Kinks Know More: Estimating Intensive and Participation Margin Responses Using Nonlinear Budget Sets

Stefan Pollinger

SciencesPo
DEPARTMENT OF ECONOMICS

Discussion Paper

KINKS KNOW MORE: ESTIMATING INTENSIVE AND PARTICIPATION MARGIN RESPONSES USING NONLINEAR BUDGET SETS

Stefan Pollinger

SCIENCES PO ECONOMICS DISCUSSION PAPER

No. 2025-09

Kinks Know More: Estimating Intensive and Participation Margin Responses Using Nonlinear Budget Sets*

Stefan Pollinger †

April 15, 2025 (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 kinks in 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. A second application to notches in the preferential income tax for foreigners in Denmark demonstrates the wider applicability of the methodology.

Keywords: Kinks, Notches, Bunching, Extensive Margin.

1 Introduction

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.

*This paper was formerly circulated under the title "Kinks Know More: Policy Evaluation Beyond Bunching with an Application to Solar Subsidies."

†Stefan Pollinger, Assistant Professor, Department of Economics, Sciences Po, 75007 Paris, France; email address: stefan.pollinger@sciencespo.fr; website: www.stefanpollinger.com.

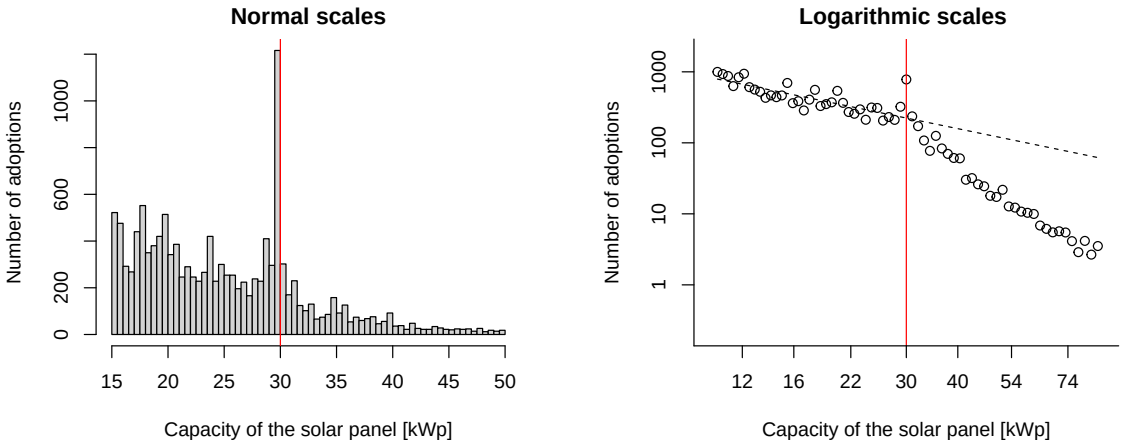
[1]The participation margin is also called the extensive margin.

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

This paper proposes an estimator for agents' responses at both margins. The estimator leverages the effect of kinks or notches in an incentive scheme to identify the two responses jointly. In the main application, the paper evaluates the kinks in the German subsidy for rooftop solar panels. The programme 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

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



subsidy rate for adopting a unit of solar capacity decreased discontinuously at 30 kWp (see Figure 2 in Section 2 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 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

[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 the efficiency of the adopted solar panel. The number of observations in a bin is normalised by the bin-size, but not the total number of observations.

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 2.1). Importantly, these observable effects are absent in years when the German subsidy was linear (see Figure 7 in Section 4.1). 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 paper uses the model with isoelastic intensive margin responses customary in the bunching literature (e.g., see Saez, 2010, Blomquist et al., 2021, Bertanha, McCallum and Seegert, 2023), and augments it with isoelastic participation margin responses. As in the bunching literature, 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 estimator in five dimensions to alleviate these concerns. First, the paper shows that both margins are point-identified under the assumption that the counterfactual distribution is infinitely differentiable and the derivatives are sufficiently regular. This smoothness assumption is weaker than the parametric assumption in the classic bunching literature. 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, the 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 infinite differentiability 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. Fourth, the paper shows how to relax the infinite differentiability assumption when additional observations that do not face a kink are available. In this case, standard regularity conditions on the counterfactual, such as in Newey (1997), suffice for identifying both margins. The estimates using this alternative estimation procedure are not statistically different from the main estimates. Fifth, building on the partial identification results for the intensive margin in Blomquist et al. (2021), Bertanha, McCallum and Seegert (2023), and Goff (2022), the paper derives conditions for the partial identification of both response margins. To this end, it relaxes the infinite differentiability assumption, using shape restrictions on the counterfactual instead, as in the aforementioned papers. The bounds are tight around the point-estimates of the main estimation procedure.

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., discontinuities in incentive schemes, is straightforward. These low informational and identifying requirements enhance its applicability for estimating behavioural responses to taxes, subsidies, transfers, regulations, and product prices.

Applying the novel estimator to multiple kinks in the German subsidy 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. In a companion paper, Pollinger (2025) uses the estimates to conduct further counterfactual exercises. A first exercise in Pollinger (2025) shows that the most cost-effective nonlinear scheme would more than triple the gain mentioned above; however, due to the sizeable participation margin responses, cost savings remain modest overall. A second exercise in Pollinger (2025) shows that ignoring the participation margin when optimising the subsidy would increase costs substantially. These results 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 subsidies are not a panacea for cost-effectiveness and can even trigger detrimental effects. Estimating both response margins is essential for their optimal design.

A second application to the preferential income tax for foreigners in Denmark demonstrates the applicability of the methodology to taxes instead of subsidies and notches instead of kinks. Furthermore, the application serves as a validation exercise: the estimates of the participation margin are statistically equal to the estimates in Kleven et al. (2013), who exploit the introduction of the preferential scheme in a difference-in-differences design. A policy reform, parallel trends, and a time dimension of the data are necessary for applying difference-in-differences. While difference-in-differences is a powerful and well-established method, these features are lacking in some applications. The methodology proposed in this paper solely relies on the effect of a kink or notch on the cross-section of agents' choices. Therefore, it is complementary to difference-in-differences.

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 the classic bunching design does not consider 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 applications, data often lacks a panel dimension.[5] 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. Blomquist et al. (2021), Bertanha,

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

McCallum and Seegert (2023), and Goff (2022), propose conditions for the partial identification of intensive margin responses in bunching designs. Building on these results, my paper proposes conditions for the partial identification of both margins. Bertanha, McCallum and Seegert (2023) show that point-identification of the intensive margin in the bunching design can also be achieved using covariates.[6] My applications lack such variables.

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 is inapplicable because it suffers from endogeneity. Ignoring this effect biases the results by 5 percent in my application. The simultaneous estimation in my paper addresses these biases of the bunching and regression kink designs. Gerard, Rokkanen and Rothe (2020), Bachas and Soto (2021), and Caetano, Caetano and Nielsen (2024) propose methods to correct for endogeneity due to intensive responses in the regression discontinuity design and 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.

From an applied perspective, the paper contributes to the literature evaluating subsidies for solar panels. This literature has not yet focused on evaluating and optimising the cost-effectiveness of nonlinear solar subsidies. One strand of the solar literature uses structural models to study the dynamics of the adoption decision[7] Another strand of the solar literature uses reduced-form methods.[8] For example, Germeshausen (2018) uses a difference-in-differences 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. My estimates imply a treatment effect equal to the effect estimated by Germeshausen (2018), which validates the methodology in both papers. 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. These papers use a difference-in-differences design, controlling for or using bunching.[9] Difference-in-differences relies on a parallel trend assumption, which is unnecessary in my approach and violated in my main application.

The rest of the paper proceeds as follows. Section 2 outlines the model and discusses identification. Section 3 presents the estimator. Section 4 presents the main empirical application

[6]Building on the identification results in this paper and Bertanha, McCallum and Seegert (2023), the recent contribution by Song (2024) explores point and partial identification of the intensive margin combining an analyticity assumption with covariates.

[7]For example, De Groote and Verboven, 2019; Feger, Pavanini and Radulescu, 2022; Langer and Lemoine, 2022; Gerarden, 2022.

[8]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.

[9]In the same vein, Myhre (2021) combines bunching with a regression discontinuity design in the time dimension.

of the method to the German solar subsidy. Section 5 presents the application to the Danish income tax. Section 6 discusses further applications. Section 7 concludes.

2 The model

Consider the standard model used in the bunching literature (e.g., see Saez (2010), Blomquist et al. (2021), and Bertanha, McCallum and Seegert, 2023). Agents choose a quantity $q \in \mathbb{R}^+$ for which they receive a payment $S(q)$. They solve the maximisation problem:

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

The variable cost $c_v(q, \nu)$ is twice continuously differentiable in all arguments and increasing and convex in quantity q. The marginal cost $\frac{\partial c_v(q,\nu)}{\partial q}$ is continuous and increasing in $\nu \in \mathbb{R}^+$, which denotes the variable cost type of an agent. The variable profit is denoted by π_v. 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 $\pi_v \geq c_f$, where c_f denotes an agent's fixed cost of participation. The fixed and variable costs are unobservable and may also contain unobservable benefits. In general, the fixed cost will be heterogeneous among agents. However, the model also nests the case when they are homogeneous and equal to zero, as in the standard bunching model. In an income tax context, q is gross income, $S(q)$ is the net of tax income, $c_v(q, \nu)$ is the disutility of generating gross income q, and c_f is the fixed cost of participating in the labour market. In the main application of this paper, agents are the adopters of solar panels, q is the capacity of the solar panel, $S(q)$ is the subsidy payment, $c_v(q, \nu)$ is the variable cost of adopting capacity q, and c_f is the fixed cost of adopting a solar system. For ease of exposition, consider the example of solar subsidies from now on.

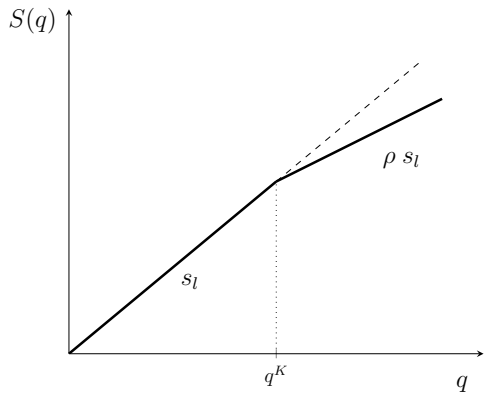
The subsidy $S(.)$ in (1) can take two forms: the observed kinked subsidy $S_k(.)$ and the counterfactual linear subsidy $S_l(.)$. Comparing adopters' reactions under the kinked subsidy to reactions under the linear subsidy is useful for deriving the estimator. However, this comparison is a thought experiment. 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

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.

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 illustrates both subsidies.

Denote by $q(s, \nu)$ the capacity-choice q of adopter ν under a linear subsidy with rate s. Note that it is continuous in both arguments, increasing in s and decreasing in ν. Consider an interval $(\underline{q}, \overline{q})$ around the kink point q^K.

Assumption 1 (Intensive margin)**.** *Close to the kink point, intensive margin responses are isoelastic and bounded:*

$$c_v(q, \nu) = \nu q^{1+\frac{1}{\epsilon}}, \tag{5}$$

where ϵ denotes the intensive margin elasticity. Formally, close to the kink point means the assumption applies for all agents ν such that $q(s_l, \nu) > \underline{q}$ and $q(\rho s_l, \nu) < \overline{q}$, and for all quantities $q \in [q(\rho s_l, \nu), q(s_l, \nu)]$.[10]

Assumption 1 is standard in the bunching literature (e.g., see Saez (2010), Blomquist et al. (2021), and Bertanha, McCallum and Seegert, 2023).[11] Intuitively, the assumption implies that, for adopters close to the kink point, reducing the marginal subsidy rate s_l by one precent reduces their capacity choice by ϵ percent, i.e., $\frac{\partial \ln q(s_l,\nu)}{\partial \ln s_l} = \epsilon$. The assumption restricts the function

[10] These are the agents ν that locate in the interval $(\underline{q}, \overline{q})$ under the kinked subsidy, and this is the range of their responses $q(s, \nu)$ to marginal subsidy rates s between the marginal kinked schedule $S_k'(.)$ and the marginal linear schedule $S_l'(.)$. See Appendix A.2 for more details.

[11] Exceptions are Goff (2022), who derives partial identification results when responses are heterogeneous, and Kleven (2016), who allows for heterogeneous responses when the counterfactual choice distribution is constant for agents who bunch.

$c_v(.,.)$ on the part of its domain that is close to the kink point under the kinked subsidy, while not restricting $c_v(.,.)$ on the rest of its domain. Hence, Assumption 1 is a local parametric approximation to the nonparametric variable cost function. Next, I impose a corresponding assumption on the participation margin. To this end, denote the choice of an adopter under the counterfactual subsidy $S_l(.)$ by q_l, where l stands for linear: $q_l = q(s_l, \nu)$. Similarly, denote the total cost of an adopter under the counterfactual subsidy $S_l(.)$ by c_t, where t stands for total: $c_t = c_v(q_l, \nu) + c_f$. Note that there is a one-to-one mapping from the variable and fixed cost type (ν, c_f) to the choice and total cost under the counterfactual subsidy (q_l, c_t).[12] Therefore, without loss of generality, (q_l, c_t) can be used as an alternative type parameter and the total cost function can be written as

$$c(q, q_l, c_t) = \underbrace{\frac{s_l}{q_l{}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}}}_{\text{variable cost}} + \underbrace{c_t - q_l s_l \frac{\epsilon}{1+\epsilon}}_{\text{fixed cost}}. \tag{6}$$

Using (q_l, c_t) as a type parameter is necessary for extending the isoelasticity assumption to the participation margin. Moreover, it has the advantage that the type has direct economic meaning: it is equal to the choice and total cost under the counterfactual subsidy. Denote by $F_{t|q_l}(.|q_l)$ the CDF of c_t conditional on q_l. Define the interval of counterfactual quantities $(\underline{q_l}, \overline{q_l}) = (\underline{q}, \overline{q}\rho^{-\epsilon})$.

Assumption 2 (Participation margin)**.** *Close to the kink point, participation margin responses are isoelastic and bounded:*

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{g(q_l)}\right)^{\eta}, \tag{7}$$

where η denotes the participation margin elasticity. Formally, close to the kink point means the assumption applies for all $q_l \in (\underline{q_l}, \overline{q_l})$ and all $c_t \in [S_k(q_l), S_l(q_l)]$. The function $g(q_l) > S_l(q_l)$ denotes a normalisation term assuring $F_{t|q_l}(c_t|q_l) \leq 1$.

Intuitively, the assumption implies that, close to the kink point, reducing the total subsidy payment S_l by one percent reduces participation by η percent, i.e., $\frac{\partial \ln F_{t|q_l}(S_l|q)}{\partial \ln S_l} = \eta$. Note that the assumption on the participation margin, which is the novel margin considered in this paper, exactly mirrors the standard assumption on the intensive margin used in the literature. Like Assumption 1, Assumption 2 restricts the function $F_{t|q_l}(.|.)$ on the part of its domain that is close to the kink point under the kinked subsidy, but not on the rest of its domain. Hence, like Assumption 1, Assumption 2 is a local parametric approximation to the nonparametric distribution of total costs. Note that the normalisation term $g(q_l)$ in Assumption 2 assures that $F_{t|q_l}$ is a CDF. Besides $g(q_l) > S_l(q_l)$, the assumption does not further restrict $g(q_l)$. In particular, it may depend on q_l. Therefore, Assumption 2 does not rule out a correlation

[12] Using the definition of $c_v(.,.)$ and the agents' FOC renders $\nu = \frac{s_l}{q_l{}^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon}$ and $c_f = c_t - q_l s_l \frac{\epsilon}{1+\epsilon}$. 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 does not poses a problem.

between the variable cost type q_l and the total cost type c_t, and, hence, between ν and c_f. Appendix D.5.10 shows that the Assumption 2 can be relaxed to allow η to depend on q_l. The appendix also shows that this generalisation does not change the results in the main application of this paper.

Denote by $f_k(.)$ the observable distribution of adopters' choices under the kinked subsidy and by $f_l(.)$ the counterfactual distribution of adopters' choices under the linear counterfactual subsidy. Both functions describe the distribution of quantities conditional on participation. Technically, $f_l(.)$ and $f_k(.)$ are measures; for example, for any interval $[q_1, q_2]$, $\int_{q_1}^{q_2} f_l(q)dq$ is the mass of adopters in the interval under subsidy $S_l(.)$. Intuitively, $f_l(.)$ and $f_k(.)$ are densities that may not integrate to one.[13] Typically, due to participation responses, the mass under $f_l(.)$ is larger than the mass under $f_k(.)$. Note that, conveniently, $f_l(.)$ is also the distribution of the parameter q_l conditional on participation.[14]

2.1 The graphical intuition behind identification

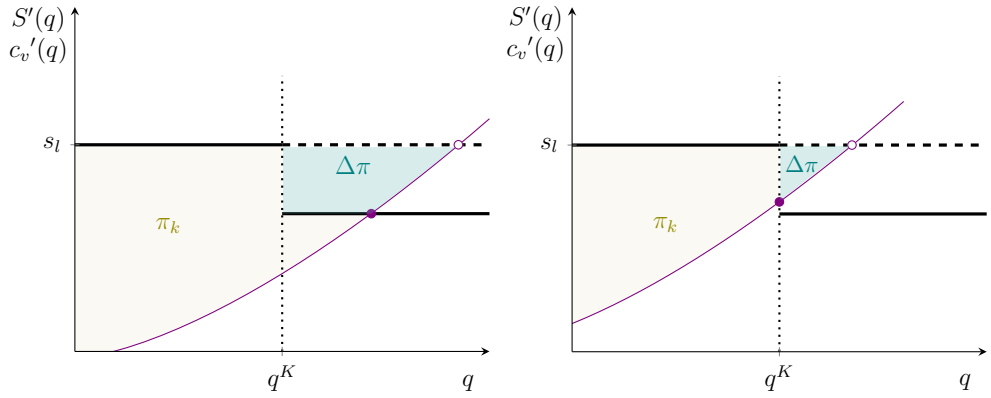
This section gives graphical intuition on how the distribution of adoptions under the kinked scheme S_k depends on the intensive and the participation margin. It explains the effect on the distribution using a hypothetical change in the subsidy schedule from S_l to S_k. Depending on their production 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 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

[13] Without loss of generality, one of the two may be normalised to one.

[14] The unconditional distribution is defined in Appendix A.2.

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 $c_v{}'$ 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.

marginal subsidy precisely between the two marginal rates.[15] 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.

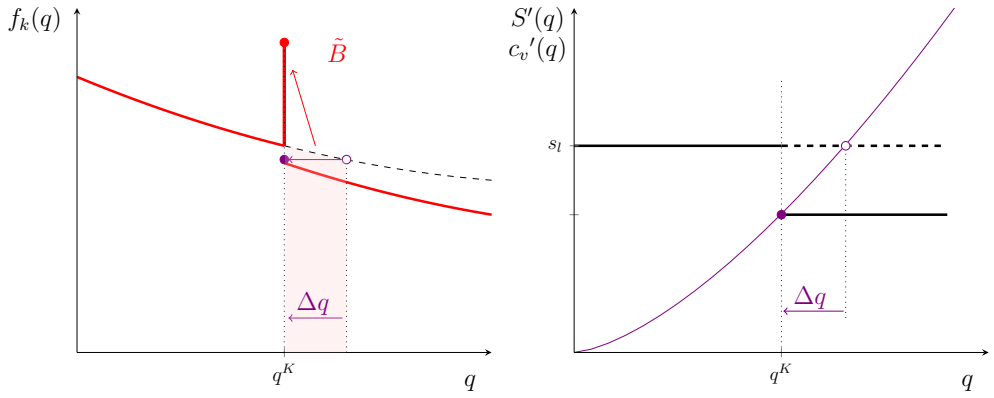
The distinct effect of the kinked subsidy on these three groups of adopters affects the distribution of adoptions. First, to better understand the effect on the distribution, consider the case where fixed costs are homogeneous and equal to zero. As a consequence, there are no participation responses. 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.[16] At the kink

[15] Note that these adopters are at a corner solution; their first order condition does not hold.

