“BETTER BUNCHING, NICER NOTCHING”

Marinho Bertanha, Andrew McCallum, Nathan Seegert

B Supplemental Appendix for Online Publication

B.1 General Utility Maximization Problem with Multiple Kinks and Notches

To generalize the objective function in Equation 1, we update the budget set to have J different tax regimes that change at cutoff points $0 < K_1 < \ldots < K_J$ on pre-tax labor income Y. Each tax regime has income tax t_j such that $0 \leq t_0 \leq t_1 \leq \ldots \leq t_J < 1$. There are two possible tax changes. A change in tax rate is a kink. A lump-sum tax change is called a notch. Agent type N^* maximizes utility $U(C, Y; N^*)$ as follows

$$\max_{C,Y} \quad C - \frac{N^*}{1+1/\varepsilon}\left(\frac{Y}{N^*}\right)^{1+\frac{1}{\varepsilon}} \tag{B.1}$$

$$s.t. \quad C = \sum_{j=0}^{J} \mathbb{I}\{K_j < Y \leq K_{j+1}\}\left[I_j + (1-t_j)(Y-K_j)\right], \tag{B.2}$$

where $K_0 = 0$, $K_{J+1} = \infty$, $\mathbb{I}\{\cdot\}$ is the indicator function, the solution is always on the budget frontier (Equation B.2), and we assume the agent resolves indifference by choosing the smallest value of Y. The elasticity of income Y with respect to $(1-t_j)$ is equal to ε when the solution is interior.

The budget frontier is continuous except when there is a notch. The limit of the budget frontier when $Y \downarrow K_j$ is equal to I_j, but equal to $I_{j-1} + (1-t_{j-1})(K_j - K_{j-1})$ when $Y \uparrow K_j$. The size of the jump discontinuity at a notch location K_j is equal to $I_j - I_{j-1} - (1-t_{j-1})(K_j - K_{j-1})$. The intercepts I_j and I_{j-1} are assumed to be such that jump discontinuities at notches are negative.

B.2 General Solution with Multiple Kinks and Notches

Lemma B.1 below provides a general solution to Problem B.1 with any combination of kinks and notches.

Lemma B.1. *Define $\mathcal{N} = \cup_{j=0}^{J}\left(K_j(1-t_j)^{-\varepsilon}; K_{j+1}(1-t_j)^{-\varepsilon}\right]$ as the set of N^* values for which the indifference curves are tangent to the budget frontier. The function $Y^* : \mathcal{N} \to \mathbb{R}$, $Y^*(N^*) = \sum_{j=0}^{J} \mathbb{I}\{K_j(1-t_j)^{-\varepsilon} < N^* \leq K_{j+1}(1-t_j)^{-\varepsilon}\} N^*(1-t_j)^{\varepsilon}$, maps N^* values to the Y values corresponding to such tangency points. Similarly, $C^*(N^*)$ is consumption on the budget frontier (Equation B.2) when $Y = Y^*(N^*)$. Let C_j be the value of $C^*(N^*)$ whenever $Y^*(N^*) = K_j$, $j = 1, \ldots, J$. For a notch-point K_j, define the value of N_j^I to be that of the first indifference curve tangent to the budget frontier on the right of $Y = K_j$, such that the utility level is equal to the utility of the notch-point K_j,*

$$N_j^I = \min\left\{N^* \in \mathcal{N} \ : \ U(C_j, K_j) = U(C^*(N^*), Y^*(N^*))\right\}. \tag{B.3}$$

In the case of a kink, the bunching interval is defined as $[\underline{N}_j, \overline{N}_j]$*, where* $\underline{N}_j = K_j(1-t_{j-1})^{-\varepsilon}$*, and* $\overline{N}_j = K_j(1-t_j)^{-\varepsilon}$*. In the case of a notch, the expression for* $\underline{N}_j$ *equals that of the kink case, but* $\overline{N}_j$ *changes to* N_j^I*.*

Note that the bunching intervals of two consecutive kinks do not overlap, that is, $K_j(1-t_j)^{-\varepsilon} < K_{j+1}(1-t_j)^{-\varepsilon}$*. The same is not true for a kink or a notch* K_{j+1} *that comes right after a notch* K_j*, because* N_j^I *may be greater than* $K_{j+1}(1-t_j)^{-\varepsilon}$ *depending on* ε*. In this case,* $Y = K_{j+1}$ *does not appear in the solution. To account for that, construct a subsequence* $\{j_l\}_{l=1}^L$ *of* $\{1,\ldots,J\}$ *such that: (i)* $j_1 = 1$*; and (ii) for* $l \geq 2$*, set* j_l *to be the smallest* j *such that* $\underline{N}_j > \overline{N}_{j_{l-1}}$*. Then, the solution to the maximization problem in (B.1) is given by*

$$Y = \begin{cases} N^*(1-t_{j_1-1})^{\varepsilon} & , \text{ if } \ 0 < N^* < \underline{N}_{j_1} \\ K_{j_1} & , \text{ if } \ \underline{N}_{j_1} \leq N^* \leq \overline{N}_{j_1} \\ N^*(1-t_{j_2-1})^{\varepsilon} & , \text{ if } \ \overline{N}_{j_1} < N^* < \underline{N}_{j_2} \\ \vdots & \\ N^*(1-t_{j_L-1})^{\varepsilon} & , \text{ if } \ \overline{N}_{j_{L-1}} < N^* < \underline{N}_{j_L} \\ K_{j_L} & , \text{ if } \ \underline{N}_{j_L} \leq N^* \leq \overline{N}_{j_L} \\ N^*(1-t_J)^{\varepsilon} & , \text{ if } \ \overline{N}_{j_L} < N^* < \infty. \end{cases} \tag{B.4}$$

