# Evaluating net energy savings

## EPATEE topical case study illustrated with examples

### Table of contents



### Summary – Key ideas

The aim of energy efficiency policies is to promote a more efficient use of energy through actions that can be technical, organisational or behavioural. These actions are meant to consume less energy, thereby reducing energy costs and most often also GHG emissions. These are reasons, why some energy efficiency actions would also occur without any targeted policy. Hence, one of the frequent issues about energy efficiency policy evaluation is trying to find out what energy savings can be attributed to the policy and **what would have occurred anyways**. This is where net savings come in. They account only for that part of savings that are due to the policy.

Policymakers need this information to decide upon policy success, especially when looking at the **efficiency of the policy**. In other words to determine whether the money and other resources spent on the policy are spent well and have a strong impact on savings performance or whether the policy design should be changed to use the money and other resources more effectively. For example, Courts of Auditors frequently consider net savings to scrutinize the use of public money.

The evaluation of net energy savings is not trivial and depends on a series of factors including the type of data collection, access to participants or stakeholders, calculation methodology, baseline, inclusion and exclusion of effects and intended usage of the study results. Experience shows that there is **no silver bullet** to tackle this issue. Evaluators face the difficulty that results can vary largely depending on the methodology and that evaluation results are not comparable between each other. After all, the interested reader of the evaluation study might face difficulties to interpret the results because of the multitude of methodologies and complexity of data analysis.

On a broad level, net savings can be determined using two ways. The one being to define or assess a **baseline that represents the scenario that "would have happened" without the policy**. The other one is **adjusting the gross savings for a series of effects**. Gross savings are those savings that occurred among participants of the policy, independently of whether they would have acted the same or not in the absence of the policy. The most common adjustment effects are free-rider and spill-over effects. **Free-riders** are those participants who would have invested also without the policy but participated to benefit from a financial incentive or other types of support. **Spill-over effects** lead to savings in other

The EPATEE project has received funding from the European Union's Horizon 2020 Research and innovation programme under grant agreement No 746265.

areas or in the future due to the informative character of the policy, word-of-mouth effects, market transformation effects or other effects beyond the scope of the policy.

This topical case study presents commonly used approaches to calculate net savings. Each of the methods has its strong and weak points, which are discussed in more detail on the following pages. The following table sums up the methods.

| Type of method and related conditions | Pros and Cons |
|---|---|
| **Randomised Control Trial (RCT)**<br>• Experimental Design<br>• Treatment (exposed) and control group (not exposed to policy)<br>• Random assignment to groups<br>• Measured savings data available | + Random assignment reduces bias<br>+ Increases reliability and validity<br>+ Widely accepted<br>- Expensive in terms of time and money<br>- Needs to be planned<br>- Ethical problems to bar control group from participation |
| **Quasi-experimental design**<br>• Experimental Design<br>• Treatment (exposed) and comparison group (not exposed to policy)<br>• Matching of treatment and comparison groups by defined properties<br>• Measured savings data available | + Reliable method<br>+ Limited bias if comparison group well assigned<br>- Assignment of comparison group difficult<br>- No statistical means to determine the adequacy of the comparison group |
| **Survey approach**<br>• Participants are asked how they would have acted without the policy<br>• Deemed or scaled savings possible<br>• When no access to non-participant group (or not possible to define a control or comparison group)<br>• When budget and time restrictions | + Does not require non-participant control group<br>+ Flexibility to adjust questions to policy<br>+ Relatively low costs<br>- Prone to biases (in questions and answers)<br>- Participants' inability to know what they would have done<br>- Tendency to rationalise past choices<br>- Responses cannot be validated |
| **Deemed or Stipulated Net-to-Gross-Ratios**<br>• Literature-based estimation using market averages and secondary data<br>• Good for short studies<br>• Strict budget limitations<br>• Limited data availability | + Gives a first rough estimate before exploring further with other methods<br>+ Inexpensive<br>- Inexact<br>- Heterogeneity of policies do not allow for default values |

Some lessons learned about evaluation of net effects in practice are summarised below.

Due to issues with limited data availability, budget constraints, privacy laws and policy design and implementation, **randomised controlled trials are often not feasible**. In many cases, this applies also to quasi-experimental designs. That is because savings data are often based on deemed or scaled savings or there is no access to a non-participant group. The two examples from Germany and Denmark therefore used the survey approach.

Then one of the crucial issues is the exposure to **biases in the questions and answers**. Questions about the free-rider effect face biases in both directions. The free-rider effect may be overestimated due to hindsight bias and social desirability bias. **Hindsight bias** means that respondents might tend to rationalise their decision and believe they would have acted the same without the policy after knowing that the action was successful. **Social desirability bias** means that respondents might tend to give the answers that they consider as socially desirable, e.g., acting environmentally friendly and not only for monetary reasons. In the other directions, participants have an incentive to stress the importance of the policy to be able to obtain financial incentives in the future. It appears probable that the two former effects outweigh the latter and that free-rider effects are overestimated. However, there is no statistical means to verify this argument. While absolute values of net savings are therefore disputed, the results using the survey method can be used to compare the net savings between policies that have been analysed using the same methodology.

Due to the cross-sector nature of the Danish example, differences in net savings could be analysed for businesses as opposed to households stating that the net effects are higher relative to gross effects for businesses than for households. Furthermore, programmes including a combination of informative action and financial incentives showed a better net performance than each of the two intervention types alone. Another issue with the Danish survey method was the small sample size, which made it impossible to form subsamples large enough for statistical significance.

An approach to reduce the issue with biases is to include further question for verification. Implementing discrete choice experiments into the survey can reduce response biases. Furthermore, comparison with less subjective methods like market data analyses for the construction of a baseline can be included.

## Scope and definition

**Net energy or emissions savings** are those savings that directly accrue to the policy and would not have happened if it was not for the policy intervention. In contrast, gross energy savings are all savings that occurred among the scope of the policy. Hence, the differences between the two are those savings in the scope of the policy that would have also occurred without the existing policy in place, or that have occurred due to the policy but outside its monitoring scope. The question of net savings is relevant for all types of policies and sectors and can therefore be considered a transversal issue.
Two main conceptual approaches are commonly used to assess them:

- defining a **counterfactual** (or **baseline**) used to deduct net energy savings from comparing this baseline with what has happened (and could be monitored or measured);
- applying **effect adjustments** to the **gross savings** (that correspond to the results monitored for the policy or programme, without assessing the causality between the policy measure and the actions implemented).

Two of these effects are free-rider effects and spill-over effects.

**Free riders** are those participants who would have implemented the savings measure, in full or partly, also in the absence of the policy. Hence they do not accrue to the policy intervention. They are particularly an important issue to consider when the policy measure includes financial incentives. From a policy maker’s point of view, the money spent on those participants was not successfully invested.

**Spill-over effects** are the flipside of that effect. Participation in the policy intervention may lead to a higher awareness for energy efficiency and indirectly positively influenced investment decisions. Furthermore, affected third parties or those that are introduced to energy efficiency investment opportunities by the participant are also not accounted for in the policy evaluation’s gross effects. Hence, an amount for spill-over effects can be added. In theory, spill-over effects can also be negative, as is the case with negative publicity due to unsatisfied participants or media coverage of frauds, defaults or incidents (Griechisch & Unger, 2014). However, policy measures usually include quality processes or feedback loops to monitor participants' satisfaction in order to limit or detect problems and prevent negative spill-over.

The topic of net savings evaluation has implications for all kinds of energy efficiency policies and beyond as well as for most informative and incentive policies in general. Hence, it is a transversal or general issue. The examples included in this case study provide practical illustrations on cases for the German Energy Efficiency Fund and the Danish Energy Companies' Energy-Saving Efforts.

The terms "net effects" and "**additionality**" are often used interchangeably. This is particularly the case when the term is used on the individual programme or policy level. Additionality in this sense refers to the choice of a baseline. Possibilities of baseline are for example a before-after comparison of energy consumption (Bundgaard, et al., 2013) or the consumption reduction compared to a default technology on the market (European Commission, 2016). Depending on the method of data assessment, the definition of additionality can lead to a result equivalent to gross energy savings (when the baseline is taken as the situation before implementing an energy efficiency action, and that no adjustment effect is further taken into account) or to net energy savings (when the baseline is equivalent to a counterfactual scenario representing what would have occurred in the absence of the policy).

The policy objectives can also lead to define additionality as results additional to other policies or programmes in force. This is for example the case for the Energy Efficiency Directive (EED, 2012/27/EU) for which additional energy savings are defined as additional to the effects of the other EU legislation and regulations for energy efficiency (European Commission, 2013). When the level of analysis is a superior aggregate level, then additionality is not the same as net effects, as is the case for the EED.

