Identifying Causal Effects under Kink Setting: Theory and Evidence

Yi Lu
Tsinghua University

Jianguo Wang
Renmin University of China

Huihua Xie
Zhejiang University

March 2024

Abstract

This paper develops a generalized framework for identifying causal impacts in a reduced-form manner under kinked settings when agents can manipulate their choices around the threshold. The causal estimation using a bunching framework was initially developed by Diamond and Persson (2017) under notched settings. Many empirical applications of bunching designs involve kinked settings. We first propose a model-free causal estimator in kinked settings with sharp bunching and then extend to the scenarios with diffuse bunching, misreporting, optimization frictions, and heterogeneity. The estimation method is mostly non-parametric and accounts for the interior response under kinked settings. Applying the proposed approach, we estimate how medical subsidies affect outpatient behaviors in China.

1 Introduction

In many empirical setups, agents received treatment based on whether their value of a variable (also referred to as the "assignment variable" or "running variable" in the literature) is above or below a known cutoff. For example, students with test scores above the cutoff are admitted to better schools/colleges (e.g.,Zimmerman [2019], Pop-Eleches and Urquiola [2013]); workers with annual income above the threshold are subject to higher tax rates (Saez [2010]). Such thresholds feature discontinuity in the *level* of choice sets/treatment probabilities (referred to as "notches" with level change only hereafter), or discontinuity in the *slope* of choice sets/treatment probabilities (referred to as "kinks" hereafter), or discontinuity in both the *level* and the *slope* of choice sets/treatment probabilities (referred to as "notches" with both level and slope changes hereafter). These non-linear designs facilitate treatment effects identification and policy impact evaluation. The literature distinguishes between two conceptually different non-linear designs, based on whether agents (e.g., students, workers, patients, firms) can fully manipulate their measures around the cutoff. Specifically, when agents cannot fully manipulate the assignment variable around the threshold, regression discontinuity design (RDD), regression kink design (RKD), and regression probability jump and kink design (RPJKD) are adopted depending on whether agents face the notched or kinked policies. However, when agents can fully manipulate their measure and decide whether to locate above or below the threshold, assumptions in RDD, RKD, and RPJKD are no longer valid. In such scenarios, bunching methods are used to study agents' behavior (see literature review by Kleven [2016]).

Early studies in the bunching literature focus on identifying the key elasticity (e.g., the elasticity of taxable income to the net of tax rate), which involves estimating the counterfactual density distribution of the assignment variable (when agents cannot manipulate their measure). Saez [2010] and Chetty et al. [2011] developed the bunching method in kinked settings, while Kleven and Waseem [2013] developed the bunching method in notched settings. The method has been deployed in various settings, such as R&D (Chen et al. [2021],), housing markets (Best and Kleven [2018], Best et al. [2020], Cloyne et al. [2019]). However, fewer studies focus on the impacts of the agents' manipulation behaviors due to the kinked policy on other outcome variables. Diamond and Persson [2017] proposed a causal estimator in notch settings. By assuming that manipulation only happens within a certain region around the cutoff, they recover the counterfactual density and outcome distributions within the manipulation region when there is no manipulation, by extrapolating

the corresponding distributions outside the manipulation region into the manipulation region. The difference between the average observed and counterfactual values within the manipulation region reveals the treatment effect from the agents' responses. Diamond and Persson [2017] complements the RDD method when agents can fully manipulate the assignment variable. A critical part of Diamond and Persson [2017] is that manipulation only happens within a certain region, which may not be true when there is a discontinuity in the slope of choice sets/treatment probabilities. Facing slope changes (i.e., change in marginal incentives), agents to one side of the cutoff would all adjust their assignment variable upwards/downwards, which are denoted as "interior responses" [1] Therefore, there would not be a manipulation region, which makes Diamond and Persson [2017]'s method invalid under kink settings as well as notch settings with both level and slope changes. In this paper, we develop a framework for estimating the treatment effects of agents' manipulation behavior due to the kinked policy on other outcome variables[2], which complements the RKD method when agents can manipulate the assignment variables. Our method is model-free and is based on agents' interior response behavior which draws less attention in the previous bunching literature.

Our approach is centered around agents' interior responses under kink settings. Consider a counterfactual linear policy with the tax/co-payment rate being the same as that below the kink. When a kinked policy is introduced, agents below the kink remain unchanged, while agents above the kink face a change in marginal incentives and adjust their assignment variable accordingly. Specifically, agents with counterfactual values just above the kink would bunch at the kink (denoted as *bunchers*), and agents with counterfactual values further above the kink would all reduce their value by a constant share but stay above the kink. Since the *marginal bunching agent* is also the *marginal shifting agent*, we can infer the relative change for *shifting agents* using the ratio of the *marginal bunching agent*'s counterfactual value and the kink. This important feature allows us to recover the counterfactual locations for *shifting agents* and hence obtains the counterfactual density distribution non-parametrically. [3]

[1] Chetty et al. [2011] address the interior response issue for the counterfactual density estimation by imposing the integration constraint (i.e., assuming that the number of observations under the observed and counterfactual distributions is the same) and by assuming that the observed density in the interior response part is a parallel shift of the counterfactual one.

[2] The method could also be applied to notched settings with both level and slope changes upon small modification, as the fundamental issue addressed here is the interior response.

[3] It involves a computation algorithm. First, given an initial guess of marginal bunchers' location, we can infer a counterfactual density distribution using the feature that *shifting agents* adjust their values of assignment variable by the same constant share as the *marginal bunching agent*. Second, based on the inferred counterfactual density distribution, we compute excess bunching

We then plot the conditional outcome distribution. Since we have recovered the counterfactual location for *shifting agents*, we can relocate *shifting agents* back to their counterfactual locations, which generates an auxiliary outcome distribution. This relocation process preserves the smoothness of the underlying agents' characteristics, however, the auxiliary outcome distribution still includes the impacts from the changes in the assignment variable and the changes in tax/copayments/benefits (due to the kinked policy and changes in the assignment variable). Compared to the counterfactual outcome distribution under the linear policy, the auxiliary outcome distribution would feature a level change and a slope change.[4] Agents below the kink face the same incentive and do not adjust their assignment variable, their observed outcome distribution is the same as the counterfactual outcome distribution. We estimate the level and slope changes using the observed outcome distribution below the kink and the auxiliary outcome distribution above the kink.[5] The estimated level change and slope change allow us to exactly identify two structural parameters, which reflect how changes in the assignment variable directly affect the outcome variable and how changes in tax/copayment/benefits affect outcome outcome variable. The structural parameters also work as sufficient statistics, allowing us to simulate the impacts of alternative kinked policies.

Our methodology could be used more generally to study the treatment effects of a kinked policy under bunching settings with various extensions, including diffused bunching, rounding in agents' choices, potential misreporting, the existence of stayers (unresponsive agents) due to optimization frictions or inattention, and heterogeneous treatment effects.

We apply our treatment effect estimator to study the kinked cost-sharing design under China's health insurance system. Specifically, the co-payment ratio for rural and urban non-employed patients increased from 50% to 100% when their accumulated annual medical expenses exceeded the policy statutory threshold. This generates a discontinuity in the marginal cost of treatment borne by patients when their annual eligible expenditure is above the given threshold. We found that

and find the updated value of the marginal bunchers' location by integrating over the inferred counterfactual density distribution from just above the kink till excess bunching equals missing mass. We repeat this process till it converges. Details are discussed in section 3.1.

[4]In regression kink designs, there is only slope change because agents cannot manipulate their assignment variable and the slope change in outcome distribution is driven by the kinked policy. In bunching setups, there is additional level change because agents do adjust their assignment variable, which may affect outcome variables directly and indirectly through changes in tax/copayments.

[5]Since the relocation process resolves the selection issue in bunching, we can extrapolate the counterfactual outcome distribution rightwards or the auxiliary outcome distribution leftwards.

patients adjust their annual medical expenses and bunch at the threshold. Compared to the counterfactual scenario where the co-payment rate is 50%, patients visit the hospital less often when the co-payment rate increases to 100%. This indicates a significant amount of compressed medical demand due to patients' financial concerns.

Our paper is related to three threads of literature. First, we contribute to the literature on treatment effect estimation and policy evaluation using quasi-experimental approaches. The fact that agents can fully determine their value of the assignment variable above or below the threshold indicates that the identifying assumption for RDD or RKD fails. Under the notched design with bunching mass around the cutoff, Carneiro et al. [2015] proposed the "donut" regression design method ("donut" RD) by excluding a certain manipulative region around the threshold to solve this issue. The estimation precision of the "donut" RD estimator depends on how large the excluded region is. Alternatively, Diamond and Persson [2017] proposes a Wald estimator that captures the causal impact of manipulation on the subset of agents that are chosen for manipulation. Their method shares certain similarities with "donut" RD in the sense that both assume manipulation happens within a certain range around the threshold.[6] Our paper contributes to this literature by providing the framework to estimate the average treatment effects of a kinked policy with manipulative agents, where the interior response agents lead to the un-neglectful shift of density distribution to one side of the threshold, immediately invalidating the assumption that manipulation happens within a certain region around the cutoff in previous literature.

Second, we contribute to the estimation of counterfactual density distribution in bunching estimation. The critical step in bunching estimation relies on estimating the counterfactual density distribution in the counterfactual situation absent of kinks or notches. The standard approach to obtaining such counterfactual was developed by Chetty et al. [2011] in the context of kinks and extended by Kleven and Waseem [2013] to notches. The standard approach is to fit a flexible polynomial to the observed distribution for the region slightly away from the threshold and then extrapolate the fitted distribution to the threshold, under the assumption that the counterfactual distribution is smooth around the threshold. However, agents to one side of the cutoff would adjust their locations in response to the changed marginal incentive, leading to an interior shift of density distribution. This fact results in a biased estimation of the counterfactual distribution if we do not account for the shift in the observed distribution brought by these interior response

[6] Under a notched design with discrete change in both the level and the marginal incentives at the threshold, the assumption that manipulation happens within a certain region around the threshold is invalid.

agents (also known as shifting agents). The magnitude of the interior response depends on the slope of the density distribution and the size of the change in incentives. In general, such interior response effects are larger for kinks than for notches because kinks usually feature a larger change in marginal incentives.[7]

In kinked designs, Chetty et al. [2011] addresses this interior response issue by assuming that the counterfactual density to one side of the threshold is a constant upwards/downwards movement of the observed density distribution. However, in most setups with changes in marginal incentives, interior response agents would adjust their locations by a constant percentage and hence by different magnitudes depending on their initial values. Therefore, interior responses would lead to a non-parallel shift of density distribution along the x-axis, which indicates that the commonly adopted method proposed by Chetty et al. [2011] could lead to certain biases. In this paper, we propose an algorithm for counterfactual density estimation that features the exact interior responses. In addition, our proposed estimation method also works under notched settings with different marginal incentives.

Third, we contribute to the literature which explores various extensions in the bunching methodology. Some extensions focus on causal identification when there is no discontinuity in the policy but agents' choices are truncated at 0. For example, the number of hours children spend watching TV has to be above or equal to 0. Caetano [2015] exploits such setups for identifying potential selection in reduced-form estimation. Caetano et al. [2023] propose causal estimators for identifying treatment effects at 0. Other extensions include a two-dimensional bunching approach by Cox et al. [2021], and non-identification of elasticity under a single budget set by Blomquist et al. [2021]. we distinguish our paper from these papers in the sense that we focus on different research topics and setups. Specifically, we focus on identifying causal effects under kink settings where agents can manipulate and the cutoff is not at the truncated point.

The rest of the paper is arranged as follows. Section II discusses a generalized framework under kinked settings, which covers the basic setup, interior response of shifting agents, and causal effects under sharp bunching and under bunching with diffusion. Section III discusses the estimation strategy for estimating the counterfactual density distribution and the counterfactual outcome distribution. Section IV extends the treatment effect estimation to various scenarios, such as re-labelling, rounding, and stayers due to optimization frictions and heterogeneity in the structural

[7] In notched designs, researchers often ignore the interior response and assume manipulation only happens within a certain region around the threshold. This would lead to potentially biased estimates of counterfactual density and elasticity as well.

parameter. Section V applies the proposed causal estimator to the kinked coinsurance policy in China, where the medical care system, data, bunching pattern, density and outcome distributions, and the treatment effects are presented. Section VI concludes.

2 A Generalized Framework under the Kink Setting

We elaborate a theoretical framework for the causal inference under the kink setting in this section and defer the empirical execution to the next section. Specifically, we first lay out the basic setup, and derive the optimal interior responses for the complying agents. Then we study the causal inference under the sharp bunching scenario. Finally, we consider the diffusion case for the causal impact under the kink setting.

2.1 Setup

Consider a focal kinked policy in which agents face a tax rate (or co-payment rate) of t if their value of z is below a statutory cutoff z^*, but face a higher marginal tax rate (or co-payment rate) of $t+\Delta t$ if their $z > z^*$. Denote the amount of money that agents pay under the kinked policy as $T(z)$. That is,

$$T(z) = \begin{cases} t \times z & \text{if } z \leq z^* \\ (t+\Delta t) \times z - \Delta t z^* & \text{if } z > z^* \end{cases} \tag{1}$$

Denote the optimal response function of z from agents maximizing their objective functions as $z = z(D,n)$, where $D = 1$ indicates that agents face the lower marginal tax/co-payment rate of t and $D = 0$ indicates agents face the high marginal tax/co-payment rate of $t+\Delta t$; and n is an unobserved agent heterogeneity, with $z(D,n)$ increasing in n. In general, $z(1,n) > z(0,n)$, i.e., z is higher when the tax rate (co-payment rate) on z is lower on the margin.

Agents' optimal choice z under the kinked policy can ben shown as:

$$z = \begin{cases} z(1,n) & \text{if } n \leq n_L \\ z^* & \text{if } n \in (n_L, n_H]. \\ z(0,n) & \text{if } n > n_H \end{cases} \tag{2}$$

For the counterfactual policy, we consider a linear policy where agents always face the low tax rate (co-payment rate) of t.[8] That is, $T^{ct}(z) = t \times z^{ct}$. Consequently, agents' optimal choices are $z^{ct} = z(1,n)$.

For agents with $n < n_L$, the optimal z under the kinked policy remains the same as z^{ct} in the counterfactual policy as they face the same tax rate (co-payment rate) of t. We denote these agents as "*always-takers*". Next, for agents with $n > n_H$, they reduce their z in response to the higher marginal tax rate (co-payment rate) of $t + \Delta t$ under the kinked policy, $(z = z(0,n) < z(1,n) = z^{ct})$, but stay in the interior of the upper bracket, compared to the counterfactual policy. We denote them as "*shifters*", or, agents with "interior response". Finally, for agents with $n \in (n_L, n_H]$, their optimal choice under the kinked policy is to reduce their z and bunch at the threshold z^*. We denote them as "*bunchers*", as their behavior produces excess bunching in the density distribution at the kink point z^* when the kinked policy is introduced.

Remark 1 The literature on bunching has specified the agent's objective function, in which the optimization leads to a specific function of $z(D, \phi)$. For example, Saez (2010) and subsequent studies (e.g., Chetty et al. 2011; Einav, Finkelstein, and Schrimpf 2017.) typically assume a static quasi-linear, iso-elastic preference over consumption and labor supply (or medical spending) to obtain agents' response elasticity to tax (or coinsurance) kinks. Specifically, Saez (2010) considers a quasi-linear utility function $u(c,z) = z - T(z) - \frac{n}{1+1/e} \times (\frac{z}{n})^{1+1/e}$, where $T(z)$ is the tax system; n denotes the individual heterogeneity in abilities; and e is the labor supply elasticity. The counterfactual scenario is characterized by a linear tax system with $T(z) = t \times z$, whereas the focal kinked tax policy introduces an increase in the marginal tax rate from t to $t + \Delta t$ at the earnings threshold z^*. The optimal labor supply choice can be derived as

[8] Alternatively, one can consider a counterfactual policy where agents always face the high tax rate (co-payment rate) of $t + \Delta t$. These two cases generate identical density responses (Kleven [2016]), but have different implications on the casual inference. We study the counterfactual linear policy with a high tax rate in the extensions.

$$z = \begin{cases} n \times (1-t)^e & \text{if } n \leq \frac{z^*}{(1-t)^e} \equiv n_L \\ z^* & \text{if } n \in (n_L, n_H] \\ n \times (1-t-\Delta t)^e & \text{if } n > \frac{z^*}{(1-t-\Delta t)^e} \equiv n_H \end{cases}$$

This optimal choice equation under the kinked policy corresponds to Eq.(4) in Chetty et al. (2011) and Eq.(5) in Einav, Finkelstein, and Schrimpf (2017).

To derive the key features of optimal responses, following the above literature, we make a weak assumption of the optimal response function of z:

Assumption 1 (Separability)**.** *Assume that* $z(D,n) = f(D;e)g(n;e)$.

Here, e is a structural parameter, such as the elasticity of labor supply to the net of the tax rate or a semi-elasticity that relates the probability of participation/consumption to the percentage change in financial incentives; $f(D;e)$ is a discrete function with D being 0 or 1; and $g(n;e)$ is the distribution of n. Assumption 1 states the separability of the marginal tax (or co-payment) rate ($D = 0/1$) and agents' heterogeneity n in the optimal choice function.

Remark 2 Almost all studies on the bunching estimation make Assumption 1, such as the model with no uncertainty and quasi-linear, iso-elastic preferences (Saez, 2010; Chetty et al, 2011; Einav, Finkelstein, and Schrimpf, 2017). For example, in Saez (2010), the optimal income choices $z(1,n) = (1-t)^e \times n$ and $z(0,n) = (1-t-\Delta t)^e \times n$ satisfy the Assumption 1.

Given Assumption 1, Equation (2) can be re-written as:

$$z = \begin{cases} z(1,n) & \text{if } n \leq n_L \\ z^* & \text{if } n \in (n_L, n_H] \\ z(0,n) = z(1,n)\frac{f(0;e)}{f(1;e)} & \text{if } n > n_H \end{cases} \tag{3}$$

That is, agents with $n > n_H$ who originally choose $z(1,n)$ under the counterfactual linear policy respond to the kinked policy by setting $z = z(0,n) = z(1,n)\frac{f(0;e)}{f(1;e)} > z^*$.

For marginal *bunchers* with n_H, its optimal choices under the kinked policy and under the counterfactual linear policy are, respectively, given by $z = z(0, n_H) = z^*$ and $z^{ct} = z(1, n_H) = z^* + \Delta z^*$, where Δz^* is the change in z by the marginal bunching agent with n_H due to the introduction of the kinked policy. The excess bunching at the kink point is the cumulative density of *bunchers* , i.e., agents with $n \in (n_L, n_H]$. And given the one-to-one mapping between n and z (as shown in Equation (3)), the bunching mass can be calculated as:

$$B = \int_{z^*}^{z^*+\Delta z^*} h^{ct}(z)dz, \tag{4}$$

where $h^{ct}(z)$ denotes the counterfactual density distribution of z (i.e., the one under the linear low tax/co-payment rate plan).[9]

For all agents with $n > n_H$, we have,

$$\frac{z^{ct}}{z} = \frac{z(1,n)}{z(0,n)} = \frac{f(1;e)}{f(0;e)} = \frac{z^* + \Delta z^*}{z^*}. \tag{5}$$

Equation (5) characterizes the relationship between the original location (under the counterfactual linear policy) and the new location (under the kinked policy) for each shifting agent.

Remark 3 Studies in the bunching literature largely use Equation (5) to back out the structural parameter from the estimated value of Δz^*. For example, in Saez (2010), Equation (5) is $(\frac{1-t-\Delta t}{1-t})^e = \frac{z^*+\Delta z^*}{z^*}$.

Combining Equations (3) and (5), we can summarize the change in z as

$$\frac{z}{z^{ct}} = \begin{cases} \frac{z(1,n)}{z(1,n)} = 1 & \text{if } n \leq n_H \\ \frac{z^*}{z(1,n)} & \text{if } n \in (n_L, n_H] \\ \frac{z(0,n)}{z(1,n)} = \frac{f(0;e)}{f(1;e)} = \frac{z^*}{z^*+\Delta z^*} & \text{if } n > n_H \end{cases}. \tag{6}$$

Hence, moving from the counterfactual linear scenario to the state with the kinked policy, all agents with $n > n_H$ (*shifters*) reduce their z by a constant share , i.e., $\left(\frac{z^*}{z^*+\Delta z^*}\right) - 1 < 0$, but do not bunch at the cutoff z^*.[10] Meanwhile, agents with $n \in (n_L, n_H]$ (*bunchers*) reduce their z to bunch at the

[9] The observed density distribution of z under the kinked policy is denoted by $h(z)$.

[10] Note that each shifter's adjustment $(z^{ct} - z)$ is not a constant, it depends on the initial location

cutoff z^*. By contrast, agents with the $n \leq n_L$(*always-takers*) remain unchanged.

Equation (6) enables us to estimate the counterfactual density distribution $h^{ct}()$ and the marginal bunchers' response Δz^* nonparametrically (details will be discussed in section 3.1).

2.2 Causal Inference under Kinked Bunching

The bunching literature has focused on estimating the key structural parameter e since the methodological development by Saez (2010) and Kleven and Waseam (2013). What is equally (if not more) important is whether the bunching technique can be used to do a causal inference analysis; that is, estimating the effects from the introduction of a kinked policy on other outcome variables. In this subsection, we develop a methodological framework for the causal effects under the kinked design,[11] and defer empirical details of the estimation framework in section 3.

We start with the sharp bunching case to illustrate how our estimation framework works, and then discuss the case of bunching wish diffusion, which matches the data pattern. Meanwhile, as shown in subsection 2.1, the kinked policy divides agents into three groups: *shifters*, *bunchers*, and *always-takers*. Given that *always-takers* do not respond to the kinked policy, we examine the average treatment effects for *shifters* and *bunchers*.

2.2.1 Policy effect for *Shifters*

For *shifters* (i.e., $n > n_H$), the average policy effect (denoted as $\tau_y^{TE,shifter}$) can be calculated as:

$$\tau_y^{TE,shifter} = E[y_n - y_n^{ct} | n \in shifters], \tag{7}$$

where $y_n \equiv y(z_n)$ is the observed outcome for shifting agent n under the kinked policy and $y()$ is the observed outcome distribution that maps the effort z_n to the outcome under the kinked policy; and $y_n^{ct} \equiv y^{ct}(z_n)$ is the corresponding outcome for shifting agent n under the counterfactual linear policy and $y^{ct}()$ is the outcome distribution that maps the effort z_n to the outcome under the

(z^{ct}). Alternatively, we can take the logarithm of z so that each shifter's adjustment will be a constant, i.e., $\ln z^{ct} - \ln z = \ln \frac{z^* + \Delta z^*}{z^*}$.

[11] In a companion work, Diamond and Persson (2017) study the causal estimation using the *notch* design.

counterfactual linear policy.

The policy generates two possible effects. First, in response to the kinked policy, agent n reduces its optimal effort z_n. That is, agent n sets the optimal effort as $z_n = z_n^{ct}\frac{z^*}{z^*+\Delta z^*}$ under the kinked policy, where z_n^{ct} is the optimal effort under the counterfactual linear policy. The reduction in the effort z could directly affect outcome y. For example, a reduction in taxable income could affect consumption, or a reduction in medical expenses could affect health. Define semi-elasticity $\mu_n \equiv \frac{\Delta y_n}{\Delta z_n/z_n}$. Therefore, we can calculate the direct effect of z on y as $(y_n - y_n^{ct})|_{\text{due to direct change in } z} = \mu_n\big(\frac{z^*}{z^*+\Delta z^*} - 1\big)$.

Second, the reduction in z also allows agent n to enjoy taxes (or fees) repayment, which could in turn affect outcome y. Recall that under the counterfactual policy, we have $T^{ct}(z^{ct}) = t \times z^{ct}$ and under the kinked policy, we have $T(z) = (t+\Delta t)\times z - \Delta t \times z^* = (t+\Delta t)\frac{z^*}{z^*+\Delta z^*} \times z^{ct} - \Delta t \times z^*$. From z_n^{ct} to z_n, agent n obtain $\Delta T_n \equiv T(z_n) - T^{ct}(z_n^{ct})$. Define $-\lambda_n \equiv \frac{\Delta y_n}{\Delta T_n}$. Hence, we can calculate the effect of change in tax or fees (T) on y as $(y_n - y_n^{ct})|_{\text{due to change in } T} = -\lambda_n\big((t+\Delta t)\frac{z^*}{z^*+\Delta z^*} \times z_n^{ct} - \Delta t \times z^* - t \times z_n^{ct}\big) = -\lambda_n z_n^{ct}\big((t+\Delta t)\frac{z^*}{z^*+\Delta z^*} - t\big) + \lambda_n \Delta t \times z^*$.

To estimate the policy effect, we make the following assumption:

Assumption 2 (Additive)**.** *Assume the effects of z and T on outcome y are additive.*

Remark 4. Assumption 2 satisfies most of the utility model used in the bunching literature. For example, the quasi-linear, iso-elastic preference model separates income and effort effects in an additive way. Meanwhile, Assumption 2 is largely invoked in the multiple linear regressions, i.e., multiple regressors in the separable and additive format.

Given Assumption 2, we can rewrite Equation (8) as:[12]

$$\begin{aligned}
\tau_y^{TE,shifter} &= E[y_n - y_n^{ct} | n \in shifters] \\
&= E\Big[\mu_n\big(\frac{z^*}{z^*+\Delta z^*} - 1\big) - \lambda_n z_n^{ct}\big((t+\Delta t)\frac{z^*}{z^*+\Delta z^*} - t\big) + \lambda_n \Delta t \times z^*\Big]. \\
&= E(\mu_n)\big(\frac{z^*}{z^*+\Delta z^*} - 1\big) - E(\lambda_n z_n^{ct})\left((t+\Delta t)\frac{z^*}{z^*+\Delta z^*} - t\right) + E(\lambda_n)\Delta t \times z^*
\end{aligned}$$

The change in z would generate three changes to the outcome distribution.

