Online Appendix: Estimating Adjustment Frictions Using Non-Linear Budget Sets: Method and Evidence from the Earnings Test, by Alexander M. Gelber, Damon Jones, and Daniel W. Sacks

A1. Polynomial Adjustment Costs

We now extend the adjustment cost to encompass a polynomial adjustment cost, allowing for greater generality than a fixed cost. We begin with an adjustment cost that increases linearly in the size of the adjustment, which illustrates how the method generalizes for higher-order polynomials. Assume that given an initial level of earnings z_0, agents must pay a cost of $\phi^* \cdot |z - z_0|$ when they change their earnings to a new level z. Utility $\tilde{u}$ at the new earnings level can be represented as:

$$\tilde{u}\left(c, z; n, z_0\right) = u\left(c, z; n\right) - \phi^* \cdot \left|z - z_0\right|.$$

The first order condition for earnings can be characterized as:

$$\begin{aligned} -\frac{u_z\left(c, z; n\right)}{u_c\left(c, z; n\right)} &= \left(1 - \tau - \phi^*/\lambda^* \cdot \text{sgn}\left(z - z_0\right)\right) \\ &= \begin{cases} \left(1 - \tau - \phi\right) & \text{if } z > z_0 \\ \left(1 - \tau + \phi\right) & \text{if } z < z_0 \end{cases}, \end{aligned}$$

where $\lambda^* = u_c\left(c^*, z^*; n\right)$ is the Lagrange multiplier and $\phi = \phi^*/\lambda^*$ is the dollar equivalent of the linear adjustment cost ϕ^*.

The individual chooses earnings as if he faces an effective marginal tax rate of $\tilde{\tau} = \tau + \phi{\cdot}\text{sgn}(z - z_0)$. It follows that our predictions about earnings adjustment are similar to our previous predictions, except that the effective marginal tax rate $\tilde{\tau}$ appears, rather than τ. Thus, we can solve for the elasticity of earnings as a function of the change in earnings $\triangle z^*$ due to introduction of a kink in the tax schedule and the jump in marginal tax rate $d\tau_1$:

$$\begin{aligned} \varepsilon &= \frac{\triangle z^*/z^*}{d\tilde{\tau}_1/\left(1 - \tilde{\tau}_0\right)} \\ &= \frac{\triangle z^*/z^*}{\left(d\tau_1 - 2\phi\right)/\left(1 - \tau_0 - \phi\right)}. \end{aligned}$$

Since the right-hand side is increasing in ϕ, the estimate of the elasticity increases as the linear adjustment cost increases. This makes intuitive sense: the adjustment cost attenuates bunching, so holding constant the level of bunching, the elasticity must be higher as the adjustment cost increases.

Now assume that when an individual adjusts his earnings, he incurs a linear adjustment cost ϕ^{*L} for every unit of change in earnings, as well as a fixed cost ϕ^{*F} associated with any change in earnings. Consider again bunching at z^*, with a tax rate jump of $d\tau_1 = \tau_1 - \tau_0$ at earnings level z^*. We have the following set of expressions for excess mass:

$$\begin{aligned} B &= \int_{\underline{z}}^{z^*+\triangle z^*} h_0(\zeta)\, d\zeta \\ \varepsilon &= \frac{\triangle z^*/z^*}{\left(d\tau_1 - 2\phi^L\right)/\left(1-\tau_0-\phi^L\right)} \\ \phi^{*F} + \phi^{*L}\cdot\left(\underline{z}-z^*\right) &= u\left(\left(1-\tau_1\right)z^* + R', z^*; \underline{n}\right) - u\left(\left(1-\tau_1\right)\underline{z} + R', \underline{z}; \underline{n}\right). \end{aligned}$$

In this case, we need at least three kinks to separately identify $\left(\varepsilon, \phi^F, \phi^L\right)$. A similar argument generalizes this to the case of any polynomial adjustment cost: for a polynomial adjustment cost of order n, we need $n+1$ moments to identify these parameters as well as the elasticity.

A2. Dynamic Model with Forward-Looking Behavior

We present in this appendix a version of the dynamic model in Section V.C in which we allow for forward-looking behavior. The key difference in implications is that in addition to a gradual, lagged response to policy changes, this version of the model also predicts anticipatory adjustment by agents when policy changes are anticipated in advance. We have essentially the same setting as in Section V.C, except that we will alter three of the assumptions. First, in each period, an individual draws a cost of adjustment, $\tilde{\phi}_t$, from a discrete distribution, which takes a value of ϕ with probability π and a value of 0 with probability $1-\pi$.[13] Second, individuals make decisions over a finite horizon, living until Period $\overline{\mathcal{T}}$. In period 0, the individuals face a linear tax schedule, $T_0(z) = \tau_0 z$, with marginal tax rate τ_0. In period 1, a kink, K_1, is introduced at the earnings level z^*. This tax schedule is implemented for $\mathcal{T}_1$ periods, after which the tax schedule features a smaller kink, K_2, at the earnings level z^*. The smaller kink is present until period $\mathcal{T}_2$, after which we return to the linear tax schedule, T_0. As before, the kink K_j, $j \in \{1,2\}$, features a top marginal tax rate of τ_j for earnings above z^*.[14] Finally, in each period, individuals solve this maximization problem:

$$\text{(A1)} \qquad \max_{\left(c_{a,t}, z_{a,t}\right)} v\left(c_{a,t}, z_{a,t}; a, z_{a,t-1}\right) + \delta V_{a,t+1}\left(z_{a,t}, A_{a,t}\right),$$

[13] For expositional purposes, we constrain the probability of drawing a nonzero fixed costs to be π in all periods. Thus, the terms from Section V.C of the form $\prod \pi_j$ simplify to π^j in this appendix. All results go through with the more flexible distribution of adjustment costs in Section V.C.

[14] In Section V.C, we do not specify time $\mathcal{T}_2$, when the smaller kink, K_2, is removed, as it is not relevant to the case where individuals are not forward-looking.

where $v(c_{a,t}, z_{a,t}; a, z_{a,t-1}) \equiv u(c_{a,t}, z_{a,t}; a) - \tilde{\phi}_t \cdot \mathbf{1}(z_{a,t} \neq z_{a,t-1})$, δ is the discount factor, and $V_{a,t+1}$ is the value function moving forward in Period $t+1$:

(A2)

$$V_{a,t+1}(\zeta, A_{a,t}) = \mathbb{E}_{\phi}\left[\max_{(c_{a,t+1}, z_{a,t+1})} v(c_{a,t+1}, z_{a,t+1}; a, \zeta) + \delta V_{a,t+2}(z_{a,t+1}, A_{a,t+1})\right].$$

$V_{a,t+1}$ is a function of where the individual has chosen to earn in Period t and assets $A_{a,t}$. The expectation $\mathbb{E}_{\phi}[\cdot]$ is taken over the distribution of $\tilde{\phi}_t$. The intertemporal budget constraint is:

$$A_{a,t} = (1+r)(A_{a,t-1} + z_{a,t} - T(z_{a,t}) - c_{a,t}). \quad \text{(A3)}$$

We assume that $\delta(1+r) = 1$. Because individuals have quasilinear preferences, this implies that consumption can be set to disposable income in each period: $c_{a,t} = z_{a,t} - T(z_{a,t})$. We therefore use the following shorthand:

$$\begin{aligned} u_a^j(z) &= u(z - T_j(z), z; a) \\ V_{a,t}(z) &= V_{a,t}(z, A_{a,t-1}) \end{aligned} \quad \text{(A4)}$$

Next, we define two operators that measure the utility gain (or loss) following a discrete change in earnings:

$$\begin{aligned} \triangle u_a^j(z, z') &= u_a^j(z) - u_a^j(z') \\ \triangle V_{a,t}(z, z') &= V_{a,t}(z) - V_{a,t}(z') \end{aligned} \quad \text{(A5)}$$

In each case above, the utility and utility differential depend on the tax schedule. We define z_a^j as the optimal level of earnings under a frictionless, static optimization problem, facing the tax schedule T_j. We will refer to the frictionless, *dynamic* optimum in any given period as $\tilde{z}_{a,t}$.[15] This is the optimal level of earnings when there is a fixed cost of zero drawn in the *current* period, but a nonzero fixed cost may be drawn in future periods. We will also make a distinction between two types of earnings adjustments: *active* and *passive*. An *active* earnings adjustment takes place in the presence of a nonzero fixed cost, while a *passive* earnings adjustment takes place only when a fixed cost of zero is drawn. We solve the model recursively, beginning in the regime after time $\mathcal{T}_2$, when the smaller kink, K_2, has been removed, continuing with the solution while the kink K_2 is present between times $\mathcal{T}_1$ and $\mathcal{T}_2$, and finally considering the first regime when the kink K_1 is present between time period 1 and $\mathcal{T}_1$.[16]

[15] In a model with no forward-looking behavior, $z_a^j = \tilde{z}_{a,t}$.

[16] Our recursive method can be extended to the case of multiple, successive kinks. The effect on bunching of a sequence of more kinks depends on the relative size of the successive kinks.

Earnings between $\mathcal{T}_2$ and $\overline{\mathcal{T}}$. — We will now derive the value function $V_{a,\mathcal{T}_2+1}(z)$. We begin with the following result: If an individual with initial earnings z makes an active adjustment in period $t > \mathcal{T}_2 + 1$, then it must be the case that

$$\text{(A6)} \qquad \frac{1-(\delta\pi)^{\mathcal{T}_1+1-t}}{1-\delta\pi} \triangle u_a^0\left(z_a^0, z\right) \geq \phi.$$

We demonstrate this result with a constructive proof, showing the result for periods $\overline{\mathcal{T}}$ and $\overline{\mathcal{T}}-1$. Because the tax schedule is constant throughout this terminal period, the frictionless, dynamic optimum is equal to the static optimum: $\tilde{z}_{a,t} = z_a^0$. First, consider an agent in period $\overline{\mathcal{T}}$, with initial earnings z, who is considering maintaining earnings at z or paying the fixed cost ϕ and making an active adjustment to z_a^0, the frictionless, dynamic optimum in period $\overline{\mathcal{T}}$. The agent will make the adjustment if:

$$\text{(A7)} \qquad \begin{aligned} \triangle u_a^0\left(z_a^0, z\right) &\geq \phi \\ &= \frac{1-\delta\pi}{1-\delta\pi}\phi. \end{aligned}$$

Rearranging terms, we have satisfied the inequality in (A6).

Now consider agents in period $\overline{\mathcal{T}}-1$ with initial earnings z. There are two types, those who would make an active adjustment to z_a^0 in period $\overline{\mathcal{T}}$ if the earnings z are carried forward and those who would not. Consider those who would not. If the agent remains with earnings of z, then utility will be $u_a^0(z) + \delta V_{a,\overline{\mathcal{T}}}(z) = u_a^0(z) + \delta\left[\pi\left(u_a^0(z)\right) + (1-\pi)\, u_a^0\left(z_a^0\right)\right]$. If the agent actively adjusts to z_a^0, then utility will be $u_a^0\left(z_a^0\right) - \phi + \delta u_a^0\left(z_a^0\right)$. The agent will actively adjust in period $\overline{\mathcal{T}}-1$ if:

$$\text{(A8)} \qquad \begin{aligned} \triangle u_a^0\left(z_a^0, z\right) &\geq \frac{1}{1+\delta\pi}\phi \\ &= \frac{1-\delta\pi}{1-(\delta\pi)^2}\phi. \end{aligned}$$

Once again, rearranging terms confirms that (A6) holds. Finally, consider agents who would actively adjust from z to z_a^0 if earnings level z is carried forward. In this case, the agent's utility when remaining at z is:

$$\text{(A9)} \qquad \begin{aligned} u_a^0(z) + \delta V_{a,\overline{\mathcal{T}}}(z) &= u_a^0(z) + \delta\left[\pi\left(u_a^0\left(z_a^0\right) - \phi\right) + (1-\pi)\, u_a^0\left(z_a^0\right)\right] \\ &= u_a^0(z) + \delta\left(u_a^0\left(z_a^0\right) - \pi\phi\right). \end{aligned}$$

Intuitively, the agent will receive the optimal level of utility in the next period, and with probability π the agent will have to pay the fixed cost to achieve it. Similarly, the agent's utility after actively adjusting to z_a^0 in period $\overline{\mathcal{T}}-1$ is

$u_a^0\left(z_a^0\right) - \phi + \delta u_a^0\left(z_a^0\right)$. The agent will therefore adjust in period $\overline{\mathcal{T}}$ if:

$$\triangle u_a^0\left(z_a^0, z\right) \geq (1 - \delta\pi)\,\phi. \tag{A10}$$

However, we know from (A7) that this already holds for the agent who actively adjusts in period $\overline{\mathcal{T}}$. Finally, note that (A7) implies (A8). It follows that in period $\overline{\mathcal{T}} - 1$, adjustment implies (A7). We can similarly show the result for earlier periods by considering separately: (a) those who would actively adjust in the current period, but not in any future period; and (b) those who would adjust in some future period. Both types will satisfy the key inequality. As a corollary, note that if an individual with initial earnings z makes an active adjustment in period $t > \mathcal{T}_2 + 1$, then she will also find it optimal to do so in any period t', where $\mathcal{T}_2 < t' < t$. To see this, note that if (A6) holds for t, then it also holds for $t' < t$. It follows that the agent would also actively adjust in period t'.