This topical case study focuses on the **evaluation of net savings on the individual policy level**, but the issue of additional savings is discussed in one of the examples (about the Danish Energy Efficiency Obligation scheme).

Rebound effects are not discussed in this case study, but some of the methods in the section "common approaches" can directly or indirectly take rebound effects into account.

## Insight from the literature

As pointed out by (Violette & Rathbun, 2017), evaluating net impacts is not a challenge specific to energy efficiency policies, but common to any evaluation of any type of public policy, when evaluation aims at attributing impacts to a policy. Evaluators usually proceed with caution, as ideal cases where attribution can be proven without any doubt remain rare (see e.g., Train 1994). In practice, this often requires to define and test a set of assumptions, for example about what the baseline and the effects of the policy would be.

Whereas there is a broad agreement on the general definition of "net energy savings" (i.e. energy savings that would not have occurred in the absence of the policy measure), the practical definition adapted to the evaluation of a particular policy measure might take various forms and might be perceived differently depending on stakeholders' point of view.

Clarifying, how the general concept of "net energy savings" is transcribed in practical terms for the case under evaluation, and ensuring a common understanding of the issue among the stakeholders involved in the evaluation are therefore two essential preliminary steps to the evaluation of net savings. This has been pointed out in the principles and guidance defined by NEEP (Northeast Energy Efficiency Partnerships, 2016).

One of the most discussed issues revolves around the definition of free-riders and what kind of free riders exist under which circumstances.

The most commonly used and general definition of **free-riders** is: actors that would have invested into energy efficiency measures (EEM) even in the absence of support programmes, and particularly in the absence of the related incentives (Paramovona & Thollander, 2016; Vine & Sathaye, 1999; Wade & Eyre, 2015). In contrast, Rietbergen, Farla, & Blok (2002) do not explicitly apply the term free-rider effects but discuss it in the context how policies have changed investment behaviour and stimulated energy savings. As (Skumatz, 2009) points out, some studies complement the general definition of free-riders with four criteria to make it more practical to assess:

- being aware of the energy efficiency measure before the policy/programme,
- intending to implement it before the programme,
- knowing where to purchase it and
- be willing to pay the non-subsidised price.

Moreover, some authors differentiate free riders into full, partial and deferred (Schiller, 2007; Broc, Adnot, Bourges, Thomas, & Vreuls, 2009; Collins, 2016):

- A **full free rider** would have installed the same energy efficiency measure at the same time without the policy.
- A **partial free rider** would have implemented only a part of it, or would have implemented the same type of action but with a lower energy performance (for example, someone who already planned to replace an equipment, but would have bought a new equipment of a lower energy performance than the one promoted by the programme).
- A **deferred free rider** would have installed a smaller or the same energy efficiency measure, but at a later time.

In practice, it is difficult to get reliable estimates of partial or deferred free-rider effects. It is thus important to have a shared understanding and an agreement among stakeholders about the way free-rider effects are defined and will be evaluated, so that the results are well accepted by all.

(Olsthoorn, Schleich, Gassmann, & Faure, 2018) distinguish between **strong and weak free riders**. While the first was perfectly informed and complied with the four conditions listed above and would have implemented the measure, the latter was not well informed but after knowing of the investment opportunity would not have needed the financial incentive, but received it anyways.

Evaluation frameworks or protocols often require evaluations to account for free-rider effects, especially when evaluation objectives are focused on accountability, examining if public resources have been used efficiently. Literature is discussing that free-rider effects are relevant as soon as subsidies are involved, not when minimum standards are applied (Nilsson, Lopes, Fonseca, & Araújo, 2008; Nauleau, 2014). (Breitschopf, Voswinkel, & Schlomann, 2018) advocate for a differentiated view on free rider effects in policy evaluation. While they play an important role for the evaluation of the effectiveness or the economic efficiency of a policy, they are not relevant for the evaluation of the degree of target achievement. In the end of the day, the amounts of energy are saved whether it would have happened without the policy or not. (Eichhammer, Boonekamp, Labanca, Schlomann, & Thomas, 2008) recommend accounting for free-rider effects if the aim is to evaluate energy savings additional to baseline projections. Wade, Eyre (2015) point out that free-ridership matters for some methods applied, while it can be ignored with a counterfactual based on control group approach if the propensity to make changes is the same in the participant and control groups. In this latter case, free-rider effects are assumed to be directly taken into account when comparing changes in the participant and control groups.

The **objectives and time perspective** of a programme or policy might question the way to define free-riders and the relevance to evaluate free-rider effects. Particularly in case of long term programmes aiming at market transformation: making energy efficient products more easily available and affordable will increase in the next years the share of consumers who would buy these products without incentives. But would they have been able to do so without the previous changes due to the market transformation programme? This raised a **debate** on whether evaluation rules might affect the design of programmes and create a bias towards short term results (Friedmann, 2007; Blumstein, 2009).

Overall, assessing and accounting for free-riding differs from study to study, some assume free-riders cancel out with other effects, some assume a certain percentage of participants are free-riding, or others assume zero free-ridership (Alberini & Bigano, 2015). Further, free-ridership differs by the type of EEMs (Vine & Sathaye, 1999) and sectors, and it depends on the socio-professional category, e.g. education, and income (Nauleau, 2014). Because it is difficult to account for all the individual aspects at the EEM level with a reasonable amount of effort, as (Moser, Leutgöb, & Kollmann, 2012) state, most often only average effects are taken into account.

An analysis of the EPATEE knowledge base among the empirical studies grouped studies based on both the the target group and sector as well as the type of energy efficiency measure. The results have shown that the that the calculated **percentages of free ridership differ largely both in between categories as well as within within for both type of categories** as can be seen in Figure 1. Free-rider shares found in literature.

Source: Breitschopf, Voswinkel and Schlomann (2017). Note: EEM: energy efficiency measure. This might come from many reasons: different definitions or scopes of free-rider effects considered, different methods used to evaluate these effects, differences in the objectives and design of the policy instruments, etc. An approximation of free rider share can hence not be deducted from these categories. And **the use of default values for free-rider effects should be considered with a lot of caution**.

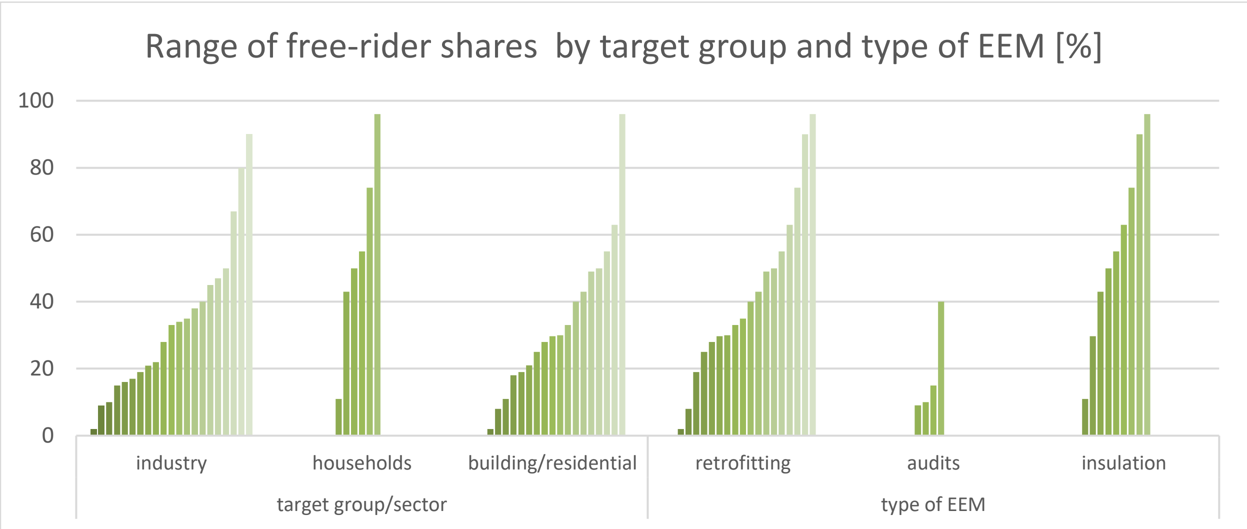


Figure 1. Free-rider shares found in literature.

*Source: Breitschopf, Voswinkel and Schlomann (2017). Note: EEM: energy efficiency measure.*

Results from the EPATEE knowledge base show that the issue is present both in theoretical and empirical studies. Empirical studies are commissioned by policy makers, policy implementers while extensive conceptual work has been performed by evaluators and researchers.

