Economic experiments as a tool for agricultural policy evaluation: insights from the European CAP

Liesbeth Colen, Sergio Gomez-Y-Paloma, Uwe Latacz-Lohmann, Marianne Lefebvre, Raphaële Préget, Sophie Thoyer

Economic experiments as a tool for agricultural policy evaluation: Insights from the European CAP

Liesbeth Colen[a, f], Sergio Gomez-Y-Paloma[a], Uwe Latacz-Lohmann[b], Marianne Lefebvre[c], Raphaële Préget[d], Sophie Thoyer[e]

[a] European Commission, Joint Research Centre, Institute for Prospective Technological Studies, ES-41092, Seville, Spain

[b] Kiel University, Department of Agricultural Economics, D-24118 Kiel, Germany, and School of Agricultural and Resource Economics, University of Western Australia, Australia

[c] Université d'Angers, GRANEM UMR-MA n°49, F-49000 Angers, France

[d] INRA, UMR LAMETA 1135, F-34000 Montpellier, France

[e] Montpellier SupAgro, UMR LAMETA 1135, F-34000 Montpellier, France

[f] LICOS Centre for Institutions and Economic Performance, KU Leuven, BE-3000 Leuven, Belgium.

Abstract

This article assesses the potential contribution of economic experiments to evidence-based policy making in the field of agriculture, with a special focus on the EU's Common Agricultural Policy (CAP). CAP evaluation mostly relies on standard evaluation tools such as farm and market simulation models, calibrated with EU-wide statistical data, statistical and econometric analysis of survey data, and a range of qualitative methods such as interviews with stakeholders, focus group or internet-based public consultation. Yet, the CAP has changed considerably over the past decades, requiring adaptations of its evaluation toolbox. A detailed review of existing studies using economic experiments for designing and evaluating agricultural policies provides the backbone for a comprehensive assessment of the complementarity of experimental approaches with standard evaluation tools. The article also provides recommendations aiming at facilitating inclusion of economic experiments into the CAP evaluation toolbox based on conclusions drawn from a workshop organized with experts, academics and policy makers of Directorate-General for Agriculture and Rural Development of the European Commission.

JEL: Q1, C9

1. Introduction

Evaluation of public policies and intervention programs has received increasing attention over the past years. Most non-governmental organizations (NGOs) and public bodies, including the European Commission (EC), commit themselves to critically evaluate whether their activities have the desired impact on society, whether spending is justified based on the performance of the interventions, and how planned actions and policies can be adjusted to improve outcomes (EC 2015). Ex-ante evaluation provides guidance on the choice of most cost-effective policies among different options. Ex-post evaluations help policy makers to measure the net impact of a policy and establish the reasons for success or failure.

In parallel, methodologies for policy evaluation have made several major advances in the past decades. First, greater attention is paid to identifying cause-effect relationships of policies, by building an experimental situation or by identifying a quasi-experimental situation for which outcomes can be compared to a proper counterfactual (Imbens and Wooldridge 2009). Second, behavioral studies, often making use of economic experiments, have highlighted the need to account for elements of the decision context beyond the simple profit maximization assumption, in order to predict economic agents' responses to different policy instruments and to design efficient and cost-effective policies (Shafir 2012).

Economic experiments are at the forefront of these recent methodological developments. Their main common feature is to rely on data that are generated in a controlled setting, with a randomized assignment of participants to treatment and control groups. In contrast with "naturally-occurring" observational data, experimental data are obtained through a rigorous experimental protocol, which allows for a clearer identification of impact and causality, and which are replicable. To ensure that participants in the experiment reveal their true preferences, they often receive financial incentives related to what their economic gains could have been in an equivalent real-life setting. Several types of economic experiments exist. They vary according to: experimental subjects (from students to policy stakeholders); the environment in which the experiment takes place (laboratory, lab-in-the-field, or the subjects' natural context); and the type of experimental setting (discrete choice experiments, experimental games with or without financial incentives, or the real-world implementation of the policy to be tested through randomized control trials).

Despite the rapidly growing number of academic publications on economic experiments, and the success of experimental approaches for policy advice in various domains of public economics, notably health (Holt, Smith, and Shobe 2006), the mobilization of such approaches in agricultural policy design and evaluation has remained surprisingly limited up to now. Evaluations of the EU's Common Agricultural Policy (CAP) are a telling example. The CAP toolbox for ex-ante policy assessment and ex-post evaluation studies includes mostly farm and market simulation models, calibrated with EU-wide statistical data (e.g. FADN or FSS), statistical and econometric analysis of survey data, and a range of qualitative methods such as interviews with stakeholders, focus group or internet-based public consultation (EC 2015).

In this paper, we investigate the potential contribution and main challenges of integrating economic experiments in the toolbox for agricultural policy evaluation. We use the EU's CAP for illustration.

Section 2 analyzes whether there is space for economic experiments in the agricultural policy evaluation toolbox, by comparing the main tools used by the European Commission (EC) with the new evaluation needs arising from successive CAP reforms. Section 3 provides an extensive literature review, which aims at identifying the type of evaluation issues in the field of agricultural policy for which economic experiments can provide suitable responses. Section 4 examines the extent to which economic experiments can make exclusive contributions for policy evaluation. We discuss the advantages and limitations of experimental and non-experimental methods and how they can complement each other in agricultural policy evaluation. In the final section of the paper, we draw conclusions and formulate recommendations, both for researchers in agricultural economics, willing to have their research relevant for policy makers, and for agricultural policy makers, looking for policy advice based on sound evidence and the use of the best available methods.

2. Is there room for economic experiments in the agricultural policy evaluation toolbox?

To the best of the authors' knowledge, economic experiments have been used only marginally by the EC or other EU institutions to conduct evaluations of CAP impacts. Yet several Directorates-General of the European Commission have already commissioned specific experimental studies to analyze policies in the fields of consumption, health, employment or the environment (Ciriolo 2011). Are these recent developments relevant for agricultural policy evaluations? Is there room for economic experiments, especially in the context of recent CAP reforms?

The current CAP evaluation toolbox

CAP evaluation is multifaceted: depending on the type of policy and the phase in the policy cycle, different methods are implemented. Before a new set of policies is proposed by the European Commission, an extensive impact assessment evaluates alternative scenarios for CAP changes (EC 2011). After implementation by Member States, at the end of the programming period, ex-post evaluation projects are set up to examine the effectiveness, efficiency and relevance of CAP measures.

Ex-ante policy assessments are mostly conducted in-house, by the appropriate EC services. These make extensive use of quantitative analysis, complemented with desk research and public consultations. Quantitative analysis consists of policy simulations using partial equilibrium models, such as CAPRI and AGLINK-COSIMO (Burrell and Nii-naate 2013; Witzke and Gocht n.d.), which are specifically designed to simulate the economic, environmental and social impacts, mostly at market and regional levels, of alternative agricultural policy scenarios in comparison to the status quo situation. Simulation models rely on simplified behavioral assumptions (such as maximization of an expected utility based on profit) and exogenous parameters, and on the quality of data available for calibration. Although sensitivity analyzes are used to test the robustness of the results in the presence of uncertainty regarding these assumptions and parameters (EC 2011), the current

emphasis is not on further sophisticating the models to include more realistic assumptions concerning farmers' behavior.

Ex-post evaluations usually combine desk research and qualitative methods such as focus groups and stakeholder interviews, with quantitative statistical analysis of existing farm survey or administrative data. Case studies are often mobilized, selecting a list of farm types and regions that are considered most relevant for studying that specific policy and comparing results across case studies.

The European Commission (EC) has developed a strong expertise in collecting market and farm-level data, such as the Farm Structure Survey or the Farm Accountancy Data Network (FADN). Derived from national surveys, the FADN is the only source of microeconomic agricultural data based on harmonized bookkeeping principles. Currently, the annual sample covers approximately 80.000 holdings across the EU. Given the high detail of data and the representativeness of the sample[1], FADN data are used by evaluators and researchers for a variety of purposes. For example, the recent evaluation of investment aid provides an example of how these data, in combination with sophisticated econometric techniques, can provide quantitative estimates of the net impact of investment aid under the CAP's Rural Development Policy (EC 2014). Another recent study used FADN data to evaluate the structural effects of the single payment scheme, in particular on farm labor force and capital intensity, although the causal link could not be clearly established (EC 2013).

