SCHRIFTEN DER GESELLSCHAFT FÜR WIRTSCHAFTS- UND SOZIALWISSENSCHAFTEN DES LANDBAUES E.V.

GEWISOLA

Moser, S., Mußhoff, O.: Reward, Punishment and Probabilities in Policy Measurements: An Extra Laboratory Experiment about Effectiveness and Efficiency of Incentives in Palm Oil Production. In: Mußhoff, O., Brümmer, B., Hamm, U., Marggraf, R., Möller, D., Qaim, M., Spiller, A., Theuvsen, L., von Cramon-Taubadel, S., Wollni, M.: Neue Theorien und Methoden in den Wirtschafts- und Sozialwissenschaften des Landbaus. Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V., Band 50, Münster-Hiltrup: Landwirtschaftsverlag (2015), S. 283-295.

Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V., Bd. 50, 2015, S. 283 – 295

REWARD, PUNISHMENT AND PROBABILITIES IN POLICY MEASUREMENTS: AN EXTRA LABORATORY EXPERIMENT ABOUT EFFECTIVENESS AND EFFICIENCY OF INCENTIVES IN PALM OIL PRODUCTION

Stefan Moser,[1] Oliver Mußhoff

Abstract

Palm oil production creates negative externalities, e.g. through intensive fertiliser application. If policy wants to determine externalities an effective and efficient measurement seems desirable. Embedded in an extra laboratory field experiment on Sumatra, a business simulation game tests several incentives for reducing the use of fertiliser. These incentives are differently designed, i.e., either reward or punishment, varying in their magnitude and probability of occurrence but constant in the effect on expected income. Results show that participants react significantly different depending on the incentive design. A high reward with a low probability to occur was found to be the most effective and sustainable incentive design. For efficiency, a low and certain reward is indicated to be the best design.

Keywords

policy measurement, effective incentive, efficient inventive, field experiment,
business simulation game, Indonesia

1 Introduction

With a production of approximately 50 million tons in the year 2012, palm oil is the most significant vegetable oil in the world (FAOSTAT, 2013). Starting in the 1960s, with a production of several million tons, palm oil production has grown exponentially, doubling every 10 years. With 23.6 million tons in 2012, Indonesia is the largest palm oil producer in the world, working towards increasing its production by up to 40 million tons by 2020 (UNCTAD, 2013).

Exaggerated fertiliser use in general and of nitrogen in particular leads to several negative externalities in palm oil production. For example, JELSMA et al. (2009) state that fertilizer use contributes a large proportion of total CO2 emissions from palm oil production. SEKHON (1995) shows for developing countries that, the increased usage of N-fertiliser causes groundwater pollution. Additionally, the use of N-fertiliser in humid, tropical climates increases NOx emissions, which are a major contributor to global warming (VELDKAMP and KELLER, 1997; REIJNDERS and HUIJBREGTS, 2008; KELLER and REINERS, 1994). Fertiliser in the tropics leads to NOx emissions (VELDKAMP et al., 2008; PALM et al., 2002) and can cause ground-level ozone in tropical oil palm plantations, high concentrations of which can be detrimental to human health (HEWITT et al., 2009) and reduced CH4 uptake (PALM et al., 2002). In summary, the exaggerated use of fertiliser in palm oil production causes negative externalities, which should be accounted for. Independent of which measurement in detail to choose, there should be an effective and efficient incentive to account for externalities.[2]

Incentives differ in several ways, making them different in efficiency and effectiveness. One differentiation is if there is a reward for desired or a punishment for undesired behaviour. An-

[1] Georg-August-Universität Göttingen, Department für Agrarökonomie und Rurale Entwicklung, Platz der Göttinger Sieben 5, 37073 Göttingen; smoser@uni-goettingen.de

[2] Effectiveness of an incentive means the decrease on fertiliser use; efficiency of an incentive means the influence on social welfare.

other differentiation is the probability of occurrence when behaviour is undesired. Incentives also differ in their magnitude. In combination, magnitude and probability of occurrence give the effect on the expected income, i.e., the cost or reward for cooperative or uncooperative behaviour, respectively. As long as the effects on expected income are held constant, a rational, risk neutral agent would behave independent of the probability and the magnitude of an incentive (BECKER 1974). BALLIET et al. (2011) lists further possible differentiations for designs, e.g., cost of giving incentives, centralized versus decentralized source of incentives, matching procedures, iterations, type of dilemma, participation payment and cost-to-fine ratio. SUTTER et al. (2010) find that endogenously chosen incentive mechanism increase cooperation more than exogenously chosen ones. A paper form HERRMANN et al. (2008) shows that the cultural context matters for the incentive efficiency. All of these differences and their interaction effects may influence the overall effectiveness and efficiency of incentives.

The objective of this paper is to analyse the effectiveness and efficiency of different incentive designs for farmers about which little is known. More specifically, we investigate in incentives for fertiliser reduction for the case of small-scale palm oil producers in Indonesia's island of Sumatra. The effectiveness and efficiency of reward and punishment with respect to the influence of magnitude versus probability of occurrence for incentives are analysed. The business simulation game applied within an extra laboratory experiment seems to be a method that achieves high internal as well as external validity for the context at hand.[3]

This paper contributes to the current literature in several ways. We are the first testing effectiveness and sustainability of particular designed incentives[4] with small scale farmers within an extra laboratory experiment. This allows for a direct comparison of the differently designed incentives effectiveness. Furthermore, this simulation also enables drawing conclusions about the efficiency of the incentives. Another contribution is the group of participants. Other business simulation games (MUSSHOFF and HIRSCHAUER, 2012) work with convenience groups for their experiment. In contrast, the study at hand works with small scale farmers in Indonesia for the extra laboratory field experiment.

