Why Don't Taxpayers Bunch at Kink Points?*

Andrew H. McCallum † Michael A. Navarrete ‡

April 6, 2023

Preliminary and Incomplete: Comments Welcome.

Abstract

We introduce a new theory and new estimation method for optimizing frictions with a piecewise linear constraint. Allowing frictions to depend on observables, we estimate why agents do not behave as standard frictionless models predict. Our methods are not limited to public finance and apply to a general class of mixture models and any of the four possible piecewise linear constraints, 1) slope increase (convex kink), 2) slope decrease (concave kink), 3) intercept increase (convex notch), or 4) intercept decrease (concave notch). We demonstrate these methods in three of these four settings. Individual income tax returns with a, 1) convex kink and, 2) concave kink implied by the EITC. New Jersey real estate transfer taxes with a 3) convex notch. We document which covariates account for a substantial share of optimizing frictions and provide elasticity estimates that explicitly control for optimizing frictions.

JEL codes: F10, F14, F65, G21
Keywords: bunching, notching, optimizing frictions, EM algorithm

*Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Board of Governors of the Federal Reserve System, or any other person associated with the Federal Reserve System. Alexis Payne and Benjamin Stagoff-Belfort provided excellent research assistance.

†Corresponding author, andrew.h.mccallum@frb.gov, Board of Governors of the Federal Reserve System, Washington, DC 20551

‡mnav@umd.edu, University of Maryland

1 Introduction

In several distributions we see a mass point where individuals make decisions that place them at or near this point in the distribution. This type of behavior is referred to as bunching in the literature. Typically when bunching is present, there is an incentive for individuals to do so. The stronger the incentive to bunch, the more likely we are to see a larger mass point in the distribution. Bunching happens because there is a change in the incentive schedule. If the slope of the incentive schedule changes, then we refer to that point as a kink. If the intercept of the incentive schedule changes, then we refer to that point as a notch. In general, notches lead to more bunching because notches have strong disincentives to be just to the right of that point by creating dominated regions. Individuals that are optimizing with perfect information should never be just to the right of a notch because their payoff will lower than if they just bunched at the notch point. Convex kinks do not create this dominated region where it would never be optimal for agents to be located just to the right of a kink point. However, concave kinks which occur less commonly than convex kinks empirically do exhibit a dominated region.

One gap in the literature is that we observe bunching only in a few cases. Several kink points exhibit no bunching despite individuals having an incentive to do so. Much of the research on bunching has focused only when bunching is especially pronounced: sharp bunching. When researchers apply their bunching estimators to non-sharp bunching cases, "fuzzy" bunching, they imprecisely estimate their parameters of interest. We construct a micro founded model that allows us to explain the observed "fuzzy" bunching, while simultaneously retrieving parameters of interest from these kink points. Although we are not the first paper to develop a bunching estimator that allows for "fuzzy" bunching (Alvero and Xiao, 2020), we are the first to develop a bunching estimator that incorporates optimizing frictions that are micro-founded with covariates both on the individuals' ability and individuals' optimizing frictions. We are also the first to use an estimator, maximum likelihood estimation (MLE), in conjunction with the Expectation Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977) extension to missing data (Ruud, 1991) to all 4 cases: (1) convex kink, (2) concave kink, (3) convex notch, and (4) concave notch.

Although the tax salience literature (Chetty, Looney, and Kroft (2009); Farhi and Gabaix (2020); Kroft, Laliberté, Leal-Vizcaíno, and Notowidigdo (2020)) directly influenced the optimizing frictions literature (Kleven and Waseem (2013); Chetty (2012); Chetty, Friedman, and Saez (2013)), we view our work as directly contributing to the optimizing friction literature; we develop a bunching estimator that incorporates these optimizing frictions. The tax salience literature typically focuses on individuals misperceiving prices by not taking sales tax into account. Our bunching estimator is applicable to settings where non-linearities in the incentive schedule are present such as taxpayers facing marginal tax rates. In our model agents will face optimizing frictions about the location of the kink (or notch) point, which will result in "fuzzy" bunching (or notching).

Workers respond to tax rates, which affect how many hours they choose to supply and whether they choose to participate in the labor force. Estimating the elasticity of earnings with respect to the net of tax rate is a central problem in public economics. Much of the literature considers a continuous univariate distribution of agents who face a piece-wise linear schedule of incentives such as the one imposed by the Earned Income Tax Credit (EITC). Other work has shown that agents face "optimization frictions" that prevent them from responding to the incentive schedule as implied by utility maximization (Chetty, 2012; Kleven and Waseem, 2013). Our contribution will be to build a micro founded model that incorporates these optimization frictions and to retrieve the elasticity of earnings even when no apparent bunching occurs.

The method of using bunching to estimate an elasticity is referred to as a bunching estimator. In the tax setting bunching estimators estimate the elasticity of earnings with respect to the net of tax rate. Bunching estimators began with Saez (2010), which was followed by several influential bunching estimator papers such as Kleven and Waseem (2013) and Chetty, Friedman, Olsen, and Pistaferri (2011). However, a limitation in this literature is that these estimators are based on sharp bunching which is not the case when optimizing frictions affect how agents bunch.

Our paper has two goals. First, we seek to explain why agents do not appear to optimize utility. We do this by introducing a micro foundation for a second distribution of heterogeneity of agents' optimizing friction that affects how close agents can bunch to the locations of kinks or notches in the incentive schedule. We estimate the effects that agents' observable characteristics have on the probability of not reporting at the kink by making these optimizing frictions a function of observable as well unobservable heterogeneity. Second, our method improves on the methods used to estimate the tax elasticity by controlling for optimizing frictions explicitly. This improvement addresses results from work that documents that estimates of the tax elasticity will be attenuated in the presence of optimizing frictions (Kostøl and Myhre, 2021).

Optimizing frictions are present in cases of bunching; often agents experience optimizing frictions in both kink and notch points. For example, Saez (2010, figure 3A) finds that individuals bunch at the first EITC kink but do not bunch at higher EITC kinks (reproduced in figure 1). The third kink point in figure 1 corresponds to a concave kink. At this point there should be a hole in the data (dominated region), but empirically we do not observe a perceptible change in the distribution. The existence of optimizing frictions are even more striking in most empirical studies of notches. Utility maximization subject to the true incentive schedule predicts that no individuals should report in a dominated region just to the right of the notch. Empirically, agents facing a notch are often observed in this dominated region Kopczuk and Munroe (2015, figure A.1) (reproduced in figure 2).

We introduce optimizing frictions as a friction that prevents an agent from bunching at a kink or notch point. These optimizing frictions could take a myriad of forms ranging from agents misperception about the location of the kink or notch point to labor market frictions that prevent a worker from controlling the exact hours that they supply (inflexible work hours). Agents in our model choose the optimal level of consumption and labor supply (consequently income) to maximize utility subject to their budget constraint. This constraint is a piece-wise function of tax rates and agents' optimizing frictions in relation to the location of where tax rates change. These optimizing frictions can be set to zero, which would result in agents having perfect knowledge of the tax schedule along with perfect flexibility on their labor supplied. In this way, we nest the prior literature that assumes agents face no optimizing frictions (Bertanha, McCallum, and Seegert, 2020). Agents will optimize utility with the incorrect constraint when they face optimizing frictions and potentially report non-optimal taxable income (labor supply).

The optimal reported income that results in our model is a function of the heterogeneous earnings ability and heterogeneous optimizing frictions. We can write the probability of observing reported earnings as a function of these two sources of heterogeneity and then develop the likelihood function. By making heterogeneous earnings ability and optimizing frictions functions of observable covariates and unobserved errors, we can simultaneously estimate the non-tax determinants of earnings and optimizing frictions while also recovering the elasticity. We then employ the expectation-maximization (EM) algorithm to maximize the likelihood and estimate all the parameters of the model. Previous work incorporating optimization frictions has often used a two step process that first cleans the data of the optimizing frictions before estimating the elasticity, but our method can do both in one step. We do not want to clean the data given that we are particularly interested in those optimizing frictions.

Recent work such as Bertanha et al. (2020) have shown that a kink does not identify an elasticity when the distribution of agents is non-parametric and continuous. The elasticity can be partially identified in the kink case with semi-parametric conditions that use covariates. By introducing uncertainty around the kink and having covariates determine agents' degree of uncertainty, we can better estimate the elasticity parameter.

The paper proceeds with an utility maximization model subject to a piecewise-linear constraint in section 2. We propose multiple estimation strategies in Section 3.

2 Optimization and the data generating process

This section presents a microeconomic theoretical framework that leads to the data generating processes (DGPs) that we study. But we emphasizes that our estimation methods are applicable to these DGPs regardless of how those DGPs come about. Each DGP is the result of heterogeneous agents optimizing their iso-elastic quasi-linear utility function subject to a piecewise-linear constraint. The constraint is defined by a sequence of intercepts and slopes that change at points that are perceived by the individual to be heterogeneous as well. We call a change in the the slope a ''kink'', and call a change in the intercept a ''notch''. We show the different DGPs can be generated for increases or decreases in the slope or intercept at each change point. To provide a concrete example, we turn to the well-known public finance settings of Saez (2010) and Kleven and Waseem (2013).

2.1 Microeconomic model

Each agent i maximizes utility, $U(C_i, L_i)$, by jointly choosing the level of consumption C_i and labor supply L_i. Utility is increasing in C_i, but the agent has disutility from working, and so utility is decreasing in L_i. As in much of the public finance literature and for simplicity, we assume the prices of labor and consumption are equal to one, such that taxable labor income Y_i is equal to L_i and the value of consumption is the same as after-tax income. The agent faces a budget constraint and may consume all of their labor income net of taxes plus an exogenous endowment or lump sum tax, I_0. We illustrate the model using only one kink or notch. Formally, this model is given by the following utility maximization problem

$$(C_i, Y_i) = \arg\max_{C_i, Y_i} \; C_i - \frac{N_i^*}{1+1/\varepsilon} \left(\frac{Y_i}{N_i^*} \right)^{1+\frac{1}{\varepsilon}} \tag{1a}$$

subject to:

$$C_i = \mathbb{1}_{Y_i \leq K_i^*} \left[I_0 + (1-t_0) Y_i \right] + \mathbb{1}_{Y_i > K_i^*} \left[I_0 + (1-t_0) K_i^* - \Delta + (1-t_1)(Y_i - K_i^*) \right], \tag{1b}$$

in which $\mathbb{1}$ is the indicator function. The constraint has intercept I_0 and slope $1 - t_0$ if $Y_i \leq K_i^*$, but intercept $I_{i1} = I_0 + K_i^*(1 - t_0) - \Delta$ with slope $1 - t_1$ if $Y_i > K_i^*$. The budget constraint in Equation (1b) has a kink because agents perceive the tax rate of t_0 for reporting income below the kink K_i^* and perceive the tax rate of t_1 for incomes above the kink. Likewise, Equation (1b) has a notch when the agent faces a lump-sum tax or transfer Δ for reporting income greater or equal to K_i^*. Without a notch, $\Delta = 0$, the constraint in Equation (1b) is continuous over Y_i. In the case of a notch, the constraint has a discontinuous change of size Δ at $Y_i = K_i$.

Ability, N_i^*, and kink heterogeneity K_i^*, are known to the agent when they choose optimal consumption and reported income but unknown to the econometrician. Importantly, we do not take a stand on the theoretical source of heterogeneity in N_i^* and K_i^*. Instead of imposing interpretations at this stage, we will let estimates from our data inform our interpretation. Heterogeneity in earning ability could, for example, be related to industry of employment or a measure of human capital possessed by the agent. Heterogeneity in the kink location, could arise, for example, from agents not being able to perfectly control their reported income because of capital gains or losses. Alternatively, agents may not know the true tax schedule, so they incorrectly guess the location of the kink point and optimize with respect to their perceived budget constraint based on that guess. As such, using a professional tax preparer may affect their proximity to the true kink.

