

**Reply to a Referee report dated Sep 2, 2019, regarding submission 2019-32 entitled “Overpricing persistence in experimental asset markets with intrinsic uncertainty” and submission 2019-33 entitled “Behavioural effects and market dynamics in field and laboratory experimental asset markets”**

We appreciate the feedback and comments provided by this referee. Given that the reviews of our two submissions (2019-32 and 2019-33) are provided together, we do not separate them in our reply as well. Below, we provide our point to point response to the comments of this review and explain how we will implement the suggested changes in our revised manuscript. The original comment by the referee is provided in *blue italic*, while our replies in black.

*MS 3107 - Behavioural Effects and Market Dynamics and Laboratory Experimental Asset Markets*

*This paper studies to what extent the behaviour observed in an experimental asset market can be replicated within a more “realistic” setting implemented via a “field” experiment. Generally speaking, this is a question of obvious importance to experimental economics, as the lack of external validity limits any contribution to the context of artificial environments.*

We agree with the referee on the importance of reproducibility in scientific work. This is the driving motivation of the work reported in our paper.

*However, this paper falls victim to the common misconception that external validity can only be increased by deprioritizing laboratory control. As an experimenter, one should only be willing to trade off the clean identification offered by the lab in return for a meaningful approximation of the real world. Contrary to the authors, I do not agree that their “field study” meets this necessary requirement.*

*To make my argument more precise, I rely on the same comparison between laboratory and field experiments that the authors refer to themselves<sup>1</sup>. In the context of financial markets, field experiments – in contrast to the lab – are characterized by:*

- 1. Participants who are financial professionals (ideally traders)*
- 2. Participants who possess sufficient experience in the task*
- 3. Real-world financial assets*
- 4. Real-world incentives (both in nature and magnitude)*
- 5. Real-world trading task*

<sup>1</sup> *The authors refer in particular to Harrison and List (2004).*

*From my understanding, the paper's "field study" falls short of all (!) the above listed conditions. By simply introducing numerous "long-lived" Arrow-Debreu securities to an otherwise quite artificial trading environment, neither condition 3 nor 5 can be satisfied.*

We do not agree with the reviewer's statement because the synthesis made there is incomplete. The referee emphasizes the term "real-world", while we propose an "intermediate-complexity" set-up that incorporate several real-world features but not all. Indeed, the guiding idea of our paper (submission 2019-32, Sornette et al 2019) is to propose a new experimental paradigm that features uncertainty about the payoffs and to move away from the laboratory setting to a more engaging environment involving real-life events. Compared to commonly used experimental asset market designs, our setup combines features that have not been put all together in one experiment. For example, in our design, the traders trade various numbers of assets, and the participants cannot determine with certainty the expected value (and the fundamental value) of the securities. These two design features play a crucial role in mimicking real markets, while they are usually absent from classical experimental asset markets. By putting the emphasis on these two features, our design allows us to study their impact, in a context which is more controlled and clean than in a full real-world situation.

Following the spirit of deprioritizing laboratory control as the reviewer suggested, we intentionally designed the special Arrow-Debreu securities so that the subjects cannot learn or estimate the probabilities of the possible outcomes with certainty. Meeting all of the conditions listed by the reviewer was not the ultimate goal of the study. What we are trying to achieve is to bring concepts from the conventional experimental asset market and prediction markets one step away from the sterile laboratory situation and one step closer to the real-life setting, instead of moving completely to the real markets. This is in line with what is the scientific method, which starts from fully controlled and constrained set-ups and then relaxes progressively one by one the constraints to get closer and closer to the real world.

We would like to point out that, according to Harrison and List (2004), the five criteria that distinguish a field experiment are:

- "the nature of the subject pool,
- the nature of the information that the subjects bring to the task,
- the nature of the commodity,
- the nature of the task or trading rules applied,
- the nature of the stakes, and
- the nature of the environment that the subject operates in."

Therefore, the five criteria brought by the reviewer are the reviewer's interpretation of the original source. The above five criteria summarise clearly and in a general way what are the fundamental elements allowing one to classify different experiments. In our experiments, participants were students educated with financial knowledge, who traded assets that had a real-life consequence, on a market whose trading rules exactly mimic the Swiss Stock Market implemented in a platform that precisely simulates professional trading platforms. While we highly appreciate constructive criticism, we believe that stretching the interpretation of academic sources for providing non-constructive comments (i.e. "trolling"<sup>2</sup>) is not a good academic review practice.

The main criticism from the reviewer is that we do not follow the "standard rule" of a field study. However, the etymology of the word research implies using previously acquired knowledge to find new applications and enlarge the knowledge. This includes inventing new designs, discovering new phenomena, developing new models, which may or may not follow the "standard rules". We believe that we have proposed a meaningful new experiment design, implemented it following the scientific approach, and found new interesting insights, which we would like to report to the community. We do not think that the definition of "field study" should be applied here and could be used as an argument to judge our paper. Therefore, in our revised papers, we will reformulate the definition of our experimental design to avoid any confusion.

