\*

**Tutor:** Nicolas Espinoza and Magnus Johannesson

**Examiner:** Örjan Sjöberg

**Discussant:** Karolina Ling-Vannerus and Björn Wallenberg

**Presentation:** June 1, 2012, 09.15-10, Room 350

\*E-mail: 21205@student.hhs.se

## ***List of Figures***

Figure 1: FOI's Risk and vulnerability analysis	p. 15
Figure 2: Procedure of the Risk and vulnerability analysis	p. 17
Figure 3: Price trend for the contract about the governing coalition	p. 39
Figure 4: Price trend for the contract about the Sweden democrats	p. 40
Figure 5: Price trend for contract about prioritized commitment 1	p. 40
Figure 6: Price trend for contract about prioritized commitment 2	p. 41
Figure 7: Price trend for contract about prioritized commitment 3	p. 41
Figure 8: Price trend for contract about prioritized commitment 4	p. 42
Figure 9: Price trend for contract about prioritized commitment 5	p. 42

## **List of Tables**

Table 1: Activity over time, number of logins and transactions presented per day	p. 25
Table 2: Number of transactions and contracts traded by contract	p. 26
Table 3: Number of transactions and contracts traded per market	p. 26
Table 4: Number of transactions, contracts traded and closing price	p. 27
Table 5: Comparison between outcomes in the prediction markets and the role plays	p. 28
Table 6: The two prediction markets, and the weighted averages and the outcome from the role plays compared to each other	p. 29
Table 7: Results from the Spearman rank order correlation test	p. 29

## ***Table of contents***

1. Introduction .....	4
2. Previous research .....	6
2.1 Foresight and information aggregation .....	6
2.2 Risk and vulnerability analyses .....	7
2.3 Prediction markets .....	9
2.4 Potential areas to implement prediction markets in .....	14
3. Description of the implemented prediction market .....	15
3.1 FOI's risk and vulnerability analysis .....	15
3.2 Experimental design.....	17
3.3 Evaluation of the experiment .....	23
4. Results .....	25
4.1 Market activity .....	25
4.2 Price results .....	27
5. Discussion.....	30
5.1 Interpretation of the outcome .....	30
5.2 Design and function of the markets .....	33
5.3 Advice and lessons to be learned for the future regarding the experimental design ..	35
5.4 Theoretical and practical implications and validity of the study .....	36
5.5 Future research.....	36
6. Conclusion .....	38
7. Appendix .....	39
8. References .....	43

## ***1. Introduction***

The security of society and its inhabitants is dependent on the premise that protective social functions work sufficiently in normal circumstances as well as in crises. A public authority, in this essay defined as divisions of government authorities or cabinet offices, has the duty of setting up and maintaining certain vital social functions and must be able to keep up these abilities even during crises. To secure this, the authority must continuously evaluate that the capacity is in accordance with the demands of the function. One way to make this evaluation is to perform a risk and vulnerability analysis with the purpose to increase the awareness and knowledge about risks, threats and vulnerability in the topical field of activities. The results are relevant for decision makers and people in a responsible position and gives useful information to the planning of preventive measures. Risk and vulnerability analysis is defined in this essay as an estimation of the risk and vulnerability of a system and a valuation of an authority's capability to solve their assignments given certain risk scenarios.<sup>1</sup> To date, as will be shown below, no clear-cut methodology of how to perform risk and vulnerability analyses is established and practices differ.

In risk and vulnerability analyses it is essential to make good predictions about future and unsettled events. To accomplish this, complicated courses of events must be mapped out and educated guesses have to be made. The event mapping process can be very costly and long-running. An approach to improve this process and make it more efficient and faster would be very welcome. This paper addresses an inherent problem in risk and vulnerability analyses, namely how to best aggregate information relevant for the capacity assessment. One approach, prediction markets, might meet the need of information gathering in risk and vulnerability analysis. Prediction markets are future markets that use the information content in current markets to make predictions about specific future events<sup>2</sup>. The aim of this essay is to evaluate an experiment that has used this new method in a risk and vulnerability analysis and discuss if it serves the purpose intended.

Prediction markets have been used in an experiment trying to make more consistent quantitative estimations of an authority's capability to complete their assignments given

---

<sup>1</sup> The valuation of the authority's capability is also called capacity assessment, which is an expression that sometimes will be used in this essay.

<sup>2</sup> For overview of the research, see for example Tziralis and Tatsiopoulos (2007), Wolfers (2006) or Arrow (2008).

certain risk scenarios. The try-out was performed at the Swedish Defense Research Agency's, FOI, which is the Swedish government agency for defense research in August 2010. The pilot markets were constructed with play-money and I participated in the work of conducting the project. This paper will first describe how the experiment was performed to enable evaluation and discussion of the results from the pilot study. Finally, suggestions of improvements of the experimental setting and some discussion about inferences follow.

The thesis is organized as follows. Section 2 will present the relevant background. In section 3 the method of applying prediction markets in capacity assessments will be presented in detail. Section 4 presents the results and section 5 offers a discussion and evaluation of the results and the experimental setting. Lastly, section 6 gives some concluding remarks.

## ***2. Previous research***

### **2.1 Foresight and information aggregation**

Good decision basis and support are essential to making good decisions. This section will introduce information aggregation and point out some of the problems in aggregating information among a group of people. These will then be presented in the context of risk and vulnerability analysis.

Information relevant to many important decisions is often fragmented and distributed among a large number of people. A single person cannot have comprehensive and complete information. Surowiecki argues in *The wisdom of crowds* (2004) that aggregation of information in groups often results in better decisions than could have been made by any single member of the group.

To elicit information from individuals a variety of approaches can be used. The traditional ways to make forecasts and decisions are conventional group meetings and surveys. Sunstein (2004) discusses different ways to obtain information from a group of people: statistical means of individual judgments, various kinds of deliberation and prediction markets. He states that deliberation has two main problems. First of all, members do not always express what they know due to other members' behavior and statements. Secondly, social pressure affects the members as fear of disapproval and fear of being inconvenient may lead to silence. To show one's hidden profile, one provides benefits to others, but at the cost of a possible private detriment. There are even more known problems; for example group think (Janis 1972), cascade effects where the judgments of initial actors are followed by the actors to come who do not disclose what they know and group polarization, the phenomenon when members of a deliberating group move toward a more extreme point in the members' predeliberation tendency (Sunstein 2002) are remaining problems for prevalent methods of group deliberation. There is also an aggregation problem; it is not clear how to weight opinions. Furthermore, it is hard to know how to allocate resources when the outcomes are binary or discrete.

Other techniques for elicitation of information are nominal group techniques, where the members first make individual assessments, then have discussions and then make final

individual estimations (Graefe and Scott 2011). Delphi is a method consisting of a multiple-round survey in which participants anonymously reveal their individual estimates as well as comments on a problem. The process is iterated and after each round comments and estimations are summarized and reported as feedback. In the next round, the participants can take into account the information and provide a new estimate. The final result is the aggregated outcome of the final round's individual estimations. It mixes anonymity, feedback, iteration and a statistical aggregation (Rowe and Wright 1999). One drawback is that the method is time-consuming and that participants' positions and estimations cannot be restated.

As shown above conventional forms of group deliberation have several weaknesses and an approach to improve the information aggregation in some areas would be very welcome.

## **2.2 Risk and vulnerability analyses**

Risk and vulnerability can be analyzed in a number of ways and different conceptions are used. This thesis will use the term risk and vulnerability analysis. What it embraces will be presented below.

Risk and vulnerability analyses assess the risks and weaknesses in a system. The purpose is to increase the awareness and knowledge of decision makers and people in a responsible position about risks, threats and vulnerability in the authority's field of activities, and to give useful information to the planning of preventive measures. The risk and vulnerability analyses function not only as a document in the interests of the own authority, but also contribute to society's accumulated assessment of vulnerability. No one can know in advance where or when serious accidents might happen, but society has a responsibility to have the capacity and readiness to handle and prevent them. In order to be able to do this, it is important to have methods to systematically identify undesired events, the consequences and the capacity to handle them. There are several methods to conduct risk and vulnerability analyses (Johansson and Jönsson, 2007).

It is essential in risk and vulnerability analyses to anticipate future and unsettled events. To achieve this, complicated courses of events must be mapped out based on educated guesses. Also, the probability of those events must be calculated. As presented above, there are many

known challenges to seeing into the future. Different methods use different techniques but they have in common that the forecasting part of the analysis has obstacles.

Risk and vulnerability analyses deal with conceivable scenarios, explicit or implicit, that may occur in the future. Broadly speaking, the risk and vulnerability of a system is the answer to the following three questions:

- What can happen? (In other words; what can go wrong?)
- How likely is it that the identified events will happen?
- If any of the identified events does happen, what are the consequences?

(Johansson and Jönsson, 2007, p.25)

The answers to these questions form scenarios, with a probability that the scenario will happen and a measure of the damage if it does happen. When all scenarios are included, the questions listed above are answered and the risk estimated (Kaplan and Garrick 1981). This thesis will focus on the second question; how likely is it that identified events will happen?

Vulnerability, or the propensity to experience harm during a crisis, will in this thesis be seen as a relationship between a system and a source of risk or event. Vulnerability according to this definition must be seen as related to sources of risk and not decoupled and as a feature in itself (Johansson and Jönsson, 2007).

There is, as previously mentioned, no consensus on how to perform risk and vulnerability analyses. Different methods have various merits and demerits and fulfill different purposes. Some methods include both risk and vulnerability, while others are developed for specific systems or stresses. Traditional methods for analyzing risks are system based, for example the “What if”-method, fault tree analysis and a preliminary risk analysis. Scenario based methods study one or more risk scenario and are based on group discussions. The results of the risk analysis give a description of different risk scenarios. Risk and/or vulnerability analyses can be both quantitative and qualitative. Other methods, models and tools are capacity assessments, RIB and dependence analysis (Johansson and Jönsson, 2007). Known problems with the existing methods are, among other things: problems with covering all conceivable consequences and vulnerabilities, problems to estimate uncertainties for different scenarios at

risk, and difficulties in providing accurate estimations of the consequences of the different scenarios at risk (Johansson and Jönsson, 2007).

This thesis will evaluate a try-out with prediction markets as an approach to solve problems with existing methods regarding estimations of uncertainties for and difficulties in providing accurate estimations of the consequences of the different scenarios of risk. A scenario based method with risk scenarios is used and prediction markets implemented in the phase with group discussions. The method used will be further described below in section 3.1.

