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

Stefan Pollinger

SciencesPo
DEPARTMENT OF ECONOMICS

Discussion Paper

KINKS KNOW MORE: POLICY EVALUATION BEYOND BUNCHING WITH AN APPLICATION TO SOLAR SUBSIDIES

Stefan Pollinger

SCIENCES PO ECONOMICS DISCUSSION PAPER

No. 2023-01

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

Stefan Pollinger †

July 31, 2024 (first version, November 12, 2020)

Abstract

This paper shows that kinks or discontinuities in economic incentive schemes, such as taxes or subsidies, simultaneously identify agents' intensive and participation margin responses. The proposed semi-nonparametric estimator enables the evaluation of such schemes when existing kink and discontinuity methods are inapplicable due to the presence of both margins. The paper applies the estimator to evaluate the German subsidy for rooftop solar panels, a cornerstone of global climate policies. Due to sizeable responses at both margins, nonlinearities in the programme only modestly increase its cost-effectiveness. The results highlight the importance of simultaneously estimating both margins for optimal policy design.

Keywords: Nonlinear Incentives, Bunching, Participation Margin, Solar Subsidies.
JEL codes: H20, H30, C14.

Nonlinear incentive schemes have a wide range of policy applications in subsidy programmes, taxation, product pricing, and public transfers. A major challenge in their evaluation and optimal design is reliably estimating how agents react to them at the intensive and participation margins.[1] When agents react solely at the participation margin, the regression kink design (Card et al., 2015) can be applied to exploit kinks in the incentive scheme.[2] Correspondingly, when there is only an intensive margin response, the bunching design (Saez, 2010) can be applied. However, agents often respond at both margins simultaneously. In such cases, estimating the participation *and* the intensive margin responses is necessary to evaluate an incentive scheme. Yet the mentioned estimators are not applicable because each margin biases the estimate of the other margin.

*I am especially grateful to Christian Hellwig, Nicolas Werquin, Christian Gollier, and Jean-Pierre Florens for their supervision and guidance. I thank Andrew Atkeson, Nicolas Bonneton, Clément de Chaisemartin, Antoine Ferey, Christophe Gaillac, Emeric Henry, Camille Landais, Stefan Lamp, Alix de Loustal, Thierry Magnac, Isabelle Mejean, Franz Ostrizek, Itzhak Rasooly, Whitney Newey, Emmanuel Saez, Jean Tirole, Augustin Tapsoba, Stefanie Stantcheva, and three anonymous referees for helpful comments and discussions. All errors are mine.

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

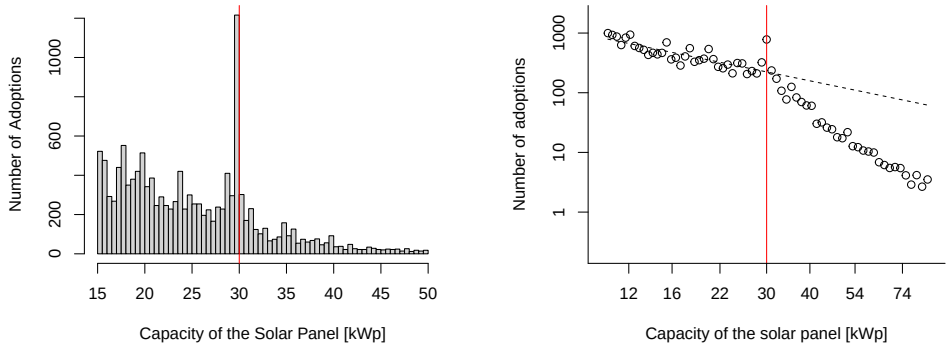
[1]The participation margin is also called the extensive margin.

[2]Kinks are discontinuities in the marginal incentive scheme.

As a methodological contribution, this paper proposes an estimator for agents' responses at both margins. The estimator leverages the effect of kinks in an incentive scheme to identify the two responses jointly. As an applied contribution, the paper evaluates the German subsidy for rooftop solar panels. It is a prominent example of a large and successful deployment subsidy for a nascent green technology.[3]

To illustrate the identification strategy, Figure 1 presents the histogram of solar panel adopters in Germany in 2004. The x-axis shows the capacity choice of adopters in kilowatt-peak (kWp), while the y-axis depicts the number of adopters.[4] Scales are linear in the left panel of the figure and logarithmic in the right panel. This year, the German subsidy had a kink: the marginal sub-

Figure 1: The histogram of solar panel adopters in Germany (2004) in linear scales (left panel) and logarithmic scales (right panel).



sidy rate for adopting a unit of solar capacity decreased discontinuously at 30 kWp (see Figure 2 in Section 1 for an illustration). The red line illustrates the location of the kink point in Figure 1. Kinks have two effects on the observable distribution of agents' choices. First, they cause bunching, i.e., a discrete mass of agents at the kink point, which is clearly apparent in the left panel of Figure 1. Intuitively, compared to a counterfactual linear subsidy, the kink reduces marginal payments to agents above the kink point. Thus, they react at the intensive margin by reducing their capacity. Consequently, the distribution above

[3]Gerarden (2022) and Nemet (2019) show that the German programme was instrumental in driving the enormous price decreases in solar panels over the last two decades, making them one of the cheapest sources of electricity in 2021 (IRENA, 2022).

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

the kink point shifts into the distribution below the kink point; at the kink point, the two parts collide and create a mass point. Second, kinked incentives induce a slope change in the choice distribution at the kink point, which is strikingly apparent in the right panel of Figure 1. Intuitively, compared to a counterfactual linear subsidy, agents receive lower total payments the further above they locate from the kink point. It triggers responses at the participation margin, which tilts the distribution above the kink point clockwise (for a detailed illustration of the two responses, see Section 1.1). Importantly, these observable effects are absent in years when the German subsidy was linear (see Figure 7 in Section 3.1). It confirms that, indeed, the kink causes them. Both observable effects, bunching and the slope change, simultaneously depend on the magnitude of both response margins, intensive and participation. However, they also depend on them distinctly. Therefore, modelling the dependence of the two observable moments on the two margins enables their simultaneous estimation.

The main identifying assumption is the smoothness of the counterfactual choice distribution absent the kink. Blomquist and Newey (2017) and Bertanha et al. (2023) point out that the classic bunching estimator implicitly relies on a parametric functional form assumption on this distribution. As a consequence, it is vulnerable to misspecification. In this paper, I extend the classic approach in three dimensions to alleviate these concerns. First, the paper shows that both margins are point-identified under a weaker assumption: the counterfactual distribution is infinitely differentiable, and the derivatives are sufficiently regular. Second, leveraging this identification result, the paper proposes a semi-nonparametric estimator that employs a data-driven selection of the specification. Compared to the classic bunching estimator, this procedure reduces bias by 11 percent and increases precision by two orders of magnitude in my application. Third, the paper proposes placebo tests to alleviate concerns of specification bias due to an eventual violation of the smoothness assumption. To this end, it exploits additional observations that do not face a kink. Notably, I do not find evidence of specification bias in my application.

The novel estimator offers several advantages. Exploiting the quasi-experimental variation created by the kink circumvents the need for exogenous supply or demand shifters, instruments, control variables, covariates, or panel data to estimate the two margins. The estimation relies solely on the easily observable cross-section of agents' choices. An adaptation to exploit notches, i.e., discon-

tinuities in incentive schemes, is straightforward. These low informational and identifying requirements enhance its applicability for estimating behavioural responses to taxes, subsidies, transfers, and product prices. A disadvantage of the estimator is that, like regression kink, regression discontinuity, and classic bunching designs, it estimates local responses. However, this limitation can often be mitigated by estimating responses at multiple kink points.

Deployment subsidies for nascent green technologies, such as the German solar subsidy, have moved to the forefront of climate action (Podesta, 2023). They can facilitate the direct displacement of polluting activities and catalyse green innovation (Acemoglu et al., 2012). However, one major caveat is their potential burden on public finances. To illustrate, the annual payments for the solar programme in Germany constituted 0.6 per cent of total government expenditure, mainly benefiting wealthy owners of rooftops. A strategy to mitigate these costs is subsidies that are nonlinear in a technology's attributes. For example, the German government implemented a subsidy that is nonlinear in the capacity of solar panels adopted by households and firms. Two questions arise. First, was the subsidy schedule effective at reducing costs without compromising aggregate adoption goals? Second, what is the most cost-efficient nonlinear schedule to achieve a certain aggregate capacity goal?[5] The answers to both questions depend critically on how adopters react to the subsidy. Applying the novel estimator to multiple kinks in the German schedule reveals sizeable responses at both margins. As a consequence, compared to a simple linear subsidy, the government's scheme only modestly reduces costs by 0.14 percent. The most cost-effective nonlinear subsidy triples these savings to 0.45 percent; however, they remain modest overall. A third counterfactual exercise reveals that the interaction of the participation margin with the intensive margin limits the scope for more substantial cost reductions via nonlinear pricing. A fourth exercise underlines the importance of considering both margins when designing nonlinear incentive schemes: optimising the subsidy based on intensive margin estimates alone would increase the programme's costs instead of decreasing them.

The results of this paper indicate that nonlinearities can reduce the public costs of deployment subsidies; however, cost savings can be seriously limited by the interaction of behavioural responses at both margins. Hence, nonlinear

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

subsidies are not a panacea for cost-effectiveness and can even trigger detrimental effects. Estimating both response margins is essential for their optimal design.

Related literature. Methodologically, this paper builds on the semi-nonparametric sieve estimation literature (see Chen, 2007). It is closely related to Gautier and Gaillac (2021) and Iaria and Wang (2024), who use the same smoothness assumption for identification in models distinct from mine: the nonparametric estimation of densities and discrete choice models. Moreover, the paper builds on the bunching and regression kink design literature. The bunching design estimates intensive margin responses using bunching at kink and notch points within the budget set, but it does not account for participation margin responses (see Saez, 2010, Chetty et al., 2011, Kleven and Waseem, 2013, and reviews by Kleven, 2016 and Bertanha et al., 2023). Exceptions are Gelber et al. (2021) and Marx (2024), who estimate both margins exploiting panel data. As in my application, data often lacks a panel dimension.[6] Kleven (2016) discusses participation margin responses as a threat to identification in bunching designs. Indeed, I find that ignoring the participation margin leads to a 12 percent downward bias in my intensive margin estimate.

The participation margin responses at kink points can be estimated using a regression kink design (see Nielsen, Sørensen and Taber, 2010 and Card et al., 2015). However, if there is a positive intensive margin, the classic regression kink design suffers from endogeneity because agents sort into treatment. Ignoring this effect biases the results by 5% in my application. The simultaneous estimation in my paper addresses these biases of the classic approaches. Gerard, Rokkanen and Rothe (2020), Bachas and Soto (2021), and Caetano, Caetano and Nielsen (2024) propose methods to correct for endogenous sorting in the regression discontinuity design and in regression models. Similarly, Kopczuk and Munroe (2015) control for, but do not estimate, standard participation margin responses when investigating market unravelling as a response to a notch. These methods do not aim at simultaneously estimating intensive and participation responses. Blomquist et al. (2021), Bertanha, McCallum and Seegert (2023), and Goff (2022), propose conditions for the partial identification of intensive margin responses in bunching designs. My paper proposes conditions for point-identification, which is necessary for conducting my counterfactual exer-

[6]Repeated cross-sectional data is insufficient to apply their approaches.

cises. Bertanha, McCallum and Seegert (2023) show that point-identification in the classic bunching design can also be achieved using covariates. As in my application, data often lacks such variables.

From an applied perspective, the paper contributes to the literature evaluating subsidies for solar panels. To the best of my knowledge, it is the first paper to evaluate and optimise the cost-effectiveness of nonlinearities in solar subsidies. One strand of the solar literature uses structural models to study the dynamics of the adoption decision (e.g., De Groote and Verboven, 2019; Feger, Pavanini and Radulescu, 2022; Langer and Lemoine, 2022; Gerarden, 2022). Another strand of the solar literature uses reduced-form methods.[7] For example, Germeshausen (2018) uses a difference-in-difference approach to estimate the treatment effect of introducing a new kink in Germany in 2012. The paper does not estimate elasticities at the two adoption margins nor evaluate the cost-effectiveness of counterfactual schemes.[8] It relies on a parallel trend assumption, which is unnecessary in my approach and is violated in my sample. Theoretically, my counterfactual exercises build on the literature on second-degree price discrimination with intensive and participation margin responses (see Rochet and Stole, 2002, Saez, 2002, and Jacquet, Lehmann and Van der Linden, 2013). Using the theoretical results in Rochet and Stole (2002), I solve for the optimal mechanism in a concrete empirical application.

The rest of the paper proceeds as follows. Section 1 outlines the model and discusses identification. Section 2 presents the estimator. Section 3 discusses the empirical application. Section 4 evaluates the policy, and Section 5 concludes.

1 The model

Consider the standard model used in the bunching literature (see Kleven, 2016 and Bertanha et al., 2023). There is a mass of heterogeneous agents indexed by i. They choose a quantity $q \in \mathbb{R}^+$ for which they receive a payment $S(q)$. They

[7] Hughes and Podolefsky (2015) use geographical discontinuities in California to study adoption behaviour. Such discontinuities are not available in Germany. Srivastav (2023) studies feed-in-tariffs and their effect on the financial frictions faced by adopters.

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

solve a standard maximisation problem:

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

where the function $c_v^i(.)$ denotes the increasing and convex variable cost of agent i; π_v^i denotes the variable profit. Contrary to the standard model, this paper adds a participation margin to the decision problem. To this end, assume that agents participate if and only if $\pi_v^i \geq c_f^i$, where c_f^i denotes the fixed cost of participation of agent i. The fixed and variable costs are unobservable and may also contain unobservable benefits. Note that, in particular, the fixed cost may be homogeneous and equal to zero; hence, the model nests the standard bunching model. In an income tax context, q is gross income, $S(q)$ is the net of tax income, $c_v^i(q)$ is the effort-cost of producing income q, and c_f^i is the fixed cost of participating in the labour market. In the application of this paper, q is the capacity of the solar panel, $S(q)$ is the subsidy payment, $c_v^i(q)$ is the variable cost of adopting capacity q, and c_f^i is the fixed cost of adopting a solar system. For ease of exposition, consider the example of solar subsidies from now on.

The subsidy $S(.)$ in (1) can take two forms: the observed kinked subsidy $S_k(.)$ and the counterfactual linear subsidy $S_l(.)$. Comparing adopters' reactions under the kinked subsidy to reactions under the linear subsidy is useful for deriving the estimator. However, this comparison is a thought experiment. The estimator does not exploit changes in a subsidy scheme over time but the effect of the kinked scheme on the cross-section of adopters in a given period. The kinked subsidy $S_k(q)$ is:

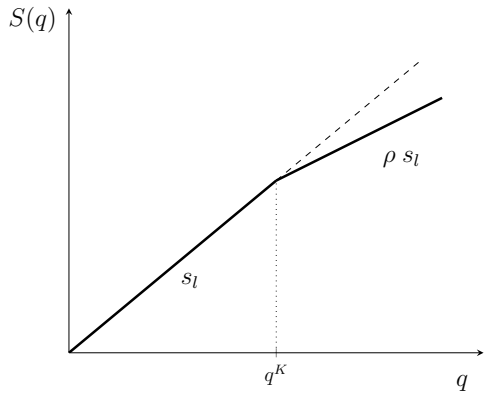
$$S_k(q) = s_l q, \qquad \text{for } q \leq q^K; \tag{2}$$

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

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

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

Figure 2: 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.

Figure 2 illustrates both subsidies. 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.[9]

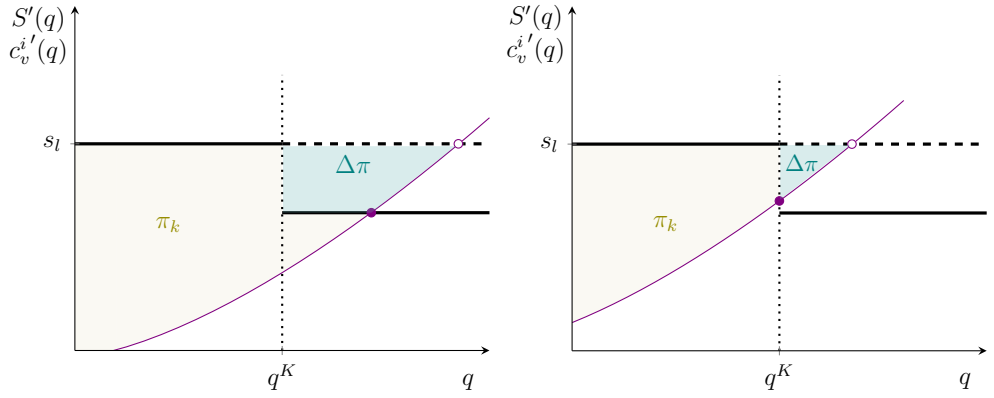
1.1 The graphical intuition behind identification

This section gives graphical intuition on how the distribution of adoptions under the kinked scheme S_k depends on the intensive and the participation margin. It explains the effect on the distribution using a hypothetical change in the subsidy schedule from S_l to S_k. Depending on their production choice under the linear subsidy, the kink affects adopters differently. There are three groups of adopters.

The first group produces more than the kink point under both subsidy schemes. The thin purple line in the left panel of Figure 3 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 these adopters. First, they face a lower marginal subsidy under the kinked scheme than under the linear scheme. Therefore, they adopt less capacity. Note

[9]For example, for any interval of capacity $[q_1, q_2]$, $\int_{q_1}^{q_2} f_k(q)dq$ is the mass of adopters in the interval under the subsidy $S_k(.)$.

Figure 3: The effect of the kinked scheme S_k on adopters 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 lines illustrate the marginal cost curves $c_v^{i\,\prime}$ of adopters who, counterfactually, adopt above the kink point. 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 adopter in the right panel bunches at the kink point. 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.

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. The left panel of Figure 3 depicts the variable profit under the linear scheme as the total coloured area. The light green area π_k is the variable profit under the kinked scheme. The dark green area $\Delta\pi$ is the reduction in profit under the kinked subsidy.

The second group of adopters produces above but close to the kink point under the linear scheme. The thin purple line in the right panel of Figure 3 illustrates the marginal cost curve of such an adopter locally around the kink

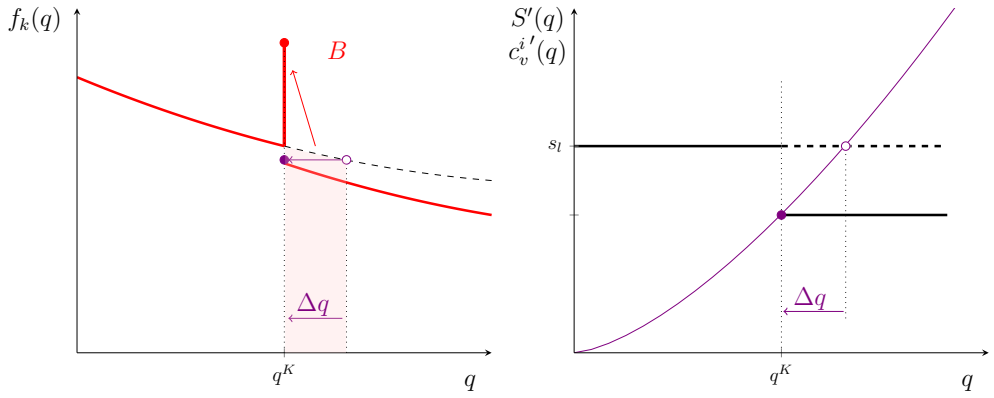
point. Their marginal cost curves cross the kinked marginal subsidy precisely between the two marginal rates. Again, the change in subsidy has two effects on them. First, they reduce production precisely to the kink point, i.e., they bunch at the kink point. Second, they lose profit $\Delta\pi$, depicted as the dark green area in the right panel of Figure 3. Again, due to heterogeneous fixed costs, some of them stop participating under the kinked scheme.

The third group of adopters produces less than the kink point under both subsidy schemes. 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 (for an illustration, see Figure 11 in Appendix B.1).

The distinct effect of the kinked subsidy on these three groups of adopters affects the distribution of adoptions. First, 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 is the case considered by Saez (2010). The left panel of Figure 4 depicts the counterfactual measure f_l as a black dashed line and the observable measure under the kinked subsidy f_k as a red line. 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.[10] At the kink point, there is a mass point B, i.e., the bunching mass. It consists of adopters from the second group. Counterfactually, their mass is the shaded red area. These adopters reduce production; however, they hit the kink point q^K when doing so. By reaching the kink point, they are no longer affected by the subsidy change. Therefore, they "bunch" precisely at the kink point. Below the kink point, the measures under the kinked and linear schemes are the same; adopters in this range are not affected by the change in schemes. How is the measure of adoptions, as illustrated in Figure 4, useful to identify the intensive margin response? Consider the adopter depicted by the thin purple marginal cost curve in the right panel of Figure 4. Her marginal cost crosses the lower marginal subsidy rate exactly at the kink point. The lit-

[10]Depending on the exact response, the mass in each interval increases or decreases because mass needs to be conserved. It is the standard effects of a change-in-variable on a measure, i.e., the effect of the Jacobian.

Figure 4: The observable measure f_k when there is only an intensive margin.

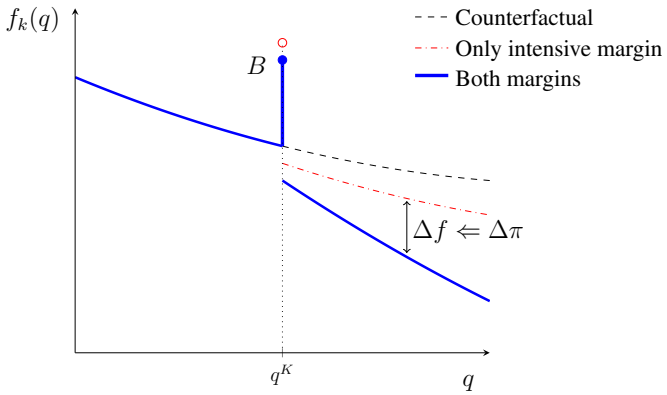


Note: The left panel shows the counterfactual measure f_l as a dashed black line and the observable measure f_k as a red line. At the kink point, there is a mass point B. The right panel shows the marginal cost curve of the marginal buncher. This agent reduces capacity by Δq. The dots in the right panel show the marginal buncher in the measure of adoptions. Adopters in the shaded area to her left bunch at the kink point.

erature calls this adopter the "marginal buncher" (see Saez, 2010). In response to the change in marginal subsidy, the adopter reduces production by Δq. The dots in the left panel of Figure 4 show the marginal buncher in the measure of adoptions. If marginal cost functions do not cross, all adopters in the shaded area to her left "bunch" at the kink point. Since mass is preserved, the bunching mass is approximately proportional to the reduction in the marginal buncher's production Δq. The bunching mass identifies the response Δq, which, under an additional assumption on the cost function, identifies the intensive margin elasticity.

Next, consider the case when fixed costs are present and heterogeneous. Figure 5 illustrates the consequent participation effects on the measure of adoptions. The blue line illustrates the measure when there are responses at both margins. In comparison, the red dash-dotted line illustrates the measure when there is a response only at the intensive margin; the black dashed line illustrates the counterfactual. Again, the range below the kink point is unaffected by the subsidy change. Above the kink point, adopters from the first group illustrated in Figure 3 suffer from a profit loss despite adopting less capacity. Due to heterogeneous fixed costs, some of them stop participating; hence, the loss $\Delta\pi$ causes a drop in

Figure 5: 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.

the participating mass Δf. The larger the capacity q, the larger the loss in profit $\Delta\pi$. Therefore, the larger the capacity q, the larger the drop in participation Δf. This effect is responsible for the slope change in the measure.[11] As in a regression kink design (e.g., see Card et al., 2015), the kink in the subsidy causes only a small change in incentives $\Delta\pi$. Consequently, the change in behaviour Δf is also small. However, the effect of interest depends on $\Delta f/\Delta\pi$, which, as in the regression kink design, makes it possible to exploit even small changes Δf and $\Delta\pi$ for identification. At the kink point, adopters from the second group also suffer from a loss in profit; consequently, some of them stop participating. The bunching mass B decreases from the empty red dot to the full blue dot. Appendix D.7 shows that this effect biases the classic bunching estimates.

Note that the theoretical prediction, illustrated in Figure 5, is strikingly similar to the observed adoption behaviour in Figure 1. Contrary to the counterfactual, the measure under the kinked subsidy is observable. The observable bunching mass and the slope change distinctly depend on the magnitudes of both margins. Under some assumptions, it is possible to formalise the depen-

[11] These statements provide intuition but implicitly assume sufficiently regular responses. See Section 1.2 below for the exact assumptions necessary for identification

dence of each part of the distribution on the magnitude of the two margins. Two observable moments, bunching and slope change, are then sufficient to identify the unknown magnitudes of the two margins. The following section carries out this exercise.

1.2 The distribution of adoptions

This section derives how the observed distribution of adoptions under the kinked subsidy depends on the three unknowns: the intensive margin elasticity, the participation margin elasticity, and the counterfactual distribution. Remember, the counterfactual distribution $f_l(.)$ is the capacity distribution under the counterfactual linear subsidy $S_l(.)$; the capacity q_l denotes the counterfactual choice of an adopter.

Assumption 1 (Intensive margin)**.** *For small variations in the subsidy and adopters close to the kink point, the intensive margin response is isoelastic and bounded.*[12]

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

Denote the intensive margin elasticity by ϵ and the participation margin elasticity by η; $\overline{\epsilon}, \overline{\eta}$ denote their upper bounds. Assumption 1 is standard in the bunching literature. It is a local parametric approximation to the nonparametric variable cost function. Assumption 2 is simply the corresponding assumption for the participation margin. It is a local parametric approximation to the nonparametric distribution of total costs.

Theory predicts that the bunching mass is located precisely at the kink point. However, in practice, the excess mass scatters around the kink point due to agents' optimisation errors. The literature calls this phenomenon non-sharp bunching. To account for non-sharp bunching, it is standard in the bunching literature to choose a bunching interval $[q_L, q_H]$ around the kink point after visual inspection of the histogram (see Kleven, 2016). The literature implicitly

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

assumes that non-sharp bunching is limited to this interval. Note that bunching in my application is very sharp (e.g., see Figure 1). Therefore, I follow the standard approach in this respect.[13]

Proposition 1 (The observed density)**.** *Under Assumptions 1 and 2, and close to the kink point, the observable measure $f_k(.)$ 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:*

$$f_k(q) = f_l(q), \qquad \text{for } q < q_L; \tag{5}$$

$$B = \frac{\int_{q_L}^{q_H \rho^{-\epsilon}} R(q_l, \epsilon)^\eta f_l(q_l) dq_l}{q_H - q_L}, \qquad \text{for } q \in [q_L, q_H]; \tag{6}$$

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

If bunching is sharp $q_L = \lim_{q \uparrow q^K}$ and $q_H = \lim_{q \downarrow q^K}$ and B is a mass point.

Note: The variable q^K denotes the kink point; $[q_L, q_H]$ denotes the bunching interval; ρ denotes the relative change in marginal subsidy rates, and the function $R(., \epsilon)$ is the net subsidy payment to an adopter under the kinked scheme relative to the subsidy payment under the counterfactual scheme. The definition of $R(., \epsilon)$ is in Appendix A.1.1.

The proof of Proposition 1 is in Appendix A.1. Below the bunching interval, the observable measure $f_k(.)$ depends only on the counterfactual measure $f_l(.)$. In the bunching interval, there is an observable mass B. The mass depends on all three unknowns. The measure above the bunching interval depends on all three unknowns as well. However, generically, all observables depend on the three unknowns distinctly, a property crucial for identification.

1.3 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-

[13] Bunching in Figure 1 is much sharper than, for example, in Chetty et al. (2011). A slight scattering of the bunching mass around the kink point is visible. It can be explained by the unavailability of the exact optimal system size at the purchase date. For papers that explicitly consider non-sharp bunching see Anagol, Davids and Lockwood (2022), Bosch, Dekker and Strohmaier (2020), and McCallum and Navarrete (2022).

dimensional. Equation (5) shows that below the bunching interval, the observable measure $f_k(.)$ is equal to $f_l(.)$. Therefore, $f_l(.)$ is identified for values smaller than q_L. However, $f_l(.)$ is part of Equation (6) and (7) evaluated at values larger than q_L. The function is unobservable at these points. For this reason, the elasticities ϵ and η are not identified without further restrictions on the counterfactual distribution $f_l(.)$. As noted by Blomquist and Newey (2017) and Bertanha, McCallum and Seegert (2023), the same problem also appears for the identification of the intensive margin in the classic bunching design. The classic design solves this problem by implicitly relying on a parametric functional form assumption on the counterfactual distribution, making it vulnerable to misspecification (see Bertanha et al., 2023 for a detailed discussion). This section shows that a weaker smoothness assumption allows for the point-identification of both elasticities without restricting $f_l(.)$ to a parametric function.

Consider a sufficiently large interval of quantities $(\underline{q}, \overline{q}) \supsetneq [q_L, q_H \rho^{-\bar{\epsilon}}]$ around the kink point and a transformation of the counterfactual measure $f_l(.)$. For ease of exposition, consider a logarithmic transformation of $f_l(.)$ and an interval such that $\ln q^K - \ln \underline{q} = \ln \overline{q} - \ln q^K$.

Assumption 3.a (Smoothness a)**.** *The logarithm of the counterfactual measure* $\ln f_l(.)$ *is infinitely differentiable in* $\ln q_l$ *at each point in* $(\underline{q}, \overline{q})$ *and the derivatives are bounded by*

$$\left| \frac{d^{(p)} \ln f_l(q_l)}{d \ln q_l^{(p)}} \right| \leq M \frac{p!}{(\ln \overline{q} - \ln q^K)^p} \quad \textit{for all } p \in \mathbb{N}, \tag{8}$$

where the bound $M > 0$ *denotes a large real number.*

Intuitively, this assumption states that the logarithm of the counterfactual measure is sufficiently smooth. It is infinitely differentiable with sufficiently well-behaved derivatives. Note that the assumption is weaker than assuming the derivatives are all bounded since the fraction $p!/(\ln \overline{q} - \ln q^K)^p$ diverges to infinity as p goes to infinity. Therefore, the assumption only states that the derivatives do not go to infinity too fast as p increases. Note that $\bar{\epsilon}$, M, $\underline{q}$, and $\overline{q}$ are assumed to fulfil the properties above but may be unknown to the econometrician.

Alternatively, consider the following assumption:

Assumption 3.b (Smoothness b)**.** *The logarithm of the counterfactual measure* $\ln f_l(.)$ *has a locally convergent power series representation:*

$$\ln f_l(q_l) = \sum_{p=0}^{\infty} \gamma_p \left(\ln \frac{q_l}{q^K} \right)^p \text{ for all } q_l \in (\underline{q}, \overline{q}), \tag{9}$$

where $\gamma_p = \frac{d^{(p)} \ln f_l(q^K)}{d \ln q_l^{(p)}} / p!$.

While Assumption 3.a is more intuitive, it implies the slightly weaker Assumption 3.b (see Lemma 10 in Appendix B.3 for then proof). The weaker Assumption 3.b is sufficient for the results in the rest of the paper. An extensive and flexible class of functions fulfils these assumptions. For example, it includes: finite polynomials; the probability distribution functions of the exponential, normal, and type-I generalised extreme value distribution; any finite mixture of these probability distribution functions; any infinitely differentiable function with bounded derivatives; any finite mixture or composition of such functions. Assumption 3.a or 3.b implies that $\ln f_l(.)$ has a convergent power series representation at each point in the interval $(\underline{q}, \overline{q})$. Such functions are called real analytic on the interval $(\underline{q}, \overline{q})$. See Gautier and Gaillac (2021) and Iaria and Wang (2024) for identification results using such functions in other applications.