Proof. For every $N^* > 0$, there exists an unique solution on the budget frontier. If the consumer is indifferent between two solutions, we assume the consumer takes the solution with less Y. The proof is by induction over $\bar{J} = 0, 1, \ldots, J$. Denote the budget frontier $BF^{\bar{J}}$ by

$$C = \sum_{j=0}^{\bar{J}} \mathbb{I}\{\bar{K}_j < Y \leq \bar{K}_{j+1}\} \left[I_j + (1-t_j)(Y - \bar{K}_j)\right].$$

where $\bar{K}_j = K_j$ for $j = 0, 1, \ldots, \bar{J}$ and $\bar{K}_{\bar{J}+1} = \infty$.

As we change the budget frontier from $BF^{\bar{J}}$ to $BF^{\bar{J}+1}$, $K_{\bar{J}+1}$ takes a finite value strictly greater than $K_{\bar{J}}$, and $K_{\bar{J}+2}$ is set to ∞. If the solution to Problem B.1 with budget frontier $BF^{\bar{J}}$ is such that $Y < K_{\bar{J}+1} < \infty$, then this is also the solution to Problem B.1 with budget frontier $BF^{\bar{J}+1}$. In fact, points on $BF^{\bar{J}}$ dominate points on $BF^{\bar{J}+1}$, and they coincide for $Y < K_{\bar{J}+1}$.

Part I: $\bar{J} = 0$*, solve Problem B.1 with budget* BF^0*.*

This is a standard consumer maximization problem where the optimal choice for Y occurs at the point the indifference curve is tangent to BF^0. Therefore, for $N^* > 0$, $Y = N^*(1-t_0)^{\varepsilon}$.

Part II: $\bar{J} = 1$*, solve Problem B.1 with budget* BF^1*.*

The budget frontier BF^1 has two segments BF_0^1 for $0 < Y \leq K_1$, and BF_1^1 for $K_1 < Y$. If $N^* < K_1(1-t_0)^{-\varepsilon}$, then the solution of Part I, $Y = N^*(1-t_0)^{\varepsilon} < K_1$, is also the solution in Part II. It remains to find the solution for $N^* \geq K_1(1-t_0)^{-\varepsilon}$. These solutions must lie on BF^1 for $Y \geq K_1$ because they strictly dominate those that lie to the left of K_1.

Case I : Suppose K_1 *is a kink.*

Assume N^* is such that $K_1(1-t_0)^{-\varepsilon} \leq N^* \leq K_1(1-t_1)^{-\varepsilon}$. If the solution is interior to BF_1^1, then it must be at a tangent point in which case $Y = N^*(1-t_1)^{\varepsilon}$. However,

$Y = N^*(1-t_1)^\varepsilon \leq K_1$, a contradiction because this Y falls outside of the interior of BF_1^1. Therefore, if N^* is such that $\underline{N}_1 = K_1(1-t_0)^{-\varepsilon} \leq N^* \leq K_1(1-t_1)^{-\varepsilon} = \overline{N}_1$, then the solution is $Y = K_1$. Suppose $N^* > \overline{N}_1$. Then, the solution is in the interior of BF_1^1, and it is equal to $Y = N^*(1-t_1)^\varepsilon$.

Case II : Suppose K_1 is a notch.

There is a jump-down discontinuity in BF^1 at K_1, and BF^1 is continuous from the left. Consider the point $(C, Y) = (C_1, K_1)$ on BF_0^1. Define Y^D to be the value of Y such that the corresponding C value on BF_1^1 is equal to C_1. The jump-down discontinuity creates a strictly dominated region on BF_1^1 because the utility of (C_1, K_1) is strictly greater than the utility of any solution with $Y \in (K_1, Y^D)$. Indifference between K_1 and Y^D is resolved towards K_1 by assumption. Therefore, we cannot have solutions to Problem B.1 with budget BF^1 such that $Y \in (K_1, Y^D]$.

Define the point $\widetilde{N}_1^I$ as being the solution of Problem B.1 with budget BF^1 (instead of BF). This is the smallest N^* for which Problem B.1 with budget BF_1^1 has solution with utility equal to $U(C_1, K_1)$.

First, a solution $\widetilde{N}_1^I$ exists. To see that, note that for small N^*, the tangent point $Y = N^*(1-t_1)^\varepsilon$ along BF_1^1 falls in the dominated region $Y \in (K_1, Y^D]$, and the utility is less than $U(C_1, K_1)$; on the other hand, the utility at this tangent point increases with N^*, and it eventually equals $U(C_1, K_1)$. The solution is such that $\widetilde{N}_1^I \geq Y^D(1-t_1)^{-\varepsilon} > K_1(1-t_1)^{-\varepsilon}$.

