

NBER WORKING PAPER SERIES

A BIG DATA APPROACH TO OPTIMAL SALES TAXATION

Christian Baker
Jeremy Bejarano
Richard W. Evans
Kenneth L. Judd
Kerk L. Phillips

Working Paper 20130
<http://www.nber.org/papers/w20130>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2014

Thanks to Laurence Kotlikoff, Kent Smetters, James McDonald, Frank Caliendo, and Aspen Gorry for helpful comments and insights. We are also grateful for comments from conference participants at the BYU Computational Public Economics Conference 2012 and the Hoover Institution Summer 2013 Computational Economics Workshop, and support from the BYU Macroeconomics and Computational Laboratory. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2014 by Christian Baker, Jeremy Bejarano, Richard W. Evans, Kenneth L. Judd, and Kerk L. Phillips. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Big Data Approach to Optimal Sales Taxation

Christian Baker, Jeremy Bejarano, Richard W. Evans, Kenneth L. Judd, and Kerk L. Phillips
NBER Working Paper No. 20130

May 2014

JEL No. C1,H2,H21

ABSTRACT

We characterize and demonstrate a solution method for an optimal commodity (sales) tax problem consisting of multiple goods, heterogeneous agents, and a nonconvex policy maker optimization problem. Our approach allows for more dimensions of heterogeneity than has been previously possible, incorporates potential model uncertainty and policy objective uncertainty, and relaxes some of the assumptions in the previous literature that were necessary to generate a convex optimization problem for the policy maker. Our solution technique involves creating a large database of optimal responses by different individuals for different policy parameters and using "Big Data" techniques to compute policy maker objective values over these individuals. We calibrate our model to the United States and test the effects of a differentiated optimal commodity tax versus a flat tax and the effect of exempting a broad class of goods (services) from commodity taxation. We find that only a potentially small amount of tax revenue is lost for a given societal welfare level by departing from an optimal differentiated sales tax schedule to a uniform flat tax and that there is only a small loss in revenue from exempting a class of goods such as services in the United States.

Christian Baker
Department of Economics
cbaker8128@gmail.com

Jeremy Bejarano
University of Chicago
jmbejara@gmail.com

Richard W. Evans
Brigham Young University
Department of Economics
167 FOB
Provo, Utah 84602
revans@byu.edu

Kenneth L. Judd
Hoover Institution
Stanford University
Stanford, CA 94305-6010
and NBER
JUDD@HOOVER.STANFORD.EDU

Kerk L. Phillips
Department of Economics
Building 130 FOB
Brigham Young University
Provo, UT 84602
kerk_phillips@byu.edu

1 Introduction

In this study, we characterize and demonstrate a solution method for an optimal commodity (sales) tax problem consisting of multiple goods, heterogeneous agents, and a nonconvex policy maker optimization problem. The contribution of our approach is to allow for more dimensions of heterogeneity than has been previously possible, to incorporate potential model uncertainty and policy objective uncertainty, and to relax some of the assumptions in the previous literature that were necessary to generate a convex optimization problem for the policy maker. Our solution technique involves creating a large database of optimal responses by different individuals for different policy parameters and using “big data” techniques to compute policy maker objective values over these individuals.

We calibrate our model to the United States and calculate the difference between an optimal differentiated sales tax schedule versus an optimal uniform (flat) tax schedule. We also compute the revenue loss for a given welfare level of having a tax-exempted commodity class, such as is the case in nearly all countries. We find that only a potentially small amount of tax revenue is lost for a given societal welfare level by departing from an optimal differentiated sales tax schedule to a uniform flat tax. We also find a small loss in revenue from exempting a class of goods such as services in the United States.

[Ramsey \(1927\)](#) studied a representative agent economy, with multiple goods, and a policy maker with a total tax revenue constraint and linear commodity tax instruments (no lump sum taxation). The main findings were that tax rates on commodities should be inversely related to the representative consumer’s price elasticity of demand and should, therefore, vary across goods. [Diamond and Mirrlees \(1971a,b\)](#) and [Deaton \(1977\)](#) extend the Ramsey approach by studying optimal commodity taxation with multiple goods, a revenue constraint and multiple consumers who have identical preferences and heterogeneous productivities (incomes). However, they make strong assumptions on the demand functions in order attain aggregation results and to keep the policy maker’s optimization problem convex. These three studies highlight the

tradeoff between equalizing incomes and maximizing social welfare. However, [Deaton \(1977, p.310\)](#) concedes:

The result rests on strong simplifying assumptions in order to avoid the complexity of the general case; in particular, consumer behaviour has been restricted by the use of linear Engel curves and by permitting only very limited substitution between commodities. Relaxing either of these could alter quite fundamentally the nature of the empirical results.