Equations (6) and (7) form a simultaneous, nonlinear system of equations. Intuitively, they provide infinitely many moments, i.e., one for each q, to identify the two parameters ϵ and η. However, as acknowledged by Newey and McFadden (1994), it is inherently difficult to prove identification in nonlinear models. A common solution in the estimation of nonlinear models is to impose an additional condition:

Condition 1 (Global rank condition)**.** *The true counterfactual measure* $f_l(.)$ *fulfils the following property: there exits a* q *such that Equations (6) and (7) have a unique intercept in the parameters* (ϵ, η).

For a formal definition of Condition 1 see Appendix B.4.1. The condition can be verified ex-post estimation. See Appendix D.6.4 for the estimates in the application of this paper. Moreover, Lemma 11 in Appendix B.4.2 provides a series of very general sufficient conditions on $f_l(.)$ that guarantee Condition 1 holds. In particular, the lemma shows that Condition 1 holds generically. Moreover, the lemma shows that the condition holds if $d \ln f_l(.)/d \ln q < -1$;

a property that holds true in the application of this paper (see Appendix D.6.4). Note that the necessity of Condition 1 is unrelated to $f_l(.)$ being parametric or nonparametric; in particular, the lemma shows that the condition is not necessary for the identification of the intensive margin when there is no participation margin. Appendix B.4.4 discusses the economic intuition behind Condition 1.

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

The proof is in Appendix A.2. Intuitively, due to Assumption 3.b, it is possible to use the pseudo-parameter $\theta = (\epsilon, \eta, \gamma)$ instead of (ϵ, η, f_l), where $\gamma = (\gamma_0, \gamma_1, ...)$ is the infinite sequence of parameters in Equation (9). Proposition 2 shows that there exists a population criterion with a minimum at the true θ.[14] Intuitively, observations below the bunching interval uniquely identify γ, which, due to analyticity, uniquely identifies the nonparametric counterfactual distribution. Condition 1 implies that observations in the bunching interval and observations above the bunching interval jointly identify the two response margins. Moreover, Assumption 3.b implies that the population criterium is continuous in θ and that the parameter space is compact. Therefore, minimising the population criterion is a well-posed problem.[15]

Proposition 2 has three important implications. First, the result allows deriving a point-estimator of the two response margins (see Section 2). Such point-estimates are necessary to conduct the counterfactual exercises in Section 4. Second, it is not necessary to restrict $f_l(.)$ to a parametric functional form, which alleviates concerns of specification bias. Third, since $f_l(.)$ is a nonparametric object, the specification of the estimator can be selected using standard nonparametric estimation techniques (see Section 2)

1.4 Further notes on Proposition 2

Note that Proposition 2 does not rely on additional assumptions on the parameters of the subsidy scheme q^K, s_l, and ρ, i.e., the location and the size of the

[14]The population criterion is the square distances of the model to the true observable measure. See Equation (17) in Appendix A.2.

[15]In particular, no additional completeness condition, such as in nonparametric instrumental variable problems, is necessary.

kink.[16] These variables may be randomly chosen by the policymaker or endogenous. Moreover, the proposition does not rely on additional assumptions on ϵ and η. In particular, η may be zero and is identified as such. This is the case when, locally, fixed costs are not heterogeneous. Therefore, the model nests the classic bunching estimator with only intensive margin responses. In principle, it is applicable whenever the classic bunching model is applicable.

Following the non-structural econometric literature, Assumptions 1 and 2 are reduced form assumptions on endogenous objects. However, Appendix B.2.1 shows that they are equivalent to the structural Assumptions 4 and 5: the variable cost function and the distribution of total costs are locally isoelastic. As a consequence, locally, the elasticities ϵ and η are structural parameters of the problem. Moreover, the counterfactual choice q_l is equivalent to the adopters' type-parameter; hence, the nonparametric counterfactual distribution $f_l(.)$ is a structural pseudo-parameter. Note that Assumptions 1 and 2 do not rule out a correlation between fixed and variable costs. See Appendix B.2.1 for a detailed derivation of these points.

Assumptions 3.a and 3.b use a logarithmic transformation of f_l and q.[17] One could also use other strictly increasing and continuous transformations. In particular, one could use the identity, i.e., no transformation. However, the logarithmic transformation has the advantage that its series expansion contains the uniform distribution, the Pareto distribution, and the log-normal distribution as special cases. These are common distributions for random variables on a positive domain. Moreover, Figures 1 and 7 show that the observed distributions are very close to linear on a logarithmic scale in the empirical application of this paper. This shape suggests that the counterfactual distribution is close to a Pareto distribution; hence, it is advantageous to use the logarithmic transformation.

The following appendices discuss Proposition 2 further. Appendix B.4.3 discusses conditions for local identification. Appendix B.6 compares Proposition 2 to the identification results with only intensive responses in Chetty et al. (2011), Blomquist and Newey (2017), Bertanha, McCallum and Seegert (2023), and

[16]The identification result in Proposition 2 does not rely on an assumption about the size of the kink. However, the size influences the properties of the estimator presented in Section 2. Large kinks have the advantage of creating stronger variation. Hence, everything else equal, estimates have lower variance. Small kinks have the advantage that, as discussed by Kleven (2016), estimates have smaller specification biases when Assumptions 1-3.b only hold approximately.

[17]The assumptions imply that $f_l(.)$ is uniformly strictly larger than zero.

Goff (2022). In particular, Blomquist and Newey (2017) show that, even when bunching is sharp, the bunching mass does not identify the intensive margin elasticity if $f_l(.)$ is differentiable only finitely many times. Appendix B.6 discusses why their non-identification result does not contradict Proposition 2. Moreover, Appendix B.7 illustrates Proposition 2 using simulations. It illustrates that, under Assumption 3.b, both elasticities are identified even when bunching overlaps with a hump in the counterfactual distribution and when the bunching interval is large. Appendix B.8 generalises the result in Proposition 2 to the case when there is a discontinuity (i.e., a notch) in the incentive scheme.

2 Estimation

Figure 6 illustrates the estimation using the distribution of solar panel adoptions in Germany in 2004. This year, the German subsidy had a kink at a capacity of 30 kWp, marked by the red line in the figure. The black dots show the normalised histogram and the black bar depicts the mass in the bunching interval. Scales are logarithmic, and the kink point is normalised to zero. Intuitively, the estimation minimises the distance between the observed log-histogram and the model in Proposition 1. The blue line depicts the estimated model. The purple dashed line depicts the estimated counterfactual distribution.

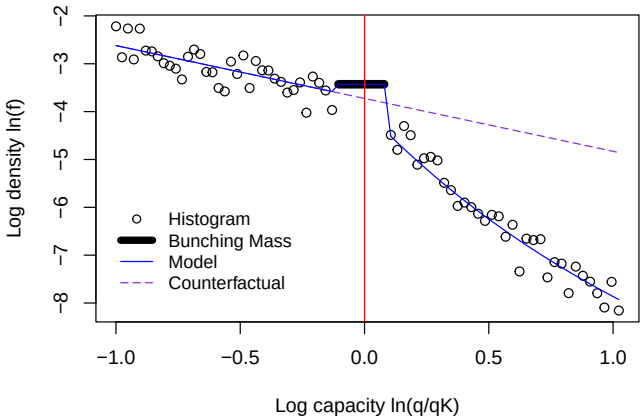
The next paragraphs describe the estimation in detail. A first step constructs the variables $\widehat{\ln f(q_j)}$, an estimate of the log-density at capacity q_j, and $\widehat{\ln B}$, the logarithm of the normalised number of adopters in the bunching interval $[q_L, q_H]$. These are the black dots and the black bar in Figure 6. See Appendix C.1 for the details of their estimation.

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

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

The vector $(\gamma_0, ..., \gamma_P)$ denotes the parameters of the series and $P(n)$ denotes the order. It converges slowly to infinity as the sample size n converges to infinity. The interval $[\underline{b}, \bar{b}] \supsetneq [q_L, q_H]$ is the bandwidth. It is the interval of values around the kink point used for estimation, i.e., the range of q_j. Note that the bandwidth

Figure 6: The distribution of adoptions in 2004 with the estimated model.



Note: The x-axis shows the normalised logarithm of capacity. The y-axis shows the logarithm of the density. The black dots show the logarithm of the histogram; the number of observations in a bin is normalised by the bin size and the total number of observations. The black bar shows the mass in the bunching interval. The red line marks the kink point. The estimation minimises the distance between the data in black and the model in blue. The dashed purple line depicts the estimated counterfactual.

is not to be confused with the bin sizes of the log-histogram.

The estimation follows a two-step least square procedure. It iteratively minimises the square distance between the model and the log-histogram outside and inside the bunching interval until the estimates converge:

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

$$\hat{\epsilon} = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln B(\epsilon, \hat{\eta}, \hat{\gamma_P}) \right)^2, \qquad (11)$$

where $f_k(q_j \mid \hat{\epsilon}, \eta, \gamma_P)$ and $B(\epsilon, \hat{\eta}, \hat{\gamma_P})$ denote the model in Proposition 1 as a function of the parameters. By a slight abuse of notation γ_P denotes the vector of coefficients of the series. Appendix C.2 shows that the estimator is consistent. Following Chetty et al. (2011), the standard errors can be estimated using the nonparametric bootstrap.

The selection of the nonparametric specification is data-driven. As standard

in nonparametric estimations, the estimates' bias and variance depend on the specification. There are two main specification-parameters. The first parameter is the order of the series P. The higher the order P, the lower the bias and the larger the variance. The second parameter is the bandwidth $[\underline{b}, \overline{b}]$. The smaller the bandwidth, the lower the bias and the larger the variance.[18] As is standard in nonparametric estimations, I choose the specification that minimises an estimate of the mean squared error (MSE). Appendix C.3 discusses the estimator of the mean squared error in detail. Appendix C.4 gives a summary and step-by-step guide of the entire estimation procedure.

3 Empirical application

3.1 Policy description

The German subsidy for solar panels was introduced on April 1st, 2000. The subsidy is a guaranteed feed-in tariff, paid per kWh (kilowatt-hour) of produced electricity sold to the grid. A fixed tariff rate is guaranteed for 20 years once a household or firm decides to adopt. Typically, agents only adopt once during the sample period. Since this paper focuses on deployment subsidies for early-stage technologies, it studies the first years of the German programme from 2000 to 2008. Over this period, households and firms sell all the produced electricity to the government at the subsidised rate. Therefore, the problem is equivalent to a procurement problem: the principal (i.e., the government) procures the installation of capacity to agents (i.e., households and firms). The tariff rates depend on the time point of adoption and the adopted capacity. The capacity of a solar panel, measured in kilowatt-peak (kWp), is the amount of electricity it produces under standardised conditions. Because produced electricity is proportional to capacity, for agents who adopted from 2000 to 2003, the present discounted value of subsidy payments was simply linear in their system's capacity.[19]

The subsidy programme was very successful at incentivising households and firms to adopt solar panels; numbers increased rapidly in most years (see Ta-

[18] For a given bandwidth, the bias and the variance of the estimates depend on the order of the polynomial P. For the estimates to be consistent, it suffices that P goes to infinity as the sample size goes to infinity. A smaller bandwidth reduces the bias for any given P; however, the bandwidth does not need to shrink with sample size for the estimates to be consistent.

[19] See Appendix D.2.1 for a formal derivation of this statement

ble 7 in Appendix D.1). However, as a consequence, the programme became very costly, mainly benefiting the wealthy owners of rooftops. For example, total yearly payments in 2016 were 9 billion euros (Übertragungsnetzbetreiber, 2016), corresponding to 0.6 percent of total government spending.[20] To curtail these costs, in 2004, the government introduced two kink points in the subsidy schedule. For agents who adopted from 2004 to 2008, the present discounted value of subsidy payments was nonlinear in their system's capacity. At a capacity of 30 kWp, the marginal subsidy rate decreased by 5%; at 100 kWp, it decreased by 1%. Note that the policy change did not affect agents who had adopted before. See Appendix D.1 for a more detailed description of the policy and data.

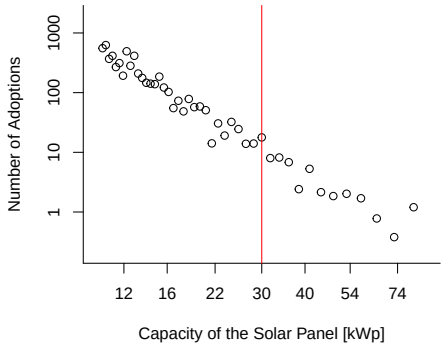
Consistent with the theoretical predictions in Section 1, the kinks affected the adopters' behaviour. Figure 1 in the introduction shows the histogram of all solar panel adoptions in the year 2004 around the kink point at 30 kWp. Many adopters bunch at the kink point, and the distribution shows an evident slope change. In contrast, Figure 7 shows the distribution of adopters who adopted in 2003 when the subsidy was linear. The distribution is remarkably smooth around the future kink point. There is no significant bunching mass or visible slope change as in 2004. It confirms that the slope change and bunching in 2004 are indeed caused by the kink, hence evidence for responses at the two margins.[21] 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. The left panel of Figure 19 in Appendix D.1 shows the histogram of adoptions from 2000 to 2002. There is no evidence for a time trend in the concavity of the distribution. Moreover, the right panel in Figure 19 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.

To evaluate and optimise the subsidy's cost-effectiveness, it is necessary to know how agents react to it at the participation and intensive margin. While subsidy rates and prices for solar panels vary over time, in a large market like Germany, this variation is endogenous to demand, making it unsuitable for identifying the two margins. Methods for estimating both margins proposed by Best

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

[21]On a side note, the subsidy rate in 2003 was lower than in 2004. While it is not directly relevant to the shape of the distributions, it explains why the number of adopters in 2004 was higher than in 2003.

Figure 7: The histogram of adoptions in 2003 (logarithmic scales).



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

and Kleven (2017), Gelber et al. (2021), and Marx (2024) are not applicable since capacity-specific time-trends in adoption are not parallel and the data lacks a panel dimension. Therefore, the following section exploits the effect of kinks on the cross-sectional distribution of adopters' behaviour in Figure 1. It directly applies the model presented in Section 1 for estimating the two margins.

In the context of this application, examples of the cost components in model (1) are: the monetary costs of the installation, warm-glow preferences for solar panels, the opportunity and aesthetic costs of using space on the roof, opportunity costs of time and money, the opportunity cost of adopting at a different point in time, and eventual direct benefits from consuming electricity produced by the solar panel.[22] The subsidy $S(.)$ in model (1) is the present discounted value of the feed-in-tariff payments. Appendix D.2.1 shows that the model in Section 1 encompasses subsidy payments via a feed-in tariff since electricity production is proportional to the adopted capacity. In particular, the model accounts for adopter-specific heterogeneity due to climate conditions or discounting of future payments. An adopter can increase capacity by using more area on

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

the roof or by adopting a system of higher efficiency (i.e., higher capacity per area). Therefore, as in model (1), the variable cost of adopting solar panels is increasing and convex.[23] Appendix D.2.2 further discusses the convexity of the variable cost function. Appendix D.2.3 discusses how the model encompasses dynamic decisions.

3.2 Empirical results

Without loss of generality, normalise the rate s_l in Equation (2) to one. It corresponds to choosing a monetary unit. Consequently, the monetary unit of all monetary variables is the present discounted value of payments to one kWp of capacity. Since the bunching literature estimates behavioural responses as elasticities, the empirical model estimates the participation margin response as the elasticity η. However, in the theoretical literature on nonlinear pricing, it is more common to use participation semi-elasticities (see Rochet and Chone, 1998 and Jacquet, Lehmann and Van der Linden, 2013). Therefore, this section reports the results at the participation margin as semi-elasticities κ. The interpretation is as follows: a lump sum payment equivalent to the present discounted value of one kWp increases participation by a factor κ.[24]

Table 1 shows the main empirical results: the estimates at the two kink points at 30kWp and 100 kWp, pooling observations from 2004 to 2008. The results show sizeable responses at the intensive margin, with statistically equal elasticities at different capacities. The participation margin semi-elasticity is sizeable at the lower kink point but decreases with capacity. For an interpretation of this pattern, see Appendix D.3. I cannot exploit data after 2008 since the kink points overlap with other policy changes. I pool yearly observations to increase sample size and to estimate the long-run elasticities over this time period.

Appendix D.4 presents and discusses the specification. Correspondingly to Figure 6, Figures 21 and 22 in Appendix D.4 depict the data and illustrate the estimation. Appendix D.5 discusses the details of selecting the specification.

[23] The opportunity costs of area on the roof are convex. Moreover, the cost of increasing capacity via efficiency is convex; the more efficient a system, the higher the cost of increasing its efficiency further.

[24] The relation between the two variables is $\eta = \kappa \times S$, where S is the subsidy payment. The standard deviation of $\hat{\kappa}$ follows from the delta method: $SD_\kappa = SD_\eta / S(q)$.

Table 1: The estimates at 30 and 100 kWp pooling observations from 2004 to 2008.

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

Note: The table reports the estimated intensive margin elasticity $\hat{\epsilon}$ and participation margin semi-elasticity $\hat{\kappa}$ at the kink points at 30 and 100 kWp. The standard errors are in brackets. The estimation pools observations from 2004 to 2008.

3.3 Robustness

One concern for the estimation is the violation of Assumption 3.b because of irregularities in the counterfactual. There might be an excessive mass or a slope change at the kink point for reasons other than the subsidy's kink. Moreover, there might be a continuous bump or a concavity in the counterfactual distribution, which is not or only hardly predictable from observations to the left and right of the kink point.

As a robustness check, I estimate the model on observations from 2000 to 2003 when there was no kink in the subsidy. I use the same specification as in 2004-2008. Figure 8 shows the distribution in 2000-2003, with the estimated model and counterfactual. The estimates in Table 2 are insignificant, alleviating the above-mentioned concerns. Appendix D.6.1 reports the robustness check for the estimates at 100 kWp. Again, the estimates are not significant.

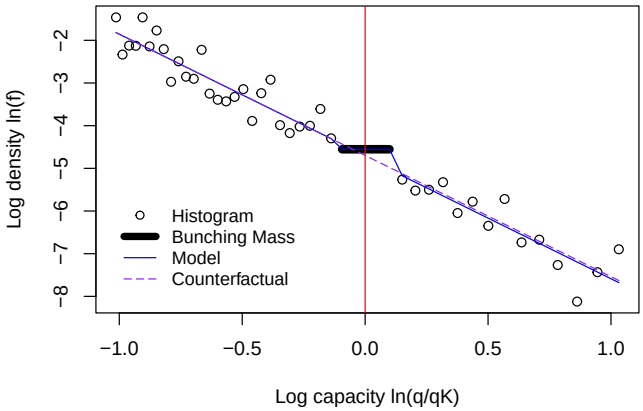
Table 2: The estimates in 2000-2003 (placebo).

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

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

As an additional robustness check, Appendix D.6.2 shows how to directly estimate an eventual specification bias caused by the violation of Assumption 3.b. Again, the estimator exploits untreated observations responding to a sub-

Figure 8: The distribution of adoptions in 2000-2003 with the estimated model.



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

sidy without kink, such as the observations from 2000-2003. Table 12 in Appendix D.6.2 reports the results. The point estimates are qualitatively small and statistically insignificant. Appendix D.6.3 reports robustness checks to the choice of the bunching interval; the estimates are robust. Appendix D.6.4 verifies the rank condition, which holds by a large amount.

3.4 Comparison to the classic bunching and regression kink design

The estimator in Section 2 differs from the classic bunching estimator in three dimensions. First, it correctly specifies the shape of the counterfactual distribution $f_l(.)$ in the range $[q^K, q^K\rho^{-\epsilon}]$, alleviating the concerns of specification bias raised by Blomquist and Newey (2017).[25] Second, it explicitly accounts for participation margin responses. Third, it estimates the counterfactual distribution nonparametrically outside the bunching interval, choosing the order of the series P optimally. Appendix D.7 further discusses these differences and quantifies them for the main estimates of this paper. It finds that the implicit assumptions

[25]Note that this is the theoretical range of agents who bunch sharply. It is a subinterval of the interval $[q_L, q_H\rho^{-\epsilon}]$ in Equation (6) .

of the classic bunching estimator downward bias the intensive margin estimate by 23 %. Half of the bias is attributable to the misspecified counterfactual and half to ignoring the participation margin (see Table 17 in Appendix D.7). Moreover, Table 18 in Appendix D.7 shows estimates using the suboptimal order of the series $P = 7$, which is the specification used in Chetty et al. (2011). The exercise gives much noisier point estimates: the standard errors of the participation margin increase by two orders of magnitude. The comparison underlines the importance of choosing the series order based on the mean squared error to avoid unnecessarily imprecise estimates.

Similarly, Appendix D.7 shows that ignoring the intensive margin introduces an upward bias of 5% in the estimate of the participation margin. These results show that it is essential to estimate both margins simultaneously. It confirms that the classic regression kink design as in Card et al. (2015) is not applicable to estimate participation margin responses when agents select into treatment via the intensive margin.

4 Policy evaluation

This section uses the estimates to evaluate and optimise the subsidy. It builds on the results of Rochet and Stole (2002), who provide a theoretical solution to the nonlinear pricing problem of a monopoly when there are intensive and participation margin responses. In line with the empirical results in Section 3, this section assumes globally isoelastic intensive margin responses and a normally distributed fixed cost to conduct the counterfactual exercises. Appendices D.3 and E.1 discuss these assumptions and the calibration. Note that, in line with the bunching literature, so far, the term counterfactual refers to the linear subsidy without a kink. This section calls any subsidy different from the observed kinked subsidy a counterfactual subsidy.

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

$$\min_{S(.)} \int S(q) dF_S(q) \quad \text{such that} \quad \int q \, dF_S(q) \geq Q^T, \tag{12}$$

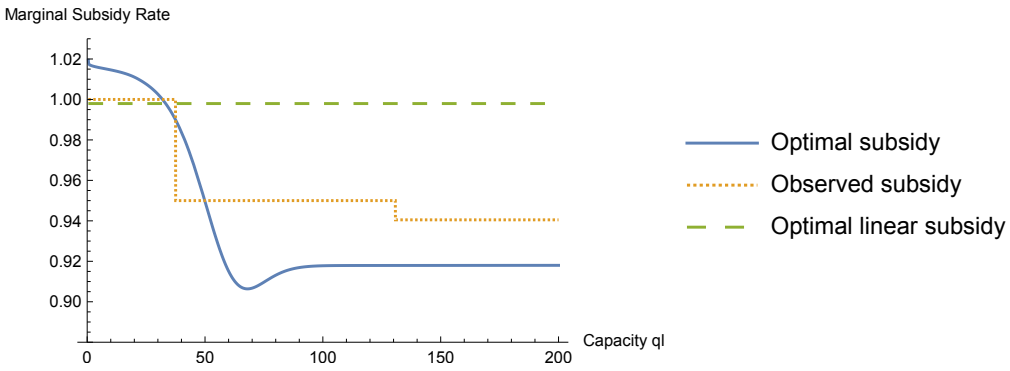
where $F_S(q)$ is the distribution of capacity under subsidy $S(.)$. This section uses the simple objective (12), which avoids taking a detailed stand on governmen-

tal preferences. For a detailed discussion and extension of this objective, see Appendix E.2.

4.1 Results of the counterfactual exercises

This section discusses four counterfactual exercises. The first exercise solves for the optimal linear subsidy, which serves as a benchmark. Appendix E.3.1 describes the procedure. Compared to this benchmark, the government's subsidy achieves a modest cost reduction of 0.14 percent. The second exercise solves for the optimal nonlinear subsidy using mechanism design. The analysis follows the screening problem in Rochet and Stole (2002) and is outlined in Appendix E.3.2. The optimal nonlinear subsidy is 0.45 percent less costly than the linear benchmark. Cost savings are by a factor 3.14 larger than the cost savings of the government's subsidy. Figure 9 compares the optimal marginal subsidy schedule $S'(q(q_l))$ to the linear benchmark and the observed marginal subsidy. Qualitatively, the optimal marginal subsidy is mostly downward sloping; hence,

Figure 9: The observed and the optimal marginal subsidies $S'(q(q_l))$.



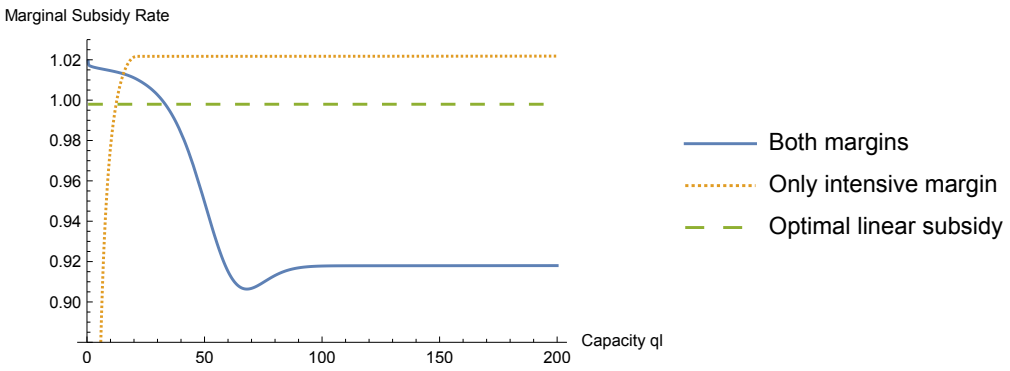
Note: The figure compares the observed marginal subsidy with the optimal linear and the optimal nonlinear marginal subsidies. Note that the first marginal subsidy rate of the observed subsidy is normalised to one.

it is similar to the subsidy used by the government. However, quantitatively, the range of optimal marginal subsidy rates is larger than the range used by the government. The optimal scheme reduces rents more than the observed scheme by paying higher rates to small adopters and lower rates to large adopters. While

the result shows room for improving the government's scheme, the overall benefits of using a nonlinear subsidy scheme remain modest.

To better understand what limits the scope for curtailing costs, a third thought experiment assumes adopters can only respond at the intensive margin. Without participation margin, the optimal nonlinear subsidy would be 8.6 percent less costly than the linear benchmark. See Appendix E.3.2 for the detailed analysis. This result shows that the interaction of the participation margin with the intensive margin limits the scope for cost reduction via a nonlinear subsidy. Figure 10 compares the optimal marginal subsidy without participation to the optimal marginal subsidy with participation. The marginal subsidy schedule is

Figure 10: The optimal marginal subsidy $S'(q(q_l))$ with and without participation margin.



Note: The dotted orange line depicts the optimal marginal subsidy when there is no participation margin. The blue line shows the optimal marginal subsidy when there are both margins.

increasing, and the distance to the linear benchmark is much more significant than when the participation margin is present. The shape of the scheme follows from simple intuition. On the one hand, it is optimal to pay a high marginal rate to large adopters. Reacting only to marginal rates, large adopters install larger capacities than under the linear benchmark. On the other hand, the marginal rate paid to small adopters is low. They adopt lower quantities than under the linear benchmark. By definition, the net capacity effect is zero, and the government achieves the fixed capacity goal. However, the net cost effect is not zero. The total payment to large adopters is the integral under the orange/dotted curve. The low marginal rates for small adopters extract profit from large adopters be-

cause they receive a lower payment for infra-marginal units. Therefore, the scheme is less costly than the linear scheme. In contrast, when, as in reality, there are participation margin responses, this strategy to extract profits is ineffective. Low marginal subsidy rates for small capacities affect larger adopters by reducing their profit margins. It triggers responses at the participation margin and, therefore, a loss in capacity. This effect limits the room for rent extraction through nonlinear pricing when the participation margin is active. Therefore, cost-savings when both margins are active remain moderate.

The results in Figure 10 show that the participation margin changes the shape of the optimal marginal subsidy. However, is it counterproductive to ignore participation when, in reality, such a margin is present? Rochet and Stole (2002) show that when the type distribution $f_l(.)$ is uniform, the optimal intensive allocation bounds the optimal allocation with both margins. Their result suggests that implementing the optimal intensive schedule in Figure 10 might reasonably approximate the optimum. Appendix E.3.3 shows that their result is not robust to more general forms of the type distribution: for the type distribution in this application, the optimal allocation is close to but outside the bounds derived by Rochet and Stole (2002). To gather additional evidence, a fourth counterfactual exercise assumes the policymaker implements the optimal intensive schedule, but adopters react at both margins. I keep aggregate capacity constant again to allow for a meaningful comparison to the other counterfactual exercises. The exercise finds that the optimal intensive schedule increases costs by 3.1 percent instead of decreasing them, i.e., ignoring participation has a sizeable adverse effect on costs. The result shows that combining the theoretical results in Rochet and Stole (2002) with estimates of the intensive *and* participation margin responses is crucial for implementing the optimal schedule. Optimising incentive schemes based on intensive margin estimates alone may even be counterproductive.

Appendix E.3.4 compares the results in this paper to the results in Germeshausen (2018). My estimates imply a treatment effect very close to the one estimated by Germeshausen (2018), which provides evidence for the validity of the respective identifying assumption in both studies. Table 20 and 21 in Appendix E.3.5 summarise the results of the four counterfactual exercises.

5 Conclusion

This paper shows how to leverage kinks or notches to estimate agents' intensive and participation margin responses simultaneously. An application to the German subsidy for rooftop solar panels demonstrates how to use these estimates to evaluate and optimise a nonlinear incentive scheme. The relatively low informational requirements of the proposed estimator enhance its potential applicability in other contexts.

Kinks and notches are common features of economic incentive schemes in taxation, health care, labour regulation, environmental regulation, education, product pricing, and finance. Consequently, the bunching estimator has been widely used to estimate intensive margin responses (see Bertanha et al., 2023 for a review). However, in many applications, not only intensive but also participation margin responses to the incentive scheme are plausible. For example, as a response to income taxation, workers may cease active participation in the formal labour market, instead relying on welfare programmes, unemployment insurance, or informal work. Certain taxpayers may decide to migrate. Secondary earners may decide to engage in household production instead of earning wage income. Indeed, the theoretical contributions by Saez (2002) and Jacquet, Lehmann and Van der Linden (2013) point out that participation margin responses are important to consider when designing optimal income taxes. The empirical evidence on participation margin responses to income taxes is mixed (e.g., see Blundell, Bozio and Laroque, 2011, Chetty et al., 2013, and Kleven, 2024). Potential simultaneous responses are not limited to income taxes. Corporate taxes may influence firms' extensive choice of corporation type and formality, along with the intensive choice of reported profit. Nonlinear product prices may influence whether consumers purchase from a certain firm, as well as the purchased quantity. Size-based regulations of products may distort firms' production decisions at an intensive and extensive margin. The results of this paper highlight the importance of estimating both margins for the evaluation of such nonlinear incentive schemes.

A Appendix: Proofs

A.1 Proof of Proposition 1

Denote by q_l the choice and by c_t the total cost of adopter i under the counterfactual subsidy S_l. Denote by $F_{t|q_l}(c_t|q_l)$ the conditional CDF of total costs. It follows by Assumptions 1 and 2 that, locally around the kink point, the cost functions of adopters and the CDF of total costs can be expressed as

$$c(q, q_l, c_t) = \underbrace{\frac{s_l}{q_l^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}}}_{\text{variable cost}} + \underbrace{c_t - q_l s_l \frac{\epsilon}{1+\epsilon}}_{\text{fixed cost}}, \tag{13}$$

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

where q_l and c_t are equivalent to type parameters and $\overline{c_t}(q_l)$ is the upper bound of total costs.[26] Proposition 1 follows from solving the adopters' maximisation problem (1) using Equations (13) and (14). For a detailed step-by-step derivation of Proposition 1, see Appendix B.2.

