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

Stefan Pollinger †

January 17, 2021

Job Market Paper

Please find the latest version of the paper here.

Abstract

This paper shows how to estimate agents' intensive and participation margin responses to nonlinear incentive schemes such as taxes, subsidies, and prices. The proposed semi-nonparametric estimator allows evaluating nonlinear incentive schedules when existing kink and discontinuity methods are inapplicable due to the presence of both margins. The paper's first contribution is to show that agents' reactions to kinks or discontinuities in an incentive scheme *identify responses at both margins simultaneously.* The only observable needed for estimation is the distribution of agents' choices, making the estimator widely applicable. The paper's second contribution is to *evaluate the German subsidy for solar panels*, which is a cornerstone in the country's energy-transition efforts. I find sizable elasticities along both margins and an optimal subsidy that is close to linear.

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

*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, Jaques Crémer, Pierre Dubois, Damon Jones, Stefan Lamp, Guy Laroque, Alix de Loustal, Thierry Magnac, Mathias Reynaert, Whitney Newey, John Rust, Emmanuel Saez, Augustin Tapsoba, Elie Tamer, 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, the Congress of the Institute of International Public Finance 2019, and the European Winter Meeting of the Econometric Society 2019. All errors are mine.

†Stefan Pollinger, Ph.D. candidate at the Toulouse School of Economics; email address: stefan.pollinger@tse-fr.eu; website: www.stefanpollinger.com.

1 Introduction

The German subsidy for rooftop solar panels is a cornerstone of the country's transition towards a carbon-free economy. Although successful in incentivizing households and firms in adopting solar panels, the subsidy is very costly. The yearly payments amount to 0.6% of total government spending.[1] To curtail costs, the government uses a subsidy that is nonlinear in adopted capacity.[2] The German subsidy is piecewise linear and has several kinks. The marginal subsidy rates decrease discontinuously at the kink points, which reduces payments to very profitable large adoptions.

In this paper, I evaluate the cost-effectiveness of the German subsidy programme. Concretely, the paper focuses on two questions. First, are the kinks in the German subsidy schedule effective at reducing the programme's costs without compromising the government's total capacity goal?[3] Second, what is the most cost-efficient nonlinear subsidy schedule to achieve a specific total capacity goal? The answers to both questions depend on how adopters react to the subsidy. There are two possible margins of reaction: the participation margin, which determines how many adopters participate in the scheme, and the intensive margin, which determines the capacity choice of participants.[4] The novel estimator proposed in this paper exploits the kinks in the subsidy schedule to estimate adopters' responses at both margins. The estimates allow me to evaluate the current subsidy scheme and to solve for the most cost-efficient nonlinear subsidy. I find that the government's strategy to curtail costs is barely effective. In fact, the most cost-efficient scheme is very close to linear, a result driven by adopters' responses at the participation margin.

More generally, 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 to reliably estimate how agents react to them at the participation and intensive margins. When agents react only at the participation margin, kinks in the incentive scheme can be used to estimate their response applying the regression kink design (Card, Lee, Pei, and Weber, 2015). Correspondingly, when there is only an intensive margin re-

[1] According to the official report from the entity responsible for the payment, Übertragungsnetzbetreiber (2016), total payments in 2016 were 9 billion euros.

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

[3] The total capacity is the sum of the capacity of installations in a given period.

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

sponse, the bunching estimator (Saez, 2010) can be used. However, when both margins are present, it is necessary to estimate both responses to evaluate an incentive scheme. 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 at 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. I exploit the distinct impact of the two reaction-margins on the two observable effects in the distribution to estimate both margins simultaneously.

Using administrative data from 2004 to 2008, I estimate both response margins at each of the two kink points in the subsidy scheme. The first kink point is at a mid-range capacity of 30 kWp, while the second kink point is at a large capacity of 100 kWp.[5] The estimate of the intensive margin elasticity is 4.8 at both points. The estimate of the participation margin elasticity is much higher at the mid-range capacity than at the large capacity (65 vs. 1). This result is intuitive. Net of fixed costs, profits increase in capacity. The higher the profitability, the lower the probability that the fixed cost is high enough to prevent an adoption. Therefore, the participation elasticity decreases in capacity.

Together with global assumptions on adopters' variable and fixed costs, the estimates allow the computation and evaluation of public costs under various counterfactual subsidies.[6] I assume the government wishes to implement a given total capacity of solar panels at minimal public costs. I use the observed total capacity as the government's capacity goal. Contrary to a pure Pigouvian analysis, where the optimal marginal subsidy is equal to the marginal environmental benefit, I explicitly take into account the social cost of distributing information rents to adopters.[7]

[5]The unit kWp (kilowatt peak) is the physical unit of capacity.

[6]As suggested by the results, I assume an isoelastic variable-cost function and a log-normal and independent distribution of fixed costs.

[7]Note that evaluating the policy choice of solar subsidies *per se* is an interesting question for future research but outside the scope of this paper.

The counterfactual exercises provide four crucial insights. First, the nonlinearities introduced by the government are barely effective in reducing costs. To arrive at this conclusion, I solve for the optimal linear subsidy and use it as a benchmark. Relative to the linear benchmark, the scheme used by the government is only 0.016 per cent less costly. Second, the optimal subsidy scheme is very close to linear. Compared to the linear benchmark, the optimal scheme reduces costs by 0.055 per cent. Optimal cost reduction is by a factor 3.5 larger than under the government's scheme; however, the relative effect is still very small. The result shows that the room for cost reductions through nonlinear pricing is very limited. Third, it is the large participation margin response that drives this result. To arrive at this conclusion, I solve for the optimal nonlinear subsidy assuming there are no responses at the participation margin. I find that the optimal scheme is 7.6 per cent less costly than the benchmark. In this case, the government implements relatively low marginal subsidy rates for small capacities and high marginal rates for large capacities. As a consequence, big adopters choose large capacities because they face high marginal incentives. Simultaneously, their aggregate subsidy payment is low because they earn less profit on inframarginal units. In contrast, when there are participation margin responses, this strategy to extract profits is not effective. Low marginal subsidy rates for small capacities affect larger adopters by reducing their profit margins and triggering responses at the participation margin. Importantly, implementing this schedule, when in fact adopters react at both margins, has sizable adverse effects. Costs increase by 3.7 per cent compared to the benchmark instead of decreasing. This result gives a fourth insight: it is crucial to consider the participation margin when designing the policy.

Methodologically, the estimation of adopters' responses poses several challenges. First, price and subsidy variation over time is unsuitable for identification because they are endogenous to adoption behaviour. Second, the solar industry is fast-changing, making it necessary to reestimate adopters' behaviour over time. Third, there is considerable unobservable heterogeneity among adopters since heterogeneous costs, discount rates, local climate conditions, and the roof characteristics influence the adoption decisions. The estimator proposed in this paper addresses these challenges.

The identification relies on the smoothness of the counterfactual distribution of agents' choices, i.e., the choice-distribution under a counterfactual linear subsidy without a kink. More precisely, I assume the distribution is locally representable by a convergent power

series. This assumption avoids the need for parametric restrictions on the counterfactual to identify the two margins. Blomquist, Newey, Kumar, and Liang (2017) criticize the use of such a parametric assumption in the bunching literature because it may lead to specification bias. The smoothness assumption in this paper addresses this critique and is robust to specification bias. The estimation method is semi-nonparametric and uses sieve estimation (see Chen, 2007 for a review).[8] The nonparametric specification is data-driven. Additionally, using untreated data, I check the robustness of the specification. The estimator circumvents the need for exogenous variation, or instrumental and control variables, and only uses the easily observable distribution of agents' choices. These low informational and identifying requirements increase its potential applicability. The adaptation to notches, i.e., discontinuities in average incentives, is straightforward.

Related literature. Methodologically, this paper contributes to the bunching and the regression kink design literature. Both methods exploit kinks in an incentive scheme. The bunching estimator (see Saez, 2010; Kleven and Waseem, 2013; Kleven, 2016) estimates intensive margin responses using bunching at kink and notch points in the budget set. The excess bunching mass at the kink or notch point identifies the intensive margin. This literature does not consider participation margin responses. I find that ignoring the participation margin leads to a 20% downward bias in the intensive margin estimate in my application. Moreover, it is essential to consider both margins because both are relevant for optimal policy-design.

In the bunching literature, Gelber, Jones, Sacks, and Song (2020) estimate intensive and participation margin responses for workers close to retirement. Marx (2018) estimates both responses for charitable organizations. Both papers use additional information, coming from their data's panel dimension. The missing panel dimension in my application makes such an approach infeasible. An advantage of the simultaneous estimation in my paper is that only the distribution of adoptions is needed to estimate both margins. Another methodological contribution in the bunching literature distinct from my paper is Cox, Liu, and Morrison (2020). They extend the approach to the case when the intensive margin is two-dimensional.

The participation margin responses at kink points can be estimated using a regression

[8] Following Chen (2007), I call a model semi-nonparametric if the parametric and the nonparametric part of the model are of interest.

kink design (RKD), proposed by Nielsen, Sørensen, and Taber (2010) and generalized by Card, Lee, Pei, and Weber (2015). They show that the difference between the slopes to the left and the right of the kink point identifies the participation margin elasticity. However, if there is a positive intensive margin, this estimator suffers from endogeneity since agents select into treatment. So far, the regression kink design literature uses bunching to detect endogenous sorting. The classical RKD is applicable only when sorting is rejected.

In the closely related context of a regression discontinuity design, Gerard, Rokkanen, and Rothe (2020), Bachas and Soto (2018), and Caetano, Caetano, and Nielsen (2020) propose methods to correct for endogenous sorting. Gerard et al. (2020) use partial identification. Combining the bunching approach with a regression discontinuity design, Bachas and Soto (2018) correct for sorting when decomposing the tax-elasticity of corporate profits into a revenue and cost elasticity. Caetano et al. (2020) correct for endogenous sorting in data with a censored assignment variable. These papers do not consider participation margin responses. Their methodologies are not applicable when both margins are present. The reason is that each margin is responsible for the bias in the estimate of the other margin. As a consequence, the simultaneous estimation of both margins is necessary.

Potentially, the bunching estimator, the regression kink design, and structural models using nonlinear budget sets suffer from specification bias. Blomquist, Newey, Kumar, and Liang (2017) and Bertanha, McCallum, and Seegert (2019) point out that the bunching literature uses parametric functional form assumptions on the counterfactual to identify the intensive margin elasticity. Under a misspecified functional form, the estimates are inconsistent. Ganong and Jäger (2018) and Ando (2017) point out that the regression kink design suffers from a related issue. Being very sensitive to the specification, the RKD can badly estimate nonlinearities in the counterfactual as a slope change. I address these issues by using a nonparametric method and checking the robustness of the specification on untreated data. In a related paper, using nonlinear budget sets, Beffy, Blundell, Bozio, Laroque, and To (2019) estimate a structural model of labour supply under hours constraints. I use a local nonparametric approach, which is robust to the potential specification bias of structural models.

This paper also contributes to the literature evaluating subsidies for solar panels. To the best of my knowledge, it is the first paper to evaluate the cost-effectiveness of nonlinear

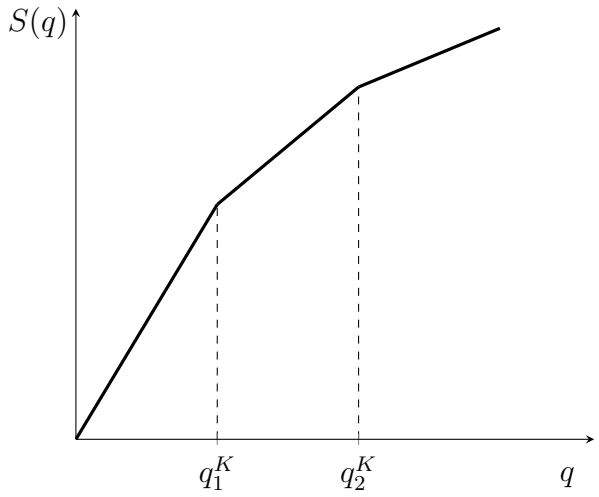
solar subsidies. I find that nonlinearities are not very effective at reducing the costs of the German programme. One strand of the solar literature focuses on the dynamics of the adoption decision (Burr, 2016; De Groote and Verboven, 2019; Feger, Pavanini, and Radulescu, 2017; Reddix, 2015; Gerarden, 2017). Methodologically, these papers use structural models. Hughes and Podolefsky (2015) use geographical discontinuities in California to study adoption behaviour. Such discontinuities are not available in Germany. Germeshausen (2018) uses a difference-in-difference approach to estimate the treatment effect of the introduction of a new kink in Germany in the year 2013. Although, possible in principle, the paper does not aim to estimate elasticities or evaluate the nonlinear scheme. Methodologically, Germeshausen (2018) follows Best and Kleven (2017). Kleven, Landais, Saez, and Schultz (2013), Slemrod, Weber, Shan, and Sachs (2012) and Besley, Meads, and Surico (2014) are similar. All these papers use a difference-in-difference approach, controlling for or using bunching. Difference-in-difference is a powerful quasi-experimental method for estimating the treatment effect after a policy reform. However, the method relies on a parallel trend assumption. If conditions for different adopters change distinctly over time, it cannot be used to reestimate adopters' behaviour over time. The solar industry is a fast-changing industry, which is confirmed by my estimates. My identification strategy does not rely on a parallel trend assumption.

Section 2 shows graphical evidence for responses at the two margins; Section 3 presents the empirical model; Section 3.3 discusses identification. Section 4 describes the German subsidy and the data in more detail. Section 5 shows the estimation method. Section 6 presents the results and discusses the robustness. Section 7 evaluates the policy, and Section 8 concludes.

2 Graphical evidence

The German subsidy for solar panels, illustrated in Figure 1, has several kink points.

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

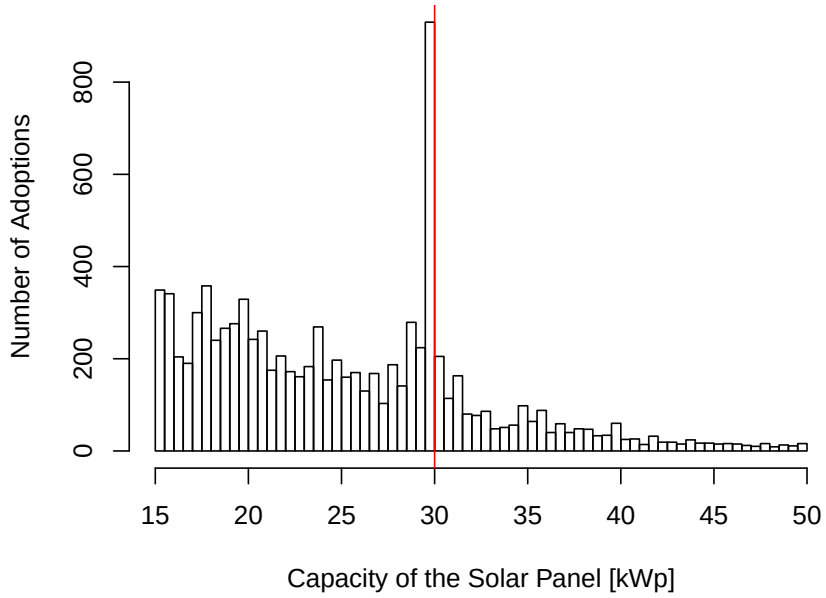


Note: The figure illustrates the kinked subsidy in Germany. The two kink points are marked by q_1^K and q_2^K.

The variable q on the x-axis is the capacity of an adopted solar panel. The subsidy payment $S(q)$ on the y-axis is the present discounted value of all subsidy payments an adopter receives when installing capacity q. The capacity of a solar panel is the amount of electricity it produces under standardized conditions. The subsidy is piecewise linear, with decreasing marginal rates. Adopters are households or firms. Over the sample period, households and firms sell the produced solar energy to the government at a subsidized rate. Therefore, you can think of them as firms producing capacity for the government. A detailed description of the policy is in Section 4.

Figure 2 shows the histogram of all solar panel adoptions in the year 2004 around the kink point at capacity 30 kWp. The unit kWp (kilowatt peak) is the physical unit of capacity. Typically, adopters only adopt once. 2004 is the first year that the government introduced kinks in the subsidy, and 30 kWp is the kink point around which there have been the most adoptions.

