EXPERIMENT DESIGN IN ECONOMIC SCIENCE

Petr Obergruber

Abstract

Although the experimental method is indisputably one of the core scientific methods, it has been not a long time it became part of economical sciences. Dominant influence of classical and neoclassical economics, with strong mathematical background and rely on sometimes problematic assumptions, pushed the experimental methods beyond the perception of the vast majority of economists. On the one hand, this scepticism persists. But the rejection of experiments is not that wide as before. More economists are searching for situations and questions in which an experiment might offer a feasible and desirable approach. In this paper, I will first offer an overview of what experiments are and what goal and design follows. Then I will offer some thoughts about the potential gains from doing economic research using experiments and problems and situations I faced during designing specific experiments (generally and specifically), how those problems should be treated and how treated the problems specifically us.

Key words: Ceteris paribus, experimental design, homo economicus, behavioural economics, reliability of results

JEL Code: C44, C91, C93

Introduction

The so-called mainstream economy is mainly based on strong mathematical apparatus. This approach has many pros, such as internal consistency, understandable language (mathematics), and includes verified statistical methods, which can provide suitable evidence regarding hypotheses. Unfortunately, this manner of economic thinking has two major issues that withstands the rising criticism. Those issues are the homo economicus assumption and the ceteris paribus assumption.

Without debate, the homo economicus assumption is more famous than the other, but at the end, both condition are very restrictive. The homo economics assumption basically says, that people can rationally decide in every possible situation and reach the maximum amount of utility. This should work independently on problem context and any emotions. At the end, the lighter version of the assumption says that there could be some deviations from the optimal decision, but in average, people behave rationally.

The ceteris paribus assumption is not that much debate, but it is built in economics science as substantially as homo economics assumption. The ceteris paribus condition makes economics the static system, and looks what happens when one or few variables are changed. This looks reasonable at first look, but raises many problems – the biggest one: economics is not static system, but on the contrary extremely dynamic environment. There are rarely situations where only few variables are changed. More often, there are situations where many inputs are moving, interact, cumulate or cancel themselves or feedback with each other. Those highly complex changes are almost if not entirely unpredictable with any statistical or mathematical tools. The only way how to at least approximately reach the ceteris paribus condition is situation or event which is as simple as it could be.

The empirical gold standard in the social sciences is to estimate a causal effect of a certain action, but amidst the complexity of the real world. This process is easier said than done. Economists have long worked on approaches that seek to separate cause and effect in naturally occurring data. A few decades ago, a standard approach was to use multiple regression analysis to hold other factors constant (basically insert the ceteris paribus condition artificially). However, economists have now internalized the old maxim that "correlation doesn't imply causation", and have, in recent decades, sought a variety of other approaches (List 2011). As we observe above, we must trade the real-world objectivity for laboratory cleanliness, at least to an extent.

Despite all that, ceteris paribus condition was not the reason why the experimental approach did not occur in economics a long time ago. There is concept of homo economicus, which is a stronger and older assumption than ceteris paribus. Additionally, from this concept directly originates the reason why economic researchers do not use experiments: we do not need them because we can calculate the results.

The concept of homo economicus did not accompany the economy as a science at the beginning. Most economists considered and included the specifics of human behaviour in their work. Adam Smith, author of The Wealth of the Nations, wrote one more important but less famous book, The Theory of Moral Sentiments. The focus of this book is primary psychological principles of individual human behaviour and decision making. One of Smith`s most interesting perceptions is the fact that people suffer much more with less than they are happy with an equal gain. This concept was greatly ahead of his time. Today, we know this effect as loss aversion, which is part of Kahneman’s and Tversky’s Prospect Theory (Kahneman a Tversky 1979;

Tversky a Kahneman 1992), which is one of the basic theories of all of behavioural economics. In addition, for example, Francis Edgeworth's Theory of Mathematical Psychics, which introduced his famous "box" diagram that shows two-person bargaining outcomes, included a simple model of social utility, in which one person's utility was affected by another person's payoff, which is a springboard for modern theories.