A.1.1 The definition of function $R(., \epsilon)$ in Proposition 1

The function $R(q_l, \epsilon)$ in Proposition 1 denotes the net subsidy payment to an adopter under the kinked scheme relative to the subsidy payment under the linear scheme. The function $R(q_l, \epsilon) = 1$, for $q_l < q^K$; $R(q_l, \epsilon) = R_B(q_l, \epsilon)$ for $q_l \in [q^K, q^K \rho^{-\epsilon}]$; and $R(q_l, \epsilon) = R_2(q_l, \epsilon)$ for $q_l > q^K \rho^{-\epsilon}$; where

$$R_B(q_l, \epsilon) = \frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 - \left(\frac{q^K}{q_l}\right)^{\frac{1+\epsilon}{\epsilon}}\right); \tag{15}$$

$$R_2(q_l, \epsilon) = (1-\rho)\frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 + \frac{\rho^{\epsilon+1}}{\epsilon}\right). \tag{16}$$

A.2 Proof of Proposition 2

This section denotes the true functions and parameters in the population with superscript "o", while functions and parameters without superscript denote the

[26] This formulation of the cost depends on the counterfactual subsidy rate s_l. However, since s_l is observable and fixed, this dependence poses no problem.

general values of these objects. Denote by $\gamma = (\gamma_0, \gamma_1, ...)$ the infinitely dimensional vector of coefficients γ_p in Assumption 3.b and by $\theta = (\epsilon, \eta, \gamma)$ the pseudo-parameter. The functional $f_l(q|\gamma)$ denotes the counterfactual measure as a function of the sequence γ. The interval $[\underline{b}, \overline{b}]$ denotes the bandwidth, i.e., a range of points around the kink point such that $[q_L, q_H] \subsetneq [\underline{b}, \overline{b}] \subsetneq (\underline{q}, \overline{q}\rho^{\overline{\epsilon}})$.

Consider the population criterion $Q(\theta)$:

$$Q(\theta) = \int_{\underline{b}}^{q_L} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \epsilon, \eta, \gamma\right)\right)^2 \, dF^w + \left(\ln B^o - \ln B(\epsilon, \eta, \gamma)\right)^2 + \int_{q_H}^{\overline{b}} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \epsilon, \eta, \gamma\right)\right)^2 \, dF^w. \tag{17}$$

The function $f_k^o(.)$ denotes the observable measure of agents in the population. The function $f_k(.|\epsilon, \eta, \gamma)$ denotes the respective function in Proposition 1, where the general function $f_l(.)$ is replaced by the power series $f_l(.|\gamma)$. B^o denotes the observable mass of agents in the bunching interval in population, while $B(\epsilon, \eta, \gamma)$ denotes the respective function in Proposition 1, where the general function $f_l(.)$ is replaced by the power series $f_l(.|\gamma)$. The function F^w denotes a weighting measure. It assigns strictly positive weight to any open interval $(q_x, q_y) \subset [\underline{b}, q_L] \cup [q_H, \overline{b}]$, where $q_x < q_y$.[27]

Lemma 1. *Under Assumptions 1, 2, 3.b, in any minimum of $Q(\theta)$ it holds that $\gamma = \gamma^o$. Under Condition 1, $\theta^o = (\epsilon^o, \eta^o, \gamma^o)$ is the globally unique minimum of $Q(\theta)$.*

PROOF Lemma 1:

By Proposition 1 and Assumption 3.b, $f_k^o(q) = f_k(q \mid \epsilon^o, \eta^o, \gamma^o)$ and $B^o = B(\epsilon^o, \eta^o, \gamma^o)$. It is easy to see that $Q(\theta^o) = 0$ and $Q(\theta) \geq 0$ for all θ. Therefore, $\theta = \theta^o$ is a minimum of $Q(.)$. If another minimum θ exists, it needs to hold that $Q(\theta) = 0$.

Suppose a pseudo-parameter θ minimizes $Q(.)$. Using Proposition 1 and Assumption 3.b, it follows that the first integral in Equation (17) is equal to

[27] The weighting measure may be used to increase efficiency. Identification will follow for any such measure. In particular, it follows in the case when there is no such measure, i.e., F^w is the uniform CDF.

zero:

$$\int_{\underline{b}}^{q_L} \left(\ln f_l(q \mid \gamma^o) - \ln f_l\left(q \mid \gamma\right)\right)^2 \, dF^w = 0.$$

Assumption 3.b implies that $\ln f_l(.)$ is continuous in this range. Note that this integral is a norm for continuous functions. Therefore, the norm can only be zero if $\ln f_l(q \mid \gamma^o) = \ln f_l\left(q \mid \gamma\right)$ for all $q \in (\underline{b}, q_L)$. Assumption 3.b implies that $\ln f_l(e^{(.)} \mid \gamma^o)$ and $\ln f_l\left(e^{(.)} \mid \gamma\right)$ are real analytic in $\ln q$ on the interval $(\ln \underline{q}, \ln \overline{q}) \supset (\ln \underline{b}, \ln q_L)$.[28] By the identity theorem of real analytic functions $\ln f_l(e^{(\ln q)} \mid \gamma^o) = \ln f_l\left(e^{(\ln q)} \mid \gamma\right)$ for all $\ln q \in (\ln \underline{q}, \ln \overline{q})$ (see Corollary 1.2.6 in Krantz and Parks, 2002). The cited results from Krantz and Parks (2002) are summarised in Appendix F. Since the power series representation of analytic functions is unique (Corollary 1.1.16 Krantz and Parks, 2002), it follows that $\gamma_p^o = \gamma_p$ for all p. The population criterion $Q(.)$ uniquely identifies γ^o, and, hence, the counterfactual distribution $f_l^o(.)$ in the interval $(\underline{q}, \overline{q})$.

Because $f_k(.|\epsilon, \eta, \gamma)$ is continuous in the interval $[q_H, \overline{b}]$ and the third part of the criterion $Q(.)$ is a norm for continuous functions, it follows that

$$f_k(q \mid \epsilon^o, \eta^o, \gamma^o) - f_k(q \mid \epsilon, \eta, \gamma^o) = 0 \; \forall q \in (q_H, \overline{b}). \tag{18}$$

By Proposition 1, the definition of $R(., \epsilon)$ in Section A.1.1, and Assumption 3.b

$$f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o} R_2(q\rho^{-\epsilon^o}, \epsilon^o)^{\eta^o} = f_l^o(q\rho^{-\epsilon})\rho^{-\epsilon} R_2(q\rho^{-\epsilon}, \epsilon)^{\eta} \;\; \forall q \in (q_H, \overline{b}), \tag{19}$$

where the function $R_2(., \epsilon)$ is defined in Equation (16) in Section A.1.1 and $f_l^o(.)$ denotes the true counterfactual. Note that both sides of this equation are real analytic in $\ln q$ since $f_l^o(.)$ and $R_2(.)$ are real analytic in $\ln q$. Therefore, by the identity theorem of real analytic functions

$$f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o} R_2(q\rho^{-\epsilon^o}, \epsilon^o)^{\eta^o} = f_l^o(q\rho^{-\epsilon})\rho^{-\epsilon} R_2(q\rho^{-\epsilon}, \epsilon)^{\eta} \; \forall q \in (\underline{q}, \overline{q}\rho^{\overline{\epsilon}}). \tag{20}$$

The second part of the criterion $Q(.)$ implies that $B(\epsilon^o, \eta^o, \gamma^o) = B(\epsilon, \eta, \gamma^o)$.

[28] Since real analyticity is assumed to hold in logarithms, it is necessary to work with $\ln f_l\left(e^{(.)} \mid \gamma\right)$. Note that $\ln f_l\left(e^{(\ln q)} \mid \gamma\right) = \ln f_l\left(q \mid \gamma\right)$.

By Proposition 1, the definition of $R(.,\epsilon)$ in Section A.1.1, and Assumption 3.b

$$\int_{q_L}^{q^K} f_l^o(q_l)dq_l + \int_{q^K}^{q^K\rho^{-\epsilon^o}} R_B(q_l,\epsilon^o)f_l^o(q_l)dq_l + \int_{q^K}^{q_H} f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o}R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o}dq =$$
$$= \int_{q_L}^{q^K} f_l^o(q_l)dq_l + \int_{q^K}^{q^K\rho^{-\epsilon}} R_B(q_l,\epsilon)f_l^o(q_l)dq_l + \int_{q^K}^{q_H} f_l^o(q\rho^{-\epsilon})\rho^{-\epsilon}R_2(q\rho^{-\epsilon},\epsilon)^{\eta}dq, \tag{21}$$

where the functions $R_B(.,\epsilon)$ and $R_2(.,\epsilon)$ are defined in Equations (15) and (16) in Section A.1.1 and $f_l^o(.)$ denotes the true counterfactual. Using Equation (20) and cancelling equal terms on both sides renders

$$\int_{q^K}^{q^K\rho^{-\epsilon^o}} R_B(q_l,\epsilon^o)f_l^o(q_l)dq_l = \int_{q^K}^{q^K\rho^{-\epsilon}} R_B(q_l,\epsilon)f_l^o(q_l)dq_l. \tag{22}$$

Denote by the functions $\eta(q,\epsilon)$ and $\eta_B(\epsilon)$ the η implicitly defined by the Equations (20) and (22). Formally, Condition 1 states that for all $\epsilon \neq \epsilon^o$ there exits a $q \in (\underline{q}, \overline{q}\rho^{\bar{\epsilon}})$ such that $\eta(q,\epsilon) \neq \eta_B(\epsilon)$ (see Appendix B.4 for a discussion of Condition 1). Therefore, the only pair (ϵ,η) that fulfils Equations (20) and (22) for all q is (ϵ^o,η^o). It follows that under Condition 1, $(\epsilon^o,\eta^o,\gamma^o)$ is the globally unique minimum of $Q(.)$.
qed.

Consider the parameter-space

$$\Theta = \{\theta \text{ s.t. } 0 \leq \epsilon \leq \bar{\epsilon}, \text{ and } 0 \leq \eta \leq \overline{\eta}, \text{ and } |\gamma_p| \leq M(\ln \overline{q} - \ln q^K)^{-p}\}. \tag{23}$$

Assume that $(\ln q^K - \ln \underline{q}) = (\ln \overline{q} - \ln q^K) > 1$.[29] Note that by Assumption 1, 2, 3.b, and Lemma 10, $\theta^o \in \Theta$. Consider the norm

$$d(\theta,\tilde{\theta}) = \mid \epsilon - \tilde{\epsilon} \mid + \mid \eta - \tilde{\eta} \mid + \sup_p \mid \gamma_p - \tilde{\gamma}_p \mid .$$

Lemma 2. *Under Assumptions 1, 2, 3.b, the parameter space Θ defined in Equation (23) is compact.*

Lemma 3. *Under Assumptions 1, 2, 3.b, the criterion $Q(\theta)$ is continuous at all*

[29] This assumption is without loss of generality since it is possible to use a transformation that rescales q such that the assumption holds. Moreover, convergence radii of power series are symmetric.

$\theta \in \Theta$.

The proofs of Lemmas 2 and 3 are in the Appendices B.5.1 and B.5.2.

By Lemma 1-3, $Q(\theta)$, θ^o, and Θ fulfil the Condition 3.1" for identification in Chen (2007), which proves Propositions 2. The cited results from Chen (2007) are summarised in Appendix F.

References

Acemoglu, Daron, Philippe Aghion, Leonardo Bursztyn, and David Hemous. 2012. "The environment and directed technical change." *American Economic Review*, 102(1): 131–166.

Anagol, Santosh, Allan Davids, and Benjamin Lockwood. 2022. "Diffuse Bunching with Frictions: Theory and Estimation." CEPR Discussion Papers.

Bachas, Pierre, and Mauricio Soto. 2021. "Corporate taxation under weak enforcement." *American Economic Journal: Economic Policy*, 13(4): 36–71.

Bertanha, Marinho, Andrew H McCallum, and Nathan Seegert. 2023. "Better bunching, nicer notching." *Journal of Econometrics*, 237(2): 105512.

Bertanha, Marinho, Carolina Caetano, Hugo Jales, and Nathan Seegert. 2023. "Bunching Estimation Methods." *Handbook of Labor, Human Resources, and Population Economics. Springer*.

Besley, Timothy, Neil Meads, and Paolo Surico. 2014. "The incidence of transaction taxes: Evidence from a stamp duty holiday." *Journal of Public Economics*, 119: 61–70.

Best, Michael Carlos, and Henrik Jacobsen Kleven. 2017. "Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK." *The Review of Economic Studies*, 85(1): 157–193.

Blomquist, Sören, and Whitney Newey. 2017. "The bunching estimator cannot identify the taxable income elasticity." National Bureau of Economic Research.

Blomquist, Sören, Whitney K Newey, Anil Kumar, and Che-Yuan Liang. 2021. "On bunching and identification of the taxable income elasticity." *Journal of Political Economy*, 129(8): 2320–2343.

Blundell, Richard, Antoine Bozio, and Guy Laroque. 2011. "Labor supply and the extensive margin." *American Economic Review*, 101(3): 482–486.

Bosch, Nicole, Vincent Dekker, and Kristina Strohmaier. 2020. "A data-driven procedure to determine the bunching window: an application to the Netherlands." *International Tax and Public Finance*, 27: 951–979.

Caetano, Carolina, Gregorio Caetano, and Eric Nielsen. 2024. "Correcting for Endogeneity in Models with Bunching." *Journal of Business & Economic Statistics*, 42(3): 851–863.

Card, David, David S Lee, Zhuan Pei, and Andrea Weber. 2015. "Inference on causal effects in a generalized regression kink design." *Econometrica*, 83(6): 2453–2483.

Chen, Xiaohong. 2007. "Large sample sieve estimation of semi-nonparametric models." *Handbook of Econometrics*, 6: 5549–5632.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2013. "Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities." *NBER macroeconomics Annual*, 27(1): 1–56.

Chetty, Raj, John N Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records." *The Quarterly Journal of Economics*, 126(2): 749–804.

Cremer, Helmuth, Firouz Gahvari, and Norbert Ladoux. 1998. "Externalities and optimal taxation." *Journal of Public Economics*, 70(3): 343–364.

De Groote, Olivier, and Frank Verboven. 2019. "Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems." *American Economic Review*, 109(6): 2137–72.

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

Einmahl, Uwe, and David M Mason. 2005. "Uniform in bandwidth consistency of kernel-type function estimators." *The Annals of Statistics*, 33(3): 1380.

Feger, Fabian, Nicola Pavanini, and Doina Radulescu. 2022. "Welfare and redistribution in residential electricity markets with solar power." *The Review of Economic Studies*, 89(6): 3267–3302.

Gautier, Eric, and Christophe Gaillac. 2021. "Estimates for the SVD of the Truncated Fourier Transform on L2 (cosh (b.)) and Stable Analytic Continuation." *Journal of Fourier Analysis and Applications*.

Gelber, Alexander M, Damon Jones, Daniel W Sacks, and Jae Song. 2021. "Using nonlinear budget sets to estimate extensive margin responses: Method and evidence from the earnings test." *American Economic Journal: Applied Economics*, 13(4): 150–193.

Gerarden, Todd. 2022. "Demanding innovation: The impact of consumer subsidies on solar panel production costs." *Managment Science*.

Gerard, François, Miikka Rokkanen, and Christoph Rothe. 2020. "Bounds on treatment effects in regression discontinuity designs with a manipulated running variable." *Quantitative Economics*, 11(3): 839–870.

Germeshausen, Robert. 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*, , (57).

Goff, Leonard. 2022. "Treatment Effects in Bunching Designs: The Impact of the Federal Overtime Rule on Hours."

Hughes, Jonathan E, and Molly Podolefsky. 2015. "Getting green with solar subsidies: evidence from the California solar initiative." *Journal of the Association of Environmental and Resource Economists*, 2(2): 235–275.

Iaria, Alessandro, and Ao Wang. 2024. "Real Analytic Discrete Choice Models of Demand: Theory and Implications." *Econometric Theory*, 1–49.

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

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden. 2013. "Optimal redistributive taxation with both extensive and intensive responses." *Journal of Economic Theory*, 148(5): 1770–1805.

Kaplow, By Louis. 2012. "Optimal control of externalities in the presence of income taxation." *International Economic Review*, 53(2): 487–509.

Kaplow, Louis. 1996. "The optimal supply of public goods and the distortionary cost of taxation." *National Tax Journal*, 513–533.

Kaplow, Louis. 2008. "Optimal policy with heterogeneous preferences." *The BE Journal of Economic Analysis & Policy*, 8(1).

Kleven, Henrik. 2024. "The EITC and the extensive margin: A reappraisal." *Journal of Public Economics*, 236: 105135.

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics*, 8: 435–464.

Kleven, Henrik Jacobsen, Camille Landais, Emmanuel Saez, and Esben Schultz. 2013. "Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in Denmark." *The Quarterly Journal of Economics*, 129(1): 333–378.

Kleven, Henrik J, and Mazhar Waseem. 2013. "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan." *The Quarterly Journal of Economics*, 128(2): 669–723.

Kopczuk, Wojciech, and David Munroe. 2015. "Mansion tax: The effect of transfer taxes on the residential real estate market." *American economic Journal: economic policy*, 7(2): 214–57.

Krantz, Steven G, and Harold R Parks. 2002. *A primer of real analytic functions.* Springer Science & Business Media.

Langer, Ashley, and Derek Lemoine. 2022. "Designing dynamic subsidies to spur adoption of new technologies." *Journal of the Association of Environmental and Resource Economists*, 9(6): 1197–1234.

Marx, Benjamin M. 2024. "Dynamic bunching estimation with panel data." *Journal of Econometric Methods*, , (0).

McCallum, Andrew H, and Michael Navarrete. 2022. "Why Don't Taxpayers Bunch at Kink Points?" *Available at SSRN 4219884*.

Myhre, Andreas. 2021. "Intensive and Extensive Margin Labor Supply Responses to Kinks in Disability Insurance Programs." *Available at SSRN 3914055*.

Nemet, Gregory F. 2019. *How solar energy became cheap: A model for low-carbon innovation.* Routledge.

Newey, Whitney K., and Daniel McFadden. 1994. "Chapter 36 Large sample estimation and hypothesis testing." In . Vol. 4 of *Handbook of Econometrics*, 2111–2245. Elsevier.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher Taber. 2010. "Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform." *American Economic Journal: Economic Policy*, 2(2): 185–215.

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

Rochet, Jean-Charles, and Lars A Stole. 2002. "Nonlinear pricing with random participation." *The Review of Economic Studies*, 69(1): 277–311.

Rochet, Jean Charles, and Philippe Chone. 1998. "Ironing, sweeping, and multidimensional screening." *Econometrica*, 783–826.

Ruh, Philippe, and Stefan Staubli. 2019. "Financial incentives and earnings of disability insurance recipients: Evidence from a notch design." *American Economic Journal: Economic Policy*, 11(2): 269–300.

Saez, Emmanuel. 2001. "Using elasticities to derive optimal income tax rates." *The Review of Economic Studies*, 68(1): 205–229.

Saez, Emmanuel. 2002. "Optimal income transfer programs: Intensive versus extensive labor supply responses." *Quarterly Journal of Economics*, 117(3): 1039 – 1073.

Saez, Emmanuel. 2010. "Do taxpayers bunch at kink points?" *American Economic Journal: Economic Policy*, 2(3): 180–212.

Samuelson, Paul A. 1954. "The pure theory of public expenditure." *The review of economics and statistics*, 387–389.

Slemrod, Joel, Caroline Weber, and Hui Shan. 2017. "The behavioral response to housing transfer taxes: Evidence from a notched change in DC policy." *Journal of Urban Economics*, 100: 137–153.

Srivastav, Sugandha. 2023. "Bringing Breakthrough Technologies to Market: Solar Power and Feed-in-Tariffs." Mimeo.

Übertragungsnetzbetreiber. 2016. "EEG-Jahresabrechnung 2016." https://www.netztransparenz.de/portals/1/EEG-Jahresabrechnung_2016.pdf, Last accessed: 2017-09-30.

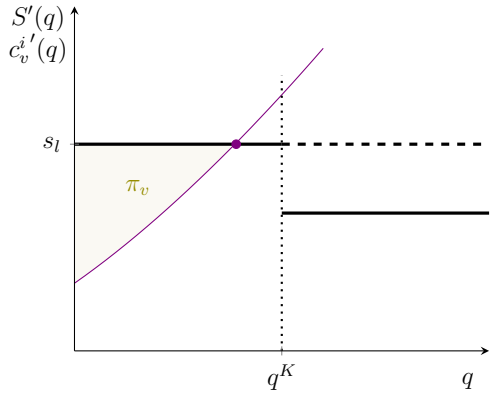
Übertragungsnetzbetreiber. 2018. "EEG-Vergütungskategorien 2018." https://www.netztransparenz.de/EEG/Verguetungs-und-Umlagekategorien, Last accessed: 2018-09-30.

Supplemental appendices for online publication

B Details model

B.1 The illustration of adopters below the kink point

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



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

B.2 A detailed proof of Proposition 1

B.2.1 An equivalent structural formulation of Assumptions 1 and 2

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

$$\epsilon^i(s) = \frac{\mathrm{d}\ln q^i(s)}{\mathrm{d}\ln s}. \tag{24}$$

Formally, using this notation, Assumption 1 states that for all marginal subsidy rates s in the interval $[\rho s_l, s_l]$ and for all adopters i such that their capacity-choice under the counterfactual subsidy $q^i(s_l)$ is in an interval $[\underline{q}, \overline{q}]$ around the kink point q^K, it holds that the elasticity $\epsilon^i(s) = \epsilon$, where ϵ is a constant. Assumption 1 a is a reduced form assumption on an endogenous object, i.e., a high-level assumption. However, it is equivalent to the structural assumption 4:

Assumption 4 (Structural assumption intensive margin)**.** *The cost function is locally isoelastic, and the elasticity is bounded. Formally, for all adopters i such that their capacity-choice under the linear subsidy $q^i(s_l)$ is in the interval $[\underline{q}, \overline{q}]$, and for all quantities q in the interval $[q^i(s_l\rho), q^i(s_l)]$, it holds that the variable cost function $c_v^i(.)$ is equal to*

$$c_v^i(q) = \theta^i q^{1+\frac{1}{\epsilon}}, \tag{25}$$

where θ^i is the variable cost type.

Lemma 4 (Equivalence intensive margin)**.** *Assumption 1 is equivalent to Assumption 4.*

The proof is in Appendix B.2.2.

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

Corollary 1 (Type parameters)**.** *For each adopter i, there is a one-to-one mapping from the variable and fixed cost type (θ^i, c_f^i) to the choice and total cost under the counterfactual subsidy (q_l^i, c_t^i):*

$$\theta^i = \frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon}, \tag{26}$$

$$c_f^i = c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}. \tag{27}$$

Therefore, locally, the total cost function is equal to

$$c(q, q_l, c_t) = \underbrace{\frac{s_l}{{q_l}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}}}_{\textit{variable cost}} + \underbrace{c_t - q_l s_l \frac{\epsilon}{1+\epsilon}}_{\textit{fixed cost}}. \tag{28}$$

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

The proof is in Appendix B.2.3. Using (q_l, c_t) has the advantage that the type parameter has direct economic meaning. The type is equal to the choice and cost under the counterfactual subsidy. The mapping between (θ, c_f) and (q_l, c_t) depends on the counterfactual subsidy rate s_l. However, since s_l is observable and fixed, this dependence poses no problem. Note that I will drop the adopter-specific index i from now on.

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

$$\eta(S, q) = \frac{f_{t|q_l}(S(q)|q)}{F_{t|q_l}(S(q)|q)} S(q), \tag{29}$$

where $S(.)$ is a general subsidy function. The participation margin elasticity under the counterfactual subsidy is $\eta(S_l, q_l)$.[30] Formally, using this notation, Assumption 2 states that for all subsidy functions $S(q)$ such that $S_l(q) \geq S(q) \geq S_k(q)$ and for all quantities q in an interval $[\underline{q}, \overline{q}]$ around the kink point, it holds that the functional $\eta(S, q) = \eta$, where η is the constant participation margin elasticity. Again, Assumption 2 is a high-level assumption. It is equivalent to the structural assumption 5:

Assumption 5 (Structural assumption participation margin)**.** *The conditional CDF of the total cost is locally isoelastic, and the elasticity is bounded. Formally, for all values c_t in the interval $[S_k(q_l), S_l(q_l)]$, the conditional CDF of the total cost is equal to*

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)} \right)^{\eta}, \tag{30}$$

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

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

Lemma 5 (Equivalence participation margin)**.** *Assumption 2 is equivalent to Assumption 5.*

The proof is in Appendix B.2.4. Note that the normalisation term $\overline{c_t}(q_l)$ in Assumption 5 may depend on q_l. Therefore, the assumption does not rule out a correlation between the variable cost type q_l and the total cost type c_t in Corollary 1. Consequently, Assumption 1 and 2 do not rule out a correlation between fixed and variable costs.

B.2.2 Proof of Lemma 4

Assumption 1 $\Rightarrow$ Assumption 4:
By Assumption 1 and the definition of the elasticity

$$\epsilon = \frac{q^{i\prime}(s)\, s}{q^i(s)}, \text{ for all } s \text{ in } [s_l\rho, s_l]. \tag{31}$$

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

$$q^{i\prime}(s) = \frac{1}{c_v^{i\,\prime\prime}(q^i(s))}. \tag{32}$$

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

$$\epsilon = \frac{c_v^{i\,\prime}(q)}{c_v^{i\,\prime\prime}(q)\, q}. \tag{33}$$

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 (33) form an initial value problem with solution

$$c_v^i(q) + c_f^i = \underbrace{\frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}}}_{\text{variable cost}} + \underbrace{c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}}_{\text{fixed cost}}. \tag{34}$$

The result follows by defining

$$\theta^i = \frac{s_l}{{q_l^i}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon}, \tag{35}$$

$$c_f^i = c_t^i - q_l^i s_l \frac{\epsilon}{1+\epsilon}. \tag{36}$$

Assumption 4 $\Rightarrow$ Assumption 1:
By the FOC

$$q^i(s) = \left(\frac{\epsilon}{(1+\epsilon)\theta^i}\right)^{\epsilon} s^{\epsilon}. \tag{37}$$

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

B.2.3 Proof of 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 (25) gives the result.
qed.

B.2.4 Proof of Lemma 5

Assumption 2 $\Leftrightarrow$ Assumption 5:
By Assumption 2, for all $S(q)$ such that $S_l(q) \geq S(q) \geq S_k(q)$, and for all q in $[\underline{q}, \overline{q}]$,

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

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

$$\eta = \frac{f_{t|q_l}(c_t|q_l)c_t}{F_{t|q_l}(c_t|q_l)}. \tag{39}$$

The solution of this partial differential equation is

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)}\right)^{\eta}, \tag{40}$$

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

B.2.5 Solving the maximisation problem (1)

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

Lemma 6. *The choice under the kinked subsidy q_k as a function of the choice under the linear subsidy q_l is*

$$q_k(q_l) = q_l, \qquad \textit{for } q_l < q^K; \tag{41}$$
$$q_k(q_l) = q^K, \qquad \textit{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; \tag{42}$$
$$q_k(q_l) = q_l\rho^{\epsilon}, \qquad \textit{for } q_l > q^K\rho^{-\epsilon}. \tag{43}$$

PROOF:
By Equation (28) and the first order condition of the adopters' maximisation problem

$$q(s, q_l) = q_l \left(\frac{s}{s_l}\right)^{\epsilon}. \tag{44}$$

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

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

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

$$q_k(q_l) = q_l\rho^{\epsilon}, \text{ for } q_l \gg q^K. \tag{46}$$

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

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

Lemma 7. *The difference in cost $\Delta c(q_l)$ of adopter q_l between the kinked and linear subsidy is*

$$\Delta c(q_l) = 0, \qquad \textit{for } q_l < q^K; \tag{47}$$
$$\Delta c(q_l) = \frac{1}{1+\epsilon^{-1}} \frac{s_l}{q_l^{1/\epsilon}} ((q^K)^{1+\epsilon^{-1}} - q_l^{1+\epsilon^{-1}}), \qquad \textit{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; \tag{48}$$
$$\Delta c(q_l) = \frac{1}{1+\epsilon^{-1}} s_l q_l (\rho^{\epsilon+1} - 1), \qquad \textit{for } q_l > q^K\rho^{-\epsilon}. \tag{49}$$

PROOF:
By Corollary 1

$$c(q, q_l, c_t) = c_t + \left[\frac{q^{1+1/\epsilon}}{q_l^{1/\epsilon}} - q_l\right] \frac{s_l}{1+1/\epsilon}. \tag{50}$$

By definition and Lemma 6

$$\Delta c(q_l) = c(q_k(q_l), q_l, c_t) - c_t = c(q_l, q_l, c_t) - c_t \qquad \text{for } q_l < q^K; \tag{51}$$
$$= c(q^K, q_l, c_t) - c_t \qquad \text{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; \tag{52}$$
$$= c(q_l\rho^{\epsilon}, q_l, c_t) - c_t \qquad \text{for } q_l > q^K\rho^{-\epsilon}. \tag{53}$$

Use Equation (50) in Equation (51)-(53) to get Equation (47)-(49).
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:

$$R(q_l) = \frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}. \tag{54}$$

Lemma 8. *Define*

$$R_1(q_l, \epsilon) = 1; \tag{55}$$
$$R_B(q_l, \epsilon) = \frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 - \left(\frac{q^K}{q_l}\right)^{\frac{1+\epsilon}{\epsilon}}\right); \tag{56}$$
$$R_2(q_l, \epsilon) = (1-\rho)\frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 + \frac{\rho^{\epsilon+1}}{\epsilon}\right). \tag{57}$$

The function $R(q_l)$ is:

$$R(q_l) = R_1(q_l, \epsilon), \qquad \textit{for } q_l < q^K; \tag{58}$$
$$R(q_l) = R_B(q_l, \epsilon), \qquad \textit{for } q_l \in [q^K, q^K\rho^{-\epsilon}]; \tag{59}$$
$$R(q_l) = R_2(q_l, \epsilon), \qquad \textit{for } q_l > q^K\rho^{-\epsilon}. \tag{60}$$

PROOF:
By definition

$$R(q_l) = \frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}, \tag{61}$$

which together with Lemma 6 and 7 gives Equation (58)-(60).
qed.

Lemma 9. *The mass of participating adopters under the kinked subsidy as a function of q_l is*

$$R(q_l)^{\eta} f_l(q_l), \tag{62}$$

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

$$\frac{F_{t|q_l}(S_k(q_k(q_l)) - \Delta c(q_l)|q_l)}{F_{t|q_l}(S_l(q_l)|q_l)} f_l(q_l), \tag{63}$$

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

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{\overline{c_t}(q_l)}\right)^\eta, \text{ for all } c_t \text{ in } [S_k(q_l), S_l(q_l)]. \tag{64}$$

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

$$\frac{F_{t|q_l}(S_k(q_k(q_l)) - \Delta c(q_l)|q_l)}{F_{t|q_l}(S_l(q_l)|q_l)} f_l(q_l) = \left(\frac{S_k(q_k(q_l)) - \Delta c(q_l)}{S_l(q_l)}\right)^\eta f_l(q_l), \tag{65}$$

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

PROOF of Proposition 1:
Change variable in Equation (62) using Lemma 6 to derive Equation (5) and (7). Integrate Equation (62) over $[q_L, q_H\rho^{-\epsilon}]$ to derive Equation (6). Remember, to avoid an overloaded notation, the rest of the paper denotes the choice under the kinked subsidy simply by q.
qed.