New evaluation tools for new policy instruments?

The CAP has experienced several reforms over the last three decades. In order to meet its evolving objectives, the type of policy interventions have changed drastically. As a result, the toolbox for evaluating these policies also needs to be adapted to respond better to the evaluation needs of these policy changes.

First, the switch to decoupled payments, targeting farmers rather than commodities has called for a change of the unit of evaluation: the farm rather than the market. Simulation models such as CAPRI were initially designed to estimate the supply responses and market impacts to changes in guaranteed prices and other market interventions, and are therefore less suited to analyze impacts at the farm level. In response to this concern, new modelling tools focusing on individual farm responses rather than on aggregate supply and demand are currently being developed by the European Commission. These will be able to capture the farm-specific implementation of policies, as well as the heterogeneity of impacts across farmers (Louhichi et al. 2015).

Second, an increasing number of CAP measures are based on farmers' voluntary enrolment, such as agri-environmental contracts in the Rural Development Policy. To evaluate the efficiency of such measures, one needs to pay more attention to the drivers of individual farm decisions and on farmers' motivations to participate in such schemes. Ex-ante evaluation should be able to provide information regarding the expected uptake of a voluntary measure, what types of farmers are expected to subscribe, what incentives are needed to encourage uptake and the risks of non-

[1] They represent a population of about 5.000.000 farms in the EU, which covers approximately 90% of the total utilized agricultural area (UAA) and account for about 90% of the total agricultural production.

compliance. This requires in turn that behavioral drivers of economic decisions be better anticipated and included in economic models. This type of ex-ante information can help to fine-tune policies so as to obtain the desired level of participation, or to optimize the environmental outcome for a given budget.

Third, the EC is committed to evaluate in a proportionate way all EU policies intended to have an impact on society or the economy, including policies with a dedicated budget (spending activities) and regulatory policies (non-spending activities) (EC 2015). Some aspects of the agricultural policy are strictly regulatory, but were not systematically evaluated in the past. Note that there is much less data available in the regulatory domain, compared to the detailed financial data related to spending activities. Data generation methods such as economic experiments could therefore have potential here. Moreover, when farmers receive no payments in return for complying with a set of rules (for example with regard to the impact of their activities on the environment), the traditional profit maximizing assumption has limited explanatory power to understand farmers' reaction to the new regulation, and to predict the compliance rate. One should account for the diversity in farmers' objectives, such as the desire to comply with the rule or to behave like the group (social norms).

Last but not least, new CAP regulations for the 2014-2020 programming period have granted more discretion to Member States for implementing both first and second-pillar payments. As a result, general EU-wide evaluation approaches will increasingly need to be replaced or complemented by more targeted evaluations, focusing on measures implemented in specific Member States or regions, and accounting for the selection criteria that are set up at regional level to target specific groups of farmers or policy outcomes (for example, Less Favored Area payments are restricted to specific farm types in specific regions). This requires better and more rapid policy-testing. Case-study approaches have already been mobilized for the evaluation of such local implementations of the Rural Development Policy. For example, Häring et al. (2004) evaluate the impact of CAP measures designed to encourage conversion to organic farming, using a case study approach in six Member States. This work would benefit from an experimental set-up allowing rapid replication in different contexts, but at the same time easy adaptation to the specificities of the policy in a particular location.

3. Policy relevant applications of experiments in the agricultural economics literature

There is a growing body of academic literature applying economic experiments to understand farmers' decisions and assess agricultural and agri-environmental policies. This section provides an overview of studies that illustrate the potential usefulness of economic experiments for informing agricultural policy – without laying claim to be complete. Beyond the experimental results of these studies, we emphasize their suitability and potential contribution to policy evaluation. This review also allows identifying the types of questions in the field of agricultural policy that can best be addressed through economic experiments.

Discrete Choice Experiments

Discrete choice experiments (DCE) assess people's preferences or decisions in hypothetical situations, e.g. before a new product is launched, a new technology becomes available, or a new policy is implemented. DCE are designed such that it is possible for the respondent to envisage the different situations in the real world even when they do not yet exist. In a DCE, respondents are presented with a series of so-called choice sets. Each choice set contains a discrete number of choice options, and respondents are asked to indicate their most preferred option. Each option is characterized by a number of characteristics, or attributes, one of which is a price (to be paid to get access to the option) or a payment (received to accept the option) while the others are non-monetary attributes. Hence, when making their choices, respondents have to trade off a lower monetary benefit (or higher cost) for more preferred non-monetary characteristics. Econometric analysis of the choice data allows quantifying the effect of the monetary and non-monetary attributes on choices and estimating in monetary terms the value of each attribute for the participants in the form of their willingness-to-pay (WTP) or willingness-to-accept (WTA) for changes in the level of each attribute (Louviere, Hensher, and Swait 2000). DCE can take place in the lab, online, or directly in the field. DCE is usually a stated-preference method[2]: respondents are asked to choose amongst hypothetical scenarios and do not receive incentive payments for their responses.

Most DCE with farmers have focused on eliciting farmers' willingness to participate in agri-environmental schemes (AES). Examples include schemes encouraging the cultivation of nitrogen-fixing crops in Spain (Espinosa-Goded et al. 2010), establishing pesticide-free buffer zones in Denmark (Christensen et al. 2011), protecting ground-nesting birds in Germany (Breustedt et al. 2013 a,b), promoting traditional farming methods in Hungary (Birol et al. 2006), encouraging the adoption of bio-diversity contracts among pastoralists in Australia (Greiner 2015) and afforestation contracts in Denmark (Broch and Vedel 2012). Typically, farmers are asked to choose among alternative contracts, each with different contract specifications and payment levels.

These studies reveal estimates of farmers' willingness to accept to participate in such schemes, i.e. the minimum compensation required in exchange of participation. They thus provide an estimate of the cost of compliance with such agri-environmental contracts. More importantly, the DCE allow to decompose the total cost of compliance into specific part-costs for each attribute of the schemes, and to reveal the heterogeneity in farmers' preferences and WTA. As such, they provide policy makers with information as to which characteristics of the contract are considered most problematic or most desirable from the farmers' point of view, and they provide information on which types of farmers are most likely to participate, allowing to better fine tune and target policies. Although the precise results are obviously specific to the scheme and location studied, the general conclusion from these studies is that introducing some flexibility in the contract considerably increases farmers' willingness to participate, as does shorter or flexible contract duration. Christensen et al. (2011) conclude that the overall flexibility of the contract might even be more important than the type of management prescriptions imposed.

[2] Real choice experiments, where discrete choice questions are combined with real economic incentives, are developing fast (e.g. Alfnes et al. 2006; Lusk and Schroeder 2004) to analyze consumer preferences and willingness to pay for new products. However, they are more difficult to set-up when the choice under study is a policy option rather than a consumption good.

In addition to assessing farmers' willingness to participate in AES, DCE can also reveal information as to how much land farmers would be willing to enroll in a conservation schemes in response to changes in the contractual terms (Breustedt et al. 2013 b). Other studies have tested the introduction of a collective dimension in AES, for example by including a bonus conditional upon reaching a common environmental target (Kuhfuss et al. 2014). They find that such a bonus can improve participation rates and the amount of land enrolled. Based on this type of information, the design of AES can be optimized in order to increase participation rates and environmental outcomes, and schemes can be targeted on those types of farmers who are willing to participate at low payment rates.

Overall, DCE seem to be particularly useful for optimizing the design of voluntary agri-environmental schemes. If the policy designer knows beforehand which contractual obligations farmers consider to be particularly demanding, this information can be taken into account in designing the contracts and calibrating payment rates. In this way, the informational imbalance between the policy designer and the farmer, which often impedes the design of efficient conservation contracts, can be mitigated. An outstanding feature of DCE in this respect is that they provide estimates of farmers' participation costs differentiated by farm type, location or resource settings, thereby allowing the policy maker to devise contracts for different types of farmers and different resource settings. This is likely to boost the cost-effectiveness of such schemes compared to the standard one-size-fits-all contract.