## Common approaches

The evaluation of net impacts is one of the most difficult issues in policy evaluation. There are a series of approaches. The Uniform Methods Project by the US-American NREL distinguishes nine methods listed below (Violette & Rathbun, 2017). This case study presents four of these approaches, which all have their strengths and weaknesses. The presented approaches are marked in **bold print**.

- **Randomised Controlled Trials (RCTs)** and options for randomised approaches
- **Quasi-experimental designs including matching**
- **Survey-based approaches**
- Market sales data analyses
- Structured expert judgement approaches
- **Deemed or stipulated Net-to-Gross ratios**
- Historical tracing (or case study) method
- Common practice baseline approaches
- Top-down evaluations (or macroeconomic models)

The four types of methods presented below were selected because they were the most frequently found in the papers gathered in the EPATEE Knowledge Base and in the EPATEE case studies. Examples of using market sales data analyses, structured expert judgement approaches, historical tracing, and top-down evaluations can be found in (Violette and Rathbun, 2017). About market sales data analyses, see also the example of Hummer (2016).

*Randomised Controlled Trials (RCT)*

This approach uses an experimental design. The full study population is randomly assigned to either the treatment group or the control group. The treatment group is then subject to the policy and their savings can be evaluated. The same applies to the control group that is not subject to the policy and is in the further process considered the baseline. The difference between the two groups are then the net savings. The actions of the control group include free-riders and other possible effects. A clear distinction between the effects is not possible with this method.

With this approach, it is either possible to compare energy usage after the policy implementation for both groups, or a **difference-in-differences** (DiD) approach to compare the change in energy usage before and after the policy for both groups. Finally, a **fixed-effects regression** can identify the effects of the policy of the treatment group using control group data as variables. Observer bias in selecting representative sample groups are eliminated by the random assignment.

However, in reality **randomised controlled trials are often not possible** (Wade & Eyre, 2015). It is hardly justifiable to deny policy participation in an energy efficiency policy to a random half of participants for the sole reason of evaluation accuracy. In this case, quasi-experimental designs can be used.

RCT has been rarely used for evaluation of energy efficiency policies in Europe. No practical example can be found in the EPATEE case studies. This has been more widely used in North America, especially for the evaluation of Home energy reports or other behavioural pilot programmes (Schellenberg, Savage, Blundell, Cook, & Smith, 2015; Dougherty & Van de Grift, 2014).

*Quasi-experimental designs*

The main difference to randomised controlled trials is that the participants are self-selecting into the treatment group. In other words, they decide to participate in the policy and implement an energy efficiency measure. The evaluator then tries to match the treatment and control groups based on certain criteria that influence energy use and propensity to implement energy efficiency measures. While more and more advanced **matching methods** are reducing the possible biases, a certain amount of uncertainty is always involved as unobservable criteria that affect energy use can be present and the evaluator can never make a perfect decision because perfect information do not exist. The subsequent analysis of the two groups is then performed in the same way as in the randomised controlled trials, leading to estimate net savings from the policy or programme.

However, also this method can often not be used. Issues include the prevalence of deemed or scaled savings in bottom-up evaluations (Labanca & Bertoldi, 2016). Deemed and scaled savings rely on engineering estimates of energy savings rather than physical measurements or metering of before and after energy usage or usage compared to a predefined baseline. This approach is often chosen because it might be easier to implement than **collecting metered energy consumption (e.g., energy bills)**. Without such measurements though, no group comparisons can be performed (or this would be limited to looking at whether individuals in the control group have also implemented actions, as a proxy to estimate free-rider effects).

This can be a major difference between the evaluation practices in Europe and in North America: most of the evaluations done in North America are about utility programmes, with a kind of guarantee for the evaluators to have an access to energy consumption data based on utility bills or metering. The access to these data is often much more challenging in European countries. It often requires a careful planning of data collection (i.e. to be prepared before the policy starts) and to take into account strict legal requirements (cf. privacy and personal data regulations).

Another issue is the **lack of available data for the control group**. While participants in a policy can sign that they will participate in an evaluation survey as part of their contract, non-participants are per definition not involved with the policy and there is hardly a basis for contacting them or acquire measurements from them. Data privacy laws furthermore protect each person's or company's data which makes the process legally challenging. Therefore, a survey-based approach can be used. Apart from that, it can be very costly to analyse energy usage among not only participants, but also non-participants. An example of experience feedback on the challenges to apply this type of method can be found in the EPATEE case study about Better Energy Homes (Ireland) and in the paper by Scheer et al. (2013).

*Survey-based approach*

In the survey-based approach, the building of groups is not obligatory. It is possible to survey only participants which solves the problem of access to non-participants in many evaluation studies. The approach works independent of the way of savings data calculation. Hence, it is compatible with deemed or scaled savings. In this approach, the participants are asked a series of questions to find out if they would have implemented the energy efficiency action in the absence of the policy and to what extent or at what time.

The main weakness of this approach is that it is prone to a series of biases including **self-selection bias** when survey participation is not obligatory and enforced, and **social desirability bias** that plays an important role when it comes to environmental policies because it is socially desired to act environmentally friendly. The challenge is therefore, to adjust the wording and order of questions so far, so that biases can be minimised. A series of partly redundant questions (but phrased differently) is often used to check the consistency of the answers from each respondent. For a good overview of survey bias for evaluation, refer to Baumgartner (2016) in section 2.5.1 (Respondent Behaviors and Responses).

The survey method is used in two examples of this case study: the evaluations of the German Energy Efficiency Fund and of the Danish Energy Efficiency Obligation scheme.

*Deemed or Stipulated Net-to-Gross-Ratios*

In the case of very low data availability or strong budget or time constraints, this method can be used to give a first rough estimate of net effects. It is the most straightforward and low-cost method, but also the least exact way of accounting for free-rider effects. It is a literature-based estimation using market averages and secondary data. A careful selection of literature used for the average value can potentially minimise biases. However, policy designs are very heterogeneous and the free rider share of one policy may not be applicable for another policy. Figure 1 above presents the strongly differing results for free-rider shares in empirical evaluation studies. A result that renders this approach rather unreliable.

Still it can be used as a first rough estimate before exploring further with other methods. See for example the EPATEE case study about Better Energy Homes (Ireland).

The following tables sum up the four presented methods and their Strengths and Limitations as can be found in Johnson (2014), also providing references for further examples and discussions.

<table>
<tr><th colspan="3">Randomised Controlled Trial (RCT)</th></tr>
<tr><td>Description and main principle of the method</td><td colspan="2">Experimental design. The full study population is randomly assigned to either the treatment group or the control group. The treatment group is then subject to the policy and their savings can be evaluated. The same applies to the control group that is not subject to the policy.</td></tr>
<tr><td colspan="2">Strengths</td><td>Limitations</td></tr>
<tr><td colspan="2">Random assignment reduces and limits bias in estimates</td><td>Ethical problems when assigning people to the control group and barring them to participate</td></tr>
<tr><td colspan="2">Increases reliability and validity</td><td>Expensive in terms of time and money</td></tr>
<tr><td colspan="2">Widely accepted method</td><td>Needs to be planned as part of a programme implementation to allow for appropriate randomisation</td></tr>
<tr><td colspan="2">Populations of participating individuals are clearly identified</td><td>Volunteer biases: participating group may not be representative of the whole</td></tr>
<tr><td>In which cases is it used?</td><td colspan="2">Evaluation is planned before the period subject to evaluation.<br>Access to energy consumption data (e.g., energy bills)<br>Access to non-participants and sample sizes are large enough for randomisation.<br>Savings have to be measured savings<br>In practice, RCT has been mostly used to evaluate behavioural programmes or experimentations (e.g., pilot programmes).</td></tr>
<tr><td>Examples and references to go further</td><td colspan="2">Examples: (James and Ambrose 2016; Kendel et al. 2017; Mukai and 2014; Tiefenbeck et al. 2016)<br>Review of studies: (Dougherty and Van de Grift 2014)</td></tr>
</table>

| | |
|---|---|
| | Further discussions and guidance: (Cooper 2018 with a response in discussion by Vine et al. 2018; Fredericks et al. 2016; Gandhi et al. 2016; Violette and Rathbun 2017) |

