Kinks Know More: Policy Evaluation Beyond Bunching with an Application to Solar Subsidies

Stefan Pollinger

Kinks Know More: Policy Evaluation Beyond Bunching with an Application to Solar Subsidies*

Stefan Pollinger †

First Version: November 12, 2020

This Version: August 16, 2023

Please find the latest version of the paper here.

Abstract

This paper demonstrates that kinks or discontinuities in incentive schemes (e.g., taxes, subsidies, or prices) simultaneously identify agents' intensive and participation margin responses. The proposed semi-nonparametric estimator enables the evaluation of such schemes when existing kink and discontinuity methods are inapplicable due to the presence of both margins. The paper applies the estimator to evaluate the German subsidy for rooftop solar panels, a cornerstone in the global efforts to transit towards a carbon-free economy. Compared to a linear scheme, the government's nonlinear subsidy reduces costs by 0.14 per cent; an optimal nonlinear scheme would more than triple this gain. Ignoring the participation margin when optimising the subsidy would increase costs substantially. The results highlight the importance of estimating both margins for optimal policy design.

Keywords: Nonlinear Incentive Schemes, Bunching, Participation Margin, Solar Subsidies.

JEL codes: H2, H30, H41.

*I am especially grateful to Christian Hellwig, Nicolas Werquin, Jean-Pierre Florens, and Christian Gollier for their supervision and guidance. I thank Andrew Atkeson, Alan Auerbach, Richard Blundell, Christian Bontemps, Nicolas Bonneton, Clément de Chaisemartin, Emeric Henry, Damon Jones, Camille Landais, Stefan Lamp, Alix de Loustal, Isabelle Mejean,Thierry Magnac, Itzhak Rasooly, Jean-Marc Robin, Whitney Newey, Emmanuel Saez, Augustin Tapsoba, Jean Tirole, and Danny Yagan for helpful comments and discussions. I thank the participants of the internal workshops at the Toulouse School of Economics, the Public Finance Seminar at UC Berkeley, the ENTER workshop at the University College London, the International Conference on Public Economic Theory 2018, the Conference of the European Association for Research in Industrial Economics 2018, the Annual Conference of the European Association of Environmental and Resource Economists 2019 and 2023, the Congress of the Institute of International Public Finance 2019 and 2023, and the European Winter Meeting of the Econometric Society 2019. All errors are mine. I declare that I have no relevant or material financial interests that relate to the research described in this paper.

†Stefan Pollinger, Assistant Professor, Sciences Po Paris; email address: stefan.pollinger@sciencespo.fr; website: www.stefanpollinger.com.

1 Introduction

Nonlinear incentive schemes have a wide range of policy applications in subsidy programmes, taxation, product pricing, and public transfers. A challenge in their optimal design is reliably estimating how agents react to them at the participation and intensive margin.[1] When agents react only at the participation margin, the regression kink design (Card, Lee, Pei, and Weber, 2015) can be used to exploit kinks, i.e., discontinuities in the marginal incentive scheme. Correspondingly, when there is only an intensive margin response, the bunching estimator (Saez, 2010) can be used. However, very often, agents respond at both margins simultaneously. In such cases, it is necessary to estimate the participation *and* the intensive margin response to evaluate an incentive scheme. Unfortunately, the mentioned estimators are not applicable because each margin biases the estimate of the other margin.

This paper's methodological contribution is to propose an estimator for agents' responses at both margins. The estimator exploits the effect of kinks in an incentive scheme to identify the two margins jointly. Kinks have two effects on the observable distribution of agents' choices. First, they are responsible for bunching, i.e., a discrete mass of agents that choose the kink point. Second, they cause a slope change in the choice-distribution at the kink point. Both effects are observable and depend on the magnitude of the two reaction-margins distinctly. Therefore, the two observable effects enable the simultaneous estimation of the two margins. While the classic bunching estimator relies on a parametric functional form assumption (Blomquist, Newey, Kumar, and Liang, 2021), this paper demonstrates that both margins are identified under a weaker assumption: the choice-distribution absent the kink is analytic (intuitively, very smooth). In light of this identification result, the proposed semi-nonparametric estimator uses sieve estimation (Chen, 2007).[2] Moreover, the paper proposes a data-driven selection of the nonparametric specification, which increases the robustness to specification bias. Adapting the estimator to exploit notches, i.e., discontinuities in incentive schemes, is straightforward. Kinks and notches are common features of economic incentive schemes. By exploiting them, the estimator circumvents the need for exogenous variation or instrumental and control variables; the

[1] The participation margin is also called the extensive margin.

[2] Following Chen (2007), I call a model semi-nonparametric if the model's parametric and nonparametric parts are of interest.

estimation only uses the easily observable distribution of agents' choices. These low informational and identifying requirements increase the potential applicability for the evaluation of taxes, subsidies, transfers, and product prices.

As an applied contribution, this paper evaluates the German subsidy for rooftop solar panels. It is a very prominent recent example of a large and successful deployment subsidy for a nascent green technology. Gerarden (2022) and Nemet (2019) argue that the programme was instrumental in driving the enormous price decreases in solar panels over the last two decades, making them one of the cheapest sources of electricity in 2021 (IRENA, 2022). More generally, green deployment subsidies have moved to the forefront of climate action (Podesta, 2023). They can facilitate the direct displacement of polluting activities and catalyse green innovation (Acemoglu, Aghion, Bursztyn, and Hemous, 2012).

However, one major caveat of such subsidies is their potential burden on public finances. To illustrate, the annual payments in the German programme constituted 0.6 per cent of total government expenditure, mainly benefiting wealthy owners of rooftops. A strategy to mitigate these costs is subsidies that are nonlinear in a technology's attributes. In Germany, the government implemented a subsidy that is nonlinear in the capacity adopted by households and firms.[3] The subsidy schedule was piecewise linear with multiple kinks. The marginal subsidy rates decreased discontinuously at the kink points, curtailing payments for profitable large adoptions. Two questions arise. First, are the kinks in the German subsidy schedule effective at reducing the programme's public costs without compromising aggregate capacity?[4] Second, what is the most cost-efficient nonlinear subsidy schedule to achieve a given aggregate capacity goal? The answers to both questions depend on how adopters react to the subsidy at the participation and intensive margin. The participation margin determines how many adopters participate in the scheme, and the intensive margin determines the capacity choice of participants.

Applying the novel estimator to the early stage of the German programme reveals that responses at both margins are significant and large. The estimates enable the evaluation of public costs under various counterfactual subsidy schemes. Assuming the government aims to im-

[3]The capacity of a solar panel is the amount of electricity it produces under standardised conditions. It depends on the size and efficiency of the adopted panel.

[4]Aggregate capacity is the sum of the capacity of installations in a given period.

plement the observed aggregate capacity at the lowest public cost, the counterfactual exercises yield four main results.[5] The first exercise solves for the linear subsidy that achieves the capacity goal. Compared to this benchmark, the government's subsidy modestly reduces costs by 0.14 per cent. The second exercise solves for the optimal nonlinear subsidy. It more than triples cost savings to 0.45 per cent. Surprisingly, while the government's scheme can be improved, the gains from using a nonlinear scheme remain modest overall. To better understand what limits the scope for curtailing costs, a third thought experiment assumes adopters can only respond at the intensive margin. Compared to the linear benchmark, the optimal nonlinear subsidy would reduce costs by 8.7 per cent in this case. The exercise reveals that the participation margin is responsible for the limited scope for cost reductions when both margins are active. A fourth counterfactual exercise assumes the policymaker wrongly ignores the participation margin. She implements the optimal intensive schedule derived in the third thought experiment while adopters react at both margins. As a consequence, costs increase by 3.1 per cent instead of decreasing. The exercise highlights that considering both margins is crucial when designing nonlinear incentive schemes. Optimising schemes based on intensive margin estimates alone may even be counterproductive.

Related literature. Methodologically, this paper contributes to the bunching and regression kink design literature. The bunching design estimates intensive margin responses using bunching at kink and notch points within the budget set (see Saez, 2010, Chetty, Friedman, Olsen, and Pistaferri, 2011, Kleven and Waseem, 2013, and reviews by Kleven, 2016 and Bertanha, Caetano, Jales, and Seegert, 2023). The classic bunching estimator does not account for participation margin responses.[6] I find that considering both margins is essential not only for the policy evaluation but also for the estimation: ignoring the participation margin leads to a 20

[5]Contrary to a pure Pigouvian analysis, where the optimal marginal subsidy is equal to the marginal environmental benefit, this objective explicitly takes into account the social cost of distributing information rents to adopters.

[6]Important exceptions are Gelber, Jones, Sacks, and Song (2020) and Marx (2018), who estimate intensive and participation margin responses exploiting their data's panel dimension. The absence of a panel dimension in my application renders such an approach infeasible. A benefit of the simultaneous estimation in my paper is that it solely relies on the cross-sectional distribution of adoptions to estimate both margins. Other methodological contributions in the bunching literature distinct from mine are Cox, Liu, and Morrison (2020), who extend the approach to a two-dimensional intensive margin and Garbinti, Goupille-Lebret, Muñoz, Stantcheva, and Zucman (2023), who exploit panel data.

per cent downward bias in my intensive margin estimate. The participation margin responses at kink points can be estimated using a regression kink design (see Nielsen, Sørensen, and Taber, 2010 and Card et al., 2015). However, if there is a positive intensive margin, the classic regression kink design suffers from endogeneity because agents select into treatment.[7] The simultaneous estimation in this paper addresses these issues of the classic approaches. Blomquist et al. (2021), Bertanha, McCallum, and Seegert (2021), Moore (2021), and Goff (2022) criticise the fact that the classic bunching estimator uses a parametric functional form assumption on the type distribution to identify the intensive margin elasticity. Therefore, like structural estimators, it may suffer from specification bias.[8] To address this critique, I assume that the type-distribution is locally representable by a convergent power series, i.e., analytic, as in Gautier and Gaillac (2021) and Iaria and Wang (2022). Moreover, I check the robustness of the specification on untreated data.

Theoretically, my counterfactual exercises build heavily on the literature on second-degree price discrimination with intensive and participation margin responses (see Rochet and Stole, 2002, Saez, 2002, and Jacquet, Lehmann, and Van der Linden, 2013). Using the theoretical results in Rochet and Stole (2002), I solve for the optimal mechanism in an empirical setting. They show that when the type distribution is uniform, the optimal allocation with only intensive margin responses bounds the optimal allocation with both margins. I find that this result is not robust to a more general type distribution. E.g., my type distribution is log-normal at the bottom and Pareto at the top; the optimal allocation is close to but outside their bounds. More worryingly, I find that the bounds provided by the optimal intensive schedule are not very tight. Implementing the schedule is even counterproductive since, compared to a simple linear schedule, it increases costs considerably instead of decreasing them. The results show that the qualitative insight in Rochet and Stole (2002) provides limited guidance to the policymaker.

[7] Gerard, Rokkanen, and Rothe (2020), Bachas and Soto (2021), and Caetano, Caetano, and Nielsen (2020) propose methods to correct for endogenous sorting. These papers do not consider participation margin responses. Their methodologies are not applicable when both margins are present because each margin biases the estimate of the other margin.

[8] Ganong and Jäger (2018) and Ando (2017) raise a similar concern for the regression kink design. Due to the lack of exogenous variation in the budget set or rich covariates, the solutions to this issue proposed by Blomquist et al. (2015) and Bertanha et al. (2021) are inapplicable in my context. See Beffy, Blundell, Bozio, Laroque, and To (2019) and Aghion, Bergeaud, and Van Reenen (2021) for papers that use nonlinear budget sets for identification within a structural model.

Combining their theoretical results with estimates of the intensive *and* participation margin responses is crucial for implementing the optimal schedule.

From an applied perspective, the paper contributes to the literature evaluating subsidies for solar panels. To the best of my knowledge, it is the first paper to evaluate and optimise the cost-effectiveness of nonlinear solar subsidies. I find that nonlinearities are modestly effective at reducing the costs of the German programme. One strand of the solar literature focuses on the dynamics of the adoption decision (Burr, 2016; De Groote and Verboven, 2019; Feger, Pavanini, and Radulescu, 2022; Reddix, 2015; Gerarden, 2022). Methodologically, these papers use structural models. Another strand of the solar literature uses reduced-form methods. For example, Germeshausen (2018) uses a difference-in-difference approach to estimate the treatment effect of the introduction of a new kink in Germany in 2012.[9] The paper does not estimate elasticities at the two adoption margins nor evaluate the cost-effectiveness of counterfactual schemes.[10] It relies on a parallel trend assumption, which is unnecessary in my approach. My estimates imply a treatment effect very close to the one estimated by Germeshausen (2018), which provides evidence for the validity of the respective identifying assumption in both studies.

The rest of the paper proceeds as follows. Section 2 provides graphical evidence for responses at the two margins; Section 3 outlines the model and demonstrates identification. Section 4 presents the estimation method. Section 5 shows the empirical results and discusses their robustness. Section 6 evaluates the policy, and Section 7 concludes.

2 Policy description and graphical evidence

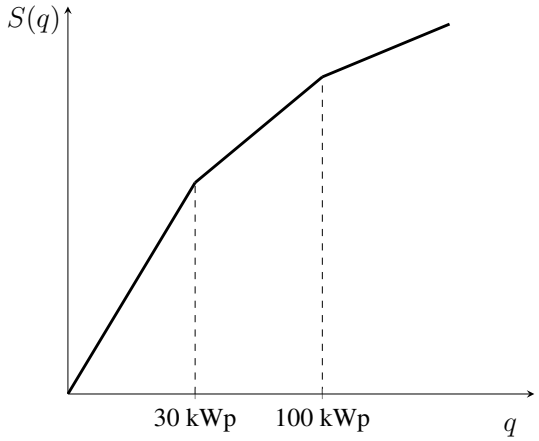
The German subsidy for solar panels was introduced on April 1st, 2000. The subsidy is a guaranteed feed-in tariff, paid per kWh (kilowatt-hour) of electricity produced.[11] A fixed tariff

[9] Hughes and Podolefsky (2015) use geographical discontinuities in California to study adoption behaviour. Such discontinuities are not available in Germany. Srivastav (2023) studies feed-in-tariffs and their effect on the financial frictions faced by adopters. Kraft-Todd, Bollinger, Gillingham, Lamp, and Rand (2018) study peer effects.

[10] Methodologically, Germeshausen (2018) follows Best and Kleven (2017). Kleven, Landais, Saez, and Schultz (2013), Ruh and Staubli (2019), Slemrod, Weber, and Shan (2017), and Besley, Meads, and Surico (2014) are similar. All these papers use a difference-in-difference approach, controlling for or using bunching. In the same vein, Myhre (2021) combines bunching with a regression discontinuity design in the time dimension.

[11] The tariffs largely surpassed market prices for electricity in this period.

Figure 1: The subsidy payment S as a function of capacity q



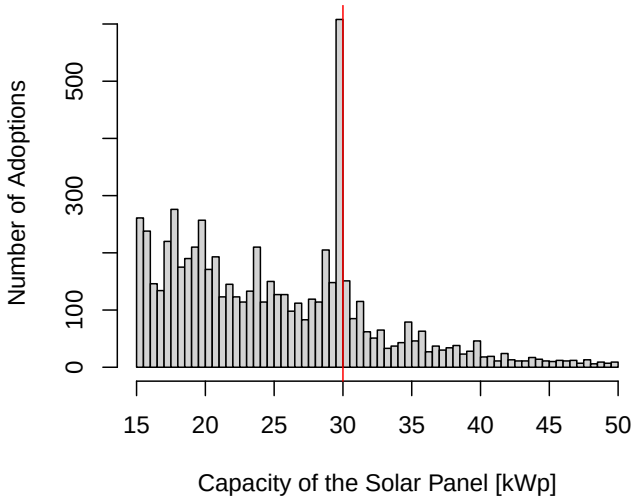
Note: The figure illustrates the kinked subsidy in Germany.

rate is guaranteed for 20 years once a household or firm decides to adopt. Typically, agents only adopt once during the sample period. Rates depend on the time point of adoption and the adopted capacity. The capacity of a solar panel, measured in kilowatt-peak (kWp), is the amount of electricity it produces under standardised conditions. Because produced electricity is proportional to adopted capacity, from 2000 to 2003, the present discounted value of payments to an adopter was simply linear in the adopted capacity (see Appendix B.1 for a formal derivation). The subsidy programme was very successful at incentivising households and firms to adopt solar panels; numbers increased rapidly in most years (see Table A1 in Appendix A). However, as a consequence, the programme became very costly. For example, total yearly payments in 2016 were 9 bio euros (Übertragungsnetzbetreiber, 2016), corresponding to 0.6 per cent of total government spending.[12] To curtail these costs, from 2004 the government introduced two kink points in the subsidy schedule. Adopters after 2004 faced the nonlinear scheme illustrated in Figure 1. Note that agents who adopted before are not affected by the policy-change. The variable q on the x-axis denotes capacity, and the variable $S(q)$ on the y-axis denotes the present discounted value of subsidy payments. At 30 kWp, the marginal subsidy rate decreased by 5% and at 100 kWp, it decreased by 1%. The rationale for this nonlinear scheme was reducing payments for profitable large adoptions.

The kinks affected the adopters' behaviour. Figure 2 shows the histogram of all solar panel

[12]The total government spending in 2016 was 1,390 billion euros (DESTATIS, 2023).

Figure 2: The histogram of adoptions in 2004

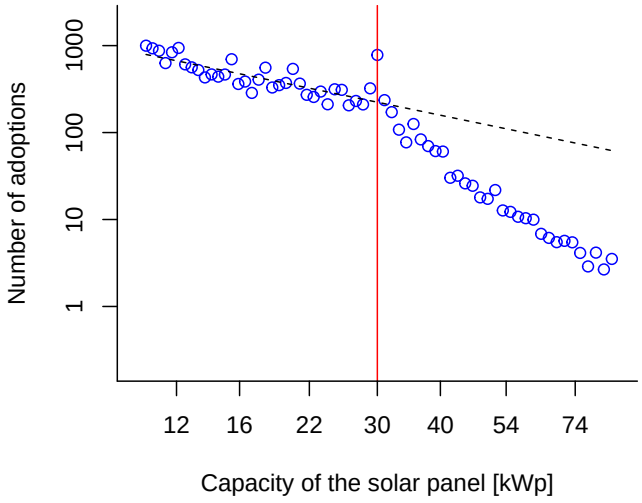


Note: The red line marks the kink point at 30 kWp. At the kink point, there is a visible mass point, i.e., bunching.

adoptions in the year 2004 around the kink point at 30 kWp. The x-axis shows the capacity in kWp. The y-axis shows the number of adopters in each column of the histogram. The red line marks the kink point at 30 kWp. The figure shows that many adoptions bunch at the kink point. Section 3 shows that bunching is reduced-form evidence for an intensive margin response, i.e., a response in the capacity-choice of adopters. The lower the marginal subsidy rate, the lower the capacity a specific adopter installs. Intuitively, adopters above the kink point reduce capacity, shifting the distribution to the left and creating excessive mass at the kink point.

Figure 3 shows the same histogram on a logarithmic scale. The number of observations in a bin is normalised by the bin size. There is a visible slope change in the distribution at the kink point. Section 3 shows that the slope change is evidence for a participation/extensive margin response. Participation depends on the total subsidy payment. The lower the payment, the fewer households and firms adopt solar panels. Intuitively, the kink in the subsidy reduces profit for adopters above the kink point, which reduces their participation.

Figure 3: Histogram of adoptions in 2004 (logarithmic scales)



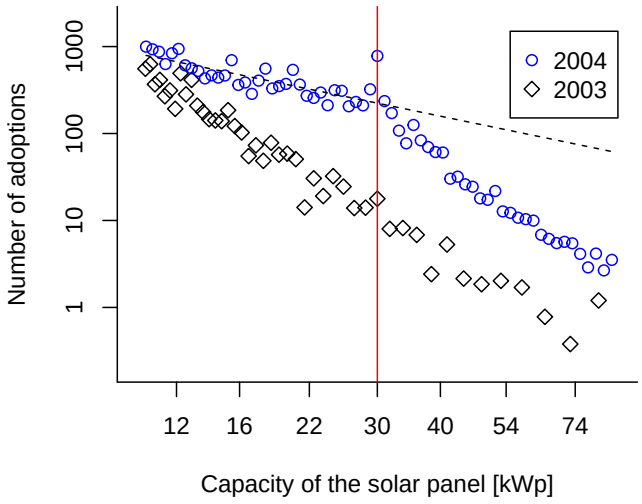
Note: The red line marks the kink point at 30 kWp. At the kink point, there is a visible mass point and slope change in the distribution. The number of observations in a bin is normalised by the bin size.

Figure 4 additionally shows the distribution of adopters who adopted in 2003 when the subsidy was linear. The distribution is smooth around the future kink point. There is no significant bunching mass or visible slope change as in 2004. It suggests that the slope change and bunching are indeed caused by the kink, hence evidence for responses at the two margins. On a side note, the subsidy rate in 2003 was lower than in 2004. While it is not directly relevant to the shape of the distributions, it explains why the number of adopters in 2004 was higher than in 2003.

One could suspect that the slope change in the distribution in 2004 is caused by a trend that adds concavity to the distribution over time. Figure 5 shows the histogram of adoptions from 2000 to 2002. It pools these years to have a sufficiently large sample size. The figure shows no graphical evidence for a time trend in the concavity of the distribution. Moreover, Figure 6 shows that the histogram in 2005 has the same pattern as the histogram in 2004. Therefore, the pattern in 2004 is not a particularity of that year. See Appendix A for a more detailed description of the data.

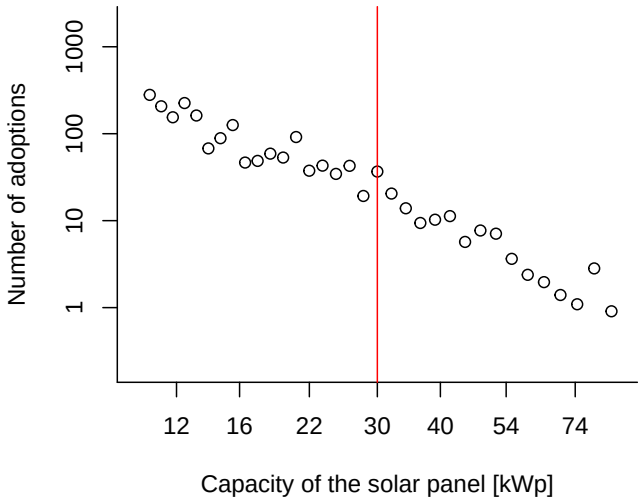
Since this paper focuses on deployment subsidies for early-stage technologies, it studies

Figure 4: Histogram of adoptions in 2003 and 2004



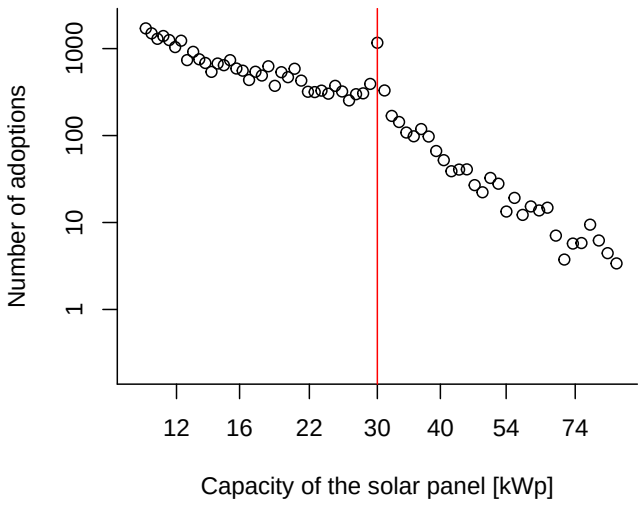
Note: The red line marks the kink point. Scales are logarithmic. Under the linear subsidy in 2003, there is no significant mass point or visible slope change in the distribution.

