Article

OXFORD

Towards more evidence-based agricultural and food policies

Nadja El Benni [ID][1,*], **Christian Grovermann** [ID][2] **and Robert Finger** [ID][3]

[1]Department Sustainability Assessment and Agricultural Management, Agroscope, 8356 Ettenhausen, Tänikon 1, Switzerland
[2]Department of Food System Sciences Science, FiBL, 5070 Frick, Switzerland
[3]Agricultural Economics and Policy Group, ETH Zürich, 8092 Zurich, Switzerland

*Corresponding author: E-mail: nadja.el-benni@agroscope.admin.ch

Abstract

The goal of this paper is to provide insights into how scientific evidence can be used for policymaking and put evidence-based agriculture and food policies at the top of research and policy agendas. We illustrate how scientific evidence can be used in a targeted manner for better policymaking and present an overview of the rich set of ex-ante and ex-post evaluation methods and tools that agricultural economists use for evaluating agricultural policies to provide evidence for policy decisions. We present insights into both established and new/emerging methods and approaches, including their advantages and disadvantages, and discuss their potential use for policy evaluation. We also discuss how methods and approaches should be combined and could be better targeted towards decision-makers. The paper also discusses the crucial role of high-quality data in supporting the science—policy interface. Finally, we present an overview of papers in this special issue titled 'Evidence-Based Agricultural and Food Policy: The Role of Research for Policy Making'.

Keywords: impact evaluation, experiment, ex-ante, agent-based model, ex-post, replication, review, policy cycle
JEL codes: Q00, Q10, Q18

1. Introduction

European agricultural policy has evolved greatly since the early 1990s concerning both the expansion of objectives and the measures to achieve them. While productivity, farm income, and affordable food prices were objectives from the very beginning, environmental and social objectives were increasingly introduced in the Common Agricultural Policy (CAP) (Pe'er et al. 2019).

Current agricultural policies aim to ensure the provision of safe, nutritional, and affordable food, to reduce the negative impact of production on the environment, to increase animal welfare, and to promote viable rural livelihoods. These objectives apply to the European Union's CAP and to the agricultural policies of other European countries, such as Switzerland (El Benni and Lehmann 2010; Pe'er et al. 2019). To reach these objectives, policy measures have been more targeted towards specific goals and have increasingly been tailored to specific farms (Matthews 2013; Finger and El Benni 2021). However, despite high governmental spending on European agricultural policies, most agri-environmental, social, and animal welfare objectives in particular, but also

most income-related objectives are not achieved (BAFU/BLW 2016; Pe'er et al. 2019; El Benni and Schmid 2022). Agriculture remains a key determinant of climate change (IPCC 2019), biodiversity loss, and associated reductions of ecosystem services (Lakner et al. 2019; Pe'er et al. 2019, 2022), while farmers' incomes highly depend on direct payments (Matthews 2016; Hill 2019).

Research can and should play a fundamental role in informing and supporting policymaking by providing the evidence necessary for the design of measures and programmes. Agricultural economists, among others, have developed a wide range of methods to generate scientific evidence to achieve policy objectives most efficiently. However, a prerequisite for evidence-based policymaking is that research provides the answers to the right questions at the right moment and in the right way, requiring that time and financial resources must be available and used throughout the entire policy cycle. In the best case, policy decisions should be closely based on rigorously established scientific evidence. Simultaneously, the extent to which research findings can actually influence political decision-making also depends on how actors use the findings. In addition to being end users, these actors may also be involved in the evaluation, itself, and the policy design to increase relevance and acceptance.

The goal of this paper is to put evidence-based (or evidence-informed) agriculture and food policies at the top of the research and policy agenda. We aim to provide insights into how scientific evidence should be used for policymaking as well as to give a coherent overview of ex-ante and ex-post evaluation methods and tools that agricultural economists use for evaluating agricultural policies to provide evidence for policy decisions. Our focus is on European policies and quantitative evaluation tools. We present insights into both established and new/emerging methods and approaches, as well as their advantages and disadvantages, and discuss their potential use for policy evaluation. In addition, we address the importance of stakeholder interests that influence the use of scientific evidence for policymaking. We conclude with an overview of papers in this special issue, titled 'Evidence-Based Agricultural and Food Policy: The Role of Research for Policy Making'.

The remainder of this paper is organized as follows. First, we present a background on the potential role of rigorous scientific evidence in European agricultural and food policies. Second, we discuss methods and approaches for ex-ante and ex-post policy assessments. We debate both currently widely used and under-utilized methods and approaches. Third, we cite data sources to provide policy-relevant evidence and discuss the need for new databases and enhanced data collection efforts. Fourth, we reveal how research should be developed to best support evidence-based policy. Finally, we provide an overview of the papers included in this special issue.

2. The role of research in policymaking is to provide evidence

Scientific evidence is particularly important in agricultural and food policies, as these policies are characterized by various (often conflicting) goals and rely on a wide spectrum of policy instruments. For instance, policies in Europe use different kinds of measures, for example, command and control instruments, information, and education, as well as economic instruments and combinations thereof, such as cross-compliance (Möhring et al. 2020). A key instrument for European agricultural policies is direct payments, which support food production and farmers' incomes (e.g. CAP Pillar I payments). Voluntary direct payment programmes support environmental and animal-friendly production (e.g. agri-environmental payments and organic farming subsidies). Market-regulating measures (e.g. tariffs) aim to protect farm-gate prices, and disincentive-based measures following the 'polluter pays' principle provide economic incentives to reduce negative environmental impacts (e.g. taxes on inputs). In addition, structural support (e.g. investment support) contributes to rural development or input regulation to reduce negative environmental impacts from production (e.g. cross-compliance obligations). This variety of measures and implementation modalities
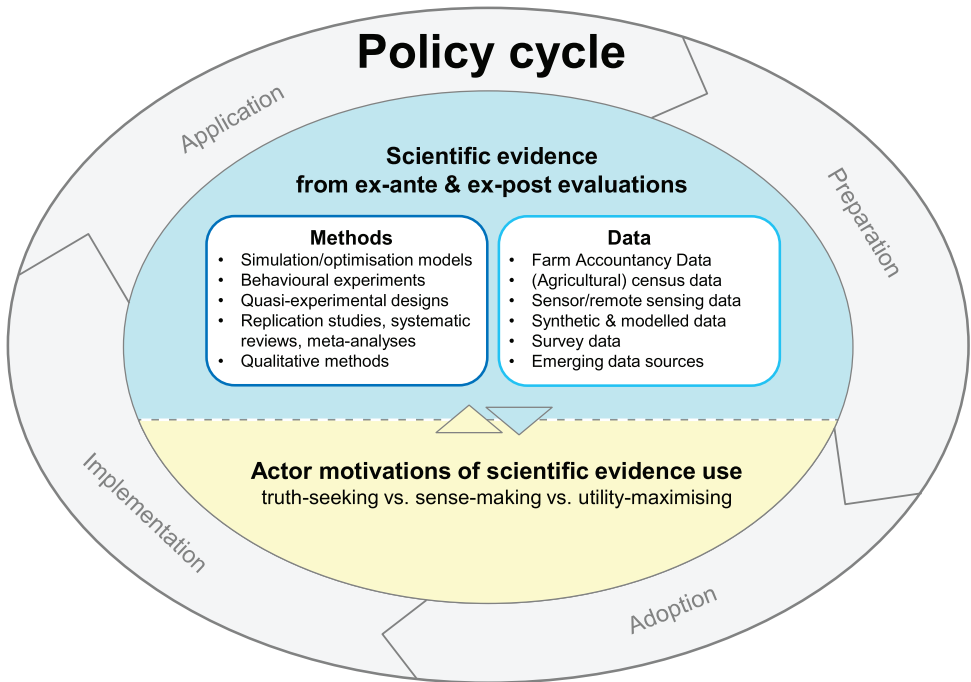
Figure 1. The role of research in policymaking. Phases of the policy cycle (it is also included in the uploaded document): policy design and preparation, adoption, implementation (transposition, complementary non-regulatory actions), application (including monitoring and enforcement), evaluation and revision (EC, 2021a).

requires a comprehensive portfolio of methods to analyse the impacts of agricultural policies and provide evidence for policymaking. To strengthen policy evaluation throughout the policy cycle, the EU's Better Regulation Agenda was introduced in 2015 (Listorti et al. 2020). According to the agenda, "Better regulation" refers to the Commission's regulatory policy, whereby it seeks to design and prepare EU policies and laws in such a way that they achieve their objectives in the most efficient way' (EC 2021a). Because evaluation[1] is among the key elements of regulatory policymaking (OECD 2018), the Better Regulation Guidelines are accompanied by a toolbox for ex-ante and ex-post evaluations (EC 2021b).

As shown in Fig. 1, researchers can provide evidence to policymakers in both ex-ante and ex-post evaluations. For example, this evidence can comprise insights on policy outcomes, impacts, and underlying behavioural changes (mechanisms), as well as provide insights for cost-benefit and cost-effectiveness analysis.[2] Methods for ex-ante evaluations include simulation and optimization models and behavioural experiments. Quasi-experimental designs, replication studies, systematic reviews, and meta-analyses are methods used in ex-post evaluations. Establishing a causal link between a specific policy and observed changes (i.e. attribution) requires quantitative analysis. However, when financial resources for an impact evaluation or the number of observations are limited, qualitative contribution analysis can be used to inform decisions (White and Raitzer 2017). Ultimately, combining insights into policy processes from qualitative analyses with the impact estimates from quantitative evaluations provides a comprehensive assessment for policymaking. The research uses and collects data for its evaluations. Commonly used data sets include farm accountancy data and agricultural census data, survey data, and synthetic and modelled data. In addition, digitalization is increasingly making new data sources available, such as high-resolution

remote sensing data, which will become more important for research in the future. Based on these data, farm- and sector-level indicators are developed to support the setup of monitoring systems and the design of new types of policy measures, such as payment schemes for environmental services (e.g. Latruffe et al. 2016; Poppe et al. 2016; Poppe and Vrolijk 2017; Elmiger et al. 2023; Gilgen et al. 2023a).

However, a broad range of scientific methods alone is not sufficient to inform policymaking. Notably, the impact of science depends on the demands of policymakers and practitioners, the evidence supplied by researchers, and the alignment between the two (McNie 2007). To ensure that scientific evidence generates real-world impacts on policymaking and to promote the uptake of measures by farmers, an actor-centred approach should be taken when constructing ex-ante and ex-post evaluations. Hofmann et al. (2022) distinguish three types of actors who influence the impact of scientific evidence on policymaking. The first is *truth-seeking actors*, who base their decisions on available scientific evidence to identify and select a pathway towards sustainable transformation. Science can support these actors by providing more and better evidence (Haas 2004; Montpetit and Lachapell 2015). The second type is *sense-making actors*, who integrate scientific evidence into their belief systems (Dewulf et al. 2020). Thus, the impact of science on preferences depends on whether the results of an evaluation match the demand, namely the beliefs and individual experiential knowledge of these actors (Raymond et al. 2010). The third type, *utility-maximizing actors*, uses scientific results strategically to pursue predefined interests, substantiate their preferences in political conflicts, and change others' perceptions (Weiss 1979; Choi et al. 2005).