A few studies have applied the DCE method to the evaluation of CAP first-pillar policies. Schulz et al. (2014) explore farmers' responses to the new 'greening' provisions of the CAP direct payments in Germany, before the policy entered into force. 'Greening' requires farmers to respect certain environmental provisions in order to obtain the full direct payment (EU 2013). The authors find that farmers perceive 'greening' as a costly constraint, but not all farmers are equally affected and not all 'greening' provisions are regarded as equally demanding: the requirement to provide Ecological Focus Area on 5 percent of a farm's arable land is considered much more demanding (and thus costly) than the requirement to cultivate at least three crops. Using a DCE, Lips and Gazzarin (2008) provide an ex-ante analysis of farmers' reactions to the planned abolishment of the milk quota in Switzerland. Their study reveals a strong tendency to stay in dairy production. There is only limited willingness of farmers to move on to other types of farming or to quit farming altogether. Paulrud and Laitila (2010) assess farmers' willingness to grow energy crops in Sweden in response to different amounts of coupled subsidies. They analyze what crops and how much of these crops farmers would cultivate at different subsidy levels.

Lab and field experiments

Lab experiments take place in a laboratory, often a classroom or computer room, where participants, often students, are invited to take part in an experiment which aims to test economic behavior in response to different (policy) interventions. Participants are financially incentivized to mimic real-world incentives: they can win money according to their decisions in the experiment and are paid in accordance with their choices or performance at the end of the experiment. In most lab experiments whose results are published in the academic literature, participants are confronted with choices that are formulated in an abstract way. Such a decontextualized setting allows the

experimenter to ensure that the results of the experiment are not influenced by the way the game is framed.

While lab experiments allow for perfect control of the experimental environment, the artificial context of lab experiments may result in behavior that is poorly correlated with naturally occurring behavior (Harrison and List 2004). When the context is believed to impact on participants' decision making, field experiments may be called for. A field experiment is typically conducted with 'real' stakeholders rather than students. If in addition to recruiting participants from the field, the experiment is designed to represent as closely as possible the field context and the policy to be assessed, this is referred to as a framed field experiment. In such experiments, the commodity used, the terminology, the task to be executed or information available to the participants, correspond to the real environment in which the task or policy takes place; but participants know they are taking part to an experiment.

Several studies have used lab and field experiments to analyze the best design for specific agricultural policies. Bahrs et al. (2008) investigate different trading mechanisms for EU decoupled payment entitlements introduced with the Fischler Reform of the CAP in 2005. Since, at the time, policies had changed only recently and the trading had just started, real-world market evidence was spurious and anecdotal. Thus, controlled laboratory experiments with students were conducted to provide first insights into the impact of different trading rules on potential market outcomes. Nagler et al. (2013) conducted a lab experiment with both students and agricultural professionals to estimate the capitalization of subsidies into land rental prices. Maart-Noelck et al. (2013) study the impact of a price floor on investment behavior. Price floors such as intervention prices for agricultural commodities or guaranteed feed-in tariffs for renewable energies are often claimed to have a stimulating effect on investment behavior. However, the key finding from the experiment with students is that investment behavior did not differ significantly with respect to the presence of a price floor. In the Unites States context, McIntosh et al. (2007) investigate supply responses to countercyclical payments given to farmers in a world of price uncertainty as part of the US Farm Bill. The experimental evidence suggests that such payments lead to greater income certainty but less efficient production decisions and (possibly) higher government payments.

Lab and field experiments have also been used extensively to inform the design of conservation auctions. Such auctions allocate conservation contracts to private landholders on the basis of competitive bidding. Bids take the form of requests for financial compensation in return for implementing the contractual obligations on one's land. Cason et al. (2003) investigated the Bush Tender trials, designed to conserve the last remaining patches of native vegetation in Victoria, Australia. Farmers were invited to submit financial bids for agreeing to conserve the land in pristine state. This study was the first to propose a combination of lab and field experiments: prior to testing the program in the field, certain design problems, in particular the amount and choice of the information to be communicated to landholders before the bidding session, were investigated experimentally in the lab with students. In a second stage, the experiment was moved to the field, with farmers as participants. Finally, the results of the experiments were used to fine-tune the design of the actual field auction. Other experiments have analyzed alternative auction formats for reducing non-point source pollution resulting from fertilizer use in Australia (Cason and Gangadharan 2005) or auctions to buy back water abstraction rights from farmers in the state of Georgia in years of drought (Cummings et al. 2004). Another key aspect of conservation contract

design that has been studied experimentally is the so-called agglomeration bonus: by rewarding conservation land use on adjacent parcels, the agglomeration bonus provides economic incentives for the creation of non-fragmented land use patterns in the landscape. Parkhurst et al. (2002) tested this experimentally (with students) and find that including an agglomeration bonus significantly decreased the fragmentation of the conserved land, highlighting its effectiveness as a policy tool.

Randomized Controlled Trials

A randomized controlled trial (RCT) is an experiment which takes place in the environment where the participants naturally make their decisions. In most cases, but not all, they are not aware that they are taking part in an experiment. The experimental design consists of the random assignment of participants to a treatment group or a control group, where the treatment consists of the implementation of a policy or program. Hence, participants are not asked to act as if they faced a certain policy, nor are they asked to play a game with financial incentives. Instead, an RCT actually implements the policy among a group of randomly selected stakeholders. It then observes the actual behavior of the participants under this policy and compares it to the group of comparable participants that did not receive the policy treatment.

Random allocation means that all participants have the same chance of being assigned to either the treatment or the control group. The purpose of random allocation of participants is to assure that, before the start of the intervention, participants in both groups have on average the same characteristics. As a result, the only expected difference between the control and experimental groups is the result of the treatment (e.g. policy intervention) being studied. Thus, any significant differences between groups in the outcome can be attributed to the policy intervention and not to some other unidentified factor (Duflo et al. 2007). Random assignment provides the best counterfactual describing what would have happened to the treatment group participants if they had not been exposed to the treatment (Cook 2003).

The RCT methodology has been developed and progressed in clinical research. Its diffusion into the social sciences and, in particular, economics is only recent, and mainly in the fields of labor and development economics. To the best of the authors' knowledge, up to now there exist no applications to EU agricultural policy programs. Yet, RCT have been applied to a variety of agricultural policy interventions in developing countries, including programs on farmers' training in Armenia (Blair et al. 2013), alternative policy interventions to incentivize Kenyan farmers to use fertilizer (Duflo et al. 2011), policies to improve the pay-back of agricultural loans (Giné et al. Yang 2012), the effect of improved seed varieties on farmer's effort and yields (Bulte et al. 2014), the impact of rainfall index insurance schemes on technology decisions and investment (Karlan et al. 2014; Mobarak and Rosenzweig 2013) and the provision of price information through mobile phones on farmers' marketing outcomes (Fafchamps and Minten 2012). Most of these interventions target specific problems of the agricultural sector at a local scale. Therefore, their evaluation encompasses different stakes and constraints than the evaluation of agricultural policies in the EU. Most of them are implemented by donor organizations as evaluation of pilot projects. But a number of more recent examples of RCT also involve evaluation of governmental programs in developing countries (Banerjee et al. 2014; Chong et al. 2010; Dizon-Ross et al. 2015) or scaling-up of pilot programs at the state level (Bouguen et al. 2014), although not in the field of agriculture.

While most RCT reviewed here are evaluations of programs supporting agriculture in the developing world, these applications do illustrate the potential of RCT to inform agricultural policies also in countries with modern agriculture and a well-developed agricultural policy.

This literature review illustrates the diversity of uses of economic experiments for agricultural policy evaluation: (i) to test different policies prior to implementation in order to compare their effectiveness and/or efficiency. Experiments are helpful in identifying potential design flaws in the policy, especially when there is no empirical evidence or when theoretical predictions are difficult if not impossible to derive; (ii) to generate fined-tuned data enabling measuring the net impact of a policy and to disentangle it from confounding factors, which is not always possible with observational data; (iii) to elicit farmers' preferences and to understand their reactions to policy in the presence of behavioral factors (risk and loss aversion, social norms, intrinsic motivations, time inconsistencies …), usually not accounted for by other evaluation methods.