[Balcer et al. \(1983\)](#) add minimum consumption levels to the household utility function, but use a simple enough functional form to still allow for aggregation and a convex optimization problem for the policy maker. Similar to [Deaton \(1977\)](#), [Slemrod and Bakija \(2008, chapters 6 and 7\)](#), [Gruber \(2011, chapters 18, 20, and 25\)](#), and [Garner \(2005\)](#), they find that commodity taxes will vary across goods according to budget shares but that their ability to redistribute income is somewhat limited. In addition, [Balcer et al. \(1983\)](#) find that the optimal differentiated commodity tax schedule only provides a small revenue increase for a particular welfare level relative to an optimal uniform (flat) tax.

In this study, we extend the [Balcer et al. \(1983\)](#) approach by using a more general utility function in which aggregation does not hold, thereby rendering the government's optimization problem nonconvex. By doing so, we are testing the observation of [Deaton \(1977, p. 310\)](#) that relaxing these simplifying assumptions "could alter quite fundamentally the nature of the empirical results." Further, the nonconvexity of the government's optimization problem requires a different set of computational tools in order to solve for the optimal sales tax schedule. Our solution method is scalable, parallelizable, customizable, and uses "big data" techniques.

Big data refers to any repository of data that is either large enough or complex enough that distributed and parallel input and output approaches must be used (see [White, 2012, p. 3](#)). [Einav and Levin \(2013\)](#) discuss the new opportunities in economics using big data, although they focus primarily on searching for important patterns in existing datasets. Our paper is a different application of the big data

approach in that we use it as a solution method to theoretical models rather than an empirical method. Our solution method is ideally suited for dealing with models with nonconvex optimization problems, a large degree of heterogeneity, rich policy structure, potential model uncertainty, and potential policy objective uncertainty.

Our big data approach involves generating a large database of optimal agent behavior for an efficiently chosen sample of points in both the space of agent types and the space of tax policies. We then use big data techniques to access this database in order to calculate social welfare and total government revenue for each point in the tax policy space. Next, we eliminate all tax policies that are strictly dominated by any other policy in terms of both social welfare and total tax revenue. This leaves us with a tax policy frontier in terms of social welfare and total tax revenue. The last step is to iteratively refine our search for optimal tax policies around the points remaining on the frontier and repeat the deletion step of strictly dominated policies. The result is a close approximation of the continuous function of tax policies that traces out the frontier of optimal social welfare and tax revenue possibilities.

Much of the current literature on optimal commodity taxation focuses on estimating price elasticities for particular classes of consumption goods. [Kaplow \(2010\)](#), [Saez \(2002\)](#), and [Naito \(1999\)](#) find that taxes on commodities should vary, not only with their elasticities of demand, but also with their complementarity with leisure. Another strand of the literature, exemplified by [Einav et al. \(2014\)](#), [Ellison and Ellison \(2009\)](#), [Ballard and Lee \(2007\)](#), [Scanlan \(2007\)](#), [Alm and Melnik \(2005\)](#), and [Goolsbee \(2000a,b\)](#), focuses on the effects and lack of uniformity of the sales taxation of goods purchased on the internet. [Bruce et al. \(2009\)](#) estimate a loss of over \$10 billion in sales tax revenue in 46 U.S. states and in the District of Columbia from inconsistent or nonexistent sales taxation on internet commerce, particularly with respect to interstate internet commerce.

The final policy experiment we perform in this study relates to the internet revenue loss studies previously mentioned. We simulate the potential tax revenue loss of exempting a category of consumption goods. [Diamond and Mirrlees \(1971a\)](#) suggest the inclusion of an untaxed sector as an interesting extension. [Einav et al. \(2014\)](#)

study this question with respect to internet commerce by empirically estimating the price elasticity of internet sales consumers to price surprises. But we know of no other studies estimating the loss from exempting a commodity group as large as the services sector in the United States (e.g., legal fees, consulting, insurance, most health care.) This question is especially important in the United States, not just in terms of welfare, because services make up just under 50 percent of U.S. GDP.

The paper proceeds as follows. Section 2 presents our model of household optimization and the policy maker's tax problem. Section 3 describes our solution method, and Section 4 presents the results of our experiments. Section 5 concludes.

2 Model

Our sales tax model is one in which heterogeneous consumers must make a static decision of how much to consume among a variety of goods available to them. We abstract from the labor decision of these individuals by assuming an inelastic labor supply. We also abstract from any ability of consumers to intertemporally substitute consumption across time using savings. The model is partial equilibrium because we do not model the production side of the economy, and consumers take the prices of goods as given.