*Therefore, the authors' "field" experiment cannot claim any more external validity than any laboratory experiment that involves inexperienced subjects trading a fictitious asset for small stakes. Needless to say that there exist valid reasons for conducting a controlled laboratory experiment to study the behaviour of financial markets.*

As discussed above, our new experiment design brings the traditional forms of experimental asset markets closer to the real environment. It contains many features that increase its external validity in comparison to the laboratory setting. For example, study participants can speak to each other, trading happens for several weeks, while the market remains closed at night, the incentives are highly relevant for the participants, just to name a few features that are not present in standard laboratory experiments. The

<sup>2</sup> [http://deliver.v.acm.org/10.1145/3000000/2998213/p1217-cheng.pdf?ip=195.176.106.251&id=2998213&acc=ACTIVE%20SERVICE&key=FC66C24E42F07228%2EA04051DB0C098788%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&acm=1571939053\\_29c5b4baa21406eed6b944fea035863d](http://deliver.v.acm.org/10.1145/3000000/2998213/p1217-cheng.pdf?ip=195.176.106.251&id=2998213&acc=ACTIVE%20SERVICE&key=FC66C24E42F07228%2EA04051DB0C098788%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&acm=1571939053_29c5b4baa21406eed6b944fea035863d)  
[https://www.researchgate.net/profile/Chris\\_Hopkinson3/publication/269652088\\_Trolling\\_in\\_Online\\_Discussions\\_From\\_Provocation\\_to\\_Community-building/links/5b8a571892851c1e123fc890/Trolling-in-Online-Discussions-From-Provocation-to-Community-building.pdf](https://www.researchgate.net/profile/Chris_Hopkinson3/publication/269652088_Trolling_in_Online_Discussions_From_Provocation_to_Community-building/links/5b8a571892851c1e123fc890/Trolling-in-Online-Discussions-From-Provocation-to-Community-building.pdf)

submission 2019-32 firstly reports the main findings of this design. As we discussed in the paper, the submission 2019-33 reports a laboratory experiment aimed at replicating the first “field” experiment, and at questioning whether moving to a less controlled setting can open opportunities for experimental investigations without distorting the relations between individual variables clearly observed in the laboratory. We did not aim to find new behavioral effects in the laboratory experiment. We successfully replicated the main findings and justified the differences, which opens the room to further relax a few strictly controlled conditions usually imposed in the laboratory. Therefore, the purpose of the laboratory replication was not only a mere robustness check, but also opening the possibilities for future investigations in more flexible environments.

*However, the lab experiment of this paper fails to provide any meaningful insights. The following list provides a collection of arguments to support this rather harsh conclusion:*

- *No clear hypothesis*

We have clearly laid out our main goal and research question in the paper, and also in the above. We acknowledge the standard protocol of hypothesis testing framework. We stress that the submission 2019-33 is methodological, in nature.

- *Vague and unsurprising results at the microstructure level:*

1. *Trading activity influenced by learning (known from existing literature)*
2. *Some mispricing at aggregate level (expected based on findings of existing literature)*
3. *Higher correlation between individual ex-ante and ex-post beliefs than with market prices (surprising?)*
4. *Higher earnings associated with higher “illusion” of control (confusing conclusion as control can exist in the presence of arbitrage opportunities!)*

As discussed above, the main goal of the paper is to test whether the key findings found in the field study would be replicated in the time-constrained more controlled laboratory setting. All these findings were related to the first paper and the exercise of this paper justifies that the less controlled “field” environment could be used instead of a laboratory environment. One of our conclusions is that the strictly controlled laboratory environment is not only unnecessary to obtain the main effect, but might provide some limitations that affect the results.

- *Lack of statistical power to draw general conclusions: only one session, i.e., three markets, with a total of 36 subjects*

It was one practice period and three experimental periods, each round lasting 10 minutes, resulting in total of 2442 orders in the experiment (average of 94 data-points per participant). In contrast, to reach that number of datapoints in a typical SSW market with 10 rounds lasting 60 seconds and 12 participants (these values are based on how a typical SSW experiment looks) in each round, each participant would have to submit an order every 3 seconds. The approximate 2.5 thousand data points constitute a data set that is sufficient and ethically appropriate for a simple study, when compared with similar studies (see Marszalek, Barber, Kohlhart & Holmes, 2011<sup>3</sup>), while the meaningfulness of “the more-the-merrier” approach in data collection approach is disputable as well (see Schäfer & Schwarz, 2019<sup>4</sup>).

The main design change compared to other laboratory-based experimental asset markets is the duration of a single round. If we wanted to implement 10 experimental rounds, each with 36 participants and repeat this exercise 10 times (as it is often done with e.g. SSW markets), we could inflate our sample size and obtain an overpowered study instead. While there are several rules of thumb for determining the sample size in psychological studies and appropriate software exists (i.e. GPower, pwr package for R statistical programming language), we were unable to find appropriate guidelines for experimental asset markets. Therefore, we think that this criticism from the reviewer is not well supported by the current state of scientific knowledge.

- *Compensation introduces potential wealth effects*

Note that for each session the initial cash and security for each participant were reset to 300 Experimental Francs and 3 units of each security. Therefore, the wealth effect mentioned by the referee does not apply here.