## **2.3 Prediction markets**

The prediction market shows potential and is a relatively new kind of forecasting mechanism (the use of which is not entirely uncontroversial<sup>3</sup>). Prediction markets are promising, not only because of the accuracy but also since they solve a great many known problems other methods fail to solve. The goal is to offer effective and accurate means of aggregation of domain-relevant information which could significantly improve the quality and cost-effectiveness of government forecasts and evaluations in the field of capacity assessment. The key realization is that otherwise difficult, time consuming, and imprecise capacity assessments can be made rather easily, quickly, and with great precision using so called prediction markets. Below the concept and characteristics of prediction markets will be further described.

### **2.3.1 The concept of prediction markets**

Prediction markets are markets created for the purpose to make predictions. Prediction markets work like fictive stock markets where expectations from future events are the contracts traded. The current prices on the market can be interpreted as either a prediction of a probability of the event or the expected value of a parameter. In a prediction market, pay offs are linked to outcome in future events to elicit the market's expectations (Wolfer & Zitzewitz, 2004).

Prediction markets have several advantages over group deliberation. Instead of a deliberation process which often amplifies individual errors, prediction markets often succeed in

---

<sup>3</sup> In 2003, a group of analysts at the Department of Defense at the Defense Advanced Research Projects Agency (DARPA) wanted to use prediction market to let ordinary people bet on macro events, e.g. economic development, likelihood of terrorist attacks and defence policy actions. The policy analysis market resulted in a lot of criticism and the plan was abstained.

correcting them. That is because individuals act individually and are not affected by social pressure. They also have the opportunity to react on the information revealed by others in the transactions. Another advantage is that well-constructed prediction markets create incentives to reveal and not withhold information. That is because the best strategy is to play on the outcome one finds most likely, and hence it is rational to use any information, private as well as well known. The existence of benefits (money, play-money or the like), creates incentive to reveal private information. The market price will be the best predictor of an event in efficient prediction markets. This holds even when not all individuals in a market are rational, only the marginal trade in the market must be motivated by rational traders. Prediction markets and the imbedded incentive mechanism can motivate participants to reveal their true beliefs. This is true because participation is beneficial when not agreeing with the current market forecast, and hence tricks participants to challenge the group opinion and look for superior information. This course of action does not take place in other methods of group deliberation (Graefe et al., 2009).

Prediction markets have been used in several areas, the most well-known is the Iowa Electronic Market (IEM) which has been in operation since 1988. The IEM has proved to be the most accurate way of tipping elections (Berg et al., 2008). A large number of studies verify the forecasting performance of information markets with examples from the fields of politics (Berg et al., 2008; Forsythe et al., 1998), sports (Debnath et al., 2003; Deschamps et al., 2007; Luckner et al., 2008), business (Spann and Skiera, 2003; Chen and Plott, 2002), medicine (Polgreen et al., 2007), and entertainment (Pennock et al., 2001). Prediction markets can potentially be used in additional fields. According to Abramowicz, 2007; Sunstein, 2006; Hahn and Tetlock, 2006 to mention a few, prediction markets have the potential to improve decisions in a wide array of applications.

### **2.3.2 The functioning of prediction markets**

Markets can be constructed in several ways. A common structure is a double auction. Double auctions are processes where potential buyers and sellers simultaneously submit bids and ask prices to a third part, the auctioneer, whom chooses a price,  $p$ , which clears the market. Then all buyers who bid more than  $p$  and all sellers who asked less than  $p$  buy at price  $p$ . The market could be constructed with or without a market maker, which is a financial actor that ensures liquidity on the market. The market maker is the actor that purchases/sells the stocks, even if there is no seller/buyer at the moment.

Hanson (2003) has developed another structure, so called Market Scoring Rules. Market scoring rules is a different scoring mechanism which has the advantage of combining the betting markets' good consensus estimates from groups with scoring rules good probability elicitation from individuals. Markets scoring rules avoid the problems from each market, by becoming market makers in "thick" markets and simple scoring rules in "thin" markets. Thick markets are markets with a high number of buyers and sellers and thin the opposite. In thin markets, fewer transactions take place and prices are often more volatile and assets are less liquid. Hanson also introduced a logarithmic market scoring rule which can help participants to simultaneously predict many combinations of outcomes. The contingent contracts work as follows: the market estimates the joint probability of choice A and outcome B. Then information can be elicited about which choice that maximizes the probability that event will occur. Care must however be taken when interpreting the outcome, the markets can underline the correlation between events but the difficulty of inferring causation between the events remain. Another type of markets are dynamic Pari-mutuel which is that all of the money that is traded goes into a common pot and then the winners share the pot (Wolfer & Zitzewitz, 2004).

According to Wolfer and Zitzewitz (2004) three types of contracts are most common in prediction markets; winner-take-all, index and spread. That means that contracts can link payoffs to occurrence of a specific event, be constructed to link payoffs to a continuous variable or be a combination of those two other types (spread betting is an example of this). The different types of contracts reveal different parameters, either the probability that an event occurs, the mean value of the outcome or the median value.

The contracts are exemplified below to explain the mechanisms. An event,  $y$ , and risk neutrality are assumed. The different types of contracts are exemplified with specified numbers below, but any number can be used.

#### *Winner-take-all*

The contract pays 1 if and only if event  $y$  occurs, 0 otherwise. The rational strategy is to bid on the probability  $p(y)$  that event  $y$  occurs. The value is interpreted as the probability the

event occur. The event could for example be outbreak of a specified infectious disease before a pre-determined date.

### *Index*

The contract pays 1 according to a ration on an index, e.g. every percentage a candidate gets in an election. The contracts pays  $y$  and the market expectation is interpreted as the mean value  $E(y)$  of outcome  $y$ , in other words the mean value of the market's expectations of a continuous number range.

### *Spread*

In spread the amount of bet is fixed and the trading is based on cut-offs that determine whether an event occurs. There is an actual outcome  $y$ , a level to yield outcome  $y^*$ , the contract costs 1 and pays 2 if  $y > y^*$ , nothing otherwise. The strategy is to bid according to the value of  $y^*$  and the market expectation of this contract will be the median value of  $y$ , in other words the market's median expected probability that an event will occur. One example of this contract could be if a candidate in an election receives more than 50 per cent of the popular vote.

Servan-Schreiber et al. (2004) come to the conclusion that play-money markets perform as well as real-money markets. They speculate that this is the result from two opposing forces: the incentives for information discovery may be better in real-money markets, but the play-money markets may yield more efficient information aggregation. Instead, they state that a more important ingredient seems to be a motivated and knowledgeable community of traders. Money and intrinsic motivators are tools to attract such traders.

### **2.3.3 Advantages with prediction markets in risk and vulnerability analysis**

Berg and Rietz (2003) showed that prediction markets can be suitable as decision support when two criteria are fulfilled: The first criterion is that optimal decision must critically depend on forecasting outcomes and conditional outcomes that are likely to result from the decision. The second criterion is that markets are more efficient in aggregating diverse information than existing methods. The first criterion is in accordance with risk and vulnerability analyses; to allocate resources in the best possible way assumptions and forecasts must be accurate. There is also reason to believe that the markets are more efficient in aggregating diverse information than existing methods and, if implemented correctly, the

optimal decision would be depending on forecasting outcomes. There are several reasons to question the prevalent results from the role play due to weaknesses with the method, and prediction markets performed in a good way could be one approach to aggregate more accurate information. Arguments in favor of this will be presented below.

Prediction markets possess several characteristics that appear to be beneficial to improve risk and vulnerability analyses. As shown above, prediction markets have been an accurate way to perform foresight. A feature of prediction markets is to work as a meta-forecasting tool (all available data is used; historical, forecasts, expectations to mention a few). Prediction markets incite participants to actively think about the future and seek and produce the best possible information (Graefe et al., 2009).

Furthermore, prediction markets offer continuous and real-time information aggregation. Due to the price mechanism that automatically aggregates all available information, it is always a reflection of the group's aggregated opinion at any time. If it is possible to maintain motivation of participants, prediction markets can continuously reflect the groups' opinions without any additional workload.

Information markets have also shown advantages over other group deliberation processes regarding the problem that members do not always express what they know due to other members' behavior and statements, and the social pressure that affects the participants if they fear disapproval or being inconvenient when expressing their views. This is because prediction markets are anonymous and the market mechanism creates incentive to reveal information, which for example Green et al. (2007) show.

Prediction markets may also reduce the cost and time requirements. This because prediction markets are easy to participate in, can involve many participants and are flexible to use when and where suitable. There are no requirements for all participants to be in the same place at the same time. Prediction markets may be effective and accurate means of aggregation of domain-relevant information, which will improve quality and cost-effectiveness of government forecasts and evaluations.

According to Berg and Rietz' (2003) criteria, prediction markets ought to be suitable as decision support in risk and vulnerability analysis.

## **2.4 Potential areas to implement prediction markets in**

Potential areas to implement prediction markets in risk and vulnerability analysis could be in deciding:

- the likelihood of accidents or outbreaks of communicable diseases
- the likelihood of terrorism
- macroeconomic factors such as inflation, bond yields and other factors
- conditional probabilities, meaning the probability of A happening given that event B occurs. One example of the last could be a terrorist attack - would the authority be able to solve its assignments given the attack?

All of the above described areas except the last one are simple to link connect to an outcome. One example could be, "the contract will yield dividend D if the event occurs before X" (under some pre-determined circumstances). Most of the contracts above are low probability events so the prices traded will be very low. This could be a problem because it might be hard to estimate low probabilities. To decide whether a specified scenario will happen with a probability of  $p=0.001$  or  $p=0.0001$  could be challenging. Snowberg and Wolfers (2005) have found a tendency that low probability events were over-priced which they call favorite-longshot bias. Still, the information elicited via the prediction markets would be possible to rank and give information about the participants' view which events that are most likely, and hence give information about resource allocation even if the low probabilities and price levels in themselves are hard to evaluate. This issue will be discussed further below.

The last potential area mentioned is a conditional probability, which means that the conditional probability of A given B is the probability of A occurring if B happens. This is denoted  $P(A|B)$ . The contracts concerning prioritized commitments used in the experiment described below were of this kind. The use of conditional probabilities in prediction markets is not a new idea; they have been used in prediction markets before. Wolfers and Zitzewitz (2004) describe one example when markets traded contracts about a re-election of President Bush, about that Osama bin Laden would be captured before the election and a combined contract when both events happened.