Figure 5: Histogram of adoptions in the years 2000 to 2002



Note: The red line marks the kink point. Scales are logarithmic. As in 2003, there is no significant mass point or visible slope change in the distribution.

Figure 6: Histogram of adoptions in 2005



Note: The red line marks the kink point. Scales are logarithmic. As in 2004, there is a visible mass point and slope change in the distribution.

the first years of the German programme from 2000 to 2008. Over this period, households and firms sell all the produced electricity to the government at the subsidised rate. Therefore, the problem is equivalent to a procurement problem: the principal (i.e., the government) procures the installation of capacity to agents (i.e., households and firms). To evaluate and optimise the subsidy's cost-effectiveness, it is necessary to know how agents react to it at the participation and intensive margin. Subsidy rates and prices for solar panels vary over time. However, in a large market like Germany, this variation is endogenous to demand, making it unsuitable for identifying the two margins. Therefore, the following section exploits the effect of kinks in the subsidy schedule on the cross-sectional distribution of adopters' behaviour. It shows that when economic agents react to incentives at the intensive and the participation margin, a kink in the incentive scheme causes a mass point and a slope change in the distribution of agents' choices. Quantitatively, the magnitude of the two responses is related to the size of the mass point and the slope change, which enables their identification.

3 The model

Assume there is a mass of heterogeneous adopters indexed by i. They produce capacity $q \in \mathbb{R}^+$ for which they receive subsidy payments. The expected present discounted value of subsidy payments $S(q)$ is a function of the adopted capacity q. Adopters solve a standard maximisation problem:

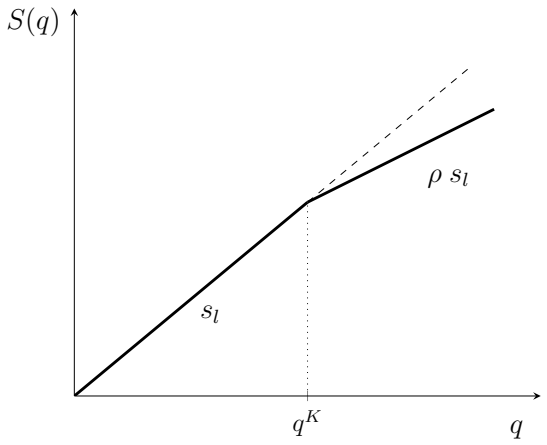
$$\pi_v^i = \max_q S(q) - c_v^i(q), \tag{1}$$

and participate if and only if $\pi_v^i \geq c_f^i$. The function $c_v^i(q)$ denotes the variable cost of adopter i to adopt capacity q, the variable π_v^i denotes the optimal variable profit of adopter i, and the variable c_f^i denotes the fixed cost of adopter i. The fixed and variable costs are heterogeneous among adopters and contain all monetary and non-monetary costs net of unobservable benefits. In practice, the German subsidy for solar panels is paid as a feed-in tariff. Appendix B.1 shows that Problem (1) implicitly accounts for subsidy payments via a feed-in tariff. In particular, the cost function accounts for all adopter-specific heterogeneity due to climate conditions and discounting of future payments.

The variable and fixed costs fully determine the adopters' behaviour. Therefore, this paper follows a sufficient statistics approach. It suffices to study the properties of the variable and fixed costs instead of their components' properties. For example, these components are: the monetary costs of the installation, warm-glow preferences for solar panels, the opportunity and aesthetic costs of using space on the roof, opportunity costs of time and money, the opportunity cost of adopting in the future, and direct benefits from consuming electricity produced by the solar panel.[13] An adopter can increase capacity by using more area on the roof or by adopting panels of higher efficiency (i.e., higher capacity per area). The variable cost function $c_v^i(.)$ is continuous and increasing. I do not assume the function is convex everywhere. There can be ranges of increasing returns to scale. However, adopters face space constraints. Moreover, the

[13]For an example of heterogeneous fixed costs, consider two firms with the same roof size. Firm one, e.g., an innovative start-up, is very productive and has high opportunity costs of time. Firm two, e.g., a traditional farm, is not very productive. The opportunity costs of time are low, and the firm is already familiar with the administrative process of receiving subsidy payments. Arguably, firm one has a higher fixed cost than firm two.

Figure 7: The kinked subsidy S_k and the counterfactual subsidy S_l



Note: The thick solid line shows the kinked subsidy S_k. The dashed line shows the counterfactual subsidy S_l. The variable s_l denotes the marginal subsidy rate below the kink point; q^K denotes the kink point, and ρ denotes the relative change in the marginal subsidy rate at the kink point.

cost of increasing capacity through more efficient panels is convex.[14] Therefore, the function $c_v^i(q)$ is convex for q large enough. Note that the optimal choice of q is always in the convex range of $c_v^i(q)$.

The subsidy $S(.)$ can take two forms: the observed kinked subsidy $S_k(.)$ and the counterfactual linear subsidy $S_l(.)$. The kinked subsidy $S_k(q)$ is:

$$S_k(q) = s_l q, \qquad \text{for } q \leq q^K; \tag{2}$$

$$S_k(q) = s_l q^K + (q - q^K)\rho s_l, \qquad \text{for } q > q^K. \tag{3}$$

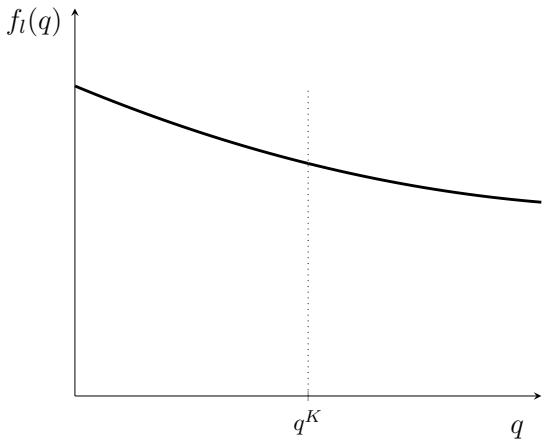
The kink point is denoted by q^K; s_l is the marginal subsidy rate below the kink point, and $\rho\ s_l$ is the marginal subsidy rate above the kink point, where $\rho \in (0, 1)$ is the relative change in subsidy rates. The counterfactual linear subsidy $S_l(q)$ is:

$$S_l(q) = s_l q, \qquad \text{for all } q. \tag{4}$$

Figure 7 illustrates both subsidies.

[14] The more efficient a panel, the higher the resource costs to increase its efficiency further. Therefore, the price of panels is convex in efficiency.

Figure 8: The counterfactual measure $f_l(q)$



Note: The counterfactual measure $f_l(.)$ is the distribution of adoptions under the counterfactual linear subsidy $S_l(.)$. Its key property is its smoothness around the hypothetical kink point.

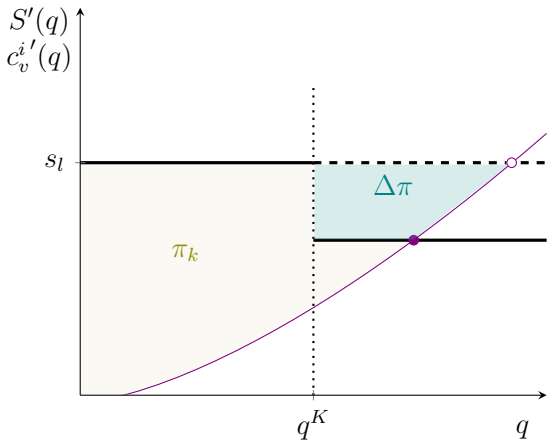
3.1 The graphical intuition behind identification

This section gives graphical intuition on how the distribution of adoptions depends on the intensive and the participation margin. Denote by $f_k(.)$ the observable distribution of adopters' choices under the kinked subsidy S_k and by $f_l(.)$ the counterfactual distribution of adopters' choices under the linear counterfactual subsidy S_l. Technically, $f_l(.)$ and $f_k(.)$ are measures. For example, for any interval of capacity $[q_1, q_2]$, $\int_{q_1}^{q_2} f_k(q)dq$ is the mass of adopters in the interval under the subsidy $S_k(.)$. Intuitively, one can think of $f_l(.)$ and $f_k(.)$ as densities that do not integrate to one.

Suppose the counterfactual linear subsidy S_l is in place. Figure 8 illustrates a possible measure of adoptions f_l. Without loss of generality, it depicts a decreasing measure. The exact shape is not essential. The only important property for the results presented later in this section is the smoothness of the counterfactual measure around the hypothetical kink point. Section 3.3 below states the precise mathematical smoothness assumption (i.e., Assumption 3).

In comparison, suppose the kinked subsidy S_k, illustrated in Figure 7, is in place. I explain its effect on the distribution of adoptions using a hypothetical change in the subsidy schedule from S_l to S_k. Note that this is a thought experiment to illustrate the effect of the kink. To estimate the two response margins, I do not exploit a change in the subsidy schedule over time

Figure 9: The effect of the kinked scheme S_k on adopters well above the kink point

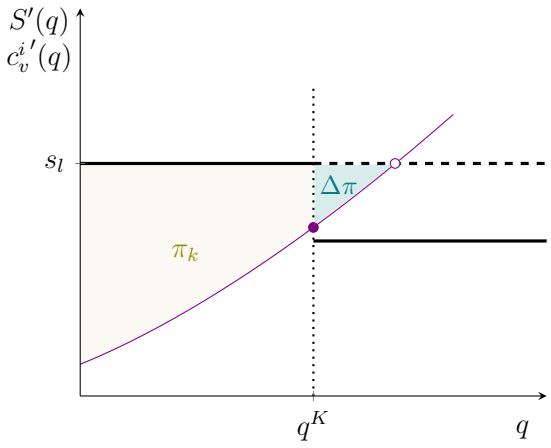


Note: The thick black line depicts the kinked marginal subsidy S'_k, and the dashed line depicts the linear marginal subsidy S'_l. The thin purple line illustrates the marginal cost curve $c_v^{i\,\prime}$ of an adopter. The capacity-choice under the kinked scheme, depicted by the full dot, is lower than the choice under the linear scheme, depicted by the empty dot. The two coloured areas depict the variable profit under the linear subsidy. The light green area π_k depicts the variable profit under the kinked subsidy, and the dark green area $\Delta\pi$ depicts the change in profit.

because these changes are endogenous to adoption behaviour. Instead, I exploit the effect of the kinked scheme on the cross-section of adopters in a given period.

Depending on their production choice under the linear subsidy, the kink affects adopters differently. There are three groups of adopters. The first group of adopters produces more than the kink point under both subsidy schemes. The thin purple line in Figure 9 illustrates the marginal cost curve of such an adopter locally around the kink point. Additionally, the figure depicts the kinked marginal subsidy as a solid black line and the linear marginal subsidy as a dashed line. The change in subsidy has two effects on the adopter. First, they face a lower marginal subsidy under the kinked scheme than under the linear scheme. Therefore, they adopt less capacity. Note that the optimal choice under each scheme is where the marginal cost curve crosses the marginal subsidy curve. The empty dot depicts the optimal choice under the counterfactual; the full dot depicts the optimal choice under the kinked scheme. The figure shows that the optimal capacity is lower under the kinked scheme than under the linear scheme. Second, the total subsidy payment under the kinked scheme is lower than under the linear scheme. Therefore, adopters earn less variable profit. Fixed costs are heterogeneous, and

Figure 10: The effect of the kinked scheme S_k on adopters just above the kink point



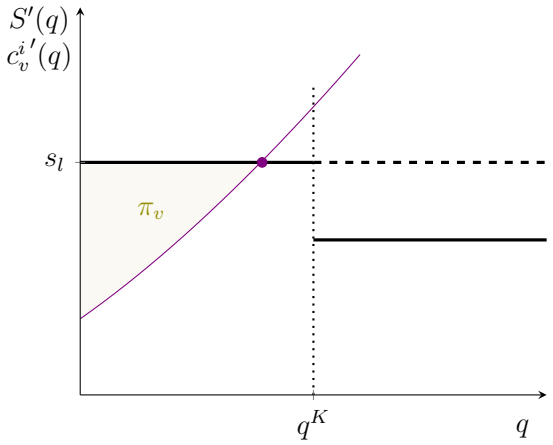
Note: The black lines depict S'_l and S'_k and purple line illustrates $c_v^{i\,\prime}$. The capacity-choice under the kinked scheme, depicted by the full dot, is exactly at the kink point. It is lower than the choice under the linear scheme, depicted by the empty dot. The dark green area $\Delta\pi$ depicts the change in profit.

therefore, some adopters stop participating. Note that the variable profit is the area between the marginal cost and the marginal subsidy curve. Figure 9 depicts the variable profit under the linear scheme as the total coloured area. The light green area π_k is the variable profit under the kinked scheme. The dark green area $\Delta\pi$ is the reduction in profit under the kinked subsidy.

The second group of adopters produces above but close to the kink point under the linear scheme. The thin purple line in Figure 10 illustrates the marginal cost curve of such an adopter locally around the kink point. Their marginal cost curves cross the kinked marginal subsidy precisely between the two marginal subsidy rates. Again, the change in subsidy has two effects on them. First, they reduce production precisely to the kink point, i.e., they bunch at the kink point. Second, they lose profit $\Delta\pi$, depicted as the dark green area in Figure 10. Again, due to heterogeneous fixed costs, some stop participating under the kinked scheme.

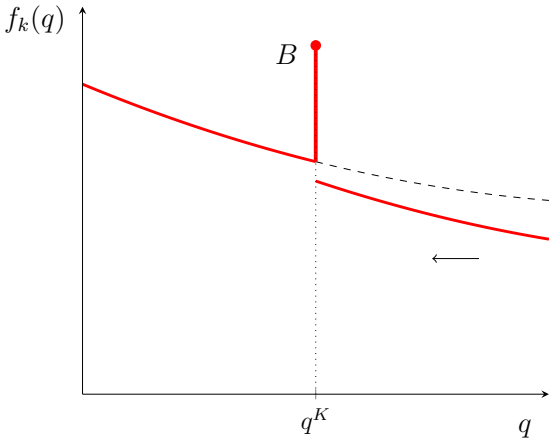
The third group of adopters produces less than the kink point under both subsidy schemes. The thin purple line in Figure 11 illustrates the marginal cost curve of such an adopter locally around the kink point. Their marginal cost curves cross both marginal subsidy schemes below the kink point. Therefore, they are not affected by a change in the scheme. They produce the same amount and earn the same profit under both schemes. Their participation does not change.

Figure 11: The effect of the kinked scheme S_k on adopters below the kink point



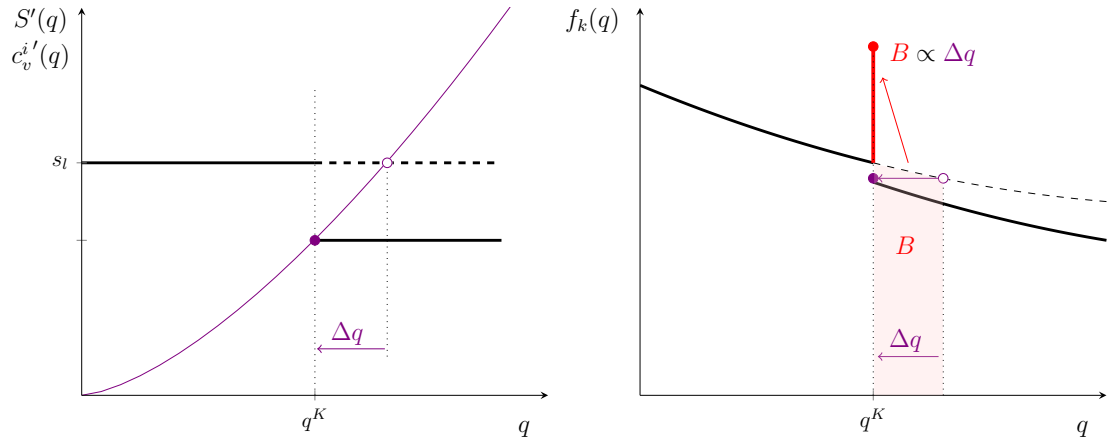
Note: The black lines depict S'_l and S'_k, and purple line illustrates $c_v^{i\,\prime}$. The full dot depicts the choice under both subsidies. The coloured area depicts the variable profit under both subsidies.

Figure 12: The observable measure f_k when there is only an intensive margin response



Note: The part of the measure above the kink point shifts to the left. It consists of adopters illustrated in Figure 9. At the kink point, there is a mass point B, i.e., the bunching mass. It consists of adopters illustrated in Figure 10. The part of the measure below the kink point is unaffected by the kink. It consists of adopters illustrated in Figure 11.

Figure 13: Identification when there is only an intensive margin response

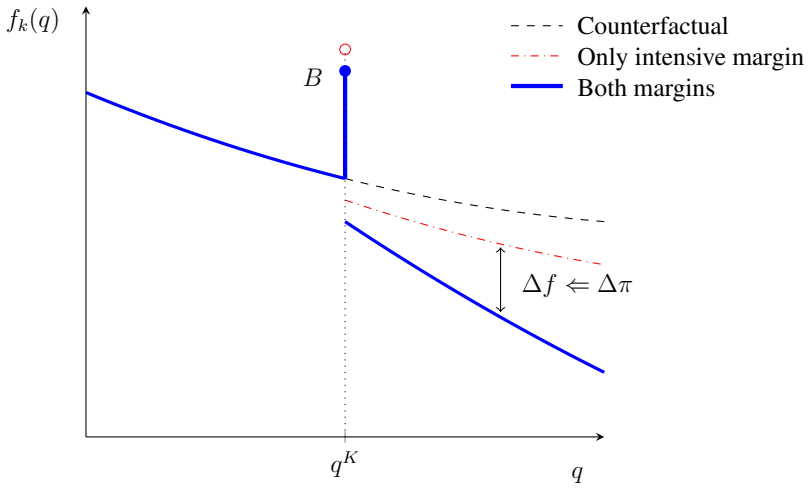


Note: The left part of the figure shows the marginal buncher. Her marginal cost curve crosses the lower marginal subsidy rate exactly at the kink point. She reduces capacity by Δq. The right part of the figure shows the marginal buncher in the measure of adoptions. Adopters to her left bunch at the kink point. The bunching mass B is approximately proportional to the response of the marginal buncher Δq.

The distinct effect of the kinked subsidy on these three groups of adopters affects the distribution of adoptions. In a first step, to better understand the effect on the distribution, consider the case where fixed costs are homogeneous and equal to zero. As a consequence, there are no participation responses. This case is considered by Saez (2010). Figure 12 depicts the observable measure of adoptions under the kinked subsidy f_k. Above the kink point, the change in schemes has two effects on the measure. First, the measure shifts to the left because adopters reduce production; second, the measure changes shape because the distribution of adopters' mass changes. Depending on the exact response, the mass in each interval increases or decreases because mass needs to be conserved. These are the standard effects of a change-in-variable on a measure.[15] At the kink point, there is a mass point B, i.e., the bunching mass. It consists of adopters from the second group. These adopters reduce production; however, they hit the kink point q^K when doing so. By reaching the kink point, they are no longer affected by the subsidy change. Therefore, they "bunch" precisely at the kink point. Below the kink point the measure under the kinked and linear scheme is the same; adopters in this range are not affected by the change in schemes.

[15]The second effect is the effect of the Jacobian.

Figure 14: The observable measure f_k when there are responses at both margins



Note: The change in profit $\Delta\pi$ causes a change in participation Δf. Above the kink point, the change in profit increases in capacity. Therefore, the change in participation increases in capacity, causing a slope change in the measure. Adopters at the kink point also react at the participation margin. Therefore, there is less bunching.

How is the measure of adoptions, as illustrated in Figure 12, useful to identify the intensive margin response? Consider the adopter depicted by the thin purple marginal cost curve in the left part of Figure 13. Her marginal cost crosses the lower marginal subsidy rate exactly at the kink point. The literature calls this adopter the "marginal buncher" (see Saez, 2010). In response to the change in marginal subsidy, the adopter reduces production by Δq. The right part of Figure 13 shows the marginal buncher in the measure of adoptions. All adopters to her left "bunch" at the kink point. Therefore, the bunching mass is approximately proportional to the reduction in the marginal buncher's production Δq. The bunching mass identifies the response Δq, which, under an additional assumption on the cost function, identifies the intensive margin elasticity.

Next, consider the case when fixed costs are present and heterogeneous. Figure 14 illustrates the consequent participation effects on the measure of adoptions. The blue line illustrates the measure when there are responses at both margins. In comparison, the red dash-dotted line illustrates the measure when there is a response only at the intensive margin; the black dashed line illustrates the counterfactual. Again, the range below the kink point is unaffected by the subsidy change. At the kink point, adopters from the second group suffer from a loss in

profit. Due to heterogeneous fixed costs, some of them stop participating. The bunching mass B decreases from the empty red dot to the full blue dot. Appendix D.4 shows that this effect biases the classic bunching estimates. Above the kink point, adopters from the first group also suffer from a profit loss. The loss $\Delta\pi$ causes a drop in the participating mass Δf. The larger is capacity q, the larger is the loss in profit $\Delta\pi$. Therefore, the larger is capacity q, the larger is the drop in participation Δf. This effect is responsible for the slope change in the measure. The theoretical prediction, illustrated in Figure 14, is perfectly consistent with the observed adoption behaviour, depicted in Figure 3 and Figure 6.

Contrary to the counterfactual, the measure under the kinked subsidy is observable. The observable bunching mass and the slope change distinctly depend on the magnitudes of both margins. Under some assumptions, it is possible to formalise the dependence of each part of the distribution on the magnitude of the two margins. Two observable features of the distribution, bunching and the slope change, are then sufficient to identify the unknown magnitudes of the two margins. The following section carries out this exercise.

3.2 The distribution of adoptions

This section derives how the observed distribution of adoptions under the kinked subsidy depends on the three unknowns: the intensive margin elasticity, the participation margin elasticity, and the counterfactual distribution. The counterfactual distribution is the capacity-distribution under a counterfactual linear subsidy without a kink. Following the non-structural econometric literature, I use the three unknowns as the primitives of my model. However, a structural formulation, explicitly specifying a cost function and a type distribution as primitives, is equivalent. It is the distinct dependence of different parts of the distribution on each of the three unknowns, which I exploit for joint identification.

Assumption 1 (Intensive margin)**.** *For small variations in the subsidy and adopters close to the kink point, the intensive margin response is iso-elastic.*[16]

[16]Formally, small variations in the subsidy means for marginal subsidy rates $s \in [\rho s_l, s_l]$ and total payment $S(q) \in [S_k(q), S_l(q)]$. Close to the kink points means for counterfactual capacity in a closed and small interval $[\underline{q}, \overline{q}]$ around the kink point.

Assumption 2 (Participation margin)**.** *For small variations in the subsidy and capacities close to the kink point, the participation margin response is iso-elastic.*

Denote the intensive margin elasticity by ϵ and the participation margin elasticity by η. Assumption 1 is standard in the bunching literature. It is a local parametric approximation to the nonparametric variable cost function. Assumption 2 is the corresponding assumption for the participation margin. It is a local parametric approximation to the nonparametric distribution of total costs. As noted by Kleven (2016), if ρ is close to 1, and q is close to the kink point, Assumption 1 is approximately true. To illustrate, suppose Assumption 1 does not hold. The closer ρ is to one, the smaller the misspecification introduced by Assumption 1. In my application, ρ is greater or equal to 0.95, hence very close to one. The same argument applies to Assumption 2.