2.1.1 Model solution for convex kink, convex notch, or concave notch

The solution to Equation (1) for C_i and Y_i are well known in the literature for a kink (Saez, 2010) and for a notch (Kleven and Waseem, 2013) when the budget constraint is convex —that is when $t_0 \leq t_1$ and $\Delta \geq 0$, respectively. To ease exposition of that well-known solution, we take the natural logarithm of all variables. Define $y_i = \ln(Y_i)$, $n_i^* = \ln(N_i^*)$, $\underline{n}_i = \ln(\underline{N}_i)$, $\bar{n}_i = \ln(\bar{N}_i)$, $k_i = \ln(K_i)$, $s_0 = \ln(1 - t_0)$, and $s_1 = \ln(1 - t_1)$ and then write the solution for one change in the tax schedule as our first DGP

$$y_i = \begin{cases} \varepsilon s_0 + n_i^* & \text{, if} \quad n_i^* < \underline{n}_i\left(k_i^*, \varepsilon, s_0, \Delta\right) \\ k_i^* & \text{, if} \quad \underline{n}_i\left(k_i^*, \varepsilon, s_0, \Delta\right) \leq n_i^* \leq \bar{n}_i\left(k_i^*, \varepsilon, s_1, \Delta\right) \\ \varepsilon s_1 + n_i^* & \text{, if} \quad n_i^* > \bar{n}_i\left(k_i^*, \varepsilon, s_1, \Delta\right). \end{cases} \tag{2}$$

The elasticity of income with respect to one minus the tax rate when the solution is interior is given by ε. The thresholds that determine the reporting cases for the agents are functions of the tax rates and the agent's heterogeneous kink locations. If an agent's log ability is lower than $\underline{n}_i\left(k_i^*, \varepsilon, s_0, \Delta\right)$, then the agent will supply labor that will place them below their individual kink point, k_i^*. Similarly, an agent of ability greater than $\bar{n}_i\left(k_i^*, \varepsilon, s_1, \Delta\right)$ will supply labor that will place them above their individual kink point. For values of $n_i^* \in \left[\underline{n}_i\left(k_i^*, \varepsilon, s_0, \Delta\right), \bar{n}_i\left(k_i^*, \varepsilon, s_1, \Delta\right)\right]$ inside the individual's bunching interval, the agent's indifference curve is never tangent to the budget constraint. Instead, it is optimal to report the non-interior solution $y_i = k_i^*$.

Thresholds for convex kinks and notches are defined by the abilities that make agents indifferent. An agent with ability $\underline{n}_i\left(k_i^*, \varepsilon, s_0, \Delta\right)$ is indifferent between earning income $y_i = \varepsilon s_0 + n_i^*$ and $y_i = k_i^*$. Likewise, an agent with ability $\bar{n}_i\left(k_i^*, \varepsilon, s_1, \Delta\right)$ is indifferent between earning income $y_i = \varepsilon s_1 + n_i^*$ and $y_i = k_i^*$. For a convex kink these thresholds can be obtained as explicit functions $\underline{n}_i\left(k_i^*, \varepsilon, s_0\right) = k_i^* - \varepsilon s_0$ and $\bar{n}_i\left(k_i^*, \varepsilon, s_1\right) = k_i^* - \varepsilon s_1$.

For a convex notch, these thresholds are implicit functions of as shown in Appendix A.1.

In the public finance literature, differences between income from Equation (2) and that same equation with $k_i^* = k$ is often called "optimizing frictions". Our estimation approach will allow us to extract these optimizing frictions and to document covariates that can explain them in Section 3.

In addition to distorting reported income, we emphasize that these optimizing frictions imply that the budget constraint Equation (1b) will not provide the actual level of consumption —which is equivalent to after tax income —attained by the agent. Instead, agents optimize considering their own kink location, k_i^*, but then pay taxes according to the tax code with the true location, k. As such, actual consumption is given by Equation (1b) with k replacing k_i^*. The next Section 2.1.2 will provide graphical depictions of these solutions with and without frictions.

2.1.2 Graphical depiction of solution for a convex kink

Figure 3 provides graphical intuition for how agents' optimizing frictions affect the optimal income to report when they face a convex kink. This intuition is clearest in levels, denoted in capital letters, but the solution from Equation (2) is clearest in logs, denoted in lower case letters.

Figure 3a shows the convex kink without optimizing frictions and Figure 3b shows the convex kink with optimizing frictions, K_i^*, along with the indifference curves. These indifference curves correspond to two individuals, one with the lowest ability, $\underline{N}\left(K, \varepsilon, \Delta\right)$, and the other with the highest ability, $\bar{N}\left(K, \varepsilon, \Delta\right)$, that report the same level of income, $Y_i = K$, in each case.

Figure 3c and Figure 3d provide histograms of reporting income that would be observed by this optimizing behavior and a distribution of N_i^* from a normal distribution with mean zero and standard deviation one. As is well known, the kink generates a mass point at K. However, the the shape of that mass point depends on whether or not one allows for optimizing frictions. In the case without optimizing frictions, Figure 3c, we observe sharp bunching where all the mass point is exactly at the kink point. In the case with optimizing frictions, Figure 3d, we observe fuzzy bunching where the mass point is diffused around the kink point. In a scenario with more optimization frictions, the diffusion could be large enough to eliminate the mass point at the kink. For example, the second kink point in Figure 1 does not exhibit a mass point.

Figure 3b shows the indifference curve of $\overline{N}$ (light blue line) and the indifference curve of $\underline{N}$ (light green line). Both of these agents will choose to bunch at the kink point, $K = 3$. Let $\tilde{N}$ denote an individual with ability between $\overline{N}$ and $\underline{N}$. All agents with ability between $\overline{N}$ and $\underline{N}$ including $\tilde{N}$ will optimally choose to bunch at the kink point, K, when there are no optimizing frictions are present. However, if $\tilde{N}$ faces optimizing frictions such that the location of the kink point with optimizing frictions moves from 3 to 7, $K^* = 7$, then their budget constraint would shift from the solid black line to the dashed black line. When $\tilde{N}$ solves Equation (2), they choose to supply labor such that they will receive an income around 5. The infeasible indifference curve for $\tilde{N}$ is shown by pink line. However, that consumption-income pair is not feasible, so $\tilde{N}$ will shift down to the original budget constrain (solid black line). This will result in $\tilde{N}$ receiving lower utility than if they had bunched at the kink point, $K = 3$.

Even though optimizing frictions such as the ones experienced by $\tilde{N}$ potentially lead agents to a lower indifference curve, the model in Figure 3b is more akin to behaviors exhibited by agents such as taxpayers bunching at EITC kink points than the model in Figure 3a. Even in the example of sharp bunching used by Saez (2010) of the first EITC kink point, Figure 1, there is diffused mass around the first kink point. In a model without optimizing frictions such as the one in Figure 3a, agents always attain the highest utility level possible, which results in all individuals with ability levels between $\overline{N}$ and $\underline{N}$ to bunch exactly at the kink point. This is unrealistic because agents could face a multitude of optimizing frictions. For example, taxpayers in Figure 1 may use a heuristic to recall the location of the first kink point. Instead of recalling the location of the first kink point as \$8,580, taxpayers may recall the location of the first kink point as \$8,600.

With only unobserved ability, agents are still optimizing correctly and should be bunching at the kink point, K, when they have ability between $\overline{N}$ and $\underline{N}$. Once taking optimizing frictions into account, agents become more dispersed around the kink points as reflected by Figure 3d. If individuals face optimizing frictions where K^* is to the right of K, then this would lead individuals to potentially work more and end up at a lower indifference curve as shown in Figure 3b. Similarly, if individuals face optimizing frictions where K^* is to the left of K, then the individual could work less than the optimal amount. It is important to note that optimizing frictions do not unambiguously change agents' allocation of consumption and income. An agent could fall within the bunching interval both with and without optimizing frictions. It could also be that optimizing frictions do not affect the agent's optimal allocation. For example, if the agent is to the left of the kink point and has optimizing frictions where the perceived kink is further to the right, then the agent will remain at the same allocation.

2.2 Model solution for a concave kink

The solution to Equation (1) for Y_i and C_i are provided by Bertanha et al. (2020) for a kink and in Appendix A.2.2 for a notch when the budget constraint is concave —that is when $t_0 > t_1$ and $\Delta \leq 0$, respectively. The solution for one change in the tax schedule is the second DGP given by

$$y_i = \begin{cases} \varepsilon s_0 + n_i^* & \text{, if} \quad n_i^* \leq \check{n}_i\left(k_i^*, \varepsilon, s_0, s_1, \Delta\right) \\ \varepsilon s_1 + n_i^* & \text{, if} \quad n_i^* > \check{n}_i\left(k_i^*, \varepsilon, s_0, s_1, \Delta\right). \end{cases} \tag{3}$$

There is only one threshold for a concave kink or notch given by $\check{n}_i\left(k_i^*, \varepsilon, s_0, s_1, \Delta\right)$ and it is defined as the ability that make the agent indifferent between incomes $\underline{y}_i = \varepsilon s_0 + \check{n}_i\left(k_i^*, \varepsilon, s_0, s_1, \Delta\right)$ and $\bar{y}_i = \varepsilon s_1 + \check{n}_i\left(k_i^*, \varepsilon, s_0, s_1, \Delta\right)$. For a concave notch, the threshold is an implicit function as shown in Appendix A.3. For a concave kink the threshold can be obtained as an explicit function given by $\check{n}_i\left(k_i^*, \varepsilon, s_0, s_1\right) = k_i^* + \ln\left(\varepsilon + 1\right) + b\left(s_0, s_1, \varepsilon\right)$ in which $b\left(s_0, s_1, \varepsilon\right) = \ln\left[\left(\exp\left(s_0\right) - \exp\left(s_1\right)\right) / \left(\exp\left(s_0\right)^{\varepsilon+1} - \exp\left(s_1\right)^{\varepsilon+1}\right)\right]$.

Importantly, even when $k_i^* = k$, for a concave budget constraint, there is no excess mass in the distribution of y_i at k. Instead, there is an region of zero mass between $\bar{y}_i$ and $\underline{y}_i$. As the elasticity increases, the region of zero mass increases as well, and Bertanha et al. (2020) show the simple formula, $\hat{\varepsilon} = \left(\bar{y}_i - \underline{y}_i\right) / \left(s_1 - s_0\right)$, non-parametrically point-identifies ε when $k_i^* = k$ for one kink. Bertanha et al. (2020) show how to non-parametrically point identification of the elasticity with a convex notch which we extend to a concave notch in Appendix A.4.

Differences between income from Equation (3) and that same equation with $k_i^* = k$ are optimizing frictions for concave budget constraints. When $k_i^* \neq k$, these non-parametric identification schemes are unavailable because the start, $\underline{y}_i$, and the end, $\bar{y}_i$, of the zero-mass region cannot be observed in the data. As such, we will apply our estimation techniques to the DGP from Equations (2) and (3).

Like the DGP in Equation (2), the optimizing frictions $k_i^* \neq k$ imply that the budget constraint Equation (1b) will not provide the actual after tax income attained by the agent because they pay taxes according to the tax code with the true location, k. Instead, actual consumption is given by Equation (1b) with Equation (2) in place of y_i but k replacing k_i^*. The next Section 2.1.2 will provide graphical depictions of these solutions with and without frictions.

2.2.1 Graphical depiction of solution for a concave kink

Figure 4 provides graphical intuition for how agents' optimizing frictions affect the optimal income to report when they face a concave kink. This intuition is clearest in levels, denoted in capital letters, but the solution from Equation (3) is clearest in logs, denoted in lower case letters.

Figure 4a shows the concave kink without optimizing frictions and Figure 4b shows the concave kink with optimizing frictions ($K_i^* = K$) along with indifference curves. These indifference curves correspond to the individual, $\check{N}$, that is indifferent at reporting at $\underline{Y}$ and $\bar{Y}$. In Figure 4a, an $\check{N}$ will be tangent at $\underline{Y}$ (1.68) and $\bar{Y}$ (4). An individual with ability less than $\check{N}$ will report income weakly less than $\underline{Y}$. Conversely, an individual with ability more than $\check{N}$ will report income weakly more than $\bar{Y}$. This results in a hole in the distribution between $\underline{Y}$ and $\bar{Y}$.

Figure 4c and Figure 4d provide histograms of reporting income that would be observed by this optimizing behavior and a distribution of N_i^* from a normal distribution with mean zero and standard deviation one. When there is no optimizing frictions, the concave kink generates a hole in the distribution, which is a region where no individual will choose to supply labor that will result in an income between $\underline{Y}$ and $\bar{Y}$. However, the the size and potential existence of said hole in the distribution depends on whether or not one allows for optimizing frictions. In the case without optimizing frictions, Figure 4c, we observe a hole in the distribution that corresponds to the difference between $\underline{Y}$ and $\bar{Y}$. In the case with optimizing frictions, Figure 4d, we see the hole disappear. There is still less mass between $\underline{Y}$ and $\bar{Y}$, but that depends on the elasticity, tax rates, and optimization frictions.