### 3. Description of the implemented prediction market

#### 3.1 FOI's risk and vulnerability analysis<sup>4</sup>

The risk and vulnerability method that FOI has developed has the following components:

*Risk identification*, which is mainly about defining scenarios. This is because it is often hard to identify sources of risk without thinking in terms of scenarios of risk.

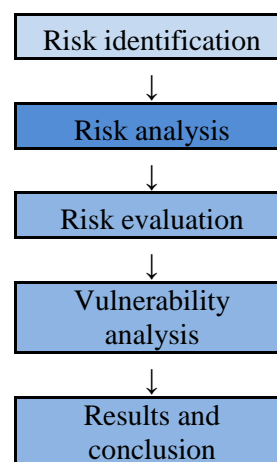
*Risk analysis*, an analysis that tries to refine the different scenarios and how probable they are. Thereafter an analysis of the consequences follows.

*Risk evaluation*, an evaluation to analyze if the level of risk is acceptable or not.

*Vulnerability analysis*, analyses how serious and far-reaching a specific event is. The vulnerability analysis is made with respect to a specific risk scenario that is identified in the analysis. The vulnerability analysis is also more detailed.

*Results and conclusions*, where the results are presented and discussed to form conclusions from the analyses.

**Figure 1:**  
FOI's Risk and  
vulnerability  
analysis



In the risk analysis, the two following components are included:

##### *Judgment of probabilities*

The judgment can be done in several ways. One common way is to use probability calculus, other to use different scales. The judgments could be both quantitative and qualitative.

##### *Judgment of consequences*

Judgment of what can happen and how probable it is, the consequences are important for the decision-making. This is either quantitative or qualitative depending on the method used.

#### 3.1.1 Basic concepts

In the method that FOI uses, three concepts are important; prioritized commitment, vital resource and undesired event. A free translation of the concepts that FOI use follows:

<sup>4</sup> For more information about the method of FOI than described below, see (in Swedish) Eriksson et al (2011).

### **Prioritized commitment**

A prioritized commitment is the organization's necessary qualities to avoid unacceptable consequences. An example is to supply somebody with a service, to maintain the leadership in operation, safeguard the health and security of specific groups or protect economic, social, historical or ecological values.

The prioritized commitments are judged on a four level scale;

*Satisfactory (reaches minimum limit), Satisfactory with some deficiency, Insufficient, Very insufficient.*

The judgment *Satisfactory* means that there are enough resources and capacity to meet the vital tasks society must achieve during crisis.

The capacity *Satisfactory with some deficiency* means that the public services to a certain degree are set aside to prioritize more urgent activities.

*Insufficient* means that the resources within the authority are substantially lower than what is needed to solve public services during crisis.

*Very insufficient* means that the activities within the authority's scope of practice are almost unprepared for any crisis.

### **Vital resource**

A vital resource is something that is necessary to maintain for one or more of the prioritized commitments. That may involve physical resources such as personnel, premises, technical systems, infrastructure or different kinds of products. It may also entail more intangible resources such as services, drilled routines and information. A vital resource can, depending on if it is a part of the organization or not, be internal or external.

### **Undesired event**

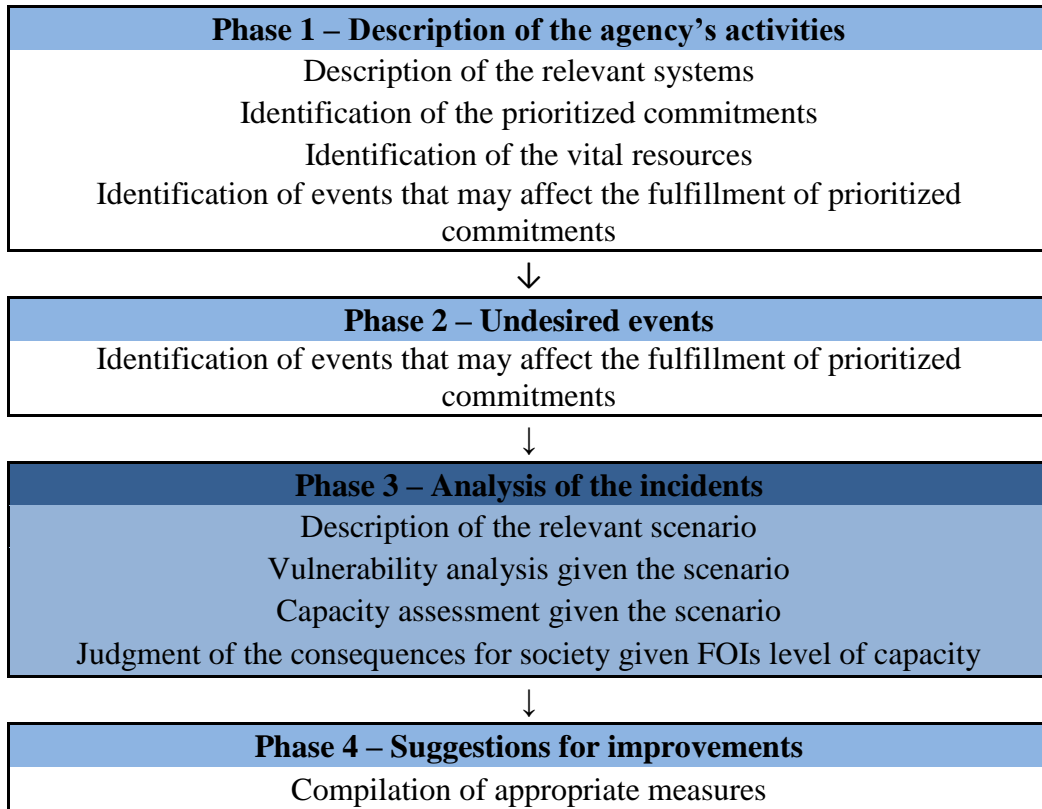
An undesired event is a course of events that has negative consequences from which it is worth protecting and preserving by obstructing or entirely eliminating the potential to maintain at least one prioritized commitment.

FOI uses role plays to outline scenarios and estimate the probabilities of them to occur.

### 3.1.2. Procedure

In more specific terms the risk and vulnerability analysis is performed as described below:

**Figure 2: Procedure of the risk and vulnerability analysis**



Prediction markets were implemented in phase 3 in the phase where role plays in ordinary cases are used. FOI uses the role playing to make predictions and judgments of different scenarios and alternative courses for action and the process is described below. Instead of the role plays prediction markets were performed to elicit the capacity. The role plays were however still necessary since they were used for the evaluation and to decide the outcome of the contracts traded on the market.

### 3.2 Experimental design

This section will show how the prediction markets in risk and vulnerability analyses were implemented. The experimental setting that was performed was just a mock-up study and was performed between August 16 and August 27 2010. The study was made with the methodology used by FOI and they also had a yearly project which was well suited for implementation of the prediction markets. Due to confidential aspects of FOI's work the scenario and commitments are not described below.

Prediction markets had not been previously used in risk and vulnerability analysis. Hence it was not possible to replicate past experiments. Therefore, the experimental setting FOI used was untested prior to the study. In short, the experiment was designed “to replace” the conventional role play in the analysis and to estimate the capacity in different scenarios of risk for the organization. Simultaneously, role plays were performed. (They were later also used to evaluate the prediction markets. Hence was it not possible to remove them in this experimental setting. This will be further examined in the discussion section.) The two prediction markets implemented were playing on the outcome of the role play. The purpose was to test if the two prediction markets arrived at the same results as the role playing. Below the experimental setting will be presented further.

### **Outline of the experimental setting**

The idea of the experiment was to elicit expectations from the staff about FOI’s ability to uphold certain functions during a specified scenario. In ordinary cases this is done by role playing but in this experiment the process used was prediction markets. It was specified in the scenario what exactly happened and how FOI was affected. To simplify the understanding of how the market worked, it will here be exemplified with a scenario that a hurricane shut off the electricity. FOI would still have to supply some services to society; one of them is to conduct defense analysis. To secure that the authority upholds this ability, a yearly risk and vulnerability analysis is performed. The analysis evaluates the ability of FOI to live up to its commitments. One part of this is role plays with superiors; the aim with them is to evaluate the ability to uphold certain commitments. One scenario is tested on several commitments. From the role plays the ability to uphold a number of commitments is estimated. The idea with the experiment described below is instead to let the staff estimate the authority’s ability.

Two prediction markets with the aim to elicit the staff’s expectations were set up. Participants were recruited, allocated on one of the markets and informed about the markets, the trading period and what they were playing on. There were seven identical contracts traded on each of the markets that paid 100 if the event happened and 0 if it did not. The trading contracts on the market were formulated as “Given a hurricane, FOI has a fair potential to still be able to make defense analyses” and the outcomes in the role plays functioned as measure points. Each participant had a portfolio with assets and was able to buy and sell the contracts. The participant had reason to sell if he or she thought the price was below what he or she thought was the probability of the event and otherwise to buy. Participants earned money when they

bought contracts cheap and sold them expensive and lost money during opposite conditions. They also earned money in the end of the game when the market closed and some of the contracts yielded dividend. The trade went on until the market closed, and then the portfolios of the participants froze and the contracts yielded dividend. The closing prices of the market were interpreted as market's expectation of the contract claim.

### **Platform**

The experiments were conducted on platforms developed by Redict, a Swedish company that works with identifying and monitoring risk. The platform consisted of a trading site with play-money markets. All participants had user accounts and were able to login from any internet connection and location. An easily accessible online platform was a condition and user-friendliness of the platform was necessary to uphold activity on the markets.

The experiment had to be performed with respect to the Swedish Defense Research Agency's, FOI's, operations and could not be conducted optimally in all respects since allowances had to be made for confidential information. Therefore, some information about the contracts and the judgment were sent out to the participants and not directly available at the platform.

### **Participants**

A total of 92 individuals (most men) participated in the experiment. All participants were employees at FOI. Participation in the prediction market was voluntary, and an offer to participate was sent by e-mail to all employees some weeks before the experiment started. All employees at FOI were encouraged to participate; both researchers and administrative personnel. The letter of invitation urged the participants to contribute at least once a day. The user accounts were distributed to ensure that every person had just one user account and was not able to manipulate the market through being in possession of many accounts.