The invariance of goods prices and the wage distribution to tax policy changes in our model is a potential issue. One solution is to assume a small open economy. However, many of the goods in the United States are nontradable and labor is often thought of as not being very mobile. So letting prices be exogenous is a potential weakness.

Consumers also face a given schedule of sales taxes across these goods. In the tradition of much of the commodity tax literature, we restrict the policy maker's tax instruments to be linear tax rates that can be differentiated according to consumption good type. This is in contrast to the Mirrleesian approach of solving for a general tax function of all observables. The policy maker's problem is to choose a tax schedule that maximizes some social welfare criterion.

Hines (2007) and Gruber (2011) document that the United States only has a few goods on which it collects excise taxes, that the U.S. collects the smallest percent of tax revenue through sales or consumption taxes among developed countries, and that U.S. sales taxes are predominantly applied heterogeneously across U.S. states. In this respect, our exercise in this paper is hypothetical in that no uniform federal sales tax schedule exists in the United States. However, the qualitative results can be applied appropriately to inform sales tax policy at the state level in the U.S. as well as potential tax reform questions at the national level.

2.1 Household problem

Let the economy be characterized by a measure of individuals characterized by type $\theta = (\eta, w)$ and I different consumption goods c_i , where $i = 1, 2, \dots, I$. Define total consumption by an individual C by the constant elasticity of substitution (CES) aggregator,¹

$$C \equiv \left(\sum_{i=1}^I \alpha_i (c_i - \bar{c}_i)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \quad \forall \eta \geq 1 \quad (1)$$

where $\eta \geq 1$ is the elasticity of substitution among all of the consumption goods, $\alpha_i \in [0, 1]$ is the weight on the consumption of each type of good with $\sum_i \alpha_i = 1$, and $\bar{c}_i \geq 0$ is a minimum level of consumption for each type of good. The constant relative risk aversion (CRRA) utility function for the individual with CES preferences over total consumption is the following,

$$u(C) = \frac{C^{1-\gamma} - 1}{1-\gamma} \quad \forall \gamma \geq 1 \quad (2)$$

where total consumption of an individual C is defined in (1) and $\gamma \geq 1$ is the coefficient of relative risk aversion.

Let the price of consumption good i be determined by the perfectly competitive equilibrium assumption of marginal cost pricing with the additional sales tax perfectly passed on to consumers of good i . If mc_i is the marginal cost of producing good i

¹See Armington (1969) and Dixit and Stiglitz (1977).

and τ_i is the sales tax rate on good i , then competitive equilibrium implies that the price of each consumption good p_i is given by.

$$p_i = (1 + \tau_i)mc_i \quad (3)$$

We can normalize the marginal cost of each good to unity $mc_i = 1$ for all i by simply changing the units of each good. So the price of each good can be simplified to its normalized version.

$$p_i = 1 + \tau_i \quad (4)$$

Given a nominal wage of w , the household budget constraint is,

$$\sum_{i=1}^I (1 + \tau_i)c_i \leq w \quad (5)$$

The household's problem is to choose a consumption basket $\{c_i\}_{i=1}^I$ to maximize utility (2) subject to the budget constraint (5). In this example, we will allow consumers to be heterogeneous in terms of their elasticity of substitution η among different consumption goods and in terms of income w . So $\theta = (\eta, w) \in \Theta = [1, \infty) \times (0, \infty)$. Let the joint density over consumer types in the economy be $f(\eta, w) = f(\eta)f(w)$, where $\eta \sim U[\eta_{min}, \eta_{max}]$ and $w \sim GG(a, b, m)$, where U is the uniform distribution and GG is the three-parameter generalized gamma distribution.²

$$(GG) : \quad f(y; a, b, m) = \frac{m}{b^a \Gamma(\frac{a}{m})} y^{a-1} e^{-\left(\frac{y}{b}\right)^m} \quad y \in [0, \infty), \quad a, b, m > 0 \quad (6)$$

Now we can write the consumer's optimization problem in terms of vectors of variables,

$$\max_{\mathbf{c}} u(\mathbf{c}; \eta, w, \boldsymbol{\tau}) \quad \text{s.t.} \quad w \geq \sum_{i=1}^i (1 + \tau_i)c_i \quad \text{and} \quad c_i \geq \bar{c}_i \quad \forall i \quad (7)$$

²McDonald (1984) started a long literature showing that the generalized beta (GB) family of distributions with three parameters or more fits the income distribution better than the two-parameter GB family distributions, such as the log normal. In particular, the GB distributions with three parameters or more capture well fat right tails that are characteristic of the income distribution. See also McDonald et al. (2013).