The remainder of the paper is structured as follows: In section two, the literature is discussed and hypotheses are generated. In section three, the method and experimental design is illustrated. Section four gives a description of the context and the sample recruitment, whereas section five states the data analysis. Section six presents the results with discussion. Finally, section seven closes with a conclusion and an outlook.

2 Literature and hypothesis generation

The effects of incentives on the expected income, which results from its probability of occurrence and magnitude, was already discussed by BECKER (1974). According to him, holding expected income constant and assuming a rational, risk neutral agent, it makes no difference for the behaviour if the incentive is designed as reward or as punishment or if the magnitude of an incentive is increased on cost of its probability of occurrence. However, the literature of experimental economics shows that the reactions on an incentive not only depends on the effects on expected income. Also the incentive design matter, i.e. rewarding or punishing and how the probability of occurrence and magnitude of an incentive is sized.

For the consequences of rewarding or punishing, extensive literature exists. KAHNEMAN (2011) gives examples for loss aversion, preconditioned there is a reference point which determines when a result is a loss. People are more motivated to avoid losses than to gain profits

---

3 ROE and JUST (2009) define internal validity as the ability of a researcher to argue that observed correlations are causal. They define external validity as the ability to generalize the relationships found in a study to other persons, times and settings.

4 This particular design incentives includes rewarding versus punishing with different magnitudes and probabilities of occurrence holding expected income effects constant.

(HEATH et al., 1999), which implies that people are more sensitive to punishment than to reward. This is in line with several examples, e.g., for cab drivers in New York (CAMERER et al., 1997), for Swiss messenger services (FEHR and GOETTE, 2007) or for professional golf players (POPE and SCHWEITZER, 2011). According to these papers, punishing should be more effective than rewarding. Contrary, a meta-analysis involving 187 effect sizes from BALLIET et al. (2011) shows reward and punishment having equivalent positive effects on cooperation, even though differences occur depending on the context. A framed field experiment from IBANEZ and MARTINSSON (2008), which first combines reward and punishment simultaneously, showing that coca farmers react stronger to increasing relative profits than to increasing probability of occurrence for punishment. Along with that, we summarize that there is evidence that reward and punishment do not substantially differ in their effectiveness in general, but it may still be that these incentives differ in certain contexts.

According to KAHNEMAN (2011), a low loss probability leads to risk avoiding behaviour, whereas a low probability to gain causes risk loving behaviour. For losses, this implies that people react more to the magnitude of the incentives than to its probability of occurrence. In line with that, ANDERSON and STAFFORD (2003) found that for punishment, raising the magnitude causes more reaction than raising the probability of occurrence. Furthermore, BLOCK and VERNON (1995) found that for punishment, the behaviour depends on the participating group, but usually the reaction to the magnitude is stronger than to its probability of occurrence. The literature about the effectiveness of rewards with low probability of occurrence is not that extensive. KAHNEMAN (2011) claims risk loving behavior for rewards with low probability. This implies that people react more to the magnitude than to the probability in the case of rewarding. In a study about weight loss, VOLPP et al. (2008) found no significant difference between reward with high magnitude or high probability of occurrence. MUSSHOFF and HIRSCHAUER (2012) found that a certain reward is more effective than a higher but uncertain one. In summary, these papers overall expect that people react stronger to the magnitude than to the probability for punishment as well as for rewarding.

This leads to the first hypothesis: *For the same effect on expected income, the effectiveness of an incentive is independent of its design*[5].

In their meta-analysis, BALLIET et al. (2011) discuss the sustainability of incentives on cooperation over time. They found that incentives became more effective over time, compared to one-shot experiments. They explain this effect on both individual learning and group learning processes. The group process is described as being driven by group norms, generating expectations of group cooperation and building up reputation. They do not indicate whether reward or punishment is more effective in terms of sustainability. Moreover, they do not discuss how the magnitude or the probabilities of incentives affect its sustainability. On the other hand,

This leads to the second hypothesis: *For the same effect on expected income, the sustainability of an incentive is independent of its design.*

The influence of incentives on social welfare, i.e. efficiency of an incentive, is another dimension to analyse. In the literature, often a public good game is applied to find out about efficiencies of incentives. In their meta-analysis BALLIET et al. (2011) summarise that punishment is efficient, at least on the long run, but they do not directly compare the efficiency of reward and punishment. RAND et al. (2009) or SEFTON et al. (2007) show that reward is more efficient than punishment. However, little is known how efficiency develops if an incentives magnitude is raised on the cost of its probability of occurrence.