### **Claim definition**

All contracts in the study were of a binary structure and yielded 100 or nothing. The play-money currency RED was used. A total of nine contracts were played in the market, of which seven were conditional probabilities events and concerned the capacity assessment and two the Swedish election September 19 2010 (one about whether the governing coalition would be furthered entrusted to govern the country and one about the xenophobic party, the Sweden democrats, would become members of the parliament). Altogether, one event was tested combined with seven different prioritized commitments. Two contracts later were cancelled

due to that no role plays on those commitments were performed. The assessment capacity contracts were formulated as “Given event x, FOI has a fair potential to reach the prioritized commitment  $PC_i$ ”. How the contracts were resolved and judged was specified in the market. For the conventional contracts, the outcome in the election resolved the contracts. The conditional probabilities events contracts were linked to the outcome in the role plays. For all contracts, the result was interpreted as the market’s expectations were ‘pass’ if a price was of 50 RED or higher and ‘fail’ if the price was below 50 RED.

The reason for including contracts other than the contracts concerning capacity assessment was to stimulate activity in the markets and to be able to compare the markets in this study to previous studies, most of which have been conducted on election results.

All the capacity assessment contracts were assigned an entry price of 75. This was partly due to the definition: “a prioritized commitment is the organization’s necessary qualities to avoid unacceptable consequences...” This ought to be fulfilled in most cases, even in unexpected and critical situations. The more conventional contracts were assigned an entry price of 50. An assumption was made that both outcomes of the event were equally possible; another possibility would have been to base entry prizes on betting sites bets.

### **Role play**

The role plays were based on the methods developed by FOI. The processes were performed with representatives of top management. Five scenarios were accomplished and analyzed with the methodology FOI uses. Two of the scenarios were tested in two rounds. The evaluation consisted of four levels and are described above in section 2.3.1; *Satisfactory (reaches minimum limit)*, *Satisfactory with some deficiency*, *Insufficient*, *Very insufficient*.

Since binary contracts were used, the two first-mentioned grades were classified and evaluated as pass and the two last-mentioned as failure. How the role plays were performed and how the results were judged and interpreted was communicated to the participants.

### **Trading mechanisms**

Knowing the scarcity of participants and with the aim to enable traders to buy and sell all the time, a decision was made to construct the market with an automatic market maker. A market maker is a financial actor that ensures liquidity on the market. The market maker is the actor

that purchases/sells the stocks, even if there is no seller/buyer at the moment. In this way, the liquidity on the market is guaranteed. The market maker did not make any profit (hence no transaction cost), and the price was the same on ask and bid prices. In the markets in this thesis, there was an insufficient amount of traders and buyers to be able to maintain sufficient activity and liquidity without the market maker. The market would have been too slow without it.

A linear scoring mechanism was used. Another type of scoring mechanism was unfeasible due to budget limitations. The scoring mechanism functioned as every traded contract affected the price with a very small constant value. The price of the next contract is hence the price plus the small constant value.

When trading three contracts:

Price of contract 1:  $P$

Price of contract 2:  $P + c$

Price of contract 3:  $(P + c) + c$

Average price:  $(3P + 3c)/3 = P + c$

The average price the actor had to pay/receive was not the amount stated on the platform but the price for all contracts plus the change divided by the number of contracts in the transaction. The value of the change was linear with an elasticity constraint  $c$  and was set so the traders were able to affect the price but not with great amounts.  $c$  was set to an arbitrary number. Redict, the company that delivered the market, said this arbitrary number of the elasticity constraint had worked well before. If the elasticity constraint is set too low, it will not move the market enough and set to high it will change the markets too much. Falsely set, it would not be able to reflect aggregate opinions.

All participants received an amount of play-money (10 000 RED) from the start along with 25 contracts of each contract type. The reason for this was to guarantee liquidity and to encourage people to be active and sell assets in the beginning of the test period. The participants were also able to go short, i. e. selling assets that have been borrowed from the market maker with the intention to buy the same contracts at a later date and return them to the lender and gain profits from a decrease. Going short is the invert of going long, i.e. the

conventional practice when one trades and gains profits from an increase. Some participants joined the experiment at a later date and also received the same amount of play-money and contracts.

### **Procedure**

A few days before the experiment started, all participants received information about the platform, the contracts and the user accounts and also a short introduction to prediction markets. All participants were given an opportunity to participate in a test market the week before the real market started with contracts similar to those in the actual market.

Two identical prediction markets were created on the platform and the participants were allocated evenly to the two markets. The reason for this was to test small groups and if the number of participants were satisfactory. The two markets also made it possible to compare them with each other. If they reached the same price level it would lend support to the theory that the accuracy of the market is not dependent on a vast number of participants. The resolutions for seven of the contracts were set by outcomes of the role plays (as previously mentioned, two were later cancelled) and for two by the outcome of the Swedish election. The participants were not told exactly when the market closed, only that the closing time was sometime between August 26 and August 28. The market closed on the morning August 27. The reason not to tell the exact closing time was to avoid volatility and speculation near the closing of the markets. When the outcomes of the contracts were declared, all the contracts that passed yielded dividend. Then the final result was presented and prizes were handed out.

### **Incentives and reminders**

It is crucial to obtain enough activity in the markets. Therefore, several extrinsic (driven by outside factors as rewards, coercion and threats of punishment) and intrinsic (driven by interest or enjoyment in the task itself) mechanisms were built-in to increase activity in the market. Several reminders were also sent out to all participants during the experiment.

The participation in the prediction market was optional for all employees at FOI. In the recruitment of participants, the e-mail sent emphasized that the participants contributed to the quality of capacity assessment and the method development at FOI as well as the opportunity to win cinema tickets.

One strategy was to make the markets visible, and remind of and arouse curiosity about the experiment by updating the top list on the front page of FOI's webpage. The five individual portfolios in both markets with the highest value were presented and the top-five list was updated every day. The three participants with highest portfolio values in each of the markets received prizes when the market ended. The prizes were 15 cinema tickets to the winner, 10 cinema tickets to the person in second place and 5 to the person on in third place.

There was also a prize, one cinema ticket, raffled to one of the persons who had been logged on the market every day. It was constructed not to be an incentive to press on activity on the market (so it did not matter if the person made any transactions or not), but to encourage that people were active and logged in and checked the latest prices. Otherwise, it would have affected the results if people had other incentives than to maximize the portfolio value.

### **3.3 Evaluation of the experiment**

To be able to evaluate the overall performance of the prediction markets, two markets were performed and two conventional contracts were included to give an indication how well the market worked together with the seven (later five) other contracts. Two contracts dropping off affected unfortunately the experiment considerably and badly.

The experiment was conducted in a small setting, which makes it hard to use statistical methods to evaluate it. A non-parametrical test was however used to compare the results between the markets and between the market and the role plays will be used. This is suitable given the small numbers of observations and that no reliable assumptions about the samples can be made. The Spearman Rank Correlation test will get a hint of the linear association between the markets. Spearman rank test evaluates the sample correlation and if it is possible to reject no association.

A comparison between the prices and trade of the conventional contracts will be made when the results are presented. Data over the activity of the market will be presented, since this is crucial for the market to function in a good way. The results from the Spearman rank correlation will be presented and examined. The discussion will include an assessment of the overall performance of the experimental setting and some suggested improvements will be presented. Also a few words will be said about the appropriateness of using prediction

markets in risk and vulnerability analysis and if other phases of the process would be more suitable to implement them in.

## 4. Results

The results presented here form the basis of the following discussion. First some statistics about the activity in the market are shown and later in the report the evaluation of the results follows. More statistics are found in the appendix.

The notation of the contracts will be *PC*, notation of prioritized commitment, notation of market. For the other contracts, the number after the letters shows which market the contract was traded in. Gov is the notation of the contract about whether the governing coalition would be furthered entrusted to govern the country and SD is the notation of the contract about if the Sweden democrats would become members of the parliament.

### 4.1 Market activity

In total, 92 people signed up for the experiments, but only 72 people logged in at least one time. The markets were running for just over ten days.

**Table 1: Activity over time, number of logins and transactions presented per day**

	Logins	Transactions
<b>Day 1</b>	171	599
<b>Day 2</b>	222	404
<b>Day 3</b>	161	979
<b>Day 4</b>	133	401
<b>Day 5</b>	120	316
<b>Day 6</b>	7	5
<b>Day 7</b>	11	19
<b>Day 8</b>	106	205
<b>Day 9</b>	91	230
<b>Day 10</b>	81	199
<b>Day 11</b>	93	325
<b>Day 12</b>	139	0

The activity fell over the time period. The high number of logins day 12 was probably to check the results, an e-mail was sent out telling that the trade was over. The market closed at 7 a.m. which is the reasons that no transactions were registered.

**Table 2: Number of transactions and contracts traded, presented by contract**

	<b>Transactions</b>	<b>Contracts traded</b>
<b>SD 1</b>	502	24810
<b>SD 2</b>	393	21795
<b>Gov 1</b>	377	14194
<b>PC 2.2</b>	279	10955
<b>PC 4.1</b>	275	10768
<b>Gov 2</b>	272	10157
<b>PC 3.1</b>	263	8446
<b>PC 5.1</b>	224	7474
<b>PC 4.2</b>	222	6939
<b>PC 2.1</b>	214	6570
<b>PC 3.2</b>	207	5313
<b>PC 1.1</b>	205	5602
<b>PC 1.2</b>	196	5989
<b>PC 5.2</b>	161	3941

The total number of transactions were 3278, average value of transactions per participant 40.5 and median value 25. The average value of traded contract per transaction was 36.8; the sum differed between 24.5 to 55.5 contracts per transaction depending on contract. The contracts were traded with varying frequency. Some contracts were traded in greater volumes per transaction than others. The conventional contracts all showed very high trade levels. In the appendix you can find figures of the price trend for each of the contracts.

**Table 3: Number of transactions and contracts traded per market**

	<b>Transactions</b>	<b>Contracts traded</b>
<b>Market 1</b>	2343	86782
<b>Market 2</b>	1989	75246

There was less activity in one of the markets; this was however almost counterbalanced by the fact that the volumes of contracts per transaction were higher when traded in the other market. To investigate if there was a relationship between the prices and how traded the contracts were a collocation of number of transactions, contracts traded and closing price put together. In the table below, the three highest numbers in every column are marked with dark grey and the three lowest with lighter grey. All contracts except one of the contracts about the Sweden

democrats have higher prices than the entry prices. It is hard to discern any trends, but the conventional contracts were more popular than the others.