1 Designing the lab experiment

As was noted, lab experiments usually study very simple decisions from limited possibilities. With this, experimental designers usually utilize surprisingly simple mechanics to study exactly what they want to study. Then, these designers gather data on the participants of the experiment.

The problem in this phase of experimental development could be the exaggerated effort to solve too many things at once. Although this problem looks harmless and at the end could be beneficial (just because you deal with wider situation, which means you can generalize your results), your first focus must be to keep the $D.V. = f(I.V.) + \varepsilon$ as clear as possible.

Otherwise, the whole experiment could generate noisy data, or in worst case scenario generate data which do not correspondent with starting $D.V. = f(I.V.) + \varepsilon$ at all. There are more reasons why complex experiment could be unsuccessful, below, I would like to show the most common problems with complex experiments:

1. Problem with unclear instructions
2. Problem with unclear evaluating
3. Problem with unclear motives

Basically, as you can see, we must deal with explaining the experiment to participants, make the experiments clear to participants to follow one or few stimulus or motives and after that been able to process the data and reward participants.

1.1 Experiment 1 – problem with unclear instructions

The instructions must contain all the information, and only the information, that the participants need to perform the experimental task. Usually the instructions are printed on paper and distributed at the beginning of the experiment. Instructions must also be dear, sharp, and of the right length. The first two requirements do not need any justification: unless your explicit goal is to create confusion in the subjects, you had better seek simplicity. Some concepts require careful formulation: it may make some difference, for instance, to use the expression "private account" instead of "individual account ": in general, it is a good idea to avoid morally charged terms like "altruism," "egoism," and so on, which might induce subjects to believe that you expect them to behave in the “right” way. It is also a good idea not to use economists’ jargon: first, most people do not know the meaning of economics' technical terms, and secondly, some of these terms are normatively laden. You usually do not want to tell subjects what is in their “rational” interest to do, or what the “equilibrium” of the game is, and so forth. You might not even want to let them know that there exists a “rational” solution to the game. On the other hand, you should not make the opposite mistake of being simplistic. The instructions must not be too short, and it is important that the subjects understand all the subtleties of the situation they will face in the experiment. (Guala 2005)

This kind of problem is easily solvable, but requires time and sacrificing a certain data sample. Organising the pilot experiment to enhance understandability of the experiment is recommended way to improve the experiment. On the other hand, there will be basically at least one experimental group which would provide you data, if would not be pilot group. This might be problem especially in the case of time constraints or a small statistical sample suitable as participants of the experiment. In this case, you can use only one or two people as pilot group, but you still risk some overlook or misunderstanding.

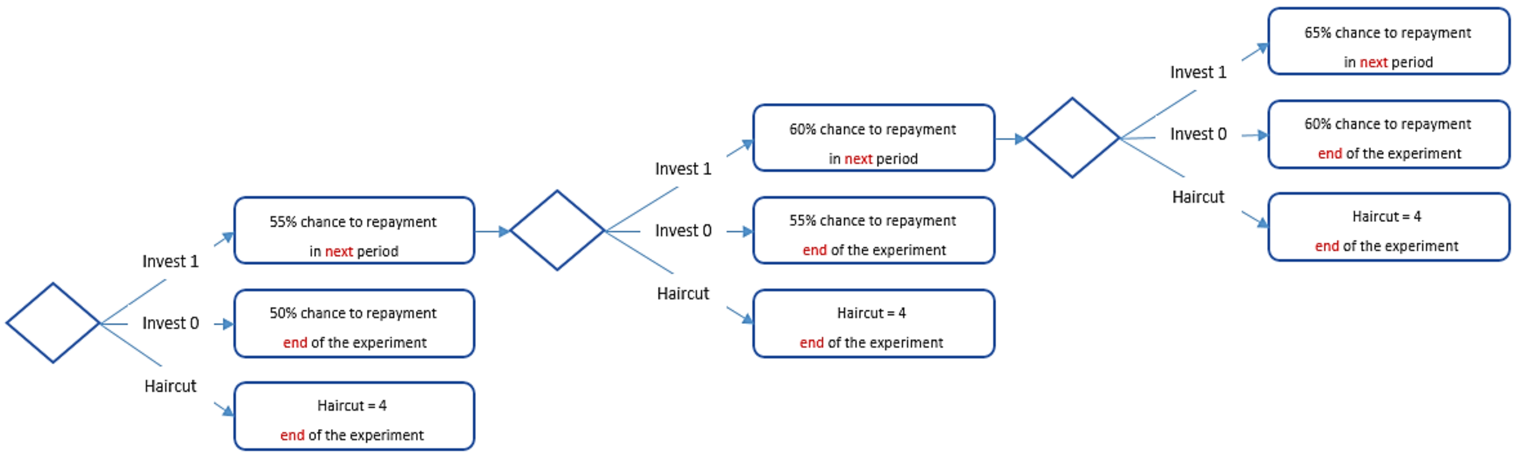
We participate at the experiment which shows this kind of mistake and unfortunately debased data from one whole group of 20 participants before the mistake has been corrected. The experiment simulated (simplified) Greek debt situation, where participants were in the position of creditor. Experiment has been designed in such a way that Greek already has debt to creditor which was partly covered by Greek’s assets and repayment was uncertain (at the beginning of the experiment the probability was 50%). Creditor had 3 options:

1. Invest more in Greek and improve his probability of repayment in next period
2. Do not invest and risk no repayment at all
3. Take haircut[1]

Repayment was at 20 points (haircut was 4 points, investment was 1 point). Picture 1 shows experimental design for first three periods.

[1] *Haircut means the difference between the market value of an asset used as loan collateral and the amount of the loan.* (Gorton a Metrick 2009)

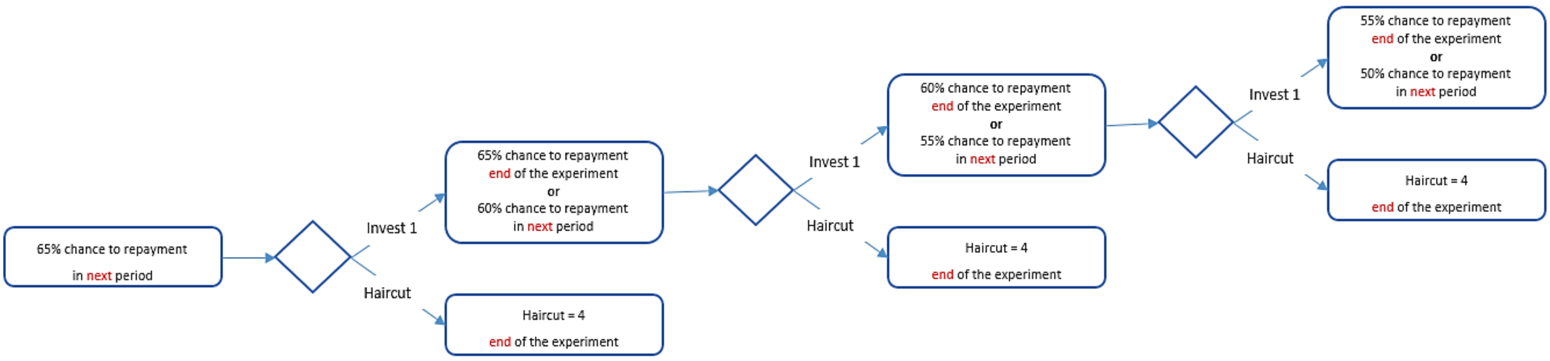
Fig. 1: Greek experiment (period 1 – 3)



Source: own

After 3 periods, the rules have changed. There were only 2 possibilities, and chance to repayment start to decrease. Investing was just a chance how to extend the possibility of being repaid. In this case you could be repaid (with some probability and in case you did not been repaid, you have lower chance in next period). Picture 2 shows experimental design for first three periods after changing the rules.

Fig 2: Greek experiment (period 4 – 6)



Source: own

After changing the rules, the experiment could run to 0% probability of being repaid (after that it would have no sense).