For rigorous scientific evidence to actually affect policy designs and decisions, all three actor types must be considered. Evidence-based policy changes can only be expected if the strategic demand of these actors is met, that is, if evidence-based supply and demand matches and if actors are interested in the necessary transformation (Hofmann et al. 2022).

The actor-centred approach not only applies to research conducted on public policies but also to the evaluation of corporate policies and policy measures by non-governmental organizations and private donors. Researchers should factor this approach into their evaluations by engaging in knowledge co-production (Norström et al. 2020) and accessing multiple evidence bases that draw on different knowledge systems (Tengö et al. 2014). Although knowledge-production activities in transdisciplinary research span disciplinary boundaries and meaningfully involve non-academic partners in research design, operation, analysis, and publication, the integration of systematic policy evaluations and transdisciplinary research remains largely unexplored (O'Donovan et al. 2022).

In the next section, we present methods commonly used for ex-ante and ex-post policy evaluations and provide examples from evaluations conducted in various European countries.

3. Methods for ex-ante and ex-post policy evaluations to support policymaking

The majority of ex-ante and ex-post policy-evaluation methods are based on the establishment of a valid counterfactual to allow for causal interpretations of policy impacts. In this section, we discuss (i) ex-ante policy-evaluation approaches based on both simulations and economic experiments and (ii) ex-post evaluations based on quasi-experimental approaches and methods to synthesize results, such as meta-analyses, systematic reviews, and replication studies. Finally, we discuss (iii) qualitative methods for both ex-ante and ex-post assessments.

3.1 Ex-ante evaluations based on optimization and simulation models

Ex-ante policy assessments are used to provide guidance on the expected costs and benefits of different policy options and their redistributive impact.

Quantitative ex-ante assessments of European agricultural, environmental, and trade policy often rely on partial equilibrium models that operate at the national, continental, or global scales, such as the partial-equilibrium sector models CAPRI, for CAP Regionalized Impact (Britz and Witzke 2008; CAPRI model documentation 2022), Aglink-Cosimo that analyses supply and demand of world agriculture (Burrell and Nii-naate 2013), or the Global Biosphere Management Model (GLOBIOM), which allows users to consider the competition for land use between the agriculture, forestry, and bioenergy sectors as well as accounting for impacts on carbon stocks and greenhouse gas emissions (Havlík et al. 2018; Pastor et al. 2019). The individual models can also be combined. For instance, Latka et al. (2021) analysed the effectiveness of consumer-side interventions towards a more sustainable agri-food system using three different economic models, including CAPRI, GLOBIOM, and the Modular Applied GeNeral Equilibrium Tool, or MAGNET (Woltjer et al. 2014). MAGNET is a multi-regional, multi-sectoral, general-equilibrium input–output model that links industries across the agri-food value chain.

These models were initially designed to estimate supply responses to changing market interventions but not to capture behaviour and impacts at the farm level and thus often cannot reflect the emerging phenomena arising from a system of farms and farm-level decisions (Colen et al. 2016). To better represent farmers' responses to policy changes, sectoral models, such as CAPRI, have increasingly been disaggregated, for example, by refining spatial aggregation and going from global to farm-type scale (Gocht and Britz 2011; Gocht et al. 2017). Heterogeneous farm-specific behaviour, however, especially that related to enrolment in voluntary agri-environmental programmes, is not captured by these aggregated (farm-type scale) models. Along these lines, most of the models available for the ex-ante evaluation of the CAP are implemented at a regional level and thus cannot capture the impacts of policy measures that are farm- and site-specific (Louhichi et al. 2015).[3] But a better understanding of farmers' behaviour in response to agricultural-policy interventions has become much more important since the change from market to decoupled direct-payment support since agricultural-policy instruments are increasingly targeted towards specific policy goals and tailored to specific farms (Finger and El Benni 2021). The individual farm is thus the most relevant unit of decision-making, and farm-level models are crucial for ex-ante agricultural-policy evaluations (Reidsma et al. 2018). Several farm-level models are currently used for policy analysis in Europe (see Reidsma et al. 2018 for an overview). Modelling some individual farms may not always be sufficient to inform policymakers, however, because the ability to upscale the potential of findings from case studies to countries or regions is often limited.

Agent-based models (ABMs) are an option to overcome this limitation and to model farm-level decisions while still being able to infer sector-level outcomes. These models allow users to better represent farmers' responses to changing policy measures (Balmann 1997; Berger 2001). Some of the distinctive features of ABMs are the heterogeneity of the agent population (e.g. diverse farms, supply chain actors, and consumers) and the ability for spatial differentiation and the consideration of interactions or transactions between agents, as well as between agents and the landscape, such as those related to social networks, markets, or water management (Schreinemachers and Berger 2006; Nolan et al. 2009; Kremmydas et al. 2018; Huber et al. 2021). The demand for the development and use of ABMs are increasing. However, currently, ABMs are rarely used to evaluate the whole portfolio of agricultural policies, however, due to their specificity, for example, because they may have been developed for a particular purpose or for a particular spatial scale and are not easily adaptable for other applications and contexts (Louhichi et al. 2010, 2013; Ciaian et al. 2013; Grovermann et al. 2017; Kremmydas et al. 2018). One of the few exceptions for a national-level ABM is the Swiss agent-based recursive-dynamic sector model 'SWISSland', which has been the standard tool used for Swiss agricultural-policy impact assessments since 2011 (Möhring et al. 2016). Along these lines, the micromodel designed for the

ex-ante economic and environmental assessment of the medium-term adaptation of individual farmers to policy and market changes, known as IFM-CAP (Individual Farm Model for CAP Analysis), uses an individual farm-level simulation-model approach to overcome the shortcomings of agricultural-modelling tools for ex-ante evaluations of the CAP (Louhichi et al. 2015).

According to Reidsma et al. (2018), one limitation of most farm-level models (as well as those used in ABMs) is their reliance on programming approaches (i.e. the constrained model is used to attempt to find the optimal solution to simulate actual behaviour) that may have limited capacities to describe actual and potential farmer behaviour. Reidsma et al. (2018) highlight positive mathematical programming (where calibrated parameters are used to model actual behaviour) as a viable strategy to overcome this drawback in policy-impact assessments. But some linear-programming-based ABMs comprise calibration steps, and actual behaviour is simulated (Troost and Berger 2015). Many ABMs are also based on heuristics and not on optimization (Becu et al. 2008). One exception is the IFM-CAP, which is based on information from the Farm Accountancy Data Network (FADN) and is calibrated for an average of 3 years, using positive mathematical programming (Louhichi et al. 2015). In one of the few studies of this kind, Mack et al. (2019) showed how linear-optimization approaches can be combined with positive mathematical programming approaches in agent-based agricultural-sector models to improve their forecasting performance. Huber et al. (2018) reviewed 20 ABMs for their representation of decision-making processes and found considerable room for improvement by combining existing modelling approaches and promoting model inter-comparisons. Despite challenges, simulation models remain an important tool for ex-ante policy evaluation.

3.2 Ex-ante evaluations based on behavioural experiments

Behavioural experiments have a large potential to provide helpful ex-ante insights into the efficacy and efficiency of policies as well as into underlying behavioural mechanisms (Palm-Forster and Messer 2021). While such experiments can be less costly for ex-ante evaluation than developing tailored, complex modelling solutions, economic experimentation with farmers for policy analysis in Europe is still in its infancy (Thoyer and Préget 2019; Dessart et al. 2021).

In economic experiments, participants are randomly assigned to either the treatment (treated with a specific policy) or control (not treated with the policy) group, and decision-making is compared between both groups afterwards to assess the effect of the policy. Variants of a policy can also be tested by including different treatment arms. Economic experiments differ in terms of the subjects considered in agricultural-policy evaluations (e.g. students versus farmers), the environment in which the experiment takes place (lab, lab-in-the-field, or natural context), and the type of experimental setting, such as discrete choice, games, or real-world policy implementation (Colen et al. 2016). Below, we discuss three major forms of experiments: randomized control trials (RCTs), laboratory experiments, and field experiments.

RCTs are experiments conducted in real-world settings and are often considered the gold standard for measuring the net impact of a policy. But RCTs also come with limitations, including resource and time intensity, limited applicability for many real-life agricultural-policy settings, and various ethical considerations, including the acceptance of RCTs by farmers (Colen et al. 2016; Morawetz and Tribl 2020). The random assignment of participation, however, ensures that the observed characteristics of the treatment and control groups are the same before treatment (unobserved characteristics are assumed to be the same), thus avoiding selection bias (Colen et al. 2016).

RCTs are thus well suited for ex-ante assessment of the causal impact of specific policy programmes before they are scaled-up to the entire population, and they have been increasingly used to assess the piloting of innovative policy measures in Development Economics (Banerjee et al. 2016; Duflo 2020). RCTs can be used to evaluate specific policy measures (such as specific agri-environmental programmes) but are unsuitable for assessing broad policy reforms and are generally not applied to evaluate large-scale policy programmes, such as the entire CAP (Colen et al. 2016; Behaghel et al. 2019). Even though RCTs could be powerful instruments to assess the effectiveness of policies by randomly introducing potentially relevant variations in design (Behaghel et al. 2019), they are often difficult and costly to implement, especially if the aim is to analyse the reasons for behavioural change in addition to the overall impact of a policy programme.[4] Maintaining randomization for evaluating long-term outcomes and the need for more than one treatment arm in many evaluations are examples of such difficulties. As with several other evaluation designs, the issue of statistical power is crucial in RCTs. If the size of the sample is insufficiently large, then the impact evaluation will be underpowered, which is a particular issue with several treatment arms (White and Raitzer 2017). This situation may lead to a high risk of not being able to find a statistically significant effect, even though the policy is actually effective. Adaptive experimental designs can help to overcome the problem (Kasy and Sautmann 2021; Jobjörnsson et al. 2022). Another key concern is related to spill-overs between the treatment group and the control group. While clustered RCTs are often used to mitigate this risk, they also aggravate the problem of lower statistical power in settings with strong cluster heterogeneity.

Laboratory experiments are used to observe the behaviour of participants in a highly controlled environment isolated from nuisance factors; tasks and choices are usually formulated in an abstract way, thus allowing replication and enhancing internal validity (Lefebvre et al. 2021). For agricultural-policy evaluation, such experiments are mainly applied among university students (Le Coent et al. 2014; Schilizzi and Latacz-Lohmann 2016), which tends to limit the external validity of their results.