Now consider an agent who earns z in period $\mathcal{T}_2$. Note that our results above imply that any active adjustment that takes place after $\mathcal{T}_2$ will only happen in period $\mathcal{T}_2 + 1$. These agents will receive a stream of discounted payoffs of $u_a^0\left(z_a^0\right)$ for $\overline{\mathcal{T}} - \mathcal{T}_2$ periods, *i.e.* $\sum_{j=0}^{\overline{\mathcal{T}}-\mathcal{T}_2-1} \delta^j u_a^0\left(z_a^0\right) = \frac{1-\delta^{\overline{\mathcal{T}}-\mathcal{T}_2}}{1-\delta} u_a^0\left(z_a^0\right)$, and pay a fixed cost of ϕ in period $\mathcal{T}_2$ with probability π. Otherwise, an agent will adjust to the dynamic frictionless optimum z_a^0 only when a fixed cost of zero is drawn. In the latter case, the agent receives a payoff of $u_a^0(z)$ until a fixed cost of zero is drawn, after which, the agent receives $u_a^0\left(z_a^0\right)$. We can therefore derive the following value function:[17]

$$V_{a,\mathcal{T}_2+1}(z) = \begin{cases} \frac{1-\delta^{\overline{\mathcal{T}}-\mathcal{T}_2}}{1-\delta} u_a^0\left(z_a^0\right) - \pi\phi & \text{if } \frac{1-(\delta\pi)^{\overline{\mathcal{T}}-\mathcal{T}_2}}{1-\delta\pi} \triangle u_a^0\left(z_a^0, z\right) \geq \phi \\ \frac{1-\delta^{\overline{\mathcal{T}}-\mathcal{T}_2}}{1-\delta} u_a^0\left(z_a^0\right) - \pi \frac{1-(\delta\pi)^{\overline{\mathcal{T}}-\mathcal{T}_2}}{1-\delta\pi} \triangle u_a^0\left(z_a^0, z\right) & \text{otherwise} \end{cases}. \tag{A11}$$

To gain some intuition for (A6), note that the left side of (A6) is the net present value of the stream of the utility differential once the agent adjusts from z to z_a^0. If this exceeds the up-front cost of adjustment, ϕ, then the agent actively adjusts. The discount factor for j periods in the future, however, is $(\delta\pi)^j$, instead of only δ^j. The reason is that current adjustment only affects future utility j periods from now if j consecutive nonzero fixed costs are drawn, which happens with probability π^j. To better understand our second result regarding the timing of active changes, note that if the gains from adjustment over $\overline{\mathcal{T}} - t$ periods exceed the up-front cost, then the agent should also be willing to adjust in period $t' < t$ and accrue $\overline{\mathcal{T}} - t'$ periods of this gain, for the same up-front cost of ϕ.

[17] The expected utility for passive adjusters is constructed recursively, working backward from period $\overline{\mathcal{T}}$ to period $\mathcal{T}_2 + 1$.

Earnings between $\mathcal{T}_1$ and $\mathcal{T}_2$. — We now derive the value function $V_{a,\mathcal{T}_1+1}(z)$. In this case, the dynamic frictionless optimum in each period, $\tilde{z}_{a,t}$, is not constant. Intuitively, the agent trades off the gains from adjusting earnings in response to K_2 with the effect of this adjustment on the value function $V_{a,\mathcal{T}_2+1}$. In general, the optimum is defined as:

$$\text{(A12)} \qquad \tilde{z}_{a,t} = \arg\max_{z\in[z_a^2, z_a^0]} \frac{1-(\delta\pi)^{\mathcal{T}_2+1-t}}{1-\delta\pi} u_a^2(z) + \delta^{\mathcal{T}_2+1-t}\pi^{\mathcal{T}_2-t} V_{a,\mathcal{T}_2+1}(z).$$

We restrict the maximization to the interval $[z_a^2, z_a^0]$, since reducing earnings below z_a^2 or raising earnings above z_a^0 weakly reduces utility in any current and all future periods for $t > \mathcal{T}_1$. From (A11), we know that $V_{a,\mathcal{T}_2+1}$ is continuous, and thus the solution in (A12) exists.[18] We present two results analogous to those in Section A.A2, without proof. The proofs, nearly identical to those in the previous section, are available upon request. First, if an individual with initial earnings z makes an active adjustment in period t, $\mathcal{T}_1 < t \leq \mathcal{T}_2$, then:

$$\text{(A13)} \qquad \frac{1-(\delta\pi)^{\mathcal{T}_2+1-t}}{1-\delta\pi} \triangle u_a^2(\tilde{z}_{a,t}, z) + \delta^{\mathcal{T}_2+1-t}\pi^{\mathcal{T}_2-t} \triangle V_{a,\mathcal{T}_2+1}(\tilde{z}_{a,t}, z) \geq \phi.$$

Furthermore, if an individual with initial earnings z makes an active adjustment in period t, $\mathcal{T}_1 < t \leq \mathcal{T}_2$, then she will also find it optimal to do so in any period t', where $\mathcal{T}_1 < t' < t$.

The condition in (A13) differs from that in (A6) because the effect of adjustment on the utility beyond period $\mathcal{T}_2$ is taken into account, in addition to the up-front cost of adjustment, ϕ. Any adjustment in this time interval, active or passive, will be to the dynamic, frictionless optimum for the current period, $\tilde{z}_{a,t}$. As before, (A13) implies that all active adjustment occurring between $\mathcal{T}_1 + 1$ and $\mathcal{T}_2$ takes place in period $\mathcal{T}_1 + 1$. Those who adjust in period $\mathcal{T}_1 + 1$ will earn $\tilde{z}_{a,\mathcal{T}_1+1}$. Thereafter, they only adjust to $\tilde{z}_{a,t}$ when a fixed cost of zero is drawn. Likewise, those who only adjust passively earn $z_{a,\mathcal{T}_1}$ in period $\mathcal{T}_1 + 1$, and thereafter adjust to $\tilde{z}_{a,t}$ when a fixed cost of zero is drawn. We can therefore derive the following

[18] Technically, we can see from (A11) that while the function V_{a,P_2+1} is continuous, it is kinked, which creates a nonconvexity. Thus, the solution in (A12) may not always be single-valued. In such cases, we define $\tilde{z}_{a,t}$ as the lowest level of earnings that maximizes utility.

value function:
(A14)

$$V_{a,\mathcal{T}_1+1}(z) = \begin{cases} \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-1} \delta^j u_a^2\left(\tilde{z}_{a,\mathcal{T}_1+1+j}\right) + \delta^{\mathcal{T}_2-\mathcal{T}_1} \triangle V_{a,\mathcal{T}_2+1}\left(\tilde{z}_{a,\mathcal{T}_2}\right) \\ - \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-2} \frac{(\delta\pi)^{\mathcal{T}_2-\mathcal{T}_1}}{\pi^{j+1}} \triangle V_{a,\mathcal{T}_2+1}\left(\tilde{z}_{a,\mathcal{T}_1+2+j}, \tilde{z}_{a,\mathcal{T}_1+1+j}\right) & \text{if (A13) is satisfied when } t = \mathcal{T}_1+1 \\ - \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-2} \frac{1-(\delta\pi)^{\mathcal{T}_2-\mathcal{T}_1-1-j}}{1-\delta\pi} \delta^{j+1}\pi \triangle u_a^2\left(\tilde{z}_{a,\mathcal{T}_1+2+j}, \tilde{z}_{a,\mathcal{T}_1+1+j}\right) \\ -\pi\phi \\ \\ \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-1} \delta^j u_a^2\left(\tilde{z}_{a,\mathcal{T}_1+1+j}\right) + \delta^{\mathcal{T}_2-\mathcal{T}_1} \triangle V_{a,\mathcal{T}_2+1}\left(\tilde{z}_{a,\mathcal{T}_2}\right) \\ - \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-2} \frac{(\delta\pi)^{\mathcal{T}_2-\mathcal{T}_1}}{\pi^{j+1}} \triangle V_{a,\mathcal{T}_2+1}\left(\tilde{z}_{a,\mathcal{T}_1+2+j}, \tilde{z}_{a,\mathcal{T}_1+1+j}\right) \\ - \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-2} \frac{1-(\delta\pi)^{\mathcal{T}_2-\mathcal{T}_1-1-j}}{1-\delta\pi} \delta^{j+1}\pi \triangle u_a^2\left(\tilde{z}_{a,\mathcal{T}_1+2+j}, \tilde{z}_{a,\mathcal{T}_1+1+j}\right) & \text{otherwise} \\ -\pi \left\{ \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-1} (\delta\pi)^j \triangle u_a^2\left(\tilde{z}_{a,\mathcal{T}_1+1}, z\right) \right. \\ \qquad \left. - \delta^{\mathcal{T}_2-\mathcal{T}_1}\pi^{\mathcal{T}_2+1-\mathcal{T}_1} \triangle V_{a,\mathcal{T}_2+1}\left(\tilde{z}_{a,\mathcal{T}_1+1}, z\right) \right\} \end{cases}$$

The first case in (A14) applies to those who actively adjust in period $\mathcal{T}_1+1$ and passively adjust thereafter. The first line is the utility that would accrue if a fixed cost of zero were drawn in each period. The next two lines represent the deviation from this stream of utility, due to nonzero fixed costs potentially drawn in periods $\mathcal{T}_1+1$ through $\mathcal{T}_2$. The final line represents the fixed cost that is paid in period $\mathcal{T}_1+1$ with probability π. The second case in (A14) applies to those who only passively adjust. The first three lines remain the same. The final two lines represent a loss in utility attributed to fact that earnings in period $\mathcal{T}_1+1$ may not be $\tilde{z}_{a,T_1+1}$. Note that earnings in period $\mathcal{T}_1$ can only affect utility through this last channel.

Earnings between Period 1 and $\mathcal{T}_1$. — Earnings during the first period, when the kink K_1 is present, can be derived similarly. The dynamic, frictionless optimum is now defined as:

$$\text{(A15)} \qquad \tilde{z}_{a,t} = \arg\max_{z\in[z_a^1,z_a^0]} \frac{1-(\delta\pi)^{\mathcal{T}_1+1-t}}{1-\delta\pi} u_a^1(z) + \delta^{\mathcal{T}_1+1-t}\pi^{\mathcal{T}_1-t} V_{a,\mathcal{T}_1+1}(z).^{19}$$

[19] Note, the objective function now features two potential nonconvexities. In cases where the solution is multi-valued, we again define $\tilde{z}_{a,t}$ as the lowest earnings level from the set of solutions.

Similar to the other cases, if an individual with initial earnings z makes an active adjustment in period t, $0 < t \leq \mathcal{T}_1$, then it must be the case that

$$\text{(A16)} \qquad \frac{1-(\delta\pi)^{\mathcal{T}_1+1-t}}{1-\delta\pi}\triangle u_a^1\left(\tilde{z}_{a,t},z\right)+\delta^{\mathcal{T}_1+1-t}\pi^{\mathcal{T}_1-t}\triangle V_{a,\mathcal{T}_1+1}\left(\tilde{z}_{a,t},z\right)\geq\phi.$$

Furthermore, if an individual with initial earnings z makes an active adjustment in period t, $0 < t \leq T_1$, then she will also find it optimal to do so in any period t', where $0 < t' < t$. Again, this implies that all active adjustment will take place in period 1. Since individuals begin with earnings of z_a^0, we know that all active adjustment will be downward. Thereafter, it can be shown that $\tilde{z}_{a,t}$ is weakly increasing, and upward adjustment will occur passively.