The experiment has been developed with one trick to investigate loss aversion and status quo bias: even if there were declared probability of being repaid, every decision which could extend the game would only extends the game – with no chance to repayment. The only options to be repaid was no investment (with declared probability) or haircut.

After all that, the first experimental group failed to go through the experiment correctly, due to misunderstanding in instructions. In those instructions were specifically said: *"…the probability varies by 5% each period"*, which looks straight forward. But participants, given to first three periods (where probability rises), took those instructions as *probability rises by 5% each period* which is evidently not true and do not follow the experimental design (even when right probability has been declared, participants just overlooked the numbers and relied on their experience). The misunderstanding has been fixed by wording *"…the probability rises or fall (depends on situation) by 5% each period"*, which surprisingly was enough to treat this problem. This kind of situation is typical example of missing experimental pilot group (at the end, the first group was unintended pilot group).

1.2 Experiment 2 – problem with unclear evaluating

One of the key elements of behavioural economics is measuring productivity at different incentives, stimulus and motives. If we accept fact, that work performance is not influenced only by reward, we can weigh up many potential things which can change human motivation.

The purpose of the following experiment was to test how performance could be affected if we reward same work by fixed fee and by variable fee. Simultaneously, same quantity of work was rewarded by the same fee.

For this type of experiment, there is obviously the need for precise measurement of work performance. The first idea could be to pick some reading tests or make participants to solve some dilemma. The problem with this approach is, that those types of work are often hard to measure, which is problem not only for rewarding participants, but for the experiment as a whole – if we cannot measure performance, we cannot see clear influence of the I.V. For this reason, the experiment was developed as a typing task. We took some digitalized statistic from Google[2] and we ask participants to rewrite those statistic to Excel. In this case we could measure precisely how much work every participant did. Technically, there organized 4 sessions with undergrad students (every session took 20 minutes). Two of those sessions have been rewarded by fix fee and two sessions have been rewarded by variable fee.

The problem with this design is probably not visible for the first look, but it would be devastating if not resolved. In this case, there is no chance to control the participants, if they rewrote the statistic precisely, or if they did more or less mistakes. At the end, there could be

[2] *There are large numbers of statistic which are digitalized by Google, but only by scanner. Therefore, there is hard manipulation with them. Excel form has obvious pros.*

participants with high performance, but also with high error rate – which is evidently the case which should not enter in the situation.

First obvious solution would be to control every number, but this would require to already have all Excel finished. The purpose here was also to gain those statistic, and this king of solution would make the experiment redundant.

At the end, I decided to have softer version of this kind of control and also leave the responsibility for the control on participants. We added in our Excels sums of every column and give those information to participants. Then, we did not have to control every column, the formula in Excel just told us the solution. Moreover, participant could fix their mistakes, which cleared the I.V. – D.V. relationship, but still punish mistakes by losing time.

In the case of interest, results of this experiment are available paper of Obergruber (Obergruber 2015)**.**

The Excel forms could help us I this case, but generally we could not rely on such kind of control. Even in this simple experiment it took hours of preparation just to make correct forms. In more difficult cases, this procedure would be impossible. There is one more solution to this dilemma, which also significantly reduces the cost of participants. The main idea of this is basically rewarding randomly only one or few participants instead all of them. Pros of this practice are not only lower cost but also requirement to check only part of the work participants did. This takes significantly less time without harming examined motivation of participants or gathered data.

1.3 Experiment 3 – problem with unclear motives

There are many studies regarding overconfidence in behaving in nearly every possible market we can imagine. Additionally, there are many studies regarding entrepreneurship (this is an independent science field) (Camerer and Lovallo, 1999). However, there is much less work regarding the influence of overconfidence in entrepreneur's decision making. The following design should have tested the relation between overconfidence and entrepreneurship attributes. The development of this experiment raised certain unexpected but serious problems, which we want to show. There are certain verified questioners that are very successful in predicting better and worse potential entrepreneurs. We used the classic Big-five[3] questioner (Cantner, Silbereisen, a Wilfling 2011). With this, we could define our final relation:

$$overconfidence = f(openness + conscientiousness + extraversion + agreeableness + \ neuroticism) + \varepsilon \quad (1)$$