Field experiments with farmers are highly relevant economic experiments for agricultural policy evaluation (Lefebvre et al. 2021). The design of the experiment can differ substantially, depending on the evaluator's interest and whether hypothetical/stated preferences or experiment with incentives are used. For instance, discrete choice experiments can be used to provide farmers with a range of hypothetical policy scenarios, of which they will choose their preferred options (Birol and Koundouri 2008; Mariel et al. 2021). A framing of the decision context can also be used as a treatment in field experiments in which participants conduct specific tasks under different treatments with monetary incentives, such as policy designs (Hermann et al. 2017; Thomas et al. 2019; Dessart et al. 2021). Recruiting large and representative samples of farmers is often a challenge (Weigel et al. 2020), and also self-selection and evaluation bias remain limitations in field experiments, in particular when incentives are small and when treatments interact with characteristics that also affect selection into the study (Krawczyk 2011; Abeler and Nosenzo 2015). Furthermore, preferences for different policy designs may differ between the hypothetical experimental setup and the real world.

Experimental approaches can suffer from evaluation bias if participants are aware of the experiment and then change their behaviour because their decisions are being recorded. Changes in behaviour can occur for a variety of reasons, either because (1) participants want to manipulate outcomes, (2) the treatment group works harder than normal (the Hawthorne effect), (3) the control group starts competing with the treatment group (the John Henry effect), or (4) participants' perception of what the evaluator is trying to test lead to behavioural changes (Colen et al. 2016). Avoiding evaluation bias is more difficult in discrete-choice, lab, and field experiments than it is in RCTs (Colen et al. 2016).

3.3 Ex-post evaluations based on quasi-experiments

Ex-post evaluations allow policymakers to measure the net impact of a policy and establish the reasons for success or failure (Colen et al. 2016). This information then serves as the basis to abandon, adjust, or upscale policy measures. Various econometric methods may be used to identify causal effects by showing whether agricultural and food policies have had an impact and by quantifying the effects of policy interventions, for example, on production, incomes, prices, and environmental effects.

Quasi-experimental designs allow users to draw causal inferences based on observational data. The most common approach is to establish treatment and control groups in an ex-post manner. Baseline and endline data, collated before and after the intervention (respectively), or even more detailed time-series information, might be available for the two groups and can facilitate causal analysis, but such information is not mandatory.

Analysis may be performed by difference-in-difference, which compares the changes before and after the treatment between the treated and untreated units over the same time period (Imbens and Wooldridge 2009; Iacus et al. 2012, 2019; Chabé-Ferret 2015). Other approaches include regression-discontinuity designs (Cattaneo and Escanciano 2017; Wuepper and Finger 2022), synthetic controls (Abadie et al. 2015; Adhikari and Alm 2016), or instrumental variables (Angrist et al. 1996). These approaches can be combined with matching/weighting, which compares the outcomes of treated versus untreated units with the same observed characteristics. These methods are all suitable for ex-post agricultural-policy evaluations and have been applied in various studies on European policies (Chabé-Ferret and Subervie 2013; Mack and Kohler 2018; Bertoni et al. 2020; Wuepper et al. 2020, 2021; Grovermann et al. 2021; Wuepper and Huber 2021). Combinations of approaches are also increasingly common. For example, differences-in-discontinuity designs combine regression-discontinuity designs with difference-in-difference approaches (Wang et al. 2022; Wuepper and Finger 2022). Other combinations such as doubly robust difference-in-difference estimators (Sant'Anna and Zhao 2020) and regression-discontinuity-like designs are available (Cattaneo and Titiunik 2022), such as regression-kink designs (Card et al. 2015, 2017), bunching, and density discontinuities (Kleven 2016; Jales and Yu 2017; Blomquist et al. 2021), but they have yet to be used in European agricultural and food-policy evaluations.

In many quasi-experimental ex-post evaluation studies, several important aspects cannot be assessed due to a lack of data. Examples include (1) the implementation costs of a given policy, (2) the unintended and deadweight effects, and (3) interactions with other policies and their interference with the behaviour of involved and affected agents, such as farmers' adoption of an agri-environmental programme (Esposti and Sotte 2013). In addition, ex-post evaluations do not always provide sufficient insights into the underlying mechanisms of change or the often-complex interactions between policy measures and farmers' behaviour. The use of mixed-method approaches can close this gap by combining evaluation with in-depth interviews and other enquiry techniques that allow researchers to address 'how' and 'why' questions more profoundly and to help with data triangulation.

3.4 Ex-post evaluations based on replication studies, systematic reviews, and meta-analyses

Replication studies are a powerful but under-utilized approach to providing evidence for policymaking. Many studies in economics and policy evaluation are not replicable (Ferraro and Shukla 2022; Finger et al. 2023), which has ramifications for the significance, direction, and effect size reported in original studies. For example, Camerer et al. (2016) showed that 40 per cent of economic experiments published in top economic journals failed to be replicable. Only a small fraction of papers are actually ever replicated; one study found that only 0.1 per cent of papers published in the 50 leading journals in economics were replication

studies (Mueller-Langer et al. 2019). This situation is also important for ensuring credible scientific knowledge, which should be used to design effective and efficient agricultural, food and environmental policies (Ferraro and Shukla 2020). Policies might be initiated or abandoned based on studies where the true impacts are actually different or less reliable than reported (Ferraro and Shukla 2020). The replication of policy-evaluation studies thus can be used to verify results, reveal underlying uncertainties, or uncover errors, and thus provide a better scientific basis for policymakers. A special issue on 'Replications in Agricultural Economics' in *Applied Economic Perspectives and Policy* provides the first examples of relevant replication studies (Finger et al. 2023). Open research principles (e.g. open data, open code, and open access) are key to enabling a required shift towards a replication culture and increasing the usability of and trust in agricultural economic research.

Systematic reviews are an important tool for synthesizing the knowledge and scientific evidence about a specific research question (Page et al. 2021). For instance, the International Initiative for Impact Evaluation (known as '3ie') was established in 2008 to systematically synthesize rigorous evaluation evidence, with evidence-gap maps on various policy issues an important product. Some widely used guidelines for systematic reviews include the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Page et al. 2021). PRISMA allows researchers to coherently extract findings from existing research, critically assess underlying studies, and synthesize the knowledge base into an overarching conclusion. PRISMA also allows users to draw from a wide range of underlying studies, including those that use different methods for the same question, such as simulation models and experimental approaches.

Among others, exemplary questions addressed in systemic reviews in agricultural policies include the following:

1. Are European farmers risk averse? (Iyer et al. 2020);
2. What behavioural factors matter for farmers' decision-making towards sustainable farming practices? (Dessart et al. 2019);
3. How is sustainability measured in agriculture and policy settings? (Latruffe et al. 2016);
4. What biodiversity indicators are most effective in results-based agri-environmental schemes? (Elmiger et al. 2023).

Meta-analyses provide a systematic approach to empirically synthesize the results from multiple studies on the same question. Researchers can thus synthesize diverging effects and can consider any underlying uncertainties in individual studies. The combining of studies also allows researchers to overcome the lack of power of individual studies and identify any evidence gaps. The use of meta-analysis can increase their understanding of policy questions by integrating a large body of research focused on policy issues. Such analyses can provide combined effect sizes and combined significance levels for the joint outcomes of multiple studies, thus providing the most reliable estimates of the effects of policy actions. For example, several researchers have conducted meta-analyses, including Böcker and Finger's (2017) meta-analysis on pesticide-demand elasticities, Condon et al.'s (2015) meta-analysis of the impacts of ethanol policy on corn prices, and Santeramo and Lamonaca's (2019) meta-analysis on the effects of non-tariff measures on agri-food trade.

The transparent presentation of data and methods, as well as the accessibility of the original data used in individual studies (such as via data repositories), are all prerequisites to enable replication studies and meta-analyses while also supporting systematic reviews. This situation highlights the importance of having open research data for better policy designs (Nosek et al. 2015).

3.5 Qualitative approaches for ex-post and ex-ante policy evaluations

The focus of impact-evaluation approaches has mostly been on large-*n* statistical designs and attribution analyses (White and Raitzer 2017), which implies that a large number of observations are available for testing the statistical significance of the difference in outcomes between treatment and control groups. In some instances, only a small number of units are exposed to a policy, for instance in the case of regional regulations for selected communities or of financial support to specific enterprises. In other instances, resources are limited, the setting is highly complex, or the evaluation focus is on the change process. In these cases, qualitative approaches are useful to analyse the contribution of a policy to the outcomes of interest. Such methods often rely on systematic analyses of the theory of change and the impact pathway (Bamberger and Mabry 2019).

Qualitative comparative analysis is a methodological approach to contribution analysis that examines patterns in the data to identify necessary and sufficient conditions for relationships between interventions and outcomes without performing any tests of statistical significance (Ragin 2008; Pattyn et al. 2017). Other methods, such as 'most significant change' or 'outcome harvesting', support the reconstruction of the impact pathway to identify and assess policy outputs, outcomes, and impacts by considering the contributions of multiple stakeholders, programmes, and contextual factors (Alvarez et al. 2010; Blundo-Canto et al. 2017; Douthwaite et al. 2017). To trace the trajectory of outcomes, qualitative data-collection methods include key-informant interviews, focus groups, ethnographic research, or simulation games with participant observation, among others (Stern et al. 2012; Hennink et al. 2020).

Qualitative research also allows for inductive analysis (i.e. to go from specific cases to the general) and for generating hypotheses regarding the underlying mechanisms of policy impacts. Such research thus complements quantitative approaches, which are usually deductive and are used to test hypotheses. For instance, based on a qualitative multi-method research design and practice theory, Kaiser and Burger (2022) identified five types of crop-routinized crop-protection practices in Swiss agriculture, which implied different responses of farmers to current incentive-based agri-environmental policy instruments.

In addition to policy advice from either a quantitative or qualitative evaluation design, evidence-based decision-making can benefit from mixed-methods approaches. Such methods incorporate a diversity of values, allow for better validation of data and participatory-evaluation elements, and extend the comprehensiveness of evaluation findings through results from different methods that can then be used to broaden and deepen the understanding of policy impacts and impact pathways (Bamberger and Mabry 2019).

4. Data requirements

A wide range of micro-level data is required to provide scientific evidence for policymaking by ex-ante and ex-post evaluations, including information on economic, environmental, and social aspects, preferably at the farm level (Poppe and Vrolijk 2017). Farm-level information is crucial for assessing and monitoring the achievement of agricultural-policy goals. The farm as a unit of decision-making contributes to several functions of agriculture, including *economic*, through the production of goods and services, *ecological*, through the management of natural resources, and *social*, by contributing to rural dynamics (Latruffe et al. 2016).

To monitor and evaluate agricultural policy regarding the economic situation at the farm and sector level, farm-level data from the FADN is available for researchers and policymakers. Although attempts have been made to expand FADN data to include environmental and social performance indicators (Andersen et al. 2007; Latruffe et al. 2016; Poppe et al. 2016; Vrolijk et al. 2016), there is as yet no systematic collection of farm-level

indicators for comprehensive analysis and monitoring of the economic, environmental, and social functions of agriculture. Kelly et al. (2018) showed that the FADN in principle is a relevant but incomplete platform to assess farm-level sustainability across the EU; it is incomplete because using FADN data implies that financial variables serve as a proxy for environmental effects, such as the expenses of fertilizer and pesticides per hectare of land (Uehleke et al. 2019; Stetter et al. 2022). Such a proxy can only be a rough approximation of the environmental effects of the input-allocation decisions of farmers; for example, expenditures on fertilizer do not necessarily reflect the amount of fertilizer applied, and even less the nutrient losses to the environment. Along these lines, pesticide expenditures do not necessarily reflect the risk that pesticides pose to the environment (Möhring et al. 2019). Nevertheless, within a set of various indicators, such approximations can add useful information for policymaking (Kelly et al. 2018). The next step will be the expansion of the FADN towards the Farm Sustainability Data Network, or FSDN (Vrolijk and Poppe 2021).