[12] We provide a formal ground-up proof in Appendix B. Specifically, the proof incorporates three features: individual heterogeneity in the initial value of y, and the impacts of z and T on y.

First, the relocation effect. Even if the change in z has no impact on y, such a "relocation" behavior (from z^{ct} to z) would change the outcome distribution. Therefore, if we directly compare y^{ct} with y along the y-axis, we are not comparing the same agent. However, we do know where each agent has moved to. Therefore, if we relocate each agent under the kinked policy back to his/her counterfactual location, then, comparing the values along the y-axis would give us the treatment effect on *"shifters"*.

Treatment Effect on "Shifters"

$$\begin{aligned}\tau_y^{TE,shifter} &= E[y_n - y_n^{ct} | n \in shifters] \\ &= \int_{z^* elta z^*}^{z^{max}} \left(y^r(z^{ct}) - y^{ct}(z^{ct}) \right) \frac{h^{ct}(z^{ct})}{\int_{z^*+\Delta z^*}^{z^{max}} h^{ct}(z^{ct}) dz^{ct}} dz^{ct} \end{aligned} \tag{8}$$

where $y^r(z^{ct}) \equiv y(z^{ct}\frac{z^*}{z^*+\Delta z^*})$ denotes the resulting auxiliary outcome distribution when we locate shifters at z back to their counterfactual location z^{ct} using the relation that $z = z^{ct}\frac{z^*}{z^*+\Delta z^*}$. That is, when we reshape the observed outcome distribution based on the changes in agents' location of z, the outcome distribution changes from $y(z)$ to $y(z^{ct}\frac{z^*}{z^*+\Delta z^*}) \equiv y^r(z^{ct})$.

Remark 4. Estimation of treatment effect on *shifters* only requires $y^{ct}(z^{ct}), h^{ct}(z^{ct})$. Therefore, the estimation is model-free. However, one might want to understand what drives the change in outcome as a result of the kinked policy. The following second and third points cover it.
Assume homogeneous preference and thus single response elasticities across agents (i.e., $\mu_n = \mu, \lambda_n = \lambda$), a condition commonly made in the bunching literature (see, e.g., Saez 2010; Chetty et al. 2011, Kleven 2016;). The above equation can be simplified as

$$\tau_y^{TE,shifter} = \mu\left(\frac{z^*}{z^*+\Delta z^*} - 1\right) - \lambda E(z_n^{ct})\left((t+\Delta t)\frac{z^*}{z^*+\Delta z^*} - t\right) + \lambda \Delta t \times z^*. \tag{9}$$

Identifying Sufficient Statistics. We claim μ and λ are sufficient statistics for estimating treatment effects under policy simulations because changes in policy cutoffs or tax/co-payment rates would result in changes in z and hence changes in outcome variables. We propose estimating these parameters by exploiting the level and slope change at z^* when comparing the distributions

$y^{ct}(z^{ct})$ and the extrapolated $y^{r}(z^{ct})$. Specifically, we have

$$\text{Level Change at } z^* = \mu\left(\frac{z^*}{z^*+\Delta z^*}-1\right)-\lambda(t+\Delta t)z^*\left(\frac{z^*}{z^*+\Delta z^*}-1\right) \tag{10}$$

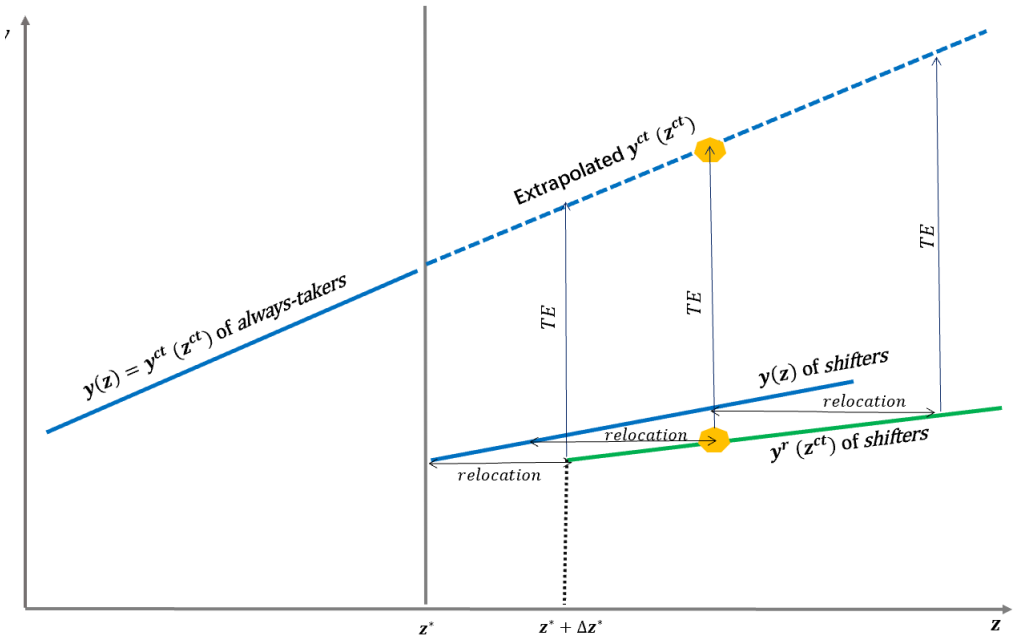
$$\text{Slope Change at } z^* = -\lambda\left((t+\Delta t)\frac{z^*}{z^*+\Delta z^*}-t\right) \tag{11}$$

where estimation of the level change and slope change at z^* is explained in the next subsection 3.2.

Remark 5 Note that calibration process identifies the parameters μ, λ for *shifters*. Because it is based on the slope and level changes at z^* by comparing the counterfactual outcome distribution with the extrapolated auxiliary distribution of *shifters*.

Figure 1 illustrates the change in outcome distribution of *shifters* when the kinked policy is introduced.

Figure 1: Change of outcome distribution for *shifters*



2.2.2 Change in outcome Distributions of *Bunchers*

As discussed in subsection 2.1, agents with $z^{ct} \in (z^*, z^*+\Delta z^*]$, i.e., $n \in (n_L, n_H]$, would reduce their value of z and bunch at the cutoff ($z = z^*$) under the kinked policy. The changes in z would also

generate changes in the outcome distribution.

Under the sharp bunching scenario, agents with $z^{ct} \in (z^*, z^* + \Delta z^*]$ relocate to $z = z^*$. As it is impossible to find a one-to-one mapping for each bunching agent, we take all the bunching agents as an entity and identify the average treatment effect on "*bunchers*" by comparing changes in the average outcome value.

Treatment Effect on "bunchers" under Sharp bunching

$$\begin{aligned} \tau_y^{TE,buncher} &= E\left[y_n - y_n^{ct} | n \in buncher\right] \\ &= \overline{Y^{buncher}} - \overline{Y^{buncher,ct}} \\ &= y^{buncher}(z^*) - \int_{z^*}^{z^*+\Delta z^*} y^{ct}(z^{ct}) \frac{h^{ct}(z^{ct})}{\int_{z^*}^{z^*+\Delta z^*} h^{ct}(z^{ct}) dz^{ct}} dz^{ct} \end{aligned} \tag{12}$$

where $y^{buncher}(z^*)$ denotes the average outcome of *bunchers* under the kinked policy, the estimation of which is shown below.

Specifically, under the kinked policy, observations at the threshold z^* contain two groups of agents: (1) bunching agents with $z^{ct} \in (z^*, z^* + \Delta z^*]$ who decrease their value to the threshold $z = z^*$ in response to the kinked policy; (2) *always-takers* with $z^{ct} = z^*$ who remain at the threshold $z = z^*$. By contrast, under the counterfactual linear policy, there is only *always-takers* at the threshold z^*. Therefore, the density of bunchers under the kinked policy is given as $h^{bunch}(z^*) = h(z^*) - h^{ct}(z^*)$. Further, the observed average outcome $y(z^*)$ is the weighted average of *bunchers* and *always-takers*, i.e., $y(z^*) = \left(y^{buncher}(z^*)h^{buncher}(z^*) + y^{ct}(z^*)h^{ct}(z^*)\right)\frac{1}{h(z^*)}$. Therefore, we obtain the average outcome of *bunchers* under the kinked policy $y^{buncher}(z^*)$. That is,

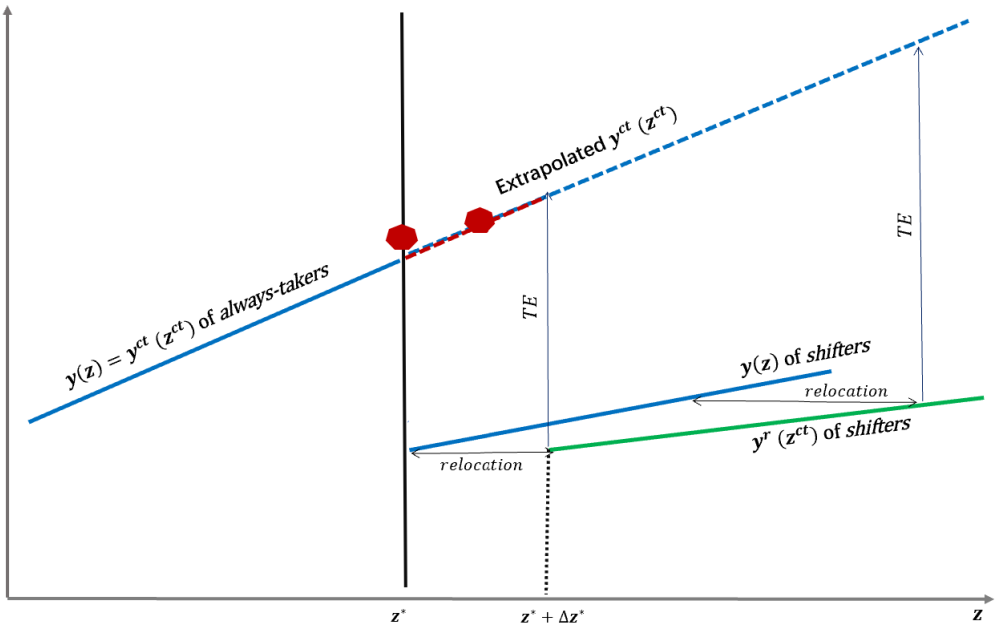
$$y^{buncher}(z^*) = \frac{y(z^*)h(z^*) - y^{ct}(z^*)h^{ct}(z^*)}{h(z^*) - h^{ct}(z^*)}. \tag{13}$$

Figure 2 illustrates the change in outcome distribution of *bunchers* when the kinked policy is introduced.

Treatment Effect on "bunchers" under Diffuse bunching

In reality, agents may bunch around the threshold (i.e., $[z^* - u_1, z^* + u_2]$) due to optimization frictions. Still, we take all the bunching agents as an entity and identify the average treatment effect on "bunchers" by comparing changes in the average outcome value. However, there is one difference. Under sharp bunching, we only need to estimate the average outcome of bunchers under

Figure 2: Change of outcome distribution for *bunchers*



the kinked state at z^* (because all bunchers relocated to z^*); by contrast, under diffuse bunching, we now need to estimate the average outcome of bunchers under the kinked state over the whole diffuse region $[z^* - u_1, z^* + u_2]$.

$$\begin{aligned}
\tau_y^{TE,buncher} &= E\left[y_n - y_n^{ct} | n \in buncher\right] \\
&= \overline{Y^{buncher}} - \overline{Y^{buncher,ct}} \\
&= \int_{z^*-u_1}^{z^*+u_2} y^{buncher}(z) \frac{h^{bunch}(z)}{\int_{z^*-u_1}^{z^*+u_2} h^{bunch}(z)dz} dz - \int_{z^*}^{z^*+\Delta z^*} y^{ct}(z^{ct}) \frac{h^{ct}(z^{ct})}{\int_{z^*}^{z^*+\Delta z^*} h^{ct}(z^{ct})dz^{ct}} dz^{ct}
\end{aligned} \tag{14}$$

where $y^{buncher}(z)$ (with $z \in [z^* - u_1, z^* + u_2]$ denotes the average outcome of *bunchers* in each bin under the kinked policy and $h^{buncher}(z)$ denotes the corresponding density. The estimation of $y^{buncher}(z)$ and $h^{buncher}(z)$ are explained below.

Specifically, consider first the left side of the diffusion region $[z^* - u_1, z^*]$. Under the kinked policy, this region contains two groups of agents: the *bunchers* and *always-takers*. Under the counterfactual policy, this region only has *always-takers*. Similar to equation (13) in the sharp bunching case, for each $z \in [z^* - u_1, z^*]$, we can back out the outcome for *bunchers* under the kinked

policy by deducting the average outcome of *always-takers* from the average observed outcome. That is,

$$y^{buncher}(z) = \frac{y(z)h(z) - y^{ct}(z)h^{ct}(z)}{h(z) - h^{ct}(z)}, \forall z \in [z^* - u_1, z^*] \tag{15}$$

Also, we have $h^{buncher}(z) = h^1(z) - h^{ct}(z), \forall z \in [z^* - u_1, z^*]$.

Next, consider the right side of the diffusion region $(z^*, z^* + u_2]$. In the observed state, it also contains two groups of agents: the *bunchers* and *shifters*. Under the counterfactual linear policy, it only contains *shifters*. Therefore, for each $z \in (z^*, z^* + u_2]$, we can back out the outcome for *bunchers* under the kinked policy by deducting the average outcome of *shifters* from the average observed outcome. That is,

$$y^{buncher}(z) = \frac{y(z)h(z) - y^{shifter}(z)h^{shifter}(z)}{h(z) - h^{shifter}(z)}, \forall z \in (z^*, z^* + u_2] \tag{16}$$

where $h^{shifter}(z)$ and $y^{shifter}(z)$ correspond to the density and outcome distributions of *shifters* for $z \in (z^*, z^* + u_2]$, which can be inferred by extrapolating the distributions of shifting agents (the observed distributions) in the region with $z > z^* + u_2$ to the diffuse region with $z \in (z^*, z^* + u_2]$. Similarily, we have $h^{buncher}(z) = h(z) - h^{shifter}(z), \forall z \in (z^*, z^* + u_2]$.

Remark 6 Note that while sharp bunching agents and diffused bunching agents may be different, it does not pose any threats to our estimation framework. This is because we consider all bunching agents as an entity, and compare their average observed outcomes under the kinked policy to the average counterfactual outcomes under the linear policy. In other words, we are comparing the same group of agents under the treated and the control states. Meanwhile, the data allow us to distinguish sharp *bunchers* from diffused *bunchers* (under-shooting or over-shooting), from which we can compare their predetermined characteristics to further shed light on the selection of diffused bunching.

Remark 7 For bunching agents, a reduction in z due to the kinked policy could directly affect y, shown as $\mu(\frac{z^*}{z^{ct}} - 1), \forall z^{ct} \in (z^*, z^* + \Delta z^*]$. Meanwhile, change in T could also affect y, shown as $-\lambda(z^* - z^{ct})t, \forall z^{ct} \in (z^*, z^* + \Delta z^*]$. Therefore, we can draw the linkage between the treatment effect on "bunchers" and structural parameters: $\tau_y^{TE,buncher} = \int_{z^*}^{z^* + \Delta z^*} \mu(\frac{z^*}{z^{ct}} - 1) - \lambda(z^* - z^{ct})t dz^{ct}$. Hence, one can also use the treatment effect on *bunchers* $\tau_y^{TE,buncher}$ to identify the parameters (μ, λ) for *bunchers*, provided that there are at least two kinks to provide enough moments.

3 Empirical Estimation

Our aforementioned estimation framework for the causal inference under the kinked bunching relies on the estimation of counterfactual density $h^{ct}()$ and outcome $y^{ct}()$ distributions under the linear policy. In this section, we elaborate on the empirical details to estimate these counterfactuals.

One important and common feature of *kink* settings is that all agents (both *shifters* and *bunchers*) to the one side of the policy threshold respond to the kinked policy. This is in contrast to the assumption under *notch* settings that the adjustment only happens within a certain range (*manipulation region*) around the threshold (see Diamond and Persson, 2017).[13] To account for the responses by *shifters*, we propose a new method to recover the counterfactual density distribution $h^{ct}()$ together with the marginal buncher's response Δz^*, and the counterfactual distribution of $y^{ct}()$.

Our estimation method of the counterfactual density distribution has several desired properties over the conventional approach used in the bunching literature. First, it automatically satisfies the *integration constraint* that the number of agents under the counterfactual and that under the observed distribution should be the same. Second, it allows for the fact that the observed and counterfactual density distribution for *shifters* can be non-parallel because the adjustment by shifting agents is non-uniform, i.e., $z - z^{ct} = z^{ct}(\frac{z^*}{z^*+\Delta z^*} - 1)$. This relaxes the assumption made by Chetty et al. (2011) that the counterfactual density distribution is a parallel upward shifting of the observed one within the range with $z > z^*$. Last, our empirical strategy is model-free and can be applied to most kink settings. The estimation of counterfactual density and outcomes distributions do not require or depend on modeling assumptions, except Assumption 1 which states that agents' choice of z depends on individual heterogeneity (n) and the tax/co-payment rate $(D = 0/1)$ and they appear in the form of multiplication. Assumption 1 is valid in most bunching settings.

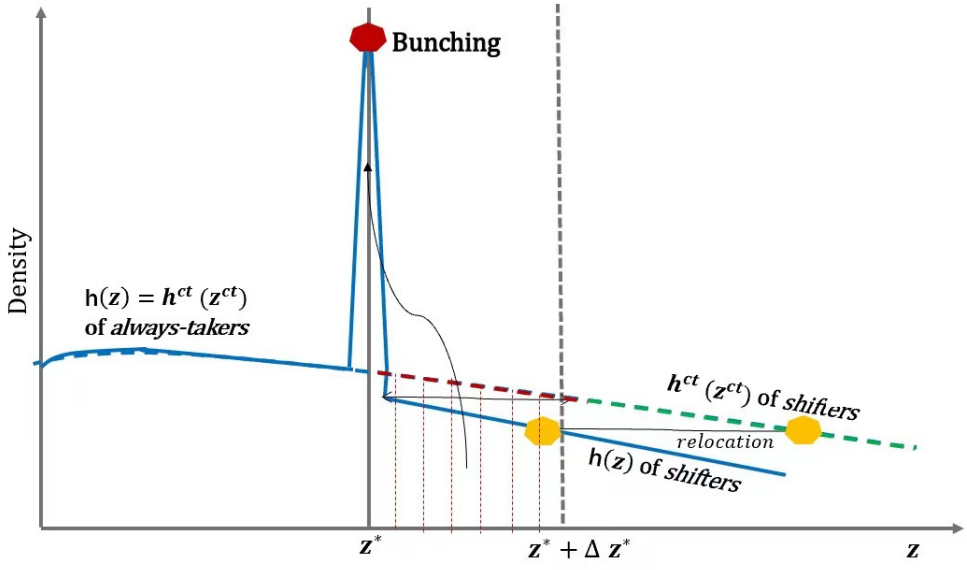
[13] In fact, there are also interior responses in the standard notch design when there is both level and slope changes of incentives around the threshold. For example, Kleven and Waseem (2013), Kleven (2016). However, such interior responses are largely ignored in the practical applications of notched designs. As pointed out in Kleven (2016), interior responses are larger for kinks than for notches because in real life changes of marginal tax rates are typically larger for the former than for the latter. Chetty el al. (2011) deals with the interior responses under kink settings, by assuming that the counterfactual density distribution is a parallel upward shifting of the observed one in the region with $z > z^*$.

3.1 Estimating Counterfactual Density Distribution

We start with the strategy to recover the counterfactual density distribution $h^{ct}(z)$, which can be applied to any kinked settings. As shown in Equation (6), agents' responses to the kinked policy can be summarized as: (i) *alwaysr-takers* with $z^{ct} \leq z^*$ remain unchanged, i.e., $z = z^{ct} \leq z^*$; (ii) *bunchers* with $z^{ct} \in (z^*, z^* + \Delta z^*]$ bunch at the threshold, i.e., $z = z^* < z^{ct}$; (iii)*shifters* with $z^{ct} > z^* + \Delta z^*$ reduce their value but do not bunch at the threshold, i.e., $z = z^{ct} \times \frac{z^*}{z^* + \Delta z^*} > z^*$.

Figure 3 illustrates the observed density distribution of z under the kinked policy (the solid curve) and the counterfactual density distribution under the linear case (the dashed curve). First, to the right of the threshold, it is the distribution of *always-takers*. As their behaviors remain unchanged in response to the kinked policy, the observed and counterfactual density distributions overlap. Second, agents with $z^{ct} \in (z^*, z^* + \Delta z^*]$ are the *bunching agents* and they move to the threshold z^* in response to the kinked policy, generating the bunching mass observed at z^* in **Figure 3**. Third, agents with $z^{ct} > z^* + \Delta z^*$ are the *shifting agents* and they reduce their z in response to the kinked policy but stay above z^* (i.e., stay in the interior of the upper bracket). These interior responses are represented by the leftward shift of the density distribution above z^*.

Figure 3: Change in the Density Distribution



To recover the counterfactual density distribution $h^{ct}()$ from the observed density distribution $h()$, we design a two-step estimation framework. First, we move *shifters* back to their counterfactual locations, which leads to the estimation of $h^{ct}(z)$ within the region $(z^* + \Delta z^*, \infty)$ for *shifters*.

Then, we extrapolate $h^{ct}()$ for bunching agents using the information of $h^{ct}()$ for *shifters* and *always-takers*, (as $h^{ct} = h()$ in the region $[z_{min}, z^*]$ for *always-takers*). Specifically, it is implemented by the following algorithm.

First, given the observed location z and an initial guess $\widehat{\Delta z^*}^{initial}$ for *shifting agents*, we infer the counterfactual choice $z^{ct,initial}$ based on the following relation derived from equation (6):

$$z^{ct,initial} = \begin{cases} z & \text{if } z < z^* - u_1 \\ z\frac{z^* + \widehat{\Delta z^*}^{initial}}{z^*} & \text{if } z > z^* + u_2 \end{cases} \tag{17}$$

where $[z^* - u_1, z^* + u_2]$ is the bunching region with diffuse, in which $u_1 = u_2 = 0$ under sharping bunching. The inferred $z^{ct,initial}$ for *shifters* forms the counterfactual density distribution $h^{ct,initial}(z), \forall z \in ((z^* + u_2)\frac{z^* + \widehat{\Delta z^*}^{initial}}{z^*}, \infty)$,[14] whereas the observed density distribution for *always-takers* is the same as counterfactual density distribution, i.e., $h^{ct,initial}(z) = h(z), \forall z \in (z_{min}, z^* - u_1)$.

Next, we obtain the counterfactual density for bunching agents based on the assumption that the counterfactual density distribution is smooth. Specifically, we use the standard approach in the bunching literature to fit a flexible polynomial to the counterfactual distribution for the *always-takers* and *shifters* outside the region $[(z^* - u_1), (z^* + u_2)\frac{z^* + \widehat{\Delta z^*}^{initial}}{z^*}]$, and extrapolate the fitted distribution inside the region. Empirically, we group agents into bins indexed by j, and estimate the following regression:

$$\begin{aligned} h_j^{ct,initial} &= \sum_{k=0}^{p} \beta_k (z_j^{ct,initial})^k + \varepsilon_j \\ \text{if } z_j^{ct,initial} < (z^* - u_1) \quad &\text{or} \quad z_j^{ct,initial} > (z^* + u_2)\frac{z^* + \widehat{\Delta z^*}^{initial}}{z^*}, \end{aligned} \tag{18}$$

where $h_j^{ct,initial}$ is the number of agents in bin j; $z_j^{ct,initial}$ is the inferred z level in bin j based on the initial guess $\widehat{\Delta z^*}^{initial}$; and p is the polynomial order. The counterfactual bin counts in the region

[14] When we relocate *shifters* back to their original location, we reshape observed density distribution $h(z), \forall z \in (z^* + u_2, \infty)$ into $h(z^{ct}\frac{z^*}{z^* + \Delta z^{*,initial}}) \equiv h^{ct,initial}(z^{ct,initial}), \forall z^{ct,initial} \in ((z^* + u_2)\frac{z^* + \widehat{\Delta z^*}^{initial}}{z^*}, \infty)$.

$[(z^*-u_1),(z^*+u_2)\frac{z^*+\widehat{\Delta z^*}^{initial}}{z^*}]$ are obtained as the predicted values from Equation (18).

