

Sornette, Didier et al.

Working Paper

Overpricing persistence in experimental asset markets with intrinsic uncertainty

Economics Discussion Papers, No. 2019-32

Provided in Cooperation with:

Kiel Institute for the World Economy (IfW)

Suggested Citation: Sornette, Didier et al. (2019) : Overpricing persistence in experimental asset markets with intrinsic uncertainty, Economics Discussion Papers, No. 2019-32, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:

<http://hdl.handle.net/10419/196165>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Overpricing persistence in experimental asset markets with intrinsic uncertainty

Didier Sornette, Sandra Andraszewicz, Ke Wu, Ryan O. Murphy, Philipp Rindler, and Dorsa Sanadgol

Abstract

To study coordination in complex social systems such as financial markets, the authors introduce a new prediction market set-up that accounts for fundamental uncertainty. Nonetheless, the market is designed so that its total value is known, and thus its rationality can be evaluated. In two experiments, the authors observe that quick consensus emerges early yielding pronounced mispricing, which however do not show the standard “bubble-and-crash”. The set-up is implemented within the xYotta collaborative platform (<https://xyotta.com>). xYotta’s functionality offers a large number of extensions of various complexity such as running several parallel markets with the same or different users, as well as collaborative project development in which projects undergo the equivalent of an IPO (initial public offering) and whose subsequent trading matches the role of financial markets in determining value. xYotta is thus offered to researchers as an open source software for the broad investigation of complex systems with human participants.

JEL C90 D47 D80

Keywords Experimental asset market; prediction market; uncertainty; experimental economics

Authors

Didier Sornette, ETH Zurich, Zurich, Switzerland and Swiss Finance Institute

Sandra Andraszewicz, ETH Zurich, Zurich, Switzerland, sandraszewicz@ethz.ch

Ke Wu, ETH Zurich, Zurich, Switzerland

Ryan O. Murphy, Morningstar

Philipp Rindler, ETH Zurich, Zurich, Switzerland

Dorsa Sanadgol, ETH Zurich, Zurich, Switzerland

Citation Didier Sornette, Sandra Andraszewicz, Ke Wu, Ryan O. Murphy, Philipp Rindler, and Dorsa Sanadgol (2019). Overpricing persistence in experimental asset markets with intrinsic uncertainty. Economics Discussion Papers, No 2019-32, Kiel Institute for the World Economy.

<http://www.economics-ejournal.org/economics/discussionpapers/2019-32>

Introduction

We propose a new experimental design for investigating asset markets. This design derives from two distinct experimental approaches - the classical asset market experiments and prediction markets. It includes several factors that should mitigate mispricing. Our experiment features true Knightian (Knight, 1921) uncertainty, which refers to a situation in which we cannot know all the information we need in order to calculate the odds. In our new set-up, dozens of securities form a complete market associated with all possible outcomes, while the underlying probability structure of the outcomes is unknown and fundamentally unknowable. This setup allows us to study how well a trading market aggregates opinions into a consensus and how stable this consensus is to changes in the environment.

In our set-up, many assets are available to trade continuously in a realistic stock market platform and over six full days. While all outcomes are known and can be priced by their corresponding assets, the subjects cannot learn or estimate the probabilities of the possible outcomes with certainty. They can make educated guesses and inference. The existence of the many assets allows us to observe self-organized emergence of consensus and its evolution over the experiment. As an additional design feature, our experimental setup eliminates the possibility that subjects distrust the information on the fundamental value. Instead, it focuses on the dynamics of pricing in an environment with limited information. Our key observation is quantification of opinion consensus, in spite of fundamental and persistent uncertainty. We observe that aggregated prior beliefs, and the exchange of opinions via the bid-ask spreads, result in the quick emergence of a general conjecture from the group, which turns out to be close to the realized outcomes.

Our new proposed experimental design is motivated by the need to understand what may lead to dysfunctions in financial markets. Indeed, the major social function that security markets fulfill is the aggregation of opinions and the channeling of investment capital to promising ventures. One of the mechanisms underlying the performance of financial (and prediction) markets is the “wisdom of crowds” (Mannes, 2009; Ray, 2006;

Surowiecki, 2005) – a phenomenon in which the weak existing information diluted over many individuals may emerge above the large noise by aggregation over the group. Another mechanism is that experts, and even insiders who have special private information, may reveal their knowledge by trading (Chesney, Cramer, & Mancini, 2015). Prediction markets have been used to successfully predict political elections (Berg, Forsythe, Nelson, & Rietz, 2008; Forsythe, Nelson, Neumann, & Wright, 1992; Forsythe, Rietz, & Ross, 1999; Hansen, Schmidt, & Strobel, 2004), outcry of infectious diseases (Polgreen, Nelson, & Neumann, 2007; Tung, Chou, & Lin, 2015), sports outcomes (Kain & Logan, 2014) and new product blockbusters (Cowgill, Wolfers, & Zitzewitz, 2009; Elberse & Eliashberg, 2003; Ho & Chen, 2007). All these markets have one common denominator: the possible outcomes are known while the underlying probability structure of the outcomes is unknown and fundamentally unknowable. Therefore, the participants of prediction markets make “educated guesses”, while the market prices emerging from aggregated traders’ beliefs should reflect the probability of future outcomes (Berg & Rietz, 2003; Manski, 2006). Deck and Porter (2013) provide a comprehensive review of the use of prediction markets in the laboratory and field studies.

There is a continuous debate among economists whether markets fulfil their role efficiently at all times (Grossman & Stiglitz, 1980). The most visible breakdown of asset markets is arguably during bubbles and the crashes that follow. The issue of understanding price bubbles has therefore been a great concern for researchers and practitioners. A key problem in studying security markets empirically is their complexity and interaction of many variables. To address this issue, V. L. Smith, Suchanek, and Williams (1988) conducted a seminal study with a simple setup where a few persons would trade one risky asset over a period of a few minutes. This setup, called the SSW design, pioneered the use of experimental asset markets to study how financial markets function and how specific mechanism changes might affect trading behavior and price outcomes.

Following on V. L. Smith et al. (1988), many researchers have used experimental asset

markets to study how such institutions function and how specific mechanism changes might affect trading behavior and price outcomes. These studies share one property - there is a well-defined fundamental value¹ of the traded security, which allows for precise calculations of deviations of the market prices from the rationally objective value. In all these studies, the actual “true” value for each period and the probability distribution is either given directly to subjects or can be calculated precisely with relative ease. A striking and robust finding of these experiments is the emergence of bubbles and crashes, even when the information about the rational price is directly provided to the participants (Powell, 2016). This phenomenon known as the “bubble-and-crash puzzle” has not been fully understood and “formal theoretical explanation is an area of future work” (V. L. Smith, van Boening, & Wellford, 2000). The pursuit for understanding these price bubbles has generated a large experimental literature (see Nuzzo & Morone, 2017; Palan, 2013; Powell & Shestakova, 2016, for review).

Palan (2013) identified a number of factors that have been found to mitigate the price bubbles in an experimental setting. These factors include: expertise of a trader, common expectations of rationality, low cash-to-asset ratio, large accrual dividend, trading teams instead of individual traders, lack of overconfidence, existence of alternatives to trading, short-selling, limit price change, non-tournament type of compensation, and comparison to the best players. Individual factors mitigating mispricing do not make the bubbles disappear completely, even after training of the participants. However, Kirchler, Huber, and Stöckl (2012) demonstrated that the declining fundamental value process in the SSW design is confusing for participants and that making this process more intuitive resolves the confusion and reduces mispricing. This motivates the question of “how well the results [of the SSW experiments] extend to more realistic market settings” (Powell & Shestakova, 2016). In psychological laboratory studies, the relation between the experimental findings and people’s behaviour “in the wild” is an important point of critique that is addressed by

¹Fundamental value is the value of a security obtained by discounting the future income generated by the security.

alternating experimental paradigms to test related theories and by conducting field studies. However, in experimental asset markets, the same research method - the SSW design - is repeatedly utilised in studies that derive their theories from previous experiments (Loewenstein, 1999).

Consequently, “many experiments are not aimed at a well-specified real-world target but rather contribute to a library of robust phenomena, a body of experimental knowledge to be applied case by case” (Guala & Mittoe, 2005). Repetitive use of the very same experimental design may lead to the “mutual internal validity of theory and experimental test” (Schram, 2005) that creates its own world, where the robust findings from experiments may not be generalisable to the outside world (Guala & Mittoe, 2005). Artificiality of laboratory experiments and lack of context may reduce their relation to the real trading situations (Loewenstein, 1999; Schram, 2005).

On the one hand, laboratory experiments offer the possibility to manipulate and measure individual variables in a fully controlled way. On the other hand, it is a crucial question whether the bubbles in the experimental markets are a characteristic artifact of the SSW design or whether it is a general phenomenon of the market players. According to Powell and Shestakova (2016), the structure of the market plays an important role in attenuating or mitigating the bubbles. “What is still missing, however, is a careful analysis of possible new experimental methods that will help increase the external validity [of the experimental asset markets]” (Schram, 2005).

Moreover, the information available to agents in standard experiments departs strongly from the situation in actual markets and in real social coordination problems, where the probabilities of possible future outcomes are generally unknown. The fundamental value of a security is a “convention” (Orléan, 1995) or a theoretical construct, which is extremely difficult to estimate in real markets. The same problem has been faced by experimental asset markets, where various measures of mispricing may lead to inconsistent results (Powell, 2016). In academic finance thinking, the logic is often turned around by taking for granted that the market price is (almost) always right and any difference from a theoretical value may be due to incorrect choices of the dividend

growth and discount rates. In this logic, Fischer Black (1986) once famously observed that “we might define an efficient market as one in which price is within a factor of 2 of value, i.e., the price is more than half of value and less than twice value”. This difficulty in quantifying what is “true” value of a security is often at the source of failures in diagnosing financial bubbles in real time. This is because, in light of one model, a security is considered overpriced, but can be considered priced correctly according to another model (Gurkaynak, 2008).

Motivated by these important questions, the new experimental design for investigating asset markets proposed here departs from the studies following V. L. Smith et al. (1988) with a single asset traded for T periods (lasting in total about 1 hour) such that subjects know the probability distribution of dividends with certainty. Including multiple assets and various number of assets in different experimental rounds over extended periods of time makes the setup closer to the real world. We present the results of two experiments in which we explored pricing behaviour in the new design. In these experiments, we address the research question of how robust the mispricing effect is, when using a design different from the SSW-design and implementing several features that previously have been shown to mitigate bubbles. In the first experiment, we test the basic experimental setup. In the second experiment, we replicate the design with a few small improvements, test for robustness of the effects found in the first experiment and conduct analysis of traders’ strategies.

For simplicity, we present the method and results of Experiment 2 in the main text, while we provide the experimental details and results of Experiment 1 in Appendix C. The key findings from Experiment 2 also hold for Experiment 1. Appendix B outlines additional analyses for Experiment 2.

In sum, this paper sets four main goals. First, we present a new design featuring inherent uncertainty, and explain its functions. The software platform xYotta that powers it is offered as an open source to researchers interested in investigating coordination between human subjects in complex environments. Second, we investigate whether mispricing does occur in a prediction-market-like setting with several features

differing from the SSW design. Third, if it does, how would this mispricing pattern differ from the bubble-and-crash scenario frequently occurring in the SSW design? Last, we raise attention to the potential of the growing research devoted to prediction markets and how they can be applied with various modifications to investigate market dynamics and players' beliefs about future ventures. But first, we briefly summarise in the next section the current status of experimental asset markets, to establish the background on which our work develops.

Current Status of Experimental Asset Markets

The Gap between Naturally Occurring and SSW Markets

As already mentioned, the SSW studies share one property - there is a well-defined fundamental value of the traded security, which allows for precise calculations of deviations of the market prices from the fundamental value. The actual "true" value for each period and the probability distribution is either given directly to participants or can be calculated precisely with relative ease.

As discussed in the introduction, the information available to agents in standard experiments departs strongly from the situation in real financial markets where the probabilities of possible future outcomes are generally unknown. Hertwig and Erev (2009) points out important differences in decision making in situations under risk (i.e. when the probabilities of events are known) and uncertainty (i.e. when the probabilities of events are unknown) leading to a "description-experience gap" in decision making. This gap is analogous to the gap between naturally occurring and experimental markets, which can have an important impact on studying mispricing.

Alternative Market Designs

Irrational mispricing has also been reported in designs other than SSW. For example, Palfrey and Wang (2012) implemented a computerized laboratory experiment with a series of eleven experimental sessions with six markets attended by 10-12 players each. Each trading period lasted 50 seconds. They investigated pricing of markets with a

single security, complete markets with six securities, and markets that allowed short selling, given good or bad public information determined either with a toss of a fair coin or a roll of a fair die. The success of each security would depend on the public information. They reported over-pricing in single and complete markets with no news and with the same amount of good and bad news, while removing short-selling partially reduced mispricing but did not eliminate it. Ball and Holt (1998) conducted a non-computerised classroom market game with 5-6 students per trading team, allowing for minute-long, incentivizing students with real money at small stakes. They report that asset prices exceed their fundamental values.

Some authors introduced changes to the SSW design to measure mispricing depending on these design changes. For example, Bostian and Holt (2009); Holt, Porzio, and Song (2017); A. Smith, Lohrenz, King, Montague, and Camerer (2014) use a double auction market implemented in Vecon Lab ², which allows for various dividend generating mechanisms, payoff schemes, transaction costs and taxes etc. All of these studies featured trading sessions lasting 1-2-minute and were repeated 10-25 times by the same group of students. Bostian and Holt (2009) introduced another change by conducting their study online, with a number of students enrolled in a finance class that could participate in the experiment at a designated time from any place they wanted as long as they had access to the Internet. In all three studies, the dividend was paid out to the stock holders at the end of the trading period and there was one or two assets available for trading. The major change to the SSW design related to the various structures of dividends and fundamental values, including flat fundamental values, random dividends, etc. In a SSW-like design with multiple short trading periods with a single asset or assets with a complete number of states with known probabilities, Plott and Sunder (1988) implemented dividends that were dependent on the state of events at the end of the trading period.

A few other studies focused on the investigating the trading dynamics in different types of markets. For example, a review by Friesen and Gangadharan (2013) outlines how

²<http://veconlab.econ.virginia.edu/da/da.php>

experimental environmental markets – markets on which one trades tickets/permits for pollution limits or use of natural resources (i.e. fishing quota) – are used to investigate the impact of regulation on individual behaviour of traders in this complex trading environment. Depending on the set of trading rules, speculative bubbles occur (i.e. allowing for permit banking, which refers to treating permits for the use of environmental resources as assets that can be bought, held or leased) or can be diminished (i.e. when “permanent transfers are allowed only after traders have had some experience with temporary lease transfers” (Friesen & Gangadharan, 2013)).