Farm-level information can also be retrieved from agricultural census data, but such data mainly comprise restricted sets of indicators such as the type of agricultural production (e.g. the number of livestock units or hectares of crops grown), land-use-allocation decisions (e.g. the share of hectares devoted to specific crops), zone of production (e.g. the 'least favourite areas' or altitude), or participation in voluntary agri-environmental and animal welfare schemes. But census data usually do not comprise specific outcomes in the economic, environmental, animal welfare, or social dimensions. Other challenges are that census data are not always available for all countries for each year, and that the coverage of agricultural censuses can be geographically and/or statistically restricted (FAO 2021).

In general, there is a need to expand the in-situ monitoring of land use, biodiversity, ecosystem services, and human well-being (Pe'er et al. 2019). Besides the established and constantly improving monitoring systems used in Europe to assess the impact of agricultural policies, remote sensing and digitalization are improving the availability of high-resolution data. The increasing availability of high-quality data and the field's ever-increasing computing capacities have allowed a revolution in the application of data-intensive evaluation methods such as regression-discontinuity-like designs to large data sets. For example, high-resolution geospatial measurements of environmental, agricultural, or socio-economic variables using remote sensing or modelled data (including data generated with machine learning) have great potential for the evaluation of agricultural and food policies (Wuepper and Finger 2022; Jain 2020; Burke et al. 2021).

5. Aligning research for evidence-based policymaking

As summarized above, numerous methods are available for the evaluation of policy measures that are constantly being developed, and policymakers increasingly demand scientific evidence. But further efforts are necessary for evidence-based policymaking, both from science and from policy: While the further development and use of methods must be geared towards the issues relevant to policymaking, the different stakeholders affected by a policy measure must be involved in the design and implementation of new measures.

We propose five directions for the improvement of research towards evidence-based policymaking. First, the rigorous scientific methods chosen for policy evaluation and design need to be balanced with the relevance of the questions that can be answered by these methods. Second, methodological developments for policy evaluation and design should increasingly focus on the combination and triangulation of different methodological approaches. Third, capacities for evidence-based policy and professional practices must be increased by recognizing the functions of different actors (such as researchers, policymakers, interest groups, and civil society) and involving them in the policy cycle, according to their responsibilities. Fourth, new data sources should be developed to be able to provide the necessary scientific evidence. Fifth, the transparency how political decisions are made shall be improved.

5.1 Balance between scientifically rigorous methods and the practical relevance of questions

The political (or practical) relevance of the questions that can be answered and the rigour of the analytical method must be balanced, both in ex-ante and ex-post policy evaluations. More precisely, even though very robust ex-post evaluation methods are available from research for identifying causal relationships, the requirements on data quality and quantity are high and are not always found in practice. As a result, not all policy-relevant questions can be addressed with the most sophisticated methods. Quasi-experimental methods also frequently focus on one specific measure or variant of a policy measure, so ex-post evaluation designs are often characterized by strong internal validity, but extrapolating outcomes to a larger population (i.e. external validity) can be an issue. This situation can lead to a lack of appreciation of this type of analysis by political decision-makers, mainly because such analyses only cover a very specific topic, no insights on the mechanisms of change can be provided, and no interactions with other policy measures are captured. But such analyses can have great value for policymaking if their results are systematically summarized (e.g. within the framework of meta-analyses or replication studies) and thereby produce universal results for a wider population. In order to generalize results through systematic summaries, a good description of the respective contextual factors in the individual studies is needed.

One challenge in practice-relevant model-based ex-ante policy evaluations is that bundles of policy measures are often modelled or evaluated as a whole, and the effects of a single measure cannot always be separated from the effects of other measures. Even if the causal effects of individual measures thus cannot be analysed, the analysis of a bundle of measures is of great importance for policy design. Researchers thus should increasingly evaluate bundles of measures, in addition to rigorously identifying causal relationships.

5.2 Combination and triangulation of different methodological approaches

Both types of evaluation, ex-post and ex-ante, as well as both types of evidence, causal and contextual, must be part of the policy cycle and need to be better connected. To this end, any policy evaluation should be guided by a theory of change and should test the postulated relationships between interventions and outcomes.

Current developments of methods are still too often focused on individual methodological approaches, such as on quasi-experimental methods, especially with new econometric approaches, or, for example, on the designs of behavioural experiments. Ideally, further development should increasingly focus on bringing together different quantitative and qualitative methodological approaches, for instance, by integrating qualitative data on the change trajectory into quasi-experimental studies. This approach is important because the various objectives and measures of agricultural and food policy cannot be analysed with a single method alone. Researchers should strive for a mix of methods and data and not only apply various methodological approaches; instead, they should integrate these approaches into an overall design for policy evaluation by means of triangulation. As with the selection and use of individual methods for answering a policy-relevant research question, care must also be taken to ensure a transparent description of the chosen methodological approach when triangulating data and methods. This is particularly important because there may be variations in the results across the different methods, which can be a challenge in applying the findings in the policy process. However, as divergent results can lead to new and better explanations for the phenomenon under investigation (Tashakkori and Teddle 2003), they should be presented and discussed transparently. Moreover, combining methods can allow revealing insights into both aggregate effects as well as the underlying mechanisms.

In addition to combining qualitative and quantitative methods, ex-post and ex-ante methodological approaches should more often be combined for developing and adapting policy instruments and thus support policymaking (Finger et al. 2017). This combination is particularly relevant in evaluations of the adoption of new sustainable production methods in agriculture, which often take place in contexts of limited data and high uncertainty (Möhring et al. 2022). Different tools and data should be combined for holistic impact evaluations, for instance by integrating data from life-cycle assessment (Gaillard and Nemecek 2009), other sustainability assessment tools such as SMART (Schader et al. 2016, 2019) or TAPE (Mottet et al. 2020) or the use of environmental monitoring data (Gilgen et al. 2023a).[5] These strategies will allow the capture of trade-offs and synergies between a range of sustainability outcomes.

Filling gaps in the understanding of farmers' behaviour is also important in order to be able to include this information in, for example, model-based ex-ante policy evaluations. The combination of behavioural factors with bio-economic and agent-based modelling of agricultural production allows for the analysis of potential effects of policy measures on farmers' behaviour and the resulting policy outcomes and impacts (Reidsma et al. 2018; Huber et al. 2021).

5.3 Strengthen capacities for evidence-based policy and professional practices

Strengthening the science-policy interface is required, and an actor-centred approach is necessary to make scientific evidence generate real-world impacts and inform policymaking (see Fig. 1).

Researchers can apply different methods to promote the use of scientific evidence in policy and practice and inform decision-makers (Hofmann et al. 2022). For instance, where truth-seeking actors are constrained by evidence gaps, the generation and accumulation of evidence can be improved through more interdisciplinary collaboration, knowledge networking, and syntheses of current knowledge (Topping et al. 2020). For meaning-oriented actors, evidence can be made more relevant by increasingly co-producing knowledge in transdisciplinary projects between interconnected actors from science, policy, and practice, also taking into account experiential knowledge (Norström et al. 2020). In the case of benefit-maximizing actors, transparency requirements can limit the strategic or even misuse of evidence (Rohr 2021). The publication of all data collected with public funds can prevent the existence of private data monopolies.

The methodological approach and stakeholder interaction must also be aligned with the phase in the policy cycle. For instance, in the preparation phase, a theory of change should be formulated to clarify intended and possible unintended consequences and better contextualize the findings. Existing knowledge from past evaluations and analyses should also be increasingly used systematically, thereby avoiding duplication and allowing a more efficient use of the limited financial and human resources in research and in the political process. For example, a meta-analysis or systematic review can produce very good results without the need to adapt and extend models, which may be subject to uncertainties, especially when new policy instruments are evaluated. Formulating the theory of change should involve different stakeholders (such as practitioners, interest groups, and civil society) so that a comprehensive understanding of the expected direct and indirect impacts of a policy measure can be formed and to ensure the greatest possible benefits of policy changes. In general, as Pe'er et al. (2019: 451) note, 'monitoring and implementation processes should engage farmers, scientists, and citizens to better evaluate the impacts of interventions, to ensure delivery, and to promote societal inclusion, innovation, and adaptation management'.

During the adoption and implementation phases (and besides ex-ante modelling exercises), the joint design of economic experiments for ex-ante impact evaluations can be valuable to adapt and refine policy instruments. In this context, we should consider that this process takes time and that this time requirement must be explicitly integrated into the policy cycle. Assessments made in real-world policy advice settings often have limited or even no time for scientifically sound analyses. Ad hoc evaluation mandates frequently prevent a systematic and rigorous scientific approach, with corresponding implications for the significance of the results. Instead, we need forward-looking and future-oriented research that anticipates future developments and develops proposals for the attention of political stakeholders, who in turn can consolidate this scientific evidence in a systematic process with other stakeholders and finalize them for implementation. A critical reflection is also needed how and by whom the transfer of scientific knowledge takes place in governmental and policy processes, e.g. ministries, political parties, or entities at the science-policy interface such as think tanks, and what useful interaction and evidence means for these different actors. Moreover, science needs a careful reflection of its role in the political process. For example, science as an honest knowledge broker provides evidence on advantages and disadvantages of possible policy choices while acknowledging that the actual policy decisions are made by policy makers and not by scientists themselves.

5.4 Analyse data from different sources with adequate methods

The availability of farm-level data is a key aspect for understanding the inter-linkages between agricultural policies, farmers' decision-making, and natural production conditions, and thus is a key for designing and refining agricultural and food policies based on rigorous ex-ante and ex-post evaluations.

In general, farm-level monitoring tools should allow stakeholders to empirically document trends in a way in which developments can be attributed to the relevant policies and separated from other influences (Poppe et al. 2016). But while the economic function of agriculture has always been an important objective of agricultural policies that aim to ensure a sufficient food supply and an adequate income for farmers (Finger and El Benni 2021), the environmental, animal welfare, and social dimensions are underrepresented in current indicator sets and monitoring systems and should be expanded. Data collections such as FADN must consider the changing data needs driven by changes of agricultural-policy goals and instruments.

There is hope that these data gaps can be closed in the future, as various emerging data sources are currently under-used. These sources comprise geospatial and remote-sensing data, such as on land use and land-use intensity (Ehlers et al. 2021; Wuepper and Finger 2022); data from sensors on machinery, such as on yields and inputs usage (Finger et al. 2019); and data from online sources extracted by web scraping (Hillen 2019), including data from social media and Google Trends, such as on societal concerns with agriculture (Schaub et al. 2020). To leverage data opportunities, a push towards open research data as well as open governmental data is required. With the dramatic increase in data availability, machine-learning approaches have also become more important for agricultural economists by helping to exploit large volumes of data more efficiently than traditional statistical methods allow (Storm et al. 2020). In policy evaluations, skills and capabilities are frequently neglected intermediate outcomes (O'Donovan et al. 2022). To assess these outcomes as part of policy-impact studies, indicators should be developed that can capture hard and soft skill development. The emerging field of behavioural experiments can contribute to filling this gap.