After recovering the $h^{ct,initial}(z)$ for the full range of z, excess bunching (with diffusion) at the threshold can then be computed as[15]

$$\widehat{B}^{initial}=\int_{z^*-u_1}^{z^*}\left(h(z)-h^{ct,initial}(z)\right)dz+\int_{z^*+1}^{z^*+u_2}\left(h(z)-h^{shift}(z)\right)dz, \tag{19}$$

where $h^{shift}(z)$ denotes the density of *shifters* under the kinked policy. Note that to the right of the bunching region, the observed density distribution contains only shifting agents, and hence, $h^{shift}(z)=h(z)$ for $z>z^*+u_2$. However, within the diffuse region $(z^*,z^*+u_2]$, the observed post-kink density distribution contains both *shifters* and diffused *bunchers*. Assuming that $h^{shift}(z)$ is smooth, we then use the observed distribution $h(z)$ in the region $z>z^*+u_2$ to extrapolate the distribution of shifting agents in the diffusion region $(z^*,z^*+u_2]$.[16]

Third, we compute the updated $\widehat{\Delta z^*}^{updated}$ based on the following relation:

$$\widehat{B}^{initial}=\int_{z^*+1}^{z^*+\widehat{\Delta z^*}^{updated}}h^{ct,initial}(z)dz, \tag{20}$$

and check whether $\widehat{\Delta z^*}^{updated}$ equals $\widehat{\Delta z^*}^{initial}$. If $\widehat{\Delta z^*}^{updated}>\widehat{\Delta z^*}^{initial}$, we increase the value of $\widehat{\Delta z^*}^{initial}$ and repeat the above steps until we have $\widehat{\Delta z^*}^{updated}=\widehat{\Delta z^*}^{initial}$. Following the above process, we obtain the estimated marginal adjustment $\widehat{\Delta z^*}$ and the counterfactual density distribution $\hat{h}^{ct}(z)$.

In addition, following the bunching literature, given the kinked policy and estimated bunching response $\widehat{\Delta z^*}$, we can calibrate e using the equation $\frac{f(D=1|e)}{f(D=0|e)}=\frac{z^*+\widehat{\Delta z^*}}{z^*}$. For example, in Saez (2010), the equivalent equation would be $\frac{(1-t)^e}{(1-t-\Delta t)^e}=\frac{z^*+\widehat{\Delta z^*}}{z^*}$.

Four remarks about our proposed method are worth noting. First, our estimation does not depend on the initial guess value $\widehat{\Delta z^*}^{initial}$, as it converges to the true unique Δz^*. The reason is as follows. Suppose our initial guess $\widehat{\Delta z^*}^{initial}<\Delta z^*$ (the true value). This means $\frac{z^*+\widehat{\Delta z^*}^{initial}}{z^*}<\frac{z^*+\Delta z^*}{z^*}$, and hence, the elasticity $\widehat{e}^{initial}<e$. In other words, our guessed $\widehat{\Delta z^*}^{initial}$ would be consistent with

[15] The excess bunching at the threshold under the sharp bunching is $\widehat{B}^{initial}=h(z^*)-h^{ct,initial}(z^*)$.

[16] Alternatively, we can use the inferred $h^{ct,initial}(z)$ and the relation that $z=z^{ct,initial}\frac{z^*}{z^*+\widehat{\Delta z^*}^{initial}}$ to obtain $h^{1,shift}(z)$ for $z\in(z^*,z^*+u_2]$.

a lower level of bunching around the cutoff, compared to the true value, i.e., $\widehat{B}^{initial} < B$. However, as B is fixed[17] and $B > \widehat{B}^{initial}$, our updated value $\widehat{\Delta z^*}^{updared} > \widehat{\Delta z^*}^{initial}$, indicating our initial guess is too low and we need to increase the value of our initial guess. The self-correcting feature is important to our estimation process and leads to the convergence of the estimated value $\widehat{\Delta z^*}$.

Second and importantly, our method accommodates the fact that *shifters* further away from the policy threshold have less adjustment in z, and therefore, the observed and the counterfactual density distributions to the right of the threshold may not be parallel. Our approach relaxes the parallel shifting assumption by Chetty et al. (2011).[18]

Third, by definition, our method satisfies the *integration constraint* that the number of agents under the observed and counterfactual density distributions should be the same, as our approach moves the exact shifting agents back to their original locations.

Fourth, our method does not depend on the assumption of the counterfactual linear policy. In the main text, we assume the counterfactual is a linear policy with a low tax/co-payment rate. However, if we assume the counterfactual is a linear high tax/co-payment rate, the analysis is still valid with corresponding adjustments. Details are shown in Appendix B. Moreover, regardless of which counterfactual policy we assume, the estimated relation between $z(D=1|e)$ and $z(D=0|e)$ and hence the elasticity remains the same.

3.2 Estimating Counterfactual Outcome Distribution and Parameters

In subsection 2.2, we lay out the framework to estimate causal effects under the kink setting, which incorporates the fact that all agents above the threshold have incentives to adjust their behaviors. Now, we discuss empirical details, in particular, the procedure to recover the counterfactual outcome distribution $y^{ct}()$, which is a crucial step to identify the causal effects of the kinked policy.

Specifically, first, given that *always-takers* do not respond to the kinked policy ($z^{ct} = z$) and

[17] Under sharp bunching, $B = h(z^*) - h^{ct}(z^*)$, where $h^{ct}(z^*)$ mainly depends on the shape of $h^{ct}(z) = h(z) \forall z < z^*$. Therefore, B does not depend much on the initial guess of $\widehat{\Delta z^*}^{initial}$.

[18] Chetty et al. (2011) estimate a regression of the following form:

$$c_j \left(1 + I_{\{z_j > z^*+u_2\}} \frac{\hat{B}}{\sum_{z^*+u_2}^{\infty} c_j}\right) = \beta_0 + \sum_{k=1}^{p} \beta_k (z_j)^k + \sum_{i=z^*-u_1}^{z^*+u_2} \gamma_i I\left[z_j = i\right] + \varepsilon_j.$$

The term $I_{\{z_j > z^*+u_2\}} \frac{\hat{B}}{\sum_{z^*+u_2}^{\infty}} c_j$ is a parallel upward shift of observed density, which captures the change in z for *shifters* such that the integration constraint is met.

pay the same amount of money (or tax) T, their observed outcomes are the same as their counterfactual outcomes, that is, $y_n^{ct} = y_n, \forall n \in$ *always-takers*. Therefore, the counterfactual outcome distribution for *always-takers* is $y^{ct}(z^{ct}) = y(z), \forall z^{ct} < z^*$.

Second, for each *shifter*, in the previous subsection, we have recovered marginal bunchers' responses Δz^* and each shifter's counterfactual location $z^{ct} = z\frac{z^*+\Delta z^*}{z^*}, \forall z > z^*$ which forms the counterfactual density distribution. To make sure that we are comparing the same *shifter* under the counterfactual and the kinked policies, we locate *shifters* back to their initial location, which generates the auxiliary outcome distribution under kinked policy $y^r(z^{ct}), \forall z^{ct} > z^* + \Delta z^*$.[19] It represents each *shifter*'s value of y under the kinked policy, including the direct impacts from changes in z and the impacts from changes in T, while excluding the relocation impacts (as we have located *shifters* back to their counterfactual locations). As shown in Equations (9), there would be both level and slope changes when comparing the counterfactual outcome distribution $y^{ct}(z^{ct})$ with the auxiliary outcome distribution under the kinked policy $y^r(z^{ct})$. Moreover, if we extrapolate the obtained auxiliary distribution $y^r(z^{ct})$ to the cutoff z^*, then the slope and the level change at z^* could be used to calibrate the sufficient statistics μ, λ as shown in Equations (10, 11).[20] These parameters represent how changes in z directly impact y and how changes in T (due to change in z and the kinked policy) impact y.

Empirically, we jointly estimate the counterfactual outcome distribution y^{ct} and the slope and level changes. Specifically, we use the observed (also the counterfactual) outcome distribution for *always-takers* ($y^{ct}(z^{ct}) = y(z), \forall z^{ct} < z^* - u_1$) and the obtained auxiliary outcome distribution for *shifters* ($y^r(z^{ct}), \forall z^{ct} > (z^* + u_2)\frac{z^*+\Delta z^*}{z^*}$) to fit a flexible polynomial distribution, allowing intercept and slope changes at the threshold. [21]

[19] Note $y^r(z^{ct}) \equiv y(z^{ct}\frac{z^*}{z^*+\Delta z^*}), \forall z^{ct} > z^* + \Delta z^*$.

[20] We could also check slope and level changes at other locations, apart from z^*, by plugging z^{ct} with the corresponding value of alternative locations. It does not affect the calibrated value of parameters μ, λ.

[21] In notch setting with just level change of incentives at the threshold, Diamond and Persson (2017) include the term $I\left[z_j^0 \geq z^*\right]$ in their estimation equation to capture the payoff (change in outcome) of just passing the threshold in a world without adjustment in z. In kinked settings, even if agents do not manipulate/adjust their z, the kinked policy would lead to slope change at the threshold. Further, agents do adjust their value of z, leading to level changes on outcome y. Therefore, even if we locate *shifters* back to their initial location, the auxiliary outcome distribution under the kinked policy would still indicate slope and level changes at the threshold, compared to the counterfactual (observed) distribution under the linear policy to the left of the threshold. Therefore, we include both the level and slope changes at the threshold, to capture change in

The estimation equation for the counterfactual outcome distribution is as follows:

$$\begin{aligned} y_j^{reg} &= \sum_{k=0}^{q} \alpha_k (z_j^{ct})^k + a_0 I\left[z_j^{ct} > z^*\right] + a_1 I\left[z_j^{ct} > z^*\right] z_j^{ct} + \varepsilon_j \\ \text{if } z_j^{ct} &< (z^* - u_1) \text{ or } z_j^{ct} > (z^* + u_2)\frac{z^* + \Delta z^*}{z^*} \end{aligned} \tag{21}$$

where j indicates the bin; and q is the polynomial order; $y_j^{reg} = y_j = y_j^{ct}$ for *always-takers* with $z_j^{ct} < (z^* - u_1)$, and $y_j^{reg} = y_j^r$ for *shifters* with $z_j^{ct} > (z^* + u_2)\frac{z^* + \Delta z^*}{z^*}$.

Further, Combined, we can calibrate structural parameters μ, λ

Note the estimated coefficients $\widehat{a_0}$ and $\widehat{a_1}$ reflect the level change and the slope change at the threshold respectively. a_0 captures the level change between the auxiliary outcome distribution and the counterfactual outcome distribution of *shifters*, while a_1 captures the corresponding slope change. Hence, following Equations (10, 11), we calibrate the values of λ, μ, based on the following equations:

$$\begin{aligned} a_0 &= \mu\left(\frac{z^* + \Delta z^*}{z^*} - 1\right) + \lambda(t + \Delta t) z^* \left(\frac{z^*}{z^* + \Delta z^*} - 1\right) \\ a_1 &= \lambda\left((t + \Delta t)\frac{z^*}{z^* + \Delta z^*} - t\right) \end{aligned}$$

With two equations and two unknowns, we can calibrate λ, μ.

Relying on the assumption that the relationship between outcome y and z would be smooth under the counterfactual policy, we obtain the counterfactual outcome distribution from Equation (21) as $\widehat{y_j^{ct}} = \sum_{k=0}^{q} \widehat{\alpha_k} (z_j^{ct})^k$.

Meanwhile, the treated outcome for *shifters* y_j^{shift} in $[z^*, z^* + u_2)$ is unobserved with diffused bunching, given that this region contains both *shifters* and diffused *bunchers* under the kinked policy. However, for the range $z > z^* + u_2$, there is only *shifters* under the kinked policy, therefore, $y_j^{shift} = y_j$ for $z > z^* + u_2$. Therefore, we fit a flexible polynomial to the observed distribution of y_j for *shifters* in the range $z > z^* + u_2$ and extrapolate the fitted distribution to obtain y_j^{shift} in $(z^*, z^* + u_2]$, with the assumption that the relationship between observed outcome $y^{shifter}$ and z under the kinked policy is smooth to the left of $z^* + u_2$.[22]

outcomes for *shifters*.

[22] Alternatively, we can use the inferred $\widehat{y_j^{ct}}$, the estimated $\widehat{a_0}, \widehat{a_1}$, the counterfactual density $\widehat{h_j^{ct}}$

Given that we have recovered the counterfactual density distribution in Equation (18), the counterfactual outcome distribution in Equation (21), and the density and outcome distributions of *shifters* within the diffuse bunching region, we can estimate the impacts of the kinked policy on *bunchers* and *shifters* following Equations (8) and (14).

Remark 8 When comparing our approach with regression kinked design (RKD), there are certain similarities and some differences as well. First, RKD does not allow adjustment of z around the threshold, indicating that agents' heterogeneity is smooth around the threshold. In our estimation process, we mimic this intuition by locating *shifters* back to the counterfactual location of z. Second, in RKD, there is no level change but there is a slope change at the threshold due to the kinked incentives T (e.g., maximum claim on unemployment insurance). That is, even if z does not change, changes in the slope of T at the threshold would lead to a change in the slope of Y at the threshold. Therefore, RKD allows us to estimate the impact of T on Y (i.e., λ in our setup). However, in bunching, even if we locate *shifters* back to the counterfactual location z^{ct}, the fact that their z did change would lead to changes in Y as well. Therefore, on top of the slope change as in RKD, we would have a level change due to (i) direct impact from Δz on y and (ii) impact from ΔT (due to Δz) on y. Therefore, we would have both slope and level changes at the threshold. It is more complex, but it also adds more calculation power in the sense that we can use the level change to identify the direct impact of z on y (i.e., μ in our setup), which is non-identified under RKD as there is no change in z to start with. In terms of policy suggestions, apart from evaluating where to set the cutoff (which is also answered by RKD), we can also evaluate to what extent we should set the difference in marginal incentives below/above the threshold. It enables us to search for the optimal policies within a large scope of choices.

3.3 Discussion on Exclusion Restriction

Recall that in Section 2, we defined the average treatment effects of the policy on bunchers and shifters and showed that estimating the treatment effects requires us to recover the counterfactual density and outcome distributions. Subsections 3.1 & 3.2 demonstrate the steps for estimating the counterfactual distributions, under the assumption that the counterfactual distributions are smooth.

However, one might be concerned whether the counterfactual distributions are correctly estimated, and if not, it would cast bias to the estimated treatment effects. Blomquist et al. (2021)

and the relation that $z = z^{ct}\frac{z^*}{z^*+\Delta z^*}$ to calculate y_j^{shift} for *shifters* with $z \in (z^*, z^* + u_2]$.

pointed out that when the distribution of agent heterogeneity is unrestricted, the estimated counterfactual density distribution (following a parametric method) could be of any form and thus the estimated extent of bunching is not informative of the policy response. Thus, Blomquist et al. (2021) suggest exploring cross-sectional or over-time variation from the policy threshold to help discipline the estimated counterfactual density distribution.

We address the concern in two ways. First, following Blomquist et al. (2021), we suggest exploring the density distribution of the same population before the focal policy threshold starts (i.e., "over-time variation") or the density distribution of another subset of the population which is not subject to the same threshold (i.e., "cross-sectional variation") to infer whether the counterfactual distribution is correctly specified. We can check whether the distribution during these placebo tests follows a similar pattern (or shape) as the estimated counterfactual distribution of the focal group. Alternatively, we can use the distribution of these placebo groups (i.e., the focal group before policy starts, or, other groups which are not subject to the same policy) as the counterfactual distributions.

Second, we would like to clarify that Blomquist et al. (2021)'s critique of the lack of information on the shape of the counterfactual distribution using a single kink policy is less severe in our setting because our method does not require assumptions on the functional form of the counterfactual density distribution on *shifters* and *bunchers*. We infer the counterfactual distribution using the non-parametric method by directly calculating how much each *shifter* has adjusted his/her value of z, which automatically forms the counterfactual density distribution. The only place we used parametric assumption is when inferring the counterfactual density distribution of *bunchers* (the middle part), for which we assume that the counterfactual distribution is smooth and can be extrapolated from the left and the right part of the distribution. In short, our method for estimating counterfactual density does not require a parametric assumption on the whole counterfactual distribution; it uses the parametric assumption only for the middle part of the counterfactual distribution. Therefore, the potential bias is supposed to be less server. Nevertheless, we suggest following Blomquist et al. (2021) by exploring the cross-sectional or over-time variation from the policy threshold. If one finds the counterfactual density and outcome distributions are correctly specified and there is no discontinuity at the kink point, then any difference between the observed and the counterfactual distribution is driven by agents' response to the kinked policy. Therefore, it alleviates the concern that the the estimated treatment effect is due to other reasons, rather than the policy response.

4 Extensions

In this section, we discuss a number of extensions to our baseline framework presented in the previous sections, including the rounding effects, unresponsive to the kinked policy by some agents due to optimization frictions (denoted as *stayers*), the heterogeneity in the structural parameter e, the relabeling behavior of z. In each scenario, we discuss the potential biases with our baseline analysis discussed in the previous sections and the remedy strategies.

4.1 Reference Points

When the policy threshold is a reference point, the excess bunching at such threshold may also capture the reference point effect, which may lead to over-estimated responses compared to the true values. In other words, with one moment (estimated excess bunching mass), there are two underlying structural parameters (i.e., the reference point effect and the policy effect). To isolate the policy effect from the reference point effect, we need an additional empirical moment to jointly identify these two structural parameters. One commonly used approach in the bunching literature is to exploit the excess bunching at similar reference points that are not thresholds to control for the bunching due to the reference point effect at the threshold (e.g., Chetty et al. 2011; Kleven & Waseem 2013; Best & Kleven 2016) with the assumption that reference point effects are same across similar reference points.[23]

Following this literature, we revise the density distribution estimation in Equation (18) by including a set of reference point fixed effects to contain the potential bias from the reference point effects:

$$\begin{aligned} h_j^{ct,initial} &= \sum_{k=0}^{p} \beta_k (z_j^{ct,initial} - z^*)^k + \sum_{r \in R} \gamma_r I\left[\frac{z_j}{r} \in \mathbb{N}\right] + \varepsilon_j \\ \text{if } z_j^{ct,initial} &< (z^* - u_1) \text{ or } z_j^{ct,initial} > (z^* + u_2) \frac{z^* + \widehat{\Delta z^*}^{initial}}{z^*} \end{aligned}$$

[23] In addition, they often assume an equal degree of excess bunching at the same reference point under the treated and counterfactual states. For instance, Chetty et al. (2011) adjust for the interior responses of shifting agents by allowing for an upward shift in the density distribution, which is equivalent to assuming that there are the same degree of excess bunching at reference points between the treated and counterfactual states. We consider the same assumption in all applications.

where $\mathbb{N}$ is the set of reference points; and R is a vector of multiples that capture similar reference points. The counterfactual density distribution is $\hat{h}_j^{ct,initial} = \sum_{k=0}^{p} \hat{\beta}_k (z_j^{ct,initial} - z^*)^k + \sum_{r \in R} \hat{\gamma}_r I\left[\frac{z_j^{ct,initial}}{r} \in \mathbb{N}\right]$.

To address the concern that the reference point effects may generate potential bias in the estimation of the outcome distribution, we take a similar remedy approach. Specifically, we revise the estimation framework of the outcome distribution (21) by including a set of reference point fixed effects:

$$\begin{aligned} y_j^{reg} &= \sum_{k=0}^{q} \alpha_k (z_j^{ct} - z^*)^k + a_0 I\left[z_j^{ct} \leq z^*\right] + a_1 I\left[z_j^{ct} \leq z^*\right](z_j^{ct} - z^*) \\ &\quad + \sum_{r \in R} \rho_r I\left[\frac{z_j}{r} \in \mathbb{N}\right] + \varepsilon_j \\ \text{if } z_j^{ct} &< (z^* - u_1) \text{ or } z_j^{ct} > (z^* + u_2)\frac{z^* + \Delta z^*}{z^*} \end{aligned}$$

The counterfactual outcome distribution is given as $\hat{y}_j^{ct} = \sum_{k=0}^{q} \hat{\alpha}_k (z_j^{ct} - z^*)^k + \sum_{r \in R} \hat{\rho}_r I\left[\frac{z_j^{ct}}{r} \in \mathbb{N}\right]$.

4.2 Stayers

Our framework as mentioned earlier implicitly assumes that all agents behave according to the optimal equation (6) without friction. However, as pointed out in the bunching literature (e.g., Kleven and Waseem 2013), optimization frictions (such as adjustment costs and inattention) may induce agents to stay at their original locations even though they would adjust z in the absence of frictions. We denote these agents as *stayers* and extend our estimation approach to incorporate *stayers* in calculating causal effects.

Specifically, bunching studies often introduce an additional parameter to characterize the adjustment costs in the presence of optimization frictions (explaining the gap between the bunching sizes with and without attenuation from frictions). It then uses additional empirical moments to uncover the parameter corresponding to optimization frictions and to estimate the underlying structural parameter that governs agents' behavior without frictions (e.g., Chetty et al. 2010; Kleven and Waseem 2013; Gelber et al. 2014; Manoli et al. 2016). For example, in the notch design with strictly dominated regions (i.e., the upward tax notches in the labor-leisure decision), Kleven and Waseem (2013) develop the approach that uses the observed density in the strictly dominated re-

gion to estimate the share of *stayers* (with the assumption of a constant share within this region).[24]

However, the change in the marginal incentives across the policy threshold in the kink setting offers only one empirical moment –the size of bunching – for estimation. Instead, kink studies often construct alternative additional moments generated by either multiple thresholds with different-sized kinks or the changes in the size of a kink at a given threshold over time to jointly identify the structural parameters of interest and the friction parameter, with the assumptions that friction and elasticity parameters are the same at multiple thresholds or over time (e.g., Chetty et al. 2010, 2011; Gelber et al 2014). These approaches also apply to our setup.

In addition, we propose a new approach to estimate the share of *stayers* by exploiting changes in the curvature of density distribution under the treated and the counterfactual states. Specifically, we follow the practice of Kleven and Waseem (2013) and others in assuming a fixed share of stayers α at each bin of z. With the introduction of the kinked policy, $(1-\alpha)$ share of shifting agents relocates from z^{ct} to z (with the relation between z and z^{ct} defined in Equation 6), and α share of agents stay unchanged (due to optimization frictions). Such relocation leads to a change of the density distribution from $h^0(z)$ to $h^1(z)$ as follows:

$$h^1(z) = \begin{cases} h^{ct}(z) & \text{, if } z < z^* \\ \int_{z^*}^{z^*+\Delta z^*} (1-\alpha) h^{ct}(z) dz + h^{ct}(z^*) & \text{, if } z = z^* \\ (1-\alpha) h^{ct}(z \times \frac{z^*+\Delta z^*}{z^*}) + \alpha h^{ct}(z) & \text{, if } z > z^*. \end{cases} \tag{22}$$

Specifically, for each bin j of *shifters* (i.e., $z > z^*$), it contains two groups of agents: α share of *stayers* and $(1-\alpha)$ share of relocated shifting agents. For *bunchers*, the density at the threshold contains those bunching from the initial range of $(z^*, z^*+\Delta z^*)$ and those with the initial value of z^*. *Always takers* remain unchanged.

We perform the following procedure to estimate the share of *stayers* and counterfactual density and outcome distributions. For a given guess of the value of α and the shape of a polynomial function of $h^{ct}()$, we use equation (22) to obtain $h()$ to fit the observed density distribution under the kinked policy. We then select the share of *stayers* and estimated polynomial coefficients that

[24] More generally, the downward tax notches and notches in contexts other than the labor-leisure decision do not always contain strictly dominated regions. In such cases, studies (see, for example, Best et al 2015; Manoli and Weber 2016) recover the constant share of *stayers* from a very narrow range above/below the threshold by ruling out extreme preferences.

minimize the mean squared error, [25] which allows us to obtain the estimated $\hat{\alpha}$ and $\hat{\Delta z^*}$. Note changes in the curvature of the density distribution are used to capture the additional parameter α.

Note that the counterfactual outcome values for *stayers* and *shifters* at the same value of z might be different. Without loss of generality, assume the relative difference is captured by β. For a given guess of the value of β and the shape of a polynomial function of $y^{ct}()$, using the estimated Δz^* and equations (10, 11), we fit the observed outcome distribution under the kinked policy. We then select the relative difference in counterfactual outcome between *stayers* and *shifters* and the estimated polynomial coefficients that minimize the mean squared error, which allows us to obtain the estimated β, the counterfactual outcome distribution. Meanwhile, the slope and level change at the threshold from the regression (with some adjustments) are used to calibrate the parameters μ, λ. Note that, similar to the density estimation part, we use changes in the curvature of the outcome distribution to capture the additional parameter β.

4.3 Heterogeneity in Structural Parameter

In our benchmark analysis, we assume homogeneous preference across agents; that is, a single structural elasticity e across agents. Given that agents may have different responses to the policy, we extend our estimation framework to account for heterogeneity in e.

Specifically, consider a joint distribution of innate agents' innate type ϕ and response elasticity e, denoted as $f(\phi, e)$, which determines a counterfactual density distribution $\tilde{h}^{ct}(z, e)$ under the linear policy and $h^{ct}(z) \equiv \int_e \tilde{h}^{ct}(z, e)\, de$. For each value of e, behavior responses can be characterized as in the benchmark model, in which the marginal bunching agents' adjustment Δz_e^* is increasing in e. In the bunching literature with homogeneous preference, the structural parameter e is inferred from the observed excess bunching mass B in the data with one empirical moment linking B to e as derived in Equation 4. However, when there is heterogeneity in e, Equation 4 be-

[25] The intuition is as follows. Suppose in reality there are stayers ($\alpha > 0$). If we impose $\alpha = 0$, we would have a maximum achievable prediction power. However, if we allow $\alpha > 0$, we would have a higher prediction power for the density distribution, by capturing the curvature change. To further help the understanding, consider a log transformation $x = \ln z$. Note $x^{ct} = x + cst$, where $cst = \ln z^* + \Delta z^* z^*$. We draw the density distribution of x. When $\alpha = 0$, we have $\widetilde{h}^{ct}(x) = \widetilde{h}(x - cst)$. However, when $\alpha > 0$, the above equality no longer holds, that is, $\widetilde{h}^{ct}$ and $\widetilde{h}$ are no longer sharing the same functional transformation. Therefore, if in reality $\alpha > 0$ but we impose $\alpha = 0$, our prediction power would be lower. Hence, information about changes in the functional form captures α.

comes $B = \int_e \int_{z^*-\Delta z_e^*}^{z^*-1} \tilde{h}^0(z,e)dzde$. Hence, linking one empirical moment B to multiple parameters e causes the empirical estimation to fall short of the identification freedom and power.