This leads to the third hypothesis: *For the same effect on expected income, the efficiency of an incentive is independent of its design.*

---

[5] Design means if the incentive is established as reward or punishment, which occur certainly or the magnitude of an incentive is increased on cost of its probability of being occurred

3 Method and Experimental Design

To test for these hypotheses, an extra-laboratory experiment (CHARNESS et al., 2013) is conducted. who classifies such experiments as extra-laboratory, discuss advantages and disadvantages of this method, for example, a better opportunity to handle the subject pool or practical drawbacks. ROE and JUST (2009) point out that field experiments enable a good combination of high internal and external validity. For the internal validity, a direct comparison of the analysed incentives is desirable. To enable this, the effect on expected income has to be independent of the incentive design. For the external validity, the setting of the extra laboratory experiment has to be as realistic as possible. A business simulation game enables meeting these conditions, guarantying high internal and external validity.

The applied business simulation game looks like follow: It is simulated that each farmer manages a one hectare palm oil plantation, already established and in production, for 10 periods. Each period is an equivalent for one year. It is assumed that the yield on the simulated plantation solely depends on the use of fertiliser in the belonging period, i.e., weather conditions, diseases etc. are shared out. Each participant gets the same yield for the same amount of fertiliser. The only decision of the participant is how much fertiliser to use in the corresponding period. This decision may be influenced through a policy measurement starting in the 6th period. Later on, periods 1-5 and periods 6-10 are named the first and second episode, respectively. At the end, the achieved profit is calculated. To give incentive to the participants, 5 percent of the gained profit in the experiment is paid to them in the form of a shopping voucher.

The proceeding for each period is constant. The first step is the participants' decision of how much fertiliser to use. The second step is to find the price for the yield, i.e., palm oil bunches, which is supported by an assistant. In the third step an assistant states if the respective public measurement takes place or not. Finally, the experiment assistant determines the achieved yield, calculates the profit, as well as the amount of the shopping voucher. A more detailed explanation follows.

In the business simulation game the total amount of fertiliser used solely determines the yield on the simulated oil palm plantation. A quadratic production function is applied to calculate the yield. The functional form is $f(x) = 10 + 0.05x – 0.000{,}055x^2$, where x represents the amount of fertiliser. To reduce the mental effort for the participants, only 10 kg multiples for fertiliser were allowed to be used. The corresponding yields are transferred into a table and a graph, which is handed out to the farmers at the beginning of the simulation. The formula behind this table was not told to the farmers. The table starts with 0 kg and ends with 590 kg. Amounts beyond 590 kg were applied very seldom, because the maximum yield is already achieved with 440 kg fertiliser. JELSMA et al. (2009) states a fertiliser use of 4.5 kg per palm oil, resulting in a similar amount per ha. Also a report (FAO 2005) names a similar amount of fertiliser to use. The possible yields range from 10 tons to a maximum of 21.4 tons, which corresponds to the approximate annual yield per hectare found in the literature (FAIRHURST and MCLAUGHLIN, 2009; JELSMA et al., 2009).

The second step in the experiment is the evaluation of the price. In the first period, the price is fixed at 1,800 Rp. As basis for this, a price of 1,710 Rp per kg of fruit bunch in the year 2008 (JELSMA et al., 2009) was found in the literature. For all of the following periods, the price rises or falls by 200 Rp with a 50% probability, always based on the previous period. Thus, the price results in an arithmetic Brownian Motion (POITRAS, 1998) with a minimum of 0 Rp and a maximum of 3,600 Rp in the last period. To evaluate whether the price rises or falls, participants have to draw a ball from a bag which contains three green and three red balls. Prices rise for drawing a green ball and fall for drawing a red ball. These prices were determined separately for each participant, so a distribution of the price developments is achieved.

The third step, i.e., the public measurement, is the solely point where the design can differ between the periods or between participants. For the first episode, the participants always receive a fixed subsidy of 10,000 Rp per period, never mind their use of fertiliser. For the second episode, this public subsidy can change to a policy measurement for reducing the use of fertiliser, and furthermore stays constant until the experiment ends. This is where we test for the reaction on different designed incentives. At the beginning, the farmers are informed that subsidies may change after the 5th period, but are not given any further information. The treatments for the second episode are structured like follow:

Control Group: The public subsidy of 10,000 Rp goes on without any restriction, there is no change.

Treatment A: The subsidy is replaced by a compensation payment of 10,000 Rp if participants use 120 kg fertiliser or less in the corresponding period.

Treatment B: The subsidy proceeds. Additionally, there is a measurement which punishes everyone who uses 130 kg fertiliser or more in the corresponding period with 10,000 Rp. This equals a deduction of the received subsidy.

Treatment C: The subsidy is replaced by a compensation payment of 100,000 Rp if participants use 120 kg fertiliser or less in the corresponding period, but if so, there is a occurrence probability of only 10%.

Treatment D: The subsidy proceeds. Additionally, there is a measurement which punishes everyone who uses 130 kg fertiliser or more in the corresponding period with 100,000 Rp, but if so, there is a occurrence probability of only 10%.

The idea of the policy measurement standing behind treatment A and treatment B is certain rewarding for desired or certain punishing for undesired behaviour. For treatment C, the idea is close to the one described by LATACZ-LOHMANN et al. (2011). Essentially, they suggest an outcome based incentive system, where the probability of occurrence reflects the possibility of failing the defined outcomes, even when serious action is taken to achieve them. Treatment D considers the situation when controlling is imperfect and only successful with a chance of 10%. For treatments C and D, the incentives are equipped with a 10% probability of being occured. To determine if the participant is imposed or not, a similar procedure as for the price determination was applied. Participants have to draw from a bag with nine blue and one yellow ball. If they draw the yellow ball, the incentive occurs, whereas nothing happens if a blue ball is drawn. It is remarkable that the expected profit is equal for all treatments, except the control group. That means a rational and risk neutral participant would behave independently of the treatment. The trigger of 120 kg, as well as the magnitudes of 10.000 Rp or 100.000 Rp, is chosen arbitrary at a height where much reaction is expected.