Characterizing Bunching. — Given these results, we can now derive expressions for excess mass at z^* analogous to (8) and (9). For notational convenience, we define $\mathcal{A}_j(z)$ as the set of individuals, a, with initial earnings z who actively adjust in period j. Again, denote B_1^t as bunching at K_1 in period $t \in [1, \mathcal{T}_1]$. We have the following generalized version of (8):

$$\begin{aligned}B_1^t \;=\; &\int_{z^*}^{z^*+\triangle z_1^*}\Bigg[\mathbf{1}\left\{\tilde{z}_{a,1}=z^*, a\in\mathcal{A}_1(\zeta)\right\}\\ &+\sum_{j=1}^{t}\left(1-\pi^j\right)\pi^{t-j}\mathbf{1}\left\{\sup\left\{l|\, l\leq t,\tilde{z}_{a,l}=z^*\right\}=j, a\notin\mathcal{A}_1(\zeta)\right\}\\ &-\sum_{j=1}^{t-1}\left(1-\pi^{t-j}\right)\mathbf{1}\left\{\sup\left\{l|\, l\leq t,\tilde{z}_{a,l}=z^*\right\}=j, a\in\mathcal{A}_1(\zeta)\right\}\Bigg]h_0(\zeta)\,d\zeta.\end{aligned} \qquad \text{(A17)}$$

We have partitioned the set of potential bunchers into three groups in (A17). In the first line, we have the set of active bunchers in period 1. In the second line, we capture individuals who are passive bunchers, *i.e.* $a \notin \mathcal{A}_1\left(z_a^0\right)$. For $j \in [1, t-1]$, the indicator function selects the individual who has $\tilde{z}_{a,j} = z^*$ but $\tilde{z}_{a,j+1} \neq z^*$. Since $\tilde{z}_{a,t}$ is weakly increasing, the optimal earnings for this individual is z^* in periods 1 through $j-1$. The probability that the individual bunches by period j is $1-\pi^j$. Thereafter, the individual will de-bunch if a fixed cost of zero is drawn. The probability of only drawing nonzero fixed costs thereafter is π^{t-j}. For $j = t$, the indicator function selects agents for whom $\tilde{z}_{a,t} = z^*$. Their probability of passively bunching by period t is $1-\pi^t$. The third line captures the outflow of active bunchers, for whom $\tilde{z}_{a,t}$ ceases to be z^* starting in period j. The probability of having drawn a nonzero fixed cost and de-bunching since period j is $1-\pi^{t-j}$.

Equation (A17) differs from (8) in three key ways. First, the set of active bunchers in period 1 is different, as can be seen by comparing (A16) and the relevant condition for active bunchers in Section V.C, $\triangle u_a^1\left(z^*, z_a^0\right) \geq \phi$. The utility gain accrues for multiple periods in the forward-looking case, increasing the probability of actively bunching, but the effect of adjustment on future payoffs via $V_{a,\mathcal{T}_1+1}$ may either reinforce or offset this incentive. Furthermore, passive bunchers are (weakly) less likely to remain bunching, as they de-bunch in anticipation of policy changes in future periods. To see this, note that the π^{t-j} factor is decreasing in t. Finally, the set of active bunchers similarly de-bunch passively, in anticipation of future policy changes. The model therefore predicts a gradual outflow from the set of bunchers, in anticipation of the shift from K_1 to K_2. Nonetheless, the overall net change in bunching over time is ambiguous.

We now turn to bunching starting in period $\mathcal{T}+1$. It can be shown, similarly to the cases above, that if an agent would be willing to actively bunch in period $\mathcal{T}_1+1$, she will also be willing to actively bunch in earlier periods. Thus, the only active adjustment occurring that affects bunching will be de-bunching. The set of individuals who actively de-bunch, $\mathcal{A}_{\mathcal{T}_1+1}\left(z^*\right)$, are those for whom (A13) is satisfied, when evaluated at $t=\mathcal{T}_1+1$ and $z=z^*$. The remaining changes in bunching between $\mathcal{T}_1$ and $\mathcal{T}_2$ consist of passive adjustment among those who were bunching at the end of period $\mathcal{T}_1$. We can thus characterize B_2^t, bunching at K_2 in period $t \in \left[\mathcal{T}_1+1, \overline{\mathcal{T}}\right]$, in a manner analogous to (9):[20]

$$\begin{aligned}
B_2^t = & \int_{z^*}^{z^*+\triangle z_1^*}\Bigg[\mathbf{1}\left\{a \notin \mathcal{A}_{\mathcal{T}_1+1}\left(z^*\right)\right\} \\
& \times\left\{\pi^{t-\mathcal{T}_1}\mathbf{1}\left\{\tilde{z}_{a,\mathcal{T}_1+1} \neq z^*\right\}+\sum_{j=\mathcal{T}_1+1}^{t}\pi^{t-j}\mathbf{1}\left\{\sup\left\{l \mid l \leq t, \tilde{z}_{a,l}=z^*\right\}=j\right\}\right\} \\
& \times\Bigg\{\mathbf{1}\left\{\tilde{z}_{a,1}=z^*, a \in \mathcal{A}_1(\zeta)\right\} \\
& +\sum_{j=1}^{\mathcal{T}_1}\left(1-\pi^j\right)\pi^{\mathcal{T}_1-j}\mathbf{1}\left\{\sup\left\{l \mid l \leq \mathcal{T}_1, \tilde{z}_{a,l}=z^*\right\}=j, a \notin \mathcal{A}_1(\zeta)\right\} \\
& -\sum_{j=1}^{\mathcal{T}_1-1}\left(1-\pi^{\mathcal{T}_1-j}\right)\mathbf{1}\left\{\sup\left\{l \mid l \leq \mathcal{T}_1, \tilde{z}_{a,l}=z^*\right\}=j, a \in \mathcal{A}_1(\zeta)\right\}\Bigg\}\Bigg] h_0(\zeta)\, d\zeta.
\end{aligned} \tag{A18}$$

The first line of this expression selects only those agents who do not actively de-bunch immediately in period $\mathcal{T}_1+1$. The second line selects the set of agents who would like to passively de-bunch beginning at some period $j>\mathcal{T}_1+1$. They are weighted by the probability of continuing to bunch due to consecutive draws

[20] When $\mathcal{T}_1=1$, we set the very last summation to zero.

of nonzero fixed costs. The final three lines select agents from the set of bunchers at the end of period T_1. As with our simpler model in Section V.C, bunching gradually decreases following a reduction in the size of the kink from K_1 to K_2. However, in this case, the reduction is due to both fixed costs of adjustment and anticipation of the removal of the kink K_2 in period $\mathcal{T}_2 + 1$.

As in Section V.C, the richer model in this appendix nests the dynamic model without forward looking behavior when we set $\delta = 0$, collapses to the comparative static model of Sections V.A-V.B if we additionally assume that $\pi = 1$ and is equivalent to the frictionless model when either $\phi = 0$ or $\pi = 0$.

A3. Derivation of Bunching Formulae with Heterogeneity

Comparative Static Model. — Under heterogenous preferences, our estimates can be interpreted as reflecting average parameters among the set of bunchers (as in Saez, 2010, and Kleven and Waseem, 2013). As described in the main text, suppose $(\varepsilon_i, \phi_i, a_i)$ is jointly distributed according to a smooth CDF, which translates to a smooth, joint distribution of elasticities, fixed costs and earnings. Let the joint density of earnings, adjustment costs and elasticities be $h_0^*(z, \varepsilon, \phi)$ under a linear tax of τ_0. Assume that the density of earnings is constant over the interval $[z^*, z^* + \Delta z^*]$, conditional on ε and ϕ. When moving from no kink to a kink, we derive a formula for bunching at K_1 in the presence of heterogeneity as follows:

$$
\begin{aligned}
B_1 &= \iiint_{\underline{z}_1}^{z^*+\Delta z_1^*} h_0^*(\zeta, \epsilon, \varphi)\, d\zeta d\epsilon d\varphi \\
&= \iint \left[z^* + \Delta z_1^* - \underline{z}_1\right] h_0^*(z^*, \epsilon, \varphi)\, d\epsilon d\varphi \\
&= h_0(z^*) \cdot \iint \left[z^* + \Delta z_1^* - \underline{z}_1\right] \frac{h_0^*(z^*, \epsilon, \varphi)}{h_0(z^*)} d\epsilon d\varphi \\
&= h_0(z^*) \cdot \mathbb{E}\left[z^* + \Delta z_1^* - \underline{z}_1\right], \qquad \text{(A19)}
\end{aligned}
$$

where we have used the assumption of constant $h_0^*(\cdot)$ in line two, $h_0(z^*) = \iint h_0^*(z^*, \epsilon, \varphi)\, d\epsilon d\varphi$, and ζ, ϵ and φ are dummies of integration. The expectation $\mathbb{E}[\cdot]$ is taken over the set of bunchers, under the various combinations of ε and ϕ throughout the support. It follows that normalized bunching can be expressed as follows:

$$b_1 = z^* + \mathbb{E}\left[\Delta z_1^*\right] - \mathbb{E}\left[\underline{z}_1\right]. \qquad \text{(A20)}$$

Under heterogeneity, the level of bunching identifies the average behavioral response, Δz^*, and threshold earnings, $\underline{z}_1$, among the marginal bunchers under each possible combination of parameters ε and ϕ. Under certain parameter values, there is no bunching, and thus, the values of the elasticity and adjustment cost in these cases do not contribute our estimates.

When we move sequentially from a larger kink, K_1 to a smaller kink, K_2, our formula for bunching under K_2 in the presence of heterogeneity is likewise derived as follows:

$$\begin{aligned}\tilde{B}_2 &= \iiint_{\underline{z}_1}^{\bar{z}_0} h_0^*(\zeta,\epsilon,\varphi)\, d\zeta d\epsilon d\varphi \\ &= \iint \left[\bar{z}_0 - \underline{z}_1\right] h_0^*(z^*,\epsilon,\varphi)\, d\epsilon d\varphi \\ &= h_0(z^*) \cdot \iint \left[\bar{z}_0 - \underline{z}_1\right] \frac{h_0^*(z^*,\epsilon,\varphi)}{h_0(z^*)} d\epsilon d\varphi \\ &= h_0(z^*) \cdot \mathbb{E}\left[\bar{z}_0 - \underline{z}_1\right]. \qquad \text{(A21)}\end{aligned}$$

Similarly, normalized bunching can now be expressed as follows:

$$\tilde{b}_2 = \mathbb{E}\left[\bar{z}_0\right] - \mathbb{E}\left[\underline{z}_1\right]. \qquad \text{(A22)}$$

Once again, the expectations are taken over the population of bunchers.

Following the approach in Kleven and Waseem (2013, pg. 682), the average value of the parameters Δz_1^*, $\underline{z}_1$ and $\bar{z}_0$ can then be related to ε and ϕ, assuming a quasi-linear utility function and using (5) and (7) and the identities $\triangle z_1^* = \varepsilon z^* d\tau_1/(1-\tau_0)$ and $\bar{z}_0 - \bar{z}_2 = \varepsilon \bar{z}_2 d\tau_2/(1-\tau_0)$.

Dynamic Model. — A similar interpretation of our results holds when we turn to our more dynamic framework in Section V.C. Suppose now that $(\varepsilon_i, \phi_i, a_i, \boldsymbol{\pi}_i)$ is jointly distributed according to a smooth CDF, which results in a smooth, joint distribution of elasticities, fixed costs, earnings, and probabilities of drawing a positive fixed cost. In order to gain tractability, we assume that the profile $\boldsymbol{\pi}_i$ is independent of the parameters $(\varepsilon_i, \phi_i, a_i)$. The result is that the joint density of these parameters, under a linear tax of τ_0, can be expressed as a product of two densities: $h_0^*(z,\varepsilon,\phi)\, g(\boldsymbol{\pi}_i)$. We maintain the assumption that the density of earnings is constant over the interval $[z^*, z^* + \Delta z^*]$, conditional on ε and ϕ.