Figure 2: The histogram of adoptions in 2004

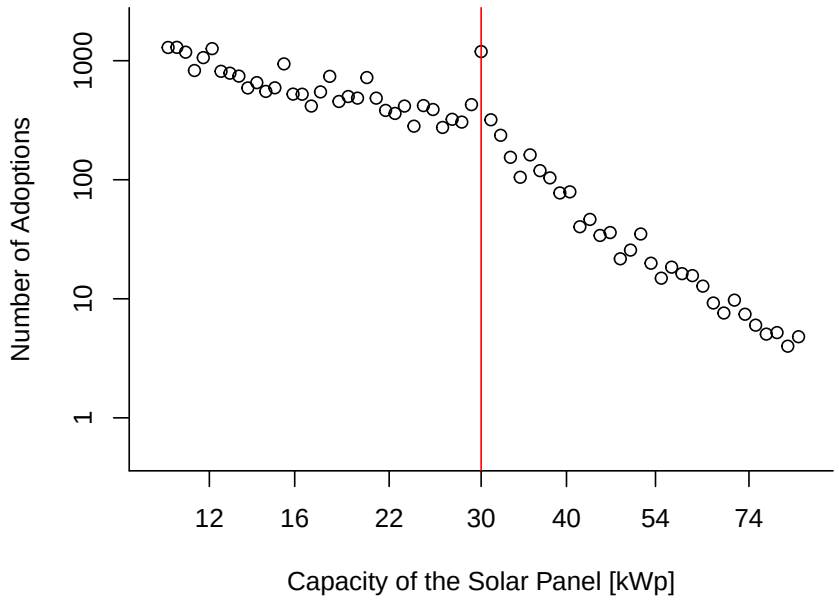


Note: The red line marks the kink point at 30 kWp. There is a visible mass point, i.e., bunching, at the kink point.

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 a large number of adoptions are bunched at the kink point. In Section 3, I show that bunching is reduced-form evidence for an intensive margin response of adopters. The intensive margin response of adopters is a response in the capacity-choice. The higher the marginal subsidy rate, the larger is the capacity a certain adopter installs.

Figure 3 shows the same histogram on a logarithmic scale. The number of observations in a bin is normalized by the bin size.

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

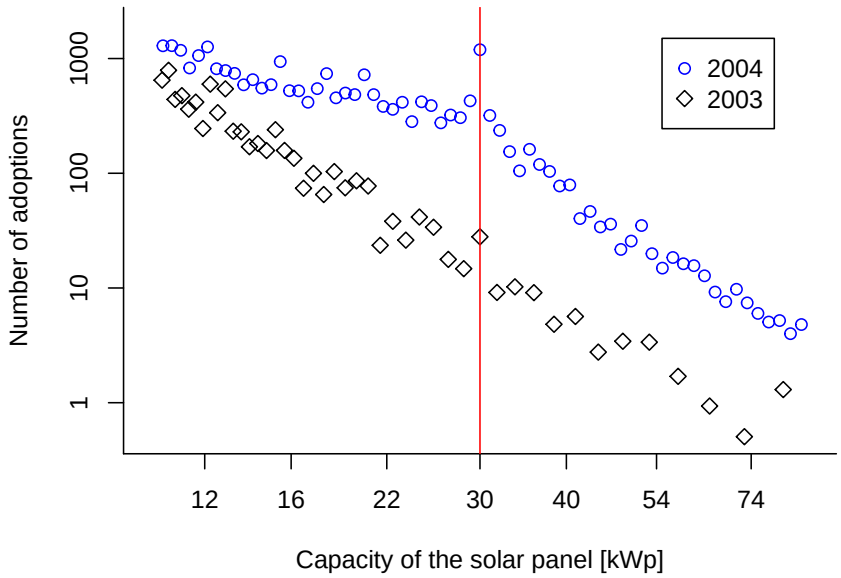


Note: The red line marks the kink point at 30 kWp. At the kink point, there is a visible mass point and slope change in the distribution. The number of observations in a bin is normalized by the bin size.

There is a visible slope change in the distribution at the kink point. In Section 3, I show that the slope change is evidence for a participation margin response. Participation depends on the total subsidy payment. The higher is the payment, the more households and firms adopt solar panels. Additionally, note that the distribution is close to linear in the logarithmic scale except for the kink point. This shape is evidence for a distribution close to a Pareto distribution.

Figure 4 compares the same plot to observations in 2003. In 2003 the subsidy is linear, and the subsidy rate is lower than in 2004.

Figure 4: Histogram of adoptions in 2003 and 2004

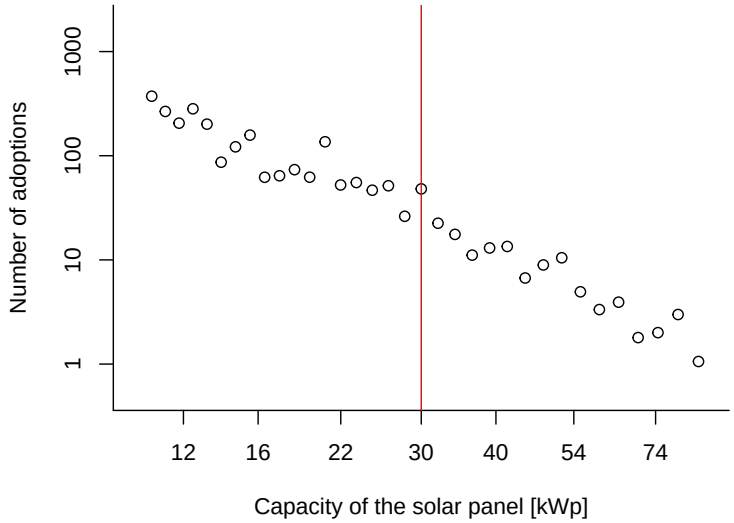


Note: The red line marks the kink point. Scales are logarithmic. Under the linear subsidy in 2003, there is no significant mass point or visible slope change in the distribution.

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, present in 2004, are indeed caused by the kink, hence evidence for responses at the two margins. Again, the distribution is close to linear in the logarithmic scale, which is more evidence for a distribution close to a Pareto distribution.

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 for the years 2000 to 2002. I pool years to have a sufficiently large sample size. The figure shows no graphical evidence for a time trend in the concavity of the distribution.

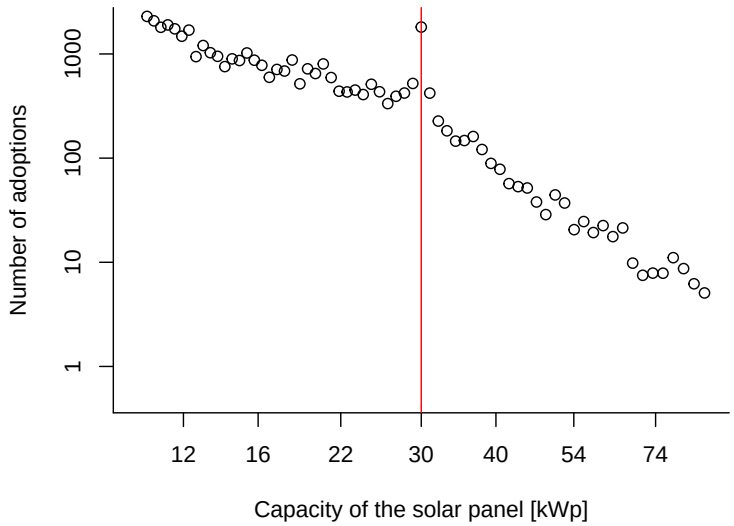
Figure 5: Histogram of adoptions in the years 2000 to 2002



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 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.

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.

3 The model

This section shows that when economic agents react to incentives at the intensive and the participation margins, a kink in the incentive scheme causes a mass point and a slope change in the distribution of agents' choices. This statement is true under a very general model of economic behaviour. Quantitatively, the magnitude of the two responses is related to the size of the mass point and the slope change. I exploit this property to identify the two margins.

There is a mass of heterogeneous adopters indexed by i. They produce capacity q 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 maximization problem:

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

and participate if and only if

$$\pi_v^i \geq c_f^i. \tag{2}$$

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 unobservable benefits. In practice, 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. Therefore, it is equivalent to a subsidy payment. See Section 4 for a detailed description of the policy. Appendix A.6 shows that Problem (1) encompasses a subsidy payment via feed-in tariffs. 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, I follow 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: warm-glow preferences for solar panels, the opportunity and aesthetic costs of using space on the roof, opportunity costs of time and money,[9] and direct benefits from consuming

[9] Opportunity costs of money are relevant if the adopter is credit constraint at the time point of adoption.

electricity produced by the solar panel.

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 $c_v^i(q)$ 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 cost of increasing capacity through more efficient panels is convex.[10] Therefore, the function $c_v^i(.)$ is convex for q large enough. Note that the optimal choice of q is always in the convex range of $c_v^i(.)$. For an example of heterogenous fixed costs, consider two firms with the same roof size. Firm one, e.g., a start-up, is very productive and has high opportunity costs of time. Firm two, e.g., a 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.

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

$$S_k(q) = s_l q, \qquad \text{for } q \leq q^K; \tag{3}$$

$$S_k(q) = s_l q^K + (q - q^K)\rho s_l, \qquad \text{for } q > q^K. \tag{4}$$

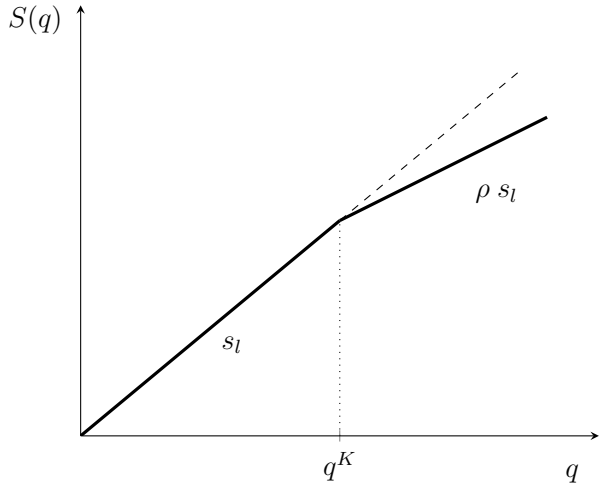
The kink point is denoted by q^K; s_l is the marginal subsidy rate below the kink point, and $\rho\ s_l$ is the marginal subsidy rate above the kink point, where $\rho < 1$ is the relative change in subsidy rates. The counterfactual linear subsidy $S_l(q)$ is:

$$S_l(q) = s_l q, \qquad \text{for all } q. \tag{5}$$

Figure 7 illustrates both subsidies.

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

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



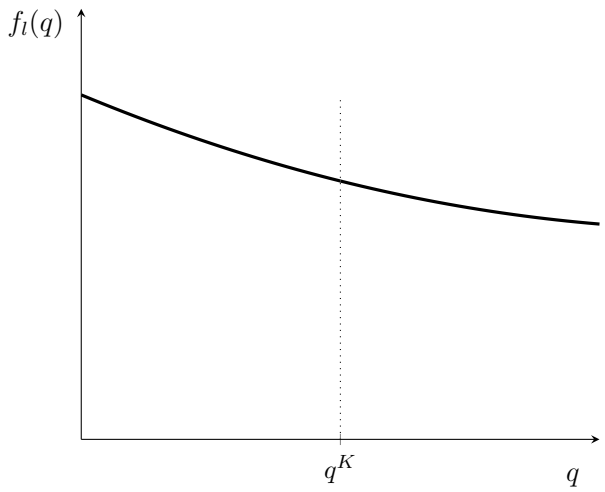
Note: The thick solid line shows the kinked subsidy S_k. The dashed line shows the counterfactual subsidy S_l. The variable s_l denotes the marginal subsidy rate below the kink point; q^K denotes the kink point, and ρ denotes the relative change in the marginal subsidy rate at the kink point.

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 any interval of capacity $[q_1, q_2]$, $\int_{q_1}^{q_2} f_x(q)dq$ is the mass of adopters in the interval under the subsidy $S_x(.)$, where x stands for l or k. Intuitively, both functions are densities but do not integrate to one.

3.1 The graphical intuition behind identification

In this section, I give graphical intuition on how the distribution of adoptions depends on the intensive and the participation margin. Suppose the counterfactual linear subsidy S_l is in place. Figure 8 illustrates a possible measure of adoptions f_l.

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

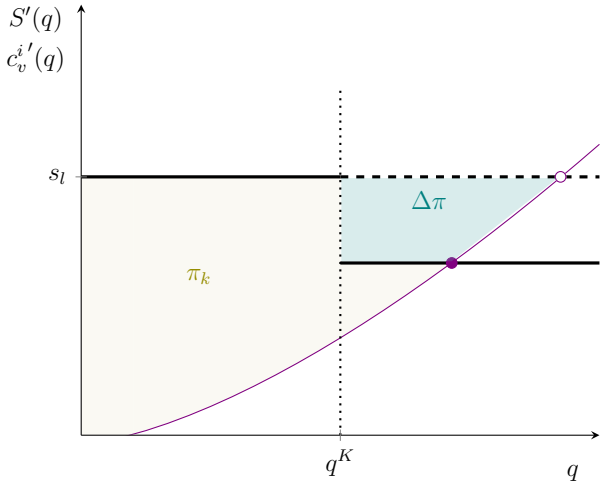


Note: The counterfactual measure $f_l(.)$ is the distribution of adoptions under the counterfactual linear subsidy $S_l(.)$. Its key property is its smoothness around the hypothetical kink point.

Without loss of generality, I depict 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. I will state the precise smoothness assumption (Assumption 3) later in this section.

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 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.

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

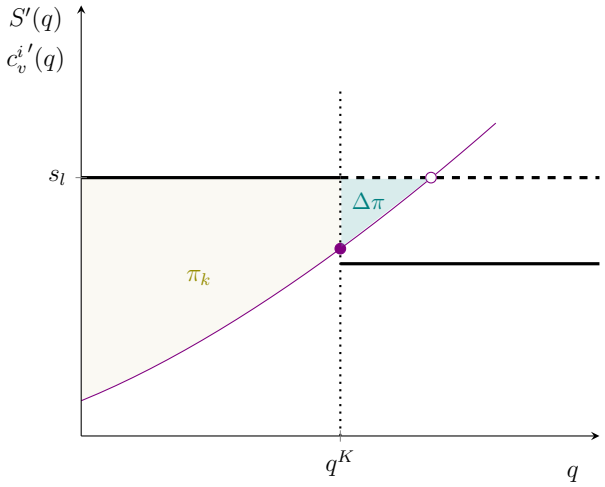


Note: The thick black line depicts the kinked marginal subsidy S'_k, and the dashed line depicts the linear marginal subsidy S'_l. The thin purple line illustrates the marginal cost curve $c_v^{i\,\prime}$ of an adopter. The capacity-choice under the kinked scheme, depicted by the full dot, is lower 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.

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 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.

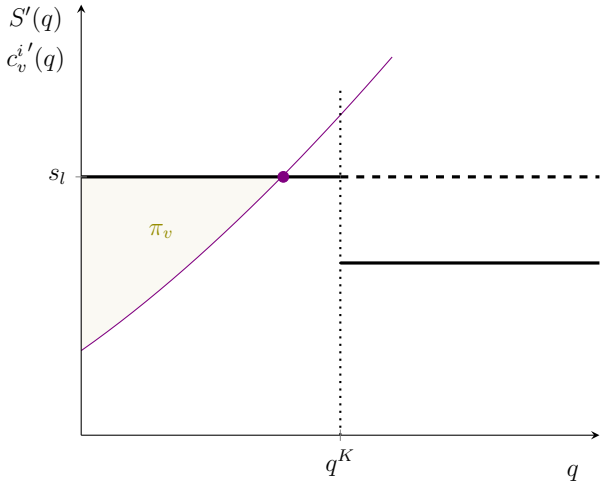
Figure 10: The effect of the kinked scheme S_k on adopters just above the kink point



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 exactly at the kink point. It 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.

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 exactly 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.

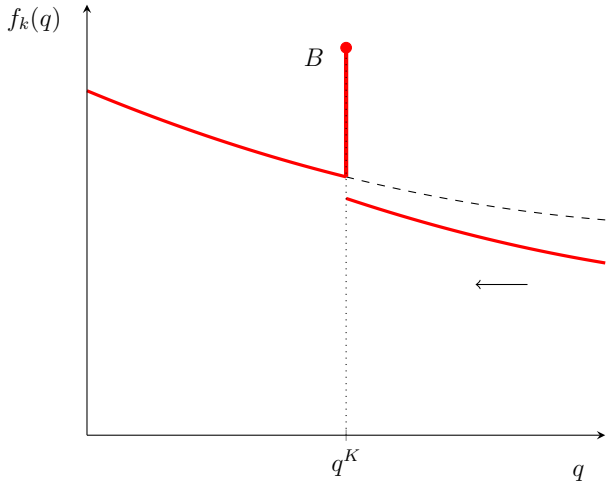
Figure 11: The effect of the kinked scheme S_k on adopters below the kink point



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 full dot depicts the choice under both subsidies. The coloured area depicts the variable profit under both subsidies.