The finally calculation of the realized profit of the period is straightforward. The revenue in a certain period equals the yield multiplied by the product price in the corresponding period. There is no option for storing, so the whole production has to be sold in the corresponding period. The cost for fertiliser is fixed at 10 Rp per kg. The revenue, costs and, if applicable, the public measurement or subsidy are added up. Then the next period starts and the participant has again to decide how much fertiliser to use. There are no interests for the gained profit, and if necessary credits are always available, making bankruptcy impossible. After 10 periods the simulation is over. If so, profits are added up and 5 percent of the achieved amount is given in the form of a shopping voucher to the participant.

4 Context and sample recruitment

The experiment is executed in the Jambi Province on Sumatra, Indonesia. Jambi has about three million inhabitants and is approximately 50,000 km² large. In last decades there was a strong transformation of the landscape towards palm oil plantations (LAUMONIER et al., 2010; WILCOVE and KOH, 2010). Therefore, it is a valuable research area for the topic of this paper.

The data was collected from 29 randomly chosen villages in the research area. For each village, an entire list of farmers was created. Depending on the size of the village in terms of inhabitants, between 10 and 18 small scale farmers are invited to participate. The business simulation game is done simultaneously for the whole group. Every participant from one village received the same treatment. At the beginning, each participant received a questionnaire. Then the experimental assistance explained the procedure supported by a visual poster and the participants had the chance to ask questions. During the experiment, the participants were divided into subgroups, so each assistant had to handle between 3 and 5 participants. This structure enabled the participants to ask questions on a more personal level. This is especially important for cultural reasons, since participants often hesitate asking during a presentation.

In total, 328 small-scale farmers participated in the experiment. Six uncompleted questionnaires were dropped out, resulting in 322 participants for the analysis. On average, participants made a profit of 37,940 Rp per period, resulting in 18,970 Rp worth of shopping vouchers for the entirety of the business simulation game. Considering that the average daily wage for a worker is around 50,000 Rp in the research area, this seems to be a good compensation for participating in this one hour experiment.

Table 1: Socioeconomic data and fertiliser use

	Sex	Age	Education years	Household size	Fertiliser (Period 1-5)	Fertiliser (Period 6-10)	Observations
Control	0,81 (0,05)	44,16 (1,49)	7,61 (0,42)	4,49 (0,18)	278 (8,18)	288 (7,88)	71
Treatment A	0,86 (0,04)	44,52 (1,35)	7,60 (0,36)	4,61 (0,24)	288 (8,39)	180 (8,07)	66
Treatment B	0,83 (0,05)	40,85 (1,41)	8,62 (0,45)	4,19 (0,16)	275 (9,63)	169 (7,80)	60
Treatment C	0,93 (0,03)	42,05 (1,61)	7,44 (0,39)	4,39 (0,21)	254 (9,21)	171 (7,56)	61
Treatment D	0,94 (0,03)	44,43 (1,38)	7,67 (0,39)	4,44 (0,20)	244 (9,08)	161 (6,02)	64
Observations	319	319	319	313	1610	1610	322
Kruskal-Wallis	0,649	0,112	0,406	0,648	0,002	0,000	

Source: Own data, Average amount and deviation per treatment, Standard deviation in parentheses
a) N varies between the variables

Table 1 gives the socioeconomic data of the participants. A Kruskal-Wallis test shows an equality of the population for the socioeconomic data for sex, age, years of education and household size of the participants. This already indicates proper sample selection.

5 Data Analysis

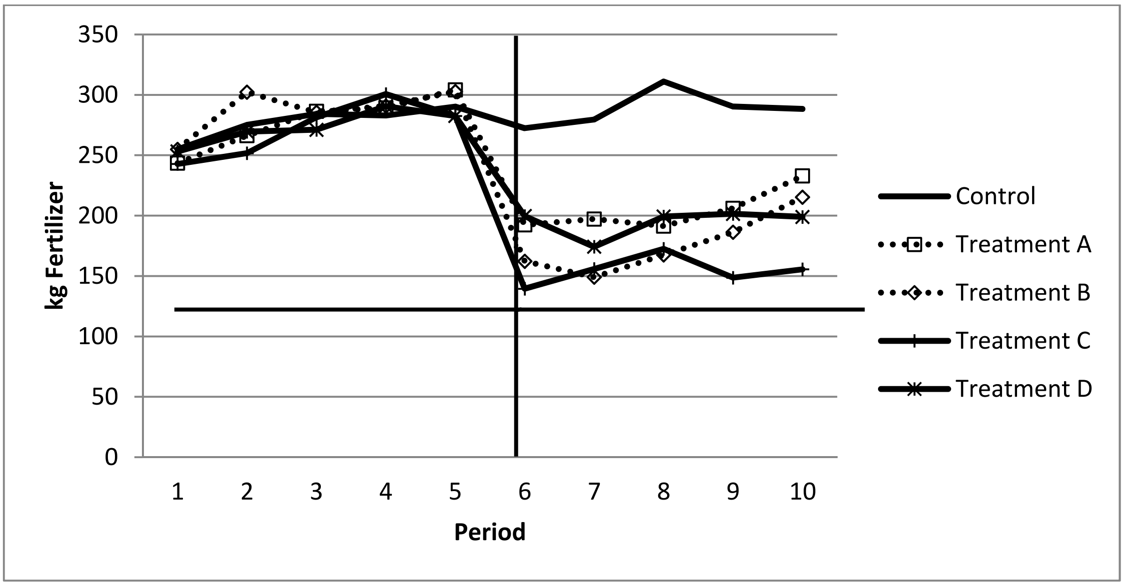
This section discusses how the collected data is analysed. To begin with, it is shown that a matching procedure is necessary, this is followed by a description of this procedure. Afterwards, the estimation of the effects is described.