Assumptions 1 and 2 are reduced form assumptions on endogenous objects, i.e., high-level assumptions. Appendix B.1 shows that they are equivalent to the structural Assumptions B1 and B2: the variable cost function and the distribution of total costs are locally iso-elastic. As a consequence, locally, the elasticities ϵ and η are structural parameters of the problem. Denote the capacity-choice of an adopter under the counterfactual subsidy S_l by q_l, where l stands for linear. Remember, $f_l(.)$ denotes the counterfactual distribution of adoptions, i.e., the distribution of q_l. Assumptions 1 and 2 imply that, locally, q_l describes all relevant adopter-specific heterogeneity. Therefore, q_l is equivalent to the adopters' type-parameter and $f_l(.)$ is equivalent to the type-distribution. As a consequence, locally, the nonparametric distribution $f_l(.)$ is a structural pseudo-parameter of the problem. See Appendix B.1 for a formal derivation of this statement. Using q_l as type-parameter has the advantage that the type has direct economic meaning: it is equal to the choice under the counterfactual subsidy.

Proposition 1 (The observed density)**.** *Close to the kink point, the observable measure $f_k(.)$ under the kinked subsidy $S_k(.)$ is a function of three unknowns: the intensive margin elasticity ϵ, the participation margin elasticity η, and the counterfactual measure $f_l(.)$. At the kink point q^K, there is a mass point with bunching mass B. Three parts of the observable measure $f_k(.)$*

depend distinctly on the three unknowns:

$$f_k(q) = f_l(q), \qquad \textit{for } q < q^K; \tag{5}$$

$$B = \int_{q^K}^{q^K \rho^{-\epsilon}} R(q_l, \epsilon)^\eta f_l(q_l) dq_l, \qquad \textit{for } q = q^K; \tag{6}$$

$$f_k(q) = R(q\ \rho^{-\epsilon}, \epsilon)^\eta f_l(q\ \rho^{-\epsilon})\rho^{-\epsilon}, \qquad \textit{for } q > q^K. \tag{7}$$

Note: The variable q^K denotes the kink point; ρ denotes the relative change in marginal subsidy rates, and the function $R(., \epsilon)$ is the net subsidy payment to an adopter under the kinked scheme relative to the subsidy payment under the counterfactual scheme. The exact definition and derivation of $R(., \epsilon)$ is in Lemma B5 in Appendix B.3.

The proof of Proposition 1 is in Appendix B.3. Below the kink point, the observable measure $f_k(.)$ depends only on the counterfactual measure $f_l(.)$. At the kink point is an observable mass point with mass B. The mass depends on all three unknowns. The measure above the kink point depends on all three unknowns as well. However, all observables depend on the three unknowns distinctly, a property crucial for identification.

3.3 Identification

This section shows under which conditions the observed measure $f_k(.)$ identifies the three unknowns. The pseudo-parameter $f_l(.)$ in Proposition 1 is infinite-dimensional. Equation (5) shows that below the kink point, the observable measure $f_k(.)$ is equal to $f_l(.)$. Therefore, $f_l(.)$ is identified for values smaller than q^K. However, $f_l(.)$ is part of Equation (6) and (7) evaluated at values larger than q^K. The function is unobservable at these points. Identification is only possible if $f_l(.)$ is smooth enough such that the observations below the kink point identify the function also for values larger than the kink point. Therefore, an additional assumption is necessary.

Assumption 3 (Smoothness)**.** *For all quantities q_l in a large enough interval $[\underline{q}, \overline{q}]$ around the kink point, there exists a transformation of the counterfactual measure $f_l(.)$ that has a convergent power series representation (i.e., is analytic).*

For example,

$$\ln f_l(q_l) = \sum_{p=0}^{\infty} \gamma_p \frac{1}{p!} \left(\ln \frac{q_l}{q^K} \right)^p \text{ for all } q_l \in [\underline{q}, \overline{q}]. \tag{8}$$

Assumption 3 is a smoothness assumption. See Gautier and Gaillac (2021) and Iaria and Wang (2022) for a use of this assumption in a different context. A set of functions satisfying Assumption 3 are functions which are complex differentiable on $[\underline{q}, \overline{q}]$.

Condition 1 (Rank condition)**.** *Equations (6) and (7) are not colinear.*

Equations (5) to (7) form a simultaneous, nonlinear system of equations. The system must not be colinear to identify the parameters of interest. Condition 1 is a rank condition that holds generically. Moreover, it can be verified ex-post estimation (see Appendix D.5 for the estimates and a detailed discussion).

Proposition 2 (Identification)**.** *Under Assumption 3 and Condition 1, the observable measure $f_k(.)$ identifies the counterfactual measure $f_l(.)$, the intensive margin elasticity ϵ, and the participation margin elasticity η.*

The proof is in Appendix B.4. Observations below the kink point identify the nonparametric counterfactual distribution. Observations at the kink point and observations above the kink point jointly identify the two response margins. Note that Proposition 2 uses no assumptions on q^K, s_l, and ρ, except for their observability. They can be random or endogenous. Without loss of generality, normalise the rate s_l to one; it corresponds to choosing a monetary unit. Consequently, the monetary unit of all monetary variables is the present discounted value of payments to one kWp of capacity.

Blomquist et al. (2021) criticise the fact that the bunching literature uses parametric functional form assumptions on f_l to identify the intensive margin elasticity. Proposition 2 proofs identification without assuming a parametric form of f_l.

Appendix B.5 generalizes the result in Proposition 2 for the case when there is a discontinuity (i.e., a notch) in the incentive scheme.

4 Estimation

Intuitively, the estimation minimises the distance between the observed log-histogram in Figure 2 and the model in Proposition 1. Formally, the empirical model is

$$\widehat{\ln f(q_j)} = \ln f_k(q_j \mid \epsilon, \eta, f_l) + u_j, \tag{9}$$

where the variable u_j denotes the noise term. The variable $\widehat{\ln f(q_j)}$ is a bias-corrected estimate of the log-histogram at capacity q_j. Appendix C.1 discusses the choice of the logarithmic model and the estimation of $\widehat{\ln f(q_j)}$.

The pseudo-parameter $f_l(.)$ in Proposition 1 is infinitely dimensional. Therefore, I use a semi-nonparametric sieve estimator (Chen, 2007):

$$\ln f_l(q) = \sum_{p=0}^{P(n)} \gamma_p \frac{1}{p!} \left(\ln \frac{q}{q^K} \right)^p. \tag{10}$$

The vector $(\gamma_0, ...\gamma_P)$ denotes the parameters of the series, and $P(n)$ denotes its order. The order converges slowly to infinity as the sample size n converges to infinity. Using this logarithmic series expansion is natural because it contains the uniform distribution, the Pareto distribution, and the log-normal distribution as special cases. These are common distributions for variables on a positive domain. Moreover, Figure 4 shows that the observed distributions are very close to linear on a logarithmic scale. This shape is evidence for a counterfactual distribution close to a Pareto distribution. Therefore, one can expect that a series expansion containing the Pareto distribution as a special case converges quickly. Appendix D.8 discusses the choice of the logarithmic expansions in further detail and shows robustness checks to this choice.

As an additional contribution to the bunching literature, I propose a data-driven method to select the nonparametric specification. As standard in nonparametric estimations, the estimates' bias and variance depend on the specification. There are two main specification-parameters. The first parameter is the order of the polynomial P. The higher the order P, the lower the bias and the larger the variance. The second parameter is the bandwidth $b = [\underline{q}, \overline{q}]$. It is the

interval of values around the kink point used for estimation, i.e., the range of q_j. The smaller b, the lower the bias and the larger the variance. Note that for the estimates to be consistent, it suffices that P goes to infinity as the sample size goes to infinity. A smaller b reduces the bias for any given P; however, b does not have to go to zero for the estimates to be consistent. Note that the bandwidth b is not to be confused with the bin size of the log-histogram. As is standard in nonparametric estimations, I choose the specification that minimises an estimate of the mean squared error. It is the sum of the variance and the square bias of an estimate. The estimate of the participation margin $\hat{\eta}$ is more sensitive to the specification than the estimate of the intensive margin $\hat{\epsilon}$. Therefore, I use an estimate of the mean squared error of $\hat{\eta}$ as a selection criterion for the specification. To estimate the variance, I use the nonparametric bootstrap. Appendix C.3.1 derives an analytical expression of the bias. It depends on the intensive margin response and the unestimated rest of the series expansion (i.e., $\gamma_{P+1}, \gamma_{P+2}, ...$). I propose an out-of-sample estimator of the bias that converges at a parametric rate. Appendix C.3 discusses the estimator of the mean squared error in detail. Appendix C.4.3 shows the estimates.

Theory predicts that the bunching mass is located precisely at the kink point. However, in practice, the excess mass scatters around the kink point due to optimisation errors. The literature calls this phenomenon non-sharp bunching. To account for non-sharp bunching, it is standard in the bunching literature to choose a bunching interval $[q_L, q_H]$ around the kink point after visual inspection of the histogram (see Kleven, 2016). Note that the bunching interval lies within the bandwidth: $[q_L, q_H] \in b$. The observed logarithmic bunching mass $\widehat{\ln B}$ denotes the logarithm of the normalised number of adopters in the bunching interval. Appendix D.6 discusses the choice of the bunching interval and reports robustness checks.[17]

The estimation follows the two-step least square procedure in Chetty et al. (2011). I iteratively minimise the square distance between the model and the log-histogram outside and

[17] Anagol et al. (2022) and McCallum and Navarrete (2022) explicitly model the optimization frictions leading to non-sharp bunching. Since bunching is relatively sharp in my application, I follow the standard approach.

inside the bunching interval until the estimates converge:

$$(\hat{\eta}, \hat{\gamma}_P) = \arg\min_{\eta,\gamma_P} \sum_{q_j \notin [q_L, q_H]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \hat{\epsilon}, \eta, \gamma_P) \right)^2, \tag{11}$$

and

$$\hat{\epsilon} = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln \int_{q_L}^{q_H} f_k(q \mid \epsilon, \hat{\eta}, \hat{\gamma_P}) dq \right)^2. \tag{12}$$

An exponential transformation constrains $\hat{\epsilon}$ and $\hat{\eta}$ to positive values. Appendix C.2 shows that the estimates are consistent and asymptotically normal. Following Chetty et al. (2011), I estimate the standard errors using nonparametric bootstrap. Appendix C.4 discusses the choice of the optimal specification.

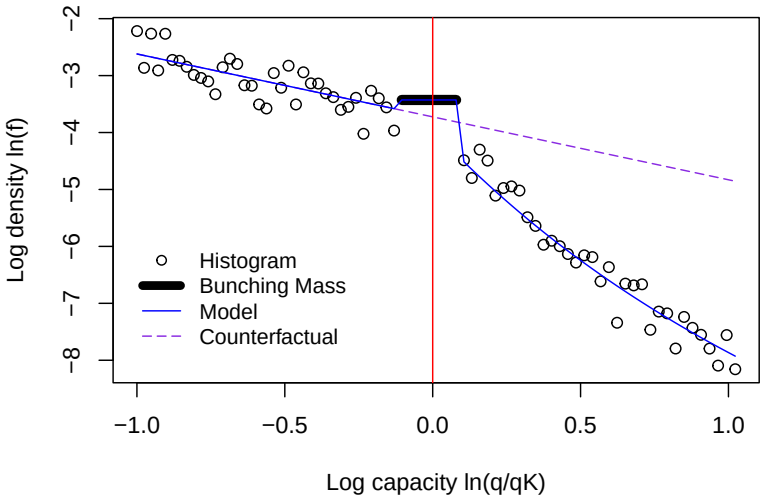
5 Empirical results

Figure 15 shows the normalised histogram at 30 kWp for 2004. The blue line depicts the estimated model. The purple dashed line depicts the counterfactual distribution. The red line marks the kink point normalised to zero; scales are logarithmic. The black bar depicts the bunching mass in the bunching interval.

Since the bunching literature estimates behavioural responses as elasticities, the empirical model estimates the participation margin response as the elasticity η. However, in the theoretical literature on nonlinear pricing, it is more common to use participation semi-elasticities (see Rochet and Chone, 1998 and Jacquet et al., 2013). Therefore, this section reports the results at the participation margin as semi-elasticities $\hat{\kappa}$. The interpretation is as follows: a lump sum payment equivalent to the present discounted value of one kWp increases participation by a factor κ. The relation between the two variables is $\eta = \kappa \times S$, where S is the subsidy payment. Appendix D.1 reports the estimates of the participation elasticity $\hat{\eta}$.

Table 1 shows the yearly estimates at 30 kWp for 2004 to 2008. Since they overlap with other policy changes, I cannot exploit kinks after 2008. The intensive margin response increases over time, while the participation margin response decreases over time. Overall, the participation margin response is substantial. To increase the sample size, I estimate the re-

Figure 15: Histogram in 2004 at 30 kWp with estimated model and counterfactual



Note: The x-axis shows the normalised logarithm of capacity. The y-axis shows the logarithm of the density. The red line marks the kink point. The estimation minimises the distance between the data in black and the model in blue.

Table 1: Yearly estimates at the kink point at 30 kWp

Year	$\hat{\epsilon}\ (SD)$	$\hat{\kappa}\ (SD)$
2004	1.92 (0.23)	3.21 (0.14)
2005	3.43 (0.26)	2.49 (0.13)
2006	3.92 (0.31)	1.88 (0.14)
2007	5.08 (0.29)	2.00 (0.12)
2008	5.33 (0.24)	1.76 (0.09)

Note: The table reports the estimated elasticity and semi-elasticity with standard errors in brackets.

Table 2: Estimates at 30 and 100 kWp pooling observations from 2004 to 2008

Capacity	$\hat{\epsilon}\ (SD)$	$\hat{\kappa}\ (SD)$
30 kWp	4.37 (0.13)	2.31 (0.06)
100 kWp	4.63 (0.84)	0.00 (0.02)

Note: The table reports the estimated elasticities with standard errors in brackets. The estimation pools observations from 2004 to 2008.

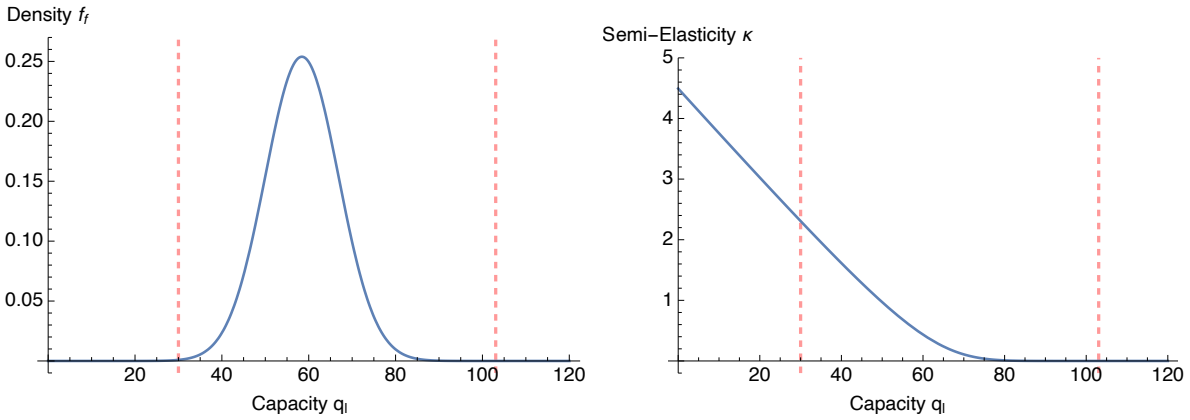
sponse at 100 kWp pooling observations from 2004 to 2008. Table 2 shows the main empirical results: the estimates at 30 and 100 kWp in the pooled data. The results suggest that the intensive margin elasticity is the same for adopters of different capacities, while the participation margin semi-elasticity decreases with capacity.

The evidence in Table 2 is not surprising. An important intensive adjustment margin is the quality of the solar panel. Low-capacity adopters have access to the same quality choices as high-capacity adopters. Therefore, their responses have the same elasticity.

It is useful to consider the underlying distribution of fixed costs to interpret the participation responses. A participation semi-elasticity that decreases in capacity is consistent with a simple normal distribution of fixed costs. The two semi-elasticities observed at the two capacity levels in Table 2 are sufficient to calibrate the distribution. Figure 16 shows the calibrated distribution and the implied semi-elasticities. The figure shows the calibrated density of fixed costs $f_f(\pi_v(q_l))$ and the semi-elasticity of participation $\kappa(\pi_v(q_l)) = \frac{f_f}{F_f}(\pi_v(q_l))$ as a function of the counterfactual variable profit $\pi_v(q_l)$ at capacity q_l. Therefore, the function shows the mass of agents indifferent to participation and the semi-elasticity of participation at a certain capacity level under the counterfactual subsidy. The two red lines show the counterfactual capacity corresponding to the variable profit at the two kink points under the kinked subsidy. Appendix E.2 discusses the calibration in detail.

Figure 16 illustrates why the participation semi-elasticity in Table 2 decreases with capacity. The semi-elasticity depends on the variable profit π_v. The higher the profit, the lower the semi-elasticity. Large capacity systems are very profitable. Therefore, the fixed cost plays only a small role in the adoption decision, i.e., only a few adopters have such high fixed costs to make adopting a large capacity unprofitable. In contrast, adopting low-capacity systems de-

Figure 16: The calibrated density of fixed costs and the implied semi-elasticity



Note: The figure on the left shows the density of fixed costs $f_f(\pi_v(q_l))$ as a function of the counterfactual variable profit π_v at capacity q_l.[18] *Therefore, the function shows the mass of agents indifferent to participation at a certain capacity level under the counterfactual subsidy. The figure on the right shows the implied participation semi-elasticity at capacity q_l under the counterfactual subsidy. The red lines illustrate the two kink points.*

pends crucially on fixed costs. In relative terms, many adopters have a fixed cost equal to the variable profit. It follows that relatively many adopters are close to indifferent to participating. Consequently, a small increase in the subsidy payment causes a large relative response. While the magnitude of the participation margin response at low capacities is surprising, there is an intuitive explanation. During the observation period, solar panels were a nascent product. They had high market potential compared to market coverage, reflected by the fact that 30 kWp is to the left of the bell curve in Figure 16. This feature explains why, compared to mature products, the participation semi-elasticities are high. Indeed, estimated participation semi-elasticities decrease over time as the market saturates. This explanation suggests that one can expect similar response patterns to other deployment subsidies for early-stage technologies.

The classic bunching estimator does not consider participation margin responses. Appendix D.4 shows that ignoring participation introduces a downward bias of 20% in the estimate of the intensive margin elasticity. The bias comes from the fact that, instead of bunching, some agents stop participating in the programme. It decreases the bunching mass. Wrongly ignoring this effect attributes the smaller bunching mass to a smaller intensive margin response. Similarly, the appendix shows that ignoring the intensive margin introduces an upward bias of 5% in the

estimate of the participation margin. These results show that it is essential to estimate both margins simultaneously.

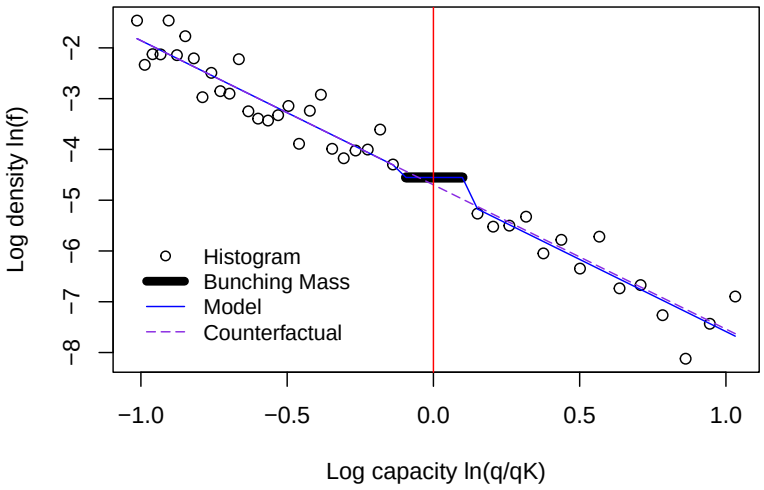
5.1 Robustness

One concern for the estimation is the violation of Assumption 2 because of irregularities in the counterfactual. On the one hand, there might be an excessive mass of adopters at the kink point for reasons other than the subsidy's kink. The irregularity may be a mass point or a continuous bump, which is not or only hardly predictable from observations to the left of the kink point. This is a general concern in the bunching literature and was raised by Blomquist et al. (2021). It leads to an upward bias in the intensive margin estimate. Additionally, even if Assumption 2 holds, choosing an inefficient series approximation may considerably bias the results in small samples.

On the other hand, there might be a slope change at the kink point for reasons other than the subsidy's kink, introducing a bias in the participation margin estimate. This effect is a general concern in the regression kink design literature. Additionally, it is easy to mistakenly estimate a concavity in the distribution as a change in slope, which would again bias the results. It is important to evaluate if the chosen specification is picking up one of these biasing effects. To address these issues, I estimate the model pooling observations from 2000 to 2003. From 2000 until 2003, there was no kink in the subsidy. If the concerns raised are an issue, the estimates in 2000-2003 should be significant. Figure 17 shows the histogram in 2000-2003, with the estimated model and the linear counterfactual. I use the same bunching interval as in 2004-2008. The selected specification does not estimate significant parameters in the untreated data. Appendix D.3 reports the same robustness check for the estimates at 100 kWp. Again, the selected specification does not estimate significant effects.

Germeshausen (2018) uses a difference-in-difference approach to estimate the treatment effect of introducing a new kink in Germany in 2012. For a comparison between his results and the results of this paper, Appendix E.4.4 uses the estimates to retrieve the implied treatment effect. The two effects are very similar: introducing the new kink reduces capacity in the 10-20 kWp range by 43% and 40%, respectively. The finding provides evidence for the validity of the

Figure 17: Histogram in 2000-2003 with estimated model



Note: The figure shows the robustness check. The x-axis shows the normalised logarithm of capacity. The y-axis shows the logarithm of the density. The red line marks the kink point. The estimated model in blue is equal to the counterfactual in purple. The estimates are not significant.

Table 3: Results for the untreated data in 2000-2003 (placebo)

Capacity	$\hat{\epsilon}\ (SD)$	$\hat{\kappa}\ (SD)$
30 kWp	0.51 (0.36)	0.00 (0.15)

Note: The table shows the results of the robustness check. The standard errors are in brackets. The estimates are not significant.

respective identifying assumption in both studies.

6 Policy evaluation

This section uses the estimates to evaluate and optimise the subsidy. It builds heavily on the results in Rochet and Stole (2002), who provide a theoretical solution to the nonlinear pricing problem of a monopoly when there are intensive and participation margin responses.

Like other kink and discontinuity estimators in the literature, Section 5 estimates the responses locally at each kink point. It is necessary to make global assumptions to use them for policy evaluation. In line with the empirical evidence in Table 2, assume an isoelastic cost function and a normally distributed fixed cost.

Assumption 4 (Isoelastic cost function)**.** *The cost function is isoelastic with intensive margin elasticity ϵ:*

$$c(q, q_l, c_f) = \frac{s_l}{q_l^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}} + c_f, \tag{13}$$

where (q_l, c_f) is the two-dimensional type-parameter.