5.5 Improving the transparency and understanding of how political decisions are made

A major challenge for science to contribute to policymaking is the political process. Even when scientific evidence is available, it does not necessarily inform policy decisions. This is because strong political pressure can maintain the status quo and block policy changes (Swinnen 2018). For instance, by providing more and better evidence, science can support *truth-seeking* and *sense-making actors*. However, *utility-maximizing actors* will only use scientific evidence if it supports their own concerns (Hofmann et al. 2022). Even if researchers can hardly influence actors to use scientific evidence, they can scientifically accompany the policy-making process and create transparency about how decisions are made. The involvement of different actors throughout the policy cycle and the co-production of knowledge can lead to the actors recognizing the mutual benefits of the policy changes being discussed. This enables the negotiation of agreements that lead to policy changes. As the negotiation of agreements is a promising path to policy change (Metz et al. 2021), the scientific analysis of the policy process is also an interesting research topic regarding agricultural policy.

6. Papers in this special issue

Based on this background, the papers in this special issue present a range of methods and data used for ex-ante and ex-post analyses of agricultural and food policies. The papers will add to the literature by (1) showing how different data sources and indicators can be used to analyse agricultural and food policies and (2) how ex-post and ex-ante assessment methods can contribute to the understanding, design, and refinement of agricultural and food policies.

In a behavioural experiment in Germany, Rommel et al. (2023) show that the willingness of farmers to cooperate in the collective implementation of agri-environmental measures was higher than experts expected. Similar studies are to be conducted in the Netherlands, Hungary, and Poland, thus creating broad-based knowledge in different contexts on how measures need to be designed in order to effectively and efficiently achieve the environmental goals of agricultural policy.

Using a quasi-experimental approach with a spatial-regression-discontinuity design, Zimmert and Zorn (2023) show that direct payments increase family farm employment in Switzerland. The analysis points to not only economic but also social side-effects of the current direct-payment system because the additional labour force often consists of non-salaried female household members. Without a wage, these family members are insufficiently protected socially, an issue that should gain importance in discussions of the further development of agricultural policy.

Fedoseeva and Irek (2023) show the added value of using online price data to obtain spatial and temporal market information on, in this case, food prices. In some countries, automated surveys of online prices are already used to measure inflation, for example, in Switzerland for books and clothing. Compared to the classic consumer price indices, which are usually calculated on a monthly basis, the advantage of this approach is that data are available in real time, and price changes are recorded in a more differentiated manner by product, location, and sales channel due to the large database (Cavallo and Rigobon 2016). In exceptionally dynamic times caused by pandemics or war, such real-time evidence can be valuable for policymakers.

Combining an ex-ante bio-economic modelling analysis with an ex-post econometric analysis based on survey data, Möhring et al. (2023) show that this combination has led to a better understanding of the adoption decisions of different types of farmers and provides stakeholders with precise information on the design and implementation of policy measures geared towards more sustainable production. They also show that the value of the different

analyses differs, depending on the project stage and the round of the policy cycle, and that the choice of the model type and aggregation level in the ex-ante assessment is crucial to generating synergies in later policy cycles.

In an Italian case study, Santeramo et al. (2023) investigate whether and to what extent there are inter-linkages between a public policy reform (namely changes in the subsidization of insurance contracts) and a private-sector reform (changes in the types of insurance contracts). Even if no causal link between farmers' behavioural changes and a specific policy could be demonstrated, this study exemplifies the challenges of evaluating policies when behavioural changes are the result of different policy and market changes. Providing contextual information in such studies is particularly important in order to yield potentially generalizable information for policymakers, for example, in meta-analyses.

Bystricky et al.'s (2023) paper shows how model-based ex-ante evaluations can contribute to the design of agricultural policy measures; the authors use as an example a currently pending revision of a voluntary agri-environmental programme in Switzerland. Applying the aforementioned ABM SWISSland, supplemented with data from life-cycle assessment and further environmental indicators, the authors evaluate the direct and indirect effects at the national scale of a reduction in protein-reduced-concentrate use in roughage-based dairy and meat production. Since the evaluation shows that feed competition will be exacerbated by the proposed design of the measures and that no improvement in environmental impacts is to be expected, the measure will not be implemented as planned, but further design options will be examined.

In a Swiss case study commissioned by Switzerland's Federal Office for Agriculture, Gilgen et al. (2023b) show how environmental indicators can be used for the development of indicator-based agri-environmental payment programmes that better take into account environmentally relevant farm structures, as compared to current agri-environmental programmes. The authors analyse the effectiveness of the proposed indicator-based policy using SWISSland, but their results show hardly any effect on the environmental performance of the sector due to the manifold interactions with other existing direct-payment programmes.

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

End Notes

1 The terms 'impact assessment' and 'evaluation' are used in OECD and EC documents for what we call 'ex-ante' and 'ex-post evaluations', respectively.

2 Cost-effectiveness would be required if a policy goal (e.g. a specific reduction target of pesticide risks) shall be achieved with the lowest possible costs; cost-benefit analysis can provide insights for prioritising areas for policy intervention and measures.

3 According to Louhichi et al. (2015), only two EU-based representative farm models include full EU coverage: the Common Agricultural Policy Regional Impact Assessment Farm Type, or CAPRI-FT

(Gocht and Britz 2011; Gocht et al. 2013) and Agriculture, Recomposition de l'Offre et Politique Agricole, or AROPAj (De Cara and Jayet 2011).

4 Other experimental setups are available to researchers who wish to learn about the reasons for behavioural change. Choice experiments, for example, provide relatively cheap and rapidly obtained information on potential ways to improve a policy (Colen et al. 2016), with the trade-off that internal validity (i.e. the isolation of the causal relationship) is more challenging to maintain (Harrison and List 2004).

5 SMART and TAPE refer to 'Sustainability Monitoring and Assessment Routine' and 'Tool for Agroecology Performance Evaluation', respectively.

References

Abadie A., Diamond A. and Hainmueller J. (2015) 'Comparative Politics and the Synthetic Control Method', *American Journal of Political Science*, 59: 495–510.

Abeler J. and Nosenzo D. (2015) 'Self-Selection into Laboratory Experiments: Pro-Social Motives versus Monetary Incentives', *Experimental Economics*, 18: 195–214.

Adhikari B. and Alm J. (2016) 'Evaluating the Economic Effects of Flat Tax Reforms using Synthetic Control Methods', *Southern Economic Journal*, 83: 437–63.

Alvarez S., Douthwaite B., Thiele G., Mackay R., Córdoba D. and Tehelen K. (2010) 'Participatory Impact Pathways Analysis: A Practical Method for Project Planning and Evaluation', *Development in Practice*, 20: 946–58.

Andersen E., Elbersen B., Godeschalk F. and Verhoog D. (2007) 'Farm Management Indicators and Farm Typologies as A Basis for Assessments in A Changing Policy Environment', *Journal of Environmental Management*, 82: 353–62.

Angrist J.D., Guido W.I. and Rubin D.B. (1996) 'Identification of Causal Effects using Instrumental Variables', *Journal of the American Statistical Association*, 91: 444–55.

BAFU/BLW (2016). *Umweltziele Landwirtschaft: Statusbericht 2016 /Environmental Goals Agriculture: Status Report 2016*, pp. 1–116. Bern: Bundesamt für Umwelt (BAFU) und Bundesamt für Landwirtschaft (BLW)/Swiss Federal Office for the Environment (FOEN) and Swiss Federal Office for Agriculture (FOAG).

Balmann A. (1997) 'Farm-Based Modelling of Regional Structural Change: A Cellular Automata Approach', *European Review of Agricultural Economics*, 24: 85–108.

Bamberger M. and Mabry L. (2019). *RealWorld Evaluation: Working under Budget, Time, Data, and Political Constraints*. Thousand Oaks, CA: Sage Publishing.

Banerjee A.V., Duflo E. and Kremer M. (2016) 'The Influence of Randomized Controlled Trials on Development Economics Research and on Development Policy', *The State of Economics, the State of the World Conference*, pp. 482–8, MIT Press.

Becu N., Neef A., Schreinemachers P. and Sangkapitux C. (2008) 'Participatory Computer Simulation to Support Collective Decision-Making: Potential and Limits of Stakeholder Involvement', *Land Use Policy*, 25: 498–509.

Behaghel L., Macours K. and Subervie J. (2019) 'How can Randomized Controlled Trials Help Improve the Design of the Common Agricultural Policy?', *European Review of Agricultural Economics*, 46: 473–93.

Berger T. (2001) 'Agent-Based Spatial Models Applied to Agriculture: A Simulation Tool for Technology Diffusion, Resource use Changes and Policy Analysis', *Agricultural Economics*, 25: 245–60.

Bertoni D., Curzi D., Aletti G. and Olper A. (2020) 'Estimating the Effects of Agri-Environmental Measures Using Difference-In-Difference Coarsened Exact Matching', *Food Policy*, 90: 101790.

Birol E. and Koundouri P. (2008) 'Choice Experiments Informing Environmental Policy: A European Perspective', in Birol, E., Koundouri, P. (eds), *New Horizons in Environmental Economics*, pp. 1–12. Cheltenham and Northampton, MA: Edward Elgar.

Blomquist S., Newey W.K., Kumar A. and Liang C.Y. (2021) 'On Bunching and Identification of the Taxable Income Elasticity', *Journal of Political Economy*, 129: 2320–43.

Blundo-Canto G., Läderach P., Waldock J. and Camacho K. (2017) 'Learning Through Monitoring, Evaluation and Adaptations of the 'Outcome Harvesting' Tool', *Cahiers Agricultures*, 26: 65004.

Böcker T. and Finger R. (2017) 'A Meta-Analysis on the Elasticity of Demand for Pesticides', *Journal of Agricultural Economics*, 68: 518–33.

Britz, W., Witzke, P., eds (2008) *CAPRI Model Documentation 2008: Version 2*. Bonn: Institute for Food and Resource Economics, University of Bonn.

Burke M., Driscoll A., Lobell D.B. and Ermon S. (2021) 'Using Satellite Imagery to Understand and Promote Sustainable Development', *Science*, 371: 6535.

Burrell A. and Nii-naate Z. (2013) 'Partial Stochastic Analysis with the European Commission's Version of the AGLINK-COSIMO Model', 46. JRC Reference Reports. EUR 25898, Luxembourg: Publications Office of the European Union, JRC76019, DOI: 10.2791/87727.

Bystricky M., Bretscher D., Schori F. and Mack G. (2023) 'Reducing Feed-Food Competition with Direct Payments? An Ex-Ante Assessment of Economic and Environmental Impacts', *Q Open* (in revision).