Figure 4b shows the indifference curve of $\check{N}$ (dotted blue line) when they face no optimization frictions or face optimization frictions such that $K^* = K = 3$ and the indifference curve of $\check{N}$ (dashed pink line) when they face optimization frictions such that $K^* = 7$. Both of these agents will choose to bunch at the K^*. However, $\check{N}$ with $K^* = 7$ will be tangent at $Y = 3.92$ and $Y = 9.3$. For this agent this corresponds to their $\underline{Y}$ and $\bar{Y}$, respectively. If this agent chooses to report income at their $\underline{Y}$ of 3.92, then they will end up in the region where there was a hole in Figure 4a. By definition, $\check{N}$ with $K^* = 7$ is indifferent between reporting income at 3.92 and 9.34. With these optimizing frictions, $\check{N}$ will end up at a lower indifference curve than if they were optimizing under no optimization frictions.

With only unobserved ability, agents are still optimizing in a manner that results in them reaching the highest utility possible. This leads all agents to choose to supply labor such that their income is weakly less than $\underline{Y}$ or weakly more than $\bar{Y}$. Once taking optimizing frictions into account, agents have individual levels of $\underline{Y}$ and $\bar{Y}$. This leads to more mass to the right of $\underline{Y}$ and to the left of $\bar{Y}$ as reflected by Figure 4d. In the scenario with no optimization friction and all agents have a homogeneous $\underline{Y}$ and $\bar{Y}$, which results in there being no mass between $\underline{Y}$ and $\bar{Y}$. It could also be that optimizing frictions do not affect the agent's optimal allocation of labor to supply. For example, if the agent has ability marginally less than $\check{N}$ and face optimization frictions such that $K^* > K$, then the agent will remain at the same allocation.

3 Estimation

3.1 General estimation problem

The optimal behavior by the agent given optimizing frictions k_i^* for convex and concave budget constraints provided by Equations (2) and (3) are the two DGP we will use for estimation. The equations are closely related and our estimation methodology encompass both of them. Our estimation methodology applies to any DGP that takes the form of Equations (2) or (3) and the specific micro-economic theory outlined above is not a necessary condition for estimation.

The main challenge for estimation created by the introduction of optimizing frictions, k_i^*, is that observations of y_i are not partitioned into being above, below, and at the kink for Equation (2), or being below or above the kink in Equation (3). In particular, this feature of the data precludes using the methods developed in Bertanha et al. (2020).

Our approach is to use the EM algorithm as the method for recovering the maximum likelihood estimates from a missing-data problem. One useful way to think of our problem is as a missing-data problem. For example, each reported income observation is the agent choosing to be below, at, or above the kink location, but we do not know with certainty to which group that observation belongs when there is heterogeneity in location, k_i^*. As such, we are missing data about to which group the agent belongs —and groups are defined as the different budget constraint slopes that the agent is trying to face. Estimation will need to estimate membership in each group and estimate the parameters that determine the income level reported within each group. Ruud (1991) shows applications of the EM algorithm to missing-data problems such as ours by writing a latent model. In our setting, we will have 3 latent variables: y_{0i}^* income below the kink without optimizing frictions, k_i^* income at the kink without optimizing frictions, and y_{1i}^* income above the kink without optimizing frictions.

The Expectation-Maximization (EM) algorithm is an iterative method that recovers the Maximum Likelihood Estimates (MLE) of the parameters of a latent variable model. The EM algorithm was developed by Dempster et al. (1977), which has approximately 64,000 citations on Google Scholar as of writing. The EM algorithm has wide application across many areas of science and has many other interpretations, not least of those is the cross-entropy loss in machine-learning.

3.1.1 Three cases for a convex kink, concave notch, or convex notch

The probability of observing a level of the outcome y_i is a joint probability of reporting a level of the outcome, y_i, in one of the three cases in Equation (2) and having ability, n_i^*, and location, k_i^* to report in that case. We separate each of those three cases in Equation (2) into the probability of two events. The first event is the probability of observing the level of the outcome variable y_i that is optimal to report conditional on the agent having ability, n_i^*, and location, k_i^*. The second event is the probability that observed income y_i comes from one of the three DGPS in Equation (2).

For the first case defined in Equation (2), the joint probability of observing the level of the outcome, y_i, and the optimal choice to face slope, s_0, is given by

$$
\begin{aligned}
&P\left[y_i \mid n_i^* < \underline{n}_i\left(k_i^*, \varepsilon, s_0, \Delta\right)\right] P\left[y_i < k_i^* \mid y_i = \varepsilon s_0 + X_i'\beta + \nu_i\right] \\
&= P\left[y_i = \varepsilon s_0 + n_i^*\right] P\left[y_i < k_i^* \mid \nu_i = y_i - \varepsilon s_0 - X_i'\beta\right] \\
&= f_{n^*}\left(y_i - \varepsilon s_0\right) G_{n^*,k^*}\left(y_i < k_i^*\right) = f_{i1} P_{i1}. \qquad (4)
\end{aligned}
$$

The probability of observing earnings, y_i, when the agent reports at the location, k_i^*, is

$$
\begin{aligned}
&P\left[y_i = k_i^*\right] P\left[n_i^* \in \left[k_i^* - \varepsilon s_1, k_i^* - \varepsilon s_0\right] \mid y_i = k_i^*\right] \\
&= f_{k^*}\left(y_i\right) G_{n^*,k^*}\left(y_i = k_i^*\right) = f_{i2} P_{i2}, , \qquad (5)
\end{aligned}
$$

The probability of observing earnings, y_i, when the agent optimizes to face the slope, s_1, is

$$\begin{aligned}
&P\left[y_i \mid n_i^* > \bar{n}_i\left(k_i^*, \varepsilon, s_1, \Delta\right)\right] P\left[y_i > k_i^* \mid y_i = \varepsilon s_1 + n_i^*\right] \\
&= P\left[y_i = \varepsilon s_1 + n_i^*\right] P\left[y_i - k - Z_i'\gamma > \psi_i \mid y_i - \varepsilon s_1 - X_i'\beta = \nu_i\right] \\
&= f_{n^*}\left(y_i - \varepsilon s_1\right) G_{n^*,k^*}\left(y_i > k_i^*\right) = f_{i3} P_{i3}, \qquad (6)
\end{aligned}$$

We assume that (n_i^*, k_i^*) is joint $i.i.d$ with cumulative distribution function (CDF) G_{n^*,k^*} and probability density function (PDF) g_{n^*,k^*}. Given g_{n^*,k^*}, the marginal ability PDF is $f_{n^*} = \int g_{n^*,k^*} dk^*$ and the marginal location PDF is $f_{k^*} = \int g_{n^*,k^*} dn^*$.

The far left hand sides of equations (4),(5), and (6) describe the joint probability of each event while the far right hand side conserves on notation by referring to the probability of observing the outcome conditional on ability and optimizing frictions as f_{ih} and probability the agent has that ability and location as P_{ih} for $h = 1, 2, 3$.

3.1.2 Two cases for a concave kink

Analogous to the steps for the convex kink or notch, we separate each of the two cases in Equation (3) into A) the probability of observing the level of the outcome variable that is optimal to report given that the agent has ability, n_i^*, and location, k_i^* and B) the probability that the agent has ability, n_i^*, and location, k_i^*, to be in that particular case.

For the first case defined in Equation (3), the joint probability of observing the level of the outcome, y_i, and the agent's choice to face slope, s_0, is given by

$$\begin{aligned}
&P\left[y_i \mid n_i^* \leq \check{n}_i\left(k_i^*, \varepsilon, s_0, s_1\right)\right] P\left[y_i < k_i^* \mid y_i = \varepsilon s_0 + X_i'\beta + \nu_i\right] \\
&= P\left[y_i = \varepsilon s_0 + n_i^*\right] P\left[y_i - k - Z_i'\gamma < \psi_i \mid \nu_i = y_i - \varepsilon s_0 - X_i'\beta\right] \\
&= f_{n^*}\left(y_i - \varepsilon s_0\right) G_{n^*,k^*}\left(y_i \leq k_i^*\right) = f_{i1} P_{i1}. \qquad (7)
\end{aligned}$$

The probability of observing earnings, y_i, when the agent chooses to face, s_1, is

$$
\begin{aligned}
& P\left[y_i \mid n_i^* > \check{n}_i\left(k_i^*, \varepsilon, s_0, s_1\right)\right] P\left[y_i > k_i^* \mid y_i = \varepsilon s_1 + n_i^*\right] \\
&= P\left[y_i = \varepsilon s_0 + n_i^*\right] P\left[y_i - k - Z_i'\gamma > \psi_i \mid y_i - \varepsilon s_1 - X_i'\beta = \nu_i\right] \\
&= f_{n^*}\left(y_i - \varepsilon s_1\right) G_{n^*,k^*}\left(y_i > k_i^*\right) = f_{i2} P_{i2}. \qquad (8)
\end{aligned}
$$

The far left hand sides of equations (7) and (8) describe the joint probability of each event while the far right hand side conserves on notation by referring to the probability of observing the level of income conditional on ability and optimizing frictions as f_{ih} and probability the agent has that ability and optimizing frictions as P_{ih} for $h = 1, 2$.

3.2 The likelihood

Define $\Omega = (\varepsilon, \theta)$ to include the tax elasticity, ε, and parameters of the distribution of ability and locations $g\left(n_i^*, k_i^* \mid \theta\right)$. The likelihood of observing any given level of income for an individual is $L\left[\Omega \mid y_i, s_0, s_1, k\right] = \sum_{h=1}^{H} f_{ih} P_{ih}$. For the convex kink case, this sum is over the $H = 3$ Equations (4),(5), and (6). In the concave kink case, this sum is over the $H = 2$ Equations (7) and (8). For readability, we abuse notation by reusing f_{ih} and P_{ih} for both convex and concave settings but it is important to note that, for example, f_{i2} from the convex case is certainly different from f_{i2} in the concave case.

The likelihood of observing an $M \times 1$ vector of M agents' incomes y is then the product of the agents' individual likelihoods given by

$$
L\left[\Omega \mid y, s_0, s_1, k\right] = \prod_{i=1}^{M} \sum_{h=1}^{H} f_{ih} P_{ih}. \qquad (9)
$$

The likelihood is not a parametric finite mixture model (FMM) that mixes H, f_{ih} distributions with probability P_{ih} for each h. This is because the P_{ih} terms do not necessarily add up 1.

Utilizing the EM algorithm is seemingly the most popular technique to recover our maximum likelihood estimates. To that end, we define an unobserved $M \times H$ matrix w as

$$w = \begin{bmatrix} w_{11} & \dots & w_{1H} \\ \vdots & \ddots & \vdots \\ w_{M1} & \dots & w_{MH} \end{bmatrix}, \quad \text{in which} \quad w_{ih} = \begin{cases} 1, & \text{observation from } f_{ih}P_{ih} \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

The unobserved matrix w tells us from which of the three cases defined by Equation (2) or two cases defined by Equation (3) is the $i-th$ observation taken. Note that each row of w contains only one "1" and $H-1$ "0"s. The elements of the matrix w convert the likelihood to multiplicative from additive form because $\sum_{h=1}^{H} f_{ih}P_{ih} = \prod_{h=1}^{H} (f_{ih}P_{ih})^{w_{ih}}$. This "augmented likelihood" and always has the same value as the likelihood for every observation when w is known but is much easier to maximize than the likelihood.