Another class of studies, uses pari-mutuel betting games, where the market players can purchase tickets for particular state of an event such that tickets are purchased at fixed prices (see Noussair & Tucker, 2013, for a discussion on this type of markets). An example of a pari-mutuel betting market is a betting market for horse races, where event can have multiple states (i.e. a given horse can end up on place, 1, 2, 3, etc.). Herding (i.e. “betting in disagreement with one’s private signal but in favour of the consensus based on prior bets” Noussair & Tucker, 2013) is a commonly observed behaviour in this type of markets, but eliciting bettors’ beliefs directs their attention more to the probability of each state. These designs, however, are not applicable for studying asset markets.

As already pointed out, given their intrinsic uncertain nature, real financial markets can be conceived as particular incarnations of prediction markets, where the possible outcomes are known while the underlying probability structure of the outcomes is unknown and fundamentally unknowable. Therefore, the participants of prediction markets make “educated guesses”, while the market prices emerging from aggregated traders’ beliefs should reflect the probability of future outcomes (Berg & Rietz, 2003; Manski, 2006). In financial markets, traders aggregate their beliefs concerning the future performance of firms, leading to prices that can be interpreted as predictions of the firm value. Indeed, in the efficient market hypothesis, the present price is equal to the discounted expectation of all future prices. The present price is thus supposed to be informed by all possible future scenarios that impact the value of the firm. It is thus

fundamentally determined from the aggregate forecasts of investors on future performance.

Prediction markets exponentially grew in popularity since the early 1990s. “Prediction markets are defined as markets that are designed and run for the primary purpose of mining and aggregating information scattered among traders and subsequently using this information in the form of market values in order to make predictions about specific future events” (Tziralis & Tatsiopoulos, 2007). The primary purpose of this strain of research is the investigation and modelling of opinion aggregation, measurement of prediction accuracy for political and sports events and eliciting group decisions over the market coordination.

In contrast to the standard SSW experiments, previous studies with prediction markets allow for purchasing hundreds or thousands of contracts. In two most widely researched prediction markets – the Iowa Electronic Market and Hollywood Stock Market – the payoff is a) dependent on the number of votes/millions of US dollars of gross revenue after the first four weeks of the release of the movie, b) there is a much larger number of shares available on the market (in the range of thousands) where each share would pay 2.5 USD or 1 Hollywood dollar for each percentage of votes/1 million USD in the box office. Despite their complexity, these markets can successfully aggregate opinions, even when conducted in laboratory conditions with very few players (see Haely, Linardi, Lowery, & Ledyard, 2010, for example) for example.

Prediction markets have also been used to predict time of occurrence of events. For example, (Othman & Sandholm, 2013) conducted a large prediction-market study involving 210 participants (169 have put at least one order) who, for eleven months, traded 365 securities corresponding to 365 days (possible states) on which the Gates Hillman Center would open, where the definition of the building opening was a vague term (i.e. it was not defined what occupancy would determine the opening). They used monetary prizes that were randomly allocated based on the number of tickets each participant collected, while participants traded with artificial money. The price distribution over the 365 possible states could also be interpreted as probability

distribution of each event coming true. A characteristic feature of this design was the automatized market-makers that were increasing liquidity on the market.

Previous research on both prediction markets (Potters & Wit, 1996) and SSW markets (Palan, 2013) indicates that high cash-to-asset ratio inflates prices. Also, designs related to the SSW design (Haruvy, Lahav, & Noussair, 2007; Palfrey & Wang, 2012, i.e.) demonstrated over-pricing and price bubbles resulting from overly-optimistic beliefs about the future price trajectory (Haruvy et al., 2007). However, the mispricing in prediction markets has neither been extensively studied nor is there a robust method for quantifying mispricing in prediction markets.

New Paradigm

The new proposed experimental paradigm is a prediction market offering a large number of securities. We first present its particular incarnation used in the experimental results reported here, before discussing possible generalisations and broader applications.

Over a period of six full days, students of a Financial Market Risks course trade financial assets that correspond to slides of a professor, to predict the page number of the final lecture slide, on which the professor will end in the next week's lecture. The professor always prepares more slides than he can cover, he does not know exactly how many he will cover and he uploads slides a week in advance to a student portal. After the market closes, only one security - the one on which the professor finishes the lecture, pays out the dividend equal to 100 units of experimental currency, while other securities are priced at 0. Therefore, to perform well in that task, one has to trade to make a lot of cash and/or correctly predict the finishing slide by holding the corresponding securities (i.e. the promising venture).

The new paradigm is derived from two distinct experimental approaches - the classical asset market experiments and prediction markets. In this setup, dozens of securities form a complete market associated with all possible outcomes, while the underlying probability structure of the outcomes is unknown and fundamentally unknowable.

While all outcomes are known and can be priced by their corresponding assets, there is

no objective way by which participants can fully learn or estimate the probabilities of the possible outcomes. This allows us to study how well a trading market aggregates opinions into a consensus and how stable this consensus is with respect to changes in the environment.

The new setup offers various improvements to classical experimental asset markets. First, the market participants trade various numbers of assets. In real markets, the number of securities varies depending on the market and traders have to adjust. This set-up thus captures the heterogeneous structure of offered securities inherent in real financial markets. In contrast to the standard experiments with trading periods lasting a few minutes, in our setup one trading period lasts six full days.

Second, participants cannot determine with certainty the expected value of the securities and they (should) know that no one knows them (i.e., there should be common knowledge of ignorance). Our experimental setup focuses on the dynamics of pricing in an environment with limited information.

Similar to the SSW studies, in our setup, the rational value of each security is determined by the dividend. However, one does not know which security will pay out the dividend, which is analogous to composing one's financial portfolio in a highly uncertain economic and political context. The securities in our setup are a type of an "all-or-nothing" option (binary option). In sum, the market structure of this new design is characterised by persistent uncertainty about the state of the event at the end of the trading period, trading restrictions by not allowing short selling, open communication among the market players and a rank-based incentive schemes. It is important to note that the securities are not independent from each other. As in (Othman & Sandholm, 2013), the securities correspond to time-dependent states of the event of the professor finishing the lecture on a particular slide. This framework thus follows the standard definition of Arrow-Debreu securities in economics. One can have many states $S_i, i = 1, \dots, N$ of the world (many different outcomes for the end slide of the lecture). By definition, one Arrow-Debreu security s_i pays 1 if state S_i occurs and 0 otherwise. In this sense, our approach and definitions are supported by a large literature in economics.

In general, the approach derived from prediction markets makes the experimental design closer to real life due to the fact that market participants bet on one's opinion about a highly discussed topic against the others, introducing additional thrill (Wolfers & Zitzewitz, 2004). Prediction markets involve taking investment decisions with partial information under uncertainty, such that participants know that the experimenters do not know the underlying true price of the asset. On the one hand, this approach offers less control than in the standard laboratory experiments with the well-defined fundamental value. On the other hand, this approach better reflects the fact that almost all decisions that we take are under uncertainty (Pennock, 1999).

Features Mitigating Bubbles

Our experimental paradigm features several mechanisms that have been shown to mitigate bubbles in previous studies. First, we use equal endowment and a fixed and deferred dividend. Results in V. L. Smith et al. (2000) and Caginalp, Porter, and Smith (2001) show that a deferred dividend payment, and a single possible dividend, reduces the incidence of bubbles by concentrating common endogenous expectations. With a single bullet dividend, participants focus more on long-term value than on short-term gains through intermittent dividend payments.

Second, our experimental setup features a constant and relatively small cash-to-asset ratio where bidding at high prices is not possible, thus curtailing bubble formation (Caginalp et al., 2001). Kirchler et al. (2012) reports that increasing the cash-to-asset ratio due to intermittent dividend payments significantly increases the likelihood and magnitude of bubbles.

Third, since the experimental market is open throughout the week, participants will not be required to continuously monitor the market. According to the *active market hypothesis* (Lei, Noussair, & Plott, 2001), irrational trading is reinforced when participants do not have any alternative to trading actively (as is the case in a standard laboratory study).

Fourth, we allow participants to openly communicate among each other and the market

features an open book. Among others, Caginalp et al. (2001) and Oechssler, Schmidt, and Schnedler (2011) show that this reduces the incidence and magnitudes of bubbles. One possible explanation is that when traders receive information about the motivations, strategies and dispositions of other players (revealed by price, bids, and order evolution), they integrate the optimisation strategy of others in their own strategy. In the game-theoretical reasoning, a trader who has access to the strategies of others will account for the reasoning of others (Caginalp et al., 2001).

Finally, our market has a large number of securities. Despite the mixed evidence with only two assets (see Ackert, Charupat, Church, & Deaves, 2006; Ackert, Charupat, Deaves, & Kluger, 2009; Caginalp, Ilieva, Porter, & Smith, 2002; Chan, Lei, & Veseley, 2013; Fisher & Kelly, 2000, for example), the overall direction of these previous results indicates that multiple assets tend to reduce overly exuberant pricing especially if the assets differ.

Experimental Software Enables Extensions of the Design

We implemented the design using the xYotta³ software developed at ETH Zurich in the Chair of Entrepreneurial Risks, with support from the rectorate of ETH Zurich as well as the ETH-Singapore Research Center, for the purpose of conducting classroom experiments and practical student education on financial markets. The functionality of the software offers a number of extensions of various complexity of the experimental design presented here. First, xYotta enables running a few parallel markets at a time with the same users or with different users. Every user can access any number of markets as long as they are granted the access code from the administrator (i.e. experimenter).

Second, beyond the market, the software has a feature that enables collaborative working in a project where, as rewards for contributions to the projects, one can earn experimental money. xYotta can make each project go through an IPO (initial public offering), where stocks are issued, which can then be traded on a financial market

³<https://xyotta.com>

reproducing exactly the Swiss Financial Market of the Swiss Stock Exchange Market ⁴. xYotta thus provides a rather faithful analog of a real economy in which entrepreneurs create start-ups, which then grow collaboratively through hard work, and which can then become publicly traded (within the xYotta ecosystem). The market values of the traded stocks reflect how the population of traders (students) buy and sell to grow successful portfolios, based on their best assessment of the quality of the projects, and also as a result of a “beauty contest” mechanism present in any financial market.

Third, the amount of cash, number of securities, values of securities, values of dividends, opening times of the market can be freely adjusted by the administrator.

This allows for, for example, investigating different compensation schemes or the effect of the number of securities on the market dynamics.

Fourth, the software is on-line based and can be accessed from any place via an internet browser, with any number of participants. Thanks to this feature, experiments can be conducted in any place at any time.

Therefore, we invite researchers interested in using the software for investigation of complex systems with human players to contact us to receive access to the xYotta software. In a similar fashion as the Marketplace software of the California Institute of Technology ⁵ revolutionised the research on experimental asset markets (see for example Noussair, Plott, & Riezman, 2007) as an example of a complex multimarket game), we believe that introducing software and design that enables creation of experimental environments of various complexity can open a new range of possibilities for empirical research of financial markets.

Goals of Experiments

The main goal of Experiment 1 was to test whether mispricing would occur in a design more realistic than the SSW design, which includes features that should mitigate overpricing. The second goal of the experiment was to investigate the emergence of opinion in a situation of inherent uncertainty and the development of that opinion in

⁴http://www.six-swiss-exchange.com/rule_book/01-RB_en.pdf

⁵<http://donut.caltech.edu/marketplace/>

the course of coordination among traders over the bid and ask offers. An additional aim of this experiment was to explore the trading behaviour over the six full days of trading in the new setup based on prediction market methodology.

Experiment 2 had three main goals. First, we aimed to measure the robustness of key results from Experiment 1 by replicating it with a different group of participants: a) quick price emergence, b) approximately constant price across the week, and c) mispricing as indicated by the market index. The second goal was to explain the emergence of consensus about the price of securities observed in the first experiment. Towards this aim, we introduced a belief elicitation mechanism before and after each trading round. Third, based on the results from Experiment 1, we improved the experimental procedure to close even more the gaps between the real markets and experimental markets.

Method

Participants

In the Fall semester 2015, 221 students were enrolled in the course of Financial Market Risks for Master students at the ETH Zurich (Swiss Federal Institute of Technology in Zurich). 122 students participated in the experiment and 95 of them took the exam at the end of the course. The experiment was voluntary and offered the possibility to obtain bonus credit points added to the exam grade of the course. The maximum grade for the course cannot be exceeded in the case when the sum of the exam grade and the bonus would be larger than 6.0. 80% of the participating students were male. Due to the personal data protection of students attending the course, we did not collect any demographic data.

Procedure

As outlined in panel B of Figure 1, the class was held on Mondays at 10:15am to noon each week. Every Monday afternoon, the professor uploaded slides for the next lecture and these were accessible by all participants. All possible outcomes were known to

participants before trading for the week began, so that the market was a complete contingent market. Based on the content of the slides, the participants could form an educated guess about the likelihoods of different outcomes. Their task was to make money by translating their expectations into prices and trading accordingly. On the next Monday, at the end of the class, the realised security was recorded, announced to all students and publicly confirmed by the lecturer. The securities that were not realised did not pay any dividend and were priced at 0. Please note, that professors with more structured lecturing style could ask students to predict the slide that will be uploaded during the lecture at some predefined time, say 45 minutes into the lecture. Even the most disciplined and well-prepared lecturer exhibits some variability in her pace, which can be used as the stimuli source.

The market was open every day from Tuesday to Sunday between 8am and 10pm. Every day, there was a market pre-opening at 6-8am. We selected the market opening times based on the trading activity in Experiment 1⁶. All buy orders had to be covered by sufficient cash in their account and sell orders were only allowed if the participant had the necessary quantity of securities in their portfolio. No short sells and no buying on margin was allowed. The trading rule follows the standard continuous double auction mechanism – a trade was successful only if there was a buyer that wanted to buy one or more units of a security for a price at least as high as a seller was offering. For Experiment 2, the trading experiment was announced in the second week of the semester. In week 4, the participants could take part in the practice period, which did not count for their final rank. The periods that counted for the final rank lasted for four weeks.

The initial prices were established in the pre-opening phase when participants could submit their orders but the orders were not executed. Most of the submitted orders were limit orders. When a sell order was lower than the buy order, the order would be executed at the lower price, making this price the official market price. A number of securities assessed by students as highly improbable were not or almost not traded. In

⁶Results regarding trading activity are provided in Appendix C.

the rare absence of trade and of pre-opening orders, the default price was set to zero. The early securities were mostly priced but they were the so-called “penny stocks.” At the beginning of each week (trading period), participants were endowed with 300 units of experimental currency and 3 units of each security, which was equal to 600 units of loan that they had to repay to the experimenters after the market closed. Results in King, Smith, Williams, and van Boening (1993) indicate that unequal endowments and the associated portfolio rebalancing motives are not necessary to induce participants to trade actively. Because the rational price of a complete set of assets is 100 (as further explained in Section Mispricing and Rationality), the market had a cash-to-asset ratio of 1. The market was completely reset every week, so no asset or cash was carried over to the following rounds. Before the experiment, the participants could take part in one practice period lasting one week. The purpose of the practice period was to let the participants familiarize with the trading software and the task. Performance during the practice period was not included in the participants’ final earnings.