In the first episode of the business simulation game, participants use 268 kg fertiliser on average. For the second episode, the control group uses 288 kg, whereas the four treatment groups use 171 kg fertiliser on average. This indicates that participants react to treatments taking place after the fifth period. Since the treatment for episode one is equal for each participant, it is expected that groups do not differ in their use of fertiliser in the first episode. A Kruskal-Wallis-test is applied to test for this assumption and shows that groups differ significantly in periods 1 to 5 at the 1% level.

The differences in fertiliser use of the groups in the first episode have consequences for the analysis. The trigger for the incentive in the second episode is 120 kg fertiliser. This does not have the same effect for each treatment group. For example, in period 5, participants with treatment D have an average fertiliser use of about 250 kg, whereas participants with treatment A have an average use of 380 kg fertiliser. Therefore, treatment D-participants have to reduce their use by 130 kg, whereas treatment A-participants have to reduce their use by 260 kg to pass the trigger level of 120 kg fertiliser. For adapting, treatment D-participants behaviour needs to change considerable less then treatment B-participants behaviour. Since we investigate in the changes through different incentives, and the basis for our analysis is the first episode, we cannot start the analysis with this structure of the data.

To overcome this problem, a minimum Euclidian distance matching is applied which is also used by other researchers (TIEDEMANN and LATACZ-LOHMANN, 2013). Therefore, the amounts of fertiliser used in the first five periods are the five matched variables. The treatments A, B, C and D are matched separately with the control group. This leads to 71, 72, 75, 79 and 83 observations for data analysis for treatment A, B, C, D and E respectively. The resulting average used fertiliser per treatment are displayed in Figure 1.

Figure 1: Average amount of fertiliser used per period and treatment after matching.



Source: Own illustration

Figure 1 shows the mean amount of fertiliser used per period and per treatment after the matching procedure. The horizontal line presents the 120 kg fertiliser trigger, whereas the vertical line stands for the policy change after period 5. As expected, after the matching a Kurskal-Wallis test indicates no significant difference between the treatments and the control group in the first episode. A look at Figure 1 already indicates differences between the treatments for the second episode.

The matched data are the basis for the evaluation of the differences in the treatments. To test for our hypothesis, we compare episode one, where each participant is treated equally, with episode two, where the different treatments take place. With the availability of panel data, a fixed effect regression allows accounting for unobserved, unchanging participant-specific effects. A Hausman-test indicates that a random effect model does not fit in this situation.

To analyse the effectiveness several variables are used for the regression. To test for the shift for used fertiliser through the policy intervention, a dummy variable is introduced for each treatment, showing the difference between the first and the second episodes. A look at Figure 1 indicates that the average used amount of fertiliser rises within the first episode. This is interpreted as a learning effect, which affects all treatments in the first episode. To account for that, a learning variable is introduced for all treatments. For the first episode, this variable starts with zero and up until the fifth period, the variable increases by one per period. Afterwards, this variable remains constant, since gained experience is not lost. For the second episode, we expect different developments throughout the treatments. So, this learning effect is split up and each treatment receives an own variable. Except for the control group, the steadily development towards the profit maximum is not to expect here, since it is restricted by the 120 kg trigger for the incentives. For the control group, it is still the learning effect to come closer to the profit maximum. For the four treatment groups, it is interpreted as a measurement of the sustainability of the treatment. The variables for the second episode start with one in period six and increases by one per each following period.

To analyse the efficiency of the treatments, the profit difference is used. The profit difference is the difference in profit between the first local profit maximum which is always at the trigger of 120 kg fertilizer use and the second local profit maximum which is around 400 kg, i.e., when participants ignore the incentive. This difference is calculated separately for each treatment in the second episode. For the first episode it is zero. It is a measurement of social costs of adapting the use of fertiliser to the incentive. If Participants react on profit difference, they reduce the amount of fertiliser when it is cheap and vice versa increase the amount of fertiliser. So, the generate the externalities when it is cheap, which increases social welfare. To enable evaluating these social welfare effects, i.e. efficiency, per treatment, these effects are estimated separately.