Bunching at K_1 in period $t \in [1, \mathcal{T}_1]$ will now be:

$$\begin{aligned}
B_1^t &= \iiiint_{\underline{z}_1}^{z^*+\Delta z_1^*} h_0^*(\zeta,\epsilon,\varphi)\, g(\boldsymbol{\pi})\, d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&\quad + \iiiint_{z^*}^{\underline{z}_1} \left(1-\Pi_{j=1}^t \pi_j\right) h_0^*(\zeta,\epsilon,\varphi)\, g(\boldsymbol{\pi})\, d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&= \iint \left[z^* + \Delta z_1^* - \underline{z}_1\right] h_0^*(z^*,\epsilon,\varphi) \left(\int g(\boldsymbol{\pi})\, d\boldsymbol{\pi}\right) d\epsilon d\varphi \\
&\quad + \iint \left[\underline{z}_1 - z^*\right] h_0^*(z^*,\epsilon,\varphi) \left(\int \left(1-\Pi_{j=1}^t \pi_j\right) g(\boldsymbol{\pi})\, d\boldsymbol{\pi}\right) d\epsilon d\varphi \\
&= h_0(z^*) \left\{ \iint \left[z^* + \Delta z_1^* - \underline{z}_1\right] \frac{h_0^*(z^*,\epsilon,\varphi)}{h_0(z^*)} d\epsilon d\varphi \right. \\
&\qquad \left. + \left(1 - \mathbb{E}\left[\Pi_{j=1}^t \pi_j\right]\right) \iint \left[\underline{z}_1 - z^*\right] \frac{h_0^*(z^*,\epsilon,\varphi)}{h_0(z^*)} d\epsilon d\varphi \right\} \\
&= h_0(z^*) \left\{ z^* + \mathbb{E}\left[\Delta z_1^*\right] - \mathbb{E}\left[\underline{z}_1\right] + \left(1 - \mathbb{E}\left[\Pi_{j=1}^t \pi_j\right]\right) \left(\mathbb{E}\left[\underline{z}_1\right] - z^*\right) \right\} \\
&= h_0(z^*) \left\{ \mathbb{E}\left[\Delta z_1^*\right] - \mathbb{E}\left[\Pi_{j=1}^t \pi_j\right] \left(\mathbb{E}\left[\underline{z}_1\right] - z^*\right) \right\}, \qquad \text{(A23)}
\end{aligned}$$

where now $h_0(z^*) = \iiint h_0^*(z^*,\epsilon,\varphi)\, g(\boldsymbol{\pi})\, d\epsilon d\varphi d\boldsymbol{\pi}$. In the second line, we have again made use of a constant $h_0^*(\cdot)$ and also the independence of $\boldsymbol{\pi}_i$. Normalized bunching at K_1 in period t will then be:

$$b_1^t = \mathbb{E}\left[\Delta z_1^*\right] - \mathbb{E}\left[\Pi_{j=1}^t \pi_j\right] \left(\mathbb{E}\left[\underline{z}_1\right] - z^*\right). \qquad \text{(A24)}$$

Using similar steps, we can show that bunching in period $t > \mathcal{T}_1$ at K_2, when moving sequentially from K_1, can be written as:

$$\begin{aligned}
B_2^t &= \iiiint_{\underline{z}_1}^{z^*+\Delta z_2^*} h_0^*(\zeta,\epsilon,\varphi)\, g(\boldsymbol{\pi})\, d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&\quad + \iiiint_{z^*+\Delta z_2^*}^{\bar{z}_0} \left(\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j\right) h_0^*(\zeta,\epsilon,\varphi)\, g(\boldsymbol{\pi})\, d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&\quad + \iiiint_{z^*}^{\underline{z}_1} \left(1 - \Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j\right) h_0^*(\zeta,\epsilon,\varphi)\, g(\boldsymbol{\pi})\, d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&= h_0(z^*) \left\{ \left(1 - \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j\right]\right) \mathbb{E}\left[\Delta z_2^*\right] + \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j\right] \mathbb{E}\left[\bar{z}_0\right] \right. \\
&\qquad \left. - \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j\right] \mathbb{E}\left[\underline{z}_1\right] - \left(\mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j\right] - \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j\right]\right) z^* \right\}. \qquad \text{(A25)}
\end{aligned}$$

Likewise, normalized bunching at K_2 will be:

$$b_2^t = \left(1 - \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j\right]\right)\mathbb{E}\left[\Delta z_2^*\right] + \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j\right]\mathbb{E}\left[\bar{z}_0\right] - \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1}\pi_j\right]\mathbb{E}\left[\underline{z}_1\right] - \left(\mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j\right] - \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1}\pi_j\right]\right) z^*. \tag{A26}$$

The levels of bunching at the kink before and after the transition are now functions of average behavioral responses, $(\Delta z_1^*, \Delta z_2^*)$, the average thresholds for marginal bunchers, $(\underline{z}_1, \bar{z}_0)$, and average survival probabilities, $\left(\Pi_{j=1}^{t}\pi_j, \Pi_{j=1}^{t-\mathcal{T}_1}\pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1}\pi_j\right)$. Relative to our baseline dynamic model in Section V.C, the number of intermediate parameters to be identified is increasing in the number of post-transition periods, due to the terms of the form $\mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1}\pi_j\right]$. A sufficient condition that allows us to retain identification while only using two transitions in kinks is that the expectation of this product simplifies to a product of expectations: $\mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1}\pi_j\right] = \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j\right]\mathbb{E}\left[\Pi_{j=1}^{\mathcal{T}_1}\pi_j\right]$. There are two cases of interest that satisfy this condition. First, if $\pi_j = 0$ for some $j < \mathcal{T}_1$, then $\Pi_{j=1}^{\mathcal{T}_1}\pi_j = 0$, and the condition holds. This empirically appears to be the case in our context: adjustment takes roughly two years, while $\mathcal{T}_1 \geq 3$ in our two main applications. Second, if there is no heterogeneity in $\boldsymbol{\pi}$ across agents, the condition also holds.

If we relax the assumption that $\mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1}\pi_j\right] = \mathbb{E}\left[\Pi_{j=1}^{t-\mathcal{T}_1}\pi_j\right]\mathbb{E}\left[\Pi_{j=1}^{\mathcal{T}_1}\pi_j\right]$, we will require additional transitions in kinks in order to achieve identification. Furthermore, if we relax the assumption that the profile $\boldsymbol{\pi}_i$ is independent of $(\varepsilon_i, \phi_i, a_i)$, identification is more complicated, as the expectations in the above expressions will then feature weights that vary with t. In that case, more parametric structure on the joint distribution of $(\varepsilon_i, \phi_i, a_i, \boldsymbol{\pi}_i)$ is needed to achieve identification. We discuss identification further in section A.A5 of the Appendix.

A4. Allowing for Frictions in Initial Earnings

In the initial period 0 (prior to the policy change), under a linear tax of τ_0, we have assumed that individuals are located at their frictionless optimum, while we have assumed in subsequent periods adjustment costs may preclude individuals from reaching their exact, interior optimum. Here, we extend the model to allow for agents to be away from their optimum in period 0, in a way that is consistent with our model of a fixed adjustment cost.

We now analyze the thought experiment previously discussed in Section V.B. That is, we demonstrate this extension in the context of the "comparative static" model. From a linear tax of τ_0 in period 0, in period 1 we introduce a kink, K_1, at z^*, and let the marginal tax rate increase to τ_1 for earnings above z^*. Finally, in period 2 we replace the first kink with a second, smaller kink, K_2, at z^*, where the marginal tax rate only increases to τ_2.

Again, agents are indexed by a. Let $z_{a,j}$ be actual earnings for individual a in period when facing tax schedule $T_j(z)$, and let $\tilde{z}_{a,j}$ be the optimal level of earnings she would choose in the absence of adjustment frictions. As in Chetty (2012), assume that earnings are not "too far" from the frictionless optimum; that is, assume that earnings are within a set such that the utility gain of adjusting to the optimum does not exceed the adjustment cost. Formally:

$$
\begin{aligned}
z_{a,j}\left(\tilde{z}_{a,0}\right) \in \left[z^{-}_{a,j}\left(\tilde{z}_{a,0}\right), z^{+}_{a,j}\left(\tilde{z}_{a,0}\right)\right], \\
\text{where } z^{-}_{a,t} \leq \tilde{z}_{a,j} \leq z^{+}_{a,t} \\
\text{and } u\left(\tilde{z}_{a,j} - T_j\left(\tilde{z}_{a,j}\right), \tilde{z}_{a,j}; a\right) - \phi^{*} = u\left(z^{-}_{a,j} - T_j\left(z^{-}_{a,j}\right), z^{-}_{a,j}; a\right) \\
= u\left(z^{+}_{a,j} - T_j\left(z^{+}_{a,j}\right), z^{+}_{a,j}; a\right)
\end{aligned}
\tag{A27}
$$

where $T_j(\cdot)$ represents a linear tax of τ_0 in period 0, reflects the kink K_1 in period 1, and reflects the kink K_2 in period 2. In words, $z^{-}_{a,j}$ and $z^{+}_{a,j}$ are the lowest and highest level of earnings, respectively, that would be acceptable before an individual chooses to adjust to their optimal earnings level. Note that we have defined $z_{a,j}\left(\tilde{z}_{a,0}\right)$ as a function of the optimal level of earnings for individual a in period 0 for notational convenience. Let the actual earnings, conditional on optimal earnings in period 0, be distributed according to the cumulative distribution function $F_{a,j}\left(z_{a,j} \,|\tilde{z}_{a,0}\right)$, with probability density function $f_{a,j}\left(z_{a,j} \,|\tilde{z}_{a,0}\right)$. Thus, individuals are distributed around their frictionless optimum in period 0.

First, consider the level of bunching at K_1. Relative to our baseline model with frictions (that assumes individuals are initially located at their frictionless optimum), there will be two differences in who bunches. First, individuals in Figure 6 Panel B area i did not bunch in the baseline because they were sufficiently close to the kink. These are agents for whom $z^{*} < \tilde{z}_{a,0} < \underline{z}_1$. Now, with some probability, a fraction of these agents will be sufficiently far from z^{*} in period 0 to justify moving to the kink in Period 1—formally, those for whom $z_{a,0} \in \left[z^{+}_{a,1}, z^{+}_{a,0}\right]$. Their initial earnings are above their interior optimum in period 0, but not far enough to outweigh the fixed cost of adjustment in Period 0. Now that the optimum in period 1 has moved to z^{*}, the utility gain to readjusting exceeds the fixed cost of adjustment. These individuals will now bunch under K_1. The second difference in this version of the model relative to our baseline model is that some individuals who had bunched under K_1 in the baseline model, i.e. areas ii, iii,and iv in Figure 6, may find themselves already close enough to z^{*} in period 0 that they do not bunch at z^{*} in period 0 (because relocating to z^{*} in period 0 does not have sufficient benefit to outweigh the fixed adjustment cost). Formally, these are individuals for whom $z_{a,0} < z^{+}_{a,1}$. These cases are illustrated in Appendix Figure B3.

Define bunching under this modified model as B_1'. Bunching under K_1 can be expressed as:

$$\begin{aligned} B_1' &= \int_{z^*}^{z^*+\triangle z_1^*}\left[\int_{z_{n,1}^+}^{z_{n,0}^+} f_{a,0}\left(v\,|\zeta\right)dv\right]h_0\left(\zeta\right)d\zeta \\ &= \int_{z^*}^{z^*+\triangle z_1^*}\left[1-F_{a,0}\left(z_{a,1}^+\,|\zeta\right)\right]h_0\left(\zeta\right)d\zeta \\ &= \int_{z^*}^{z^*+\triangle z_1^*}\Pr\left(z_{a,0}\geq z_{a,1}^+\,|\tilde{z}_{a,0}=\zeta\right)h_0\left(\zeta\right)d\zeta \end{aligned}$$

where ν and ζ are dummies of integration.