Assumption 5 (Normally distributed fixed costs)**.** *The distribution of the fixed costs c_f is normal: $c_f \sim N(\mu_f, \sigma_f)$, with CDF $F_f(c_f)$ and density $f_f(c_f)$.*[19]

Set ϵ equal to the estimated intensive margin elasticity in Table 2 and calibrate (μ_f, σ_f) using the estimated participation semi-elasticities (see Figure 16 and Appendix E.2 for details). Appendix E.2 describes the estimation of the type-distribution $f_l(.)$. It is a log-normal distribution for low capacities and a Pareto distribution for large capacities. The model can be solved for any counterfactual subsidy scheme using these assumptions. Note that, in line with the bunching literature, so far, the term counterfactual referred to the linear subsidy without a kink. This section calls any subsidy different from the observed kinked subsidy a counterfactual subsidy.

[19]The assumption implies c_f and q_l are independent, as in Rochet and Stole (2002). Appendix E.1 generalizes the assumption to allow for correlation between c_f and q_l.

6.1 The government's objective

Assume the government's objective is to incentivise the adoption of the observed aggregate capacity Q^T at a minimal public cost:

$$\min_{S(.)} \int S(q) dF_S(q), \tag{14}$$

such that

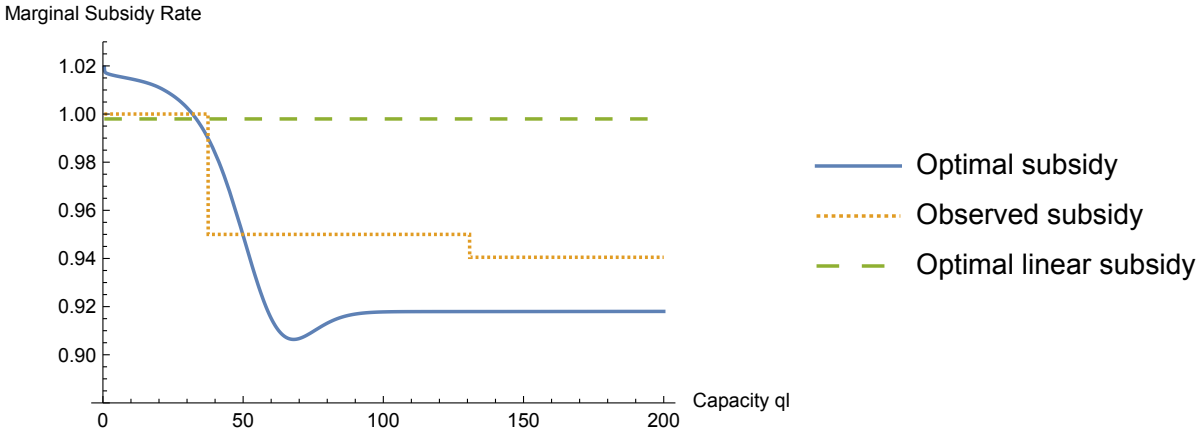
$$\int q \, dF_S(q) \geq Q^T, \tag{15}$$

where $F_S(q)$ is the distribution of capacity under subsidy $S(.)$.

While Objective (14) is appealing for its simplicity, it is a simplification of actual governmental preferences. Appendix E.3 discusses a general objective with redistributive preferences, a preference for aggregate solar capacity, and a possibly optimal taxes on other sources of income. The appendix derives the optimal nonlinear subsidy as a function of agents' responses at the two margins, the government's preferences for solar capacity, and the government's preferences for redistribution. A government can use its actual preferences together with the estimates in Section 5 and the general solution in Appendix E.3.2 to implement the optimal subsidy.

This section uses the simple objective (14), which avoids taking a detailed stand on governmental preferences. It implicitly assumes that the government does not want to distribute rents to adopters of solar panels, which is justifiable by the fact that typically only households in the upper part of the income distribution own roofs and can adopt solar panels. Moreover, the German government already uses a nonlinear subsidy scheme, showing that it is trying to curtail rents. Appendix E.3.4 shows under which conditions Objective (14) follows from the general objective. For example, it is the case if the redistributive preferences are Rawlsian, the lowest income households cannot adopt solar panels because they do not own roofs, and the government only values capacity up to an aggregate capacity goal Q^T. Note that a Pigouvian subsidy, where the optimal marginal subsidy is equal to the marginal environmental benefit, is not the optimal solution to Objective (14). Appendix E.3.3 discusses when a Pigouvian subsidy is optimal in the general objective of Appendix E.3, which, generally, is not the case (see Kaplow, 1996, Cremer, Gahvari, and Ladoux, 1998, Kaplow, 2008, and Kaplow, 2012).

Figure 18: The observed and the optimal marginal subsidies $S'(q(q_l))$



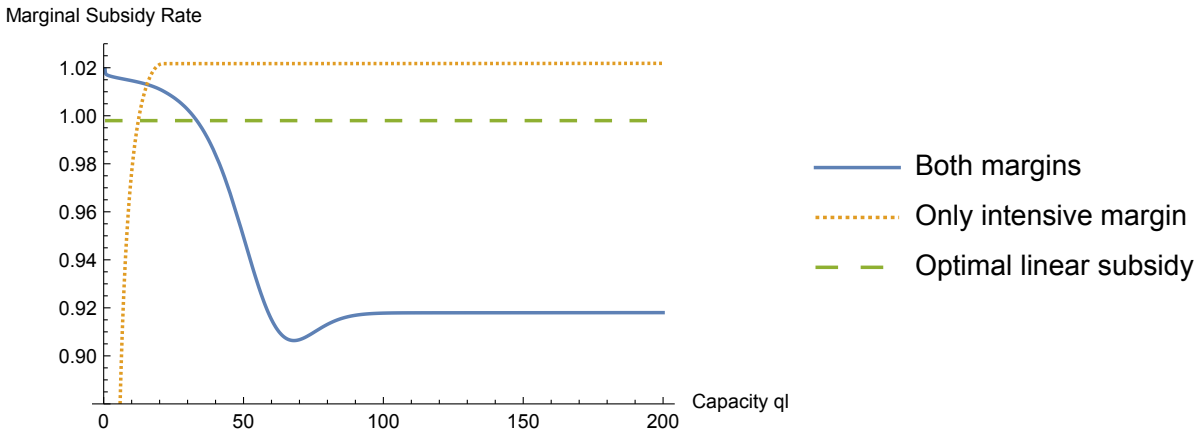
Note: The figure compares the observed marginal subsidy with the optimal linear and the optimal nonlinear marginal subsidies. Note that the first marginal subsidy rate of the observed subsidy is normalised to one.

6.2 Results counterfactual exercises

This section discusses four counterfactual exercises. The first exercise solves for the optimal linear subsidy, which serves as a benchmark. Appendix E.4.1 describes the procedure. Compared to this benchmark, the government's subsidy achieves a modest cost reduction of 0.14 per cent. The second exercise solves for the optimal nonlinear subsidy using mechanism design. The analysis follows the screening problem in Rochet and Stole (2002) and is outlined in Appendix E.4.2. The optimal nonlinear subsidy is 0.45 per cent less costly than the linear benchmark. Cost savings are by a factor 3.14 larger than the cost savings of the government's subsidy. Figure 18 compares the optimal marginal subsidy schedule $S'(q(q_l))$ to the linear benchmark and the observed marginal subsidy. Qualitatively, the optimal marginal subsidy is mostly downward sloping; hence it is similar to the subsidy used by the government. However, quantitatively, the range of optimal marginal subsidy rates is larger than the range used by the government. The optimal scheme reduces rents more than the observed scheme by paying higher rates to small adopters and lower rates to large adopters. While the result shows that there is room for improving the government's scheme, the overall benefits of using a nonlinear subsidy scheme remain modest.

To better understand what limits the scope for curtailing costs, a third thought experiment

Figure 19: The optimal marginal subsidy $S'(q(q_l))$ with and without participation



Note: The figure compares the optimal marginal subsidy without participation with the optimal marginal subsidy with participation.

assumes adopters can only respond at the intensive margin. It follows the methodology in Appendix E.4.2, assuming the participation response equals zero. Without participation margin, the optimal nonlinear subsidy would be 8.6 per cent less costly than the linear benchmark. This result shows that the participation margin limits the scope for cost reduction via a nonlinear subsidy. Figure 19 compares the optimal marginal subsidy without participation to the optimal marginal subsidy with participation. The marginal subsidy schedule is increasing, and the distance to the linear benchmark is much more significant than when the participation margin is present. The shape of the scheme follows from simple intuition. On the one hand, it is optimal to pay a high marginal rate to large adopters. Reacting only to marginal rates, large adopters install larger capacities than under the linear benchmark. On the other hand, the marginal rate paid to small adopters is low. They adopt lower quantities than under the linear benchmark. By definition, the net capacity effect is zero, and the government achieves the fixed capacity goal. However, the net cost effect is not zero. The total payment to large adopters is the integral under the orange/dotted curve. The low marginal rates for small adopters extract profit from large adopters because they receive a lower payment for infra-marginal units. Therefore, the scheme is less costly than the linear scheme. In contrast, when, as in reality, there are participation margin responses, this strategy to extract profits is ineffective. Low marginal subsidy rates for small capacities affect larger adopters by reducing their profit margins. It triggers responses

at the participation margin and, therefore, a loss in capacity. This effect limits the room for rent extraction through nonlinear pricing when the participation margin is active. Therefore, cost-savings when both margins are active remain moderate.

The results in Figure 19 show that the participation margin changes the shape of the optimal marginal subsidy. However, is it counterproductive to ignore participation when, in reality, such a margin is present? Rochet and Stole (2002) show that when the type distribution $f_l(.)$ is uniform, the undistorted and the optimal intensive allocation bound the optimal allocation with both margins. Their result suggests that implementing the optimal intensive schedule in Figure 19 might reasonably approximate the optimum. Appendix E.4.3 shows that their result is not robust to more general forms of the type distribution. E.g. for the type-distribution in this application, the optimal allocation is close to but outside the bounds derived by Rochet and Stole (2002). To gather additional evidence, a fourth counterfactual exercise assumes the policymaker implements the optimal intensive schedule, but adopters react at both margins. I keep aggregate capacity constant again to allow for a meaningful comparison to the other counterfactual exercises. The exercise finds that the optimal intensive schedule increases costs by 3.1 per cent instead of decreasing them; i.e., ignoring participation has a sizable adverse effect on costs. The result shows that the qualitative insight in Rochet and Stole (2002) provides limited guidance to the policymaker. Combining their theoretical results with estimates of the intensive *and* participation margin responses is crucial for implementing the optimal schedule.

Table E2 and E3 in Appendix E.5 summarise the results of the four counterfactual exercises.

7 Conclusion

This paper leverages a kinked incentive scheme to estimate agents' intensive and participation margin responses simultaneously. On the example of the German subsidy for rooftop solar panels, it demonstrates how to use these estimates to evaluate and optimise a nonlinear incentive scheme. The paper reveals that adopters of solar panels react at both margins. The nonlinearities in the current schedule are only modestly effective at reducing the public costs of the programme. While an optimal subsidy schedule increases cost savings by a factor of three,

overall, they remain modest. It is the participation margin responses that limit the scope for cost reductions via second-degree price discrimination. Neglecting the participation margin when optimising the policy has a sizable adverse effect on costs. The result highlights the importance of considering both margins when evaluating and optimising nonlinear incentive schemes.

Deployment subsidies for nascent green technologies have become popular climate policies. The results of this paper indicate that nonlinearities can reduce the public costs of these programmes; however, cost savings are likely moderate. Hence, nonlinear subsidies are not a silver bullet and can even trigger detrimental effects. Estimating both response margins is essential for their optimal design.

Adapting the estimator to exploit notches, i.e., discontinuities in incentive schemes, is straightforward. Kinks and notches are common features of economic incentive schemes, as are responses at the intensive and participation margin. Possible further applications of the methodology in this paper are the evaluations of taxes, subsidies, transfers, and product prices.

References

ACEMOGLU, D., P. AGHION, L. BURSZTYN, AND D. HEMOUS (2012): "The environment and directed technical change," *American Economic Review*, 102, 131–166.

AGHION, P., A. BERGEAUD, AND J. VAN REENEN (2021): "The impact of regulation on innovation," Tech. rep., National Bureau of Economic Research.

ANAGOL, S., A. DAVIDS, AND B. LOCKWOOD (2022): "Diffuse Bunching with Frictions: Theory and Estimation," Tech. rep., CEPR Discussion Papers.

ANDO, M. (2017): "How much should we trust regression-kink-design estimates?" *Empirical Economics*, 53, 1287–1322.

BACHAS, P. AND M. SOTO (2021): "Corporate taxation under weak enforcement," *American Economic Journal: Economic Policy*, 13, 36–71.

BEFFY, M., R. BLUNDELL, A. BOZIO, G. LAROQUE, AND M. TO (2019): "Labour supply and taxation with restricted choices," *Journal of Econometrics*, 211, 16–46.

BERTANHA, M., C. CAETANO, H. JALES, AND N. SEEGERT (2023): "Bunching Estimation Methods," Tech. rep., Mimeo.

BERTANHA, M., A. H. MCCALLUM, AND N. SEEGERT (2021): "Better bunching, nicer notching," Tech. rep.

BESLEY, T., N. MEADS, AND P. SURICO (2014): "The incidence of transaction taxes: Evidence from a stamp duty holiday," *Journal of Public Economics*, 119, 61–70.

BEST, M. C. AND H. J. KLEVEN (2017): "Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK," *The Review of Economic Studies*, 85, 157–193.

BLOMQUIST, S., A. KUMAR, C.-Y. LIANG, AND W. K. NEWEY (2015): "Individual heterogeneity, nonlinear budget sets, and taxable income," .

BLOMQUIST, S., W. K. NEWEY, A. KUMAR, AND C.-Y. LIANG (2021): "On bunching and identification of the taxable income elasticity," *Journal of Political Economy*, 129, 2320–2343.

BURR, C. (2016): "Subsidies and Investments in the Solar Power Market," Tech. rep., University of Colorado Boulder Working Paper.

CAETANO, C., G. CAETANO, AND E. NIELSEN (2020): "Correcting for Endogeneity in Models with Bunching," Tech. rep., Finance and Economics Discussion Series, Federal Reserve.

CARD, D., D. S. LEE, Z. PEI, AND A. WEBER (2015): "Inference on causal effects in a generalized regression kink design," *Econometrica*, 83, 2453–2483.

CHEN, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6, 5549–5632.

CHETTY, R., J. N. FRIEDMAN, T. OLSEN, AND L. PISTAFERRI (2011): "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records," *The Quarterly Journal of Economics*, 126, 749–804.

COX, N., E. LIU, AND D. MORRISON (2020): "Market Power in Small Business Lending: A Two Dimensional Bunching Approach," .

CREMER, H., F. GAHVARI, AND N. LADOUX (1998): "Externalities and optimal taxation," *Journal of Public Economics*, 70, 343–364.

DE GROOTE, O. AND F. VERBOVEN (2019): "Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems," *American Economic Review*, 109, 2137–72.

DESTATIS (2023): "Statistisches Bundesamt (DESTATIS): VGR des Bundes - Einnahmen und Ausgaben sowie Finanzierungssaldo des Staates: Deutschland, Jahre," *https://www-genesis.destatis.de/genesis/online; (Last accessed: July 25, 2023)*.

FEGER, F., N. PAVANINI, AND D. RADULESCU (2022): "Welfare and redistribution in residential electricity markets with solar power," *The Review of Economic Studies*, 89, 3267–3302.

GANONG, P. AND S. JÄGER (2018): "A permutation test for the regression kink design," *Journal of the American Statistical Association*, 113, 494–504.

GARBINTI, B., J. GOUPILLE-LEBRET, M. MUÑOZ, S. STANTCHEVA, AND G. ZUCMAN (2023): "Tax Design, Information, and Elasticities: Evidence From the French Wealth Tax," Tech. rep., National Bureau of Economic Research.

GAUTIER, E. AND C. GAILLAC (2021): "Estimates for the SVD of the Truncated Fourier Transform on L2 (cosh (b.)) and Stable Analytic Continuation," *Journal of Fourier Analysis and Applications*.

GELBER, A. M., D. JONES, D. W. SACKS, AND J. SONG (2020): "Using Kinked Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Social Security Earnings Test," *Forthcoming: American Economic Journal: Applied Economics*.

GERARD, F., M. ROKKANEN, AND C. ROTHE (2020): "Bounds on treatment effects in regression discontinuity designs with a manipulated running variable," *Quantitative Economics*, 11, 839–870.

GERARDEN, T. (2022): "Demanding innovation: The impact of consumer subsidies on solar panel production costs," *Managment Science*.

GERMESHAUSEN, R. (2018): "Effects of Attribute-Based Regulation on Technology Adoption–The Case of Feed-In Tariffs for Solar Photovoltaic," *ZEW-Centre for European Economic Research Discussion Paper*.

GOFF, L. (2022): "Treatment Effects in Bunching Designs: The Impact of the Federal Overtime Rule on Hours," Tech. rep.

HUGHES, J. E. AND M. PODOLEFSKY (2015): "Getting green with solar subsidies: evidence from the California solar initiative," *Journal of the Association of Environmental and Resource Economists*, 2, 235–275.

IARIA, A. AND A. WANG (2022): "Real Analytic Discrete Choice Models of Demand: Theory and Implications," Tech. rep., https://ssrn.com/abstract=4094866.

IRENA (2022): "International Renewable Energy Agency (IRENA): Renewable Power Generation Costs in 2021," *https://www.irena.org/publications/2022/Jul/Renewable-Power-Generation-Costs-in-2021; (Last accessed: June 21, 2023)*.

JACQUET, L., E. LEHMANN, AND B. VAN DER LINDEN (2013): "Optimal redistributive taxation with both extensive and intensive responses," *Journal of Economic Theory*, 148, 1770–1805.

KAPLOW, B. L. (2012): "Optimal control of externalities in the presence of income taxation," *International Economic Review*, 53, 487–509.

KAPLOW, L. (1996): "The optimal supply of public goods and the distortionary cost of taxation," *National Tax Journal*, 513–533.

——— (2008): "Optimal policy with heterogeneous preferences," *The BE Journal of Economic Analysis & Policy*, 8.

KLEVEN, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464.

KLEVEN, H. J., C. LANDAIS, E. SAEZ, AND E. SCHULTZ (2013): "Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in Denmark," *The Quarterly Journal of Economics*, 129, 333–378.

KLEVEN, H. J. AND M. WASEEM (2013): "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan," *The Quarterly Journal of Economics*, 128, 669–723.

KRAFT-TODD, G. T., B. BOLLINGER, K. GILLINGHAM, S. LAMP, AND D. G. RAND (2018): "Credibility-enhancing displays promote the provision of non-normative public goods," *Nature*, 563, 245.

MARX, B. M. (2018): "Dynamic Bunching Estimation with Panel Data," Tech. rep., University Library of Munich, Germany.

MCCALLUM, A. H. AND M. NAVARRETE (2022): "Why Don't Taxpayers Bunch at Kink Points?" *Available at SSRN 4219884*.

MOORE, D. T. (2021): "Evaluating Tax Reforms without Elasticities: What Bunching Can Identify," Tech. rep., Mimeo.

MYHRE, A. (2021): "Intensive and Extensive Margin Labor Supply Responses to Kinks in Disability Insurance Programs," *Available at SSRN 3914055*.

NEMET, G. F. (2019): *How solar energy became cheap: A model for low-carbon innovation*, Routledge.

NIELSEN, H. S., T. SØRENSEN, AND C. TABER (2010): "Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform," *American Economic Journal: Economic Policy*, 2, 185–215.

PODESTA, J. D. (2023): "Building a Clean Energy Economy: A Guidebook to the Inflation Reduction Act's Investment in Clean Energy and Climate Action," *https://www.whitehouse.gov/wp-content/uploads/2022/12/Inflation-Reduction-Act-Guidebook.pdf; (Last accessed: June 21, 2023)*.

REDDIX, K. (2015): "Powering Demand: Solar Photovoltaic Subsidies in California," Tech. rep., Citeseer.

ROCHET, J. C. AND P. CHONE (1998): "Ironing, sweeping, and multidimensional screening," *Econometrica*, 783–826.

ROCHET, J.-C. AND L. A. STOLE (2002): "Nonlinear pricing with random participation," *The Review of Economic Studies*, 69, 277–311.

RUH, P. AND S. STAUBLI (2019): "Financial incentives and earnings of disability insurance recipients: Evidence from a notch design," *American Economic Journal: Economic Policy*, 11, 269–300.

SAEZ, E. (2001): "Using elasticities to derive optimal income tax rates," *The Review of Economic Studies*, 68, 205–229.

——— (2002): "Optimal income transfer programs: Intensive versus extensive labor supply responses," *Quarterly Journal of Economics*, 117, 1039 – 1073.

——— (2010): "Do taxpayers bunch at kink points?" *American Economic Journal: Economic Policy*, 2, 180–212.

SAMUELSON, P. A. (1954): "The pure theory of public expenditure," *The review of economics and statistics*, 387–389.

SLEMROD, J., C. WEBER, AND H. SHAN (2017): "The behavioral response to housing transfer taxes: Evidence from a notched change in DC policy," *Journal of Urban Economics*, 100, 137–153.

SRIVASTAV, S. (2023): "Bringing Breakthrough Technologies to Market: Solar Power and Feed-in-Tariffs," Tech. rep., Mimeo.

STONE, C. J. (1990): "Large-sample inference for log-spline models," *The Annals of Statistics*, 717–741.

ÜBERTRAGUNGSNETZBETREIBER (2016): "EEG-Jahresabrechnung 2016," Https://www.netztransparenz.de/portals/1/EEG-Jahresabrechnung_2016.pdf, Last accessed: 2017-09-30.

——— (2018): "EEG-Vergütungskategorien 2018," Https://www.netztransparenz.de/EEG/Verguetungs-und-Umlagekategorien, Last accessed: 2018-09-30.

A Data description

The data used in this paper are administrative and contain all solar panels connected to the grid and receiving subsidy payments. An adopter may be a household or a firm. The unit of observation is the aggregated capacity installed by an adopter at a specific location. Therefore, it is not possible to exploit the nonlinearities in the subsidy by splitting a large system into smaller ones and asking for separate payments for each. Additionally, when an adopter adds capacity to a preexisting system, the policymaker takes the preexisting capacity into account. Therefore, it is not possible to exploit the nonlinearities by splitting up a large adoption into smaller ones over time. The data provides information on the time point of adoption, the location, the electricity production, the applied subsidy rates, and the system's capacity. Table A1 shows the yearly number of adoptions. It is increasing in most years.

Table A1: Number of adoptions per year

Year	Number of adoptions	Relative proportion in %
Until 2001	23498	10
2002	10999	5
2003	11928	5
2004	26070	11
2005	36448	15
2006	32730	13
2007	39883	16
2008	61220	25
All years	242776	100

Table A2 shows the relative number of adoptions, the total installed capacity, and the relative installed capacity by capacity ranges. The distribution of the number of adoptions is highly skewed to the left. 65% of the adoptions are smaller or equal to 10 kWp. The distribution of aggregate capacity is much less skewed to the left—installations with capacities ranging from 10 to 100 kWp account for 63% of total capacity.