4. The complementarities between experimental and non-experimental approaches

Good practices in policy evaluation include making use of the complementarities between different evaluation tools and triangulating the results drawn from different methods[3]. As stated by Harrison (2014), "There are good, scientific reasons for the popularity of experiments, but the challenge is to keep a balanced eye on what [field] experiments can do that improves on, or complements, other methods […]". To shed light on the complementarity of experiments with other methodologies currently in use by policy evaluators, the relative advantages and limitations of different evaluation methods are summarized in Table 1. The left-hand part of the table summarizes the non-experimental approaches: "qualitative methods" are mostly based on case studies and qualitative comparisons of before-after policy implementation situations. In "simulation models", we include partial/general equilibrium models, and mathematical programming tools mostly used for ex-ante evaluations. "Econometric analysis" refers mostly to ex-post evaluations and include quasi experimental approaches. Quasi experiments are sophisticated econometric methods which have been developed to identify the causal effect of a policy from observational data (EC, 2013b; Loi and Rodrigues, 2012) by artificially 'constructing' or mimicking the counterfactual. They include the following empirical strategies: instrumental variables estimations, regression discontinuity designs, difference-in-difference matching and propensity score matching.

The comparison is based on three criteria: the objective of the evaluation; the trade-off between causality and generalizability (i.e. internal and external validity); and the practical challenges in data collection and analysis. The performance indicators (low, medium, high) in Table 1 are just indications of relative performance across methodologies and criteria. They are debatable since each column refers to very heterogeneous evaluation studies. Overall, the assessment of the relative

[3] Also the evaluation policy of the European Commission stresses the role of using diverse methodologies and source (triangulation). As stipulated in the new "Better Regulation Guidelines", "Evaluations are based on the best available evidence (…), which should be drawn from a diverse and appropriate range of methods and sources (triangulation)" (EC 2015).

global performance of the different methods depends on the weight attributed to each of the selection criterion.

Table 1: Comparison of different methods according to the objective of the agricultural policy evaluation and the implementation constraints

	Empirical approaches using observational data			Economic experiments			
	Qualitative methods	Simulation models	Econometric analysis	RCT	Field experiments	Lab experiments	DCE
Objective of the evaluation	Testing a policy prior to implementation (ex-ante)						
	Low	High	N/A	High (pilot)	High	High	High
	Measuring the net impact of a specific policy (ex-post)						
	Medium	N/A	High (using quasi-experiments)	High	Low	Low	Low
	Understanding farmers' reactions to policy in the presence of behavioral factors						
	Medium	N/A	Low	Low	High	Medium	High
Trade-off between internal and external validity	Ability to identify causality						
	Low	N/A (causal effects are assumed)	Medium	High	Medium	High	Medium
	Ability to transfer the results to the real world						
	High (limited by accuracy of stated preference)	High	High	High	Medium (depending on realism of the task)	Low	High (limited by accuracy of stated preference)
	Ability to extrapolate results to other contexts/people						
	Low	High (depending on model coverage)	Medium (depending on data source)	Medium	Medium	Low (but easy to replicate)	Medium
Practical challenges	Ease of access to data or data collection						
	Medium	High (FADN, national surveys)	Medium	Low	Medium	High	Medium
	Ease of results' interpretation by policy officers						
	Medium (subjectivity of qualitative data)	Low	Low	High	High	High	High

Source: Inspired by Roe & Just (2009, fig.1 p.1268), extended and adapted for agricultural policy evaluation. ; N/A: not applicable

Objective of the evaluation

Depending on the main objective of the evaluation task, different methods may be appropriate or may need to be combined. We identify the ability of the different methods (1) to test a policy prior to implementation (ex-ante evaluation); (2) to measure the net impact of a policy (ex-post evaluation) and (3) to understand farmers' reactions to policy in the presence of behavioral factors.

For *ex-ante* evaluations, experimental data and simulation models are highly complementary. Simulation models are clearly more useful when assessing broad policy reforms, such as reducing price support or abolishing production quotas at the EU-level. Yet, when it comes to new policies that are very different from existing ones (e.g. launch of a conservation auction, introduction of novel insurance tools), it may be difficult for simulation models to make realistic assumptions as to the impacts of those policy changes on farmer behavior. Nevertheless, economic experiments may provide some insights on what response to expect. For example, the DCE related to the introduction of the greening provisions of the CAP (Schulz et al. 2014), informs policy makers of farmers' likely responses to this new element of the CAP. When it comes to targeted policy interventions (on specific farm types or in a limited number of regions), simulation models are not the most appropriate method given the cost of making the model flexible enough to capture policy specificities at regional or individual level and because of the lack of data to feed such a model. Qualitative assessment based on stakeholders' interviews in a number of carefully selected case studies, can provide useful insights there, but economic experiments can offer more robust evidence. For example, DCE can help assess how farmers would respond in aggregate to different designs, but also how responses might differ among different types of farmers. RCT can provide a powerful tool for ex ante analysis in the form of pilot programs: provided proper randomization has been set in place, the causal impact of a specific policy program can be reliably assessed before scaling it up to the entire population. On the contrary, RCT are not suitable to test broad policy reforms (Goldin et al. 2012; Rodrik 2008). For example, in the case of market price interventions it is simply impossible to exclude a (random) part of the population from the policy. RCT can therefore be useful to test specific policy programs in a specific context, but cannot offer an evaluation of broad or wide-ranging policy reforms.[4] Also, when behavioral drivers of farmers' responses to policy are important, simulation models may need some input from experimental data. Field experiments can help elicit farmers' preferences and behavioral parameters (e.g. risk aversion, time preferences), which can then be plugged into simulation models. With respect to policy measures with voluntary enrolment, the review of the literature above clearly illustrated the usefulness of all types of economic experiments in an ex-ante evaluation perspective.

When considering the *ex-post* evaluation of policy impacts, statistical analysis of observational data and RCT can provide reliable estimates of the impact of a policy or program. Yet, they usually provide limited information on the reasons underlying the outcomes. When it comes to *understanding* why a policy did not work as expected and how to improve it, qualitative methods based on stakeholders' perceptions, and discrete choice, lab or field experiments are often more

[4] Opponents of experimentation usually claim that biased answers to big explanatory questions are more important than unbiased answers to smaller casual-descriptive questions (Plott 1989).

useful. Indeed, understanding the impact of a policy requires analyzing both the results of the decisions taken by farmers in response to the policy (as can be observed from micro-level data such as FADN), but also to understand the factors underlying such decisions and behaviors. Experiments can help explain unanticipated effects of the policy or intervention (Gneezy and Rustichini 2000). For example, DCE allow analyzing in detail the mechanism of farmers' response to a program, to explain low participation rates in schemes that have failed to attract the envisaged numbers of farmers despite high financial incentives. Hence, choice data will enable the analyst not only to identify the obstacles to participation but also to devise effective remedies.

Causality versus generalizability

The second key issue in the selection of evaluation methods is the trade-off between internal validity and external validity of the evaluation results. Internal validity or causality reflects the extent to which the causal relation between two variables (for example the policy and the outcome) is properly demonstrated. External validity or generalizability refers both to the ability to transfer the results to the real world and to other contexts. While both internal and external validity are desired, the evaluator is usually confronted with a trade-off (Roe and Just 2009).

Qualitative methods have a low internal validity, since they generally do not allow the analyst to conclude decisively whether the observed changes are due to the policy. Their external validity can be high, if the relevant key stakeholders are interviewed. Nevertheless, stakeholders' stated preferences and future intentions (before a policy is implemented) or perceptions of what has happened (ex-post) can be biased and misreported, voluntarily (strategic bias) or not. Moreover, qualitative methods are often applied in a case study setting, and the transferability of findings to a different context is dependent on the specificities of the case studies.

As long as representative and carefully collected data are used, statistical analysis of observational data presents high external validity. It may be difficult, however, to establish internal validity when the analyst suspects that unobserved variables may have affected outcomes. Michalek, Ciaian, and Kancs (2015) provide an example of how sophisticated, quasi-experimental econometric techniques can be used to estimate the net impact of investment support under the CAP's rural development policy based on FADN data, but such techniques require detailed databases and cannot always provide a convincing solution. On the contrary, random assignment of the participants to the different treatments and high control of the decision-making environment, as used in most economic experiments, allow the researcher to prevent systematic differences in treatment and control groups and to limit any concurrent third elements that could confound the outcome. However, this strong internal validity comes at the cost of reduced external validity, especially for lab experiments in an abstract setting. Indeed, the choices that individuals make in the lab depend not just on financial implications, but also on the particular context in which a decision is embedded, and the manner in which participants are selected to participate. Because the lab systematically differs from most naturally occurring environments on these dimensions, experiments may not always yield results that are readily generalizable (Levitt and List 2007).