where $\mathbf{c} = \{c_i\}_{i=1}^I$ and $\boldsymbol{\tau} = \{\tau_i\}_{i=1}^I$. If the budget constraint binds and $c_i \geq \bar{c}_i$ for all i , the solution to (7) can be reduced to $I - 1$ Euler equations,

$$\alpha_i (c_i - \bar{c}_i)^{-\frac{1}{\eta}} = \alpha_I \frac{1 + \tau_i}{1 + \tau_I} (c_I - \bar{c}_I)^{-\frac{1}{\eta}} \quad \text{for } i = \{1, 2, \dots, I - 1\} \quad (8)$$

where w and all the τ_i are introduced into (8) because good c_I is substituted out of the utility function using the budget constraint. The solution to this problem is individual consumption functions $c_i(\eta, w, \boldsymbol{\tau})$ that are a functions of consumer type (η, w) and tax rates $\boldsymbol{\tau}$. This household has equilibrium utility $u(\mathbf{c}(\eta, w, \boldsymbol{\tau}))$ and total taxes paid of $r(\eta, w, \boldsymbol{\tau}) = \sum_{i=1}^I \tau_i c_i(\eta, w, \boldsymbol{\tau})$.

However, it is also the case that both the budget constraint and $c_i \geq \bar{c}_i$ are not satisfied for some sales tax schedules $\boldsymbol{\tau}$.

$$\exists \boldsymbol{\tau} : \quad w < \sum_{i=1}^I (1 + \tau_i) \bar{c}_i \quad (9)$$

We want to examine the effects of a wide range of tax policies τ , but we do not want to worry about redistribution issues or have individual utilities with consumption at or near \bar{c}_i dominate any social welfare function in the policy maker's problem.

Balcer et al. (1983) address this same issue in their model by transferring revenue (collected from taxes) to the poorest individuals so that they can meet the minimum consumption constraint. However, this method still poses the problem of how far above \bar{c}_i should the transfer allow the poorest to consume. The chosen level still affects their utility and their corresponding weight in a social welfare function. To avoid this problem, we provide a small functional variation to the total individual consumption aggregator (1) and utility function (2) so that utility is globally defined and concave, even for individual consumption levels at or below the minimum $c_i \leq \bar{c}_i + \varepsilon$. However, the purpose of this functional adjustment is to cause as little distortion as possible while providing a remedy to a computational issue. Appendix A-1 details our adjustment.

2.2 Policy maker’s problem

The policy maker’s problem is to choose a schedule of sales tax rates on the different consumption goods $\boldsymbol{\tau} = \{\tau_1, \tau_2, \dots, \tau_I\}$ in order to maximize total utility in the economy subject to total tax revenues being greater than or equal to some exogenously determined amount \bar{R} ,

$$\begin{aligned} \max_{\boldsymbol{\tau}} \int_{\eta} \int_w f(\eta) f(w) u(\mathbf{c}(\eta, w, \boldsymbol{\tau})) d\eta dw \\ \text{s.t.} \int_{\eta} \int_w f(\eta) f(w) \sum_{i=1}^I \tau_i c_i(\eta, w, \boldsymbol{\tau}) d\eta dw \geq \bar{R} \end{aligned} \tag{10}$$

This is equivalent to a specific weighting of both total utility and total revenue in the policy maker’s objective function. However, this approach can accommodate any well defined objective function with any weighting scheme.³

The optimization problem (10) is not convex because the individual demand functions cannot be aggregated. This is due to the general specification of the CES total consumption aggregator (1) as well as the heterogeneity across household types (η, w) . The solution to the policy maker’s problem is a tax schedule $\boldsymbol{\tau}$ will be a function of the distribution of consumer types in the economy $\boldsymbol{\tau}(f(\eta, w))$. However, finding this solution to the nonconvex optimization problem (10) requires a new solution technique.

2.3 Calibration

For the experiment that follows, we calibrate our sales tax model to fit U.S. data as well as some parameter values taken from other studies. Table 4 summarizes the calibration. The first dimension of heterogeneity among the households in our model is the elasticity of substitution among consumption goods η . We calibrate the distribution of the different elasticities of substitution across individuals in our model to be uniformly distributed $\eta \sim U[3.7, 4.5]$. This is consistent with an estimated

³Notice that policy maker’s objective function (10) implies that each individual in the economy is weighted equally because we use $f(\eta, w)$ as the weighting function. Saez and Stantcheva (2013) derive endogenous welfare weights that are a function of society’s preferences for redistribution.

average price over marginal cost markup in the U.S of about 1.32 or 32%, and markup range from 1.25 to 1.37. [Christopoulou and Vermeulen \(2012\)](#) estimate the price over marginal cost markup in the U.S. to have been about 1.32 or 32% in the 1990s. In DSGE models with imperfect competition, the equilibrium markup is $\eta/(\eta - 1)$, a function of the constant elasticity of substitution among differentiated consumption goods.