| **Quasi-experimental designs** | | |
|---|---|---|
| **Description and main principle of the method** | Similar to RCT, but without randomisation. A participant group is evaluated and a matching non-participant group with equal characteristics related to energy use has to be found. Energy savings are then compared like in RCT | |
| **Strengths** | | **Limitations** |
| Limits bias if a matched comparison group can be identified regarding the actions that influence energy-use | | Difficulty to identify a matched comparison group |
| Unlike RCT can be applied after programme implementation | | There is no empirical means to determine the adequacy of the comparison group |
| Widely accepted when random assignment cannot be used | | Without randomisation, statistical tests can be meaningless |
| **In which cases is it used?** | Access to energy consumption data (e.g., energy bills) and data needed to match participant and non-participant groups.<br>Access to non-participants but randomisation is not possible.<br>Savings have to be measured savings. | |
| **Examples and references to go further** | Examples: EPATEE case studies about Better Energy Homes (Ireland) and about the Weatherization Assistance Program (US); (Asensio and Delmas 2017; Bertholet et al. 2014; Granell et al. 2017; McClure and Provencher 2014; Schleich et al. 2015)<br>Further discussions and guidance: (Gaffney et al. 2015; Hannigan and Cook 2015; Stuart 2010 ; Violette and Rathbun 2017) | |

| **Survey approach** | | |
|---|---|---|
| **Description and main principle of the method** | Participants and possibly non-participants are surveyed about how they would have reacted in the absence of the policy and further questions to reduce biases. | |
| **Strengths** | | **Limitations** |
| Does not require a non-participant control group | | Prone to a long array of biases common to surveys |
| Flexibility to adjust questions to programme | | Participants' inability to know what they would have done in the absence of the policy |
| Relatively low costs | | Tendency to rationalise past choices |
| Can be combined with participant satisfaction survey | | There is no way to validate the accuracy of the responses. |
| **In which cases is it used?** | When starting from deemed or scaled savings.<br>No access to non-participant group.<br>Budget and time restrictions. | |
| **Examples and references to go further** | Examples: German and Danish examples presented further on; (McClaren and Bliss 2018; Sulyma and Tiedemann 2016)<br>Further discussions and guidance: (Galhotra and Randazzo 2015; Meurice et al. 2014; Ridge et al. 2009; Violette and Agapay-Read 2016; Violette and Rathbun 2017) | |

| Deemed or Stipulated Net-to-Gross-Ratios | | |
|---|---|---|
| **Description and main principle of the method** | Literature-based estimation using market averages and secondary data. | |
| **Strengths** | | **Limitations** |
| Simplicity | | Inexact |
| Gives a first rough estimate before exploring further with other methods | | Heterogeneity of policies do not allow for default values |
| Low costs | | |
| **In which cases is it used?** | Short studies; strict budget limitations; no access to neither participant nor non-participant data; top-down estimations | |
| **Examples and references to go further** | Examples: EPATEE case study about Better Energy Homes (Ireland); about the experience in California with DEER (Database of Energy Efficiency Resources) and the development of TRM (Technical Reference Manual) (Miller et al. 2014)<br>Further discussions and guidance: (Violette and Rathbun 2017) | |

## Concrete example n°1: [Germany] Energy Efficiency Fund

**Background**

This example is based on the EPATEE Case Study "[Germany] Energy Efficiency Fund": https://epatee.eu/sites/default/files/epatee_case_study_germany_energy_efficiency_fund.pdf
The **Energy Efficiency Fund (EEF)** is a special budget of the German Ministry of Economic Affairs and Energy and part of the **"Energiewende".** The fund consists of currently **23 policy measures** including **funding schemes** and **educational activities**. They target at contributing to a highly energy-efficient economy, the achievement of climate protection targets and existing energy saving potentials, and the decrease of energy costs. The funding schemes are administered by different federal agencies or the development bank KfW. Beneficiaries are **businesses, households and municipalities**.

The Energy Efficiency Fund has been introduced in 2011 and a detailed evaluation project of all policy measures will finish in late 2018 (Schlomann, et al., 2017). It approaches the heterogeneous measures by establishing a detailed common methodology that includes methods to calculate net effects.

**Evaluation into practice**

Because deemed savings are accounted in the majority of policies, RCT or quasi-experimental designs are not possible. Therefore, the calculation of net effects in the Energy Efficiency Fund is performed using the **survey approach**. Using this method, free-rider effects as well as spill-over effects were calculated using a set of questions to all participating companies. Free-rider effects were calculated at a high level of detail distinguishing between full, partial and deferred free-riders as well as strong and weak free-riders as described above.

For distinguishing between full, partial and deferred free-riders, the participants were asked whether they would have implemented the actions without the policy in full, partially, later or partially and later or not at all. Each of these answers were assigned a percentage of free-ridership where not at all equalled 0% and in full 100%. "Later" was calculated relative to the action's technical lifetime under the assumption that investments in the longer run are planned further in advance. Partial free-riders are distinguished using another question about their motivation for programme participation. If the financial incentive was a main motivation, a significantly lower investment is assumed and the free-rider share is reduced. In a second step, a question about the importance of the information content is asked. For this, the evaluators tried to find out how far the participant was already informed about the investment opportunities before the policy. Free-rider share of those who were not well informed before the policy was then reduced. A more detailed account of the approach can be found in Voswinkel (2018).

**Lessons learnt (about the evaluation method/practice)**

The survey approach lead to results that are in line with literature and can be considered accurate in this sense. Most survey participants answered the relevant questions. However, biases in the responses cannot be ruled out. While a robustness check has shown that the respondents are representative for the full sample based on sector and company size, a self-selection bias is possible.

The results can hence be used to compare evaluation results for policies that use the same calculation methodologies, but the absolute number may be skewed. The evaluators expect that they overestimate the free-rider effect because after implementation of the action, participants attribute the success to their own rationality, not to the financial incentive. Furthermore, for energy efficiency as a "green" topic, the social desirability is that the participant did not invest only for the incentive money, but out of conviction.

A possibility to reduce such biases may be to further work on the questions. A possibility may be to include discrete choice experiments to reveal the participant's real preferences.
However, keeping in mind that absolute numbers for net effects may be skewed, the possibility to compare policies in a reliable way can often already be crucial for evaluation.

**Added value of the evaluation/study/project (for the policy and/or about the topic)**

The developed detailed evaluation methodology for a wide range of policies in the Energy Efficiency Fund is an important step in harmonisation efforts for energy efficiency evaluation.

The results make it possible for policymakers to identify strengths and weaknesses of the different policies in the fund. These results can be compared and decisions upon them can be made.
In a row of policies in the fund, recommendations were put into practice.

## Concrete example n°2: [Denmark] Energy Companies' Energy-Saving Efforts

**Background**

This example is based on the EPATEE Case Study "[Denmark] Energy Companies' Energy-Saving Efforts":
https://epatee.eu/sites/default/files/epatee_case_study_denmark_eeo_scheme_vfinal3.pdf

This is the Danish Energy Efficiency Obligation scheme, started in 2006. The objective of the scheme is to promote cost-effective energy savings that would otherwise not have been realized. It is implemented through an agreement between the Danish Energy Agency (DEA) and the energy distributors (electricity, natural gas, district heating and heating oil). Energy distributors are required to achieve yearly energy savings targets, and must report each year their achievements to DEA. Actions saving final energy can be done in all sectors. Eligibility criteria include minimum energy performance requirements and rules about additionality (e.g., CFLs and household appliances were excluded from 2010).

An ex-post evaluation has been done for each period, meaning every 3rd or 4th year (i.e. in 2009, 2012 and 2015). This provides the ground for the renegotiations of the agreement, including the framework and specific rules of the scheme. The evaluation is commissioned by DEA. The framework and content of the evaluation is defined after discussions with the Technical Working Group. A Steering Committee provides an external view along the implementation of the evaluation.

Cost-effectiveness and net impact/additionality were the two aspects of the scheme that receive the highest attention in the ex-post evaluations (ENSPOL, 2015). This can be due to the fact that the obligated parties are energy distribution companies, and therefore the cost recovery mechanism is set and supervised by the regulatory agency (DERA). The assessment of cost-effectiveness and additionality is thus key to ensure that the scheme is delivering a net benefit to the whole end users.

**Evaluation into practice**

Energy savings are calculated by the obligated parties (or their subsidiaries or contractors) using:
- either **deemed energy savings**, i.e. pre-defined energy savings ratios per types of standardised actions (mostly used for actions in the residential sector);
- or **scaled energy savings** based on specific calculation methods (63% of the savings reported over 2006-2013, based on Deloitte et Grontmij, 2015).

The **baseline** for these calculations is the energy consumption before implementing the action, except for replacement of equipment where repair work cost is higher than 25% of replacement cost (then the baseline is based on market average or legal requirements).