B.3 Lemma 10

Lemma 10. .

- *Assumption 3.a implies Assumption 3.b.*
- *Assumption 3.b implies that*

$$|\gamma_p| \leq M \frac{1}{(\ln \overline{q} - \ln q^K)^p} \text{ for all } p\,, \tag{66}$$

where $\gamma_p = \frac{d^{(p)} \ln f_l(q^K)}{d \ln q_l^{(p)}}/p!$. *Moreover,* $\ln f_l(.)$ *is real analytic in* $\ln q_l$ *on the interval* $(\underline{q}, \overline{q})$.

Note that Assumption 3.b is slightly weaker than Assumption 3.a because it does not imply explicit bounds on the derivatives for all $q \in (\underline{q}, \overline{q})$.

PROOF:
Suppose Assumption 3.a holds. By Proposition 1.2.12 in Krantz and Parks (2002), $\ln f_l(.)$ is real analytic on the interval $(\underline{q}, \overline{q})$. The cited results from Krantz and Parks (2002) are summarised in Appendix F. By Lemma 1.1.8 in Krantz and Parks (2002), and due to the bounds on the derivatives, the power series representation at the kink point converges for all $q \in (\underline{q}, \overline{q})$. By Corollary 1.2.4 in Krantz and Parks (2002), the function defined by the power series representation is real analytic on $(\underline{q}, \overline{q})$. By definition, all derivatives of the power series representation and of $\ln f_l(.)$ are equal at the kink point. Therefore, by Corollary 1.2.5 in Krantz and Parks

(2002), the power series representation and $\ln f_l(.)$ are equal on $(\underline{q}, \overline{q})$. It follows that Assumption 3.b holds.

Suppose Assumption 3.b holds. By Corollary 1.2.4 in Krantz and Parks (2002), the function defined by the power series representation is real analytic on $(\underline{q}, \overline{q})$. Since it is equal to $\ln f_l(.)$, also $\ln f_l(.)$ is real analytic on $(\underline{q}, \overline{q})$. Corollary 1.1.10 in Krantz and Parks (2002) gives the bounds of γ_p.
qed.

B.4 Rank conditions

This section denotes the true functions and parameters in population with superscript "o", while functions and parameters without superscript denote the general values of these objects.

B.4.1 A formal version of Condition 1

Note that by Proposition 1, the mass exactly at the kink point q^K as a function of the parameters ϵ and η and the true counterfactual $f_l^o(.)$ is

$$\int_{q^K}^{q^K\rho^{-\epsilon}} R_B(q_l, \epsilon)^\eta f_l^o(q_l) dq_l, \tag{67}$$

where $R_B(., \epsilon)$ is defined in Equation (15) in Section A.1.1.[31] Note that the measure above the kink point in Equation (7) in Proposition 1 as a function of the parameters ϵ and η and the true counterfactual $f_l^o(.)$ is

$$f_l^o(q\rho^{-\epsilon})\rho^{-\epsilon} R_2(q\rho^{-\epsilon}, \epsilon)^\eta, \tag{68}$$

where $R_2(., \epsilon)$ is defined in Equation (16) in Section A.1.1.

Use the expressions (67) and (68) and evaluate them at the true and candidate values of ϵ and η. For expression (68), consider its real analytic continuation. Denote by $\eta_B(\epsilon)$ and $\eta(q, \epsilon)$ the η implicitly defined by the equations

$$\int_{q^K}^{q^K\rho^{-\epsilon^o}} R_B(q_l, \epsilon^o) f_l^o(q_l) dq_l = \int_{q^K}^{q^K\rho^{-\epsilon}} R_B(q_l, \epsilon) f_l^o(q_l) dq_l; \tag{69}$$

$$f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o} R_2(q\rho^{-\epsilon^o}, \epsilon^o)^{\eta^o} = f_l^o(q\rho^{-\epsilon})\rho^{-\epsilon} R_2(q\rho^{-\epsilon}, \epsilon)^\eta \; \forall q \in (\underline{q}, \overline{q}\rho^{\overline{\epsilon}}). \tag{70}$$

Formally, Condition 1 states that for all $\epsilon \neq \epsilon^o$ there exits a $q \in (\underline{q}, \overline{q}\rho^{\overline{\epsilon}})$ such that $\eta(q, \epsilon) \neq \eta_B(\epsilon)$.

B.4.2 Sufficient conditions for Condition 1

Lemma 11. *Each of the following properties are sufficient but not necessary for Condition 1 to hold:*

[31] Note that this mass is part of but not equal to B in Proposition 1, which is the mass in the entire interval $[q_L, q_H]$ and not only at the point q^K.

(a) For all $\epsilon \neq \epsilon^o$, the true counterfactual $f_l^o(.)$ does not solve the functional equation:

$$\ln f_l^o(q) = \ln f_l^o(q\rho^{-(\epsilon-\epsilon^o)}) + \ln \rho^{-(\epsilon-\epsilon^o)} + \eta_B(\epsilon) \ln R_2(q\rho^{-(\epsilon-\epsilon^o)}, \epsilon) - \eta^o \ln R_2(q, \epsilon^o) \;\; \forall q \in (\underline{q}, \overline{q}\rho^{\overline{\epsilon}}), \tag{71}$$

where $R_2(., \epsilon)$ is defined in Equation (16) in Section A.1.1 and $\eta_B(\epsilon)$ is implicitly defined by Equation (69). Generically, the functional equation (71) does not hold. Consequently, Condition 1 holds generically.

(b) $\frac{d \ln f_l^o(q)}{d \ln q} \leq -1$ for all $q \in (\underline{q}, \overline{q})$.

(c) The participation elasticity η^o is known, which, in particular, includes the case $\eta^o = 0$.

(d) $\frac{d \ln f_{k2}^o(q)}{d \ln q} \geq 0$ for all $q \in (\underline{q}, \overline{q}\rho^{\overline{\epsilon}})$, where $f_{k2}^o(.)$ denotes the observable measure above the kink point and its analytic continuation to points below the kink point.[32]

(e) The function $f_l^o(.)$ is real analytic in q on the interval $(0, \overline{q})$ and there exits an order of derivative P such that

$$\lim_{q \downarrow 0} \frac{d^P f_l^o(q)}{dq^P} \neq 0 \text{ or } \pm \infty. \tag{72}$$

(f) The functions $f_l^o(.)$ or $\ln f_l^o(.)$ are real analytic in q on the interval $[0, \overline{q})$.

(g) There exists an order P such that

$$\ln f_l^o(q_l) = \sum_{p=0}^{P} \gamma_p^o \left(\ln \frac{q_l}{q^K} \right)^p \text{ for all } q_l \in (\underline{q}, \overline{q}). \tag{73}$$

B.4.3 The local rank condition and local identification

The following local version of Condition 1 allows discussing local identification.

Condition 2 (Local rank condition)**.** *There exists a $q \in (\underline{q}, \overline{q}\rho^{\overline{\epsilon}})$ such that*

$$\eta_B'(\epsilon^o) - \frac{\partial \eta(q, \epsilon^o)}{\partial \epsilon} \neq 0, \tag{74}$$

where $\eta_B(\epsilon)$ and $\eta(q, \epsilon)$ are implicitly defined by Equations (69) and (70).

Under Condition 2, Condition 1 holds locally at (ϵ^o, η^o).

Lemma 12. *Each of the following properties is sufficient but not necessary for Condition 2 to hold:*

(a) Properties (b), (c), and (d) in Lemma 11.

[32] Since $\frac{d \ln f_{k2}^o(q)}{d \ln q} = \frac{d \ln f_l^o(q\rho^{-\epsilon^o})}{d \ln q} + \eta^o \frac{d \ln R(q\rho^{-\epsilon^o})}{d \ln q}$ the property indirectly restricts the shape of the function $f_l^o(.)$.

(b)

$$\left(\frac{d\ln f_l^o(\hat{q}(\epsilon^o)\,\rho^{-\epsilon^o})}{d\ln q}+1\right)\ln(1/\rho)+\eta^o\frac{(1-\rho^{1+\epsilon^o})-(1+\epsilon^o)\ln(1/\rho)}{(1+\epsilon^o)^2}\neq 0, \qquad (75)$$

where $f_l^o(.)$ is the true counterfactual, $R_2(.,\epsilon)$ is defined in Equation (16) in Section A.1.1, and $\hat{q}(\epsilon^o)$ is the unique q defined by $R_2(\hat{q}(\epsilon^o)\rho^{-\epsilon^o},\epsilon^o)=1$. This property holds generically. Consequently, Condition 2 holds generically.

Proposition 3 (Local identification)**.** *Under Assumptions 1, 2, and 3.b, the observable measure $f_k(.)$ identifies the counterfactual measure $f_l(.)$ for all q. Under Condition 2, the intensive margin elasticity ϵ and the participation margin elasticity η are locally identified.*

The proof is in Section B.5.3.

B.4.4 Some economic intuition why the rank conditions hold

There is an economic argument as to why Conditions 1 and its local counterpart Condition 2 defined in Section B.4.3 hold. Denote the right-hand side of Equation (69) in Section B.4.1 as a function of its parameters by $\tilde{B}(\epsilon,\eta)$; denote the right-hand side of Equation (70) as a function of its parameters by $f_{k2}(q|\epsilon,\eta)$. On the one hand, the mass at the kink point $\tilde{B}$ depends strongly on ϵ, as ϵ determines the mass of adopters who potentially bunch. For this reason, the upper bound of the integral in Equation (69) is a function of ϵ. Additionally, $\tilde{B}$ depends only weakly on η. The dependence is through the power of R_B, and R_B is close to one. This is because R_B is roughly one minus the profit loss from re-optimisation. Due to the Envelope Theorem, the profit loss is of second order and hence relatively small. The strong dependence on ϵ and the weak dependence on η imply that

$$\frac{\frac{\partial\tilde{B}(\epsilon,\eta)}{\partial\epsilon}}{-\frac{\partial\tilde{B}(\epsilon,\eta)}{\partial\eta}} \qquad (76)$$

is large. On the other hand, the measure above the kink point f_{k2} mainly depends on η. To see that, consider the elasticity of the function f_{k2}. From Equation (70) it follows that:

$$\frac{d\ln f_{k2}(q|\epsilon,\eta)}{d\ln q}=\frac{d\ln f_l^o(q\,\rho^{-\epsilon})}{d\ln q}+\eta\,\frac{d\ln R_2(q\,\rho^{-\epsilon},\epsilon)}{d\ln q}. \qquad (77)$$

Usually, $\frac{d\ln f_l^o}{d\ln q}$ is approximately constant. Moreover, for q close to q^K, $\frac{d\ln R_2}{d\ln q}$ is approximately $-(1-\rho)$. Therefore, it is approximately constant as well. It follows that $\frac{\partial\frac{d\ln f_{k2}(q^K|\epsilon,\eta)}{d\ln q}}{\partial\epsilon}$ is close to zero. It implies that

$$\frac{\frac{\partial\left(\frac{d\ln f_{k2}(q^K|\epsilon,\eta)}{d\ln q}\right)}{\partial\epsilon}}{-\frac{\partial\left(\frac{d\ln f_{k2}(q^K|\epsilon,\eta)}{d\ln q}\right)}{\partial\eta}}$$

is close to zero. These properties of $\tilde{B}$ and $\frac{d\ln f_{k2}}{d\ln q}$ assure that Conditions 1 and 2 hold. Section D.6.4 evaluates the conditions at the estimates of the application of this paper. It confirms the intuition outlined above: the rank condition holds by a large amount.

B.4.5 Proof of Lemma 11 and Lemma 12

As a first step, consider the following intermediate result:

Lemma 13. *The function $\eta'_B(\epsilon) > 0$ and continuous. In particular, $\eta'_B(\epsilon_o)$ is a strictly positive real number.*

PROOF Lemma 13:
Denote the right-hand side of Equation (69) as a function of its parameters by $\tilde{B}(\epsilon, \eta)$. Implicitly differentiate Equation (69) to derive

$$\eta'_B(\epsilon) = \frac{\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \epsilon}}{-\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \eta}}. \tag{78}$$

The numerator

$$-\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \eta} = \int_{q^K}^{q^K\rho^{-\epsilon}} R_B(q_l,\epsilon)^\eta f_l^o(q_l)(-\ln R_B(q_l,\epsilon))dq_l > 0, \tag{79}$$

since $1 > R_B(q_l, \epsilon) > 0$ for $q^K < q_l < q^K\rho^{-\epsilon}$. Simple algebra using an appropriate software shows that $\frac{\partial R_B(q_l,\epsilon)}{\partial \epsilon} > 0$ for all $q_l > q^K$. It follows that

$$\begin{aligned}\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \epsilon} = &\int_{q^K}^{q^K\rho^{-\epsilon}} \eta R_B(q_l,\epsilon)^{(\eta-1)} f_l^o(q_l)\frac{\partial R_B(q_l,\epsilon)}{\partial \epsilon}dq_l + \\ &R_B(q^K\rho^{-\epsilon},\epsilon)^\eta f_l^o(q^K\rho^{-\epsilon})q^K\rho^{-\epsilon}\ln(1/\rho) > 0,\end{aligned} \tag{80}$$

since $1 > R_B(q_l, \epsilon) > 0$ for $q^K < q_l < q^K\rho^{-\epsilon}$. Both functions are continuous in ϵ and η and well defined at ϵ^o and η^o.
qed.

PROOF, Point (c), Lemma 11:
By Lemma 13, $\eta'_B(\epsilon) > 0$ and continuous. Moreover, $\eta_B(\epsilon^o) = \eta^o$. Therefore, $\eta_B(\epsilon)$ intersects with the known η^o once.
qed.

PROOF, Point (b), Lemma 12:
Denote the right-hand side of Equation (70) as a function of its parameters by $f_{k2}(q|\epsilon, \eta)$. Implicitly differentiate the logarithm of Equation (70) to derive

$$\frac{\partial \eta(q,\epsilon)}{\partial \epsilon} = \frac{\frac{\partial \ln f_{k2}(q|\epsilon,\eta)}{\partial \epsilon}}{-\frac{\partial \ln f_{k2}(q|\epsilon,\eta)}{\partial \eta}} = \frac{\left(\frac{d\ln f_l^o(q\ \rho^{-\epsilon})}{d\ln q}+1\right)\ln(1/\rho) + \eta\frac{\frac{\partial R_2(q\rho^{-\epsilon},\epsilon)}{\partial q}q\rho^{-\epsilon}\ln(1/\rho)+\frac{\partial R_2(q\rho^{-\epsilon},\epsilon)}{\partial \epsilon}}{R_2(q\rho^{-\epsilon},\epsilon)}}{-\ln R_2(q\rho^{-\epsilon},\epsilon)}. \tag{81}$$

Note that R_2 is strictly decreasing in its first argument. For each ϵ there exists a unique $\hat{q}(\epsilon)$ such that $R_2(\hat{q}(\epsilon)\rho^{-\epsilon}, \epsilon) = 1$ and $q^K < \hat{q}(\epsilon)\rho^{-\epsilon} < q^K\rho^{-\epsilon}$. Evaluate the numerator at $\hat{q}(\epsilon^o)$, ϵ^o and η^o, which after some algebra using an appropriate software renders

$$\left(\frac{d\ln f_l^o(\hat{q}(\epsilon^o)\ \rho^{-\epsilon^o})}{d\ln q}+1\right)\ln(1/\rho) + \eta^o\frac{(1-\rho^{1+\epsilon^o})-(1+\epsilon^o)\ln(1/\rho)}{(1+\epsilon^o)^2}. \tag{82}$$

Consider the case when this condition is not equal to zero, which is true generically. First, consider the case when the condition is strictly negative. It follows that

$$\lim_{q \downarrow \hat{q}(\epsilon^o)} \frac{\partial \eta(\epsilon^o, q)}{\partial \epsilon} = -\infty \neq \eta'_B(\epsilon_o), \tag{83}$$

since $-\ln\left(R_2(q\rho^{-\epsilon}, \epsilon)\right) > 0$ for all $q > \hat{q}(\epsilon^o)$ and $\lim_{q \downarrow \hat{q}(\epsilon^o)} -\ln\left(R_2(q\rho^{-\epsilon}, \epsilon)\right) = 0$. Second, consider the case when the condition is strictly positive. It follows that

$$\lim_{q \uparrow \hat{q}(\epsilon^o)} \frac{\partial \eta(\epsilon^o, q)}{\partial \epsilon} = -\infty \neq \eta'_B(\epsilon_o), \tag{84}$$

since $-\ln\left(R_2(q\rho^{-\epsilon}, \epsilon)\right) < 0$ for all $q < \hat{q}(\epsilon^o)$ and $\lim_{q \uparrow \hat{q}(\epsilon^o)} -\ln\left(R_2(q\rho^{-\epsilon}, \epsilon)\right) = 0$.
qed.

PROOF, Point (d), Lemma 11:
Note that the left-hand side of Equation (70) is equal to $f^o_{k2}(q)$. Change variable in Equation (70) and solve for η to derive the function

$$\eta_l(q_l, \epsilon) = \frac{\ln f^o_{k2}(q_l \rho^\epsilon) + \ln \rho^\epsilon - \ln f^o_l(q_l)}{\ln R_2(q_l, \epsilon)} \text{ for all } q_l \in (\underline{q}\rho^{-\bar{\epsilon}}, \overline{q}). \tag{85}$$

Consider functions $q_l(\epsilon)$ such that $R_2(q_l(\epsilon), \epsilon) = const.$. Differentiate the function above and note that the denominator is constant:

$$\frac{d\ \eta_l(q_l(\epsilon), \epsilon)}{d\epsilon} = \frac{\frac{d \ln f_{k2}(q_l(\epsilon)\rho^\epsilon)}{d \ln q}\left(\frac{q'_l(\epsilon)}{q_l(\epsilon)} + \ln \rho\right) + \ln \rho - \frac{d \ln f^o_l(q_l(\epsilon))}{d \ln q}\left(\frac{q'_l(\epsilon)}{q_l(\epsilon)}\right)}{\ln R_2(q_l(\epsilon), \epsilon)}. \tag{86}$$

Denote by $\hat{q}_l(\epsilon)$ the unique q_l such that $R_2(\hat{q}_l(\epsilon), \epsilon) = 1$. Note that $q^K \leq \hat{q}_l(\epsilon) \leq q^K \rho^{-\epsilon}$. Simple algebra using appropriate software reveals that

$$\frac{\hat{q}'_l(\epsilon)}{\hat{q}_l(\epsilon)} + \ln \rho < 0 \text{ and } \frac{\hat{q}'_l(\epsilon)}{\hat{q}_l(\epsilon)} > 0. \tag{87}$$

The elasticity $\frac{d \ln f_{k2}(\hat{q}_l(\epsilon)\rho^\epsilon)}{d \ln q} \geq 0$ by assumption. It also implies $\frac{d \ln f^o_l(\hat{q}_l(\epsilon))}{d \ln q} > 0$ since $\frac{d \ln f^o_l(\hat{q}_l(\epsilon))}{d \ln q} = \frac{d \ln f_{k2}(\hat{q}_l(\epsilon)\rho^{\epsilon^o})}{d \ln q} - \frac{d \ln R_2(\hat{q}_l(\epsilon), \epsilon^o)}{\partial \ln q}$ and $\frac{d \ln R_2(\hat{q}_l(\epsilon), \epsilon^o)}{\partial \ln q} < 0$. It follows that $\lim_{q_l(\epsilon) \uparrow \hat{q}_l(\epsilon)} \frac{d\ \eta_l(q_l(\epsilon), \epsilon)}{d\epsilon} < 0$ for all ϵ since $\ln \rho < 0$, and $\ln R_2(q_l(\epsilon), \epsilon) > 0$ for all $q_l(\epsilon) < \hat{q}_l(\epsilon)$.[33] Therefore, since $\eta'_B(\epsilon) > 0$, Conditions 2 and 1 hold.
qed.

PROOF, Point (b), Lemma 11:
Denote the left-hand side of Equation (70) by $f^o_{k2}(q)$. Use Equation (70) and explicitly solve

[33] Note that $\overline{q} > \hat{q}_l(\epsilon) > \underline{q}\rho^{-\bar{\epsilon}}$, $\overline{q}\rho^{\bar{\epsilon}} > \hat{q}_l(\epsilon)\rho^\epsilon > \underline{q}$, and $\overline{q}\rho^{\bar{\epsilon}} > \hat{q}_l(\epsilon)\rho^{\epsilon^o} > \underline{q}$. Therefore, all used relations hold in the relevant range. The three inequalities follow since: by assumption, $\frac{\overline{q}}{q^K} > \rho^{-\bar{\epsilon}}$; by the property of the convergence radius of power series $\frac{\overline{q}}{q^K} = \frac{q^K}{\underline{q}} > \rho^{-\bar{\epsilon}}$; by the definition of $R_2(.)$, $q^K \leq \hat{q}_l(\epsilon) \leq q^K \rho^{-\epsilon}$; by assumption $0 \leq \epsilon \leq \bar{\epsilon}$, $0 \leq \epsilon^o \leq \bar{\epsilon}$; combining these inequalities renders the result.

for η to derive

$$\eta(q,\epsilon) = \frac{\ln f^o_{k2}(q) - f^o_l(q\rho^{-\epsilon}) - \ln \rho^{-\epsilon}}{\ln R_2(q\rho^{-\epsilon},\epsilon)} \text{ for all } q \in (\underline{q}, \overline{q}\rho^{\overline{\epsilon}}). \tag{88}$$

Consider functions $q(\epsilon)$ such that $R_2(q\rho^{-\epsilon},\epsilon) = const.$. Differentiate the function above to derive that

$$\frac{d\,\eta(q(\epsilon),\epsilon)}{d\epsilon} = \frac{\frac{d\ln f^o_{k2}(q(\epsilon))}{d\ln q}\frac{q'(\epsilon)}{q(\epsilon)} - \frac{d\ln f^o_l(q(\epsilon)\rho^{-\epsilon})}{d\ln q}\left(\frac{q'(\epsilon)}{q(\epsilon)} + \ln\rho^{-1}\right) - \ln\rho^{-1}}{\ln R_2(q(\epsilon)\rho^{-\epsilon},\epsilon)}. \tag{89}$$

Denote by $\hat{q}(\epsilon)$ the q such that $R_2(\hat{q}(\epsilon)\rho^{-\epsilon},\epsilon) = 1$. Evaluate the numerator at this function. Note that by assumption $\frac{d\ln f^o_l(\hat{q}(\epsilon)\rho^{-\epsilon})}{d\ln q} \leq -1$. It also implies $\frac{d\ln f^o_{k2}(\hat{q}(\epsilon))}{d\ln q} < -1$ since $\frac{d\ln f^o_{k2}(\hat{q}(\epsilon))}{d\ln q} = \frac{d\ln f^o_l(\hat{q}(\epsilon)\rho^{-\epsilon^o})}{d\ln q} + \frac{d\ln R_2(\hat{q}(\epsilon)\rho^{-\epsilon^o},\epsilon^o)}{d\ln q}$ and $\frac{d\ln R_2(\hat{q}(\epsilon)\rho^{-\epsilon^o},\epsilon^o)}{d\ln q} < 0$. Using these inequalities and some algebra using an appropriate software reveals that the numerator is greater than zero for all ϵ. It follows that $\lim_{q(\epsilon)\downarrow\hat{q}(\epsilon)} \frac{d\,\eta(q(\epsilon),\epsilon)}{d\epsilon} < 0$ for all ϵ since $\ln R(q(\epsilon)\rho^{-\epsilon},\epsilon) < 0$ for all $q(\epsilon) > \hat{q}(\epsilon)$.[34] Therefore, since $\eta'_B(\epsilon) > 0$, Condition 2 and 1 hold.
qed.

PROOF, Point (e), Lemma 11:
Suppose Condition 1 does not hold. It follows that there exists an $\epsilon \neq \epsilon^o$ and a η such that

$$f^o_l(q\rho^{-\epsilon^o})\rho^{-\epsilon^o}R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o} = f^o_l(q\rho^{-\epsilon})\rho^{-\epsilon}R_2(q\rho^{-\epsilon},\epsilon)^{\eta} \text{ for all } q \in (\underline{q}, \overline{q}\rho^{\overline{\epsilon}}). \tag{90}$$

Since all functions on both sides are real analytic on $(0,\overline{q})$, by the identity theorem of real analytic functions (see Corollary 1.2.6 in Krantz and Parks, 2002), this equation holds for all $q \in (0,\overline{q})$. The cited results from Krantz and Parks (2002) are summarised in Appendix F. Rearranging and taking the limit of q going to zero, it follows that

$$\lim_{q\downarrow 0} \frac{f^o_l(q\rho^{-\epsilon^o})\rho^{-\epsilon^o}}{f^o_l(q\rho^{-\epsilon})\rho^{-\epsilon}} = \lim_{q\downarrow 0} \frac{R_2(q\rho^{-\epsilon},\epsilon)^{\eta}}{R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o}} \text{ for all } q. \tag{91}$$

By assumption, there exists an order of derivative P such that

$$\lim_{q\downarrow 0} \frac{d^P f^o_l(q)}{dq^P} \neq 0 \text{ or } \pm\infty. \tag{92}$$

Using l'Hopital's rule, it follows that the left-hand side is a strictly positive real number. Next,

[34] Note that $\underline{q} < \hat{q}(\epsilon) < \overline{q}\rho^{\overline{\epsilon}}$, $\underline{q} < \hat{q}(\epsilon)\rho^{-\epsilon} < \overline{q}$, and $\underline{q} < \hat{q}(\epsilon)\rho^{-\epsilon^o} < \overline{q}$. Therefore, all used relations hold in the relevant range. The three inequalities follows since: by assumption, $\frac{\overline{q}}{q^K} > \rho^{-\overline{\epsilon}}$; by the property of the convergence radius of power series $\frac{\overline{q}}{q^K} = \frac{q^K}{\underline{q}} > \rho^{-\overline{\epsilon}}$; by the definition of $R_2(.)$, $q^K \leq \hat{q}(\epsilon)\rho^{-\epsilon} \leq q^K\rho^{-\epsilon}$; by assumption $0 \leq \epsilon \leq \overline{\epsilon}$, $0 \leq \epsilon^o \leq \overline{\epsilon}$; combining these inequalities renders the result.

consider the right-hand side

$$\lim_{q\downarrow 0}\frac{R_2(q\rho^{-\epsilon},\epsilon)^{\eta}}{R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o}}=\lim_{q\downarrow 0}\frac{\left((1-\rho)\frac{q^K}{q\rho^{-\epsilon}}+\frac{\epsilon}{1+\epsilon}\left(1+\frac{\rho^{\epsilon+1}}{\epsilon}\right)\right)^{\eta}}{\left((1-\rho)\frac{q^K}{q\rho^{-\epsilon^o}}+\frac{\epsilon^o}{1+\epsilon^o}\left(1+\frac{\rho^{\epsilon^o+1}}{\epsilon^o}\right)\right)^{\eta^o}}= \tag{93}$$

$$=\frac{\left((1-\rho)\frac{q^K}{\rho^{-\epsilon}}+\lim_{q\downarrow 0}q\frac{\epsilon}{1+\epsilon}\left(1+\frac{\rho^{\epsilon+1}}{\epsilon}\right)\right)^{\eta}}{\left((1-\rho)\frac{q^K}{\rho^{-\epsilon^o}}+\lim_{q\downarrow 0}q\frac{\epsilon^o}{1+\epsilon^o}\left(1+\frac{\rho^{\epsilon^o+1}}{\epsilon^o}\right)\right)^{\eta^o}}\lim_{q\downarrow 0}q^{\eta^o-\eta}. \tag{94}$$

The first factor is a constant. Since by Lemma 13 $\eta\neq\eta^o$, the second factor is ∞ or 0. Therefore, the right-hand side in Equation (91) is ∞ or 0, which gives a contradiction. It follows that Condition 1 holds.
qed.

PROOF, Point (f), Lemma 11:
Since the composition of real analytic functions is real analytic, $\ln f_l^o(q)$ real analytic implies $f_l^o(q)$ real analytic since the exponential is real analytic everywhere. As $f_l^o(q)$ is real analytic on $[0,\overline{q})$, $\frac{d^P f_l^o(0)}{dq^P}\in\mathbb{R}$ for all P. Moreover, there exists a P such that $\frac{d^P f_l^o(0)}{dq^P}\neq 0$; otherwise, by the identity theorem of real analytic functions, $f_l^o(q)=0$ for all q. Therefore, the sufficient condition in point (b) is satisfied.
qed.