With this, a person with entrepreneurship potential should score high in *openness*, *conscientiousness* and *extraversion* and score low in *neuroticism* and *agreeableness* (Cantner, Silbereisen, a Wilfling 2011)

The overconfidence should have been tested in a market game (more participants were in the same market). Every participant should have decided (after he received certain information) whether he wanted to enter a market in which he could earn money but did not need to. The decision to enter the market cost him certain sources. This decision should be made based on whether he believes he is better or worse than his competitors; with product[4] he has.

Every participant is, at this point, visible for certain customers in the market. Customers are programmed. Customers can solely observe 2 participants in the market randomly and buy products from the better one (for price = 1). In that case, there is no guarantee that the payment will always reflect the quality of the product. A high payment could mean a good product or luck and low payment vice versa.

After payment, every participant in the market should make decision regarding whether he wanted to remain in the market or leave the market. The payment was easy to calculate probability of getting payment = 1 when we knew all imputes.

In addition, we need to differentiate those entrants who believed they were lucky or unlucky from those who were lucky or unlucky. This need has been partly solved by the possibility to leave the market; however, as we observe, there could be participants who obtained a payment, but should have expected more. Therefore, the participant has no reason to leave market; however, his beliefs are different from the situation. Thus, we developed the last part of our experiment. After the payment, we request the follow-up action: 'Assign the probability to ranks you believe you are in the market'. Unfortunately, we could not directly request the rank the participant believes; therefore, we needed to use probability. The reason was that, with the direct question on rank, the outcome data could solely be *right* or *wrong*,

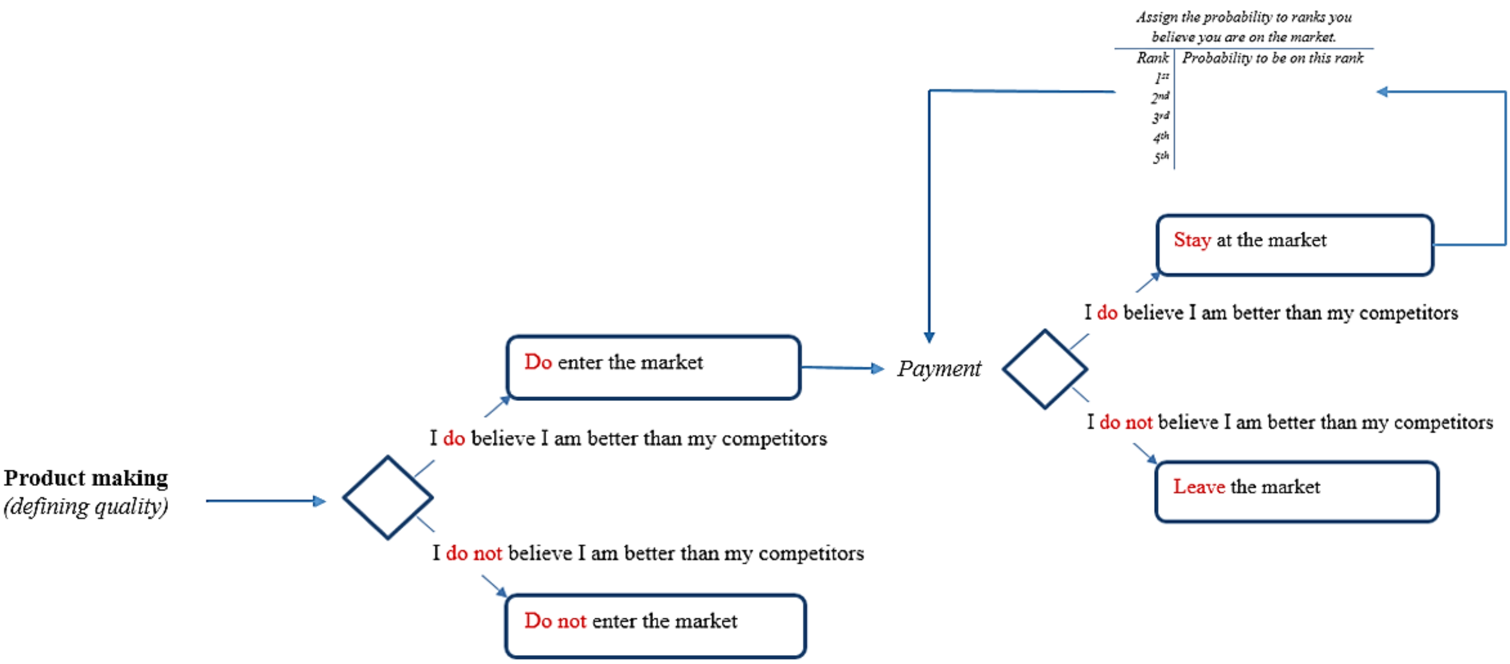
[3] *Big Five personality traits are five broad domains or dimensions of personality that are used to describe human personality, otherwise described as the five-factor model. The five factors are openness, conscientiousness, extraversion, agreeableness, and neuroticism (Costa & McCrae, 1992)*