Table A2: Relative number of adoptions, total installed capacity, and relative installed capacity by capacity range

Capacity [kWp]	Rel. number [%]	Tot. capacity [MWp]	Rel. tot. capacity [%]
(0, 5]	33.7	267	9
(5, 10]	31.1	535	18
(10, 30]	29.0	1269	43
(30, 100]	5.6	596	20
(100, 150]	0.3	89	3
> 150	0.3	196	7
Total	100.0	2953	100

Note: The column "Rel. number" shows the fraction of adopters with an installed capacity in the range in column "Capacity." The column "Tot. capacity" shows the total installed capacity in each range. The column "Rel. tot. capacity" shows the installed capacity by range relative to the overall capacity.

Table A3 shows the subsidy schedule in Euro cents per kWh per capacity range and over time.

Table A3: The subsidy schedule in Euro cents per kWh

Year	< 30 kWp	30-100 kWp	>100 kWp
Until 2001	50.62	50.62	50.62
2002	48.10	48.10	48.10
2003	45.70	45.70	45.70
2004	57.40	54.60	54.00
2005	54.53	51.87	51.30
2006	51.80	49.28	48.74
2007	49.21	46.82	46.30
2008	46.75	44.48	43.99

Note: Source Übertragungsnetzbetreiber (2018)

B Details model

B.1 Heterogeneous discounting and radiation exposure

The German subsidy for solar panels is paid as a feed-in tariff. A feed-in tariff is a guaranteed fixed price for produced electricity. The subsidy payment depends on the installed capacity and the produced electricity. Electricity production is a function of the adopter-specific location and capacity. The location matters since climate conditions vary across locations. Moreover,

adopters may have heterogeneous discount rates when evaluating future income streams. Discounting matters because adopters take the adoption decision based on the present discounted value of the income stream produced by the solar panel.

A household installing capacity q^i produces electricity e_{it} in a given year: $e_{it} = w_{it}q^i$, where w_{it} is the efficiency of the panel in year t, which depends on weather conditions and the location. Suppose electricity e_{it} is remunerated according to the following kinked subsidy scheme, which depends on the installed capacity:

$$S_k(q, e_{it}) = s_l \; e_{it}, \qquad \text{for } q \leq q^K; \tag{16}$$

$$S_k(q, e_{it}) = s_l \; e_{it}\frac{q^K}{q} + s_l \rho e_{it}\frac{q - q^K}{q}, \qquad \text{for } q > q^K. \tag{17}$$

It follows that the subsidy payment in a certain year as a function of q is:

$$S_k(q, w_{it}) = s_l \; w_{it} \; q, \qquad \text{for } q \leq q^K; \tag{18}$$

$$S_k(q, w_{it}) = s_l w_{it} q^K + s_l \rho w_{it}(q - q^K), \qquad \text{for } q > q^K. \tag{19}$$

It follows that $S_k(q, e_{it}) = w_{it}S_k(q)$. An agent evaluates the present discounting value of all future subsidy payments when taking the adoption decision. The expected present discounted value of all payments is

$$\mathbb{E}_i \left[\sum_{t=0}^{20} \beta_i^t w_{it} S_k(q) \right] = S_k(q) \mathbb{E}_i \left[\sum_{t=0}^{20} \beta_i^t w_{it} \right] = S_k(q)\zeta^i. \tag{20}$$

The subsidy is paid for 20 years; assume that panels break down afterwards. Setting $\rho = 1$ shows that the equivalent is true for a linear scheme S_l.

The decision problem of an adopter is

$$\tilde{\pi}_v^i = \max_q \{\zeta^i S(q) - \tilde{c}_v^i(q)\}, \tag{21}$$

and participate if and only if

$$\tilde{\pi}_v^i \geq \tilde{c}_f^i, \tag{22}$$

where ζ^i captures the individual-specific discounting and location. Normalisation by ζ^i shows the equivalence of Problem (21) and Problem (1). Therefore, the model outlined in Section 3 implicitly accounts for subsidy payments via a feed-in tariff. In particular, it accounts for individual-specific discounting and location.

B.2 Formal derivations Section 3.2

Denote by $q^i(s)$ the optimal choice of adopter i under a linear subsidy with rate s. Define the intensive margin elasticity of adopter i under a linear subsidy with rate s by

$$\epsilon^i(s) = \frac{\mathrm{d} \ln q^i(s)}{\mathrm{d} \ln s}. \tag{23}$$

Formally, using this notation, Assumption 1 states that for all marginal subsidy rates s in the interval $[s_l, s_l\rho]$ and for all adopters i such that their capacity-choice under the counterfactual subsidy $q^i(s_l)$ is in an interval $[\underline{q}, \overline{q}]$ around the kink point q^K, it holds that the elasticity $\epsilon^i(s) = \epsilon$, where ϵ is a constant. Assumption 1 a is a reduced form assumption on an endogenous object, i.e., a high-level assumption. However, it is equivalent to the structural assumption B1:

Assumption B1 (Structural assumption intensive margin)**.** *The cost function is locally isoelastic. Formally, for all adopters i such that their capacity-choice under the linear subsidy $q^i(s_l)$ is in the interval $[\underline{q}, \overline{q}]$, and for all quantities q in the interval $[q^i(s_l\rho), q^i(s_l)]$, it holds that the variable cost function $c_v^i(.)$ is equal to*

$$c_v^i(q) = \theta^i q^{1+\frac{1}{\epsilon}}, \tag{24}$$

where θ^i is the variable cost type.

Lemma B1 (Equivalence intensive margin)**.** *Assumption 1 is equivalent to Assumption B1.*

The proof is in Appendix B.2.1.

Denote the choice of adopter i under the counterfactual subsidy S_l by q_l^i, where l stands for linear: $q_l^i = q^i(s_l)$. Similarly, denote the total cost of adopter i under the counterfactual subsidy S_l by c_t^i, where t stands for total: $c_t^i = c_v^i(q_l^i) + c_f^i$.

Corollary B1 (Type parameters)**.** *For each adopter i, there is a one-to-one mapping from the variable and fixed cost type (θ^i, c_f^i) to the choice and total cost under the counterfactual subsidy (q_l^i, c_t^i):*

$$\theta^i = \frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon}, \tag{25}$$

$$c_f^i = c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}. \tag{26}$$

Therefore, locally, the total cost function is equal to

$$c(q, q_l, c_t) = \underbrace{\frac{s_l}{{q_l}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}}}_{\textit{variable cost}} + \underbrace{c_t - q_l s_l \frac{\epsilon}{1+\epsilon}}_{\textit{fixed cost}}. \tag{27}$$

The type parameter (q_l, c_t) captures all relevant adopter-specific heterogeneity.

The proof is in Appendix B.2.2. Using (q_l, c_t) has the advantage that the type parameter has direct economic meaning. The type is equal to the choice and cost under the counterfactual subsidy. The mapping between (θ, c_f) and (q_l, c_t) depends on the counterfactual subsidy rate s_l. However, since s_l is observable and fixed, this dependence poses no problem. Note that I will drop the adopter-specific index i from now on.

The next paragraph imposes the structural iso-elasticity assumption on the participation margin. Denote by $f_{t|q_l}(.|q_l)$ and $F_{t|q_l}(.|q_l)$ the density and the CDF of the total cost c_t conditional on the type q_l. Define a functional $\eta(S, q)$ as

$$\eta(S, q) = \frac{f_{t|q_l}(S(q)|q)}{F_{t|q_l}(S(q)|q)} S(q), \tag{28}$$

where $S(.)$ is a general subsidy function. The participation margin elasticity under the counterfactual subsidy is $\eta(S_l, q_l)$.[20] Formally, using this notation, Assumption 2 states that for all subsidy functions $S(q)$ such that $S_l(q) \geq S(q) \geq S_k(q)$ and for all quantities q in an interval $[\underline{q}, \overline{q}]$ around the kink point, it holds that the functional $\eta(S, q) = \eta$, where η is the constant participation margin elasticity. Again, Assumption 2 is a high-level assumption. It is equivalent to the structural assumption B2:

Assumption B2 (Structural assumption participation margin)**.** *The conditional CDF of the total cost is locally isoelastic. Formally, for all values c_t in the interval $[S_k(q_l), S_l(q_l)]$, the conditional CDF of the total cost is equal to*

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)} \right)^{\eta}, \tag{29}$$

where η is the constant participation margin elasticity, and $\overline{c_t}(q_l)$ is a normalisation term.

Lemma B2 (Equivalence participation margin)**.** *Assumption 2 is equivalent to Assumption B2.*

The proof is in Appendix B.2.3.

B.2.1 Proof of Lemma B1

Assumption 1 $\Rightarrow$ Assumption B1:
By Assumption 1 and the definition of the elasticity

$$\epsilon = \frac{q^{i\prime}(s)\, s}{q^i(s)}, \text{ for all } s \text{ in } [s_l \rho, s_l]. \tag{30}$$

[20] Note that, in general, $\eta(S, q)$ is not the participation margin elasticity under subsidy $S(q)$ because c_t is defined with respect to the counterfactual subsidy S_l.

By the first order condition of the adopters' problem: $c_v^{i\,\prime}(q^i(s)) = s$, and by differentiating the FOC

$$q^{i\prime}(s) = \frac{1}{c_v^{i\,\prime\prime}(q^i(s))}. \tag{31}$$

It follows that for all q in $[q^i(s_l\rho), q^i(s_l)]$

$$\epsilon = \frac{c_v^{i\,\prime}(q)}{c_v^{i\,\prime\prime}(q)\, q}. \tag{32}$$

Denote the choice of adopter i under the counterfactual subsidy S_l by $q_l^i = q^i(s_l)$, where l stands for linear. Denote the total cost of adopter i under the counterfactual subsidy S_l by $c_t^i = c_v^i(q_l^i) + c_f^i$, where t stands for total. By the FOC $c_v^{i\,\prime}(q_l^i) = s_l$ and by definition $c_v^i(q_l^i) + c_f^i = c_t^i$. These two equalities together with the ordinary differential equation (32) form an initial value problem with solution

$$c_v^i(q) + c_f^i = \underbrace{\frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}}}_{\text{variable cost}} + \underbrace{c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}}_{\text{fixed cost}}. \tag{33}$$

The result follows by defining

$$\theta^i = \frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon}, \tag{34}$$

$$c_f^i = c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}. \tag{35}$$

Assumption B1 $\Rightarrow$ Assumption 1:
By the FOC

$$q^i(s) = \left(\frac{\epsilon}{(1+\epsilon)\theta^i}\right)^{\epsilon} s^{\epsilon}. \tag{36}$$

Using the definition of $\epsilon^i(s)$, it follows that $\epsilon^i(s) = \epsilon$.
qed.

B.2.2 Proof of Corollary B1

The cost function is equal to

$$c_v^i(q) = \theta^i q^{1+\frac{1}{\epsilon}}.$$

By the first order condition and the definition of q_l^i

$$\theta^i = \frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon}.$$

By the definition of c_t^i and plugging q_l^i and θ^i into the cost function

$$c_f^i = c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}.$$

Changing variable in Equation (24) gives the result.
qed.

B.2.3 Proof of Lemma B2

Assumption 2 $\Leftrightarrow$ Assumption B2:
By Assumption 2, for all $S(q)$ such that $S_l(q) \geq S(q) \geq S_k(q)$, and for all q in $[\underline{q}, \overline{q}]$,

$$\eta = \frac{f_{t|q_l}(S(q)|q)S(q)}{F_{t|q_l}(S(q)|q)}. \tag{37}$$

It is equivalent to the following statement: For all c_t such that c_t is in $[S_k(q_l), S_l(q_l)]$

$$\eta = \frac{f_{t|q_l}(c_t|q_l)c_t}{F_{t|q_l}(c_t|q_l)}. \tag{38}$$

The solution of this partial differential equation is

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)}\right)^{\eta}, \tag{39}$$

where $\overline{c_t}(q_l)$ is a normalisation term.
qed.

B.3 Proof of Proposition 1

The variable q_l is the choice of an adopter under the linear subsidy. To illustrate the dependence of capacity on the subsidy scheme, in this section, denote q_k the choice of the same adopter under the kinked subsidy. Note that the rest of the paper simply denotes it by q to avoid an overloaded notation.

Lemma B3. *The choice under the kinked subsidy q_k as a function of the choice under the linear subsidy q_l is*

$$q_k(q_l) = q_l, \qquad \textit{for } q_l < q^K; \tag{40}$$
$$q_k(q_l) = q^K, \qquad \textit{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; \tag{41}$$
$$q_k(q_l) = q_l\rho^{\epsilon}, \qquad \textit{for } q_l > q^K\rho^{-\epsilon}. \tag{42}$$

PROOF:
By Equation (27) and the first order condition of the adopters' maximisation problem

$$q(s, q_l) = q_l \left(\frac{s}{s_l}\right)^{\epsilon}. \tag{43}$$

Below the kink point $s = s_l$. Therefore,

$$q_k(q_l) = q_l, \text{ for } q_l < q^K. \tag{44}$$

Adopters well above the kink point produce the same as under a linear subsidy with the marginal rate $s = s_l\rho$. It follows that

$$q_k(q_l) = q_l\rho^{\epsilon}, \text{ for } q_l \gg q^K. \tag{45}$$

Generally, adopters above the kink point reduce their production and produce $q_l\rho^{\epsilon}$. However, for adopters in the interval $q_l \in (q^K, q^K\rho^{-\epsilon})$ it would mean to reduce the production below q^K. As soon as they reduce production to q^K, they are not affected by the lower marginal price any more, and therefore it cannot be optimal to reduce below q^K. It follows that all adopters in this interval chose to produce exactly q^K; they "bunch" at q^K.
qed.

Denote the difference in cost of an adopter q_l between the kinked and the linear subsidy by $\Delta c(q_l) = c(q_k(q_l), q_l, c_t) - c_t$.

Lemma B4. *The difference in cost $\Delta c(q_l)$ of adopter q_l between the kinked and linear subsidy is*

$$\Delta c(q_l) = 0, \qquad \textit{for } q_l < q^K; \tag{46}$$
$$\Delta c(q_l) = \frac{1}{1+\epsilon^{-1}} \frac{s_l}{q_l^{1/\epsilon}} ((q^K)^{1+\epsilon^{-1}} - q_l^{1+\epsilon^{-1}}), \qquad \textit{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; \tag{47}$$
$$\Delta c(q_l) = \frac{1}{1+\epsilon^{-1}} s_l q_l (\rho^{\epsilon+1} - 1), \qquad \textit{for } q_l > q^K\rho^{-\epsilon}. \tag{48}$$

PROOF:

By Corollary B1

$$c(q, q_l, c_t) = c_t + \left[\frac{q^{1+1/\epsilon}}{q_l^{1/\epsilon}} - q_l\right] \frac{s_l}{1+1/\epsilon}. \tag{49}$$

By definition and Lemma B3

$$\begin{aligned}
\Delta c(q_l) = c(q_k(q_l), q_l, c_t) - c_t &= c(q_l, q_l, c_t) - c_t && \text{for } q_l < q^K; && (50)\\
&= c(q^K, q_l, c_t) - c_t && \text{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; && (51)\\
&= c(q_l\rho^{\epsilon}, q_l, c_t) - c_t && \text{for } q_l > q^K\rho^{-\epsilon}. && (52)
\end{aligned}$$

Use Equation (49) in Equation (50)-(52) to get Equation (46)-(48).
qed.

Define the function $R(q_l)$ as the net subsidy of adopter q_l under the kinked scheme as a fraction of the subsidy under the linear scheme:

$$R(q_l) = \frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}. \tag{53}$$

Lemma B5. *The function $R(q_l)$ is:*

$$\begin{aligned}
R(q_l) &= 1, && \textit{for } q_l < q^K; && (54)\\
R(q_l) &= \frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 - \left(\frac{q^K}{q_l}\right)^{\frac{1+\epsilon}{\epsilon}}\right), && \textit{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; && (55)\\
R(q_l) &= (1-\rho)\frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 + \frac{\rho^{\epsilon+1}}{\epsilon}\right), && \textit{for } q_l > q^K\rho^{-\epsilon}. && (56)
\end{aligned}$$

PROOF:
By definition

$$R(q_l) = \frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}, \tag{57}$$

which together with Lemma B3 and B4 gives Equation (54)-(56).
qed.

Lemma B6. *The mass of participating adopters under the kinked subsidy as a function of q_l is*

$$R(q_l)^{\eta} f_l(q_l), \tag{58}$$

where f_l is the measure of capacity under the linear subsidy.

PROOF:
An adopter participates if its cost is smaller than the received subsidy: $c(q_k(q_l), q_l, c_t) \leq S_k(q_k(q_l))$. Using definitions, this is equivalent to $c_t \leq S_k(q_k(q_l)) - \Delta c(q_l)$. Given a certain q_l, the mass of adopters participating as a function of q_l is

$$\frac{F_{t|q_l}(S_k(q_k(q_l)) - \Delta c(q_l)|q_l)}{F_{t|q_l}(S_l(q_l)|q_l)} f_l(q_l), \tag{59}$$

where f_l is the hypothetical measure of q_l under the linear subsidy. By Assumption 2

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)}\right)^{\eta}, \text{ for all } c_t \text{ in } [S_k(q_l), S_l(q_l)]. \tag{60}$$

Note that by revealed preference $S_k(q_l) \leq S_k(q_k(q_l)) - \Delta c(q_l)$. It follows that

$$\frac{F_{t|q_l}(S_k(q_k(q_l)) - \Delta c(q_l)|q_l)}{F_{t|q_l}(S_l(q_l)|q_l)} f_l(q_l) = \left(\frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}\right)^{\eta} f_l(q_l), \tag{61}$$

which, together with the definition of R, gives the result.
qed.

PROOF of Proposition 1:
Change variable in Equation (58) using Lemma B3 to derive Equation (5) and (7). Integrate Equation (58) over $\left[q^K, q^K\rho^{-\epsilon}\right]$ to derive Equation (6). Remember, to avoid an overloaded notation, the rest of the paper denotes the choice under the kinked subsidy simply by q.
qed.

B.4 Proof of Proposition 2

By Assumption 3

$$g_f(f_l(q)) = \sum_{p=0}^{\infty} \gamma_p \frac{1}{p!} \left(g_q(q) - g_q(q^K)\right)^p, \tag{62}$$

where g_f is the transformation of f_l and g_q is the transformation of q. Assume the transformation is known. Section D.8 discuss the choice of transformation. The measure f_l is identified because all γ_p are identified from the left limit

$$\gamma_p = \lim_{q- \to q^K} \frac{\mathrm{d}^p g_f(f_k(q))}{\mathrm{d}(g_q(q) - g_q(q^K))^p}, \text{ for all } p. \tag{63}$$

The elasticities ϵ and η are jointly identified as the simultaneous solution to Equation (6)

and (7). By Condition 1, the solution is locally unique. Because Condition 1 holds generically, ϵ, η, and f_l are generically identified. The response margins ϵ and η are locally identified. I verify the uniqueness of the solution to Equations (5) to (7) ex-post estimation, which provides strong evidence for global identification. It is important that the interval $[\underline{q}, \overline{q}]$ is large enough. It needs to hold that the lower bound $\underline{q} < q^K$ and the upper bound $\overline{q} > q^K \rho^{-\epsilon}$. The upper bound depends on the unknown parameter. For most applications, it is safe to assume that ϵ is bounded, i.e., the elasticity ϵ is an element of a bounded interval $[0, \overline{\epsilon})$. It follows that it suffices to assume $\overline{q} = q^K \rho^{-\overline{\epsilon}}$. For unbounded ϵ, it is necessary to assume $\overline{q} = \infty$.
qed.

B.5 Identification in a discontinuous (i.e., notched) incentive scheme

This section generalizes the result in Proposition 2 for the case when there is a discontinuity (i.e., a notch) in the incentive scheme.

Consider the discontinuous subsidy schedule $S_d(.)$:

$$S_d(q) = s_l q, \qquad \text{for } q \leq q_D; \tag{64}$$

$$S_d(q) = s_l q - \Delta S \; s_l, \qquad \text{for } q > q_D; \tag{65}$$

where q_D denotes the notch point and ΔS denotes the size of the notch relative to s_l.

Proposition B1 (The observed density in case of a notch)**.** *Close to the notch point, the observable measure $f_d(.)$ under the notched subsidy $S_d(.)$ is a function of three unknowns: the intensive margin elasticity ϵ, the participation margin elasticity η, and the counterfactual measure $f_l(.)$. At the notch point q_D, there is a mass point with bunching mass B. Four parts of the observable measure $f_d(.)$ depend distinctly on the three unknowns:*

$$f_d(q) = f_l(q), \qquad \textit{for } q < q_D; \tag{66}$$

$$B = \int_{q_D}^{q_B(\epsilon, \Delta S)} R(q_l, \epsilon)^\eta f_l(q_l) dq_l, \qquad \textit{for } q = q_D; \tag{67}$$

$$f_d(q) = 0, \qquad \textit{for } q_D < q < q_B; \tag{68}$$

$$f_d(q) = \left(1 - \frac{\Delta S}{q}\right)^\eta f_l(q), \qquad \textit{for } q > q_B. \tag{69}$$

Note: The variable q_D denotes the notch point; ΔS denotes the size of the notch relative to s_l; $R(., \epsilon)$ defined in Equation (55) is the net subsidy payment to an adopter under the notched scheme relative to the subsidy payment under the counterfactual scheme. The variable $q_B(\epsilon, \Delta S)$ defined in Equation (73) defines the quantity-choice of the marginal buncher. It is a function of the intensive margin elasticity ϵ and the size of the notch ΔS.

PROOF:
Use the FOC of adopters to derive that

$$q = q_l, \qquad \text{for } q_l < q_D; \tag{70}$$

$$q = q_D, \qquad \text{for } q_D \leq q_l \leq q_B; \tag{71}$$

$$q = q_l, \qquad \text{for } q_l > q_B. \tag{72}$$

The variable $q_B(\epsilon, \Delta S)$ denotes the quantity-choice of the marginal buncher. It is implicitly defined by the indifference condition of that agent:

$$q_D - \frac{\epsilon}{1+\epsilon} \frac{1}{q_B^{\frac{1}{\epsilon}}} {q_D}^{1+\frac{1}{\epsilon}} = q_B \frac{1}{1+\epsilon} - \Delta S \tag{73}$$

The rest of the proof follows the steps in Section B.3.
qed.

Proposition B2 (Identification in case of a notch)**.** *The observable measure $f_d(.)$ identifies the counterfactual measure $f_l(.)$, the intensive margin elasticity ϵ, and the participation margin elasticity η.*

PROOF:
As in Section B.4. Three parts of the measure (66), (67), and (69) are sufficient to identify the three unknowns.
qed.

C Details estimation

C.1 The estimation of $\widehat{\ln f(q_j)}$

As a first step, construct the empirical histogram $\widehat{f(q_j)}$ by choosing bins and counting the number of adopters in each bin. Normalisation, by the bin size and the total number of adopters, gives the observed density $\widehat{f(q_j)}$ at point q_j, where the index j in $\{-N_-, ..., -1, 0, 1, ..., N_+\}$ is the index of the bin and $N = N_- + N_+$ is the total number of bins.

The empirical model (9) in Section 4 uses the logarithm of $\widehat{f(q_j)}$ as the dependent variable. Using the logarithm does not affect consistency (see Appendix C.1.1); however, it introduces a small sample bias. Appendix C.1.2 applies bias-correction techniques to reduce the bias's impact. The variable $\widehat{\ln f(q_j)}$ denotes the bias-corrected dependent variable. The logarithmic model (9) has several advantages over the additive model $\widehat{f(q_j)} = f_k(q_j \mid \eta, \gamma, \epsilon) + u_j$. First, the observed density is positive by definition in the logarithmic model. The additive model's

density is negative for large negative u_j. Therefore, only the logarithmic model is logically consistent. Second, the noise term u_j may capture additional random disturbances besides sampling noise. I.e. for some random reasons, specific capacities may be more or less frequent in the data than predicted by f_k. It is more natural to model these additional disturbances as proportional to f_k.