At the end of each week after the realised state had been announced, the cash holdings and any earned dividends were added together to form a ranking of participants. The ranking was based on the following formula for the earnings of participant i in week t :

$$\text{Earnings}_{i,t} = \max \{ \text{Cash}_{i,t} + \text{Dividends}_{i,t} - 600, 0 \}. \quad (1)$$

The total ranking was not published but the online platform allowed participants to see their weekly ranking. The instructions of the experiment are outlined in Appendix A. In order to activate their trading accounts, every week, the participants had to submit their belief about the success of each security in the market. We used the modified *roulette prior belief elicitation* method (Gore, 1987; Johnson et al., 2010; Morris, Oakley, & Crowe, 2014). Participants were asked to allocate 100% of their belief among all available securities before and after each week’s period. They were presented with a dynamic bar diagram with all securities listed on the x-axis and the belief expressed in percentages on the y-axis. The participants could allocate any natural number between 0-100 to any security and to any number of securities, as long as the sum of the allocated belief was equal 100. By default, the participants were presented with uniform

allocation of the probability and could drag and block the bar for each security, while the bars of the remaining securities would automatically adapt to ensure normalisation. After the professor uploaded the new stack of slides on Monday afternoon, he submitted his own belief in the same way as the participants did.

The belief elicitation was not incentivised separately from the trading task. The participants were asked to give their honest belief and were told that their submitted belief is anonymous and will have no influence on their final grade from the course. However, submission of beliefs was the necessary requirement for opening and closing of the portfolio. Therefore, the participants had no incentive to provide false or misleading information during the belief submission, but they were aware that the experimenters had access to their submitted beliefs during the course of the experiment. However, we cannot rule out the fact that some students provided dishonest belief. While submitting the second belief, the participants were presented with their pre-trading belief for reference. The participants could enter the market and activate their account by submitting the pre-trading belief at any time between Tuesday 6am and Saturday 10pm. In order to have one's portfolio in a given week included in the final ranking, the participants had to submit their second belief between 10pm on Sunday (when the market closes) and 10am on the following Monday.

After completion of the four week trading sessions, the participants were asked to fill out a questionnaire that included questions regarding the cues that they used to predict the end-slide, whether they realised that the sum of prices should not exceed 100, whether they used the opportunity when the prices exceeded 100 to apply arbitrage and what other trading strategies they used. 114 participants responded to the questionnaire.

Materials and Apparatus

Due to the large number of slides used by the professor during each lecture, we grouped them into sets of 3. In other words, one security corresponded to 3 consecutive slides, such that security 1 covered slides 1, 2 and 3, security 2 covered slides 4, 5 and 6, etc. For instance, if the uploaded presentation contained 69 slides, this would give 23

securities, the first security for slides 1-3, the second one for slides 4-6, ..., and the 23rd security for slides 67-69. In each of the four weeks for Experiment 2, the lecture slide decks had 168, 157, 144 and 83 slides, which corresponds to 69, 54, 49 and 29 securities. Within each set of securities, there was one security that corresponded to the class not taking place due to unexpected events (e.g. when the professor being sick). To avoid ambiguity, the number of slides referred to the actual number of the pdf page of the slide shown in the lecture.

The set-up was “double blind,” in the sense that the participants did not know on which slide the professor would end, and the professor did not know how participants were betting. Moreover the professor himself did not know precisely on which slide he would finish the class. The professor covered 44 (22%), 67 (43%), 61 (42%) and 60 (72%) of all the slides, which corresponds to 15, 23, 21 and 21 securities. Slides that were not covered in one lecture, were added to the new stack of slides for the next lecture ⁷.

The schedule and the course content are the same every year, but the professor adapts his slides depending on the important financial and political events in the world, new research or other changes that should be implemented in the course.

Compensation

At the end of the four trading periods (weeks), participants were rewarded with bonus grade points that was added to their final grade from the course. According to the ranking, the top quartile (best 25%) of participants with the highest earnings received 0.5 point bonus grade to their final exam grade. The next quartile (second best 25%) received a bonus of 0.25 grade point. The rest of the participants did not receive any bonus point. The grade system is based on a 6-point system ⁸, such that 6 is the higher grade that is usually obtained in case of exceptionally good student performance. The

⁷For a taste of the professor’s teaching style, see the video lectures at <http://www.er.ethz.ch/media/presentations/Videos.html> or the TED Global talk at http://www.ted.com/talks/didier_sornette_how_we_can_predict_the_next_financial_crisis?language=en that was scheduled to last 15 minutes, took 18 minutes and the producers reduced it to 17 minutes.

⁸See <https://www.swissuniversities.ch/en/higher-education-area/swiss-education-system/grading-system/> for the explanation of the Swiss grading system.

minimum grade required to pass a course is 4 and the final grade is awarded with 0.25 steps (i.e. if a student receives 4.15 from an exam, the final grade will be rounded up to 4.25). Therefore, half a point grade bonus can help a student with an exam grade of 3.5 pass the course and is highly valued by the students. Note that students could also receive the maximum grade from the course without participating in the experiment and by performing well in the exam.

During the experiments, we observed that participants had the intrinsic motivation to experience a realistic trading experience, which often is part of the curriculum of many finance courses. Loewenstein (1999) outlines that monetary incentives in experiments do not fully relate to the real monetary incentives and can be powerfully influenced by other motives, such as social aspect or a desire to appear smart, etc. We direct a curious reader to Andraszewicz, Wu, and Sornette (under review), where we provide a discussion on the incentive compatibility in different experimental settings.

Results

Market Prices and Beliefs

To examine the distribution of prices over each week, we use the median price of all transactions within an epoch (e.g., a 4-hour block) as the price per security. As presented in Figure 2 where the securities on the x -axis are sorted consecutively, relative prices reflect the market assessment of the likelihood ratio of two states (the dividend paying out or not). Hence, the distribution of prices provides a direct representation of the market assessments for each week of the likelihood of the lecture stopping at a particular block of slides.

Each week, there are a number of securities with essentially zero prices and almost no price fluctuations. For these securities, the participants seem to agree that the corresponding block of slides is very unlikely to be realised. In each round, the peak of the distribution falls close, but not exactly on the realised security indicating high predictive power of the market. The market price distribution is more spiky than both belief distributions (as indicated by the smaller entropy of the distributions listed Table

1), but this discrepancy between the market and the beliefs disappears progressively from week 1 to week 4, as indicated by decreasing kurtosis differences between the market and the two beliefs across the four weeks (see Table 1).

The prices emerged early during each week and stayed relatively constant until the end of the trading period. This pattern occurs in all trading periods (see Figure 3 in the main text and Figure C.3 in Appendix C). We computed Jensen-Shannon Divergence (JSD) for the end-of-day prices from all six trading days. JSD is a measure of similarity between two distributions (see Table 3). JSD values across the week are close to 0 indicating very high similarity. It appears as if participants use existing prices to inform their probability assessments, which can be interpreted as a behaviour consistent with the status quo bias that prevents deviations from initial prices even in the presence of persistent uncertainty (Fleming, Thomas, & Dolan, 2010; Kahneman, Knetsch, & Thaler, 1991; Samuelson & Zeckhauser, 1988).

Further, according to Figure 2, the distribution of average pre- and post-trading beliefs are very strongly aligned with the price distribution in each week. The post-trading distribution is more strongly correlated with the price distribution ($r = .98, .98, .97, .98, p < .001$ for weeks 1-4) than the pre-trading belief ($r = .91, .97, .93, .98, p < .001$ for weeks 1-4) and both belief distributions are slightly less correlated with each other than with the market ($r = .88, .97, .96, .98, p < .001$ for weeks 1-4). We conducted a multiple correlation analysis with the pre-trading and the market price distribution as two independent variables correlated with the post-trading belief (the dependent variable in the regression). According to Table 2, the sum of the two coefficients of the two factors is close to one in all 4 weeks, with high R^2 values, showing that the two factors explain the post-trading belief very well. More importantly, the market impact on the post-trading belief decreased across the four weeks, while the impact of the pre-trading belief increased.

Additionally, to account for multicollinearity in the multiple correlation analysis, we conducted a linear regression analysis, in which we use the difference between the pre-training belief and the market as a predictor of the pos-trading belief (see Table 2).

This analysis indicates an increasing impact of the difference between the pre-trading belief and the market on the post-trading belief. Overall, these results indicate that the participants learned that there is no new information in the market and learned to ignore the opinion of the others. In other words, we observe the emergence of a communal ignorance across the four trading periods. These strong correlations support the notion that the price distribution was a result of the aggregated a priori beliefs of the market participants, which was then further consolidated in the post-trading beliefs.

Figure 4 shows that the individual beliefs of each participant were close to the security that paid the dividend. Overall, all participants assigned belief weights to the same group of securities. However, there is a large variability in the number of securities that each participant assigned weights to - some participants diversified their beliefs among many securities, while some participants indicated that only 2-3 securities as likely to pay out the dividend. Participants were consistent in applying the same strategy of assigning their belief to very few or many securities in pre- and post-trading belief elicitation. However, they adjusted their beliefs after experiencing the market, such that their post-trading beliefs became less divergent according to the average JSD measures for the pre-trading and post-trading beliefs (Pre-trading: 1.12, 0.88, 0.91, 0.77; Post-trading: 1.05, 0.75, 0.92, 0.72 for weeks 1-4). These adjustments resulted in shifts of the belief by a few securities only.

The belief of the professor was also elicited, following the same procedure as for the participants. The professor's belief was divided among 4 to 5 securities, each having 15%, 20% or 25% of the assigned weight. In weeks 1, 3 and 4, the security that paid dividend turned out to be either the first or the second security to which he assigned any weight. As the securities are ordered by increasing slides, this corresponds to the security paying the dividend being the lowest or second lowest guess of the professor on which slide he will end up the class with. In week 2, the security that paid dividend was before any of the securities indicated by the professor. Overall, the professor was over-confident about the number of slides that he would be able to cover during the lecture. The average distance between the expectation of the distribution in each week

and the security that paid dividend was higher for the professor ($M = 5.14$, averaged across all four weeks) than for the participant pre-trading belief ($M = 4.64$), participant post-trading belief ($M = 4.31$) and the market ($M = 3.56$). This confirms the fact that the experiment was double-blind.

Mispricing and Market Rationality

It is important to note that it cannot be said whether the prices at which securities are traded are rational, as there is no knowable fundamental value. However, in our setup, we are still able to make normative statements about the total price level of the market. Since one and only one security pays off 100 units of currency, the sum of the security prices – *the market index* – should equal 100 at all times. If the sum was above 100, it would be profitable to sell one unit of every security and vice versa (assuming that the mispricing would eventually to zero).

Figure 5 presents the progression of the sum of all prices for each week in real time. The two other indices included in Figure C.4 provide information on whether the market was a so-called “buyer’s market” or “seller’s market”. Most of the time, one can observe that the sum of highest bid prices is much closer to the smoothed sum of prices than is the sum of lowest ask prices. The latter often tends to be much larger, suggesting that this market was mostly a buyer’s market, i.e. supply for securities exceeded demand so that buyers can buy at low prices.

The sums of highest bids and lowest asks can also help identify periods of blatant arbitrage opportunities – if index 2 (the sum of highest bids) is larger than 100, the arbitrage opportunity can be exploited by selling one share of each security (with the sum of the sale values being above 100) and thus obtain a certain profit at maturity. Such strategy would tend to push down the overall price level. If index 3 (the sum of lowest asks) is lower than 100, an arbitrage opportunity would also occur, which could be implemented by buying one share of each security, with the sum of the paid prices being smaller than 100. There is a clear but short-lived arbitrage opportunity in week 1, while arbitrage was possible most of the time in weeks 2-4. Also, every week starts with

very high ask prices, which stabilise across the week, apart from week 1, which is characterised by high price fluctuation. In weeks 2-4, the sum of highest bids is almost always above 100 revealing the presence of overpricing.

As shown in Figure 5, the market is persistently over-priced in all trading periods. However, in week 4, the index is much closer to the 100-level than in earlier weeks, indicating a learning effect across periods. The common pattern is that the over-pricing is larger at the beginning of each trading period and it decreases towards the end of the period. Each week starts with very high ask prices. The overpricing decreases throughout the week but remains until the end of the trading. This mispricing resulted in two arbitrage opportunities in week 1.

To quantify the mispricing of the market, we computed the Relative Deviation (RD) (Stöckl, Huber, & Kirchler, 2010) of the three indices. This measure, outlined in Table 4, indicates that in both experiments, the over-pricing decreases from week 1 to week 4 (according to index 1). The Relative Deviation values for indices 2 and 3 indicate that the arbitrage opportunities in all four weeks in the two experiments were very limited.

Grades and Performance

There was a strong correlation between the number of submitted orders and total earnings by participants from all four weeks: $r = .51, p < .001$. Also, the participants with the bonus points had higher grades in the exam (Kruskal-Wallis, $p < .001$), such that the median grade of the participants with 0-bonus was 4.25 (Experiment 1 and 2), of participants with .25 bonus it was 4.50 and 4.75 in Experiments 1 and 2, while the participants with .5 bonus points from the trading would obtain a median grade of 5.63 and 5.5 in Experiments 1 and 2 (without counting the bonus). This means that performance was related to the traders' activity, knowledge and involvement in the course, involving a cumulative effect reminiscent of the Matthew effect (Merton, 1968, 1988), which states that "the rich get richer and the poor get poorer", or here in the context of grades, those were the "richest" in their grade obtained the bonus, the "poorest" grades did not get anything.

Extensions of the New Experimental Paradigm

As a follow-up of the initial experimental design outlined here, we collected data from 17 additional experimental rounds. We report the first four experimental rounds in the doctoral thesis of Ke Wu⁹. This experiment offers interesting insights but remains too preliminary to be reported in full in a scientific journal. Next, three experimental rounds conducted in the laboratory are reported in detail in Andraszewicz et al. (under review). We do not publish the last ten experimental rounds due to the fact that we experienced technical details problems during the data collection and we expect that, despite interesting insights, the experiment would need a repetition to claim robust findings from this experiment. Nevertheless, here, we summarise the variations of the experimental design introduced in this paper and outline the key findings to provide the complete picture of the work we have done to test this new experimental design.