Second, the solution $\widetilde{N}_1^I$ is unique. To see that, solve for N^* in the equation below.

$$U(C_1, K_1) = U\left(I_1 + N^*(1-t_1)^{\varepsilon+1} - K_1(1-t_1)\ ,\ N^*(1-t_1)^\varepsilon\right)$$

where $C = I_1 + N^*(1-t_1)^{\varepsilon+1} - K_1(1-t_1)$ is consumption on BF_1^1 when $Y = N^*(1-t_1)^\varepsilon$. Evaluating and rearranging the equality gives

$$N^*(1-t_1)^{1+\varepsilon} + \varepsilon(N^*)^{-1/\varepsilon}(K_1)^{\frac{1+\varepsilon}{\varepsilon}} = (1+\varepsilon)\left[C_1 - I_1 + K_1(1-t_1)\right]$$

The solution is unique because the derivative of the right-hand side is strictly positive given $N^* > K_1(1-t_1)^{-\varepsilon}$. Note that $\widetilde{N}_1^I$ is the unique solution to Problem B.1 when the budget is BF^1.

Call $\widetilde{Y}_1^I = \widetilde{N}_1^I(1-t_1)^\varepsilon$. Suppose there is a solution to Problem B.1 with budget BF^1 such that $Y^D < Y \leq \widetilde{Y}_1^I$. This solution is interior to budget BF_1^1, so we must have $Y = N^*(1-t_1)^\varepsilon$ for some N^*. But such a solution cannot be a solution to Problem B.1 with budget BF^1 because $Y \leq \widetilde{Y}_1^I$ and so dominated by (C_1, K_1). Therefore, we cannot have solutions to Problem B.1 with budget BF^1 such that $Y \in (K_1, \widetilde{Y}^I]$.

It remains to characterize the solution when N^* is such that $K_1(1-t_0)^{-\varepsilon} \leq N^*$. If N^* is such that $\underline{N}_1 = K_1(1-t_0)^{-\varepsilon} \leq N^* \leq \widetilde{Y}^I(1-t_1)^{-\varepsilon} = \overline{N}_1$, the solution cannot be in the interior of BF_0^1 since $Y = N^*(1-t_0)^\varepsilon \geq K_1$; it cannot be in $(K_1, \widetilde{Y}^I]$ either. Assume it is in the interior of BF_1^1 with $Y > \widetilde{Y}_I$. Since it is interior, it satisfies $Y = N^*(1-t_1)^\varepsilon$, but $N^* \leq \widetilde{Y}^I(1-t_1)^{-\varepsilon}$ which makes $Y \leq \widetilde{Y}^I$, a contradiction. Therefore, the solution to Problem B.1 with budget BF^1 when $N^* \in [\underline{N}_1; \overline{N}_1]$ is $Y = K_1$. Finally, suppose $N^* > \overline{N}_1$. Then, the solution is in the interior of BF_1^1, and it is equal to $Y = N^*(1-t_1)^\varepsilon$.

Part III: *Assume the solution of Problem B.1 with budget $BF^{\bar{J}}$ and $1 \leq \bar{J} < J$ is as* *in Equation B.4 with $\bar{J}$. Show that* (B.4) *with $\bar{J}+1$ solves Problem B.1 with budget $BF^{\bar{J}+1}$.*

Consider Problem B.1 with budget $BF^{\bar{J}}$ and solution B.4 with L being $\bar{L}$. If N^* is such that $Y < K_{\bar{J}+1} < \infty$, then Y also solves Problem B.1 with budget $BF^{\bar{J}+1}$. Therefore, the solution to Problem B.1 with budget $BF^{\bar{J}+1}$ or budget $BF^{\bar{J}}$ coincide for those values of N^*. Note also that, if K_j is a notch and $j < j_{\bar{L}}$, then the value of $\overline{N}_j$ (defined in (B.3)) does not change when the budget changes from $BF^{\bar{J}}$ to $BF^{\bar{J}+1}$. If $K_{j_{\bar{L}}}$ is a notch, then the value $\overline{N}_{j_{\bar{L}}}$ may change (case IV below). In what follows, consider the last two budget segments of $BF^{\bar{J}+1}$: $BF^{\bar{J}+1}_{\bar{J}}$ and $BF^{\bar{J}+1}_{\bar{J}+1}$.

Case I : $K_{j_{\bar{L}}}$ is a kink, $K_{\bar{J}+1}$ is a kink

In this case, $j_{\bar{L}+1} = \bar{J}+1$ because $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} > \overline{N}_{j_{\bar{L}}}$, so that $\bar{J}+1$ is the smallest j such that $\underline{N}_j > \overline{N}_{j_{\bar{L}}}$. It is also true that $j_{\bar{L}} = \bar{J}$. To see that, note that consecutive intervals $[\underline{N}_j, \overline{N}_j]$ never overlap for kinks because $\overline{N}_j = K_j(1-t_j)^{-\varepsilon} < K_{j+1}(1-t_j)^{-\varepsilon} = \underline{N}_{j+1}$. The upper limit of a kink interval j is strictly smaller than the lower limit of a notch interval $j+1$. However, the upper limit of a notch interval j may be bigger than the lower limit of the next interval $j+1$. Suppose $j_{\bar{L}} = \bar{J}$ were not true, that is, $j_{\bar{L}} < \bar{J}$. Then, any j such that $j_{\bar{L}} < j \leq \bar{J}$ is not in the subsequence $\{j_l\}$ because $K_{j_{\bar{L}}}$ is a notch, and its interval overlaps with the j interval. But this is a contradiction with $K_{j_{\bar{L}}}$ being a kink point.

