Erwin Schmid, Stefan Vogel (Hg.)

The Common Agricultural Policy in the 21st Century

Europäische Agrarpolitik im 21. Jahrhundert

Festschrift für Markus F. Hofreither

facultas.wuv

Adresse der Herausgeber:
Institut für nachhaltige Wirtschaftsentwicklung
Department für Wirtschafts- und Sozialwissenschaften
Universität für Bodenkultur Wien
Feistmantelstraße 4 / 1090 Wien / Austria
T: ++43/1/47654/3670 F: ++43/1/47654/3692 W: www.wiso.boku.ac.at/inwe.html

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Umschlagbild: © saicle – fotolia.com
Druck: Gorenjski tisk storitve
Printed in Slovenia
ISBN 978-3-7089-1083-3

Content

Foreword

Markus F. Hofreither studied economics at the Johannes Kepler University in Linz, Austria. He graduated with a doctorate in economics in 1984 and received his tenure as a full-time associate professor in 1989. He wrote his dissertation on economic effects of working hour reductions and his habilitation thesis on science-based policy support as a research field in economics. In September 1991, Markus F. Hofreither was appointed full professor for the chair in economics, economic policy and agricultural policy at the University of Natural Resources and Life Sciences, Vienna (BOKU). At that time, the accession of Austria to the European Community as well as the consequences of the GATT Uruguay Round were on the political agenda. Scientific expertise and quantitative analyses were demanded by stakeholders, among others from the ministry of agriculture, European Commission, agricultural lobbying groups, and civil society. Issues such as EU enlargement and integration, WTO negotiation and trade liberalisation, and multifunctional agriculture as well as stakeholder requests for scientific, quantitative analyses on agricultural policy topics played the main part in the research work of Markus F. Hofreither in the last 25 years.

In this book, colleagues and collaborators of Markus F. Hofreither contribute their scientific articles on the analysis of the European agricultural policy and the political decision-making process. In particular, Alan Matthews elaborates on the question whether there are lessons from the European experience for other countries pursuing regional integration by including agriculture in the integration process. In his article, he also provides a clear and concise overview of the Common Agricultural Policy (CAP) development since its implementation in the 1960s. Stefan Tangermann investigates the question whether direct payments, which were originally introduced to compensate for the reductions in the price supports, are going to become a permanent feature of the CAP. He provides a comprehensible discussion on modifications and justifications of direct payments in the political debate. Emil Erjavec and Karmen Erjavec complement the previous contribution providing a discourse analysis to the CAP measures and budgetary distribution of the recent reform. The authors describe which discourses and discourse strategies predominate in the political documents and explain how they were implemented into measures and budget distribution.

The following articles provide evaluations of agricultural policy measures as well as methodological advances in policy evaluation and multi-sectoral modelling. In their article, Paul Feichtinger, Klaus Salhofer, Franz Sinabell and Stanley R. Thompson explore the extent to which agricultural subsidies granted to farm operators are capitalized in land rental prices and eventually benefit land owners. They show that the capitalization has increased with the introduction of Single Farm Payments, remains high for disadvantaged area payments and is not significant for agri-environmental payments. Christoph R. Weiss discusses two problem areas of scientific policy evaluation, i.e. heterogenous effects of policies and social interactions. He develops a simple spatial model and shows that common evaluation approaches can produce biased results. The biases are driven by (i) the relative intensity of social interactions, (ii) the degree of neighbourhood, and (iii) the extent of the policy or programme. He also discusses strategies to reduce the biases. Ulrich Morawetz proposes a concept for a randomized evaluation of agri-environmental measures to better quantify the impact of public spending on environmentally friendly farming. This concept is expected to increase the acceptability of farmers and administration. Kurt Kratena and Gerhard Streicher present a methodological framework for a family of regional econometric Input-Output models. This model framework allows consistent economic linkage between agriculture and other sectors of economy as well as between regional and global economy levels. Walter Schneeberger provides an overview and description of the natural production conditions as well as the structural and income development in Austrian agriculture since the 1950s. In addition, figures on domestic agricultural production and consumption as well as self-sufficiency rates complement the structural farm and sector analysis. Finally, Friedrich Schneider presents up-to-date facts of the shadow economy in Germany, Austria and Switzerland and discusses policy elements of good governance and measures that support reducing the shadow economy.

This latter contribution reminds of the beginning of Markus F. Hofreither's academic career as the analysis of the shadow economy was among his first research topics. Describing and explaining behavioural phenomena of economic agents in a shadow economy requires knowledge and means of measuring, analysing and conceptualizing. Transferring and

applying economic knowledge to the fields of agricultural policy analysis in research and teaching belongs to the core competence of Markus F. Hofreither. In doing this, he has influenced a generation of students at BOKU as well as the political debate on agricultural policy in Austria. The scientific work of Markus F. Hofreither is based on the conviction that science shall support the political decision-making process providing evidence and objectivity in order to advance solutions which sustainably promote the welfare of society as a whole.