The profit difference is mainly driven by the product price, whereas the adoption of fertiliser use has a minor role. Therefore, estimating a price effect for the treatments would lead to multi-collinearity. For the control group no profit difference exists, therefore the influence of price can be taken into account. For the first episode, the price effect is estimated for all treatments together. Due to the design of the business simulation game, the average profit difference is negative. Therefore, a part of the shift, which represents the measurement in effectiveness, will be explained through the profit difference. This leads to a bias for the of the shifts of the treatment. To avoid this bias, the mean value of profit shifts per treatment is added. This measure corrects the shift for the described bias and has no influence on the other results.

6 Results and Discussion

The results of the conducted estimation are shown in Table 2.

Table 2 shows the results of the estimation. For the estimation 3800 observations are used and the adjusted R-squared is 0.63. For all treatments, the shift in the second period is significant negative at 5% level, whereas the shift for the control group is insignificant. This indicates that each treatment has a significant negative influence on the amount of fertiliser used. The shifts for the treatment groups have different sizes. Treatment A and treatment D with 47.9 kg and 47.7 kg respectively show the lowest shifts. Wald tests clearly indicates that treatment B and treatment C with 102.5 kg and 84 kg respectively have significantly higher absolute shifts, whereas between these two pairs of treatments no significant difference is found. So for the same effect on expected income, treatment B and C are significantly more effective. It can be stated that the size of the shift depends on the design of the incentive. It seems for punishing a high probability is more effective than a high magnitude effective, whereas for rewarding a high magnitude is more effective than a high probability. This means that the first hypothesis: *For the same effect on expected income, the effectiveness of an incentive is independent of its design.*" has to be rejected.

Table 2: Estimation results of Fixed Effect model.

	Coefficient	Standard Error	p-Value	
Shift, Control	26.6	(31.6)	0.399	
Shift, Treatment A	-47.9	(24.0)	0.040	*
Shift, Treatment B	-102.5	(23.8)	0.000	***
Shift, Treatment C	-84.0	(23.9)	0.000	***
Shift, Treatment D	-47.7	(23.7)	0.035	*
Learning, Episode 1, All	9.95	(1.51)	0.000	***
Learning, Episode 2, Control	4.18	(3.48)	0.230	
Sustainability, Treatment A	5.67	(3.47)	0.103	
Sustainability, Treatment B	14.33	(3.38)	0.000	***
Sustainability, Treatment C	2.02	(3.30)	0.541	
Sustainability, Treatment D	2.65	(3.22)	0.409	
Price, Episode 1, All	0.031	(0.011)	0.006	**
Price, Episode 2, Control	0.007	(0.014)	0.613	
Profit Difference, Treatment A	0.018	(0.002)	0.000	***
Profit Difference, Treatment B	0.011	(0.002)	0.000	***
Profit Difference, Treatment C	0.004	(0.002)	0.034	*
Profit Difference, Treatment D	-0.003	(0.002)	0.247	
Constant	200.6	(20.6)	0.000	***

Source: Own estimation

a) * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

For punishment, ANDERSON and STAFFORD (2003) and BLOCK and VERNON (1995) found magnitude has a higher effect than probability of occurence. Also KAHNEMAN (2011) states that low probabilities of occurrence lead to risk avoiding behaviour for losses. For punishment, this implies higher effectiveness if the magnitude is increased on cost of the probability to occur. This contradicts our research, since we found the certain punishment, i.e. treatment B, to be more effective compared to the uncertain punishment with high magnitude, i.e. treatment D.

For reward, VOLPP et al. (2008) found no difference in the effectiveness of magnitude or probability to occur. KAHNEMAN (2011) states that low probabilities to occur lead to risk loving behaviour for gains, i.e. reward, which implies higher effectiveness if the magnitude is increased on cost of the probability. This is in line with our results since treatment C with the high and uncertain magnitude is significantly more effective than the certain treatment A.

Comparing certain punishment with certain reward, we find punishing to be more effective. This is in line with the literature (HEATH et al., 1999; CAMERER et al., 1997; FEHR and GOETTE, 2007; POPE and SCHWEITZER, 2011). In their meta-analysis about cooperation in social dilemmas BALLIET et al. (2011) found slightly higher but not significant effectiveness of punishment compared to reward. Also in line with our results, KAHNEMAN (2011) states, preconditioned there is a reference point, that people are loss averse. These statements are in line with our result that certain punishment is more effective than certain reward. Interestingly, if we increase the magnitude of the incentives on cost of their probability to occur, the effect turns around and reward becomes more effective than punishment. It seems that the willingness to take risks in case of reward is stronger than the risk-averse behaviour for punishment (KAHNEMAN, 2011). Our results indicate that an incentives probability to occur may play an important role.

For the first episode, the learning effect is estimated for all participants at once. The average amount of used fertiliser increases significantly by about 10 kg per period, on average, for the first episode. The maximum profit would be reached with a fertiliser use of around 400 kg, depending on the price. Since most of the participants are far below that, the direction, as well as the size, of the effect is reasonable.

For the second episode, the learning effect is split up into the learning effect for the control group and the sustainability effect for the treatment groups, showing different developments. For the control group, the effect of 4.18 kg per period is positive, but not significant. This shows us that the approach to the profit maximum slowed down or might have even stopped for the second episode. This remaining learning effect is used as the benchmark for sustainability. If treatment groups have a significantly higher upward movement, this indicates a lack of sustainability of the shift through the treatment.