Table 1: Calibration parameters, matching targets, and sources

Parameter	Value/Distribution	Matching target
η	$U[3.7, 4.5]$	1.32 average U.S. markup
w	$GG(a, b, m)$, $a = 1.67, b = 20510, m = 0.74$	CPS 2011 household income bracket count data
γ	2	Estimates of CRRA between 2 and 10

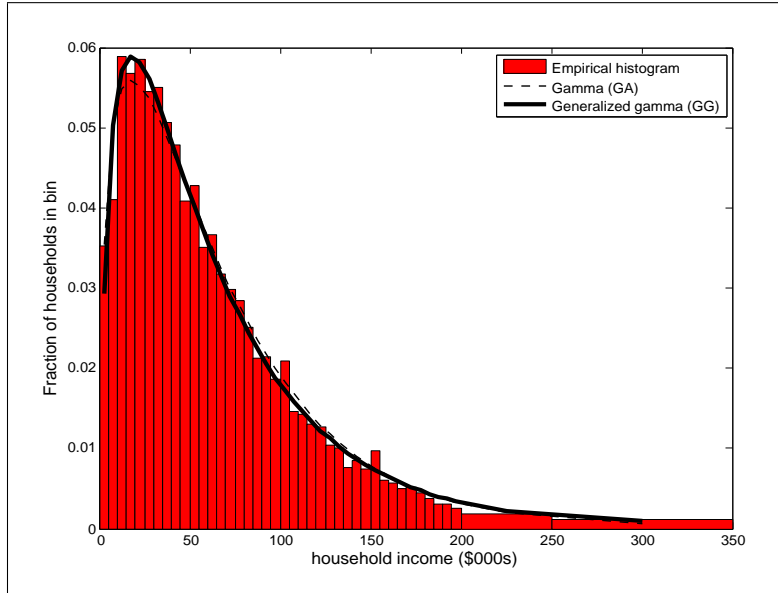
Average markup estimates for η come from [Christopoulou and Vermeulen \(2012\)](#). Our GG approximation of the distribution of U.S. household income was estimated by generalized method of moments following [McDonald et al. \(2013\)](#) and uses Current Population Survey income bracket count data from [Current Population Survey \(2012, Table HINC-01\)](#). For estimates of the constant coefficient of relative risk aversion, see [Mankiw and Zeldes \(1991\)](#), [Blake \(1996\)](#), [Campbell \(1996\)](#), [Kocherlakota \(1996\)](#), and [Mehra and Prescott \(1985\)](#).

For the distribution of income w in our model we follow [McDonald et al. \(2013\)](#) and use the simplest distribution from the generalized beta (GB) family that captures well the fat right tail of the income distribution. We use 42 moments from the U.S. household income distribution of 2011 reported in [Current Population Survey \(2012, Table HINC-01\)](#). We use generalized method of moments to test a number of distributions from the GB family, and the three-parameter generalized gamma (GG) distribution had the best fit. [Table 4](#) shows the estimated values for parameters a , b , and m of the GG distribution, and [Figure 1](#) shows how the estimated GG distribution fits the empirical data.⁴

We use the Consumer Expenditure Survey (CEX) 2011 broad category summary tables data to calibrate the expenditure shares α_i in the model, as well as the minimum consumptions \bar{c}_i . We aggregate expenditures into eight categories ($I = 8$):

⁴The Technical Appendix details how we estimated the generalized gamma (GG) distribution, how well it fit the moments, and how it outperformed other distributions.

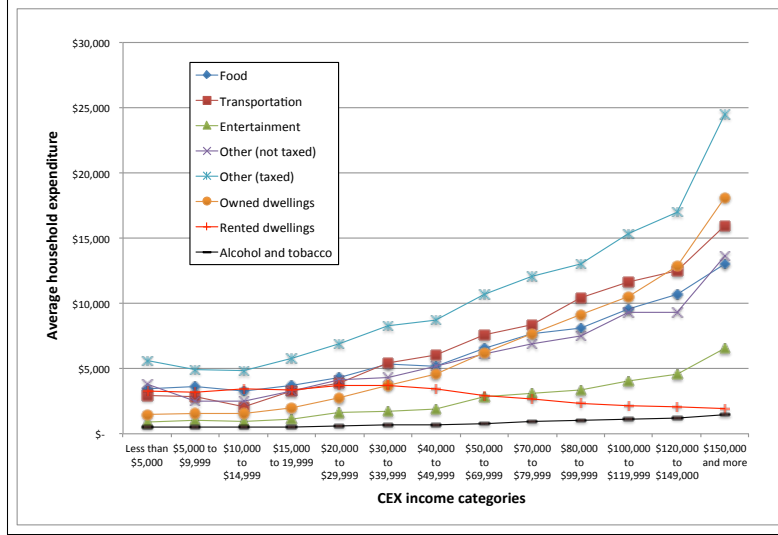
Figure 1: Histogram of U.S. household income and estimated generalized gamma (GG) distribution