Camerer C. F., Dreber A., Forsell E., Ho T. H., Huber J., Johannesson M., Kirchler M., Almenberg J., Altmejd A., Chan T., Heikensten E., Holzmeister F., Imai T., Isaksson S., Nave G., Pfeiffer T., Razen M. and Wu H. (2016) 'Evaluating Replicability of Laboratory Experiments in Economics', *Science*, 351: 1433–6.

CAPRI Model Documentation (2022). Version 18/01/2022, fetch.php (capri-model.org); capri: docs[CAPRI]; (capri-model.org), last accessed 10 September 2022.

Card D., Lee D.S., Pei Z. and Weber A. (2015) 'Inference on Causal Effects in A Generalized Regression Kink Design', *Econometrica*, 83: 2453–83.

——— (2017) 'Regression Kink Design: Theory and Practice', in Cattaneo, M.D., Escanciano, J.C. (eds), *Regression Discontinuity Designs: Theory and Applications*, pp. 341–82. Bingley: Emerald.

Cattaneo, M.D., Escanciano, J.C. eds (2017). *Regression Discontinuity Designs: Theory and Applications*. Bingley: Emerald.

Cattaneo M.D. and Titiunik R. (2022) 'Regression Discontinuity Designs', *Annual Review of Economics*, 14: 821–51.

Cavallo A. and Rigobon R. (2016) 'The Billion Prices Project: Using Online Prices for Measurement and Research', *Journal of Economic Perspectives*, 30: 151–78.

Chabé-Ferret S. (2015) 'Analysis of the Bias of Matching and Difference-In-Difference Under Alternative Earnings and Selection Processes', *Journal of Econometrics*, 185: 110–23.

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: 12–27.

Choi B.C.K., Pang T., Lin V., Puska P., Sherman G., Goddard M., Ackland M.J., Sainsbury P., Stachenko S., Morrison H. and Clottey C. (2005) 'Can Scientists and Policy Makers Work Together?', *Journal of Epidemiology & Community Health*, 59: 632–7.

Ciaian P., Espinosa Goded M., Gomez y Paloma S., Heckelei T., Sckokai P., Langrell S., Elouhichi K., Thomas A. and Vard T. Langrell, S. (ed.). (2013). *Farm Level Modelling of CAP: A Methodological Review*. EUR 25873. Luxembourg: Publications Office of the European Union. JRC79969.

Colen L., Gomez y Paloma S., Latacz-Lohmann U., Lefebvre M., Préget R. and Thoyer S. (2016) 'Economic Experiments as a Tool for Agricultural Policy Evaluation: Insights from the European CAP', *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 64: 667–94.

Condon N., Klemick H. and Wolverton A. (2015) 'Impacts of Ethanol Policy on Corn Prices: A Review and Meta-Analysis of Recent Evidence', *Food Policy*, 51: 63–73.

De Cara S. and Jayet P.A. (2011) 'Marginal Abatement Costs of Greenhouse Gas Emissions from European Agriculture, Cost Effectiveness, and the EU Non-ETS Burden Sharing Agreement', *Ecological Economics,* 70: 1680–90.

Dessart J.F., Barreiro-Hurlé J. and van Bavel R. (2019). Behavioural Factors Affecting the Adoption of Sustainable Farming Practices: A Policy-Oriented Review, *European Review of Agricultural Economics* 46: 417–71.

Dessart F., Rommel J., Barreiro-Hurlé J., Thomas F., Rodríguez-Entrena M., Espinosa-Goded M., Zagórska K., Czajkowski M. and van Bavel R. (2021). *Farmers and the New Green Architecture of the EU Common Agricultural Policy: A Behavioural Experiment*. EUR 30706 EN. Luxembourg: Publications Office of the European Union. JRC123832. DOI: 10.2760/718383.

Dewulf A., Klenk N., Wyborn C. and Lemos M.C. (2020) 'Usable Environmental Knowledge from the Perspective of Decision-Making: the Logics of Consequentiality, Appropriateness, and Meaningfulness', *Current Opinion in Environmental Sustainability* 42: 1–6.

Douthwaite B., Mayne J., McDougall C. and Paz-Ybarnegaray R. (2017) 'Evaluating Complex Interventions: A Theory-Driven, Realist-Informed Approach', *Evaluation*, 23: 294–311.

Duflo E. (2020) 'Field Experiments and the Practice of Policy', *American Economic Review*, 110: 1952–73.

EC [European Commission] (2021a). *Better Regulation Guidelines*. SWD(2021) 305 final, pp. 1–43. Brussels: EC.

——— (2021b). *Better Regulation Toolbox*. November 2021 edition, pp. 1–608. Brussels: EC, br_toolbox-nov_2021_en_0.pdf (europa.eu).

Ehlers M.-H., Huber R. and Finger R. (2021) 'Agricultural Policy in the Era of Digitalisation', *Food Policy*, 100: 102019.

El Benni N. and Lehmann B. (2010) 'Swiss Agricultural Policy Reform: Landscape Changes in Consequence of National Agricultural Policy and International Competition Pressure', in Primdahl, J., Swaffield, S. (eds), *Globalisation and Agricultural Landscapes: Change Patterns and Policy Trends in Developed Countries*, pp. 73–94. Cambridge: Cambridge University Press.

El Benni N. and Schmid D. (2022) 'Off-Farm Income and Direct Payments: An Indispensable Diversification Strategy of Swiss Farmers', *Q Open* 2: qoab019.

Elmiger N., Finger R., Ghazoul J. and Schaub S. (2023) 'Biodiversity Indicators for Result-Based Agri-Environmental Schemes: Current State and Future Prospects', *Agricultural Systems*, 204: 103538.

Esposti R. and Sotte F. (2013) 'Evaluating the Effectiveness of Agricultural and Rural Policies: An Introduction', *European Review of Agricultural Economics*, 40: 535–9.

FAO [Food and Agriculture Organization of the United Nations] (2021). *Global Review of Agricultural Census Methodologies and Results (2006–2015). World Programme for the Census of Agriculture 2010*. FAO Statistical Development Series No. 18. Rome: FAO. https://doi.org/10.4060/cb2650en.

Fedoseeva S. and Irek J. (2023) 'Within-Retailer Price Dispersion in E-Commerce: Prevalence, Magnitude, and Determinants', *Q Open*, qoac021, https://doi.org/10.1093/qopen/qoac021.

Ferraro P. J. and Shukla P. (2020) 'Feature: Is A Replicability Crisis on the Horizon for Environmental and Resource Economics?', *Review of Environmental Economics and Policy*, 14: 339–51.

——— (2022) 'Credibility Crisis in Agricultural Economics', *Applied Economic Perspectives and Policy*, 14: 339–51. https://doi.org/10.1002/aepp.13323.

Finger R. and El Benni N. (2021) 'Farm Income in European Agriculture: New Perspectives on Measurement and Implications for Policy Evaluation', *European Review of Agricultural Economics*, 48: 253–65.

Finger R., Grebitus C. and Henningsen A. (2023) 'Replications in Agricultural Economics', *Applied Economics Policy and Perspectives* (submitted).

Finger R., Listorti G. and Tonini A. (2017) 'The Swiss Payment for Milk Processed into Cheese: Ex-Post and Ex-Ante Analysis', *Agricultural Economics*, 48: 437–48.

Finger R., Swinton S., El Benni N. and Walter A. (2019) 'Precision Farming at the Nexus of Agricultural Production and the Environment', *Annual Review of Resource Economics*, 11: 313–35.

Gaillard G. and Nemecek T. (2009) 'Swiss Agricultural Life Cycle Assessment (SALCA): An Integrated Environmental Assessment Concept for Agriculture', *Integrated Assessment of Agriculture and Sustainable Development: Setting the Agenda for Science and Policy (AgSAP) Conference*, pp. 134–5. Egmond aan Zee: AgSAP Office, Wageningen University.

Gilgen A., Blaser S., Schneuwly J., Liebisch F. and Merbold L. (2023a) 'The Swiss Agri-Environmental Data Network (SAEDN): Description and Critical Review of the Dataset', *Agricultural Systems*, 205: 103576. https://doi.org/10.1016/j.agsy.2022.103576.

Gilgen A., Drobnik T., Mann S., Flury C., Mack G., Ritzel C., Roesch A. and Gaillard G. (2023b) 'Can Agricultural Policy Achieve Environmental Goals Through an Indicator-Based Direct Payment System?', *Q Open*, qoac034, https://doi.org/10.1093/qopen/qoac034.

Gocht A. and Britz W. (2011) 'EU-Wide Farm Type Supply Models in CAPRI: How to Consistently Disaggregate Sector Models into Farm Type Models', *Journal of Policy Modeling* 33: 146–67.

Gocht A., Britz W., Ciaian P. and Gomez y Paloma S. (2013) 'Farm Type Effects of an Eu-Wide Direct Payment Harmonisation', *Journal of Agricultural Economics*, 64: 1–32.

Gocht A., Ciaian P., Bielza M., Terres J.-M., Röder N., Himics M. and Salpultra G. (2017) 'EU-Wide Economic and Environmental Impacts of CAP Greening with High Spatial and Farm-Type Detail', *Journal of Agricultural Economics*, 68: 651–81.

Grovermann C., Quiédeville S., Muller A., Leiber F., Stolze M. and Moakes S. (2021) 'Does Organic Certification Make Economic Sense for Dairy Farmers in Europe? A Latent Class Counterfactual Analysis', *Agricultural Economics*, 52: 1001–12.

Grovermann C., Schreinemachers P. and Berger T. (2017) "'Smart' Policies to Reduce Pesticide use and Avoid Income Trade-Offs: An Agent-Based Model Applied to Thai Agriculture', *Ecological Economics*, 132: 91–103.

Haas P. (2004) 'When Does Power Listen to Truth? A Constructivist Approach to the Policy Process', *Journal of European Public Policy*, 11: 569–92. https://doi.org/10.1080/1350176042000248034.

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

Havlík P., Valin H., Mosnier A., Frank S., Lauri P., Leclère D., Palazzo A., Batka M., Boere E., Brouwer A., Deppermann A., Ermolieva T., Forsell N., di Fulvio F. and Obersteiner M. (2018) 'GLO-BIOM Documentation', *International Institute for Applied Systems Analysis (IIASA) Conference*, pp. 1–38.

Hennink M., Hutter I. and Bailey A. (2020). *Qualitative Research Methods*. Thousand Oaks, CA: Sage Publishing.

Hermann D., Sauthoff S. and Musshoff O. (2017) 'Ex-Ante Evaluation of Policy Measures to Enhance Carbon Sequestration in Agricultural Soils', *Ecological Economics*, 140: 241–50.

Hill B. (2019). *Farm Incomes, Wealth and Agricultural Policy*. London: Routledge.

Hillen J. (2019) 'Web Scraping for Food Price Research', *British Food Journal*, 121: 3350–61.