Erwin Schmid and Stefan Vogel

Anschrift des Verfassers

Univ. Prof. Dr. Christoph R. Weiss
Institut für Volkswirtschaftspolitik und Industrieökonomie
Wirtschaftsuniversität Wien
Welthandelsplatz 1, Gebäude D4, A-1020 Wien, Österreich
Tel.: +43 1 31336 4503
eMail: cweiss@wu.ac.at

A concept for a randomized evaluation of agri-environment measures

Ulrich B. MORAWETZ

Abstract

Approaches to program evaluation introduced in development and labor economics provide ample opportunities to evaluate agri-environment measures (AEMs). Randomized controlled trails, among the most discussed approaches, have not yet fully reached the practice of evaluation of AEMs. One difficulty with randomized controlled trails is the lack of acceptance in the target population. In this article I propose a concept for randomized evaluation of AEMs, which I hope is practicable and acceptable at least for selected AEMs.
Keywords: Agri-envrionment measures, evaluation, randomized controlled trails, additionality

1. Introduction

Agri-environment measures (AEMs) are part of the Common Agricultural Policy (CAP) of the EU. In Austria, over 20% of the CAP funds go to AEMs. In 2012 this corresponded to a total of over 526 million Euros of public spending for AEMs (BMLFUW, 2013, 114).
In principle, the evaluation of AEMs should be guided by the "Common Monitoring and Evaluation Framework" (DIRECTORATE GENERAL FOR AGRICULTURE AND RURAL DEVELOPMENT, 2006). Though evaluation is difficult and, as was concluded in a study commissioned by the European Court of Auditors, "the objectives [of AEMs] were overall too vague to be useful for assessing the extent to which they have been achieved; the policy was not designed and monitored so as to deliver tangible envi-

Published 2014 in: Schmid, E. and Vogel, S. (eds.): The Common Agricultural Policy in the 21st Century. Vienna: Facultas, p. 113-130.

ronmental benefits" (EUROPEAN COURT OF AUDITORS, 2011). This conclusion suggests that EU member states, which are in charge of designing AEMs, do not consider the evaluation of measures sufficiently when designing AEMs.

Agricultural economists are aware of the difficulties to evaluate AEMs, and some progress has been made in the debate. One approach that makes evaluation (almost) redundant is to change from the current practice of action-oriented schemes to result-oriented schemes. In result-oriented schemes the environmental impact is directly measured and paid for (e.g. for change in a biodiversity index). The survey by BURTON and SCHWARZ (2013) gives an overview about experiences with, mostly efficiency increasing, result-oriented methods. Unfortunately, result-oriented schemes are limited to AEMs with easily measurable outcomes and experience is also still limited. But first research results and experience, e.g. in Germany in the area of animal welfare and Switzerland, are promising.

For ex-ante evaluation, a frequently chosen approach is to make behavioral assumptions about farms behavior and combine results with bio-physical models to estimate the impact of a suggested policy on the environment (see, SCHÖNHART et al. (2011b) for an example). The strength of these models is the ability to make ex-ante evaluations and to link economic with bio-physical models and thereby integrating two scientific disciplines.

For ex-post program evaluation, behavioral assumptions can be substantially reduced by using statistical and econometric methods. Sophisticated econometric methods have been developed in several sub-disciplines of economics (mainly development and labor economics), and the methods have reached a level of maturity that makes them an important tool for evaluation (IMBENS and WOOLDRIDGE, 2009). Several authors have called to adopt these tools in environmental economics (FERRARO, 2009; GARROD et al., 2012; FRONDEL and SCHMIDT, 2005; GREENSTONE and GAYER, 2009). Evaluators of AEMs though, have only begun applying current developments in program evaluation (see CHABÉ-FERRET and SUBERVIE (2013) and PUFAHL and WEISS (2009) for two examples where recent program evaluations methods have been applied). The objective of this article is to give examples where econometric evaluation methods have been applied so far and to present a novel concept how a randomized controlled trail might be conducted for selected AEMs.

2. Evaluation framework

Cost benefit analysis of an agri-environment measure (AEM) involves comparison of discounted costs and benefits with and without AEM. A key part for evaluation is therefore to establish the counterfactual: what would have happened without the AEM (PEARCE, 2005). Several authors have argued that with uniform payments those farms with the lowest opportunity costs of entering an AEM could be expected to do so first (HANLEY et al., 1999; KLEIJN and SUTHERLAND, 2003). Some farms may well be able to meet the conditions of the management agreement at no extra cost, implying they receive payments for doing nothing additional (windfall gains) with no benefit for the environment. Thus, knowing the counterfactual will help to improve efficiency of AEMs.

A precondition to estimate the impact of AEMs is to have a precise environmental objective. Unfortunately, many AEMs are lacking a precise and measurable environmental objective (EUROPEAN COURT OF AUDITORS, 2011). Instead, management practices are frequently the objectives. Unfortunately though, the link between management practice and environmental conditions is far from trivial. FERRARO (2009) and HOLE et al. (2005) list non-linearity, thresholds, large spatial scales and spillovers as some of the reasons for difficulties to estimate environmental effects. Explicit econometric modeling of biophysical relations is difficult and rare (for one encouraging approach see HOFREITHER and PARDELLER (1996)). Therefore, FERRARO (2009) suggests restricting attention on measuring behavioral change as a first step. Having estimated the counterfactual behavior, economists would have to rely on environmental scientists to infer the effects on the environmental condition. Thus, where the