If the dimensions of heterogeneity are known, we can split the whole sample into subsamples according to these determinants, as conducted by Best et al. (2015b). This allows the unbiased estimation within subsamples with relatively homogeneous preferences. However, without the knowledge of heterogeneity, one approach commonly used in the bunching literature to address the freedom issue in the presence of preference heterogeneity is to estimate the average response $E[\Delta z_e^*]$. Specifically, using the procedure proposed in section 3.1 and replacing Δz^* by $E[\Delta z_e^*]$, we can estimate the counterfactual density distribution together with the average response $E[\Delta z_e^*]$ level, and then estimate the auxiliary outcome distribution (by locating agents back to their counterfactual location) and the counterfactual outcome distribution using the procedure discussed in section 3.2 and hence, the treatment impacts.

However, the estimated elasticity and treatment effects essentially represent *the elasticity and treatment effects at the average response*, instead of *the average elasticity and treatment effects*, creating potential aggregation biases. In the following, we use a simple example to discuss the aggregation bias from heterogeneous preference in the kink design and how it affects our estimations of the counterfactual density distribution and the policy effects.[26]

Specifically, consider a case with two groups of agents at each level of z, denoted as L, S. Their shares are denoted as α^L, α^S, with $\alpha^L + \alpha^S = 1$. They hold different structural parameters; without loss of generality, we assume $e^L > e^S$. A larger e implies a larger bunching response, i.e., $\Delta z^{*,L} > \Delta z^{*,S}$. We first discuss the potential biases in the estimation of counterfactual density distribution $h^{ct}(z)$ and the marginal *buncher*'s response Δz^* with the heterogeneity in e.

Consider *shifters* with a value $z_x < z^*$. Suppose we ignore the heterogeneity, we obtain the estimated average response level $\widetilde{\Delta z^*} \equiv \widehat{E[\Delta z_e^*]}$ from the excess mass B and hence would have $\tilde{z}_x^{ct} = z_x\left(\frac{z^*+\widetilde{\Delta z^*}}{z^*}\right)$. Therefore, the estimated counterfactual density at $\tilde{z}_x^{ct}$ is given by $\hat{h}^{ct}(\tilde{z}_x^{ct}) = h(\tilde{z}_x^{ct}\frac{z^*}{z^*+\widetilde{\Delta z^*}}) = h(z_x) = \alpha^L h^{ct}(z_x^{ct,L}) + \alpha^S h^{ct}(z_x^{ct,S})$, where $z_x^{ct,L} = z_x\left(\frac{z^*+\Delta z^{*,L}}{z^*}\right)$ and $z_x^{ct,S} = z_x\left(\frac{z^*+\Delta z^{*,S}}{z^*}\right)$. However, the true counterfactual density at $\tilde{z}_x^{ct}$ should be $h^{ct}(\tilde{z}_x^{ct}) = \alpha^L h^{ct}\left(z_x(\frac{z^*+\widetilde{\Delta z^*}}{z^*})\right) + \alpha^S h^{ct}\left(z_x(\frac{z^*+\widetilde{\Delta z^*}}{z^*})\right)$. Hence, using the average response $\widetilde{\Delta z^*}$ generates a bias in the estimation of counterfactual density h^{ct} at $\tilde{z}_x^{ct}$ as

[26] In the notch design, the literature generally considers such aggregation bias to be small (Kleven, 2016). For example, Kleven and Waseem (2013) discuss the bound of such aggregation bias in the case of notch design and Best et al. (2015b) conduct a rich set of subsample analyses and show that such aggregation bias is very small under the notch design.

$$\begin{aligned}\text{Aggregation Bias in } h^{ct}\left(\tilde{z}_x^{ct}\right) &= \hat{h}^{ct}(\tilde{z}_x^{ct}) - h^{ct}(\tilde{z}_x^{ct}) \\ &= \alpha^L\left[h^{ct}(z_x^{ct,L}) - h^{ct}(\tilde{z}_x^{ct})\right] - \alpha^S\left[h^{ct}(z_x^{ct,S}) - h^{ct}(\tilde{z}_x^{ct})\right].\end{aligned}$$

The degree of the bias in the density estimation depends on three factors: (1) the slope of the counterfactual density $h^{ct}(z)$, which determines the number of agents at $z_x^{ct,L}, \tilde{z}_x^{ct}$ and $z_x^{ct,S}$ under the counterfactual linear state; (2) the relative size of heterogeneous groups in the sample: α^L and α^S; (3) the degree of heterogeneity e^L, e^S, which determines $\Delta z^{*,S}$ and $\Delta z^{*,L}$. When the slope of the counterfactual density $h^{ct}(z)$ of the *shifters* is relatively small (i.e., $h^{ct}(z_x^{ct,L}) \approx h^{ct}(\tilde{z}_x^{ct}) \approx h^{ct}(z_x^{ct,S})$ $\forall n \in shifters$), the bias in the density estimation can be ignored. In this scenario, the estimated average response $\widetilde{\Delta z^*}$ is the weighted average of each heterogeneous group's response, with the relative share of each group as the weights, i.e., $\widetilde{\Delta z^*} = \frac{\alpha^S \Delta z^{*,S} + \alpha^L \Delta z^{*,L}}{\alpha^S + \alpha^L}$. [27]

Next, we consider the potential bias in the average policy effects from the heterogeneity in e. Note that a crucial step in our proposed estimation framework of the policy effects is to use the observed outcome distribution of *shifters* to estimate an auxiliary outcome distribution $y^r(z)$, which relies on the estimation of Δz^*. When there is heterogeneity in the structural parameter e and yet we ignore the heterogeneity by adjusting each *shifter*'s location using the estimated average response level $\widetilde{\Delta z^*}$ (i.e., $z^{ct} = z\left(\frac{z^* + \widetilde{\Delta z^*}}{z^*})\right)$, there would be biases in the estimation of the auxiliary

[27] To see this, the excess bunching is composed of bunching agents from both L and S group: i.e., $B = \alpha^L \int_{z^*}^{z^* + \Delta z^{*,L}} h^{ct}(z)dz + \alpha^S \int_{z^*}^{z^* + \Delta z^{*,S}} h^{ct}(z)dz$. Given the estimated counterfactual density $\hat{h}^{ct}(z)$ and excess bunching $\hat{B} = B$, we estimate the average response Δz^* using $\hat{B} = \int_{z^*}^{z^* + \widetilde{\Delta z^*}} \hat{h}^{ct}(z)dz$. Thus, we have $\alpha^S \int_{z^* + \Delta z^{*,S}}^{z^* + \widetilde{\Delta z^*}} h^{ct}(z)dz - \alpha^L \int_{z^* + \widetilde{\Delta z^*}}^{z^* + \Delta z^{*,L}} h^{ct}(z)dz = 0$. When $h^{ct}(z)$ is approximately locally linear, the above equation can be approximated as:

$$\begin{aligned}&\alpha^S \beta^S \left(\widetilde{\Delta z^*} - \Delta z^{*,S}\right) - \alpha^L \beta^L \left(\Delta z^{*,L} - \widetilde{\Delta z^*}\right) \\ =\ &\left(\alpha^S \beta^S + \alpha^L \beta^L\right) \widetilde{\Delta z^*} - \left(\alpha^S \beta^S \Delta z^{*,S} + \alpha^L \beta^L \Delta z^{*,L}\right) = 0\end{aligned}$$

where $\beta^S = \frac{h^{ct}(z^* + \Delta z^{*,S}) + h^{ct}(z^* + \widetilde{\Delta z^*})}{2}$ and $\beta^L = \frac{h^{ct}(z^* + \Delta z^{*,L}) + h^{ct}(z^* + \widetilde{\Delta z^*})}{2}$. Hence, the estimated average response $\widetilde{\Delta z^*} = \frac{\alpha^S \beta^S \Delta z^{*,S} + \alpha^L \beta^L \Delta z^{*,L}}{\alpha^S \beta^S + \alpha^L \beta^L}$. If the slope of the counterfactual density $h^{ct}(z)$ to the left of z^* is relatively small, $\beta^S = \beta^L = h^{ct}(z^* + \widetilde{\Delta z^*})$, the estimated average response is the weighted average of each heterogeneous group's response, with the relative share of each group as the weights, i.e., $\widetilde{\Delta z^*} = \frac{\alpha^S \Delta z^{*,S} + \alpha^L \Delta z^{*,L}}{\alpha^S + \alpha^L}$.

outcome distribution and policy impacts.

Similarly, consider *shifters* with a value $z_x > z^*$. Suppose we ignore heterogeneity in the parameter e and adjust *shifters* at z_x to their counterfactual locations using the estimated average response $\widetilde{\Delta z^*}$, we have $\tilde{z}_x^{ct} = z_x \left(\frac{z^* + \widetilde{\Delta z^*}}{z^*} \right)$. At the point $\tilde{z}_x^{ct}$, the estimated auxiliary outcome distribution would be

$$\widehat{E\left[y^r(\tilde{z}_x^{ct})\right]} = y(\tilde{z}_x^{ct} \frac{z^*}{z^* + \widetilde{\Delta z^*}}) = y(z_x) = \alpha^L y^{L,r}(z_x^{ct,L}) + \alpha^S y^{S,r}(z_x^{ct,S})$$

where $y^{L,r}(z_x^{ct,L})$ denote the outcome under the kinked policy for *shifters* of group L whose counterfactual values are $z_x^{ct,L} = z_x \frac{z^* + \Delta z^{*,L}}{z^*}$; and, vice versa for $y^{S,r}(z_x^{ct,S})$.
However, the true auxiliary outcome at $\tilde{z}_x^{ct}$ should be

$$E\left[y^r | \tilde{z}_x^{ct}\right] = \alpha^L y^{L,r}(z_x \frac{z^* + \Delta z^*}{z^*}) + \alpha^S y^{S,r}(z_x \frac{z^* + \Delta z^{*,S}}{z^*}).$$

Hence, using the average response $\widetilde{\Delta z^*}$ generates the bias in the estimation of y^r at $\tilde{z}_x^{ct}$ as

$$\begin{aligned}
\text{Aggregation Bias in } \left[y^r | \tilde{z}_x^{ct}\right] &= \widehat{E\left[y^r | \tilde{z}_x^{ct}\right]} - E\left[y^r | \tilde{z}_x^{ct}\right] \\
&= \alpha^L \left(y^{L,r}(z_x \frac{z^* + \Delta z^{*,L}}{z^*}) - y^{L,r}(z_x \frac{z^* + \widetilde{\Delta z^*}}{z^*}) \right) \\
&+ \alpha^S \left(y^{S,r}(z_x \frac{z^* + \Delta z^{*,S}}{z^*}) - y^{S,r}(z_x \frac{z^* + \widetilde{\Delta z^*}}{z^*}) \right).
\end{aligned}$$

Similarly, the degree of bias in the outcome estimation depends on: (1) the slope of the auxiliary outcome distribution $y^r(z)$, which determines the values at $z_x^{ct,L}$, $\tilde{z}_x^{ct}$ and $z_x^{ct,S}$; (2) the share of heterogeneous groups α^S and α^L; (3) the degree of heterogeneity which determines $\Delta z^{*,L}$ and $\Delta z^{*,S}$; and (4) the bias in the counterfactual density estimation $h^{ct}(z)$. When we have small aggregation biases in the estimation of counterfactual density distribution and the auxiliary outcome distribution holds a small slope, we can obtain a good approximation of the outcome distribution and of the average treatment effects.

In addition, we propose another approach to address the aforementioned aggregation bias with an alternative identifying assumption. Specifically, consider a logarithm transformation of z, denoted as $r \equiv ln(z)$. When the density distribution of x and the outcome distribution of y against r

are linear,[28] we can obtain unbiased estimates of counterfactual density and outcome distributions, and thus the unbiased estimate of average treatment effects in the presence of heterogeneity. The reasoning is as follows.

With the logarithm transformation, each *shifter*'s adjustment of r in response to the introduction of kinked policy becomes a constant, i.e., $r - r^{ct} = ln\frac{z^*}{z^*+\Delta z^*}$, which leads to a parallel-rightward shift of the density curve for the region to the left of cutoff $r^* \equiv lnz^*$. [29]In other words, the post-kink density distribution has the same slope as the counterfactual one, but with different intercepts. We use the same illustrative example: two groups of agents at each level of r, with the shares and structural parameters being α^L, α^S and e^L, e^S, respectively. For *shifters* with a value $r_x > r^*$, we have:

$$\begin{aligned} h(r_x) &= \alpha^L h^{ct}(r_x^{ct,L}) + \alpha^S h^{ct}(r_x^{ct,S}) \\ &= \alpha^L h^{ct}(r_x - \Delta r^{*,L}) + \alpha^S h^{ct}(r_x - \Delta r^{*S}) \\ &= h^{ct}(r_x) - \frac{dh}{dr} \times (\alpha^L \Delta r^{*,L} + \alpha^S \Delta r^{*,S}) \end{aligned}$$

Given that the amount $-\frac{dh}{dr} \times (\alpha^L \Delta r^{*,L} + \alpha^S \Delta r^{*,S})$ is a constant, the counterfactual density distribution to the left of r^* is also a downward shift of the observed one. Hence, using the observed density distribution for *shifters* and for *always-takers* to fit a linear distribution and allowing an intercept change at the threshold r^*, we can still recover an unbiased counterfactual density distribution $h^{ct}(r)$ in the presence of heterogeneity. In addition, based on the value of change at the threshold, we can recover the value of $(\alpha^L \Delta r^{*,L} + \alpha^S \Delta r^{*,S})$.

Similarly, when the outcome distribution of y against r is linear, under the kinked policy the outcome distribution of *shifters* is a parallel shift of the counterfactual outcome distribution along the x-axis. We then have:

[28] r is linear when z is exponentially distributed with a parameter within (0,1). In the data, variables often follow such a pattern under which there are more numbers of small values and only a few large values. One can plot the density of $r \equiv \ln z$ and check whether the density is close to linear in the estimation region.

[29] In terms of notations, we use r here to represent the equivalent terms of $\ln z$.

$$
\begin{aligned}
y(r_x) &= \alpha^L y^{r,L}(r_x^{ct,L}) + \alpha^S y^{r,L}(r_x^{ct,S}) \\
&= \alpha^L y^{r,L}(r_x - \Delta r_x^{*,L}) + \alpha^S y^{r,L}(r_x - \Delta r_x^{*,S}) \\
&= y^r(r_x) - \frac{dy}{dr} \times (\alpha^L \Delta r_x^{*,L} + \alpha^S \Delta r_x^{*,S})
\end{aligned}
$$

where $\frac{dy}{dr}$ is the slope of outcome distribution; and $-\frac{dy}{dr} \times (\alpha^L \Delta r_x^{*,L} + \alpha^S \Delta r_x^{*,S})$ is the constant amount of outcome distribution shift for *shifters*. Given the observed outcome distribution for *always-takers*, we can recover $\frac{dy}{dr}$. Given the estimated value of $(\alpha^L \Delta r_x^{*,L} + \alpha^S \Delta r_x^{*,S})$ from the density distributions, we can obtain the value of $\frac{dy}{dr} \times (\alpha^L \Delta r_x^{*,L} + \alpha^S \Delta r_x^{*,S})$, which allows us to obtain an unbiased estimate of the auxiliary outcome distribution for *shifters*. Combing the observed outcome distribution of *always-takers* and the auxiliary outcome distribution of *shifters*, we can follow the procedures in the main analysis to calibrate the structural parameters (μ, λ) and estimate the treatment effects. These estimates are unbiased because the auxiliary outcome distribution is unbiased. Meanwhile, the corresponding identifying assumptions of the density distribution of $r \equiv \ln z$ and the outcome distribution being linear are testable by directly examining the distribution figures.

To sum up, in the presence of heterogeneity in e, here are several potential solutions according to the following scenarios:

1. If the dimensions of heterogeneity are known, we can split the whole sample into subsamples according to these determinants, as conducted in Best et al. (2015b). This allows the estimation within subsamples with relatively homogeneous preferences.

2. If density distribution $h^0(z)$ has a small slope and with small group heterogeneity, we can obtain a good approximation of the average bunching response and achieve small aggregation bias in the estimation of counterfactual density distribution. Furthermore, if the outcome distribution also holds a small slope, we can obtain a good approximation of the average treatment effects.

3. If the density and outcome distribution of the logarithm transformation of z is linear, we can obtain an unbiased estimation of the counterfactual density distribution and the auxiliary outcome distribution as well as the average treatment effects.

4.4 Relabelling

Faced with monetary incentives, agents may engage in misreporting or other relabelling behavior, causing their reported value of z to be potentially different from the real response. For example, in the study of tax incidence on R&D investment, Chen et al. (2021) point out that relabelling is an important channel in which firms adjust their R&D expenditure upwards, to benefit from tax reduction. To investigate whether and how relabeling may affect our causal analysis, we extend our proposed estimation approach to incorporate the relabelling behavior.

Agents' optimal degree of relabelling is determined by the marginal cost (e.g., cost of cooking the books and potential risk of being caught, related to the cost function) and the marginal benefit of it (e.g., tax saving, related to the policy). We first consider a setting where agents share the same cost function and then extend our framework to a more general situation where different groups of agents may hold different cost functions (e.g. it is easier for self-employed to misreport their income than wage-earners).

Specifically, we assume that relabeling cost depends on the absolute value and the relative degree of relabeling following Chen et al. (2021). That is, we assume $c \times z^{rl} \times g(\delta)$, where c is a fixed parameter; $\delta \equiv \frac{z^{rl}-z^{rp}}{z^{rl}}$ summarizes the relabeling behavior by the agents; z^{rp}, z^{rl} are the reported and real values of z respectively; and $g'(\delta) > 0, g''(\delta) > 0, g(0) = 0$. Hence, the marginal cost of an additional degree of relabelling is $cz^{rl}g'(\delta)$.

Consider the counterfactual linear policy with a low tax/co-payment rate at t. The benefit of relabelling is the money saved, i.e. $(z_n^{rl,ct} - z_n^{rp,ct})t \equiv \delta_n^{ct} z_n^{rl,ct} t$, therefore, the marginal benefit of an additional degree of relabelling is $z_n^{rl,ct} t$. Recall the marginal cost of an additional degree of relabelling is $cz^{rl,ct}g'(\delta_n^{ct})$. Agent n optimally chooses his/her degree of relabelling δ_n^{ct} by setting marginal benefit equalizing marginal cost, i.e., $g'(\delta_n^{ct}) = \frac{t}{c}$, which implies $\delta_n^{ct} = g'^{-1}(\frac{t}{c})$. Note $g'^{-1}(\frac{t}{c})$ is constant for all agents, therefore, we can rewrite it as $\delta^{ct} = g'^{-1}(\frac{t}{c}), \forall n$.

Next, consider the kinked policy, which sets a higher tax rate at $t + \Delta t$ if $z_n > z^*$ and leaves the tax/co-payment rate unchanged at t if $z_n \leq z^*$. Similar to the analysis in Section 2, the introduction of the kinked policy divides agents into three groups. First, agents with $z_n^{rp,ct} \leq z^*$ (i.e., *always-takers*) face no change in marginal incentives and set $z_n^{rp,1} = z_i^{rp,ct} \leq z^*$ and $\delta = \delta^{ct} = g'^{-1}(\frac{t}{c})$. Second, agents with $z_n^{rp,ct} > z^* + \overline{\Delta z^*}$ (i.e., *shifters*) face a change in the marginal benefit and adjust their optimal responses accordingly. Specifically, all *shifters* set $\delta = g'^{-1}(\frac{t+\Delta t}{c})$. Also, *shifters* change their reported value of z by a constant percentage with $\frac{z_n^{rp}}{z_n^{rp,ct}} = \frac{z^*}{z^* + \overline{\Delta z^*}}$, where $\overline{\Delta z^*}$

denotes the response in the reported value of z by the marginal bunching agent. Third, agents with $z_n^{rp,ct} \in (z^*, z^* + \overline{\Delta z^*}]$ (i.e., *bunchers*) also face a change in their marginal incentives but are subject to a corner solution. However, these agents bunch at the cutoff $z_n^{rp} = z^*$, and choose different optimal degrees of relabelling δ_n, depending on how far away their counterfactual value $z_n^{rp,ct}$ is from the cutoff z^*. Detailed proofs are shown in Appendix D**??**.

To summarize, the optimal reported value z_n^{rp}, the optimal degree of relabeling δ_n and the optimal real value z_n^{rl} under the kinked policy are given as

$$z_n^{rp} = \begin{cases} z_n^{rp,ct} & \text{if } z_n^{rp,ct} \leq z^* \\ z^* & \text{if } z_n^{rp,ct} \in (z^*, z^* + \overline{\Delta z^*}] \\ z_n^{rp,ct} \frac{z^*}{z^* + \overline{\Delta z^*}} & \text{if } z_n^{rp,ct} > z^* + \overline{\Delta z^*} \end{cases} \tag{23}$$

$$\delta_n = \begin{cases} g'^{-1}(\frac{t}{c}) & \text{if } z_n^{rp,ct} \leq z^* \\ \left(0, g'^{-1}(\frac{t+\Delta t}{c})\right] & \text{if } z_n^{rp,ct} \in (z^*, z^* + \overline{\Delta z^*}] \\ g'^{-1}(\frac{t+\Delta t}{c}) & \text{if } z_n^{rp,ct} > z^* + \overline{\Delta z^*} \end{cases} \tag{24}$$

$$z_n^{rl} = z_n^{rp} \frac{1}{(1-\delta_n)} = \begin{cases} z_n^{rl,ct} & \text{if } z_n^{rp,ct} \leq z^* \\ z^* \frac{1}{(1-\delta_n)} & \text{if } z_i^{rp,ct} \in (z^*, z^* + \overline{\Delta z^*}] \\ \frac{z^*}{z^* + \overline{\Delta z^*}} \frac{1-g'^{-1}(\frac{t}{c})}{1-g'^{-1}(\frac{t+\Delta t}{c})} z_n^{rl,ct} & \text{if } z_n^{rp,ct} > z^* + \overline{\Delta z^*} \end{cases} \tag{25}$$

The estimation of the treatment effect crucially depends on inferring the counterfactual density and outcome distributions, i.e., $h^{ct}(z)$ and $y^{ct}(z)$. Since all *shifters* adjust their reported (observed) value of z by the same percentage, we can apply the same algorithm as in the baseline analysis to recover the counterfactual density distribution of the reported z and the reported marginal *buncher*'s response. Hence, $h^{ct}(z)$, $\overline{\Delta z^*}$ is unbiasedly estimated. Therefore, we can locate the agents back to their counterfactual locations and compare the same agents under the kinked pol-

icy and the counterfactual policy, giving us unbiased estimates of treatment effects on *shifters* and *bunchers*.[30]

However, even though the estimation procedures on marginal bunching response and the treatment effects on shifters and bunchers are still correct under potential relabelling or misreporting, given that the real responses are smaller, we would anticipate a reduction in the magnitude of the impacts. Redefine $\mu \equiv \frac{\Delta y}{\Delta z^{rl}/z^{rl}}$. Mathematically, the reasons are as follows:

$$\begin{aligned}\tau_y^{TE,shifter} &= E[y_n - y_n^{ct} | n \in shifters] \\ &= \mu\left(\frac{z^*}{z^* + \overline{\Delta z^*}} \frac{1 - g'^{-1}(\frac{t}{c})}{1 - g'^{-1}(\frac{t+\Delta t}{c})} - 1\right) - \lambda E(z_n^{rp,ct})\left((t+\Delta t)\frac{z^*}{z^* + \overline{\Delta z^*}} - t\right) + \lambda \Delta t \times z^*\end{aligned}$$

The last equality is based on the assumption that agents have the same preferences (and parameters).

Recall we use the slope and level changes at the threshold z^* when comparing the observed outcome of *always-takers* and the obtained auxiliary outcome of *shifters* (when being located back to the counterfactual locations). Accordingly, the equations for calibrating our structural parameters μ, λ would change. Specifically, Equation (11) for the slope change would remain the same, but Equation (10) for the level change would be:

$$\text{Level Change at } z^* = \mu_n\left(\frac{z^*}{z^* + \overline{\Delta z^*}} \frac{1 - g'^{-1}(\frac{t}{c})}{1 - g'^{-1}(\frac{t+\Delta t}{c})} - 1\right) \tag{26}$$

$$- \lambda(t+\Delta t)z^*\left(\frac{z^*}{z^* + \overline{\Delta z^*}} - 1\right) \tag{27}$$

Under potential relabelling, to calibrate parameters μ, λ, we need an additional moment to identify the relabelling cost parameter c. One possibility is to exploit variations in changes in the marginal incentives across different thresholds.

While the previous analysis assumes that all agents share the same cost function of relabelling (i.e., $c \times z^{rl} \times g(\delta)$, where c is constant for all agents), it could be possible in reality that relabeling cost functions are different across agents due to differential predetermined character-

[30]Note estimating counterfactual outcome distribution is based on the observed distribution of *always-takers* and the extrapolation via assumptions on smooth counterfactual outcome distribution. It does not require information on observed outcome distribution of *shifters*. Therefore, it is also unbiased.

istics. If we know how to classify agents into subgroups with the homogeneous cost function of relabelling within each subgroup, we can then analyze by subgroups, and there is no bias for each subgroup estimation. However, without the knowledge of which agents belong to which group, we have to conduct the analysis using the full sample. In this scenario, the estimated average response level $\widetilde{\Delta z^*}$ based on one empirical moment B contains the potential aggregation bias, which leads to the same aggregation bias as discussed in Subsection 4.3 on the heterogeneity in response elasticity. The solution to the situation with heterogeneous relabeling costs across agents is the same as the solutions to the heterogeneous preference in Subsection 4.3.