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

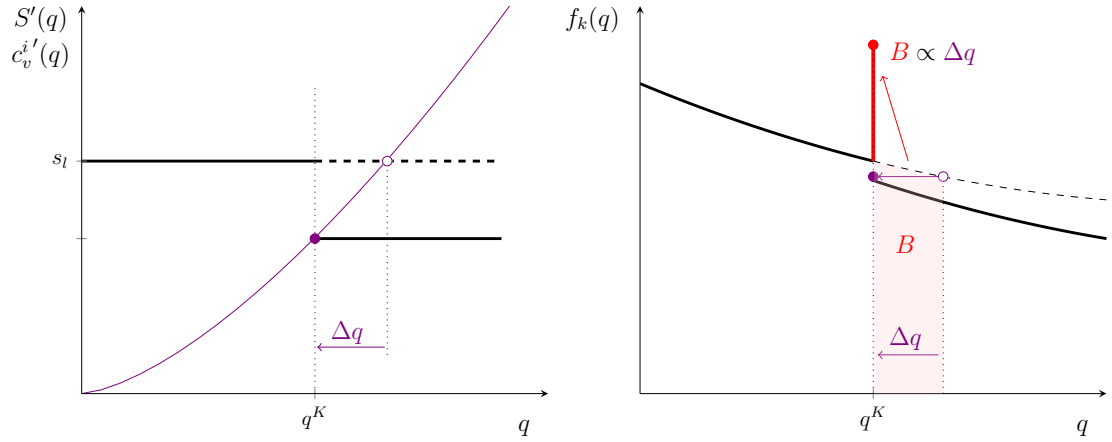


Note: The part of the measure above the kink point shifts to the left. It consists of adopters illustrated in Figure 9. At the kink point, there is a mass point B, i.e., the bunching mass. It consists of adopters illustrated in Figure 10. The part of the measure below the kink point is not affected by the kink. It consists of adopters illustrated in Figure 11.

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.[11] 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, when doing so, they hit the kink point q^K. By reaching the kink point, they are no longer affected by the subsidy change. Therefore, they "bunch" precisely at the kink point. The measure is the same under the two subsidy schemes below the kink point because adopters in this range are not affected by the change in schemes.

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

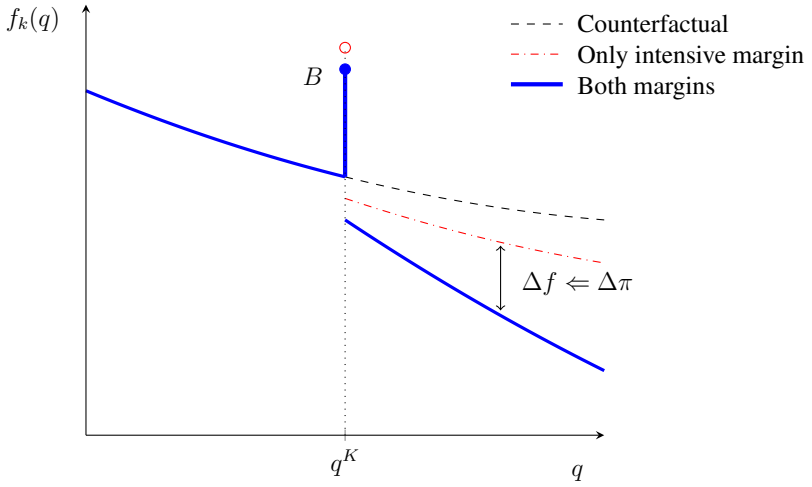
Figure 13: Identification when there is only an intensive margin response



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.

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.

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.

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 not affected 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. Above the kink point, adopters from the first group suffer from a loss in profit as well. 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 formalize 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 next section carries out this exercise.

3.2 The formal derivation of the distribution of adoptions

This section derives formally 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, I show that a structural approach, 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.

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

$$\epsilon^i(s) = \frac{\mathrm{d} \ln q^i(s)}{\mathrm{d} \ln s}. \tag{6}$$

In the next step, I impose an assumption on the cost function. I formulate the assumption as a reduced form and a structural assumption and show that the two formulations are equivalent.

Assumption 1 (Locally isoelastic intensive margin response)**.**

1a *(Reduced form assumption):*
For small variations in the subsidy and for points close to the kink point, the intensive margin elasticity is constant.

Mathematically, 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.

***1b** (Equivalent structural assumption):*
The cost function is locally isoelastic.

Mathematically, 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

$$c_v^i(q) = \theta^i q^{1+\frac{1}{\epsilon}}, \tag{7}$$

where θ^i is the variable cost type.

Assumption 1a is a reduced form assumption on an endogenous object, i.e., a high-level assumption. However, Lemma 1 in appendix A.1 shows that it is equivalent to Assumption 1b, which is a structural assumption on the cost function. Assumption 1 is standard in the bunching literature.[12] It is a local parametric approximation to the nonparametric cost function. As noted by Kleven (2016), if ρ is close to 1, and $q^i(s_l)$ is close to the kink point, Assumption 1 is approximately true.[13] Suppose Assumption 1 does not hold. The closer ρ is to one, the smaller is the misspecification introduced by Assumption 1. In my application, ρ is greater or equal to 0.95. Additionally, in my application, I estimate the intensive margin elasticity ϵ at two kink points. I find that the elasticity is not significantly different at the two kink points, which is strong evidence that Assumption 1 holds.

Denote the choice of adopter i under the counterfactual subsidy S_l by q_l^i, where l stands for linear:

$$q_l^i = q^i(s_l). \tag{8}$$

Similarly, denote the total cost of adopter i under the counterfactual subsidy S_l by c_t^i, where t stands for total:

$$c_t^i = c_v^i(q_l^i) + c_f^i. \tag{9}$$

Corollary 1. *For each adopter i, there is a one to one mapping from the variable and fixed*

[12] In the bunching literature, the assumption can be generalized to heterogeneous ϵ. I leave such a generalization in my case for future research.

[13] The approximation still relies on homogeneity in the elasticity, i.e., for all i with equal $q^i(s_l)$, ϵ^i is equal.

cost type (θ^i, c_f^i) *to the choice and total cost under the counterfactual subsidy* (q_l^i, c_t^i)*:*

$$\theta^i = \frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon}, \tag{10}$$

$$c_f^i = c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}. \tag{11}$$

Therefore, locally, the total cost function is equal to

$$c(q, q_l, c_t) = \underbrace{\frac{s_l}{{q_l}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}}}_{\textit{variable cost}} + \underbrace{c_t - q_l s_l \frac{\epsilon}{1+\epsilon}}_{\textit{fixed cost}}. \tag{12}$$

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

The proof of Corollary 1 is in appendix A.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, s_l is observable and fixed. Note that from now on I drop the adopter specific index i.

The next paragraph imposes the corresponding isoelasticity assumption for 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 function $\eta(S, q)$ as

$$\eta(S, q) = \frac{f_{t|q_l}(S(q)|q)}{F_{t|q_l}(S(q)|q)} S(q), \tag{13}$$

where $S(.)$ is a general subsidy function. The participation margin elasticity under the counterfactual subsidy is $\eta(S_l, q_l)$. 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.

Assumption 2 (Locally isoelastic participation margin response)**.**

2a *(Reduced form assumption):*
For small variations in the subsidy and for points close to the kink point, the participation margin elasticity is constant.

Mathematically, 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 function $\eta(S, q) = \eta$, where η is the constant participation margin elasticity.

2b *(Equivalent structural assumption):*
The conditional CDF of the total cost is locally isoelastic.

Mathematically, for all values c_t in the interval $[S_k(q_l), S_l(q_l)]$, the conditional CDF of of the total cost is equal to

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)}\right)^{\eta}, \tag{14}$$

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

Lemma 2 in appendix A.3 shows the equivalence.

Proposition 1. *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 $f_l(.)$. Three parts of the observable measure $f_k(.)$ depend distinctly on the three unknowns:*

$$f_k(q) = f_l(q), \qquad \text{for } q < q^K; \tag{15}$$

$$f_k(q) = f_l(q) + B \qquad \text{for } q = q^K; \tag{16}$$

$$f_k(q) = R(q\ \rho^{-\epsilon}, \epsilon)^{\eta} f_l(q\ \rho^{-\epsilon})\rho^{-\epsilon}, \qquad \text{for } q > q^K. \tag{17}$$

At the kink point q^K there is a mass point with bunching mass B:

$$B = \int_{q^K}^{q^K \rho^{-\epsilon}} R(q_l, \epsilon)^{\eta} f_l(q_l) dq_l. \tag{18}$$

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 of 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 5 in appendix A.4.

The proof of Proposition 1 is in appendix A.4. Below the kink point, the observable

measure $f_k(.)$ depends only on the counterfactual measure $f_l(.)$. At the kink point, there 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, both 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 (15) 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 (17) and (18) 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. *The counterfactual measure $f_l(.)$ or one of its transformations is locally analytic: Mathematically, for all quantities q in a large enough interval $[\underline{q}, \overline{q}]$ around the kink point it holds that the counterfactual measure $f_l(.)$ or one of its transformations has a convergent power series representation.*

Note that Assumption 3 is not a high level assumption on an endogenous object; by Corollary 1, q_l is a structural parameter. Therefore, the measure $f_l(.)$ is a structural parameter. Assumption 3 is a smoothness assumption. A set of functions satisfying Assumption 3 are functions which are complex differentiable on $[\underline{q}, \overline{q}]$. Blomquist, Newey, Kumar, and Liang (2017) point out that the bunching literature uses parametric functional form assumptions on f_l to identify the intensive margin elasticity. They show that the assumption of only finite differentiability of f_l at the kink point does not identify the intensive margin elasticity. Proposition 2 below shows that under Assumption 3 the intensive margin elasticity ϵ, the participation margin elasticity η, and the counterfactual choice-distribution f_l are jointly identified.

Equations (15)-(18) 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 (17) as a function $f(q, \epsilon, \eta, f_l)$ and the bunching mass in Equation (18) as a

function $B(\epsilon, \eta, f_l)$:

$$f(q, \epsilon, \eta, f_l) = R(q\ \rho^{-\epsilon}, \epsilon)^{\eta} f_l(q\ \rho^{-\epsilon})\rho^{-\epsilon}, \tag{19}$$

$$B(\epsilon, \eta, f_l) = \int_{q^K}^{q^K \rho^{-\epsilon}} R(q_l, \epsilon)^{\eta} f_l(q_l) dq_l. \tag{20}$$

Condition 1. *Equation (17) and Equation (18) are not colinear. Mathematically, there exists a q such that*

$$\frac{\frac{\partial B(\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial B(\epsilon,\eta,f_l)}{\partial \eta}} \neq \frac{\frac{\partial f(q,\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial f(q,\epsilon,\eta,f_l)}{\partial \eta}}. \tag{21}$$

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

$$\frac{\frac{\partial B(\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial B(\epsilon,\eta,f_l)}{\partial \eta}} = \frac{\frac{\partial f(q,\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial f(q,\epsilon,\eta,f_l)}{\partial \eta}} \text{ for all } q. \tag{22}$$

Equation (22) 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 ex-post estimation and find that it holds by a large amount (see Section 6.1.1 for the estimates and a more detailed discussion).

Proposition 2. *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. Observations below the kink point nonparametrically identify the counterfactual. Observations at the kink point and observations above the kink point jointly identify the two response margins. They are locally identified.*

The proof is in appendix A.5. Contrary to methods using an exogenous variation, I make no assumption for q^K, s_l, and ρ, except for their observability. They can be random or endogenous. Because Condition 1 holds generically, ϵ, η, and f_l are generically identified.

Theoretically, I cannot prove global identification, i.e., for a given observable $f_k(.)$ several pairs (ϵ, η) may solve Equation (17) and (18). However, I verify the uniqueness of the solution ex-post estimation. Additionally, simulating the system for many values of

f_l, ϵ, and η suggests that the system has a unique solution. The simulations show that the bunching mass B depends mainly on the intensive margin elasticity, while the slope change in the measure depends mainly on the participation margin elasticity.

4 Description of the German policy

The German subsidy for solar panels was introduced on 1 April 2000. The subsidy is a guaranteed feed-in-tariff, paid per kWh (kilowatt-hour) of electricity produced. Once an agent decides to adopt, a constant tariff rate is guaranteed for 20 years. Typically, agents only adopt once. Rates depend on the time point of adoption. They decreased over time, as did the price for solar panels. From 2004 onwards, rates also depended on the capacity of the system. The subsidy is paid as a fixed rate per kWh if the system has a capacity smaller than 30 kWp. Because production is linear in peak capacity, the subsidy is equivalent to a subsidy linear in peak capacity. If the system has a peak capacity higher than 30 kWp, the proportion of electricity produced by the higher capacity is remunerated by a smaller rate per kWh. Imagine the installed system has 60 kWp, then 50% of the produced electricity is paid the higher rate and 50% the lower rate. A subsidy of this form is piecewise linear in capacity, with a kink at 30 kWp (see appendix A.6.1 for the exact formula). A second kink was located at 100 kWp. In 2009, an additional kink was introduced at 1000 kWp. In 2012 the kinks changed to 10, 40, 1,000, and 10,000 kWp. From 2009 onwards the kink overlaps with other policy changes at 30 kWp and 100 kWp. Therefore, I focus on the years 2004 to 2008.

The data used in this paper are administrative and contain all solar panels connected to the grid and receiving subsidy payments. Since there is no reason for investing in solar panels and not claiming the subsidy, the data set is very likely complete. An adopter may be a household or a firm. The unit of observation is the aggregated capacity installed by an adopter at a certain 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 of them. 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 1 shows the number

of adoptions in different years. The number is increasing for most years.

Table 1: Number of adoptions per year

Year	Number of adoptions	Relative proportion [%]
Until 2001	30,934	9.5
2002	14,073	4.3
2003	15,137	4.6
2004	35,077	10.8
2005	49,678	15.3
2006	44,990	13.8
2007	54,160	16.6
2008	81,508	25.0
All years	325,557	100.0

Column 2 in Table 2 shows some values of the CDF of the number of adoptions. The distribution is highly skewed to the left. 65% of adoptions are smaller than 10 kWp. The third column in Table 2 shows the proportion in the aggregate capacity of adopters smaller than the threshold. For example, adoptions smaller than 10 kWp account for 27 % of aggregate capacity. The table shows that the distribution of aggregate capacity is less skewed to the left than the distribution of the number of adoptions.

Table 2: Relative number of adoptions and aggregate capacity below certain thresholds

Threshold [kWp]	Relative number of adoptions [%]	Relative aggregate capacity [%]
5	33	8.9
10	65	27
30	93	70
100	99.5	91
150	99.8	94

Note: The number in the column "Relative number of adoptions" shows the fraction of adopters with a capacity smaller than the value in the column "Threshold." The number in the column "Relative aggregate capacity" shows the fraction of total capacity from installations smaller than the value in the column "Threshold."

5 Estimation

I closely follow Chetty, Friedman, Olsen, and Pistaferri (2011) in the estimation procedure. First, I construct an empirical histogram by choosing bins and counting the number of adopters in each bin. Normalization, 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 $\{1, 2, ..., N\}$ is the index of the bin and N is the total number of bins.

Theory predicts that the bunching mass is located precisely at the kink point. However, in practice, the excess mass is scattered around the kink point. 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). The observed bunching mass $\widehat{B}$ is the normalized number of adopters in the bunching interval. I check for the robustness of the estimates regarding the choice of the bunching interval in appendix A.12.4

The next step is the choice of transformation for the infinite-dimensional parameter $f_l(.)$. I use a logarithmic transformation on $f_l(.)$ and q:

$$\ln f_l(q) = \sum_{p=0}^{\infty} \gamma_p \frac{1}{p!} \ln(q/q^K)^p. \tag{23}$$

The logarithmic transformation of the measure and the argument is natural because both variables are defined over a positive domain.[14] 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 (23) contains the uniform distribution, the Pareto distribution, and the log-normal distribution as special cases. These are common distributions for variables with a positive domain. Figure 4 illustrates that, except at the kink point, the capacity-distribution is very close to linear in the logarithmic scale. This shape is visual evidence for a counterfactual distribution, which is close to a Pareto distribution. I check the fit of the log-log transformation against a specification without transformation and a logarithmic transformation on $f_l(.)$ only. To this end, I use pretreatment data and treated data, excluding observations around the kink. The details are in appendix A.11. The log-log transformation approximates the distribution best.[15]

[14] As suggested by the empirical evidence, I assume the the measure of adoptions is strictly positive on the relevant domain.

[15] 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 the expansion, are very uncommon. Additional restrictions need to 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 on the argument (log-density

It is impossible to estimate infinitely dimensional pseudo-parameters such as $f_l(.)$ because, in practice, sample sizes are finite. Nonparametric methods address this problem by estimating a restricted number of parameters that grow with the sample size. I use a semi-nonparametric sieve estimator (see Chen, 2007 for a review). The idea behind this estimation method is to use a finite series, where the order of the series increases in sample size:

$$\ln f_l(q) = \sum_{p=0}^{P(N)} \gamma_p \frac{1}{p!} \ln(q/q^K)^p. \tag{24}$$

$P(N)$ is the order of the series and goes slowly to infinity as the number of bins N goes to infinity. I discuss the choice of $P(N)$ below. For a given P, I use a parametric M-estimator.