[4] *In this part of the experiment, participants received the product that they could sell in the market. The quality of product was the sole difference between all products of all participants, and it has been based on a simple skill game participants played previously.*

which is very restrictive. We could not assume that, if the participant would miss, for example, by one, that the outcome would me *wrong, but close*. The reason is that those beliefs regarding ranking do not need to be normally distributed. Therefore, we were required to request the probability of the beliefs regarding each rank.

The entire process has been planned to repeat five times. The idea for measuring the overconfidence is as follows: When the participant shows more willingness to enter the market than objective information says, then he should be (based on difference of information and decision to enter or not enter) identify as overconfident. Picture 3 shows the process of every participant. The squares identify the decision-making process. As picture 3 shows, there is more than 1 decision-making process, which is the point where all data became highly complicated and unclear. Moreover, the understanding of payment rules, which are crucial for participant`s decisions, is also not easy.

Fig. 3: Big5/overconfidence experimental design (complex)



Source: own

As the entire picture 3 shows, this experiment has been designed as extremely difficult to understand and more difficult to process potential data (moreover, the product defining part and payment part has its own specific rules). Ultimately, this design has been (after many corrections) abandoned!

In this example, we wanted to show and share the mistakes that an experimental design could have. The effort to solve as many problems as possible cannot prevent the simplicity and comprehensibility of the experiment. At the same time, researchers must consider data they will gather. With design, we showed that data could be so complicated and interrelated that any statistical analysis could be beyond the scope of the researchers. The introduced experiment would gather data regarding the beliefs of product quality, beliefs regarding luck and beliefs regarding the likely rank, 5x in the row. This process is something that few people (if any) could reliably compile. We choose to redesign whole structure of experimental design. You can find the design with results in Obergruber paper (Obergruber a Hrubcova 2016)

Conclusion

When economic are struggling to understand a key causal relationship of variables or cause-causality chain, they have long used this semiautomatic approach: write down a theoretical model and start looking for available naturally occurring data. To economists, research has often meant chatting with the cab driver on the way from the airport to another academic seminar. But more and more empirical economists are opening their eyes and searching for situations and questions in which an experiment might offer a feasible and desirable approach (List 2011).

In this paper, we introduce you a brief introduction to experimental methods in economics. Moreover, we tried to show the way the experimentally acquired data should be treated and how an experimental design should look like to make the data relevant. As it was noticed, experiment still represent unorthodox way to obtain data not only because short tradition of experimental research in economics, but also because they force the researcher to understand everyday phenomena, many of which we stumble upon frequently.

As it was mentioned, the very structure of the experiments is and must be the same as in other sciences. We always want to describe one variable through the other variable - in other words: find causality. The statistical prescription for this is known for centuries: find changing *independent variable* and look how your *dependent variable* changes. Than find possible regularity in this relationship: $D.V. = f(I.V.) + \varepsilon$.