4.5 Diffusion

In our analyses mentioned above, we consider diffusion behavior for bunchers; that is, there is no sharp bunching exactly at the kink point as bunching agents cannot target at the kink point precisely. One may then be concerned whether diffusion behavior also happens for other agents. Specifically, shifters adjust their values of z when a kinked policy is introduced, and may not target precisely as well.[31] Whether and how the diffusion by shifters biases our estimated counterfactual density distribution and then the causal estimates? In this subsection, we discuss sources of potential biases in the estimated counterfactual density distribution, excess bunching, the counterfactual outcome distribution, and the treatment effect, when there is diffusion for both shifters and bunchers.[32]

Denote the observed effort choice as z and the optimal targeted effort choice as $z^{targetted}$, with $z_i = z_i^{targetted} + \varepsilon_i$, where ε_i denote the degree of diffusion for agent i. Hence, $\varepsilon_i > 0$ indicates overshooting behavior, $\varepsilon_i < 0$ indicates undershooting behavior, and $\varepsilon_i = 0$ suggests precise targeting. We start with the case that the degree of diffusion for each shifter is a random draw from a common i.i.d. distribution, and then investigate the case that different groups of agents draw their diffusion degrees from different i.i.d. distributions.

Under the first scenario, the degree of diffusion for each shifter is a random draw from the same distribution $g(\varepsilon)$ with the mean value being $\mu(\varepsilon) = 0$ and variance being $Var(\varepsilon) = \sigma^2$. The observed density distribution can be written as $h(z) = \int_\varepsilon h^{ct}\left((z-\varepsilon)\frac{z^*+\Delta z^*}{z^*}\right)g(\varepsilon)d\varepsilon$, where $h^{ct}()$

[31] Note there are over-shooting and under-shooting at each point of z; therefore, we may not observe excess bunching in the shifter's distribution even if shifters are subject to diffusion.

[32] The general practice to deal with only the diffusion bunching is to treat the overall amount of excess bunching around the kink point as the policy response (e.g., Saez, 2010).

denotes the counterfactual density distribution and $\frac{z^*}{z^*+\Delta z^*}$ denotes the marginal buncher's relative change in z under the kinked policy.

Hence, bias arises from diffusion as $h^{ct}\left((z-\varepsilon)\frac{z^*+\Delta z^*}{z^*}\right) \neq h^{ct}\left(z\frac{z^*+\Delta z^*}{z^*}\right)$. Specifically, the bias in the estimated counterfactual density is small when $Var(\varepsilon_i)$ is small (i.e., less degree of diffusion) or when the slope of $h^{ct}()$ is small. To illustrate this point, consider a special situation where $\varepsilon_i = 0$ with 60% probability, $\varepsilon_i = \varepsilon$ with 20% probability, and $\varepsilon_i = -\varepsilon$ with 20% probability. Then, we have $h(z) = h^{ct}\left(z\frac{z^*+\Delta z^*}{z^*}\right) * 60\% + h^{ct}\left((z-\varepsilon)\frac{z^*+\Delta z^*}{z^*}\right) * 20\% + h^{ct}\left((z+\varepsilon)\frac{z^*+\Delta z^*}{z^*}\right) * 20\%$. If $h^{ct}\left((z-\varepsilon)\frac{z^*+\Delta z^*}{z^*}\right) \approx h^{ct}\left((z+\varepsilon)\frac{z^*+\Delta z^*}{z^*}\right) \approx h^{ct}\left(z\frac{z^*+\Delta z^*}{z^*}\right)$, there is no bias even if we ignore shifters' diffusion (i.e., by assuming $h(z) = h^{ct}\left(z\frac{z^*+\Delta z^*}{z^*}\right)$). Hence, when $Var(\varepsilon_i)$ is small or the slope of $h^0()$ is small, we have less bias when ignoring shifters' diffusion.

Similarly, the outcome distribution is $y(z) = \int_\varepsilon y^r\left((z-\varepsilon)\frac{z^*+\Delta z^*}{z^*}\right)g(\varepsilon)d\varepsilon$, where $y^r()$ denotes the auxiliary outcome distribution (by locating *shifters* back to their counterfactual locations). The bias from diffusion is due to that $y^r\left((z-\varepsilon)\frac{z^*+\Delta z^*}{z^*}\right) \neq y^r\left(z\frac{z^*+\Delta z^*}{z^*}\right)$. When $Var(\varepsilon_i)$ is small (i.e., less degree of diffusion) or when the slope of $y^r()$ is small, the bias in the estimated auxiliary outcome distribution (when ignoring the diffusion) is small.

Note that the treatment effects on shifters and bunchers depend on the estimation of the counterfactual density, the auxiliary outcome distribution, and the counterfactual outcome distribution. Hence, the bias from diffusion is small when (i) $Var(\varepsilon_i)$ is small or (ii) when the slopes of h^{ct} and y^r are small. In practice, one can check the diffusion variance by exploring the diffusion pattern around the cutoff, check the slope of h^{ct} by exploring the slope of the density for *always-takers*, and check the slope of the auxiliary outcome distribution y^r of *shifters* to understand the potential degree of bias.

Then, we consider a more general setting in which different groups of agents randomly draw their degree of diffusion from different distributions. For example, the self-employed are better at targeting their annual income at the cutoff than the wage-earners. Specifically, assume there are M groups of agents at each value of z in the counterfactual state (i.e., the linear policy), with the share of each group denoted as α_m. Each agent i belonging to group m randomly draws his/her degree of diffusion ε_i from the density distribution $g_m(\varepsilon)$, with mean value $\mu_m(\varepsilon) = 0$ and variance as $Var_m(\varepsilon) = \sigma_m^2$.

The observed density distribution is shown as $h(z) = \sum_m \alpha_m \int_\varepsilon h^{ct}\left((z-\varepsilon)\frac{z^*+\Delta z_m^*}{z^*}\right)g_m(\varepsilon)d\varepsilon$. The bias from diffusion is due to two reasons: first, $h^{ct}\left((z-\varepsilon)\frac{z^*+\Delta z_m^*}{z^*}\right) \neq h^{ct}\left(z\frac{z^*+\Delta z_m^*}{z^*}\right)$; and second, $\sum_m \alpha_m h^{ct}\left(z\frac{z^*+\Delta z_m^*}{z^*}\right) \neq h^{ct}\left(z\frac{z^*+\Delta z}{z^*}\right)$, where Δz is the estimated marginal buncher's response when

ignoring preference heterogeneity. That is, biases come from neglecting the shifters' diffusion and neglecting the heterogeneity in the structural parameter. When $\sum_m \alpha_m \sigma_m$ is small (i.e., the average dispersion in the degree of diffusion is small) and when the slope of $h^{ct}()$ is small, we are back to the scenario with preference heterogeneity discussed in Section 4.3.

The observed outcome distribution is $y(z) = \frac{1}{\sum_m \alpha_m} \sum_m \alpha_m \int_\varepsilon y_m^r\big((z-\varepsilon)\frac{z^*+\Delta z_m^*}{z^*}\big) g_m(\varepsilon) d\varepsilon$, where $y_m^r()$ denote the auxiliary outcome distribution of group m. similarly, the bias from diffusion is generated due to two reasons: first, $y_m^r\big((z-\varepsilon)\frac{z^*+\Delta z_m^*}{z^*}\big) \neq y^r\big(z\frac{z^*+\Delta z_m^*}{z^*}\big)$; second, $\frac{1}{\sum_m \alpha_m} \sum_m \alpha_m y_m^r\big(z\frac{z^*+\Delta z_m^*}{z^*}\big) \neq y^r\big(z\frac{z^*+\Delta z^*}{z^*}\big)$. Therefore, when $\sum_m \alpha_m \sigma_m$ is small and when the slope of $y_m^r()$ is small, we are back to the scenario with heterogeneity, which is discussed in Subsection 4.3.

4.6 Alternative Counterfactual Policy: linear high tax/co-payment rate

Our baseline analysis assumes that the counterfactual policy is a linear low tax/co-payment rate, i.e., the same as the policy below the cutoff. As agents with values above the cutoff face a higher marginal tax/co-payment rate under the kinked policy, they will reduce their value, leading to a "bunching down" design. Meanwhile, agents with values below the above face the same marginal incentive and pay the same amount of fees under the kinked policy. They are denoted as *always-takers*. Accordingly, we have proposed an estimator for quantifying the impact of the kinked policy on these agents: *bunchers* and *shifters*

Alternatively, we might consider an alternative counterfactual policy with a linear high tax/co-payment rate $(t+\Delta t)$. Compared to this new counterfactual policy, agents below the cutoff face a lower tax/co-payment under the kinked policy and hence adjust their values of z upwards, leading to a "bunching up" design. Meanwhile, as agents above the cutoff face the same marginal incentive under the kinked policy, they won't change their values of z. We denote them as *never-takers*. However, in terms of outcomes, because *never-takers* do receive a lump-sum transfer under the kinked policy (compared to the new counterfactual policy)[33], their outcome values might change. This composes the key difference for analyzing the policy impacts under "bunching up" and "bunching down" designs.

Specifically, we cannot take the observed outcome distribution of *never-takers* as the new counterfactual outcome distribution; instead, we need to adjust the impact from the lump-sum

[33]Denote the new counterfactual policy as $T^{ct,new}(z) = (t+\Delta t)z$. For agents above the cutoff, under the kinked policy, we have $T(z) = (t+\Delta t)z - \Delta t z^*, \forall z > z^*$. Therefore, the lump-sum transfer between the kinked policy and the new counterfactual policy is $T^{ct,new}(z) - T(z) = \Delta t z^*, \forall z > z^*$.

transfer. This is doable because we have a parameter λ which captures the impact of money T on outcome y and we also know how large the money change is (i.e., $\Delta t \times z^*$). Therefore, with modifications, we can still use the level change and the slope change at the cutoff z^* between the observed outcome distribution of *never-takers* and the auxiliary outcome distribution of *shifters* to calibrate the parameters μ, λ and estimate the policy impacts (after addressing the impact from lump-sum transfer). Details are shown in Appendix B. One thing to note is that, in the "bunching up" setup, we assume that the impact from the lump-sum transfer shares the same parameter λ. This assumption is more likely to be valid when the level change $\Delta t \times z^*$ is relatively small.

5 Application: Coinsurance Policy In China

We apply our aforementioned bunching technique to identify the causal impacts of the coinsurance policy on the patients' outpatient behaviors in China. Specifically, we first introduce the healthcare system in China and the medical claim data for our empirical analysis. Next, we present the bunching evidence to examine patients' responses to the coinsurance policy. Then, we apply our causal inference framework to study the policy effect.

5.1 Healthcare System in China

China established the current health insurance system since the late 1990s, and gradually achieved universal health insurance coverage. The Urban Employee Basic Medical Insurance (UEBMI) was first introduced in 1998, covering formal sector workers in the urban area. This was followed by the gradual introduction of the New Cooperative Medical Scheme (NRCMS) during the period of 2003-2008 targeting the rural population, and then the Urban Resident Basic Medical Insurance (URBMI) launched in 2007 targeting urban residents who were not covered by the UEBMI (i.e., the unemployed, children, students and the disabled in urban areas). Starting in 2010, the Chinese government gradually integrated NRCMS and URBMI and established a unified Urban and Rural Residents Basic Medical Insurance Scheme (URRBMI) to bridge the gap in medical care between rural residents and urban residents who are not working. These basic health insurance programs (i.e., URRBMI and UEBMI) expanded at a remarkable pace, covering more than 92% of the urban

population and 97% of the rural population in 2011 in China (Yu, 2015).[34]

The benefits depend on the medical insurance catalogs and the program's cost-sharing design. Specifically, the medical insurance catalogs specify the payment scopes and prices of drugs, items of diagnosis treatment, and standards of medical service facility, which are the same for both UEBMI and URRBMI. The cost-sharing design consists of the deductibles, copayment rates (τ), and the maximum amounts payable ($z^*(1-\tau)$), which are designed separately for outpatient and inpatient care, vary across different tiers of hospitals and are different under UEBMI and URRBMI schemes. Specifically, the insurance benefits (Benefits) and hence the annual out-of-pocket expenses (Out-of-Pocket) under the insurance scheme are shown as:

$$\text{Benefits} = \begin{cases} z \times (1-\tau) & \text{if } z \leq z^* \\ z^* \times (1-\tau) & \text{if } z > z^* \end{cases} \tag{28}$$

$$\text{Out-of-Pocket} = \begin{cases} z \times \tau & \text{if } z \leq z^* \\ z \times 1 - z^* \times (1-\tau) & \text{if } z > z^* \end{cases} \tag{29}$$

where z denotes the annual medical expenses eligible for insurance coverage (i.e., annual medical expenses within the medical insurance catalog with the total deductibles subtracted); z^* denotes a statutory cutoff; τ denotes the co-payment rate (hence $1-\tau$ denotes the reimbursement rate when $z < z^*$). The values of z^* and τ depend on the insurance schemes, with a lower reimbursement rate (hence a larger copayment rate τ) and a lower maximum amount payable (i.e., a smaller threshold z^*) under the URRBMI, compared to the UEBMI.

[34]The premiums for UEBMI are usually determined by the employee's average monthly wages in the previous year and are jointly borne by the employer and the employees concerned. It is usually 2% of the salary for employees and 6% of the salary for employers. As for URRBMI, a large portion of the premiums are subsidized by the government, with enrollees contributing a small part.

5.2 Data and Analysis Sample

Our empirical analysis draws on a dataset covering the universe of visit-level outpatient medical claims in a city in the eastern part of China of all the enrollees under the city's public health insurance programs in 2011 and 2012. There were around 26 million residents (99% of urban unemployed residents and 100% of rural residents who are not employees) enrolled under the URRBMI and around 21 million (98.6% of urban employees) enrolled under the UEBMI. Our medical claim data contain approximately 19 million outpatient visits in 2011 and more than 21 million outpatient visits in 2012. For each visit, the data provide detailed information regarding expenditures on the drugs, diagnosis, and treatment, the type of insurance, the eligible expenditure, and patient ID. For each patient, we aggregate the visit-level medical expenditure data to the annual level to obtain annual eligible expenditures and the total number of visits in a year.

The cutoff of annual reimbursement (z^*) under the URRBMI was 600 RMB in 2011 and 800 RMB in 2012, respectively, and the reimbursement rate (δ) is 50% at the Tier 1 community health services institutions and 40% at the Tier 2 and 3 hospitals. By contrast, the upper bound of annual reimbursement (z^*) under the UEBMI is much higher: at 2500, 3000, 3500, 4500 RMB in 2011 (depending on whether the patient is on-the-job or retired and whether the disease is chronic or not) and at 3500, 4000, 4500, 5500 RMB in 2012. The reimbursement rate (δ) under the UEBMI is also higher: 70% for on-the-job workers and 85% for retired workers. Details of the medical insurance plan are shown in Table 1. Given the policy complexity in the UEBMI, we focus our empirical analysis on the sample of the URRBMI which contains one policy threshold each year, and use the sample of the UEBMI for placebo analyses to support our empirical identification.

[Insert Table 1 Here]

5.3 Bunching Evidence

To examine whether patients respond to the medical expenses deduction limits, we first plot the density distribution of annual eligible expenses (z) for patients under the URRBMI. Results are shown in **Figure 4a** for 2011 and **Figure 4b** for 2012, respectively. There is a clear bunching at the policy threshold for both figures; that is, a significant and sharp bunching mass at 600 for 2011 and 800 for 2012. These results suggest that consistent with our theoretical analysis in Section 2,

patients comply with the kinked policy by optimally choosing their medical consumption. Meanwhile, the fact that there is no excess bunching at the 2011 threshold of 600 in 2012 indicates that patients incur low adjustment costs when the threshold changes, relieving the concern of stayers.

To alleviate the concern that bunching at the policy thresholds in **Figures 4a-4b** may be spurious due to other confounding factors, we repeat the analysis for the sample of patients under the UEBMI in **Figure A1** in Appendix A. Given that the policy thresholds of the UEBMI were much higher than those of the URRBMI, we should not expect any bunching mass at the policy thresholds of the URRBMI. Indeed, we do not spot any bunching behavior at 600 in 2011 and at 800 in 2012. These findings lend support to our argument that patients indeed adjust their medical expenses in response to the kinked reimbursement policy.

Another common concern related to bunching analyses is whether the adjustment is real or just a relabelling behavior, which may generate estimation complexity and potential biases as illustrated in Section 4.4. For example, studies detect a certain share of bunching response due to relabelling in settings where agents self-report the values (e.g., Saez, 2010; Chen et al. 2021). However, in our setting, the eligible medical expenses are not self-reported. Instead, the numbers are aggregated from visit-level medical transactions, and hence, relabeling or misreporting is very unlikely in this setup.

[Insert Figure 4 Here]

5.4 Causal Impacts on Patient Behavior

We have shown that patients adjust their eligible expenses to take advantage of the kinked reimbursement scheme, resulting in excess bunching at the threshold of the reimbursement limit. We now explore the potential impacts of such adjustments on patients' outpatient behaviors.

5.4.1 Stylized Facts

Before a formal estimation of the causal impacts, we first plot the raw relation between the eligible annual medical expenses (z) and medical outcomes (y) to gain some direct evidence. We consider whether a change in co-payment rate could affect patients' choice of outpatient visits. **Figure 5** reports the total number of outpatients at each bin level of eligible annual expenses for outpatients under the URRBMI. Green triangles represent the distribution for 2011 and blue squares represent

the distribution for 2012, where sizes of triangles and squares are proportional to the sample size in each bin for the corresponding group.

Let us first study the distribution of 2012, which is represented by the blue rectangles. To the left of the policy threshold, there is a clear upward relation between the total number of outpatient visits and the eligible annual expenses. This trend carries on to the policy threshold and is then followed by a significant drop in the total number of visits which becomes much flat afterward. Meanwhile, overall the total number of visits to the left of the policy threshold is larger than those to the right of the policy threshold. These results provide direct visual support to our theoretical analysis in Section 2: compared to patients located to the left of the threshold (with a marginal copayment rate of τ), patients to the right of the threshold respond to the increase in copayment rate (of 100%) by paying fewer outpatient visits.

The distribution in 2011 shows a similar pattern as those in 2012, in which values to the left of the policy threshold are generally larger than those to the right of the policy threshold. These further lend support to our theoretical framework elaborated in Section 2 that changes in the co-payment rate significantly impacted patients' decisions for outpatient visits. Combining the distributions of 2011 and 2012, it is interesting to note that jumps only happen at the corresponding policy thresholds. Specifically, there is no clear jump at 600 in the 2012 distribution when the policy threshold was at 800; and vice, versa. These results are consistent with the bunching behavior in **Figures 4**, which further confirm that induced manipulation behavior and its impacts were caused by the kinked policy.

To further alleviate the concern of spurious responses due to other factors, we examine the distributions of 2011 and 2012 for patients under the UEBMI. As the policy thresholds under the UEBMI were much higher than those under the URRBMI in both 2011 and 2012, we should not expect any significant behavioral changes around 600 or 800. As shown in **Figure A2** in Appendix A, we find smooth relations between the total number of outpatient visits and eligible expenses throughout the whole region in both years, with no systematic changes below and at the placebo policy thresholds. These results lend further support to the argument that patients adjust their number of outpatient visits in response to the kinked medical insurance plan.

[Insert Figure 5 Here]

5.4.2 Counterfactual Density Distribution

A crucial element in formally estimating the magnitude of policy impact is the counterfactual density distribution; that is, the density distribution under the counterfactual linear scheme with the low co-payment rate. To this end, we estimate the counterfactual density distribution following our proposed method in Section 3.1 and compare it with the commonly used approach by Chetty et al. (2011).[35]

Figure 6a shows the observed density distribution $h()$ and our estimated counterfactual density distribution $h^{ct}()$ based on outpatients under the URRBMI in 2012[36]. Specifically, the solid green curve represents the observed density distribution, and the dashed red curves represent the estimated counterfactual density distribution. Meanwhile, the solid vertical line indicates the policy threshold (annual reimbursement limit z^*), the long vertical dashed line in the upper part of the density distributions shows the estimated marginal *buncher*'s response Δz^*, and two short dashed vertical lines around the threshold specify the diffuse range that is visually determined.

[Insert Figure 6 Here]

Three groups of patients under the kinked reimbursement policy are clearly shown in the figures: (a) *always-takers* with counterfactual expenses $z^{ct} \leq z^*$ remain unchanged with $z = z^{ct}$ and located to the left of the threshold; (b) *bunchers* with counterfactual expenses $z^{ct} \in (z^*, z^* + \Delta z^*]$ adjust their expenses downwards and bunch at the threshold, i.e., $z = z^*$; (a) *shifters* with counterfactual expenses $z^{ct} > z^* + \Delta z^*$ reduce their expenses to $z = z^{ct} \times \frac{z^*}{z^* + \Delta z^*}$, resulting in leftwards shifting in the counterfactual density distribution and located to the right of the threshold z^*.

In magnitude, the estimated marginal *buncher*'s response Δz^* is 260 RMB and significant at the 1% level. This number indicates that the counterfactual values of annual eligible expenses are around 1.3 times the observed values under the kinked policy for the marginal bunching agents and the shifting agents (i.e., $\frac{z^{ct}}{z} = \frac{z^* + \Delta z^*}{z^*} = 132.5\%$).

It is worth noting that the counterfactual density distribution to the left of the policy threshold is not an upward parallel shifting of the observed density distribution. This is because patients with different counterfactual expenses shift leftwards with different magnitudes in response to the kinked policy as elaborated in Section 2, resulting in the counterfactual and observed density

[35] We control for the reference points effect using the method in Section 4.1.

[36] The results remain similar if we use the 2011 sample, as shown in **Figures 4 & 5** earlier. For illustration purposes, we focus on the 2012 sample hereafter.

distribution having different shapes in the region to the left of the threshold. These are in contract with the assumption under the estimation framework by Chetty et al. (2013). Specifically, by assuming an upward parallel shift from the observed to the counterfactual density distributions, estimates following Chetty et al. (2011) end up overestimating the marginal buncher's response. As shown in **Figure 6b**, the estimated marginal *buncher*'s response Δz^* is 400 RMB, which is larger than our estimates of 260 RMB.

5.4.3 Estimation of Policy Impacts and Structural Parameters

We now use the causal estimation framework proposed in Section 3.2 to quantify the impact of the kinked medical reimbursement scheme on outpatient behaviors for shifting and bunching patients separately.

Figure 7 plots the empirical results using the 2012 outpatient data with the policy threshold at 800. Consistent with the outlay in **Figure 6**, the solid vertical line indicates the policy threshold, the long vertical dashed line in the upper part of the distribution shows the estimated marginal *buncher*'s response Δz^*, and two short dashed vertical lines around the threshold specify the diffuse range. Green dots present the observed distribution of the total number of outpatient visits annually (in logarithm) (y) against eligible annual expenses (z). Blue dots represent the auxiliary outcome distribution for shifting patients when we locate *shifters* back to their counterfactual location of z. Following Equation (21) in section 3.2, we obtain the counterfactual outcome distribution (represented by the red dashed curve) and calibrate the structural parameters (μ, λ) as shown in **columns 1 & 2 of Table 2**. It indicates that when eligible annual expenses z increase by 1%, the number of outpatient visits annually increase by 14.384, significant at 1% level, capturing the direct impact from changes in z; meanwhile, when the annual out-of-pocket increases by 100 RMB (as a result of changes in z), the number of outpatient visits annually increase by 1.9, significant at 1% level, capturing the indirect impact from changes in z. As discussed in section 2, the introduction of the kinked policy leads to a reduction in z for *shifting patients* and *bunching patients*, given that the estimated values $\hat{\mu} > 0, \hat{\lambda} > 0$, therefore, we would anticipate a negative effect on the number of outpatient visits annually for both *shifters* and *bunchers*.

To verify this, we can compare the observed outcome distribution (green dots) and the counterfactual outcome distribution (red dashed line) in **Figure 7.** There are significant decreases in the number of outpatient visits for *shifting agents* (those with $z^{ct} \in (1060, 1275)$ and $z \in (800, 980)$) and a substantial decrease for *bunching agents*(those with $z^{ct} \in (800, 1060]$ and $z = z^*$). In terms

of the economic magnitude, **column 3 of Table 2** shows that the policy effect for shifting agents is -2.110, significant at 1%, implying that the kinked medical insurance policy causes shifting agents to pay around two times fewer outpatient visits, compared to the counterfactual linear policy with a low co-payment rate. The estimation results for *bunchers* are shown in **column 4 of Table 2**. We find a negative average treatment effect on bunching patients as well, although the magnitude is smaller because *bunching patients* encounter a smaller reduction in z compared to *shifting patients*.

[Insert Table 2, Figure 7 Here]

5.4.4 Heterogeneous Impacts

Patients in different age groups may respond heterogeneously to the kinked reimbursement scheme. We next split the full sample into three subgroups based on patients' age at the time of treatment and explore the heterogeneous impacts.

Figure 8 compares the degree of bunching for the three subgroups: children (patients aged under age 15), middle-aged adults (patients aged between 16 and 54), and elders (patients aged above 55). We find excess bunching in all three subgroups, indicating that patients indeed adjust their eligible expenses as a response to the kinked medical insurance scheme. In addition, we find similar level of excess bunching for all age groups,with the marginal *buncher*'s response at 260 RMB.