We do not observe w but we can use the EMA to estimate w and Ω iteratively. The EM algorithm is comprised of an expectation step that computes the expected value of the unobserved indicator about whether the observation is from the $h-th$ distribution conditional on the data and current iteration's parameter estimates as

$$E[w_{ih} \mid y_i, \Omega] = P[w_{ih} = 1 \mid y_i, \Omega] = \frac{f_{ih}P_{ih}}{\sum_{h=1}^{H} f_{ih}P_{ih}} = \hat{w}_{ih} \tag{11}$$

Given this expectation step, the maximization step which computes the MLE of the parameters $\hat{\Omega}$ for the convex DGP given by Equation (2) using the log of this augmented likelihood by solving

$$\hat{\Omega} = \arg\max_{\Omega} \sum_{i=1}^{M} \sum_{h=1}^{3} \hat{w}_{ih} \ln (f_{ih}P_{ih}) \tag{12a}$$

subject to:

$$\sum_{h=1}^{3} P_{ih} = 1 \quad \forall i, \qquad \underline{n}_i = \underline{n}_i (k_i^*, \varepsilon, s_0) \quad \forall i, \qquad \bar{n}_i = \bar{n}_i (k_i^*, \varepsilon, s_1) \quad \forall i. \tag{12b}$$

The maximization step for the concave DGP given by Equation (3) come from solving

$$\hat{\Omega} = \arg\max_{\Omega} \sum_{i=1}^{M} \sum_{h=1}^{2} \hat{w}_{ih} \ln\left(f_{ih} P_{ih}\right) \tag{13a}$$

$$\text{subject to:} \qquad \sum_{h=1}^{2} P_{ih} = 1 \quad \forall i, \qquad \check{n}_i = \check{n}_i\left(k_i^*, \varepsilon, s_0, s_1\right) \quad \forall i. \tag{13b}$$

To be explicit, the EM algorithm steps are to start with an initial parameter vector $\Omega^{(0)}$, compute the expectations step for the auxiliary parameters $\hat{w}_{ih}^{(0)}$ using Equations (11). Then, using these estimates of $\hat{w}_{ih}^{(0)}$, maximize the augmented log-likelihood in Equation (12b) or (13b) to get the next parameter vector $\Omega^{(1)}$. Iterate these steps J times until the augmented log-likelihood evaluated at $\Omega^{(J-1)}$ and $\Omega^{(J)}$ are the same so that J is the last iteration. These iterative steps are guaranteed to converge to a local maximum.

Intuition for the EM algorithm for a convex kink is as follows. A small observed income level is likely, but not certainly, from the agents having low ability and having a kink location such that they choose to face s_0. Thus the probability that the observation should be classified as coming from $y_i = s_0 + n_i^*$ is relatively high. Computing similar probabilities of classifications for all the observations and then holding those fixed, we can re-estimate the parameters. Those new and more precise parameter estimates can then improve the classification of observations, which can again be used to estimate even more precise parameters. This approach nests the setting without optimizing frictions ($k^* = k$) because observations are either below, at, or above k so classification is directly observed instead of estimated.

3.2.1 Numerical optimization

Solving the maximization step of the EM algorithm in Equations (12b) or (13b) requires a numerical optimization algorithm. Many of these algorithms can be classified broadly into three types based on the number of derivatives of the objective function that are used. These types are: 1) first derivative methods (for example, Gauss-Newton, Levenberg

(1944)-Marquardt (1963), and (Berndt, Hall, Hall, and Hausman, 1974, BHHH), 2) second derivative methods (for example, Newton-Raphson (NR), Broyden (1970)-Fletcher (1970)-Goldfarb (1970)-Shanno (1970) (BFGS), and Goldfeld, Quandt, and Trotter (1966), and 3) derivative free methods (for example, (Nelder and Mead, 1965, Simplex) and (Goffe, Ferrier, and Rogers, 1994, Simulated Annealing).

We undertake the maximization step of the EM algorithm using the BHHH algorithm. The NR algorithm has a quadratic convergent rate but requires computing the inverse of the second derivative (Hessian) of the likelihood. The BHHH is similar to the NR but uses the outer product of the gradient (OPG) of the likelihood in place of the Hessian. This substitution is only valid when the function being optimized is a likelihood function because the expectation of the OPG is the information matrix, which is equal to the negative of the Hessian under general regularity conditions. The BHHH algorithm has the advantages of being computationally simple because only the first derivative of the likelihood is needed and it is guaranteed to converge to a local minimum because the OPG is always positive semi-definite. One BHHH disadvantage is that it can be slow to converge to the optimum because it has an approximately linear rate of convergence near the solution. In our application in Section 4, we calculate the gradient analytically to further improve the precision and speed of each BHHH maximization step.

3.2.2 Inference

The EM alogirthm recovers the MLE parameters which are unbiased, efficient (they achieve the Cramér–Rao lower bound), and asymptotically normal. We report robust standard errors for the parameter estimates using the typical "sandwich" form of the likelihood (extensions to correct for clusters are straightforward). These standard errors can be used for inference individually on the parameters of the model.

We can use the likelihood ratio (LR) to test hypothesis on many parameters jointly. We are particularly interested in testing the null hypothesis that the determinants of optimizing frictions do not belong in the model or if we can reject that null hypothesis. The LR test is performed by estimating two models and checking for a statistically different fit between the models. Estimating two models is computationally expensive, however, and the Wald test approximates the LR test and can be calculated after estimating the model only once. We report both Wald and LR test results. All inference likelihoods or truncated likelihoods with parameters estimated using augmented likelihood. This approach avoids the need to correct for the fact that the w from Equation (10) are themselves estimates.

4 Application to EITC

The microeconomic theory and estimation methodology presented above is extremely general. Before applying these methods to a specific dataset, we make a few more assumptions. Assume that ability is $n_i^* = X_i'\beta + \nu_i$, in which X_i are observable characteristics and ν_i is an error term. Also assume that the friction is $k_i^* = k + Z_i'\gamma + \psi_i$, a function of the true location k, observable characteristics related to frictions, Z_i, and an error term ψ_i. Substituting these definitions into Equation (2) provides the DGP for the convex kink,

$$y_i = \begin{cases} \varepsilon s_0 + X_i'\beta + \nu_i & \nu_i - \psi_i < k + Z_i'\gamma - \varepsilon s_0 - X_i'\beta \\ k + Z_i'\gamma + \psi_i & \nu_i - \psi_i \in [k + Z_i'\gamma - \varepsilon s_1 - X_i'\beta, k + Z_i'\gamma - \varepsilon s_0 - X_i'\beta] \\ \varepsilon s_1 + X_i'\beta + \nu_i & \nu_i - \psi_i > k + Z_i'\gamma - \varepsilon s_1 - X_i'\beta, \end{cases} \tag{14}$$

and substituting these definitions into Equation (3) provides the DGP for the concave kink,

$$y_i = \begin{cases} \varepsilon s_0 + X_i'\beta + \nu_i & \nu_i - \psi_i \le k + Z_i'\gamma + \ln(\varepsilon + 1) + b(s_0, s_1, \varepsilon) - \varepsilon s_0 - X_i'\beta \\ \varepsilon s_1 + X_i'\beta + \nu_i & \nu_i - \psi_i > k + Z_i'\gamma + \ln(\varepsilon + 1) + b(s_0, s_1, \varepsilon) - \varepsilon s_1 - X_i'\beta. \end{cases} \tag{15}$$

in which $b(s_0, s_1, \varepsilon) = \ln\left[(\exp(s_0) - \exp(s_1)) / \left(\exp(s_0)^{\varepsilon+1} - \exp(s_1)^{\varepsilon+1}\right)\right]$.

Simple definitions for f_{ih}, P_{ih}, and F_{ih} from these two Equations (14) and (15) that are ingredients for the unobserved matrices, w, Equations (11), (12b), (13b), (**??**), and (**??**) come from assuming that (ν_i, ψ_i) is $i.i.d$ from a joint bivariate normal with zero means,

$V[\nu_i] = \sigma_\nu^2$, $V[\psi_i] = \sigma_\psi^2$, and $COR[\nu_i, \psi_i] = \rho$. Explicit definitions for the ingredients for the convex and concave problems —with or without truncation —are provided in Appendix B.1.1 and B.2.1, respectively.

Kline and Walters (2019) Say that Heckit models are numerically equivalent to instrumental variables in some cases.

4.1 Data

We use annual cross-section data from the Individual Public Use Tax Files, constructed by the Internal Revenue Service (IRS) initially used by Saez (2010). The data are individual tax returns for each year 1995 to 2004 that have been inflation-adjusted to 2008 dollars using the IRS inflation parameters. The data include basic demographic information and sampling weights which allow interpretation of any estimates as being based on the population of U.S. individual income tax returns.

During 1995 to 2004, the EITC schedule for taxpayers with one child provided taxpayers that earn between 0 and \$8,580 a marginal subsidy of −34 percent. The maximum subsidy is $\$2,917 = \$8,580 \times 0.34$. Between \$8,580 and \$15,740 of reported income, the marginal EITC rate is zero, between \$15,740 and \$33,995, the marginal rate is 0.11. For reported incomes higher than \$33,995, the marginal rate is zero again. The change in marginal rates generates a convex kink at each of \$8,580 and \$15,740 and a concave kink at \$33,995.

We use two sub-samples for estimation. What we will call the "self-employed" sub-sample includes taxpayers that report self-employment income but no wage income and have one child. The "wage earners" sub-sample includes taxpayers that report wage income but no self-employment income and have one child. We do not restrict the sample based on other sources of income such as capital gains, dividends, or pension benefits.

Figure 5a displays a histogram of earnings for the self-employed sample. The EIC amount (based on the EITC schedule) for the taxpayer is depicted by the black dashed line. The solid red vertical lines depict the locations of the convex or concave kink in the tax schedule. The literature commonly refers to the first kink point at \$8,580 as an example of sharp bunching, but the bunching mass is diffused around the convex kink point. This histogram is more similar to figure 3d than figure 3c. Within the framework of our model these self-employed tax filers are facing optimization frictions that prevent them from bunching exactly at $8,580.

However, the optimization frictions for the self-employed are not uniform within this group of tax filers; one covariate where we see a large difference in optimization frictions is marriage. In figure 5b, the solid black kernel density is self-employed not married taxpayers and the dashed black line is the self-employed married taxpayers. Figure 5b shows that not being married amongst tax filers who only report self-employment income dramatically increases the probability of reporting at the $8,580 kink relative to being married, while leaving much of the rest of the distribution unchanged. The similarity of the kernel densities away from the kink suggests that optimization frictions are playing a large role. If the married distribution had a distribution shifted to the right, then we would expect less bunching given that tax filers would be less willing to bunch at the kink point because it would represent a larger loss in income.

Figure 5c shows the percentage of taxpayers that are married with one dependent child and only self-employment income peaks very near the first kink and rapidly drops just after it. From prior work we know that many government programs such as the EITC have a marriage penalty, Alm, Dickert-Conlin, and Whittington (1999). These marriage penalties affect whether a spouse decides to join the labor market Eissa and Hoynes (2004) to increase their EIC. However, most of these effects of the EITC on the marriage market is on the extensive margin whether a spouse decides to work. Given that, it's surprising that the share of not married tax filers almost peaks near the first kink point. One might expect that the married households have a stronger incentive to bunch at the kink point.

Similar to figure 5a, figure 5d displays a histogram of earnings, but for the wage earners subsample. Unlike figure 5a, figure 5d does not display sharp nor diffused bunching at any of the three kink points. Given that the first and second kink points are convex kinks, we might expect to see bunching in one or both of these locations. The third kink point at

$33,995 corresponds to a concave kink where we would expect to see less mass in the earnings distribution. The third kink point is a concave kink because instead of the marginal tax rate increasing at $33,995, it decreases. Theoretically, we should see a hole in the distribution at the concave kink.

However, the optimization frictions for the wage earners are also not uniform within this group of tax filers; one covariate where we see a large difference in optimization frictions is whether a tax filer receives social security benefits. Figure 5e shows that receiving social security benefits amongst tax filers who only report wage income dramatically increases the probability of reporting at the $33,995 kink compared to not receiving social security benefits, while leaving much of the rest of the distribution unchanged. The solid black kernel density corresponds to wage earners without social security benefits and the dashed black line corresponds to wage earners with social security benefits. Wage earners with social security benefits appear to face larger optimization frictions given that they have less control over their social security benefits than their wage income. Receiving social security benefits is an important covariate for wage earning taxpayers and has the potential to be an important covariate in explaining why tax filers fail to move away from the third kink point.

Because the tax rate decreases at the $33,995 theory on concave kinks without frictions predicts that individual would try to report away from the that kink (as we discussed in Figure **??**). At first glance if one were to only look at Figure 5d, then one would conclude that taxpayers do not report away from the concave kink as theory would predict. Figure 5d does not depict the "hole" in the distribution but conditioning on receipt of social security benefits, paints a different picture. Figure 5e shows that receiving social security benefits dramatically increases the probability of reporting at the $33,995 kink compared to not getting benefits. The solid black kernel density is wage-earners who also receive social security benefits and the dashed black line is wage earners that do not receive social security benefits. Individuals that do not get benefits could have more control of their income and find it easier to move away from the concave kink.