**Additionality** is not assessed for each action or project implemented, but is one of the key issue addressed in the ex-post evaluations. In practice, additionality is addressed in several ways:
- increasing the overall annual targets from one period to the other, to take into account when part of the energy savings reported for the previous period were not considered additional based on the ex-post evaluation;
- applying **reduction factors** for action types evaluated as having a lower additionality based on the ex-post evaluations;
- not accounting energy savings that would last less than one year;
- requiring that in case of offering grants, the grants do not lower the payback time to less than one year (rule introduced after the evaluation done in 2012, to avoid too high incentives).

The reduction factor can be up to 100%, meaning that the corresponding action type is removed from the list of eligible actions (case of CFL and most household appliances from 2010, after the 2008 evaluation).

Moreover, the scheme requires a chain of agreements from the end-user to the obligated party, equivalent to the materiality criteria required by the EED. This requirement is meant to ensure that there was a demonstrable support from the obligated party (or its subsidiary or contractor) to the end-user, and to prevent double counting within the obligation scheme (same energy savings being reported by several obligated parties).

The definition of additionality used for the ex-post evaluations has been as follows: "*energy savings are additional if the energy savings actions had not been implemented (today or within the next few years) in the absence of the obligation scheme*".

Assessing additionality was considered as one part of the evaluation of net energy savings. Net energy savings were calculated in the ex-post evaluations according to the following formula:

net energy-savings =

reported savings * error corrections * actions' additionality * rebound effect * spill-over effect

With:

- reported savings: savings as reported by the obligated parties;
- errors corrections: corrections of errors identified in the calculation or reporting of energy savings by the obligated parties;
- actions' additionality: factors assessed through surveys of participants (see more details below)
- rebound effect and spill-over effect: in practice these factors were not or could not be evaluated or taken into account (results not reliable or representative enough).

*Method used about additionality in the ex-post evaluation done in 2008*

This evaluation was about the entire Danish energy efficiency portfolio, therefore the evaluation of the EEO scheme went less in-depth than the successive ones. This means that it had to partly rely on desk research (existing literature, databases and earlier evaluations) combined with dialogue with stakeholders and targeted empirical data collection (through questionnaires and telephone surveys) (Togeby et al. 2009).

Additionality was evaluated through a survey of 100 end-users who implemented large projects (i.e. high energy savings) reported for the scheme. The three corresponding questions in the survey were:

- To what extent were you, before you were in contact with the obligated party, thinking about realising the energy saving project?
- How likely is it that you, without contact to the obligated party or other actors, had implemented the energy within 1 year?
- How likely is it that you, without contact to the obligated party or other actors, had implemented the energy within 3 years?

Neither spill-over nor rebound effect were quantified in this evaluation done in 2008.

*Method used about additionality in the ex-post evaluation done in 2012*

Additionality was evaluated through **phone interviews** with a sample of 209 end-users (46 in the residential sector and 163 in non-residential sectors) that had received subsidies or advice from an obligated party (or its subsidiary or contractor). The projects analysed with these interviews represented 16 % of the energy savings target for 2010 (Bundgaard et al. 2013).

A question about additionality was added compared to the survey done in 2008:

- How critical to the implementation of the project was the subsidy you received? (when relevant)

Complementary questions were also used for cross checking the answers.

In parallel, a **case study** was used to compare changes in energy consumption of 166 households who received an action through the scheme, and 165 households forming a non-participant group (**quasi-experimental approach**), both groups supplied with the same district heating. All houses in the non-participant group were selected as the neighbour house to a specific house in the participants group (matching based on similarities of buildings and location). Energy consumption over 12 months before and after 2009 were compared for both groups, including adjustments of heating consumption for weather conditions based on heating degree days. The results showed a reduction of 10.9 kWh/m² on average for the non-participant group (i.e. reduction by about 7%) and an average reduction of 24.8 kWh/m² for the participants group (i.e. reduction by about 15%). This would mean that the net effect for the participants group would be about 56% of the energy savings achieved. This is a much higher additionality than the additionality that was assessed with the survey of households on a project basis (between 6 to 9%). The evaluators highlighted that these results should be seen as a case study and not as a general results for all energy savings in the residential sector (cf. small sample not meant to be representative).

Neither spill-over nor rebound effect were quantified in this evaluation done in 2012. The case study could have provided results about rebound effect. But as the sample was not representative, this was not used for the evaluation of the overall impact of the scheme.

*Method used about additionality in the ex-post evaluation done in 2015*

The objective of this study was also to enable a broader coverage of the different factors to take into account for evaluating net energy savings (see formula above), as well as to provide more robust results about additionality (see discussion on this point in “Lessons learnt” later on).

The table below provides a summary of the analytical methods used, showing what factors they are meant to assess. X mark in parentheses indicate that the analytical method is used to assess the factor only indirectly, partially or to a limited extent (according to evaluators’ description, Deloitte et Grontmij 2015).

Table 1. Analytical method used to assess the different elements taken into account when calculating net energy savings (about the evaluation done in 2015).

| **Type of method used** | **Reported savings** | **Error corrections** | **Actions’ additionality** | **Rebound effect** | **Spill-over effect** |
|---|---|---|---|---|---|
| Review and assessment of annual reports (of obligated parties) | X | | | | |
| Survey of end users (participants and non-participants) | X | | X | | |
| Survey of contractors | | | | | X |
| Review and evaluation of the documentation of a sample of energy savings projects | X | X | (X) | | |
| Econometric time-series analysis of energy consumption at the macro level of end-use sectors | X | (X) | (X) | (X) | X |

Analyses of actions' additionality based on the **survey of end-user groups** (participants and non-participants):

- combination of direct questions about additionality with complementary questions to clarify whether there is consistency in the answers and linearity effects;
- answers used to estimate average additionality percentage per type of end-use sector;
- assumption that a web survey provides more reliable answers compared to the previous evaluation (phone survey), as respondents get opportunity to thoroughly consider the answers they submit;
- **statistical inference analysis** to verify that results from the samples can be extrapolated to the whole population (all energy savings projects reported for the same end-user group in that period);
- **regression analysis** to investigate what variables (within each group) may influence the additionality rates.

Analyses of spill-over effect based on the **survey of contractors** (installers and engineering consultancies):

- questions about the extent to which the obligation scheme has led to increased supply/demand and changes in prices or in the products they buy or recommend to their customers;
- answers used to assess whether there have been spill-over effects in the supply chain;
- conclusions that this approach does not provide an accurate estimation of spill-over effects, but brought some insights.

**Review and evaluation of the documentation** of a sample of energy savings projects:

- verification that the project and its documentation have been correctly completed;
- verification of the calculation of the energy savings;
- analysis of the description about additionality in the documentation;
- results used to update the average correction factor applied to reported energy savings.

**Econometric time-series analysis** of energy consumption at the macro-level of end-use sectors:

- statistical model designed to include the main variables affecting energy consumption in each end-use sector;
- analysis of long time series from national statistics (energy prices, consumption per end-use sector, etc.), complemented by data provided by the trade associations of the energy companies for actions done in 2013 and data collected by the Energy Regulatory Authority (costs induced for the obligated parties);
- used to estimate the overall net effects (assuming that this method enables to capture directly or indirectly all factors listed in the table above);

See Annex 4 of (Deloitte et Grontmij, 2015) for more details about the econometric method used.

Overall, this evaluation surveyed 1746 respondents, compared to 260 respondents (participants + stakeholders) in the evaluation done in 2012. The evaluators of the study done in 2015 thus argued to have gathered a more robust basis, which was a key requirement for this study.

**Lessons learnt (about the evaluation method/practice)**

The stakeholders as well as the evaluators acknowledge that assessing additionality is a **challenge**. And issues were raised about the evaluation methodologies used and to which extent results with high **uncertainty** can be used to guide decision on improvements.

ENSPOL (2015) reported in particular about the evaluation done in 2012 that the accuracy of the evaluation methodology used to assess additionality in households was heavily disputed, as this led to applying reduction factors for some deemed savings (e.g., replacement of oil boilers, windows and cavity wall insulation).

The results of the evaluation in 2012 showed **differences in the additionality depending on the type of support offered** by the obligated parties to the end-users. According to these results, advice in combination with subsidies would have an additionality of 52-61%, subsidy alone 29-44% and advice alone 25-31%. Similarly, the evaluation found differences in the additionality **depending on the end-use sector** where the projects were implemented: according to the answers of the surveyed end-users, about 45% of energy savings in businesses and 80% of energy savings in households would have been achieved within three years in the absence of the support from the obligated parties or their contractors.

The evaluation method leading to these results was **criticised**, in particular for two main reasons pointed out in ENSPOL (2015):

- Answers to a counterfactual question may include significant bias.
- Sample size for the evaluation was too small and statistically insignificant.