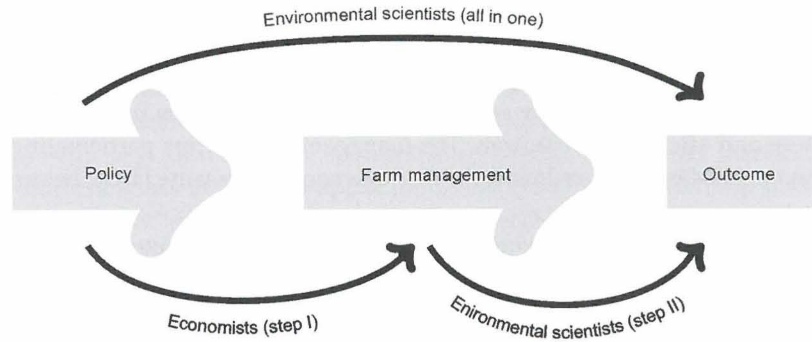

Fig. 1: Direct and indirect evaluation of environmental impact
Source: own

direct environmental effect of AEMs is not measurable (be it due to unclear environmental objectives or too complex links) a second best option is to measure behavior. The idea is illustrated in Figure 1. Economists can focus on estimating behavior change and, in a second step, environmental scientists can estimate the impact of the changed in behavior on the environment. In what follows, I restrict the discussion and examples to the first step as this is where agricultural economists benefit most from developments in other economic sub-disciplines. Of course, the first step has to be done in a way that provides environmental scientists with information detailed enough for the second step.

3. Empirical approaches

The difference between how a farm would have been managed without participation in an AEM and how the farm was actually managed under an AEM is called "treatment effect" of an AEM in statistics. Estimation of the "average treatment effect" can be done from observational data (under relatively strong assumptions) or from experiments (under weaker assumptions). The following sections describe possible approaches to observational and experimental studies and provide examples for applications in the evaluation of AEMs. The examples are restricted to articles using farm management as outcome (Step I). Several articles using environmental indicators as outcome (Step II) are cited in the survey by BURTON and SCHWARZ (2013). For a more formal, though still well readable description of the methods I recommend FRONDEL and SCHMIDT (2005), HENNING and MICHALEK (2008) or ANGRIST and PISCHKE (2009) on which the following is partly based on.

3.1 Before-after estimator

The before-after estimator is based on a comparison of the outcome before and after an intervention. The management of farms participating in an AEM is compared with the management of the same farms before participation.

The assumption underlying the estimator of the average treatment effect is that the management of the farm would not have changed if the farm had not participated in the AEM. For example, the amount of manure applied to the fields would have remained constant if the farm had not participated. Though, if e.g. farms change their crop mix due

to changes in relative prices, this might not be the case. Additionally it is necessary to assume that the outcome before the participation is not influenced by the anticipation of the AEM.

Pure before-after estimators are rarely found in econometric publications as usually effort is undertaken to adjust for a likely bias. Though, if longitudinal data are available it is a logical starting point.

3.2 Cross-section estimator

In many cases, conditions change too much for the assumption of the before-after estimator to hold. In such cases the mean of the outcome of the non-participants might be used to replace the mean of the unobservable counterfactual. Hence, participants are compared to non-participants.

The assumption necessary for an unbiased estimation of the average treatment effect is that no unobserved factors, such as environmental consciousness of the farmer or unobserved site characteristics, influence the decision to participate in the AEM. Otherwise, this omitted variable leads to a selection bias and a biased average treatment effect.

An example of a cross-section estimator where the outcome is a management practice is found in the article by PRIMDAHL et al. (2003) who compare Swiss AEMs participants with non-participants and show how they differ with respect to farm management.

3.3 Difference-in-difference estimator

The difference-in-difference estimator is based on observations over time for participants and non-participants. In comparison to the before-after estimator it accounts for changes of the circumstances, which affect participants and non-participants alike. This is achieved by comparing the difference in the outcome of participants before and after an intervention to the difference of non-participants before and after an intervention. The estimator therefore accounts for unobservable farm heterogeneity (e.g. difference in environmental consciousness and therefore different levels in fertilizer application). For example, if lower crop price expectations lead to reduced fertilizer applications by participants and non-participants, the difference-in-difference estimator is unbiased. But this is only true, as long as the effect of reduced crop price expectations is of the same magnitude for participants and non-participants.

Examples of difference-in-difference estimators where outcome is management practice include the article by PRIMDAHL et al. (2003) mentioned above and the article by CHABÉ-FERRET and SUBERVIE (2013) and UDAGAWA et al. (2013). CHABÉ-FERRET and SUBERVIE calculate the effects for five French AEMs and derive their cost-effectiveness. Among other things, they find that the AEM subsidizing the planting of cover crops has increased cover crops by 10 hectares on the average recipient farm at the expense of almost 7 hectares of windfall gains. Subsidizing conversion to organic farming, in contrast, has low windfall gains and high additionality. UDAGAWA et al. (2013) find for their difference-in-difference estimation that the UK Entry-Level-Stewardship scheme payment broadly compensates for losses of cereal farm income without providing over compensation. Both, CHABÉ-FERRET and SUBERVIE and UDAGAWA et al. combine their difference-in-difference estimators with matching, which is discussed next.

3.4 Matching estimators

The basic idea of the matching estimators is to compare similar farms. Each farm that is participating in the AEM is compared with one or more farms that have similar observable characteristics but do not participate. The central assumption is that the participation in the AEM is independent from unobservable farm characteristics (i.e. all reasons why a farm participates have to be captured by the observable characteristics used to find the matched pairs). If the characteristics get too numerous, it is difficult to find similar farms. Therefore, Propensity Score Matching has been developed where similarity is replaced by likelihood to participate in the program. If all assumptions are met, matched pairs represent treatment and control group and average treatment effects can be calculated. The assumptions of the matching estimators are less restrictive than the difference-in-difference estimator as the average change is allowed to differ across various outcomes. Though it cannot deal with anticipation effects of the introduction of AEMs (FRONDEL and SCHMIDT, 2005) and often the results are sensitive to the choices the researchers have to take (CALIENDO and KOPEINIG, 2008).