Figure 5f shows the percentage of taxpayers without benefits declines as it gets closer to the $33,995 kink and then rapidly recovers afterwards. This decline in the fraction without benefits echoes the region of zero mass that we would predict at a concave kink in a model without frictions. This example with social security benefits highlights the importance of our model in taking optimization frictions into account given that after conditioning on relevant covariates that affect optimization frictions, drastically different patterns are observed. Covariates on the ability distribution matter as Bertanha et al. (2020) showed, but so do covariates on the optimization frictions distribution.

4.2 Estimation results

4.3 Convex taxes estimation results

In figure **??**, we apply the truncated version of our model to one of the subsamples that Saez (2010) uses where sharp bunching is present. We look at self-employed tax filers with 1 child. We truncate the model to only use values in a vicinity of the first kink point of the EITC ($8,580) and omit covariates on ability and on optimization frictions. Specifically we use 8% of the distribution to the left of the first kink point and 22% of the distribution to the right of the kink point. The distribution implied by our model closely matches the empirical distribution with slightly less mass at the kink point in our model. This closeness of fit is better reflected in the CDF.

In table **??**, we can compare the estimates of the taxable income with respect to the net-of-tax rate from the previous subsample of self-employed tax filers with 1 child. We find a higher elasticity estimate than Saez (2010). This difference can be explained in part by the introduction of heterogeneous optimization frictions. Even when we do not include any covariates, we find that the coefficient on γ is statistically significant, which implies that optimizing frictions matter even when sharp bunching is present. Our elasticity estimates are about 1.4 and Saez's estimate are about 1.1.

In figure 6, we apply our model to another subsample that Saez (2010) uses where where "fuzzy" bunching is present. We look at wage-earners tax filers with 1 child using 22 covariates in the model. We use all 22 covariates in both the ability distribution and the optimization frictions distribution. We use the full subsample and focus on the second kink point of the EITC ($15,740). Again our model closely fits the empirical distribution both in the PDF and CDF. At first glance there is no bunching present, but with our model that accounts for optimizing friction we are able to retrieve an elasticity estimate and other covariates of interest.

In table **??**, we can compare the estimates of the taxable income with respect to the net-of-tax rate from the previous subsample of wage-earners tax filers with 1 child using our model. Unlike Saez (2010) we find a large elasticity estimate of about 1.2 when he finds an estimate close to 0. We also provide the coefficient on the constant term on *gamma* as well as coefficient values on covariates on optimizing frictions. We illustrate 4 binary variables with the largest coefficients such as receiving a tax credit. We have standardized the variables, so the magnitudes of each of the gammas are comparable. The coefficient should be compare to the constant to see if tax filers with these characteristics are more likely to be to the left or right of the kink point when they bunch due to optimizing frictions. One should interpret the coefficient of 0.3 on have capital gains income as a 1 standard deviation increase in capital gains is associated with being 30% to the right of the kink.

Note that in our examples of applying our model to the EITC we looked at two cases: (1) self-employed workers with one child at the first kink point and (2) wage earners with one child at the second kink point. The bunching literature focuses on the first kink point where one sees the most excess mass (Saez (2010) and Chetty et al. (2013)). However, our model can estimate an elasticity at other kink points where an excess mass point may be less obvious. Not only are elasticity estimates away from the first kink more interesting, but they may also be preferred.

4.4 Concave taxes estimation results

4.5 Determinants of concave optimizing frictions

Likelihood ratio test and Wald test results go here.

Mortenson and Whitten (2020) show that approximately two-thirds of bunchers always bunch at the refund maximizing kink. This is problematic because our model as well as other models in the literature have the wrong utility function for agents (equation 1a). This type of behavior of refund maximizing occurs at the tax refund maximum, which is the first kink point of the EITC and occurs much more for self-employed individuals than wage earners; self-employed individuals can more easily misreport their true income than wage earners. We are staying within the framework of the literature by assuming the same utility function that is quasi-linear in consumption and isoelastic in labor. However, our second case with wage earners at the second case should draw less concern given that these bunchers are not located at the refund maximizing kink and are less likely to misreport their income.

5 Welfare and optimal taxes

6 Conclusion

When we run our model on the same sample as Saez (2010) where he observes sharp bunching, we are able to retrieve similar elasticity estimates to Saez (2010). We are also able to retrieve an elasticity estimate that are non-zero for other kink points with "fuzzy" bunching. Our model explicitly takes these optimizing frictions into account and estimates an elasticity for wage earners with one child in the second kink point. These taxpayers are bunching at second kink point, but due to optimizing frictions, they are less precise in bunching exactly at the kink point.

In addition to being able to derive elasticity estimates that incorporate optimizing frictions and increasing the set of kinks and notches where we can estimate an elasticity, we can estimate the impact of optimizing frictions directly and see whether they matter and if so by how much. We find that these additional covariates help determine where an agent falls in the distribution of optimizing friction. We find that about 13% of the variance in optimizing frictions is explained by our covariates in optimizing frictions distribution. In addition we can determine whether these covariates shift agents to the left or right of the kink or the notch.

We see our contribution as developing a model that allows researchers to estimate an elasticity even in scenarios without sharp bunching. Our model can be generalized and applied to non-tax situations where bunching occurs. We also provide the Stata package to allow other researchers to easily apply our model to their settings.

References

Alm, James, Stacy Dickert-Conlin, and Leslie A. Whittington (1999). "Policy watch: The marriage penalty." *Journal of Economic Perspectives*, 13(3), pp. 193--204. doi:10.1257/jep.13.3.193. URL https://www.aeaweb.org/articles?id=10.1257/jep.13.3.193.

Alvero, Adrien and Kairong Xiao (2020). "Fuzzy bunching." *Available at SSRN 3611447*.

Berndt, Ernst R., Bronwyn H. Hall, Robert E. Hall, and Jerry A. Hausman (1974). *Estimation and Inference in Nonlinear Structural Models*, pp. 653--665. NBER. URL http://www.nber.org/chapters/c10206.

Bertanha, Marinho, Andrew H. McCallum, and Nathan Seegert (2020). "Better Bunching, Nicer Notching." Working Paper 3144539, SSRN. doi:http://dx.doi.org/10.2139/ssrn.3144539.

Broyden, C. G. (1970). "The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations." *IMA Journal of Applied Mathematics*, 6(1), pp. 76--90. doi:10.1093/imamat/6.1.76. URL https://doi.org/10.1093/imamat/6.1.76.

Chetty, Raj (2012). "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *ECONOMETRICA*, 80(3), pp. 969--1018. doi:10.3982/ECTA9043.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri (2011). "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics*, 126(2), pp. 749--804.

Chetty, Raj, John N. Friedman, and Emmanuel Saez (2013). "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review*, 103(7), pp. 2683--2721. doi:10.1257/aer.103.7.2683. URL http://www.aeaweb.org/articles?id=10.1257/aer.103.7.2683.

Chetty, Raj, Adam Looney, and Kory Kroft (2009). "Salience and taxation: Theory and evidence." *American Economic Review*, 99(4), pp. 1145--77. doi:10.1257/aer.99.4.1145. URL https://www.aeaweb.org/articles?id=10.1257/aer.99.4.1145.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum likelihood from incomplete data via the em algorithm." *Journal of the Royal Statistical Society Series B (Methodological)*, 39(1), pp. 1--38. URL http://www.jstor.org/stable/2984875.

Eissa, Nada and Hilary Williamson Hoynes (2004). "Taxes and the labor market participation of married couples: the earned income tax credit." *Journal of Public Economics*, 88(9), pp. 1931--1958. doi:https://doi.org/10.1016/j.jpubeco.2003.09.005. URL https://www.sciencedirect.com/science/article/pii/S0047272703001440.

Farhi, Emmanuel and Xavier Gabaix (2020). "Optimal taxation with behavioral agents."

American Economic Review, 110(1), pp. 298--336. doi:10.1257/aer.20151079. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151079.

Fletcher, R. (1970). ''A new approach to variable metric algorithms.'' *The Computer Journal*, 13(3), pp. 317--322. doi:10.1093/comjnl/13.3.317. URL https://doi.org/10.1093/comjnl/13.3.317.

Goffe, William L., Gary D. Ferrier, and John Rogers (1994). ''Global optimization of statistical functions with simulated annealing.'' *Journal of Econometrics*, 60(1), pp. 65--99. doi:https://doi.org/10.1016/0304-4076(94)90038-8. URL https://www.sciencedirect.com/science/article/pii/0304407694900388.

Goldfarb, Donald (1970). ''A family of variable-metric methods derived by variational means.'' *Mathematics of Computation*, 24(109), pp. 23--26. URL http://www.jstor.org/stable/2004873.

Goldfeld, Stephen M., Richard E. Quandt, and Hale F. Trotter (1966). ''Maximization by quadratic hill-climbing.'' *Econometrica*, 34(3), pp. 541--551. URL http://www.jstor.org/stable/1909768.

Kleven, Henrik J. and Mazhar Waseem (2013). ''USING NOTCHES TO UNCOVER OPTIMIZATION FRICTIONS AND STRUCTURAL ELASTICITIES: THEORY AND EVIDENCE FROM PAKISTAN.'' *QUARTERLY JOURNAL OF ECONOMICS*, 128(2), pp. 669--723. doi:10.1093/qje/qjt004.

Kline, Patrick and Christopher R. Walters (2019). ''On heckits, late, and numerical equivalence.'' *Econometrica*, 87(2), pp. 677--696. doi:https://doi.org/10.3982/ECTA15444. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15444.

Kopczuk, Wojciech and David Munroe (2015). ''Mansion Tax: The Effect of Transfer Taxes on the Residential Real Estate Market.'' *AMERICAN ECONOMIC JOURNAL-ECONOMIC POLICY*, 7(2), pp. 214--257. doi:10.1257/pol.20130361.

Kostøl, Andreas R. and Andreas S. Myhre (2021). ''Labor supply responses to learning the tax and benefit schedule.'' *American Economic Review*, 111(11), pp. 3733--66. doi:10.1257/aer.20201877. URL https://www.aeaweb.org/articles?id=10.1257/aer.20201877.

Kroft, Kory, Jean-William P Laliberté, René Leal-Vizcaíno, and Matthew J Notowidigdo (2020). ''Salience and taxation with imperfect competition.'' Working Paper 27409, National Bureau of Economic Research. doi:10.3386/w27409. URL http://www.nber.org/papers/w27409.

Levenberg, Kenneth (1944). ''A method for the solution of certain non-linear problems in least squares.'' *Quarterly of Applied Mathematics*, 2(2), pp. 164--168. doi:https://doi.org/10.1090/qam/10666. URL http://www.jstor.org/stable/43633451.

Marquardt, Donald W. (1963). ‘‘An algorithm for least-squares estimation of nonlinear parameters.’’ *Journal of the Society for Industrial and Applied Mathematics*, 11(2), pp. 431--441. doi:10.1137/0111030. URL https://doi.org/10.1137/0111030.

Mortenson, Jacob A. and Andrew Whitten (2020). ‘‘Bunching to maximize tax credits: Evidence from kinks in the us tax schedule.’’ *American Economic Journal: Economic Policy*, 12(3), pp. 402--32. doi:10.1257/pol.20180054. URL https://www.aeaweb.org/articles?id=10.1257/pol.20180054.

Nelder, John A and Roger Mead (1965). ‘‘A simplex method for function minimization.’’ *The computer journal*, 7(4), pp. 308--313.

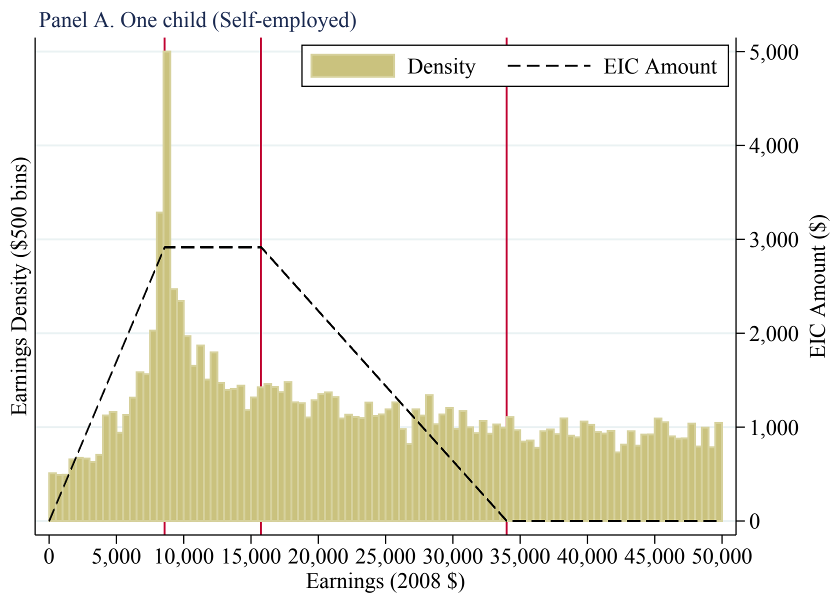
Ruud, Paul A. (1991). ‘‘Extensions of estimation methods using the em algorithm.’’ *Journal of Econometrics*, 49(3), pp. 305--341. doi:https://doi.org/10.1016/0304-4076(91)90001-T. URL https://www.sciencedirect.com/science/article/pii/030440769190001T.