food, transportation, entertainment, owned dwellings, rented dwellings, alcohol and tobacco, other (taxed), and other (not taxed). We divided expenditures into eight categories for the following two reasons. First, the more consumption categories increases the computational requirements for solving the model. Second, we wanted to separate out some of the key consumption categories that appear in tax policy debates. In particular, we will focus one of our experiments on the commodity category of “other (not taxed)”. This category includes many of the services that have no sales tax in the United States, such as personal services, vehicle finance charges, vehicle insurance, health care, and education. A more complete description of the consumption categories and what they include is in Appendix [A-2](#).

Figure [2](#) shows average total expenditure by household income category on each of the eight categories of consumption goods. One problem with the measure of income on the x -axis of Figure [2](#) is that it does not match up well with the concept of income in our model. Our static model includes no savings decisions. However, income in the CEX is used for both consumption and savings. For this reason, we compute total

Figure 2: Consumption expenditure by household income category of eight goods



expenditure by households in a particular income category as our empirical analog of income in our model. Figure 3 shows the expenditures from Figure 2 as a percent of average total expenditure by households in a particular income category.

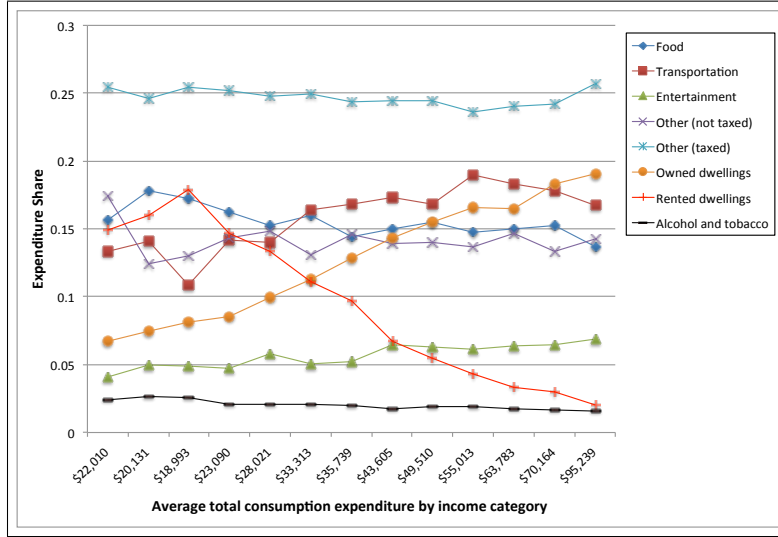
We estimate the minimum consumption of the poorest individuals \bar{c}_i and the consumption expenditure share of the wealthiest individuals α_i as the coefficients from the following regression,

$$c_{i,j} = \bar{c}_i + \alpha_i \text{AVGTOTEXP}_j + \varepsilon_{i,j} \quad \forall i, j \quad (11)$$

where $c_{i,j}$ is the average consumption of the i th good by the average household in the j th income category, AVGTOTEXP_j is the average total consumption expenditure by households in income category j , and $\varepsilon_{i,j}$ is a zero-mean, i.i.d. random variable.

Table 2 presents the estimated values for \bar{c}_i and α_i obtained by running the OLS regression in (11) for each of the eight consumption goods categories. Because negative values for minimum consumption levels \bar{c}_i and expenditure shares α_i do not make sense, we adjust our calibration of the minimum consumption of goods that are projected to be negative for $\text{AVGTOTEXP}_j = 0$ to be $\bar{c}_i = 0$. This adjustment was made for the four categories of Transportation, Entertainment, Other (taxed), and

Figure 3: Expenditure share by average total household consumption



Owned dwellings. The adjustment of $\alpha_i = 0.01$ for any initially estimated $\alpha_i < 0$ only had to be implemented for the one consumption category of Alcohol and tobacco. For our calibration of \bar{c}_i and α_i , we use the adjusted values in the rightmost two columns of Table 2.⁵