We now turn to bunching in period 2, under K_2. Note that because this kink is smaller, anyone sufficiently close to z^* that they did not bunch under K_1 will continue not to bunch under K_2. Thus, the only change in bunching in period 2 will be those who now move away from the kink. Under the baseline model, these were individuals for whom $\bar{z}_0 \leq \tilde{z}_{a,0} \leq z^* + \triangle z_1^*$, i.e. area *iv* in Figure 6, Panel B. These individuals will still find it worthwhile to move away from the kink, but the difference from the baseline model is that only a subset of them bunched in period 1. Thus, the decrease in bunching will be related to the share of people in area v who actually bunched under K_1. What remains are those individuals with $z^* \leq \tilde{z}_{a,0} \leq \bar{z}_0$ who actually bunched in period 1. Formally, bunching in period 2 under K_2 can be expressed as follows:

$$\begin{aligned} \tilde{B}_2' &= \int_{z^*}^{\bar{z}_0}\left[\int_{z_{n,1}^+}^{z_{n,0}^+} f_{a,0}\left(v\,|\zeta\right)dv\right]h_0\left(\zeta\right)d\zeta \\ &= \int_{z^*}^{\bar{z}_0}\left[1-F_{a,0}\left(z_{a,1}^+\,|\zeta\right)\right]h_0\left(\zeta\right)d\zeta \\ &= \int_{z^*}^{\bar{z}_0}\Pr\left(z_{a,0}\geq z_{a,1}^+\,|\tilde{z}_{a,0}=\zeta\right)h_0\left(\zeta\right)d\zeta \end{aligned}$$

We can rewrite the level of bunching in this setting in terms of bunching amounts derived above:

$$\begin{aligned}
B_1' &= \int_{z^*}^{z*+\triangle z_1^*} \Pr\left(z_{a,0} \geq z_{a,1}^+ \left|\tilde{z}_{a,0} = \zeta\right.\right) h(\zeta)\, d\zeta \\
&= \int_{z^*}^{z^*+\triangle z_1^*} h(\zeta)\, d\zeta \cdot \int_{z^*}^{z*+\triangle z_1} \Pr\left(z_{a,0} \geq z_{a,1}^+ \left|\tilde{z}_{a,0} = \zeta\right.\right) \frac{h(\zeta)}{\int_{z^*}^{z*+\triangle z_1} h(\zeta)\, d\zeta} d\zeta \\
&= B_1^* \cdot \int_{z^*}^{z^*+\triangle z_1^*} \Pr\left(z_{a,0} \geq z_{a,1}^+ \left|\tilde{z}_{a,0} = \zeta\right.\right) h\left(\zeta | z^* < \zeta \leq z^* + \triangle z_1^*\right) d\zeta \\
&= B_1^* \cdot \mathbb{E}\left[\Pr\left(z_{a,0} \geq z_{a,1}^+\right) | z^* < \tilde{z}_{a,0} \leq z^* + \triangle z_1^*\right]
\end{aligned}$$

where $B_1^* = \int_{z^*}^{z^*+\triangle z_1^*} h_0(\zeta)\, d\zeta$ is defined in equation (2) when $j = 1$. This is the bunching that would occur in a model of no frictions under K_1, i.e. areas $i - iv$ in Figure 6, Panel B. Likewise, we have:

$$\begin{aligned}
\tilde{B}_2' &= \int_{z^*}^{\bar{z}_0} \Pr\left(z_{a,0} \geq z_{a,1}^+ \left|\tilde{z}_{a,0} = \zeta\right.\right) h_0(\zeta)\, d\zeta \\
&= \left[\tilde{B}_2 + B_1^* - B_1\right] \cdot \mathbb{E}\left[\Pr\left(z_{n,0} \geq z_{n,1}^+\right) | z^* < \tilde{z}_{a,0} \leq \bar{z}_0\right]
\end{aligned}$$

where $\tilde{B}_2$ is defined in equation (6), and B_1 is defined in equation (4). It follows that $\tilde{B}_2 + B_1^* - B_1 = \int_{z^*}^{\bar{z}_0} h_0(\zeta)\, d\zeta$, i.e. areas $i - iii$ in Figure 6.

Without further restrictions on the distribution of optimal earnings under a linear tax, $H_0(z)$, or distribution of earnings about the frictionless optimum in period 0, $F_{a,j}(z_{a,j} | \tilde{z}_{a,0})$, we cannot make further simplifications of these expressions. However, if we assume that the initial actual earnings level is distributed uniformly about optimal earnings in period 0, following Chetty *et al.* (2011) or Kleven and Waseem (2013), then we have:

$$z_{a,0} \sim U\left[z_{a,0}^-, z_{a,0}^+\right]$$

which implies that:

$$\Pr\left(z_{a,0} \geq z_{a,1}^+ \left|\tilde{z}_{a,0} = \zeta\right.\right) = \min\left(\frac{z_{a,0}^+(\zeta) - z_{a,1}^+(\zeta)}{z_{a,0}^+(\zeta) - z_{a,0}^-(\zeta)}, 1\right)$$

Using our definitions above for $z_{a,0}^+(\cdot)$, $z_{a,0}^-(\cdot)$ and $z_{a,1}^+(\cdot)$ we can calculate this probability conditional on initial frictionless earnings in period 0, the elasticity ε and the adjustment cost ϕ. Note that the uniform distribution of actual earnings is not generally centered at the optimal earnings level in period 0, since the lower and upper limits of the support in period 0, i.e. $\left[z_{a,0}^-, z_{a,0}^+\right]$, will tend to be different distances from the frictionless optimum. We can also calculate B_1^*, B_1, and $\tilde{B}_2$, conditional on the counterfactual distribution $H_0(z)$ and a value of ε and ϕ. We are therefore able to calculate predicted values for B_1' and $\tilde{B}_2'$ and use these in a modified version of the estimation procedure outlined in Section V.E.

Although it is not necessary for our estimation procedure, if we further assume that the optimal earnings density, $h_0(\cdot)$, is constant over the range $[z^*, z^* + \triangle z_1^*]$, as is common in the literature (e.g., Chetty *et al.* 2011 or Kleven and Waseem 2013), then we have the following:

$$\begin{aligned} B_1' &= B_1^* \cdot \mathbb{E}\left[\Pr\left(z_{a,0} \geq z_{a,1}^+\right) | z^* < \tilde{z}_{a,0} \leq z^* + \triangle z_1^*\right] \\ &= \triangle z_1^* h_0(z^*) \cdot \mathbb{E}\left[\Pr\left(z_{a,0} \geq z_{a,1}^+\right) | z^* < \tilde{z}_{a,0} \leq z^* + \triangle z_1^*\right] \end{aligned}$$

and likewise:

$$\begin{aligned} \tilde{B}_2' &= \left[\tilde{B}_2 + B_1^* - B_1\right] \cdot \mathbb{E}\left[\Pr\left(z_{n,0} \geq z_{n,1}^+\right) | z^* < \tilde{z}_{a,0} \leq \bar{z}_0\right] \\ &= [\bar{z}_0 - z^*] \, h_0(z^*) \cdot \mathbb{E}\left[\Pr\left(z_{n,0} \geq z_{n,1}^+\right) | z^* < \tilde{z}_{a,0} \leq \bar{z}_0\right] \end{aligned}$$

It also follows that bunching normalized by the height of the density at the kink will be:

$$\begin{aligned} b_1' &= \triangle z_1^* \cdot \mathbb{E}\left[\Pr\left(z_{a,0} \geq z_{a,1}^+\right) | z^* < \tilde{z}_{a,0} \leq z^* + \triangle z_1^*\right] \\ b_2' &= [\bar{z}_0 - z^*] \cdot \mathbb{E}\left[\Pr\left(z_{n,0} \geq z_{n,1}^+\right) | z^* < \tilde{z}_{a,0} \leq \bar{z}_0\right] \end{aligned}$$

A5. Identification

Our estimator is a minimum distance estimator (MDE); Newey and McFadden (1994) give conditions for identification, consistency, and asymptotic normality. An MDE is defined as:

$$\begin{aligned} \hat{\theta} &= \arg\min_{\theta} \hat{Q}(\theta) \\ \hat{Q}(\theta) &= [B - m(\theta)]'\hat{W}[B - m(\theta)] \end{aligned}$$

In our case, B is a vector of L estimated bunching amounts from before and after a policy change, and $m(\theta)$ is a vector of predicted bunching amounts. $\hat{W}$ is a weighting matrix. We consider our comparative static, and dynamic, models, in turn.

Comparative Static Model. — We focus on the exactly identified case with two bunching moments, which is relevant in our empirical application of the comparative static model. We have:

$$\begin{aligned} m(\theta) &= (B_1(\varepsilon,\phi), \tilde{B}_2(\varepsilon,\phi)) \\ B_1 &= \int_{\underline{z}_1}^{z^*+\Delta z_1^*} h(\xi)d\xi \\ \tilde{B}_2 &= \int_{\underline{z}_1}^{\bar{z}_0} h(\xi)d\xi \end{aligned}$$

where B_1 and $\tilde{B}_2$ refer to bunching before and after the policy change, and $\theta \equiv (\varepsilon,\phi)$.

The upper cutoff in B_1 is defined as

$$z^* + \Delta z_1^* = z^* \left(\frac{1-\tau_0}{1-\tau_1}\right)^{\varepsilon}.$$

A necessary condition for identification is that solutions for $\underline{z}_1$ and $\bar{z}_0$ exist; if they do not, then no bunching occurs. It is straightforward to show that a solution for $\underline{z}_1$ exists if

$$z^* \left[(1-\tau_1) - \left(\frac{1-\tau_0}{1-\tau_1}\right)^{\varepsilon} \left((1-\tau_1) - \varepsilon\left(\tau_1-\tau_0\right)\right)\right] > \phi\left(\varepsilon+1\right).$$

This ensures that the "top" buncher wants to adjust to the kink. A solution for $\bar{z}_0$ exists as long as some debunching occurs. It is straightforward to show that this requires that:

$$z^* \left[\frac{(1-\tau_2)^{\varepsilon+1} - (1-\tau_1)^{\varepsilon+1}}{(1-\tau_1)^{\varepsilon}}\right] > \phi\left(\varepsilon+1\right).$$

As long as $\tau_0 < \tau_2 < \tau_1$, $\varepsilon > 0$, and $\phi > 0$, there exists a range of values of ε and ϕ for which these inequalities hold.

Provided that $\bar{z}_0$ and $\underline{z}_1$ exist, identification requires that $m(\theta) = B$ has a unique solution. Following previous literature (*e.g.* Kline and Walters 2016), we establish local uniqueness by linearizing $m(\cdot)$ around a solution $m(\theta_0) = B$. Let θ_0 be a solution to $m(\theta) = B$. Linearizing $m(\cdot)$ around θ_0, we have:

$$m(\theta) \approx m(\theta_0) + \nabla m(\theta_0)(\theta - \theta_0).$$

It follows that a unique solution requires $\mathbf{J_m}(\theta_0)$ to have full rank, where $\mathbf{J_m}(\theta_0)$ is the Jacobian of $m(\cdot)$ evaluated at θ_0:

$$\mathbf{J_m}(\theta_0) = \begin{bmatrix} \frac{\partial B_1}{\partial \varepsilon} & \frac{\partial B_1}{\partial \phi} \\ \frac{\partial \tilde{B}_2}{\partial \varepsilon} & \frac{\partial \tilde{B}_2}{\partial \phi} \end{bmatrix}.$$

We calculate the elements of this matrix analytically by differentiating the expressions above for B_1 and $\tilde{B}_2$, which is straightforward.[21] Thus, given $\hat{\theta}$, $\underline{z}_1$, and $\bar{z}_0$, we can calculate the Jacobian analytically (although $\underline{z}_1$ and $\bar{z}_0$ must be found numerically).

$\mathbf{J_m}$ has full rank only if it has a non-zero determinant. We find in all of our bootstrap iterations that $\det(\mathbf{J_m}) < 0$, demonstrating that the determinant is significantly different from zero. We have also shown analytically that the determinant is generically non-zero (results available upon request).