Saez, Emmanuel (2010). ‘‘Do Taxpayers Bunch at Kink Points?’’ *AMERICAN ECONOMIC JOURNAL-ECONOMIC POLICY*, 2(3), pp. 180--212. doi:10.1257/pol.2.3.180.

Shanno, D. F. (1970). ‘‘Conditioning of quasi-newton methods for function minimization.’’ *Mathematics of Computation*, 24(111), pp. 647--656. URL http://www.jstor.org/stable/2004840.

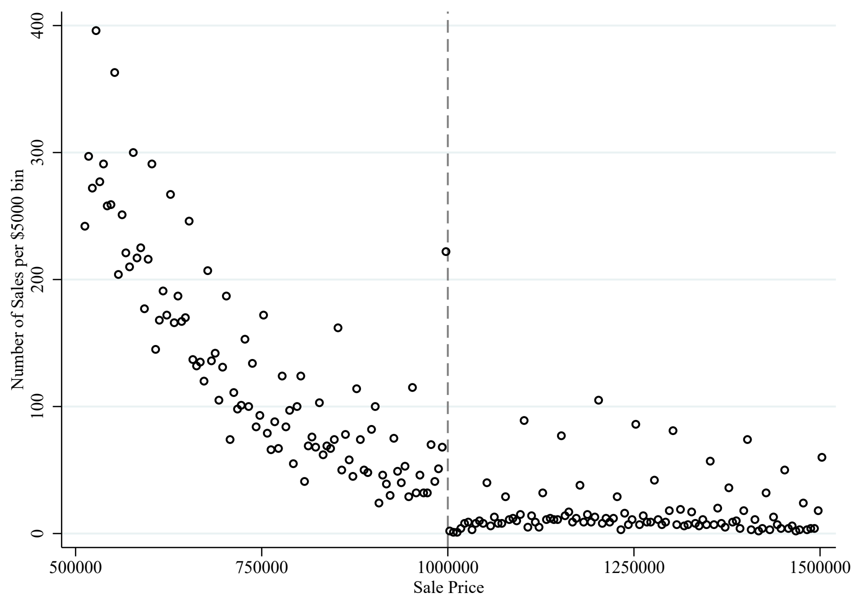
Figure 1: Lack of Bunching at Kinks

(a) Earnings Density Distributions and the EITC



Notes: This figure displays a histogram of earnings density (by $500 bins) for tax filers with one dependent child and who are self-employed. The EITC schedule is depicted by the black dashed line. When the black dashed line, changes slope, a kink point is formed. There are three kink points in this figure, which are depicted by red vertical lines. The literature commonly refers to the first kink point at $8,580 as an example of sharp bunching with no bunching apparent in the second or third kink points. However, even the first kink point is not a traditional example of sharp bunching given that the bunching mass is diffused around the kink point with the bins adjacent to the kink point having more mass than the counterfactual distribution would imply. Our model can account for this diffused bunching through optimizing frictions. This figure and data is similar to Saez (2010) where we also look at all years from 1995 to 2004 and inflate earnings to 2008 dollars using the IRS inflation parameters.

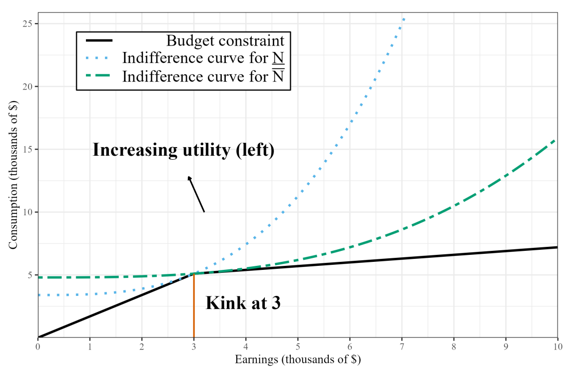
Figure 2: Mass in the Dominated Region after the Tax Notch



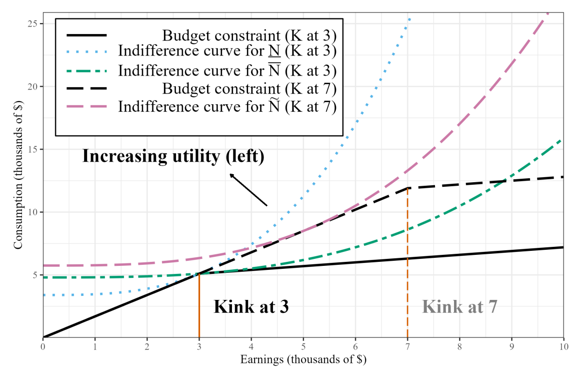
Notes: This figure is drawn from Figure 1 in Kopczuk and Munroe (2015) . In order to avoid a tax notch at $5000,000 where the tax rate changes from 1% to 1.425%, we exclude New York City from our analysis. We look at the remaining sales from New York state for 2004. The only tax notch these sales face are at $1,0000,000 where the tax rate changes from 0% to 1%. We also restrict to sales that arm's length transactions and one family year-round residence.

Figure 3: Convex kink with and without optimizing frictions

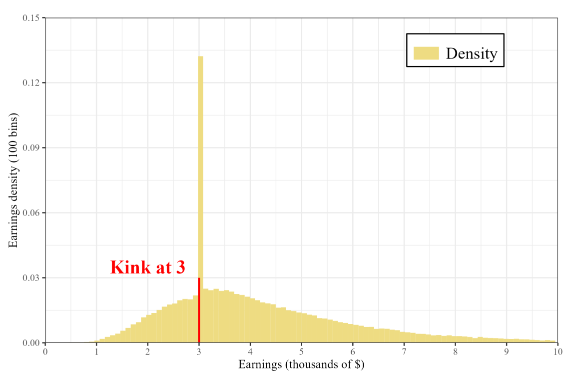
(a) Kink without Optimizing Frictions



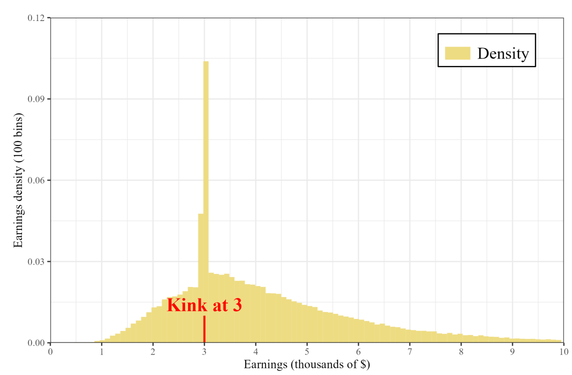
(b) Kink with Optimizing Frictions



(c) PDF of Kink without Optimizing Frictions



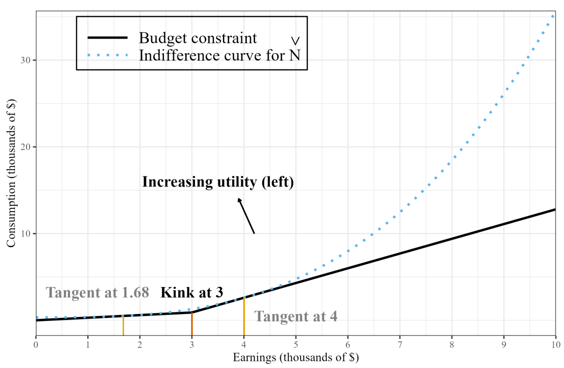
(d) PDF of Kink with Optimizing Frictions



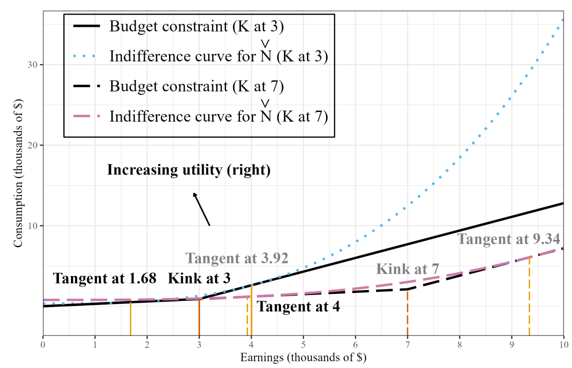
Notes: Panels A and B are related to convex kink points where panel A depicts two individuals without optimization frictions, while panel B depicts individual $\tilde{N}$ and how their decision varies depending on what optimization frictions they face. Panels C and D correspond to histograms of reporting that correspond to the optimizing behavior depicted in panels A and B, respectively.

Figure 4: Concave kink with and without optimizing frictions

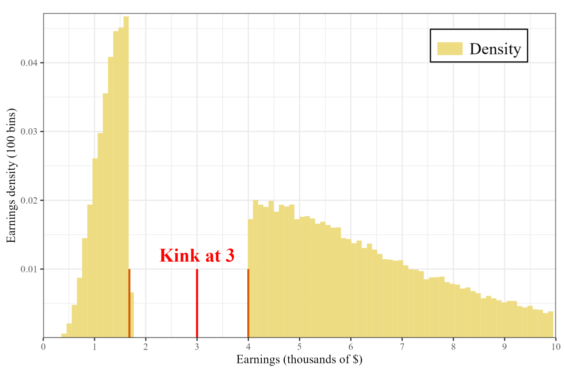
(a) Concave kink w/o frictions



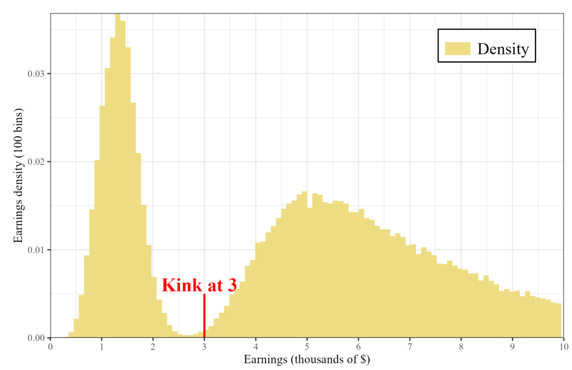
(b) Concave kink with frictions



(c) PDF of Concave kink w/o frictions



(d) PDF of Concave kink with frictions



Notes: Panels A and B are related to concave kink points where panel A depicts one individual without optimization frictions who is indifferent between reporting at the start and end of the zero reporting region. Panel B depicts an additional individual and how their decision varies depending on what optimization frictions they face. Panels C and D correspond to histograms of reporting that correspond to the optimizing behavior depicted in panels A and B, respectively.