Because the coefficient of relative risk aversion γ does not change the household's optimal consumption decision \mathbf{c} , we did not include it as a dimension of heterogeneity across households. However, variation in this parameter does affect the optimal tax policy τ because it affects the utility levels of households, and total utility is one of the inputs to the government objective. We choose the parameter value $\gamma = 2$ for all households, which is a common value in within the range of 2 to 10 found in the literature.⁶

⁵Note from Figure 2 that a linear specification as in (11) is likely not fully appropriate. However, a linear specification is the only one that allows us to estimate, in a consistent manner, the consumption minimum \bar{c}_i and the expenditure share α_i for each consumption category i . If we did a log specification for $AVGTOTEXP_j$, we would not be able to estimate \bar{c}_i by taking the predicted value of $c_{i,j}$ at $AVGTOTEXP_j = 0$. We would also not be able to estimate the consumption share of the wealthiest individuals α_i by letting $AVGTOTEXP_j$ go to infinity because the share would necessarily go to zero. The same problems exist for a quadratic specification of (11). For this reason, we use the linear specification to calibrate our model.

⁶See Mankiw and Zeldes (1991), Blake (1996), Campbell (1996), Kocherlakota (1996), and Mehra and Prescott (1985).

Table 2: Estimated \bar{c}_i and α_i

Consumption category (i)	initial \bar{c}_i	initial α_i	adjusted \bar{c}_i	adjusted α_i	Implied demand elasticity ranking (lowest to highest) ^a
Food	\$691	0.135	\$691	0.135	3
Transportation	-\$957	0.189	\$0	0.189	6
Entertainment	-\$605	0.074	\$0	0.074	5
Other (not taxed)	\$29	0.140	\$29	0.140	4
Other (taxed)	-\$123	0.249	\$0	0.249	8
Owned dwellings	-\$3,269	0.223	\$0	0.223	7
Rented dwellings	\$4,012	-0.025	\$4,012	0.010	1
Alcohol and tobacco	\$222	0.014	\$222	0.014	2

Source: Consumer Expenditure Survey 2011 Broad Summary Table.

^a Demand elasticities are inversely related to $\bar{c}_i \alpha_i^{-\eta}$. We calculated ranking for average $\eta = 4.1$. See Technical Appendix for derivation.

3 Solution Method

Our big data approach to solving for the optimal tax policy $\tau(f(\theta))$ is an iterative process. We first specify the type space of households, the tax policy space, and the type space distribution. We then create a large database containing a diverse collection of household responses to a diverse collection of tax policies. Using the distribution for the type space, we compute societal utility and total revenue for each tax policy in the sample. We then strategically add new tax policies to the database near the area of interest and repeat the process to find societal utility and total revenue for these new policies. We can continue to add new policies to the database until a sufficient level of refinement is achieved.

We define the type space to be $\Theta = [3.7, 4.5] \times [8000, 500000]$, where a point $\theta \in \Theta$, $\theta = (\eta, w)$ describes a household. We choose a set of M equidistributed points $\{\theta_m\}_{m=1}^M$ to represent a simulated sample of households.⁷ For our eight good model, the tax policy space is defined to be $\mathbf{T} = (0, 1)^8$, where a point $\tau \in \mathbf{T}$, $\tau = (\tau_1, \tau_2, \dots, \tau_8)$ describes a potential tax policy. We choose a set of N equidistributed points $\{\tau_n\}_{n=1}^N$ to represent a simulated approximation of all possible tax policies. In order to ensure that the simulated households θ_m and policies τ_n are representative

⁷See Technical Appendix T-2 for more information on equidistributed sequences and quasi-Monte Carlo integration.

of the space, M and N must be large. Note that more sample households and tax policies can be added to the database at a later time if further refinement is desired.

We now consider the set $\{\boldsymbol{\tau}_n\}_{n=1}^N \times \{\boldsymbol{\theta}_m\}_{m=1}^M$. This represents the N tax policies that may be faced by each of the M household types. We must solve for the household optimal choice $\mathbf{c}(\boldsymbol{\theta}, \boldsymbol{\tau})$ for every possible combination of $\boldsymbol{\tau}_n$ and $\boldsymbol{\theta}_m$, which requires solving a total of $M \times N$ optimization problems. We use parallel processing to divide these independent optimization problems evenly across nodes on supercomputer. We compute the optimal household consumption of each good $\mathbf{c}(\boldsymbol{\theta}_m, \boldsymbol{\tau}_n)$. From this we get household utility $u(\mathbf{c}(\boldsymbol{\theta}_m, \boldsymbol{\tau}_n))$ as well as the revenue generated by that household $r(\boldsymbol{\theta}_m, \boldsymbol{\tau}_n) = \sum_{i=1}^8 \tau_i c_i(\boldsymbol{\theta}_m, \boldsymbol{\tau}_n)$.