Dynamic Model. — To identify the dynamic model, we need to observe at least as many moments as the number of parameters we seek to estimate. In our case this means that we must observe bunching across multiple policy changes, specifically the reductions in the benefit reduction rate above the exempt amount in 1990 and at age 70. Let l index different such policy changes (in our case, $l \in \{1990, 70\}$). Let $B^t_{1,l}$ be bunching at kink l and period t *before* the policy change, let $B^t_{2,l}$ be bunching at kink l and period t *after* the policy change, let time t measure the time since the introduction of the first kink, $K_{1,l}$, and let the policy change at kink l take place at time $\mathcal{T}_{1,l}$. The parameter vector θ now consists of $(\varepsilon, \phi, \pi_1, \pi_2, ..., \pi_5)$. We match 12 bunching amounts in our estimates: 1987 to 1992 (pooling 66 to 68 year olds) and ages 67 to 72 (pooling years 1990 to 1999).

Bunching before the policy change is

$$B^t_{1,l} = \Pi^t_{j=1}\pi_j \cdot B_{1,l} + (1 - \Pi^t_{j=1}\pi_j)B^*_{1,l}$$

where $B_{1,l} = \int_{\underline{z}_{1,l}}^{z^*_l+\Delta z^*_{1,l}} h(\xi)d\xi$ and $B^*_{1,l} = \int_{z^*_l}^{z^*_l+\Delta z^*_{1,l}} h(\xi)d\xi$, and the limits of integration are defined similarly to the static case (but with the additional subscript l to allow for analysis across multiple policy changes, as in our empirical application of the dynamic model). If the policy change happens $\mathcal{T}_{1,l}$ periods after the kink is initially introduced, then bunching under the new policy in period t is

$$B^t_{2,l} = \Pi^{t-\mathcal{T}_{1,l}}_{j=1}\pi_j \cdot \tilde{B}_{2,l} + \left(1 - \Pi^{t-\mathcal{T}_{1,l}}_{j=1}\pi_j\right) B^*_{2,l} + \Pi^{t-\mathcal{T}_{1,l}}_{j=1}\pi_j \left(1 - \Pi^{\mathcal{T}_{1,l}}_{j=1}\pi_j\right) \left(B^*_{1,l} - B_{1,l}\right)$$

where $\tilde{B}_{2,l} = \int_{\underline{z}_{1,l}}^{\bar{z}_{0,l}} h(\xi)d\xi$, $B^*_{2,l} = \int_{z^*_l}^{z^*_l+\Delta z^*_{2,l}} h(\xi)d\xi$, and the limits of integration again are defined similarly to the static case but with the additional subscript l.

[21] We can specify functions implicitly defining the lower and upper cutoffs $\underline{z}_1$ and $\bar{z}_0$, respectively, as functions of the other parameters, given our quasilinear and isoelastic case. These enter the expressions for each element of the Jacobian (more details are available upon request).

We calculate the elements of the resulting Jacobian analytically by differentiating the expressions above for $B^t_{1,l}$ and $B^t_{2,l}$ with respect to ε, ϕ, π_1, π_2, π_3, π_4, and π_5, which is again straightforward. Thus, given $\hat{\theta}$, $\underline{z}_{1,l}$ and $\bar{z}_{0,l}$, we can again calculate the Jacobian analytically.

Identification requires that this Jacobian have full rank. To test for full rank of the Jacobian, we use the method of Kleibergen and Papp (2006). We use the bootstrap to obtain an estimate of $Var[\mathbf{J_m}(\hat{\theta})]$. In each iteration of our bootstrap, we also calculate $\mathbf{J_m}(\hat{\theta})$, and we estimate $Var[\mathbf{J_m}(\hat{\theta})]$ from the bootstrap variance-covariance matrix. The RK test easily rejects under-identification, with $p < 0.001$.

A6. Econometric Estimation

We begin by describing our econometric estimation procedure under our basic comparative static model of Sections V.A and V.B. Let $B = (B_1, B_2, \ldots, B_L)$ be a vector of (estimated) bunching amounts, using the method described in Section II. Let $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_L)$ be the tax schedule at each kink. The triplet $\boldsymbol{\tau}_l = (\tau_{0,l}, \tau_{1,l}, \tau_{2,l})$ denotes the tax rate below the kink $(\tau_{0,l})$, above the kink $(\tau_{1,l})$, and the *ex post* marginal tax rate above the kink after it has been reduced $(\tau_{2,l})$, as in Section V.B. Let $\mathbf{z}^* = (z^*_1, \ldots, z^*_L)$ be the earnings levels associated with each kink. In principle, it would be possible to estimate bunching separately for each age group at a given kink. In practice and for simplicity, we pool across a constant set of ages to estimate bunching at a given kink—for example, when examining the 1990 policy change we examine 66-68 year-olds both before and after the change. Thus, the bunching amounts are not indexed by age.[22]

n our baseline, we use a non-parametric density for the counterfactual earnings distribution, H_0. Once H_0 is known, we use (4) and (6) to obtain predicted bunching from the model. To recover H_0 non-parametrically we take the empirical earnings distribution for 72 year-olds in \$800 bins as the counterfactual distribution. 72 year-olds' earnings density represents a reasonable counterfactual because they no longer face the Earnings Test, no longer show bunching, and are close in age to those aged 70 or 71. Letting z_i index the bins, our estimate of the distribution is $\hat{H}_0(z_i) = \sum_{j \leq i} Pr(z \in z_j)$. This function is only defined at the midpoints of the bins, so we use linear interpolation for other values of z. In a robustness check, we instead assume that the earnings distribution over the range $[z^*, z^* + \Delta z]$ is uniform, a common assumption in the literature (*e.g.* Chetty *et al.*, 2011, Kleven and Waseem, 2013). Using the nonparametrically-estimated distribution of earnings from age 72 is helpful because it does not entail distributional assumptions, but relative to assuming a uniform distribution, using the age-72 distribution comes at the cost of using a different age (*i.e.* 72) to generate

[22] Analogously, when we examine bunching at each age around 70 when the AET is eliminated, we pool across calendar years (namely 1990-1999) to estimate bunching, so that we do not also have to index the bunching amounts by calendar year. We find comparable results when we estimate bunching separately at each age and year.

the earnings distribution.[23]

To estimate (ε, ϕ), we seek the values of the parameters that make predicted bunching $\hat{B}$ and actual (estimated) bunching B as close as possible on average. Letting $\hat{B}(\varepsilon, \phi) \equiv (\hat{B}(\boldsymbol{\tau}_1, z_1^*, \varepsilon, \phi), \ldots, \hat{B}(\boldsymbol{\tau}_L, z_L^*, \varepsilon, \phi))$, our estimator is:

$$\text{(A28)} \qquad \left(\hat{\varepsilon}, \hat{\phi}\right) = \operatorname{argmin}_{(\varepsilon,\phi)} \left(\hat{B}(\varepsilon, \phi) - B\right)' W \left(\hat{B}(\varepsilon, \phi) - B\right),$$

where W is a $K \times K$ identity matrix. This estimation procedure runs parallel to our theoretical model, as the bunching amounts $\hat{B}$ are those predicted by the theory (and the estimated counterparts B are found using the procedure outlined in Section II).[24] When we pool data across multiple time periods, we assume that ε and ϕ are constant across these time periods.

We obtain our estimates by minimizing (A28) numerically. Solving this problem requires evaluating $\hat{B}$ at each trial guess of (ε, ϕ).[25] Our estimator assumes a quasilinear utility function, $u(c, z; a) = c - \frac{a}{1+1/\varepsilon}\left(\frac{z}{a}\right)^{1+1/\varepsilon}$, following Saez (2010), Chetty *et al.* (2011) and Kleven and Waseem (2013). Note that because we have assumed quasilinearity, $\Delta z_{1,l} = z_l^*\left(\left(\frac{1-\tau_{1,l}}{1-\tau_{0,l}}\right)^{\varepsilon} - 1\right)$ and $a = z(\tau) / (1-\tau)^{\varepsilon}$, where $z(\tau)$ are the optimal, interior earnings under a linear tax of τ. Typically there is no closed form solution for $\underline{z}_{1,l}$ or $\bar{z}_{0,l}$. Instead, given ε and ϕ, we find $\underline{z}_{1,l}$ and $\bar{z}_{0,l}$ numerically as the solution to the relevant indifference conditions in (5) and (7). For example, $\underline{z}_{1,l}$ is defined implicitly by:

$$\text{(A29)} \quad \underbrace{u((1-\tau_{1,l})z_l^* + R_{1,l}, z_l^*; \underline{z}_{1,l}/(1-\tau_{0,l})^{\varepsilon})}_{\text{utility from adjusting to kink}} - \underbrace{u((1-\tau_{1,l})\underline{z}_{1,l} + R_{1,l}, \underline{z}_{1,l}; \underline{z}_{1,l}/(1-\tau_{0,l})^{\varepsilon})}_{\text{utility from not adjusting}} = \phi,$$

This equation is continuously differentiable and has a unique solution for $\underline{z}_{1,l}$.[26]

Dynamic Model. — Our estimation method is easily amended to accommodate the dynamic extension of our model in Section V.C. As in (8) and (9), the bunching expressions in the dynamic model are weighted sums of B_1 and $\tilde{B}_2$,

[23]Because we use the age-72 density as our counterfactual density – unlike most bunching papers bunching that estimate the counterfactual from the same density that is used to estimate bunching – our method is not subject to the Blomquist and Newey (2017) point that the functional form of preference heterogeneity cannot be simultaneously estimated with the taxable income elasticity.

[24]Without loss of generality, we use normalized bunching, $\hat{b} = \delta\hat{B}/h_0(z^*)$, so that the moments are identical to what is reported elsewhere in the text.

[25]In solving (A28), we impose that $\phi \geq 0$. When $\phi < 0$, every individual adjusts her earnings by at least some arbitrarily small amount, regardless of the size of ϕ. This implies that ϕ is not identified if it is less than zero. Inattention or the difficulty of negotiating new contracts should be associated with positive adjustment costs (which could distinguish this context from the firm context studied in Garicano *et al.*, 2016).

[26]Note that some combinations of $\boldsymbol{\tau}_l$, z_l^*, ε, and ϕ imply $\underline{z}_{1,l} > z_l^* + \triangle z_{1,l}$. In this case, the lowest-earning adjuster does not adjust to the kink. Whenever this happens, we set $\hat{B}_l = 0$.

which are calculated as in Section V.E, and two measures of frictionless bunching, B_1^* and B_2^*. Frictionless bunching under either kink can be calculated conditional on H_0 and ε using (2).

We must also estimate the probability of drawing a positive fixed cost as a function of the time since the last policy shock, π_{t-t^*}.[27] For given values of ε, ϕ, and the vector $\boldsymbol{\pi}$ of π_{t-t^*}'s, we can evaluate (8) and (9). Our vector of predicted bunching, $\hat{B}$, will now be a function of these additional parameters, as well as the relevant time indices: $\hat{B}(\varepsilon, \phi, \boldsymbol{\pi}) \equiv (\hat{B}(\boldsymbol{\tau}_1, z_1^*, t_1, \mathcal{T}_{1,l}, \varepsilon, \phi, \boldsymbol{\pi}), \ldots, \hat{B}(\boldsymbol{\tau}_L, z_L^*, t_L, \mathcal{T}_{1,L}, \varepsilon, \phi, \boldsymbol{\pi}))$, where t_l is the time elapsed since the first kink, $K_{1,l}$, was introduced, and $\mathcal{T}_{1,l}$ is the length of time before the second kink, $K_{2,l}$, is introduced. Once again we use the minimum distance estimator (A28).

Equations (8) and (9) illustrate how we estimate the elasticity and adjustment cost in this richer setting. We require as many observations of bunching as the parameters, $(\varepsilon, \phi, \pi_1, ..., \pi_J)$, and these moments must span a change in $d\tau$.[28] Suppose we observe the pattern of bunching over time around two or more different policy changes. Loosely speaking, the $\boldsymbol{\pi}$'s are estimated relative to one another from the time pattern of bunching over time: a delay in adjustment in a given period will generally correspond to a higher probability of facing the adjustment cost (all else equal). Note that the relationship is linear; the degree of "inertia" in bunching in (for example) period 1 increases linearly in π_1. Meanwhile, a higher ϕ implies a larger amount of inertia in *all* periods until bunching has fully dissipated (in a way that depends on the earnings distribution, the elasticity, and the size of the tax change). Finally, a higher ε will correspond to a larger amount of bunching once bunching has had time to adjust fully to the policy changes. Intuitively, these features of the data help us to identify the parameters using our dynamic model.