I use the logarithm of $\widehat{f(q_j)}$ as the dependent variable. Using the logarithm does not affect consistency; however, it introduces a small sample bias. To reduce the impact the bias, I use bias correction techniques. The details are in appendix A.8. The variable $\widehat{\ln f(q_j)}$ denotes the bias-corrected dependent variable. The transformation corresponds to an econometric model of the form

$$\widehat{\ln f(q_j)} = \ln f_k(q_j \mid \eta, \gamma, \epsilon) + u_j, \tag{25}$$

where u_j is the noise term. Note that, because I normalized the histogram by the total number of observations, $\widehat{\ln f(q_j)}$ is an estimate of the density. Without loss of generality, I assume the total mass of adopters under the observed subsidy is equal to one. Therefore, the measure $f_k(q)$ is a density. The empirical model (25) relates it to the observed density $\widehat{\ln f(q_j)}$.

The logarithmic model has several advantages over the additive model

$$\widehat{f(q_j)} = f_k(q_j \mid \eta, \gamma, \epsilon) + u_j. \tag{26}$$

First, the observed density is positive by definition in the logarithmic model. In the addi-

estimation; see Stone (1990)). This transformation is a natural approach when the argument's domain is the real line, which is not the case in my application. It contains the exponential distribution and the normal distribution as special cases.

tive model, the density is negative for large negative errors. Therefore, only the logarithmic model is logically consistent. Second, besides sampling noise, the noise term u_j may capture additional random disturbances. For some random reasons certain capacities may be more or less frequent in the data than predicted by f_k. It is more natural to think about these additional disturbances as proportional to f_k, corresponding to specification (25), and not as additive, corresponding to (26).

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, I use bins with a bin-size that increases in capacity.[16] The exact specification of the binning function is in Equation (110) in appendix A.9. An additional advantage of this binning procedure is that it equalizes the variance of the dependent variable $\widehat{\ln f(q_j)}$. Therefore, it avoids the need for a weighting matrix in the estimation.

For the estimation, I follow the same two-step least square procedure as Chetty et al. (2011). First, I minimize the square distance between the model and the observed density outside the bunching interval:

$$\min_{\eta,\gamma} \sum_{q_j \notin [q_L, q_H]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \eta, \gamma, \epsilon_1) \right)^2. \tag{27}$$

I choose an initial guess for ϵ denoted by ϵ_1. Second, I use the estimates $\hat{\eta_1}$ and $\hat{\gamma_1}$, and I minimize the square distance between the observed bunching mass and the model

$$\min_{\epsilon} \left(\widehat{\ln B} - \ln \int_{q_L}^{q_H} f_k(q \mid \hat{\eta_1}, \hat{\gamma_1}, \epsilon) dq \right)^2. \tag{28}$$

The minimum is the estimate $\hat{\epsilon_2}$. I repeat the procedure until $\hat{\eta}, \hat{\gamma}, \hat{\epsilon}$ converge. The estimates are consistent and asymptotically normal; the proof is in appendix A.9. Following Chetty et al. (2011), I estimate the standard errors using nonparametric bootstrap.

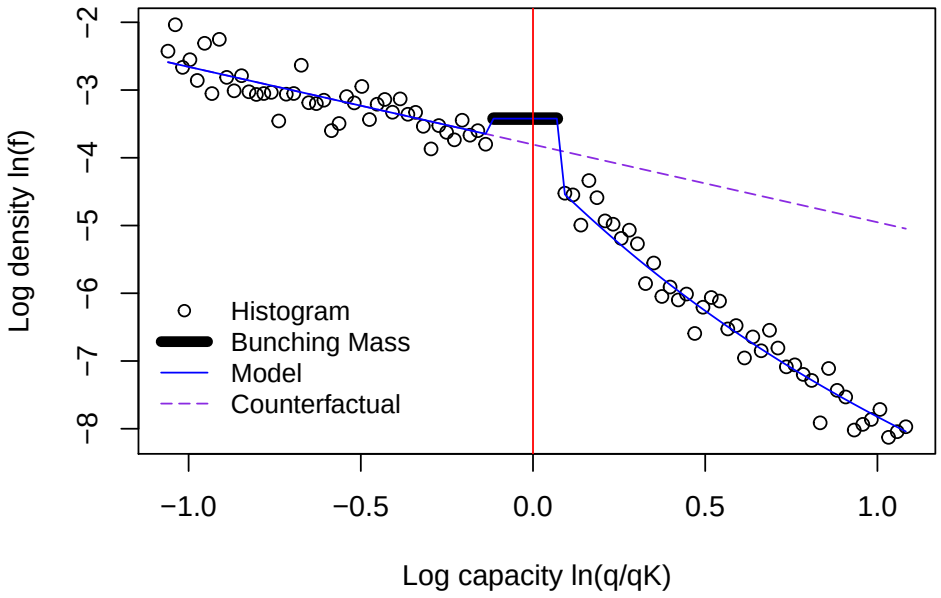
[16] Note that such a binning procedure does not affect consistency because $\widehat{f(q_j)}$ is equal to the number of observations in bin j normalized by the bin-size and the total number of observations.

As in any nonparametric estimation, the estimates have a bias and a variance. Both depend on the specification of the nonparametric estimation, giving rise to the classic bias-variance trade-off. There are two specification-parameters: The first parameter is the order of the polynomial P. The higher is the order P, the lower is the bias, and the larger is the variance of the estimates. The second parameter is the bandwidth $b = [\underline{q}, \overline{q}]$. It is the interval of values around the kink point used for estimation. The smaller is b, the lower is the bias, and the larger is the variance of the estimates. 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. It standard in nonparametric estimations to choose the specification that minimizes an estimate of the mean squared error. 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 to choose the specification b and P. Appendix A.10 discusses the estimator of the mean squared error. Appendix A.12 shows the estimates and the optimal specification.

6 Results

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

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



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

Table 3 shows the results at 30 kWp for all years until 2011.

Table 3: Results for the kink point at 30 kWp

Year	ϵ (σ_ϵ)	η (σ_η)
2004	2.3 (0.2)	91 (3)
2005	3.9 (0.2)	71 (3)
2006	4.4 (0.3)	55 (4)
2007	5.2 (0.3)	60 (3)
2008	5.2 (0.2)	49 (2)
2009	7.7 (0.2)	30 (2)
2010	7.8 (0.2)	15 (1)
2011	7.9 (0.1)	0.00 (0.01)
2004-2011	6.72 (0.06)	20.2 (0.5)

Note: The table reports the estimated elasticities with standard errors in brackets. The last row reports the aggregate estimates pooling all years.

The intensive margin elasticity is increasing over time, while the participation margin

elasticity is decreasing over time. As already mentioned in the introduction, the participation margin in the first years is substantial. Since 2004 is early in the solar adoption programme, the mass of adopters just on the margin of participation is large compared to the mass of adopters that already adopted. Although the magnitude of the elasticity is surprising, the fact that the elasticity is larger than elasticities measured in saturated markets is not.[17] The fact that the elasticity decreases over time as the market saturates is consistent with this explanation. However, the aggregated long-run elasticity is still considerable. Additionally, the profitability of the programme increases rapidly over time since costs for solar panels decrease. It is reflected by the rapid increase of adoptions from 2000 to 2011. The fixed cost becomes less and less relevant, reflected in the decrease in the participation margin elasticity.

Table 4 shows the results at the 100 kWp kink point for the aggregate data in the years 2004 to 2008. It compares them with the results at 30 kWp for the same period. The scatter plots and exact specifications of the two estimations are in appendix A.13.

Table 4: Aggregate results for the two kink points in the years 2004 to 2008

Capacity	ϵ (σ_ϵ)	η (σ_η)
30 kWp	4.8 (0.1)	65 (1)
100 kWp	4.4 (0.8)	1.1 (10)

Note: The table reports the estimated elasticities with standard errors in brackets. The estimation pools all years.

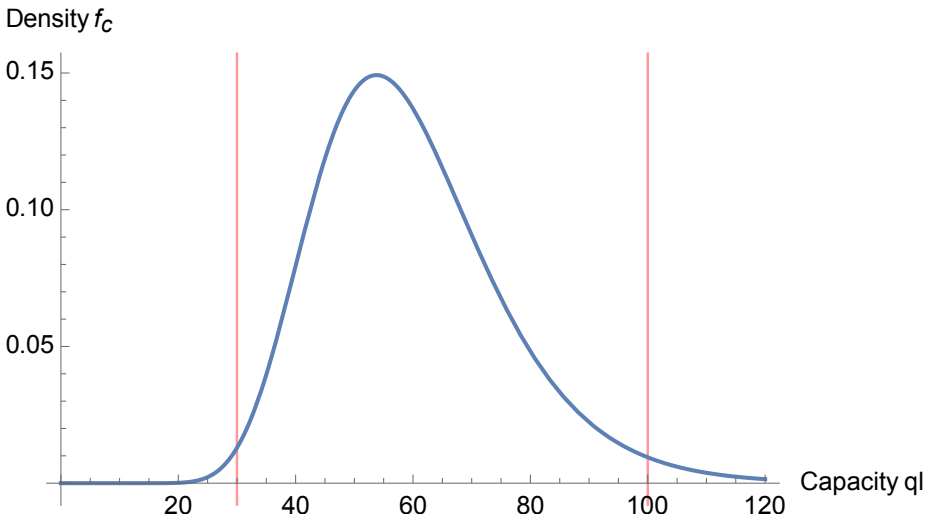
I estimate the elasticities at 100 kWp on the aggregated data from 2004 to 2008 because the sample size is small. I cannot use later years because the cutoff of 100 kWp is used as a cutoff for other policies influencing the adoption capacity. For the same reasons, I do not use any other kink points or years. The results suggest that the intensive margin elasticity is the same for adopters of different capacities, while the participation margin elasticity rapidly decreases with capacity. This evidence is not surprising. On the one hand, the primary 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. On the other hand, the participation margin elasticity depends on the profit net of the fixed cost. The higher this profit, the lower is the participation margin elasticity. Large capacity systems are very profitable. Therefore, the

[17] By a saturated market I mean a market where the number of potential buyers relative to the number of buyers is small.

fixed cost plays only a small role in the adoption decision. Only very few adopters have such high fixed costs to make adoption of a large capacity unprofitable. In contrast, the adoption of low capacity systems depends crucially on fixed costs. Many adopters have a fixed cost equal to the net profit. It follows that many adopters are close to indifferent in terms of participating. As a consequence, a small increase in the subsidy payment incentivizes many adopters to participate.

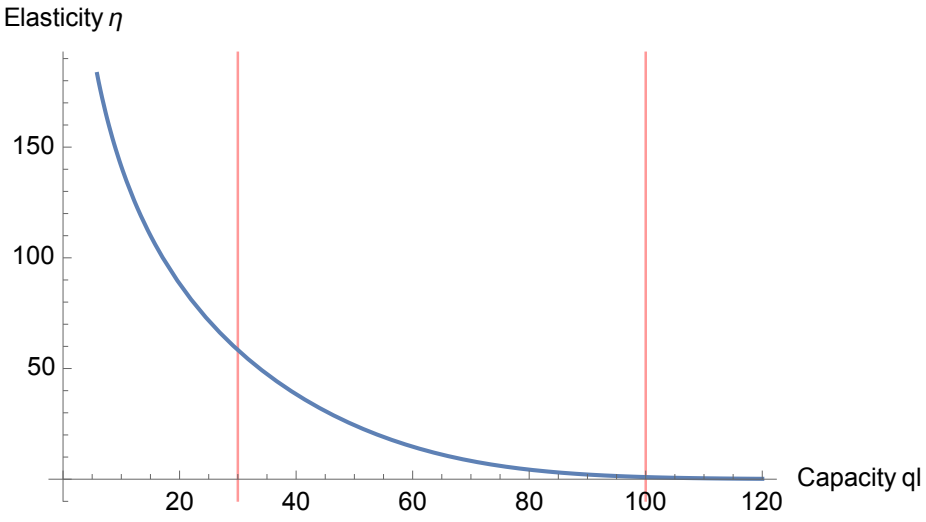
The assumption of a locally constant participation elasticity is convenient for the estimation, but it is at odds with the empirical evidence. The estimates show that the participation elasticity is decreasing in variable profits. The most straightforward assumption consistent with this evidence is a log-normal and independent distribution of fixed costs. By definition, fixed costs are positive. The log-normal density is a simple bell curve defined over the positive support. Another advantage of the log-normal distribution is that it has only two parameters. Two elasticities, observed at two different capacity levels, are sufficient to calibrate the distribution. I use the results in Table 4 to calibrate the distribution in Figure 16.

Figure 16: The calibrated density of fixed costs



Note: The figure shows the density of fixed costs $f_c(\pi_v(q_l))$ 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.*

Figure 17: The implied elasticity of the calibrated distribution of fixed costs



Note: The figure shows the participation elasticity at capacity q_l under the counterfactual subsidy, implied by the calibrated distribution of fixed costs.

The calibrated mean and variance of the logarithmic fixed costs are equal to $(2.3, 0.26)$. Figure 16 shows the calibrated density of fixed costs $f_c(\pi_v(q_l))$ as a function of the counterfactual variable profit π_v at capacity q_l. Therefore, the function shows the mass of agents indifferent to participate at a certain capacity level under the counterfactual subsidy. Figure 17 illustrates the implied participation elasticity. It decreases in variable profit, and therefore, in capacity.

6.0.1 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. This effect is due to the participation margin response. Instead of bunching, some agents stop participating in the programme, which decreases the bunching mass. Wrongly, ignoring this effect attributes the smaller bunching mass to a smaller intensive margin response. 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 (27):

$$\tilde{\epsilon} = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln \int_{q_L}^{q_H \rho^{-\epsilon}} f_l(q \mid \hat{\gamma}, \epsilon) dq \right)^2, \tag{29}$$

where $\hat{\gamma}$ is the same estimate as in (27). Correspondingly, ignoring the intensive margin biases the estimate of the participation margin elasticity. To evaluate the bias, I proceed correspondingly to above:

$$\tilde{\eta} = \arg\min_{\eta} \left(\sum_{j} \widehat{\ln f(q_j)} - \ln f_k(q_j | \epsilon = 0, \eta, \widehat{\gamma}) \right)^2 . \tag{30}$$

Table 5 summarizes the results for the estimates at 30 kWp in the years 2004 to 2008:

Table 5: Biased estimates ignoring the other margin

Parameter	Unbiased Estimate	Biased Estimate	Relative Difference
ϵ	4.76 (0.11)	3.82 (0.09)	- 20 %
η	65 (1)	69 (1)	+ 6 %

Note: The table shows the correct 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.

The result shows that ignoring participation introduces a downward bias of 20 % in the estimate of the intensive margin elasticity. It suggests that considering participation is essential, at least in this application. Ignoring the intensive margin introduces an upward bias of 6 % in the estimate of the participation margin elasticity.

6.1 Robustness

I report robustness checks for the aggregate estimates of the two margins.

6.1.1 Rank condition and global identification

Table 6 shows the rank condition evaluated at the estimated values. The standard errors are in brackets:

Table 6: Co-linearity condition evaluated at the estimated values

	$\frac{\frac{\partial B(\hat{\epsilon},\hat{\eta},\hat{f}_l)}{\partial \epsilon}}{-\frac{\partial B(\hat{\epsilon},\hat{\eta},\hat{f}_l)}{\partial \eta}}$	$\neq$	$\frac{\frac{\partial f(q^K,\hat{\epsilon},\hat{\eta},\hat{f}_l)}{\partial \epsilon}}{-\frac{\partial f(q^K,\hat{\epsilon},\hat{\eta},\hat{f}_l)}{\partial \eta}}$
$q^K = 30$	124 (5)	$\neq$	0.67 (0.30)
$q^K = 100$	3290 (1121)	$\neq$	51 (9)

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

The table shows that Condition 1 holds in sample.

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 bunch. That is the reason why the upper bound of the integral in Equation (18) 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 the case because R is roughly one minus the profit loss from re-optimization. As a consequence of 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

$$\frac{\frac{\partial B(\epsilon,\eta,f_l)}{\partial \epsilon}}{-\frac{\partial B(\epsilon,\eta,f_l)}{\partial \eta}} \tag{31}$$

is large. On the other hand, f mainly depends on η. To see that, consider the elasticity $\mathcal{E}$ of the function f_k.[19] From equation (17) it follows that:

$$\mathcal{E} f_k(q) = \mathcal{E} f_l(q\ \rho^{-\epsilon}) + \eta\ \mathcal{E} R(q\ \rho^{-\epsilon}). \tag{32}$$

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 (32) 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.