The logarithm makes it necessary to have at least one observation in each bin because the logarithm of zero is not defined. Additionally, even after bias correction, the small sample bias caused by the logarithm decreases in each bin's number of observations. Therefore, it is preferable to avoid bins with a small number of observations. The observed density is decreasing in capacity q. As a consequence, the expected number of observations in an interval decreases in capacity. To counteract this effect, use bins with a bin size that increases in capacity. Such a binning procedure does not affect consistency because $\widehat{f(q_j)}$ equals the number of observations in bin j normalised by the bin size and the total number of observations. An additional advantage of this binning procedure is that it equalises the variance of the dependent variable $\widehat{\ln f(q_j)}$. Therefore, it avoids the need for a weighting matrix in the estimation. Concretely, use the function

$$\overline{q_j} = q^K (1 - j c_0(n))^{-\frac{1}{\omega}}, \tag{74}$$

where $\overline{q_j}$ is the right border of a bin and $c_0(n)$ is a constant that goes to zero as n goes to infinity. This binning function is advantageous for the following reason. The variance of the log-histogram depends on the number of observations in each bin. Approximately the log-histogram has constant variance if the number of observations is approximately constant. The distribution of observations is close to a Pareto-distribution. It follows that the expected number of observations in a bin is approximately equal to

$$\phi_0 \left(\frac{\overline{q_j}}{q^K} \right)^{-\phi_1} - \phi_0 \left(\frac{\overline{q}_{j+1}}{q^K} \right)^{-\phi_1}, \tag{75}$$

where ϕ_0 and ϕ_1 are the parameters of the Pareto-distribution. If ω is equal to ϕ_1, the above expression is equal to $\phi_0 c_0(n)$, which is constant.

C.1.1 Consistency of $\widehat{\ln f(q)}$

Denote by n the sample size and denote by Q_ι a single observation. The variable q_j is the centre of a bin. Denote by $\widehat{N_j}$ the number of observations in a bin. The variable h_j is the bin size. It is a function of n and goes to zero as n goes to infinity. The constructed dependent variable is

$$\widehat{f(q_j)} = \frac{\widehat{N_j}}{n h_j}. \tag{76}$$

Lemma C1. *$\widehat{f(q_j)}$ converges to $f_k(q_j)$ in probability, which is also true in logarithms:*

$$\widehat{f(q_j)} \xrightarrow{p} f_k(q_j); \tag{77}$$

$$\ln \widehat{f(q_j)} \xrightarrow{p} \ln f_k(q_j). \tag{78}$$

The corresponding result holds for $\widehat{B}$.

PROOF:
Write $\widehat{f(q_j)}$ as

$$\widehat{f(q_j)} = \frac{1}{n} \sum_{\iota=1}^{n} \frac{\mathbb{1}[q_j - h_j/2 \leq Q_\iota < q_j + h_j/2]}{h_j}. \tag{79}$$

Using the law of large numbers, it follows that

$$\frac{1}{n} \sum_{\iota=1}^{n} \frac{\mathbb{1}[q_j - h_j/2 \leq Q_\iota < q_j + h_j/2]}{h_j} \xrightarrow{p} \frac{1}{h_j} \int_{q_j - h_j/2}^{q_j + h_j/2} f_k(q) dq. \tag{80}$$

As h_j goes to zero when n goes to infinity

$$\frac{1}{h_j} \int_{q_j + h_j/2}^{q_j - h_j/2} f_k(q) dq = f_k(q_j). \tag{81}$$

By the continuous mapping theorem, it follows that

$$\ln \widehat{f(q_j)} \xrightarrow{p} \ln f_k(q_j). \tag{82}$$

The same arguments hold for $\widehat{B}$.
qed.

C.1.2 Bias and bias-correction of $\ln \widehat{f(q_j)}$

While $\ln \widehat{f(q_j)}$ converges to $\ln f_k(q_j)$, $\mathbb{E}\left[\ln \widehat{f(q_j)}\right]$ is not equal to $\ln f_k(q_j)$. Therefore, using the logarithm introduces a small-sample bias. To counteract this effect, model (9) in Section 4 uses the bias-corrected dependent variable

$$\widehat{\ln f(q_j)} = \ln \widehat{f(q_j)} + \frac{1}{2\widehat{N_j}}, \tag{83}$$

where $\widehat{N_j}$ denotes the number of observations in a bin. Note that $\widehat{\ln f(q_j)}$ denotes the bias-corrected dependent variable, while $\ln \widehat{f(q_j)}$ denotes the logarithm of the histogram. Using

(83) reduces bias since

$$\mathbb{E}\left[\widehat{\ln f(q_j)}\right] = \ln f_k(q_j) + O\left(\frac{1}{N_j^2}\right), \tag{84}$$

where N_j is the expected value of $\widehat{N_j}$.

PROOF:

Taylor approximate all functions of random variables in Equation (83) around their expected values and use $f_k(q_j)\ n\ h_j = N_j$:

$$\begin{aligned}\widehat{\ln f(q_j)} \approx &\ln f_k(q_j) + \frac{1}{N_j}(\widehat{N_j} - N_j) - \frac{1}{2N_j^2}(\widehat{N_j} - N_j)^2 + \frac{1}{3N_j^3}(\widehat{N_j} - N_j)^3 + \\ &+ \frac{1}{2N_j} - \frac{1}{2N_j^2}(\widehat{N_j} - N_j) + \frac{1}{2N_j^3}(\widehat{N_j} - N_j)^2.\end{aligned} \tag{85}$$

Take the expectation on both sides above. Note that, because $\widehat{N_j}$ follows a Binomial distribution, $\mathbb{E}(\widehat{N_j} - N_j) = 0$, $\mathbb{E}(\widehat{N_j} - N_j)^2 = N_j$, and $\mathbb{E}(\widehat{N_j} - N_j)^3 = N_j$.
qed.

C.2 Asymptotic normality of $\hat{\epsilon}$ and $\hat{\eta}$

The histogram is asymptotically normal:

$$\lim_{n\to\infty} \sqrt{nh_j(n)}\left(\widehat{f(q_j)} - f_k(q_j)\right) \sim N(0, f_k(q_j)). \tag{86}$$

By the delta method, the log-histogram is asymptotically normal:

$$\lim_{n\to\infty} \sqrt{nh_j(n)}\left(\widehat{\ln f(q_j)} - \ln f_k(q_j)\right) \sim N(0, \frac{1}{f_k(q_j)}). \tag{87}$$

The binning function keeps the variance of the log-histogram constant. Approximately $h_j \approx h_0(n)\frac{1}{f_k(q_j)}$, and $h_0(n)$ decreases slowly with sample size. It follows that asymptotically

$$\lim_{n\to\infty} \sqrt{nh_0(n)}\left(\widehat{\ln f(q_j)} - \ln f_k(q_j)\right) \sim N(0, 1). \tag{88}$$

Write

$$\widehat{\ln f(q_j)} = \ln f_k(q_j|\epsilon, \eta, f_l) + u_j \tag{89}$$

where

$$u_j \sim N(0, \frac{1}{nh_0(n)}). \tag{90}$$

The errors u_j are approximately uncorrelated for n large enough. Chen (2007) shows that the nonlinear least square estimate of Equation (89), where the nonparametric function $\ln f_k(q_j|\epsilon, \eta, f_l)$ is replaced by $\ln f_k(q_j|\epsilon, \eta, \gamma_P)$ and P goes slowly to infinity as sample size goes to infinity, gives consistent and asymptotically normal estimates $\hat{\epsilon}$ and $\hat{\eta}$. This result does also hold when $\hat{\epsilon}$ and $\hat{\eta}$ are constraint to positive values using an exponential transformation, i.e., $\hat{\epsilon} = \exp(\hat{\beta}_\epsilon)$ and $\hat{\eta} = \exp(\hat{\beta}_\eta)$. I use the auxiliary parameters $\hat{\beta}_\epsilon$ and $\hat{\beta}_\eta$ to minimise the sum of squares. By the same argument as above, they are asymptotically normal. It implies that $\hat{\epsilon}$ and $\hat{\eta}$ are asymptotically normal.

C.3 Estimation of the mean squared error

The bias and the variance of the estimates depend on the order of the polynomial P. It is standard in nonparametric estimations to choose a specification that minimises an estimate of the mean squared error. As the estimate of the participation margin $\hat{\eta}$ is more sensitive to the specification than the intensive margin $\hat{\epsilon}$, I use an estimate of its mean squared error to choose the specification. The estimate $\hat{\eta}(P, n)$ is a function of the specification parameter and the sample size n. The mean squared error is defined as

$$MSE = \mathbb{E}\left[(\hat{\eta}(P, n) - \eta)^2\right], \tag{91}$$

where η is the true value of the parameter. A standard bias-variance decomposition renders

$$MSE = \underbrace{\mathbb{E}\left[(\hat{\eta}(P, n) - \mathbb{E}\left[\hat{\eta}(P, n)\right])^2\right]}_{\text{variance}} + \underbrace{\left(\mathbb{E}\left[\hat{\eta}(P, n)\right] - \eta\right)^2}_{\text{bias}^2}. \tag{92}$$

Define

$$\tilde{\eta}(P) := \lim_{n \to \infty} \hat{\eta}(P, n), \tag{93}$$

where P is kept constant.[21] Intuitively, $\tilde{\eta}(P)$ is the "true" value of the parameter under the parametric specification P. Therefore, $\tilde{\eta}(P) - \eta$ is the estimator's specification bias. Note that in large enough samples $\mathbb{E}\left[\hat{\eta}(P, n)\right] \approx \tilde{\eta}(z)$. Therefore, the mean squared error is the sum of the "parametric variance" and the estimator's squared specification bias. All parts of the MSE are unknown and need to be estimated. The nonparametric bootstrap provides a consistent estimate for the variance.

Estimating the bias is more challenging since it depends on $\tilde{\eta}(P)$ and the true value η. A consistent estimate of $\tilde{\eta}(P)$ is the estimate $\hat{\eta}(P, n)$ itself. However, a consistent estimate of η, which converges fast, is challenging to find. Therefore, I estimate the bias out of sample on untreated data. This approach has the advantage that, out of sample, the true value of η is

[21] Note that if P changes accordingly with sample size, $\tilde{\eta}(P)$ converges to the true value η. However, that is not true if the specification is kept constant.

known. It is equal to zero because there is no treatment. The estimate of the bias is

$$\widehat{bias(\hat{\eta}(P,n))} = \hat{\eta}_{os}(P,n) - \eta_{os} = \hat{\eta}_{os}(P,n), \tag{94}$$

where $\hat{\eta}_{os}$ is the estimate of the participation margin out of sample. This estimate of the bias has the advantage that it converges at a parametric rate. This approach relies on the assumption that absent the treatment, the treated data is similar to the out-of-sample data. Intuitively, any effect estimated on the untreated data is a bias due to the specification. The next section discusses the derivation of the estimate of the bias in detail.

C.3.1 The estimator of the bias of $\hat{\eta}$

Denote by $(\tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P)$ the estimate of $(\epsilon, \eta, \gamma_P)$ under specification P and when sample sizes goes to infinity, e.g., $\tilde{\eta} := \lim_{n\to\infty} \hat{\eta}(P,n)$, where P is kept constant. The parameter vector γ_P denotes $(\gamma_0, ..., \gamma_P)$. The parameter $\tilde{\gamma}_P$ is identified from the derivatives of the observed distribution to the left of the kink point. It can be estimated without bias because the bandwidth $[\underline{q}, \overline{q}]$ converges to q^K as n goes to infinity. Therefore, assume $\tilde{\gamma}_P = \gamma_P$. The parameter ϵ is much less sensitive to the specification than the parameter η. Assume that P is large enough such that $\tilde{\epsilon} \approx \epsilon$. The parameter $\tilde{\eta}$ is estimated from points just to the right of the kink point. Because the bandwidth $[\underline{q}, \overline{q}]$ converges to q^K as n goes to infinity,

$$\lim_{q\downarrow q^K} \ln f_k(q|\epsilon, \eta, \gamma) = \lim_{q\downarrow q^K} \ln f_k(q|\tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P) \tag{95}$$

Therefore, using Equation (7) renders

$$\sum_{p=0}^{\infty} \gamma_p \frac{1}{p!} \left(\ln \left(\frac{q^K \rho^{-\epsilon}}{q^K} \right) \right)^p + \eta \ln R(q^K \rho^{-\epsilon}, \epsilon) + \ln \left(\rho^{-\epsilon} \right) \tag{96}$$

$$= \sum_{p=0}^{P} \tilde{\gamma}_p \frac{1}{p!} \left(\ln \left(\frac{q^K \rho^{-\tilde{\epsilon}}}{q^K} \right) \right)^p + \tilde{\eta} \ln R(q^K \rho^{-\tilde{\epsilon}}, \tilde{\epsilon}) + \ln \left(\rho^{-\tilde{\epsilon}} \right). \tag{97}$$

Use $\gamma_P = \tilde{\gamma}_P$, $\epsilon = \tilde{\epsilon}$, and rearrange to derive that

$$\tilde{\eta} - \eta = \frac{\sum_{p=P}^{\infty} \gamma_p \frac{1}{p!} \left(\ln \left(\rho^{-\epsilon} \right) \right)^p}{\ln R(q^K \rho^{-\epsilon}, \epsilon)} \tag{98}$$

The bias in η depends only on ϵ and on the un-estimated rest of the parameter γ, i.e., $(\gamma_{P+1}, \gamma_{P+2}, ...)$.

Estimating the bias on out-of-sample data is possible if the counterfactual distributions in the treated and untreated data are similar. More specifically, assume that there exists a certain order of the series expansion of the two distributions, such that all coefficients above that order

are equal. Mathematically,

$$\ln(f_l^x(q)) = \sum_{p=0}^{\infty} \gamma_p^x \frac{1}{p!} \left(\ln \left(\frac{q}{q^K} \right) \right)^p, \tag{99}$$

where x is either is (in sample) or os (out of sample) and denotes the respective counterfactual measure. Assume there exists an order j^* such that for all $j \geq j^*$ the coefficients are equal: i.e., $\gamma_j^{is} = \gamma_j^{os}$. Furthermore, assume that $P \geq j^*$.

Additionally, Equation (98) shows that the bias depends on the intensive margin response ϵ. To consider this dependence, first estimate $\tilde{\epsilon}$ on the treated data using an auxiliary specification. Second, simulate the intensive margin response in the untreated data using the auxiliary estimate $\tilde{\epsilon}$. Third, estimate $\eta_{os}(P)$ in the untreated data. I do not constrain $\eta_{os}(P)$ to positive values and I use the inequality constraint $\epsilon_{os}(P) \geq 0$. A discussion of the choice of the out-of-sample data and the estimates of the MSE are in Section C.4.

C.4 The selection of the optimal specification

The counterfactual analysis in Section 6 uses the estimates from the pooled sample from 2004 to 2008. Therefore, this section chooses the optimal specification for this sample. For simplicity, I use this specification also for the yearly estimates in Table 1. The estimation of the variance uses the nonparametric bootstrap with 200 repetitions at 30kWp and 1000 repetitions at 100kWp and 250 kWp.

C.4.1 The untreated data for estimating the bias

As discussed in Section C.3, for each kink point, it is necessary to choose a range of untreated data to estimate the bias. A natural choice is observations in the years 2000 to 2003. In these years, the subsidy was linear. Therefore, the counterfactual distribution is directly observable. I use it to estimate the bias at 30kWp.

For the kink point at 100 kWp, the pre-treatment data is not a satisfactory choice to estimate the bias for two reasons. First, the number of observations around 100 kWp is very low these years. Second, in 2000-2003, the data do not specify whether a solar panel was installed on a rooftop or the ground. From 2004 onwards, the data specifies where a panel is installed. This paper only considers rooftop panels. Overall, ground panels are only a very small share of installations. Also, the subsidy for ground panels is linear in all years. After 2004 and close to capacity 30kWp, only very few panels are ground panels. Therefore, the fact that the sample from 2000-2003 contains ground panels does not pose a concern for using these years to estimate the bias at 30 kWp. However, this is not true for capacities close to 100 kWp. There is a significant number of ground panel installations exactly at 100 kWp in the years after 2004. For these reasons, I cannot use observations around 100 kWp in the years 2000-2003 to

estimate the bias at 100 kWp. To be conservative, I remove the observations at and around 100kWp when using the data to estimate the bias of the parameter at 30 kWp.

Therefore, to estimate the bias at 100 kWp, I use observations around a point similar to 100 kWp in 2004-2008. On the one hand, for the counterfactual distribution to be similar, the point should be close to 100 kWp. On the other hand, it should be far enough from 100 not to include observations affected by the kink. I choose the point 250 kWp because it satisfies these requirements. Like the point 100 kWp, the point 250 kWp is a focal point (i.e., it is a quarter of 1,000 kWp).

The out-of-sample data needs to be similar to the treated data in the sense discussed in Section C.3. Section D.7 checks the similarity of the treated and untreated data for the estimates at 30 and 100 kWp. Section C.4.3 reports the estimates of the mean squared error.

C.4.2 The parameters of the binning function

As discussed in Section C.2, the binning function in Equation (74) should guarantee a constant variance of the log-histogram and sufficiently many observations in each bin. To select its parameters, I pre-select initial maximal bandwidths. For the estimation at 30 kWp, I use a maximal bandwidth of $[9.5, 95]$. I use this interval because it is symmetric around 30 on the logarithmic scale, and the upper limit is such that the sample does not contain observations from the second kink point at 100.

For the estimation at 100 kWp, I use the interval $[42, 1600]$. I use this interval for the following reasons. The proportional interval around 250 kWp, which I use to estimate the bias, is $[105, 4000]$.[22] The lower limit for the interval around 250 is 105 - and therefore, in proportion, 42 around 100 - to keep a distance from the next kink point at 100. The upper limit is 4,000 - and therefore, in proportion, 1,600 - because there are only very few observations above 4,000 kWp. The interval is asymmetric to increase the sample size.

In the next step, I choose the size of the bin at the kink point h_0 and the scaling parameter ω. I choose the two parameters such that there are at least four observations in each bin, and the variance of the log-histogram in an auxiliary estimation is approximately constant. By Equation (84), four observations guarantee that the bias introduced by the logarithm is of the order $\frac{1}{16}$. The procedure gives the following specification:

[22] $42\frac{250}{100} = 105$ and $1600\frac{250}{100} = 4000$.

Table C1: Selected bin size h_0 and scaling parameter ω.

Years	Interval	Bin Size	Scaling Parameter
2004-08 pooled	[9.5, 95]	0.18	-0.35
2004-08 pooled	[42, 1600]	1.7	1.2
2004-08 pooled	[105, 4000]	10	1.5
2000-03 pooled	[9.5, 95]	1.5	0.6
2004-11 yearly	[9.5, 95]	0.8	0.1

Note: The table shows the selected bin size h_0 and scale parameter ω of the binning function in Equation (74). The parameters guarantee a minimum of 4 observations in each bin and an approximately constant variance.

C.4.3 Estimates of the mean squared error

As discussed above, this section estimates the mean squared error using the variance estimate from the treated sample and the bias estimate from the untreated sample. To estimate the variance, it uses the nonparametric bootstrap. Table C2 shows the estimated mean squared error (MSE) for different bandwidths and orders P of the series at 30 kWp.

Table C2: Estimated mean squared error of $\widehat{\eta_{30}}$

Bandwidth	P=1	P=2	P=3
[16.4, 55]	2611	3960	40268
[15.0, 60]	3956	490	40129
[13.8, 65]	1779	457	1880
[12.9, 70]	764	23681	10150
[12.0, 75]	14	19428	18436
[11.2, 80]	121	15078	17047
[10.6, 85]	6	1855	4265
[10.0, 90]	74	1871	4742
[9.5, 95]	139	120	2708

Note: The table reports the estimated mean squared error (MSE) of $\widehat{\eta_{30}}$ for different bandwidths b and orders of the series P. The estimate is the lowest for $P = 1$ and $b = [10.6, 85]$.

The optimal bandwidth for the estimation at 30 kWp is $[10.6, 85]$; the optimal order of the series is $P = 1$.

For the kink at 100 kWp, Table C3 shows the estimated mean squared error (MSE) for different bandwidths and orders P of the series.

Table C3: Estimated mean squared error of $\widehat{\eta_{100}}$

Bandwidth	P=1	P=2	P=3
[67, 150]	23007	40025	40287
[50, 200]	478	24513	28034
[42, 400]	41	3159	4459
[42, 700]	701	9751	3541
[42, 1000]	700	9750	2212
[42, 1300]	369	733	9046
[42, 1600]	379	755	8990

Note: The table reports the estimated mean squared error (MSE) of $\widehat{\eta_{100}}$ for different bandwidths b and orders of the series P. The estimate is the lowest for $P = 1$ and $b = [42, 400]$.

The optimal bandwidth for the estimation at 100 kWp is $[42, 400]$; the optimal order of the series is $P = 1$.

For both kink points, the optimal bandwidth is relatively large, and the series' optimal order is low. Consistent with the graphical evidence in Figure 4, this result shows that the counterfactual distribution is very close to a Pareto distribution.

D Additional results and robustness

D.1 The estimates of the participation margin elasticity $\hat{\eta}$

Since $\kappa = \eta/S(q)$, it follows that $\hat{\kappa}_{30} = \frac{\hat{\eta}_{30}}{30}$ and $\hat{\kappa}_{100} = \frac{\hat{\eta}_{100}}{96.5}$. The standard deviation of $\hat{\kappa}$ follows from the delta method: $SD_\kappa = SD_\eta/S(q)$.

Table D1: Estimates $\hat{\eta}$ at various kink points

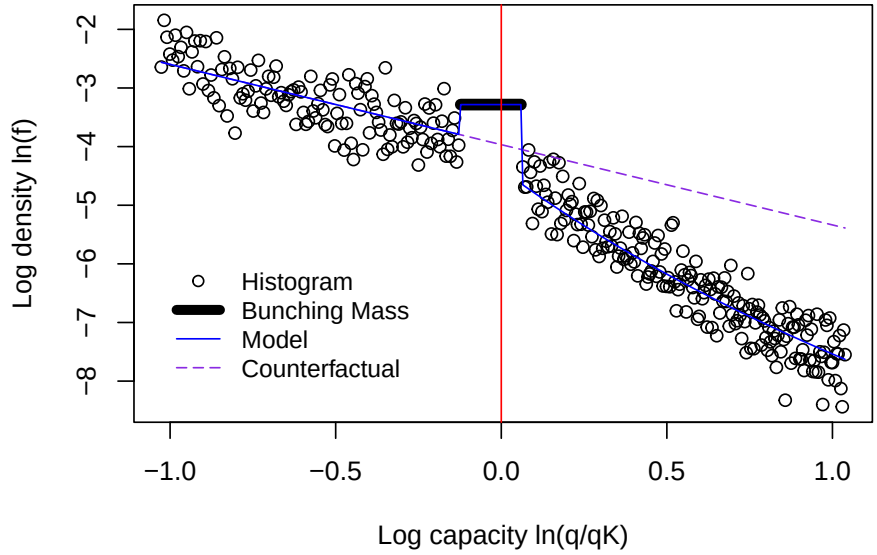
Kink Point	$\hat{\eta}$ (SD)
30 kWp 2004	96.2 (4.3)
30 kWp 2005	74.7 (3.9)
30 kWp 2006	56.4 (4.2)
30 kWp 2007	60.1 (3.5)
30 kWp 2008	52.7 (2.6)
30 kWp 2004-08	69.2 (1.7)
100 kWp 2004-08	0.0 (2.4)

Note: The table reports the estimated participation elasticity $\hat{\eta}$ with standard errors in brackets.