PROOF, Point (g), Lemma 11:
Suppose Condition 1 does not hold. It follows that there exists an $\epsilon\neq\epsilon^o$ and a η such that

$$f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o}R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o}=f_l(q\rho^{-\epsilon})\rho^{-\epsilon}R_2(q\rho^{-\epsilon},\epsilon)^{\eta}\text{ for all }q\in(\underline{q},\overline{q}\rho^{\bar{\epsilon}}). \tag{95}$$

Using logarithms, it follows that

$$\ln f_l^o(q\rho^{-\epsilon^o})+\ln\rho^{-\epsilon^o}+\eta^o\ln R_2(q\rho^{-\epsilon^o},\epsilon^o)=\ln f_l^o(q\rho^{-\epsilon})+\ln\rho^{-\epsilon}+\eta\ln R_2(q\rho^{-\epsilon},\epsilon)\text{ for all }q\in(\underline{q},\overline{q}\rho^{\bar{\epsilon}}). \tag{96}$$

Changing variable, rearranging, and defining $\rho^{-(\epsilon-\epsilon^o)}=a$ renders

$$0=\ln f_l^o(qa)-\ln f_l^o(q)+\ln a+\eta\ln R_2(qa,\epsilon)-\eta^o\ln R_2(q,\epsilon^o)\text{ for all }q\in(\underline{q}\rho^{-\bar{\epsilon}},\overline{q}\rho^{\bar{\epsilon}}). \tag{97}$$

Rewrite $\eta\ln R_2(qa,\epsilon)$ defining $\tilde{q}=\ln q-\ln q^K$ and $\tilde{a}=\ln a$ as

$$\eta\ln R_2(qa,\epsilon)=\eta\ln\left((1-\rho)\frac{q^K}{qa}+\frac{\epsilon}{1+\epsilon}\left(1+\frac{\rho^{\epsilon+1}}{\epsilon}\right)\right)=$$
$$=-\eta(\tilde{q}+\tilde{a}-\ln(1-\rho))+\eta\ln\left(1+e^{\tilde{q}+\tilde{a}-\ln(1-\rho)}g(\epsilon)\right), \tag{98}$$

where to shorten the notation $g(\epsilon)=\frac{\epsilon}{1+\epsilon}\left(1+\frac{\rho^{\epsilon+1}}{\epsilon}\right)$. Use Equations (98) and (73) in Equation (97) to get that

$$0 = \sum_{p=0}^{P} \gamma_p^o \left[(\tilde{q}+\tilde{a})^p - \tilde{q}^p\right] + \tilde{a} - \eta(\tilde{q}+\tilde{a}-\ln(1-\rho)) + \eta \ln\left(1 + e^{\tilde{q}+\tilde{a}-\ln(1-\rho)} g(\epsilon)\right) + \\ + \eta^o(\tilde{q}-\ln(1-\rho)) - \eta^o \ln\left(1 + e^{\tilde{q}-\ln(1-\rho)} g(\epsilon^o)\right) \tag{99}$$

Both sides of the equation are real analytic for all $\tilde{q}$. Therefore, by the identity theorem of real analytic functions, the equation holds for all $\tilde{q}$. Differentiate with respect to $\tilde{q}$:

$$\eta - \eta^o = \sum_{p=1}^{P} \gamma_p^o p \left[(\tilde{q}+\tilde{a})^{p-1} - \tilde{q}^{p-1}\right] + \eta \frac{e^{\tilde{q}+\tilde{a}-\ln(1-\rho)} g(\epsilon)}{1 + e^{\tilde{q}+\tilde{a}-\ln(1-\rho)} g(\epsilon)} - \eta^o \frac{e^{\tilde{q}-\ln(1-\rho)} g(\epsilon^o)}{1 + e^{\tilde{q}-\ln(1-\rho)} g(\epsilon^o)} \tag{100}$$

Take the limit of $\tilde{q}$ going to $-\infty$:

$$\eta - \eta^o = \lim_{\tilde{q}\to-\infty} \sum_{p=1}^{P} \gamma_p^o p \left[(\tilde{q}+\tilde{a})^{p-1} - \tilde{q}^{p-1}\right] \tag{101}$$

As a first case, consider $P = 1$. The right-hand side is zero, while, by Lemma 13 $\eta \neq \eta^o$, which gives a contradiction. As a second case, consider $P > 2$. Using the Binomial Theorem and focusing on the terms with the highest and second highest order, it follows that

$$\eta - \eta^o = \lim_{\tilde{q}\to-\infty} \sum_{p=2}^{P} \gamma_p^o p \left[(\tilde{q}+\tilde{a})^{p-1} - \tilde{q}^{p-1}\right] = \\ \lim_{\tilde{q}\to-\infty} \left(\gamma_P^o P \tilde{q}^{P-1} + \gamma_P^o P(P-1)\tilde{q}^{P-2}\tilde{a} + \gamma_{P-1}(P-1)\tilde{q}^{P-2} - \gamma_P^o P \tilde{q}^{P-1} - \gamma_{P-1}^o (P-1)\tilde{q}^{P-2}\right) = \\ \lim_{\tilde{q}\to-\infty} \gamma_P^o P(P-1)\tilde{q}^{P-2}\tilde{a} = \pm\infty, \tag{102}$$

since $\tilde{a} \neq 0$, which gives a contradiction. As a last case, consider $P = 2$. Rearrange Equation (100) to

$$0 = \gamma_2^o 2\tilde{a} - \eta \frac{1}{1 + e^{\tilde{q}+\tilde{a}-\ln(1-\rho)} g(\epsilon)} + \eta^o \frac{1}{1 + e^{\tilde{q}-\ln(1-\rho)} g(\epsilon^o)}. \tag{103}$$

Taking $\tilde{q}$ to ∞ gives a contradiction.
qed.

PROOF, Point (a), Lemma 11:
Rearrange Equation (96) to

$$\ln f_l^o(q) = \ln f_l^o(qa) + \ln a + \eta \ln R_2(qa, \epsilon) - \eta^o \ln R_2(q, \epsilon^o) \text{ for all } q \in (\underline{q}\rho^{-\overline{\epsilon}}, \overline{q}\rho^{\overline{\epsilon}}). \tag{104}$$

Remember, $a = \rho^{-(\epsilon-\epsilon^o)} \neq 1$. This is a functional equation that, generically, does not hold. Therefore, Condition 1 holds generically.

To see that, consider the point $\hat{q}$ such that $\ln R_2(\hat{q}, \epsilon^o) = 0$. Without loss of generality,

normalise $f_l^o(\hat{q})$ to one. It follows that the only compatible η is

$$\eta = \frac{-\ln f_l^o(\hat{q}a) - \ln a}{\ln R_2(\hat{q}a, \epsilon)}. \tag{105}$$

Next, consider the points $\hat{q}a^J$, where J is an integer larger than 1. Assume that $(\underline{q}, \overline{q})$ is large enough such that at least two such points exist. It follows that, if Equation (96) holds, the value of $f_l^o(.)$ at these points needs to be

$$\ln \tilde{f}_l^o(\hat{q}a^J) = \sum_{j=1}^{J} \left(-\ln a - \frac{-\ln f_l^o(\hat{q}a) - \ln a}{\ln R_2(\hat{q}a, \epsilon)} \ln R_2(\hat{q}a^j, \epsilon) + \eta^o \ln R_2(\hat{q}a^{j-1}, \epsilon^o) \right). \tag{106}$$

Take two such points. These hypothetical values $\tilde{f}_l^o$ are functions of the true parameter values ϵ^o, η^o, and $f_l^o(\hat{q}a)$ and ϵ. Denote the hypothetical values at these points as functions of ϵ by $\tilde{f}^1(\epsilon)$ and $\tilde{f}^2(\epsilon)$. The triple $(\epsilon, \tilde{f}^1(\epsilon), \tilde{f}^2(\epsilon))$ describes a one dimensional line in $\mathbb{R}^3$. The true values at these points $f^1(\epsilon), f^2(\epsilon)$ are also functions of ϵ and the triple $(\epsilon, f^1(\epsilon), f^2(\epsilon))$ also describes a line. By construction, since $a = 1$ when $\epsilon = \epsilon^o$, these lines cross at $\epsilon = \epsilon^o$. Generically, the lines do not cross again in $\mathbb{R}^3$. Note that the argument can be strengthened using $J > 2$ points. In this case, the true and implied values describe one-dimensional lines in $\mathbb{R}^J$; generically, lines do not cross twice in such high-dimensional spaces.
qed.

PROOF, Point (a), Lemma 12:
See proofs of the respective points of Lemma 11.

B.5 Details identification

Denote the true functions and parameters in the population with superscript "o", while functions and parameters without superscript denote the general values of these objects.

B.5.1 Proof of Lemma 2

The parametric part of Θ is compact because it is closed and bounded. By Lemma 10 and since $(\ln \overline{q} - \ln q^K) > 1$, the series γ converges. Therefore, it is an element of the complete space l^∞. Since $|\gamma_p| \leq M(\ln \overline{q} - \ln q^K)^{-p}$ for all p and $\lim_{p\to\infty} M(\ln \overline{q} - \ln q^K)^{-p} = 0$, the space is a closed and totally bounded subspace of l^∞. Therefore, it is complete and totally bounded. It follows that Θ is compact.
qed.

B.5.2 Proof of Lemma 3

Consider a sequence $\theta^n \in \Theta$ such that $\lim_{n\to\infty} d(\theta, \theta^n) = 0$. It follows that $\lim_{n\to\infty} \epsilon^n = \epsilon$, $\lim_{n\to\infty} \eta^n = \eta$, and $\lim_{n\to\infty} \gamma_p^n = \gamma_p$ for all p.

The next step shows that $\ln f_l(q \mid \gamma^n)$ converges uniformly to $\ln f_l(q \mid \gamma)$ for all $q \in [\underline{b}, \overline{b}\rho^{-\overline{\epsilon}}]$. Consider a $\tilde{q}$ such that

$$\left(\ln \overline{q} - \ln q^K\right) > \left(\ln \tilde{q} - \ln q^K\right) > \max\left\{1, (\ln q^K - \ln \underline{b}), (\ln \overline{b}\rho^{-\overline{\epsilon}} - \ln q^K)\right\}.$$

It follows that

$$\lim_{n\to\infty}\sup_q |f_l(q \mid \gamma^n) - f_l(q \mid \gamma)| = \lim_{n\to\infty}\sup_q \left|\sum_{p=0}^{\infty}(\gamma_p^n - \gamma_p)\left(\ln\frac{q}{q^K}\right)^p\right| \leq$$
$$\leq \lim_{n\to\infty}\sum_{p=0}^{\infty}\left|\gamma_p^n - \gamma_p\right|\left(\ln\frac{\tilde{q}}{q^K}\right)^p = 0, \tag{107}$$

where the last step follows from Tannery's Theorem. Tannery's Theorem applies since:

$$\lim_{n\to\infty}\left|\gamma_p^n - \gamma_p\right|\left(\ln\frac{\tilde{q}}{q^K}\right)^p = 0 \;\forall\; p; \tag{108}$$

$$\left|\gamma_p^n - \gamma_p\right|\left(\ln\frac{\tilde{q}}{q^K}\right)^p \leq 2M\left(\frac{\ln\tilde{q} - \ln q^K}{\ln\overline{q} - \ln q^K}\right)^p; \tag{109}$$

$$\sum_{p=0}^{\infty} 2M\left(\frac{\ln\tilde{q} - \ln q^K}{\ln\overline{q} - \ln q^K}\right)^p = 2M\frac{1}{1 - \frac{\ln\tilde{q} - \ln q^K}{\ln\overline{q} - \ln q^K}}. \tag{110}$$

$Q(.)$ is continuous in ϵ and η. Moreover, $\ln f_l(q \mid \gamma^n)$ converges uniformly to $\ln f_l(q \mid \gamma)$. Therefore, $\lim_{n\to\infty} Q(\theta^n) = Q(\theta)$. By the continuous mapping theorem $Q(\theta)$ is continuous. qed.

B.5.3 Proof of Proposition 3:

Denote the right-hand sides of Equations (20) and (22) in Section A.2 as a function of their parameters by $\tilde{B}(\epsilon,\eta)$ and $f_{k2}(q|\epsilon,\eta)$. Taylor expand the right hand side of Equations (20) and (22) around ϵ^o, η^o and cancel equal terms to derive

$$\frac{\partial\tilde{B}(\epsilon^o,\eta^o)}{\partial\epsilon}(\epsilon - \epsilon^o) + \frac{\partial\tilde{B}(\epsilon^o,\eta^o)}{\partial\eta}(\eta - \eta^o) + h_1(\epsilon,\eta)\left(\mid \epsilon - \epsilon^o \mid + \mid \eta - \eta^o \mid\right) = 0, \tag{111}$$

$$\frac{\partial f_{k2}(q \mid \epsilon^o,\eta^o)}{\partial\epsilon}(\epsilon - \epsilon^o) + \frac{\partial f_{k2}(q \mid \epsilon^o,\eta^o)}{\partial\eta}(\eta - \eta^o) + h_2(\epsilon,\eta,q)\left(\mid \epsilon - \epsilon^o \mid + \mid \eta - \eta^o \mid\right) = 0; \tag{112}$$

where $\lim_{(\epsilon,\eta)\to(\epsilon^o,\eta^o)} h_1(\epsilon,\eta) = 0$ and $\lim_{(\epsilon,\eta)\to(\epsilon^o,\eta^o)} h_2(\epsilon,\eta,q) = 0$. W.l.o.g. assume $\epsilon \geq \epsilon^o$ and $\eta \geq \eta^o$. Rearrange the equations above to

$$\frac{\frac{\partial\tilde{B}(\epsilon^o,\eta^o)}{\partial\epsilon} + h_1(\epsilon,\eta)}{-\frac{\partial\tilde{B}(\epsilon^o,\eta^o)}{\partial\eta} - h_1(\epsilon,\eta)}(\epsilon - \epsilon^o) = (\eta - \eta^o), \tag{113}$$

$$\frac{\frac{\partial f_{k2}(q|\epsilon^o,\eta^o)}{\partial\epsilon} + h_2(\epsilon,\eta,q)}{-\frac{\partial f_{k2}(q|\epsilon^o,\eta^o)}{\partial\eta} - h_2(\epsilon,\eta,q)}(\epsilon - \epsilon^o) = (\eta - \eta^o). \tag{114}$$

By Condition 2, there exists a q such that for all (ϵ,η) sufficiently close to (ϵ^o,η^o), the coefficients of this linear system of equations are not collinear, which renders a contradiction (see Appendix B.4 for a discussion of Condition 2). Therefore, $(\epsilon,\eta) = (\epsilon^o,\eta^o)$. It follows that $(\epsilon^o,\eta^o,\gamma^o)$ is the locally unique minimum of $Q(.)$. By this result together with Lemmas 2 and 3, $Q(\theta)$, θ^o, and Θ fulfil the Condition 3.1" for identification in Chen (2007), which proves

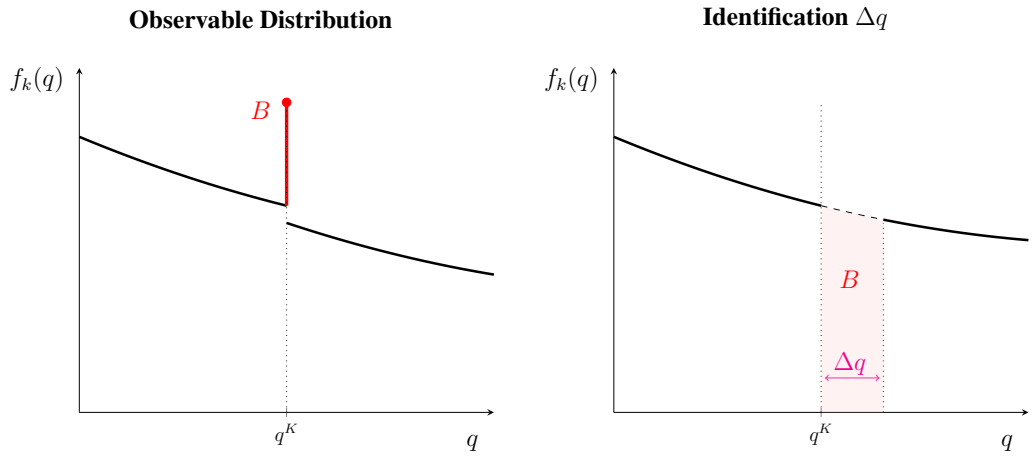
Propositions 3. The cited results from Chen (2007) are summarised in Appendix F.
qed.

B.6 A comparison of Proposition 2 to the results in Blomquist and Newey (2017), Blomquist et al. (2021), Bertanha, McCallum and Seegert (2023), Chetty et al. (2011), and Goff (2022)

Because the properties of the counterfactual distribution $f_l(.)$ necessary for identification are not specific to the presence of participation margin responses, I only consider intensive margin responses in this section. Similarly, because the identification result does not depend on whether bunching is sharp or scattered, I consider only sharp bunching in this section.

Figure 4 illustrates that bunching depends on the intensive margin response of the marginal buncher Δq. But why is the observable measure $f_k(.)$, illustrated in the left panel of Figure 12, useful to infer Δq? Intuitively, one can use the observable mass B and fill it into the distribution to the right of the kink point, i.e., one can transfer the mass back to where it would be located in the counterfactual. The right panel of Figure 12 illustrates the transfer of the mass. In this way, one can infer the Δq necessary to reconstruct a "smooth" counterfactual.

Figure 12: The identification of intensive margin responses.



Note: The left panel illustrates the observable measure $f_k(.)$. The right panel illustrates how the bunching mass B is related to the intensive margin response Δq. The dashed line illustrates the segment of the counterfactual distribution $f_l(.)$ that is not observable.

However, to carry out this procedure, one needs to take a stand on the values of the counterfactual in the interval $[q^K, q^K + \Delta q]$. The counterfactual $f_l(.)$ is neither directly nor indirectly observable in this interval. A different shape of $f_l(.)$ in this interval implies a different value of Δq; i.e., if $f_l(.)$ is relatively large in this range, the implied Δq is small, if $f_l(.)$ is small, the implied Δq is large.

The classic bunching estimator (e.g., see Saez, 2010, Chetty et al., 2011, and Kleven, 2016) assumes the counterfactual is constant or linear over this range. Under this assumption, the value of $f_l(.)$ can be inferred from points just to the left and the right of the kink point, and Δq is identified. However, this distributional assumption can lead to a substantial asymptotic bias in the estimator (see Section 2.2.2 in Bertanha et al., 2023 for a detailed discussion).

Alternatively, Bertanha et al. (2023) and Goff (2022) show that restricting $f_l(.)$ to a parametric function allows identifying the intensive margin elasticity. Intuitively, points to the left of the kink point identify the parameters of $f_l(.)$, which, together with the bunching mass, identifies Δq. However, again, whenever the true counterfactual does not have the assumed parametric form, the estimator is asymptotically biased. The size of the bias depends on the accuracy of the parametric assumption.

Is it possible to relax the parametric assumption on $f_l(.)$, and, if yes, under which conditions? Theorem 1 in Blomquist and Newey (2017) shows that, when $f_l(.)$ is continuously differentiable of order D but otherwise unrestricted, the intensive margin elasticity is not identified. Intuitively, since $f_l(.)$ is unobserved in a strictly positive interval $[q^K, q^K + \Delta q]$, observing D derivatives of $f_l(.)$ outside of this interval is insufficient to infer the value of the function in the strict interior of the interval. Any value of the function in the strict interior of the interval is compatible with the values of the D derivatives of the function outside the interval; it suffices that the derivatives of order higher than D are sufficiently extreme. This result implies that a stronger restriction than the differentiability of order D is necessary.

Proposition 2 in this paper shows that intensive and participation margin responses are identified if the counterfactual $f_l(.)$ is nonparamteric, but fulfils Assumption 3.a or 3.b. Note that this result does not contradict Theorem 1 in Blomquist and Newey (2017), because Assumption 3.a or 3.b are stronger than assuming differentiability of order D. Under Assumption 3.a or 3.b, the counterfactual $f_l(.)$ is uniquely determined by the sequence of its derivatives at the kink point; this sequence is identified from points outside the bunching interval. Intuitively, the function is sufficiently smooth such that it can be interpolated from points outside the interval $[q^K, q^K + \Delta q]$.

Blomquist et al. (2021), Bertanha, McCallum and Seegert (2023), and Goff (2022), show that the intensive margin elasticity can be partially identified if the econometrician knows bounds of the slope of $f_l(.)$ or other restrictions on the shape of $f_l(.)$. Proposition 2 shows that under Assumption 3.a or 3.b, both elasticities are point-identified.[35] Point-estimates are necessary to conduct the counterfactual exercises in Section 4. Moreover, no ex-ante knowledge about the bounds of the slope of $f_l(.)$ or its shape is necessary. Bertanha, McCallum and Seegert (2023) show that the intensive margin elasticity can be point-identified without restricting $f_l(.)$ to a parametric functional form when a rich set of covariates is available; Proposition 2 does not rely on the availability of covariates. In my application, no rich set of covariates is available.

B.7 An illustration of Proposition 2 using simulations

This section illustrates the identification in Proposition 2 using simulations. The section uses superscript "o" for the true values of the parameters or measures, while objects without superscript denote the corresponding general values of these objects. I simulate the observable measure $f_k^o(.)$ using three different functions for the underlying counterfactual measure $f_l^o(.)$ of the data-generating process. Then, I show that the minimum of the population criterion

$$\theta_P = \arg\min_{\theta \in \Theta_P} Q(\theta) \tag{115}$$

converges to the true parameter θ^o as the order of the series P increases. The population criterion $Q(\theta)$ is defined in Equation (17) in Section A.2. Intuitively, it is the square distance between the logarithm of the observed measure $f_k^o(.)$ and the corresponding object in the model.

[35] Note that when the interval $(\underline{q}, \overline{q})$ in Assumption 3.b is not large enough, point-identification is not possible.

Remember, $\theta = (\epsilon, \eta, \gamma_0, \gamma_1, ...)$, where $(\gamma_0, \gamma_1, ...)$ is the sequence of coefficients of the power series in Equation (9). The minimisation (115) restricts γ to dimension P, i.e., it is an element of the P dimensional sieve-space Θ_P. For a formal definition of the sieve-space Θ_P see Section C.2.

Figures 7 and 1 suggest that the counterfactual measure in the empirical application is very close to linear in the log-log scale. Accordingly, in the first example, I assume the counterfactual distribution of the data-generating process is linear in logarithmic scales:

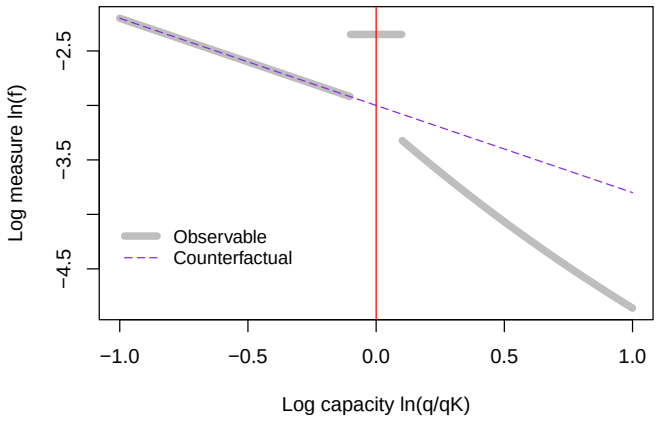
$$\ln f_l^o(q) = \lambda_0 + \lambda_1 \ln\left(\frac{q}{q^K}\right).$$

It corresponds to a Pareto distribution. In the model, I use

$$\ln f_l(q_l) = \sum_{p=0}^{P} \gamma_p \left(\ln \frac{q_l}{q^K}\right)^p, \tag{116}$$

where P converges to infinity. In this first example, the functional form of the true counterfactual is a special case of the finite series expansion of $\ln f_l(.)$ as long as $P \geq 1$. Figure 13 depicts the observable measure $\ln f_k^o(.)$ as the thick grey line and the underlying counterfactual measure $\ln f_l^o(.)$ as the dashed purple line. The grey bar at zero depicts the normalised mass in the bunching interval $[q_L, q_H]$, i.e., it depicts $\ln B^o$. Table 6 summarises the parameters of the simulation. Unsurprisingly, Table 3 shows that $(\epsilon_P, \eta_P) = (\epsilon^o, \eta^o)$ for $P \geq 1$. Figure 14 shows the theoretical and inferred counterfactual distribution for points in and to the right of the bunching interval. For $P \geq 1$, the two functions are equal even though, at this range, the distribution is not directly observable. Again, this is no surprise since the functional form of $\ln f_l^o(.)$ is a special case of the functional form of $\ln f_l(.)$ if $P \geq 1$. The example shows that the elasticities η and ϵ are identified when the functional form of $f_l^o(.)$ is a special case of $f_l(.)$.

Figure 13: Simulation; example 1, linear $\ln f_l^o(.)$.



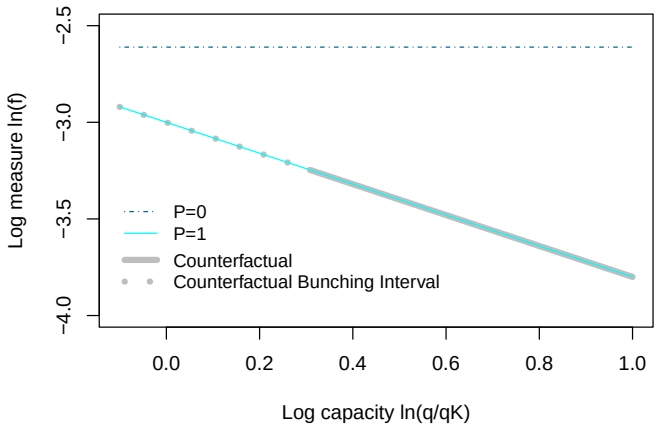
Note: The dashed purple line shows the true counterfactual measure $\ln f_l^o(.)$ *of the data-generating process. The thick grey line shows the implied observable measure* $\ln f_k^o(.)$.

Table 3: Inferred parameters; example 1, linear $\ln f_l^o(.)$.

	Epsilon	Bias Epsilon [%]	Eta	Bias Eta [%]
True Value	0.30000	0.000	3.0000	0.000
P=0	0.11691	-61.029	6.5562	118.540
P=1	0.30000	0.000	3.0000	0.000
P=2	0.30000	0.000	3.0000	0.000

Note: Columns 2 and 4 show the elasticities ϵ and η. The first row shows their true value, while rows two to four show the inferred values using different polynomial orders P. Columns 3 and 5 show the relative difference to the true value in percent. For $P \geq 1$, the inferred values are equal to the true values.

Figure 14: The true and inferred counterfactual; example 1, linear $\ln f_l^o(.)$.



Note: The thick grey line shows the true counterfactual measure $\ln f_l^o(.)$ in the range where it is not directly observable. The dotted line shows the bunching interval, while the solid line shows points to the right of the bunching interval. The blue lines show the inferred counterfactual measure using different polynomial orders P. For $P = 1$ the functions coincide.

Bertanha, McCallum and Seegert (2023) point out that the intensive margin elasticity ϵ is identified when the parametric functional form of $f_l^o(.)$ is known. The first example illustrates that this result is also true when the participation margin is present. However, they also point out that, without restrictions on $f_l^o(.)$, the intensive elasticity is not identified. The next two examples illustrate that the real analyticity of $\ln f_l^o(.)$ is sufficient to identify the two elasticities.

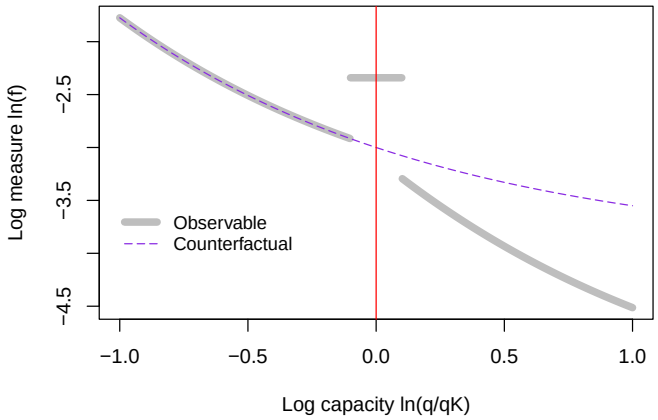
In the second example, I assume the true counterfactual distribution is an exponential function in logarithmic scales:

$$\ln f_l^o(q) = \lambda_0 + e^{-\lambda_1 \ln\left(q/q^K\right)}.$$

Importantly, in this case, the true counterfactual measure is not a special case of the series expansion used in the model, i.e., $\ln f_l^o(.) \neq \ln f_l(.)$ for all P. However, the exponential function

is real analytic over the entire real line, i.e., its power series expansion at the kink point converges on the entire real line. Figure 15 shows the observable in grey and the true counterfactual in purple; the grey bar at zero illustrates the mass in the bunching interval. Table 4 shows that for all P, η_P and ϵ_P are not equal to the true values. The reason is that a finite polynomial cannot represent the exponential function. However, as P increases, η_P and ϵ_P converge to their true values. Already at $P = 4$, the small sample bias in both elasticities is smaller than 1 %. Figure 16 shows that the model infers $\ln f^o(.)$ in intervals where it is not observable - the bunching interval - or only indirectly observable - to the right of the bunching interval. The inferred function converges to the true function as P increases; already at $P = 4$ the two functions are visually indistinguishable. The example illustrates that, if $\ln f_l^o(.)$ is analytic, ϵ^o and η^o are identified since $\ln f_l(.)$ converges to $\ln f_l^o(.)$ as P increases.

Figure 15: Simulation; example 2, exponential $\ln f_l^o(.)$.



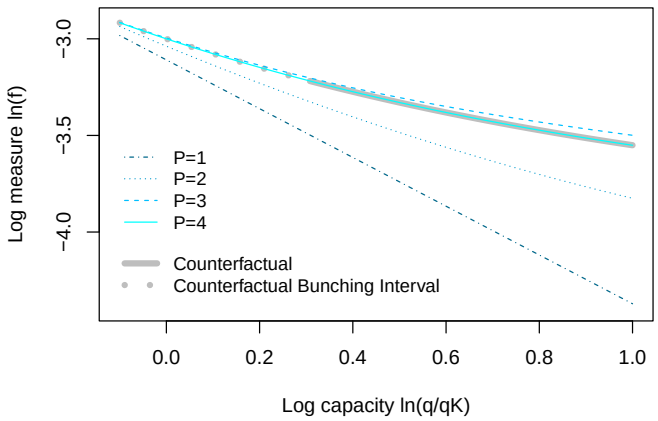
Note: The dashed purple line shows the true counterfactual measure $\ln f_l^o(.)$ *of the data-generating process. The thick grey line shows the implied observable measure* $\ln f_k^o(.)$. *The grey bar at zero depicts the bunching mass* $\ln B^o$.

Table 4: Inferred parameters; example 2, exponential $\ln f_l^o(.)$.

	Epsilon	Bias Epsilon [%]	Eta	Bias Eta [%]
True Value	0.30000	0.000	3.0000	0.000
P=1	0.36239	20.795	0.3976	-86.747
P=2	0.32369	7.895	2.1423	-28.591
P=3	0.29792	-0.694	3.1557	5.190
P=4	0.30000	0.000	2.9996	-0.012

Note: Columns 2 and 4 show the elasticities ϵ *and* η. *The first row shows their true values, while rows two to five show the inferred values using different polynomial orders* P. *Columns 3 and 5 show the relative difference to the true value in percent. The inferred values converge to the true values as* P *increases.*

Figure 16: The true and inferred counterfactual; example 2, exponential $\ln f_l(.)$.



Note: The thick grey line shows the true counterfactual measure $\ln f_l^o(.)$ *in the range where it is not directly observable. The dotted line shows the bunching interval, while the solid line shows points to the right of the bunching interval. The blue lines show the inferred counterfactual measure using different polynomial orders* P*. The inferred function converges to the true function as* P *increases.*

One may argue that the exponential function has a benign shape since it is monotonically decreasing. Moreover, the bunching interval in example two is relatively narrow; hence, it is easier to infer the function. Are the two elasticities identified when bunching coincides with a peak in $\ln f_l^o(.)$, or when using a very large bunching interval is necessary because the bunching mass is very scattered? The following example illustrates that the answer is yes. I assume $\ln f_l^o(.)$ has the shape of a normal distribution with a maximum at the kink point:

$$\ln f_l^o(q) = \lambda_0 + \frac{1}{\sqrt{2\pi}\lambda_1} e^{-\frac{1}{2}\left(\frac{\ln\left(q/q^K\right)}{\lambda_1}\right)^2}.$$

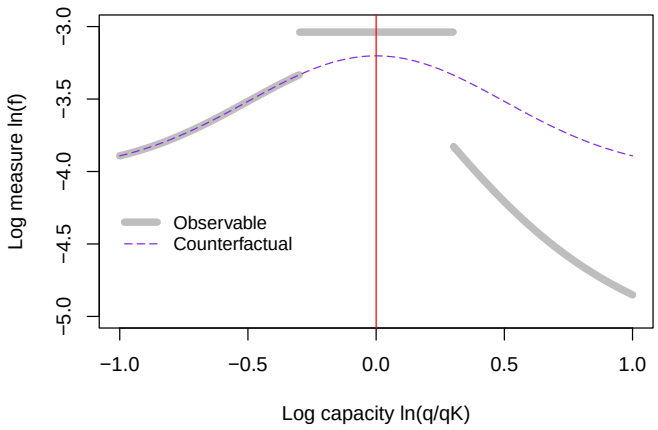
Moreover, I choose a sizeable bunching interval. As in example two, no finite polynomial can represent the pdf of the normal distribution, but the function is analytic. Figure 17 shows the observable, the counterfactual, and the bunching mass. Table 5 shows the true and inferred values of the elasticities. As in example two, η_P and ϵ_P converge to their true values as P increases; at $P = 10$, the biases are below 1 percent.[36] The convergence in this example is slower than in example two since $\ln f_l^o(.)$ is less regular, and the bunching window is larger. Figure 18 shows how the inferred counterfactual converges to the true counterfactual. The figure only shows the bunching interval and not the points to the left since, for these points, the convergence is too fast to distinguish the functions visually.