We introduced you 3 experiments and show the most common and most harmful problems in experimental design or execution: Problem with unclear instructions, problem with unclear motives and problem with unclear evaluating.

We introduced on first experiment possible ways how to treat with unclear instructions. It shows that even experienced designer could overlook crucial detail, which can change message of whole instruction purpose. Moreover, motivation of participants to behave in certain way in the experiment could be significantly changed, which obviously harm the D.V. – I.V. relationship and at the end could ruin whole experiment. On this example, we purpose to organize pilot group to treat this problem. The costs are sacrifice data from one whole group and longer time to execute whole experiment.

In the case of the second experiment, we introduced the problem of evaluating of participants and measuring their performance or results. We argue that the evaluating and rewarding of participants must be designed in effective and most importantly fast way. There are basically two approaches to do so. First, come up with system which can quickly measure participant´s performance for you (like algorithm or prepared check points). Second possibility is basically bypass this need for quick measurement of large amount of data and decide to measure only one of few performances (selected randomly). The second approach contraintuitively do not decrease validity of dataset and allow to reduce cost of experiment and shrink time to execute the experiment.

In the third experiment, we introduced the problem of unclear motives. Same as in the first experiment, this situation damages the D.V. – I.V. relationship. In this case it is not because wrong explanation what we want from participants, but because of some possible reasons or motives, which can drive participant´s behaviour, but which are not treated or considered in experimental design at all. In this type of situation, we do not propose to treat every possible variant in the case that there are too much possible scenarios. On the contrary, our (and not only our) best experience is to redesign whole experiment and make him less intuitive and more straightforward. We argue that this approach would generate cleaner data and it would improve explanatory power of the experiment. In case of treating every scenario in complex experiment, there is high risk not only of unclear motives, but also noisy dataset, which could blur the D.V. – I.V. relationship.

Acknowledgment

This article was created with the help of the Internal Grant Agency of University of Economics in Prague No. 80/2014 under the title “Non-monetary Incentives on the Labour Market”.

Bibliography

Camerer, Colin, a D Lovallo. 1999. „Overconfidence and Excess Entry: An Empirical Approach". *The American Economic Review* 89(1): 306–18. http://www.researchgate.net/publication/4980840_Overconfidence_and_Excess_Entry_An_Empirical_Approach.

Cantner, Uwe, Rainer K. Silbereisen, a Sebastian Wilfling. 2011. „Which Big-Five personality traits drive entrepreneurial failure in highly innovative industries?" *Friedrich Schiller University Jena, Department of Economics* (April): 1–36.

Gorton, GB, a A Metrick. 2009. „Haircuts". *NBER Working Paper Series*. http://www.nber.org/papers/w15273.pdf.

Guala, Francesco. 2005. *The Methodology of Experimental Economics*. Cambridge University Press. https://books.google.com/books?id=5em5qjqFf0wC&pgis=1 (27. leden 2016).

Kahneman, Daniel, a Amos Tversky. 1979. „Prospect Theory: An Analysis of Decision under Risk". *Econometrica* 47(2): 263–92.

List, John A. 2011. „Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off". *Journal of Economic Perspectives* 25(3): 3–16.

Obergruber, Petr. 2015. „Statistic and experimental economics". *The conference proceedings: The 9th International Days of Statistics and Economics*: 1233–44.

Obergruber, Petr, a Gabriela Hrubcova. 2016. „Driving Forces of Entrepreneurship; an Experimental Approach". *Procedia Economics and Finance* 39: 493–501. www.sciencedirect.com (18. 2. 2017).

Tversky, Amos, a Daniel Kahneman. 1992. „Advances in Prospect Theory: Cumulative Representation of Uncertainty". *Journal of Risk and Uncertainty* 5(4): 297–323.

Contact

Petr Obergruber

University of Economics, Prague

The Faculty of Business Administration

Department of Microeconomics

W. Churchill Sq. 4

130 67 Prague 3

Czech Republic

petr.grub@gmail.com