**Bias** in the answers can introduce uncertainties in both directions, for example:

- decision makers in businesses may fear to being perceived as running inefficiently their business, and may therefore answer that they would have implemented projects with payback time between 1 and 3 years anyway, even if not the case in practice;
- due to social desirability bias, households may want to show their energy (and/or environmental) awareness by answering that they would have implemented projects anyway;
- at the opposite, end-users (both, households and businesses) may fear that the support they received be withdrawn in case they answer that they would have implemented the project anyway.

More generally, using a survey to assess additionality means the implicit assumption that surveyed end-users would have a clear idea and memory of what they would have done in a virtual scenario.

About the **sample size**, the main problem is that a small sample makes that any disaggregation of the results leads to very small sub-samples per categories (for example, per type of support offered, or per type of action). This was indeed still the case for the evaluation done in 2015, despite a much larger sample (see below).

The evaluators of the studies done in 2008 and 2012 acknowledged that using a few questions can be a cost-effective approach, but can be questioned for validity reasons. Even if complementary questions were used for cross checking the answers. They therefore recommended for the next evaluations to add qualitative questions regarding the additionality, as well as **triangulation** with less subjective methods for determining the baseline such as market data analysis.

The criticism about the evaluation done in 2012 made that DEA prepared a memorandum related to methodological challenges for assessing net impact and additionality, in order to include specific requirements in the specifications for the evaluation done in 2015. This memorandum included the demand of testing a market data analysis. But due to data issues, the study done in 2015 could finally not do it. For an example of evaluation study comparing the survey approach and market data analysis see Sulyma and Tiedemann (2016).

The table below provides an overview of the requirements included in the specifications and how it was implemented by the evaluators (study done in 2015).

The main novelty in the evaluation done in 2015 was the econometric analyses per end-use sector. Based on these analyses, the evaluators concluded that the obligation scheme did make a difference in terms of reducing the total energy consumption in Denmark compared with a situation where it would not have been implemented.

The additionality (or relative net effects) of actions in private companies would have been about 74% based on the econometric analyses, and about 46% based on the surveys of private companies. According to the evaluators, the assessment based on the econometric analyses takes into account possible spill-over effects. The 46% from the surveys correspond to the share of actions that the respondents mentioned they would not have implemented on their own initiative within 3 years.

The 2015 evaluation could not assess the net effects of the actions in households based on econometric analyses, because the data about energy consumption in households were not disaggregated enough. The effects of these actions represent a too small share of the total energy consumption in the residential sector to make possible to significantly distinguish variations due to these actions compared to other sources of variations. Based on surveys, the additionality would be about 22% (share of actions that the respondents mentioned they would not have implemented on their own initiative within 3 years).

Table 2. Comparison between the memorandum for the specifications and the methods used by the evaluators (for the evaluation done in 2015).

| **Requirements defined in the memorandum**<br>(source: ENSPOL, 2015) | **Method used by the selected evaluators**<br>(source: Deloitte et Grontmij, 2015) |
|---|---|
| Asking the end-users e.g. "is it likely that you would have implemented the energy saving measure without the involvement (e.g. subsidy or consultancy) of the operator within 1/3 years?" | Question included in the survey of the end-users<br>+ similar question asked to obligated parties and contractors |
| Econometric analysis of data - if data on the sales of different technologies related to energy usage and these technologies' energy efficiency is both available before and after implementation/increase of EEO targets. | Not included in the evaluation (data needed not available or too costly to collect) |
| Statistical analysis of a group exposed to a measure (e.g. subsidy or consultancy) and a control group that is not exposed to such measures. | Survey of a control group, however crossing or comparison with the survey of participant end-users is not included in the evaluation report |
| Statistical analysis without a control adjusting for the impact of changes in e.g. energy prices that would influence changes in energy efficiency. | Econometric analysis including variables related to the end-users and the actions (but not to changes in energy prices) |
| Top down approach macroeconomic models are used to isolate the expected changes in energy consumption as a result of firstly economic growth and changes in energy prices using e.g. price elasticity and secondly other trends such as an EEO. | Top-down regression analysis per end-use sector, based on main macro variables affecting energy consumption, using long time series. |

The evaluators noted that an explanation of the significantly higher additionality observed in businesses (compared to households) could be due to the **differences of supporting approaches** mostly used in both sectors: the use of comprehensive energy advice (like energy screening) was common in industry while very rare in households. This was confirmed by comparing the additionality rate between projects in businesses depending on the type of support they received: the average additionality rate was about 55% for projects supported by energy screening, whereas only about 15% for projects without energy screening.

The additionality rates were also analysed depending on the action type. However, the same problem arises as encountered in the evaluation done in 2012: the results per action type are difficult to analyse due to the small size of the sub-samples for each action type (for example, some sub-samples include only 1 project for the corresponding action type).

Despite the limit of very small size of the sub-samples per action type, the evaluators of the assessments done in 2015 considered that due to the larger samples used, their results were more robust than the ones of the previous evaluation (2012). They pointed in particular one of the consequences of the 2012 evaluation: deemed savings for window replacement and cavity insulation were reduced due to the estimations that their additionality was relatively low. These initiatives thus became less interesting for obligated parties, as the energy savings they may report from them were decreased. According to the evaluators, this meant that obligated parties reduced the grants and energy advice in relation to these actions, reducing end-user incentives to implement them. According to them, the estimations done in 2012 could not be completely confirmed by the evaluation done in 2015, as windows had a lower rate of answers “would have been implemented anyway” compared to most of the other action type for the residential sector. However, as the size of the sub-samples is also very small in the evaluation done in 2015, the main argument would more be that none of the evaluations provides a reliable quantitative basis for decision-making.

An attempt to improve the additionality assessment was to ask questions to a group of non-participants. About half of the households in the non-participant group answered that they were willing to implement energy efficiency improvements in their home within the next three years without subsidies, which confirms that the influence of grants would be limited for a large share of households (while it was the main approach used to support projects in the residential sector). About 40% of the households in the non-participant group would be influenced by a grant. But the answers do not show a clear threshold of grants (in %) that would have a triggering effect. Which shows that the appreciation of the grant level is rather subjective.

About the spill-over effect, results from the survey of 153 contractors were that 30% said that they experienced a “large increase” in demand from end users about energy savings projects since 2006, 38% said they experienced a “moderate increase” and 32% said they experienced no increase. However, it is not clear to what extent these changes would be due to the obligation scheme or to other actors. This is why the evaluators concluded that this approach does not provide an accurate estimation of spill-over effects, but brought some insights.

The evaluation report does not provide clear explanations about how the results of the different methods used to assess additionality would have been crossed or combined. Indeed, while combining different methods seem to be an interesting approach, cross-analyses appear difficult to perform. And even combining the results on a qualitative basis does not seem obvious.

**Added value of the evaluations for the policy**

As discussed above, the quantification of additionality raised questions about the related uncertainties. However, the successive evaluations did bring clear qualitative findings, improving the understanding of the scheme. One clear conclusion from the evaluation done in 2012 and that was confirmed and deepened in the 2015 evaluation is that the additionality rates depend not only on the type of sector (or action), but also (and maybe first) on the type of support offered (tailored energy advice showing higher additionality than grants alone).

Likewise, the findings of the evaluations were essential to improve the rules of the scheme. One example is the introduction after the second evaluation (2012) of a rule that actions with a payback time of less than 1 year could not receive a grant, as this was raised as an additionality issue.

The evaluations thus proved to be very useful to help ensuring that the rules of the scheme encourage cost-effective approaches to achieve energy savings target, and prevent money to be spent inefficiently.

The evaluations also brought an evidence base about the impacts and mechanisms of the scheme, as a valuable input to the discussions preparing the next period.

## Bibliography

### General references

Alberini, A., & Bigano, A. (2015). How effective are energy-efficiency incentive programs? Evidence from Italian homeowners. *Energy Economics*, pp. 76-85. doi:10.1016/j.eneco.2015.08.021

Asensio, O. I., & Delmas, M. A. (2017). The effectiveness of US energy efficiency building labels. *Nature Energy, 4*. Retrieved 08 16, 2018, from https://cloudfront.escholarship.org/dist/prd/content/qt9m11b9rf/qt9m11b9rf.pdf

Baumgartner, R. (2016). Chapter 12: Survey Design and Implementation for Estimating Gross Savings Cross-Cutting Protocol. *The Uniform Methods Projects: Methods for Determining Energy-Efficiency Savings for Specific Measures*. Retrieved 08 07, 2018, from https://www.nrel.gov/docs/fy17osti/68568.pdf

Bertholet, J., Cabrera, D., & Patel, M. (2014). Evaluation of an Energy Efficiency Program for Small Customers in Geneva (Switzerland). *Proceedings of IEPPEC 2014*. Retrieved 08 16, 2018, from https://www.iepec.org/conf-docs/papers/2014/Jean-Luc%20Bertholet.pdf

Blumstein, C. (2009). Program evaluation and incentives for administrators of energy efficiency programs: can evaluation solve the principal/agent problem? *Proceedings to the ECEEE 2009 Conference*.