Note that if one were to pick an example where $\ln f_l^o(.)$ is even less regular or the bunching window is even larger, P would need to be even larger to reduce the bias below one percent. However, as long as Assumption 3.b holds, there always exists an order P such that the biases are smaller than some arbitrary small number $\delta > 0$. In the estimation, increasing P is costly

[36]The table only shows values for even P because, for the normal distribution, the true parameters at uneven powers in the series expansion are equal to zero. Therefore, values for uneven P are similar to values for $P - 1$.

since it increases the estimates' variance, giving rise to a standard bias-variance trade-off in nonparametric estimations. However, this is a problem of estimation and not of identification. Section 2 discusses the optimal choice of P in further detail.

Figure 17: Simulation; example 3, normal $\ln f_l$.



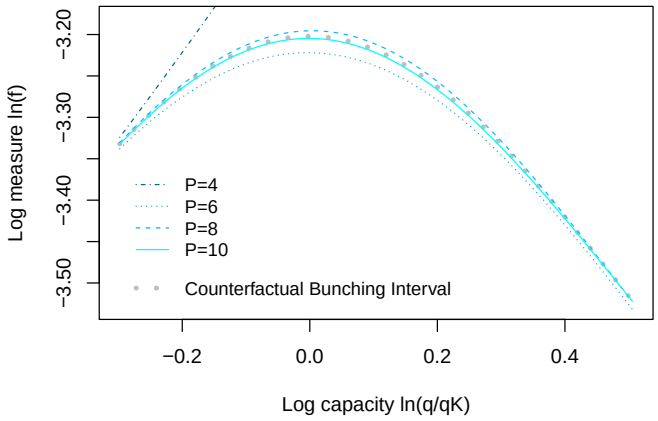
Note: The dashed purple line shows the true counterfactual measure $\ln f_l^o(.)$ *of the data-generating process. The thick grey line shows the implied observable measure* $\ln f_k^o(.)$. *The grey bar at zero depicts the bunching mass* $\ln B^o$.

Table 5: Inferred parameters; example 3, normal $\ln f_l$.

	Epsilon	Bias Epsilon [%]	Eta	Bias Eta [%]
True Value	0.30000	0.000	3.0000	0.000
P=4	0.19170	-36.101	9.1228	204.093
P=6	0.31168	3.893	2.9096	-3.013
P=8	0.29439	-1.869	2.9933	-0.223
P=10	0.30248	0.827	3.0015	0.050

Note: Columns 2 and 4 show the elasticities ϵ *and* η. *The first row shows their true values, while rows two to five show the inferred values using different polynomial orders* P. *Columns 3 and 5 show the relative difference to the true value in percent. The inferred values converge to the true values as* P *increases.*

Figure 18: The true and inferred counterfactual; example 3, normal $\ln f_l$.



Note: The dotted grey line shows the true counterfactual measure $\ln f_l^o(.)$ *in the bunching interval, where it is not observable. The blue lines show the inferred counterfactual measure using different polynomial orders* P*. The inferred function converges to the true function as* P *increases.*

Table 6: Parameters simulations

	λ_0	λ_1	Bandwidth	Bunching Interval
Example one	- 0.3	-0.8	[-1,1]	[-0.1,0.1]
Example two	-0.4	0.8	[-1,1]	[-0.1,0.1]
Example three	-0.4	0.5	[-1,1]	[-0.3,0.3]

Note: Columns 1 and 2 show the parameters of the counterfactual distribution $f_l^o(.)$*. Column 3 shows the bandwidth, i.e., the points around the kink point used for the estimation. Column 4 shows the bunching interval. Bandwidth and bunching interval are expressed as normalised logarithmic capacities (i.e.,* $[\ln(\underline{b}/q^K), \ln(\overline{b}/q^K)]$*.) The parameters of the incentive scheme* $S_k(.)$ *are* $q^K = 30$ *and* $\rho = 0.5$*.*

B.8 The identification in a discontinuous (i.e., notched) incentive scheme

This section generalizes the result in Proposition 2 for the case when there is a discontinuity (i.e., a notch) in the incentive scheme. For simplicity, I focus on the case where bunching is sharp.

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

$$S_d(q) = s_l q, \qquad \text{for } q \leq q_D; \tag{117}$$
$$S_d(q) = s_l q - \Delta S \; s_l, \qquad \text{for } q > q_D; \tag{118}$$

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

Proposition 4 (The observed density in the case of a notch)**.** *Close to the notch point, the observable measure $f_d(.)$ under the notched subsidy $S_d(.)$ is a function of three unknowns: the intensive margin elasticity ϵ, the participation margin elasticity η, and the counterfactual measure $f_l(.)$. At the notch point q_D, there is a mass point with bunching mass B. Four parts of the observable measure $f_d(.)$ depend distinctly on the three unknowns:*

$$f_d(q) = f_l(q), \qquad \textit{for } q < q_D; \tag{119}$$
$$B = \int_{q_D}^{q_B(\epsilon,\Delta S)} R_B(q_l,\epsilon)^\eta f_l(q_l) dq_l, \qquad \textit{for } q = q_D; \tag{120}$$
$$f_d(q) = 0, \qquad \textit{for } q_D < q < q_B; \tag{121}$$
$$f_d(q) = \left(1 - \frac{\Delta S}{q}\right)^\eta f_l(q), \qquad \textit{for } q > q_B. \tag{122}$$

Note: The variable q_D denotes the notch point; ΔS denotes the size of the notch relative to s_l; $R(.,\epsilon)$ defined in Equation (15) is the net subsidy payment to an adopter under the notched scheme relative to the subsidy payment under the counterfactual scheme. The variable $q_B(\epsilon, \Delta S)$ defined in Equation (126) denotes the quantity-choice of the marginal buncher. It is a function of the intensive margin elasticity ϵ and the size of the notch ΔS.

PROOF:
As in Section A.1. Use the cost function and the FOC of adopters to derive that

$$q = q_l, \qquad \text{for } q_l < q_D; \tag{123}$$
$$q = q_D, \qquad \text{for } q_D \leq q_l \leq q_B; \tag{124}$$
$$q = q_l, \qquad \text{for } q_l > q_B. \tag{125}$$

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

$$q_D - \frac{\epsilon}{1+\epsilon} \frac{1}{q_B^{\frac{1}{\epsilon}}} {q_D}^{1+\frac{1}{\epsilon}} = q_B \frac{1}{1+\epsilon} - \Delta S \tag{126}$$

The rest of the proof follows the steps in Section B.2.5.
qed.

Proposition 5 (The identification in the case of a notch)**.** *The observable measure $f_d(.)$ identifies the counterfactual measure $f_l(.)$, the intensive margin elasticity ϵ, and the participation margin elasticity η.*

PROOF:
As in Section A.2. Three parts of the measure (119), (120), and (122) are sufficient to identify the three unknowns.
qed.

C Details estimation

This section denotes the true functions and parameters in the population with superscript "o", while functions and parameters without superscript denote the general values of these objects.

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

As a first step, construct the empirical histogram $\widehat{f(q_j)}$ by choosing bins and counting the number of adopters in each bin. The normalisation by the bin size h_j and the total number of adopters n gives the observed density $\widehat{f(q_j)}$ at point q_j, where q_j is the location of the bin. The index j in $\{-N_-, ..., -1, 0, 1, ..., N_+\}$ is the index of the bin and $N = N_- + N_+$ is the total number of bins. Each bin size h_j is a function of the sample size n and goes to zero as n goes to infinity; moreover, $\frac{nh_j}{\ln n}$ goes to infinity as n goes to infinity. As a second step, take the logarithm of the histogram $\ln \widehat{f(q_j)}$. The location and the size of the bins q_j and h_j may be chosen equidistantly or otherwise. In general, the estimator is more efficient if the bins are chosen such that the variance of the histogram is approximately constant. Section C.1.2 and C.1.3 discuss binning and bias correction procedures that improve the small sample properties and efficiency. Note that the steps proposed in these sections are not necessary for the consistency and asymptotic normality of the estimator.

The estimate $\widehat{B}$ is simply the number of observations in the bunching interval $[q_L, q_H]$. The next section shows that $\ln \widehat{f(q_j)}$ and $\ln \widehat{B}$ are consistent estimates of $\ln f_k^o(q_j)$ and $\ln B^o$, where B^o is the mass in the bunching interval in population and $f_k^o(q_j)$ is the measure at point q_j in population.

C.1.1 The consistency of $\ln \widehat{f(q_j)}$ and $\ln \widehat{B}$

Denote by Q_ι the value of the observation number ι. Denote by $\widehat{N_j}$ the number of observations in a bin. The constructed dependent variable is

$$\widehat{f(q_j)} = \frac{\widehat{N_j}}{nh_j}. \tag{127}$$

Lemma 14. *$\widehat{f(q_j)}$ converges to $f_k^o(q_j)$ almost surely, which is also true in logarithms:*

$$\widehat{f(q_j)} \overset{a.s.}{\to} f_k^o(q_j); \tag{128}$$

$$\ln \widehat{f(q_j)} \overset{a.s.}{\to} \ln f_k^o(q_j). \tag{129}$$

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

PROOF:
The result $\widehat{f(q_j)} \xrightarrow{a.s.} f_k^o(q_j)$ is standard (e.g., see Einmahl and Mason, 2005). Write $\widehat{B}$ as

$$\widehat{B} = \frac{1}{n}\sum_{\iota=1}^{n} \frac{\mathbb{1}[q_L \leq Q_\iota < q_H]}{q_H - q_L}. \tag{130}$$

Using the strong law of large numbers, it follows that

$$\frac{1}{n}\sum_{\iota=1}^{n} \frac{\mathbb{1}[q_L \leq Q_\iota < q_H]}{q_H - q_L} \xrightarrow{a.s.} \left(\int_{q_L}^{q_H \rho^{-\epsilon^o}} R(q_l, \epsilon^o)^{\eta^o} f_l^o(q_l) dq_l \right) / (q_H - q_L) = B^o, \tag{131}$$

as n and goes to infinity. By the continuous mapping theorem, it follows that

$$\ln \widehat{f(q_j)} \xrightarrow{a.s.} \ln f_k^o(q_j). \tag{132}$$

The same argument holds for $\ln \widehat{B}$.
qed.

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

The empirical model in Section 2 uses the logarithm of $\widehat{f(q_j)}$ as the dependent variable. Using the logarithm does not affect consistency (see Appendix C.1.1). However, it introduces a small sample bias: $\mathbb{E}\left[\ln \widehat{f(q_j)}\right]$ is not equal to $\ln f_k^o(q_j)$. To counteract this effect, model (11) in Section 2 uses the bias-corrected dependent variable

$$\widehat{\ln f(q_j)} = \ln \widehat{f(q_j)} + \frac{1}{2\widehat{N_j}}, \tag{133}$$

where $\widehat{N_j}$ denotes the number of observations in a bin. Note that $\widehat{\ln f(q_j)}$ denotes the bias-corrected dependent variable, while $\ln \widehat{f(q_j)}$ denotes the logarithm of the histogram. Using (133) reduces bias since

$$\mathbb{E}\left[\widehat{\ln f(q_j)}\right] = \ln f_k^o(q_j) + O\left(\frac{1}{N_j^2}\right), \tag{134}$$

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

PROOF:
Taylor approximate all functions of random variables in Equation (133) around their expected values and use $f_k^o(q_j)\, n\, h_j = N_j$:

$$\begin{aligned}\widehat{\ln f(q_j)} \approx & \ln f_k^o(q_j) + \frac{1}{N_j}(\widehat{N_j} - N_j) - \frac{1}{2N_j^2}(\widehat{N_j} - N_j)^2 + \frac{1}{3N_j^3}(\widehat{N_j} - N_j)^3 + \\ & + \frac{1}{2N_j} - \frac{1}{2N_j^2}(\widehat{N_j} - N_j) + \frac{1}{2N_j^3}(\widehat{N_j} - N_j)^2. \end{aligned} \tag{135}$$

Take the expectation on both sides above. Note that, because $\widehat{N_j}$ follows a Binomial distribu-

tion, $\mathbb{E}(\widehat{N_j} - N_j) = 0$, $\mathbb{E}(\widehat{N_j} - N_j)^2 = N_j$, and $\mathbb{E}(\widehat{N_j} - N_j)^3 = N_j$.
qed.

C.1.3 Binning

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 in this paper's application is decreasing in capacity q. Consequently, 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. Such a binning procedure does not affect consistency because $\widehat{f(q_j)}$ is normalised by the bin size. Moreover, the variance of the log-histogram depends on the number of observations in each bin. The log-histogram has constant variance if the number of observations in each bin is approximately constant. Therefore, an additional advantage of a well-chosen binning procedure is that it equalises the variance of the dependent variable $\widehat{\ln f(q_j)}$. Hence, it increases efficiency and avoids the need for a weighting matrix in the estimation.

Concretely, use the function

$$\overline{q_j} = q^K (1 - j c_0(n))^{-\frac{1}{\omega}}, \tag{136}$$

where $\overline{q_j}$ is the right border of a bin. The bin size $h_j = \overline{q_j} - \overline{q}_{j-1}$. The specification parameter $c_0(n)$ goes to zero as the sample size n goes to infinity at a rate such that $\frac{h_j n}{\ln n}$ goes to infinity for all j. The variable ω is another specification parameter. It determines how fast the bin sizes increase in capacity. In the application of this paper, the distribution of observations is close to a Pareto distribution. In this case, the function (136) keeps the number of observations in a bin approximately constant if ω is chosen appropriately. Section D.5.1 discusses the choice of c_0 and ω.

C.2 The consistency of $\hat{\epsilon}$ and $\hat{\eta}$

The estimates $\hat{\epsilon}$ and $\hat{\eta}$ are consistent because Conditions 3.1", 3.2, 3.4, and 3.5 i) in Chen (2007) hold. The cited results from Chen (2007) are summarised in Appendix F. The paragraphs below verify these conditions.

By Proposition 2 and its proof in Section A.2, Condition 3.1" is fulfilled. This paragraph shows that Conditions 3.2, 3.3 i), and 3.4 in Chen (2007) are fulfilled. Consider the sieve spaces $\Theta_P = \{\theta \text{ such that } 0 \leq \epsilon \leq \overline{\epsilon}, \text{ and } 0 \leq \eta \leq \overline{\eta}, \text{ and } |\gamma_p| \leq M(\ln \overline{q} - \ln q^K)^{-p} \text{ for all } p \leq P, \text{ where } \overline{\epsilon}, \overline{\eta}, M > 0; \gamma_p = 0 \text{ for all } p > P.\}$. Consider the norm

$$d(\theta, \tilde{\theta}) = \mid \epsilon - \tilde{\epsilon} \mid + \mid \eta - \tilde{\eta} \mid + \sup_p \mid \gamma_p - \tilde{\gamma}_p \mid .$$

The sieve spaces Θ_P are compact because they are closed and bounded; $\Theta_P \subseteq \Theta_{P+1} \subseteq \Theta$; the sequence $\pi_P \theta^o = (\epsilon^o, \eta^o, \gamma_0^o, ..., \gamma_P^o, 0, ...) \in \Theta_P$ and $\lim_{P \to \infty} d(\theta^o, \pi_P \theta^o) = 0$, where superscript "o" denotes the true parameters. Moreover, $Q(\theta)$ is continuous on Θ_P because it is a continuous function of its parameters.

This paragraph shows that Condition 3.5 i) in Chen (2007) is fulfilled. Consider a $\theta \in \Theta_P$

and rewrite the sample criterion $\widehat{Q_n}(\theta)$ (11) as:

$$\widehat{Q}_n(\theta) = \frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \bar{b}]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \theta) \right)^2 + \left(\widehat{\ln B} - \ln B(\theta) \right)^2. \tag{137}$$

The interval $[\underline{b}, q_L] \cup [q_H, \bar{b}]$ is partitioned into N bins of potentially unequal widths $h_j(n)$; $q_j(n)$ denotes the location of these bins; assume that $h_j(n)$ goes to zero and $\frac{h_j(n)n}{\ln n}$ goes to infinity for all j as the sample size n goes to infinity. Note that the number of bins N goes to infinity as n goes to infinity. The log-histogram $\widehat{\ln f(q_j)}$ may be constructed following Section C.1.3; alternatively, one could also use a different binning function as long as the bin sizes converge to zero at the appropriate rates. Consider

$$\begin{aligned} |\widehat{Q_n}(\theta) - Q(\theta)| = |\frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \bar{b}]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \theta) \right)^2 - \\ - \int_{[\underline{b}, q_L] \cup [q_H, \bar{b}]} \left(\ln f_k^o(q) - \ln f_k(q \mid \theta) \right)^2 dF^w + \left(\widehat{\ln B} - \ln B(\theta) \right)^2 - \left(\ln B^o - \ln B(\theta) \right)^2 |. \end{aligned} \tag{138}$$

Subtract and add $\frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \bar{b}]} \left(\ln f_k^o(q_j) - \ln f_k(q_j \mid \theta) \right)^2$ within this absolute value, use the triangle inequality, and take the limit:

$$\begin{aligned} &\lim_{n \to \infty} |\widehat{Q_n}(\theta) - Q(\theta)| \leq \\ &\leq \lim_{n \to \infty} |\frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \bar{b}]} \left[\left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \theta) \right)^2 - \left(\ln f_k^o(q_j) - \ln f_k(q_j \mid \theta) \right)^2 \right] | + \\ &+ \lim_{n \to \infty} | \int_{[\underline{b}, q_L] \cup [q_H, \bar{b}]} \left(\ln f_k^o(q) - \ln f_k(q \mid \theta) \right)^2 dF_N^w - \int_{[\underline{b}, q_L] \cup [q_H, \bar{b}]} \left(\ln f_k^o(q) - \ln f_k(q \mid \theta) \right)^2 dF^w | \\ &+ \lim_{n \to \infty} | \left(\widehat{\ln B} - \ln B(\theta) \right)^2 - \left(\ln B^o - \ln B(\theta) \right)^2 |. \end{aligned} \tag{139}$$

The third line uses
$\frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \bar{b}]} \left(\ln f_k^o(q_j) - \ln f_k(q_j \mid \theta) \right)^2 = \int_{[\underline{b}, q_L] \cup [q_H, \bar{b}]} \left(\ln f_k^o(q) - \ln f_k(q \mid \theta) \right)^2 dF_N^w$, where F_N^w is a CDF with steps of size $1/N$ at each q_j and constant anywhere else. Define F^w as the limit of this CDF when N goes to infinity. It follows that the expression in line 3 in (139) converges to zero by definition. Note that this line is not a random variable. The expression in line 2 converges to zero almost surely (see Einmahl and Mason, 2005).[37] The expression in line 4 converges to zero almost surely (see Lemma 14). It follows that $\widehat{Q_n}(\theta) \xrightarrow{a.s.} Q(\theta)$ for all $\theta \in \Theta_P$. Therefore, Condition 3.5 i) in Chen (2007) is fulfilled (see footnote 29, Chen, 2007.)

C.3 The estimation of the mean squared error

It is standard in nonparametric estimations to choose a specification that minimises an estimate of the mean squared error. As the estimate of the participation margin $\hat{\eta}$ is more sensitive to the specification than the intensive margin $\hat{\epsilon}$ (see the simulation in Table 4 in Section B.7), I use an estimate of its mean squared error to choose the specification. The estimate $\hat{\eta}(P, n)$ is

[37] Note that by Assumption 3.a or 3.b, $\mid \ln f_k^o(q_j) \mid$ is bounded uniformly.

a function of the specification parameter and the sample size n. The mean squared error is defined as

$$MSE = \mathbb{E}\left[(\hat{\eta}(P,n) - \eta^o)^2\right], \tag{140}$$

where η^o denotes the true value of the parameter. A standard bias-variance decomposition renders

$$MSE = \underbrace{\mathbb{E}\left[(\hat{\eta}(P,n) - \mathbb{E}\left[\hat{\eta}(P,n)\right])^2\right]}_{\text{variance}} + \underbrace{(\mathbb{E}\left[\hat{\eta}(P,n)\right] - \eta^o)^2}_{\text{bias}^2}. \tag{141}$$

Define

$$\tilde{\eta}(P) := \lim_{n\to\infty} \hat{\eta}(P,n), \tag{142}$$

where P is kept constant.[38] Intuitively, $\tilde{\eta}(P)$ is the biased value of the parameter under the parametric specification P. Therefore, $\tilde{\eta}(P) - \eta^o$ is the estimator's specification bias. Note that in large enough samples $\mathbb{E}\left[\hat{\eta}(P,n)\right] \approx \tilde{\eta}(P)$. Therefore, the mean squared error is the sum of the variance and the estimator's squared specification bias. All parts of the MSE are unknown and need to be estimated. Following Chetty et al. (2011), the variance can be estimated by the nonparametric bootstrap. Estimating the bias is more challenging since it depends on $\tilde{\eta}(P)$ and the true value η^o. A consistent estimator of $\tilde{\eta}(P)$ is $\hat{\eta}(P,n)$ itself. However, a consistent estimator of η^o, which converges fast, is challenging to find. The next section discusses the estimator of the bias.

C.3.1 The estimator of the bias

The first part of this section derives an approximate analytical expression of the bias. The second part discusses the estimation of the bias. In this section, I assume bunching is sharp. Note that in the application of this paper, bunching is relatively sharp, and the bunching interval is narrow (see Footnote 13 and Appendix D.6.3.) I leave a generalisation to non-sharp bunching for future research. Denote by $(\tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P)$ the estimate of $(\epsilon, \eta, \gamma_P)$ under specification P and when sample sizes goes to infinity, e.g., $\tilde{\eta} := \lim_{n\to\infty} \hat{\eta}(P,n)$, where P is kept constant. The parameter vector γ_P^o denotes the true $(\gamma_0^o, ..., \gamma_P^o)$. The parameter $\tilde{\gamma}_P$ is identified from the derivatives of the observed distribution to the left of the kink point. It is estimated without bias if the bandwidth $[\underline{b}, \overline{b}]$ converges to q^K as n goes to infinity. Assume the bandwidth is sufficiently narrow such that $\tilde{\gamma}_P^o \approx \gamma_P$. The simulation in Table 4 shows that the parameter ϵ is less sensitive to the specification than the parameter η. Assume that P is large enough such that $\tilde{\epsilon} \approx \epsilon^o$. Given ϵ^o and γ_P^o, the parameter $\tilde{\eta}$ is identified from points just to the right of the kink point. Again, assume the bandwidth is sufficiently narrow such that $\tilde{\eta}$ is estimated from the moment

$$\lim_{q\downarrow q^K} \ln f_k(q|\epsilon^o, \eta^o, \gamma^o) \approx \lim_{q\downarrow q^K} \ln f_k(q|\tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P). \tag{143}$$

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

Using Equation (7) renders

$$\sum_{p=0}^{\infty} \gamma_p^o \frac{1}{p!} \left(\ln \left(\frac{q^K \rho^{-\epsilon^o}}{q^K} \right) \right)^p + \eta^o \ln R(q^K \rho^{-\epsilon^o}, \epsilon^o) + \ln \left(\rho^{-\epsilon^o} \right) \tag{144}$$

$$\approx \sum_{p=0}^{P} \tilde{\gamma}_p \frac{1}{p!} \left(\ln \left(\frac{q^K \rho^{-\tilde{\epsilon}}}{q^K} \right) \right)^p + \tilde{\eta} \ln R(q^K \rho^{-\tilde{\epsilon}}, \tilde{\epsilon}) + \ln \left(\rho^{-\tilde{\epsilon}} \right). \tag{145}$$

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

$$\tilde{\eta} - \eta^o \approx \frac{\sum_{p=P+1}^{\infty} \gamma_p^o \frac{1}{p!} \left(\ln \left(\rho^{-\epsilon^o} \right) \right)^p}{\ln R(q^K \rho^{-\epsilon^o}, \epsilon^o)} \tag{146}$$

The bias in $\hat{\eta}$ depends mainly on ϵ and on the un-estimated rest of the parameter γ, i.e., $(\gamma_{P+1}, \gamma_{P+2}, ...)$. The formula motivates the estimation of the bias using out-of-sample data, which I discuss in the next paragraph.

Suppose the econometrician observes untreated data, i.e., there is no kink in the incentive scheme; otherwise, the untreated data is similar to the treated data. This section derives an estimate of the bias using such data. Denote by subscript ut and tr variables in the untreated and treated data respectively. In the untreated data, the true value η_{ut}^o is known: it is equal to zero because there is no treatment. Running the estimation on such data, therefore, estimates the specification bias. The estimate of the bias is

$$\widehat{bias(\hat{\eta}(P, n))} = \hat{\eta}_{ut}(P, n) - \eta_{ut}^o = \hat{\eta}_{ut}(P, n), \tag{147}$$

where $\hat{\eta}_{ut}$ is the estimate of the participation margin in the untreated data.[39] This estimate of the bias has the advantage that it converges at a parametric rate. What are the requirements on the untreated data such that $\hat{\eta}_{ut}$ is indeed an estimate of the bias $\tilde{\eta}_{tr} - \eta_{tr}^o$? Equation (160) above provides the answer. First, estimating the bias on untreated data is possible if the counterfactual distributions in the treated and untreated data are similar. More specifically, assume that there exists a certain order of the series expansion of the two distributions, such that all coefficients above that order are equal. Mathematically,

$$\ln(f_{l,ut}^o(q)) = \sum_{p=0}^{\infty} \gamma_{p,ut}^o \frac{1}{p!} \left(\ln \left(\frac{q}{q^K} \right) \right)^p, \tag{148}$$

$$\ln(f_{l,tr}^o(q)) = \sum_{p=0}^{\infty} \gamma_{p,tr}^o \frac{1}{p!} \left(\ln \left(\frac{q}{q^K} \right) \right)^p, \tag{149}$$

where $f_{l,ut}^o$ and $f_{l,tr}^o$ denotes the true counterfactual measure in the treated and untreated data respectively; $\gamma_{p,ut}^o$ and $\gamma_{p,tr}^o$ denotes the respective parameters γ. Assume there exists an order p^* such that for all $p \geq p^*$ the coefficients of these two distributions are equal: i.e., $\gamma_{p,tr}^o = \gamma_{p,ut}^o$, $\forall p \geq p^*$. Furthermore, assume that $P \geq p^*$.

Second, Equation (160) shows that the bias depends on the intensive margin response ϵ. To consider this dependence, first estimate $\tilde{\epsilon}$ on the treated data using an auxiliary specification. Second, simulate the intensive margin response in the untreated data by shifting observations

[39] It is important to not constrain $\hat{\eta}_{ut}$ to positive values.

by the intensive margin response using the auxiliary estimate $\tilde{\epsilon}$. Third, estimate $\eta_{ut}(P)$ in the untreated data with the simulated intensive margin response. A discussion of the choice of the untreated data is in Section D.5.2. A discussion of the selected specification is in Section D.5.3.

C.4 A summary of the empirical approach

This section gives a step-by-step guide of the estimation.

1. Choose a bunching interval $[q_L, q_H]$ by visually inspecting the histogram, count the number of choices in the bunching interval, and take the logarithm to estimate $\widehat{\ln B}$.

2. Estimate the log-histogram $\widehat{\ln(f(q_j))}$ by partitioning the data into bins, counting the choices in each bin, and applying the logarithm. To improve the small-sample properties of the estimator, it is possible to apply the bias-correction described in Section C.1.2. Moreover, it is advantageous to bin the data such that the variance of $\widehat{\ln(f(q_j))}$ is approximately constant (e.g., see Section C.1.3).

3. Estimate $\hat{\epsilon}, \hat{\eta}, \hat{\gamma_P}$ by minimising the square distance between $\widehat{\ln(f(q_j))}$ and $\widehat{\ln B}$ and the respective objects in the model.

4. Choose the bandwidth $[\underline{b}, \overline{b}]$, i.e., the range of the data used for the estimation, and the order of the series P, such that the estimate of the mean-squared error described in Section C.3 is at its minimum.

5. Check the robustness to the choice of the bunching interval (see Section D.6.3). Moreover, check if the rank condition holds (see Section D.6.4).

D Details empirical application

D.1 Data description

The data used in this paper are administrative and contain all solar panels connected to the grid and receiving subsidy payments. An adopter may be a household or a firm. The unit of observation is the aggregated capacity installed by an adopter at a specific location. Therefore, it is not possible to exploit the nonlinearities in the subsidy by splitting a large system into smaller ones and asking for separate payments for each. Additionally, when an adopter adds capacity to a preexisting system, the policymaker takes the preexisting capacity into account. Therefore, it is not possible to exploit the nonlinearities by splitting up a large adoption into smaller ones over time. The data provides information on the time point of adoption, the location, the electricity production, the applied subsidy rates, and the system's capacity. Table 7 shows the yearly number of adoptions. It is increasing in most years. Table 8 shows the subsidy schedule in Euro cents per kWh per capacity range and over time. Figure 19 shows additional evidence for responses at the two margins. The patterns are the same as in Figures 7 and 1.

Table 7: The number of adoptions per year.

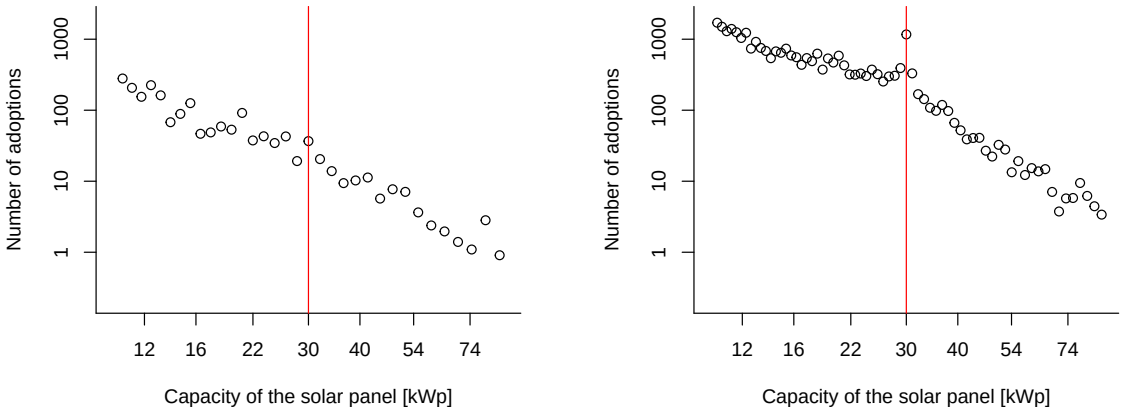
Year	Number of adoptions	Relative proportion in %
Until 2001	23498	10
2002	10999	5
2003	11928	5
2004	26070	11
2005	36448	15
2006	32730	13
2007	39883	16
2008	61220	25
All years	242776	100

Table 8: The subsidy schedule in Euro cents per kWh.

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

Note: Source Übertragungsnetzbetreiber (2018)

Figure 19: The histogram of adoptions in the years 2000 to 2002 (left panel) and 2005 (right panel).



Note: The red line marks the kink point. Scales are logarithmic. The left panel shows the histogram of adoptions in 2000 to 2002, when the subsidy was linear. As in 2003, there is no significant mass point or slope change in the distribution. The left panel pools years to have a sufficiently large sample size. The right panel shows the histogram in 2005, when there was a kink in the schedule. As in 2004, it shows a visible mass point and slope change.