[16] 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.

point, there is a mass point $\tilde{B}$, 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 unaffected by the change in schemes. How is the measure of adoptions, as

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



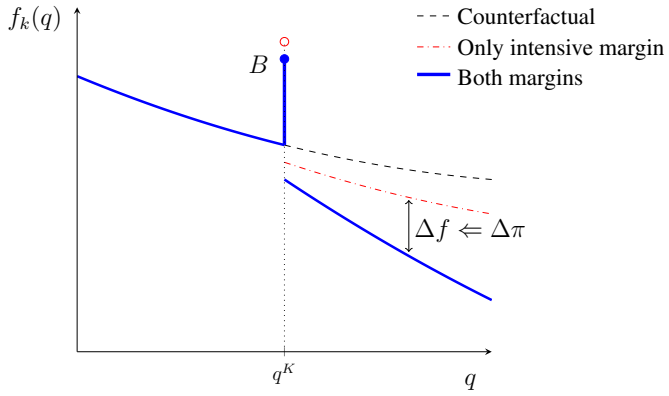
Note: The left panel shows the counterfactual measure f_l as a dashed black line and the observable measure f_k as a red line. At the kink point, there is a mass point $\tilde{B}$. The right panel shows the marginal cost curve of the marginal buncher. This agent reduces capacity by Δq. The dots in the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

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 literature calls this adopter the "marginal buncher" (see Saez, 2010). In response to the change in marginal subsidy, the adopter reduces production by Δq. The dots in the left panel of Figure 4 show the marginal buncher in the measure of adoptions. 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 can be used to identify the response Δq, which can be used to identify 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. The mass of agents below the kink point is unaffected by the subsidy change.[17] Above the kink point, adopters from the first group illustrated in Figure 3

[17]Note that participation responses trigger a decrease in overall mass. However, the measure $f_k(.)$ is not renormalised by the overall mass. Therefore, there is no effect on the measure below the kink point.

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. Adopters at the kink point also react at the participation margin. Therefore, there is less bunching.

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 the participating mass Δf. As depicted by the red dash-dotted line in Figure 5, the slope of the measure above the kink point is already altered by the intensive margin responses. As depicted by the blue solid line, participation margin responses are responsible for an additional slope change in the measure. 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 $\tilde{B}$ decreases from the empty red dot to the full blue dot. Appendix D.6 shows that this effect biases the classic bunching estimates. For further intuition on the effect of participation responses, Appendix A.1 discusses the case when there is only a participation margin. It also discusses how, in that case, the regression kink design can be applied.

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. As illustrated above, the slope change in the observable measure depends on the intensive *and* participation margin responses. The same is true for the observable bunching mass. Therefore, on its own, neither of the two observables is sufficient for identifying the two responses. However, the two observables distinctly depend on the magnitudes of both margins. Therefore, two observable moments, bunching and slope change, are sufficient to *jointly* identify the unknown magnitudes of the two margins. The following section carries out this exercise.

2.2 Identification

Proposition 1 (The observable measure)**.** *Under Assumptions 1 and 2, for all $q \in (\underline{q}, \overline{q})$, 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:

$$f_k(q) = f_l(q), \qquad \textit{for } q < q_L; \tag{8}$$

$$\tilde{B} = \int_{q^K}^{q^K \rho^{-\epsilon}} R_B(q_l, \epsilon)^\eta f_l(q_l) dq_l, \qquad \textit{for } q = q^K; \tag{9}$$

$$f_k(q) = R_2(q\ \rho^{-\epsilon}, \epsilon)^\eta f_l(q\ \rho^{-\epsilon})\rho^{-\epsilon}, \qquad \textit{for } q > q_H, \tag{10}$$

where $\tilde{B}$ is the mass of a mass point at q^K. The functions $R_B(., \epsilon)$ and $R_2(., \epsilon)$ denote the net subsidy payment to an adopter under the kinked scheme relative to the subsidy payment under the counterfactual scheme:

$$R_B(q_l, \epsilon) = \frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 - \left(\frac{q^K}{q_l}\right)^{\frac{1+\epsilon}{\epsilon}}\right); \tag{11}$$

$$R_2(q_l, \epsilon) = (1-\rho)\frac{q^K}{q_l} + \frac{\epsilon}{1+\epsilon}\left(1 + \frac{\rho^{\epsilon+1}}{\epsilon}\right). \tag{12}$$

The proof of Proposition 1 is in Appendix A.2. Below the kink point, the observable measure $f_k(.)$ depends only on the counterfactual measure $f_l(.)$.[18] At the kink point, there is an observable mass point $\tilde{B}$. It depends on all three unknowns. The measure above the kink point depends on all three unknowns as well. However, generically, all observables depend on the three unknowns distinctly, a property crucial for identification.

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 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.[19] Considering non-sharp bunching does not necessitate a significant change to the analysis above except that: (i) $(\underline{q}, \overline{q}) \supsetneq [q_L, q_H]$ (which also implies $(\underline{q_l}, \overline{q_l}) \supsetneq [q_L, q_H\rho^{-\epsilon}]$) in Assumption 1, 2, and Proposition 1; and (ii), the point mass $\tilde{B}$ in Proposition 1 becomes the mass in the bunching

[18] Note that $f_k(.)$ does neither depend on the unconditional distribution of q_l nor the normalisation term $g(.)$, because the latter cancels out in the derivation.

[19] 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).

interval $[q_L, q_H]$:

$$B = \frac{\int_{q_L}^{q^K} f_l(q_l)dq_l + \tilde{B} + \int_{q^K\rho^{-\epsilon}}^{q_H\rho^{-\epsilon}} R_2(q_l,\epsilon)^\eta f_l(q_l)dq_l}{q_H - q_L}, \tag{13}$$

and $f_k(q) = B$ for all $q \in [q_L, q_H]$. Note that B denotes the mass in the bunching interval, while $\tilde{B}$ denotes the sharp bunching mass. The mass in the bunching interval B is equal to the sharp bunching mass $\tilde{B}$ plus the mass of agents locating below and above the kink point in the bunching interval.

The pseudo-parameter $f_l(.)$ in Proposition 1 is infinite-dimensional. Equation (8) 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 (10) and (13) 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 the interval of counterfactual quantities $(\underline{q_l}, \overline{q_l})$ and a transformation of the counterfactual measure $f_l(.)$. For ease of exposition, consider a logarithmic transformation of $f_l(.)$ and an symmetric interval $\ln q^K - \ln \underline{q_l} = \ln \overline{q_l} - \ln q^K$.

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

$$\left| \frac{d^{(p)} \ln f_l(q_l)}{d \ln q_l^{(p)}} \right| \leq M \frac{p!}{(\ln \overline{q_l} - \ln q^K)^p} \quad \textit{for all } p \in \mathbb{N}, \tag{14}$$

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 $p!/(\ln \overline{q_l} - \ln q^K)^p$ 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 M, $\underline{q_l}$, and $\overline{q_l}$ are assumed to fulfil the properties above but may be unknown to the econometrician.

Lemma 1 (Smoothness)**.** *Assumption 3 implies that the logarithm of the counterfactual measure* $\ln f_l(.)$ *is real analytic in* $\ln q_l$ *on the interval* $(\underline{q_l}, \overline{q_l})$*. It has the convergent power series*

representation:

$$\ln f_l(q_l) = \sum_{p=0}^{\infty} \gamma_p \left(\ln \frac{q_l}{q^K} \right)^p \textit{ for all } q_l \in (\underline{q_l}, \overline{q_l}), \tag{15}$$

where $\gamma_p = \frac{d^{(p)} \ln f_l(q^K)}{d \ln q_l^{(p)}} / p!$ *and* $|\gamma_p| \leq \frac{M}{(\ln \overline{q_l} - \ln q^K)^p}$.

The proof is in Appendix A.3. While Assumption 3 is more intuitive, it implies the slightly weaker Condition (15). The weaker Condition (15) is sufficient for the results in the rest of the paper. Due to Lemma 1, 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 (15). Intuitively, due to Equations (8) and (15), observations below the bunching interval uniquely identify the infinitely dimensional pseudo-parameter γ, which, due to Lemma 1, identifies the nonparametric counterfactual distribution $f_l(.)$ for all $q_l \in (\underline{q_l}, \overline{q_l})$. See Proposition 2 below for the formal statement.

Assumption 3 or Condition (15) imply that $\ln f_l(.)$ has a convergent power series representation at each point in the interval $(\underline{q_l}, \overline{q_l})$. Such functions are called real analytic on the interval $(\underline{q_l}, \overline{q_l})$. See Gautier and Gaillac (2021) and Iaria and Wang (2024) for identification results using such functions in other applications. Note that an extensive and flexible class of functions is real analytic. 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.

Once $f_l(.)$ is identified, Equations (9) and (10) form a simultaneous, nonlinear system of equations in the two parameters (ϵ, η). 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 such nonlinear models since it is difficult to prove that the nonlinear equations have a unique intercept in the parameters. A common approach to address this issue is to impose as an additional rank condition that such a unique intercept exists (see Newey and McFadden, 1994); the condition can be verified in sample. See point (a) in Lemma 5 in Appendix A.5.1 for a formal definition of the condition and see Appendix D.5.9 for the verification in sample. Moreover, importantly, Lemma 5 in Appendix A.5.1 provides a long list of very general conditions on $f_l(.)$, each of which is sufficient for the existence of a unique intercept. Since Equations (9) and (10) provide infinitely many moments, it is intuitive that a unique intercept exists. Indeed, point (b) in Lemma 5 shows that generically it is the case. Moreover, the lemma shows that it is the case if $d \ln f_l(.)/d \ln q < -1$; Appendix D.5.9 shows this property holds in the application of this paper. Examples of other sufficient conditions in Lemma 5 are that $f_l(.)$ is real analytic on $[0, \overline{q_l})$ or that the series (15) is of finite order. In particular, the latter implies a unique intercept exists if $f_l(.)$ is Pareto or log-normal. Note that the necessity of these additional conditions is not directly related to $f_l(.)$

being parametric or nonparametric; in particular, the lemma shows that the condition is not necessary to identify the intensive margin when there is no participation margin. Appendix A.5.2 discusses the economic intuition behind the existence of a unique intercept.

Proposition 2 (Identification)**.** *Under Assumptions 1, 2, 3 and one of the conditions in Lemma 5, 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.4. Proposition 2 shows that there exists a population criterion with a minimum at the true pseudo-parameter (ϵ, η, γ). The population criterion is the square distance of the model to the true observable measure. See Equation (18) in Appendix A.4 for the definition. Observations below the bunching interval uniquely identify γ, which, due to Lemma 1, identifies the nonparametric counterfactual distribution. Observations in the bunching interval and observations above the bunching interval jointly identify the two response margins ϵ and η. Moreover, Lemma 1 implies that the population criterium is continuous in (ϵ, η, γ) and that the parameter space is compact (see Lemma 3 and 4 in Appendix A.4). Therefore, minimising the population criterion is a well-posed problem.[20] Appendix E.1 generalises the result in Proposition 2 to the case when there is a discontinuity (i.e., a notch) in the incentive scheme.

Proposition 2 has three important implications. First, the result allows deriving a point-estimator of the two response margins (see Section 3). Such point-estimates are necessary to conduct the counterfactual exercises in Appendix D.4 and Pollinger (2025). 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 3).

While Assumption 3 is weaker than the parametric functional form assumption in the classic bunching literature (e.g., Saez, 2010, Chetty et al., 2011), it is stronger than assumptions typically used in nonparametric estimations. Therefore, Appendix D.5.4 and D.5.6 further relax the assumption. Appendix D.5.4 shows how to achieve point-identification when additional observations, which do not face a kink, are available. In this case, standard regularity conditions on the counterfactual, such as in Newey (1997), suffice for identification. Building on the partial identification results for intensive margin responses in Blomquist et al. (2021), Bertanha, McCallum and Seegert (2023), and Goff (2022), Appendix D.5.6 derives partial identification results for both margins assuming the first derivative of the counterfactual is Lipschitz continuous. Both exercises show that the results in the main application of this paper are robust to relaxing Assumption 3.

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

2.3 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 kink. These variables may be randomly chosen by the policymaker or endogenous. 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 $\Delta\pi$ illustrated in Figure 3. Consequently, the change in behaviour Δf illustrated in Figure 5 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.[21] Moreover, the proposition does not rely on additional assumptions on ϵ and η. In particular, η may be small or even zero and it is identified as such. This is the case when, locally, the CDF of fixed costs is constant. 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.

Assumptions 3 and Lemma 1 use a logarithmic transformation of f_l and q.[22] 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 main 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 A.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 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 A.6 discusses why their non-identification result does not contradict Proposition 2. Moreover, Appendix A.7 illustrates Proposition 2 using simulations. In particular, it illustrates that, under the conditions in Lemma 1, both elasticities are identified even when bunching overlaps with a hump in the counterfactual distribution and when the bunching interval is large.

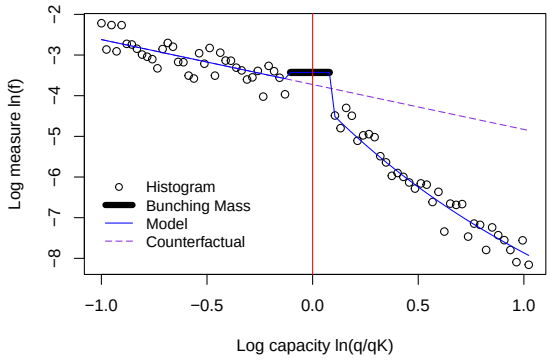
[21] The identification result in Proposition 2 does not rely on an assumption about the size of the kink. However, the size influences the properties of the estimator presented in Section 3. Large kinks have the advantage of creating stronger variation. Hence, everything else equal, the estimates have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

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

3 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 histogram normalised by the bin-size and the total number of observations; the black bar depicts the mass in the bunching interval.[23] Scales are logarithmic, and the kink point is normalised to zero. Intuitively, the estimation minimises the distance between the observed log-measure and the model in Proposition 1. The blue line depicts the estimated model. The purple dashed line depicts the estimated counterfactual distribution.

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 measure. 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.

The next paragraphs describe the estimation in detail. The first step constructs the variables $\widehat{\ln f(q_j)}$, an estimate of the log-measure 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):

$$\ln f_l(q) = \sum_{p=0}^{P(n)} \gamma_p \left(\ln \frac{q}{q^K} \right)^p . \tag{16}$$

[23] Note that the total number of observations does not include agents who do not adopt.

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 $[\underline{b}, \overline{b}] \supsetneq [q_L, q_H]$ 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 is not to be confused with the bin sizes of the log-histogram.

The estimation minimises the square distance between the model and the log-measure:

$$(\hat{\epsilon}, \hat{\eta}, \hat{\gamma}_P) = \arg \min_{(\epsilon,\eta,\gamma_P)\in\Theta_P} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \overline{b}]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \epsilon, \eta, \gamma_P) \right)^2 + \left(\widehat{\ln B} - \ln B(\epsilon, \eta, \gamma_P) \right)^2, \tag{17}$$

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 and Equation (13) as a function of the parameters. By a slight abuse of notation, γ_P denotes the vector of coefficients of the series. The set Θ_P denotes the sieve space defined in Equation (59) in Appendix C.2. Appendix C.2 shows that the estimator is consistent. Following Chetty et al. (2011), I estimate the standard errors using the nonparametric bootstrap.[24]

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.[25] 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 proses an alternative method for selecting the specification.

4 Main empirical application: the German solar subsidy

4.1 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 a procurement problem: the prin-

[24] Appendix C.2 discusses the asymptotic properties of the estimator.

[25] 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.

cipal (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.[26]

The subsidy programme was very successful at incentivising households and firms to adopt solar panels; numbers increased rapidly in most years. 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.[27] 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 A.9 for a more detailed description of the policy and data.

Consistent with the theoretical predictions in Section 2, 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, the left panel in 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.[28] 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 right panel in Figure 7 shows the histogram of adoptions from 2000 to 2002. There is no evidence for a time trend in the concavity of the distribution.

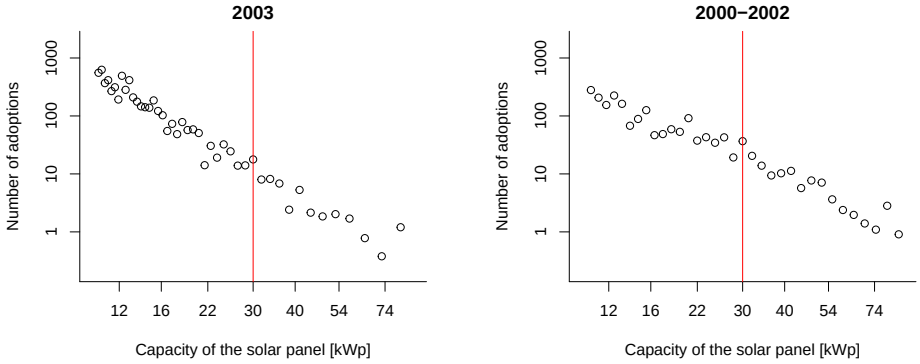
In order to evaluate the subsidy's cost-effectiveness, it is necessary to know how agents react to it at the participation and intensive margin. While subsidy rates and prices for solar panels vary over time, in a large market like Germany, this variation is endogenous to demand, making it unsuitable for identifying the two margins. Methods for estimating both margins proposed by Best 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 2 for estimating the two margins.

[26] See Appendix A.8.1 for a formal derivation of this statement

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

[28] 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 (left) and 2000-2002 (right); logarithmic scales.



Note: The red line marks the kink point. Scales are logarithmic. The number of observations in a bin is normalised by the bin-size, but not the total number of observations. The left panel shows the histogram of adoptions in 2003. The right panel shows the histogram of adoptions in 2000 to 2002. The right panel pools years to have a sufficiently large sample size. In all these years, the subsidy was linear. There is no significant mass point or slope change in the distribution.

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.[29] The subsidy $S(.)$ in model (1) is the present discounted value of the feed-in-tariff payments. Appendix A.8.1 shows that the model in Section 2 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 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.[30] Appendix A.8.2 further discusses the convexity of the variable cost function. Appendix A.8.3 discusses how the model encompasses dynamic decisions.

[29]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.

[30]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.

4.2 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 it is customary in the bunching literature to estimate 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 advantageous to use participation semi-elasticities (see Rochet and Chone, 1998 and Jacquet, Lehmann and Van der Linden, 2013). Hence, the counterfactual exercises in Appendix D.4 and Pollinger (2025) rely on semi-elasticities. 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.[31]

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 a discussion 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.1 presents and discusses

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.

the specification. Appendix D.2 discusses the details of selecting the specification.