**Table 4: Number of transactions, contracts traded and closing price**

	Transactions	Contracts	Closing price
<b>PC 1.1</b>	205	5602	79,1
<b>PC 1.2</b>	196	5989	79,45
<b>PC 2.1</b>	214	6570	82,15
<b>PC 2.2</b>	279	10955	80,05
<b>PC 3.1</b>	263	8446	77,9
<b>PC 3.2</b>	207	5313	78,1
<b>PC 4.1</b>	275	10768	82,5
<b>PC 4.2</b>	222	6939	75,25
<b>PC 5.1</b>	224	7474	86,45
<b>PC 5.2</b>	161	3941	75,5
<b>Gov 1</b>	377	14194	79,2
<b>Gov 2</b>	272	10157	81,55
<b>SD 1</b>	502	24810	33,5
<b>SD 2</b>	393	21795	58,55

**4.2 Price results**

For some of the contracts the prices congregated under the test period, but overall it is hard to generalize the price trend. Several of the contracts showed higher volatility in the end of the period, statistics about the price trends are found in the appendix. The weighted averages with respect to number of transactions for each PC on the two markets showed below were calculated to combine the results. The weighted averaged is compared to the results from the role plays.

**Table 5: Comparison between outcomes in the prediction markets and the role plays.**

<b>Contract</b>	<b>Closing price</b>	<b>Diff</b>	<b>Weighted averages closing price</b>	<b>Outcome role play**</b>	<b>Number of Satisfactory and Satisfactory with some deficiency</b>
<b>PC 1.1</b>	79.1		79.3	Passed	0.55
<b>PC 1.2</b>	79.45	-0.35			
<b>PC 2.1</b>	82.15		81.0	Passed	0.75
<b>PC 2.2</b>	80.05	2.1			
<b>PC 3.1</b>	77.9		78.0	Passed	0.625
<b>PC 3.2</b>	78.1	-0.2			
<b>PC 4.1</b>	82.5		79.3	Failed	0.45
<b>PC 4.2</b>	75.25	7.25			
<b>PC 5.1</b>	86.45		81.9	Passed	0.75
<b>PC 5.2</b>	75.5	10.95			
<b>Gov 1</b>	79.2		-	Passed	Outcome of election
<b>Gov 2</b>	81.55	-2.35			yielded dividend
<b>SD 1</b>	33.5		-	Passed	Outcome of election
<b>SD 2*</b>	58.55	-25.05			yielded dividend

\*SD 2 was traded heavily just before the market closed which raised the price

\*\* The outcome of the role play was judged as pass if the number of Satisfactory and Satisfactory with some deficiency was over 50

As shown above in table 5, the prices of identical contracts were not equal on the two markets for several of the contracts. Nor did they have the same order of how likely similar contracts were relative to each other. Both the price itself and the rank of the contracts give information and are worth examining. Regarding the closing price the difference between the markets was not especially large except for PC 5 (and SD). It is worth mentioning that prices varied more in one of the markets (the one with higher activity). With respect to the rating relative to each other is the relevant parameter, the markets agree on PC5 > PC4 and PC2 > PC1 > PC3, but rate the first cluster in the top in one of the markets and in the bottom in the other and vice versa. The weighted average of the PC with respect to the number of transactions provides yet another interpretation; then the order of the contracts is PC1, PC3, PC4, PC2 and PC5. One pair of the conventional contracts ended on very alike closing price while the other pair was the one that was most different. This contract was traded heavily on one of the markets just before the market closed which raised the price.

**Table 6: The two prediction markets, and the weighted averages and the outcome from the role plays compared to each other**

Market 1		Rank	Market 2		Rank	Weighted averages	Rank	Outcome role plays	Rank
PC 1.1	79.1	5 (4)	PC 1.2	79.45	3 (2)	79.3	3	55	4
PC 2.1	82.15	3 (3)	PC 2.2	80.05	2 (1)	81	2	75	1
PC 3.1	77.9	6 (5)	PC 3.2	78.1	4 (3)	78	5	62.5	3
PC 4.1	82.5	2 (2)	PC 4.2	75.25	6 (5)	79.3	3	45	5
PC 5.1	86.45	1 (1)	PC 5.2	75.5	5 (4)	81.9	1	75	1
Gov 1	79.2	4	Gov 2	81.55	1				
SD 1	33.5	7	SD 2	58.55	7				

In order to examine whether there was any association between the ranks of the markets and the ranks of the weighted averages and the role plays Spearman rank correlation test was conducted. Two hypotheses (whether the ranks of the tested variable did not co-vary with each other at 10 per cent level) were tested to evaluate the association between the markets and the weighted averages and the role plays (and also the co-variation with the role plays and each market separate). All results were tested at the 10 per cent level.

**Table 7: Results from the Spearman rank order correlation test**

Rank	N	Spearman rank correlation coefficient
Market 1 – Market 2	7	0.11
Weighted average – Role plays	5	0.50
Weighted average – Market 1	5	0.15
Weighted average – Market 2	5	0.35

The hypothesis of no association between the two variables could not be rejected for any of the comparisons at the 10 per cent level. The coefficient of the weighted average compared to the role plays was the one closest to the cut-off point. Worth mentioning though is that the sample is very small, hence the statistical strength is very limited and the point estimate of the correlation high. Therefore conclusions will be refrained from.

## ***5. Discussion***

### **5.1 Interpretation of the outcome**

#### **Results from the experiments**

The experiment was performed in a far too small setting to be able to draw any certain conclusions, but some tendencies can be discussed. How to interpret the result must be clear. Prediction markets seem, according to most research (Wolfers and Zitzewitz 2006 b), to provide accurate (but sometimes biased) predictions. The binary claim structure with winner-take-all dividend determines the probability that an event will occur. Wolfers and Zitzewitz (2004) state that the price on a winner-take-all market represents the market's expectation of the probability that an event will occur. The interpretation used below is that the group betting disposition showed the groups' degree of belief of how likely it was that FOI would reach the prioritized commitment PC<sub>i</sub>. This interpretation will be used in the discussion.

The price levels differed between the markets with one market having significantly higher prices than the other. Nor were the prices on identical contracts on the two markets particularly analogous, since both the order and price levels differed. It is worth noting that the participants judged all contracts alike; all contracts except the Sweden democrats either passed or failed (the result were interpreted as pass if prices were higher than 50 and fail if prices were lower than 50). Market 2 was the one most analogous to the results in the role plays in respect to the price levels, but the weighted average of the two markets were closer to the ranking order of the role plays. The Spearman rank order correlation test could not reject hypotheses of no associations between the ranks of the markets, the weighted averages and the role plays at the 10 per cent level. The result closest to be significant was when the ranks of the weighted averages were compared to the ranks of the role plays. This might imply that the number of participants in the two markets were too low.

One of the prioritized commitments failed in the role playing; still the closing prices on that PC were 82.5 and 75.25 and higher than the entry price. Ottavani and Norman Sorensen (2007) claim that the equilibrium price when traders have a limited amount of money, heterogeneous prior beliefs, and private information, under-reacts to information. Thus even if

the rational expectations equilibrium price reveals private information; it does not manage to reveal all information. This could be one explanation why the prices were so high.

Another explanation why the judgment differed between the markets and the role plays could be that the personnel at FOI who participated in the prediction markets may have been biased and thought too highly of the agency's capability and competence, even if some deficiencies exist. They also might have had confidence in the superiors. The superiors that participated in the role plays may have had an altogether different picture than other personnel. This may have been yet another reason for differences in judgment. Still, a heterogenic group with varying levels of knowledge and experiences as well as different backgrounds and research areas is important for the prediction markets to work as intended. The personnel at FOI are assuredly a homogenous group, but there are reasons to believe that the superiors are even more homogenous. In the role plays each person exerts significant influence on the outcome, so when the views of one person deviates strongly, this will affect the judgment considerably. This may also have influenced the results in one additional way, since the personnel at FOI are a fairly homogenous group. To involve even more stakeholders in the future may result in more accurate results, at least when implementing prediction markets slightly differently.

In the experiment performed, a linear scoring mechanism and a market maker were used. The market maker upheld activity and made people able to trade all the time. Without the market maker there would have been too little action during the test period. If bids have to be met with a counterbid, it is vital that activity is very high. In this experiment, too few participated to uphold the activity in these actual markets. The linear scoring mechanism likely affected the result more than the automated market maker and had several disadvantages. The mechanism was not as sensitive and well calibrated as would have been desirable. Linear increases and decreases affected the market via a fixed sum and the prices were not the same for all contracts traded in a transaction but changed for every contract. It was communicated, but there is still reason to believe that not all participants were conscious of that fact. This linear scoring mechanism and the elasticity constraint could be the reason of the rapid price change of one of the SD contracts in the end of the test period. It is vital to find a scoring mechanism that is not too slow nor too sensitive. It is preferable to have a system that functions similarly to conventional financial markets, since some of the participants are

familiar with them and they have been proven over time. This could unfortunately not be done in this mock-up study but is a strong recommendation for future studies.

As mentioned in the description of the platform, the entry value was not straightforward to set. Most of the contracts also ended close to the entry value. According to previous research on prediction markets, it is of great importance to think about the implications of psychology for economics (Sonneman et. al 2008). One implication of this in this experiment was how to set entry prices. There was no previous experience and hence no possibility to make educated guesses on past outcomes to base entry prices on. In the choice between estimating starting values based on expectations of the closing value and to start at the same level for all contracts, the latter was chosen. The possible impact of a different entry value on the outcome is debatable. The value used may have had an anchoring effect connected with the entry price. Sonneman et al. (2008) discuss effects of framing. They conducted an experiment using a different set of parameters and thus their results are not directly applicable, but in their experiments framing was seen to have affected the result.

The prices recorded were more volatile in the beginning of the test period for most of the contracts. This is in line with what would have been expected, especially with regard to arbitrary entry prices. This could be explained by several factors; the contracts were new and were launched at a specific price and the prices seemed to converge a little after a while. The activity was also highest in the beginning which also could be an explanation.