A7. Policy Simulations

In this Appendix, we describe how we simulate the effect of various policy changes on earnings. These calculations are designed to be illustrative of the attenuation of earnings responses to policy changes that can result from incorporating adjustment frictions in the analysis. Nonetheless, we highlight that these calculations are done in the context of a highly stylized model making a number of assumptions, as well as a particular sample of earners. One key (extreme) assumption is that everyone has the same elasticity and adjustment cost. Moreover, these estimates are specific to a particular context, and they are not intended to be an exhaustive account of the implications of adjustment costs for earnings responses to taxation. Rather, they are intended simply to illustrate the attenuation of earnings responses to policy changes that can result from incorporating

[27] We have also tried using a flexible, logistic functional form, $\pi_j = \exp(\alpha + \beta \cdot j) / (1 + \exp(\alpha + \beta \cdot j))$, and we found comparable results (available upon request).

[28] The number of moments is not itself sufficient. We also require non-trivial variation in bunching before and after the tax change in order to point identify ϕ. As in footnote 8, this requires $\bar{z}_0 < z^* + \triangle z_1^*$.

adjustment frictions in the analysis in such contexts.

We assume that utility is isoelastic and quasi-linear with elasticity ε. Individuals must pay an adjustment cost ϕ to change their earnings. Individuals are heterogeneous in their ability n_i. Individuals are therefore distributed according to their "counterfactual" earnings z_{0i} that they would have under a linear tax schedule. (Despite the absence of heterogeneity in the elasticity and adjustment cost, there is still heterogeneity in the gains from re-optimizing earnings, due to heterogeneity in z_{0i}.) We use the 1989 earnings distribution for 60-61 year-olds (from the MEF data) as the counterfactual earnings distribution, *i.e.* the earnings distribution under a linear tax schedule in the region of the exempt amount. We incorporate the key features of the individual income tax code, including individual federal income taxes, state income taxes, and FICA (all from Taxsim applied in 1989), and the Earnings Test. Our estimates of elasticities and adjustment costs apply to a population earning near the exempt amount; to avoid extrapolating too far out of sample, our simulations examine only those whose counterfactual earnings is from \$10,000 under to \$10,000 over the exempt amount (and is greater than \$0). (While the Earnings Test should only affect people whose counterfactual earnings are over the exempt amount, we also include the group earning up to \$10,000 under the exempt amount in order to illustrate the fact that some individuals could be unaffected by a policy change.)

We consider two periods, 1 and 2. In period 1, in the region of the Earnings Test exempt amount, the mean tax rate below the exempt amount is 27.21 percent, and the mean tax rate above the exempt amount is 77.21 percent. Note that these tax rates mimic those faced by 62-64 year-old Social Security claimants.[29] In period 2, the tax rate below the exempt amount remains 27.21 percent, but the tax rate above the exempt amount changes according to the policy changes we specify below. (We assume that in the counterfactual individuals face a linear schedule with a mean tax rate of 27.21 percent.)

For a given counterfactual earnings level z_{0i}, we calculate optimal frictionless earnings z^*_{1i} in period 1, and we calculate whether the individual with counterfactual earnings z_{0i} wishes to adjust her earnings from the frictionless optimum because the gains from doing so outweigh the adjustment cost. (Optimal "frictionless" earnings refers to the individual's optimal earnings in the absence of adjustment costs.) We then determine the individual's optimal frictionless earnings z^*_{2i} under the new tax schedule in period 2. We assess whether given the adjustment cost, the individual obtains higher utility by staying at her period 1 earnings level, or by paying the adjustment cost and moving to a new earnings level in period 2.

We perform these calculations alternatively under the assumptions that (a) the elasticity ε is 0.35 and the adjustment cost ϕ is \$280 (our baseline estimates);

[29] As we note elsewhere, 62-64 year-olds technically face a notch in the budget constraint at the exempt amount, as opposed to a kink. However, we find no evidence that they behave as if they faced a notch, as the earnings distribution for this age group 1) does not show bunching just above the exempt amount and 2) does not show a "hole" in the earnings distribution just under the exempt amount.

or (b) the elasticity ε is 0.35 and the adjustment cost ϕ is zero. Thus, our simulations illustrate the difference between incorporating adjustment costs and not incorporating them, holding the elasticity constant.

Under these alternative assumptions, we can perform a number of experiments to simulate the effects of changing the effective tax schedule. These calculations are shown in Appendix Table B6 below.

We calculate that if the marginal tax rate above the exempt amount were reduced by 17.22 percentage points, so that the tax rate above the exempt amount were reduced from 77.21 percent to 59.99 percent, mean earnings in the population under consideration would be unchanged at $9,371.9 under our baseline estimates of the elasticity and adjustment cost. In this case, adjustment is not optimal for anyone when we assume the adjustment cost. In fact, earnings would be unchanged for any reduction in the marginal tax rate above the exempt amount up to 17.22 percentage points; 17.22 percentage points is the largest percentage point marginal tax rate decrease above the exempt amount for which there is no adjustment. Since the gains are second-order near the kink, even a modest adjustment cost of $280 prevents adjustment with an 17.22 percentage point (or smaller) cut in marginal tax rates. By contrast, when assuming ε=0.35 and ϕ=0, we predict that mean earnings would rise from $9,340.3 to $10,166.3, an increase of 8.84 percent.

At the same time we calculate that if the 50 percent Earnings Test above the exempt amount were eliminated, so that the tax rate above the exempt amount were reduced from 77.21 percent to 27.21 percent, mean earnings in the population under consideration would rise from $9,371.9 to $11,566.7.7, or 23.4 percent, under our baseline estimates of the elasticity and adjustment cost. When assuming ε=0.35 and ϕ=0, we predict that mean earnings would rise from $9,340.3 to $11,639.2, a nearly identical increase of 24.6 percent. The slight discrepancy between the two estimates arises because there are individuals whose counterfactual earnings is just above the exempt amount who choose to adjust without adjustment costs, but for whom the gains from adjustment do not outweigh the adjustment cost when we assume the friction.

It is worth noting an additional caveat to these results: they apply to those with counterfactual earnings in the range from $10,000 below to $10,000 above the exempt amount. If we allowed unbounded counterfactual earnings, there would be some individuals with very large counterfactual earnings for whom the gains from adjustment would outweigh the adjustment cost, even in the presence of adjustment costs. However, this is less relevant to the Earnings Test because as we have noted, the Social Security benefit phases out entirely at very high earnings levels. Moreover, considering such individuals would involve extrapolating the estimates much farther out of sample. Finally, the results are qualitatively robust to considering other earnings ranges within the range we measure in our study, such as the range of individuals earning from $10,000 below to $30,000 above the exempt amount. In fact, under all of the other choices we have explored, the results always show that the maximum tax cut that leads to no earnings change is quite substantial (and larger than the changes in marginal tax rates envisioned in most tax reform proposals)—including when we use other ages to specify the counterfactual earnings density; use a different baseline marginal tax rate; and use the constrained estimate of the elasticity (0.58) when performing the simulations (which actually leads to still starker results).

All of these simulations use the static model. If we were to use our estimates of the dynamic model instead to perform these simulations, we would still find that the immediate reaction even to large taxes changes is greatly attenuated, since the estimates of the dynamic model still show that most individuals are constrained from adjusting immediately.

Additional Empirical Results

Table B1—: Robustness of normalized bunching to alternative birth month restrictions

	b_{68}	b_{69}	b_{70}	b_{71}	b_{72}
A) Born January-March	3545.4	4036.2	565.0	881.6	-236.8
	[2750.8, 4340.0]***	[2712.6, 5359.8]***	[-108.5, 1238.5]	[-67.3, 1830.6]	[-847.6, 374.0]
B) Born any month	3992.2	3552.3	1203.9	941.4	-231.4
	[3360.4, 4624.0]***	[3111.4, 3993.2]***	[895.2, 1512.6]***	[503.2, 1379.7]***	[-548.6, 85.8]

Notes: The table shows excess normalized bunching and its confidence interval at each age from 68 to 72 for two samples: those born January to March (Row A), and those born in any month (Row B). The data are pooled over the period from 1983-1999. The table shows that we continue to estimate significant bunching at age 70 (and in some cases 71) when the sample is restricted to those born in January to March. Limiting the sample only to those born in January yields insignificant results, with little statistical power. *** indicates p<0.01; **p<0.05; * p<0.10.

Table B2—: Heterogeneity in Estimates of Elasticity and Adjustment Cost across Samples

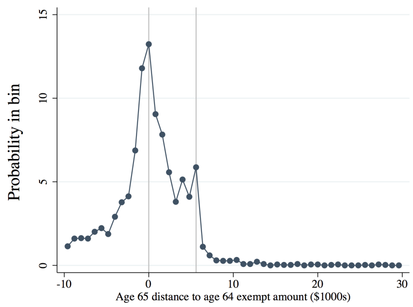
	(1)	(2)	(3)	(4)
	ε	p-value for ε equality	ϕ	p-value for ϕ equality
Men	0.44	0.39	\$62	0.00
	[0.38, 0.52]***		[14, 167]***	
Women	0.42		\$489	
	[0.32, 0.50]***		[165, 720]***	
High lifetime earnings	0.48	0.05	\$24	0.00
	[0.41, 0.58]***		[2, 90]***	
Low lifetime earnings	0.44		\$538	
	[0.32, 0.51]***		[217, 688]***	
High lifetime earnings variability	0.39	0.25	\$116	0.16
	[0.35, 0.46]***		[37, 315]***	
Low lifetime earnings variability	0.38		\$178	
	[0.33, 0.46]***		[55, 378]***	

Notes: This table implements our "comparative static" method separately in each of several groups shown in each row. "High/low lifetime earnings" refers to the group of individuals with mean real earnings from 1951 (when the data begin) to 1989 that are above/below the median level in our study population. "High/low lifetime earnings variability" refers to the group of individuals for whom the standard deviation of real earnings from 1951 to 1989 is above/below the median level in our study population. Columns 2 and 4 show the p-values for the two-sided test of equality in the estimates between each set of groups (i.e. men vs. women, high vs. low lifetime earnings, and high vs. low earnings variability), for ε and ϕ, respectively. We pool data from two policy changes: (a) around the 1989/1990 transition analyzed in Table 2, and (b) around the age 69/70 transition analyzed in Table 3. We pool the transitions because this gives us the maximum power to detect differences across groups. The results are generally comparable when we investigate each transition separately. See also notes from Tables 2 and 3.

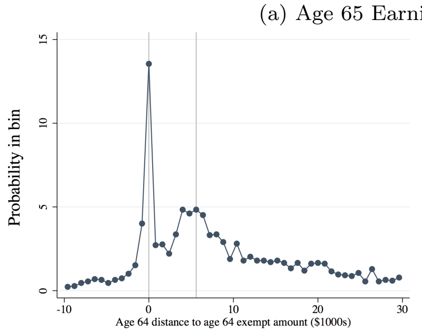
Table B3—: Robustness to alternative empirical choices

Binsize	Degree	Excluded Bins	b_{68}	b_{69}	b_{70}	b_{71}	b_{72}
			Panel A: Baseline				
$800	7	4	3442.3	2868.4	657.9	1068.8	70.1
			[2767.3, 4117.3]***	[2313.5, 3423.3]***	[172.1, 1143.8]***	[547.9, 1589.6]***	[-341.7, 481.8]
			Panel B: Robustness to bin size				
$400	7	8	3107.8	2606.3	462.2	923.0	-55.3
			[2598.8, 3616.8]***	[2047.7, 3164.9]***	[65.9, 858.5]**	[556.2, 1289.7]***	[-382.2, 271.5]
$1,600	7	2	3047.0	2941.0	601.4	1210.9	241.2
			[2336.8, 3757.2]***	[2422.8, 3459.2]***	[-59.4, 1262.2]	[668.6, 1753.2]***	[-378.6, 861.1]
			Panel C: Robustness to degree				
$800	6	4	3677.2	3267.3	1310.6	993.7	224.6
			[3129.2, 4225.2]***	[2758.1, 3776.5]***	[736.9, 1884.3]***	[529.6, 1457.8]***	[-168.0, 617.2]
$800	8	4	3535.1	2948.0	710.8	1084.3	82.4
			[2850.7, 4219.5]***	[2547.8, 3348.2]***	[317.9, 1103.7]***	[534.2, 1634.5]***	[-311.1, 475.9]
			Panel D: Robustness to excluded region				
$800	7	3	2170.2	2182.4	202.2	191.2	-55.0
			[1619.3, 2721.2]***	[1694.3, 2670.5]***	[-120.3, 524.8]	[-322.8, 705.1]	[-446.3, 336.3]
$800	7	5	3610.6	2651.3	298.1	1103.3	579.9
			[2742.3, 4478.9]***	[2084.2, 3218.3]***	[-400.7, 996.8]	[266.3, 1940.3]***	[-152.9, 1312.7]

Notes: The table shows the estimated bunching amount at each age from 68 to 72, varying the bin size, degree of the polynomial of the smooth density, or number of excluded bins around the exempt amount. Note that varying the bin size but fixing the number of excluded bins automatically changes the width of the excluded region, so to (approximately) fix the width of the excluded region when changing the bin size, we also change the number of excluded bins. *** indicates p<0.01; **p<0.05; * p<0.10.