First, we conducted four rounds following Experiment 2 conducted in Fall semester 2015 using the same participants as in Experiment 2. After the four rounds in Experiment 2, we asked participants to reply to a survey in which we asked them whether they were aware of the fact that the sum of prices should not exceed 100. 66% of the participants indicated that they realized that the sum of prices should be exactly 100 but only 44% of them knew that they could use this fact to apply arbitrage. After these four weeks, we implemented a quantitative tool showing live values of three indices (index 1 – sum of prices, index 2 – sum of lowest bids, index 3 – sum of highest asks) and the participants received a half-an-hour training about how to use these indices to apply arbitrage. Over the course of two experimental rounds (weeks), we observed lower mispricing than in the four previous weeks. In the next two weeks, we sent by e-mail to a randomly selected half of all participants the subjective opinion of the professor's teaching assistants about the security that they believe the professor would end the lecture with. Participants, who did not receive the information in one week, received it in the next week, to treat all participants fairly. All participants knew

⁹available at https://www.ethz.ch/content/dam/ethz/special-interest/mtec/chair-of-entrepreneurial-risks-dam/documents/dissertation/master\%20thesis/phd_thesis_kewu.pdf

that we would distribute the information, at what time and that the information is subjective and may or may not be true. In the first week of this treatment, we observed a price bubble on the indicated security and on the neighboring securities, directly after the release of the information. In the second week of this treatment, participants tried to anticipate which securities would be indicated in the distributed information and the prices of these securities increased just before the release of the information. In both weeks of the treatment, the predictions of the teaching assistants turned out to be close to but different from the correct security.

In three rounds in the laboratory, which we report in detail in Andraszewicz et al. (under review), 1) we investigated whether the same main effects in the laboratory and in the classroom settings could be observed and how the market dynamics changes for the two experimental environments and 2) we compared the effectiveness of monetary versus grade compensation schemes in obtaining reliable experimental results. In sum, we concluded that the laboratory and the classroom settings result in the same main effects, offering the possibility to relax some of the laboratory control to experiment with larger samples in more natural environments. However, we found some differences in the market dynamics resulting mainly from the task complexity and compressing this relatively difficult task to a short laboratory setting. Also, we concluded that grades provide sufficient motivation to students participating in a classroom experiment. Further, we found that competitive monetary payments are incentive compatible for students participating in laboratory experiments.

Finally, over the five weeks (rounds) in the Fall Semester 2016, we implemented two parallel markets in which students of the Financial Market Risks at ETH Zurich were asked to predict the next week percentage change of the S&P500 index and the next day percentage change of the FTSE100 index. Each week, the number of securities was the same, such that each security corresponded to the range of the percentage change (i.e. participants' confidence interval). These price change intervals were generated based on 20 years of historical weekly/daily price changes such that, based on these historical data, the likelihood of each security coming true was uniform. The

experimental S&P500-based market was open for twelve hours each day over six days, while the experimental FTSE100-based market was open for only two hours the day before which the price change was predicted. The main aim of implementing two markets was to compare the market dynamics when the trading is only allowed at strictly designated time and when trading is allowed at various times over a few days. Also, in the S&P500-based market, each trade could only be submitted after stating one's motivation for making the particular trade. In the trading platform, we implemented four choice options that participants could indicate: 1) fundamental analysis of S&P500, 2) technical analysis of S&P500, 3) technical analysis of the market performed during the class, 4) gut feeling. Participants were also allowed to insert a free comment. The key findings indicated that people tend to center their beliefs about the success of each security around the middle prices of the available securities, resulting in bell-shaped price and belief distributions. The S&P500 and the FTSE100 markets did not differ in the number of submitted orders and the number of active traders. Also, the most frequent reason to submit a trade was gut feeling, followed by technical and fundamental analyses from the S&P500.

Discussion

In two experiments, we tested a new design for investigating experimental asset markets. This design substantially differs from the well-established SSW design (V. L. Smith et al., 1988). The aim of implementing the new design was twofold. First, we investigated whether the “bubble-and-crash” pattern is a typical phenomenon only found in this type of experimental markets, or it is a general bias that is reflected in other artificial and real markets. Second, we aimed at testing the coordination of opinions in a situation of intrinsic uncertainty.

In the new design, the market players not only have to agree on the “right” price of each of dozens securities, but also have to predict a real future uncertain event whose outcome affects the market. This experimental setup employs a prediction market approach to study stylised results observed in real financial markets and classical asset

market experiments. In contrast to standard prediction market studies, the result of our market directly impacts the traders and is experienced by them after the market closes. To do well in such a market, one has to position his/her portfolio by anticipating the outcome of this event, while using the opportunities in the market to obtain cash. Such situations are very common in real markets. As recent examples, let us mention Brexit in June 2016, the US elections in November 2016 and the Italian referendum in December 2016, whose anticipations influenced the asset allocations of investors according to their beliefs and how they would impact the security prices. Evidence for this can be found in the large price impact that these events triggered on financial markets as a result of the reassessment of investment opportunities following each event, leading to significant changes in portfolio allocations (Wu, Wheatley, & Sornette, 2018). In the two reported experiments, we observed a few robust effects. First, market players quickly approached a consensus price in spite of the intrinsic Knightian uncertainty, i.e. the lack of well-defined fundamental value resulting from the impossibility to know the probabilities of different outcomes. The market emerged despite the lack of initial price and showed higher predictive accuracy than the professor himself - who was the underlying stochastic process.

The price consensus occurred already in the pre-opening phase as a result of the convergent beliefs of individual players and the price was merely fine-tuned during the actual trading. In contrast to Othman and Sandholm (2013), we observed much smaller price fluctuations across the duration of the experiment. In their setup, human participants were exploiting high volatility in the market for arbitrage purposes. This is an important finding that reaches beyond standard prediction market experiments, in which researchers traditionally focus on the change of the price over the trading period. Instead, our report of an early agreement on the price reveals the effect of coordination facilitated by the information flow provided by the order book of bid and ask quotes. This transcends the “wisdom of crowd” phenomenon, whose mechanism is based on the averaging of distributed noise to make a small systematic signal emerge by aggregation. Second, the price emergence was strongly influenced by the prior beliefs of individual

market participants, whose initial beliefs were remarkably convergent, despite the intrinsic uncertainty. Participants based their beliefs on vague information in the historical data to extrapolate the future events.

Third, substantial overpricing occurred, despite the features of the new design mitigating bubbles. This shows a general bias that persistently occurs in the markets. However, the overpricing pattern shows much more variability and departs from the typical “bubble-and-crash” scenario found in previous experiments. We found that roughly half of the participants were aware of the mispricing but this situation could not have been always arbitrated due to insufficient liquidity in the market (Appendix B provides more detailed analysis.). This could be one possible explanation for why mispricing is so persistent in experimental settings, despite the large number of market participants. Also, not all mispricing observed in the markets may be irrational but instead, it may partly result from various constraints as discussed in a vast literature (e.g. Palan, 2013; Powell & Shestakova, 2016).

By replicating the experiment a year later with a different group of participants, we conclude that all observed behavioural effects are independent of the content of the lecture slides. Our two experiments used different decks of slides and started at two different time-points of the semester. Nevertheless, the mispricing is the most pronounced in week 1 of both experiments, the mispricing diminishes over time and the price distribution over time stays approximately constant across the week. The participants were able to adjust their expectations and strategies, given the new securities and differing number of securities. Most of the participants spent relatively little time on analysing the content of the slides, but rather developed technical analysis tools.

In our setup, paying a high price for securities that one likes would lead to under-performance, because the final value that counts for the grade bonus is the sum of cash plus the dividends of only one security (reflecting the real-world “all-or-nothing” security) and does not take into account the value of the portfolio at the last trading price. In other words, the value of the portfolio of each participant was just determined

by the payoff of the held securities at maturation, i.e. when the winning slide / security was revealed: namely, 0 value for all securities except the one containing the slide ending the lecture. Therefore, inflated prices that increase the instantaneous values of participants' portfolios were irrelevant for the final valuation of the portfolio of each participant.

The current setup suggests numerous extensions. Most of the design changes we introduced that can be compared to traditional experimental asset markets are such that excessive speculation and bubble formation are decreased. Recall that our market features delayed, bullet dividends, a low cash-asset-ratio, a large number of assets (30-60), a long time horizon (one week), equal endowments among participants and reward unrelated to (or even deterring) high prices. Our two experiments were conducted in a semi-controlled environment, where the trading environment was fully controlled by the experimenters and no non-course related events had an impact on the market. However, the conditions in which the participants traded were not controlled. This approach mimics well the real-life situation of individual investors. Varying these features would be interesting to understand the conditions under which even more bubbly at one extreme or rational markets at the other extreme can emerge. Also, making the market fully multi-periods where capital can accumulate could result in different price distributions and evolutions during each trading period.

The new experimental approach proposed here does not rely on the specific teaching style of the lecturer, even though its design was inspired by that of the inventor of the design, who is also the lecturer and the first author of the current paper. Indeed, Prof. Sornette has a particularly unpredictable teaching style, which we tested during Experiment 2 by taking notes and conducting a quantitative analysis on the time devoted to each slide, number of slides covered during one lecture, number of skipped slides etc. (for a more elaborate discussion on the analysis of the professor's lecturing style, see Andraszewicz et al., under review). This analysis showed no predictive power of any of these measures. This may not be the case for other lecturers, for which a simple change is to make the dividend be dependent on the slide, which the lecturer

discusses at, for example, 30th minute in the lecture. Even for very controlled lecturers, it may be difficult to always discuss the same number of slides within the designated time, thus allowing the design to exploit an always present degree of randomness. Moreover, the generalisation discussed above involving the prediction of a financial index over some fixed future time horizon removes the need to link the design to teaching and lecturing style if desired.

The proposed experimental design still does not completely mimic real financial markets. For example, the step-wise payoff function as well as the highly-valued dividends are not very common in real markets. However, the experiments described here present an example among many others that can be developed of how one can implement ideas from prediction markets within our xYotta software platform for the investigation of market dynamics and potential mispricing. Payoff schemes and stochastic processes underlying the success of each event/asset on the market can be varied at will, depending on the research question.

Finally, it is important to note that despite the fact that our market participants had a quite accurate a priori belief about the success of each security, there were persistent errors, suggesting that prediction markets involving real Knightian uncertainty, and financial markets in particular, are useful in the face of intrinsic uncertainty but are not panacean oracles. This could help explain why, in real markets, sometimes resources are allocated to losing ventures and biases are persistent across a substantial number of the market players and over long trading periods.

Acknowledgments

This work was partially supported by a Seed Project Grant from the Risk Center at ETH Zurich. This work was also partially supported by funding from the European Union's Horizon 2020 research and innovation programme for the ITN CRITICS under Grant Agreement Number 643073. We would like to thank Christoph Hölscher and attendants of the Conference in Experimental Finance in Mannheim in 2016 for very useful comments. The data is available upon request.

References

- Ackert, L. F., Charupat, N., Church, B. K., & Deaves, R. (2006). Margin, short selling, and lotteries in experimental asset markets. *Southern Economic Journal*, 73, 419 – 436.
- Ackert, L. F., Charupat, N., Deaves, R., & Kluger, B. D. (2009). Probability judgment error and speculation in laboratory asset market bubbles. *Journal of Financial and Quantitative Analysis*, 3, 719 – 744.
- Andraszewicz, S., Wu, K., & Sornette, D. (under review). Behavioural effects and market dynamics in field and laboratory experimental asset markets.
- Ball, S. B., & Holt, C. A. (1998). Classroom games: Speculation and bubbles in an asset market. *Journal of Economic Perspectives*, 12(1), 207 – 218.
- Berg, J. E., Forsythe, R., Nelson, F., & Rietz, T. (2008). Results from a dozen years of election futures markets research. In C. R. Plott & V. L. Smith (Eds.), (Vol. 1, pp. 724 – 751). Elsevier, North Holland.
- Berg, J. E., & Rietz, T. A. (2003). Prediction markets as decision support systems. *Information Systems Frontiers*, 5, 79 – 93.
- Black, F. (1986). Noise. *The Journal of Finance*, 41, 529 – 543.
- Bostian, A. A., & Holt, C. A. (2009). Price bubbles with discounting: A web-based classroom experiment. *Journal of Economic Education*, 40(1), 27 – 37.
- Caginalp, G., Ilieva, V., Porter, D. P., & Smith, V. L. (2002). Do speculative stocks lower prices and increase volatility of value stocks? *Journal of Psychology and Financial Markets*, 3, 118 – 132.
- Caginalp, G., Porter, D. P., & Smith, V. L. (2001). Financial bubbles: excess cash, momentum, and incomplete information. *Journal of Psychology and Financial Markets*, 2, 80 – 99.
- Chan, K. S., Lei, V., & Veseley, F. (2013). Differentiated assets: An experimental study on bubbles. *Economic Enquiry*, 51, 1731 – 1749.
- Chesney, M., Cramer, R., & Mancini, L. (2015). Detecting abnormal trading activities in option markets. *Swiss Finance Institute Research Paper No. 11-42*, Available at

SSRN: <http://ssrn.com/abstract=1522157>.

- Cowgill, B., Wolfers, J., & Zitzewitz, E. (2009). Using prediction markets to track information flows: Evidence from Google. In S. Das, M. Ostrovsky, D. Pennock, & B. Szymanski (Eds.), *Auctions, market mechanisms and their applications* (Vol. 14, p. 3-3). Springer Berlin Heidelberg.
- Deck, C., & Porter, D. (2013). Prediction markets in the laboratory. *Journal of Economic Surveys*, 27(3), 589 – 603.
- Elberse, A., & Eliashberg, J. (2003). Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Science*, 22, 329 – 354.
- Fisher, E. O., & Kelly, F. S. (2000). Experimental foreign exchange markets. *Pacific Economic Review*, 5, 365 – 387.
- Fleming, S., Thomas, C., & Dolan, R. J. (2010). Overcoming status quo in the human brain. *Proceedings of National Academy of Sciences, USA*, 107, 6005 – 6009.
- Forsythe, R., Nelson, F., Neumann, G. R., & Wright, J. (1992). Anatomy of an experimental political stock market. *The American Economic Review*, 82, 1142 – 1161.
- Forsythe, R., Rietz, T. A., & Ross, T. W. (1999). Wishes, expectations and actions: a survey on price formation in election stock markets. *Journal of Economic Behavior and Organization*, 39, 83 – 110.
- Friesen, L., & Gangadharan, L. (2013). Enviromental markets: Wwhat do we learn from the lab? *Journal of Economic Surveys*, 27(3), 515 – 535.
- Gore, S. (1987). Biostatistics and the medical research council. *Medical Reseach Council News*, 35, 19 – 20.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70, 393 – 408.
- Guala, F., & Mittoe, L. (2005). Experiments in economics: External validity and the robustness of phenomena. *Journal of Economic Methodology*, 12, 495 – 515.
- Gurkaynak, R. S. (2008). Econometric tests of asset price bubbles: taking stock.