Field experiments, RCT and DCE present a compromise. They lessen the inherent tension between establishing causality and the generalizability of results. Field experiments give in some control

compared to lab experiments, when the experimental design does not allow controlling for the numerous factors that are at work in the field; but results are likely to be more representative of real world decision making. Moreover, even when field experiments are usually conducted on limited samples, they constitute a structured evaluation tool, which allows replication in different contexts or with other participants to test the generalizability of the results. This facilitates the comparison of results across different farm types and/or Member States. As such, field experiments provide a bridge between lab and observational data (List 2007, 2011). A combination of lab experiments with small-scale field experiments can ensure both the correctness and relevance of the results for policy-making.

Since DCE rely on stated instead of revealed preferences, they may suffer from a lower internal validity than lab experiments. However, because respondents are typically drawn from the targeted population and scenarios proposed are usually as close as possible to what could be observed in reality, external validity may be higher than in an artificial laboratory setting with students. The ability to extrapolate discrete choice experiments' results to other contexts can nevertheless be limited if the survey (e.g. the set and level of attributes) is designed to fit a specific context.

Finally, the RCT methodology combines high internal validity (thanks to randomization) and high transferability of results to the real world since data come from the natural environment and stakes are real, in contrast to lab and field experiments which are merely games. RCT are therefore often referred to as the "gold standard" to measure the net impact of a specific program. However, external validity in the sense of transferability of results to a different context or country, or even to a larger scale, may be limited, especially when the policy response is very much dependent on specific local factors or when general equilibrium effects play a role.

Practical challenges in data collection and analysis

The choice of an evaluation method is also often affected by practical issues, due to time and budget constraints. Any method requiring the collection of new data will generate additional budget costs and delays. Since the common feature of all economic experiments is that data must be generated under a controlled process, it is expected that they engender greater hassle and data-related costs than simulation models and econometric analysis exploiting existing EU-wide databases such as FADN or Eurostat data.

The time, difficulty and costs involved in collecting observational data or generating experimental data vary greatly. Collection of qualitative data from stakeholders in a case study, or experimental data collected in a discrete choice, lab or field experiment, is usually faster since it involves fairly small sample sizes. For randomized controlled trials, the establishment of the baseline situation and the preparation of the experimental set-up is often a costly and time-consuming precondition for effective experimentation, especially since it may require specific authorization by member states or by European authorities.

Practical challenges in data collection also arise from the requirement to constitute representative samples and, in the case of economic experiments, to control for assignment and strategic biases. The first concern of any evaluator is to make sure that the group under study is representative of the

population of interest in order to ensure that results can be generalized to the entire population. The sample selection problem exists in all evaluation methods requiring data but it can be more acute in some cases. For laboratory experiments conducted with students, the risk of selection bias is fairly well controlled thanks to the generally large number of participants and to the sophisticated and normalized recruitment procedures followed by most experimental laboratories. However, this issue is more problematic for field and DCE, especially when conducted with farmers. Randomization of experiments conducted with farmers requires that an up-to-date list of farmers in the region of interest be made available to the experimenter who can pick participants at random in the list provided. This is rarely the case since national administrative and statistical services are reluctant to supply this information for privacy protection reasons and authorization procedures are often slow. Experimenters therefore tend to resort to other recruitment procedures, for example through calls for participation in farm magazines or newsletters, or through a network of farm advisors. These recruitment methods do not preclude the risk of self-selection. Even if efforts are made to invite farmers to participate on a random basis, one can suspect that their propensity to accept to be enrolled in an experiment is correlated to variables such as their familiarity with administrative or technical staff, their interest for policy-making, their time availability etc. – factors that make them statistically different from the population of interest. With RCT, the cases of self-selection bias are less frequent since participation is usually imposed by the experimental design.

Overall, achieving a sample that is representative of all EU farmers in the 28 Member States with the use of experimental methodology is generally not feasible within existing budget constraints. Moreover, comparison of results from experiments replicated in several regions or Member States is not straightforward if the selection bias is of different nature in the different contexts.

Economic experiments are particularly prone to the risk of assignment bias (Berriet-Solliec, Labarthe, and Laurent 2014). Since most experiments require setting up sub-groups receiving different treatments, particular attention must be paid to the procedure assigning participants to treatment and control groups. To ensure that the average effect of the program or policy measure under study is properly measured, it is necessary that the treated group and the control group have the same observed and unobserved characteristics before the treatment. Assume for example an RCT conducted to evaluate farm advisory service policies. Farmers are selected to benefit from a tailored individualized technical support program. The objective is to measure whether such policy can improve the take-up of innovations at the farm level. However, if selected farmers are systematically more skilled or are more prone to take risks than farmers of the control group, the probability that they adopt innovations is greater even in the absence of the individualized technical support program under study.

Randomization at the assignment stage is a way to guarantee that the assignment of each participant either to treatment groups or to control groups is not correlated with any of his/her characteristics that could affect the outcomes of the experiment. Often, however, the assignment bias is introduced (consciously or not) by the investigator himself. The ideal procedure to ensure correct randomization is to use assignment concealment techniques: the investigator in charge of contacting potential participants is not aware of the procedure allocating participants to one group or another and does not take part in the assignment decision procedure (Jaddad and Enkin 2007).

Evaluation bias is a specific drawback of economic experiments, which is not observed in non-experimental approaches. It occurs when participants change behavior because they are aware that their decisions are recorded. They may thus anticipate that their responses or observed behavior will have an impact on future policy choices and may want to manipulate the outcomes of the experiment by adopting an insincere behavior. Another example is the 'Hawthorne effect', when the treatment group works harder than normal or the 'John Henry effect' when the control group starts competing with the treatment group. There may also be a 'warm glow bias' when the participants are informed of the purpose of the study and change their behavior in response to their perception of what the evaluator is trying to test. Each of these biases can affect the reliability of results and may limit the possibility to compare results across studies because such biases are often group or context-dependent. Controlling for these biases requires the experiment to be carefully designed and, if possible, participants are not made aware of their participation in the experiment. While this is possible for RCT, it is more difficult for discrete choice, lab and field experiments. Increasing incentives or making the experimental tasks/surveys less prone to manipulation by participants are also ways of mitigating biases. At the same time, this will increase complexity and raise costs.

The ease of interpretation of results can also make some evaluation methods more attractive than others. Experimental results are often easier to interpret than the results from statistical methods or complex simulation models. As indicated by Burtless (1995), "the simplicity of experiments offers notable advantages in making results convincing […] and understandable to policy makers." Qualitative data are also easy to interpret since they require limited technical knowledge, although drawing a conclusion from numerous interviews with a diverse pool of stakeholders does not always allow a straightforward interpretation.

Complementarities across approaches

In what follows we illustrate the complementarity between methods relying on observational data and economic experiments with two examples.

Assessment of the new greening provisions of the CAP

In the 2014-2020 CAP, a so-called greening premium is paid to farmers respecting three mandatory agricultural practices, namely maintenance of permanent grassland, presence of ecological focus areas and crop diversification. Before the final vote of the CAP reform was made, the ex-ante assessment question was: how will EU farmers respond to the greening policy? We illustrate how a simulation model and DCE provide complementary insights into this question.

Louhichi et al. (2015) developed an EU-wide Individual Farm Model for CAP Analysis (IFM-CAP) to simulate the effects of the crop diversification requirement on land use and farmers' income. Based on FADN data, the decision making of individual farms in response to this policy change is modelled. Yet, as any simulation model, its results depend on specific assumptions and on the data used for calibration. For example, farmers are assumed to maximize income and to be risk neutral, land is assumed to be reallocated only within and not between farms, and several assumptions on supply elasticities, input cost etc. had to be made.