Treatment B has the only sustainability coefficient which is significant different from zero. Moreover, Wald tests indicates that it is the only treatment with a significantly higher coefficient than the control group. For treatment A, treatment C and treatment D the sustainability coefficients are neither significant different from the control group nor significant different from the zero, which indicates sustainability. This shows that the shift through the treatment is not sustainable for treatment B. Therefore, the second hypothesis, i.e., "*For the same effect on expected income, the sustainability of an incentive is independent of its design.*" has to be rejected.

Sustainability can be compared with iterated dilemmas. BALLIET et al. (2011) found punishment to be insignificantly more effective than reward. This contradicts our results, since certain punishment, i.e. treatment B, is the only treatment which is significantly less sustainable compared to the other treatments. For creating a high and sustainable reduction for fertiliser use, treatment C, i.e. high but uncertain reward, is the most effective design for an incentive.

The profit difference is the measurement for efficiency in this experiment. For one Rp in profit difference, participants adopt their fertiliser use for 0.018 kg, 0.011 kg, 0.004 kg and 0.003 kg for treatment A, treatment B, treatment C and treatment D respectively. A Wald test indicates a different adoption for all of these treatments. Since profit difference is the measurement for our social welfare efficiency, this order also represents the efficiency of the different treatments in this experiment. For the certain incentives, i.e. treatment A and treatment B, as well as for reward, i.e. treatment A and treatment C, participants show a significant higher reaction to the profit difference than to the uncertain or punishing incentives. The third hypothesis, i.e., *"For the same effect on expected income, the efficiency of an incentive is independent of its design."* has to be rejected.

In terms of efficiency, the cost here is not-realised production, whereas the gains are realised externalities. It is reasonable to gain externalities if they are cheap to have, and abdicate if the realisation of externalities is expensive. Participants react strongest to this trade-off for secure incentives and reward compared to insecure and punishment, respectively. This is in line with papers who found that reward is more efficient than punishment (RAND et al. 2009, SEFTON et al. 2007). In these papers punishing can crucial lower the social welfare through generating costs. In our experiments, no such costs occur, which even strengthen our results.

The price effects are related to the efficiency. These are positive in a reasonable extend, so participants seem to anticipate the price developments. For this simulation, a price uptick of 200 Rp causes the use of more fertiliser by 6.3 kg for the first period and 1.4 kg for the control group in the second episode.

7 Conclusion and Outlook

The use of fertiliser in palm oil production creates negative externalities. If policy wants to restrict such externalities, an effective and efficient incentive seems desirable. The aim of this paper is to test differently design incentives for policy measurements on their effectiveness and efficiency.

Results indicate a difference between the incentives, depending on their design. Participants react strongest either to certain punishment or to uncertain reward with high magnitude, holding effects on expected income constant. For the later design, the reaction is also sustainable, making uncertain reward with high magnitude to the most effective design. Furthermore, certain incentives are significant more efficient than uncertain ones as well as reward is significant more efficient than punishment. To sum it up, our findings suggest adapting the magnitude and probability of incentives can make a substantial difference for the effectiveness and efficiency of the measurement.

There are some implementations to improve external validity in the presented business simulation games. If corresponding data available the production function, prices for the harvest or for fertiliser can be further specified to get a more realistic setting. In addition, for deeper insights about efficiency in social welfare, the assumption of zero control cost can be weakened.

It is likely that participants would react similar if they are confronted with incentives for other measurements, e.g. against erosion, reduction for pesticide use or also wildlife protection. This is an interesting point to prove. However, forthcoming research should develop the method to create a more realistic setting for further increasing the validity of the method.

References

ANDERSON, L.R. and S.L. STAFFORD (2003): Punishment in a regulatory setting: experimental evidence from the VCM. In: Journal of Regulatory Economics 24 (1): 91-110.

BALLIET, D.; MULDER, L. B. and P.A.M. VAN LANGE (2011): Reward, punishment, and cooperation: A meta-analysis. In: Psychological Bulletin 137 (4): 594-615.

BECKER, G.S. (1974): Crime and punishment: An economic approach. In: Essays in the Economics of Crime and Punishment UMI: 1-54.

BLOCK, M.K. and G.E. VERNON (1995): Some Experimental Evidence on Differences between Student and Prisoner Reactions to Monetary Penalties and Risk. In: The Journal of Legal Studies 24 (1): 123-138.

CAMERER, C.; BABCOCK, L.; LOEWENSTEIN, G. and R. THALER (1997): Labor Supply of New York City Cabdrivers: One Day at a Time. In: The Quarterly Journal of Economics (112): 407-441.

CHARNESS, G.; GNEEZY, U. and M.A. KUHN (2013): Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics. In: Journal of Economic Behavior & Organization 91: 93-100.

FAIRHURST, T. and D. MCLAUGHLIN (2009): Sustainalbe Oil Palm Development on Degraded Land in Kalimantan. A study commissioned by World Wildlife Fund.

FAO (2005): Fertilizer use by crop in Indonesia. Food and Agriculture Organization of the United Nations.

FAOSTAT. http://faostat3.fao.org/faostat-gateway/go/to/download/Q/*/E. 18.October 2013.

FEHR, E. and L. GOETTE (2007): Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment. In: The American Economic Review 97 (1): 298-317.