- Journal of Economic Survey*, 22, 166 – 186.
- Haely, P. J., Lopardi, S., Lowery, R., & Ledyard, J. (2010). Prediction markets: Alternative mechanisms for complex environments with few traders. *Management Science*, 56(11), 1977 – 1996.
- Hansen, J., Schmidt, C., & Strobel, M. (2004). Manipulation in political stock markets - preconditions and evidence. *Applied Economics Letters*, 11, 459 – 463.
- Haruvy, E., Lahav, Y., & Noussair, C. N. (2007). Traders' expectations in asset markets: Experimental evidence. *The American Economic Review*, 97(5), 1901 – 1920.
- Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13, 517 – 523.
- Ho, T. H., & Chen, K.-Y. (2007). New product blockbusters: The magic and science of prediction markets. *California Management Review*, 50, 144 – 158.
- Holt, C. A., Porzio, M., & Song, M. Y. (2017). Price bubbles, gender, and expectations in experimental asset markets. *European Economic Review*, 100, 72 – 94.
- Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J., Grosbein, H. A., & Feldman, B. (2010). A valid and reliable belief elicitation method for Bayesian priors. *Journal of Clinical Epidemiology*, 63, 370 – 383.
- Kahneman, D., Knetsch, J., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5, 193 – 206.
- Kain, K. J., & Logan, T. D. (2014). Are sports betting markets prediction markets?: Evidence from a new test. *Journal of Sports Economics*, 1, 45 – 63.
- King, R. R., Smith, V. L., Williams, A. W., & van Boening, M. V. (1993). Nonlinear dynamics and evolutionary economics. In R. Day & P. Chen (Eds.), (pp. 183 – 200). Oxford University Press, New York.
- Kirchler, M., Huber, J., & Stöckl, T. (2012). Thar she bursts, reducing confusion reduces bubbles. *The American Economic Review*, 102, 865 – 883.
- Knight, F. H. (1921). *Risk, uncertainty and profit*. Boston, New York: Houghton

- Mifflin Company.
- Lei, V., Noussair, C., & Plott, C. R. (2001). Non-speculative bubbles in experimental asset markets: Lack of common knowledge of rationality vs. actual irrationality. *Econometrica*, *69*, 831 – 859.
- Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioural economics. *The Economic Journal*, *109*, F25 – F34.
- Mannes, A. E. (2009). Are we wise about the wisdom of crowds? the use of group judgments in belief revision. *Management Science*, *55*, 1267 – 1279.
- Manski, C. F. (2006). Interpreting the predictions of prediction markets. *Economics Letters*, *91*, 425 – 429.
- Merton, R. K. (1968). The matthew effect in Science. *Sciencel*, *159*(3810), 56-63.
- Merton, R. K. (1988). The matthew effect in Science, ii. *ISISl*, *79*, 606-623.
- Morris, E., David, Oakley, E., Jeremy, & Crowe, A., John. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental and Modelling Software*, *52*, 1 – 4.
- Noussair, C., Plott, C., & Riezman, R. (2007). Production, trade, prices, exchange rates and equilibration in large experimental economies. *European Economic Review*, *51*, 49 – 76.
- Noussair, C., & Tucker, S. (2013). Experimental research on asset pricing [Discussion Paper]. In *Discussion paper center* (pp. 1 – 26). Tilburg University.
- Nuzzo, S., & Morone, A. (2017). Asset markets in the lab: A literature overview. *Journal of Behavioural and Experimental Finance*, *In press*.
- Oechssler, J., Schmidt, C., & Schnedler, W. (2011). On the ingredients for bubble formation: Informed traders and communication. *Journal of Economic Dynamics and Control*, *35*, 1831 – 18951.
- Orléan, A. (1995). Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion. *Journal of Economic Behavior and Organization*, *28*, 257 – 274.
- Othman, A., & Sandholm, T. (2013). The Gates Hillman prediction market. *Review of*

- Economic Design*, 17, 95 – 128.
- Palan, S. (2013). A review of bubbles and crashes in experimental asset markets. *Journal of Economic Surveys*, 27, 570 – 588.
- Palfrey, T. R., & Wang, S. W. (2012). Speculative overpricing in asset markets with information flows. *Econometrica*, 80(5), 1937 – 1976.
- Pennock, D. M. (1999). *Aggregating probabilistic beliefs: Market mechanisms and graphical representation* (Unpublished doctoral dissertation). The University of Michigan, Computer Science and Engineering.
- Plott, C. R., & Sunder, S. (1988). Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*(5), 1085 – 1118.
- Polgreen, P. M., Nelson, F. D., & Neumann, G. R. (2007). Use of prediction markets to forecast infectious disease activity. *Clinical Infectious Diseases*, 44, 272 – 279.
- Potters, J., & Wit, J. (1996).
In *Center discussion paper* (Vol. 1996-04). Tilburg.
- Powell, O. (2016). Numeraire independence and the measurement of mispricing in experimental asset markets. *Journal of Behavioral and Experimental Finance*, 9, 56 – 62.
- Powell, O., & Shestakova, N. (2016). Experimental asset markets: A survey of recent developments. *Journal of Behavioral and Experimental Finance*, 12, 14 – 22.
- Ray, R. (2006). Prediction markets and the financial “wisdom of crowds”. *Journal of Behavioral Finance*, 7, 2 – 4.
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1, 7 – 59.
- Schram, A. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12, 225 – 237.
- Smith, A., Lohrenz, T., King, J., Montague, P. R., & Camerer, C. F. (2014). Irrational exuberance and neural crash warning signals during endogenous experimental asset market bubbles. *Proceedings in National Academy of Sciences*, 111(29), 10503 – 10508.

- Smith, V. L., Suchanek, G., & Williams, A. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica*, *65*, 1119-1151.
- Smith, V. L., van Boening, M. V., & Wellford, C. P. (2000). Dividend timing and behavior in laboratory asset markets. *Economic Theory*, *16*, 567 – 583.
- Stöckl, T., Huber, J., & Kirchler, M. (2010). Bubble measures in experimental asset markets. *Experimental Economics*, *13*, 284 – 298.
- Surowiecki, J. (2005). *The wisdom of crowds*. Anchor Books.
- Tung, C.-Y., Chou, T.-C., & Lin, J.-W. (2015). Using prediction markets of market scoring rule to forecast infectious diseases: a case study in Taiwan. *BMC Public Health*, *15*, 766 – 777.
- Tziralis, G., & Tatsiopoulos, I. (2007). Prediction markets: An extended literature review. *The Journal of Prediction Markets*, *1*, 75 – 91.
- Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, *18*(2), 107 – 126.
- Wu, K., Wheatley, S., & Sornette, D. (2018). Brexit vs trump: quantifying pound and peso market efficiency with a natural experiment. *International Journal of Forecasting (submitted)*, (<http://ssrn.com/abstract=2940173>).

Tables and Figures

Table 1

Experiment 2: standard deviation, skewness, kurtosis and entropy of the market price distribution, pre-trading belief distribution and post-trading distribution in each of the four weeks of the experiment.

	SD	Skewness	Kurtosis	Entropy
Week 1				
Pre-trading belief	16.22	0.86	2.62	3.76
Post-trading belief	14.96	1.18	3.51	3.62
Market	11.98	.1.52	4.37	3.16
Week 2				
Pre-trading belief	9.67	1.30	4.86	3.46
Post-trading belief	8.99	6.30	3.51	3.25
Market	6.06	2.75	12.95	2.78
Week 3				
Pre-trading belief	9.32	0.50	2.63	3.55
Post-trading belief	9.45	0.71	3.60	3.51
Market	6.20	1.09	4.81	3.11
Week 4				
Pre-trading belief	5.99	-0.72	3.29	3.11
Post-trading belief	5.75	-0.81	3.56	3.07
Market	3.83	-0.28	3.37	2.73

Table 2

Experiment 2: Multiple linear correlation coefficients showing the dependence of the post-trading belief as a function of the pre-trading belief and the market prices, according to the formula $Belief_{post} = \beta_0 + \beta_1 Belief_{pre} + \beta_2 Market$ (Model 1), and coefficients of the linear regression according to the formula $Belief_{post} = \beta_0 + \beta_1 (Belief_{pre} - Market)$ (Model 2).

Model 1			
Week	β_1	β_2	R^2
1	0.152*	0.737***	0.9628
2	0.503***	0.485***	0.9721
3	0.580***	0.381***	0.9835
4	0.716***	0.268***	0.9755
Model 2			
Week	β_1	R^2	
1	0.345***	0.3650	
2	0.522***	0.6180	
3	0.646***	0.8010	
4	0.756***	0.8330	

* $p < 0.05$

*** $p < 0.001$

Table 3

Jansen-Shannon Divergence of end-of-day market price distributions in Experiments 1 and 2.

Week	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Experiment 1						
1	0.29	0.16	0.13	0.21	0.13	0.18
2	0.28	0.02	0.02	0.01	0.01	0.05
3	0.48	0.05	0.02	0.03	0.00	0.03
4	0.27	0.04	0.04	0.02	0.00	0.01
Experiment 2						
1	0.46	0.07	0.05	0.03	0.04	0.07
2	0.42	0.03	0.02	0.12	0.14	0.07
3	0.37	0.12	0.09	0.03	0.09	0.08
4	0.22	0.03	0.01	0.02	0.03	0.03

Table 4

A market mispricing measure - Relative Deviation for the three market indices: the sum of prices (index 1), the sum of highest bid prices (index 2) and the sum of lowest ask prices (index 3) in each trading period (week) in Experiment 2.

Index	Week 1	Week 2	Week 3	Week 4
1	0.3	0.42	0.27	0.21
2	-0.5	0.28	0.06	0.16
3	1.78	0.87	0.82	1.07

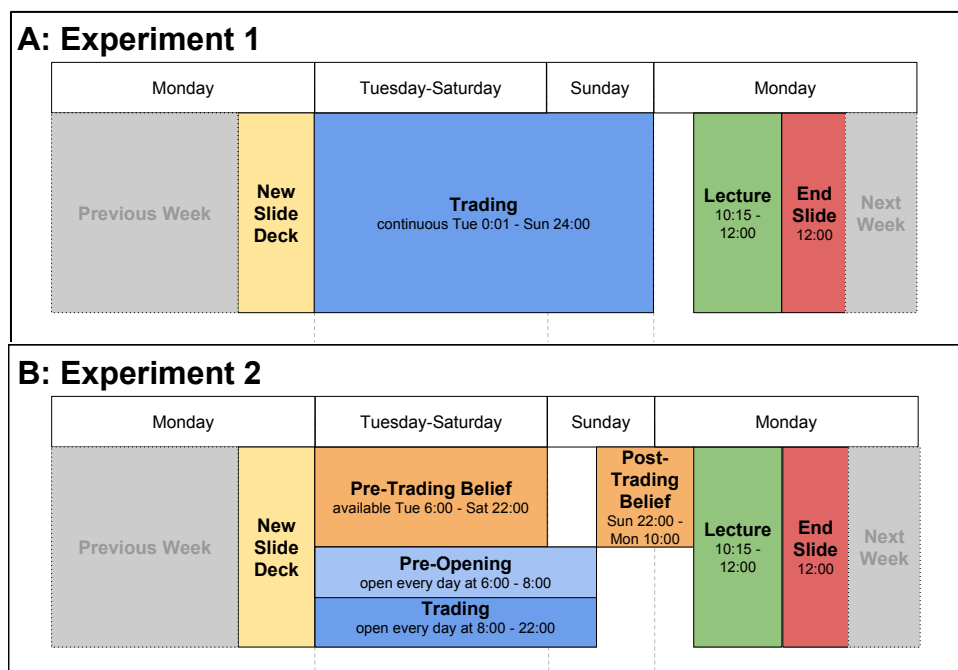


Figure 1. A: Timeline of one week of the procedure in Experiment 1. The market was continuously open from Tuesday morning until Sunday night; B: Timeline of one week of the procedure in Experiment 2. In Experiment 2, entering of the market was only possible after submitting pre-trading belief. Participants who did not enter their pre-trading belief could not see the prices on the market. Similarly, accruing the results of one’s trading and banking on the payment of the dividend were only possible after submitting one’s post-trading belief.

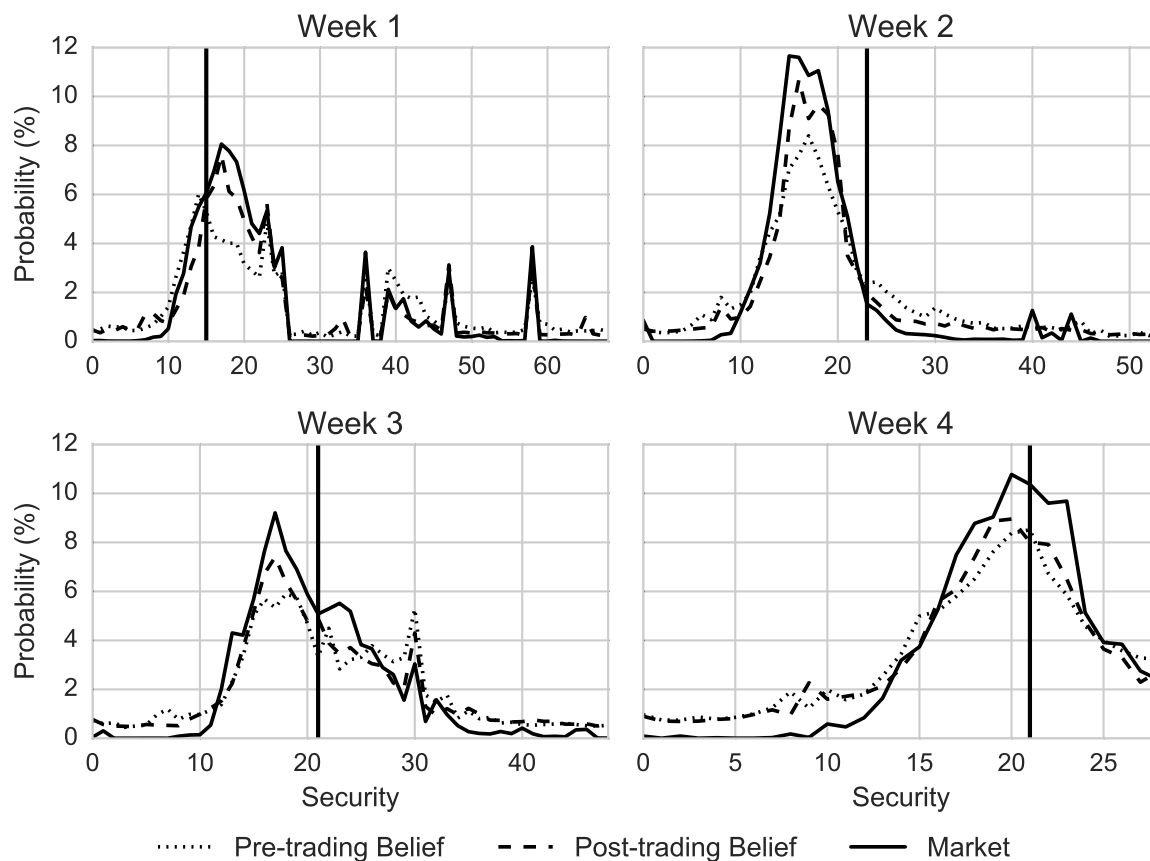


Figure 2. Experiment 2: Average distribution of prices based on the median of all transactions in each epoch and pre-trading and post-trading beliefs averaged across all participants in the market. The price is normalised to ensure that the sum of all security prices is 100. The security whose state was realised at the end of the week (payout of 100 units) is marked with the vertical line.

