Reply to a Referee report, submission 2019-33 entitled "Behavioural Effects and Market Dynamics in Field and Laboratory Experimental Asset Markets"

Dear Referee,

Thank you for your useful feedback and comments. Below, we provide our response to your comments and we 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 are typed in black.

This paper studies a lab version of the classroom experiment described in their earlier paper (Overpricing persistence...). Their goal is to study whether results obtained in laboratory experimental asset markets persist in the field.

The second sentence of this summary inverts the logic of our paper: indeed, the goal of our study was to replicate a field study in a laboratory (not the other way around) for the purpose of testing not only replicability of the main findings of the field study, but also to test the differences and similarities of the market dynamics, depending on the experimental environment.

Assessment

The paper does not offer a substantial contribution. The authors replicate their field experiment in a lab setup. In doing so, they introduce a number of differences in the two experimental setups: duration, immersive experience, monetary vs grade payoffs, number of rounds and a few others. This is problematic because it is not clear which difference in the experimental setup contributes to the differences in results.

We respect your opinion. It is true that we had to introduce a few changes to adapt the field experiment to the laboratory conditions. However, we do not agree that the paper does not bring a substantial contribution. First, it demonstrates the differences in market dynamics between field/classroom-like and laboratory experiments. Second, it highlights the robustness of bubble formation in markets of different designs and different experimental environments. Third, it shows that the behavioral effects replicate independently of the grade vs. monetary compensation, where in economics (but not in psychology), not using monetary compensation could be considered as not incentive-compatible.

Nevertheless, we are very grateful for this comment because it shows that we did not sufficiently highlight the contribution of the paper. In the revised manuscript, we will highlight its contribution.

My main concern, however, is that the paper is of limited interest. It does not really tell us whether standard lab experiments are externally valid since their lab experiment is quite different from the standard lab asset market experiment. Nor does it tell us much about the field, since their field experiment is very different from real financial markets.

We understand this concern. However, we believe that the replication presented in our paper

provides a “next-step” to the replication procedure vividly exercised in many fields dealing with human behavior. In particular, in the context of the “replication crisis” (Pashler & Wagenmakers, 2015), as noted in the work by the Open Science Collaboration (2015), the fact that a study is conducted in exactly the same manner as the original study is a confirmation of the reliability of the results when the results are reproduced. However, failure to replicate may stem from factors others than failing to exactly reproduce the experimental conditions, for instance when experimental stimuli or environment may not be relevant any more (i.e. a movie that was funny 10 years ago, and was used for a mood induction in an experiment, may not be funny now to most participants). As Open Science Collaboration (2015) proposes, alternative ways of replicating the effect are important and necessary to estimate the robustness of an effect.

In my opinion, this paper is more a robustness check of their earlier paper.

We respect this opinion. However, in face of the research contributing to solving the replication crisis, we do not agree with this opinion. A number of studies have been conducted in as much similar way as it could be done. Replicability in experimental economics is higher (about 60%) than in psychology (below 40%). This would result from the fact that certain social and psychological aspects of an experiment may be more prone to change over time or to be impacted by the general variability in the environment.

The scientists leading the #NARPS project (https://www.narps.info) tried to investigate the replicability of the fMRI studies, where they first replicated a neuroimaging study and later asked a number of science labs to analyze the same data set according to the procedure that the specialists in a given lab find the best. It turns out that 60% of the teams replicated the main effect claimed in the original paper, some teams found no effect, while others found the opposite effect. This again shows the importance of replicating studies, but also highlights that replicability is prone to the analysis and experimental setting.

Therefore, more studies, using a similar, but not exactly the same design are needed to prove the validity of effects. This idea has only arisen very recently (i.e. during the annual world meeting of the Society for Experimental Finance in Copenhagen). Therefore, we understand that this approach may feel unusual for the time being. However, we see it as a next stepping stone in the replicability community.

As a small comment, the citation of their earlier work does not have the same title as the one that can be found on e-conomics.

Thank you for this remark! We will correct this error in the updated version of the manuscript.