4.3 Counterfactual exercises

Appendix D.4 uses the estimates in Table 1 to evaluate the nonlinearities in the German schedule. The counterfactual exercise solves for the linear subsidy that achieves the same aggregate capacity as the subsidy used by the government. Compared to this simple linear benchmark, the government's scheme only modestly reduces costs by 0.14 percent.

[31] 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: $SD_\kappa = SD_\eta / S(q)$.

Pollinger (2025) uses the estimates in Table 1 to conduct three more counterfactual exercises. First, the paper shows that the most cost-effective nonlinear subsidy triples these cost savings to 0.45 percent; however, they remain modest overall. To better understand what limits the scope for curtailing costs, a second experiment assumes adopters can only respond at the intensive margin. Compared to the linear benchmark, the optimal nonlinear subsidy would reduce costs by 8.6 per cent in this case. The exercise reveals that the participation margin is responsible for the limited scope for cost reductions when both margins are active. A third counterfactual exercise assumes the policymaker wrongly ignores the participation margin. She implements the optimal intensive schedule derived in the third thought experiment while adopters react at both margins. As a consequence, costs increase by 3.1 per cent instead of decreasing. These results 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 subsidies are not a panacea for cost-effectiveness and can even trigger detrimental effects. Estimating both response margins is essential for their optimal design.

The counterfactual exercise in Appendix D.4.4 serves at comparing the results of this paper to the results in Germeshausen (2018). It solves for the treatment effect of introducing a new kink at 10kWp on total capacity. 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.

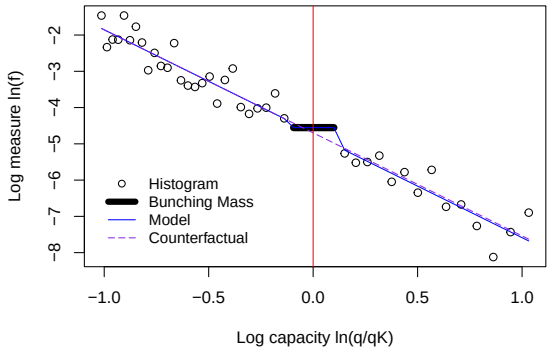
4.4 Robustness

One concern for the estimation is the violation of Assumption 3 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. The following five robustness checks alleviate these concerns.

A first robustness check selects the specification of the estimator using block-cross validation. This second selection method described in Appendix C.4 is an alternative to the main selection method in Appendix C.3. Note that both methods are data-driven. Reassuringly, the selected specification and, hence, the estimates do not change. This exercise shows that the estimates are robust to changing the selection method of the specification.

A second robustness check estimates the model on observations from 2000 to 2003, when there was no kink in the subsidy. It uses the specification selected by the second method, relying only on observations from 2004-2008. Figure 8 shows the distribution in 2000-2003, with the estimated model and the counterfactual. The estimates in Table 2 are insignificant, alleviating the above-mentioned concerns. Appendix D.5.1 reports the robustness check for the estimates at 100 kWp. Again, the estimates are not significant.

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 measure. 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 red line marks the kink point. The estimated model in blue is equal to the estimated counterfactual in purple. 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.

As a third robustness check, Appendix D.5.2 derives a direct estimator for the specification bias caused by an eventual violation of Assumption 3. Again, the estimator exploits untreated observations facing a subsidy without kink, such as the observations from 2000-2003. Table 7 in Appendix D.5.2 reports the results. The point estimates of the bias are small and statistically insignificant. The fourth robustness check in Appendix D.5.4 exploits the untreated observations to estimate both response margins without relying on Assumption 3. The estimates in Table 8 are not statistically different from the main estimates in Table 1. A fifth robustness check in Appendix D.5.6 relaxes Assumption 3 and estimates bounds for both elasticities. The bounds in Figure 17 and 18 are tight around the main estimates in Table 1.

The following appendices conduct additional robustness checks. Appendix D.5.8 reports robustness checks to the choice of the bunching interval; the estimates are robust. Appendix D.5.9 verifies the rank condition in sample; it holds by a large amount. Appendix D.5.10 relaxes Assumption 2, allowing the participation margin elasticity to vary in capacity; the generalisation does not change the main estimates.

4.5 Comparison to the classic bunching and regression kink design

The estimator in Section 3 differs from the classic bunching estimator in three dimensions. First, it estimates the shape of the counterfactual distribution $f_l(.)$ in the range $[q^K, q^K\rho^{-\epsilon}]$ using observations outside the bunching interval instead of assuming a parametric functional form in this range.[32] Second, it explicitly accounts for participation margin responses. Third, it estimates the counterfactual distribution outside the bunching interval nonparametrically, using a data-driven selection of the order of the series P. Appendix D.6 further discusses these differences and quantifies them for the main estimates of this paper. It finds that the implicit assumptions of the classic bunching estimator downward bias the intensive margin estimate by 23 %. Half of the bias is attributable to assuming a parametric functional form and half to ignoring the participation margin (see Table 13 in Appendix D.6). Moreover, Table 14 in Appendix D.6 shows estimates using the suboptimal order of the series $P = 7$, which is the specification used in Chetty et al. (2011) for estimating the bunching mass. The exercise renders 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 order of the series based on the mean squared error to increase precision.

Similarly, Appendix D.6 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 intensive margin responses are present.

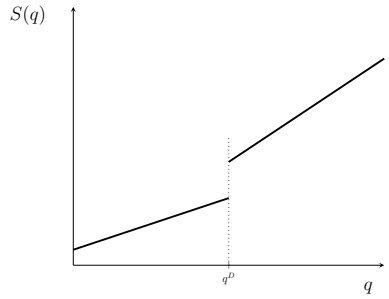
5 An application to the Danish income tax for foreigners

This section applies the method to the Danish income tax for foreigners. It serves three goals. First, it demonstrates that the methodology can be applied to taxes and, in particular, to income taxes instead of subsidies. Second, it demonstrates the applicability of the methodology to notches instead of kinks. Third, the section serves as a validation exercise. Without relying on the time dimension of the data, it recovers the same estimates as in Kleven et al. (2013), who use a difference-in-differences design in the same context.

The preferential income tax for foreigners in Denmark taxes new immigrants with high earnings at a preferential flat rate. It creates a notch in the income tax for this group. See Figure 9 for an illustration. Kleven et al. (2013) use a dif-in-dif and the introduction of the notch in 1991 to estimate the intensive and participation margin responses of foreigners in Denmark. The participation margin is the foreigners' decision to participate in the Danish labour market (i.e., to migrate to Denmark and work there).

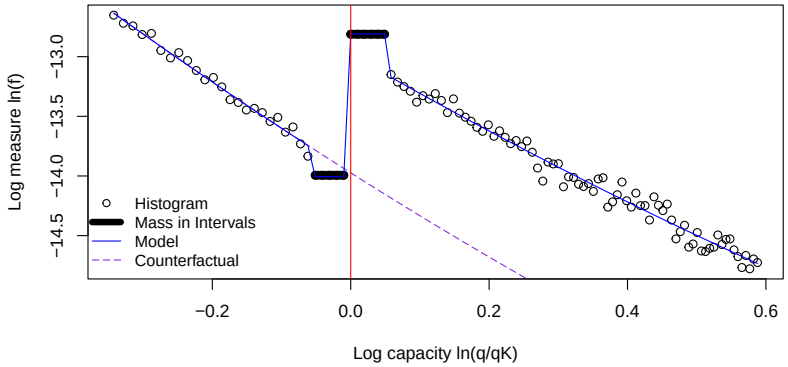
[32] Note that this is the range of agents who bunch sharply. It is a subinterval of the agents in the bunching interval $[q_L, q_H\rho^{-\epsilon}]$.

Figure 9: Net earnings $S_d(.)$ under the notched tax scheme.



Note: The thick solid line shows net earnings $S_d(.)$ under the notched scheme.

Figure 10: The distribution of earnings in 1995-2000 with the estimated model.



Note: The x-axis shows the normalised logarithm of earnings. The y-axis shows the logarithm of the measure of earnings. 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 bars show the mass in the bunching interval and the interval containing the missing mass. The red line marks the notch point.

This section shows that both margins can be estimated by relying solely on the cross-section of observations after the introduction of the notch. Figure 10 shows the measure of foreigners' earnings on logarithmic scales with the estimated model and the counterfactual. The red vertical line illustrates the location of the notch. The black bar below the notch shows the mass in this interval. As is well known in the bunching literature, there is missing mass in this interval since agents have a strong incentive to locate above the notch. Moreover, as is common in the presence of notches, some agents are located in this interval due to adjustment frictions. Above the notch, there is excessive mass in the bunching interval. Moreover, above the notch, the measure jumps upwards due to participation margin responses. As in Section 3, the estimation minimises the square distance between the model and the observable measure. The estimates in Table 15 in Appendix E.2 are statistically equal to the estimates in Kleven et al. (2013) (see

Appendix E.2 for the comparison). Therefore, the exercise validates both methodologies. The methodology proposed in this paper can be used whenever there is a kink or notch in the budget set. It does not rely on parallel trends and a policy reform, which are necessary for applying dif-in-dif. Therefore, the methodology is complementary to the methodology in Kleven et al. (2013). For a more detailed discussion of this application, see Appendix E. Appendix E.1 discusses identification. Appendix E.2 discusses the details of the estimation and robustness checks.

6 Further potential applications

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). As discussed in Section 2.3, in principle, the method in this paper can be applied whenever the bunching estimator can be applied.

Importantly, 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. Moreover, participation responses to income taxes have been documented empirically (e.g., see Blundell, Bozio and Laroque, 2011 and Chetty et al., 2013).[33] Indeed, Figures 3 and A2 in Mortenson and Whitten (2020) show bunching and visible changes in slope in the US earnings distribution at kink points in the US income tax schedule. It suggests simultaneous responses are relevant in this context.[34]

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. Estimating both response margins using the proposed estimator allows for the evaluation and optimisation of these incentive schemes.

[33]Typically, the literature finds very small participation elasticities for mid-age primary earners. The method of this paper can still be applied to estimate them as such (see Section 2.3). The other groups mentioned above usually show larger participation elasticities.

[34]The data in this paper are not generally available to researchers. See the Data Availability Statement in Mortenson and Whitten, 2020.

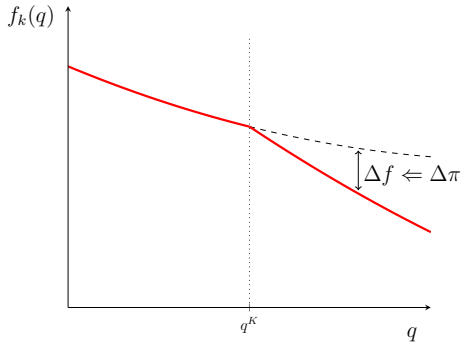
7 Conclusion

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

A Appendix

A.1 The effect of a kink when there are only participation margin responses

Figure 11: The observable measure f_k when there is only a participation margin.



Note: The figure shows the counterfactual measure f_l as a dashed black line and the observable measure f_k as a thick red line. The change in profit $\Delta\pi$ causes a change in participation Δf. There is no bunching.

Consider the case when there are only participation margin responses. This is the case when marginal cost curves are locally vertical. Again, adopters who counterfactually adopt below the kink point are not affected by a change in subsidy schemes. Adopters who counterfactually adopt above the kink point do not adjust the adopted capacity. Therefore, there is no bunching. The difference in the subsidy payment between the kinked and linear scheme ΔS equals their profit loss $\Delta\pi$. Due to heterogeneous fixed costs, it triggers a loss in participation Δf. To the right and close to the kink point, ΔS and $\Delta\pi$ increase in q. Therefore, Δf increases in q. Figure 11 illustrates the consequent effect on the observable measure of adoptions. The kink in the subsidy triggers a kink in the observable measure. Without intensive margin responses, the kink in the measure could be exploited to estimate participation margin responses using a regression kink design with q as the running variable and an estimate of the measure as the

outcome variable (e.g., see Gelber et al., 2021).

A.2 Proof of Proposition 1

Consider the interval $(\underline{q}, \overline{q})$ around the kink point and consider agents ν such that $q(s_l, \nu) > \underline{q}$ and $q(\rho s_l, \nu) < \overline{q}$. Since $q(s, \nu)$ is continuous in both arguments, increasing in the first argument and decreasing in the second argument, the two inequalities define an open interval of types $(\underline{\nu}, \overline{\nu})$. It is equal to the set of all types ν locating in the interval $(\underline{q}, \overline{q})$ under the kinked subsidy $S_k(.)$. Therefore, Assumption 1 applies to all agents of type ν locating in the interval $(\underline{q}, \overline{q})$ under the kinked subsidy $S_k(.)$. Moreover, the assumption applies to the domain of quantities q for which the marginal cost functions have the range $[s_l\rho, s_l]$. In particular, the assumption applies for all $q(s_l, \nu)$, where s_l is the counterfactual linear subsidy rate and $\nu \in (\underline{\nu}, \overline{\nu})$. Defining $q_l = q(s_l, \nu)$ and changing variable from ν to q_l in Equation (5), it follows that Assumption 1 applies to all $(\underline{q_l}, \overline{q_l}) = (\underline{q}, \overline{q}\rho^{-\epsilon})$.

In this section, denote by $q_k(q_l)$ the choice of type q_l under the kinked subsidy $S_k(.)$. Using the cost function (6) and solving the optimization problem (1) renders: $q_k(q_l) = q_l$ for $q_l < q^K; q_k(q_l) = q^K$ for $q_l \in [q^K, q^K\rho^{-\epsilon}]; q_k(q_l) = q_l\rho^{\epsilon}$ for $q_l > q^K\rho^{-\epsilon}$. The relative change in profit of adopter q_l is $R(q_l, \epsilon) = \frac{S_k(q_k(q_l)) - c((q_k(q_l), q_l, c_t) + c_t}{S_l(q_l)}$, which renders the function $R(., \epsilon)$ in Proposition 1. Remember, $f_l(.)$ denotes the measure of q_l conditional on participation; the unconditional measure is $\frac{f_l(q_l)}{F_{t|q_l}(S_l(q_l)|q_l)}$. The mass of participating adopters under the kinked scheme as a function of q_l is $F_{t|q_l}(S_k(q_k(q_l)) - c((q_k(q_l), q_l, c_t) + c_t|q_l) \frac{f_l(q_l)}{F_{t|q_l}(S_l(q_l)|q_l)} = R(q_l, \epsilon)^{\eta} f_l(q_l)$, where the equality follows from Assumption 2. Changing variable to q_k renders the result. Note that $f_k(.)$ does neither depend on the unconditional distribution of q_l nor the normalisation term $g(.)$, because the latter cancels out in the derivation.

A.3 Proof of Lemma 1

By Proposition 1.2.12 in Krantz and Parks (2002), $\ln f_l(.)$ is real analytic on the interval $(\underline{q_l}, \overline{q_l})$. 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 $q_l \in (\underline{q_l}, \overline{q_l})$. By Corollary 1.2.4 in Krantz and Parks (2002), the function defined by the power series representation is real analytic on $(\underline{q_l}, \overline{q_l})$. 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 $(\underline{q_l}, \overline{q_l})$.

A.4 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 general values of these objects. Denote by $\gamma = (\gamma_0, \gamma_1, ...)$ the infinitely dimensional vector of coefficients γ_p in Lemma 1

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 $[q_L, q_H] \subsetneq [\underline{b}, \overline{b}] \subsetneq (\underline{q}, \overline{q})$. Also note that $(\underline{q_l}, \overline{q_l}) = (\underline{q}, \overline{q}\rho^{-\epsilon^o})$.

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

$$Q(\theta) = \int_{\underline{b}}^{q_L} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \epsilon, \eta, \gamma\right)\right)^2 \, dF^w + \left(\ln B^o - \ln B(\epsilon, \eta, \gamma)\right)^2 + \\ + \int_{q_H}^{\overline{b}} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \epsilon, \eta, \gamma\right)\right)^2 \, dF^w. \tag{18}$$

The function $f_k^o(.)$ denotes the observable measure of agents in the population. The function $f_k(.|\epsilon, \eta, \gamma)$ denotes $f_k(.)$ in Proposition 1 as a function of the parameters and $f_l(.)$ 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 Equation (13), where $f_l(.)$ 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 $(q_x, q_y) \subset [\underline{b}, q_L] \cup [q_H, \overline{b}]$. The weighting measure may be used to increase the efficiency of the estimator. Identification follows for any such F^W. In particular, it follows in the case when there is no such measure, i.e., F^w is the uniform CDF.

Lemma 2. *Under Assumptions 1, 2, 3, in any minimum of $Q(\theta)$ it holds that $\gamma = \gamma^o$. Under one of the conditions in Lemma 5, $\theta^o = (\epsilon^o, \eta^o, \gamma^o)$ is the unique minimum of $Q(\theta)$.*[35]

PROOF Lemma 2: By Proposition 1 and Lemma 1, $f_k^o(q) = f_k(q \mid \epsilon^o, \eta^o, \gamma^o)$ 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 $Q(.)$. It follows that the first integral in Equation (18) is equal to zero. Using Proposition 1 and Lemma 1, it follows that: $\int_{\underline{b}}^{q_L} \left(\ln f_l(q \mid \gamma^o) - \ln f_l\left(q \mid \gamma\right)\right)^2 \, dF^w = 0$. Lemma 1 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 $q \in (\underline{b}, q_L)$. Lemma 1 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 $(\ln \underline{q_l}, \ln \overline{q_l}) \supset (\ln \underline{b}, \ln q_L)$.[36] 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 $\ln q \in (\ln \underline{q_l}, \ln \overline{q_l})$ (see Corollary 1.2.6 in Krantz and Parks, 2002). 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 $(\underline{q_l}, \overline{q_l})$.

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 $f_k(q \mid \epsilon^o, \eta^o, \gamma^o) - f_k(q \mid \epsilon, \eta, \gamma^o) =$

[35] Under conditions (i)-(j) in Lemma 5, the minimum is locally unique.

[36] 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 $\ln f_l\left(e^{(\ln q)} \mid \gamma\right) = \ln f_l\left(q \mid \gamma\right)$.

$0\ \forall q \in (q_H, \overline{b})$. By Proposition 1

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

It follows that

$$\epsilon < \frac{\ln \overline{q_l} - \ln \overline{b}}{-\ln \rho}. \tag{20}$$

Otherwise, $f_l(.|\gamma^o)$ in the right-hand side of (19) does not converge, and the equation does not hold.[37] Both sides of Equation (19) 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

$$f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o}R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o} = f_l(q\rho^{-\epsilon}|\gamma^o)\rho^{-\epsilon}R_2(q\rho^{-\epsilon},\epsilon)^{\eta}\ \forall q \in (\underline{q}, \overline{b}). \tag{21}$$

The second part of the criterion $Q(.)$ implies that $B(\epsilon^o, \eta^o, \gamma^o) = B(\epsilon, \eta, \gamma^o)$. By Equation (13) and since $f_l^o(.) = f_l(.|\gamma^o)$

$$\begin{aligned}\int_{q_L}^{q^K} f_l^o(q_l)dq_l + \int_{q^K}^{q^K\rho^{-\epsilon^o}} R_B(q_l,\epsilon^o)^{\eta^o} f_l^o(q_l)dq_l + \int_{q^K}^{q_H} f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o}R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o}dq = \\ = \int_{q_L}^{q^K} f_l^o(q_l)dq_l + \int_{q^K}^{q^K\rho^{-\epsilon}} R_B(q_l,\epsilon)^{\eta} f_l^o(q_l)dq_l + \int_{q^K}^{q_H} f_l^o(q\rho^{-\epsilon})\rho^{-\epsilon}R_2(q\rho^{-\epsilon},\epsilon)^{\eta}dq,\end{aligned} \tag{22}$$

Using Equation (21) and cancelling equal terms on both sides renders

$$\int_{q^K}^{q^K\rho^{-\epsilon^o}} R_B(q_l,\epsilon^o)^{\eta^o} f_l^o(q_l)dq_l = \int_{q^K}^{q^K\rho^{-\epsilon}} R_B(q_l,\epsilon)^{\eta} f_l^o(q_l)dq_l. \tag{23}$$

The parameter η in Equations (21) and (23) is equal to the functions $\eta(q,\epsilon)$ and $\eta_B(\epsilon)$ defined in Section A.5.1. Under one of the conditions (a)-(h) in Lemma 5, for all $\epsilon \neq \epsilon^o$ and $\epsilon < \frac{\ln \overline{q_l} - \ln \overline{b}}{-\ln \rho}$, there exits a $q \in (\underline{q}, \overline{b})$ such that $\eta(q,\epsilon) \neq \eta_B(\epsilon)$. Therefore, the only pair (ϵ, η) that fulfils Equations (21) and (23) for all q is (ϵ^o, η^o). It follows that $(\epsilon^o, \eta^o, \gamma^o)$ is the globally unique minimum of $Q(.)$. Under one of the conditions (i)-(j) in Lemma 5, the minimum is locally unique.
qed.