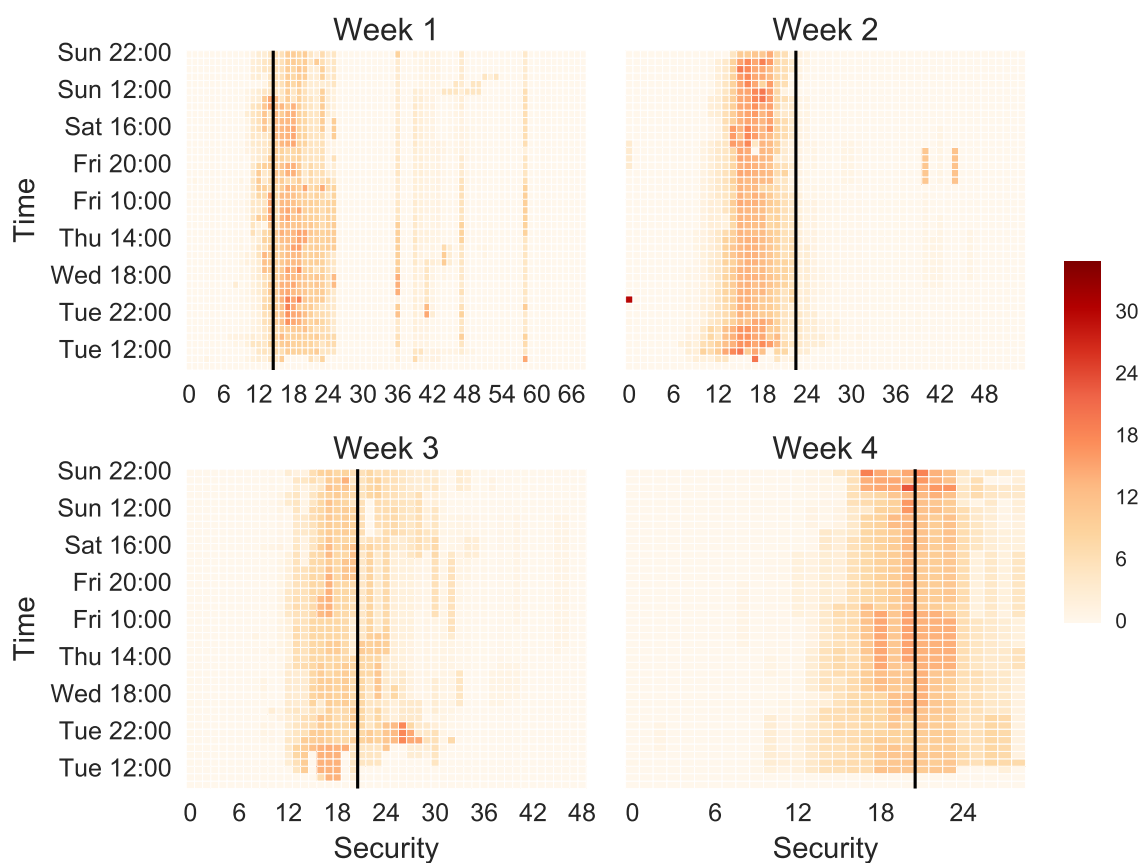


Figure 3. Experiment 2: The evolution of security prices over time. Within a trading period (one week) the prices are generally stable; price levels are established rather early in the week and remain relatively constant. In week one, the distribution of prices is multi-peaked, with four securities emerging as favorites. The tendency disappears in week 2, which has a single-peaked distribution and further the distribution becomes wider and less-peaked, indicating larger uncertainty about the success of each security.

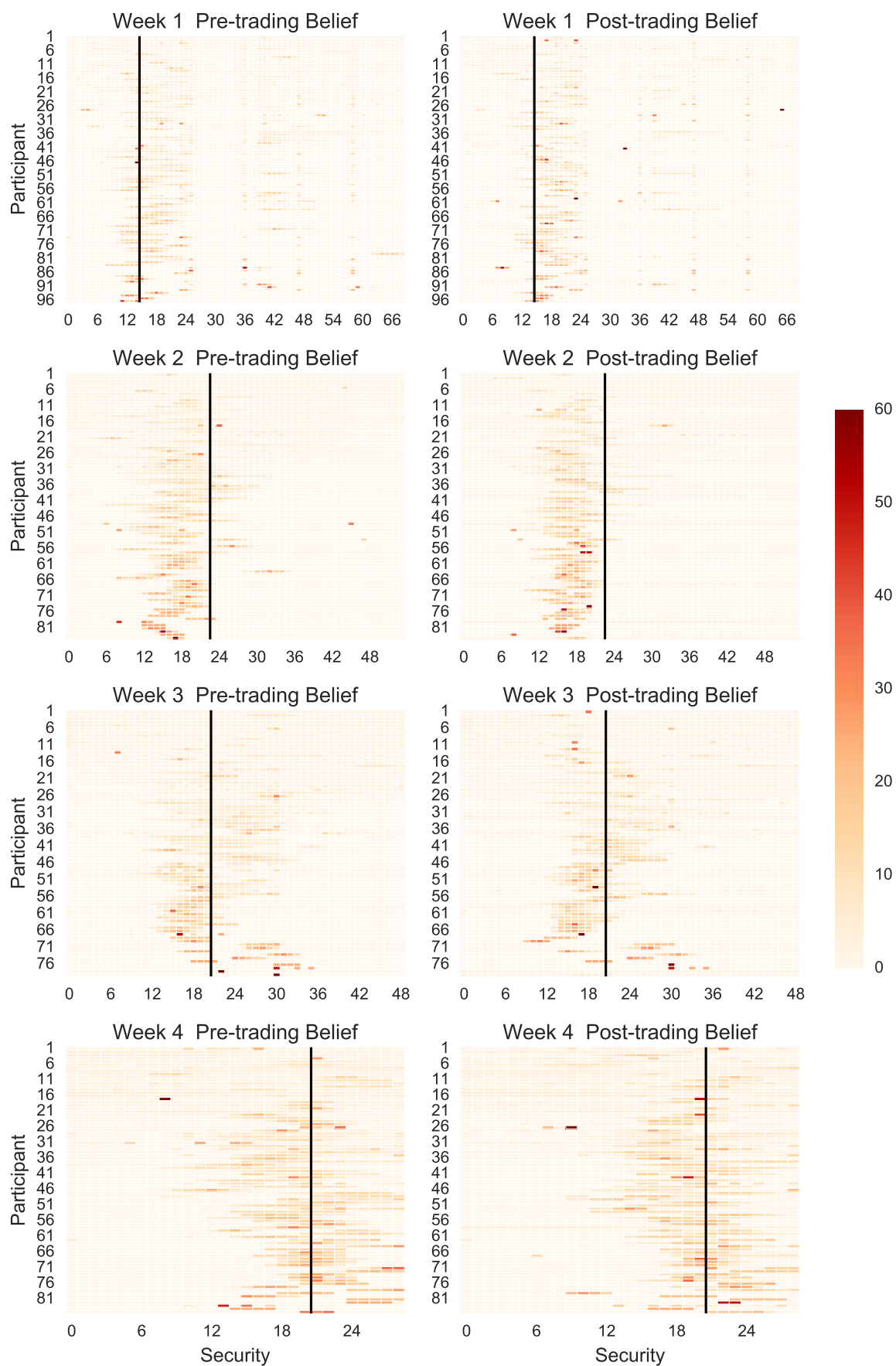


Figure 4. Left: Individual pre-trading beliefs; Right: individual post-trading beliefs.

The black line indicates the security that paid dividend.

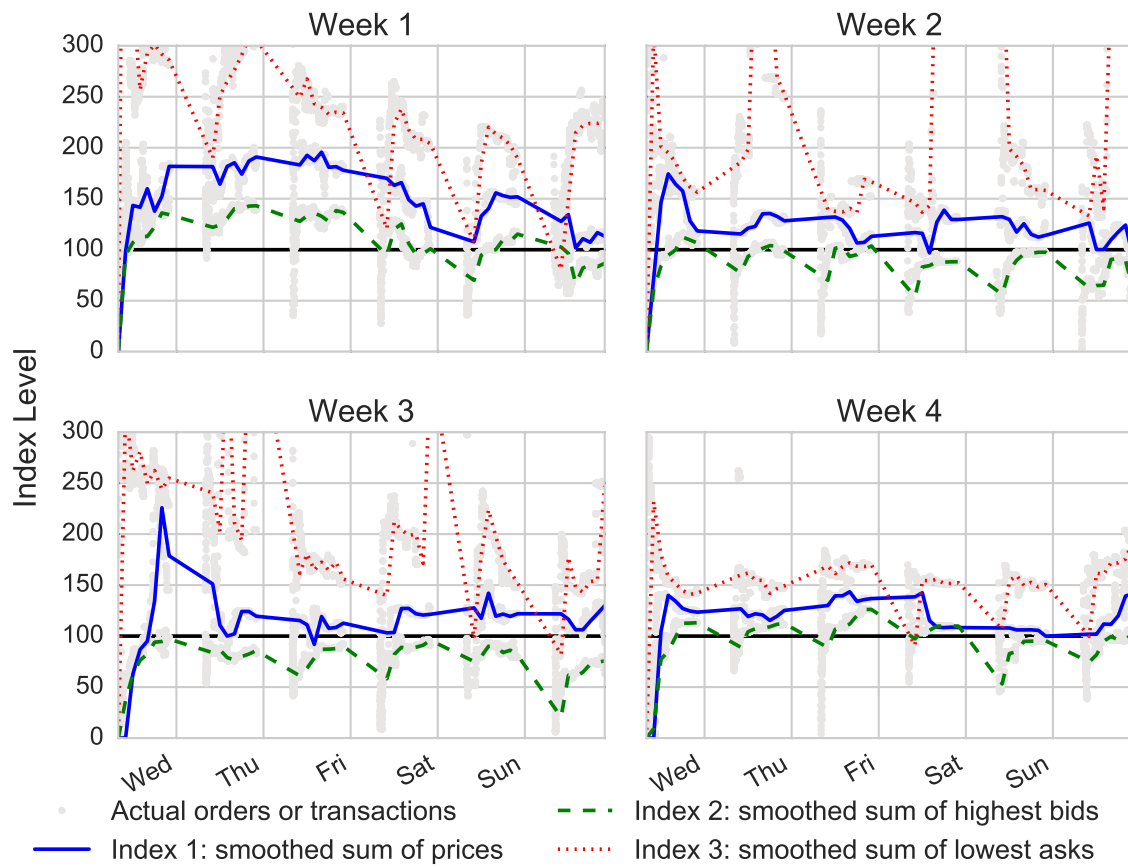


Figure 5. Experiment 2: Evolution of the sum of all security prices for the four weeks of trading. The sum of all security prices should be 100 but there are several pronounced deviations from this normative prediction. The indices are smoothed using a 2-hour moving average.

Appendix A

Experimental Instructions

Dear Students, This document will provide you with the necessary information for the trading market for this class. Should you have any questions that are not answered in here, please contact Philipp Rindler at prindler@ethz.ch. In the interest of fairness, please address all questions by e-mail. Answers will be sent to all students so that everybody has the same basic information.

Goals of the Experiment

There are three primary goals that we pursue with this experiment. First, we are scientifically interested in the results of the market and how you will trade. Second, we want to offer you a pedagogical experience in a real trading environment so that you can apply some of the concepts learned in class. Finally, it is an opportunity for you to earn extra points in addition to the final exam.

Your Compensation

Each week, your final earnings are recorded. At the end of the experiment (end of the semester), your total earnings over all weekly sessions will be used to compile a ranking of all students. The top 25% students with the highest earnings will receive a bonus of 0.5 grade points. The next 25% will receive a bonus of 0.25 grade points. These grades will be added to your grade on the final exam (whereby 6.0 cannot be exceeded).

General Information

Every week, you will be asked to predict the page number of the final lecture slide that Prof. Sornette will talk about in next week's lecture. To do so, you will trade securities that pay out 100 FMRF (FMR francs) if and only if a particular number is realised. Your goal is to accumulate as many FMRF as possible in order to gain bonus points to increase your grade in the class. You can gain FMRFs by trading during the week and by the final pay off at the end of each week.

To be concrete, page number refers to the actual page number in the pdf file. At the end of the class, Prof. Sornette will publicly announce the realised state. Tradable securities are available for possible results only. Securities for pages that have already

been covered are not available in any given week.

Prof. Sornette will not be aware of the trading results during any week. Each week, he will be completely ignorant of the trading results and your portfolio holdings. The flow of the lecture will not be, in any way, affected by your trading.

Similarly, market manipulation, in the form of excessive questioning or interruption of the lecture, will not be tolerated. Prof. Sornette will resist any such attempts during the lecture.

Market Description

You will trade on a market platform on Innovwiki. In order to gain grade points, you have to participate in the market. To do that, you will need to create an Innovwiki account. You need to have an account by November 9th.

The market consists of all students in this class and you each trade individually from your own account. Every week, you will receive a 300 FMRF and 3 units of every security. You may buy or sell securities at any time during the week. You are allowed to input market orders or limit orders. When you issue a market order, the trade (buy or sell) is executed at the current price (if you can afford it). A limit order is entered into the trading book until someone else agrees to the trade. Note that you cannot enter into a trade that you cannot afford: your cash balance and asset balance cannot drop below zero.

During the week, you can make as many trades as you like, subject to the limit that you have to be able to honor your commitments: if you offer to sell a certain number of securities, you have to own them at the time that you submit that offer. If you want to buy a certain number of securities at a certain price, you have to have the cash balance available for that trade.

Each trading week, the market opens after class at midnight between Monday and Tuesday and remains open until the end of the week until midnight between Sunday and Monday. During this time, you are free to use the trading mechanism to obtain your optimal portfolio. You may rebalance your current portfolio should market situations change or your opinion changes.

At the end of each class, the realised number of slides is announced and your account will be credited with your payoff for that week. Since the securities pay off 100 FMRF if their respective page number is realised, your final payoff is equal to 100 FMRF times the number of units of the correct state security that you have in your portfolio at the closing of the market.

The Innovwiki trading platform provides information on the time series of the prices of all securities as well as the order book, which presents all the standing orders by you and other students to buy or to sell that are waiting to be fulfilled. On the basis of this information and your own analysis, you will trade along the week to position your portfolio in your assessed optimal way.

Each week, the market is reset completely. No money or asset positions are carried over. Your earnings for the week are recorded by the system. You can access information about your own earnings but not of others.