(a) Age 65 Earnings Distribution | Near Kink at 64



(b) Age 64 Earnings Distribution | Near Kink at 65

Figure B1. : Inertia in Bunching from 64 to 65

Notes: Using data from 1990 to 1999, Panel A of the figure shows that when they are age 65, those previously bunching at age 64 tend to either (a) remain near the age 64 exempt amount or (b) move to the age 65 exempt amount. Panel B of the figure shows that those bunching at age 65 were usually bunching at age 64 in the previous year, or were near the age 65 exempt amount in the previous year. Having earnings "near the kink" at a given age is defined as having earnings within $1,000 of the kink at that age. The first vertical line at zero shows the age 64 exempt amount, and the second vertical line shows the average location of the age 65 exempt amount.

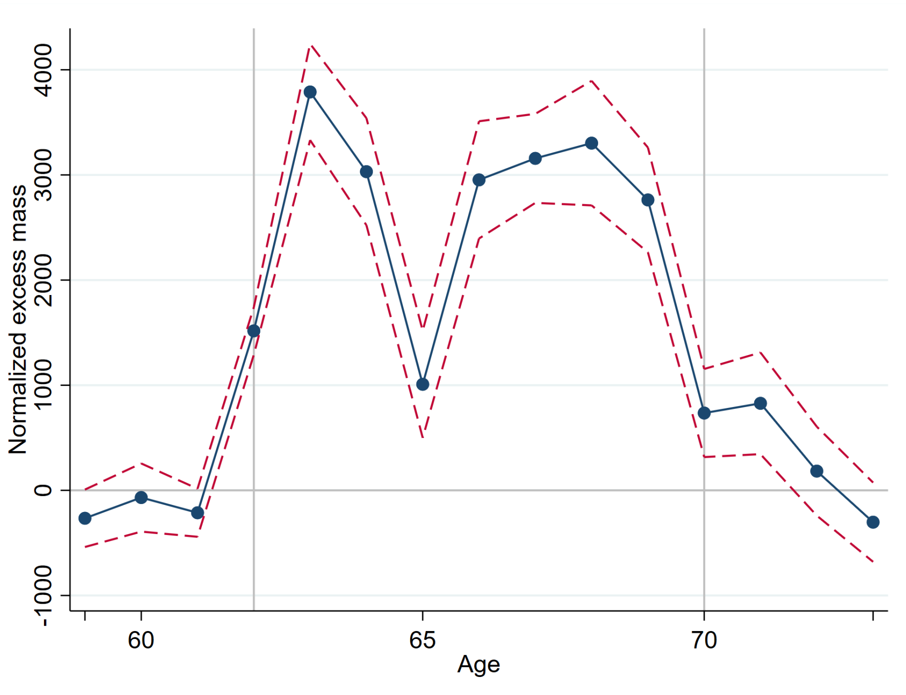


Figure B2. : Normalized Excess Mass of Claimants, Ages 59 to 73, 1990 to 1999

Notes: See notes to Figure 2 Panel B. This figure differs from Figure 2 Panel B because here the sample in year t consists only of people who have claimed Social Security in year t or before (whereas in Figure 2 Panel B it consists of those who claimed by age 65).

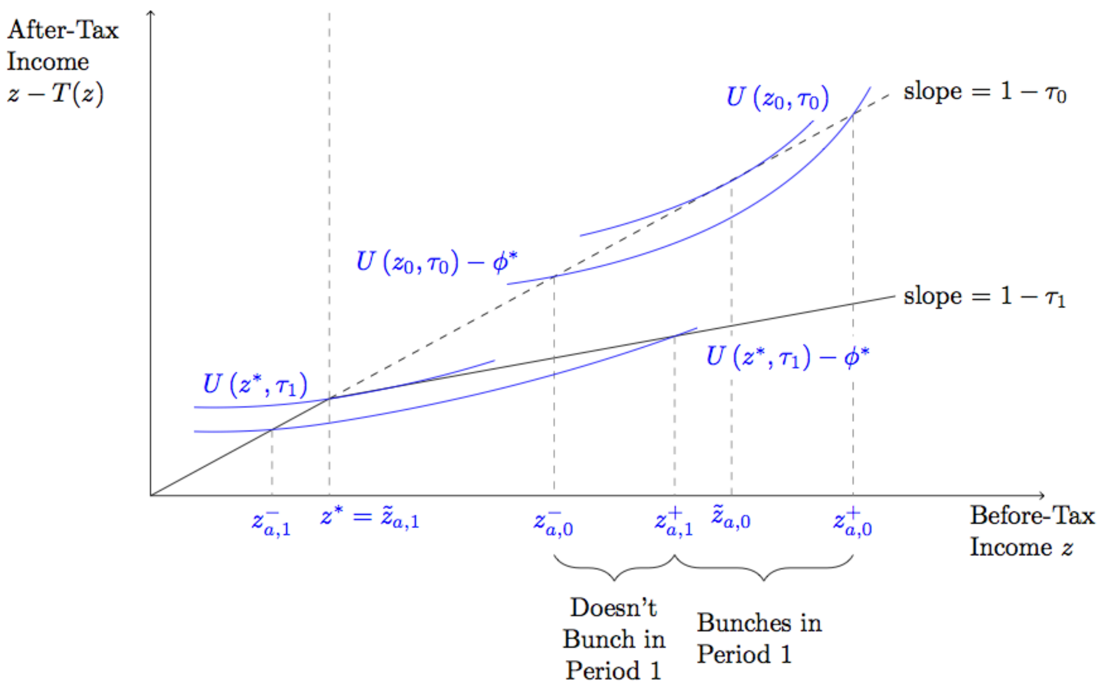


Figure B3. : Bunching Response to a Convex Kink, with Frictions in Initial Earnings

Notes: See Section A.A4 for an explanation of the figure.

Table B4—: Estimates of Elasticity and Adjustment Cost 1990 Policy Change, Assuming Pre-Period Bunching may not be at Frictionless Optimum

	(1)	(2)	(3)	(4)
	ε	ϕ	$\varepsilon\|\phi = 0$	
			1990	1989
Baseline	0.28	$193	0.43	0.24
	[0.25, 0.32]***	[56, 299]***	[0.36, 0.53]***	[0.20, 0.28]***
Uniform Density	0.24	$163	0.39	0.22
	[0.21, 0.28]***	[54, 268]***	[0.33, 0.48]***	[0.18, 0.25]***
Benefit Enhancement	0.47	$103	0.66	0.41
	[0.41, 0.54]***	[21, 172]***	[0.54, 0.80]***	[0.33, 0.48]***
Excluding FICA	0.39	$165	0.56	0.34
	[0.34, 0.45]***	[41, 270]***	[0.46, 0.68]***	[0.27, 0.39]***
Bandwidth = $400	0.37	$123	0.53	0.33
	[0.31, 0.46]***	[6, 383]**	[0.42, 0.70]***	[0.27, 0.42]***

Notes: The table examines the 1990 policy change, using data from 1989 and 1990, but assumes that bunching in 1989 may not be at the frictionless optimum, as described in the text. See also notes to Table 2.

Table B5—: Estimates of Changes in Bunching Around 1990

Sample	Old only	Old only, linear trend	DD	DD, separate linear trend
old x 1990 dummy	28.9	-165.1	-107.3	-69.2
	(249.1)	(411.0)	(306.7)	(411.7)
old x 1991 dummy	-1728.9	-1966.0	-1824.5	-1777.9
	(249.1)***	(500.6)***	(306.7)***	(481.3)***
old x 1992 dummy	-1648.8	-1928.9	-1130.2	-1075.1
	(249.1)***	(594.9)***	(306.7)***	(558.1)*
old x 1993 dummy	-2123.8	-2447.1	-2131.2	-2067.6
	(249.1)***	(692.1)***	(306.7)***	(639.7)***
Ages	66-68	66-68	62-64, 66-68	62-64, 66-68
Year FE?	No	No	Yes	Yes
Linear time trend (in year)	No	Yes	No	No
Separate linear trend for "old"	No	No	No	Yes

Notes: The table shows that the estimated change in bunching amounts from before to after 1990 in the age 66-68 age group are similar under several specifications. The dummy variable "old" indicates the older age group (66-68). The sample in Columns (1) and (2) includes only 66-68 year-olds, and in Columns (3) and (4) it also includes 62-64 year-olds. Additional controls include a linear time trend (in year) in column (2), year fixed effects in columns (3) and (4), and the linear time trend interacted with the "old" dummy in column (4). Robust standard errors are in parentheses. Under all the specifications, the coefficient on old x 1990 is insignificantly different from zero: bunching in 1990 is not significantly different from prior bunching, indicating that adjustment does not immediately occur. However, the coefficients on old x 1991, old x 1992, old x 1993 are negative and significant, indicating that bunching falls significantly after 1990—i.e. a reduction in bunching does eventually occur (but not immediately in 1990). The fact that the results are similar under all these various specifications indicates that the results are little changed by controlling for a linear trend (Column 2), comparing 66-68 year-olds to a reasonable control group of 62-64 year-olds (Column 3), and additionally controlling for a separate linear trend for the older group (Column 4). In Columns 1 and 3, the standard errors are the same across all of the interaction coefficients shown because there is only one observation underlying each dummy, and the dummies are exactly identified. See also notes from Table 2.

Table B6—: Policy Simulations

	(1)	(2)
	Panel A: Eliminate Earnings Test for 62-64 year olds	
	With adjustment costs	Without adjustment costs
Period 1 mean earnings	\$9,371.9	\$9,340.1
Mean earnings change	\$2,194.8	\$2,298.9
Share affected	50.4	50.4
Share who adjust	41.9	50.4
Mean change among adjusters	\$5,239.6	\$4,563.7
Percent change among adjusters	42.6	37.3
	Panel B: Reduce Earnings Test BRR by 17.22 percentage points	
	With adjustment costs	Without adjustment costs
Period 1 mean earnings	\$9,371.9	\$9,340.3
Mean earnings change	\$0	\$826.0
Percent earnings change	50.4	50.4
Share who adjust	0.0	37.4
Mean change among adjusters	0.0	\$2,207.6
Percent change among adjusters	0.0	17.7

Notes: Each panel shows the results of a different policy simulation. Column 1 shows the results when we assume $\varepsilon = 0.35$ and $\phi = \$280$, and Column 2 shows the results when we assume $\varepsilon = 0.35$ and $\phi = 0$. "Mean earnings change" refers to the change in mean earnings from Period 1 to Period 2 predicted in the full study population (i.e. the population with counterfactual earnings between -\$10,000 below and \$10,000 above the exempt amount). "Percent earnings change" is the percent change in mean earnings predicted in the full study population. "Share who adjust" refers to the percent of the full study population whose earnings does not change in response to the policy change. Note that only 50.4 percent of the full study population has counterfactual earnings above the exempt amount and therefore has incentives that are potentially affected by the policy change in our model. "Mean change among adjusters" refers to the change in mean earnings predicted among those who change earnings in response to the policy change. "Percent change among adjusters" refers to the percent change in mean earnings among those who change earnings in response to the policy change. "BRR" is the benefit reduction rate. See Appendix A.A7 for further explanation.