### **Activity on the markets**

The activity on the markets was quite low, especially when looking at the two markets separately. This together with the results from the correlation test indicate that the number of participants were too low. Low activity makes the market thin and could be a reason that the prices were so different on the two markets. Group size is a question that must be further examined, and if implementation of prediction markets in risk and vulnerability analysis works in small groups. Christiansen (2007) examines trading behavior and shows that even markets with few participants uphold the ability to make accurate predictions. However, accuracy decreases when the number of people involved falls. The choice of price mechanism seems to play an important role in prediction markets in small groups. The mechanism to reflect the price as the true judgment is hard to uphold entirely when there are few

participants. Another option is to use market scoring rules which were presented in section 2.3.1.

Wolfers and Zitzewitz (2006 b) suggest that markets with events that are widely discussed and have high entertainment value and an abundance of information will work better, which was the reason to include election contracts and can explain why the election contracts had higher levels of trade. The results support that it might be a good idea to include more engaging contracts since the conventional contracts were more traded than the others. Hopefully this spilled over and led to more activity on the other contracts as well. This could be a reason to mix different kinds of engaging contracts, so all participants have contracts they find engaging. One disclaimer is that the liquidity from each participant may be invested in contracts not concerning the agency capacity assessments. Worth mentioning is also the difference in price between the SD two contracts even if they were traded the most. While the Gov contracts ended on prices very close to each other the SD did not. One problem with interpreting the SD prices is that one person traded heavily on one SD contract shortly before the closing which affected the price a lot. However, there are too few observations to draw any conclusions.

A lot of effort was spent to market the experiment to uphold activity and create awareness of it. Prediction market in this context might have had the charm of novelty which might have increased the willingness to participate, but this may not work in the long run. High activity is vital for the markets to function properly. The trading was still very low at some of the contracts. This could be the case due to several reasons; the contracts were hard to understand, participants knew very little about some of the contracts, participants did not understand entirely how the markets work (i.e. short selling), lack of other information etc. How to best uphold activity is a vital question that must be investigated further, for example by surveying.

## **5.2 Design and function of the markets**

Claim definition is of great importance for the market to be reliable and accurate. Wolfers and Zitzewitz (2004) among several other researchers state the importance of clearly defining all claims for the market to be efficient, effective and manageable for participants. People must understand what they are trading on for the market to be well-functioning and a useable means to elicit information. This must be further investigated and one idea is to perform

surveys together with prediction markets in future research to ensure that the contracts are correctly understood.

The outcomes of the contracts were set based on the outcomes in role plays. This was not optimal, because it made the participants to trade on the results of the role plays and not the conditional probability that “FOI, given event X, had a fair potential to reach the prioritized commitment  $PC_i$ ”. Still, one can argue that the best guess still is to trade on the likely outcome of the scenario which will make this less of a problem. However, finding other grounds for judgment would be better. That might lead to an implementation of prediction markets in a slightly other phase of the risk and vulnerability analysis. Examples could be the likelihood of accidents rather than the outcome of the role plays. The aim to implement prediction markets in this phase of the risk and vulnerability analysis was to replace an imperfect method and to improve the process and make it less costly and time-consuming. This could in some aspects still be true even if the method has some deficiencies. Even if the prices in themselves might be inappropriate to base decisions on, the relative prices and rank among them still make it useful for resource allocation and the like. Another motive was also to improve the predictions; this is not solved when playing on the outcome of the role plays. One problem with the ground for judgment used in this experiment is that the role plays cannot be removed. It might give more reliable results to combine the results from the role plays and the prediction markets, but it will fail in saving any time or money. To find another ground for judgment would be most welcome, one solution could be to only predict outcomes that could be judged by real outcomes.

Participation was voluntary. This ought to have attracted participants that were curious and more likely to actively participate in the market and appears to goad some into being more active and also seems to lead to a bias of participants, which could be a problem. According to Christiansen (2007) the passion incentive, when traders participate due to interest and passion, is as relevant as other incentives. He suggests that in a market traders find interesting and where they are passionate about the outcome, traders do not need any incentives at all. He also attaches less importance to anxieties that all traders must be active in the market. Even if the majority of the participants do not participate regularly, they still set an initial landscape of pricing in the market. This research and lessons from it must be further investigated to create markets with high activity. One objection to the reward system used is also that it

mostly rewarded participants that reach the top per cent. It was a minor reward to log in and be active, but this prize was less than the prize of topping the market. This kind of reward system has the disadvantage that it may reward risky strategies with great variance. The reward system must be designed not to encourage risky strategies. Studies on both financial market and empirical game theory discuss how to deal with this problem and how to design more neutral reward systems. To improve the reward system, guidance can be taken from prediction markets used in business.

### **5.3 Advice and lessons to be learned for the future regarding the experimental design**

Some changes of the experimental setting have already been mentioned and this section will discuss this further. A great deal of work can be done to improve the experimental setting as many aspects of the experiment worked well but some did not function optimally. The experiment must be made on a larger scale and another market mechanism must be implemented and preferably one similar to conventional financial markets. The market mechanism has been found to depend considerably on the number of participants and how the contracts are formulated. The optimal or at least essential number of participants must be found out. Markets scoring rules must be explored further because it may be feasible and able to compensate for a low participation rate. The impact of the starting value must be investigated further to avoid an inappropriate influence. This could for example be done with parallel markets with identical contracts but different entry prices. It is also possible to make improvements on incentives for participation. The reward system must be amended not to encourage risky strategies. Another important question is how to do to increase activity in the market setting.

There may also be a risk that some actors have problems to communicate their well-informed predictions. One way to improve the prediction markets could be to involve direct interaction or adding the possibility to leave comments or the like. There might otherwise be a problem that participants that hold minority opinions and are well-informed not are able to convince the group. If the prediction market would allow the traders to communicate with each other, those people could by reasoning with each other be more reliable.

## **5.4 Theoretical and practical implications and validity of the study**

This study contributes to the growing literature on prediction markets and their implementation. Previous research has tested prediction markets in a number of areas but the area of risk analysis remains relatively unexplored. The research question about prediction markets validity and eligibility to be a tool for risk analysis is partly answered. The experiments described in this paper have not uncovered any evidence contradicting the use of prediction markets in this area. Admittedly, this thesis was not able to draw firm conclusions from the experiments that prediction markets will be a valuable tool, the opposite cannot be said to apply either. The theoretical research supports the implementation and this experiment could form the basis for more ambitious and larger scale experiments. There are promising prospects which provide sufficient reasons to continue the test. If prediction markets stand up to the test of further research, there is much to be gained in saved time and money in capacity assessment. Whether or not prediction markets are the only used tool to forecast they will be of value and contribute to improve analyses. Still, a vast body of future research on the design of the experimental set up needs to be done.

As previously mentioned the experiment was conducted in far too small a setting to validate the results. The empirical study this essay is based on can only be seen as a pilot study and function as inspiration for studies to come. The market mechanism, entry prices and experimental design were not optimal and questions where to best implement prediction markets remain. Therefore, the validity of the experimental setting is low. The lessons to be learned from this thesis are recommendations for future studies. However, the purpose of this thesis was to discuss the idea itself and examine if prediction markets are well-suited to implement in capacity assessments. As such, this thesis is more valid than the experiment in itself.

## **5.5 Future research**

Some suggestions for improvements and future research have already been presented, the most important of which is to show that the result from the prediction markets is reliable and relevant for implementation in risk and vulnerability analyses. The use of prediction markets in risk and vulnerability analysis is dependent on this. This may infer to implement prediction markets in a slightly different way than the experimental setting that this essay is based on.

Furthermore, a rather thorough exploration into the markets and discussion of how to exactly interpret the outcome of them will be necessary in future studies.

To obtain more favorable and reliable results from the experiment some improvements and changes must be made. Suggestions for improvements have been outlined in section 5.3, some examples of research questions that have to be solved. To be able to draw conclusions from the experiment it must be conducted on a much larger scale and with a better design.

## ***6. Conclusion***

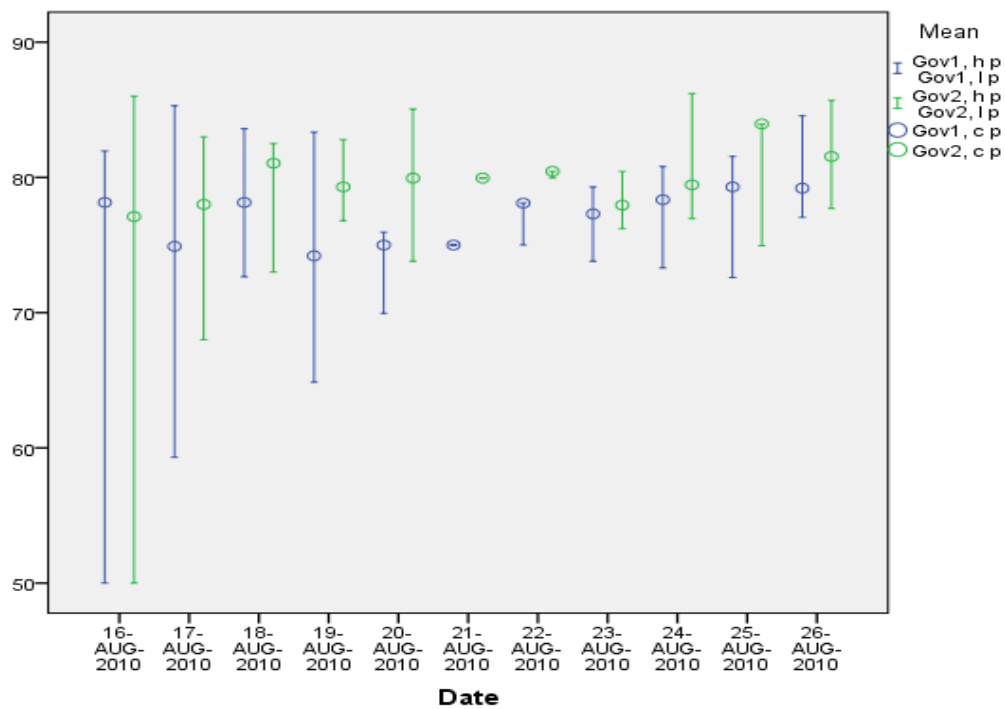
The aim of this essay was to evaluate an experiment that uses prediction markets in a risk and vulnerability analysis. The research presented in this paper regarding implementation of prediction markets and the characteristics of risk and vulnerability analyses indicate that risk and vulnerability analyses are a promising field to implement prediction markets in. There are many challenges remaining to make risk and vulnerability analyses more reliable and effective regarding the forecasting. The characteristics and requirements of risk and vulnerability analysis, with the need for forecasting, point towards that it might be a suitable field to implement prediction markets in. This thesis was not able to draw firm conclusions from the experiments that prediction markets will be a valuable tool, the opposite cannot be said to apply either. Prediction markets may or may not be implemented to use solely, however, it will likely, if implemented properly, help generating better prediction and decisions. This thesis has served as a first critical step to examine the potential use of prediction markets in risk and vulnerability analysis. Still, a vast body of future research on prediction markets in this context and the design of the experimental set up needs to be done.