Consider the parameter-space

$$\Theta = \{\theta \text{ s.t. } 0 \leq \epsilon \leq \overline{\epsilon},\ \text{and } 0 \leq \eta \leq \overline{\eta},\ \text{and }\ |\gamma_p| \leq M(\ln \overline{q_l} - \ln q^K)^{-p}\}, \tag{24}$$

where $\overline{\epsilon} > 0$ and $\overline{\eta} > 0$ are upper bounds on ϵ and η. Consider $(\ln q^K - \ln \underline{q_l}) = (\ln \overline{q_l} -$

[37] Note that by assumption ϵ^o is below the bound in Equation (20).

$\ln q^K) > 1$.[38] Note that by Assumption 1, 2, and Lemma 1, $\theta^o \in \Theta$. 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 3. *Under Assumptions 1, 2, 3, the parameter space Θ defined in Equation (24) is compact.*

Lemma 4. *Under Assumptions 1, 2, 3, the criterion $Q(\theta)$ is continuous at all $\theta \in \Theta$.*

The proofs of Lemmas 3 and 4 are in the Appendices B.2 and B.3.

By Lemma 2-4, $Q(\theta)$, θ^o, and Θ fulfil the Condition 3.1" for identification in Chen (2007), which proves Propositions 2.

A.5 Rank condition

A.5.1 Lemma 5

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. Use (9) and (10) and evaluate them at the true and candidate values of ϵ and η. For expression (10), consider its real analytic continuation.

Definition 1. *The functions $\eta_B(\epsilon)$ and $\eta(q, \epsilon)$ are the η implicitly defined by the equations*

$$\int_{q^K}^{q^K \rho^{-\epsilon^o}} R_B(q_l, \epsilon^o)^{\eta^o} f_l^o(q_l) dq_l = \int_{q^K}^{q^K \rho^{-\epsilon}} R_B(q_l, \epsilon)^{\eta} f_l^o(q_l) dq_l; \tag{25}$$

$$f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o} R_2(q\rho^{-\epsilon^o}, \epsilon^o)^{\eta^o} = f_l(q\rho^{-\epsilon}|\gamma^o)\rho^{-\epsilon} R_2(q\rho^{-\epsilon}, \epsilon)^{\eta} \; \forall q \in (\underline{q}, \overline{b}). \tag{26}$$

Note that, by Proposition 1, the left-hand side of Equation (25) is $\tilde{B}^o$; the left-hand side of Equation (26) is the true $f_k^o(q)$ above the kink point and its real analytic continuation to points below the kink point. Denote it by $f_{k2}^o(.)$. Also note that $(\underline{q_l}, \overline{q_l}) = (\underline{q}, \overline{q}\rho^{-\epsilon^o})$.

Lemma 5. *Condition (a) below is necessary and sufficient such that Equation (9) and (10) have a unique intercept at (ϵ^o, η^o). Each of the conditions (b)-(h) below is sufficient but not necessary for a unique intercept at (ϵ^o, η^o). Each of the conditions (i) and (j) is sufficient but not necessary for a locally unique intercept.*

(a) The true counterfactual $f_l^o(.)$ is such that, for all $\epsilon \neq \epsilon^o$ and $\epsilon < \frac{\ln \overline{q_l} - \ln \overline{b}}{-\ln \rho}$, there exits a $q \in (\underline{q}, \overline{b})$ such that $\eta(q, \epsilon) \neq \eta_B(\epsilon)$. Note that this condition is simply a formal definition of Equation (9) and (10) having a unique intercept at (ϵ^o, η^o). Also note that, by Equation (20) in Appendix A.4, such a q exists for all $\epsilon \geq \frac{\ln \overline{q_l} - \ln \overline{b}}{-\ln \rho}$.

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

(b) For all $\epsilon \neq \epsilon^o$ and $\epsilon < \frac{\ln \overline{q_l} - \ln \overline{b}}{-\ln \rho}$, the true counterfactual $f_l^o(.)$ does not solve the functional equation:

$$\ln f_l^o(q) = \ln f_l^o(q\rho^{-(\epsilon-\epsilon^o)}) + \ln \rho^{-(\epsilon-\epsilon^o)} + \eta_B(\epsilon) \ln R_2(q\rho^{-(\epsilon-\epsilon^o)}, \epsilon) - \eta^o \ln R_2(q, \epsilon^o), \tag{27}$$

$\forall q \in (\underline{q}\rho^{-\epsilon^o}, \overline{b}\rho^{-\epsilon^o})$. *Generically, the functional equation (27) does not hold.*

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

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

(e) $\frac{d \ln f_{k2}^o(q)}{d \ln q} \geq 0$ *for all* $q \in (\underline{q}, \overline{q})$. *Remember, $f_{k2}^o(.)$ denotes the observable measure above the kink point and its analytic continuation to points below the kink point.*[39]

(f) The function $f_l^o(.)$ is real analytic in q on the interval $(0, \overline{q_l})$ and there exists an order of derivative P such that $\lim_{q \downarrow 0} \frac{d^P f_l^o(q)}{dq^P} \neq 0$ or $\pm \infty$.

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

(h) There exists an order P such that

$$\ln f_l^o(q_l) = \sum_{p=0}^{P} \gamma_p^o \left(\ln \frac{q_l}{q^K} \right)^p \quad \text{for all } q_l \in (\underline{q_l}, \overline{q_l}). \tag{28}$$

(i) There exists a $q \in (\underline{q}, \overline{q})$ such that $\eta_B'(\epsilon^o) - \frac{\partial \eta(q, \epsilon^o)}{\partial \epsilon} \neq 0$.

(j)

$$\left(\frac{d \ln f_l^o(\hat{q}(\epsilon^o)\, \rho^{-\epsilon^o})}{d \ln q} + 1 \right) \ln(1/\rho) + \eta^o \frac{(1 - \rho^{1+\epsilon^o}) - (1 + \epsilon^o) \ln(1/\rho)}{(1 + \epsilon^o)^2} \neq 0, \tag{29}$$

where $f_l^o(.)$ is the true counterfactual and $\hat{q}(\epsilon^o)$ is the unique q defined by $R_2(\hat{q}(\epsilon^o)\rho^{-\epsilon^o}, \epsilon^o) = 1$. This condition holds generically.

The proof is in Section B.1.

A.5.2 The economic intuition why the rank conditions hold

Equations (9) and (10) form a simultaneous, nonlinear system of equations in the parameters (ϵ, η). They provide infinitely many moments, i.e., one for each q, to identify the two parameters ϵ^o and η^o. Therefore, this system of equations is usually overidentified. Identification can

[39] 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 $f_l^o(.)$.

only fail if, for all q, equations (9) and (10) have a second intercept in (ϵ, η). This can only be the case if $f_l^o(.)$ fulfils the functional equation (27) in Lemma 5. Since this equation is difficult to solve, it is difficult to show that no function fulfils this equation. However, point (b) in Lemma 5 shows that generically, $f_l^o(.)$ does not fulfil Equation (27).

Moreover, there is an economic argument as to why Equations (9) and (10) have a unique intercept at the true parameter values. Denote the right-hand side of Equation (9) as a function of its parameters by $\tilde{B}(\epsilon, \eta)$; denote the right-hand side of Equation (10) 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 (9) 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 $\frac{\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \epsilon}}{-\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \eta}}$ 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 (26) it follows that:

$$\frac{d \ln f_{k2}(q|\epsilon, \eta)}{d \ln q} = \frac{d \ln f_l^o(q\ \rho^{-\epsilon})}{d \ln q} + \eta\ \frac{d \ln R_2(q\ \rho^{-\epsilon}, \epsilon)}{d \ln q}. \tag{30}$$

Usually, $\frac{d \ln f_l^o(q)}{d \ln q}$ is approximately constant. Moreover, for q close to q^K, $\frac{d \ln R_2}{d \ln q}$ is approximately $-(1-\rho)$ and 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 $\frac{d \ln f_{k2}(q)}{d \ln q}$ are strong evidence that conditions (a) and (i) in Lemma 5 hold. Section D.5.9 evaluates the conditions in sample. It confirms the intuition outlined above: the rank condition holds by a large amount.

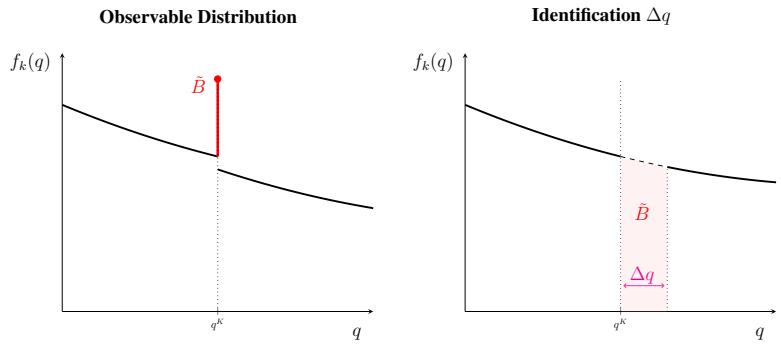
A.6 A comparison of Proposition 2 to the results in Blomquist and Newey (2017), Bertanha, McCallum and Seegert (2023), Chetty et al. (2011), and Goff (2022)

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 $\tilde{B}$ 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 $\tilde{B}$ 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 nonparametric but fulfils Assumption 3 or the conditions in Lemma 1. Note that this result does not contradict Theorem 1 in Blomquist and Newey (2017) because Assumption 3 or Lemma 1 are stronger than assuming differentiability of order D. Under Assumption 3 or the conditions in Lemma 1, 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 restrictis the shape of $f_l(.)$. Proposition 2 shows that under Assumption 3 or Lemma 1, both elasticities are point-identified.[40] Point-estimates are necessary to conduct the counterfactual exercises in Section D.4 and Pollinger (2025). Building on Blomquist et al. (2021), Bertanha, McCallum and Seegert (2023), and Goff (2022), Appendix D.5.6 derives partially identification results for both margins. 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.

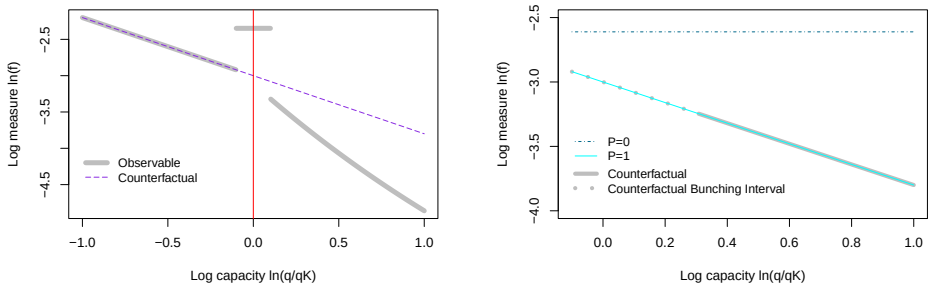
A.7 An illustration of Proposition 2 using simulations

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 $\theta_P = \arg\min_{\theta \in \Theta_P} Q(\theta)$ converges to the true parameter θ^o as the order of the series P increases. The population criterion $Q(\theta)$ is defined in Equation (18) in Section A.4. Intuitively, it is the square distance between the logarithm of the observed measure $f_k^o(.)$ and the corresponding object in the model. Remember, $\theta = (\epsilon, \eta, \gamma_0, \gamma_1, ...)$, where $(\gamma_0, \gamma_1, ...)$ is the sequence of coefficients of the power series in Equation (15). The minimisation 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 econometric model, I use

[40] Note that when the interval $(\underline{q_l}, \overline{q_l})$ in Lemma 1 is not large enough, point-identification is not possible.

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



Note Left Panel: 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* $\ln f_k^o(.)$. *Note Right Panel: 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.*

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

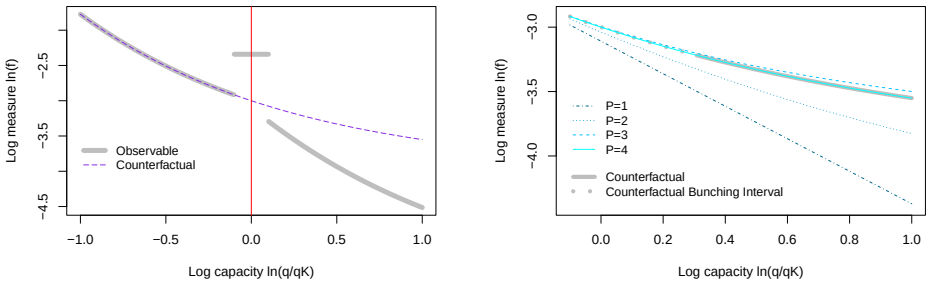
	Epsilon	Bias Epsilon [%]	Eta	Bias Eta [%]
True Value	0.30000	0.000	3.0000	0.000
P=0	0.11691	-61.029	6.5562	118.540
P=1	0.30000	0.000	3.0000	0.000
P=2	0.30000	0.000	3.0000	0.000

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.*

$\ln f_l(q_l) = \sum_{p=0}^{P} \gamma_p \left(\ln \frac{q_l}{q^K}\right)^p$, 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 $P \geq 1$. The left panel in 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 $\ln B^o$. Unsurprisingly, Table 3 shows that $(\epsilon_P, \eta_P) = (\epsilon^o, \eta^o)$ for $P \geq 1$. The right panel in Figure 13 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 $f_l(.)$.

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.

Figure 14: Simulation; example 2, exponential $\ln f_l^o(.)$.



Note Left Panel: 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* $\ln B^o$*.*
Note Right Panel: 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.*

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

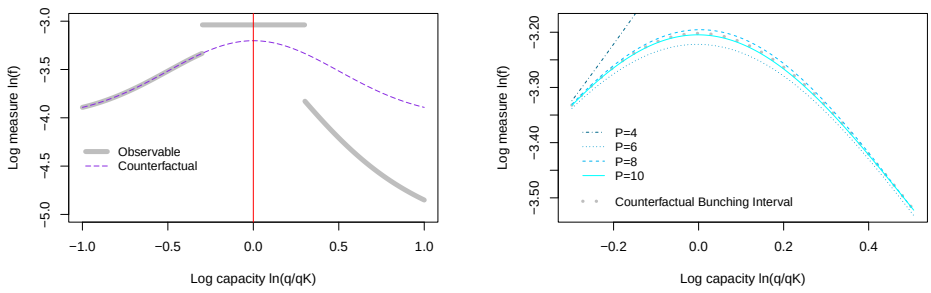
	Epsilon	Bias Epsilon [%]	Eta	Bias Eta [%]
True Value	0.30000	0.000	3.0000	0.000
P=1	0.36239	20.795	0.3976	-86.747
P=2	0.32369	7.895	2.1423	-28.591
P=3	0.29792	-0.694	3.1557	5.190
P=4	0.30000	0.000	2.9996	-0.012

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.*

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. The left panel in Figure 14 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 %. The right panel in Figure 14 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 3, normal $\ln f_l$.



Note Left Panel: 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* $\ln B^o$.
Note Right Panel: 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.*

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. The left panel in Figure 15 shows the observable, the counterfactual, and the bunching mass. Table 5 shows the true and inferred values of the elasticities. As in example

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.

two, η_P and ϵ_P converge to their true values as P increases; at $P = 10$, the biases are below 1 percent.[41] 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. The right panel in Figure 15 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 Lemma 1 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 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 3 discusses the choice of P in further detail.

A.8 Generalisations of the model

A.8.1 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 make the adoption decision based on the present discounted value of the income stream produced by the solar panel.

[41] 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 $P - 1$.

A household i 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: $S_k(q, e_{it}) = s_l\ e_{it},\ \text{ for } q \leq q^K;\ S_k(q, e_{it}) = s_l\ e_{it}\frac{q^K}{q} + s_l\rho e_{it}\frac{q-q^K}{q},\ \text{ for } q > q^K$. It follows that the subsidy payment in a certain year as a function of q is: $S_k(q, w_{it}) = s_l\ w_{it}\ q,\ \text{ for } q \leq q^K;\ S_k(q, w_{it}) = s_l w_{it} q^K + s_l\rho w_{it}(q - q^K),\ \text{ for } q > q^K$. 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

$$\mathbb{E}_i\left[\sum_{t=0}^{20}\beta_i^t w_{it}S_k(q)\right] = S_k(q)\mathbb{E}_i\left[\sum_{t=0}^{20}\beta_i^t w_{it}\right] = S_k(q)\zeta^i. \tag{31}$$

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

$$\tilde{\pi}_v = \max_q\{\zeta S(q) - \tilde{c}_v(q, \tilde{\nu})\},\ \text{and participate if and only if } \tilde{\pi}_v \geq \tilde{c}_f, \tag{32}$$

where ζ captures the individual-specific discounting and location. Normalisation by ζ and applying Assumption 1 shows the equivalence of Problem (32) and Problem (1). Therefore, the model outlined in Section 2 implicitly accounts for subsidy payments via a feed-in tariff. In particular, it accounts for individual-specific discounting and location.

A.8.2 The convexity of the variable cost function

There are at least two reasons why the cost function $c_v(., \nu)$ in Section 2 is convex. First, for a given adopter, 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 $c_v(q, \nu)$ of a specific adopter ν is convex for q large enough. Note that the optimal choice of q is always in the convex range of $c_v(q, \nu)$.

A.8.3 Dynamic decisions

Suppose there are two time periods. For simplicity, assume there is no discounting. Denote the decision problem in the current period by $\pi = \max \underbrace{S(q) - c_v(q, \nu)}_{\pi_v} - \tilde{c}_f$ Denote with superscript $-t$ these variables in the other time period. Note that $-t$ may be before or after the current period: $\pi^{-t} = \max S^{-t}(q) - c_v^{-t}(q, \nu) - \tilde{c}_f^{-t}$ 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.

A.9 Data description

The data used in this paper are administrative and contain all solar panels connected to the grid and receiving subsidy payments. 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.[42] The data provides information on the time point of adoption, the electricity production, the applied subsidy rates, and the system's capacity.

Acknowledgements

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

[42] 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.

References

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.

Bergmeir, Christoph, and José M Benítez. 2012. "On the use of cross-validation for time series predictor evaluation." *Information Sciences*, 191: 192–213.