If $N^* < K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$, then the solution B.4 with budget $BF^{\bar{J}}$ is $Y < K_{\bar{J}+1}$, and Y also solves Problem B.1 with budget $BF^{\bar{J}+1}$ for that same value of N^*. It remains to characterize the solution when $N^* \geq K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$

Assume N^* is such that $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} \leq N^* \leq K_{\bar{J}+1}(1-t_{\bar{J}+1})^{-\varepsilon} = \overline{N}_{\bar{J}+1}$. As seen in Part II, Case I, the solution cannot be interior to $BF^{\bar{J}+1}_{\bar{J}+1}$. The solution must be at $K_{\bar{J}+1}$. Assume $N^* > \overline{N}_{\bar{J}+1}$. Then, the solution is interior to $BF^{\bar{J}+1}_{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$.

Case II : $K_{j_{\bar{L}}}$ is a kink, $K_{\bar{J}+1}$ is a notch

As seen in Part III, Case I, $j_{\bar{L}} = \bar{J}$. We also have $j_{\bar{L}+1} = \bar{J}+1$ because the j interval $[\underline{N}_j, \overline{N}_j]$ of a kink does not overlap with the $j+1$ interval of a notch.

If $N^* < K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$, then the solution B.4 with budget $BF^{\bar{J}}$ is $Y < K_{\bar{J}+1}$, and Y also solves Problem B.1 with budget $BF^{\bar{J}+1}$ for that same value of N^*. It remains to characterize the solution when $N^* \geq K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$

Assume N^* is such that $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} \leq N^* \leq \overline{N}_{\bar{J}+1}$, where $\overline{N}_{\bar{J}+1}$ is the solution of Problem B.3 when the budget is $BF^{\bar{J}+1}$. As seen in Part II, Case II, the solution Y cannot be in $(K_{\bar{J}+1}, \overline{N}_{\bar{J}+1}(1-t_{\bar{J}+1})^{\varepsilon}]$ or in the interior of $BF^{\bar{J}+1}_{\bar{J}+1}$. Therefore, the solution is $Y = K_{\bar{J}+1}$. Assume $N^* > \overline{N}_{\bar{J}+1}$. Then, the solution is interior to $BF^{\bar{J}+1}_{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$.

Case III : $K_{j_{\bar{L}}}$ is a notch, $\overline{N}_{j_{\bar{L}}} < \underline{N}_{\bar{J}+1}$

For the notch $K_{j_{\bar{L}}}$, the solution $\overline{N}_{j_{\bar{L}}}$ to Problem B.3 when the budget is $BF^{\bar{J}}$ does not change when the budget becomes $BF^{\bar{J}+1}$ precisely because $\overline{N}_{j_{\bar{L}}} < \underline{N}_{\bar{J}+1}$. In this case, $j_{\bar{L}+1} = \bar{J}+1$. For N^* such that $\overline{N}_{j_{\bar{L}}} < N^* < K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$, the solution B.4 with budget

$BF^{\bar{J}}$ is $Y < K_{\bar{J}+1}$, and Y also solves Problem B.1 with budget $BF^{\bar{J}+1}$ for that same value of N^*. It remains to characterize the solution when $N^* \geq K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$.

Assume $K_{\bar{J}+1}$ is a kink, and that N^* is such that $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} \leq N^* \leq K_{\bar{J}+1}(1-t_{\bar{J}+1})^{-\varepsilon} = \overline{N}_{\bar{J}+1}$. As seen in Part II, Case I, the solution cannot be interior to $BF^{\bar{J}+1}_{\bar{J}+1}$. The solution must be at $K_{\bar{J}+1}$. Assume $N^* > \overline{N}_{\bar{J}+1}$. Then, the solution is interior to $BF^{\bar{J}+1}_{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$.

Assume $K_{\bar{J}+1}$ is a notch, and that N^* is such that $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} \leq N^* \leq \overline{N}_{\bar{J}+1}$, where $\overline{N}_{\bar{J}+1}$ is the solution of Problem B.3 when the budget is $BF^{\bar{J}+1}$. As seen in Part II, Case II, the solution Y cannot be in $(K_{\bar{J}+1}, \overline{N}_{\bar{J}+1}(1-t_{\bar{J}+1})^{\varepsilon}]$ or in the interior of $BF^{\bar{J}+1}_{\bar{J}+1}$. Therefore, the solution is $Y = K_{\bar{J}+1}$. Assume $N^* > \overline{N}_{\bar{J}+1}$. Then, the solution is interior to $BF^{\bar{J}+1}_{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$.