Matching estimators where the outcome is management practice are relatively popular in agricultural economics. Examples include an article by OSTERBURG (2006) who assesses land use intensity in Germany, and SAUER et al. (2012) who assess the impact of AEMs on production intensity in the UK. PUFAHL and WEISS (2009) use a matching method to analyze the influence of AEMs on fertilizer and pesticide application in

Germany and CHABÉ-FERRET and SUBERVIE (2013) as well as UDAGAWA et al. (2013) combine their above mentioned difference-in-difference estimators with matching.

3.5 Regression discontinuity design estimators

Regression discontinuity design estimators can be used if the probability to take part in an AEM is a discontinuous function of the variables that determine participation in the AEM. For example, if an AEM is only available for farms from certain municipalities, the farms on both sides of an administrative border might be very similar with the only difference that only those on one side of the border can participate in the AEM. A comparison of farms on both sides of the border of the municipality will produce an estimate of the average treatment effect.

I am not aware of an application of regression discontinuity design to the evaluation of AEMs. An example for the application to the EU regional policy where income levels are the discontinuing variable is the work by BECKER et al. (2013) on regional growth.

3.6 Instrumental variable regression

A widely applied approach in econometrics to identify causal relationships (or treatment effects) is the instrumental variable regression. This approach is applicable if there is an instrument variable (e.g. distance to agricultural extension center) that is correlated with the treatment variable (e.g. participation in the AEM) but not with unobserved variables (e.g. unobserved soil quality). Obviously, the difficulty with this approach is to find suitable instrument variables. For example, assume agricultural extension centers are randomly distributed over the country and those farms closer to extension centers are more likely to participate in AEMs. All other farm characteristics are independent form the distance to the extension center. Then, closeness to the agricultural extension center is correlated with participation in AEMs but not with unobservable variables such as soil quality. Therefore, closeness to the extension center can be used as instrument to estimate the effect of an AEM. For unbiased estimates though, the necessary assumptions are hard to fulfill, some claim practically impossible (DEATON, 2010).

ROBERTS and BUCHOLTZ (2005) use an instrument to estimate whether the US Conservation Reserve Program to retired cropland led to unintended new plantings. Their instrument is the proportion of cropland in 1982 classified as highly erodible which serves as proxy for the proportion of land that was eligible for the program.

3.7 Randomized controlled trails

The methods described so far are based on observational data. In comparison to experimental data, they are not the outcome of a planned experiment. While a plant breeder can run an experiment by fully controlling the conditions under which his plants grow, economists do not have this possibility. Therefore, a key concept in economic evaluation experiments is randomization: the treatment is applied randomly, and consequently there is no self-selection of participation in the program. The omitted variable bias disappears for the average treatment effect. To collect experimental data, though, it is necessary to design and run an experiment, which in many cases is limited by costs and acceptance by the target population. Though, in development as well as in labor economics procedures have been introduced which allow "close to random" experiments.

All close to random procedures have some elements of random program participation. DUFLO et al. (2007) and SHADISH et al. (2002, 269) describe how randomization can be done to increase acceptance. Even though none of these approaches is a perfect randomization it will help to reduce the selection bias. Some of these methods could be adopted for the evaluation of AEMs.

- **Randomization as part of a pilot project**: Randomly offering farms to participate in a pilot study before an AEM is introduced. Those not participating are the control group.
- **Over-subscription**: If more farms want to participate in a program than can be financed, a random choice of who can participate introduces the necessary randomization.
- **During the phase-in of a program**: If it is not possible that all farms start participating in an AEM at the same time, the starting point can be randomly chosen and, until all participate, the difference between participants and non-participants can be measured.
- **Encouragement design**: A random sample of farms can be targeted by an information campaign to participate in a voluntary AEM. The farms targeted would be more likely to participate in the program than others.

KLEIJN and SUTHERLAND (2003) argue that for small-scale measures it may be practical to ask farms to identify a pair of sites on the farm and

then allocate one at random to be managed under the AEM and the other conventionally. An example of this approach is described in the article by FIRBANK et al. (2003) who test if there is a difference between the management of genetically modified herbicide-tolerant crops and conventionally varieties in the UK.

The main reason why randomized controlled trails are not used to evaluate AEMs is that it is politically difficult to randomly exclude farms from participating in AEMs. This might be perceived as unfair. I therefore suggest a mechanism that might make randomized controlled trails feasible and thereby add an additional option to the approaches discussed so far.

4. A concept for a randomized evaluation of an AEM

We can easily calculate the average difference in the outcome y (e.g. amount of fertilizer used) between those farms who participate in the AEM ($y \mid D=1$) and those who do not ($y \mid D=0$). This observed difference in the average outcome is thus $E(y \mid D=1)-E(y \mid D=0)$.