HEATH, C.; LARRICK, R.P. and G. WU (1999): Goals as Reference Points. In: Cognitive Psychology 38 (1): 79-109.

HERRMANN, B.; THÖNI, C. and S. GÄCHTER (2008): Antisocial punishment across societies. In: Science 319: 1362-1367.

HEWITT, C.N.; MACKENZIE, A.R.; DI CARLO, P.; DI MARCO, C. F.; DORSEY, J. R.; EVANS, M. et al. (2009): Nitrogen management is essential to prevent tropical oil palm plantations from causing ground-level ozone pollution. In: Proceedings of the National Academy of Sciences 106 (44): 18447-18451.

IBANEZ, M. and P. MARTINSSON (2008). Can we do policy recommendations from a framed field experiment? The case of coca cultivation in Colombia. Working Papers in Economics 306. University of Gothenburg. School of Business, Economics and Law.

JELSMA, I.; GILLER, K. and T. FAIRHUST (2009): Smallholder Oil Palm Production Systems in Indonesia: Lessons from the NESP Ophir Project. In: Plant Productin Systems, Plant Sciences Group, Wagening University Wageningen.

KAHNEMAN, D. (2011): Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

KELLER, M. and W.A. REINERS (1994): Soil-atmosphere exchange of nitrous oxide, nitric oxide, and methane under secondary succession of pasture to forest in the Atlantic lowlands of Costa Rica. In: Global Biogeochem. Cycles 8 (4): 399-409.

LATACZ-LOHMANN, U.; SCHILIZZI, S. and G. BREUSTEDT (2011): Auctioning Outcome-Based Conservation Contracts: German Association of Agricultural Economists (GEWISOLA). 28.-30. September 2011, Halle.

LAUMONIER, Y.; URYU, Y.; STÜWE, M.; BUDIMAN, A.; SETIABUDI, B. and O. HADIAN (2010): Eco-floristic sectors and deforestation threats in Sumatra: identifying new conservation area network priorities for ecosystem-based land use planning. In: Biodiversity and conservation 19 (4): 1153-1174.

MUSSHOFF, O. and N. HIRSCHAUER (2012): Planspiele als experimentelle Methode der Politikfolgenabschätzung: Das Beispiel der Stickstoffextensivierung: German Association of Agricultural Economists (GEWISOLA). 26.-28. September 2012, Hohenheim.

PALM, C. A.; ALEGRE, J.C.; AREVALO, L.; MUTUO, P.K.; MOSIER, A.R. and R. COE (2002): Nitrous oxide and methane fluxes in six different land use systems in the Peruvian Amazon. In: Global Biogeochem. Cycles 16 (4): 1-21.

POITRAS, G. (1998): Spread options, exchange options, and arithmetic Brownian motion. In: Journal of Futures Markets 18 (5): 487-517.

POPE, D.G. and M. SCHWEITZER (2011): Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. In: The American Economic Review 101 (1): 129-157.

RAND, D.G.; DREBER, A.; ELLINGSEN, T.; FUDENBERG, D. and M.A. NOWAK (2009): Positive interactions promote public cooperation. In: Science 325: 1272-1275.

REIJNDERS, L. and M.A.J. HUIJBREGTS (2008): Palm oil and the emission of carbon-based greenhouse gases. In: Journal of Cleaner Production 16 (4): 477-482.

ROE, B.E. and D.R. JUST (2009): Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. In: American Journal of Agricultural Economics 91 (5): 1266-1271.

SEFTON, M.; SHUPP, R. and J.M. WALKER (2007): The effect of rewards and sanctions in provision of public goods. In: Economic Inquiry 45 (4): 671-690.

SEKHON, G. S. (1995): Fertilizer-N use efficiency and nitrate pollution of groundwater in developing countries. In: Journal of Contaminant Hydrology 20 (3): 167-184.

SUTTER, M.; HAIGNER, S. and M.G. KOCHER (2010): Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations. In: Review of Economic Studies 77 (4): 1540-1566.

TIEDEMANN, T. and U. LATACZ-LOHMANN (2013): Production Risk and Technical Efficiency in Organic and Conventional Agriculture - The Case of Arable Farms in Germany. In: Journal of Agricultural Economics 64 (1): 73-96.

UNCTAD (2013): Palm oil: www.unctad.info/en/Infocomm/AACP-Products/Palm-oil/. 21.October.2013.

VELDKAMP, E. and M. KELLER (1997): Nitrogen oxide emissions from a banana plantation in the humid tropics. In: Journal of Geophysical Research: Atmospheres (1984–2012) 102: 15889-15898.

VELDKAMP, E.; PURBOPUSPITO, J.; CORRE, M.D.; BRUMME, R. and D. MURDIYARSO (2008): Land use change effects on trace gas fluxes in the forest margins of Central Sulawesi, Indonesia. In: Journal of Geophysical Research: Biogeosciences 113(G2).

VOLPP, K.G.; JOHN, L.K.; TROXEL, A.B.; NORTON, L.; FASSBENDER, J. and G. LOEWENSTEIN (2008): Financial incentive–based approaches for weight loss. In: JAMA: the journal of the American Medical Association 300 (22): 2631-2637.

WILCOVE, D.S. and L.P. KOH (2010): Addressing the threats to biodiversity from oil-palm agriculture. In: Biodiversity and conservation 19 (4): 999-1007.