## 7. Appendix

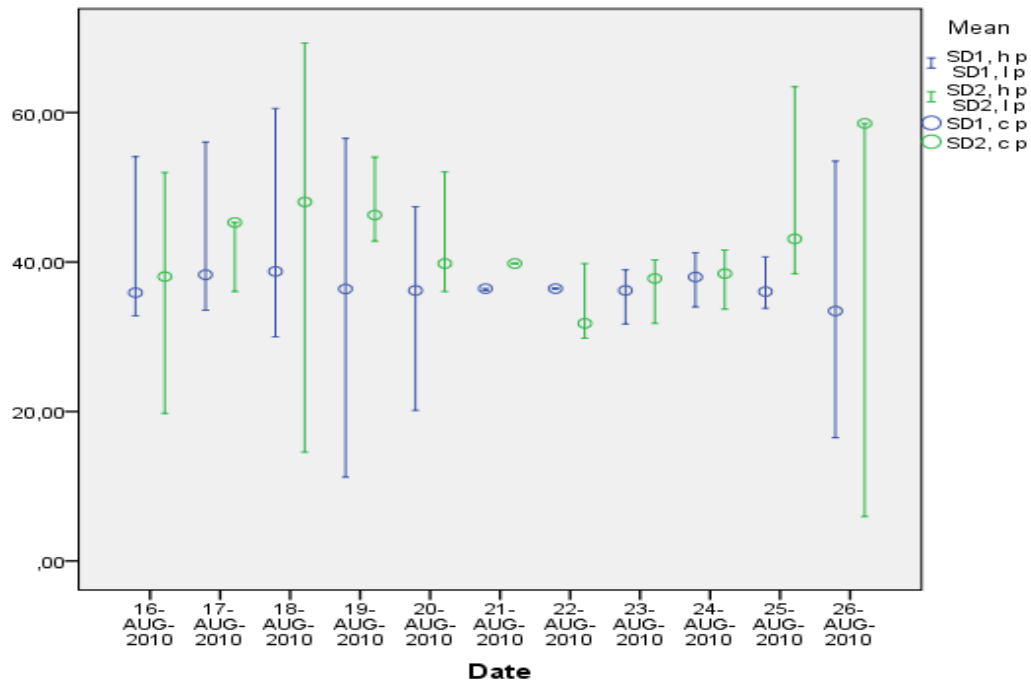
In this part further results from the experiment will be presented.

Below each contract and the development of the two markets is compared. Stock charts, which show the highest price, lowest price and closing price for every day, are used. The low-active part in the model was due to a weekend in the middle of the test period.

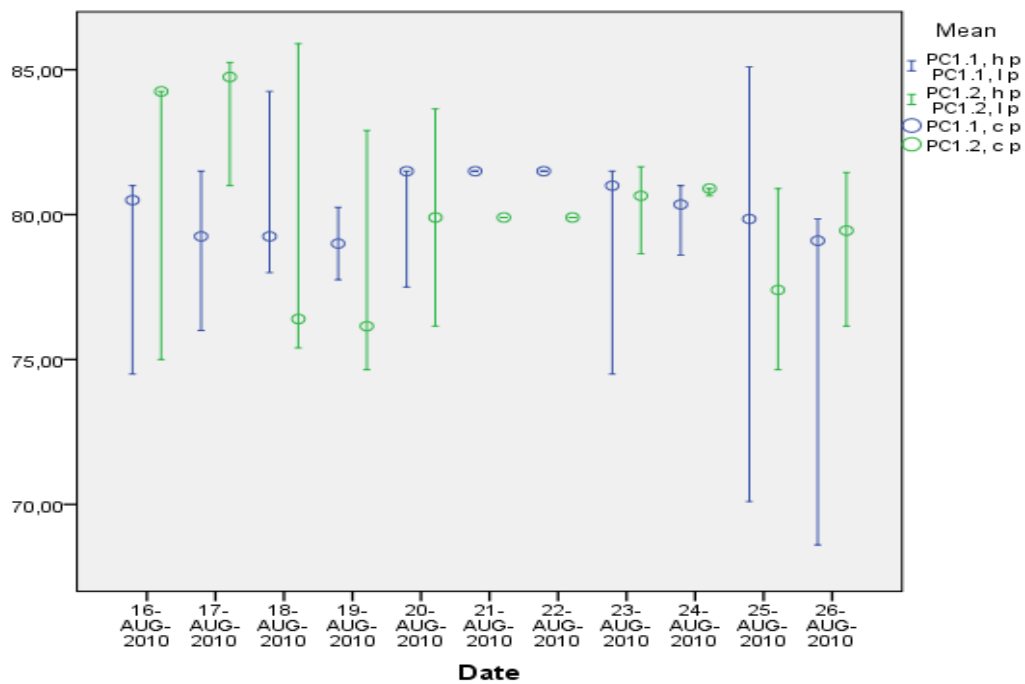
**Figure 3: Price trend for the contract about the governing coalition**



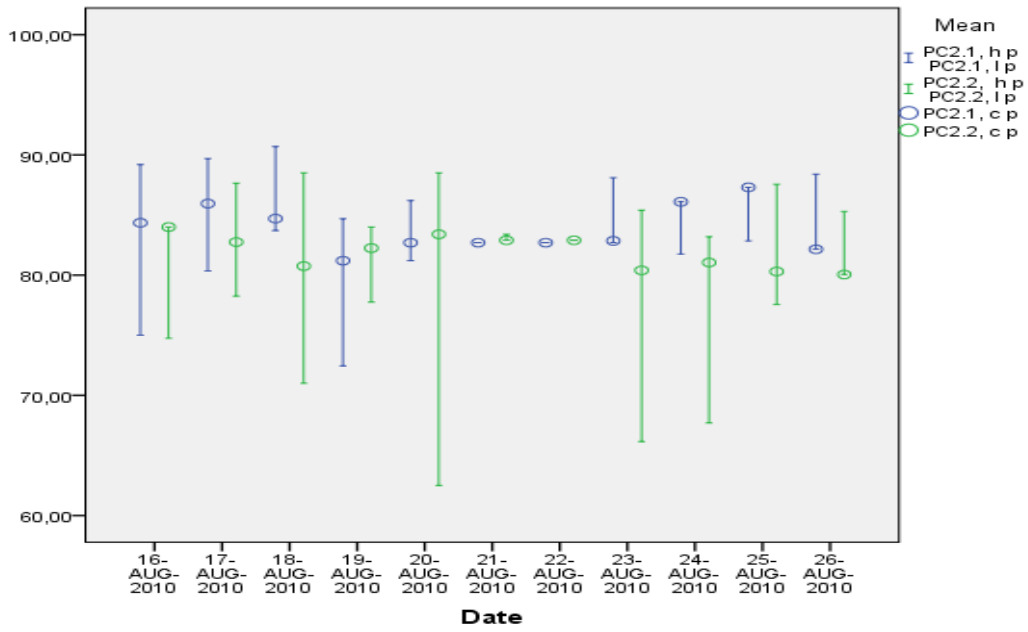
**Figure 4: Price trend for the contract about the Sweden democrats**



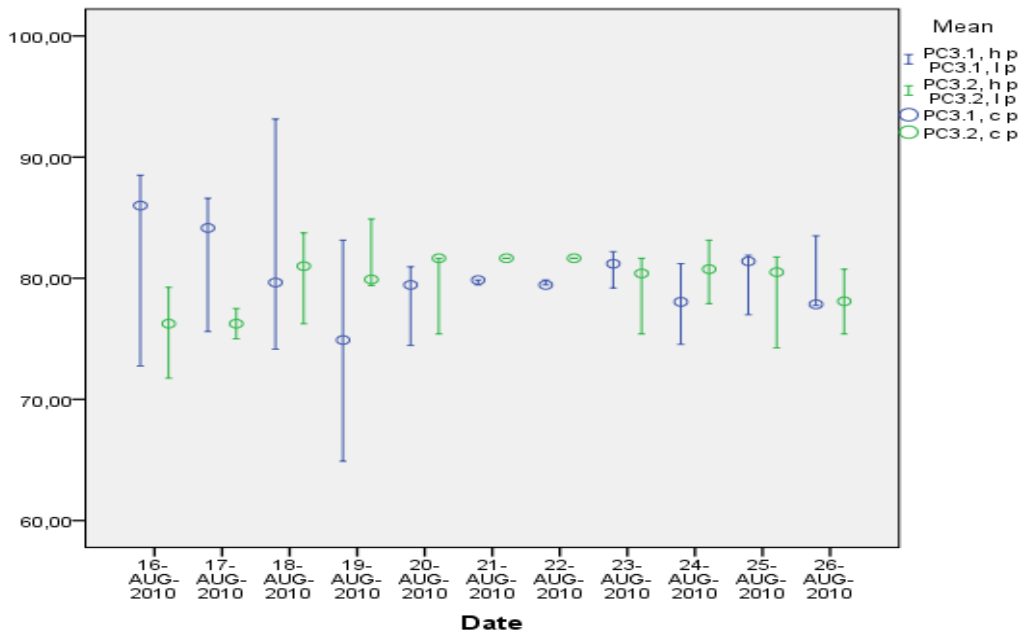
**Figure 5: Price trend for contract about prioritized commitment 1**