The difference can be split into the average treatment effect on the treated (how much was the effect of the AEM on the outcome y) and the selection bias. To do so, it is useful to define potential outcomes: they represent an outcome, even if this particular outcome is not observable. Each farm could participate in the AEM or not. Hence, for each farm there are two potential outcomes. One can be observed, the other one cannot (also called counterfactual). Following ANGRIST and PISCHKE (2009, 14), I write the potential outcomes as y_0 (for non-participation) and y_1 (for participation) and decompose the observed difference as:

$$\underbrace{E(y|D=1)-E(y|D=0)}_{\substack{\text{Observed difference}\\ \text{in average outcome}}} = \underbrace{E(y_1|D=1)-E(y_0|D=1)}_{\substack{\text{average treatment}\\ \text{effect on treated}}} + \underbrace{E(y_0|D=1)-E(y_0|D=0)}_{\substack{\text{selection}\\ \text{bias}}}$$

where $y=y_0+(y_1-y_0)D$.

The average treatment effect on the treated measures how much the outcome changed due to the AEM. This measures "additionality". Additionality is a key measure in evaluation as it measures the extent to which payments made to farms are buying changes which otherwise would not occur (MORRIS and POTTER, 1995).

Unfortunately, $E(y_0 \mid D{=}1)$ is "counterfactual" (we cannot observe outcome without AEM participation for those farms which participate). Hence, the effect of the AEM cannot be calculated for individual farms. If participation in the AEM is randomized, though, it is possible to calculate the average treatment effect on the treated because the selection bias is eliminated (ANGRIST and PISCHKE, 2009, 15).
Above, approaches from development and labor economics literature summarized by DUFLO et al. (2007) and SHADISH et al. (2002, 269) gave examples how randomization can be achieved. An approach, which I call "free-lunch randomization" is described next.
Imagine a new AEM is about to be introduced. From all farms that are eligible to participate a lottery selects free-lunch farms. These free-lunch farms are granted the agri-environment payment, irrespectively whether they comply with the requirements of the AEM and irrespectively whether they had originally applied for the measure. Figure 2 illustrates how from all (eligible) farms (independent whether they applied to participate in the AEM) free-lunch farms are chosen randomly.
This lottery is to be hold after the application for participation in the AEM but before the program period starts to leave time to inform the free-lunch farms that they do not have to comply with the requirements of the AEM even though they receive the full payment. The lottery allows to make observations about how farms are managed that do not have to comply with the rules (those drawn in the lottery) even though they would have participated in the AEM.
In terms of the above equation, the first term of the "average treatment

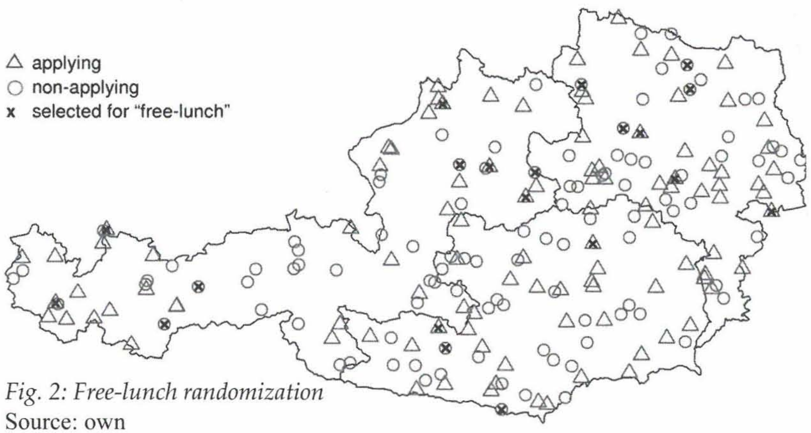

Fig. 2: Free-lunch randomization
Source: own

effect on the treated" is the mean outcome of those farms applying and participating in the AEM and the second term is interpreted as the outcome of the free-lunch farms that were applying, but were drawn in the lottery. We can thus calculate the average treatment effect on the treated as

$$\underbrace{E(y_1 \mid D{=}1)}_{\substack{\text{applying for AEM,} \\ \text{no-free-lunch}}} - \underbrace{E(y_0 \mid D{=}1)}_{\substack{\text{applying for AEM,} \\ \text{free-lunch}}}$$

where D=1 are the farms willing to participate in the AEM, y_1 is the outcome of farms that have to comply with the program requirements and y_0 is the outcome of farms which do not need to comply to the requirements.
Before discussing the practical feasibility of free-lunch randomization it is important to stress theoretical limitations. One immediate concern is that the pool of applicants to the AEM is altered through the introduction of a free-lunch lottery. This would introduce a randomization bias. Though, if farms are rational this is not the case: all farms (applying or non-applying) are participating in the free-lunch lottery and therefore the expected value of winning in the free-lunch lottery is independent from the application for the AEM. Only if farms do not belief that this is actually the case, the pool of applications to the AEM will be altered through the free-lunch lottery.
A second limitation is that it is not known whether being drawn in the free-lunch lottery changes behavior: those farms that did apply might comply with the rules (even though they do not have to) to show good will. The procedure does not allow differentiating between this motivation from other motivations (e.g. environmental consciousness). This phenomenon has been observed in experimental auctions and is termed "reciprocal obligation" (CORRIGAN and ROUSU, 2006).
A third limitation is that there might also be an income effect. Farms might be managed differently if the income is increased through the windfall gain from the free-lunch lottery. If the windfall gain is relatively small, though, this impact is likely to be small as well. Also, this effect can be quantified by comparison with farms which were not drawn in the lottery (e.g. through a propensity score matching).
Finally there is a possible indirect effect via the markets due to the free-lunch income. PUFAHL and WEISS (2009) reflect on some of these possible impacts. Such general equilibrium effects cannot be estimated as part of a randomized controlled trail but must be accounted for by using a model of the respective markets.