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.

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.

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*.

Mortenson, Jacob A, and Andrew Whitten. 2020. "Bunching to maximize tax credits: Evidence from kinks in the US tax schedule." *American Economic Journal: Economic Policy*, 12(3): 402–432.

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. 1997. "Convergence rates and asymptotic normality for series estimators." *Journal of econometrics*, 79(1): 147–168.

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.

Pollinger, Stefan. 2025. "Optimal Nonlinear Solar Subsidies." Mimeo: www.stefanpollinger.com/research.

Roberts, David R, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, et al. 2017. "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure." *Ecography*, 40(8): 913–929.

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. 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.

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.

Song, Myunghyun. 2024. "Identification and Inference in General Bunching Designs." *arXiv preprint arXiv:2411.03625*.

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. 2017. "Anlagenstammdaten und Bewegungsdaten." Distributer: WWW.NETZTRANSPARENZ.DE for 50Hertz Transmission GmbH, Berlin; Amprion GmbH, Dortmund; TenneT TSO GmbH, Bayreuth; TransnetBW GmbH, Stuttgart. URL: "https://www.netztransparenz.de/EEG", Last accessed on 2018-09-01.

Supplemental appendices for online publication

B Details Appendix A

In this section, 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.1 Proof of Lemma 5

As a first step, consider the following intermediate result:

Lemma 6. *The function $\eta'_B(\epsilon) > 0$ and continuous. In particular, $\eta'_B(\epsilon_o)$ is a strictly positive real number.*

PROOF Lemma 6: Denote the right-hand side of Equation (25) as a function of its parameters by $\tilde{B}(\epsilon, \eta)$. Implicitly differentiate Equation (25) to derive $\eta'_B(\epsilon) = \frac{\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \epsilon}}{-\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \eta}}$. The numerator $-\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \eta} = \int_{q^K}^{q^K \rho^{-\epsilon}} R_B(q_l, \epsilon)^\eta f_l^o(q_l)(-\ln R_B(q_l, \epsilon))dq_l > 0$, 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 $\frac{\partial \tilde{B}(\epsilon,\eta)}{\partial \epsilon} = \int_{q^K}^{q^K \rho^{-\epsilon}} \eta R_B(q_l, \epsilon)^{(\eta-1)} f_l^o(q_l) \frac{\partial R_B(q_l,\epsilon)}{\partial \epsilon} dq_l + R_B(q^K \rho^{-\epsilon}, \epsilon)^\eta f_l^o(q^K \rho^{-\epsilon}) q^K \rho^{-\epsilon} \ln(1/\rho) > 0$, 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 (a), Lemma 5: By definition.
PROOF, Point (i), Lemma 5: At (ϵ^o, η^o), the functions $\eta(q, \epsilon)$ and $\eta_B(\epsilon)$ are equal but they have different slopes. Therefore, locally, the intercept is unique.
PROOF, Point (d), Lemma 5: By Lemma 6, $\eta'_B(\epsilon) > 0$ and continuous. Moreover, $\eta_B(\epsilon^o) = \eta^o$. Therefore, $\eta_B(\epsilon)$ intersects with the known η^o once.
PROOF, Point (j), Lemma 5: Denote the right-hand side of Equation (26) as a function of its parameters by $f_{k2}(q|\epsilon, \eta)$. Implicitly differentiate the logarithm of Equation (26) to derive

$$\frac{\partial \eta(q,\epsilon)}{\partial \epsilon} = \frac{\frac{\partial \ln f_{k2}(q|\epsilon,\eta)}{\partial \epsilon}}{-\frac{\partial \ln f_{k2}(q|\epsilon,\eta)}{\partial \eta}} = \frac{\left(\frac{d \ln f_l^o(q\ \rho^{-\epsilon})}{d \ln q} + 1\right) \ln(1/\rho) + \eta \frac{\frac{\partial R_2(q\rho^{-\epsilon},\epsilon)}{\partial q} q\rho^{-\epsilon} \ln(1/\rho) + \frac{\partial R_2(q\rho^{-\epsilon},\epsilon)}{\partial \epsilon}}{R_2(q\rho^{-\epsilon},\epsilon)}}{-\ln R_2(q\rho^{-\epsilon}, \epsilon)}. \tag{33}$$

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

$$\left(\frac{d \ln f_l^o(\hat{q}(\epsilon^o)\ \rho^{-\epsilon^o})}{d \ln q} + 1\right) \ln(1/\rho) + \eta^o \frac{(1 - \rho^{1+\epsilon^o}) - (1 + \epsilon^o) \ln(1/\rho)}{(1 + \epsilon^o)^2}. \tag{34}$$

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 $\lim_{q\downarrow\hat{q}(\epsilon^o)} \frac{\partial\eta(\epsilon^o,q)}{\partial\epsilon} = -\infty \neq \eta'_B(\epsilon_o)$, 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 $\lim_{q\uparrow\hat{q}(\epsilon^o)} \frac{\partial\eta(\epsilon^o,q)}{\partial\epsilon} = -\infty \neq \eta'_B(\epsilon_o)$, since $-\ln\left(R_2(q\rho^{-\epsilon},\epsilon)\right) < 0$ for all $q < \hat{q}(\epsilon^o)$ and $\lim_{q\uparrow\hat{q}(\epsilon^o)} -\ln\left(R_2(q\rho^{-\epsilon},\epsilon)\right) = 0$. Therefore, the sufficient condition (i) holds.
qed.

PROOF, Point (e), Lemma 5: Remember, the left-hand side of Equation (26) denoted by $f^o_{k2}(q)$. Change variable in Equation (26) and solve for η to derive the function

$$\eta_l(q_l,\epsilon) = \frac{\ln f^o_{k2}(q_l\rho^\epsilon) + \ln\rho^\epsilon - \ln f^o_l(q_l)}{\ln R_2(q_l,\epsilon)} \text{ for all } q_l \in (\underline{q}\rho^{-\epsilon}, \overline{b}\rho^{-\epsilon}). \tag{35}$$

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:

$$\frac{d\,\eta_l(q_l(\epsilon),\epsilon)}{d\epsilon} = \frac{\frac{d\ln f^o_{k2}(q_l(\epsilon)\rho^\epsilon)}{d\ln q}\left(\frac{q'_l(\epsilon)}{q_l(\epsilon)} + \ln\rho\right) + \ln\rho - \frac{d\ln f^o_l(q_l(\epsilon))}{d\ln q}\left(\frac{q'_l(\epsilon)}{q_l(\epsilon)}\right)}{\ln R_2(q_l(\epsilon),\epsilon)}. \tag{36}$$

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 $\frac{\hat{q}'_l(\epsilon)}{\hat{q}_l(\epsilon)} + \ln\rho < 0$ and $\frac{\hat{q}'_l(\epsilon)}{\hat{q}_l(\epsilon)} > 0$. The elasticity $\frac{d\ln f^o_{k2}(\hat{q}_l(\epsilon)\rho^\epsilon)}{d\ln q} \geq 0$ by assumption. It also implies $\frac{d\ln f^o_l(\hat{q}_l(\epsilon))}{d\ln q} > 0$ since $\frac{d\ln f^o_l(\hat{q}_l(\epsilon))}{d\ln q} = \frac{d\ln f^o_{k2}(\hat{q}_l(\epsilon)\rho^{\epsilon^o})}{d\ln q} - \frac{d\ln R_2(\hat{q}_l(\epsilon),\epsilon^o)}{\partial\ln q}$ 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 $q_l(\epsilon) < \hat{q}_l(\epsilon)$.[43] Therefore, since $\eta'_B(\epsilon) > 0$, the sufficient conditions (i) and (a) hold.
qed.

PROOF, Point (c), Lemma 5: Remember, left-hand side of Equation (26) is denoted by $f^o_{k2}(q)$. Use Equation (26) and explicitly solve for η to derive

$$\eta(q,\epsilon) = \frac{\ln f^o_{k2}(q) - f^o_l(q\rho^{-\epsilon}) - \ln\rho^{-\epsilon}}{\ln R_2(q\rho^{-\epsilon},\epsilon)} \text{ for all } q \in (\underline{q},\overline{b}). \tag{37}$$

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

[43] Note that $\overline{b}\rho^{-\epsilon} > \hat{q}_l(\epsilon) > \underline{q}\rho^{-\epsilon}$, $\overline{q} > \hat{q}_l(\epsilon)\rho^\epsilon > \underline{q}$, and $\overline{q} > \hat{q}_l(\epsilon)\rho^{\epsilon^o} > \underline{q}$. Therefore, all relations used are in the relevant range. The three inequalities follow since: by assumption, $\frac{\overline{q_l}}{q^K} > \rho^{-\epsilon^o}$; by the property of the convergence radius of power series $\frac{\overline{q_l}}{q^K} = \frac{q^K}{\underline{q_l}}$; by the definition of $R_2(.)$, $q^K \leq \hat{q}_l(\epsilon) \leq q^K\rho^{-\epsilon}$; by definition $(\underline{q_l},\overline{q_l}) = (\underline{q},\overline{q}\rho^{-\epsilon^o})$; by Equation (20) $\overline{b}\rho^{-\epsilon} < \overline{q_l}$; combining these inequalities renders the result.

$$\frac{d\,\eta(q(\epsilon),\epsilon)}{d\epsilon} = \frac{\frac{d\ln f^o_{k2}(q(\epsilon))}{d\ln q}\frac{q'(\epsilon)}{q(\epsilon)} - \frac{d\ln f^o_l(q(\epsilon)\rho^{-\epsilon})}{d\ln q}\left(\frac{q'(\epsilon)}{q(\epsilon)} + \ln\rho^{-1}\right) - \ln\rho^{-1}}{\ln R_2(q(\epsilon)\rho^{-\epsilon},\epsilon)}. \tag{38}$$

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 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 $q(\epsilon) > \hat{q}(\epsilon)$.[44] Therefore, since $\eta'_B(\epsilon) > 0$, the sufficient conditions (i) and (a) hold.
qed.

PROOF, Point (f), Lemma 5: Suppose condition (a) does not hold. It follows that there exists an $\epsilon \neq \epsilon^o$ and a η such that

$$f^o_l(q\rho^{-\epsilon^o})\rho^{-\epsilon^o}R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o} = f^o_l(q\rho^{-\epsilon})\rho^{-\epsilon}R_2(q\rho^{-\epsilon},\epsilon)^{\eta} \text{ for all } q \in (\underline{q},\overline{b}). \tag{39}$$

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

$$\lim_{q\downarrow 0}\frac{f^o_l(q\rho^{-\epsilon^o})\rho^{-\epsilon^o}}{f^o_l(q\rho^{-\epsilon})\rho^{-\epsilon}} = \lim_{q\downarrow 0}\frac{R_2(q\rho^{-\epsilon},\epsilon)^{\eta}}{R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o}} \text{ for all } q. \tag{40}$$

By assumption, there exists an order of derivative P such that $\lim_{q\downarrow 0}\frac{d^P f^o_l(q)}{dq^P} \neq 0$ or $\pm\infty$. Using l'Hopital's rule, it follows that the left-hand side is a strictly positive real number. Next, consider the right-hand side

$$\lim_{q\downarrow 0}\frac{R_2(q\rho^{-\epsilon},\epsilon)^{\eta}}{R_2(q\rho^{-\epsilon^o},\epsilon^o)^{\eta^o}} = \lim_{q\downarrow 0}\frac{\left((1-\rho)\frac{q^K}{q\rho^{-\epsilon}} + \frac{\epsilon}{1+\epsilon}\left(1+\frac{\rho^{\epsilon+1}}{\epsilon}\right)\right)^{\eta}}{\left((1-\rho)\frac{q^K}{q\rho^{-\epsilon^o}} + \frac{\epsilon^o}{1+\epsilon^o}\left(1+\frac{\rho^{\epsilon^o+1}}{\epsilon^o}\right)\right)^{\eta^o}} = \tag{41}$$

$$= \frac{\left((1-\rho)\frac{q^K}{\rho^{-\epsilon}} + \lim_{q\downarrow 0} q\frac{\epsilon}{1+\epsilon}\left(1+\frac{\rho^{\epsilon+1}}{\epsilon}\right)\right)^{\eta}}{\left((1-\rho)\frac{q^K}{\rho^{-\epsilon^o}} + \lim_{q\downarrow 0} q\frac{\epsilon^o}{1+\epsilon^o}\left(1+\frac{\rho^{\epsilon^o+1}}{\epsilon^o}\right)\right)^{\eta^o}}\lim_{q\downarrow 0} q^{\eta^o-\eta}. \tag{42}$$

The first factor is a constant. Since by Lemma 6 $\eta \neq \eta^o$, the second factor is ∞ or 0. Therefore, the right-hand side in Equation (40) is ∞ or 0, which gives a contradiction. It follows that

[44] Note that $\underline{q} < \hat{q}(\epsilon) < \overline{b}$, $\underline{q_l} < \hat{q}(\epsilon)\rho^{-\epsilon} < \overline{q_l}$, and $\underline{q_l} < \hat{q}(\epsilon)\rho^{-\epsilon^o} < \overline{q_l}$. Therefore, all relations used are in the relevant range. The three inequalities follows since: by assumption, $\frac{\overline{q_l}}{q^K} > \rho^{-\epsilon^o}$; by the property of the convergence radius of power series $\frac{\overline{q_l}}{q^K} = \frac{q^K}{\underline{q_l}}$; by the definition of $R_2(.)$, $q^K \leq \hat{q}(\epsilon)\rho^{-\epsilon} \leq q^K\rho^{-\epsilon}$; by Equation (20) $\overline{b}\rho^{-\epsilon} < \overline{q_l}$; combining these inequalities renders the result.

Condition (a) holds.

qed.

PROOF, Point (g), Lemma 5: 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 $[0, \overline{q_l})$, $\frac{d^P f_l^o(0)}{dq^P} \in \mathbb{R}$ 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 (f) is satisfied.

qed.

PROOF, Point (h), Lemma 5: Suppose Condition (a) does not hold. It follows that there exists an $\epsilon \neq \epsilon^o$ and a η such that

$$f_l^o(q\rho^{-\epsilon^o})\rho^{-\epsilon^o} R_2(q\rho^{-\epsilon^o}, \epsilon^o)^{\eta^o} = f_l(q\rho^{-\epsilon})\rho^{-\epsilon} R_2(q\rho^{-\epsilon}, \epsilon)^{\eta} \text{ for all } q \in (\underline{q}, \overline{b}). \tag{43}$$

Using logarithms, it follows that

$$\ln f_l^o(q\rho^{-\epsilon^o}) + \ln \rho^{-\epsilon^o} + \eta^o \ln R_2(q\rho^{-\epsilon^o}, \epsilon^o) = \ln f_l^o(q\rho^{-\epsilon}) + \ln \rho^{-\epsilon} + \eta \ln R_2(q\rho^{-\epsilon}, \epsilon) \text{ for all } q \in (\underline{q}, \overline{b}). \tag{44}$$

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

$$0 = \ln f_l^o(qa) - \ln f_l^o(q) + \ln a + \eta \ln R_2(qa, \epsilon) - \eta^o \ln R_2(q, \epsilon^o) \text{ for all } q \in (\underline{q}\rho^{-\epsilon^o}, \overline{b}\rho^{-\epsilon^o}). \tag{45}$$

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

$$\begin{aligned} \eta \ln R_2(qa, \epsilon) &= \eta \ln \left((1-\rho)\frac{q^K}{qa} + \frac{\epsilon}{1+\epsilon}\left(1 + \frac{\rho^{\epsilon+1}}{\epsilon}\right)\right) = \\ &= -\eta(\tilde{q} + \tilde{a} - \ln(1-\rho)) + \eta \ln\left(1 + e^{\tilde{q}+\tilde{a}-\ln(1-\rho)}\tilde{g}(\epsilon)\right), \end{aligned} \tag{46}$$

where to shorten the notation $\tilde{g}(\epsilon) = \frac{\epsilon}{1+\epsilon}\left(1 + \frac{\rho^{\epsilon+1}}{\epsilon}\right)$. Use Equations (46) and (28) in Equation (45) to get that

$$\begin{aligned} 0 = \sum_{p=0}^{P} \gamma_p^o \left[(\tilde{q} + \tilde{a})^p - \tilde{q}^p\right] + \tilde{a} - \eta(\tilde{q} + \tilde{a} - \ln(1-\rho)) + \eta \ln\left(1 + e^{\tilde{q}+\tilde{a}-\ln(1-\rho)}\tilde{g}(\epsilon)\right) + \\ + \eta^o(\tilde{q} - \ln(1-\rho)) - \eta^o \ln\left(1 + e^{\tilde{q}-\ln(1-\rho)}\tilde{g}(\epsilon^o)\right) \end{aligned} \tag{47}$$

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 $\tilde{q}$:

$$\eta - \eta^o = \sum_{p=1}^{P} \gamma_p^o p \left[(\tilde{q} + \tilde{a})^{p-1} - \tilde{q}^{p-1} \right] + \eta \frac{e^{\tilde{q}+\tilde{a}-\ln(1-\rho)} \tilde{g}(\epsilon)}{1 + e^{\tilde{q}+\tilde{a}-\ln(1-\rho)} \tilde{g}(\epsilon)} - \eta^o \frac{e^{\tilde{q}-\ln(1-\rho)} \tilde{g}(\epsilon^o)}{1 + e^{\tilde{q}-\ln(1-\rho)} \tilde{g}(\epsilon^o)} \tag{48}$$

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

$$\eta - \eta^o = \lim_{\tilde{q} \to -\infty} \sum_{p=1}^{P} \gamma_p^o p \left[(\tilde{q} + \tilde{a})^{p-1} - \tilde{q}^{p-1} \right] \tag{49}$$

As a first case, consider $P = 1$. The right-hand side is zero, while, by Lemma 6 $\eta \neq \eta^o$, 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

$$\begin{aligned} &\eta - \eta^o = \lim_{\tilde{q} \to -\infty} \sum_{p=2}^{P} \gamma_p^o p \left[(\tilde{q} + \tilde{a})^{p-1} - \tilde{q}^{p-1} \right] = \\ &\lim_{\tilde{q} \to -\infty} \left(\gamma_P^o P \tilde{q}^{P-1} + \gamma_P^o P(P-1) \tilde{q}^{P-2} \tilde{a} + \gamma_{P-1}(P-1) \tilde{q}^{P-2} - \gamma_P^o P \tilde{q}^{P-1} - \gamma_{P-1}^o (P-1) \tilde{q}^{P-2} \right) = \\ &\lim_{\tilde{q} \to -\infty} \gamma_P^o P(P-1) \tilde{q}^{P-2} \tilde{a} = \pm\infty, \end{aligned} \tag{50}$$

since $\tilde{a} \neq 0$, which gives a contradiction. As a last case, consider $P = 2$. Rearrange Equation (48) to

$$0 = \gamma_2^o 2\tilde{a} - \eta \frac{1}{1 + e^{\tilde{q}+\tilde{a}-\ln(1-\rho)} \tilde{g}(\epsilon)} + \eta^o \frac{1}{1 + e^{\tilde{q}-\ln(1-\rho)} \tilde{g}(\epsilon^o)}. \tag{51}$$

Taking $\tilde{q}$ to ∞ gives a contradiction.
qed.
PROOF, Point (b), Lemma 5: Rearrange Equation (44) to

$$\ln f_l^o(q) = \ln f_l^o(qa) + \ln a + \eta \ln R_2(qa, \epsilon) - \eta^o \ln R_2(q, \epsilon^o) \text{ for all } q \in (\underline{q}\rho^{-\epsilon^o}, \overline{b}\rho^{-\epsilon^o}). \tag{52}$$

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

To see that, consider the point $\hat{q}$ such that $\ln R_2(\hat{q}, \epsilon^o) = 0$.[45]Without loss of generality,

[45]Note that $\hat{q} \in (\underline{q}\rho^{-\epsilon^o}, \overline{b}\rho^{-\epsilon^o})$, since by the definition of $R_2(.,.)$, $q^K < \hat{q} < q^K \rho^{-\epsilon^o}$, by assumption, $\frac{\overline{q_l}}{q^K} > \rho^{-\epsilon^o}$, by the property of the convergence radius of power series $\frac{\overline{q_l}}{q^K} = \frac{q^K}{\underline{q_l}}$, by definition $\underline{q} = \underline{q_l}$.