Hofmann B., Ingold K., Stamm C., Ammann P., Eggen R.I.L., Finger R., Fuhrimann S., Lienert J., Mark J., McCallum C., Probst-Hensch N., Reber U., Tamm L., Wiget M., Winkler S., Zachmann L. and Hoffmann S. (2022) 'Barriers to Evidence use for Sustainability: Insights from Pesticide Policy and Practice', *Ambio*, 52: 425–39.

Huber R., Bakker M., Balmann A., Berger T., Bithell M., Brown C., Grêt-Regamey A., Xiong H., Le Q.B., Mack G., Meyfroidt P., Millington J., Müller B., Polhill J.G., Sun Z., Seidl R., Troost C. and Finger R. (2018) 'Representation of Decision-Making in European Agricultural Agent-Based Models', *Agricultural Systems*, 167: 143–60.

Huber R., Xiong H., Keller K. and Finger R. (2021) 'Bridging Behavioural Factors and Standard Bio-Economic Modelling in an Agent-Based Modelling Framework', *Journal of Agricultural Economics*, 73: 35–63.

Iacus S.M., King G. and Porro G. (2012) 'Causal Inference Without Balance Checking: Coarsed Exact Matching', *Political Analysis*, 20: 1–24.

——— (2019) 'A Theory of Statistical Inference for Matching Methods in Causal Research', *Political Analysis*, 27: 46–68.

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

IPCC [Intergovernmental Panel on Climate Change] (2019). *Summary for Policy Makers. Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*. Geneva: IPCC. https://bit.ly/2U1gzza.

Iyer P., Bozzola M., Hirsch S., Meraner M. and Finger R. (2020) 'Measuring Farmer Risk Preferences in Europe: A Systematic Review', *Journal of Agricultural Economics*, 71: 3–26.

Jain M. (2020) 'The Benefits and Pitfalls of Using Satellite Data for Causal Inference', *Review of Environmental Economics and Policy* 14: 157–69.

Jales H. and Yu Z. (2017) 'Identification and Estimation using a Density Discontinuity Approach', in Cattaneo, M.D., Escanciano, J.C. (eds), *Regression Discontinuity Designs: Theory and Applications*, pp. 29–72. Bingley: Emerald.

Jobjörnsson S., Schaak H., Musshoff O. and Friede T. (2022). 'Improving the Statistical Power of Economic Experiments using Adaptive Designs', *Experimental Economics*, https://doi.org/10.1007/s10683-022-09773-8.

Kaiser A. and Burger P. (2022) 'Understanding Diversity in Farmers' Routinized Crop Protection Practices', *Journal of Rural Studies*, 89: 149–60.

Kasy M. and Sautmann A. (2021) 'Adaptive Treatment Assignment in Experiments for Policy Choice', *Econometrica* 89: 113–32.

Kelly E., Latruffe L., Desjeux Y., Ryan M., Uthes S., Diazabakana A., Dillon A. and Finn J. (2018) 'Sustainability Indicators for Improved Assessment of the Effects of Agricultural Policy Across the EU: Is FADN the Answer?', *Ecological indicators*, 89: 903–11.

Kleven H.J. (2016) 'Bunching', *Annual Review of Economics*, 8: 435–64.

Krawczyk M. (2011) 'What Brings Your Subjects to the Lab? A Field Experiment', *Experimental Economics*, 14: 482–9.

Kremmydas D., Athanasiadis I.N. and Rozakis S. (2018) 'A Review of Agent Based Modeling for Agricultural Policy Evaluation', *Agricultural Systems*, 164: 95–106.

Lakner S., Holst C., Dittrich A., Hoyer C. and Pe'er G. (2019) 'Impacts of the EU's Common Agricultural Policy (CAP) on Biodiversity and Ecosystem Services', in Schröter, M., Bonn, A., Klotz, S., Seppelt, R., and Baessler, C. (eds), *Atlas of Ecosystem Services: Drivers, Risks, and Societal Responses*, pp. 383–9. Cham: Springer

Latka C., Kuiper M., Frank S., Heckelei T., Havlík P., Witzke H.P., Leip A., Cui H.D., Kuijsten A., Geleijnse J.M. and van Dijk M. (2021) 'Paying the Price for Environmentally Sustainable and Healthy EU Diets', *Global Food Security*, 28: 100437.

Latruffe L., Diazabakana A., Bockstaller C., Desjeux Y., Finn J., Kelly E., Ryan M. and Uthes S. (2016) 'Measurement of Sustainability in Agriculture: A Review of Indicators', *Studies in Agricultural Economics*, 118: 123–30.

Le Coent P., Thoyer S. and Préget R. (2014) 'Why Pay for Nothing? An Experiment on a Conditional Subsidy Scheme in a Threshold Public Good Game', *Conférence annuelle de la FAERE (French Association of Environmental and Resource Economists)*. Montpellier, France. 13: ffhal-02739425.

Lefebvre M., Barreiro-Hurlé J., Blanchflower C., Colen L., Kuhfuss L., Rommel J., Sumrada T., Thomas F. and Thoyer S. (2021) 'Can Economic Experiments Contribute to a more Effective CAP?', *EuroChoices*, 20: 42–9.

Listorti G., Basyte-Ferrari E., Acs S. and Smits P. (2020) 'Towards an Evidence-Based and Integrated Policy Cycle in the EU: A Review of the Debate on the Better Regulation Agenda', *Journal of Common Market Studies*, 58: 1558–77.

Louhichi K., Ciaian P., Espinosa Goded M., Colen L., Perni Llorente A. and Gomez y Paloma S. (2015). *An EU-Wide Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP)*. EUR 26910. Luxembourg: Publications Office of the European Union. JRC92574.

Louhichi K., Espinosa M., Ciaian P. and Gomez y Paloma S. (2013) 'Farm-Level Models for EU Policy Analysis: Review of Recent Literature and Comparison of most Relevant Models', in Langrell, S. (ed.), *Farm Level Modelling of CAP: A Methodological Overview*. JRC Scientific and Policy Reports Series, pp. 1–88. Seville: European Commission, Joint Research Centre.

Louhichi K., Kanellopoulos A., Janssen S., Flichman G., Blanco M., Hengsdijk H., Heckelei T., Berentsen P., Lansink A.O. and Van Ittersum M. (2010) 'FSSIM, A Bio-Economic Farm Model for Simulating the Response of EU Farming Systems to Agricultural and Environmental Policies', *Agricultural Systems*, 103: 585–97.

McNie E.C. (2007) 'Reconciling the Supply of Scientific Information with user Demands: An Analysis of the Problem and Review of the Literature', *Environmental Science & Policy*, 10: 17–38.

Mack G., Ferjani A., Möhring A., von Ow A. and Mann S. (2019) 'How did Farmers Act? Ex-Post Validation of Linear and Positive Mathematical Programming Approaches for Farm-Level Models Implemented in an Agent-Based Agricultural Sector Model', *Bio-Based and Applied Economics*, 8: 3–19.

Mack G. and Kohler A. (2018) 'Short- and Long-Run Policy Evaluation: Support for Grassland-Based Milk Production in Switzerland', *Journal of Agricultural Economics*, 70: 215–40.

Mariel P., Hoyos D., Meyerhoff J., Czajkowski M., Dekker T., Glenk K., Bredahl Jaobsen J., Liebe U., Olsen S.B., Sagebiel J. and Thiene M. (2021). *Environmental Valuation with Discrete Choice Experiments*, SpringerBriefs in Economics, pp. 1–136. Cham: Springer Nature Switzerland AG.

Matthews A. (2013) 'Greening Agricultural Payments in the EU's Common Agricultural Policy', *Bio-based and Applied Economics*, 2: 1–27.

——— (2016) 'The Dependence of EU Farm Income on Public Support', Progressive Economy @ TASC.

Metz F., Lieberherr E., Schmucki A. and Huber R. (2021) 'Policy Change Through Negotiated Agreements: the Case of Greening Swiss Agricultural Policy', *Policy Studies Journal*, 49: 731–56.

Möhring N., Huber R. and Finger R. (2023) 'Combining Ex-ante and E-post Assessments to Support the Sustainable Transformation of Agriculture: The Case of Swiss Pesticide-free Wheat Production', *Q Open*, qoac022, , https://doi.org/10.1093/qopen/qoac022.

Möhring A., Mack G., Zimmermann A., Ferjani A., Schmidt A. and Mann S. (2016) 'Agent-Based Modeling on a National Scale: Experiences from SWISSland', *Agroscope Science*, 30: 1–56.

Möhring N., Gaba S. and Finger R. (2019) 'Quantity Based Indicators Fail to Identify Extreme Pesticide Risks', *Science of the Total Environment*, 646: 503–23

Möhring N., Huber R. and Finger R. (2022) 'Combining Ex-Ante and Ex-Post Assessments to Support the Sustainable Transformation of Agriculture: The Case of Swiss Pesticide-Free Wheat Production', *Q Open*, qoac022. https://doi.org/10.1093/qopen/qoac022.

Möhring N., Ingold K., Kudsk P., Martin-Laurent F., Niggli U., Siegrist M., Studer B., Walter A. and Finger R. (2020) 'Pathways for Advancing Pesticide Policies', *Nature Food*, 1: 535–40.

Montpetit E. and Lachapelle E. (2015) 'Can Policy Actors Learn from Academic Scientists?', *Environmental Politics*, 24: 661–80.

Morawetz U.B. and Tribl C. (2020) 'Randomised Controlled Trials for the Evaluation of the CAP: Empirical Evidence about Acceptance by Farmers', *German Journal of Agricultural Economics*, 69: 183–99.

Mottet A., Bicksler A., Lucantoni D., De Rosa F., Scherf B., Scopel E., López-Ridaura S., Gemmil-Herren B., Bezner Kerr R., Sourisseau J.M. and Petersen P. (2020) 'Assessing Transitions to Sustainable Agricultural and Food Systems: A Tool for Agroecology Performance Evaluation (TAPE)', *Frontiers in Sustainable Food Systems*, 4: 579154.

Mueller-Langer F., Fecher B., Harhoff D. and Wagner G. G. (2019) 'Replication Studies in Economics— How Many and Which Papers are Chosen for Replication, and Why?', *Research Policy*, 48: 62–83.

Nolan J., Parker D., Van Kooten G.C. and Berger T. (2009) 'An Overview of Computational Modeling in Agricultural and Resource Economics', *Canadian Journal of Agricultural Economics*, 57: 417–29.

Norström A. V., Cvitanovic C., Löf M. F., West S., Wyborn C., Balvanera P., Bednarek A. T., Bennett E. M., Biggs R., de Bremond A., Campbell B.M., Canadell J.G., Carpenter S.R., Folke C., Fulton E.A., Gaffney O., Gelcich S., Jouffray J.-B., Leach M., Le Tissier M., Martín-López B., Louder E., Loutre M.-F., Meadow A.M., Nagendra H., Payne D., Peterson G.D., Reyers B., Scholes R., Speranza C.I., Spierenburg M., Stafford-Smith M., Tengö M., can der Hel S., van Putten I. and Österblom H. (2020) 'Principles for Knowledge Co-Production in Sustainability Research', *Nature Sustainability* 3: 182–90.