4.1 Applicability to the Austrian AEMs "ÖPUL"

Certainly, there are many practical challenges to a free-lunch randomization. I want to share some initial thoughts about a free-lunch randomization in the context of the Austrian AEMs. The current (2007-2013) Austrian AEMs "ÖPUL" consists of 28 measures (the 29th on animal welfare is not discussed here). Farms can voluntarily sign contracts for 5-7 years during which they have to comply with farm-management requirements. In 2012 about 76% of Austrian farms, or 89% of all agricultural land, was covered by at least one ÖPUL-measure (BMLFUW, 2013, 114).

Sampling: A practical approach to sampling would be to draw a sample from the farms participating in the Farm Accountancy Data Network (FADN). This stratified sample of about 2000 voluntarily book-keeping farms provides a valuable data base about input use and outputs. Possibly, it could be extended by further variables of interest relevant for the free-lunch evaluation. Additionally, the chance to become a "free-lunch" farm could be interpreted as a bonus for those farms voluntarily participating in FADN and therefore increase acceptance of free-lunch randomization.

Measure to evaluate: Clearly, as a randomized controlled trail runs only for a limited time, it does not make sense to randomize on measures that require a substantial change to the system of farming. In a recent report of the Institute for European Environmental Policy (IEEP) all EU AEMs were categorized as either "entry-level" or "higher-level" measures (KEENLEYSIDE et al., 2011). Entry-level measures do not require significant changes to the system of farming and are achievable by most of the target farms. Only farms with entry-level measures are suitable as candidate for free-lunch randomization. Table 1 shows that 17 of the Austrian AEMs are categorized as entry-level measures (the other 11 AEMs are categorized as higher-level measures). Of course, this is only a first, rough selection and a more careful look has to be taken on each measure individually to decide whether it really requires a significant change to the system of farming.

Outcome: Another central aspect is which outcome should be evaluated. As economists mostly are not trained in measuring environmental outcomes, I have argued above that farm management might be a more suitable outcome to be evaluated than environmental indicators. KEENLYSIDE et al. (2011) have linked AEMs with management actions required. Table 2 gives an overview about which management actions are targeted by the entry-level ÖPUL measures.

Tab 1: Entry-level measures of ÖPUL based on the IEEP report (Program numbers in brackets)

(2)	Environmentally-friendly management of arable- and grassland
(3)	Refraining from use of inputs on arable land
(4)	Refraining from use of inputs on arable forage land and grassland
(5)	Refraining from use of fungicides on area under cereals
(6)	Environmentally-friendly management of medicinal & spice plants, seed reproduction
(8)	Protection against erosion in fruit and hop growing
(10)	Protection against erosion in wine-growing
(13)	Refrain from using silage
(14)	Maintaining extensive fruit tree cultivation on grassland
(15)	Mowing of steep slopes
(16)	Cultivation of alpine meadows
(17)	Alpine farming and herding
(18)	Region Lower Austria: Eco Points
(19)	Green cover on arable land (dates for sowing and first tillage)
(20)	Mulch seed and direct seed
(24)	Cultivation of catch-crops/under-sown crops in maize cultivation
(25)	Accurate spreading of liquid manure & liquid biogas manure

Source: adopted from personal information by Clunie Keenleyside on details about the IEEP report (KEENLYSIDE et al., 2011)

Tab. 2: Management action targeted by entry-level ÖPUL measures according to Keenlyside et al. (2011)

Management action	Entry-level measures of ÖPUL																
	(2)	(3)	(4)	(5)	(6)	(8)	(10)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(24)	(25)
Maintain permanent pasture	X									X			X				
Traditional management (grass)										X	X	X					
Grazing Regime										X		X					
No grazing											X						
Restricted management dates (grass)										X							
Sheperding												X					
Hay making								X									
Cutting regime	X										X						
Specified grass or seeding regime									X								
No fertilizer application		X	X								X	X					
Limits to fertilizer application or specified regimes	X				X			X				X	X				
No plant production products (PPP)		X	X	X							X	X	X				
Limits to PPP or specified regimes																	X
No growth regulators					X												
Record keeping					X												
Grass cover in permanent crops						X	X										
Green or vegetative cover													X	X		X	
Mulching regime															X		
No tillage															X		
Rotation		X			X								X				
Maintenance or traditional orchards									X								
Specified crop varieties/or seeding regime	X																
Management of non-aquatic landscape features											X		X				
Strips or patches for wildlife	X																
Maintain area of land out of production						X	X										

Source: personal information by Clunie Keenleyside on details about the IEEP report (KEENLYSIDE et al., 2011)

If the management action required in the ÖPUL measures is compared to data available from the FADN, it is possible to identify measures where data on the outcome is readily available. Input use (limits to fertilizer or plant production products) would be an obvious choice, but quite possibly it would not be too complicated to gather data about green cover, hay making or cutting regimes.

Relevance, financial and environmental costs: Ultimately, the choice of which measure to use for a free-lunch randomization should be based on the relevance of the measure, uncertainty about additionality, the financial costs of a free-lunch randomization and the environmental costs. For example, the measure "Maintaining extensive fruit tree cultivation on grassland" would have high environmental costs as, once the fruit trees are cut, this cannot be undone after the free-lunch randomization (SCHÖNHART et al., 2011a).