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

$$\eta = \frac{-\ln f_l^o(\hat{q}a) - \ln a}{\ln R_2(\hat{q}a, \epsilon)}. \tag{53}$$

Next, consider the points $\hat{q}a^J$, where J is an integer larger than 1. Consider a $(\underline{q}, \overline{b})$ large enough such that at least two such points exist. It follows that, if Equation (44) holds, the value of $f_l^o(.)$ at these points needs to be

$$\ln \tilde{f}_l^o(\hat{q}a^J) = \sum_{j=1}^{J} \left(-\ln a - \frac{-\ln f_l^o(\hat{q}a) - \ln a}{\ln R_2(\hat{q}a, \epsilon)} \ln R_2(\hat{q}a^j, \epsilon) + \eta^o \ln R_2(\hat{q}a^{j-1}, \epsilon^o) \right). \tag{54}$$

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.

B.2 Proof of Lemma 3

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

B.3 Proof of Lemma 4

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 $q \in [\underline{b}, \overline{b}\rho^{-\epsilon^o}]$. Consider a $\tilde{q}$ such that

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

It follows that

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

$$\leq\lim_{n\to\infty}\sum_{p=0}^{\infty}\left|\gamma_p^n-\gamma_p\right|\left(\ln\frac{\tilde{q}}{q^K}\right)^p=0, \tag{55}$$

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

$$\lim_{n\to\infty}\left|\gamma_p^n-\gamma_p\right|\left(\ln\frac{\tilde{q}}{q^K}\right)^p=0\;\forall\,p; \tag{56}$$

$$\left|\gamma_p^n-\gamma_p\right|\left(\ln\frac{\tilde{q}}{q^K}\right)^p\leq 2M\left(\frac{\ln\tilde{q}-\ln q^K}{\ln\overline{q}_l-\ln q^K}\right)^p; \tag{57}$$

$$\sum_{p=0}^{\infty}2M\left(\frac{\ln\tilde{q}-\ln q^K}{\ln\overline{q}_l-\ln q^K}\right)^p=2M\frac{1}{1-\frac{\ln\tilde{q}-\ln q^K}{\ln\overline{q}_l-\ln q^K}}. \tag{58}$$

$Q(.)$ 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.

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 $\widehat{\ln f(q_j)}$ 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 measure $\widehat{f(q_j)}$ at point q_j, where q_j is the location of the bin. Note that the total number of adopters n is the number of adopters located within the bandwidth. In particular, it excludes households or firms that do not adopt. 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, $\widehat{\ln f(q_j)}=\ln\widehat{f(q_j)}+\frac{1}{2\widehat{N}_j}$, where $\widehat{N_j}$ denotes the number of observations in a bin. The second summand reduces the small sample bias of $\widehat{\ln f(q_j)}$ and does not affect consistency.[46] The location and the size of the bins q_j and h_j may be chosen equidistantly or otherwise.[47]

[46]Since $\widehat{N_j}$ follows a Binomial distribution it follows that $\mathbb{E}\left[\widehat{\ln f(q_j)}\right]=\ln f_k^o(q_j)+O\left(\frac{1}{N_j^2}\right)$, where N_j is the expected value of $\widehat{N_j}$.

[47]Since it increases the efficiency of the estimator, the bins are chosen such that the variance of the histogram is approximately constant.

C.2 The asymptotic properties 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. By Proposition 2 and its proof in Section A.4, Condition 3.1" is fulfilled. The next two paragraphs below verify the other conditions. The third paragraph discusses asymptotic normality.

This paragraph shows that Conditions 3.2, 3.3 i), and 3.4 in Chen (2007) are fulfilled. Consider the sieve spaces

$$\Theta_P = \{\theta \text{ s.t.: } 0 \leq \epsilon \leq \overline{\epsilon},\ 0 \leq \eta \leq \overline{\eta},\ |\gamma_p| \leq M(\ln \overline{q_l} - \ln q^K)^{-p}\ \forall p \leq P,\ \gamma_p = 0\ \forall p > P\}, \tag{59}$$

where $\overline{\epsilon} > 0$ and $\overline{\eta} > 0$ are upper bounds on ϵ and η. 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 $\widehat{Q_n}(\theta)$ (17) as:

$$\widehat{Q}_n(\theta) = \frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \overline{b}]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \theta)\right)^2 + \left(\widehat{\ln B} - \ln B(\theta)\right)^2. \tag{60}$$

The interval $[\underline{b}, q_L] \cup [q_H, \overline{b}]$ 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. Consider $|\widehat{Q_n}(\theta) - Q(\theta)| = |\frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \overline{b}]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \theta)\right)^2 - \int_{[\underline{b}, q_L] \cup [q_H, \overline{b}]} (\ln f_k^o(q) - \ln f_k(q \mid \theta))^2 dF^w + \left(\widehat{\ln B} - \ln B(\theta)\right)^2 - (\ln B^o - \ln B(\theta))^2|$. Subtract and add $\frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \overline{b}]} (\ln f_k^o(q_j) - \ln f_k(q_j \mid \theta))^2$ within this absolute value, use the triangle inequality, and take the limit:

$$\lim_{n\to\infty} |\widehat{Q_n}(\theta) - Q(\theta)| \leq \lim_{n\to\infty} |\frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \overline{b}]} \left[\left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \theta)\right)^2 - (\ln f_k^o(q_j) - \ln f_k(q_j \mid \theta))^2\right]| +$$
$$+ \lim_{n\to\infty} |\int_{[\underline{b}, q_L] \cup [q_H, \overline{b}]} (\ln f_k^o(q) - \ln f_k(q \mid \theta))^2 dF_N^w - \int_{[\underline{b}, q_L] \cup [q_H, \overline{b}]} (\ln f_k^o(q) - \ln f_k(q \mid \theta))^2 dF^w|$$
$$+ \lim_{n\to\infty} |\left(\widehat{\ln B} - \ln B(\theta)\right)^2 - (\ln B^o - \ln B(\theta))^2|. \tag{61}$$

The second line uses
$\frac{1}{N} \sum_{q_j \in [\underline{b}, q_L] \cup [q_H, \overline{b}]} (\ln f_k^o(q_j) - \ln f_k(q_j \mid \theta))^2 = \int_{[\underline{b}, q_L] \cup [q_H, \overline{b}]} (\ln f_k^o(q) - \ln f_k(q \mid \theta))^2 dF_N^w$, 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 (61) 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).[48] The expression in line 4 converges to zero almost surely. 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.)

This paragraph discusses asymptotic normality. Consider the case when P in Condition (15) is finite and ϵ^o and η^o are strictly positive.[49] The estimator (17) is a nonlinear least square estimator. It is asymptotically normal, and the bootstrap is valid.

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 A.7), I use an estimate of its mean squared error to choose the specification. The estimate $\hat{\eta}(P, n)$ is a function of the specification parameter and the sample size n. The mean squared error is defined as $MSE = \mathbb{E}\left[(\hat{\eta}(P, n) - \eta^o)^2\right]$, where η^o denotes the true value of the parameter. A standard bias-variance decomposition renders

$$MSE = \underbrace{\mathbb{E}\left[(\hat{\eta}(P, n) - \mathbb{E}\left[\hat{\eta}(P, n)\right])^2\right]}_{\text{variance}} + \underbrace{(\mathbb{E}\left[\hat{\eta}(P, n)\right] - \eta^o)^2}_{\text{bias}^2}. \tag{62}$$

Define $\tilde{\eta}(P) := \lim_{n\to\infty} \hat{\eta}(P, n)$, where P is kept constant.[50] 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), I estimate the variance by 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

[48] Note that by Assumption 3 or Lemma 1, $\mid \ln f_k^o(q_j) \mid$ is bounded uniformly.

[49] I leave a generalisation of this result to $\lim_{n\to\infty} P = \infty$ for future research.

[50] 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.

is narrow (see Footnote 19 and Appendix D.5.8.) 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. By a slight abuse of notation, 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 $[\underline{b}, \bar{b}]$ 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 $\lim_{q\downarrow q^K} \ln f_k(q|\epsilon^o, \eta^o, \gamma^o) \approx \lim_{q\downarrow q^K} \ln f_k(q|\tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P)$. Using Equation (10) renders $\sum_{p=0}^{\infty} \gamma_p^o \frac{1}{p!} \left(\ln\left(\frac{q^K \rho^{-\epsilon^o}}{q^K}\right)\right)^p + \eta^o \ln R(q^K \rho^{-\epsilon^o}, \epsilon^o) + \ln\left(\rho^{-\epsilon^o}\right)$
$\approx \sum_{p=0}^{P} \tilde{\gamma}_p \frac{1}{p!} \left(\ln\left(\frac{q^K \rho^{-\tilde{\epsilon}}}{q^K}\right)\right)^p + \tilde{\eta} \ln R(q^K \rho^{-\tilde{\epsilon}}, \tilde{\epsilon}) + \ln\left(\rho^{-\tilde{\epsilon}}\right)$. Use $\gamma_P^o \approx \tilde{\gamma}_P$, $\epsilon^o \approx \tilde{\epsilon}$, and rearrange to derive that

$$\tilde{\eta} - \eta^o \approx \frac{\sum_{p=P+1}^{\infty} \gamma_p^o \frac{1}{p!} \left(\ln\left(\rho^{-\epsilon^o}\right)\right)^p}{\ln R(q^K \rho^{-\epsilon^o}, \epsilon^o)} \tag{63}$$

The bias in $\hat{\eta}$ depends on ϵ^o and on the un-estimated rest of the parameter γ^o, i.e., $(\gamma_{P+1}^o, \gamma_{P+2}^o, ...)$. 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. In the untreated data, the true 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 $\widehat{bias(\hat{\eta}(P, n))} = \hat{\eta}_{ut}(P, n) - \eta_{ut}^o = \hat{\eta}_{ut}(P, n)$, where $\hat{\eta}_{ut}$ is the estimate of the participation margin in the untreated data.[51] 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 (63) 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, $\ln f_{l,ut}^o(q) = \sum_{p=0}^{\infty} \gamma_{p,ut}^o \frac{1}{p!} \left(\ln\left(\frac{q}{q^K}\right)\right)^p$, and $\ln f_{l,tr}^o(q) = \sum_{p=0}^{\infty} \gamma_{p,tr}^o \frac{1}{p!} \left(\ln\left(\frac{q}{q^K}\right)\right)^p$, 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

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

are equal: i.e., $\gamma^o_{p,tr} = \gamma^o_{p,ut},\ \forall p \geq p^*$. Furthermore, assume that $P \geq p^*$.

Second, Equation (63) shows that the bias depends on the intensive margin response ϵ^o. 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 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.2.1. A discussion of the selected specification is in Section D.2.2.

C.4 An alternative method for selecting the specification

The method described in Section C.3 selects the specification based on the mean squared error of the parameters of interest. However, the method relies on the availability of untreated data. In some applications, such untreated data may not be available. Moreover, one may want to use the untreated data for robustness checks and not for the selection of the specification. Therefore, this section proposes an alternative method to select the specification. It will use block cross-validation, which is frequently used in time series analysis and performs well in similar estimation exercises (e.g., see Bergmeir and Benítez, 2012 and Roberts et al., 2017).

The procedure is as follows. Split the data into 4 blocks (i.e., intervals). Estimate the model leaving out one block and validate using the left-out block. Repeat this step for all blocks. Choose the specification P for which the model's estimated integrated mean squared error is at its minimum.

D Details main empirical application

D.1 The selected specification for the estimates in Section 4.2

The selected bandwidth for the estimates at 30 kWp in Table 1 is $[\underline{b}, \overline{b}] = [10.6, 85]$; the selected order of the series is $P = 1$; the bunching interval is $[q_L, q_H] = [26.5, 31.75]$. The selected bandwidth for the estimates at 100 kWp is $[\underline{b}, \overline{b}] = [42, 400]$; the selected 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.2 and D.5.8. At both kink points, the selected bandwidth is relatively large, and the series' selected 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 100 kWp and 250 kWp. An exponential transformation constrains $\hat{\epsilon}$ and $\hat{\eta}$ to positive values.

D.2 The selection of the specification

D.2.1 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. 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).[52] Section D.2.2 discusses the selected specification.

D.2.2 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. In order 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]}. I select the specification where the estimated MSE is at its minimum. They are: $[10.6, 85]$ and $P = 1$ at 30 kWp, $[42, 400]$ and $P = 1$ at 100 kWp.

D.3 A discussion of the empirical results

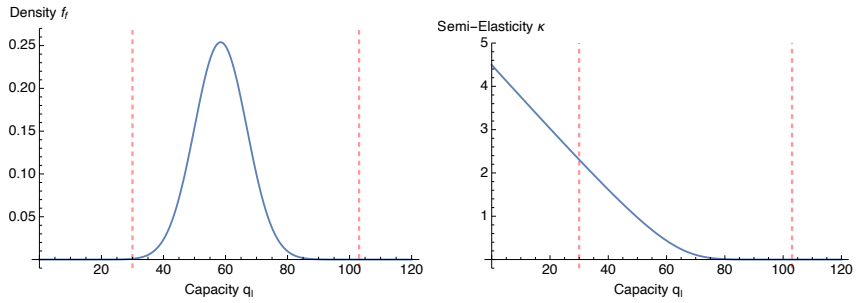
The results in Table 1 suggest that the intensive margin elasticity is the same for adopters of different capacities, while the participation margin semi-elasticity decreases with capacity. This pattern is not surprising. An important intensive adjustment margin is the quality of the solar panel. Low-capacity 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

[52] 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. 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.

responses. A participation semi-elasticity that decreases in capacity is consistent with a simple normal distribution of fixed costs. Denote the density and CDF of fixed costs by $f_f(.)$ and $F_f(.)$. The two semi-elasticities observed at the two capacity levels in Table 1 are sufficient to calibrate the distribution. Figure 16 shows the calibrated distribution and the implied semi-elasticities. The figure shows the calibrated density of fixed costs $f_f(\pi_v(q_l))$ and the semi-elasticity of participation $\kappa(\pi_v(q_l)) = \frac{f_f}{F_f}(\pi_v(q_l))$ as a function of the counterfactual variable profit $\pi_v(q_l)$ at capacity q_l.[53] 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 D.4.3 discusses the calibration in detail.

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



Note: The figure on the left shows the density of fixed costs $f_f(\pi_v(q_l))$ as a function of the counterfactual variable profit π_v at capacity q_l. 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.

Figure 16 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.[54] 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 the left of the bell curve in Figure 16. This feature explains why, compared to mature products, the participation semi-elasticities are high. Yearly estimates at 30 kWp show that,

[53] I.e., $\pi_v(q_l) = s_l q_l - c(q_l, q_l, c_t) + c_t = s_l q_l/(1+\epsilon)$; it increases in q_l.

[54] See Figure 3 in Section 2.1 for an graphical illustration or footnote 53 for a mathematical derivation.

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.

D.4 Policy evaluation

D.4.1 Additional assumptions and calibration

Like other kink and discontinuity estimators in the literature, Section 4 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 4 (Isoelastic cost function)**.** *The cost function is isoelastic with a constant intensive margin elasticity ϵ: $\tilde{c}(q, q_l, c_f) = \frac{s_l}{q_l^{\frac{1}{\epsilon}}} \frac{\epsilon}{1+\epsilon} q^{1+\frac{1}{\epsilon}} + c_f$, where (q_l, c_f) is the two-dimensional type-parameter.*

Note that $\tilde{c}(.,.,.)$ in Assumption 4 is a function of c_f, while $c(.,.,.)$ in Equation 6 is a function of c_t. As discussed in Section D.3, the estimated participation margin responses in Table 1 are in line with a normally distributed fixed cost:

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

Assumption 5 implies c_f and q_l are independent, as in Rochet and Stole (2002). The footnote below generalises the assumption to allow for correlation between c_f and q_l.[55] 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 16 and Appendix D.4.3 for details). Appendix D.4.3 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.

D.4.2 The linear subsidy

This section solves for a linear subsidy rate ρ_l that incentivises the adoption of the same aggregate capacity as the observed kinked subsidy.[56] By the first order condition of the agents'

[55] Assumption 5 implies that q_l and c_f are independent. However, the assumption can be easily extended to allow for a correlation between q_l and c_f. Assume c_f follows a truncated normal distribution: $c_f \sim N(\mu_f, \sigma_f, \underline{c_f}(q_l), \overline{c_f}(q_l))$, where $\underline{c_f}(q_l)$, $\overline{c_f}(q_l)$ are the truncation bounds which can vary with q_l. Denote by $F_f(.)$ the CDF of the normal distribution. It follows that the CDF of c_f is $\frac{F(c_f)}{F(\overline{c_f}(q_l)) - F(\underline{c_f}(q_l))}$; hence, c_f and q_l 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 D.4 remain unchanged.

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

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 $\tilde{c}_v(q, q_l)$ the variable part of the cost function in Assumption 4: $\tilde{c}_v(q, q_l) = \tilde{c}(q, q_l, c_f) - c_f$. Denote the variable profit of type q_l under subsidy rate ρ_l as $\pi_v(q_l, \rho_l) = \rho_l q(q_l, \rho_l) - \tilde{c}_v(q(q_l, \rho_l), q_l)$. Given the estimate of $f_l(q_l)$, the unconditional type distribution $f_u(q_l)$ is $f_u(q_l) = \frac{f_l(q_l)}{F_f(q_l - \tilde{c}_v(q_l, q_l))}$. It follows that ρ_l is the solution to $\int q(q_l, \rho_l) F_f(\pi_v(q_l, \rho_l)) f_u(q_l) dq_l = Q^T$ 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.

D.4.3 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 $\rho_2 = 0.99$. $S_k(q) = q,$ for $q \leq q_1^K$; $S_k(q) = q_1^K(1 - \rho_1) + q\rho_1,$ for $q \in (q_1^K, q_2^K]$; $S_k(q) = q_1^K(1 - \rho_1) + q_2^K \rho_1 (1 - \rho_2) + q\rho_1\rho_2,$ for $q > q_2^K$. Using Assumption 4 , the choice q as a function of type q_l is: $q(q_l) = q_l,$ for $q_l \in [q_l^{min}, q_1^K]$; $q(q_l) = q_1^K,$ for $q_l \in [q_1^K, q_1^K \rho_1^{-\epsilon}]$; $q(q_l) = q_l \rho_1^\epsilon,$ for $q_l \in [q_1^K \rho_1^{-\epsilon}, q_2^K \rho_1^{-\epsilon}]$; $q(q_l) = q_2^K,$ for $q_l \in [q_2^K \rho_1^{-\epsilon}, q_2^K (\rho_1\rho_2)^{-\epsilon}]$; $q(q_l) = q_l (\rho_1\rho_2)^\epsilon,$ for $q_l \in [q_2^K (\rho_1\rho_2)^{-\epsilon}, q_l^{max}]$, 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

$$\frac{f_f}{F_f}\left(S_k(q) - \tilde{c}_v(q, q_l(q)) | \mu_f, \sigma_f\right) = \kappa(q),$$

where $q_l(q)$ denotes the inverse of $q(q_l)$ and $\tilde{c}_v(q, q_l)$ denotes the variable part of the cost function in Assumption 4.