Nosek B. A. et al. (2015) 'Scientific Standards: Promoting an Open Research Culture', *Science*, 348: 1422–5.

O'Donovan C., Michalec A. and Moon J.R. (2022) 'Capabilities for Transdisciplinary Research', *Research Evaluation*, 31: 145–58.

OECD [Organisation for Economic Co-operation and Development] (2018). *OECD Regulatory Policy Outlook 2018*. Paris: OECD Publishing.

Page M. J., McKenzie J. E., Bossuyt P. M., Boutron I., Hoffmann T. C., Mulrow C. D., Tetzlaff J. M., Akl E. A., Brennan S. E., Chou R., Glanville J., Grimshaw J. M., Hróbjartsson A., Lalu M. M., Li T., Loder E. W., Mayo-Wilson E., McDonald S., McGuinness L. A., Stewart L. A., Thomas J., Tricco A. C., Welch V. A., Whiting P. and Moher D. (2021) 'The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews', *Systematic Reviews*, 10: 1–11.

Palm-Forster L. and Messer K.D. (2021) 'Chapter 80—Experimental and Behavioral Economics to Inform Agri-Environmental Programs and Policies', in Barrett, C. B., Just, D. R. (eds), *Handbook of Agricultural Economics*, 5, pp. 4331–406. Elsevier.

Pastor A.V., Palazzo A., Havlík P., Biemans H., Wada Y., Obersteiner M., Kabat P. and Ludwig F. (2019) 'The Global Nexus of Food–Trade–Water Sustaining Environmental Flows by 2050', *Nature Sustainability*, 2: 499–507.

Pattyn V., Molenveld A. and Befani B. (2019) 'Qualitative Comparative Analysis as an Evaluation Tool: Lessons from an Application in Development Cooperation', *American Journal of Evaluation*, 40: 55–74.

Pe'er G., Bonn A., Bruelheide H., Dieker P., Eisenhauer N., Feindt P.H., Hagedorn G., Hansjürgens B., Herzon I., Lomba Â., Marquard E., Moreira F., Nitsch H., Oppermann R., Perino A., Röder N., Schleyer C., Schindler S., Wolf C., Zinngrebe Y. and Lakner S. (2019) 'Action Needed for the EU Common Agricultural Policy to Address Sustainability Challenges', *People and Nature*, 2020: 305–16.

Pe'er G., Finn J.A., Díaz M., Birkenstock M., Lakner S., Röder N., Kazakova Y., Šumrada T., Bezák P., Concepción E.D., Dänhardt J., Morales M.B., Rac I., Špulerová J., Schindler S., Stavrinides M., Targetti S., Viaggi D., Vogiatzakis I.N. and Guyomard H. (2022) 'How can the European Common Agricultural Policy Help Halt Biodiversity Loss? Recommendations by Over 300 Experts', *Conservation Letters*, 2022: 1–12.

Poppe K. and Vrolijk H. (2017) 'Microdata: A Critical Source for Policy Evaluation', *EuroChoices*, 17: 28–35.

Poppe K., Vrolijk H., Dolman M. and Silvis H. (2016) 'FLINT: Farm-Level Indicators for New Topics in Policy Evaluation: An Introduction', *Studies in Agricultural Economics*, 118: 116–22.

Ragin C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.

Raymond C.M., Fazey I., Reed M.S., Stringer L.C., Robinson G.M. and Evely A.C. (2010) 'Integrating Local and Scientific Knowledge for Environmental Management', *Journal of Environmental Management*, 91: 1766–77.

Reidsma P., Janssen S., Jansen J. and can Ittersum M.K. (2018) 'On the Development and use of Farm Models for Policy Impact Assessment in the European Union—A Review', *Agricultural Systems*, 159: 111–25.

Rohr J. R. (2021) 'The Atrazine Saga and Its Importance to the Future of Toxicology, Science, and Environmental and Human Health', *Environmental Toxicology and Chemistry*, 40: 1544–58.

Rommel J., Schulze C., Matzdorf B., Sagebiel J. and Wechner V. (2023) 'Learning about German Farmers' Willingness to Cooperate from Public Goods Games and Expert Predictions', *Q Open*, qoac023, https://doi.org/10.1093/qopen/qoac023.

Sant'Anna P. H. and Zhao J. (2020) 'Doubly Robust Difference-In-Differences Estimators', *Journal of Econometrics*, 219: 101–22.

Santeramo F.G. and Lamonaca E. (2019) 'The Effects of Non-Tariff Measures on Agri-Food Trade: A Review and Meta-Analysis of Empirical Evidence', *Journal of Agricultural Economics*, 70: 595–617.

Santeramo F.G., Russo I. and Lamonaca E. (2023) 'Italian Subsidised Crop Insurance: What the Role of Policy Change', *Q Open*, qoac031, https://doi.org/10.1093/qopen/qoac031.

Schader C., Baumgart L., Landert J., Muller A., Ssebunya B., Blockeel J., Weisshaidinger R., Petrasek R., Mészáros D., Padel S., Gerrard C., Smith L., Lindenthal T., Niggli U. and Stolze M. (2016) 'Using the Sustainability Monitoring and Assessment Routine (SMART) for the systematic analysis of Trade-Offs and Synergies between Sustainability Dimensions and Themes at Farm Level', *Sustainability*, 8: 274.

Schader C., Curran M., Heidenreich A., Landert J., Blockeel J., Baumgart L., Ssebunya B., Moakes S., Marton S., Lazzarini G., Niggli U. and Stolze M. (2019) 'Accounting for Uncertainty in Multi-Criteria Sustainability Assessments at the Farm Level: Improving the Robustness of the SMART-Farm tool', *Ecological Indicators*, 106: 105503.

Schaub S., Huber R. and Finger R. (2020) 'Tracking Societal Concerns on Pesticides: A Google Trends Analysis', *Environmental Research Letters*, 15: 084049.

Schilizzi S. and Latacz-Lohmann U. (2016) 'Incentivizing and Tendering Conservation Contracts: the Trade-Off between Participation and Effort Provision', *Land Economics*, 92: 273–91.

Schreinemachers P. and Berger T. (2006) 'Land use Decisions in Developing Countries and Their Representation in Multi-Agent Systems', *Journal of Land Use Science*, 1: 29–44.

Stern E., Stame N., Mayne J., Forss K., Davies R. and Befani B. (2012) 'Broadening the Range of Designs and Options for Impact Evaluations', Working Paper 38, Department for International Development.

Stetter C., Mennig P. and Sauer J. (2022) 'Using Machine Learning to Identify Heterogeneous Impacts of Agri-Environment Schemes in the EU: A Case Study', *European Review of Agricultural Economics*, 49: 723–59.

Storm H., Baylis K. and Heckelei T. (2020) 'Machine Learning in Agricultural and Applied Economics', *European Review of Agricultural Economics*, 47: 849–92.

Swinnen J. (2018). 'The Political Economy of Agricultural and Food Policies', In Barrett, C. (ed.). *Palgrave Studies in Agricultural Economics and Policy*, Springer.

Tashakkori A. and Teddle C. (2003). *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, CA: Sage.

Tengö M., Brondizio E.S., Elmqvist T., Malmer P. and Spierenburg M. (2014) 'Connecting Diverse Knowledge Systems for Enhanced Ecosystem Governance: the Multiple Evidence Base Approach', *Ambio*, 43: 579–91.

Thomas F., Midler E., Lefebvre M. and Engel S. (2019) 'Greening the Common Agricultural Policy: A Behavioural Perspective and Lab-In-The-Field-Experiment in Germany', *European Review of Agricultural Economics*, 46: 367–92.

Thoyer S. and Préget R. (2019) 'Enriching the CAP Evaluation Toolbox with Experimental Approaches: Introduction to the Special Issue', *European Review of Agricultural Economics*, 46: 347–66.

Topping C. J., Aldrich A. and Berny P. (2020) 'Overhaul Environmental Risk Assessment for Pesticides', *Science*, 367: 360–3.

Troost C. and Berger T. (2015) 'Dealing with Uncertainty in Agent-Based Simulation: Farm-Level Modeling of Adaptation to Climate Change in Southwest Germany', *American Journal of Agricultural Economics*, 97: 833–54.

Vrolijk H. and Poppe K. (2021) 'Cost of Extending the Farm Accountancy Data Network to the Farm Sustainability Data Network: Empirical Evidence', *Sustainability*, 13: 8181.

Vrolijk H., Poppe K. and Keszthelyi S. (2016) 'Collecting Sustainability Data in Different Organisational Settings of the European Farm Accountancy Data Network', *Studies in Agricultural Economics*, 118: 138–44.

Uehleke R., Petrick M. and Hüttel S. (2019). 'Agricultural Policy Evaluation with Large-Scale Observational Farm Data: Environmental Impacts of Agri-Environmental Schemes', *SiAg-Working Paper 21 (2019)*, DFG-Forschergruppe 986, Humboldt-Universität zu Berlin, pp. 1–37.

Wang Y., Schaub S., Wuepper D. and Finger R. (2022) 'Culture and Agricultural Biodiversity Conservation', https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4043727 (accessed 25 Feb 2022).

Weigel C., Paul L.A., Ferraro P.J. and Messer K.D. (2020). 'Challenges in Recruiting U.S. Farmers for Policy-Relevant Economic Field Experiments', *Applied Economic Perspectives and Policy*, 43: 556–72.

Weiss C.H. (1979) 'The Many Meanings of Research Utilization', *Public Administration Review* 39: 426–31.

White H. and Raitzer D. (2017). *Impact Evaluation of Development Interventions: A Practical Guide*. Manila: Asian Development Bank (ADB).

Woltjer G.B., Kuiper M., Kavallari A., van Meijl H., Powell J.P., Rutten M.M., Shutes L. and Tabeau A.A. (2014). *The MAGNET Model: Module Description* (no. 14–57). pp. 1–148. Wageningen: LEI Wageningen University and Research Centre.

Wuepper D. and Finger R. (2022) 'Regression Discontinuity Designs in Agricultural and Environmental Economics', *European Review of Agricultural Economics*, 50: 1–28. https://doi.org/10.1093/erae/jbac023.

Wuepper D. and Huber R. (2021) 'Comparing Effectiveness and Return on Investment of Action- and Results-Based Agri-Environmental Payments in Switzerland', *American Journal of Agricultural Economics*, 104: 1585–604.

Wuepper D., Wimmer S. and Sauer J. (2020) 'Is Small Family Farming more Environmentally Sustainable? Evidence from A Spatial Regression Discontinuity Design in German', *Land Use Policy*, 90: 104360, https://doi.org/10.1016/j.landusepol.2019.104360.

Wuepper D.J., Wimmer S. and Sauer J. (2021) 'Does Family Farming Reduce Rural Unemployment?', *European Review of Agricultural Economics*, 48: 315–37.

Zimmert F. and Zorn A. (2023) 'Direct Payments and On-Farm Employment: Evidence from a Spatial Regression Discontinuity Design', *Q Open*, qoac024, https://doi.org/10.1093/qopen/qoac024.