D.2 Estimation pooling observations from 2004 to 2008

The optimal specification, derived in Section C.4, for the estimation at 30 kWp is: bin-size $h_0 = 0.18$, scaling parameter $\omega = -0.35$, bandwidth $b = [10.6, 85]$, bunching window $[26.5, 31.75]$, and series order $P = 1$. Figure D1 shows the histogram with the estimated model and the counterfactual. The results are in Table 2.

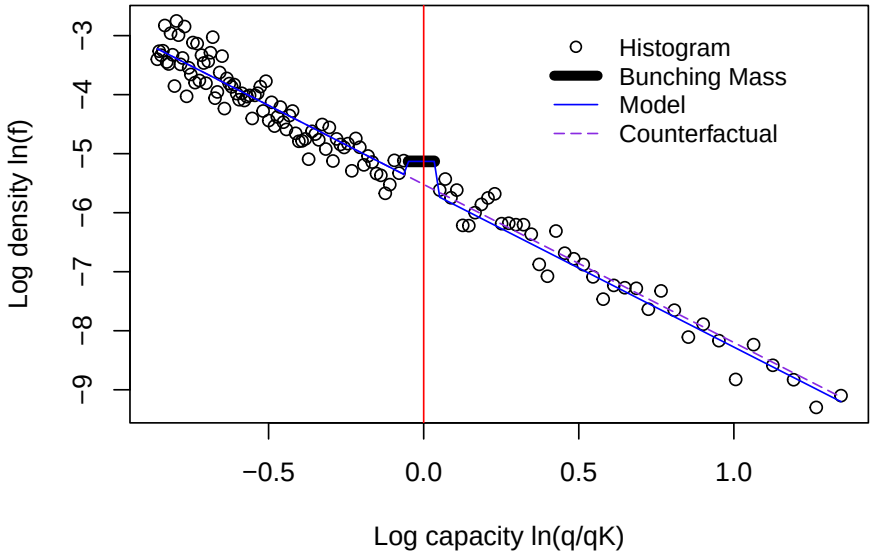
Figure D1: Histogram for observations from 2004 to 2008 at 30 kWp with estimated model and counterfactual



Note: The x-axis shows the normalised logarithm of capacity. The y-axis shows the logarithm of the density. The red line marks the kink point. The estimation minimises the distance between the data in black and the model in blue.

The optimal specification, derived in Section C.4, for the estimation at 100 kWp is: bin-size $h_0 = 1.7$, scaling parameter $\omega = 1.2$, bandwidth $b = [42, 400]$, bunching window $[95, 102.5]$, and series order $P = 1$. Figure D2 shows the histogram with the estimated model and the counterfactual. The results are in Table 2.

Figure D2: Histogram for observations from 2004 to 2008 at 100 kWp with estimated model and counterfactual

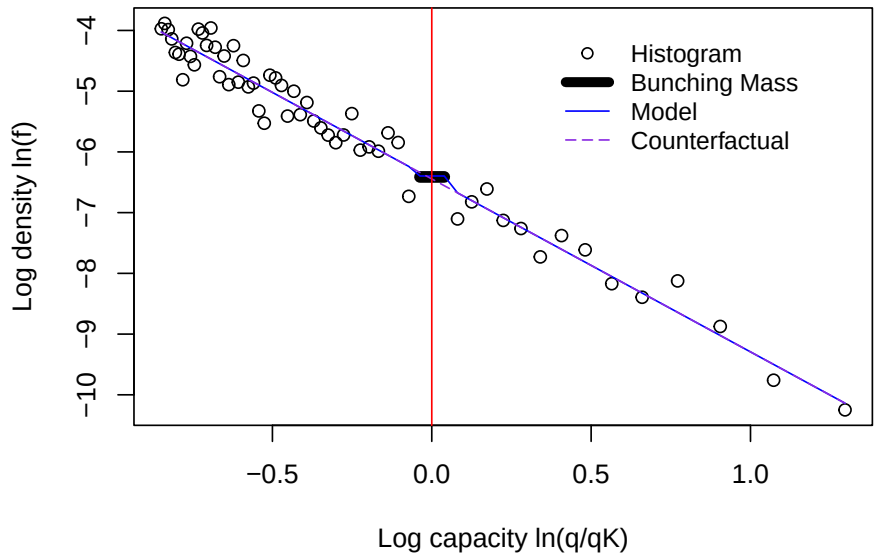


Note: The x-axis shows the normalised logarithm of capacity. The y-axis shows the logarithm of the density. The red line marks the kink point. The estimation minimises the distance between the data in black and the model in blue.

D.3 Robustness check 100 kWp

I run the robustness check for the estimation at 100 kWp on the data around 250 kWp pooling observations from 2004 to 2008. For a discussion of this choice, see Section C.4. Both estimates are insignificant.

Figure D3: Histogram at 250 kWp with estimated model



Note: The figure shows the robustness check. The x-axis shows the normalised logarithm of capacity. The y-axis shows the logarithm of the density. The red line marks the kink point. The model in blue is equal to the counterfactual in purple. The estimates are not significant.

Table D2: Results for the untreated data at 250 kWp

Capacity	$\hat{\epsilon}$ (SD)	$\hat{\kappa}$ (SD)
250 kWp	0.00 (1.04)	0.00 (0.18)

Note: The table shows the results of the robustness check. The standard errors are in brackets. The estimates are not significant.

D.4 Ignoring the participation margin or the intensive margin in the estimation

Given a certain amount of bunching, ignoring the participation margin biases the estimate of the intensive margin elasticity downwards. To evaluate this bias, I estimate an intensive margin elasticity $\tilde{\epsilon}$, ignoring the participation margin, but using the rest of the parameters from the correct estimation (11):

$$\tilde{\epsilon} = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln \int_{q_L}^{q^H} f_k(q \mid \eta = 0, \hat{\gamma}, \epsilon)\, dq \right)^2, \tag{100}$$

where $\hat{\gamma}$ is the same estimate as in (11). Correspondingly, ignoring the intensive margin biases the estimate of the participation margin elasticity. To evaluate the bias, I proceed correspondingly to the above:

$$\tilde{\eta} = \arg\min_{\eta} \sum_{j} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \eta, \widehat{\gamma}, \epsilon = 0) \right)^2. \tag{101}$$

Table D3 summarises the results for the estimates at 30 kWp pooling 2004 to 2008. The result shows that ignoring participation introduces a downward bias of 20% in the estimate of the intensive margin elasticity. Ignoring the intensive margin introduces an upward bias of 5% in the estimate of the participation margin elasticity.

Table D3: Biased estimates ignoring the other margin

Parameter	Unbiased Estimate (SD)	Biased Estimate (SD)	Relative Difference in %
ϵ	4.37 (0.13)	3.49 (0.11)	-20
η	69.24 (1.71)	72.97 (1.50)	5

Note: The table shows the unbiased estimates in the second column. The third column shows the biased estimates ignoring the other margin. The fourth column shows the relative magnitude of the bias.

D.5 Rank condition and global identification

Equations (5)-(7) form a simultaneous, nonlinear system of equations. The system must not be colinear to identify the parameters of interest. Denote the right hand side of Equation (7) as a function $f(q, \epsilon, \eta, f_l)$ and the bunching mass in Equation (6) as a function $B(\epsilon, \eta, f_l)$:

$$f(q, \epsilon, \eta, f_l) = R(q\ \rho^{-\epsilon}, \epsilon)^{\eta} f_l(q\ \rho^{-\epsilon}) \rho^{-\epsilon}, \tag{102}$$

$$B(\epsilon, \eta, f_l) = \int_{q^K}^{q^K \rho^{-\epsilon}} R(q_l, \epsilon)^{\eta} f_l(q_l) dq_l. \tag{103}$$

Formally, Condition 1 states that there exists a q such that

$$\frac{\frac{\partial B(\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial B(\epsilon,\eta,f_l)}{\partial \eta}} - \frac{\frac{\partial f(q,\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial f(q,\epsilon,\eta,f_l)}{\partial \eta}} \neq 0. \tag{104}$$

Condition 1 is a rank condition that holds generically. To see that, note that the condition does not hold if

$$\frac{\frac{\partial B(\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial B(\epsilon,\eta,f_l)}{\partial \eta}} = \frac{\frac{\partial f(q,\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial f(q,\epsilon,\eta,f_l)}{\partial \eta}} \text{ for all } q. \tag{105}$$

Equation (105) implicitly defines a function $\tilde{f}_l$. Condition 1 is violated only if $f_l = \tilde{f}_l$. Since any particular function f_l has measure zero, this is a zero probability event. However, even if Condition 1 holds, identification could be weak if the two sides of the condition are almost equal. I verify Condition 1 at the estimated values and find that it holds by a large amount. Table D4 shows the result; the standard errors are in brackets:

Table D4: Rank condition evaluated at the estimated values and $q = q^K$.

Capacity	Rank Condition (SD)
30 kWp	143 (8)
100 kWp	2833 (1158)

Note: The table shows the rank condition (Condition 1) evaluated at the estimated values and $q = q^K$. The standard errors are in brackets. The condition holds by a large amount.

Moreover, there is an economic argument why Condition 1 holds. On the one hand, the bunching mass B depends strongly on ϵ, as ϵ determines the mass of adopters who potentially bunch. That is why the upper bound of the integral in Equation (6) is a function of ϵ. Additionally, B depends only very weakly on η. The dependence is through the power of R, where R is very close to one. This is because R is roughly one minus the profit loss from re-optimisation. Due to the Envelope Theorem, the profit loss is of second order and hence very small. The strong dependence on ϵ and the weak dependence on η imply that

$$\frac{\frac{\partial B(\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial B(\epsilon,\eta,f_l)}{\partial \eta}} \tag{106}$$

is large. On the other hand, f mainly depends on η. To see that, consider the elasticity $\mathcal{E}$ of the function f_k.[23] From Equation (7) it follows that:

$$\mathcal{E} f_k(q) = \mathcal{E} f_l(q\ \rho^{-\epsilon}) + \eta\ \mathcal{E} R(q\ \rho^{-\epsilon}). \tag{107}$$

In my application, f_l is close to a Pareto distribution, and therefore, $\mathcal{E} f_l$ is approximately constant. For reasonable values of ϵ, and q close to q^K, $\mathcal{E} R$ is approximately $-(1-\rho)$. Therefore, it is approximately constant as well. Denoting the right hand side of Equation (107) by $\mathcal{E} f$, it follows that $\frac{\partial \mathcal{E} f(q,\epsilon,\eta,f_l)}{\partial \epsilon}$ is close to zero. These properties of B and $\mathcal{E} f$ make it very likely that Condition 1 holds.

I find no evidence for multiple solutions of the estimation, which is strong evidence for global identification.

[23] The definition of the elasticity of a function $g(q)$ is $\mathcal{E} g(q) = \frac{\mathrm{d}\ \ln g(q)}{\mathrm{d}\ \ln q}$.

D.6 Robustness bunching interval

After the histogram's visual inspection, I choose the two bunching intervals $[26.5, 31.5]$ and $[95, 102.5]$. The intervals are asymmetric because there is more non-sharp bunching before the kink point than after. This section reports the robustness of the estimates to changes in the bunching interval.

Table D5: Estimates at 30 kWp for various bunching intervals

Interval	$\hat{\eta}\ (SD)$	$\hat{\epsilon}\ (SD)$
[25.50, 32.250]	70.1 (1.7)	4.18 (0.14)
[25.75, 32.125]	69.4 (1.7)	4.34 (0.14)
[26.00, 32.000]	69.4 (1.7)	4.34 (0.14)
[26.25, 31.875]	69.1 (1.7)	4.38 (0.14)
[26.50, 31.750]	69.2 (1.7)	4.37 (0.13)
[26.75, 31.625]	69.9 (1.7)	4.23 (0.12)
[27.00, 31.500]	69.2 (1.7)	4.34 (0.12)
[27.25, 31.375]	70.8 (1.6)	4.00 (0.10)
[27.50, 31.250]	71.3 (1.6)	3.89 (0.10)

Note: The table shows estimates of the two elasticities for different bunching intervals. The estimates are robust to changes in the bunching interval.

The estimates at 30 kWp are robust to changes in the bunching interval.

Table D6: Estimates at 100 kWp for various bunching intervals

Interval	$\hat{\eta}\ (SD)$	$\hat{\epsilon}\ (SD)$
[92.00, 104.00]	0.0 (1.3)	5.12 (0.99)
[93.50, 103.25]	0.0 (1.7)	5.05 (0.89)
[95.00, 102.50]	0.0 (2.4)	4.63 (0.84)
[96.50, 101.75]	0.0 (2.8)	4.51 (0.80)
[98.00, 101.00]	0.0 (2.5)	4.67 (0.68)

Note: The table shows estimates of the two elasticities for different bunching intervals. The estimates are robust to changes in the bunching interval.

The estimates at 100 kWp are robust to changes in the bunching interval.

D.7 Robustness check out of sample data

At both kink points, the optimal order of the series is $P = 1$. To check if the untreated data can be used to estimate the bias, I test if the differences in the second-order parameters $\gamma_2^{is} - \gamma_2^{os}$ of

the two series expansions are not significantly different from zero. The indices is and os denote the parameter of the in-sample and out-of-sample data, respectively. I restrict the elasticities to their value in Table 2 and use a series of order $P = 2$ to estimate $\gamma_2^{is} - \gamma_2^{os}$. Table D7 shows the estimates.

Table D7: Estimates of $\gamma_2^{is} - \gamma_2^{os}$ at the two kink points with standard errors in brackets.

Capacity	$\gamma_2^{is} - \gamma_2^{os}$ (SD)
30 kWp	0.13 (0.11)
100 kWp	-0.03 (0.15)

Note: The table shows the difference $\gamma_2^t - \gamma_2^{nt}$ at the two kink points with standard errors in brackets. It is not significantly different from zero.

The difference is not significantly different from zero. It suggests the distributions fulfil the requirement.

D.8 The choice of the series expansion

This section discusses the choice of the series expansion of $f_l(.)$. By Assumption 3

$$g_f(f_l(q)) = \sum_{p=0}^{\infty} \gamma_p \frac{1}{p!} \left(g_q(q) - g_q(q^K)\right)^p, \tag{108}$$

where $g_f(.)$ is the transformation of $f_l(.)$ and $g_q(.)$ is the transformation of q. It is efficient to use transformations $g_f(.)$ and $g_q(.)$ such that the Series (108) converges fast. I use a logarithmic transformation on $f_l(.)$ and q:

$$\ln f_l(q) = \sum_{p=0}^{\infty} \gamma_p \frac{1}{p!} \left(\ln \frac{q}{q^K}\right)^p. \tag{109}$$

The logarithmic transformation of the measure and the argument is natural because both variables are defined over a positive domain.[24] To use the logarithm guarantees that the measure is non-negative and restricted to the domain of q. Additionally, the logarithmic series expansion in Equation (109) contains the uniform distribution, the Pareto distribution, and the log-normal distribution as special cases. As already mentioned in Section 4, these are common distributions for variables on a positive domain and Figure 4 suggests that the observed distributions

[24]The empirical evidence suggests the measure of adoptions is strictly positive on the domain.

are very close to a Pareto distribution.[25]

This section checks the fit of the log-log transformation against a specification without transformation and a logarithmic transformation on $f_l(.)$ only. To this end, it uses pre-treatment and treated data, excluding observations around the kink. Formally, it regresses

$$g_f(\widehat{f(q_j)}) = \sum_{p=0}^{P} \gamma_p \frac{1}{p!} \left(g_q(q_j) - g_q(q^K)\right)^p + u_j, \tag{110}$$

where u_j is the error term. The section runs the regression for $P = 1$ and $P = 2$, which gives a low-order approximation of the Series (108). It considers three combinations of transformations:

1. id-id transformation: $g_f(.)$ and $g_q(.)$ are equal to the identity function (no transformation).
2. log-id transformation: $g_f(.)$ is equal to the natural logarithm and $g_g(.)$ is equal to the identity function.
3. log-log transformation $g_f(.)$ and $g_q(.)$ are equal to the natural logarithm.

It runs the estimations on the pre-treatment data containing observations from 2000 to 2003. It uses data in the capacity range 10.6 to 85 because it corresponds to the optimal bandwidth selected in Section C.4. For the treated data, it excludes capacities close to the kink points. Consistent with the optimal bandwidths selected in Section C.4, it runs the regression on the intervals $[10.6, 26.5]$, $[35, 95]$, and $[105, 400]$. It uses R^2 as a measure of fit. Tables D8-D11 summarise the results.

[25] f_l could also be developed using a different transformation. Typically, the bunching literature does not use a transformation and directly assumes a power series. This approach has the disadvantage that many common distributions, such as the exponential distribution, the normal distribution, the Pareto distribution, or the log-normal distribution, are not special cases of the specification. Polynomial densities, which would be special cases of that expansion, are very uncommon. Additional restrictions must be implemented to ensure that the expansion fulfils a measure's standard properties, such as non-negativity and integrability. Alternatively, one can use a logarithmic transformation on the measure but not the argument (log-density estimation; see Stone, 1990). This transformation is a natural approach when the argument's domain is the real line, which is not true in my application. It contains the exponential distribution and the normal distribution as special cases.

Table D8: R-squared for untreated data in the years 2000-2003 and capacity range $[10.6, 85]$

Order of the series P	Transformation	R-squared
P=1	id-id	0.48
P=1	log-id	0.79
P=1	log-log	0.85
P=2	id-id	0.66
P=2	log-id	0.85
P=2	log-log	0.85

Note: Number of bins: N=54. Number of observations: n=3447.

Table D9: R-squared for treated data in the years 2004 to 2008 and capacity range $[10.6, 26.5]$

Order of the series P	Transformation	R-squared
P=1	id-id	0.3
P=1	log-id	0.31
P=1	log-log	0.33
P=2	id-id	0.37
P=2	log-id	0.35
P=2	log-log	0.36

Note: Number of bins: N=124. Number of observations: n=48452.

Table D10: R-squared for treated data in the years 2004 to 2008 and capacity range $[35, 85]$

Order of the series P	Transformation	R-squared
P=1	id-id	0.57
P=1	log-id	0.74
P=1	log-log	0.77
P=2	id-id	0.7
P=2	log-id	0.77
P=2	log-log	0.77

Note: Number of bins: N=182. Number of observations: n=7106.

Table D11: R-squared for treated data in the years 2004 to 2008 and capacity range $[105, 400]$

Order of the series P	Transformation	R-squared
P=1	id-id	0.21
P=1	log-id	0.22
P=1	log-log	0.23
P=2	id-id	0.23
P=2	log-id	0.23
P=2	log-log	0.23

Note: Number of bins: N=54. Number of observations: n=1073.

The log-log transformation performs particularly well. It outperforms the other transformations in most specifications and over most data ranges. Moreover, contrary to the other transformations, the log-log transformation performs almost as well with $P = 1$ as with $P = 2$. It is evidence that the series expansion converges fast in the log-log transformation.

E Details policy evaluation

E.1 Extension Assumption 5

Assumption 5 implies that q_l and c_f are independent. However, the assumption can be easily extended to allow for a correlation between q_l and c_f. Assume c_f follows a truncated normal distribution: $c_f \sim N(\mu_f, \sigma_f, \underline{c_f}(q_l), \overline{c_f}(q_l))$, where $\underline{c_f}(q_l)$, $\overline{c_f}(q_l)$ are the truncation bounds which can vary with q_l. Denote by $F_f(.)$ the CDF of the normal distribution. It follows that the CDF of c_f is $\frac{F(c_f)}{F(\overline{c_f}(q_l))-F(\underline{c_f}(q_l))}$; hence, c_f and q_l can be correlated. Assume that the bounds $\underline{c_f}(q_l)$, $\overline{c_f}(q_l)$ are large enough such that the variable profit of an agent implied by any of the counterfactual exercises lies within the bounds. As a consequence, all results in Section 6 remain unchanged.

E.2 The calibration and estimation of the type distributions

The observed subsidy $S_k(.)$ has two kink points: $q_1^K = 30$ and $q_2^K = 100$. The relative slope change at the kink points is $\rho_1 = 0.95$ and $\rho_2 = 0.99$.

$$S_k(q) = q, \qquad \text{for } q \leq q_1^K; \tag{111}$$

$$S_k(q) = q_1^K(1-\rho_1) + q\rho_1, \qquad \text{for } q \in (q_1^K, q_2^K]; \tag{112}$$

$$S_k(q) = q_1^K(1-\rho_1) + q_2^K\rho_1(1-\rho_2) + q\rho_1\rho_2, \qquad \text{for } q > q_2^K. \tag{113}$$

Using Assumption 4 , the choice q as a function of type q_l is:

$$q(q_l) = q_l, \qquad \text{for } q_l \in [q_l^{min}, q_1^K]; \tag{114}$$
$$q(q_l) = q_1^K, \qquad \text{for } q_l \in [q_1^K, q_1^K \rho_1^{-\epsilon}]; \tag{115}$$
$$q(q_l) = q_l \rho_1^{\epsilon}, \qquad \text{for } q_l \in [q_1^K \rho_1^{-\epsilon}, q_2^K \rho_1^{-\epsilon}]; \tag{116}$$
$$q(q_l) = q_2^K, \qquad \text{for } q_l \in [q_2^K \rho_1^{-\epsilon}, q_2^K (\rho_1 \rho_2)^{-\epsilon}]; \tag{117}$$
$$q(q_l) = q_l (\rho_1 \rho_2)^{\epsilon}, \qquad \text{for } q_l \in [q_2^K (\rho_1 \rho_2)^{-\epsilon}, q_l^{max}], \tag{118}$$

where q_l^{min} and q_l^{max} denote the highest and lowest type respectively.

The semi-elasticity of participation at the observed capacity q is

$$\frac{f_f}{F_f} \left(S_k(q) - c_v(q, q_l(q)) | \mu_f, \sigma_f \right) = \kappa(q), \tag{119}$$

where $q_l(q)$ denotes the inverse of $q(q_l)$ and $c_v(q, q_l)$ denotes the variable part of the cost function in Assumption 4.

Using Assumptions 4, 5, the results in Table 2, and inverting this equation at the two kink points gives (μ_f, σ_f) in Table E1.[26] Figure 16 shows the calibrated density of fixed costs and the implied semi-elasticities of participation at counterfactual capacities q_l. Note that the two red lines illustrate the counterfactual capacity with corresponding variable profit as under the observed subsidy, i.e., q_l solves $q_l - c_v(q_l, q_l) = S(q^K) - c_v(q^K, q_l(q^K))$. Because at the first kink point $q = q_l$, the first red line is at $q_1^K = 30$, while the second red line is slightly above $q_2^K = 100$.