Earnings

Each week, you get an endowment of 300 FMRF and 3 units of each security. This loan has to be repaid at the end of the week at 600 FMRF. This is the payout you would receive if you do nothing during the week. Hence, the total earnings at the end of each week is your cash balance at the end of the week, plus the pay offs from your securities minus 600 FMRF. Therefore, you can be ahead of the rest of the class by buying and selling intelligently or by predicting the outcome correctly, or both.

Definition of Securities

In the market, securities are available for each outcome that can happen every week. Each security pays off if class ends on one of three consecutive slides. The last security may cover fewer depending on how many slides there are in total. Securities in the market are named in the following manner: Li.Pa-b. The letter i refers to the lecture, the numbers a-b to the page number that needs to be realised for the security to pay off. For example, L3.P4-6 is the security that pays off if class ends on page 4, 5, or 6 of lecture 3.

Each week, the following securities are available for trading. Starting from the

ending slide of the last lecture, there is a security for each following page in the same lecture. In addition, if a new lecture is uploaded that week (which will always happen on Monday after class), securities that cover all pages of that lecture will also be available for trading. At the end of each lecture, Professor Sornette will announce whether the next lecture will continue exactly from he left off or whether he will start from a slide further ahead, for instance at the first slide of the next lecture notes. The securities available for trading will always reflect this information.

For the number of slides, the page number in the pdf document is relevant, not the page number shown on the slides! Please note that on display, some slides appear to be “animated” but are in fact a set of different pdf pages. Each pdf page counts as a different slide!

Example

As an example, let's consider the following made up situation: class on a Monday ends on page 12 of lecture 3. No additional lectures are uploaded that week. The total number of pages in lecture 3 is 104. So for the next class the following Monday, there are 92 possible pages on which class could end: page 13 to page 104 of lecture 3. Therefore, there would be 31 different securities available for trading: L3.P13-15, L3.P16-18, and so forth until L3.P100-102, L3.P103-104 that each pay off if and only if the final page number is among their respective pages indicated by their names.

At the beginning of the week, you would receive 300 FMRF and 3 units of each of the 31 securities. You are free to trade any of these at any price. Of course, for trades to occur, someone else has to take the other side of the bargain.

As an alternative scenario, assume again that class on Monday ends on page 12 of lecture 3 but now lecture 4 is uploaded on that Monday as well. In this case, you will be able to trade on each page in lecture 3 and 4. Lecture 4 contains 90 pages. Therefore, in this market there would be 61 different securities available for trading:

- 31 Securities that pay off if and only if the final page number is among their respective pages of lecture 3: L3.P13-15, L3.P16-18, and so forth until L3.P100-102, L3.P103-104.

- 30 Securities that pay off if and only if the final page number is among their respective pages of lecture 4: L4.P1-3, L4.P4-6, and so forth until L4.P85-87, L4.88-90.

In this case, you would receive the same 300 FMRF and 3 units of each of the 61 securities.

Your Task

Your goal is to hold a portfolio that you find optimal. To do so, you will have to estimate the probabilities for the different states. You can use the market prices unfolding over the week pushed by buys and sells by yourself and your fellow students to infer the markets assessment of these probabilities in a kind of a wisdom of crowds mechanism. Use the market to buy state securities that you find undervalued and sell state securities that you find overvalued.

To help you with your task, the online platform provides you with a number of tools. For example, for each security you can see the price history and transaction volume over time, both graphically and numerically.

Literature

In order to arrive at your optimal decisions, you will have to compare the market behavior to your own expectations and adjust your portfolio accordingly. You can find further information on designing an optimal portfolio in a setup such as in this experiment (state-preference approach) in the following articles:

- Hens, T. & M. Rieger (2010). Two Period Model: State-Preference Approach In T. Hens, T. & M. Rieger (Eds.), *Financial Economics* (pp. 141-209). Heidelberg, Springer.
- Hirshleifer, J. (1966). Investment Decision Under Uncertainty: Applications of the State-Preference Approach. *Quarterly Journal of Economics*, 80(2), pp. 252-277.
- Kerruish, A. (1983). Can we use State Preference Theory? *Managerial Finance*, 9(3/4), pp. 52-57.

- Sauer, R. (1998). The Economics of Wagering Markets. *Journal of Economic Literature*, 36(4), pp. 2021-2064.

Appendix B

Additional Analyses for Experiment 2

Experiment 2: Trading Activity

The market was very active – 128, 102, 102 and 97 participants submitted pre-trading beliefs and 99 (77%), 86 (84%), 82 (80%) and 87 (90%) post-trading beliefs. This means that some traders had access to the market by submitting their initial beliefs but did not want to have their portfolios included in the final ranking, or some simply forgot to submit their second belief, as reported by some of them. We did not exclude any of these from the analysis because this structure reflects the real-life imperfections of complex social systems. Also, all students that had access to the market formed it. Therefore, excluding activity of the students that did not submit their final belief would not reflect the market the way the participants experienced it.

As in experiment 1, the activity of the market decreased from week 1 to week 4. First, the number of submitted orders decreased across weeks, from 4443, through 2700 and 2006 to 1493 in the last week. Also, the number of active traders (i.e. the traders whose number of submitted orders was 1.5 times the distance between the 75% and 25% quantile) was 12, 15, 12 and 9 in the consecutive weeks, indicating a stability in the activity, except for the last week.

As in Experiment 1, the market was the most active on Tuesday, when it opened and on Sunday, just before it closed. Different from Experiment 1, the highest daily activity was in the morning.

Experiment 2: Self-reported Measures

The post-trading questionnaire revealed that 83% of the participants used the number of slides covered in the previous lectures to predict the next end slide (Please, note that more than strategy was allowed in the questionnaire.). However, only 50% of the participants studied the professor's slides before trading and the majority (60%) would spend less than 30 minutes on studying the slides. The second most popular cue (45%) was participant's own belief submitted before the trading. Other strategies were

used by 24-37% of the participants. These included the average time spent by the professor on each slide, the time spent on presenting slides, the number of topics usually covered by the professor, bid-ask prices offered by the traders and security prices in the market. Two participants indicated using a model based on the previous lectures. It is important to note that most of the participants claimed to use more than one strategy.

66% of the participants realised that the sum of the prices should be equal to 100 at any time. However, only 44% of the participants realised that this mispricing can be used to arbitrage the market. Out of these, 45% applied the arbitrage strategy explained in the description of the results of experiment 1. The main reason for not applying the arbitrage strategy was insufficient market liquidity, while only less than one third of those who did not apply the arbitrage strategy did not know how to do this. Participants used multiple trading strategies at a time, where the most popular were “buy-and-hold” (62%) and “mean-reverting” (50%).

The main motivation to participate in the voluntary trading experiment was to gain additional credit points (76% of the participants), 53% of the respondents participated to gain trading experience, while 61% found the task interesting. This provides an additional argument that grades can be very motivating to participate in experiments and also that many participants have motivations different from purely monetary to participate in financial experiments.

Experiment 2: Trading Strategies

Depending on the percentage of dividend in their total earnings, we distinguish three types of traders: a) *fundamental traders* whose earnings come more from the dividends than from the cash accumulated from trading, b) *technical traders* whose earnings come from cash more than from the dividends and c) *silent traders* who do not apply any of these strategies and whose ratio of earnings from dividends to earnings from cash equals 1. According to this measure, we identified 63 fundamental traders, 56 technical traders and 17 silent traders.

Technical traders put the largest median number of orders out of the three groups:

$Me_{Technical} = 72$, $Me_{Fundamental} = 21$, $Me_{Silent} = 0$. According to a Kruskal-Wallis test, these group differences were significant between all types of traders: technical vs. fundamental, $p = .021$, fundamental vs. silent, $p < .001$, technical vs. silent, $p < .001$

Also, the technical traders had the highest grade from the exam (Median grades: $Me_{Technical} = 5.7$, $Me_{Fundamental} = 5.2$, $Me_{Silent} = 5.0$). However, this difference was statistically significant only between technical and silent traders, and fundamental and silent ($p < .001$, according to a Kruskal-Wallis test). Further, technical traders had the highest earnings ($Me_{Technical} = 3660.67$, $Me_{Fundamental} = 2366.56$, $Me_{Silent} = 600$, where the difference was significant between technical and silent traders, and fundamental and silent traders $p < .001$ according to Kruskal-Wallis test).

From these findings, we derive that the more active traders gained more money and those with the highest predictive skills were the most successful. Better financial knowledge was related to “profiting from the market”, but the causality in this relation is not clear.

Appendix C

Experiment 1

Method

Participants. In a class of 234 students with different majors at the MSc level at ETH Zurich (the Swiss Federal Institute of Technology in Zurich, Switzerland), enrolled in the course “Financial Markets Risks” in Fall 2014, the students were asked to take part in a trading experiment. Participation in the experiment was voluntary and 102 (44 %) of the total students actively participated. The ratio of men to women was approximately 4:1.

Materials. The specific numbers of securities for each of these four weeks were 46, 26, 48 and 53, associated with respectively 138, 78, 144 and 159 slides. The professor finished on slide 60 (43% coverage), 68 (87%), 47 (33%) and 57 (36%), in four weeks.

Procedure. Trading started after the 9th lecture allowing the participants to familiarize with the professor’s teaching style and lasted 4 weeks. During lecture 6, the trading task was announced and explained in detail. As outlined in panel A of Figure 1, each week, the market was continuously open from Tuesday at 00:01 until Sunday midnight before the class on Monday. Orders could be put at any point during this period. All buy orders had to be covered by sufficient cash in their account and sell orders were only allowed if the participant had the necessary quantity of securities in their portfolio. No short sells and no buying on margin was allowed. The trading rule follows the standard continuous double auction mechanism (A trade was successful only if there was a buyer that wanted to buy one or more units of a security for a price at least as high as a seller was offering.). The experiment started after the 9th lecture allowing the participants to familiarize with the professor’s teaching style and lasted 4 weeks. During lecture 6, the trading task was announced and explained in detail.

Results

Market Formation and Prices. To examine the distribution of prices over each week, we use the median price of all transactions within an epoch (e.g., a 4-hour

block) as the price per security. As presented in Figure C.2 where the securities on the x -axis are sorted consecutively, relative prices reflect the market assessment of the likelihood ratio of two states (the dividend paying out or not). Hence, the distribution of prices provides a direct representation of the market assessments for each week of the likelihood of the lecture stopping at a particular block of slides.

Each week, there are a number of securities with essentially zero prices and almost no price fluctuations. For these securities, the participants seem to agree that the corresponding block of slides is very unlikely to be realised. In the first week, the distribution of prices has two pronounced peaks whereas in weeks two and three the distributions are much smoother and single peaked. In week four, we observe an almost flat, and perhaps bi-modal distribution. In three out of four weeks, the peak of the distribution falls very close to the realised security indicating high predictive power of the market. By comparing the prices between the weeks, we observe that the overall price level declined week for week. The flat distribution of prices in the final week suggests that participants learned to diversify their investments across a range of state realizations and not speculate on the realization of a particular security.

In order to examine the dynamics of the security prices over time, we use heat maps presented in Figure C.3 that progress temporally from bottom to top. Grey cells represent securities that do not have an available price because they have not been traded up to that point. The heat maps show that price levels do not change much throughout a given week. Once a market opinion about the value of a security emerges, other participants seem to anchor on the existing price. It appears as if participants use existing prices to inform their probability assessments, which can be interpreted as a behavior consistent with the status quo bias that prevents deviations from initial prices even in the presence of persistent uncertainty (Fleming et al., 2010; Kahneman et al., 1991; Samuelson & Zeckhauser, 1988).

Mispricing and Market Rationality. It is important to note that it cannot be said whether the prices at which securities are traded are rational, as there is no knowable fundamental value. However, in our setup, we are still able to make normative

statements about the total price level of the market. Since one and only one security pays off 100 units of currency, the sum of the security prices – *the market index* – should equal 100 at all times. If the sum was above 100, it would be profitable to sell one unit of every security and vice versa.

Figure C.4 presents the progression of the sum of all prices for each week in real time. The market index is much inflated in every week. In the first week, there are two pronounced bubbles in the prices with the index reaching almost 200 each time! In the second week, prices are inflated at the beginning with a peak of 180 on Wednesday at noon, followed by a steady decline. In the last week, the index reaches its equilibrium value of 100 in the last trading hours. There are only isolated instances lasting only a few minutes during which the index dropped below 100. When the overall price level decreases, the volume of trading of highly-valued securities increases. This price behavior may stem from participants hedging their bets more extensively.

The two other indices included in Figure C.4 provide information on whether the market was a so-called “buyer’s market” or “seller’s market”. Most of the time, one can observe that the sum of highest bid prices is much closer to the smoothed sum of prices than is the sum of lowest ask prices. The latter often tends to be much larger, suggesting that this market was mostly a seller’s market, i.e. demand for securities exceeded supply. The sums of highest bids and lowest asks can also help identify periods of blatant arbitrage opportunities – if index 2 (the sum of highest bids) is larger than 100, the arbitrage opportunity can be exploited by selling one share of each security (with the sum of the sale values being above 100) and thus obtain a certain profit at maturity. Such strategy would tend to push down the overall price level. If index 3 (the sum of lowest asks) is lower than 100, an arbitrage opportunity would also occur, which could be implemented by buying one share of each security, with the sum of the paid prices being smaller than 100. There is a clear but short-lived arbitrage opportunity in week 1, while arbitrage was possible most of the time in weeks 2-4. Also, every week starts with very high ask prices, which stabilise across the week, apart from week 1, which is characterised by high price fluctuation. In weeks 2-4, the sum of

highest bids is almost always above 100 revealing the presence of overpricing.

Grades and Performance. 47 out of the 102 participants in the trading experiment received a final grade bonus. Participants who received the bonus had higher scores on the exam (before bonus) ($M = 4.85$, $Md = 5$ on the scale 1 – 6) than the participants who did not receive the bonus ($M = 4.22$, $Md = 4.25$). This difference and the difference in final grades between the students were statistically significant according to the Kruskal-Wallis test ($p < .001$) indicating that students who were better traders and were potentially better at guessing about the winning slide, also did better in the course (and likely they were more interested in the class itself). There was a strong correlation between number of submitted orders and earnings: $r = .48$, $p < .001$.

Trading activity

The market was very active with the largest numbers of trades occurring just after the market opened and just before it closed, similarly to real financial markets. According to Figure C.1, there is substantial trading activity during each day, although the number of transactions decreased towards the end of the week. During each week, the largest number of transactions occurred in the afternoon after the new set of slides was uploaded by the professor and in late evening hours just before the closing of the market. When we look at intra-day pattern, there were always more trades in the evening (i.e., after 8 p.m.), and almost no transactions occurred between 2.00am and 6.00am. After we exclude these times with very few trades, we observe an average of 19.4 transactions per hour with a median value of 6.