Using a DCE, Schulz et al. (2014) assessed farmers' willingness to comply with the greening requirements for a sample of 128 German farmers. Their study identifies the variables affecting the likelihood of compliance. In particular, the DCE allowed the authors to assess farmers' prospective responses to different designs of the policy, including radically different types of greening provisions. The study also identified behavioral factors (such as attitudes towards the environment) that are likely to affect compliance decisions. However, a key limitation lies in extrapolating the results from a small localized DCE based on a non-representative sample of German farmers to the EU level. By contrast, the results of simulation models can be easily applied to all EU Member States as long as calibration data are available. Simulation provides an estimation of the EU-wide impact of the greening policy, and enables the analyst to differentiate the impacts across Member States or farm types. By contrast, the DCE provides insights into farmers' preferences specifically where the experiment has been conducted. In addition, estimates of farmers' willingness to accept can be a useful reference to establish financial sanctions for non-complying farmers.

Evaluation of agri-environmental schemes

Another example illustrates the potential complementarity of statistical methods based on observational data and economic experiments for evaluating agri-environmental schemes (AES). AES are considered to be effective only if they have an additional effect on farmers' adoption of environmentally friendly practices. Paying for practices that would have been adopted in the absence of an AES would merely represent a windfall effect. Such windfall effects should be excluded when measuring the net impact of the policy. Using observational data for a representative sample of French farmers, Chabé-Ferret and Subervie (2013) estimate the net impact of five AES. A simple comparison of the situation before and after the launch of the AES would suffer from a potential time trend bias. In addition, farmers who volunteer to enroll in AES may not be similar to those not enrolling, implying that the no-policy counterfactual cannot be approximated by the farmers who did not enroll in the AES (selection bias). Therefore, the authors rely on a quasi-experimental approach, namely difference-in-difference matching, which allows the analyst to control for both types of biases.

While this method offers a solution to identify the true causal effect of a policy, quasi-experimental methods do require appropriate data and some specific assumptions need to be fulfilled. An alternative way to determine the causal effect of AES would be to implement an RCT where only half of the farmers, randomly chosen, can decide to enroll in AES, while the other half serves as comparison group. However, RCT also have limits and raise practical challenges. For instance, it may be politically difficult to offer the AES only to a subsample of the population and to refuse enrolment to the rest of the population.

Yet, economic experiments could still be useful for testing variations in the design of AES. As illustrated by the many studies reviewed in section 3, DCE can provide cues as to which characteristics of an AES are perceived most restrictive and could be relaxed in order to increase scheme uptake. Such methods generate relatively cheap and rapid information on potential ways to improve the policy.

A further step would be to set-up an RCT to compare farmers' reactions to a set of alternative AES. Such an RCT would most likely face less ethical concerns, as it would still offer all potentially interested farmers the option to subscribe. They would only be offered slightly different designs of

the program. In this way, RCT could provide reliable results on which types of AES provide the most desirable outcomes.

5. Conclusions and policy recommendations

This article provides an overview of the usefulness and advantages as well as the limitations and challenges of economic experiments as new and complementary tools for agricultural policy evaluation. We conclude with a list of recommendations aiming at facilitating the incorporation of economic experiments into the CAP evaluation toolbox: these recommendations arise from the presentations and discussions at the workshop "*How can economic experiments inform EU agricultural policy?"*, organized in January 2015 at the Directorate-General for Agriculture and Rural Development of the European Commission (best known as DG-Agri). The objectives of the workshop were twofold: to inform DG-Agri staff on the potential of economic experiments; and to share views on ways to promote efficiently these approaches in European agricultural policy-making and evaluation. The recommendations are formulated both with a view of enhancing the use of economic experiments in CAP evaluations conducted or commissioned by the European Commission, but also for the evaluation of other national or local agricultural policies.

Be open to innovative methods

The first conclusion addressed to evaluators and experts is to be open to innovative methodologies and to make full use of the numerous complementarities that can be found in the evaluation toolbox. A general observation is that, although the current EC tendering process for agricultural policy evaluation does not exclude economic experiments, responses to calls for tender are often biased in favor of traditional methods (case studies, econometric analysis, simulation models). This is often due to path dependency and false expectations. We understand that bidders who have demonstrated their competence in mastering traditional evaluation techniques fear that introducing innovative techniques may reduce their chances of success in the tender. Yet, the EC procedures ensure that evaluators are free to propose the methodological approaches they wish to mobilize.

Relying on a combination of experimental and non-experimental approaches and on a gradual approach in the use of economic experiments can increase the cost efficiency of an evaluation proposal and improve the quality of conclusions.

For example, laboratory experiments are typically appropriate methods to pre-test individual and group responses to different incentives or policy designs. If the policy does not work as anticipated in the lab, it is very likely to fail in a field application as well (Plott, 1997). Moreover, different treatments of a lab experiment allow comparing several initial versions of a policy easily. The experiment plays the equivalent role of wind-tunnel testing for a newly designed aircraft. Then, as emphasized by Hellerstein et al. (2015) from USDA's Economic Research Service, "the next logical step after laboratory testing is a field test—the equivalent of a test flight" (Hellerstein et al. 2015). Lab experiments, initially run with students, can then be repeated with farmers, in a more contextualized setting; in order to check the sensitivity of outcomes to the context in which the

policy is applied. In a final step, pilot studies and RCT can then be envisaged to confirm results, fine-tune implementation and engage a policy dialogue with stakeholders and policy makers. Conducting model simulations is extremely useful to capture the indirect effects and market implications of new policies. Behavioral parameters or elasticities estimated through lab or field experiments can be plugged into models to improve their external validity. DCE can also provide estimates which are useful for models, such as the expected adoption rates of various policy designs.

Reconsider the search for representativeness

Policy officers generally look for evidence general enough to be valid for the entire population affected by the policy. Given the limitations in sample sizes and the sampling issues involved in experimental studies, their representativeness can be questioned. Yet, given the high interval validity of most lab experiment results, they can make a solid contribution to the policy-making process: for example, behavioral findings, replicated over time and across domains (e.g. evidence of loss aversion), can safely be assumed to be valid everywhere and at any time and can therefore help understand reactions to policy of a large share of the EU farming population. Some experimental results can be easily extended to a wide range of policy issues and contexts and remobilized in other evaluation processes (van Bavel et al. 2015). For example, experiments to identify efficient designs for agri-environmental auctions provide results on the acceptability and efficiency of different auction formats which are useful in various contexts (Cason and Gangadharan, 2005; Hellerstein et al., 2015).

Nevertheless, other behavioral components such as sensitivity to social norms are more local. When the policy change might be affected by behavioral factors which are culturally entrenched, it is essential to plan for experiments in different Member States. A careful selection of the case studies areas and samples of participants can help ensure that some of the heterogeneity among farmers across EU regions is captured. If no 'country effects' are identified, results can probably be generalized to the EU as a whole (van Bavel et al. 2015)

Find responses to moral and ethical obstacles

All aspects of EU action, including evaluation procedures, require the absence of discrimination between stakeholders. The design of experiments must respect this principle. As long as experiments do not involve effective policy interventions (as is the case for most discrete choice, lab and field experiments), discrimination is not an issue. The case of RCT is more controversial because treatment is effective: assignment can create ethical problems since it means that treatment is made available to some (the treated group) and denied to others (the control group). Many authors have thus questioned fairness and morality of RCT (see Baele, 2013 for a review). Moreover, unequal access to policy may be challenged in the court, or by European competition policies when the experimental program creates undue inequity between citizens (e.g. in the case of a subsidy program). This problem can be particularly problematic in Europe's agricultural sector since the CAP edicts fairly strict rules (especially on first pillar's payments) to avoid competition distortions across European farmers. Indeed, experimental programs may require a notification procedure and obtain

approval by the relevant European Commission services. Recent works, especially in the development literature, have devised innovative ways of introducing randomization into existing programs with minimal disruption. Duflo et al. (2007), Shadish et al. (2002) and Morawetz (2014) describe close-to-random procedures for randomization designed to both mitigate self-selection bias and increase acceptance compared to pure randomization. One close-to-random procedure is to apply the treatment to participants in successive waves. Financial and administrative constraints can lead to phase-in programs over time, and randomization can be argued to provide the fairest way of determining the order of phase-in. Another solution is the "encouragement design" which is randomly assigned to participants. For example, a random sample of farmers receives by mail an invitation to participate in a program. The farms targeted by the campaign are therefore more likely to participate in the program than others, but this probability is not equal to one as it would be in a classical assignment procedure. Econometric techniques, using the dummy variable "has received an encouragement" as a natural instrumental variable, allow the analyst to measure the treatment effect (see Duflo et al. 2007 for more details). Other innovative proposals exist such as the free lunch randomization proposed by Morawetz (2014).