Case IV : $K_{j_{\bar{L}}}$ is a notch, $\overline{N}_{j_{\bar{L}}} \geq \underline{N}_{\bar{J}+1}$

The indifference value for Y at $\overline{N}_{j_{\bar{L}}}$ is $Y^I_{j_{\bar{L}}} = \overline{N}_{j_{\bar{L}}}(1-t_{\bar{J}})^{\varepsilon} \geq \underline{N}_{\bar{J}+1}(1-t_{\bar{J}})^{\varepsilon} = K_{\bar{J}+1}$. If $\overline{N}_{j_{\bar{L}}} = \underline{N}_{\bar{J}+1}$, the solution to Problem B.3 when the budget is $BF^{\bar{J}}$ remains unchanged when the budget becomes $BF^{\bar{J}+1}$. If $\overline{N}_{j_{\bar{L}}} > \underline{N}_{\bar{J}+1}$, then $Y^I_{j_{\bar{L}}} > K_{\bar{J}+1}$, and the solution to Problem B.3 when the budget is $BF^{\bar{J}}$ changes when the budget becomes $BF^{\bar{J}+1}$. The value of $\overline{N}_{j_{\bar{L}}}$ increases such that the new indifference point satisfies $Y^I_{j_{\bar{L}}} = \overline{N}_{j_{\bar{L}}}(1-t_{\bar{J}+1})^{\varepsilon}$.

There does not exist a j such that $\underline{N}_j > \overline{N}_{j_{\bar{L}}}$ because $K_{\bar{J}+1}$ is the last tax-change point available and $\underline{N}_{\bar{J}+1} \leq \overline{N}_{j_{\bar{L}}}$. Therefore, when constructing the solution of Problem B.1 with budget $BF^{\bar{J}+1}$, the last term in the subsequence $\{j_l\}$ remains $j_{\bar{L}}$.

The point $K_{j_{\bar{L}}}$ is a notch, so Part II, Case II says that for N^* such that $\underline{N}_{j_{\bar{L}}} = K_{j_{\bar{L}}}(1-t_{j_{\bar{L}}-1})^{-\varepsilon} \leq N^* \leq \overline{N}_{j_{\bar{L}}}$, the solution Y cannot be in $(K_{j_{\bar{L}}}, \overline{N}_{j_{\bar{L}}}(1-t_{\bar{J}+1})^{\varepsilon}]$ or in the interior of $BF^{\bar{J}+1}_{\bar{J}+1}$. Therefore, the solution is $Y = K_{j_{\bar{L}}}$. Assume $N^* > \overline{N}_{j_{\bar{L}}}$. Then, the solution is interior to $BF^{\bar{J}+1}_{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$. □

B.3 Friction Errors and Failure of the "Polynomial Strategy"

This section presents a counterexample that illustrates the failure of a common identification strategy used in applied work to estimate the elasticity using kinks. For a review, see Kleven (2016).

First, we set the parameters of the model. The true values are: $\varepsilon = 1.5$ (elasticity); $t_0 = .2$ and $t_1 = 0.3$ (before and after tax rates); kink-point $k = 0$. The bunching interval is $[\underline{n}, \overline{n}] = [0.335, .535]$. The distribution of the ability variable is assumed uniform, $n^* \sim U[-.565; 1.435]$; that is, the support is centered at 0.435 and has length equal to 2. The probability of bunching, or bunching mass B, is equal to 10% in this example. The friction error e is also assumed uniformly distributed $e \sim U[-0.5; 0.5]$. The value of labor income observed by the researcher is $\widetilde{y} = y + e$, where y is a function of n^*, ε, t_0, and t_1, as described in Equation 4.

In the counterfactual scenario of no tax change, we have $\underline{n} = \overline{n}$, and the counterfactual income with friction error is denoted $\widetilde{y}_0$. The counterfactual income without friction error is y_0. Figure B.1a depicts the PDF of $\widetilde{y}$ and $\widetilde{y}_0$.

A common identification strategy used in applied work is to fit a polynomial to the PDF of $\widetilde{y}$ excluding observations in the neighborhood of the kink $k = 0$, that corresponds to the support of the measurement error (i.e. $[-0.5; 0.5]$). The estimated bunching mass is the area between the PDF of $\widetilde{y}$ and the polynomial fit extrapolated to the excluded neighborhood around the kink. Figure B.1b illustrates the procedure. The figure shows that such strategy fails to identify the true bunching mass, even when the polynomial fit of 7th order is perfect, and we assume the researcher knows the support of e.

The last part of the estimation strategy uses the extrapolated polynomial to predict the counterfactual PDF of y_0. Following Equation 6, identification of ε requires the counterfactual PDF of y_0, without measurement error. Figure B.1c shows that the polynomial strategy fails to retrieve the PDF of y_0. The PDF predicted by the polynomial regression does not integrate to one, and thus it is not a PDF. If we divide the polynomial-based PDF in Figures B.1b and B.1c by its integral, the PDF shifts up in the graphs. The re-normalized PDF still misses the true f_{y_0}, and the underestimation of B is larger than before.

The polynomial strategy fails for two reasons:

1. The PDF of $\widetilde{y}$ is not simply the PDF of y plus the PDF of e (Figure B.1a), but the convolution between the two PDFs. While y_0 and e have uniform distributions, with a flat PDF, their convolution does not have a flat PDF. As a result, extrapolating the polynomial to find the bunching mass and to predict the PDF of y_0 is misleading;

2. The counterfactual distribution required for identification of the elasticity is the PDF of y_0, and not the PDF of $\widetilde{y}_0$ (Equation 6). Moreover, even if friction errors were not a problem, it is not possible to use the distribution of y to back out the distribution of y_0 for values of y_0 inside $[k, k + (s_0 - s_1)\varepsilon]$. The shape of the distribution of y_0 is unidentified when n^* falls in the bunching interval (Figure 1).