Figure 5: Determinants that affect reporting at kinks

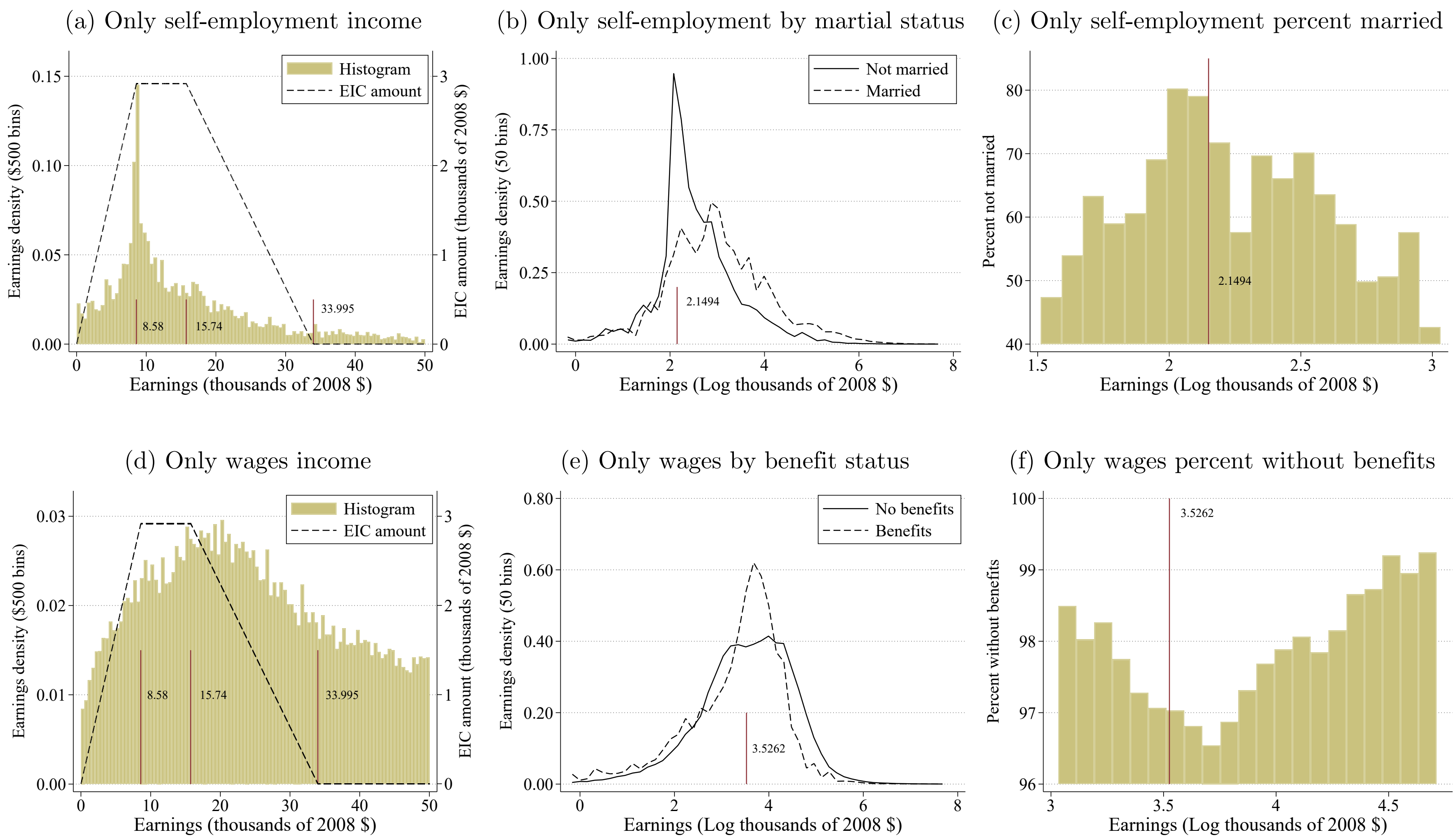
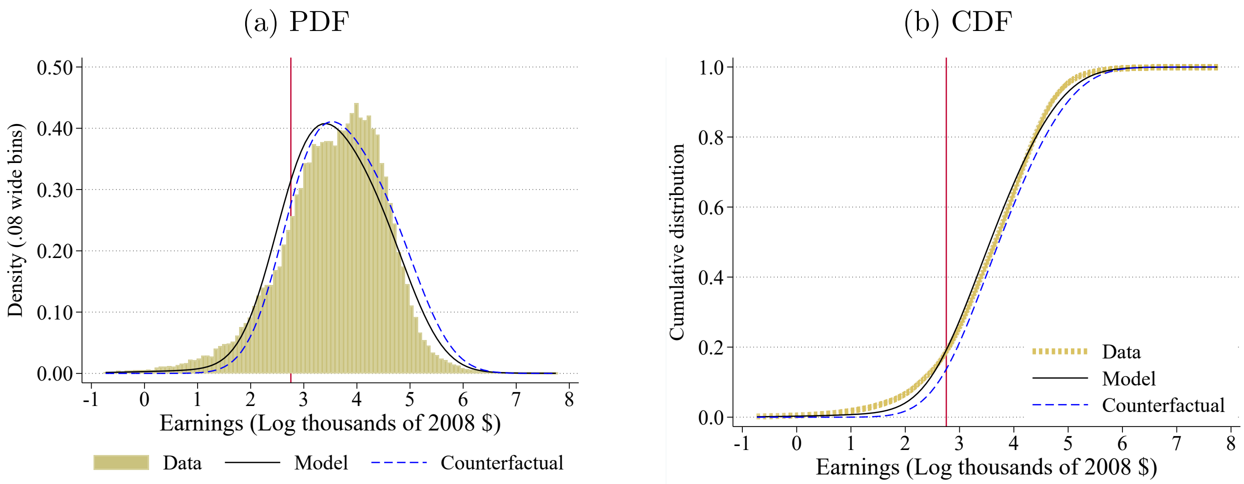


Figure 5 includes tax filers with one dependent child and who report only self-employment income and no wages (Figures 5a, 5b, 5c) or only wages and no self-employment income (Figures 5d, 5e, and 5f). There is diffuse bunching in the unconditional histogram of earnings at the $8,580 convex kink in Figure 5a but no missing mass at the $33,995 concave kink in Figure 5d. The black dashed line is the dollar amount of the EIC for reporting the income on the x-axis and the red vertical lines are two concave kinks created by the EITC at $8,580 and $15,740 and the one concave kink at $33,995. This figure and data is similar to Saez (2010) where we also look at all years from 1995 to 2004 and inflate earnings to 2008 $ using the IRS inflation parameters.

Figure 6: Wage earners, one child, $15,740 kink, many covariates



Note: Unobserved ability and knowledge are assumed to be uncorrelated so that ρ is constrained to zero. The data is the SOI PUFs (IRS) 1995-2004, wage earners, with one child and has 181.8 million weighted observations. These figures depict the second kink point of the EITC at $15,740. Panel A depicts the pdf of the data, our model, and the counterfactual distribution for this sample. Panel B depicts the cdf of the data, our model, and the counterfactual distribution for this sample. The model used does not incorporate any covariates.

A Theory appendix

A.1 Thresholds for convex notch

A.2 Indifference curves

A.2.1 Convex kink

The agent solves Equation (1a) and reports labor supply given by Equation (2). Resulting optimal consumption substitutes optimal labor supply NOT into the budget constraint in Equation (1b), because this is what the agent thinks they will be able to consume, but into the actual budget constraint that takes the actual kink and tax schedule. An $\underline{n}_i$ ability agent, indifferent between reporting at and below the kink, k_i^*, substitutes optimal income $\underline{y}_i \leq k_i^*$ into the budget constraint. In levels, that is

$$\underline{N} = \exp\left(\underline{n}\right) = \exp\left(k_i^* - \varepsilon s_0\right) \tag{16}$$

$$\underline{Y} = \exp\left(\underline{y}\right) = \exp\left(\varepsilon s_0 + \underline{n}\right) = \exp\left(k_i^*\right) \tag{17}$$

$$\underline{C} = I_0 + \exp\left(s_0\right)\exp\left(\underline{y}\right) = I_0 + \exp\left(s_0\right)\exp\left(k_i^*\right) \tag{18}$$

To solve for the agent's optimal utility, $\underline{U}$, we substitute $\underline{C}$ into Equation (1a). We can then recover the indifference curve for the agent with ability $\underline{n}_i$.

$$\underline{U} = I_0 + \left(\exp\left(s_0\right) - \frac{\exp\left(\varepsilon s_0\right)^{\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}\right)\exp\left(k_i^*\right) \tag{19}$$

$$\underline{C} = \underline{U} + \left(\frac{1}{1+\frac{1}{\varepsilon}}\right)\left(\frac{1}{\exp\left(k_i^* - \varepsilon s_0\right)^{\frac{1}{\varepsilon}}}\right)\left(\exp\left(y\right)\right)^{1+\frac{1}{\varepsilon}} \tag{20}$$

We repeat this process for an $\bar{n}_i$ ability agent, indifferent between reporting at and above the kink, k_i^*. The agent substitutes optimal income $\bar{y}_i \geq k_i^*$ into the budget constraint. We have

$$\overline{N} = \exp\left(\overline{n}\right) = \exp\left(k_i^* - \varepsilon s_1\right) \tag{21}$$

$$\overline{Y} = \exp\left(\overline{y}\right) = \exp\left(\varepsilon s_1 + \overline{n}\right) = \exp\left(k_i^*\right) \tag{22}$$

$$\overline{C} = I_0 + \exp(s_0)\exp(k_i^*) + \exp(s_1)(\exp(\overline{y}) - \exp(k_i^*)) = I_0 + \exp(s_0)\exp(k_i^*) \tag{23}$$

$$\overline{U} = I_0 + \left(\exp(s_0) - \frac{\exp(\varepsilon s_1)^{\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}\right)\exp(k_i^*) \tag{24}$$

$$\overline{C} = \overline{U} + \left(\frac{1}{1+\frac{1}{\varepsilon}}\right)\left(\frac{1}{\exp(k_i^* - \varepsilon s_1)^{\frac{1}{\varepsilon}}}\right)(\exp(y))^{1+\frac{1}{\varepsilon}} \tag{25}$$

A.2.2 Concave kink

The agent solves Equation (1a) and reports labor supply given by Equation (3). As in the convex case, resulting optimal consumption substitutes optimal labor supply NOT into the budget constraint in Equation (1b), but into the actual budget constraint that takes the actual kink and tax schedule.

There is only one threshold for a concave kink. An $\check{n}_i$ ability agent, indifferent between reporting above and below the kink, k_i^*, substitutes optimal income $\check{y}_i \le k_i^*$ into the budget constraint. We have

$$b(s_0, s_1, \varepsilon) = \ln\left(\frac{\exp(s_0) - \exp(s_1)}{\exp(s_0)^{\varepsilon+1} - \exp(s_1)^{\varepsilon+1}}\right) \tag{26}$$

$$\check{N} = \exp(\check{n}) = \exp(k_i^* + \ln(\varepsilon+1) + b(s_0, s_1, \varepsilon)) \tag{27}$$

$$\check{Y} = \exp(\check{y}) = \exp(\varepsilon s_0 + \hat{n}) = \exp(\varepsilon s_0 + k_i^* + \ln(\varepsilon+1) + b(s_0, s_1, \varepsilon)) \tag{28}$$

$$\check{C} = I_0 + \exp(s_0)\exp(\check{y}) = I_0 + \exp(s_0)\exp(\varepsilon s_0 + k_i^* + \ln(\varepsilon+1) + b(s_0, s_1, \varepsilon)) \tag{29}$$

We can solve for the agent's optimal utility, $\check{U}$, by substituting $\check{C}$ into Equation (1a) and recovering the indifference curve for the agent with ability $\check{n}_i$.

$$\check{U} = I_0 + \left(exp(s_0) - \frac{exp(\varepsilon s_0)^{\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}\right) exp(\varepsilon s_0 + k_i^* + \ln(\varepsilon+1) + b(s_0, s_1, \varepsilon)) \tag{30}$$

over

$$\check{C} = \check{U} + \left(\frac{1}{1+\frac{1}{\varepsilon}}\right)\left(\frac{1}{\exp\left(k_i^* + \ln\left(\varepsilon+1\right) + b\left(s_0, s_1, \varepsilon\right)\right)}\right)^{\frac{1}{\varepsilon}} \left(\exp\left(y\right)\right)^{1+\frac{1}{\varepsilon}} \tag{31}$$

End here - Ben
Note

$$C_i = \mathbb{I}\left(Y_i \leq K_i\right)\left[I_{i0} + \left(1-t_0\right)Y_i\right] + \mathbb{I}\left(Y_i > K_i\right)\left[I_{i1} + \left(1-t_1\right)\left(Y_i - K_i\right)\right] \tag{32}$$

$$y_i = \begin{cases} \varepsilon s_0 + n_i^* & \text{, if} \quad n_i^* < \underline{n}_i\left(k_i^*, \varepsilon, s_0\right) \\ k_i^* & \text{, if} \quad \underline{n}_i\left(k_i^*, \varepsilon, s_0\right) \leq n_i^* \leq \bar{n}_i\left(k_i^*, \varepsilon, s_1\right) \\ \varepsilon s_1 + n_i^* & \text{, if} \quad n_i^* > \bar{n}_i\left(k_i^*, \varepsilon, s_1\right). \end{cases} \tag{33}$$