As suggested by the empirical evidence, assume the distribution $f_l(.)$ of the variable cost type q_l is log-normal in its lower part and Pareto in its upper part:

$$f_l(q_l) = \exp\left(\gamma_{0l} + \gamma_{1l} \ln(q_l) + \gamma_{2l} \ln(q_l)^2\right) \qquad \text{for } q_l \in [q_l^{min}, q_{lb}] \tag{120}$$
$$f_l(q_l) = \exp\left(\gamma_{0u} + \gamma_{1u} \ln(q_l)\right) \qquad \text{for } q_l \in (q_{lb}, q_l^{max}]. \tag{121}$$

The parameters are such that $f_l(.)$ has a continuous first derivative. I use $[q_l^{min}, q_l^{max}]$ such that $q \in [0.5, 4000]$, which covers 99.9991% of observed aggregate capacity. The observed capacity distribution $f_k(.)$ is

$$f_k(q) = f_l(q_l(q)) \frac{F_f(S_k(q) - c_v(q, q_l(q)))}{F_f(q_l(q) - c_v(q_l(q), q_l(q)))} \frac{dq_l}{dq}. \tag{122}$$

I estimate the parameters γ_{1l} and γ_{2l} using the observed capacity range $[0.5, 21]$ and γ_{1u} using the observed range $[105, 400]$. I follow the same estimation procedure as in Section 4. I calibrate γ_{0l} so that total capacity equals the observed $Q^T = 2.7GWp$. The parameters γ_{0u} and q_{lb}

[26] The unrounded point estimate of the participation semi-elasticity at 100 kWp is 2.1×10^{-7}.

are determined by the smoothness of $f_l(.)$. Table E1 summarises the parameter values:

Table E1: Parameters distribution of fixed cost and variable cost types

Parameter	Value
μ_f	10.88
σ_f	1.57
γ_{0l}	7.70
γ_{1l}	3.12
γ_{2l}	-1.01
γ_{0u}	17.15
γ_{1u}	-3.07
q_{lb}	21.24

Note: The table shows the parameters of the distribution of fixed costs and of the type distribution $f_l(.)$.

E.3 The general welfare function and the optimal subsidy

Assume the utility of an adopter is equal to

$$u\left(\left(S(q) - c_v(q, q_l) - c_f\right) \mathbb{1}\left(S(q) - c_v(q, q_l) \geq c_f\right) + y - T(y) - c_l(y, a)\right), \tag{123}$$

where the utility function $u(.)$ is increasing and concave, $S(.)$ is the subsidy function, $c_v(q, q_l)$ is the variable cost of type q_l to adopt capacity q, and c_f is the fixed cost. The symbol $\mathbb{1}(.)$ denotes the indicator function. Agents only adopt if they make positive profit. The first part of the expression within the utility function is the net income from participating in the subsidy scheme. The variable y denotes other income, such as labour income. The function $T(.)$ is an income tax, and $c_l(y, a)$ is the effort-cost of producing income y for an agent with ability a. The second part of the expression within the utility function is the net labour income. The type θ of an adopter is three dimensional: $\theta = (q_l, c_f, a)$ with density $f_\theta(\theta)$. Note that q is a function of adopter type q_l. I will not explicitly denote this dependence to avoid an overloaded notation. The adopter type q_l contains all characteristics determining the intensive margin decision. In particular, it is determined by the characteristics of the adopter's roof and the adopter's preference for using her roof. Income y is a function of ability a. Again, I will not explicitly denote this dependence. For simplicity, I use a quasi-linear utility function. It rules out income effects and complementaries between income and solar adoption.

Consider the following objective function of the government:

$$\max_{S(.)} \int G\big(\left(S(q) - c_v(q, q_l) - c_f\right) \mathbb{1}\left(S(q) - c_v(q, q_l) \geq c_f\right) + \\ + y - T(y) - c_l(y, a)\big) f_\theta(\theta) d\theta + V(Q), \tag{124}$$

such that

$$\int q \; \mathbb{1}\left(S(q) - c_v(q, q_l) \geq c_f\right) \; f_\theta(\theta) d\theta = Q, \tag{125}$$

and

$$\int T(y) - S(q) \mathbb{1}\left(S(q) - c_v(q, q_l) \geq c_f\right) f_\theta(\theta) d\theta - R = 0. \tag{126}$$

The variable Q denotes the aggregate capacity; $V(.)$ is the government's value of aggregate capacity. The function $G(.)$ weights agents' utilities and represents the redistributive preferences of the government. It is increasing and concave. Its argument is the same as the argument of the agents' utility function. In the special case where $G(.) = u(.)$, the government is utilitarian. Equation (126) is the government's budget constraint. The variable R denotes other government spending.

Objective (124) assumes that the government sets the subsidy $S(q)$ independently of income y. This assumption is not without loss of generality. Generally, a subsidy $S(q, y)$ that depends on adopted capacity and income may achieve higher social welfare than a subsidy that only depends on capacity. However, I follow this approach for three reasons. First, I do not observe the income of adopters. Joint information about adoption decisions q and income y is necessary to solve for the optimal joint subsidy $S(q, y)$. Due to this data limitation, I will focus on the optimal separable problem, where subsidy payments are only a function of capacity q. Second, the observed subsidy is independent of income. Arguably, a subsidy that depends on income is complicated to implement. Therefore, the government did choose a subsidy that only depends on capacity q. Third, Problem (124) is a multidimensional screening problem. The type parameter determining the choice of capacity q and income y is two-dimensional. Theoretically, these problems are challenging to solve because local incentive-compatibility constraints are generally insufficient to determine the optimal schedule (see Rochet and Chone, 1998 for a detailed discussion). Treatment of the multidimensional screening problem is an interesting direction for future research beyond this paper's scope. I do not make any assumption on the income tax $T(y)$ except for the requirement that the government's budget is balanced. In particular, the income tax may be optimal.

E.3.1 Derivation of the optimality condition

I use a mechanism design approach to solve for the optimal subsidy. Denote by $q(q_l)$ the capacity q produced by type q_l. Define the variable profit $\pi_v(q_l)$ of type q_l as

$$\pi_v(q_l) = S(q(q_l)) - c_v(q(q_l), q_l). \tag{127}$$

The government chooses functions $q(.)$ and $\pi_v(.)$ instead of choosing $S(.)$ directly. The interpretation is as follows. Imagine the government asks an agent to reveal her type q_l. The agent reports the type; the government asks the agent to produce $q(q_l)$ and pays variable profit $\pi_v(q_l)$ as compensation. An incentive-compatible mechanism is two functions $q(.), \pi_v(.)$, under which each an agent is incentivised to report her type truthfully. Therefore, the government's objective is to find two such functions which give the highest payoff. Using standard mechanism design, it follows that a mechanism is incentive compatible if and only if

$$\pi_v'(q_l) = -\frac{\partial c_v(q(q_l), q_l)}{\partial q_l}, \tag{128}$$

and $q(.)$ non-decreasing. As standard in the literature, I neglect the monotonicity constraint on $q(.)$. It can be verified ex-post. Equation (128) defines a function $q(\pi_v')$. Therefore, the government's problem reduces to choosing a function $\pi_v(.)$:

$$\max_{\pi_v(.)} \int G\left(\left(\pi_v - c_f\right) \mathbb{1}\left(\pi_v \geq c_f\right) + y - T(y) - c_l(y, a)\right) f_\theta(\theta) d\theta + V(Q), \tag{129}$$

s.t.

$$\int q \; \mathbb{1}\left(\pi_v \geq c_f\right) \; f_\theta(\theta) d\theta = Q, \tag{130}$$

and

$$\int T(y) - (\pi_v + c_v(q, q_l)) \mathbb{1}\left(\pi_v \geq c_f\right) f_\theta(\theta) d\theta - R = 0. \tag{131}$$

For better readability, I suppress the arguments of the functions $\pi_v(.)$, $\pi_v'(.)$, and $q(.)$. I solve Problem (129) using calculus of variation.

E.3.2 The general optimality condition

It follows that in the optimum

$$\int_{q_l}^{q_l^{max}} \Bigg[\int_{-\infty}^{\pi_v} \left(\int_0^{\infty} \frac{G'(\pi_v - c_f + y - T(y) - c_l(y,a))}{\lambda} f_a(a|c_f, \tilde{q}_l) da - 1 \right) f_f(c_f|\tilde{q}_l)\, dc_f + \\ + \left(\frac{V'(Q)}{\lambda} q - \pi_v - c_v(q(\pi_v'), q_l) \right) f_f(\pi_v|q_l) \Bigg] f_l(\tilde{q}_l) d\tilde{q}_l + \\ + \left(\frac{V'(Q)}{\lambda} - \frac{\partial c_v(q, q_l)}{\partial q} \right) \frac{dq}{d\pi_v'} F_f(\pi_v|q_l) f(q_l) = 0 \quad (132)$$

and

$$\left(\frac{V'(Q)}{\lambda} - \frac{\partial c_v(q, q_l)}{\partial q} \right) \frac{dq}{d\pi_v'} F_f(\pi_v|q_l) f_l(q_l) = 0 \text{ for } q_l = q_l^{min} \text{ and } q_l = q_l^{max} . \quad (133)$$

The variable λ is the marginal cost of public funds. Equation (132) is a second-order differential equation in the function $\pi_v(.)$ with two boundary conditions (133). The optimal rent $\pi_v(.)$ is the solution to this system.

E.3.3 The Pigouvian subsidy (i.e., the Samuelson rule)

The Pigouvian subsidy is a linear subsidy where the marginal subsidy rate is equal to the marginal social benefit of the public good, i.e., $S'(q) = \frac{V'(Q)}{\lambda}$ for all q. This solution is also known as the Samuelson rule (Samuelson, 1954). To implement the Pigouvian subsidy, knowing how adopters react to the subsidy is not necessary. It suffices to know $V'(Q)$. However, the Pigouvian subsidy is only optimal if the government is indifferent in distributing rents to adopters. Even if an optimal income tax is available, Kaplow (1996) and Kaplow (2012) show that the Pigouvian subsidy is optimal only if preferences are separable, and the only relevant heterogeneity is earnings ability. Intuitively, in this case, the income tax is sufficient to redistribute optimally, and the choice of the public good is, therefore, not distorted. Kaplow (2008) shows that this result breaks down when agents' heterogeneity is more than one-dimensional. Importantly, in my application, the heterogeneity determining the capacity choices of adopters and income from other sources is two-dimensional. Therefore, the Pigouvian subsidy is not optimal even if income is taxed optimally. Intuitively, if the heterogeneity determining income and capacity choices correlate positively, agents in the upper part of the income distribution profit more from the subsidy programme. Because they have a low marginal social welfare weight, limiting their rents through a nonlinear subsidy is optimal.

To see this point formally, note that the Pigouvian subsidy is optimal only if

$$\int_{-\infty}^{\pi_v} \left(\int_0^{\infty} \frac{G'(\pi_v - c_f + y - T(y) - c_l(y,a))}{\lambda} f_a(a|c_f, \tilde{q}_l) da - 1 \right) f_f(c_f|\tilde{q}_l)\, dc_f = 0, \quad (134)$$

for all capacity types q_l. The see this result, guess and verify the solution, using the first order

condition of adopters $\frac{\partial c_v(q,q_l)}{\partial q} = \frac{V'(Q)}{\lambda}$. However, in general, Condition (134) does not hold. The condition particularly depends on the marginal social weight of adopters $G'(.)$ relative to the marginal cost of public funds λ. If the possibility to adopt solar panels is positively correlated with ability a, and the tax $T(y)$ does not fully equalise marginal social welfare weights, then Condition (134) does not hold. The term to the left of the condition is smaller than zero in this case. Importantly, in general, even an optimal income tax does not equalise marginal social welfare weights. Consider the optimality condition for the optimal income tax. Following Saez (2001), I solve for the optimal tax using the variational approach:

$$\int_y^\infty \Bigg(1 - \int \frac{G'\left(\left(\pi_v - c_f\right)\mathbb{1}\left(\pi_v \geq c_f\right) + \tilde{y} - T(\tilde{y}) - c_l(\tilde{y}, a^{-1}(\tilde{y}))\right)}{\lambda} \times \\ \times f(c_f, q_l|\tilde{y}) dc_f dq_l \Bigg) f_y(\tilde{y}) d\tilde{y} + T'(y)\frac{dy}{dT'}\, f(y) = 0 \tag{135}$$

As long as there are behavioural responses to taxation, the government does not equalise marginal welfare weights $G'(.)$.

E.3.4 The relation to the simple objective (14)

Consider redistributive preferences $G(.)$ such that the marginal welfare weight for income above a certain level $\underline{y}$ is zero. Additionally, assume only agents with income higher than $\underline{y}$ can adopt solar panels. For instance, this could be the case since only agents with income higher than $\underline{y}$ own buildings. It follows that Condition (134) is equal to -1. Importantly, only high-income agents adopt solar panels because of the correlation of earning ability a and capacity type q_l, not because of an income effect. There are no income effects since I use quasi-linear preferences. These preferences and correlation patterns, together with the assumption that the government values aggregate capacity only up to the capacity goal Q^T, reduce Problem (124) to Problem (14). For example, it is the case if the redistributive preferences are Rawlsian, and households with the lowest incomes cannot adopt solar panels.

E.4 Counterfactual experiments

E.4.1 The optimal linear subsidy

A first experiment solves for the optimal linear subsidy. The exercise solves for a linear subsidy rate ρ_l that incentivises the adoption of the same aggregate capacity as the observed kinked subsidy.[27] By the first order condition of the agents' problem, the choice q of type q_l under subsidy ρ_l is $q(q_l, \rho_l) = q_l \rho_l^\epsilon$. Denote by $c_v(q, q_l)$ the variable part of the cost function $c_v(q, q_l) = c(q, q_l, c_f) - c_f$. Denote the variable profit of type q_l under subsidy rate ρ_l as $\pi_v(q_l, \rho_l) = \rho_l q(q_l, \rho_l) - c_v(q(q_l, \rho_l), q_l)$. Given the estimate of $f_l(q_l)$, the unconditional type distribution

[27] Note that s_l is normalised to one; ρ_l's interpretation is relative to s_l.

$f_u(q_l)$ is

$$f_u(q_l) = \frac{f_l(q_l)}{F_f(q_l - c_v(q_l, q_l))}. \tag{136}$$

It follows that ρ_l is the solution to

$$\int q(q_l, \rho_l) F_f(\pi_v(q_l, \rho_l)) f_u(q_l) dq_l = Q^T, \tag{137}$$

where Q^T is the observed aggregate capacity. I find that $\rho_l = 0.998$. The public cost of the linear policy is $Q^T \rho_l$. The policy is 0.14 per cent more expensive than the actual subsidy.

E.4.2 The optimal nonlinear subsidies

The second counterfactual experiment solves for the optimal nonlinear policy using mechanism design. The analysis follows the screening problem in Rochet and Stole (2002). Rewrite the government's objective (14) as a Lagrangian and a mechanism design problem. The government maximises

$$\max_{\psi, q(.), \pi_v(.)} \int \left(\psi q(q_l) - \pi_v(q_l) - c_v(q(q_l), q_l)\right) F_f(\pi_v(q_l)) f_u(q_l) dq_l - \psi Q^T, \tag{138}$$

such that for all q_l

$$\pi_v'(q_l) = -\frac{\partial c_v(q(q_l), q_l)}{\partial q_l} \text{ and } q(.) \text{ is not decreasing.} \tag{139}$$

(139) is the incentive-compatibility constraint. The variable ψ denotes the Lagrange multiplier. Substitute the subsidy paid to type q_l using the definition: $S(q(q_l)) = \pi_v(q_l) + c_v(q(q_l), q_l)$. Problem (138) is equivalent to Problem (14). The government chooses functions $q(q_l)$ and $\pi_v(q_l)$ instead of a subsidy $S(q)$. The interpretation is as follows. Imagine the government asks an agent to reveal her type q_l. The agent reports the type; the government asks the agent to produce $q(q_l)$ and pays variable profit $\pi_v(q_l)$ as compensation. An incentive-compatible mechanism is two functions $q(.), \pi_v(.)$, under which each agent is incentivised to report her type truthfully. Therefore, the government's objective is to find two such functions which maximise its objective. The incentive-compatibility constraint (139) follows from the standard revealed preference argument in mechanism design. As standard in the literature, neglect the monotonicity constraint on $q(.)$ and verify it ex-post. Define as

$$L(\pi_v, \pi_v', q_l) = \left(q_l \psi (\pi_v'(1+\epsilon))^{\frac{\epsilon}{1+\epsilon}} - \pi_v - q_l \epsilon \pi_v'\right) F_f(\pi_v) f_u(q_l), \tag{140}$$

which is the integrand of Problem (138). Use Equation (139) to substitute for the function $q(.)$. The problem simplifies to finding an optimal function $\pi_v(.)$. Suppress the arguments of the functions $\pi_v(.), \pi_v'(.)$ for better readability. By calculus of variation, it follows that the optimal

function π_v satisfies

$$\frac{\partial L(\pi_v, \pi_v', q_l)}{\partial \pi_v} = \frac{d}{dq_l} \frac{\partial L(\pi_v, \pi_v', q_l)}{\partial \pi_v'} \quad \text{for all } q_l, \tag{141}$$

and

$$\frac{\partial L(\pi_v, \pi_v', q_l)}{\partial \pi_v'} = 0 \text{ for } q_l = q_l^{min} \text{ and } q_l = q_l^{max}. \tag{142}$$

The values q_l^{min} and q_l^{max} denote the boundaries of the type distribution $f_l(.)$. For each ψ, the Equations (141) and (142) define a nonlinear second-order differential equation with boundary values. Fix ψ, solve the differential equation numerically, and evaluate the capacity constraint $Q = Q^T$. Iterate over ψ until the constraint holds. Using the first order condition and the solution $\pi_v'(q_l)$, solve for the optimal nonlinear marginal subsidy $S'(q(q_l))$ in Figure 18. Using the definition of the variable profit, it follows that the total public costs are

$$\int \left(\pi_v(q_l) + c_v(q(q_l), q_l)\right) F_f(\pi_v(q_l)) f_u(q_l) dq_l. \tag{143}$$

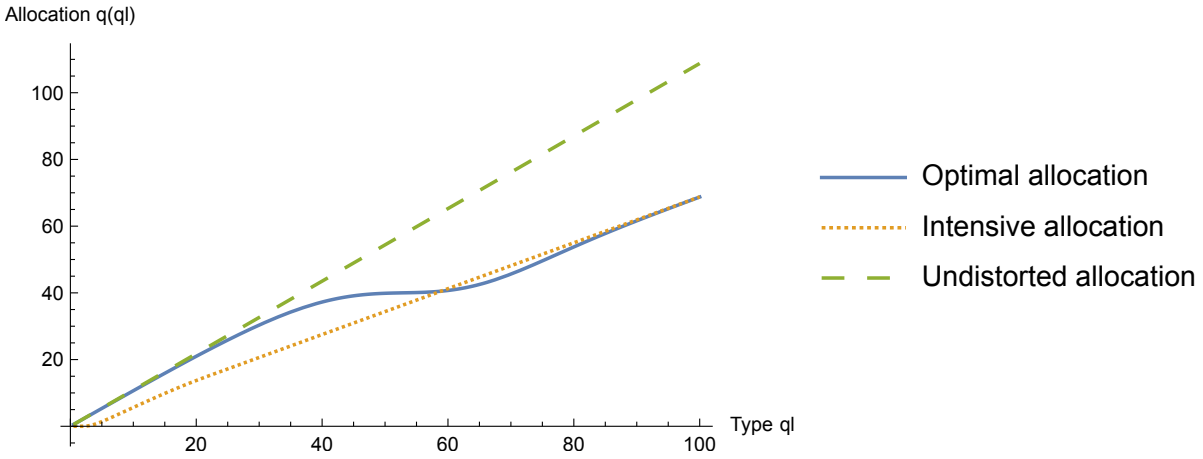
The optimal subsidy is 0.45 per cent less costly than the linear benchmark.

The third counterfactual experiment assumes there is no participation margin response. It solves the problem using the same methodology as above, but assuming $F_f(\pi_v) = 1$ for all π_v. The fourth counterfactual experiment implements the optimal intensive schedule from the third experiment but lets agents react at both margins. Both experiments keep aggregate capacity constant by adjusting the Lagrange multiplier.

E.4.3 Comparison to Rochet and Stole (2002)

Figure E1 compares the optimal allocation to the bounds derived in Proposition 4 in Rochet and Stole (2002). They show that when the type distribution $f_l(.)$ is uniform, the optimal allocation $q(q_l)$ is bounded from above by the undistorted allocation $q^{fb}(q_l) = \psi^\epsilon q_l$ and from below by the optimal intensive allocation $q^{it}(q_l)$ derived in the third counterfactual experiment. Note that Rochet and Stole (2002) also assume the distribution of fixed costs is log-concave, which is the case in my application. Contrary to the third counterfactual experiment and consistent with the result in Rochet and Stole (2002), I do not adjust the Lagrange multiplier to fulfil the capacity constraint, but I use the multiplier from the second counterfactual experiment. The figure shows that there exist regions where the optimal allocation lies outside the bounds. It shows that the result in Rochet and Stole (2002) is not robust to a more general form of the type distribution $f_l(.)$. E.g., the type distribution in my application is log-normal at the bottom and Pareto at the top.

Figure E1: The optimal allocation $q(q_l)$ and the bounds derived by Rochet and Stole (2002)



Note: The figure shows the optimal allocation and the bounds derived by Rochet and Stole (2002). In some regions, the optimal allocation lies outside the bounds.

E.4.4 Comparison to Germeshausen (2018)

Germeshausen (2018) uses a difference-in-difference approach to estimate the treatment effect of introducing a new kink of 5% at 10 kWp in Germany in 2012. Methodologically, Germeshausen (2018) follows Best and Kleven (2017) and controls for self-selection due to bunching in the dif-in-dif. He estimates the treatment effect of introducing the new kink; he finds it reduces capacity installed in the interval of 10-20 kWp by 43%. I cannot use his methodology because my data suggest that the parallel trend assumption necessary for a difference-in-difference approach is not satisfied.

Germeshausen does not estimate intensive and participation margin elasticities. To compare the results in the two studies, I use my estimates to calculate the implied treatment effect of introducing a new kink of 5% at 10 kWp in my data, i.e., I calculate:

$$\frac{\int_{q_1^K}^{q^{max}\rho_1^{-\epsilon}} q(q_l) F_f\left(S_k(q(q_l)) - c_v(q(q_l), q_l)\right) f_u(q_l) dq_l}{\int_{q_1^K}^{q^{max}} q_l f_l(q_l) dq_l}, \tag{144}$$

where $q(.)$ and $S_k(.)$ are defined in Section E.2, $f_u(.)$ is defined in Section E.4.1, q_1^K=10 kWp, $\rho_1 = 0.95$, $\rho_2 = 1$, and q^{max}= 20 kWp.

I find introducing the kink would reduce capacity in the range 10-20 kWp range by 40%. The similarity of the two treatment effects provides evidence for the validity of the respective identifying assumption in both studies.

E.5 Summary results policy evaluation

Table E2 and E3 summarise the results of the counterfactual exercises.

Table E2: Cost of the counterfactual scenarios compared to the optimal linear subsidy

Subsidy	Relative Cost Compared to Optimal Linear in %
Optimal Linear	0
Observed	-0.14
Optimal Nonlinear	-0.45
Only Intensive Margin Hypothetical	-8.71
Only Intensive Margin Real	+3.08

Note: The table shows the cost of the counterfactual scenarios relative to the cost of the optimal linear subsidy in per cent.

Table E3: Subsidy payment to the lowest type (0.5 kWp) under the counterfactual scenarios

Subsidy	Subsidy Payment to Lowest Type (0.5 kWp)
Optimal Linear	0.499
Observed	0.5
Optimal Nonlinear	0.323
Only Intensive Margin Hypothetical	0.323
Only Intensive Margin Real	0.323

Note: The table shows the subsidy payment to the lowest type at 0.5 kWp under the counterfactual subsidies.