B.4 Parametric Gaussian Family Identifies the Elasticity

We demonstrate how to verify conditions (11) - (13) in the parametric Gaussian case. Suppose the distribution of n^* follows a normal distribution with unknown mean μ and unknown variance σ^2, such that $F_{n^*}(n) = G_{n^*}(n; \mu, \sigma^2) = \Phi\left(\frac{n-\mu}{\sigma}\right)$ where Φ denotes the standard normal CDF.

Take $(k, s_0, s_1, \varepsilon, \mu, \sigma^2)$ arbitrary. The goal is to show that $\bar{\varepsilon} = \varepsilon$, $\bar{\mu} = \mu$, and $\bar{\sigma}^2 = \sigma^2$ are the only solutions to the equalities below:

$$\Phi\left(\frac{k - \varepsilon s_1 - \mu}{\sigma}\right) - \Phi\left(\frac{k - \varepsilon s_1 - \mu}{\sigma}\right) = \Phi\left(\frac{k - \bar{\varepsilon} s_1 - \bar{\mu}}{\bar{\sigma}}\right) - \Phi\left(\frac{k - \bar{\varepsilon} s_1 - \bar{\mu}}{\bar{\sigma}}\right) \tag{B.5}$$

$$\Phi\left(\frac{u - \varepsilon s_0 - \mu}{\sigma}\right) = \Phi\left(\frac{u - \bar{\varepsilon} s_0 - \bar{\mu}}{\bar{\sigma}}\right) \quad \text{for } \forall u < k \tag{B.6}$$

$$\Phi\left(\frac{u - \varepsilon s_1 - \mu}{\sigma}\right) = \Phi\left(\frac{u - \bar{\varepsilon} s_1 - \bar{\mu}}{\bar{\sigma}}\right) \quad \text{for } \forall u > k \tag{B.7}$$

Take (B.5), and apply $\Phi^{-1}(\cdot)$ to both sides.

$$\frac{u - \varepsilon s_0 - \mu}{\sigma} = \frac{u - \bar{\varepsilon} s_0 - \bar{\mu}}{\bar{\sigma}}, \quad \forall u < k.$$

These are two lines that must have the same slope, $1/\sigma = 1/\bar{\sigma}$, and the same intercept $(\varepsilon s_0 + \mu)/\sigma = (\bar{\varepsilon} s_0 + \bar{\mu})/\bar{\sigma}$. These imply that $\bar{\sigma} = \sigma$, and $\bar{\varepsilon} s_0 + \bar{\mu} = \varepsilon s_0 + \mu$.

Similarly, (B.6) implies that $\bar{\varepsilon} s_1 + \bar{\mu} = \varepsilon s_1 + \mu$. Subtracting this last equation from the previous one gives $\bar{\varepsilon}(s_1 - s_0) = \varepsilon(s_0 - s_1)$, which yields $\bar{\varepsilon} = \varepsilon$. Finally, $\varepsilon s_1 + \bar{\mu} = \varepsilon s_1 + \mu$ gives $\bar{\mu} = \mu$.

□

B.5 Implementation of Censored Quantile Regressions

The optimization problem in Equation 21 is computationally difficult. For the left (or right) censored case, Chernozhukov and Hong (2002) proposed a fast and practical estimator that consists of three steps. First, you fit a flexible Probit model that explains the probability of no censoring; then, you select observations whose values of X lead to a predicted probability of no censoring that is greater than $1 - \tau$. Second, you fit a quantile regression of y on X using the selected observations in the first step; then, you select observations whose values of X lead to a predicted quantile that is greater than k. Third, repeat the second step using the observations selected at the end of the second step. Chernozhukov and Hong (2002) demonstrate consistency and asymptotic normality of their three-step estimator. Moreover, they show that the standard errors computed by the quantile regression in the third step are valid.

Our case of middle censoring requires a straightforward modification of the method proposed by Chernozhukov and Hong (2002). Inspired by their algorithm, we propose the following implementation steps.

1. Create dummies $\delta_i^- = \mathbb{I}\{y_i < k\}$ (not censored, left of k) and $\delta_i^+ = \mathbb{I}\{y_i > k\}$ (not censored, right of k). Fit two Probit models to estimate $\mathbb{P}[\delta_i^+ | X_i] = \Phi(X_i g^+)$ and $\mathbb{P}[\delta_i^- | X_i] = \Phi(X_i g^-)$, where Φ denotes the cdf of a standard normal distribution, and $g^{\pm}$ are vectors of parameters. You may use powers and interactions of X_i to make this stage as flexible as possible. Select two subsamples as follows. Compute the 10th quantile of the empirical distribution of $\Phi(X_i \hat{g}^+) - (1 - \tau)$ conditional on $\Phi(X_i \hat{g}^+) > 1 - \tau$. Let $\kappa_0^+(\tau)$ be the 10th quantile of that distribution. The first subsample is $J_0^+(\tau) = \{i : \Phi(X_i \hat{g}^+) > 1 - \tau + \kappa_0^+(\tau)\}$. The second subsample is $J_0^-(\tau) = \{i : \Phi(X_i \hat{g}^-) > \tau + \kappa_0^-(\tau)\}$, where $\kappa_0^-(\tau)$ is the 10th quantile of the empirical distribution of $\Phi(X_i \hat{g}^-) - \tau$ conditional on $\Phi(X_i \hat{g}^-) > \tau$. Create a dummy $W_i^0 = \mathbb{I}\{i \in J_0^+(\tau)\}$.