The estimation converges to the same estimates for starting values $\epsilon_0 \times \eta_0 \in [0, 200] \times [0, 1000]$, which is strong evidence for global identification.

6.1.2 Robustness check on untreated data

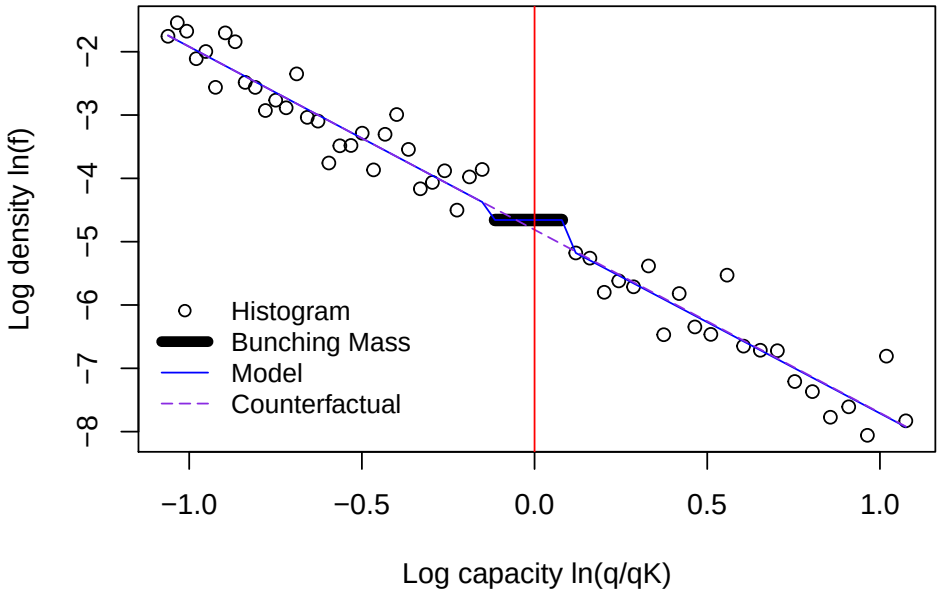
One concern is the violation of Assumption 2 because of irregularities in the counterfactual. On the one hand, there might be an excessive mass of adopters at the kink point for other reasons than the subsidy's kink. Not or only hardly predictable from observations to the left of the kink point, the irregularity may be a mass point or a continuous bump. This is a general concern in the bunching literature and was raised by Blomquist, Newey, Kumar,

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

and Liang (2017). It leads to an upward bias in the intensive margin estimate. Additionally, even if Assumption 2 holds, in small samples, choosing an inefficient series approximation may considerably bias the results.

On the other hand, there might be a slope change at the kink point for other reasons 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 crucial to evaluate if the chosen specification is picking up one of these biasing effects. To address these issues, I estimate the model on the observations in the years 2000 to 2003. From 2000 until 2003, there was no kink in the subsidy. If any of the concerns just raised are an issue, the estimates in 2000 to 2003 should be significant. The Figure 18 shows the histogram in 2000 to 2003, with the estimated model and the linear counterfactual. I use the same bunching interval as in 2004 to 2008.

Figure 18: Histogram in the years 2000 to 2003 with estimated model



Note: The figure shows the robustness check. The x-axis shows the normalized 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 7: Results for the untreated data in the years 2000 to 2003 (placebo)

Year	ϵ (σ_ϵ)	η (σ_η)
2000-2003	0.2 (0.3)	−0.5 (6)

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

The selected specification does not estimate significant parameters in the untreated data. It shows that the chosen specification estimates the effects on the distribution correctly. Additionally, it shows that the counterfactual distribution is smooth enough not to mistakenly estimate an effect. The result is robust to changes in the bandwidth. Appendix A.14 reports the same robustness check for the estimates at 100 kWp. Again, ϵ and η are not significant.

7 Policy evaluation

This section uses the estimates to evaluate and optimize the subsidy. As in existing kink and discontinuity methods, the estimates are locally correct at each kink point. It is necessary to make global assumptions to use them for policy evaluation. I restrict the policy analysis to the years 2004 to 2008 since the local estimates at two different capacity levels give the best available information about global properties. In line with the empirical evidence in Table 4, I assume an isoelastic cost function and a log-normally distributed fixed cost.

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

$$c(q, q_l, c_f) = \frac{s_l}{q_l^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}} + c_f, \tag{33}$$

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

Assumption 5 (Log-normal fixed cost)**.** .
The distribution of the fixed cost c_f is log-normal and independent of q_l:

$$\ln(c_f) \sim N(\mu_f, \sigma_f), \tag{34}$$

with CDF $F_f(c_f)$ and density $f_f(c_f)$.

I set ϵ equal to the observed intensive elasticity and calibrate (μ_f, σ_f) to match the observed participation elasticities. It follows that $(\mu_f, \sigma_f) = (2.3, 0.26)$. The model can be solved for any counterfactual subsidy scheme using these assumptions. Note that, in line with the bunching literature, I used the term counterfactual for the linear subsidy without a kink. In this section, I call any subsidy different from the observed kinked subsidy, a counterfactual subsidy. I normalize the linear rate s_l to one, which corresponds to a choice of monetary unit. As a consequence, all monetary variables are to be interpreted relative to s_l. For simplicity, I assume that the distribution of q_l is log-normal. The assumption is simple and fits the lower part of the distribution well. This part is responsible for the bulk of adopted capacity. Note, however, this assumption is easy to relax when needed, as one can invert any estimate of $f_k(q)$ by using the global assumptions about the reaction margins.

7.1 The government's objective

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

$$\min_{S(.)} \int S(q) dF_c(q), \tag{35}$$

such that

$$\int q \, dF_c(q) \geq Q^T, \tag{36}$$

where $F_c(q)$ is the implied counterfactual distribution of adoptions. While Objective (35) is appealing for its simplicity, it is a simplification of real governmental preferences. Objective (35) implicitly assumes that the government does not want to distribute rents to adopters of solar panels. Typically, only households in the upper part of the income distribution can adopt solar panels, which gives a rationale for such governmental preferences.

The estimates of both reaction margins can also be used to maximize a much more general governmental objective. Appendix A.15 discusses an objective with general redistributive preferences, a general valuation of aggregate capacity, and an optimal tax on other sources of income. I derive under which conditions Objective (35) 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. The appendix derives the theoretical solution to the general objective. It is necessary to know the government's redistributive preferences and its valuation of aggregate capacity to evaluate the general solution quantitatively. Such an exercise is beyond the scope of this paper. Therefore, I use the simpler objective (35) for the quantitative exercises in the following sections. However, a government can use its real preferences and valuation together with the general solution in appendix A.15 to implement a subsidy.

A prominent possible subsidy scheme is the Pigouvian subsidy. The Pigouvian subsidy is a linear subsidy where the marginal subsidy rate is equal to the marginal social benefit of the public good. It is also known as the Samuelson rule (Samuelson, 1954). To implement the Pigouvian subsidy, it is not necessary to know how adopters react to the subsidy. However, the optimal solution to Problem (35) is not the Pigouvian subsidy. For the general objective in appendix A.15, the Pigouvian subsidy is only optimal if the government is indifferent in distributing rents to adopters. The German government already uses a nonlinear subsidy scheme, showing that it is not indifferent in distributing rents. Even if an optimal income tax is available, Kaplow (1996) shows 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 (see appendix A.15 for an explicit model of this heterogeneity). 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, it is optimal to limit their rents through a nonlinear subsidy. A detailed discussion on why the Pigouvian subsidy is not optimal is in appendix A.15.

7.2 The optimal linear subsidy

In a first experiment, I solve for the optimal linear subsidy that keeps aggregate installed capacity constant. I compare its public cost with the cost of the observed subsidy. Denote by ρ_l the marginal subsidy rate. Note that, as I normalized s_l to one, ρ_l's interpretation

is relative to s_l. It means that a linear subsidy with the rate of $\rho_l s_l$ would incentivize for the same aggregate adoption behaviour as the actual kinked subsidy. By the first order condition of the decisions problem, the choice q of type q_l under subsidy ρ_l is

$$q(q_l, \rho_l) = q_l \rho_l^{\epsilon}. \tag{37}$$

Denote by $c_v(q, q_l)$ the variable part of the cost function:

$$c_v(q, q_l) = c(q, q_l, c_f) - c_f. \tag{38}$$

Denote the variable profit of type q_l under subsidy rate ρ_l as

$$\pi_v(q_l, \rho_l) = \rho_l q(q_l, \rho_l) - c_v(q(q_l, \rho_l), q_l). \tag{39}$$

Given the estimate of $f_l(q_l)$, the unconditional type distribution $f_u(q_l)$ is

$$f_u(q_l) = \frac{f_l(q_l)}{F_f(q_l - c_v(q_l, q_l))}. \tag{40}$$

It follows that ρ_l is the solution to

$$\int_0^{\infty} q(q_l, \rho_l) F_f(\pi_v(q_l, \rho_l)) f_u(q_l) dq_l = Q^T, \tag{41}$$

where Q^T is the observed aggregate capacity. Numerically, I find that $\rho_l = 0.9998$. The public cost of the linear policy is $Q^T \rho_l$. The policy is only 0.016 per cent more expensive than the actual subsidy. Therefore, it shows that the nonlinearities introduced by the government are barely effective in reducing costs.

7.3 The optimal nonlinear subsidy

In this section, I solve for the optimal nonlinear policy using mechanism design. The analysis follows the monopoly pricing problem in Rochet and Stole (2002). I rewrite the government's objective (35) as a Lagrangian and a mechanism design problem. The gov-

ernment maximizes

$$\max_{\psi, q(.), \pi_v(.)} \int_0^\infty \left(\psi q(q_l) - \pi_v(q_l) - c_v(q(q_l), q_l)\right) F_c(\pi_v(q_l)) f_u(q_l) dq_l - \psi Q^T, \tag{42}$$

such that for all q_l

$$\pi_v'(q_l) = -\frac{\partial c_v(q(q_l), q_l)}{\partial q_l} \text{ and } q(.) \text{ is not decreasing.} \tag{43}$$

(43) is the incentive-compatibility constraint. The variable ψ denotes the Lagrange multiplier. I substituted 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 (42) is equivalent to Problem (35). 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 a compensation. An incentive-compatible mechanism are two functions $q(.), \pi_v(.)$, under which each agent has the incentive to report her type truthfully. Therefore, the objective of the government is finding two such functions, which maximize its objective. The incentive-compatibility constraint (43) follows from the standard revealed preference argument in mechanism design. As standard in the literature, I neglect the monotonicity constraint on $q(.)$ and verify it ex-post. Define as

$$L(\pi_v, \pi_v', q_l) = \left(q_l \psi (\pi_v'(1+\epsilon))^{\frac{\epsilon}{1+\epsilon}} - \pi_v - q_l \epsilon \pi_v'\right) F_c(\pi_v) f_u(q_l), \tag{44}$$

which is the integrand of Problem (42). I used Equation (43) to substitute for the function $q(.)$. The problem simplifies to finding an optimal function $\pi_v(.)$. I 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

$$\frac{\partial L(\pi_v, \pi_v', q_l)}{\partial \pi} = \frac{d}{dq_l} \frac{\partial L(\pi_v, \pi_v', q_l)}{\partial \pi'} \quad \text{for all } q_l, \tag{45}$$

and

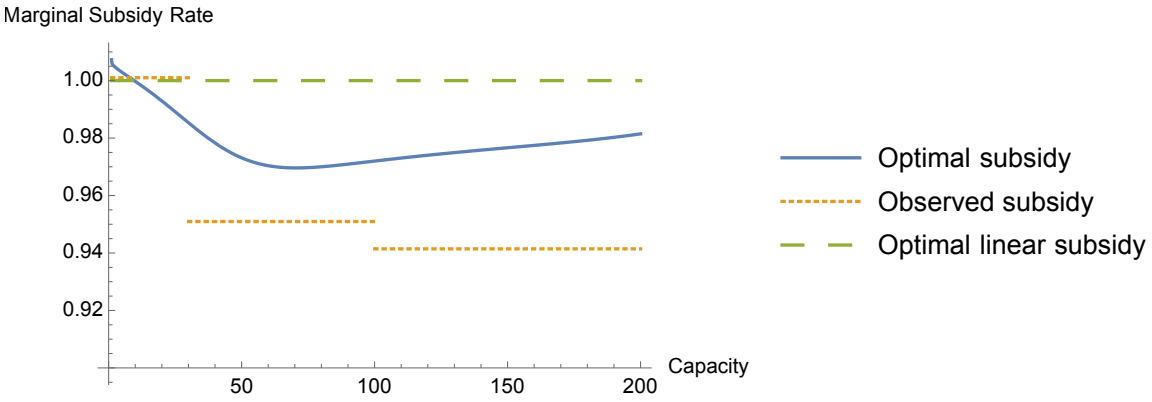
$$\frac{\partial L(\pi, \pi', q_l)}{\partial \pi'} = 0 \text{ for } q_l = 0 \text{ and } q_l = \infty. \tag{46}$$

For each ψ, the Equations (45) and (46) define a nonlinear second order differential equation with boundary values. I fix ψ, solve the differential equation numerically, and evaluate the capacity constraint $Q = Q^T$. I iterate over ψ until the constraint holds.

7.3.1 The optimal solution

Using the first order condition and the solution $\pi'_v(q_l)$, I solve for the optimal nonlinear marginal subsidy $S'(q)$. Figure 19 shows the optimal marginal subsidy schedule in comparison to the linear benchmark and the observed marginal subsidy. The optimal marginal subsidy is u-shaped and very close to constant. Therefore, the optimal subsidy $S(q)$ is close to linear.

Figure 19: The observed and the optimal marginal subsidies



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

Using the definition of the variable profit, it follows that the total public costs are

$$\int_0^\infty \left(\pi_v(q_l) + c_v(q(q_l), q_l)\right) F_c(\pi_v(q_l)) f_u(q_l) dq_l. \tag{47}$$

Compared to the benchmark, cost savings are by a factor 3.5 larger than the cost savings of the actual subsidy. However, the optimal subsidy is still only 0.055 per cent less costly than the benchmark. This result shows that the benefits of nonlinear prices are very small due to the high reaction margins.

7.4 Zero participation

In this section, I assume the participation elasticity is equal to zero. This thought experiment will answer several questions: What is the optimal subsidy without a participation margin? Would a nonlinear subsidy perform well under these circumstances? Can the participation margin be safely ignored for the counterfactual analysis?

Without participation it is possible to use standard mechanism design to solve for the optimal subsidy. It follows that the optimal capacity $q(q_l)$ as a function of type q_l solves

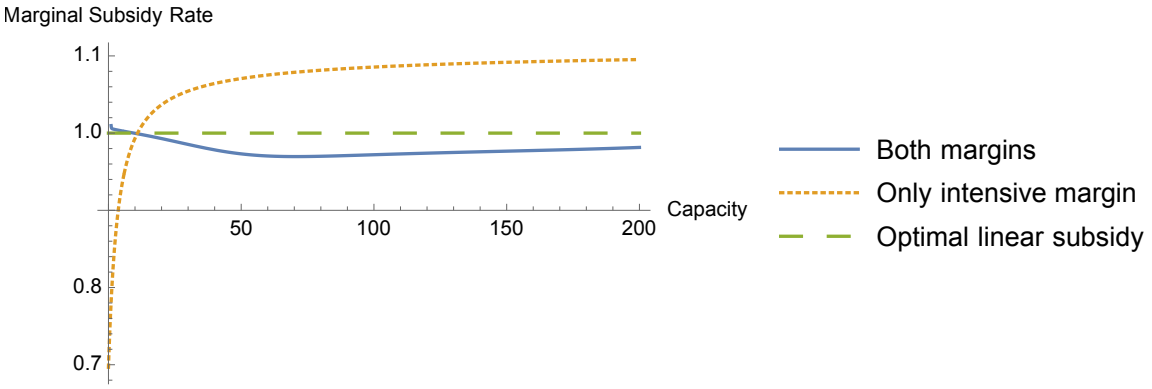
$$\frac{\partial c_v(q(q_l), q_l)}{\partial q} = \psi + \frac{1 - F_l(q_l)}{f_l(q_l)} \frac{\partial^2 c_v(q(q_l), q_l)}{\partial q \partial q_l}. \tag{48}$$

Inverting the function $q(.)$ and using the first order condition gives the optimal marginal subsidy:

$$S'(q) = \frac{\partial c_v(q, q_l^{-1}(q))}{\partial q}. \tag{49}$$

Figure 20 compares the optimal marginal subsidy without participation to the optimal marginal subsidy with participation.

Figure 20: The optimal marginal subsidy with and without participation



Note: The figure compares the optimal marginal subsidy without participation with the optimal marginal subsidy with participation. The marginal subsidy without participation is much farther from the linear benchmark. When both margins are present, the participation margin is responsible for an optimal scheme that is close to linear.