- *Belief elicitation is not incentivized (because of hedging concerns?)*

We intentionally decided not to incentivize the belief elicitation, and found that it achieved the same goal of forming the belief.

- *Questionnaire is not included in the provided instructions*

Questionnaire was by design provided at the end of the experiment to understand the behavior of traders, and not influence them a priori. The questionnaire included standardized scales that are listed in the Methods Section and in Appendix A.

*Finally, neither the similarities nor the differences between the results from the laboratory experiment and the “field study” are unexpected. On the one hand, the closely-related design mechanically introduces a somewhat similar*

<sup>3</sup> <https://pdfs.semanticscholar.org/ab75/e988e0544e01eec517a5d56d948ff6c3abc5.pdf>

<sup>4</sup> <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00813/full>

*structure. On the other hand, changing some important parameters unsurprisingly causes certain discrepancies. For instance, a much shorter trading period naturally makes it harder to stabilize trading prices, to decrease (relative) mispricing, and to trade numerous assets simultaneously.*

As discussed above, the main goal is to evaluate the generalisability of behavioural effects obtained both in more realistic and in artificial experimental environments. As importantly pointed out by the reviewer, the much shorter trading period and the strictly controlled laboratory environment introduce other factors that may influence the experimental results, given that these factors make the experiments far from the real world. One of the goals of the study was to demonstrate this phenomenon, which is possible using the experimental paradigm proposed in submission 2019-32. The same exercise could not be done with, for example, the SSW design. Saying that a study demonstrates the “obvious” or “foreseeable” effect is a vague criticism. There are plenty of behavioral studies that scientifically demonstrate evidence for certain psychological phenomena for which one could foresee the effect, based on some non-documented observations (see for example Kosinski’s work on predicting one’s intimate traits and personality based on online behavior). In the revised version of the manuscript, we will improve the presentation of our goals to make them clearer as well as the presentation of the conclusions.

*MS 3106 - Overpricing persistence in experimental asset markets with intrinsic uncertainty*

*This paper studies (mis)pricing patterns in an experimental asset market with inherent uncertainty. In contrast to the seminal work by Smith, Suchanek and Williams (1988), the paper makes use of a prediction market setting in which the underlying probabilities are unknown. More specifically, market prices depend on students’ beliefs about the teaching style (speed) of the lecturer of a university course on “Financial Market Risks”. As this paper largely relies on the same “field” experiment as MS 3107 (see above), the same criticism with respect to its design and result interpretations applies.*

Thank you for the summary of the paper. We have provided our answers regarding the comments above.

*Moreover, I am concerned about the relatively low participation rate of 55% (reported in MS 3107). This could indicate a lack of incentives generated by the chosen bonus credit point scheme.*

Note that the experiment was completely voluntary for the students, and the incentive was the bonus credit points added to the exam grade of the course.

First, for the majority of the students enrolled in the course, this course was not mandatory to obtain their degree and a significant fraction took it for their own education out of pure interest in the subject. Second, the students could choose to not participate in the experiment and achieve the maximum credit points 100% through the final exam. It is in the context that one should judge the 55% participation rate, and we conclude that it is actually quite large. Also, the number of participants was 122, which is sufficient for such a trading experiment, compared to many laboratory experiments.

*In the case of the final experiment (predicting the future performance of real-world stock indices) two additional confounding factors are introduced: (i) the implied signal from predefined return intervals (of different length), and (ii) students' subjective beliefs about future market performance.*

We respectfully argue that they are not confounding factors but helping factors to simplify or augment the value of the experiments. The predefined return intervals of different length have been designed to help the subjects to match their subjective appreciations that the range of price changes in probability space is not uniform. In other words, we specifically designed the return intervals with lengths such that a fully rational and informed subject using unconditional probabilities would assign the same probabilities for each segment. Said differently, we designed the return intervals to correspond to uniform priors. The second factor on subjects' subjective beliefs was added to provide further insight in their decision making process, with the goal to obtain a clearer interpretation of the obtained price dynamics.

*Generally, the paper fails to acknowledge the existing experimental literature on (mis)pricing, bubbles, and crashes that has already introduced alternative environments to the one proposed by Smith et al. (1988): e.g., Moinas and Pouget (2013) on bubbles, Asparouhova, Bossaerts, Eguia and Zame (2015) on mispricing under asymmetric reasoning, and Crockett, Duffy and Izhakian (2018) on the Lucas asset pricing model. While Asparouhova et al. (2015) introduce a setting of intrinsic uncertainty that allows for a well-defined mispricing measure at the security level, Crockett et al. (2018) study a more realistic consumption-based trading economy. Additionally, there exists a vast and mature literature on experimental prediction markets involving Arrow-Debreu securities (see report on MS 3017).*

Thank you for pointing out this useful literature, which we will incorporate in the revised version of the paper.

*Lastly, in its current state, this paper is not well-enough written to be published in an academic journal.*

We will carefully review and revise the written following the academic standard in the next revised version. We will also employ a native English speaker to proofread the paper before the final submission.