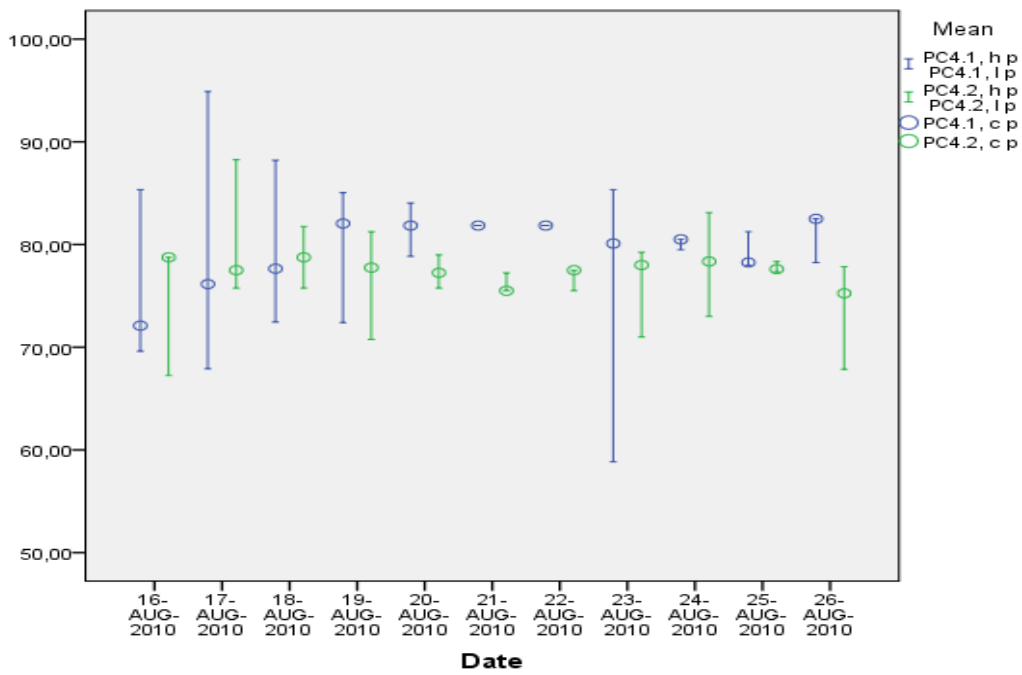
**Figure 6: Price trend for contract about prioritized commitment 2**



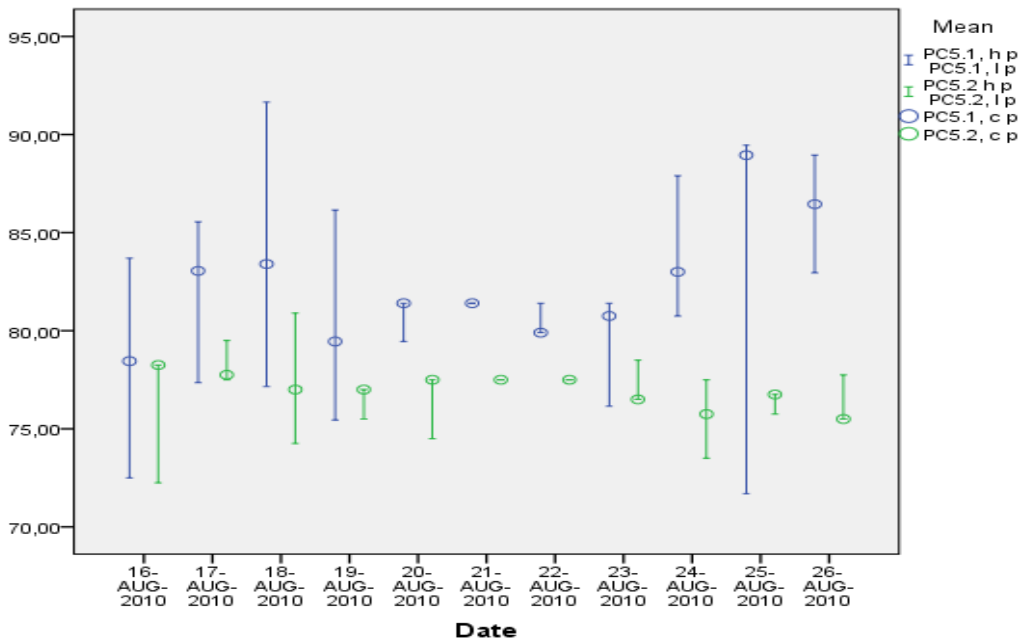
**Figure 7: Price trend for contract about prioritized commitment 3**



**Figure 8: Price trend for contract about prioritized commitment 4**



**Figure 9: Price trend for contract about prioritized commitment 5**



## 8. References

- Arrow, K. J. (2008). The Promise of Prediction Markets. *Science* 320: 877-888.
- Berg, J. E., and Rietz, T. A. (2003). Prediction Markets as Decision Support Systems. *Information Systems Frontiers*, 5 (1): 79–93.
- Berg, J., Nelson, F. and Rietz, T. A. (2008). Prediction Market Accuracy in the Long Run. *International Journal of Forecasting* 24: 285–300.
- Chen, K. Y. and Plott, C. R. (2002). Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem. *Social Science Working Paper No. 1131*, California Institute of Technology
- Christiansen JD. (2007). Prediction markets: practical experiments in small markets and behaviours observed. *Journal of Prediction Markets* 1: 17–41.
- Debnath, S., Pennock, D. M., Lawrence, S. and Giles C. L. (2003). Information incorporation in online in-game sports betting markets. *Proceedings of the 4th Annual ACM Conference on Electronic Commerce (EC' 03)*: 258–259.
- Deschamps, B. and Gergaud, O. (2007). Efficiency in Betting Markets: Evidence From English Football. *The Journal of Prediction Markets* 1: 61–73.
- DeWispelare, A. R., Herren, L. T. and Clemen, R. T. (1995). The use of probability elicitation in the high-level nuclear waste regulation program. *International Journal of Forecasting*, 11 (1): 5–24.
- Eriksson, J., Juhl, A-K., Wikström, T. and Rinne, A. (2011). *Vägledning för risk- och sårbarhetsanalyser*. Karlstad. Myndigheten för samhällsskydd och beredskap (MSB)
- Forsythe, R., Frank, M., Krishnamurthy, V. and Ross, T. W. (1998). Markets as Predictors of Election Outcomes: Campaign Events and Judgement Bias in the 1993 UBC Election Stock Market. *Canadian Public Policy* 24: 329–351.
- Graefe, A., Luckner, S. and Weinhardt, C. (2009). Prediction Markets for Foresight. *Futures* (forthcoming).
- Graefe, A. and Scott, A. J. (2011). Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task, *International Journal of Forecasting*, 27.1: 183-195.
- Green, K.C., Armstrong, J. and Graefe, S. (2007): Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared. *Foresight: The International Journal of Applied Forecasting*, Fall issue.
- Hahn, R. W. and Tetlock, P. C. (2006). *Information Markets*, AEI Press.

- Hanson, R. (1999). Decision Markets. *IEEE Intelligent Systems*, 14(3): 16-19.
- Hanson, R. (2000). Shall We Vote on Values, But Bet on Beliefs?. *mimeo*, George Mason University.
- Hanson, R. (2003). Combinatorial Information Market Design. *Information Systems Frontiers* 5: 107-119.
- Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decision and fiascoes*. Boston: Houghton Mifflin.
- Johansson, H. and Jönsson, H. (2007). Metoder för risk- och sårbarhetsanalys ur ett systemperspektiv. LUCRAM, Lunds universitet, Lund.
- Kaplan, S. (1997). The words of risk analysis. *Risk Analysis* 17: 407-417.
- Kaplan, S. and Garrick, B.J, (1981). On the Quantitative Definition of Risk. *Risk Analysis* 1.1: 11-27.
- Luckner, S., Schröder, J. and Slamka, C. (2008). On the Forecast Accuracy of Sports Prediction Markets. in Gimpel, H. et al., (Eds), *Negotiation, Auctions & Market Engineering, Lecture Notes in Business Information Processing (LNBIP)*, Springer-Verlag (2008), 227–234.
- Ottaviani, M. and Norman Sorensen, P. (2007). Aggregation of Information and Beliefs in Prediction Markets.
- Ortner, G. (1997). Forecasting Markets—An Industrial Application: Part I, TU Vienna, Working Paper.
- Pennock, D. M., Lawrence, S., Giles, C. L., and Nielsen, F. A. (2001). The Power of Play: Efficiency and Forecast Accuracy of Web Market Games. Technical Report 2000-168, NEC Research Institute.
- Polgreen, P. M., Nelson, F. D., and Neumann, G. R. (2007). Use of Prediction Markets to Forecast Infectious Disease Activity. *Clinical Infectious Diseases* 44: 272-279.
- Rowe, G. and Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting* 15: 353–375.
- Servan-Schreiber, E., Wolfers, J., Pennock, D. M. and Galebach, B. (2004). Prediction Markets: Does Money Matter?. *Electronic Markets* 14: 243-251.
- Sonnemann, U, Camerer, C F., Fox, C R. and Langer, T (2008). Partition-Dependent Framing Effects in Lab and Field Prediction Markets.
- Spann, M. and Skiera, B. (2003). Internet-Based Virtual Stock Markets for Business Forecasting. *Management Science* 49: 1310–1326.

- Sunstein, C. (2002). The Law of Group Polarization. *Journal of Political Philosophy* 10: 175-195.
- Sunstein C. (2004). Group Judgments: Deliberations, Statistical Means, and Information Markets. *AEI-Brookings Joint Center for Regulatory Studies*. working paper No. 04-17.
- Sunstein, C. (2006). *Infotopia*, Oxford University Press.
- Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. New York: Doubleday.
- Tziralis, G. and Tatsiopoulos, I. (2007). Prediction Markets: An Extended Literature Review. *The Journal of Prediction Markets* 1: 75–91.
- Wolfers, J, and Zitzewitz E. (2004). Prediction Markets. *Journal of Economic Perspectives* 18: 107-126.
- Wolfers, J. and Zitzewitz E. (2006 a). Interpreting Prediction Market Prices as Probabilities. *National Bureau of Economic Research Inc*. No. 12200.
- Wolfers, J. and Zitzewitz E. (2006 b). Five Open Questions About Prediction Markets. in R. W. Hahn and P. C. Tetlock (eds) *Information Markets: A New Way of Making Decisions*. Washington D.C., AEI Press, 2006: 13-36.
- Wolfers, J. and Zitzewitz E (2006 c). Prediction Markets in Theory and Practice. in S. Durlauf and L. Blume (eds) *The New Palgrave Dictionary of Economics* (McMillan, 2nd ed., 2006): forthcoming.