The marginal subsidy schedule is increasing, and the distance to the linear benchmark is much more significant. This result shows that it is the large participation margin which limits the role of nonlinearities. Without participation margin, the optimal nonlinear subsidy would be 7.6 per cent less costly than the benchmark. Again, it is the large participation margin, which makes nonlinearities ineffective at reducing costs. The shapes of the marginal subsidy rates in Figure 20 follow from simple intuition. First, consider the optimal schedule without participation, depicted by the orange dotted line. The marginal rates

are increasing with capacity. 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. In contrast, the marginal rate paid to small adopters is low. They adopt lower quantities than under the linear benchmark. The total payment to large adopters is the integral under the orange curve. The low marginal rates for small adopters extract profit from large adopters without affecting their choice of adoption. For this reason, it is optimal to pay low rates to small adopters, and therefore, to have an increasing marginal schedule. When there are participation margin responses, the same logic does not apply anymore. Due to the participation margin, the total subsidy payments and the marginal subsidy rates are relevant for the adoption decision. Reducing the marginal rate for small adopters reduces the profit for larger adopters. Therefore, it affects participation negatively and is not optimal. This effect limits the room for rent extraction through nonlinear pricing.

The results in Figure 20 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? To answer this question, I calculate the real responses - intensive and participation - to the schedule in Equation (49). I keep aggregate capacity constant, which allows for a meaningful comparison to the other experiments. I find that the schedule in Equation (49) would increase costs by 3.7 per cent instead of decreasing them. The experiment shows that ignoring participation would have a sizable adverse effect on costs. Therefore, it is crucial to take the participation margin into account.

8 Conclusion

This paper proposes an estimator, which simultaneously estimates the intensive and the participation margin response of solar panel adopters to the nonlinear subsidy. I find that reactions along both margins are important. The nonlinearities in the current schedule are not effective in reducing the public costs of the programme. The optimal subsidy schedule is very close to linear. Not taking the participation margin into account when optimizing the policy has a sizable adverse effect on costs. The results are driven by adopters' large participation margin reactions, limiting the scope of cost-savings through price discrimination.

The method proposed here is applicable whenever the bunching estimator is insufficient because there is a participation margin response of interest, or the regression kink

design cannot be used because there are reactions along both margins. Kinked incentive schemes are a popular form of mechanisms, and often, both margins are present and of interest. Governments or other principals have to implement nonlinear mechanisms without having detailed information on the agents' behaviour. The limited information results in suboptimal choices, even within the limited class of piecewise linear schemes. The method outlined in this paper can be used to increase welfare in these situations. Possible further applications are problems of optimal taxation, optimal nonlinear pricing, and policy evaluation.

The method is not limited to kinked incentive schemes. Close relatives of the bunching estimator using kinks and the regression kink design are the bunching estimator using notches and the regression discontinuity design. These methods are applicable when there is a discrete jump in the incentive scheme instead of a kink. However, whenever both margins are present, the problems are the same as with the kink bunching estimator and the regression kink design: the notch bunching estimator cannot estimate both margins; the regression discontinuity design is inapplicable when there is an intensive margin response. An extension of the method of this paper to discontinuous incentive schemes is straightforward.

References

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

BACHAS, P. J. AND M. SOTO (2018): "Not (ch) Your Average Tax System: Corporate Taxation Under Weak Enforcement," Tech. Rep. 8524, World Bank Policy Research Working Paper.

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., A. H. MCCALLUM, AND N. SEEGERT (2019): "Better Bunching, Nicer Notching," Tech. rep., nathanseegert.com.

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., W. NEWEY, A. KUMAR, AND C.-Y. LIANG (2017): "On Bunching and Identification of the Taxable Income Elasticity," Working Paper 24136, National Bureau of Economic Research.

BURR, C. (2016): "Subsidies, Tariffs 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," .

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.

FEGER, F., N. PAVANINI, AND D. RADULESCU (2017): "Welfare and redistribution in residential electricity markets with solar power," Tech. rep., CEPR Discussion Paper No. DP12517.

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

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. (2017): "Demanding innovation: The impact of consumer subsidies on solar panel production costs," Tech. rep., Working paper, Harvard University.

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*.

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.

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.

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

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.

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.

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

——— (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, H. SHAN, AND G. SACHS (2012): "The Lock-In Effect of Housing Transfer Taxes: Evidence From a Notched Change in DC Policy," *University of Michigan: Michigan*.

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, Accessed: 2017-09-30.

A Appendix

A.1 Lemma 1

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

PROOF Lemma 1:

Assumption 1a $\Rightarrow$ Assumption 1b:

By Assumption 1a and the definition of the elasticity

$$\epsilon = \frac{q^{i\prime}(s)\, s}{q^i(s)}, \text{ for all } s \text{ in } [s_l\rho, s_l]. \tag{50}$$

By the first order condition of the adopters' problem:

$$c_v^{i\,\prime}(q^i(s)) = s, \tag{51}$$

and by differentiating the FOC

$$q^{i\prime}(s) = \frac{1}{c_v^{i\,\prime\prime}(q^i(s))}. \tag{52}$$

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

$$\epsilon = \frac{c_v^{i\,\prime}(q)}{c_v^{i\,\prime\prime}(q)\, q}. \tag{53}$$

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 (53) form an initial value problem with solution

$$c_v^i(q) + c_f^i = \underbrace{\frac{s_l}{q_l^{i\,\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}}}_{\text{variable cost}} + \underbrace{c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}}_{\text{fixed cost}}. \tag{54}$$

The result follows by defining

$$\theta^i = \frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon}, \tag{55}$$

$$c_f^i = c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}. \tag{56}$$

Assumption 1b $\Rightarrow$ Assumption 1a:

By the FOC

$$q^i(s) = \left(\frac{\epsilon}{(1+\epsilon)\theta^i} \right)^{\epsilon} s^{\epsilon}. \tag{57}$$

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

A.2 Proof Corollary 1

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 (7) gives the result.
qed.

A.3 Lemma 2

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

PROOF:

Assumption 2a $\Leftrightarrow$ Assumption 2b:

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

$$\eta = \frac{f_{t|q_l}(S(q)|q)S(q)}{F_{t|q_l}(S(q)|q)}. \tag{58}$$

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

$$\eta = \frac{f_{t|q_l}(c_t|q_l)c_t}{F_{t|q_l}(c_t|q_l)}. \tag{59}$$

The solution of this partial differential equation is

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)}\right)^{\eta}, \tag{60}$$

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

qed.

A.4 Proof of Proposition 1

The variable q_l is the choice of an adopter under the linear subsidy and q_k is the choice of the same adopter under the kinked subsidy.

Lemma 3. *The choice under the kinked subsidy q_k as a function of the choice under the linear subsidy q_l is*

$$q_k(q_l) = q_l, \qquad \textit{for } q_l < q^K; \tag{61}$$

$$q_k(q_l) = q^K, \qquad \textit{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; \tag{62}$$

$$q_k(q_l) = q_l\rho^{\epsilon}, \qquad \textit{for } q_l > q^K\rho^{-\epsilon}. \tag{63}$$

PROOF:
By Equation (12) and the first order condition of the adopters' maximization problem

$$q(s, q_l) = q_l \left(\frac{s}{s_l} \right)^{\epsilon}. \tag{64}$$

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

$$q_k(q_l) = q_l, \text{ for } q_l < q^K. \tag{65}$$

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

$$q_k(q_l) = q_l\rho^{\epsilon}, \text{ for } q_l \gg q^K. \tag{66}$$

In general, 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 4. *The difference in cost $\Delta c(q_l)$ of adopter q_l between the kinked and linear subsidy is*

$$\Delta c(q_l) = 0, \qquad \textit{for } q_l < q^K; \tag{67}$$

$$\Delta c(q_l) = \frac{1}{1+\epsilon^{-1}} \frac{s_l}{q_l^{1/\epsilon}} ((q^K)^{1+\epsilon^{-1}} - q_l^{1+\epsilon^{-1}}), \qquad \textit{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; \tag{68}$$

$$\Delta c(q_l) = \frac{1}{1+\epsilon^{-1}} s_l q_l (\rho^{\epsilon+1} - 1), \qquad \textit{for } q_l > q^K\rho^{-\epsilon}. \tag{69}$$

PROOF:

By Corollary 1

$$c(q, q_l, c_t) = c_t + \left[\frac{q^{1+1/\epsilon}}{q_l^{1/\epsilon}} - q_l \right] \frac{s_l}{1 + 1/\epsilon}. \tag{70}$$

By definition and Lemma 3

$$\begin{aligned} \Delta c(q_l) = c(q_k(q_l), q_l, c_t) - c_t &= c(q_l, q_l, c_t) - c_t && \text{for } q_l < q^K; && (71)\\ &= c(q^K, q_l, c_t) - c_t && \text{for } q_l \in [q^K, q^K \rho^{-\epsilon}]; && (72)\\ &= c(q_l \rho^{\epsilon}, q_l, c_t) - c_t && \text{for } q_l > q^K \rho^{-\epsilon}. && (73) \end{aligned}$$

Use Equation (70) in Equation (71)-(73) to get Equation (67)-(69).
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:

$$R(q_l) = \frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}. \tag{74}$$

Lemma 5. *The function $R(q_l)$ is:*

$$\begin{aligned} R(q_l) &= 1, && \textit{for } q_l < q^K; && (75)\\ R(q_l) &= \frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 - \left(\frac{q^K}{q_l}\right)^{\frac{1+\epsilon}{\epsilon}}\right), && \textit{for } q_l \in [q^K, q^K \rho^{-\epsilon}]; && (76)\\ R(q_l) &= (1-\rho)\frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 + \frac{\rho^{\epsilon+1}}{\epsilon}\right), && \textit{for } q_l > q^K \rho^{-\epsilon}. && (77) \end{aligned}$$

PROOF:
By definition

$$R(q_l) = \frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}, \tag{78}$$

which together with Lemma 3 and 4 gives Equation (75)-(77).
qed.

Lemma 6. *The mass of participating adopters under the kinked subsidy as a function of q_l is*

$$R(q_l)^{\eta} f_l(q_l), \tag{79}$$

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

$$\frac{F_{t|q_l}(S_k(q_k(q_l)) - \Delta c(q_l)|q_l)}{F_{t|q_l}(S_l(q_l)|q_l)} f_l(q_l), \tag{80}$$

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

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)}\right)^{\eta}, \text{ for all } c_t \text{ in } [S_k(q_l), S_l(q_l)]. \tag{81}$$

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

$$\frac{F_{t|q_l}(S_k(q_k(q_l)) - \Delta c(q_l)|q_l)}{F_{t|q_l}(S_l(q_l)|q_l)} f_l(q_l) = \left(\frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}\right)^{\eta} f_l(q_l), \tag{82}$$

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

PROOF of Proposition 1:
Change variable in Equation (79) using Lemma 3 to derive Equation (15) and (17). Integrate Equation (79) over $\left[q^K, q^K \rho^{-\epsilon}\right]$ to derive Equation (18).
qed.

A.5 Proof of Proposition 2

By Assumption 3

$$g_f(f_l(q)) = \sum_{p=0}^{\infty} \gamma_p \frac{1}{p!} \left(g_q(q) - g_q(q^K)\right)^p, \tag{83}$$

where g_f is the transformation of f_l and g_q is the transformation of q. I assume the transformation is known. I discuss the choice of transformation in the estimation stage. The measure f_l is identified because all γ_p are identified from the left limit

$$\gamma_p = \lim_{q_- \to q^K} \frac{\mathrm{d}^p g_f(f_k(q))}{\mathrm{d}(g_q(q) - g_q(q^K))^p}, \text{ for all } p. \tag{84}$$

The elasticities ϵ and η are jointly identified as the simultaneous solution to Equation (18) and (17). By Condition 1, the solution is locally unique. 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.

A.6 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's specific location and capacity. The location matters, since climate conditions vary across locations. Moreover, adopters may have heterogeneous discount rates when evaluating future streams of income. Discounting matters because adopters take the adoption decision based on the present discounted value of the income stream produced by the solar panel.

All these factors transform the decision problem of adopters to

$$\pi_v^i = \max_q \{\zeta^i S(q) - c_v^i(q)\}, \tag{85}$$

and participate if and only if

$$\pi_v^i \geq c_f^i, \tag{86}$$

where ζ^i is an additional individual-specific factor that depends on individual-specific discounting and location. The derivation of Equation (85) is in Section A.6.1 below. Normalization by ζ^i shows the equivalence of Problem (85) and Problem (1). Therefore, the model outlined in Section 3 encompasses individual-specific discounting and location.

A.6.1 Derivation Equation (85)

A household installing capacity q^i produces electricity e_{it} in a given year:

$$e_{it} = w_{it} q^i, \tag{87}$$

where w_{it} is the efficiency of the panel in year t which depends on weather conditions and the location. The electricity e_{it} is remunerated according to the following subsidy scheme, which depends on the installed capacity:

$$S_k(q, e_{it}) = s_l \; e_{it}, \qquad \text{for } q \leq q^K; \tag{88}$$

$$S_k(q, e_{it}) = s_l \; e_{it} \frac{q^K}{q} + s_l \rho e_{it} \frac{q - q^K}{q}, \qquad \text{for } q > q^K. \tag{89}$$

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

$$S_k(q, w_{it}) = s_l \; w_{it} \; q, \qquad \text{for } q \leq q^K; \tag{90}$$

$$S_k(q, w_{it}) = s_l w_{it} q^K + s_l \rho w_{it} (q - q^K), \qquad \text{for } q > q^K. \tag{91}$$

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

$$\mathbb{E}_i\left[\sum_{t=0}^{20}\beta_i^t w_{it} S_k(q)\right] = S_k(q)\mathbb{E}_i\left[\sum_{t=0}^{20}\beta_i^t w_{it}\right] = S_k(q)\zeta^i. \tag{92}$$

The subsidy is payed for 20 years and I assume that panels break down after that period. Note that the discount factor, weather conditions, and the expectations may be agent specific. I assume agents are risk neutral. This assumption is reasonable for two reasons. Random weather conditions are observable and we can expect that financial markets are complete. Risk averse adopters can insure against weather fluctuations. For small adopters the income fluctuation generated from random solar earnings is very small compared to other income shocks. We expect them to self insure against these shocks using precautionary savings. The adopter's decision problem is

$$\pi_v^i = \max_q\{\zeta^i S_k(q) - c_v^i(q)\}, \tag{93}$$

and participate if and only if

$$\pi_v^i \geq c_f^i, \tag{94}$$

which is Problem (85). Since the individual factor ζ^i is multiplicative, normalize Problem (85) to reduce it to Problem (1).

A.7 Convergence of $\ln \widehat{f(q)}$, Lemma 7:

Denote by n the sample size and denote by Q_ι a single observation. The variable q_j is the center 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

$$\widehat{f(q_j)} = \frac{\widehat{N_j}}{nh_j}. \tag{95}$$

Lemma 7. *$\widehat{f(q_j)}$ converges to $f_k(q_j)$ in probability, which is also true in logarithms:*

$$\widehat{f(q_j)} \xrightarrow{p} f_k(q_j); \tag{96}$$

$$\ln \widehat{f(q_j)} \xrightarrow{p} \ln f_k(q_j). \tag{97}$$

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

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

$$\widehat{f(q_j)} = \frac{1}{n}\sum_{\iota=1}^{n}\frac{\mathbb{1}[q_j - h_j/2 \leq Q_\iota < q_j + h_j/2]}{h_j}. \tag{98}$$

Using the law of large numbers it follows that

$$\frac{1}{n}\sum_{\iota=1}^{n}\frac{\mathbb{1}[q_j - h_j/2 \leq Q_\iota < q_j + h_j/2]}{h_j} \xrightarrow{p} \frac{1}{h_j}\int_{q_j-h_j/2}^{q_j+h_j/2} f_k(q)dq. \tag{99}$$

As h_j goes to zero when n goes to infinity

$$\frac{1}{h_j}\int_{q_j+h_j/2}^{q_j-h_j/2} f_k(q)dq = f_k(q_j). \tag{100}$$

By the continuous mapping theorem it follows that

$$\ln\widehat{f(q_j)} \xrightarrow{p} \ln f_k(q_j). \tag{101}$$

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

A.8 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. I use bias correction techniques to correct for this bias. Instead of directly using $\ln\widehat{f(q_j)}$ I use

$$\widehat{\ln f(q_j)} = \ln\widehat{f(q_j)} + \frac{1}{2\widehat{N}_j}. \tag{102}$$

Note that $\widehat{\ln f(q_j)}$ denotes the bias corrected variable, while $\ln\widehat{f(q_j)}$ denotes the logarithm of the histogram. Using a Taylor approximation of the functions of random variables

around the expected value of the random variable it follows that

$$\mathbb{E}\left[\widehat{\ln f(q_j)}\right] = \ln f_k(q_j) + O\left(\frac{1}{N_j^2}\right), \tag{103}$$

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

PROOF:

Taylor approximate all functions of random variables in Equation (102) around their expected values and use $f_k(q_j)\ n\ h_j = N_j$:

$$\begin{aligned}\widehat{\ln f(q_j)} \approx &\ln f_k(q_j) + \frac{1}{N_j}(\widehat{N_j} - N_j) - \frac{1}{2N_j^2}(\widehat{N_j} - N_j)^2 + \frac{1}{3N_j^3}(\widehat{N_j} - N_j)^3 + \\ &+ \frac{1}{2N_j} - \frac{1}{2N_j^2}(\widehat{N_j} - N_j) + \frac{1}{2N_j^3}(\widehat{N_j} - N_j)^2.\end{aligned} \tag{104}$$

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.