2. Fit the quantile regression model $Q_\tau(y_i | X_i, W_i^0) = X_i b(\tau) + W_i^0 \delta(\tau)$ using observations in $J_0^-(\tau) \cup J_0^+(\tau)$. Use the estimates of this quantile regression, that is $\hat{b}^0(\tau)$ and $\hat{\delta}^0(\tau)$, to create two subsamples as follows. The first subsample is $J_1^+(\tau) = \{i : X_i \hat{b}^0(\tau) + \hat{\delta}^0(\tau) > k + \kappa_1^+(\tau)\}$, where $\kappa_1^+(\tau)$ is the 3rd quantile of the empirical distribution of $X_i \hat{b}^0(\tau) + \hat{\delta}^0(\tau) - k$ conditional on $X_i \hat{b}^0(\tau) + \hat{\delta}^0(\tau) > k$. The second subsample is $J_1^-(\tau) = \{i : X_i \hat{b}^0(\tau) < k + \kappa_1^-(\tau)\}$, where $\kappa_1^-(\tau)$ is the 97th

quantile of the empirical distribution of $X_i\hat{b}^0(\tau) - k$ conditional on $X_i\hat{b}^0(\tau) < k$. Create a dummy $W_i^1 = \mathbb{I}\{i \in J_1^+(\tau)\}$.

3. Fit the quantile regression model $Q_\tau(y_i|X_i, W_i^1) = X_i b(\tau) + W_i^1 \delta(\tau)$ using observations in $J_1^-(\tau) \cup J_1^+(\tau)$ to obtain estimates $\hat{b}^1(\tau)$ and $\hat{\delta}^1(\tau)$. The elasticity estimator is $\hat{\varepsilon} = \hat{\delta}^1(\tau)/(s_1 - s_0)$.

B.6 Estimates with the Filtering Method of Saez (2010)

In this section, we recompute the estimates of Table 1 using a different filtering method. Specifically, we employ the procedure used by Saez (2010) to obtain the bunching mass and the side limits of the distribution of income without friction error Y. The procedure implicitly defines a way to estimate the unobserved distribution of Y given the observed distribution of income with friction error $\tilde{Y}$. We refer the reader to Figure 2 by Saez (2010).

The first step is to construct a histogram-based estimate of the PDF $f_{\tilde{Y}}$, and then average $f_{\tilde{Y}}$ for $\tilde{Y} \in [K - 2\delta, K - \delta] \cup [K + \delta, K + 2\delta]$, where $K = 8,580$ is the kink point, and $\delta = 1,500$ defines the excluded region. Call that average $\bar{f}$. The bunching mass is estimated by the area between two curves, $f_{\tilde{Y}}$ and $\bar{f}$. The continuous portion of f_Y equals $f_{\tilde{Y}}$, except for the excluded region $[K - \delta, K + \delta]$, where f_Y equals $\bar{f}$. We obtain the CDFs F_Y and $F_{\tilde{Y}}$ from their PDF estimates. Finally, we rely on $Y = F_Y\left(F_{\tilde{Y}}^{-1}(\tilde{Y})\right)$ to transform $\tilde{Y}$ into Y.

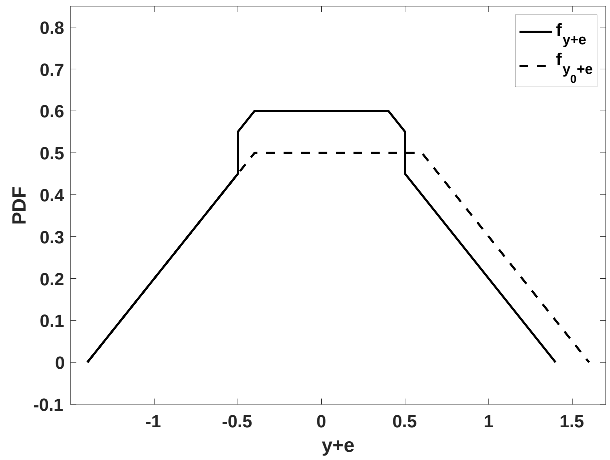
Table B.1: Estimates Using U.S. Tax Returns 1995--2004