Using Assumptions 4, 5, the results in Table 1, and inverting this equation at the two kink points gives (μ_f, σ_f).[57] Figure 16 shows the calibrated density of fixed costs and the implied semi-elasticities of participation at counterfactual capacities q_l. Note that the two red lines illustrate the counterfactual capacity with corresponding variable profit as under the observed subsidy, i.e., q_l solves $q_l - \tilde{c}_v(q_l, q_l) = S(q^K) - \tilde{c}_v(q^K, q_l(q^K))$. Because at the first kink point $q = q_l$, the first red line is at $q_1^K = 30$, while the second red line is slightly above $q_2^K = 100$.

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

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

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

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

distribution $f_k(.)$ is

$$f_k(q) = f_l(q_l(q)) \frac{F_f(S_k(q) - \tilde{c}_v(q, q_l(q))}{F_f(q_l(q) - \tilde{c}_v(q_l(q), q_l(q)))} \frac{dq_l}{dq}.$$

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 3. 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 $f_l(.)$.

D.4.4 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:

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

where $q(.)$ and $S_k(.)$ are defined in Section D.4.3, $f_u(.)$ and $\tilde{c}_v(.,.)$ is defined in Section D.4.2, 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.

D.5 Robustness checks

D.5.1 The robustness check at 100 kWp

Table 6: 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.

I estimate the model selected for observations at 100 kWp on observations at 250 kWp, which do not face a kink in the incentive scheme. For a discussion of this choice of data, see Section D.2.1. Both estimates in Table 6 are insignificant.

D.5.2 An estimator of the specification bias

This section proposes an estimator of the specification bias caused by an eventual violation of Assumption 3. To this end, consider an estimator similar to Section 3, but with fixed polynomial order P and a bandwidth $[\underline{b}, \bar{b}]$ that converges to q^K as n goes to infinity. It iteratively minimises:

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

$$\hat{\epsilon} = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln B(\epsilon, \hat{\eta}, \hat{\gamma_P}) \right)^2. \tag{64}$$

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 3. *Assume bunching is sharp,* $\ln f_l^o(q_l)$ *is* P *times differentiable in* $\ln q_l$ *at point* $\ln q^K$*, and* $\tilde{\epsilon} = \epsilon^o$.[58] *It follows that the bias*

$$|\tilde{\eta} - \eta^o| = \left| \frac{h(q^K \rho^{-\epsilon^o}) \left(\ln \left(\frac{q^K \rho^{-\epsilon^o}}{q^K} \right) \right)^P}{\ln R(q^K \rho^{-\epsilon^o}, \epsilon^o)} \right|, \tag{65}$$

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.5.3. The bias of $\hat{\eta}$ depends on ϵ^o and on the rest of the Taylor series $h(.)$. The formula motivates the estimation of the bias using untreated 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. Denote by subscripts ut and tr variables in the untreated and treated data. Equation (65) shows that the bias depends on the intensive margin response ϵ^o. To consider this dependence, simulate the intensive margin response in the untreated data by shifting observations using the estimate $\hat{\epsilon}_{tr}$ in the treated data. Then estimate $\hat{\eta}_{ut}$ in the untreated data using the same specification as in the treated data.[59] In

[58] Note that in the application of this paper, bunching is relatively sharp, and the bunching interval is narrow (see Footnote 19 and Appendix D.5.8). I leave a generalisation to non-sharp bunching for future research. Also note that the simulation in Table 4 in Section A.7 shows that the parameter ϵ is less sensitive to the specification than the parameter η, which motivates assuming $\tilde{\epsilon} = \epsilon^o$. I leave a generalisation to non-sharp bunching for future research.

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

Table 7: 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.

the untreated data, the true value η_{ut}^o is known: it is equal to zero because there is no treatment. Therefore, by Proposition 3

$$|\tilde{\eta}_{ut} - \eta_{ut}^o| = |\tilde{\eta}_{ut}| = |\frac{h_{ut}(q^K \rho^{-\epsilon^o}) \left(\ln\left(\frac{q^K \rho^{-\epsilon^o}}{q^K}\right)\right)^P}{\ln R(q^K \rho^{-\epsilon^o}, \epsilon^o)}|, \tag{66}$$

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 3 that $|\tilde{\eta}_{ut}| = |\tilde{\eta}_{tr} - \eta_{tr}^o|$. 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.2.1. The order P is the one used in the main estimation, selected by the method in Appendix C.4 using only treated observations from 2004-2008. The estimates of the bias are in Table 7. The point-estimates are small and statistically insignificant.

D.5.3 Proof of Proposition 3

By Taylor's Theorem

$$\ln f_l^o(q_l) = \sum_{p=0}^{P} \gamma_p^o \left(\ln \frac{q_l}{q^K}\right)^p + h(q_l) \left(\ln \frac{q_l}{q^K}\right)^P, \text{ where } \lim_{q_l \to q^K} h(q_l) = 0. \tag{67}$$

Following the corresponding steps of the proof in Section C.2, the first and third part of the sample criterion (64) converges to the population criterion

$$\min_{\tilde{\eta}, \tilde{\gamma_P}} \lim_{n \to \infty} \int_{\underline{b}(n)}^{q^K} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)\right)^2 \, dF^w + \int_{q^K}^{\bar{b}(n)} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)\right)^2 \, dF^w. \tag{68}$$

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 (16). The function F^w denotes

a known weighting measure.

Step 1: $\tilde{\gamma}_0 = \gamma_0$. Differentiate Equation (68) with respect to $\tilde{\gamma}_0$ and assume $\bar{b}$ converges sufficiently faster to q^K than $\underline{b}$. It follows that

$$0 = \lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)\right) \frac{\frac{\partial f_k(q|\tilde{\epsilon},\tilde{\eta},\tilde{\gamma}_P)}{\partial \tilde{\gamma}_0}}{f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)} \, dF^w. \tag{69}$$

Use Equation (67), Proposition 1, and Equation (16):

$$0 = \lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \left[\sum_{p=0}^{P} \left(\gamma_p^o - \tilde{\gamma}_p\right) \frac{1}{p!} \left(\ln\left(\frac{q}{q^K}\right)\right)^p + h(q) \left(\ln\left(\frac{q}{q^K}\right)\right)^P\right] \frac{1}{0!} \left(\ln\left(\frac{q}{q^K}\right)\right)^0 dF^w \tag{70}$$

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

$$\tilde{\gamma}_0 - \gamma_0^o = \frac{\lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \left[\sum_{p=1}^{P} \left(\gamma_p^o - \tilde{\gamma}_p\right) \frac{1}{p!} \left(\ln\left(\frac{q}{q^K}\right)\right)^p + \; h(q) \left(\ln\left(\frac{q}{q^K}\right)\right)^P\right] \frac{1}{0!} \left(\ln\left(\frac{q}{q^K}\right)\right)^0 dF^w}{\lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \frac{1}{0!0!} \left(\ln\left(\frac{q}{q^K}\right)\right)^{2*0} dF^w} \tag{71}$$

Take the limit using l'Hopital:

$$\tilde{\gamma}_0 - \gamma_0^o = \frac{\sum_{p=1}^{P} \left(\gamma_p^o - \tilde{\gamma}_p\right) \frac{1}{p!} \left(\ln\left(\frac{q^K}{q^K}\right)\right)^p + h(q^K) \left(\ln\left(\frac{q^K}{q^K}\right)\right)^P}{1} = 0 \tag{72}$$

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 (68) with respect to $\tilde{\gamma}_\pi$ and assume $\bar{b}$ converges sufficiently faster to q^K than $\underline{b}$. It follows that

$$0 = \lim_{n\to\infty} \int_{\underline{b}(n)}^{q^K} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)\right) \frac{\frac{\partial f_k(q|\tilde{\epsilon},\tilde{\eta},\tilde{\gamma}_P)}{\partial \tilde{\gamma}_\pi}}{f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)} \, dF^w \tag{73}$$

Use Equation (67), Proposition 1, Equation (16), $\gamma_p^o = \tilde{\gamma}_p$ for all $p < \pi$, rearrange and use l'Hopital to derive that

$$\tilde{\gamma}_\pi - \gamma_\pi^o = \lim_{q\to q^K} \frac{\sum_{p=\pi+1}^{P} \left(\gamma_p^o - \tilde{\gamma}_p\right) \frac{1}{p!} \left(\ln\left(\frac{q}{q^K}\right)\right)^p + h(q) \left(\ln\left(\frac{q}{q^K}\right)\right)^P}{\frac{1}{\pi!} \left(\ln\left(\frac{q}{q^K}\right)\right)^\pi} = 0. \tag{74}$$

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

Step 3: $\tilde{\eta} - \eta^o$. Differentiate Equation (68) with respect to $\tilde{\eta}$:

$$0 = \lim_{n\to\infty} \int_{q^K}^{\bar{b}(n)} \left(\ln f_k^o(q) - \ln f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)\right) \frac{\frac{\partial f_k(q|\tilde{\epsilon},\tilde{\eta},\tilde{\gamma}_P)}{\partial \tilde{\eta}}}{f_k\left(q \mid \tilde{\epsilon}, \tilde{\eta}, \tilde{\gamma}_P\right)} \, dF^w \tag{75}$$

Assume $\tilde{\epsilon} = \epsilon^o$, use $\gamma_p^o = \tilde{\gamma}_p$, Equation (67), Proposition 1, and Equation (16):

$$0 = \lim_{n\to\infty} \int_{q^K}^{\bar{b}(n)} \left[h(q\rho^{-\epsilon^o}) \left(\ln\left(\frac{q\rho^{-\epsilon^o}}{q^K}\right)\right)^P + (\eta^o - \tilde{\eta}) \ln R\left(q\rho^{-\epsilon^o}\right)\right] \ln R\left(q\rho^{-\epsilon^o}\right) \, dF^w \tag{76}$$

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

$$\tilde{\eta} - \eta^o = \frac{h(q^K\rho^{-\epsilon^o}) \left(\ln\left(\frac{q^K\rho^{-\epsilon^o}}{q^K}\right)\right)^P}{\ln R(q^K\rho^{-\epsilon^o}, \epsilon^o)} \tag{77}$$

D.5.4 Identification and estimation using additional observations

In many applications, additional observations that are exposed to a linear budget set are available. For example, it is the case if observations before the introduction of a kink in a budget set are available or a second contemporaneous cross-section of observations faces linear incentives. This section shows that Assumption 3 can be considerably weakened in this case. Denote by $f_{l1}(.)$ the observable measure of observations facing a linear budget set. I will refer to these observations as dataset 1. Denote by $f_{k2}(.)$ the observable measure of observations facing the kink budget set, which I will refer to as dataset 2. Denote by $f_{l2}(.)$ the counterfactual measure of observations in dataset 2 without the kink. Consider the following assumption:

Assumption 6. *Assume that* $\ln f_{l2}(q) = \ln f_{l1}(q) + \sum_{g=0}^{G^*} \lambda_g \left(\ln \frac{q}{q^K}\right)^g$, *where* G^* *is finite but unknown to the econometrician. Moreover, assume that* $\ln f_{l1}(.)$ *fulfils the regularity conditions in Newey (1997).*

Intuitively, this assumption states that absent the kink, the distributions of agents in datasets 1 and 2 are similar but can differ in level, slope, curvature, etc. Note that this assumption is weaker than Assumption 3, since, if $\ln f_{l1}(.)$ is not analytic, $\ln f_{l2}(.)$ is not analytic. Note also that the assumption does not impose restrictions on the budget sets. In particular, the linear counterfactual rates may vary between them.

Proposition 4 (Identification)**.** *Under Assumptions 1, 2, 6 and one of the conditions in Lemma 5, the observable measures* $f_{l1}(.)$ *and* $f_{k2}(.)$ *identify the counterfactual measure* $f_{l2}(.)$*, the intensive margin elasticity* ϵ*, and the participation margin elasticity* η*.*

Table 8: The estimates at 30 and 100 kWp.

Capacity	$\hat{\epsilon}$ (SD)	$\hat{\kappa}$ (SD)
30 kWp	4.37 (0.13)	2.31 (0.06)
100 kWp	5.12 (0.92)	0.00 (0.01)

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.

Several papers in the bunching literature have used the assumption $f_{l1}(.) = f_{l2}(.)$ for identifying the intensive margin elasticity (see Bertanha et al. (2023) Section 2.3.2 for a review). Assumption 6 and Proposition 4 generalise this approach since the measure's level, slope, curvature, etc. can vary between the two datasets. This generalisation is relevant in practice. For example, dataset 1 may be observations before the introduction of the kink, but other factors driving demand or supply change over time. In particular, this is the case in the main application of this paper discussed below, where demand for solar panels increases over time.

For estimation, I use a two-step series estimator. In the first step, I estimate $\ln f_{l1}(.)$ in dataset 1 using the series $\ln f_{l1}(q) = \sum_{p=0}^{P} \gamma_p \left(\ln \frac{q}{q^K}\right)^p$, where P goes to infinity as n goes to infinity and is selected by block cross-validation described in Section C.4. Note that this estimator is consistent under the conditions lined out in Newey (1997). Analyticity is not necessary for consistency since $\ln f_{l1}(.)$ is observable for all q. In the second step, use the series $\ln f_{l2}(q) = \widehat{\ln f_{l1}}(q) + \sum_{g=0}^{G} \lambda_g \left(\ln \frac{q}{q^K}\right)^g$ for the counterfactual measure $\ln f_{l2}(.)$ in the sample criterion (17), where G goes to G^* as n goes to infinity and is selected by block cross-validation. The estimator $\hat{\lambda}$, $\hat{\epsilon}$, and $\hat{\eta}$ are consistent since step 2 is a parametric M-estimator. I estimate the standard errors by nonparametric bootstrap following Chetty et al. (2011).

In my application, I use the data described in Section D.2.1 as dataset 1. The selected orders at 30kWp are $P = 1$ and $G = 1$, and at 100 kWp $P = 1$ and $G = 0$. The point estimates in Table 8 are not significantly different from the main estimates in Table 1.

D.5.5 Proof of Proposition 4

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. The function $\ln f^o_{l1}(.)$ is observable for all q, and, hence, it is identified and can be consistently estimated using the series estimator in Newey (1997). The finite parameter vector $\lambda = (\lambda^o_0, ..., \lambda^o_{G^*})$ is identified from observations of $f^o_{k2}(.)$ to the left of the kink. Therefore, due to Assumption 6, $f^o_{l2}(.)$ is identified for all q. The rest of the proof follows the steps of the proof of Proposition 2 in Section A.4.

D.5.6 Partial Identification

Building on the partial identification results in Bertanha, McCallum and Seegert (2023), this section derives a partial identification result under the assumption that the first derivative of the counterfactual measure is Lipschitz continuous. It 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. Denote by

$$\tilde{\eta} = \frac{\ln f_k^o(q_H) - \ln(\rho^{-\epsilon^o}) - \ln f_k^o(q_L) - \frac{d\ln f_k^o(q_L)}{d\ln q} \ln\left(\frac{q_H\rho^{-\epsilon^o}}{q_L}\right)}{-\ln R(q_H\rho^{-\epsilon^o}, \epsilon^o)}. \tag{78}$$

Note that $\tilde{\eta}$ is equal to the true η^o if the true counterfactual measure $\ln f_l^o(.)$ is linear within the bunching interval.

Proposition 5. *Assume $\frac{d\ln f_l^o(q_l)}{d\ln q_l}$ is Lipschitz continuous with constant C. It follows that*

$$\tilde{\eta} - \frac{C}{2}\frac{\left(\ln\left(\frac{q_H\rho^{-\epsilon^o}}{q_L}\right)\right)^2}{-\ln R(q_H\rho^{-\epsilon^o}, \epsilon^o)} \leq \eta^o \leq \tilde{\eta} + \frac{C}{2}\frac{\left(\ln\left(\frac{q_H\rho^{-\epsilon^o}}{q_L}\right)\right)^2}{-\ln R(q_H\rho^{-\epsilon^o}, \epsilon^o)}. \tag{79}$$

Denote the lower and upper bound of ϵ^o by ϵ_L and ϵ_U. Define upper and lower bounds for the counterfactual as $\ln f_{lL}(q_l) = \ln f_k^o(q_L) + \frac{d\ln f_k(q_L)^o}{d\ln q_l}(\ln q_l - \ln q_L) - \frac{C}{2}(\ln q_l - \ln q_L)^2$ and $\ln f_{lU}(q_l) = \ln f_k^o(q_L) + \frac{d\ln f_k^o(q_L)}{d\ln q_l}(\ln q_l - \ln q_L) + \frac{C}{2}(\ln q_l - \ln q_L)^2$. It follows that ϵ_L and ϵ_U are implicitly defined by

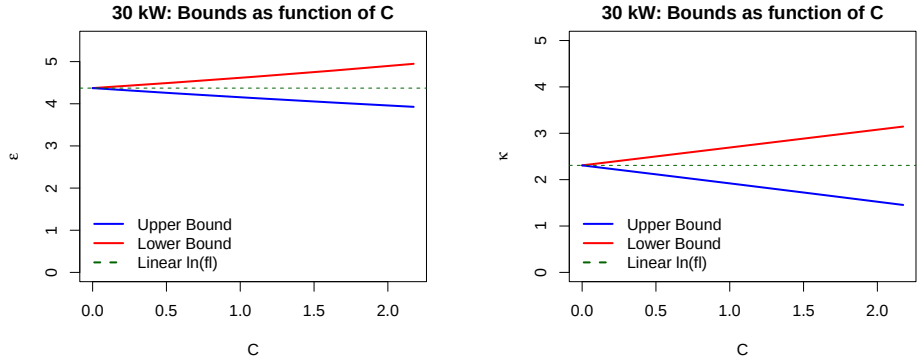
$$B^o = B(\epsilon_L, \eta^o, f_{lU}) \text{ and } B^o = B(\epsilon_H, \eta^o, f_{lL}), \tag{80}$$

where B^o is the observed mass in the bunching interval and $B(\epsilon, \eta, f_l)$ is defined in Equation (13).

The proof is in Section D.5.7. Following Bertanha, McCallum and Seegert (2023) I estimate the Lipschitz constant C as the maximal convexity of $f_l^o(.)$ outside the bunching interval and within the bandwidth.[60] Figures 17 and 18 show the bounds. Note that the bounds of η in Proposition 5 depend on ϵ^o, while the bounds of ϵ depend on η^o. Since the observable moments depend jointly on both margins, it is not possible to express joint bounds of ϵ, η as functions of the Lipschitz constant C and observables. However, as a robustness check, I solve for such joint bounds numerically. The joint bounds lie within or extremely close to the bounds in Figure 17 and 18.