As a general conclusion, it seems to us that giving more space to economic experiments in the CAP evaluation toolbox is not only feasible but also advisable. Currently mentioned obstacles, such as representativeness concerns, ethical issues associated with randomization, or supposed disinterest of policy makers and evaluators in this type of approach can be overcome. As mentioned by Herberich et al. (2009) in their paper titled "Can field experiments return agricultural economics to the glory days?" economic experiments can be an opportunity to open a new line of research stimulating for researchers in agricultural economics, willing to have their research relevant for policy advice. It would complement the useful advice already provided by Brink (2013). Such methodological advance is also highly relevant for agricultural policy makers looking for policy advice based on sound evidence and the use of best available methods.

6. References

Alfnes, F., A. G. Guttormsen, G. Steine, and K. Kolstad. 2006. "Consumers' Willingness to Pay for the Color of Salmon: A Choice Experiment with Real Economic Incentives." *Americal Journal of Agricultural Economics* 88(4): 1050–61.

Baele, S. 2013. "The Ethics of New Development Economics: Is the Experimental Approach to Development Economics Morally Wrong?" *The Journal of Philosophical Economics* VII(1). http://www.jpe.ro/pdf.php?id=4998 (July 25, 2015).

Bahrs, E., S. Kroll, and M. Sutter. 2008. "Trading Agricultural Payment Entitlements: An Experimental Investigation of Bilateral Negotiations." *American Journal of Agricultural Economics* 90(5): 1201–7.

Banerjee, Abhijit, E. Duflo, C. Imbert, S. Mathew and R. Pande. 2014. "Can E-Governance Reduce Capture of Public Programs? Experimental Evidence from a Financial Reform of India's Employment Guarantee." *Working paper Poverty Action Lab*.

http://www.povertyactionlab.org/fr/publication/can-e-governance-reduce-capture-public-programs-experimental-evidence-financial-reform-i (July 25, 2015).

Van Bavel, R., N. Rodríguez-Priego and I. Maghiros. 2015. “Seven Points to Remember When Conducting Behavioural Studies in Support of EU Policy-Making.” *JRC Scientific and Policy Report*.

Berriet-Solliec, M, P. Labarthe and C. Laurent. 2014. “Goals of Evaluation and Types of Evidence.” *Evaluation* 20: 195–2013.

Birol, E., M. Smale and Á. Gyovai. 2006. “Using a Choice Experiment to Estimate Farmers’ Valuation of Agrobiodiversity on Hungarian Small Farms.” *Environmental and Resource Economics* 34(4): 439–69.

Blair, R., K. Fortson, J. Lee and A. Rangarajan. 2013. “Should Foreign Aid Fund Agricultural Training ? Evidence from Armenia August 2013.” *Mathematica Policy Research Working Paper* 19 (August).

Bouguen, A., F. Deon, K. Macours and S. Naudeau. 2014. “Preschools and Early Childhood Development in a Second Best World: Evidence from a Scaled-up Experiment in Cambodia.” *CEPR Discussion Paper* 10170. http://www.povertyactionlab.org/publication/preschools-and-early-childhood-development-second-best-world-evidence-scaled-experiment- (July 25, 2015).

Breustedt, G., N. Schulz and U. Latacz-Lohmann. 2013a. “Calibrating Agri-Environmental Schemes Using a Two-Stage Discrete Choice Experiments.” *German Journal of Agricultural Economics* 62(4): 259–75.

Breustedt, G., N. Sshulz and U. Latacz-Lohmann. 2013b. “Factors Affecting Participation and Compensation Requirements in Agri-Environmental Schemes: Insights from a Discrete Choice Experiment.” *German Journal of Agricultural Economics* 62(4): 244–58.

Brink, L. 2013. "Making Agricultural Economics Research Relevant for Policy Advice." *Canadian Journal of Agricultural Economics* 61: 15–36

Broch, S.W. and S.E. Vedel. 2012. “Using Choice Experiments to Investigate the Policy Relevance of Heterogeneity in Farmer Agri-Environmental Contract Preferences.” *Environmental and Resource Economics* 51(4): 561–81.

Bulte, E., G. Beekman, S. Di Falco, J. Hella and P. Lei (2014). Behavioral responses and the impact of new agricultural technologies: Evidence froma double-blind field experiment in Tanzania. *American Journal of Agricultural Economics* 96(3): 813–30.

Burrell, A. and Z. Nii-naate. 2013. “Partial Stochastic Analysis with the European Commission’s version of the AGLINK-COSIMO Model.” *JRC Reference Reports* (Report EUR 25898 EN, European Commission, JRC-IPTS).

Burtless, G. 1995. “The Case for Randomized Field Trials in Economic and Policy Research.” *Journal of Economic Perspectives* 9(2): 63–84.

Cason, T.N. and L. Gangadharan. 2005. “A Laboratory Comparison of Uniform and Discriminative Price Auctions for Reducing Non-Point Source Pollution.” *Land Economics* 81(1): 51–70.

Cason, T.N., L. Gangadharan and C. Duke. 2003. “A Laboratory Study of Auctions for Reducing Non-Point Source Pollution.” *Journal of Environmental Economics and Management* 46(3): 446–71.

Chabé-Ferret, S. and J. Subervie. 2013. “How Much Green for the Buck? Estimating Additional and Windfall Effects of French Agro-Environmental Schemes by DID-Matching.” *Journal of Environmental Economics and Management* 65(1): 12–27.

Chong, A., A.L. De La O, D. Karlan and L. Wantchekro. 2010. “Information Dissemination and Local Governments’ Electoral Returns, Evidence from a Field Experiment in Mexico.” *Working paper Poverty Action Lab*. http://www.povertyactionlab.org/fr/publication/information-dissemination-and-local-governments%E2%80%99-electoral-returns-evidence-field-experi (July 25, 2015).

Christensen, T., A.B. Pedersen, H.O. Nielsen, M.R. Mørkbak, B. Haslerand S. Denver. (2011). "Determinants of farmers’ willingness to participate in subsidy schemes for pesticide-free buffer zones - A choice experiment study". *Ecological Economics*, 70(8): 1558–1564.

Ciriolo, E. 2011. “Behavioural Economics in the European Commission: Past, Present and Future.” *Oxera*.

Cook, T. D. 2003. “Why Have Educational Evaluators Chosen Not to Do Randomized Experiments?” *The Annals of the American Academy of Political and Social Science* 589(1): 114–49.

Cummings, R.G., C.A. Holt and S.K. Laury. 2004. “Using Laboratory Experiments for Policy Making: An Example of the Georgia Irrigation Reduction Auction.” *Journal of Policy Analysis and Management*.

DG AGRI. “Evaluation of Policy Measures in Agriculture.” http://ec.europa.eu/agriculture/evaluation/index_en.htm (July 24, 2015).

Dizon-Ross, R., P. Dupas and J. Robinson. 2015. “Governance and the Effectiveness of Public Health Subsidies.” *NBER Working Paper No. 21324*. http://www.povertyactionlab.org/fr/publication/governance-and-effectiveness-public-health-subsidies (July 25, 2015).

Duflo, E., R. Glennester and M. Kremer. 2007. “Using Randomization in Development Economics Research: A Toolkit.” *Handbook of Development Economics* 4: 3895–3962.

Duflo, E., M. Kremer and J. Robinson. 2011. “Nudging Farmers to Utilize Fertilizer: Theory and Experimental Evidence from Kenya.” *American Economic Review* 101(October): 2350–90.

EC 2011. Commission Staff Working Paper *IMPACT ASSESSMENT Common Agricultural Policy towards 2020*. http://ec.europa.eu/agriculture/policy-perspectives/impact-assessment/cap-towards-2020/report/full-text_en.pdf.

EC 2013. *Evaluation of the Structural Effects of Direct Support*. http://ec.europa.eu/agriculture/evaluation/market-and-income-reports/structural-effects-direct-support-2013_en.htm (July 24, 2015).

EC (2013b). Designing and Commissioning of Counterfactual Impact Evaluations. Retrieved from http://ec.europa.eu/social/BlobServlet?docId=10805&langId=en

EC 2014. *Investment Support under Rural Development Policy*. http://ec.europa.eu/agriculture/evaluation/rural-development-reports/investment-support-rdp-2014_en.htm.