Statistical Model	(1) Saez (2010)	(2) Theorem 2 Bounds M = 0.5	(3) Theorem 2 Bounds M = 1	(4) Tobit Full Sample	(5) Tobit Trunc. 75%	(6) Tobit Trunc. 50%	(7) Tobit Trunc. 25%	(8) Sample details	
All								Obs.	189.1m
Elasticity (ε)	0.235	$[0.223, 0.250]$	$[0.210, 0.282]$	0.138	0.165	0.170	0.197	Avg.	$54.1k
	(0.0311)			(0.0001)	(0.0001)	(0.0001)	(0.0002)	Std.	$131.1k
Self-employed								Obs.	33.5m
Elasticity (ε)	0.933	$[0.768, 1.304]$	$[0.685, \infty]$	0.632	0.759	0.764	0.847	Avg.	$61.8k
	(0.0759)			(0.0006)	(0.0007)	(0.0007)	(0.0009)	Std.	$168.2k
Self-employed, married								Obs.	24.0m
Elasticity (ε)	0.391	$[0.330, 0.441]$	$[0.290, \infty]$	0.185	0.254	0.288	0.318	Avg.	$75.0k
	(0.0823)			(0.0004)	(0.0006)	(0.0007)	(0.0008)	Std.	$185.6k
Self-employed, not married								Obs.	9.6m
Elasticity (ε)	1.260	$[1.130, 1.519]$	$[1.019, \infty]$	1.074	0.978	1.013	1.275	Avg.	$28.7k
	(0.1193)			(0.0011)	(0.0011)	(0.0012)	(0.0017)	Std.	$106.3k

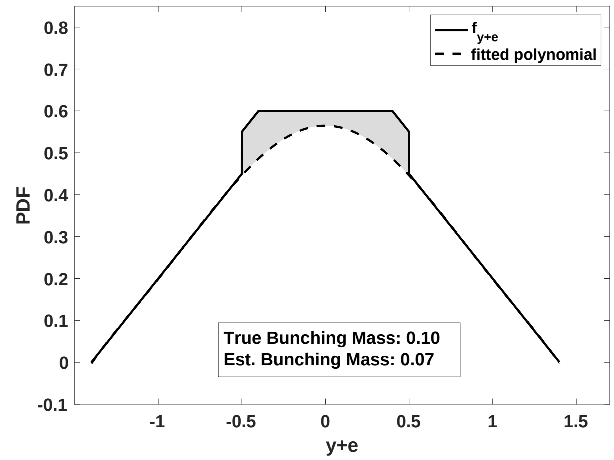
Notes: The table shows estimates of the elasticity for four different subsamples of the IRS data, and using three different approaches discussed in the paper. The first approach (column 1) uses the trapezoidal approximation to point-identify the elasticity (Example 1). Estimates and standard errors were computed using the publicly available code by Saez (2010) at the website of the American Economic Journal, Economic Policy. The second approach (columns 2 and 3) computes partially identified sets for the elasticity (Theorem 2), using non-parametric estimates of the side limits of f_y at the kink, and the bunching mass. Side limits were estimated using the method of Cattaneo et al. (2019). The estimate for the bunching mass equals the sample proportion of y observations that equals the kink point (see discussion in Section B.6 on friction errors). Upper and lower bounds are calculated for two choices of M, that is, the maximum slope of the PDF of the unobserved heterogeneity n^*. Column 4 has Tobit MLE estimates of the elasticity that utilizes the full sample of data, along with robust standard errors. Columns 5 through 7 report truncated Tobit MLE estimates. As we move from column 5 to column 7, we restrict the estimation sample to shrinking symmetric windows around the kink that utilizes 75% to 25% of the data. The set of covariates that enters the Tobit estimation is kept constant across different truncation windows. It includes dummy variables such as marital and employment status, year effects, types of deductions or social security benefits received, and whether the filer used a tax prep software.

Figure B.1: Counterexample where "Polynomial Strategy" Fails

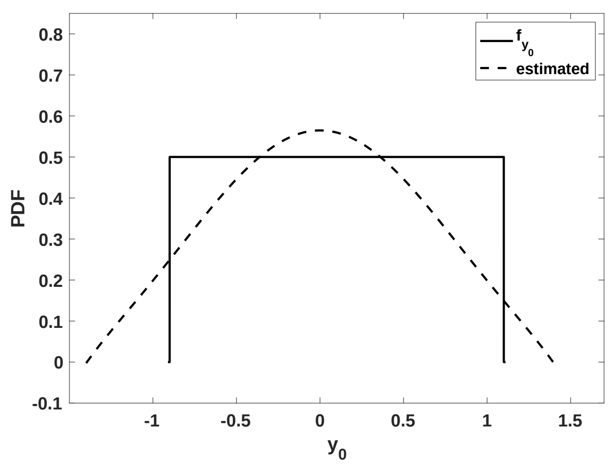
(a) Distribution of Income with Friction Error



(b) Estimation of Bunching Mass



(c) Counterfactual Distribution of Income
without Friction Error



Notes: The population model of this example has $\varepsilon = 1.5$, $t_0 = .2$, and $t_1 = 0.3$ at kink $k = 0$. The distribution of ability is assumed uniform, $n^* \sim U[-.565; 1.435]$. The probability of bunching is equal to 10%, and the distribution of the friction error is $e \sim U[-0.5; 0.5]$. The researcher observes $\widetilde{y} = y + e$, where y is a function of n^*, ε, t_0, and t_1, as described in Equation 4. Figure B.1a displays the PDF of $\widetilde{y}$ and $\widetilde{y}_0$. Figure B.1b displays the fitted 7th-order polynomial to the PDF of $\widetilde{y}$ using observations in $(-\infty, -0.5) \cup (0.5, \infty)$. The bunching mass is estimated by the integral of the difference between $f_{\widetilde{y}}$ and the fitted polynomial, inside the excluded region. The polynomial strategy understimates the true bunching mass, and does not retrieve the PDF of y_0 (Figure B.1c).