Breitschopf, B., Voswinkel, F., & Schlomann, B. (2018). How relevant are free-rider effects for taget achievement? *Proceedings of the IEPPEC 2018 Conference*. Retrieved 08 07, 2018, from http://www.ieppec.org/wp-content/uploads/2018/06/Breitschopf_paper_vienna.pdf

Broc, J.-S., Adnot, J., Bourges, B., Thomas, S., & Vreuls, H. (2009). The development process for harmonised bottom-up evaluation methods of energy savings. (W. I. EMEEES. Edited by IEE, Ed.)

Bundgaard, S. S., Togeby, M., Dyhr-Mikkelsen, K., Sommer, T., Kjærbye, V. H., & Larsen, A. E. (2013). Spending to save: evaluation of the energy efficiency obligation in Denmark. *ECEEE Summer Study 2013*. Retrieved 08 16, 2018, from https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2013/7-monitoring-and-evaluation/

Collins, M. C. (2016). Willingness-to-Pay and Free-Riding in a National Energy Efficiency Retrofit Grant Scheme: A Revealed Preference Approach. *ESRI Working Paper (551)*. Retrieved 08 07, 2018, from http://www.esri.ie/pubs/WP551.pdf

Cooper, A. (2018). Evaluating energy efficiency policy: understanding the 'energy policy epistemology'may explain the lack of demand for randomised controlled trials. *Energy Efficiency, 4*, pp. 997-1008.

Dougherty, A., & Van de Grift, S. C. (2014). Behavioral Energy Feedback Program Evaluations: A Survey of Current Knowledge and a Call to Action. *Proceedings to the 2014 IEPPEC Conference*. Retrieved 08 07, 2018, from https://www.iepec.org/wp-content/uploads/2018/02/Anne-Dougherty.pdf

Eichhammer, W., Boonekamp, P., Labanca, N., Schlomann, B., & Thomas, S. (2008). Distinction of energy efficiency improvement measures by type of appropriate evaluation method. Final Report on Evaluation and Monitoring for the EU Directive on Energy End-Use Efficiency and Energy Services. (F. ISI, E. e. savings, & W. Institute, Eds.)

Eto, J., Vine, E., Sonnenblick, L., & Payne, C. (1996). The Total Cost and Measurement performance of Utliity-Sponsored Energy Efficiency Programs. *The Energy Journal*, pp. 31-51.

European Commission. (2016). Directive of the Eurpean Parliament and of the Council amending Directive 2012/27/EU on energy efficiency. proposed EED 2016. *Official Journal of the European Union*. Retrieved 08 07, 2018, from http://eur-lex.europa.eu/resource.html?uri=cellar:efad95f3-b7f5-11e6-9e3c-01aa75ed71a1.0009.02/DOC_1&format=PDF

Frederiks, E. R., Stenner, K., Hobman, E. V., & Fischle, M. (2016). Evaluating energy behavior change programs using randomized controlled trials: Best practice guidelines for policymakers. *Energy Research & Social Science*, pp. 147-164.

Friedmann, R. (2007). Maximizing Societal Uptake of Energy Efficiency in the New Millenium: Time for Net-to-Gross to Get Out of the Way? *Proceedings to the Energy Program Evaluation Conference*. Retrieved 08 07, 2018, from https://www.iepec.org/conf-docs/papers/2007PapersTOC/papers/63_1065_ab_586.pdf

Gaffney, K., Papadamou, & Agnew, K. (2015). State-of-the-art in behaviour change program evaluation. *Proceedings of the ECEEE 2015 Summer Study*. Retrieved from https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2015/8-monitoring-and-evaluation-building-confidence-and-enhancing-practices/

Galhotra, J., & Randazzo, K. (2015). Show Me the Attribution: Recommended Methods for Estimating Impacts for Finance Programs. *Proceedings of IEPEC 2015*. Retrieved 08 16, 2018, from https://www.iepec.org/wp-content/uploads/2018/02/2015paper_galhotra_randazzo-1.pdf

Gandhi, R., Knittel, C. R., Pedro, P., & Wolfram, C. (2016). Running randomized field experiments for energy efficiency programs: A practitioner's guide. *Economics of Energy & Environmental Policy, 2*, pp. 7-26. Retrieved 08 16, 2018, from http://ceepr.mit.edu/files/papers/2016-015.pdf

Granell, R., Wallom, D., Janda, K. B., & Bright, S. (2017). Quantifying the impact of green leasing on energy use in a retail portfolio: limits to big data analytics. *Proceedings of ECEEE 2017 Summer Study*, pp. 1849-1859. Retrieved 08 16, 2018, from https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2017/8-monitoring-and-evaluation-building-confidence-and-enhancing-practices/

Griechisch, D., & Unger, C. (2014). Spreading Myths around Buillding Renovation: Information Failure Prevails in Germany. *2014 International Energy Policy & Programme Evaluation Conference (IEPPEC)*.

Hannigan, E., & Cook, J. (2015). Matching and VIA: Quasi-Experimental Methods in a World of Imperfect Data. *Proceedings of IEPEC 2015*. Retrieved 08 16, 2018, from https://www.iepec.org/wp-content/uploads/2018/02/2015paper_hannigan_cook-1.pdf

Hummer, J. (2016). Improving Attribution Analysis through Trade Ally Research. *Proceedings of IEPPEC 2016*. Retrieved 08 16, 2018, from http://www.ieppec.org/wp-content/uploads/2016/05/Paper-Hummer-2.pdf

James, M., & Ambrose. (2016). Evaluating the Impact of Retrofit and Behaviour Change Interventions on Energy Efficiency in Low Income Households. *Proceeding of IEPPEC 2016*. Retrieved 08 16, 2018, from http://www.ieppec.org/wp-content/uploads/2016/05/Paper-James.pdf

Johnson, K. (2014). Decision-Framework for Determining Net Savings Approach - Supplemental Document #2 to Principles and Guidance. *Northeast Energy Efficiency Partnerships*.

Kendel, A., Lazaric, N., & & Maréchal, K. (n.d.). What do people 'learn by looking'at direct feedback on their energy consumption? Results of a field study in Southern France. *Energy Policy*, pp. 593-605.

Labanca, N., & Bertoldi, P. (2016). Energy Savings Calculation Methods under Article 7 of the Energy Efficiency Directive. *JRC Science for Policy Report*. Retrieved 08 07, 2018, from https://e3p.jrc.ec.europa.eu/publications/energy-savings-calculation-methods-under-article-7-energy-efficiency-directive

McClaren, M., & Bliss, R. (2018). Avoiding Being (Too Much of) A Victim of Your Own Success: Mitigating Free-ridership Losses Through Better Spillover Assessment. *Proceedings of IEPPEC 2018*. Retrieved 08 16, 2018, from http://www.ieppec.org/wp-content/uploads/2018/05/Bliss_paper_vienna.pdf

McClure, C., & Provencher, B. (2014). Energy Savings Over a 3-year Opt-in Rewards-based Residential Behaviour Programme. *Proceedings of IEPPEC 2014*. Retrieved 08 16, 2018, from https://www.iepec.org/conf-docs/papers/2014/Carly%20McClure.pdf

Meurice, J., Beaman, N., Renta, L., Jerozal, J., Dimetrosky, S., & Warren, N. (2014). Reducing Evaluation Uncertainty in Large Custom Projects. *Proceedings of IEPPEC 2014*. Retrieved 08 16, 2018, from https://www.iepec.org/conf-docs/papers/2014/Julianne%20Meurice.pdf

Miller, P., Beitel, A., & Mejia, A. (2014). California Technical Forum (Cal TF): A New Collaborative for Developing Savings Estimates for California. *Proceedings of the ACEEE 2014 Summer Study*. Retrieved 08 16, 2018, from http://aceee.org/files/proceedings/2014/data/papers/8-1074.pdf

Moser, S., Leutgöb, K. R., & Kollmann, A. (2012). Making the results of bottom-up energy savings comparable. *Therm sci, 3*, pp. 687-702. doi:10.2298/TSCI120106119M

Mukai, T., & K., N. (2014). Evaluating a Behavioral Demand Response Trial in Japan: Evidence from the Summer of 2013. *Proceedings of IEPPEC 2014*. Retrieved 08 16, 2018, from https://www.iepec.org/conf-docs/papers/2014/Toshihiro%20Mukai.pdf

Nauleau, M.-L. (2014). Free-riding on tax credits for home insulation in France. An econometric assessment using panel data. *Energy Economics*, pp. 78-92. doi:10.1016/j.eneco.2014.08.011

Nilsson, L., Lopes, C., Fonseca, P., & Araújo, M. L. (2008). Assessment of existing evaluation practice and experience. (E. D. D1, Ed.)