The measure "Refrain from using silage" might be more appropriate. The measure supports production of hay, as grass is cut later compared to silage production, and it is hoped that this positively impacts on biodiversity. Since in the last years hay-milk was introduced and consumers pay premiums for hay-milk, additionality can be questioned. The environmental costs of free-lunch randomization would be limited as persistent negative impacts are not expected if the free-lunch farms do produce silage instead of hay for a couple of years. Monetary costs depend on the number of free-lunch farms and can be calculated without too much uncertainty. Unfortunately, FADN data do not provide information about hay produced, hence an extra effort would be necessary to collect these data (maybe as part of FADN). Finally, with over 10.000 participants, over 114.000 ha and expenses of over 18 million Euros in the year 2009 alone (BMLFUW, 2010, 203), the measure has some relevance. In total, refrain from using silage might be a candidate for a free-lunch randomization.

A word of caution might be in order. Even if it should turn out that the measure "refrain from using silage" leads to windfall gains after it has been in place for a couple of years, this does not mean that it generated windfall gains all years: A market for hay-milk might only have established, because the program made hay-milk production feasible.

Ultimately, the selection of appropriate measures for free-lunch randomization must be a joint effort of economists, environmental scientists, administration and farm representatives.

5. Conclusions

Additionality is crucial to compare the outcome of an AEM with a situation without this AEM. Therefore, in cost-benefit analysis, it is essential to estimate the additionality of the AEM. New developments in program evaluation provide ample opportunities to evaluate additionality of AEMs. Interestingly, even though EU countries provide rich sets of farm level data, examples of evaluation studies applying these methods to evaluate AEMs are relatively rare. Most likely, this has to do with a slow adoption rate by agricultural economists and with special challenges related to the evaluation of environmental outcomes. To circumvent the second problem, it has been suggested that agricultural economists focus on management as outcome instead of environmental outcomes. Additionally, to improve evaluations, elements of randomization could be included in the design of AEMs. This opens the opportunity for more robust results of ex-post evaluation that are less dependent on model assumptions. The AEMs which are suitable for elements of randomization should be determined in a joint effort of economists, environmental scientists, administration and farm representatives.

The free-lunch randomization, suggested in this paper, is one possibility of introducing randomization. Its main objective is to make a randomized controlled trail for AEMs politically feasible. It might be well acceptable among farms as no farm is worse off compared to a situation without free-lunch evaluation. It should be perceived as an opportunity to measure the impact of public spending on environmentally less harmful management by farms and their representatives. Unless there are windfall gains to hide, they should support the evaluation. The costs of free-lunch randomization have to be carried by the public. But, ultimately, the evaluation should be in the interest of the public as it reduces uncertainty about which AEMs actually increase environmentally-less harmful farming.

Acknowledgments

One of the strength of the Institute of Sustainable Economic Development is a tradition of a critical assessment of agricultural policy. In my view, this tradition can be attributed to Markus Hofreither who, unlike most other Austrian professors of economics, sought to bridge the gap

between academic research and political decision making. In the last ten years I was in the fortunate situation to work with Markus Hofreither at the above mentioned institute. This article was inspired by his attitude towards applied economic research. I also thank Sylvain Chabé-Ferre, Georg Lehecka and Martin Schönhart for their feedback on earlier versions of this text. All errors are mine.

Literature

ANGRIST, J. D. and PISCHKE, J.-S. (2009): *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

BECKER, S. O., EGGER, P. and VON EHRLICH, M. (2013): Absorptive Capacity and the Growth Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects. *American Economic Journal: Economic Policy*, 5, 4, p. 29-77.

BMLFUW (2010): Evaluierungsbericht 2010. Halbzeitbewertung des Österreichischen Programms für die Entwicklung des ländlichen Raums. Teil B. Bewertung der Einzelmaßnahmen. Vienna: Federal Ministry of Agriculture, Forestry, Environment and Water Management (BMLFUW).

BMLFUW (2013): *Der Grüne Bericht*. Vienna: Federal Ministry of Agriculture, Forestry, Environment and Water Management (BMLFUW).

BURTON, R. J. F. and SCHWARZ, G. (2013): Result-oriented agri-environmental schemes in Europe and their potential for promoting behavioural change. *Land Use Policy*, 30, 1, p. 628–641.

CALIENDO, M. and KOPEINIG, S. (2008): Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22, 1, p. 31–72.

CHABÉ-FERRET, S. and SUBERVIE, J. (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, p. 12-27.

CORRIGAN, J. R. and ROUSU, M. C. (2006): The Effect of Initial Endowments in Experimental Auctions. *American Journal of Agricultural Economics*, 88, 2, p. 448–457.

DEATON, A. (2010): Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48, 2, p. 424–455.

DIRECTORATE GENERAL FOR AGRICULTURE AND RURAL DEVELOPMENT (2006): Rural Development 2007-2013. Handbook on common monitoring and evaluation framework - Guidance document.

DUFLO, E., GLENNERSTER, R. and KREMER, M. (2007) Using Randomization in Development Economics Research: A Toolkit. In: Schultz, T. P. and Strauss, J. A.: *Handbook of Development Economics*, 4, p. 3895–3962.

EUROPEAN COURT OF AUDITORS (2011): Is agri-environment support well designed and managed? Special Report.

FERRARO, P. J. (2009): Counterfactual thinking and impact evaluation in environmental policy. *Environmental Program and Policy Evaluation. New Directions for Evaluation*, 122, p. 75–84.