For the further analysis, we bin the trading data in four-hour intervals. Excluding the time from 2am to 6am with virtually no transactions, there are no four-hour intervals during which no transaction occurred. Every trading period (week) is therefore subdivided into 31 epochs of four hours each, with only the first epoch of each week lasting for 2 hours, from midnight to 2.00am on Tuesday night. On average, there were 75.1 transactions in any four-hour interval with a median of 56.

The activity of participants decreased from week 1 to week 4 – 34%, 32%, 41%

and 53% of the participants did not make any trade in weeks 1 – 4. The number of participants that were very active traders (i.e. the traders whose number of submitted orders was 1.5 times the distance between the 75% and 25% quantile) decreased from the first week and stabilised over the three following weeks (15, 18, 15, 10 for weeks 1 – 4). The number of traders in each week decreased from week 1 to week 4 and was 72, 72, 68 and 55.

The number of trades decreased from 2696, through 2378, 2088 to 1618 across weeks. This decrease was not only due to the decrease in the number of market participants, but the activity of the participants also decreased – The average number of trades per participant in each week was 37, 33, 31, 29.

Order Book Summary

To provide more insights into how the opinion about the prices was formed, we analysed the order book. For each week, we split the securities into good (indicating participants' high expectation to pay the dividend) and bad (indicating participants' low expectation to pay the dividend), according to the median prices at the closing of the market. Therefore, the good securities were the securities around the peak of the price distributions in Figure C.2 in the main text, whereas the bad securities were in the tails of the distributions. There are three main results that we can highlight from these data.

First, the number of orders was very strongly correlated with the final prices of the securities ($r = .64$, $r = .78$, $r = .73$ and $r = .85$, $p < .001$ for weeks 1 – 4 consecutively), such that the distribution of prices had the same shape as the distribution of the number of transactions. This implies that people traded the good securities more ($N_{\text{orders}} = 1987, 1949, 1721, 1454$ for weeks 1 – 4) than the bad ones ($N_{\text{orders}} = 1118, 853, 795, 579$ for weeks 1 – 4) and this correlation became more pronounced from week 1 to week 4, indicating a learning effect.

Second, on average, we did not observe a significant difference between the spreads between asks and bids for the good securities compared to the bad ones (Mean difference between spreads of good and bad securities is $\Delta_M = .32, .18, -2.72$ and $.81$ for

weeks 1 – 4), apart from week 3, in which the spreads were relatively high for the first 20 securities. Overall, we did not observe differences in variance of spreads between good and bad securities. Median prices for both asks and bids were higher for the good securities (Ask: $Md = 6, 5.01, 1, 1.5$; Bid: $Md = 3.07, 4.7, 0.5, 0.41$) than for the bad securities (Ask: $Md = 0.03, 0.15, 0.01, 0.1$; Bid: $Md = 0.04, 0.02, 0.01, 0.01$ for weeks 1 – 4) (We calculate the median prices from the whole trading period.). Median bid-ask spreads during the first hours after the market opened were much smaller for the bad securities than for the good securities. During the first few hours, the spreads were negative for the bad securities and positive for the good securities. Spreads between asks and bids were approaching 0 at the end of each week. These results reflect the agreement among the participants about which securities are the most valuable. In the very first orders, the good securities had higher prices and these securities were traded more frequently until the end of each week.

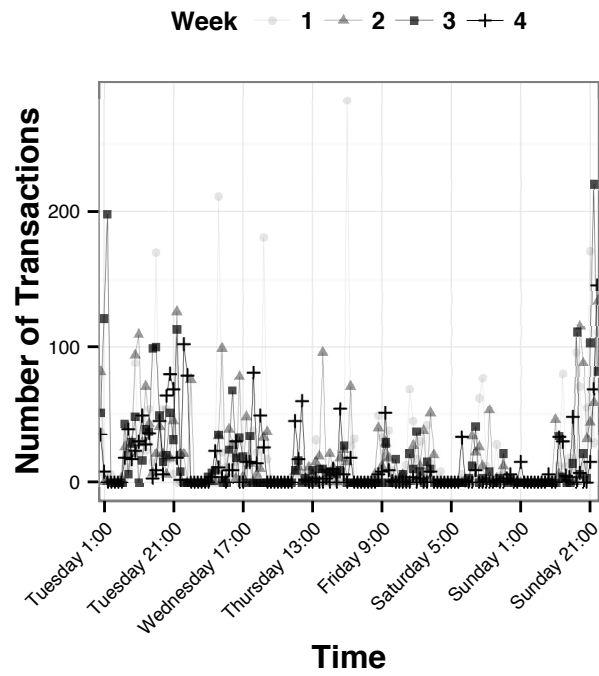


Figure C.1. Experiment 1: Trading Activity during the four repetitions over the four weeks.

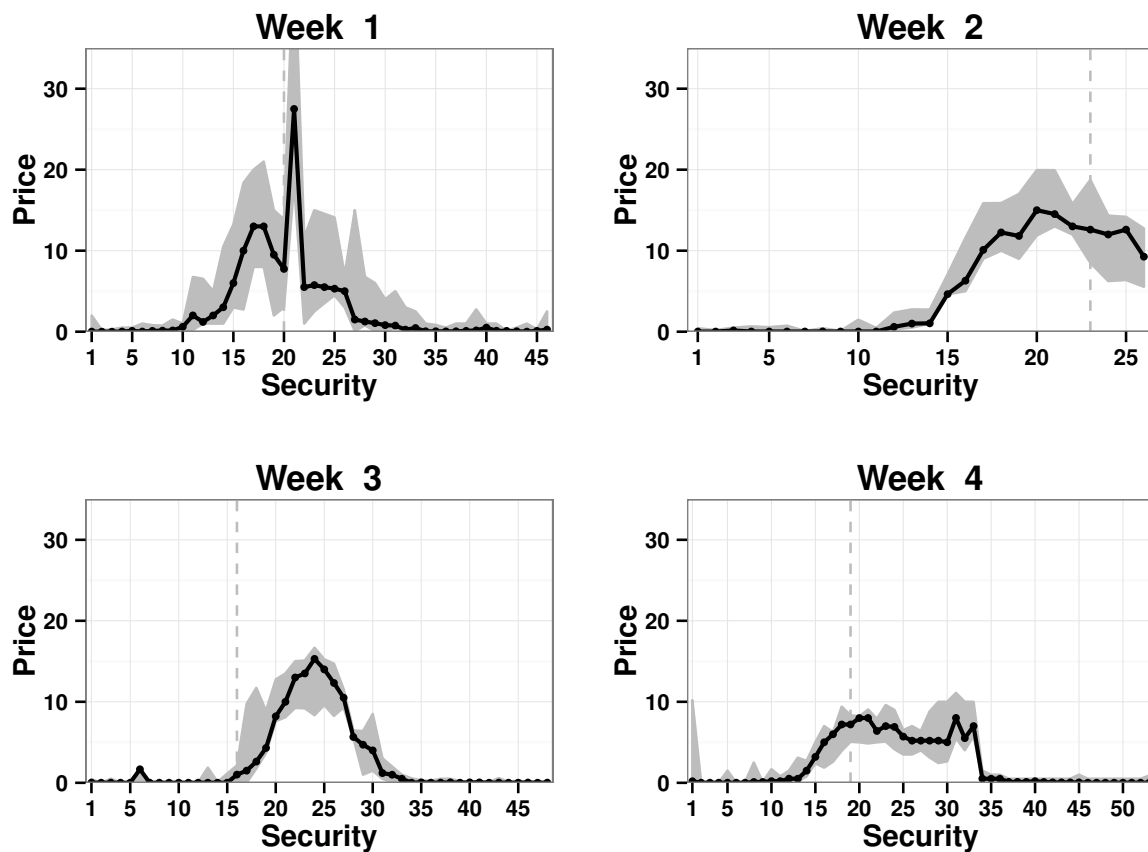


Figure C.2. Experiment 1: For a given week and a given epoch of four hours, we take the median price of all transactions within that epoch. Then we construct the average price over the 31 median prices of the 31 epochs of each week with 95% confidence intervals. The four panels show the resulting distribution of prices over all securities in each week. The security whose state was realised at the end of the week (payout of 100 units) is marked with the dashed line.

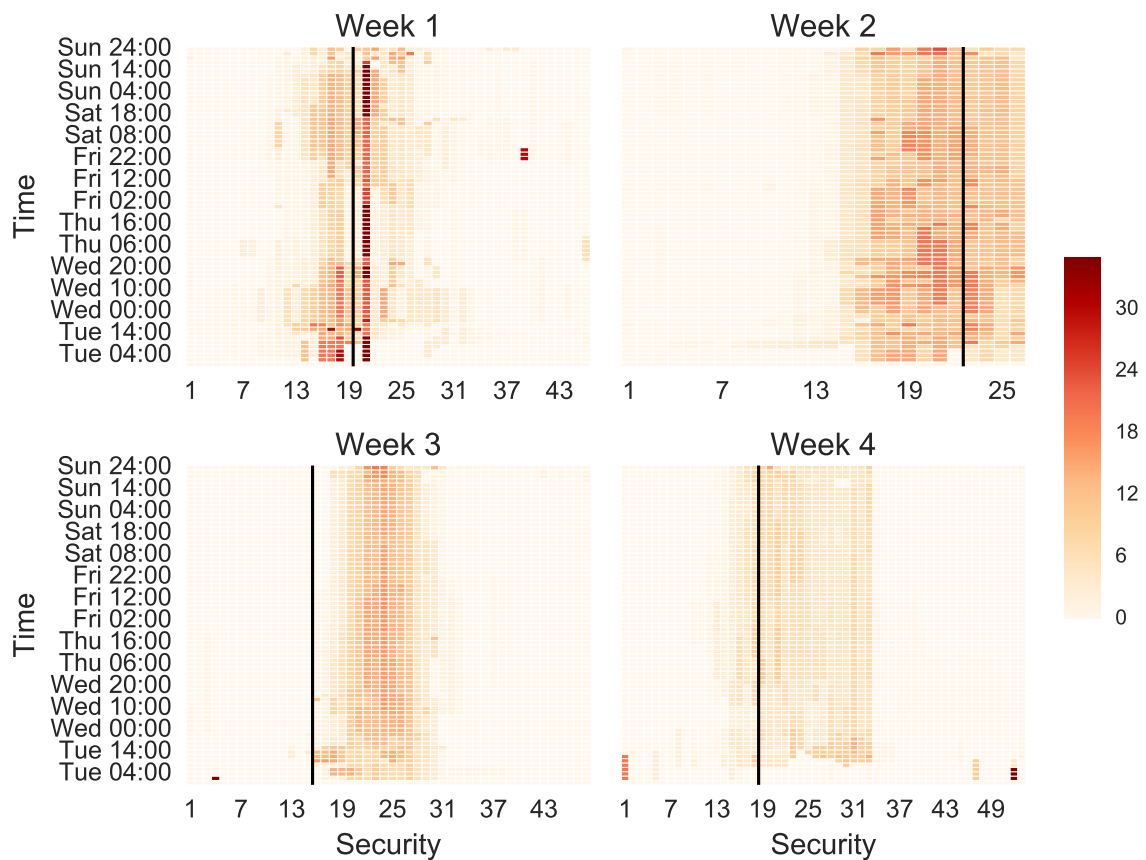


Figure C.3. Experiment 1: The evolution of security prices over time. Within a trading period (one week) the prices are generally stable; price levels are established rather early in the week and remain relatively constant. In week one, the distribution of prices is very peaked, with one security emerging as a clear favorite. The tendency however diminishes over periods, and by the last week of the experiment, aggregate price levels are much flatter.

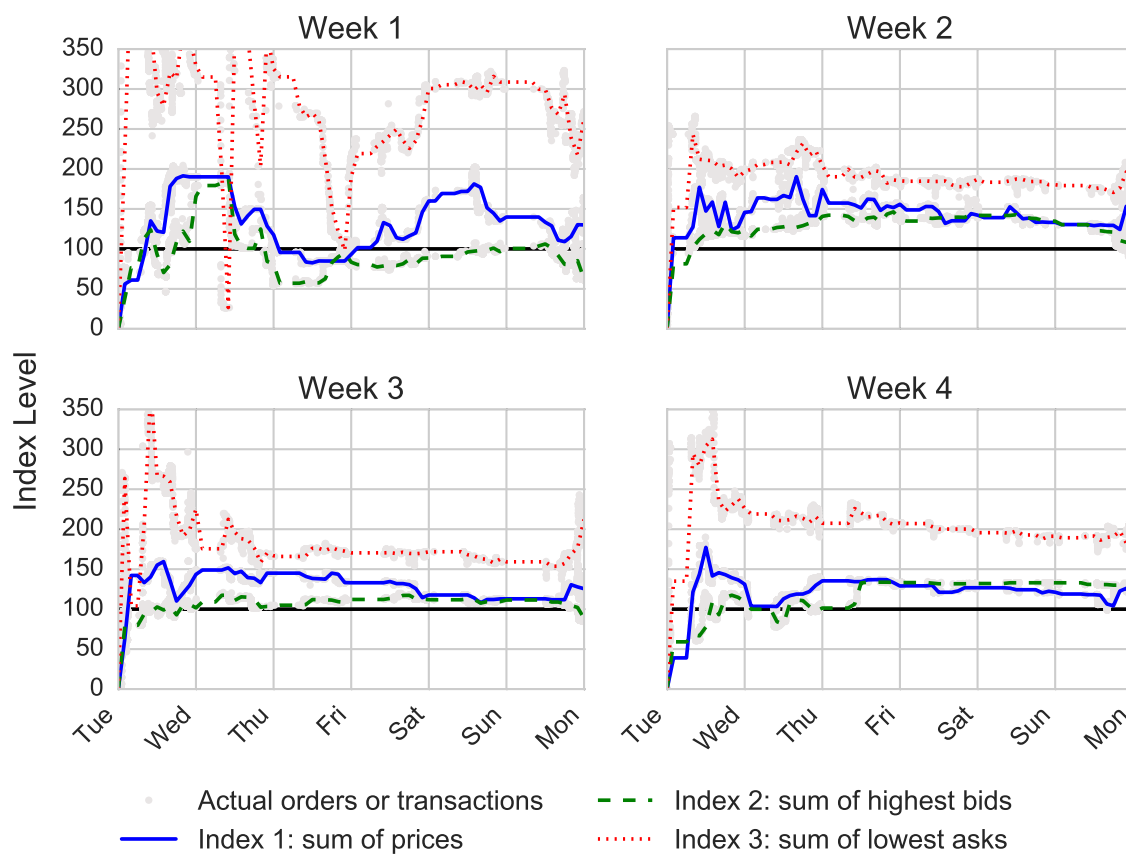


Figure C.4. Experiment 1: Evolution of the sum of all security prices for the four weeks of trading. The sum of all security prices should be 100 but there are several pronounced deviations from this normative prediction. The indices are smoothed using a 2-hour moving average.

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

<http://www.economics-ejournal.org/economics/discussionpapers/2019-32>

The Editor