Northeast Energy Efficiency Partnerships. (2016). Gross Savings and Net Savings: Principles and Guidance. Retrieved 08 07, 2018, from http://www.neep.org/sites/default/files/FINAL%20GS%20and%20NS%20Principles%20and%20Guidance%20Document_2016May17.pdf

Olsthoorn, M., Schleich, J., Gassmann, X., & Faure, C. (2018). Free riding and rebates for residential energy efficiency upgrades: A multi-country contingent valuation experiment. *Energy Economics*.

Paramovona, S., & Thollander, P. (2016). Ex-post impact and process evaluation of the Swedish energy audit policy programme for small and medium-sized enterprises. *Journal of Cleaner Production*, pp. 932-949. doi:10.1016/j.clepro.2016.06.139

Ridge, R., Willems, P., Fagan, J., & Randazzo, K. (2009). The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating the Net-To-Gross Ratio. *Proceedings of IEPEC 2009*. Retrieved 08 16, 2018, from https://www.iepec.org/wp-content/uploads/2018/02/015-8.pdf

Rietbergen, M. G., Farla, J. C., & Blok, K. (2002). Do agreements enhance energy efficiency improvement? Analysing the actual outcome of long-term agreements on industrial energy efficiency improvement in the Netherlands. *Journal of Cleaner Production*, pp. 153-163.

Scheer, J., Clancy, M., & Hógáin, S. N. (2013). Quantification of energy savings from Ireland's Home Energy Saving scheme: an ex post billing analysis. *Energy Efficiency, 1*, pp. 35-48.

Schellenberg, J., Savage, A., Blundell, M., Cook, J., & Smith, B. A. (2015). Losing Control: What Will Happan if Randomized Controlled Trials are Phased Out of Behavioral Program Evaluation? *2015 International Program Evaluation Conference*. Retrieved 08 07, 2018, from https://www.iepec.org/wp-content/uploads/2018/02/2015paper_schellenberg_savage_blundell_cook_smith-1.pdf

Schiller, S. R. (2007). National Action Plan for Energy Efficiency. Model Energy Efficiency Program Impact Evaluation Guide. (S. C. Inc, Ed.) *National Action Plan for Energy Efficiency*. Retrieved 08 07, 2018, from https://www.epa.gov/sites/production/files/2017-06/documents/evaluation_guide.pdf

Schleich, J., Fleiter, T., Hirzel, S., Schlomann, B., Mai, M., & Gruber, E. (2015). Effect of energy audits on the adoption of energy-efficiency measures by small companies. *Proceedings of the ECEEE 2015 Summer Study*, pp. 1827-1836. Retrieved 08 16, 2018, from https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2015/8-monitoring-and-evaluation-building-confidence-and-enhancing-practices/

Skumatz, L. (2009). Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior. With assistance of Edward Wine. *CIEE Berkeley*. Retrieved 08 07, 2018, from ttps://library.cee1.org/sites/default/files/library/10517/CIEE_Behavior_White_Paper_-_Skumatz_2009.pdf

Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics, 1*, pp. 1-21.

Sulyma, I., & Tiedemann, K. (2016). Estimating Lighting Programme Impacts Using Self Report Surveys and. *Proceedings of IEPPEC 2016*. Retrieved 08 16, 2018, from http://www.ieppec.org/wp-content/uploads/2016/05/Paper-Sulyma.pdf

Tiefenbeck, V., Staake, T., Schöb, S., & Götte, L. (2016). Better get focused: How feedback on a specific behavior can reduce energy consumption. *Proceedings of IEPPEC 2016*. Retrieved 08 16, 2018, from http://www.ieppec.org/wp-content/uploads/2016/05/Paper-Tiefenbeck-1.pdf

Train, K. (1994). Estimation of net savings from energy-conservation programs. *Energy, 4*, pp. 423-441.

Vine, E., & Sathaye, J. (1999). Guidelines for the Monitoring, Evaluation, Reporting, Verification, and Certification of Energy-Efficiency Projects for Climate Change Mitigation. Retrieved 08 07, 2018, from https://www.osti.gov/scitech/servlets/purl/7288,

Vine, E., Sullivan, M., Lutzenhiser, L., Blumstein, C., & Miller, B. (2018). Response to: "Evaluating energy efficiency policy: understanding the 'energy policy epistemology' may explain the lack of demand for randomized controlled trials," by Adam Cooper, Energy Efficiency, published online 26 January 2018. *Energy Efficiency*, pp. 1-2. Retrieved from https://link.springer.com/article/10.1007%2Fs12053-018-9650-8

Violette, D. M., & Rathbun, P. (2017). Chapter 21: Estimating Net Savings - Common Practices. *The Uniform Methods Projects: Methods for Determining Energy-Efficiency Savings for Specific Measures*. Retrieved 08 07, 2018, from https://www.nrel.gov/docs/fy17osti/68578.pdf

Violette, D., & Agapay-Read, L. (2016). Programme Attribution Analysis : Use of Self-Reports and Triangulation. *Proceedings of IEPPEC 2016*. Retrieved from http://www.ieppec.org/wp-content/uploads/2016/05/Paper-Violette-1.pdf

Wade, J., & Eyre, N. (2015). Energy Efficiency Evaluations: The evidence for real energy savings from energy efficiency programmes in the household sector. *UKERC Technology & Policy Assessment Function (REF UKERC/RR/TPA/2015/001*. Retrieved 08 07, 2018, from http://www.ukerc.ac.uk/programmes/technology-and-policy-assessment/energy-efficiency-evaluation.html

**References about example n°1 ([GERMANY] Energy Efficiency Fund)**

Voswinkel, F. (2018). Accounting of energy savings in policy evaluation. How to get at least 8 different (correct!) results from the same data. *Proceedings of the IEPPEC 2018 Conference*. Retrieved 08 07, 2018, from http://www.ieppec.org/wp-content/uploads/2018/05/Voswinkel_paper_vienna.pdf

Schlomann, B., Hirzel, S., Nabitz, L., Heinrich, S., Jessing, D., Paar, A., Pehnt, M. & Antoni, O. (2017). From Targets to Impacts: Eight Steps for Evaluating Energy Efficiency Policies. Proceedings of IEPEC 2017. Retrieved 16 08, 2018, from http://www.iepec.org/2017-proceedings/polopoly_fs/1.3718027.1502993147!/fileserver/file/798246/filename/104.pdf

**References about example n°2 ([DENMARK] Energy Companies' Energy-Saving Efforts)**

Bundgaard, S. S., Togeby, M., Dyhr-Mikkelsen, K., Sommer, T., Kjærbye, V. H., & Larsen, A. E. (2013). Spending to save: evaluation of the energy efficiency obligation in Denmark. Proceedings of the ECEEE 2013 Summer Study. Retrieved 16 08, 2018, from https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2013/7-monitoring-and-evaluation

Deloitte et Grontmij, 2015. Evaluering af energiselska-bernes energispareindsats (Evaluation of energy savings effort for utilities). Final evaluation report, March 2015. Retrieved 16 08, 2018, from https://ens.dk/ansvarsomraader/energibesparelser/energiselskabers-energispareindsats/kontrol-og-evalueringer

ENSPOL, 2015. Report on existing and planned EEOs in the EU – Part I: Evaluation of existing schemes. Deliverable D2.1.1 of the ENSPOL project, March 2015 (see pp.181-199 about the Danish EEO scheme). Retrieved 16 08, 2018, from http://enspol.eu/results

Sulyma, I. & Tiedemann, K. (2016). Estimating Lighting Programme Impacts Using Self Report Surveys and Market Sales Analysis. *Proceedings of IEPPEC 2016*. Retrieved 16 08, 2018, from http://www.ieppec.org/wp-content/uploads/2016/05/Paper-Sulyma.pdf

Togeby, M., et al., 2009. Danish energy efficiency policy: revisited and future improvements. Proceedings of the ECEEE 2009 Summer Study, pp.299-310. http://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2009/Panel_2/2.043

### How to cite this case study

Voswinkel, F., Broc, J.S., Breitschopf, B., Schlomann, B., 2018. Evaluating net energy savings. Topical case study of the EPATEE project, funded by the European Union's Horizon 2020 programme.