Then, we move on to the policy impact on the number of outpatient visits annually. **Figure 9** shows the observed and counterfactual outcome distributions for each subgroup. **Table 3** shows the calibrated parameters and the estimated policy impact for each subgroup. *bunching patients* and *shifting patients* of all age groups have decreased their number of outpatient visits when the co-payment rate decreased due to the kinked policy. The impact is slightly larger on patients aged between 16 to 54, compared to other groups. The consistency in results indicates that financial incentives matter for patients' outpatient behaviors across all age groups.

One thing to note is that the estimated causal impact on the full sample is close to the weighted average of the causal impacts on these subgroups. This is consistent with our discussion in section 4.4 that when the heterogeneity in bunching response is relatively small, there is very limited bias when locating shifting agents back to their original location under the homogeneous parameter assumption. Therefore, the average bunching response, average calibrated structural

parameters, and the average causal effects under the homogeneous approach are a close approximation to the average estimates of each subgroup when heterogeneity is taken into consideration.

5.5 Alternative Policies: Changes in Thresholds or Co-payment Rates

Given the calibrated value of structural parameters (μ, λ, e) and our understanding of patients' behavior under kinked policies, we can study the impact of alternative policy designs, by varying the location of the kink and by changing the difference in co-payment rates below and above the threshold. These analyses could shed light on policy designs by exploring questions like these: fixing the overall cost of medical insurance, what kind of policy design generates the highest outcomes for the overall population? Further, who benefits from such a policy? Our approach allows us to conduct certain welfare analyses using a reduced-form approach, however, we do note that the analysis rules out potential changes in price due to general equilibrium effects (e.g., changes in patient behavior might affect the price of seeing the doctors).

As an illustration example, we analyze the impact of increasing the cutoff z^* from 600 RMB to 800 RMB on the medical insurance burden and the overall number of outpatient visits. When the cutoff increases, patients in the middle of the distribution of annual eligible expenses (z) would see a surge in outpatient visits due to the reduction in the co-payment rate, while other patients remain the same. This is consistent with our findings in **Panel A of Table 4**, where the overall number of outpatient visits and insurance burden increase as the cutoff moves rightwards.

Our current policy imposes a 100% co-payment rate once the expenses exceed the cutoff (i.e., $z > z^*$). If we are willing to reduce the co-payment rate for $z > z^*$, to maintain a constant insurance burden, we need to reduce the threshold by imposing more patients on the higher marginal copayment scheme. Which policy is better, a higher cutoff with a larger jump in copayment rate at the cutoff, or, a lower cutoff with a smaller jump in copayment rate at the cutoff? This question would be more of interest if there are relabelling or misreporting [37] In our setup, there is no misreporting. Here we study the distributional impact, measured by the dispersion of patient outcomes. If we are imposing a high cutoff, high jump of copayment rate across the cutoff (denoted as Policy I), then, a small group of patients with high medical demand would see a big reduction in outpatient visits. By contrast, if we are imposing a low cutoff, small jump of copayment rate across the cutoff

[37] Chen et al. (2021) pointed out that when there is misreporting, a larger threshold is better at stimulating R&D expenses using tax incentives, as the share of firms who misreporting is smaller.

(denoted as Policy II), then, a large group of patients with medium to high medical demand would see a mild reduction in outpatient visits. Both Policy I and Policy II affect outpatient behaviors, but Patient II imposes a more balanced impact than Policy I. This is shown in Panel B of Table 4, where Policy II has a higher dispersion of outpatient expenses and visits, holding the insurance burden (total reimbursement) constant. These distributional analyses could be of interest to policymakers in various setups where a kinked policy is relevant.

6 Conclusion

In this paper, we develop a reduced-form estimator for identifying treatment effects under kink settings when agents manipulate or adjust their values of the assignment variable in response to the non-linear policy. The method is model-free and makes use of agents' interior response behavior. Specifically, under kinked settings, agents to one side of the cutoff face a change in marginal incentives and adjust their assignment variable by a constant share. Such interior response allows us to recover the counterfactual density and outcome distribution, which facilitates the estimation of treatment effects on bunching agents and shifting agents. Extensions with diffuse bunching, rounding in assignment variable values, potential misreporting/relabelling, optimization frictions, and heterogeneity in structural parameters are also explored. We apply the proposed causal estimator to a medical insurance setting in China where patients are subject to a much higher co-insurance rate when their cumulative annual medical expenses cross a statutory threshold. Based on administrative visit-level outpatient data from a city in China, we show that patients adjust their outpatient behavior in response to the kinked policy, indicating a health and financial tradeoff by patients.

References

Michael Carlos Best and Henrik Jacobsen Kleven. Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK. *The Review of Economic Studies*, 85(1):157–193, 2018.

Michael Carlos Best, James S Cloyne, Ethan Ilzetzki, and Henrik J Kleven. Estimating the elasticity of intertemporal substitution using mortgage notches. *The Review of Economic Studies*, 87 (2):656–690, 2020.

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

Carolina Caetano. A test of exogeneity without instrumental variables in models with bunching. *Econometrica*, 83(4):1581–1600, 2015.

Carolina Caetano, Gregorio Caetano, and Eric Nielsen. Correcting for endogeneity in models with bunching. *Journal of Business & Economic Statistics*, pages 1–13, 2023.

Pedro Carneiro, Katrine V Løken, and Kjell G Salvanes. A flying start? maternity leave benefits and long-run outcomes of children. *Journal of Political Economy*, 123(2):365–412, 2015.

Zhao Chen, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu. Notching r&d investment with corporate income tax cuts in china. *American Economic Review*, 111(7):2065–2100, 2021.

Raj Chetty, John N Friedman, Tore Olsen, and Luigi Pistaferri. 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, 2011.

James Cloyne, Kilian Huber, Ethan Ilzetzki, and Henrik Kleven. The effect of house prices on household borrowing: a new approach. *American Economic Review*, 109(6):2104–36, 2019.

Natalie Cox, Ernest Liu, and Daniel Morrison. Market power in small business lending: A two-dimensional bunching approach. Technical report, 2021.

Rebecca Diamond and Petra Persson. The long-term consequences of teacher discretion in grading of high-stakes tests. Technical report, National Bureau of Economic Research, 2017.

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

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

Cristian Pop-Eleches and Miguel Urquiola. Going to a better school: Effects and behavioral responses. *American Economic Review*, 103(4):1289–1324, 2013.

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

Seth D Zimmerman. Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, 109(1):1–47, 2019.

Table and Figures

Table 1. Medical Insurance Plan

	2011		2012	
	z^*	δ	z^*	δ
Insurance Plans	(1)	(2)	(3)	(4)
$URRBMI^a$	**600**	50%	**800**	50%
$URRBMI^b$	**600**	40%	**800**	50%
$UEBMI^a$	3500	70%	4500	70%
$UEBMI^b$	2500	70%	3000	70%
$UEBMI^c$	4500	85%	5500	85%
$UEBMI^d$	3000	85%	4000	85%

Notes: z^* denotes the upper bound of annual reimbursement, and δ denotes the reimbursement rate in the city where we have access to all the medical data. $URRBMI_a$ and $URRBMI_b$ correspond to patients under URRBMI (including urban unemployed and rural) treated at Tier 1 community health institutions and Tier 2/3 hospitals, respectively. $UEBMI^a$ and $UEBMI^b$ correspond to employed workers under UEBMI for non-chronic and chronic diseases respectively. $UEBMI^c$ and $UEBMI^d$ correspond to retired workers under UEBMI for non-chronic and chronic diseases respectively.

Table 2. Estimates of Policy Effects on Patients' Behavior

	Number of Outpatient Visits (log)			
	Shifting Patients			*Bunching Patients*
	μ	λ	TE	TE
	(1)	(2)	(3)	(4)
Treatment Effect	16.503***	0.019***	-2.110***	-0.463***
	(1.139)	(0.002)	(0.122)	(0.037)
Density Order	4	4	4	4
Outcome Order	2	2	2	2
Observations	184,202	184,20	184,202	184,20

Significance: *.10; **.05; ***.01.

Notes: Column (1) reports changes in Δz^* for the marginal bunching patient, estimated from the density distribution. As for patient outcomes, we focus on the number of outpatient visits annually. Column (2) reports the calibrated value of structural parameter μ, reflecting the direct impact of percentage changes in z on outcome y. Column (3) reports the calibrated value of structural parameter λ, reflecting the impact of changes in T(due to the introduction of kinked policy and changes in z) on outcome y. Column(4) reports the impacts of introducing the kinked policy (compared to the linear low copayment counterfactual policy) on *shifting patients*. Column (5) reports the impacts on *bunching patients*. Details are shown in section 3. Standard errors are computed via bootstrap.

Table 3. Estimates of Policy Effects on Patients of Different Age Groups

	Number of Outpatient Visits (log)			
	Shifting Patients			*Bunching Patients*
	μ	λ	TE	TE
	(1)	(2)	(3)	(4)
(a) Aged$\leq$15				
Treatment Effect	12.359***	0.012***	-1.730***	-0.128**
	(1.970)	(0.003)	(0.182)	(0.058)
Density Order	4	4	4	4
Outcome Order	2	2	2	2
Observations	45,401	45,401	45,401	45,401
(b) Aged: 16$\sim$54				
Treatment Effect	14.863***	0.016***	-1.986***	-0.493***
	(2.478)	(0.004)	(0.258)	(0.078)
Density Order	4	4	4	4
Outcome Order	2	2	2	2
Observations	35,947	35,947	35,947	35,947
(C) Aged: $\geq$55				
Treatment Effect	15.050***	0.018***	-1.816***	-0.480***
	(1.671)	(0.002)	(0.153)	(0.047)
Density Order	4	4	4	4
Outcome Order	2	2	2	2
Observations	100,874	100,874	100,874	100,874

Significance: *.10; **.05; ***.01.

Notes: For patient outcomes, we focus on the number of outpatient visits annually. Column (1) reports the calibrated value of structural parameter μ, reflecting the direct impact of percentage changes in z on outcome y. Column (2) reports the calibrated value of structural parameter λ, reflecting the impact of changes in T(due to the introduction of kinked policy and changes in z) on outcome y. Column(3) reports the impacts of introducing the kinked policy (compared to the linear low co-payment counterfactual policy) on *shifting patients*. Column (4) reports the impacts on *bunching patients*. Panels (a), (b), and (c) present the analysis for patients aged below 15, aged between 16 to 54, and patients aged above 55. Details are shown in section 3. Standard errors are computed via bootstrap.

Table 4. Cost and Benefits Analyses

Cutoff z^*	Number of Outpatient Visits (log) 600	650	700	800
	(1)	(2)	(3)	(4)
A: holding $\Delta t = 1-t$ for $z > z^*$				
Total Outpatient Visits	521.18	668.62	832.53	1184.304
Total Eligible Expenses	110810.21	113710.39	116755.20	122585.19
Total Reimbursement	49634.53	52391.85	54790.30	58509.31
Observations	184,202	184,202	184,202	184,202

(B) Varying z^* and Δt

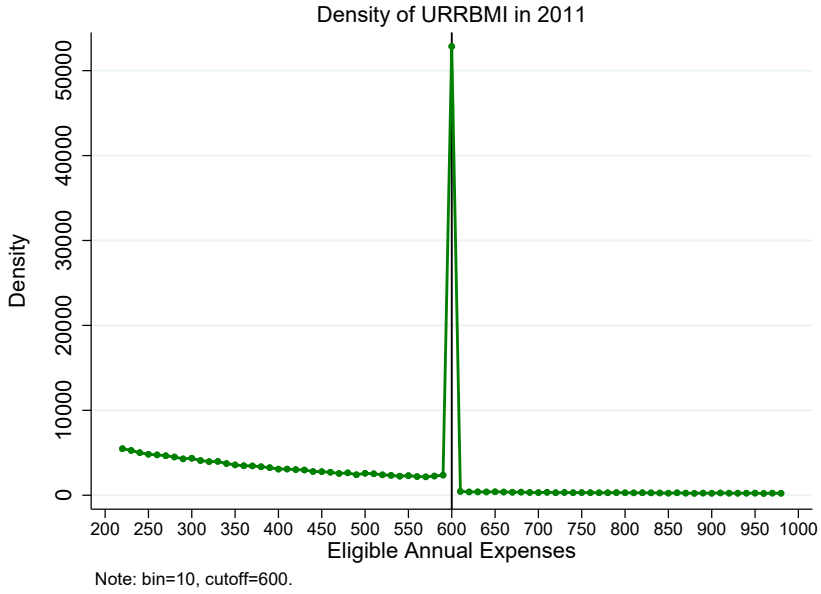
	Policy I		*Policy II*	
	z^*	Δt	z^*	Δt
	800	$1-t$	650	$0.253 * (1-t)$
	(1)	(2)	(3)	(4)
Dispersion of "Outpatient Visits (log)"	0.096		.102	
Mean Value of "Outpatient Visits (log)"	1.857		1.744	
Dispersion of "Eligible Expenses"	0.203		0.255	
Mean Value of "Eligible Expenses"	673.741		693.668	
Total Reimbursement	58509.312		58509.299	
Observations	184,202		184,202	

Significance: *.10; **.05; ***.01.

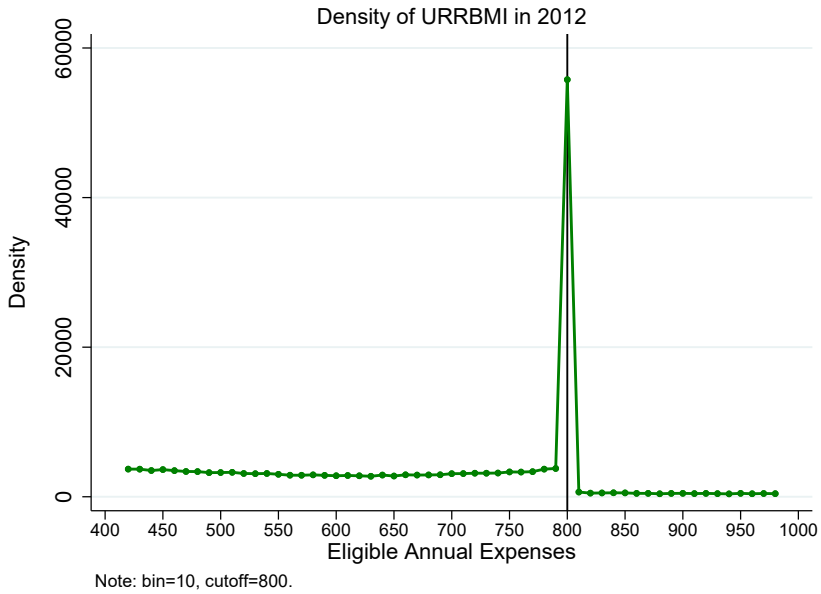
Notes:

Figure 4 Density Distribution of Eligible Annual Medical Expenses for Patients under the URRBMI Plan

(a). URRBMI in 2011

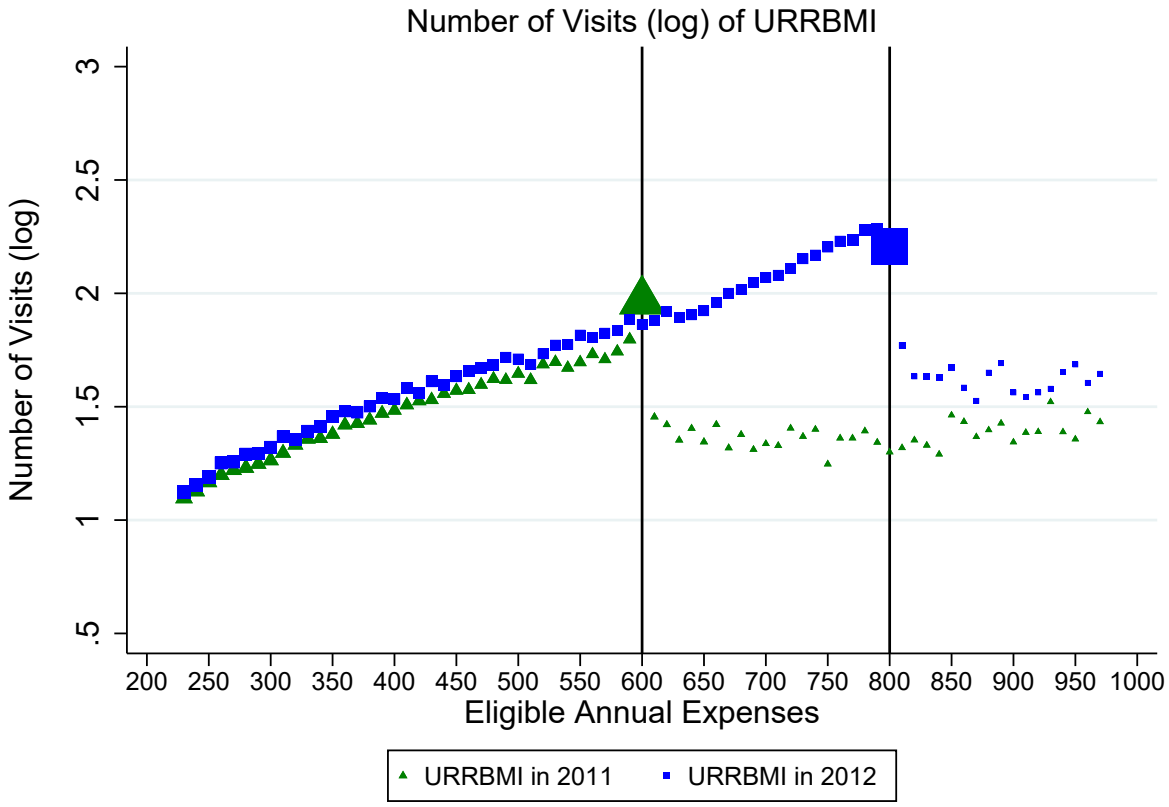


(b). URRBMI in 2012



Note: The cutoff of eligible annual medical expenses for outpatients under the URRBMI plan was 600 in 2011 and increased to 800 in 2012.

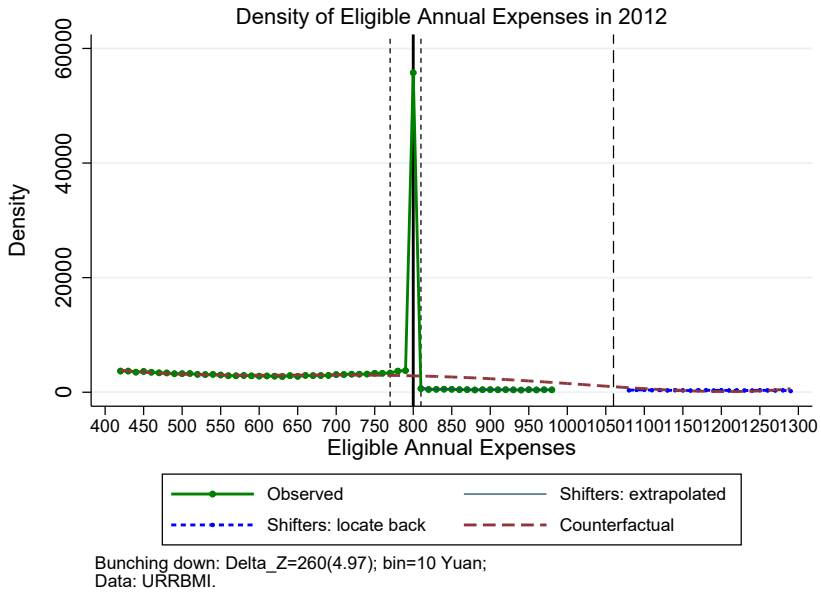
Figure 5 Conditional Distribution of Patient Outcomes – Number of Annual Outpatient Visits (log) under the URRBMI Plan



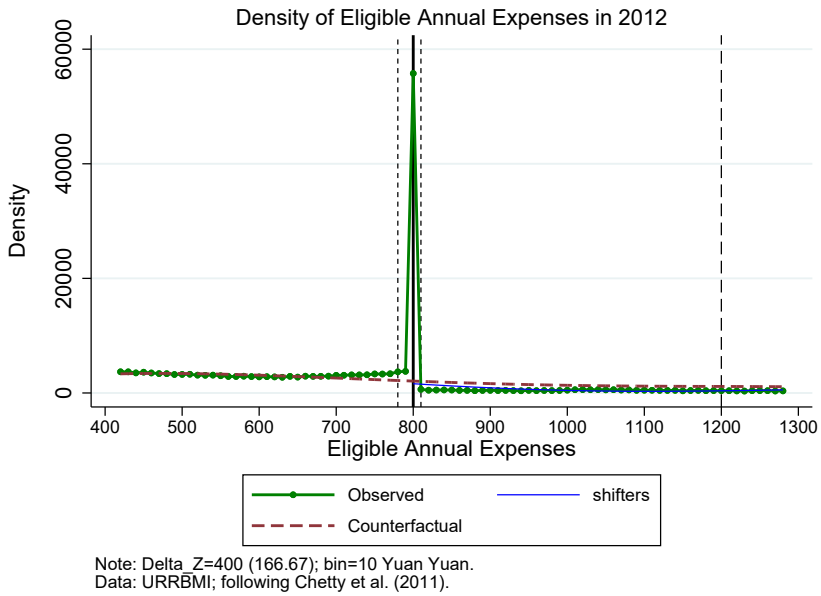
Note: The cutoff of the eligible annual medical expenses for outpatients under the URRBMI plan was 600 in 2011 and increased to 800 in 2012.

Figure 6 Density Distribution of Eligible Annual Medical Expenses for Patients under the UR-RBMI Plan in 2012: Observed and Counterfactual (using the proposed method)

(a). Our Method

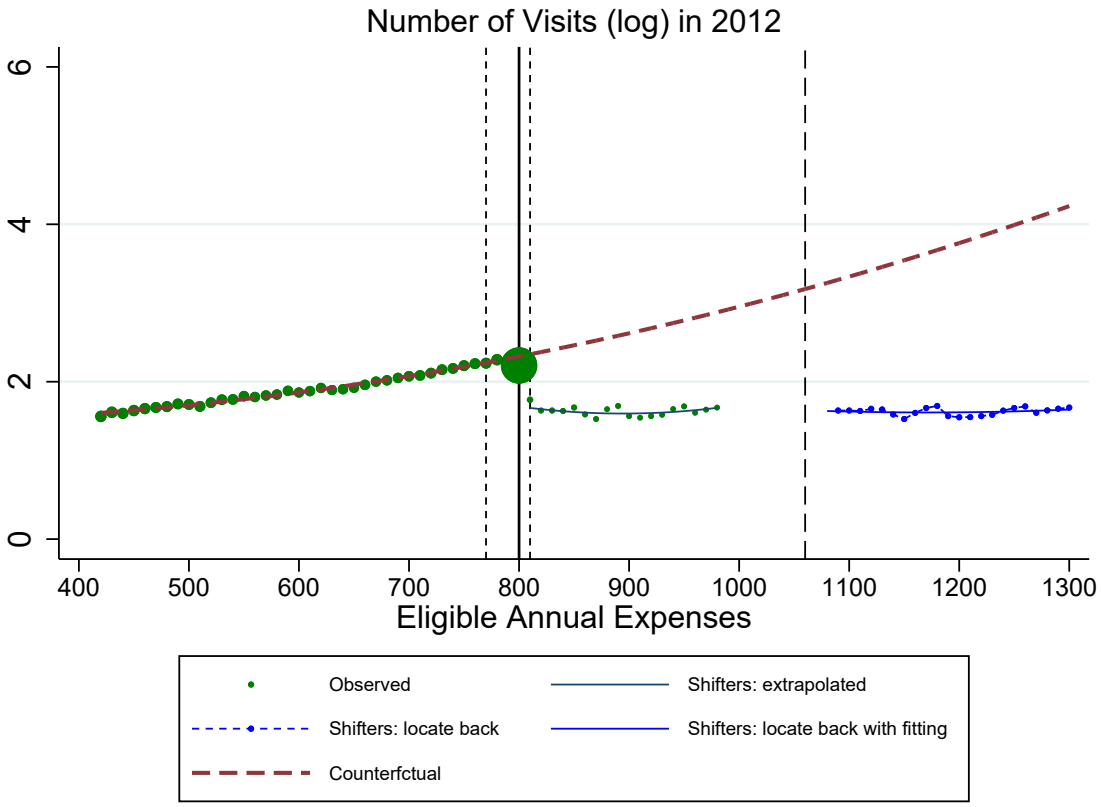


(b). Following Chetty et al. (2011)



Note: The cutoff of eligible annual medical expenses for outpatients under the URRBMI plan was 800 in 2012. The counterfactual distribution (when patients are subject to a linear low co-payment rate, marked by the dashed line) in panel (a) is estimated following the method proposed in section 3.1, while that in panel (b) is estimated following Chetty et al. (2011).