EC 2015. *Better Regulation Guidelines*.

Espinosa-Goded, M., J. Barreiro-Hurlé and E. Ruto. 2010. "What Do Farmers Want from Agri-Environmental Scheme Design? A Choice Experiment Approach." *Journal of Agricultural Economics* 61(2): 259–73.

EU. 2013. "Regulation (EU) No 1307/2013 Establishing Rules for Direct Payments to Farmers under Support Schemes within the Framework of the Common Agricultural Policy and Repealing Council Regulation (EC) No 637/2008 and Council Regulation (EC) No 73/2009." *Official Journal of the European Union* OJ L 347/6.

Fafchamps, M. and B. Minten. 2012. "Impact of SMS-Based Agricultural Information on Indian Farmers." *World Bank Economic Review* 26(3): 383–414.

Giné, X., J. Goldberg and D. Yang. 2012. "Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi." *The American Economic Review* 102(6): 2923–54.

Gneezy, U. and A. Rustichini. 2000. "A Fine Is a Price." *The Journal of Legal Studies* 29(1): 1-17.

Goldin, I., F. Halsey Rogers and N. Stern. 2012. "Making Aid Work." *Boston Review*. http://bostonreview.net/goldin-rogers-stern-economy-whole (July 26, 2015).

Greiner, R. 2015. "Factors Influencing Farmers' Participation in Contractual Biodiversity Conservation: A Choice Experiment with Northern Australian Pastoralists." *Australian Journal of Agricultural and Resource Economics* 58(1): 1-21

Häring, A.M., S. Dabbert, J. Aurbacher, B. Bichler, C. Eichert, D. Gambelli, N. Lampkin, F. Offermann, S.Olmos, J. Tuson and R. Zanoli. 2004. Organic Farming in Europe: Economics and Policy. *Organic Farming and Measures of European Agricultural Policy*. Volume 11. University of Hohenheim.

Harrison, G. W. 2014. "Cautionary Notes on the Use of Field Experiments to Address Policy Issues." *Oxford Review of Economic Policy* 30(4): 753–63.

Harrison, G.W., and J. A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42(4): 1009–55.

Hellerstein, D., N. Higgins and M. Roberts. 2015. "Options for Improving Conservation Programs: Insights from Auction Theory and Economic Experiments." *U.S. Department of Agriculture, Economic Research Service,* ERR-181(January). http://www.ers.usda.gov/media/1732610/err181.pdf (August 17, 2015).

Herberich, D.H., S.D. Levitt and J.A. List. 2009. "Can Field Experiments Return Agricultural Economics to the Glory Days?" *American Journal of Agricultural Economics* 91(5): 1259–65.

Holt, C.A., A.M. Smith and W. Shobe. 2006. “An Experimental Basis for Public Policy Initiatives.” In *Promoting the General Welfare: New Perspectives on Government Performance*, ed. E. Patashnik and A. Gerber. Washington D.C.: DC: Brookings Institution Press, 174–99.

Imbens, G.W. and J.M. Wooldridge. 2009. “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature* 47(1): 5–86.

Jaddad, A.R. and M. Enkin. 2007. *Randomized Controlled Trials: Questions, Answers, and Musings.* Second Edi. Oxford, UK: Blackwell Publishing Ltd.

Karlan, D., R. Osei, I. Osei-Akoto and C. Udry. 2014. “Agricultural Decisions after Relaxing Credit and Risk Constraints.” *The Quarterly Journal of Economics* 129(2): 597–652.

Kuhfuss, L., R. Preget and S. Thoyer. 2014. “Individual Preferences and Collective Incentives: What Design for Agri-Environmental Contracts? The Case of Wine-Growers’ Herbicide Use Reduction.” *Review of Agricultural and Environmental Studies* 95(1): 111–43.

Levitt, S.D. and J.A. List. 2007. “What Do Laboratory Experiments Tell Us About the Real World?” *Journal of Economic Perspectives* 21(2): 153–74.

Lips, M. and C. Gazzarin. 2008. “What Are the Preferences of Dairy Farmers Regarding Their Work? A Discrete Choice Experiment in the Eastern Part of Switzerland.” *Aa*: 1–6.

List, J.A. 2007. “Field Experiments: A Bridge between Lab and Naturally Occurring Data.” *The B.E. Journal of Economic Analysis & Policy* 5(2): 1–47.

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

Loi, M. and M. Rodrigues. 2012. A note on the impact evaluation of public policies: the counterfactual analysis. *JRC Scientific and Policy Report*. Retrieved from http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/26891/1/lbna25519enn.pdf

Louhichi, K., P. Ciaian, M. Espinosa, L. Colen, A. Perni and S. Gomez y Paloma. 2015. *An EU-Wide Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP)*. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92574/jrcreport_jrc92574.pdf.

Louviere, J.J., D.A. Hensher and J.D. Swait. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge, U.K.: Cambridge University Press.

Lusk, J. L., and T. C. Schroeder. 2004. “Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks.” *American Journal of Agricultural Economics* 86(2): 467–82.

Maart-Noelck, S.C., O. Musshoff and M. Maack. 2013. “The Impact of Price Floors on Farmland Investments: A Real Options Based Experimental Analysis.” *Applied Economics* 45(35): 4872–82.

McIntosh, C.R., J.F. Shogren and E. Dohlman. 2007. “Supply Response to Countercyclical Payments and Base Acre Updating Uncertainty: An Experimental Study.” *American Journal of Agricultural Economics* 89(4): 1046-57.

Michalek, J., P. Ciaian and D'Artis Kancs. 2015. "Investment Crowding Out: Firm-Level Evidence from Northern Germany." *Regional Studies*: 1–16. http://www.tandfonline.com/doi/full/10.1080/00343404.2015.1044957.

Mobarak, A.M. and M.R. Rosenzweig. 2013. "Informal Risk Sharing , Index Insurance , and Risk Taking." *American Economic Review* 103(3): 375–80.

Morawetz, U.B. 2014. "A Concept for a Randomized Evaluation of Agri-Environment Measures." In *The Common Agricultural Policy in the 21st Century*, eds. E. Schmid and S. Vogel. Vienna, 113–30.

Nagler, A.M., D.J. Menkhaus, C.T. Bastian, M.D. Ehmke and K.T. Coatney.2013. "Subsidy Incidence in Factor Markets: An Experimental Approach." *Journal of Agricultural and Applied Economics* 45(1): 17–33.

Parkhurst, G.M., J.F. Shogren, C. Bastian, P. Kivi, J. Donner and R.W.B. Smith. 2002. "Agglomeration Bonus: An Incentive Mechanism to Reunite Fragmented Habitat for Biodiversity Conservation." *Ecological Economics* 41(2): 305–28.

Paulrud, S. and T. Laitila. 2010. "Farmers' Attitudes about Growing Energy Crops: A Choice Experiment Approach." *Biomass and Bioenergy* 34(12): 1770–79.

Plott, C.R. 1989. Chapter 19: An updated review of industrial organization: applications of experimental methods, In: R. Schmalensee and R. Willig, Editors: Handbook of Industrial Organization, Elsevier, 1989, Volume 2, 1109-1176

Plott, C. R. 1997. "Laboratory Experimental Testbeds: Application to the PCS Auction". *Journal of Economics and Management Strategy*, 6(3): 605–638.

Rodrik, D. 2008. *We Shall Experiment, but How Shall We Learn?* Faculty Research Working Papers Series *The New Development Economics: John F . Kennedy School of Government - Harvard University October 2008*.

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." *American Journal of Agricultural Economics* 91(5): 1266–71.

Schulz, N., G. Breustedt and U. Latacz-Lohmann. 2014. "Assessing Farmers' Willingness to Accept 'Greening': Insights from a Discrete Choice Experiment in Germany." *Journal of Agricultural Economics* 65(1): 26-48.

Shadish, W. R, T. D. Cook and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Ho.

Shafir, E. 2012. *The Behavioral Foundations of Public Policy*. Princeton University Press.

Witzke, H.P. and A. Gocht. "CAPRI Modelling System." http://www.capri-model.org/dokuwiki/doku.php?id=capri:contact (January 1, 2015).