FIRBANK, L. G., HEARD, M. S., WOIWOD, I. P., et al. (2003): An introduction to the Farm-Scale Evaluations of genetically modified herbicide-tolerant crops. *Journal of Applied Ecology*, 40, 1, p. 2–16.

FRONDEL, M. and SCHMIDT, C.M. (2005): Evaluating environmental programs: The perspective of modern evaluation research. *Ecological Economics*, 55, 4, p. 515–526.

GARROD, G., RUTO, E., WILLIS, K. and POWE, N. (2012): Heterogeneity of preferences for the benefits of Environmental Stewardship: A latent-class approach. *Ecological Economics*, 76, p. 104–11.

GREENSTONE, M. and GAYER, T. (2009) Quasi-experimental and experimental approaches to environmental economics. *Journal of Environmental Economics and Management*, 57, 1, p. 21–44.

HANLEY, N., WHITBY, M. and SIMPSON, I. (1999): Assessing the success of agri-environmental policy in the UK. *Land Use Policy*, 16, 2, p. 67–80.

HENNING, C. and MICHALEK, J. (2008): Ökonometrische Methoden der Politikevaluation: Meilenstein für eine sinnvolle Agrarpolitik der 2. Säule oder akademische Fingerübung? *Agrarwirtschaft*, 57, 3-4, p. 232–243.

HOFREITHER, M. F. and PARDELLER, K. (1996): An econometric analysis of the relationship between farm production and nitrate pollution of groundwater in Austria. *Die Bodenkultur*, 47, 4, p. 279–290.

HOLE, D. G., PERKINS, A. J., WILSON, J. D., ALEXANDER, I. H., GRICE, P. V. and EVANS, A. D. (2005): Does organic farming benefit biodiversity? *Biological Conservation*, 122, 1, p. 113–130.

IMBENS, G. W. and WOOLDRIDGE, J. M. (2009): Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47, 1, p. 5–86.

KEENLEYSIDE, C., ALLEN, B., HART, K., MENADUE, H., STEFANOVA, V., PRAZAN, J., HERZON, I., CLEMENT, T., POVELLATO, A., MACIEJCZAK, M. and BOATMAN, N. (2011): Delivering environmental benefits through entry level agri-environment schemes in the EU. London: Institute for European Environmental Policy. Report Prepared for DG Environment.

KLEIJN, D. and SUTHERLAND, W. J. (2003): How effective are European agri-environment schemes in conserving and promoting biodiversity? *Journal of Applied Ecology*, 40, 6, p. 947–969.

MORRIS, C. AND POTTER, C. (1995) Recruiting the New Conservationists: Farmers' Adoption of Agri-environmental Schemes in the U.K. *Journal of Rural Studies*, 11, 1, p. 51–63.

OSTERBURG, B. (2006): Assessing Long-term Impacts of Agri-environmental Measures in Germany. In: *Evaluating Agri-environmental Policies*. p. 187–205. Paris: Organisation for Economic Co-operation and Development.

PEARCE, D. (2005): What Constitutes a Good Agri-environmental Policy Evaluation? In: *Evaluating Agri-Environmental Policies*. p. 71–97. Paris: OECD Publishing.

PRIMDAHL, J., PECO, B., SCHRAMEK, J., ANDERSEN, E. and OÑATE, J. J. (2003): Environmental effects of agri-environmental schemes in Western Europe. *Journal of Environmental Management*, 67, 2, p. 129–138.

PUFAHL, A. and WEISS, C. R. (2009): Evaluating the effects of farm programmes: results from propensity score matching. *European Review of Agricultural Economics*, 36, 1, p. 1–23.

ROBERTS, M. J. and BUCHOLTZ, S. (2005): Slippage in the Conservation Reserve Program or Spurious Correlation? A Comment. *American Journal of Agricultural Economics*, 87, 1, p. 244–250.

SAUER, J., WALSH, J. and ZILBERMAN, D. (2012): Behavioural Change through Agri-Environmental Policies? A Distance Function based Matching Approach. 86th Annual Conference, April 16-18, 2012, Warwick University, Coventry, UK. (Agricultural Economics Society).

SCHÖNHART, M., SCHAUPPENLEHNER, T., SCHMID, E. and MUHAR, A. (2011a): Analysing the maintenance and establishment of orchard meadows at farm and landscape levels applying a spatially explicit integrated modelling approach. *Journal of Environmental Planning and Management*, 54, 1, p. 115–143.

SCHÖNHART, M., SCHAUPPENLEHNER, T., SCHMID, E. and MUHAR, A. (2011b): Integration of bio-physical and economic models to analyze management intensity and landscape structure effects at farm and landscape level. *Agricultural Systems* 104, 2, p. 122–134.

SHADISH, W. R., COOK, T. D. and CAMPBELL, D. T. (2002): Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

UDAGAWA, C., HODGE, I. and READER, M. (2013): Farm Level Costs of Agri-environment Measures: The Impact of Entry Level Stewardship on Cereal Farm Incomes. *Journal of Agricultural Economics*. In press.

Affiliation

Ulrich B. Morawetz
Institute for Sustainable Economic Development
University of Natural Resources and Life Sciences, Vienna
Feistmantelstraße 4, 1180 Vienna, Austria
Tel.: +43 1 47654 3672
eMail: ulrich.morawetz@boku.ac.at