A.9 Asymptotic normality

The histogram is asymptotically normal:

$$\lim_{n\to\infty} \sqrt{nh_j(n)}\left(\widehat{f(q_j)} - f_k(q_j)\right) \sim N(0, f_k(q_j)). \tag{105}$$

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

$$\lim_{n\to\infty} \sqrt{nh_j(n)}\left(\widehat{\ln f(q_j)} - \ln f_k(q_j)\right) \sim N(0, \frac{1}{f_k(q_j)}). \tag{106}$$

I choose the binning function such that the variance of the log-histogram is 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

$$\lim_{n\to\infty} \sqrt{nh_0(n)}\left(\widehat{\ln f(q_j)} - \ln f_k(q_j)\right) \sim N(0, 1). \tag{107}$$

Write

$$\widehat{\ln f(q_j)} = \ln f_k(q_j|\epsilon,\eta) + u_j \tag{108}$$

where

$$u_j \sim N(0, \frac{1}{nh_0(n)}). \tag{109}$$

The errors u_j are approximately uncorrelated for n large enough. Chen (2007) shows that the nonlinear least square estimate of Equation (108), where the nonparametric function $\ln f_k(q_j|\epsilon,\eta)$ 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 $\hat{\eta}$.

In practice I use the following binning function

$$\overline{q_j} = q^K(1 - jc_0(n))^{-\frac{1}{\omega}}, \tag{110}$$

where $\overline{q_j}$ is the left border of a bin and $c_0(n)$ is a constant that goes to zero as n goes to infinity. The justification for this binning function is the following. 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

$$\phi_0 \left(\frac{\overline{q_j}}{q^K}\right)^{-\phi_1} - \phi_0 \left(\frac{\overline{q}_{j+1}}{q^K}\right)^{-\phi_1}, \tag{111}$$

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

A.10 Estimation of the mean squared error

The bias and the variance of the estimates depend on the order of the polynomial P and the bandwidth b. It is standard in nonparametric estimations to choose a specification that minimizes 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. Denote by $z = (P, b)$ the vector of specification parameters. The estimate $\hat{\eta}(z, n)$ is a function of the specification-parameters

and the sample size n. The mean squared error is defined as

$$MSE = \mathbb{E}(\hat{\eta}(z,n) - \eta)^2, \tag{112}$$

where η is the true value. Define

$$\tilde{\eta}(z) := \lim_{n\to\infty} \hat{\eta}(z,n), \tag{113}$$

where z is kept constant. Note that if z changes accordingly with sample size, $\tilde{\eta}$ converges to the true value η. However, that is not true if the specification is kept constant. In a sense, $\tilde{\eta}$ is the true value of the parametric estimation. As I use bias correction methods for all small-sample biases arising from the parametric part of the estimation, it is true that $\mathbb{E}\hat{\eta}(z,n) \approx \tilde{\eta}(z)$. It follows that the mean squared error is equal to

$$MSE = \underbrace{\mathbb{E}(\hat{\eta}(z,n) - \tilde{\eta}(z))^2}_{\text{variance}} + \underbrace{(\tilde{\eta}(z) - \eta)^2}_{\text{bias}^2}. \tag{114}$$

The mean squared error is the sum of the "parametric variance" and the estimator's square-bias. Note that all parts of the expression are unknown and need to be estimated. The bootstrap provides an estimate for the variance. The main challenge is that to estimate the bias, one needs an estimate of $\tilde{\eta}$ and the true value η. An unbiased and consistent estimate of $\tilde{\eta}$ is the estimate $\hat{\eta}$ itself. The big challenge is to estimate the true value η. I use an out of sample estimation on untreated data to estimate the bias. This approach has the advantage that absent the treatment, the true value of η is known and equal to zero. The estimate of the bias is

$$\widehat{bias(\hat{\eta})} = \hat{\eta}_{nt}, \tag{115}$$

where $\hat{\eta}_{nt}$ is the estimate of the participation margin on untreated data. This approach relies on the assumption that absent the treatment, the treated data is similar to the untreated data. Intuitively, any effect estimated on the untreated data is a bias due to the specification.

More specifically, the distributions are similar if the two counterfactual distributions of the treated and untreated data are similar. Suppose there exists a certain order of the series expansion of the two distributions, such that all coefficients above that order are equal. In that case, one can use the estimated coefficient on the untreated data to estimate the bias on

the treated data.[20] Mathematically, by Assumption 3

$$g_f(f_l^x(q)) = \sum_{p=0}^{\infty} \gamma_p^x \frac{1}{p!} \left(g_q(q) - g_q(q^K)\right)^p, \tag{116}$$

where x is either t (treated) or nt (untreated) and denotes the respective counterfactual measure. Assume there exists an order j^* such that for all $j \geq j^*$ the coefficients are equal: $\gamma_j^t = \gamma_j^{nt}$. In this case, the estimated coefficient $\widehat{\eta}^{nt}$ is an estimate for the bias of the coefficient $\widehat{\eta}^t$. I use observations from 2000 to 2003 and observations far from the kink points as untreated data. The details of the estimation are in Section A.12.

A.11 The choice of the series expansion

In this section, I discuss the choice of the series extension for $f_l(.)$. By Assumption 3

$$g_f(f_l(q)) = \sum_{p=0}^{\infty} \gamma_p \frac{1}{p!} \left(g_q(q) - g_q(q^K)\right)^p, \tag{117}$$

where $g_f(.)$ is the transformation of $f_l(.)$ and $g_q(.)$ is the transformation of q. The goal is to find transformations $g_f(.)$ and $g_q(.)$ such that the Series (117) converges fast. In order to choose the transformations, I regress a low order approximation of the above series on the observed histogram of capacities on pre-treatment data and on treated data excluding the kink points.

Formally, I regress

$$g_f(\widehat{f(q_j)}) = \sum_{p=0}^{P} \gamma_p \frac{1}{p!} \left(g_q(q_j) - g_q(q^K)\right)^p + u_j, \tag{118}$$

where u_j is the error term. I run the regression for $P = 1$ and $P = 2$, which gives a low order approximation of the Series (117). I consider three combinations of transformations:

1. id-id transformation: $g_f(.)$ and $g_q(.)$ are equal to the identity function (no transformation).

[20] An additional problem arises from eventual nonlinearities in the estimation. Strictly speaking, the above statement is only correct for linear estimations. However, because my estimation is close to linear, I neglect this problem.

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.

I run the estimations on the pre-treatment data containing observations from 2000 to 2003. I use data in the capacity-range 10 to 90, because it corresponds to the optimal bandwidth selected in Section A.12. For the treated data, I exclude capacities close to the kink points. Consistent with the optimal bandwidths selected in Section A.12, I run the regression on the intervals $[10, 26.5]$, $[35, 95]$, and $[105, 4000]$. I use R^2 as a measure of fit. Tables 8-11 summarize the results.

Table 8: R^2 for untreated data in the years 2000-2003 and capacity range $[10, 90]$

Order of the series P	Transformation	R^2
P=1	id-id	0.4938
	log-id	0.8120
	log-log	0.8811
P=2	id-id	0.6911
	log-id	0.8769
	log-log	0.8812

Note: Number of bins: N=57. Number of observations: n=4,979.

Table 9: R^2 for treated data in the years 2004 to 2008 and capacity range $[10, 26.5]$

Order of the series P	Transformation	R^2
P=1	id-id	0.3254
	log-id	0.3647
	log-log	0.3982
P=2	id-id	0.4265
	log-id	0.4205
	log-log	0.4343

Note: Number of bins: N=131. Number of observation: n=73,693.

Table 10: R^2 for treated data in the years 2004 to 2008 and capacity range $[35, 95]$

Order of the series P	Transformation	R^2
P=1	id-id	0.5631
	log-id	0.7659
	log-log	0.7928
P=2	id-id	0.7031
	log-id	0.7966
	log-log	0.7968

Note: Number of bins: N=210. Number of observation: n=10,263.

Table 11: R^2 for treated data in the years 2004 to 2008 and capacity range $[105, 4000]$

Order of the series P	Transformation	R^2
P=1	id-id	0.2323
	log-id	0.3655
	log-log	0.5444
P=2	id-id	0.3630
	log-id	0.4679
	log-log	0.5535

Note: Number of bins: N=63. Number of observation: n=1,515.

The log-log transformation has the highest measure of fit for all specifications. It outperforms the other transformations particularly well for $P = 1$ and when the capacity-range is large.

A.12 Optimal specification

I use the estimates from the pooled sample of the years 2004 to 2008 for the counterfactual analysis. Therefore, I choose the optimal specification for this time-range. The two kink points are at 30 kWp and 100 kWp. For simplicity, I use this specification also for the yearly estimates in Table 3. 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. I check for the robustness of this choice in section A.12.4

As discussed in Section A.10, for each kink point, I need a range of untreated data to estimate the biases of the main estimates. A natural choice is observations in the years

2000 to 2003. In these years, the subsidy was linear. Therefore, there should be no detectable impact of the kink on the distribution of capacities when estimating the model. Any detected impact is an estimate of the bias coming from the specification. This bias is a valid estimate for the main estimate's bias as long as the counterfactual distributions are similar in the sense discussed in Section A.10. I test the similarity of the two distributions in Section A.12.3.

For the kink point a 100 kWp, the pre-treatment data is not a satisfactory choice to estimate the bias. It is due to two reasons. First, the number of observations above 100 kWp is very low in these years. Second, in the years 2000-2003, I do not observe if a solar panel is installed on a rooftop or the ground. From 2004 onwards, I observe where a panel is installed. I do only consider rooftop panels in my analysis. Ground panels are only a very small share of overall installations. Also, the subsidy for ground panels is linear in all years. In 2004, I observe that close to capacity 30, only very few panels are ground panels. Therefore, the fact that the sample from 2000-2003 contains ground panels does not pose a big problem for using these years to estimate the bias discussed above. 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. It indicates that there is round number bunching for ground panels. For these reasons, I cannot use observations around 100 kWp in the years 2000-2003 for the estimation of the bias.

Alternatively, I will use observations around a point similar to 100 kWp in the years 2004-2008. On the one hand, for the counterfactual distribution to be similar, the point should be close enough to 100 kWp. On the other hand, it should be far enough from 100 to not 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 (it is a quarter of 1,000 kWp). As discussed in Section A.10, the estimate of the bias is valid if the higher-order terms in the counterfactual distribution are equal above a certain order of the series. For observations from the same sample at different points, this property can only be true if the terms above a certain order a zero. I test this hypothesis in Section A.12.3.

A.12.1 The binning function

As discussed in Section A.9, the binning function should guarantee a constant variance and sufficiently many observations in each bin. I use the binning function in Equation (110). To select its parameters, I preselect a maximal range of bandwidths for the analysis. For the estimation at 30 kWp, I use a maximal bandwidth of $[9.5, 95]$. The interval is symmetric around 30 in 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 $[42, 1600]$. The proportional interval around 250 kWp is $[105, 4000]$.[21] 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. To increase sample size, I use asymmetric intervals.

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 five observations in each bin, and the variance of the residuals of an auxiliary estimation is constant. By Equation (103), five observations guarantee that the bias introduced by the logarithm is of the order $\frac{1}{25}$. The procedure gives the following specification:

Table 12: Selected bin size h_0 and scaling parameter ω.

Years	Interval	Bin size h_0	Scaling parameter ω
2004-2008	[9.5, 95]	0.18	-0.35
	[42, 1600]	1.7	1.2
	[105, 4000]	10	1.5
2000-2003	[9.5, 95]	1.2	0.35

Note: The table shows the selected bin size h_0 and scale parameter ω of the binning function in Equation (110). The parameters guarantee a minimum of 5 observations in each bin and a variance that is approximately constant.

A.12.2 Estimates of the mean squared error

As discussed above, I estimate the mean squared error by using the variance estimate from the treated sample and the bias estimate from the untreated sample. To estimate the variance, I use the nonparametric bootstrap. The untreated sample for the estimation at 30

[21] $42\frac{250}{100} = 105$ and $1600\frac{250}{100} = 4000$.

kWp is the sample with the same bandwidth in the years 2000 to 2003. Table 13 shows the estimated mean squared error (MSE) for different bandwidths and orders P of the series.

Table 13: Estimated mean squared error of $\widehat{\eta_{30}}$

Bandwidth	MSE, P=1	MSE, P=2	MSE, P=3
[16, 55]	113	14,722	1,013
[15, 60]	504	6,906	457
[14, 65]	146	1,491	1,176
[13, 70]	299	300	184
[12, 75]	195	142	598
[11, 80]	149	572	209
[10.5, 85]	13	1,230	2,479
[10, 90]	**1.8**	1,079	2,501
[9.5, 95]	546	546	3,831

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 mean squared error is the lowest for $P = 1$ and $b = [10, 90]$.

The optimal bandwidth for the estimation at 30 kWp is $[10, 90]$; the optimal order of the series is $P = 1$.

For the kink at 100 kWp, I use data around the point 250 kWp as untreated data. I use a proportional bandwidth for the estimation by multiplying the lower bound and the upper bound of the bandwidth used at 100 with $\frac{250}{100}$. Table 14 shows the estimated mean squared error (MSE) for different bandwidths and orders P of the series.

Table 14: Estimated mean squared error of $\widehat{\eta_{100}}$

Bandwidth	MSE, P=1	MSE, P=2	MSE, P=3
[67, 150]	5,782	36,975	46,739
[50, 200]	1,544	16,171	20,575
[42, 400]	3,276	7,952	11,771
[42, 700]	5,528	1,973	10,743
[42, 1000]	5,904	1,434	15,163
[42, 1300]	4,746	544	**163**
[42, 1600]	4,765	569	226

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 mean squared error is the lowest for $P = 3$ and $b = [42, 1300]$.

The optimal bandwidth for the estimation at 100 kWp is $[42, 1300]$; the optimal order of the series is $P = 3$.[22]

For both kink points, the optimal bandwidth is relatively large, and the series' optimal order is low. It shows that the counterfactual distribution is very close to a Pareto distribution.

A.12.3 Test untreated data

For the estimation at 30 kWp, the optimal order of the series is $P = 1$. To check if the untreated counterfactual in the years 2000 to 2003 can be used to estimate the estimate's bias in 2004 to 2008, I test if the second-order parameters γ_2^t and γ_2^{nt} of the two series are not significantly different in both samples. The indices t and nt denote the parameter of the treated and untreated data, respectively. I restrict the elasticities to their optimally estimated values and use a series of order $P = 2$ in an additional estimation. I cannot find a significant different between the two estimates: $\widehat{\gamma_2^t} - \widehat{\gamma_2^{nt}} = 0.125\ (0.09)$. In brackets is the standard error of the difference. The p-value for testing the null hypothesis $H_0 : \gamma_2^t - \gamma_2^{nt} = 0$ is 0.184. The hypothesis cannot be rejected.

As the discussed above, to use the data around 250 kWp, the last significant parameters in both series need to be equal. Table 15 shows the estimates of γ for the treated and the untreated data using the optimal specification:

Table 15: Estimates of the series parameters γ for the estimation at 100 kWp

Kink Point	$\widehat{\gamma_1}$	$\widehat{\gamma_2}$	$\widehat{\gamma_3}$
100 kWp (treated)	-2.8 (0.06)	-0.06 (0.06)	0.03 (0.04)
250 kWp (untreated)	-2.9 (0.4)	0.2 (0.3)	0 (0.1)

Note: The table shows the estimates of the parameters of the series $\hat{\gamma}$. The last significant estimates in the two series need to be equal to use data from the point around 250 kWp as untreated data for the estimation at 100 kWp. The table shows that the last significant parameter in both series is $\widehat{\gamma_1}$. The two estimates of $\widehat{\gamma_1}$ are not significantly different from each other.

The two parameters $\widehat{\gamma_1}$ are not significantly different from each other. All other parameters are not significantly different from 0. The distribution fulfils the requirement.

[22] Note that for the estimation of the variance the estimate $\hat{\eta}$ is restricted to be positive. Theoretically, the elasticity η cannot be negative. Therefore, the restriction reduces the variance of the estimate.