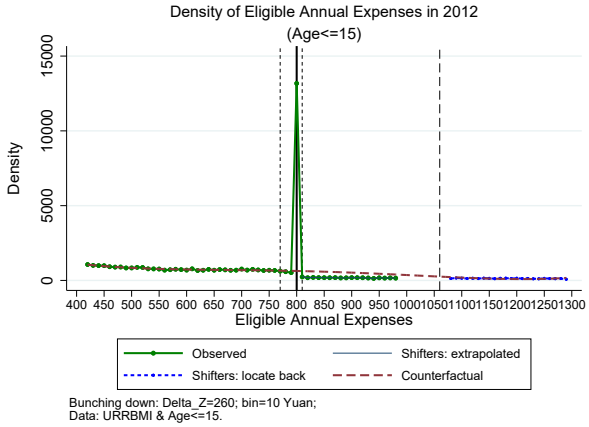
Figure 7 Conditional Distribution of Patient Outcomes – Number of Outpatient Visits Annually under the URRBMI Plan: Observed and Counterfactual



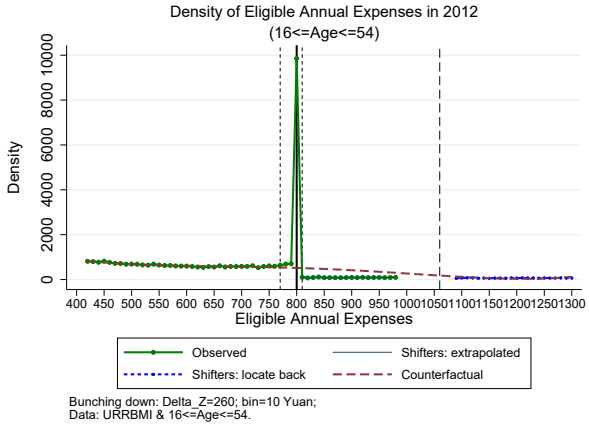
Note: The cutoff of the eligible annual medical expenses for outpatients under the URRBMI plan was 800 in 2012. The counterfactual distribution (when patients are subject to a linear low co-payment rate, marked by the dashed line) is estimated following the method proposed in section 3.2.

Figure 8 Density Distribution of Eligible Annual Medical Expenses for Patients at Different Age Groups under the URRBMI Plan

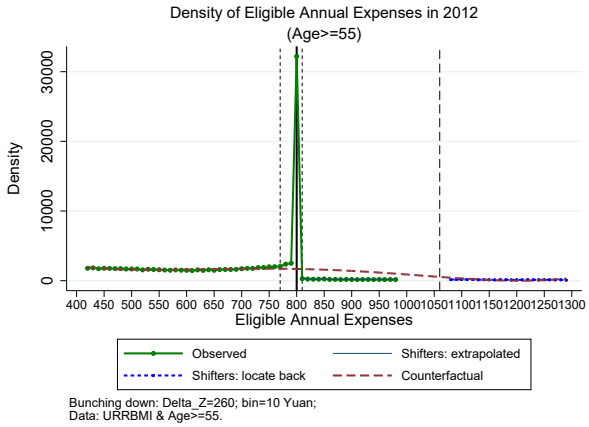
(a). Aged below 15



(b). Aged between 16 to 54



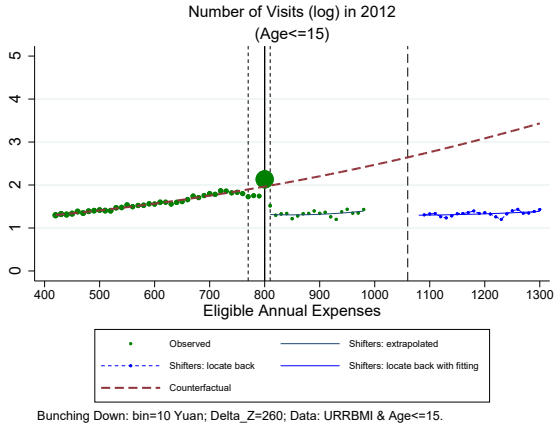
(c). Aged above 55



Note: The cutoff of eligible annual medical expenses for outpatients under the URRBMI plan was 800 in 2012. The counterfactual distribution (i.e., when medical expenses are not deductible, marked by the dashed line) is estimated following the method proposed in section 3.1.

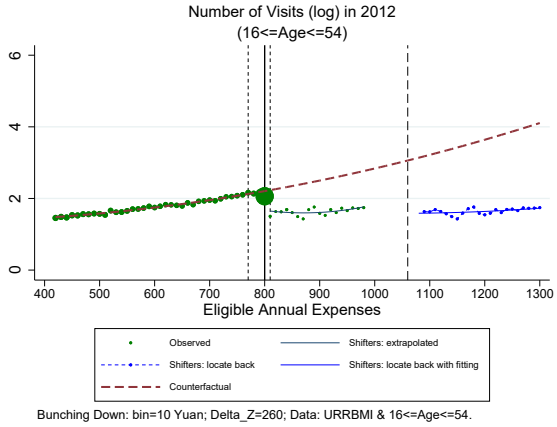
Figure 9 Conditional Distribution of Patient Outcomes – Number of Outpatient Visits Annually for Patients at Different Age Groups under the URRBMI Plan

(a). Aged below 15



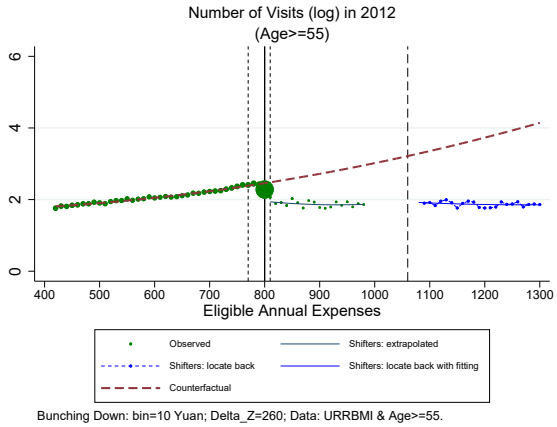
Bunching Down: bin=10 Yuan; Delta_Z=260; Data: URRBMI & Age<=15.

(b). Aged between 16 to 54



Bunching Down: bin=10 Yuan; Delta_Z=260; Data: URRBMI & 16<=Age<=54.

(c). Aged above 55



Bunching Down: bin=10 Yuan; Delta_Z=260; Data: URRBMI & Age>=55.

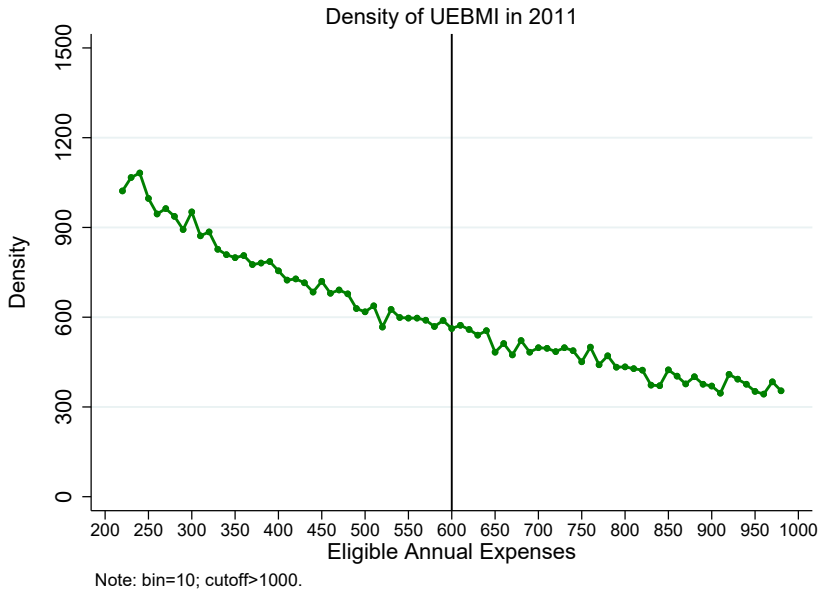
Note: The cutoff of the eligible annual medical expenses for outpatients under the URRBMI plan was 800 in 2012. The counterfactual distribution (when patients are subject to a linear low co-payment rate, marked by the dashed line) is estimated following the method proposed in section 3.2.

Appendices

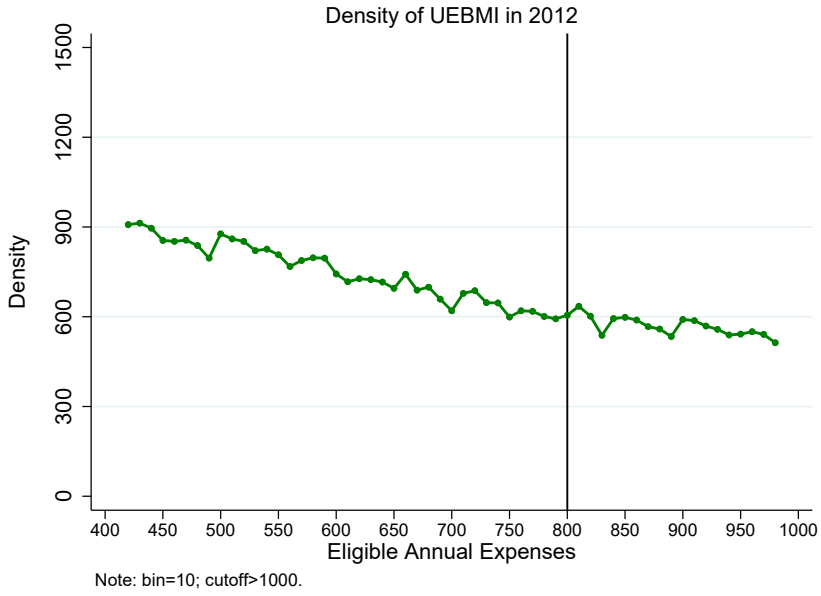
Appendix A: Additional Tables and Figures

Figure A1 Density Distribution of Eligible Annual Medical Expenses for Patients under the UEBMI Plan

(a). Placebo: UEBMI in 2011

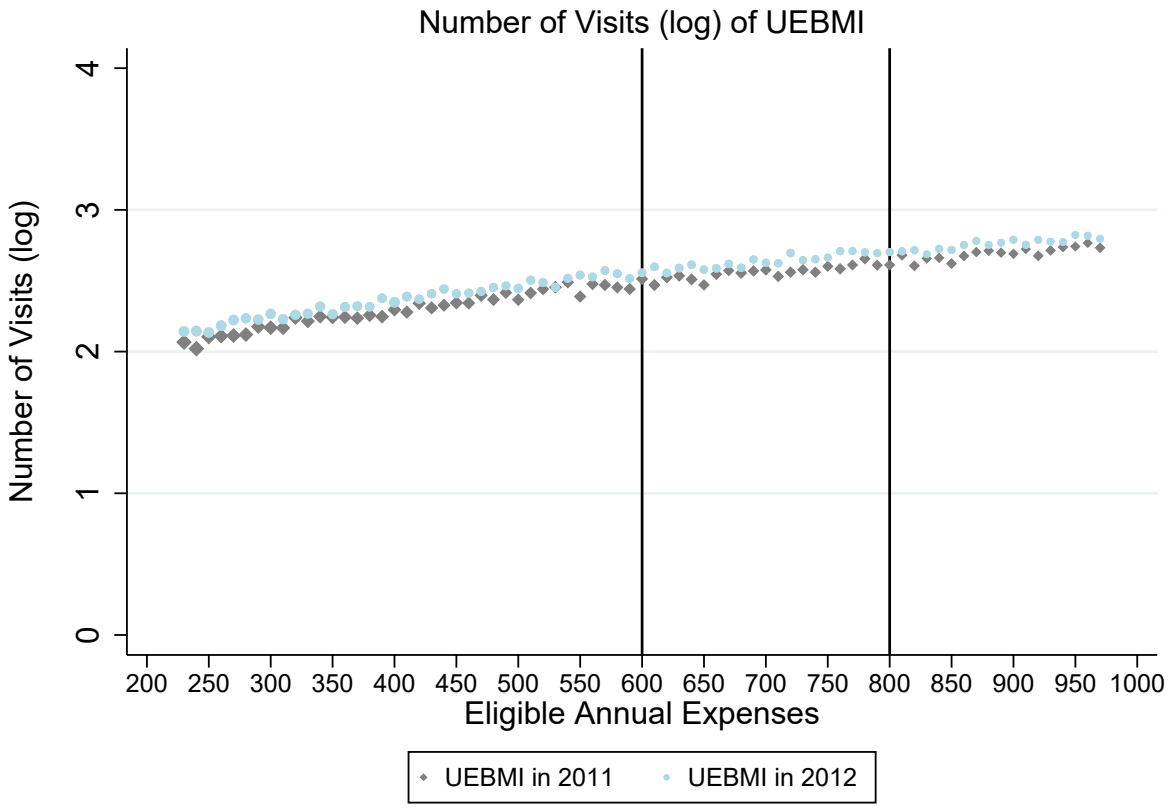


(b). Placebo: UEBMI in 2012



Note: The cutoff of eligible annual medical expenses for outpatients under the UEBMI plan was above 1000 in both 2011 and 2012. This serves as a robustness check.

Figure A2 Conditional Distribution of Patient Outcomes under the UEBMI Plan



Note: The cutoff of eligible annual medical expenses for outpatients under the UEBMI plan was above 1000 in 2011 and 2012. This serves as a robustness check.

Appendix B: Alternative Counterfactual Policy with a high linear rate – bunch up

B.1 A Generalized Framework Under the Kink Setting

B.1.1 Setup

Recall the focal kinked policy that states agents face a tax rate (or co-payment rate) of t if their value of z is below a statutory cutoff z^* but face a higher marginal tax rate (or co-payment rate) of $t+\Delta t$ if their $z > z^*$. Agents' optimal choice z under the kinked policy is given as:

$$z = \begin{cases} z(1,n) & \text{if } n \leq n_L \\ z^* & \text{if } n \in (n_L, n_H]. \\ z(0,n) & \text{if } n > n_H \end{cases}$$

where $D = 1$ indicates that agents face the lower marginal tax/co-payment rate of t and $D = 0$ indicates agents face the high marginal tax/co-payment rate of $t+\Delta t$; and n is an unobserved agent heterogeneity, with $z(D,n)$ increasing in n.

Consider an alternative counterfactual policy where agents always face the high tax/co-payment rate of $t+\Delta t$. That is, $T^{act}(z) = (t+\Delta t) \times z^{act}$. Consequently, agents' optimal choices are $z^{act} = z(0,n)$.

For agents with $n < n_L$, they increase their z in response to the lower marginal tax rate (co-payment rate) of t under the kinked policy, ($z = z(1,n) > z(0,n) = z^{act}$), but stay in the interior of the lower bracket, compared to the alternative counterfactual policy. We denote them as "*shifters*", or, agents with "interior response". Next, for agents with $n \geq n_H$, the optimal z under the kinked policy remains the same as z^{act} in the alternative counterfactual policy as they face the same marginal tax (co-payment) rate of $t+\Delta t$. We denote these agents as "*never-takers*". Finally, for agents with $n \in [n_L, n_H)$, their optimal choice under the kinked policy is to increase their z and bunch at the threshold z^*. We denote them as "*bunchers*", as their behavior produces excess

bunching in the density distribution at the kink point z^* when the kinked policy is introduced.

Following the Separability assumption that $z(D,n) = f(D;e)g(n;e)$, where e is a structural parameter, we would have

$$z = \begin{cases} z(0,n)\frac{f(1;e)}{f(0;e)} & \text{if } n < n_L \\ z^* & \text{if } n \in [n_L, n_H) \\ z(0,n) & \text{if } n \geq n_H \end{cases} \tag{30}$$

That is, agents with $n < n_L$ who originally choose $z(0,n)$ under the alternative counterfactual linear policy respond to the kinked policy by setting $z = z(1,\phi) = z(0,n)\frac{f(1;e)}{f(0;e)} < z^*$.

Note that for marginal *bunchers* with n_L, the optimal choice under the kinked policy is given by $z = z(1,n_L) = z^*$, and their location under the alternative counterfactual linear policy is $z(0,n_L) = z^* - \Delta z^{*,act}$, where $\Delta z^{*,act}$ is the change in z by the marginal bunching agent with n_L due to the introduction of the kinked policy. Hence, the excess bunching at the kink point is the cumulative density of *bunchers* (i.e., agents with $n \in [n_L, n_H)$, and can be derived as:

$$B = \int_{z^* - \Delta z^{*,act}}^{z^*} h^{act}(z)dz,$$

where $h^{act}(z)$ denotes the counterfactual density distribution of z (i.e., the one under the alternative linear high tax/co-payment rate plan).[38]

Therefore, for all agents with $n < n_L$, we have,

$$\frac{z}{z^{act}} = \frac{z(1,\phi)}{z(0,\phi)} = \frac{f(1;e)}{f(0;e)} = \frac{z^*}{z^* - \Delta z^{*,act}}. \tag{31}$$

Equation (31) characterizes the relationship between the original location (under the alternative counterfactual linear policy) and the new location (under the kinked policy) for each shifting agent.

Remark 3. Studies in the bunching literature largely use Equation (31) to back out the structural parameter from the estimated value of $\Delta z^{*,act}$. For example, in Saez (2010), the equivalent of Equation (31) is $(\frac{1-t}{1-t-\Delta t})^e = \frac{z^*}{z^* - \Delta z^{*,act}}$.

Combining Equations (30) and (31), we can summarize the change in z as

[38] The observed density distribution of z is denoted by $h(z)$.

$$\frac{z}{z^{act}} = \begin{cases} \frac{z(1,\phi)}{z(0,\phi)} = \frac{f(1;e)}{f(0;e)} = \frac{z^*}{z^* - \Delta z^{*,act}} & \text{if } n < n_L \\ \frac{z^*}{z(0,\phi)} & \text{if } n \in [n_L, n_H) \\ \frac{z(0,\phi)}{z(0,\phi)} = 1 & \text{if } n \geq n_H \end{cases} . \tag{32}$$

Hence, moving from the alternative counterfactual linear scenario to the state with the kinked policy, all agents with $n < n_L$ (*shifters*) increase their z by a constant share $\frac{z^*}{z^{*,act} - \Delta z^*} - 1$, but do not bunch at the cutoff z^*.[39] Meanwhile, agents with $n \in [n_L, n_H)$ (*bunchers*) increase their z to bunch at the cutoff z^*. By contrast, agents with $n \geq n_H$ (the *never-takers*) remain unchanged.

Equation (32) enables us to estimate the counterfactual density distribution $h^{act}()$, the excess bunching mass B at the kink point, and the marginal bunchers' response $\Delta z^{*,act}$ nonparametrically. Details will be discussed later in subsubsection B.2.1.

B.1.2 Causal Inference under Kinked Bunching

Denote $y()$ as the observed outcome distribution under the kinked policy and $h()$ the observed density distribution. Meanwhile, denote $y^{act}()$ as the counterfactual outcome distribution under the alternative high tax/co-payment rate, the estimation of which will be discussed in subsubsection B.2.2. Meanwhile, denote $h^{act}()$ as the corresponding counterfactual density distribution, the estimation of which will be discussed in subsection B.2.1.

Change in Outcome Distribution of Shifters

As discussed in the previous subsection 2.1, agents with $z^{act} < z^* - \Delta z^{*,act}$ (i.e., $n < n_L$) would increase their values of z when faced with a lower tax/co-payment rate under the kinked policy. That is, under the kinked policy, shifters would set $z = z^{act} \frac{z^*}{z^* - \Delta z^{*,act}}$. The change in z would generate three changes to the outcome distribution:

First, the relocation effect. Even if the change in z has no impact on y, such a "relocation" behavior (from z^{act} to z) would change the outcome distribution. Therefore, if we directly compare y^{act} with y along the y-axis, we are not comparing the same agent. However, we do know where each agent has moved to. Therefore, if we relocate each agent under the kinked policy back to his/her counterfactual location, then, comparing the values along the y-axis would give us the

[39] Note that each shifter's adjustment $(z - z^{act})$ is not a constant, it depends on the initial location (z^{act}). Alternatively, we can take the logarithm of z so that each shifter's adjustment will be a constant, i.e., $\ln z - \ln z^{act} = \ln \frac{z^*}{z^* - \Delta z^{*,act}}$.

treatment effect on *"shifters"*.

Treatment Effect on "Shifters"

$$\begin{aligned}\tau_y^{TE,shifter} &= E[y_n - y_n^{act} | n \in shifters] \\ &= \int_{z^{min}}^{z^*-\Delta z^{*,act}} \Big(y^r(z^{act}) - y^{act}(z^{act}) \Big) \frac{h^{act}(z^{act})}{\int_{z^{min}}^{z^*-\Delta z^{*,act}} h^{act}(z^{act}) dz^{act}} dz^{act} \end{aligned} \tag{33}$$

where $y^r(z^{act}) \equiv y(z^{act} \frac{z^*}{z^*-\Delta z^{*,act}})$ denotes the resulting auxiliary outcome distribution when we locate shifters at z back to their alternative counterfactual locations z^{act} using the relation that $z = z^{act} \frac{z^*}{z^*-\Delta z^{*,act}}$. That is, when we reshape the observed outcome distribution based on the changes in the agents' location of z, the outcome distribution changes from $y(z)$ to $y(z^{act} \frac{z^*}{z^*-\Delta z^{*,act}}) \equiv y^r(z^{act})$.

Second, an increase in the value from z^{act} to z could directly affect outcome y. For example, an increase in taxable income could affect consumption, or an increase in medical expenses could affect health. Define semi-elasticity $\mu_n \equiv \frac{\Delta y_n}{\Delta z_n / z_n}$, where n denotes agent heterogeneity as defined previously in subsection 2.1. Recall for *shifters*, we have $\frac{z}{z^{act}} = \frac{z^*}{z^*-\Delta z^{*,act}}$. Therefore,

$$(y_n - y_n^{act})|_{\text{due to direct change in } z} = \mu_n \big(\frac{z^*}{z^* - \Delta z^{*,act}} - 1 \big)$$

That is, the change in the value of z would directly lead to a level change in y.

Third, changes in taxes or fees (T) that *shifters* pay could also affect outcome y. Recall that under the alternative counterfactual policy, we have $T^{act}(z^{act}) = (t + \Delta t) \times z_n^{act}$, and under the kinked policy, we have $T(z) = t \times z = t \times \frac{z^*}{z^*-\Delta z^{*,act}} \times z_n^{act}$. Define $-\lambda_n \equiv \frac{\Delta y_n}{\Delta T_n}$. Hence, we have

$$\begin{aligned} y_n - y_n^{act}|_{\text{due to change in } T} &= -\lambda_n \big(t \times \frac{z^*}{z^* - \Delta z^{*,act}} \times z_n^{act} - (t + \Delta t) \times z_n^{act} \big) \\ &= -\lambda_n z_n^{act} \big(t \times \frac{z^*}{z^* - \Delta z^{*,act}} - (t + \Delta t) \big) \end{aligned}$$

That is, the change in tax or fees (T) would lead to both a level change and slope change in y.

Following Assumption 2 that the impacts from z and T on outcome y are additive and com-

bining the second and third points, we have

$$\begin{aligned}\tau_y^{TE,shifter} &= E[y_n - y_n^{act} | n \in shifters] \\ &= E[\mu_n]\big(\frac{z^*}{z^* - \Delta z^{*,act}} - 1\big) - E[\lambda_n z_n^{act}]\big(t \times \frac{z^*}{z^* - \Delta z^{*,act}} - (t + \Delta t)\big)\end{aligned}$$

Assume homogeneous preference and thus single response elasticities across agents (i.e., $\mu_n = \mu, \lambda_n = \lambda$), a condition commonly made in the bunching literature (see, e.g., Saez 2010; Chetty et al. 2011, Kleven 2016;). The above equation can be simplified as

$$\tau_y^{TE,shifter} = \mu\big(\frac{z^*}{z^* - \Delta z^{*,act}} - 1\big) - \lambda E[z_n^{act}]\big(t \times \frac{z^*}{z^* - \Delta z^{*,act}} - (t + \Delta t)\big) \tag{34}$$

Identifying Sufficient Statistics. We claim μ and λ are sufficient statistics for estimating treatment effects under policy simulations because changes in policy cutoffs or tax/co-payment rates would result in changes in z and hence changes in outcome variables. We propose estimating these parameters by exploiting the level and slope change at z^* when comparing the distributions $y^{act}(z^{act})$ and the extrapolated auxiliary outcome distribution $y^r(z^{act})$. Specifically, we have

$$\text{Level Change at } z^* = \mu\big(\frac{z^*}{z^* - \Delta z^{*,act}} - 1\big) - \lambda z^*\big(t \times \frac{z^*}{z^* - \Delta z^{*,act}} - (t + \Delta t)\big) \tag{35}$$

$$\text{Slope Change at } z^* = -\lambda\big(t \times \frac{z^*}{z^* - \Delta z^{*,act}} - (t + \Delta t)\big) \tag{36}$$

where estimation of the level change and slope change at z^* is explained in the next subsubsection B.2.2.

Change in Outcome Distribution of Bunchers

Agents with $z^{act} \in [z^* - \Delta z^{*,act}, z^*)$, i.e., $n \in [n_L, n_H)$, would increase their value of z and bunch at the cutoff ($z = z^*$) under the kinked policy. The changes in z would also generate changes in the outcome distribution.

Under the sharp bunching scenario, agents with $z^{act} \in [z^* - \Delta z^{*,act}, z^*)$ relocate to $z = z^*$. As it is impossible to find a one-to-one mapping for each bunching agent, we take all the bunching agents as an entity and identify the average treatment effect on "*bunchers*" by comparing changes in the average outcome value.

Treatment Effect on "bunchers" under Sharp bunching

$$\begin{aligned}\tau_y^{TE,buncher} &= E\left[y_n^{ct} - y_n | n \in buncher\right] \\ &= \overline{Y^{buncher}} - \overline{Y^{buncher,act}} \\ &= y^{buncher}(z^*) - \int_{z^*-\Delta z^{*,act}}^{z^*} y^{act}(z^{act}) \frac{h^{act}(z^{act})}{\int_{z^*-\Delta z^{*,act}}^{z^*} h^{act}(z^{act}) dz^{act}} dz^{act} \end{aligned} \tag{37}$$

where $y^{buncher}(z^*)$ denotes the average outcome of *bunchers* under the kinked policy, estimation of which are shown below.