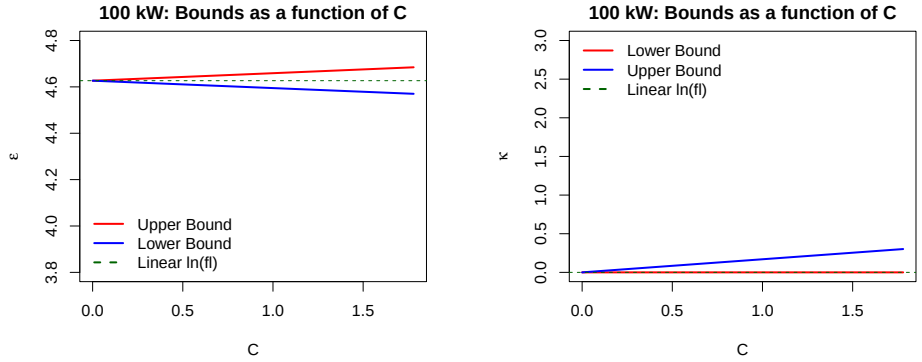
[60] For the bounds of η, I use the estimate $\hat{\eta}$ in Section 3 with $P = 1$ as an estimate for $\tilde{\eta}$ and $\hat{\epsilon}$ as an estimate for ϵ^o. For the bounds of ϵ, I use $\hat{\gamma}$ to calculate $\ln f_k^o(q_L)$ and $\frac{d\ln f_k(q_L)^o}{d\ln q_l}$ and $\hat{\eta}$ as an estimate for η^o.

Figure 17: Bounds of ϵ_{30} (left) and κ_{30} (right) for various choices of maximum convexity



Note: The figure shows the upper and lower bound of ϵ (left) and κ (right) at 30kWp as a function of the Lipschitz constant C.

Figure 18: Bounds of ϵ_{100} (left) and κ_{100} (right) for various choices of maximum convexity



Note: The figure shows the upper and lower bound of ϵ (left) and κ (right) at 100kWp as a function of the Lipschitz constant C.

D.5.7 Proof Proposition 5

Due to Lipschitz-continuity, $|\frac{d\ln f_l^o(q_H\rho^{-\epsilon^o})}{d\ln q} - \frac{d\ln f_l^o(q_L)}{d\ln q}| \leq C\ln\left(\frac{q_H\rho^{-\epsilon^o}}{q_L}\right)$. Integrating from q_L to $q_H\rho^{-\epsilon^o}$ renders $|\ln f_l^o(q_H\rho^{-\epsilon^o}) - \ln f_l^o(q_L) - \frac{d\ln f_l^o(q_L)}{d\ln q}\ln\left(\frac{q_H\rho^{-\epsilon^o}}{q_L}\right)| \leq \frac{C}{2}\left(\ln\left(\frac{q_H\rho^{-\epsilon^o}}{q_L}\right)\right)^2$. Use $f_k^o(q_L) = f_l^o(q_L)$ from Proposition 1, rearrange, and divide by $-\ln R(q_H\rho^{-\epsilon^o}, \epsilon^o)$:

$$\left|\frac{\ln f_l^o(q_H\rho^{-\epsilon^o}) - \ln f_k^o(q_H) + \ln(\rho^{-\epsilon^o})}{-\ln R(q_H\rho^{-\epsilon^o}, \epsilon^o)} + \frac{\ln f_k^o(q_H) - \ln(\rho^{-\epsilon^o}) - \ln f_k^o(q_L) - \frac{d\ln f_k^o(q_L)}{d\ln q}\ln\left(\frac{q_H\rho^{-\epsilon^o}}{q_L}\right)}{-\ln R(q_H\rho^{-\epsilon^o}, \epsilon^o)}\right| \leq$$

$$\leq \frac{C}{2}\frac{\left(\ln\left(\frac{q_H\rho^{-\epsilon^o}}{q_L}\right)\right)^2}{-\ln R(q_H\rho^{-\epsilon^o}, \epsilon^o)}.$$

Define $\tilde{\eta} = \frac{\ln f_k^o(q_H) - \ln(\rho^{-\epsilon^o}) - \ln f_k^o(q_L) - \frac{d \ln f_k^o(q_L)}{d \ln q} \ln\left(\frac{q_H \rho^{-\epsilon^o}}{q_L}\right)}{-\ln R(q_H \rho^{-\epsilon^o}, \epsilon^o)}$, and by Proposition 1, $\eta^o = \frac{\ln f_k^o(q_H) - \ln(\rho^{-\epsilon^o}) - \ln f_l^o(q_H \rho^{-\epsilon^o})}{-\ln R(q_H \rho^{-\epsilon^o}, \epsilon^o)}$. It follows that $|\tilde{\eta} - \eta^o| \leq \frac{C}{2} \frac{\left(\ln\left(\frac{q_H \rho^{-\epsilon^o}}{q_L}\right)\right)^2}{-\ln R(q_H \rho^{-\epsilon^o}, \epsilon^o)}$, which proves (79). Note that by Lipschitz continuity $f_{lL}(q_l) \leq f_l^o(q_l) \leq f_{lU}(q_l)\ \forall q_l \in [q_L, q_H \rho^{-\epsilon^o}]$. The bounds on ϵ follow since the function $B(\epsilon, \eta, f_l)$ is strictly increasing in $f_l(q_l)$ for all q_l and, as shown in the proof of Lemma 6, strictly increasing in ϵ.

D.5.8 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 19, non-sharp bunching can be explained by the unavailability of the exact optimal system size at the purchase date. Table 9 and 10 report the robustness of the estimates to changes in the bunching interval. All estimates are robust.

Table 9: 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.

Table 10: 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.

D.5.9 The empirical verification of the rank conditions

Consider Condition (a) and its local counterpart Condition (i) in Lemma 5 in Section A.5.1. By Lemma 5, both conditions hold generically. Additionally, this section verifies the two conditions in sample and shows that they hold by a large amount. First, consider Condition (i) in Lemma 5. Table 11 shows the result; the standard errors are in brackets. Condition (a) holds

Table 11: The rank condition evaluated at $q = q^K$.

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

Note: The table shows Condition (i) in Lemma 5 evaluated at $q = q^K$. The standard errors are in brackets. The condition holds by a large amount.

since, in sample, there is no evidence of a second local minimum of the sample criterion (17). Moreover, Condition (c) in Lemma 5 holds in sample. Table 12 shows the result; the standard errors are in brackets:

Table 12: Condition (c) in Lemma 5.

Capacity	$\max \frac{d \ln \widehat{f}_l(q_l)}{d \ln q_l}$ (SD)
30 kWp	-1.37 (0.02)
100 kWp	-2.68 (0.03)

Note: The table shows the maximum value of $\frac{d \ln \widehat{f}_l(q_l)}{d \ln q_l}$ in the bandwidth and at the two capacity levels. Standard errors are in brackets. The condition (c) in Lemma 5 holds by a large amount.

D.5.10 Generalisation of Assumption 2

Assumption 2 can be generalised to

Assumption 7 (Generalised assumption participation margin)**.** *For all $q_l \in (\underline{q_l}, \overline{q_l})$ and all $c_t \in [S_k(q_l), S_l(q_l)]$, the conditional CDF of the total cost is equal to*

$$F_{t|q_l}(c_t|q_l) = \left(\frac{c_t}{g(q_l)}\right)^{\eta(q_l)}, \quad \textit{where} \quad \eta(q_l) = \sum_{j=0}^{J} \eta_j \left(\frac{q_l}{q^K}\right)^j \quad \textit{and} \; g(q_l) > S_l(q_l). \tag{81}$$

I estimate this generalized model selecting the orders P and J using the procedure in Section C.3. The selected orders are $P = 1$ and $J = 0$, which is the same specification as used in the main estimation. Therefore, the estimates do not change.

D.6 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 range of agents who bunch sharply. Note that it is not equal to the range of agents in the bunching interval $[q_L, q_H\rho^{-\epsilon}]$ in Equation (13). 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 3 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 3 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_{B1} assuming the counterfactual is constant in the bunching region as in Chetty et al. (2011). In order to isolate the biasing effect of this assumption, I use the rest of the parameters from the main estimation (17) and I do not restrict the counterfactual to a constant outside the bunching range:

$$\epsilon_{B1} = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln B_{B1}(\epsilon, \eta = \hat{\eta}, \gamma = \hat{\gamma})\right)^2, \text{ where} \quad (82)$$

$$B_{B1}(\epsilon, \eta, \gamma) = \int_{q_L}^{q^K} f_l(q_l|\gamma)dq_l + \int_{q^K}^{q^K\rho^{-\epsilon}} f_l(q^K|\gamma)dq_l + \int_{q^K\rho^{-\epsilon}}^{q_H\rho^{-\epsilon}} f_l(q_l|\gamma)R(q_l, \epsilon)^\eta dq_l, \quad (83)$$

and $\hat{\gamma}$ and $\hat{\eta}$ are the estimates from (17). As a comparison, in the my main estimation, the second integral in Equation (83) is

$$\int_{q^K}^{q^K\rho^{-\epsilon}} f_l(q_l|\gamma)R(q_l, \epsilon)^\eta dq_l. \quad (84)$$

Row 1 in Table 13 shows the result. The intensive margin elasticity has a downward bias of 23 %.

Second, to distinguish the bias from assuming the constant counterfactual from the bias of ignoring participation, I estimate an intensive margin elasticity ϵ_{B2} by using the specification of the counterfactual from the main estimation in the bunching range, but by ignoring the participation margin in this range. Again, I use the rest of the parameters from the main estimation

(17), and I do not ignore the participation margin outside the bunching range:

$$\epsilon_{B2} = \arg\min_{\epsilon} \left(\widehat{\ln B} - \ln B_{B2}(\epsilon, \eta = \hat{\eta}, \gamma = \hat{\gamma}) \right)^2, \text{ where} \quad (85)$$

$$B_{B2}(\epsilon, \eta, \gamma) = \int_{q_L}^{q^K} f_l(q_l|\gamma) dq_l + \int_{q^K}^{q^K \rho^{-\epsilon}} f_l(q_l|\gamma) dq_l + \int_{q^K \rho^{-\epsilon}}^{q_H \rho^{-\epsilon}} f_l(q_l|\gamma) R(q_l, \epsilon)^{\eta} dq_l, \quad (86)$$

and $\hat{\gamma}$ and $\hat{\eta}$ are the estimates from (17). Again, as a comparison, in the main specification, the second integral in Equation (86) is Equation (84). Row 2 in Table 13 shows the result. The intensive margin elasticity has a downward bias of 12 %. It reveals that around half of the bias in ϵ_{B1} can be attributed to ignoring the participation margin.[61]

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

$$\eta_B = \arg\min_{\eta} \sum_{q_j \notin [q_L, q_H]} \left(\widehat{\ln f(q_j)} - \ln f_k(q_j \mid \eta, \gamma = \hat{\gamma}, \epsilon = 0) \right)^2. \quad (87)$$

Row 3 in Table 13 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 13: A comparison of the biased and unbiased estimates.

Parameter	Unbiased Estimate	Biased Estimate	Relative Difference in %
ϵ_{B1}	4.37	3.39	-23
ϵ_{B2}	4.37	3.87	-12
κ_B	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.

In order 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 estimation of the counterfactual in Section 3, a common point 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 series; however, they change quantitatively. An advantage of the nonparametric estimator in Section 3 is that the series' order is selected in a data-driven way based on the estimates' mean squared errors. The selected order is low, i.e., $P = 1$. See Section D.1 for a discussion. To illustrate the importance of the choice of P, Table 14 compares the

[61] Note that, for both exercises, I do not use the misspecified model outside the bunching range and inside the bunching interval, i.e., for $q_l \in [q_L, q^K] \cup [q^K \rho^{-\epsilon}, q_H \rho^{-\epsilon}]$. The biases become more severe if one uses the misspecified models for all q in $[q_L, q_H]$.

estimates using the selected specification to those using $P = 7$ like in Chetty et al. (2011). As in Chetty et al. (2011), the estimates do not differ qualitatively. 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 order based on the estimates' mean squared errors.

Table 14: A comparison of the data-driven order of the series and order $P = 7$.

Parameter	Data-Driven Specification	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 data-driven order of the series in the second column. The third column shows the estimates using order $P = 7$.

E Details application to notch in Denmark

The model and assumptions in this section are the same as in Section 2. The variable q denotes gross earnings, $S(q)$ denotes net of tax earnings, $c_v(.,.)$ is the effort cost of earnings, the participation margin is the foreigners' decision to participate in the Danish labour market (i.e., to migrate to Denmark and work there), and c_f is the fixed cost of participation. Before 1991, foreigners in Denmark paid the progressive Danish income tax. After 1991, foreigners with earnings larger than q^D paid a flat tax that was lower than the Danish income tax. For a detailed policy description, see Kleven et al. (2013). Figure 9 in Section 5 illustrates net earning $S_d(q)$ under the tax scheme.[62] Equation (88) in Section E.1 below describes the scheme.

E.1 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. I focus on the case where the marginal and average rates jump upwards to the right of the discontinuity. Derivations for other variations of the notch are equivalent.

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

$$S_d(q) = s_l q + S, \text{ for } q \leq q_D; \text{ and } S_d(q) = s_l \rho q, \text{ for } q > q_D; \tag{88}$$

where q_D denotes the notch point, $\rho > 1$, (i.e., the marginal rate increases at the right of the

[62] Net earnings $S(q)$ are gross earnings q minus the income tax $T(q)$.

discontinuity), and $s_l q^D + S < s_l \rho q^D$ (i.e., the payment is higher to the right of the discontinuity).

Denote the observable measure under the notched scheme by $f_d(.)$. Denote by $f_l(.)$ the counterfactual measure under the counterfactual linear scheme $S_l(q) = s_l q + S$ for all q. Compared to a counterfactual measure $f_l(.)$, the notch has several effects on the observable measure $f_d(.)$. First, agents below the notch bunch at the notch point due to intensive margin responses. This behaviour creates a hole in the distribution below the notch. It is well known in the bunching literature that not all agents below the notch follow this behaviour since they face optimisation frictions (see Kleven, 2016). Therefore, following this literature, I assume a fraction μ of agents is constrained by optimization frictions. The interval $[q_L, q_D]$ denotes the part below the discontinuity containing the missing mass; $[q_D, q_H]$ denotes the bunching interval above the notch. Second, agents above the discontinuity increase q due to intensive margin responses. Third, mass above the notch increases due to participation margin responses. The following proposition describes these effects mathematically using the assumptions in Section 2.

Proposition 6 (The observed measure in the case of a notch)**.** *Under Assumptions 1 and 2, for all $q \in (\underline{q}, \overline{q})$, the observable measure $f_d(.)$ under the notched subsidy $S_d(.)$ is a function of four unknowns: the intensive margin elasticity ϵ, the participation margin elasticity η, the counterfactual measure $f_l(.)$, and the fraction of agents with optimization frictions μ. Above the notch point q_D, there is the bunching mass B. Below the notch point, there is an interval with mass M; it lacks the missing mass. Four parts of the observable measure $f_d(.)$ depend distinctly on the four unknowns:*

$$
\begin{aligned}
f_d(q) =& f_l(q), && \text{for } q < q_L; && (89)\\
M =& \frac{\int_{q_L}^{q_D} f_l(q_l)dq_l + \int_{q_B}^{q_D} \mu f_l(q_l)dq_l}{q_D - q_L}, && \text{for } q_L < q < q_D; && (90)\\
B =& \frac{\int_{q_B}^{q_D} R_B(q_l, \epsilon)^\eta f_l(q_l)dq_l - \int_{q_B}^{q_D} \mu f_l(q_l)dq_l}{q_H - q_D} + \\
&+ \frac{\int_{q_D\rho^{-\epsilon}}^{q_H\rho^{-\epsilon}} R_2(q_l, \epsilon)^\eta f_l(q_l)dq_l}{q_H - q_D}, && \text{for } q_D \leq q < q_H; && (91)\\
f_d(q) =& R_2(q\rho^{-\epsilon}, \epsilon)^\eta f_l(q\rho^{-\epsilon})\rho^{-\epsilon}, && \text{for } q > q_H. && (92)
\end{aligned}
$$

The functions $R_B(., \epsilon)$ and $R_2(., \epsilon)$ denote the net subsidy payments to an adopter under the notched scheme relative to the subsidy payment under the counterfactual scheme for bunching agents and agents above the notch. They are $R_B(q_l, \epsilon) = \frac{s_l\ \rho\ q_D - \left(\frac{q_D^{1+1/\epsilon}}{q_l^{1/\epsilon}} - q_l\right)\frac{s_l\ \epsilon}{1+\epsilon}}{S + s_l\ q_l}$ and $R_2(q_l, \epsilon) = \frac{s_l\ \rho^{1+\epsilon}\ q_l - \left(\rho^{1+\epsilon}\ q_l - q_l\right)\frac{s_l\ \epsilon}{1+\epsilon}}{S + s_l\ q_L}$. The variable q_B denotes the quantity-choice of the marginal buncher and is implicitly defined by $R_B(q_B, \epsilon) = 1$.

PROOF: As in Section A.2.

Proposition 7 (The identification in the case of a notch)**.** *Under Assumptions 1, 2, 3 and a corresponding rank condition, the observable measure $f_d(.)$ identifies the counterfactual measure $f_l(.)$, the fraction of agents with optimization frictions μ, the intensive margin elasticity ϵ, and the participation margin elasticity η.*

PROOF: As in Section A.4. Four parts of the measure (89), (90), (91), and (92) are sufficient to identify the four unknowns.

E.2 Estimation and robustness

The estimation follows Section 3. The data from Kleven et al. (2013) contains the distribution of foreigners' earnings pooled from 1995 to 2000.[63] Figure 10 in Section 5 shows the data on logarithmic scales with the estimated model and the counterfactual. The selection of the specification follows Section C.4 and selects $P = 2$. I use the same interval to estimate the bunching mass and missing as in Kleven et al. (2013). Standard errors are estimated following the bootstrap procedure in Chetty et al. (2011). I estimate a participation elasticity of $\hat{\eta} = 1.76\ (0.05)$, which is not statistically different from the result in Kleven et al. (2013) using dif-in-dif, i.e., $\hat{\eta} = 1.6\ (0.2)$ (see Table II column 1, A1 in Kleven et al. (2013)). I estimate an intensive margin elasticity of $\hat{\epsilon} = 0.0033\ (0.0003)$, which is in line with Kleven et al. (2013) who report that their elasticity is below 0.01. Table 15 summarises the results.

To estimate the bias from ignoring the participation margin, I use points above the notch for estimating counterfactual earnings following Kleven et al. (2013). The bias from ignoring participation is -17%. This is because bunching is costly for agents; compared to the counterfactual, where agents face a flat tax also below the notch, some agents cease to participate.

Table 15: Results Application Denmark.

$\hat{\epsilon}$ (SD)	$\hat{\eta}$ (SD)	Relative Bias ϵ [%]
0.0030 (0.0003)	1.76 (0.05)	-17

Note: The table shows the estimates and relative bias.

Kleven et al. (2013) also consider a model where bunching is partly due to general equilibrium wage reductions for agents above the notch. Therefore, as a robustness check, I consider a model where bunching is due to such wage responses. The estimated participation margin elasticity is not significantly different.[64]

[63] The data is retrieved using image processing software from Figure IV in Kleven et al. (2013). The parameters of the tax function are in sections II.A and IV.A in Kleven et al. (2013): the marginal net of tax rate tax jumps from 0.32 to 0.7 and the average net of tax rate from 0.4 to 0.7 at 765,600 Danish kroner.

[64] I first estimate the intensive margin response assuming all bunching comes from above due to wage effects. Then, I estimate the participation margin, excluding observations in the bunching and missing mass interval. The point estimate is 1.86 which is within the 90% confidence interval of the main estimate in Table 15.