A.12.4 Robustness bunching interval

This section reports the robustness of the estimates to changes in the bunching interval.

Table 16: Estimates at 30 kWp for various bunching intervals

Bunching Window	$\widehat{\eta}$	$\widehat{\epsilon}$
[25.50, 32.250]	65.9 (1.3)	4.57 (0.11)
[25.75, 32.125]	65.1 (1.2)	4.76 (0.12)
[26.00, 32.000]	65.3 (1.6)	4.71 (0.12)
[26.25, 31.875]	64.8 (1.4)	4.78 (0.12)
[26.50, 31.750]	65.0 (1.4)	4.76 (0.11)
[26.75, 31.625]	65.8 (1.3)	4.60 (0.10)
[27.00 31.500]	65.2 (1.3)	4.72 (0.10)
[27.25, 31.375]	66.6 (1.4)	4.35 (0.10)
[27.50, 31.250]	67.3 (1.4)	4.22 (0.09)

Note: The table shows estimates of the two elasticities for different bunching intervals. The estimates do only react weakly to changes in the bunching interval.

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

Table 17: Estimates at 100 kWp for various bunching intervals

Bunching Window	$\widehat{\eta}$	$\widehat{\epsilon}$
[90, 105.0]	0.1 (3)	5.8 (1.0)
[91, 104.5]	0 (1)	5.5 (0.9)
[92, 104.0]	0 (1)	5.5 (0.9)
[93, 103.5]	0.1 (2)	5.6 (0.9)
[94, 103.0]	0.9 (8)	4.4 (0.8)
[95, 102.5]	1.1 (10)	4.4 (0.8)
[96, 102.0]	0.9 (8)	4.5 (0.7)
[97, 101.5]	0.6 (6)	4.9 (0.6)
[98, 101.0]	0.6 (6)	4.9 (0.6)

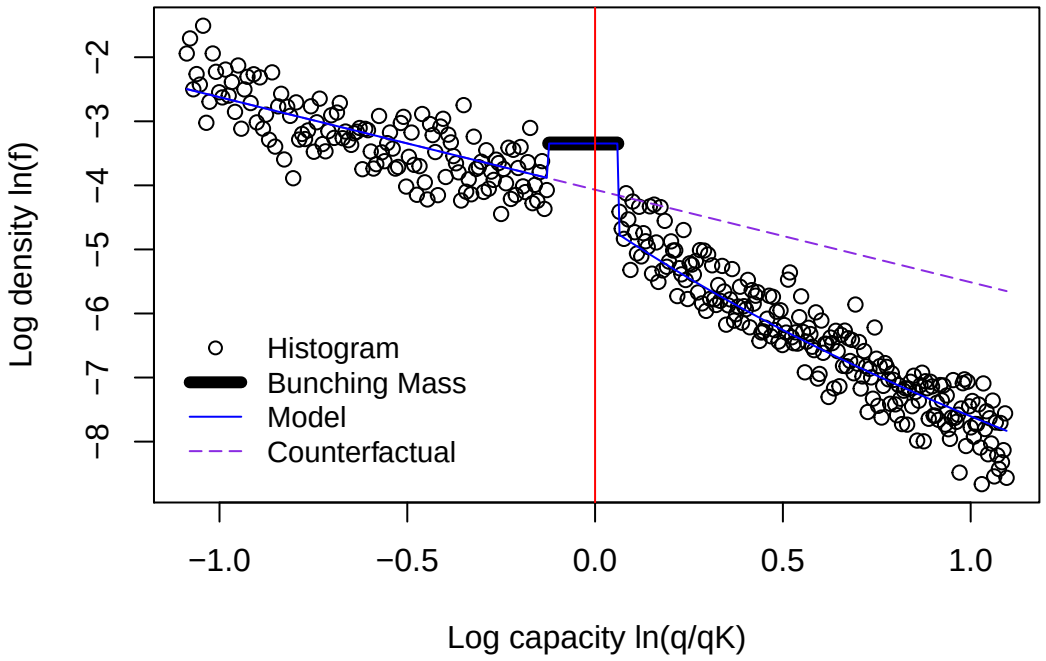
Note: The table shows estimates of the two elasticities for different bunching intervals. The estimates do only react weakly to changes in the bunching interval.

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

A.13 Estimation aggregate data years 2004 to 2008

The optimal specification, derived in Section A.12, for the estimation at 30 kWp is: bin-size $h_0 = 0.18$, scaling parameter $\omega = -0.35$, bandwidth $b = [10, 90]$, bunching window $[26.5, 31.74]$, and series order $P = 1$. Figure 21 shows the histogram with the estimated model and the counterfactual. The results are in Table 4.

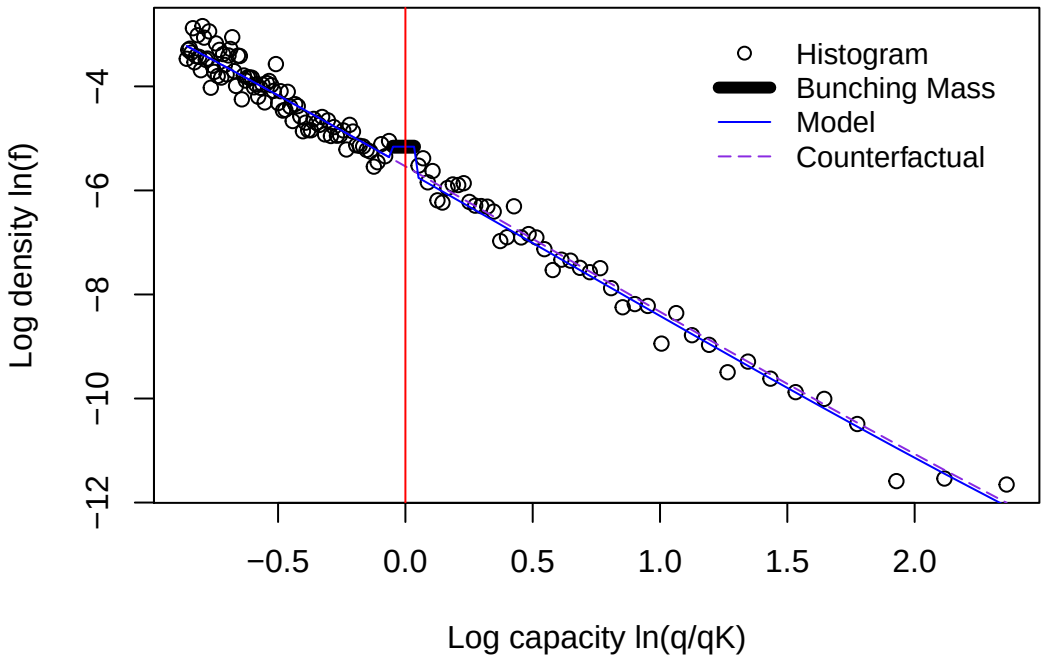
Figure 21: Histogram in the years 2004 to 2008 at 30 kWp with estimated model and counterfactual



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

The optimal specification, derived in Section A.12, for the estimation at 100 kWp is: bin-size $h_0 = 1.7$, scaling parameter $\omega = 1.2$, bandwidth $b = [42, 1600]$, bunching window $[95, 102.5]$, and series order $P = 3$. Figure 22 shows the histogram with the estimated model and the counterfactual. The results are in Table 4.

Figure 22: Histogram in the years 2004 to 2008 at 100 kWp with estimated model and counterfactual

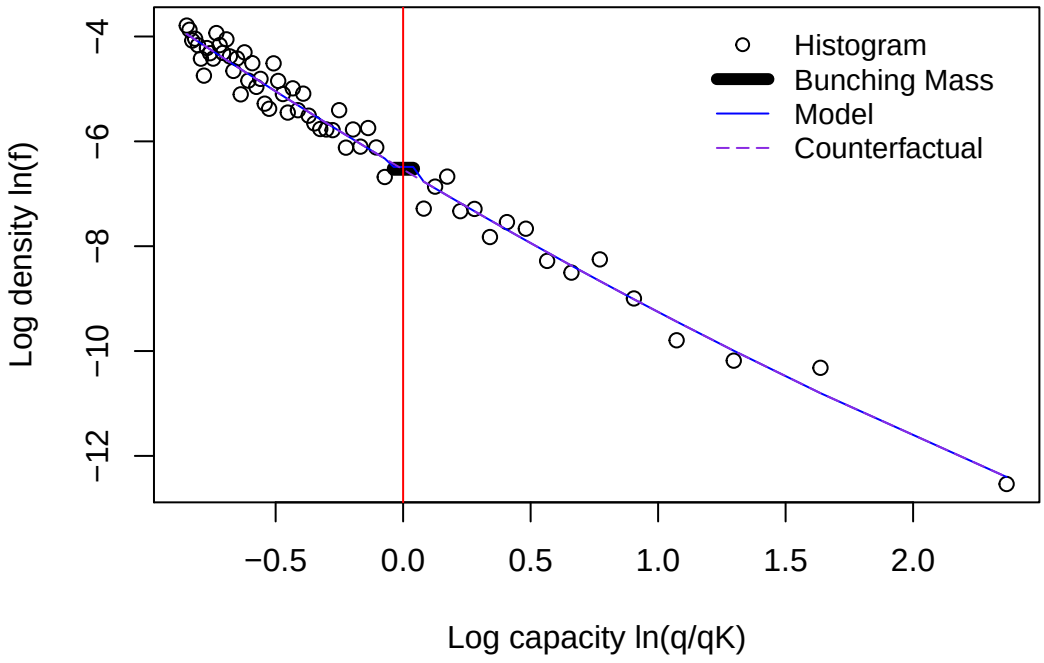


Note: The x-axis shows the normalized logarithm of capacity. The y-axis shows the logarithm of the density. The red line marks the kink point. The estimation minimizes the distance between the data in black and the model in blue.

A.14 Robustness check 100 kWp

I run the robustness check for the estimation at 100 kWp on the data around 250 kWp in the years 2004 to 2008. For a discussion of this choice, see Section A.12.

Figure 23: Histogram at 250 kWp with estimated model



Note: The figure shows the robustness check. The x-axis shows the normalized 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 18: Results for the untreated data at 250 kWp

Data	ϵ (σ_ϵ)	η (σ_η)
250 kWp	0 (1)	0 (99)

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

Both estimates are insignificant.

A.15 The general welfare function and the optimal subsidy

Assume the utility of an adopter is equal to

$$u\left(\left(S(q)-c_v(q,q_l)-c_f\right)\mathbb{1}\left(S(q)-c_v(q,q_l)\geq c_f\right)+y-T(y)-c_l(y,a)\right), \tag{119}$$

where $u(.)$ 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 variable y denotes other income such as labour income. The function $T(.)$ is an income tax, and $c_l(y,a)$ is the cost of producing income y for an agent with ability a. The type θ of an adopter is three dimensional: $\theta=(q_l,c_f,a)$ with density $f_\theta(\theta)$. The symbol $\mathbb{1}(.)$ denotes the indicator function. Agents only adopt if they make positive profit. 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 which determine 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:

$$\max_{S(.)}\int G\Big(\left(S(q)-c_v(q,q_l)-c_f\right)\mathbb{1}\left(S(q)-c_v(q,q_l)\geq c_f\right)+ \\ +y-T(y)-c_l(y,a)\Big)f_\theta(\theta)d\theta+V(Q), \tag{120}$$

such that

$$\int q\;\mathbb{1}\left(S(q)-c_v(q,q_l)\geq c_f\right)\;f_\theta(\theta)d\theta=Q, \tag{121}$$

and

$$\int T(y)-S(q)\mathbb{1}\left(S(q)-c_v(q,q_l)\geq c_f\right)f_\theta(\theta)d\theta-R=0. \tag{122}$$

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. In the special case where $G(.)=u(.)$, the government is utilitarian. Equation (122) is the government's budget constraint. The variable R denotes other government's spending.

Objective (120) assumes that the government sets the subsidy $S(q)$ independently of income y. This assumption is not without loss of generality. In general, a subsidy $S(q, y)$ that depends on adopted capacity and income may achieve higher social welfare than a subsidy that does only depend 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 conditional on income is complicated to implement. Therefore, the government did choose a subsidy that only depends on capacity q. Third, Problem (120) is a multidimensional screening problem. The type parameter determining the choice of capacity q and income y is two dimensional. Theoretically, these problems are difficult to solve because local incentive-compatibility constraints are, in general, not sufficient 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.

A.15.1 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

$$\pi_v(q_l) = S(q(q_l)) - c_v(q(q_l), q_l). \tag{123}$$

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 a compensation. An incentive-compatible mechanism are two functions $q(.), \pi_v(.)$, under which each agent has the incentive to truthfully report her type. Therefore, the objective of the government is finding two such functions, which give the highest payoff. Using standard mechanism design, a mechanism is incentive compatible if and

only if

$$\pi_v'(q_l) = -\frac{\partial c_v(q(q_l), q_l)}{\partial q_l}, \tag{124}$$

and $q(.)$ non-decreasing. As standard in the literature, I neglect the monotonicity constraint on $q(.)$. It can be verified ex-post. Equation (124) defines a function $q(\pi_v')$. Therefore, the government's problem reduces to choosing a function $\pi_v(.)$:

$$\max_{\pi_v(.)} \int G\left(\left(\pi_v - c_f\right) \mathbb{1}\left(\pi_v \geq c_f\right) + y - T(y) - c_l(y,a)\right) f_\theta(\theta) d\theta + V(Q), \tag{125}$$

s.t.

$$\int q\ \mathbb{1}\left(\pi_v \geq c_f\right)\ f_\theta(\theta) d\theta = Q, \tag{126}$$

and

$$\int T(y) - \left(\pi_v + c_v(q, q_l)\right) \mathbb{1}\left(\pi_v \geq c_f\right) f_\theta(\theta) d\theta - R = 0. \tag{127}$$

For better readability, I suppress the arguments of the functions $\pi_v(.)$, $\pi_v'(.)$, and $q(.)$. I solve Problem (125) by using calculus of variation.

A.15.2 The general optimality condition

It follows that in the optimum

$$\begin{aligned}\int_{q_l}^{\infty} \Bigg[\int_0^{\pi_v} \left(\int_0^{\infty} \frac{G'(\pi_v - c_f + y - T(y) - c_l(y,a))}{\lambda} f(a|c_f, \tilde{q}_l) da - 1 \right) f(c_f|\tilde{q}_l)\ dc_f + \\ + \left(\frac{V'(Q)}{\lambda} q - \pi_v - c_v(q(\pi_v'), q_l) \right) f(\pi_v|q_l) \Bigg] f(\tilde{q}_l) d\tilde{q}_l = \\ = \left(\frac{V'(Q)}{\lambda} - \frac{\partial c_v(q, q_l)}{\partial q} \right) \frac{dq}{d\pi_v'} F(\pi_v|q_l) f(q_l)\end{aligned} \tag{128}$$

and

$$\left(\frac{V'(Q)}{\lambda} - \frac{\partial c_v(q, q_l)}{\partial q} \right) \frac{dq}{d\pi_v'} F(\pi_v|q_l) f(q_l) = 0 \text{ for } q_l = 0 \text{ and } q_l = \infty\ . \tag{129}$$

The variable λ is the marginal cost of public funds. Equation (128) is a second order differential equation in the function $\pi_v(.)$ with two boundary conditions (129). The optimal rent $\pi_v(.)$ is the solution to this system.

A.15.3 The Pigouvian subsidy (i.e., the Samuelson rule)

The Pigouvian subsidy is the linear subsidy where $S'(q) = \frac{V'(Q)}{\lambda}$ for all q. This solution is also known as the Samuleson rule (Samuelson, 1954). It is optimal only if

$$\int_0^{\pi_v} \left(\int_0^{\infty} \frac{G'(\pi_v - c_f + y - T(y) - c_l(y,a))}{\lambda} f(a|c_f, \tilde{q}_l) da - 1 \right) f(c_f|\tilde{q}_l)\, dc_f = 0, \tag{130}$$

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 (130) does not hold. The condition depends in particular 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 equalize marginal social welfare weights, then Condition (130) 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 equalize 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:

$$\int_y^{\infty} \left(1 - \int \frac{G'\left((\pi_v - c_f)\, \mathbb{1}\left(\pi_v \geq c_f\right) + \tilde{y} - T(\tilde{y}) - c_l(\tilde{y}, a^{-1}(\tilde{y})) \right)}{\lambda} \times \right.$$
$$\left. \times\, f(c_f, q_l|\tilde{y}) dc_f dq_l \right) f(\tilde{y}) d\tilde{y} + T'(y) \frac{dy}{dT'}\, f(y) = 0 \tag{131}$$

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

A.15.4 The relation to the simple objective (35)

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 (130) 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 (120) to Problem (35). For example, it is the case if the redistributive preferences are Rawlsian, and households with the lowest incomes cannot adopt solar panels.