A.3 Thresholds for concave notch

A.4 Non-parametric point identification for a concave notch

A.5 Defining a convex notch

Let $N_i^* = \bar{N}$

From equation (1a), the agent solves

$$C_k - \frac{\bar{N}}{1+\frac{1}{\varepsilon}}\left(\frac{Y_k}{\bar{N}}\right)^{1+\frac{1}{\varepsilon}} - \bar{C} + \frac{\bar{N}}{1+\frac{1}{\varepsilon}}\left(\frac{\bar{Y}}{\bar{N}}\right)^{1+\frac{1}{\varepsilon}} = 0 \tag{34}$$

and from equations (1b) and (2) the following equations are true by assumption.

$$\bar{Y} - \bar{N}\left(1-t_1\right)^{\varepsilon} = 0 \tag{35}$$

$$\bar{C} - I_0 - \left(1-t_0\right)K + \Delta - (1-t_1)(\bar{Y} - K) = 0 \tag{36}$$

$$C_k - I_0 - \left(1-t_0\right)K = 0 \tag{37}$$

$$Y_k - K = 0 \tag{38}$$

Substituting equation (35) into equation (36), we get

$$\bar{C} = I_0 + (1 - t_0) K - \Delta + (1 - t_1) \left(\bar{N} (1 - t_1)^{\varepsilon} - K\right) \tag{39}$$

Setting utility evaluated at $\bar{N}$ equal to utility specified by equation (35) results in:

$$I_0 + (1 - t_0) K - \frac{\bar{N}}{1 + \frac{1}{\varepsilon}} \left(\frac{K}{\bar{N}}\right)^{1+\frac{1}{\varepsilon}} - \bar{C} + \frac{\bar{N}}{1 + \frac{1}{\varepsilon}} \left(\frac{\bar{N} (1 - t_1)^{\varepsilon}}{\bar{N}}\right)^{1+\frac{1}{\varepsilon}} = 0 \tag{40}$$

Replacing $\bar{C}$ with (39) and simplifying, we arrive at a nonlinear expression in terms of $\bar{N}$:

$$-\left(\frac{K^{1+\frac{1}{\varepsilon}}}{1 + \frac{1}{\varepsilon}}\right) \bar{N}^{-\frac{1}{\varepsilon}} + \left(-(1 - t_1)^{\varepsilon+1} + \frac{(1 - t_1)^{\varepsilon+1}}{1 + \frac{1}{\varepsilon}}\right) \bar{N} + (1 - t_1) K + \Delta = 0 \tag{41}$$

We solve this nonlinear equation for $\bar{N}$ using the Newton-Raphson method (henceforth NR). NR solves the equation by better approximating the solution to a function following an initial guess of the root:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \tag{42}$$

NR iteratively updates the approximation to the solution through successive applications of equation (42), where

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{43}$$

and stops after a fixed point x of equation (43) is found relative to a given tolerance level: in our case 0. We define the left hand side of equation (41) equal to 0 as $f(x)$. Iterations are faster when we use the derivative of (41), which is

$$\left(\frac{1}{\varepsilon}\right) \left(\frac{K^{1+\frac{1}{\varepsilon}}}{1 + \frac{1}{\varepsilon}}\right) \bar{N}^{\frac{-1-\varepsilon}{\varepsilon}} - (1 - t_1)^{\varepsilon+1} + \frac{(1 - t_1)^{\varepsilon+1}}{1 + \frac{1}{\varepsilon}} + (1 - t_1) K + \Delta \tag{44}$$

A.6 Defining a concave notch

Let $N_i^* = \underline{N}$

From equation (1a), the agent solves

$$C_k - \frac{\underline{N}}{1 + \frac{1}{\varepsilon}} \left(\frac{Y_k}{\underline{N}}\right)^{1+\frac{1}{\varepsilon}} - \underline{C} + \frac{\underline{N}}{1 + \frac{1}{\varepsilon}} \left(\frac{\underline{Y}}{\underline{N}}\right)^{1+\frac{1}{\varepsilon}} = 0 \tag{45}$$

and from equations (1b) and (3) the following equations are true by assumption.

$$\underline{Y} - \underline{N}\,(1-t_0)^{\varepsilon} = 0 \tag{46}$$

$$\underline{C} - I_0 - (1-t_0)\,\underline{Y} = 0 \tag{47}$$

$$C_k - I_0 - (1-t_0)\,K + \Delta = 0 \tag{48}$$

$$Y_k - K = 0 \tag{49}$$

Substituting equation (46) into equation (47), we get

$$\underline{C} = I_0 + (1-t_0)\,(1-t_0)^{\varepsilon}\,\underline{N} = I_0 + (1-t_0)^{\varepsilon+1}\,\underline{N} \tag{50}$$

Setting utility evaluated at $\underline{N}$ equal to utility specified by equation (46) results in:

$$I_0 + (1-t_0)\,K - \Delta - \frac{\underline{N}}{1+\frac{1}{\varepsilon}}\left(\frac{K}{\underline{N}}\right)^{1+\frac{1}{\varepsilon}} - \underline{C} + \frac{\underline{N}}{1+\frac{1}{\varepsilon}}\left(\frac{\underline{N}\,(1-t_0)^{\varepsilon}}{\underline{N}}\right)^{1+\frac{1}{\varepsilon}} = 0 \tag{51}$$

Replacing $\underline{C}$ with (51) and simplifying, we arrive at a nonlinear expression in terms of $\underline{N}$:

$$[(1-t_0)\,K - \Delta] - \left(\frac{K^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}\right)\underline{N}^{-1/\varepsilon} + \underline{N}\left(\frac{(1-t_0)^{\varepsilon+1}}{1+\frac{1}{\varepsilon}} - (1-t_0)^{\varepsilon+1}\right) = 0 \tag{52}$$

We solve this nonlinear equation for $\underline{N}$ using NR, defining the left hand side of equation (52) equal to 0 as $f(x)$, through steps like those shown in A.5 in which the derivative is given by

$$[(1-t_0)\,K - \Delta] + \left(\frac{K^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}\right)\underline{N}^{\frac{-1-\varepsilon}{\varepsilon}} + \left(\frac{(1-t_0)^{\varepsilon+1}}{1+\frac{1}{\varepsilon}} - (1-t_0)^{\varepsilon+1}\right) \tag{53}$$

B Estimation appendix

B.1 Convex taxes and bivariate normal errors

B.1.1 Convex taxes augmented log-likelihood

This section provides the ingredients for the augmented log-likelihood given in Equation (12b). First, consider the probability of observing a data set generated by Equation (14). The probability that agent i has ability level n_i^* and location k_i^* that places them to the left of the kink with bivariate normal errors is denoted by Q_{i1}. The probability of observing earnings for agent i when they believe they will face the lower tax rate, s_0, is $f_{i1}Q_{i1}$:

$$f_{i1} = \frac{1}{\sigma_\nu}\phi\left(\frac{y_i - \varepsilon s_0 - X_i'\beta}{\sigma_\nu}\right) \text{ and } Q_{i1} = \Phi\left(\frac{k - \varepsilon s_0 - X_i'\beta + Z_i'\gamma}{\sigma_{\nu-\psi}}\right).$$

The probabilities of f_{i2} and Q_{i2} are defined similarly to f_{i1} and Q_{i1}. The probability that the agent finds it optimal to bunch at k_i^* and therefore report $y_i = k_i^*$ is $f_{i2}Q_{i2}$. Each of these components are defined by,

$$f_{i2} = \frac{1}{\sigma_\psi}\phi\left(\frac{y_i - k - Z_i'\gamma}{\sigma_\psi}\right) \text{ and } Q_{i2} = \Phi\left(\frac{k - \varepsilon s_1 - X_i'\beta + Z_i'\gamma}{\sigma_{\nu-\psi}}\right) - \Phi\left(\frac{k - \varepsilon s_0 - X_i'\beta + Z_i'\gamma}{\sigma_{\nu-\psi}}\right).$$

The probabilities of f_{i3} and Q_{i3} are defined similarly to f_{i1} and Q_{i1}. The probability of observing earnings for agent i when they optimize trying to face the higher tax rate, s_1, is $f_{i3}Q_{i3}$:

$$f_{i3} = \frac{1}{\sigma_\nu}\phi\left(\frac{y_i - \varepsilon s_1 - X_i'\beta}{\sigma_\nu}\right),\ Q_{i3} = 1 - \Phi\left(\frac{k - \varepsilon s_1 - X_i'\beta + Z_i'\gamma}{\sigma_{\nu-\psi}}\right).$$

B.1.2 Convex taxes truncated augmented log-likelihood

All of the ingredients for the truncated augmented log-likelihood from Equation (**??**) with a convex budget constraint are the same as in Section B.1.1 with the addition that we need to calculate F_ih $\forall h$, which is just the CDF of the PDF of the corresponding f_{ih}. With join-normal errors, these are given by

$$F_{i1} =$$

$$F_{i2} =$$

$$F_{i3} =$$

B.1.3 Convex taxes gradient for augmented log-likelihood

B.1.4 Convex taxes gradient for truncated augmented log-likelihood

B.2 Concave taxes and bivariate normal errors

B.2.1 Concave taxes augmented log-likelihood

B.2.2 Concave taxes truncated augmented log-likelihood

B.2.3 Convex taxes gradient for augmented log-likelihood

B.2.4 Convex taxes gradient for truncated augmented log-likelihood

B.3 Properties of Bivariate Normal Distribution

The random variables ν and ψ follow a joint bivariate normal distribution

$$\begin{pmatrix} \nu \\ \psi \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_\nu \\ \mu_\psi \end{pmatrix}, \begin{pmatrix} \sigma_\nu^2 & \rho_{\nu\psi}\sigma_\nu\sigma_\psi \\ \rho_{\nu\psi}\sigma_\nu\sigma_\psi & \sigma_\psi^2 \end{pmatrix}\right]$$

This implies that the errors for the selection equation $\nu - \psi$ have the following moments and properties

$$\begin{aligned} E[\nu] = \mu_\nu,\ E[\psi] = \mu_\psi, \qquad & COV[\nu, \nu-\psi] = \sigma_\nu^2 - \rho_{\nu\psi}\sigma_\nu\sigma_\psi, \\ E[\nu-\psi] = \mu_\nu - \mu_\psi, \qquad & COR(\nu, \nu-\psi) = \rho_{\nu,\nu-\psi}, \\ V[\nu-\psi] = \sigma_\nu^2 + \sigma_\psi^2 - 2\rho_{\nu\psi}\sigma_\nu\sigma_\psi = \sigma_{\nu-\psi}^2, \qquad & \rho_{\nu,\nu-\psi} = \frac{\sigma_\nu - \rho_{\nu\psi}\sigma_\psi}{\sqrt{\sigma_\nu^2 + \sigma_\psi^2 - 2\rho_{\nu\psi}\sigma_\nu\sigma_\psi}} \end{aligned}$$

Therefore, the random variables ν and $\nu - \psi$ follow the joint bivariate normal distribution

$$\begin{pmatrix} \nu \\ \nu-\psi \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_\nu \\ \mu_\nu - \mu_\psi \end{pmatrix}, \begin{pmatrix} \sigma_\nu^2 & \rho_{\nu,\nu-\psi}\sigma_\nu\sigma_{\nu-\psi} \\ \rho_{\nu,\nu-\psi}\sigma_\nu\sigma_{\nu-\psi} & \sigma_{\nu-\psi}^2 \end{pmatrix}\right]$$

It is useful to note the following conditional distributions

$$(\nu \mid \nu-\psi) \sim N\left[\mu_\nu + \frac{\sigma_\nu}{\sigma_{\nu-\psi}}\rho_{\nu,\nu-\psi}(\nu-\psi-\mu_\nu+\mu_\psi), \left(1-\rho_{\nu,\nu-\psi}^2\right)\sigma_\nu^2\right]$$

and

$$(\nu-\psi \mid \nu) \sim N\left[\mu_\nu - \mu_\psi + \frac{\sigma_{\nu-\psi}}{\sigma_\nu}\rho_{\nu,\nu-\psi}(\nu-\mu_\nu), \left(1-\rho_{\nu,\nu-\psi}^2\right)\sigma_{\nu-\psi}^2\right].$$

The marginal distributions of ν and ψ are given by

$$(\nu) \sim N\left[\mu_\nu, \sigma_\nu^2\right],\ (\nu-\psi) \sim N\left[\mu_\nu - \mu_\psi, \sigma_{\nu-\psi}^2\right].$$

C Gradient appendix