D.2 Generalisations of the model

D.2.1 Heterogeneous discounting and radiation exposure

The German subsidy for solar panels is paid as a feed-in tariff. A feed-in tariff is a guaranteed fixed price for produced electricity. The subsidy payment depends on the installed capacity and the produced electricity. Electricity production is a function of the adopter-specific location and capacity. The location matters since climate conditions vary across locations. Moreover, adopters may have heterogeneous discount rates when evaluating future income streams. Discounting matters because adopters take the adoption decision based on the present discounted value of the income stream produced by the solar panel.

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

$$S_k(q, e_{it}) = s_l \, e_{it}, \qquad \text{for } q \leq q^K; \tag{150}$$

$$S_k(q, e_{it}) = s_l \, e_{it}\frac{q^K}{q} + s_l \rho e_{it}\frac{q - q^K}{q}, \qquad \text{for } q > q^K. \tag{151}$$

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

$$S_k(q, w_{it}) = s_l \, w_{it} \, q, \qquad \text{for } q \leq q^K; \tag{152}$$

$$S_k(q, w_{it}) = s_l w_{it} q^K + s_l \rho w_{it}(q - q^K), \qquad \text{for } q > q^K. \tag{153}$$

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

$$\mathbb{E}_i \left[\sum_{t=0}^{20} \beta_i^t w_{it} S_k(q) \right] = S_k(q) \mathbb{E}_i \left[\sum_{t=0}^{20} \beta_i^t w_{it} \right] = S_k(q) \zeta^i. \tag{154}$$

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

The decision problem of an adopter is

$$\tilde{\pi}_v^i = \max_q \{\zeta^i S(q) - \tilde{c}_v^i(q)\}, \tag{155}$$

and participate if and only if

$$\tilde{\pi}_v^i \geq \tilde{c}_f^i, \tag{156}$$

where ζ^i captures the individual-specific discounting and location. Normalisation by ζ^i shows the equivalence of Problem (155) and Problem (1). Therefore, the model outlined in Section 1 implicitly accounts for subsidy payments via a feed-in tariff. In particular, it accounts for individual-specific discounting and location.

D.2.2 The convexity of the variable cost function

There are at least two reasons why the cost function $c_v^i(.)$ in Section 1 is convex. First, for a given adopter i, an important margin of adjustment for increasing the capacity of her solar system is covering a larger area of her roof with solar panels. Typically, the opportunity cost of using space on the roof is convex. It is for two reasons: i) the larger the area already covered with solar panels, the more valuable the remaining space for alternative uses (e.g., windows, chimneys, solar heating systems, etc.); ii) the larger the area covered, the less aesthetic is the roof.

A second margin of adjustment is the average efficiency of the solar system. For any fixed area of the roof dedicated to solar panels, adopters can increase the total capacity of their system by using solar panels with a higher efficiency, i.e., solar panels with a higher peak capacity per area. More efficient solar panels are more expensive, and the higher the efficiency, the larger the price differences. Therefore, increasing the capacity of a system using the efficiency margin has a convex cost. Denote by A a dedicated area on the roof in square meters, by e_f the system's efficiency in kWp per square meter, and by p the price in euros per square meter. The price is a function of efficiency, where $p(.) > 0$, $p'(.) > 0$, and $p''(.) > 0$. It follows that the monetary cost of the system is $p(e_f)A = p(q/A)A$. It is easy to see that the cost of increasing capacity via efficiency is convex for any fixed area.

Moreover, I do not assume the cost function is convex everywhere. There can be ranges of increasing returns to scale. However, due to the constraints outlined above, the function $c_v^i(q)$ of a specific adopter i is convex for q large enough. Note that the optimal choice of q is always in the convex range of $c_v^i(q)$.

D.2.3 Dynamic decisions

Suppose there are two time periods. For simplicity, assume there is no discounting. Denote the decision problem in the current period by

$$\pi = \max \underbrace{S(q) - c_v(q, \theta)}_{\pi_v} - \tilde{c}_f \tag{157}$$

Denote with superscript $-t$ these variables in the other time period. Note that $-t$ may be before or after the current period:

$$\pi^{-t} = \max S^{-t}(q) - c_v^{-t}(q, \theta) - \tilde{c}_f^{-t} \tag{158}$$

An adopter participates in the current period if $\pi \geq \max(0, \mathbb{E}\pi^{-t})$. It is equivalent to $\pi_v \geq \tilde{c}_f + \max(0, \mathbb{E}\pi^{-t})$. Denote $c_f = \tilde{c}_f + \max(0, \mathbb{E}\pi^{-t})$. The problem is equivalent to Problem (1). The participation margin is the participation of adopters in the period under consideration.

D.3 A discussion of the empirical results

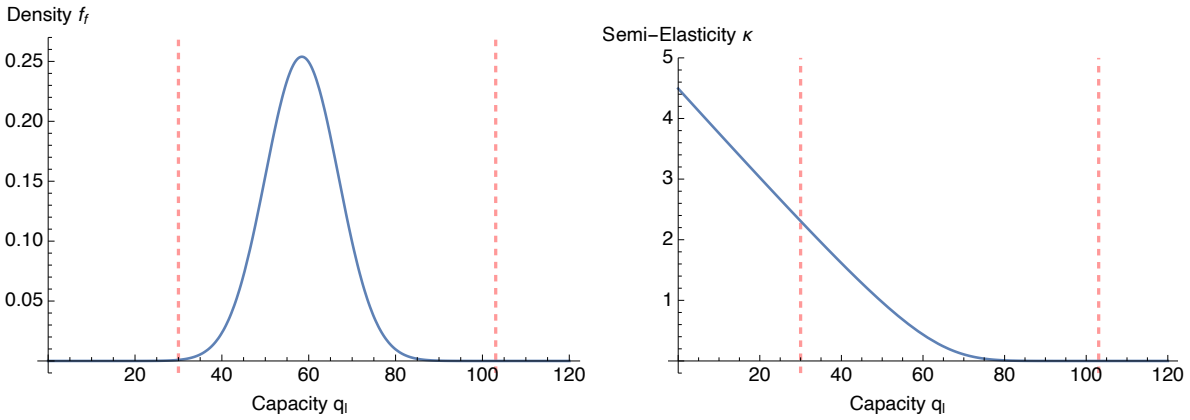
The results in Table 1 suggest that the intensive margin elasticity is the same for adopters of different capacities, while the participation margin semi-elasticity decreases with capacity. This pattern is not surprising. An important intensive adjustment margin is the quality of the solar panel. Low-capacity adopters have access to the same quality choices as high-capacity adopters. Therefore, their responses have the same elasticity at the intensive margin. It is useful to consider the underlying distribution of fixed costs to interpret the participation responses. A participation semi-elasticity that decreases in capacity is consistent with a simple normal distribution of fixed costs. The two semi-elasticities observed at the two capacity levels in Table 1 are sufficient to calibrate the distribution. Figure 20 shows the calibrated distribution and the implied semi-elasticities. The figure shows the calibrated density of fixed costs $f_f(\pi_v(q_l))$ and the semi-elasticity of participation $\kappa(\pi_v(q_l)) = \frac{f_f}{F_f}(\pi_v(q_l))$ as a function of the counterfactual variable profit $\pi_v(q_l)$ at capacity q_l.[40] Therefore, the function shows the mass of agents indifferent to participation and the semi-elasticity of participation at a certain capacity level under the counterfactual subsidy. The two red lines show the counterfactual capacity corresponding to the variable profit at the two kink points under the kinked subsidy. Appendix E.1.2 discusses the calibration in detail.

Figure 20 illustrates why the participation semi-elasticity in Table 1 decreases with capacity. The semi-elasticity depends on the variable profit π_v. The higher the profit, the lower the semi-elasticity. Large capacity systems have large variable profits.[41] Therefore, the fixed cost plays only a small role in the adoption decision, i.e., only a few adopters have such high fixed costs to make adopting a large capacity unprofitable. In contrast, adopting low-capacity systems depends crucially on fixed costs. In relative terms, many adopters have a fixed cost equal to the variable profit. It follows that relatively many adopters are close to indifferent to participating. Consequently, a small increase in the subsidy payment causes a large relative response. While the magnitude of the participation margin response at low capacities is surprising, there is an intuitive explanation. During the observation period, solar panels were a nascent product. They had high market potential compared to market coverage, reflected by the fact that 30 kWp is to

[40] I.e., $\pi_v(q_l) = s_l q_l - c_v(q_l, q_l) = s_l q_l/(1+\epsilon)$; it increases in q_l. See Equation (13) in Appendix A.1 for the definition of $c_v(.,.)$.

[41] See Figure 3 in Section 1.1 for an graphical illustration or footnote 40 for a mathematical derivation.

Figure 20: The calibrated density of fixed costs and the implied semi-elasticity.



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

the left of the bell curve in Figure 20. This feature explains why, compared to mature products, the participation semi-elasticities are high. Table 9 shows the yearly estimates at 30 kWp. Indeed, estimated participation semi-elasticities decrease over time as the market saturates. This explanation suggests that one can expect similar response patterns to other deployment subsidies for early-stage technologies.

Table 9: Yearly estimates at the kink point at 30 kWp.

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

Note: The table reports the estimated intensive margin elasticity $\hat{\epsilon}$ and participation margin semi-elasticity $\hat{\kappa}$ at the kink point at 30 kWp from 2004 to 2008. The standard errors are in brackets.

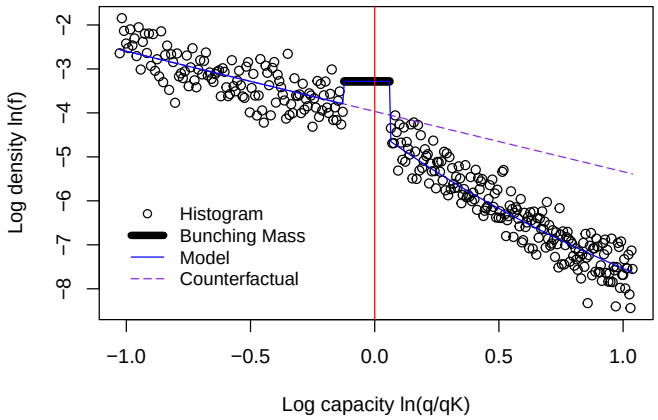
D.4 The selected specification and figures for the estimates in Section 3.2

The optimal bandwidth for the estimates at 30 kWp in Table 1 is $[\underline{b}, \overline{b}] = [10.6, 85]$; the optimal order of the series is $P = 1$; the bunching interval is $[q_L, q_H] = [26.5, 31.75]$. The optimal bandwidth for the estimates at 100 kWp is $[\underline{b}, \overline{b}] = [42, 400]$; the optimal order of the series is $P = 1$; the bunching interval is $[q_L, q_H] = [95, 102.5]$. For the detailed procedure to select this specification, see Sections D.5 and D.6.3. At both kink points, the optimal bandwidth is

relatively large, and the series' optimal order is low. This is not surprising; the graphical evidence in Figures 7 and 1 shows that the counterfactual distribution is very close to a Pareto distribution. The estimation of the standard errors uses the nonparametric bootstrap with 200 repetitions at 30kWp and 1000 repetitions at 100kWp and 250 kWp. An exponential transformation constrains $\hat{\epsilon}$ and $\hat{\eta}$ to positive values.

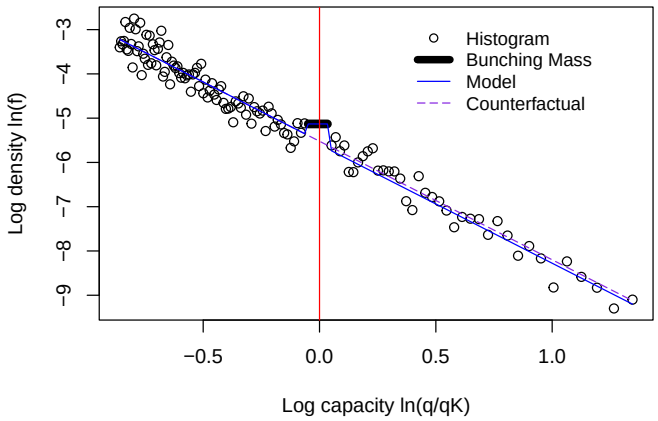
Figures 21 and 22 show the histogram with the estimated model and the counterfactual.

Figure 21: The distribution of adoptions from 2004 to 2008 at 30 kWp with the estimated model and counterfactual.



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

Figure 22: The distribution of adoptions from 2004 to 2008 at 100 kWp with the estimated model and counterfactual.



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

D.5 The selection of the specification

D.5.1 The parameters of the binning function

The binning function in Equation (136) in Section C.1.3 should guarantee a constant variance of the log-histogram and sufficiently many observations in each bin. To select its parameters, I pre-select initial maximal bandwidths. For the estimation at 30 kWp, I use a maximal bandwidth of $[9.5, 95]$.[42] For the estimation at 100 kWp, I use the interval $[42, 1600]$.[43] Note that these are the bandwidths for selecting the parameters of the binning function. The bandwidths for the main estimates are selected in Section D.5.3 below.

In the next step, I choose the size of the bin at the kink point h_0 and the scaling parameter ω. Note that there is a one-to-one mapping between the parameter c_0 in function (136) and h_0.[44] I choose the two parameters such that there are at least four observations in each bin, and the variance of the log-histogram in an auxiliary estimation is approximately constant. By Equation (134), four observations guarantee that the bias introduced by the logarithm is of the order $\frac{1}{16}$. The procedure gives the following specification:

Table 10: The selected bin sizes h_0 and scaling parameters ω.

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

Note: The table shows the selected bin size h_0 and scale parameter ω of the binning function in Equation (136). The parameters guarantee a minimum of 4 observations in each bin and an approximately constant variance.

D.5.2 The untreated data for estimating the bias

As discussed in Section C.3.1, for each kink point, it is necessary to choose a range of untreated data to estimate the bias. A natural choice is observations in the years 2000 to 2003. In these years, the subsidy was linear. I use it to estimate the bias at 30kWp.

For the kink point at 100 kWp, the pre-treatment data is not a satisfactory choice to estimate the bias. It is for two reasons. First, the number of observations around 100 kWp is very low in these years. Second, in 2000-2003, the data do not specify whether a solar panel was installed on a rooftop or the ground. From 2004 onwards, the data specifies where a solar system is installed. This paper only considers rooftop solar panels. Overall, ground panels are only a very small share of installations. Also, the subsidy for ground panels is linear in all years.

[42] I use this interval because it 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.

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

[44] $h_0 = q^K(1 - c_0)^{-\frac{1}{\omega}} - q^K$.

After 2004 and close to capacity 30kWp, only very few panels are ground panels. Therefore, the fact that the sample from 2000-2003 contains ground panels does not pose a concern for using these years to estimate the bias at 30 kWp. However, this is not true for capacities close to 100 kWp. There is a significant number of ground panel installations exactly at 100 kWp in the years after 2004. For these reasons, I cannot use observations around 100 kWp in the years 2000-2003 to estimate the bias at 100 kWp. To be conservative, I remove the observations at and around 100kWp when using the data to estimate the bias of the parameter at 30 kWp.

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

Section D.5.3 discusses the selected specification.

D.5.3 The order of the series and the bandwidth

As discussed in Section C.3, this section estimates the mean squared error using the variance estimate from the treated sample and the bias estimate from the untreated sample. To estimate the variance, it uses the nonparametric bootstrap. I estimate the MSE for P={1,2,3} and the bandwidths {[16.4, 55], [15.0, 60], [13.8, 65], [12.9, 70], [12.0, 75], [11.2, 80], [10.6, 85], [10.0, 90], [9.5, 95]}. At 100 kWp I use the bandwidths {[67, 150], [50, 200], [42, 400], [42, 700], [42, 1000], [42, 1300], [42, 1600]}. The optimal bandwidth and order are: $[10.6, 85]$ and $P = 1$ at 30 kWp, $[42, 400]$ and $P = 1$ at 100 kWp.

D.6 Robustness checks

D.6.1 The robustness check at 100 kWp

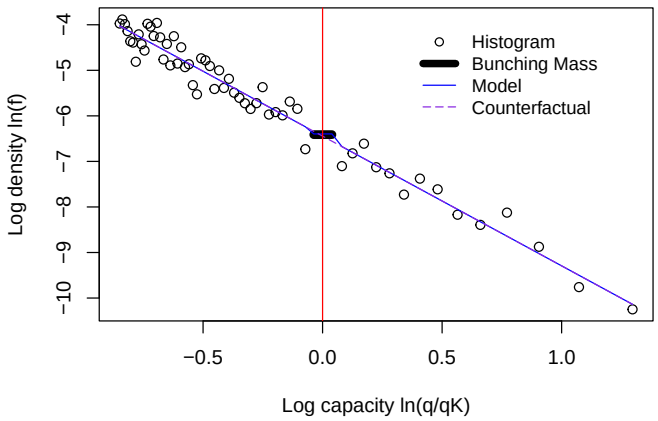
I run the robustness check for the estimation at 100 kWp on the data around 250 kWp pooling observations from 2004 to 2008. For a discussion of this choice of data, see Section D.5.2. Both estimates are insignificant.

Table 11: The results for the untreated data at 250 kWp.

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

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

Figure 23: The distribution of adoptions at 250 kWp with the estimated model.



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

D.6.2 An estimator of the specification bias

This section proposes an estimator of the specification bias caused by an eventual violation of Assumption 3.b. To this end, consider an estimator as in Section 2, but with fixed polynomial order P and a bandwidth $[\underline{b}, \bar{b}]$ that converges to q^K as n goes to infinity. Denote by $(\tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P) = \lim_{n\to\infty}(\hat{\epsilon}, \hat{\eta}, \hat{\gamma}_P)_n$ the asymptotic values of the estimates as the sample size n goes to infinity. By a slight abuse of notation, the vector γ_P denotes the parameters $(\gamma_0, ..., \gamma_P)$. The superscript "o" denotes the true values in population. Note that $\tilde{\eta}$ may be asymptotically biased, i.e., not equal to η^o. However, under some assumptions, it is possible to characterise and estimate this bias:

Proposition 6. *Assume bunching is sharp,* $\ln f_l^o(q_l)$ *is* P *times differentiable in* $\ln q_l$ *at point* $\ln q^K$*, and* $\tilde{\epsilon} = \epsilon^o$*.* [45] *It follows that the bias*

$$|\tilde{\eta} - \eta^o| = |\frac{h(q^K \rho^{-\epsilon^o}) \left(\ln\left(\frac{q^K \rho^{-\epsilon^o}}{q^K}\right)\right)^P}{\ln R(q^K \rho^{-\epsilon^o}, \epsilon^o)}|, \tag{159}$$

where the numerator is the rest of the P-th order Taylor approximation of $\ln f_l^o(.)$ *at* $\ln q^K$*.*

The proof is in Section D.6.5. The specification bias of $\hat{\eta}$ depends on ϵ and on the rest of the Taylor series. The formula motivates the estimation of the bias using untreated data, which I discuss in the next paragraph.

[45] Note that in the application of this paper, bunching is relatively sharp, and the bunching interval is narrow (see Footnote 13 and Appendix D.6.3). I leave a generalisation to non-sharp bunching for future research. Also note that the simulation in Table 4 in Section B.7 shows that the parameter ϵ is less sensitive to the specification than the parameter η, which motivates assuming $\tilde{\epsilon} = \epsilon^o$.

Suppose the econometrician observes untreated data, i.e., there is no kink in the incentive scheme. Otherwise, the untreated data is similar to the treated data. Denote by subscripts ut and tr variables for the untreated and treated data respectively. Equation (160) shows that the bias depends on the intensive margin response ϵ. To consider this dependence, simulate the intensive margin response in the untreated data by shifting observations by the intensive margin response using the estimate $\hat{\epsilon}_{tr}$ from the treated data. Then estimate $\hat{\eta}_{ut}$ in the untreated data using the same specification as in the treated data.[46] In the untreated data, the true value η^o_{ut} is known: it is equal to zero because there is no treatment. Therefore, by Proposition 6

$$|\tilde{\eta}_{ut} - \eta^o_{ut}| = |\tilde{\eta}_{ut}| = |\frac{h_{ut}(q^K\rho^{-\epsilon^o})\left(\ln\left(\frac{q^K\rho^{-\epsilon^o}}{q^K}\right)\right)^P}{\ln R(q^K\rho^{-\epsilon^o}, \epsilon^o)}|, \tag{160}$$

where the numerator is the rest of the P-th order Taylor approximation of the true counterfactual distribution in the untreated data. Under the assumption that the treated and untreated counterfactual distributions are similar in the sense that $h_{ut}(q^K\rho^{-\epsilon^o}) = h_{tr}(q^K\rho^{-\epsilon^o})$, it follows from Proposition 6 that $|\tilde{\eta}_{ut}| = |\tilde{\eta}_{tr} - \eta^o_{tr}|$. Therefore, $|\hat{\eta}_{ut}|$ is an estimate of the specification bias of $\hat{\eta}_{tr}$.

A discussion of the choice of the untreated data is in Section D.5.2. The estimates of the bias are in Table 12. The estimated biases are small and statistically insignificant.

Table 12: The estimates of the specification bias.

Capacity	$\mid$ Bias$(\hat{\kappa})\mid$ (SD)
30 kWp	0.03 (0.32)
100 kWp	0.06 (0.49)

Note: The table shows the estimates of the specification bias of the participation semi-elasticity κ. The bias is small and statistically insignificant.

D.6.3 Robustness bunching interval

After the histogram's visual inspection, I choose the two bunching intervals $[26.5, 31.5]$ and $[95, 102.5]$. The intervals are asymmetric because there is more non-sharp bunching before the kink point than after. As discussed in Footnote 13, non-sharp bunching can be explained by the unavailability of the exact optimal system size at the purchase date. This section reports the robustness of the estimates to changes in the bunching interval.

[46] It is important to not constrain $\hat{\eta}_{ut}$ to positive values.

Table 13: The estimates at 30 kWp for various bunching intervals.

Interval	$\hat{\kappa}$ (SD)	$\hat{\epsilon}$ (SD)
[25.50, 32.250]	2.34 (0.06)	4.18 (0.14)
[25.75, 32.125]	2.31 (0.06)	4.34 (0.14)
[26.00, 32.000]	2.31 (0.06)	4.34 (0.14)
[26.25, 31.875]	2.30 (0.06)	4.38 (0.14)
[26.50, 31.750]	2.31 (0.06)	4.37 (0.13)
[26.75, 31.625]	2.33 (0.06)	4.23 (0.12)
[27.00, 31.500]	2.31 (0.06)	4.34 (0.12)
[27.25, 31.375]	2.36 (0.05)	4.00 (0.10)
[27.50, 31.250]	2.38 (0.05)	3.89 (0.10)

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

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

Table 14: The estimates at 100 kWp for various bunching intervals.

Interval	$\hat{\kappa}$ (SD)	$\hat{\epsilon}$ (SD)
[92.00, 104.00]	0.00 (0.01)	5.12 (0.99)
[93.50, 103.25]	0.00 (0.02)	5.05 (0.89)
[95.00, 102.50]	0.00 (0.02)	4.63 (0.84)
[96.50, 101.75]	0.00 (0.03)	4.51 (0.80)
[98.00, 101.00]	0.00 (0.03)	4.67 (0.68)

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

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

D.6.4 The empirical verification of the rank conditions

Consider Condition 1 and its local counterpart Condition 2 in Section B.4.3. By Lemma 11 and 12, both conditions hold generically. However, identification could be weak. For example, Condition 2 could be almost equal to zero. This section verifies the two conditions empirically at the estimated values and shows that they hold by a large amount. First, consider Condition 2. Table 15 shows the result; the standard errors are in brackets:

Condition 1 holds in the application of this paper since property (b) in Lemma 11 holds. Table 16 shows the result; the standard errors are in brackets:

Table 15: The rank condition evaluated at the estimated values and $q = q^K$.

Capacity	Rank Condition (SD)
30 kWp	148 (9)
100 kWp	2866 (1164)

Note: The table shows Condition 2 evaluated at the estimated values and $q = q^K$. The standard errors are in brackets. The condition holds by a large amount.

Table 16: Property (b) in Lemma 11.

Capacity	$\hat{\gamma}_1$ (SD)
30 kWp	-1.37 (0.02)
100 kWp	-2.68 (0.03)

Note: The table shows $\gamma_1 = d \ln f_l(d)/d \ln q$ at the two capacity levels. The condition in Lemma 11 holds by a large amount.

D.6.5 Proof of Proposition 6

By Taylor's Theorem

$$\ln f_l^o(q_l) = \sum_{p=0}^{P} \gamma_p^o \left(\ln \frac{q_l}{q^K} \right)^p + h(q_l) \left(\ln \frac{q_l}{q^K} \right)^P, \text{ where } \lim_{q_l \to q^K} h(q_l) = 0. \tag{161}$$

Following the corresponding steps of the proof in Section C.2, the first part of the sample criterion (11) converges to the population criterion

$$\min_{\tilde{\eta}, \tilde{\gamma_P}} \lim_{n \to \infty} \int_{\underline{b}(n)}^{q^K} \left(\ln f_k^o(q) - \ln f_k \left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P \right) \right)^2 \, dF^w + \int_{q^K}^{\bar{b}(n)} \left(\ln f_k^o(q) - \ln f_k \left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P \right) \right)^2 \, dF^w. \tag{162}$$

The function $f_k^o(.)$ denotes the observable measure of agents in the population. The function $f_k \left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P \right)$ denotes the respective function in Proposition 1, where the function $f_l(.)$ is approximated by the finite power series of order P in Equation (10). The function F^w denotes a known weighting measure.

Step 1: $\tilde{\gamma}_0 = \gamma_0$.
Differentiate Equation (162) with respect to $\tilde{\gamma}_0$ and assume $\bar{b}$ converges sufficiently faster to q^K than $\underline{b}$. It follows that

$$0 = \lim_{n \to \infty} \int_{\underline{b}(n)}^{q^K} \left(\ln f_k^o(q) - \ln f_k \left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P \right) \right) \frac{\frac{\partial f_k(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P)}{\partial \tilde{\gamma}_0}}{f_k \left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P \right)} \, dF^w. \tag{163}$$

Use Equation (161), Proposition 1, and Equation (10):

$$0 = \lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \left[\sum_{p=0}^{P} \left(\gamma_p^o - \tilde{\gamma_p}\right) \frac{1}{p!} \left(\ln\left(\frac{q}{q^K}\right)\right)^p + h(q) \left(\ln\left(\frac{q}{q^K}\right)\right)^P \right] \frac{1}{0!} \left(\ln\left(\frac{q}{q^K}\right)\right)^0 dF^w \tag{164}$$

Solve for $(\tilde{\gamma}_0 - \gamma_0^o)$:

$$\tilde{\gamma}_0 - \gamma_0^o = \frac{\lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \left[\sum_{p=1}^{P} \left(\gamma_p^o - \tilde{\gamma_p}\right) \frac{1}{p!} \left(\ln\left(\frac{q}{q^K}\right)\right)^p + h(q) \left(\ln\left(\frac{q}{q^K}\right)\right)^P \right] \frac{1}{0!} \left(\ln\left(\frac{q}{q^K}\right)\right)^0 dF^w}{\lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \frac{1}{0!0!} \left(\ln\left(\frac{q}{q^K}\right)\right)^{2*0} dF^w} \tag{165}$$

Take the limit using l'Hopital:

$$\tilde{\gamma}_0 - \gamma_0^o = \frac{\sum_{p=1}^{P} \left(\gamma_p^o - \tilde{\gamma_p}\right) \frac{1}{p!} \left(\ln\left(\frac{q^K}{q^K}\right)\right)^p + h(q^K) \left(\ln\left(\frac{q^K}{q^K}\right)\right)^P}{1} = 0 \tag{166}$$

Step 2: $\tilde{\gamma}_p = \gamma_p^o$ for all $p \leq P$.
Suppose $\gamma_p^o = \tilde{\gamma}_p$ for all $p < \pi$. Show that $\gamma_\pi^o = \tilde{\gamma}_\pi$. To this end, proceed as in step 1. First, differentiate Equation (162) with respect to $\tilde{\gamma}_\pi$ and assume $\overline{b}$ converges sufficiently faster to q^K than $\underline{b}$. It follows that

$$0 = \lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)\right) \frac{\frac{\partial f_k(q|\tilde{\epsilon},\tilde{\eta},\tilde{\gamma}_P)}{\partial \tilde{\gamma}_\pi}}{f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)} dF^w \tag{167}$$

Use Equation (161), Proposition 1, Equation (10), $\gamma_p^o = \tilde{\gamma}_p$ for all $p < \pi$, rearrange and use l'Hopital to derive that

$$\tilde{\gamma}_\pi - \gamma_\pi^o = \lim_{q\to q^K} \frac{\sum_{p=\pi+1}^{P} \left(\gamma_p^o - \tilde{\gamma_p}\right) \frac{1}{p!} \left(\ln\left(\frac{q}{q^K}\right)\right)^p + h(q) \left(\ln\left(\frac{q}{q^K}\right)\right)^P}{\frac{1}{\pi!} \left(\ln\left(\frac{q}{q^K}\right)\right)^\pi} = 0. \tag{168}$$

By induction, $\tilde{\gamma}_p = \gamma_p^o$ for all $p \leq P$.

Step 3: $\tilde{\eta} - \eta^o$.
Differentiate Equation (162) with respect to $\tilde{\eta}$:

$$0 = \lim_{n\to\infty} \int_{q^K}^{\bar{b}(n)} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)\right) \frac{\frac{\partial f_k(q|\tilde{\epsilon},\tilde{\eta},\tilde{\gamma}_P)}{\partial \tilde{\eta}}}{f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)} dF^w \tag{169}$$

Assume $\tilde{\epsilon} = \epsilon^o$, use $\gamma_p^o = \tilde{\gamma}_p$, Equation (161), Proposition 1, and Equation (10):

$$0 = \lim_{n\to\infty} \int_{q^K}^{\bar{b}(n)} \left[h(q\rho^{-\epsilon^o}) \left(\ln \left(\frac{q\rho^{-\epsilon^o}}{q^K} \right) \right)^P + (\eta^o - \tilde{\eta}) \ln R\left(q\rho^{-\epsilon^o}\right) \right] \ln R\left(q\rho^{-\epsilon^o}\right) \, dF^w \tag{170}$$

Solve for $(\tilde{\eta} - \eta^o)$ and take the limit using l'Hopital:

$$\tilde{\eta} - \eta^o = \frac{h(q^K\rho^{-\epsilon^o}) \left(\ln \left(\frac{q^K\rho^{-\epsilon^o}}{q^K} \right) \right)^P}{\ln R(q^K\rho^{-\epsilon^o}, \epsilon^o)} \tag{171}$$

qed.

D.7 A comparison to the classic estimators

The classic bunching estimator uses a parametric functional form assumption on the counterfactual distribution over the bunching range $[q^K, q^K\rho^{-\epsilon}]$ to estimate the intensive margin elasticity (e.g., see Chetty et al., 2011). The bunching range $[q^K, q^K\rho^{-\epsilon}]$ is the theoretical range of agents who bunch. Note that it is not equal to the bunching interval $[q_L, q_H\rho^{-\epsilon}]$ in Equation (6). Depending on the true shape of the counterfactual, this parametric assumption may introduce substantial specification bias. See Blomquist and Newey (2017), Bertanha, McCallum and Seegert (2023), and Bertanha et al. (2023), for a detailed discussion of this bias. To alleviate the concerns of specification bias, Section 2 estimates the counterfactual distribution nonparametrically. Moreover, the classic bunching estimator ignores participation margin responses. Given a certain amount of bunching, ignoring the participation margin downward biases the estimate of the intensive margin elasticity. Intuitively, relatively little bunching is wrongly attributed to a small intensive margin response instead of the participation margin response. The estimator in Section 2 specifically accounts for the participation margin.