Specifically, under the kinked policy, observations at the threshold z^* contain two groups of agents: (1) bunching agents with $z^{act} \in [z^* - \Delta z^{*,act}, z^*)$ who increase their value to the threshold $z = z^*$ in response to the kinked policy; (2) *never-takers* with $z^{ct} = z^*$ who remain at the threshold $z = z^*$. By contrast, under the counterfactual linear policy, there is only *never-takers* at the threshold z^*. Therefore, the density of bunchers under the kinked policy is given as $h^{bunch}(z^*) = h(z^*) - h^{act}(z^*)$. Further, the observed average outcome $y(z^*)$ is the weighted average of *bunchers* and *never-takers*, i.e., $y(z^*) = \left(y^{buncher}(z^*)h^{buncher}(z^*) + y^{act}(z^*)h^{act}(z^*)\right)\frac{1}{h(z^*)}$. Therefore, we obtain the average outcome of *bunchers* under the kinked policy $y^{buncher}(z^*)$. That is,

$$y^{buncher}(z^*) = \frac{y(z^*)h(z^*) - y^{act}(z^*)h^{act}(z^*)}{h(z^*) - h^{act}(z^*)}. \tag{38}$$

Treatment Effect on "bunchers" under Diffuse bunching is omitted here. One can follow the same logic as the baseline while adjusting for the alternative counterfactual policy.

Change in Outcome Distribution of Never-Takers

Agents with $z^{act} \geq z^*$ (i.e., $n \geq n_H$) would not adjust their z under the kinked policy, because the marginal tax/co-payment rate remains the same, with $z = z^{act}$. However, compared to the alternative counterfactual policy, never-takers pay less money under the kinked policy, i.e., $T = (t + \Delta t)z - \Delta t \times z^* = T^{act} - \Delta t \times z^*$. It acts like a lump-sum transfer.

We analyze the potential changes in outcome distribution for *never-takers* under the kinked policy. First, there is no relocation effect as $z = z^{act}$. Second, there is no impact on outcome y from direct changes in z, as there is no change in z. Third, changes in taxes or fees (T) that *never-takers*

pay could affect outcome y. Recall $-\lambda_n \equiv \frac{\Delta y_n}{\Delta T_n}$. Hence, we have

$$y_n - y_n^{act}|_{\text{due to change in } T} \quad = \quad \lambda_n \Delta t \times z^*$$

That is, the change in tax or fees (T) would lead to a level change in y.

Combined, we have

$$\begin{aligned} \tau_y^{TE,never-takers} &= E[y_n - y_n^{act} | n \in shifters] \\ &= E\left[\lambda_n \Delta t \times z^*\right] \\ &= E(\lambda_n)\Delta t \times z^* \\ &= \lambda \Delta t \times z^* \end{aligned} \tag{39}$$

where the last equality is based on the assumption of homogeneous preference and thus single response elasticity across agents (i.e., $\mu_n = \mu, \lambda_n = \lambda$), a condition commonly made in the bunching literature (see, e.g., Saez 2010; Chetty et al. 2011, Kleven 2016;). Estimation of the level change for *never-takers* is explained in the next subsubsection B.2.2.

B.2 Empirical Estimation

Our aforementioned estimation framework for the causal inference under the kinked bunching relies on the estimation of the counterfactual density $h^{act}()$ and outcome $y^{act}()$ distributions under the alternative linear policy. In this section, we elaborate on the empirical details to estimate these counterfactuals.

B.2.1 Estimating Counterfactual Density Distribution

We start with the strategy to recover the alternative counterfactual density distribution $h^{act}(z)$, which can be applied to any kinked settings. As shown in Equation (32), agents' responses to the kinked policy can be summarized as follows: (i) *shifters* with $z^{ct} < z^* - \Delta z^{*,act}$ increase their value but do not bunch at the threshold, i.e., $z = z^{ct} \times \frac{z^*}{z^* - \Delta z^{*,act}} < z^*$. (ii) *bunchers* with $z^{ct} \in [z^* - \Delta z^{*,act}, z^*)$ bunch at the threshold, i.e., $z = z^* > z^{ct}$; (iii) *never-takers* with $z^{ct} \geq z^*$ remain unchanged, i.e., $z = z^{ct} \geq z^*$;.

To recover the alternative counterfactual density distribution $h^{act}()$ from the observed density

distribution $h()$, we design a two-step estimation framework. First, we move *shifters* back to their counterfactual locations, which leads to the estimation of $h^{act}(z)$ within the region $(z_{min}, z^* - \Delta z^{*,act})$ for *shifters*. Then, we extrapolate $h^{act}()$ for bunching agents using the information of $h^{act}()$ for *shifters* and *never-takers*, (as $h^{act} = h()$ in the region $[z^*, z_{max}]$ for *always-takers*). Specifically, it is implemented by the following algorithm.

First, given the observed location z and an initial guess $\widehat{\Delta z^{*,act}}^{initial}$ for *shifting agents*, we infer the counterfactual choice $z^{act,initial}$ based on the following relation derived from equation (32):

$$z^{act,initial} = \begin{cases} z\frac{z^* - \widehat{\Delta z^{*,act}}^{initial}}{z^*} & \text{if } z < z^* - u_1 \\ \\ z & \text{if } z > z^* + u_2 \end{cases} \tag{40}$$

where $[z^* - u_1, z^* + u_2]$ is the bunching region with diffusion, in which $u_1 = u_2 = 0$ under sharping bunching. The inferred $z^{act,initial}$ for *shifters* forms the alternative counterfactual density distribution $h^{act,initial}(z), \forall z < (z^* - u_1)\frac{z^* - \widehat{\Delta z^{*,act}}^{initial}}{z^*}$,[40] whereas the observed density distribution for *always-takers* is the same as the counterfactual density distribution, i.e., $h^{act,initial}(z) = h(z), \forall z > z^* + u_2$.

Next, we obtain the counterfactual density for bunching agents based on the assumption that the alternative counterfactual density distribution is smooth. Specifically, we use the standard approach in the bunching literature to fit a flexible polynomial to the counterfactual distribution for the *never-takers* and *shifters* outside the region $[(z^* - u_1)\frac{z^* - \widehat{\Delta z^{*,act}}^{initial}}{z^*}, (z^* + u_2)]$, and extrapolate the fitted distribution inside the region. Empirically, we group agents into z bins indexed by j, and estimate the following regression:

$$\begin{aligned} h_j^{act,initial} &= \sum_{k=0}^{p} \beta_k (z_j^{act,initial})^k + \varepsilon_j \\ \text{if } z_j^{act,initial} < (z^* - u_1) & \quad \frac{z^* - \widehat{\Delta z^{*,act}}^{initial}}{z^*} \text{ or } z_j^{act,initial} > (z^* + u_2), \end{aligned} \tag{41}$$

[40] When we relocate *shifters* back to their original location, we reshape observed density distribution $h(z), \forall z \in (z_{min}, z^* - u_1)$ into $h(z^{ct}\frac{z^*}{z^* - \widehat{\Delta z^{*,act}}^{initial}}) \equiv h^{ct,initial}(z^{ct,initial}), \forall z^{act} \in \left(z_{min}, (z^* - u_1)\frac{z^* - \widehat{\Delta z^{*,act}}^{initial}}{z^*}\right)$.

where $h_j^{act,initial}$ is the number of agents in bin j; $z_j^{act,initial}$ is the inferred z level in bin j based on the initial guess $\widehat{\Delta z^{*,act}}^{initial}$; and p is the polynomial order. The counterfactual bin counts in the region $\left[(z^*-u_1)\frac{z^*-\widehat{\Delta z^{*,act}}^{initial}}{z^*},(z^*+u_2)\right]$ are obtained as the predicted values from Equation (41).

After recovering the $h^{act,initial}(z)$ for the full range of z, excess bunching (with diffusion) at the threshold can then be computed as[41]

$$\widehat{B}^{initial}=\int_{z^*-u_1}^{z^*-1}\left(h(z)-h^{shift}(z)\right)dz+\int_{z^*}^{z^*+u_2}\left(h(z)-h^{act,initial}(z)\right)dz, \tag{42}$$

where $h^{shift}(z)$ denotes the density of *shifters* under the kinked policy. Note that to the left of the bunching region, the observed density distribution contains only shifting agents, and hence, $h^{shift}(z)=h(z)$ for $z<z^*-u_1$. However, within the diffuse region $[z^*-u_1,z^*)$, the observed post-kink density distribution contains both *shifters* and diffused *bunchers*. Assuming that $h^{shift}(z)$ is smooth, we then use the observed distribution $h(z)$ in the region $z<z^*-u_1$ to extrapolate the distribution of shifting agents in the diffusion region $[z^*-u_1,z^*)$.

Third, we compute the updated $\widehat{\Delta z^{*,act}}^{updated}$ based on the following relation:

$$\widehat{B}^{initial}=\int_{z^*-\widehat{\Delta z^{*,act}}^{updated}}^{z^*-1}h^{act,initial}(z)dz, \tag{43}$$

and check whether $\widehat{\Delta z^{*,act}}^{updated}$ equals $\widehat{\Delta z^{*,act}}^{initial}$. If $\widehat{\Delta z^{*,act}}^{updated}>\widehat{\Delta z^{*,act}}^{initial}$, we increase the value of $\widehat{\Delta z^{*,act}}^{initial}$ and repeat the above steps until we have $\widehat{\Delta z^{*,act}}^{updated}=\widehat{\Delta z^{*,act}}^{initial}$. Following the above process, we obtain the estimated marginal adjustment $\widehat{\Delta z^{*,act}}$ and the counterfactual density distribution $\hat{h}^{act}()$.

B.2.2. Estimating Counterfactual Outcome Distribution and Parameters

First, *never-takers* ($z>z^*$) do not respond to the kinked policy ($z=z^{act}$) but pay less money (like a lump-sum transfer, $\Delta T=\Delta t\times z^*$) under the kinked policy. Hence, their outcome distribution is a parallel shifting along the y-axis compared to their alternative counterfactual distribution. Specifically, from Equation (39), we have $y_n=y_n^{act}+\lambda\Delta t\times z^*$. Given that $z=z^{act}$ for *never-takers*, we have $y^{act}(z^{act})=y(z)-\lambda\Delta t\times z^*,\forall z^{act}=z>z^*$.

Second, for *shifters*, we have recovered marginal bunchers' responses $\Delta z^{*,act}$ and each shifter's

[41]The excess bunching at the threshold under the sharp bunching is $\widehat{B}^{initial}=h(z^*)-h^{act,initial}(z^*)$.

counterfactual location $z^{act} = z\frac{z^* - \Delta z^{*,act}}{z^*}, \forall z < z^*$ which forms the alternative counterfactual density distribution. To make sure that we are comparing the same *shifter* under the counterfactual and the kinked policies, we locate *shifters* back to their initial location, which generates the auxiliary outcome distribution under kinked policy $y^r(z^{act}), \forall z^{act} < z^* - \Delta z^{*,act}$.[42] It represents each *shifter*'s value of y under the kinked policy, including the direct impacts from changes in z and the impacts from changes in T, while excluding the relocation impacts (as we have located *shifters* back to their counterfactual locations).

As shown in Equations (35, 36), there would be both level and slope changes when comparing the counterfactual outcome distribution $y^{act}(z^{act}), \forall z^{act} < z^* - \Delta z^{*,act}$ with the auxiliary outcome distribution under the kinked policy $y^r(z^{act}), \forall z^{act} < z^* - \Delta z^{*,act}$. Moreover, if we extrapolate the obtained auxiliary distribution $y^r(z^{act})$ to the cutoff z^*, then the slope and the level change at z^* could be used to calibrate the sufficient statistics μ, λ as shown in Equations (35, 36). These parameters represent how changes in z directly impact y and how changes in T (due to change in z and the kinked policy) impact y.

Empirically, we jointly estimate the alternative counterfactual outcome distribution y^{act} and the slope and level changes. Specifically, we use the observed outcome distribution for *never-takers*, $y(z), \forall z = z^{act} > z^* + u_2$, and the obtained auxiliary outcome distribution for *shifters*, $y^r(z^{act}), \forall z^{act} < (z^* - u_1)\frac{z^* - \Delta z^{*,act}}{z^*}$ to fit a flexible polynomial distribution, allowing intercept and slope changes at the threshold. The estimation equation for the counterfactual outcome distribution is as follows:

$$\begin{aligned} y_j^{reg} &= \sum_{k=0}^{q} \alpha_k (z_j^{act})^k + a_0 I\left[z_j^{act} < z^*\right] + a_1 I\left[z_j^{act} < z^*\right] z_j^{act} + \varepsilon_j \\ \text{if} \quad & z_j^{act} < (z^* - u_1)\frac{z^* - \Delta z^{*,act}}{z^*} \text{ or } z_j^{act} > (z^* + u_2) \end{aligned} \tag{44}$$

where j indicates the bin; and q is the polynomial order; $y_j^{reg} = y_j$ for *never-takers* with $z_j^{ct} > (z^* + u_2)$, and $y_j^{reg} = y_j^r$ for *shifters* with $z_j^{ct} < (z^* - u_1)\frac{z^* - \Delta z^{*,act}}{z^*}$.

It is important to note that *never-takers* encounter a level change in outcome due to the lump-sum transfer ($\Delta t \times z^*$). Therefore,a_0 captures the level change between two distributions: (i) the auxiliary outcome distribution, (ii) the alternative counterfactual outcome distribution with the impact from the lump-sum transfer ($-\lambda \Delta t \times z^*$). Hence, we calibrate the values of λ, μ, based on

[42] Note $y^r(z^{act}) \equiv y(z^{act}\frac{z^*}{z^* - \Delta z^{*,act}}), \forall z^{act} < z^* - \Delta z^{*,act}$.

the following equations:

$$\begin{aligned} a_0 + \lambda \Delta t \times z^* &= \mu\left(\frac{z^*}{z^* - \Delta z^{*,act}} - 1\right) - \lambda z^*\left(t \times \frac{z^*}{z^* - \Delta z^{*,act}} - (t + \Delta t)\right) \\ a_1 &= -\lambda\left(t \times \frac{z^*}{z^* - \Delta z^{*,act}} - (t + \Delta t)\right) \end{aligned}$$

With two equations and two unknowns, we can calibrate $\widehat{\lambda}, \widehat{\mu}$.

Relying on the assumption that the relationship between outcome y and z would be smooth under the counterfactual policy, we obtain the counterfactual outcome distribution from Equation (44) as $\widehat{y_j^{ct}} = \sum_{k=0}^{q} \widehat{\alpha_k} (z_j^{act})^k - \widehat{\lambda}\widehat{\Delta t} \times z^*$.

Meanwhile, the treated outcome for *shifters* y_j^{shift} in $[z^* - u_1, z^*)$ is unobserved with diffused bunching, given that this region contains both *shifters* and diffused *bunchers* under the kinked policy. However, for the range $z < z^* - u_1$, there are only *shifters* under the kinked policy, therefore, $y_j^{shift} = y_j$ for $z < z^* - u_1$. Therefore, we fit a flexible polynomial to the observed distribution of y_j for *shifters* in the range $z < z^* - u_1$ and extrapolate the fitted distribution to obtain y_j^{shift} in $[z^* - u_1, z^*)$, with the assumption that the relationship between observed outcome $y^{shifter}$ and z under the kinked policy is smooth to the left of z^*.

Given that we have recovered the counterfactual density distribution, the counterfactual outcome distribution, and the density and outcome distributions of *shifters* within the diffuse bunching region, we can estimate the impacts of the kinked policy on *bunchers*, *shifters*, and *never-takers* following Equations (33) , (37) and (39).

Appendix C: Formation of Counterfactual Outcome Distribution and Treatment Effects

Assume each agent n may have different initial values of y, denoted as $y^{pre}(n)$, which could be of any functional form.

Card et al. (2015) set up the assumptions on the constant-effect, additive model for identifying the causal effects in regression kink designs. We follow their practice when identifying causal effects in kink settings with manipulative agents as in the bunching framework. Specifically, we assume

$$y(n) = y^{pre}(n) + \mu \ln z(n) - \lambda T\big(z(n)\big) \tag{45}$$

where $T\big(z(n)\big)$ is a deterministic and continuous function of z. Under the kinked policy, $T\big(z(n)\big)$ has a kink at z^*. $\mu \equiv \frac{\Delta y_n}{\Delta \ln z_n}$, reflecting the direct impact of percentage changes in z on outcome y. $-\lambda_n \equiv \frac{\Delta y_n}{\Delta T_n}$, reflecting the impact of changes in T on outcome y.

Under the counterfactual policy with a linear low tax/co-payment rate, we have $z^{ct} = n(1-t)^e$ and $T^{ct}(z) = tz^{ct} = tn(1-t)^e$. Therefore, from Equation (45), we have $y^{ct}(n) = y^{pre}(n) + \mu \ln n(1-t)^e - \lambda tn(1-t)^e$. Rewrite it in the form of the relation between y and z, we have

$$y^{ct}(z^{ct}) = y^{pre}\left(\frac{z}{(1-t)^e}\right) + \mu \ln z^{ct} - \lambda t z^{ct} \tag{46}$$

Under the kinked policy, agents with $z^{ct} \leq z^*$ (i.e., *always-takers*) remain unchanged. Agents with $z^{ct} > z^*$ (i.e., *shifters*) will reduce their values of z. Specifically, for *shifters*, we have $z = n(1-t-\Delta t)^e$ and $T(z) = (t+\Delta t)z - \Delta t z^*$. Therefore, from Equation (45) we have $y(n) = y^{pre}(n) + \mu \ln n(1-t-\Delta t)^e - \lambda(t+\Delta t)n(1-t-\Delta t)^e + \lambda \Delta t z^*$. Rewrite it in the form of the relation between y and z, we have

$$y(z) = y^{pre}\left(\frac{z}{(1-t-\Delta t)^e}\right) + \mu \ln z - \lambda(t+\Delta t)z + \lambda \Delta t z^* \tag{47}$$

When we relocate *shifters* back to their counterfactual location to obtain the auxiliary out-

come distribution, the above Equation (47) would become

$$y^r(z^{ct}) = y^{pre}\left(\frac{z^{ct}}{(1-t)^e}\right) + \mu \ln z^{ct}\frac{(1-t-\Delta t)^e}{(1-t)^e} - \lambda(t+\Delta t)z^{ct}\frac{(1-t-\Delta t)^e}{(1-t)^e} + \lambda \Delta t z^* \tag{48}$$

After relocating *shifters* back to their counterfactual locations, comparing the values along the y-axis would give us the treatment effect on *"shifters"*. That is,

$$\begin{aligned}
\tau_y^{TE,shifter} &= y_n - y_n^{ct} \\
&= y^r(z^{ct}) - y^{ct}(z^{ct}) \\
&= \mu \ln \frac{(1-t-\Delta t)^e}{(1-t)^e} - z^{ct}\lambda(t+\Delta t)\frac{(1-t-\Delta t)^e}{(1-t)^e} + \lambda \Delta t z^* + z^{ct}\lambda t, \\
&= \mu \ln \frac{z^*}{z^*+\Delta z^*} - z^{ct}\lambda(t+\Delta t)\frac{z^*}{z^*+\Delta z^*} + \lambda \Delta t z^* + z^{ct}\lambda t, \\
&\quad (\forall z^{ct} > z^* + \Delta z^*, i.e., \forall n \in shifters)
\end{aligned}$$

Comparing the difference between $y^{ct}(z^{ct})$ and $y^r(z^{ct})$ from Equations (46, 48), we would notice that the unknown underlying distribution $y^{pre}\left(\frac{z^{ct}}{(1-t)^e}\right)$ is canceled out, and we are left with $\mu \ln \frac{z^*}{z^*+\Delta z^*} - z^{ct}\lambda(t+\Delta t)\frac{z^*}{z^*+\Delta z^*} + \lambda \Delta t z^* + z^{ct}\lambda t$. Further, the difference could be decomposed into slope change and level change. Specifically, by extrapolating $y^{ct}(z^{ct})$ and $y^r(z^{ct})$ to the point with $z^{ct} = z^*$, we would have[43]

$$\text{Level Change at } z^* = \mu \ln\left(\frac{z^*}{z^*+\Delta z^*}\right) - \lambda(t+\Delta t)z^*\left(\frac{z^*}{z^*+\Delta z^*} - 1\right) \tag{49}$$

$$\text{Slope Change at } z^* = -\lambda\left((t+\Delta t)\frac{z^*}{z^*+\Delta z^*} - t\right) \tag{50}$$

[43] we could check slope and level changes at other values of z^{ct} as well, by extrapolating $y^{ct}(z^{ct})$ and $y^r(z^{ct})$ to the corresponding location and plug into Equation (49)

Appendix D: Extension with Relabelling under Homogeneous Cost Function

Faced with monetary incentives, agents might engage in misreporting or other relabelling behavior. Denote z^{rp}, z^{rl} as the reported and real values of z. Denote the degree of misreporting as $\delta \equiv \frac{z^{rl}-z^{rp}}{z^{rl}}$. Assume that relabeling cost depends on the absolute value and the relative degree of relabeling, that is, $c \times z^{rl} \times g(\delta)$, where c is a fixed parameter, and $g'(\delta) > 0, g''(\delta) > 0, g(0) = 0$.[44] Therefore, the marginal cost of an additional degree of relabelling is $cz^{rl}g'(\delta)$.

For illustration purposes, we incorporate relabelling in the setup by Saez (2010) (also illustrated in section 2.1). Specifically, the agent's utility function under a counterfactual linear tax system is given as follows:

$$U = z^{rl} - tz^{rp} - \frac{n}{1+1/e}(\frac{z^{rl}}{n})^{1+1/e} - cz^{rl}g(\delta),$$

where n denotes individual heterogeneity. First-order conditions give us the following optimal choices:

$$\begin{aligned} \delta^{ct} &= g'^{-1}(\frac{t}{c}) \equiv \delta_t \\ z^{rl,ct} &= [1-t(1-\delta_t)-cg(\delta_t)]^e n \\ \textit{implying } z^{rp,ct} &= z^{rl,ct}(1-\delta_t) \end{aligned}$$

This indicates that all agents engage in the same degree of relabelling δ^{ct} and their optimal real response $z^{rl,ct}$ and reported response $z^{rp,ct}$ are proportional to their ability n.

When a kinked policy is introduced, agents' real and relabelling behaviors would change. Following the previous example by Saez (2010) (also introduced in section 2.1), consider a kinked policy that leaves the marginal tax rate at t for income $z \leq z^*$ and sets the marginal tax rate $t + \Delta t$ for income $z_{>}z^*$. Similar to the baseline analysis, we have three groups of agents.

First, agents with $z^{rp,ct} \leq z^*$ (i.e., *always-takers*) face no change in marginal incentives and

[44] Chen et al. (2021) adopt the same assumption on relabelling cost.

hence set

$$\begin{aligned}
\delta &= \delta^{ct} = g'^{-1}(\frac{t}{c}) \equiv \delta_t \\
z^{rl} &= z^{rl,ct} = \left[1 - t(1-\delta_t) - cg(\delta_t)\right]^e n \\
z^{rp} &= z^{rp,ct} = z^{rl}(1-\delta_t) \\
\forall n &\in \textit{always-takers}
\end{aligned}$$

Second, agents with $z^{rp,ct} > z^* + \overline{\Delta z^*}$ (i.e., *shifters*) face a change in the marginal benefit and adjust their optimal responses accordingly. Specifically, for *shifters*, we have:

$$\begin{aligned}
\delta &= g'^{-1}(\frac{t+\Delta t}{c}) \equiv \delta_{t+\Delta t} \\
z^{rl} &= \left[1 - (t+\Delta t)(1-\delta_{t+\Delta t}) - cg(\delta_{t+\Delta t})\right]^e n \\
z^{rp} &= z^{rl}(1-\delta_{t+\Delta t}) \\
\forall n &\in \textit{shifters}
\end{aligned}$$

This indicates that all shifters engage in the same degree of relabelling $\delta_{t+\Delta t}$ and their optimal real response is proportional to their ability. Moreover, we have

$$\begin{aligned}
\frac{z^{rl}}{z^{rl,ct}} &= \left(\frac{1-(t+\Delta t)(1-\delta_{t+\Delta t}) - cg(\delta_{t+\Delta t})}{1-t(1-\delta_t) - cg(\delta_t)}\right)^e \\
\frac{z^{rp}}{z^{rp,ct}} &= \frac{1-\delta_{t+\Delta t}}{1-\delta_t}\frac{z^{rl}}{z^{rl,ct}} = \frac{z^*}{z^* + \overline{\Delta z^*}} \\
\forall n &\in \textit{shifters}
\end{aligned}$$

Therefore, *shifters* change their reported values of z by a constant percentage.[45]

Third, agents with $z^{rp,ct} \in (z^*, z^* + \overline{\Delta z^*}]$ (i.e., *bunchers*) also face a change in their marginal incentives but are subject to a corner solution. These agents set $z^{rp} = z^*$ and bunch at the cutoff. They choose different optimal degrees of relabelling δ depending on how far their counterfactual

[45]The last equality follows the baseline analysis that a marginal shifting agent is also a marginal bunching agent.

value $z^{rp,ct}$ is away from the cutoff. Specifically, we have:

$$\begin{aligned} z^* &= \left(1 - cg(\delta) - cg'(\delta)(1-\delta)\right)^e (1-\delta)n, \\ z^{rp} &= z^*, \\ z^{rl} &= \frac{z^*}{1-\delta} \\ \forall n &\in \textit{bunchers} \end{aligned}$$