I evaluate these biases in my application in two steps, using the pooled data from 2004 to 2008 at 30 kWp. First, I estimate an intensive margin elasticity ϵ_C assuming the counterfactual is constant in the bunching region as in Chetty et al. (2011). I use the rest of the parameters from the correctly specified estimation (11), and I specify the model correctly outside the bunching range:

$$\epsilon_C = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln B_C(\epsilon, \eta = \hat{\eta}, \gamma = \hat{\gamma}) \right)^2, \tag{172}$$

where

$$B_C(\epsilon, \eta, \gamma) = \int_{q_L}^{q^K} f_l(q_l|\gamma) dq_l + \int_{q^K}^{q^K\rho^{-\epsilon}} f_l(q^K|\gamma) dq_l + \int_{q^K\rho^{-\epsilon}}^{q_H\rho^{-\epsilon}} f_l(q_l|\gamma) R(q_l, \epsilon)^\eta dq_l, \tag{173}$$

and $\hat{\gamma}$ and $\hat{\eta}$ are the unbiased estimates from (11). As a comparison, in the correct specification, the second integral in Equation (173) is

$$\int_{q^K}^{q^K\rho^{-\epsilon}} f_l(q_l|\gamma) R(q_l, \epsilon)^\eta dq_l. \tag{174}$$

Row 1 in Table 17 shows the result. The intensive margin elasticity has a downward bias of 23 %.

Second, to distinguish the bias from the misspecified counterfactual from the bias of ignoring participation, I estimate an intensive margin elasticity $\tilde{\epsilon}_C$ using the right specification of the counterfactual but ignoring the participation margin in the bunching region. Again, I use the rest of the parameters from the correctly specified estimation (11), and I specify the model correctly outside the bunching range:

$$\tilde{\epsilon}_C = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln \tilde{B}_C(\epsilon, \eta = \hat{\eta}, \gamma = \hat{\gamma}) \right)^2, \quad (175)$$

where

$$\tilde{B}_C(\epsilon, \eta, \gamma) = \int_{q_L}^{q^K} f_l(q_l|\gamma) dq_l + \int_{q^K}^{q^K \rho^{-\epsilon}} f_l(q_l|\gamma) dq_l + \int_{q^K \rho^{-\epsilon}}^{q_H \rho^{-\epsilon}} f_l(q_l|\gamma) R(q_l, \epsilon)^\eta dq_l, \quad (176)$$

and $\hat{\gamma}$ and $\hat{\eta}$ are the unbiased estimates from (11). Again, in the correct specification, the second integral in Equation (176) is Equation (174). Row 2 in Table 17 shows the result. The intensive margin elasticity has a downward bias of 12 %. It reveals that around half of the bias in ϵ_C can be attributed to ignoring the participation margin.[47]

Correspondingly, ignoring intensive margin responses biases the estimate of the participation margin elasticity. To evaluate the bias, I proceed correspondingly to the above:

$$\tilde{\eta} = \arg\min_{\eta} \sum_{q_j \notin [q_L, q_H]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \eta, \gamma = \hat{\gamma}, \epsilon = 0) \right)^2. \quad (177)$$

Row 3 in Table 17 shows the result. Ignoring the intensive margin introduces an upward bias of 5% in the estimate of the participation semi-elasticity. The exercises illustrate that simultaneously estimating the two margins is necessary to obtain unbiased estimates.

Table 17: A comparison of the biased and unbiased estimates.

Parameter	Unbiased Estimate	Biased Estimate	Relative Difference in %
ϵ_C	4.37	3.39	-23
$\tilde{\epsilon}_C$	4.37	3.87	-12
κ	2.31	2.43	5

Note: The table shows the unbiased estimates in the second column. The third column shows the biased estimates of the misspecified model. The fourth column shows the relative magnitude of the bias.

To estimate the bunching mass, Chetty et al. (2011) estimate the counterfactual distribution using observations outside the bunching interval. To this end, they fit a polynomial of order seven to points outside the bunching interval. While their procedure differs in several dimensions from the theoretically consistent estimation of the counterfactual in Section 2, one similarity is the need to choose the order of the series P. They show that, qualitatively, their results do not change if one changes the order of the polynomial; however, they change quantitatively. An advantage of the nonparametric estimator in Section 2 is that the series order is chosen optimally. The optimal order is low, i.e., $P = 1$. See Section D.4 for a discussion. To illustrate the importance of the choice of P, Table 18 compares the optimal estimates to the

[47] Note that, for both exercises, I specify the model correctly outside the bunching range and inside the bunching interval, i.e., for $q_l \in [q_L, q^K] \cup [q^K \rho^{-\epsilon}, q_H \rho^{-\epsilon}]$. If one uses the misspecified models for all q in $[q_L, q_H]$, the biases become more severe.

estimates using $P = 7$ like in Chetty et al. (2011). As in Chetty et al. (2011), the estimates do not differ qualitatively. Statistically, they are not significantly different from each other because the estimates using $P = 7$ have very large standard errors. However, the point estimates show large quantitative differences, and the estimates using $P = 7$ have much larger standard errors. The comparison underlines the importance of choosing the polynomial order optimally to avoid unnecessarily imprecise estimates.

Table 18: A comparison of the optimal polynomial order and order $P = 7$.

Parameter	Optimal Estimate	P=7
ϵ	4.37 (0.13)	3.78 (0.39)
κ	2.31 (0.06)	3.25 (1.02)

Note: The table shows the estimates using the optimal order of the series in the second column. The third column shows the estimates using order $P = 7$.

E Details policy evaluation

E.1 Additional assumptions and calibration

Like other kink and discontinuity estimators in the literature, Section 3 estimates the responses locally at each kink point. It is necessary to make global assumptions to use them for counterfactual exercises. In line with the empirical evidence in Table 1, assume an isoelastic cost function:

Assumption 6 (Isoelastic cost function)**.** *The cost function is isoelastic with a constant intensive margin elasticity ϵ:*

$$c(q, q_l, c_f) = \frac{s_l}{q_l^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}} + c_f, \tag{178}$$

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

As discussed in Section D.3, the estimated participation margin responses in Table 1 are in line with a normally distributed fixed cost:

Assumption 7 (Normally distributed fixed costs)**.** *The distribution of the fixed costs c_f is normal: $c_f \sim N(\mu_f, \sigma_f)$, with CDF $F_f(c_f)$ and density $f_f(c_f)$.*

Assumption 7 implies c_f and q_l are independent, as in Rochet and Stole (2002). Appendix E.1.1 generalises the assumption to allow for correlation between c_f and q_l.

Set ϵ equal to the estimated intensive margin elasticity in Table 1 and calibrate (μ_f, σ_f) using the estimated participation semi-elasticities (see Figure 20 and Appendix E.1.2 for details). Appendix E.1.2 describes the estimation of the type-distribution $f_l(.)$. It is a log-normal distribution for low capacities and a Pareto distribution for large capacities. The model can be solved for any counterfactual subsidy scheme using these assumptions.

E.1.1 Extension Assumption 7

Assumption 7 implies that q_l and c_f are independent. However, the assumption can be easily extended to allow for a correlation between q_l and c_f. Assume c_f follows a truncated normal distribution: $c_f \sim N(\mu_f, \sigma_f, \underline{c_f}(q_l), \overline{c_f}(q_l))$, where $\underline{c_f}(q_l)$, $\overline{c_f}(q_l)$ are the truncation bounds which can vary with q_l. Denote by $F_f(.)$ the CDF of the normal distribution. It follows that the CDF of c_f is $\frac{F(c_f)}{F(\overline{c_f}(q_l))-F(\underline{c_f}(q_l))}$; hence, c_f and q_l may be correlated. Assume that the bounds $\underline{c_f}(q_l)$, $\overline{c_f}(q_l)$ are large enough such that the variable profit of an agent implied by any of the counterfactual exercises lies within the bounds. As a consequence, all results in Section 4 remain unchanged.

E.1.2 The calibration and estimation of the type distributions

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

$$S_k(q) = q, \qquad \text{for } q \leq q_1^K; \tag{179}$$
$$S_k(q) = q_1^K(1-\rho_1) + q\rho_1, \qquad \text{for } q \in (q_1^K, q_2^K]; \tag{180}$$
$$S_k(q) = q_1^K(1-\rho_1) + q_2^K\rho_1(1-\rho_2) + q\rho_1\rho_2, \qquad \text{for } q > q_2^K. \tag{181}$$

Using Assumption 6 , the choice q as a function of type q_l is:

$$q(q_l) = q_l, \qquad \text{for } q_l \in [q_l^{min}, q_1^K]; \tag{182}$$
$$q(q_l) = q_1^K, \qquad \text{for } q_l \in [q_1^K, q_1^K\rho_1^{-\epsilon}]; \tag{183}$$
$$q(q_l) = q_l\rho_1^{\epsilon}, \qquad \text{for } q_l \in [q_1^K\rho_1^{-\epsilon}, q_2^K\rho_1^{-\epsilon}]; \tag{184}$$
$$q(q_l) = q_2^K, \qquad \text{for } q_l \in [q_2^K\rho_1^{-\epsilon}, q_2^K(\rho_1\rho_2)^{-\epsilon}]; \tag{185}$$
$$q(q_l) = q_l(\rho_1\rho_2)^{\epsilon}, \qquad \text{for } q_l \in [q_2^K(\rho_1\rho_2)^{-\epsilon}, q_l^{max}], \tag{186}$$

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

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

$$\frac{f_f}{F_f}\left(S_k(q) - c_v(q, q_l(q))|\mu_f, \sigma_f\right) = \kappa(q), \tag{187}$$

where $q_l(q)$ denotes the inverse of $q(q_l)$ and $c_v(q, q_l)$ denotes the variable part of the cost function in Assumption 6.

Using Assumptions 6, 7, the results in Table 1, and inverting this equation at the two kink points gives (μ_f, σ_f) in Table 19.[48] Figure 20 shows the calibrated density of fixed costs and the implied semi-elasticities of participation at counterfactual capacities q_l. Note that the two red lines illustrate the counterfactual capacity with corresponding variable profit as under the observed subsidy, i.e., q_l solves $q_l - c_v(q_l, q_l) = S(q^K) - c_v(q^K, q_l(q^K))$. Because at the first kink point $q = q_l$, the first red line is at $q_1^K = 30$, while the second red line is slightly above $q_2^K = 100$.

As suggested by the empirical evidence, assume the distribution $f_l(.)$ of the variable cost

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

Table 19: The parameters of the distribution of the fixed and variable cost types.

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

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

type q_l is log-normal in its lower part and Pareto in its upper part:

$$f_l(q_l) = \exp\left(\gamma_{0l} + \gamma_{1l}\ln(q_l) + \gamma_{2l}\ln(q_l)^2\right) \quad \text{for } q_l \in [q_l^{min}, q_{lb}] \tag{188}$$

$$f_l(q_l) = \exp\left(\gamma_{0u} + \gamma_{1u}\ln(q_l)\right) \quad \text{for } q_l \in (q_{lb}, q_l^{max}]. \tag{189}$$

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

$$f_k(q) = f_l(q_l(q))\frac{F_f(S_k(q) - c_v(q, q_l(q))}{F_f(q_l(q) - c_v(q_l(q), q_l(q)))}\frac{dq_l}{dq}. \tag{190}$$

I estimate the parameters γ_{1l} and γ_{2l} using the observed capacity range $[0.5, 21]$ and γ_{1u} using the observed range $[105, 400]$. I follow the same estimation procedure as in Section 2. I calibrate γ_{0l} so that total capacity equals the observed $Q^T = 2.7GWp$. The parameters γ_{0u} and q_{lb} are determined by the smoothness of $f_l(.)$. Table 19 summarises the parameter values.

E.2 A discussion and extension of Objective (12)

Objective (12) implicitly assumes that the government does not want to distribute information rents to adopters of solar panels, which is justifiable by the fact that, typically, only households in the upper part of the income distribution own roofs and can adopt solar panels. Moreover, the German government already uses a nonlinear subsidy scheme, showing that it is indeed trying to curtail these rents. Section E.2.1 generalises the objective to include general redistributive preferences, a preference for aggregate solar capacity, and possibly optimal taxes on other sources of income. Note that a Pigouvian subsidy, where the optimal marginal subsidy is equal to the marginal environmental benefit, is not the optimal solution to Objective (12). Section E.2.2 discusses when a Pigouvian subsidy is optimal in the general objective of Section E.2.1. Generally, it is not the case (see Kaplow, 1996, Cremer, Gahvari and Ladoux, 1998, Kaplow, 2008, and Kaplow, 2012). Section E.2.3 shows under which conditions the simple Objective (12) in Section 4 follows from the general objective in Section E.2.1. 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.

Moreover, this paper focuses on the evaluation and optimisation of nonlinearities in the subsidy schedule. Therefore, it uses the static Objective (12) and the long-run elasticities in Table 1 to calibrate the model. For an analysis of the optimal time path of subsidies, see Langer and Lemoine (2022). Additionally, this paper focuses on second-degree price discrimination. The analysis of third-degree price discrimination by considering, for example, adopters' location or electricity consumption is beyond the scope of this paper.

E.2.1 The general welfare function and the optimal subsidy

Assume the utility of an adopter is equal to

$$u\left(\left(S(q)-c_v(q,q_l)-c_f\right)\mathbb{1}\left(S(q)-c_v(q,q_l)\geq c_f\right)+y-T(y)-c_l(y,a)\right), \tag{191}$$

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

Consider the following objective function of the government:

$$\max_{S(.)} \int G\Big(\left(S(q)-c_v(q,q_l)-c_f\right)\mathbb{1}\left(S(q)-c_v(q,q_l)\geq c_f\right)+ \\ +y-T(y)-c_l(y,a)\Big)f_\theta(\theta)d\theta+V(Q), \tag{192}$$

such that

$$\int q\;\mathbb{1}\left(S(q)-c_v(q,q_l)\geq c_f\right)\;f_\theta(\theta)d\theta=Q, \tag{193}$$

and

$$\int T(y)-S(q)\mathbb{1}\left(S(q)-c_v(q,q_l)\geq c_f\right)f_\theta(\theta)d\theta-E=0. \tag{194}$$

The variable Q denotes the aggregate capacity; $V(.)$ is the government's value of aggregate capacity. The function $G(.)$ weights agents' utilities and represents the redistributive preferences of the government. It is increasing and concave. Its argument is the same as the argument of the agents' utility function. In the special case where $G(.)=u(.)$, the government is utilitarian. Equation (194) is the government's budget constraint. The variable E denotes other government spending.

Objective (192) assumes that the government sets the subsidy $S(q)$ independently of income y. This assumption is not without loss of generality. Generally, a subsidy $S(q,y)$ that

depends on adopted capacity and income may achieve higher social welfare than a subsidy that only depends on capacity. However, I follow this approach for three reasons. First, I do not observe the income of adopters. Joint information about adoption decisions q and income y is necessary to solve for the optimal joint subsidy $S(q, y)$. Due to this data limitation, I will focus on the optimal separable problem, where subsidy payments are only a function of capacity q. Second, the observed subsidy is independent of income. Arguably, a subsidy that depends on income is complicated to implement. Therefore, the government chose a subsidy that only depends on capacity q. Third, Problem (192) is a multidimensional screening problem. The type parameter determining the choice of capacity q and income y is two-dimensional. Theoretically, these problems are challenging to solve because local incentive-compatibility constraints are generally insufficient to determine the optimal schedule (see Rochet and Chone, 1998 for a detailed discussion). Treatment of the multidimensional screening problem is an interesting direction for future research beyond the scope of this paper. 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.

The 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

$$\pi_v(q_l) = S(q(q_l)) - c_v(q(q_l), q_l). \tag{195}$$

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

$$\pi_v'(q_l) = -\frac{\partial c_v(q(q_l), q_l)}{\partial q_l}, \tag{196}$$

and $q(.)$ non-decreasing. As standard in the literature, I neglect the monotonicity constraint on $q(.)$. It can be verified ex-post. Equation (196) defines a function $q(\pi_v')$. Therefore, the government's problem reduces to choosing a function $\pi_v(.)$:

$$\max_{\pi_v(.)} \int G\left((\pi_v - c_f)\,\mathbb{1}\left(\pi_v \geq c_f\right) + y - T(y) - c_l(y, a)\right) f_\theta(\theta) d\theta + V(Q), \tag{197}$$

s.t.

$$\int q\;\mathbb{1}\left(\pi_v \geq c_f\right)\; f_\theta(\theta) d\theta = Q, \tag{198}$$

and

$$\int T(y) - (\pi_v + c_v(q, q_l))\mathbb{1}\left(\pi_v \geq c_f\right) f_\theta(\theta) d\theta - R = 0. \tag{199}$$

For better readability, I suppress the arguments of the functions $\pi_v(.)$, $\pi_v'(.)$, and $q(.)$. I solve Problem (197) using calculus of variation.

The general optimality condition:
It follows that in the optimum

$$\int_{q_l}^{q_l^{max}} \Bigg[\int_{-\infty}^{\pi_v} \left(\int_0^{\infty} \frac{G'(\pi_v - c_f + y - T(y) - c_l(y,a))}{\lambda} f_a(a|c_f, \tilde{q}_l) da - 1 \right) f_f(c_f|\tilde{q}_l)\, dc_f + \\ + \left(\frac{V'(Q)}{\lambda} q - \pi_v - c_v(q(\pi_v'), q_l) \right) f_f(\pi_v|q_l) \Bigg] f_l(\tilde{q}_l) d\tilde{q}_l + \\ + \left(\frac{V'(Q)}{\lambda} - \frac{\partial c_v(q,q_l)}{\partial q} \right) \frac{dq}{d\pi_v'} F_f(\pi_v|q_l) f(q_l) = 0 \quad (200)$$

and

$$\left(\frac{V'(Q)}{\lambda} - \frac{\partial c_v(q,q_l)}{\partial q} \right) \frac{dq}{d\pi_v'} F_f(\pi_v|q_l) f_l(q_l) = 0 \text{ for } q_l = q_l^{min} \text{ and } q_l = q_l^{max} \,. \quad (201)$$

The variable λ is the marginal cost of public funds. Equation (200) is a second-order differential equation in the function $\pi_v(.)$ with two boundary conditions (201). The optimal rent $\pi_v(.)$ is the solution to this system.

E.2.2 The Pigouvian subsidy (i.e., the Samuelson rule)

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

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

$$\int_{-\infty}^{\pi_v} \left(\int_0^{\infty} \frac{G'(\pi_v - c_f + y - T(y) - c_l(y,a))}{\lambda} f_a(a|c_f, \tilde{q}_l) da - 1 \right) f_f(c_f|\tilde{q}_l)\, dc_f = 0, \quad (202)$$

for all capacity types q_l. To see this result, guess and verify the solution in Equation (200), using the first order condition of adopters $\frac{\partial c_v(q,q_l)}{\partial q} = \frac{V'(Q)}{\lambda}$. However, in general, Condition (202) does not hold. Particularly, the condition depends on the marginal social weight of adopters $G'(.)$ relative to the marginal cost of public funds λ. If the possibility to adopt solar panels is positively correlated with ability a, and the tax $T(y)$ does not fully equalise marginal social welfare weights, then Condition (202) does not hold. The term to the left of the condition is smaller than zero in this case. Importantly, in general, even an optimal income tax does not equalise marginal social welfare weights. Consider the optimality condition for the optimal

income tax. Following Saez (2001), I solve for the optimal tax using the variational approach:

$$\int_{y}^{\infty} \left(1 - \int \frac{G'\left(\left(\pi_v - c_f\right)\mathbb{1}\left(\pi_v \geq c_f\right) + \tilde{y} - T(\tilde{y}) - c_l(\tilde{y}, a^{-1}(\tilde{y}))\right)}{\lambda} \times \right.$$
$$\left. \times f(c_f, q_l|\tilde{y}) dc_f dq_l \right) f_y(\tilde{y}) d\tilde{y} + T'(y)\frac{dy}{dT'}\, f(y) = 0 \tag{203}$$

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

E.2.3 The relation to the simple Objective (12)

This section shows under which conditions the simple Objective (12) follows from the general Objective (192). 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 (202) 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 (192) to Problem (12). For example, it is the case if the redistributive preferences are Rawlsian, and households with the lowest incomes cannot adopt solar panels.

E.3 Counterfactual experiments

E.3.1 The optimal linear subsidy

A first experiment solves for the optimal linear subsidy. The exercise solves for a linear subsidy rate ρ_l that incentivises the adoption of the same aggregate capacity as the observed kinked subsidy.[49] By the first order condition of the agents' problem, the choice q of type q_l under subsidy ρ_l is $q(q_l, \rho_l) = q_l \rho_l^{\epsilon}$. Denote by $c_v(q, q_l)$ the variable part of the cost function $c_v(q, q_l) = c(q, q_l, c_f) - c_f$. Denote the variable profit of type q_l under subsidy rate ρ_l as $\pi_v(q_l, \rho_l) = \rho_l q(q_l, \rho_l) - c_v(q(q_l, \rho_l), q_l)$. Given the estimate of $f_l(q_l)$, the unconditional type distribution $f_u(q_l)$ is

$$f_u(q_l) = \frac{f_l(q_l)}{F_f(q_l - c_v(q_l, q_l))}. \tag{204}$$

It follows that ρ_l is the solution to

$$\int q(q_l, \rho_l) F_f(\pi_v(q_l, \rho_l)) f_u(q_l) dq_l = Q^T, \tag{205}$$

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

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

E.3.2 The optimal nonlinear subsidies

The second counterfactual experiment solves for the optimal nonlinear policy using mechanism design. The analysis follows the screening problem in Rochet and Stole (2002). Rewrite the government's objective (12) as a Lagrangian and a mechanism design problem. The government maximises

$$\max_{\psi, q(.), \pi_v(.)} \int \left(\psi q(q_l) - \pi_v(q_l) - c_v(q(q_l), q_l)\right) F_f(\pi_v(q_l)) f_u(q_l) dq_l - \psi Q^T, \tag{206}$$

such that for all q_l

$$\pi_v'(q_l) = -\frac{\partial c_v(q(q_l), q_l)}{\partial q_l} \text{ and } q(.) \text{ is not decreasing.} \tag{207}$$

Condition (207) is the incentive-compatibility constraint. The variable ψ denotes the Lagrange multiplier. Substitute the subsidy paid to type q_l using the definition: $S(q(q_l)) = \pi_v(q_l) + c_v(q(q_l), q_l)$. Problem (206) is equivalent to Problem (12). The government chooses functions $q(q_l)$ and $\pi_v(q_l)$ instead of a subsidy $S(q)$. The interpretation is as follows. Imagine the government asks an agent to reveal her type q_l. The agent reports the type; the government asks the agent to produce $q(q_l)$ and pays variable profit $\pi_v(q_l)$ as compensation. An incentive-compatible mechanism is two functions $q(.), \pi_v(.)$, under which each agent is incentivised to report her type truthfully. Therefore, the government's objective is to find two such functions which maximise its objective. The incentive-compatibility constraint (207) follows from the standard revealed preference argument in mechanism design. As standard in the literature, neglect the monotonicity constraint on $q(.)$ and verify it ex-post. Define

$$L(\pi_v, \pi_v', q_l) = \left(q_l \psi(\pi_v'(1+\epsilon))^{\frac{\epsilon}{1+\epsilon}} - \pi_v - q_l \epsilon \pi_v'\right) F_f(\pi_v) f_u(q_l), \tag{208}$$

which is the integrand of Problem (206). Use Equation (207) to substitute for the function $q(.)$. The problem simplifies to finding an optimal function $\pi_v(.)$. Suppress the arguments of the functions $\pi_v(.), \pi_v'(.)$ for better readability. By calculus of variation, it follows that the optimal function π_v satisfies

$$\frac{\partial L(\pi_v, \pi_v', q_l)}{\partial \pi_v} = \frac{d}{dq_l} \frac{\partial L(\pi_v, \pi_v', q_l)}{\partial \pi_v'} \quad \text{for all } q_l, \tag{209}$$

and

$$\frac{\partial L(\pi_v, \pi_v', q_l)}{\partial \pi_v'} = 0 \text{ for } q_l = q_l^{min} \text{ and } q_l = q_l^{max}. \tag{210}$$

The values q_l^{min} and q_l^{max} denote the boundaries of the type distribution $f_l(.)$. For each ψ, the Equations (209) and (210) define a nonlinear second-order differential equation with boundary values. Fix ψ, solve the differential equation numerically, and evaluate the capacity constraint $Q = Q^T$. Iterate over ψ until the constraint holds. Using the first order condition and the solution $\pi_v'(q_l)$, solve for the optimal nonlinear marginal subsidy $S'(q(q_l))$ in Figure 9. Using the definition of the variable profit, it follows that the total public costs are

$$\int \left(\pi_v(q_l) + c_v(q(q_l), q_l)\right) F_f(\pi_v(q_l)) f_u(q_l) dq_l. \tag{211}$$

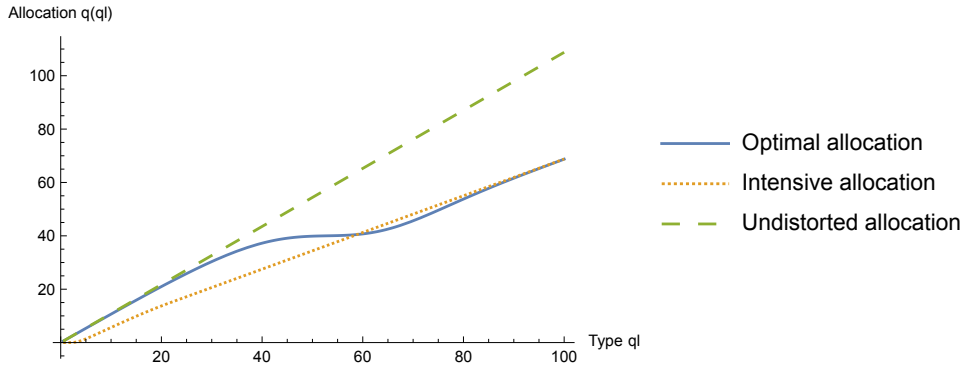
The optimal subsidy is 0.45 percent less costly than the linear benchmark.

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

E.3.3 A comparison to Rochet and Stole (2002)

Figure 24 compares the optimal allocation to the bounds derived in Proposition 4 in Rochet and Stole (2002). They show that when the type distribution $f_l(.)$ is uniform, the optimal allocation $q(q_l)$ is bounded from above by the undistorted allocation $q^{fb}(q_l) = \psi^{\epsilon} q_l$ and from below by the optimal intensive allocation $q^{it}(q_l)$ derived in the third counterfactual experiment. Note that Rochet and Stole (2002) also assume the distribution of fixed costs is log-concave, which is the case in my application. Contrary to the third counterfactual experiment and consistent with the result in Rochet and Stole (2002), I do not adjust the Lagrange multiplier to fulfil the capacity constraint, but I use the multiplier from the second counterfactual experiment. The figure shows that there exist regions where the optimal allocation lies outside the bounds. It shows that the result in Rochet and Stole (2002) is not robust to a more general form of the type distribution $f_l(.)$. E.g., the type distribution in my application is log-normal at the bottom and Pareto at the top.

Figure 24: The optimal allocation $q(q_l)$ and the bounds derived by Rochet and Stole (2002).



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

E.3.4 A comparison to Germeshausen (2018)

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

Germeshausen does not estimate intensive and participation margin elasticities. To compare the results in the two studies, I use my estimates to calculate the implied treatment effect of

introducing a new kink of 5% at 10 kWp in my data, i.e., I calculate:

$$\frac{\int_{q_1^K}^{q^{max}\rho_1^{-\epsilon}} q(q_l)F_f\left(S_k(q(q_l)) - c_v(q(q_l), q_l)\right) f_u(q_l)dq_l}{\int_{q_1^K}^{q^{max}} q_l f_l(q_l)dq_l}, \tag{212}$$

where $q(.)$ and $S_k(.)$ are defined in Section E.1.2, $f_u(.)$ is defined in Section E.3.1, q_1^K=10 kWp, $\rho_1 = 0.95$, $\rho_2 = 1$, and q^{max}= 20 kWp.

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

E.3.5 A summary of the results of the policy evaluation

Tables 20 and 21 summarise the results of the counterfactual exercises.

Table 20: The cost of the counterfactual scenarios compared to the optimal linear subsidy.

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

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

Table 21: The subsidy payment to the lowest type (0.5 kWp) under the counterfactual scenarios.

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

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

F The cited results from Krantz and Parks (2002) and Chen (2007)

F.1 Krantz and Parks (2002)

Proposition 1.2.12 Let f be an infinitely differentiable function for some interval I. The function is in fact real analytic on I if and only if, for each $x \in I$, there are an open interval J, with $x \in J \subseteq I$, and constant $M > 0$ and $R > 0$ such that the derivatives of f satisfy

$$|f^{(p)}(x)| \leq M \frac{p!}{R^p}, \ \forall x \in J. \tag{213}$$

Lemma 1.1.8 For the power series

$$\sum_{p=0}^{\infty} \gamma_p (x - x^0)^p \tag{214}$$

define A and ρ by $A = \lim_{p\to\infty} \sup |\gamma_p|^{1/p}$ and $\rho = 1/A$. Then ρ is the radius of convergence of the power series about x^0.

Corollary 1.2.4 Let (214) be a power series with open interval of convergence I. Let $f(x)$ be the function defined by the series on I. Then f is real analytic at every point of I.

Corollary 1.2.5 If f and g are real analytic functions on I and if there is a point in I where all their derivatives are equal, then $f(x) = g(x)$ for all $x \in I$.

Corollary 1.1.10 The power series (214) has radius of convergence ρ if and only if, for each $0 < R < \rho$, there exists a constant $0 < M_R$ such that

$$\gamma_p \leq \frac{M_R}{R^p}. \tag{215}$$

Corollary 1.2.6 If f and g are real analytic functions on I and there is an open set $J \subseteq I$ such that $f(x) = g(x)$ for all $x \in J$, then $f(x) = g(x)$ for all $x \in I$.

Corollary 1.1.16 The power series representation of a real analytic function at a point x^0 is unique.

F.2 Chen (2007)

Condition 3.1".
i) Θ is compact under $d(.,.)$, and $Q(\theta)$ is upper semicontinuous on Θ under $d(.,.)$;
ii) $Q(\theta)$ is uniquely maximised at θ^o in Θ, $Q(\theta^o) > -\infty$. **Condition 3.2**. $\Theta_P \subseteq \Theta_{P+1} \subseteq \Theta$ for all $P \geq 1$; and there exists a sequence $\pi_P \theta^o \in \Theta_P$ such that $d(\pi_P \theta^o, \theta^o) \to 0$ as $P \to \infty$.
Condition 3.3 i). For each $P \geq 1$, $Q(\theta)$ is upper semicontinuous on Θ_P under the metric $d(.,.)$.
Condition 3.4. The sieve spaces, Θ_P, are compact under $d(.,.)$.
Condition 3.5 i). For all $P \geq 1$, $\mathrm{p}\lim_{n\to\infty} \sup_{\theta\in\Theta_P} |\hat{Q}_n(\theta) - Q(\theta)| = 0$.
Footnote 29 Condition 3.3 i) and 3.4 and the pointwise convergence over Θ_P imply Condition 3.5 i).
Statement Page 5591 We obtain $d(\hat{\theta}_n, \theta^o) = o_P(1)$ under Condition 3.1", 3.2, 3.4, and 3.5 i).