These values for utility and revenue are then stored as a list of length $M \times N$. Since the equidistributed sequences are deterministic and thus perfectly replicable, we can save on storage space by not storing the actual values of each $\boldsymbol{\theta}_m = (\eta_m, w_m)$.

We store the values in an HDF (Hierarchical Data Format) database. HDF is a database format that allows very large databases to be stored in shared fashion across multiple machines. It also allows for parallel input and output, and is optimized for non-contiguous access. This is necessary, as during the computation process there are numerous processes that need simultaneous access to non-adjacent portions of the database. [Collette \(2013\)](#) is a great reference for using HDF5 with the Python programming language.

We access the HDF database and compute a societal welfare measure $U(\boldsymbol{\tau}_n)$ and a total revenue measure $R(\boldsymbol{\tau}_n)$ for each tax policy $\boldsymbol{\tau}_n$ using the joint density function $f(\eta, w)$,

$$U(\boldsymbol{\tau}_n) = \sum_{m=1}^M f(\boldsymbol{\theta}_m) u(\mathbf{c}(\boldsymbol{\theta}_m, \boldsymbol{\tau}_n)) \quad \forall \boldsymbol{\tau}_n \quad (12)$$

$$R(\boldsymbol{\tau}_n) = \sum_{m=1}^M f(\boldsymbol{\theta}_m) r(\boldsymbol{\theta}_m, \boldsymbol{\tau}_n) \quad \forall \boldsymbol{\tau}_n. \quad (13)$$

This step is relatively cheap computationally, and amounts to performing quasi-Monte Carlo integration.

We use a linesweeping algorithm to find any tax policies that are strictly dominated in terms of both social welfare U_n and total revenue R_n . We make a list of tax policies that are not dominated. Call these dominating points $\mathbf{D}_1 = \{\boldsymbol{\tau}_n\}_{n=1}^{N_1}$, where $N_1 \leq N$ is the number of tax policies that were not strictly dominated.

We next refine our analysis of tax policies around the policies that were not strictly dominated. We expand the policy space by adding in new tax policies that are ε_p away from the elements of \mathbf{D}_1 . The complete set of new tax policies to analyze is given by the following tensor product:

$$\mathbf{D}_2 = \{\boldsymbol{\tau}_n\}_{n=1}^{N_2} = \{\boldsymbol{\tau} \otimes [-\varepsilon_p, \varepsilon_p] \mid \boldsymbol{\tau} \in \mathbf{D}_1\} \quad (14)$$

We solve the household problem for all household types θ_m for each $\boldsymbol{\tau} \in \mathbf{D}_2$ and add the corresponding $u(\mathbf{c}(\theta_m, \boldsymbol{\tau}_n))$ and $r(\theta_m, \boldsymbol{\tau}_n)$ to the database. We then repeat the linesweep algorithm to find the points that are strictly dominating, \mathbf{D}_3 . We continue expanding the database using this iterative process until we have traced out a smooth total welfare-total revenue possibilities frontier of tax policies.

Note that after each iteration we keep all of the solutions and do not delete the dominated points, since changing the distribution will change the points that are dominated. This allows us to quickly examine how the results would change for a different distribution.

Table 3 shows the size of the type space and policy space for our database, as well as the computational time and final database size, for three iterations.

Table 3: Computation time and database size for three iterations utilizing 96 processors

Refine- ment iteration p	Points in type space M	Points in policy space N	$M \times N$	Incremental database computation time	Total revenue-utility computation time	Size of database
1	5,100	12,000	61,200,000	5.8 hours	25.1 seconds	1.0 GB
2	5,100	34,426	175,572,600	12.1 hours	61.4 seconds	2.8 GB
3	5,100	57,786	294,708,600	11.6 hours	120.1 seconds	4.7 GB

4 Results

We use our solution method from Section 3 on the model with heterogeneous households and eight consumption goods described in Section 2 to perform three experiments. The first experiment is to compare the total utility-revenue frontiers and the corresponding sets of optimal tax policies when the tax policy is allowed to be optimally set for each good (differentiated tax policy) versus a tax policy in which the rate on each of the goods is constrained to be equal (flat tax policy). Next, we show how the optimally differentiated tax policy is affected by a change in the underlying distribution of income.

Our last experiment is to study how the optimal total utility-revenue frontier and the corresponding set of optimal differentiated tax policies change when one class of goods is exempted from the tax. For this experiment, we use the “Other (not taxed)” category from our calibration.

4.1 Optimal differentiated sales tax versus optimal flat tax

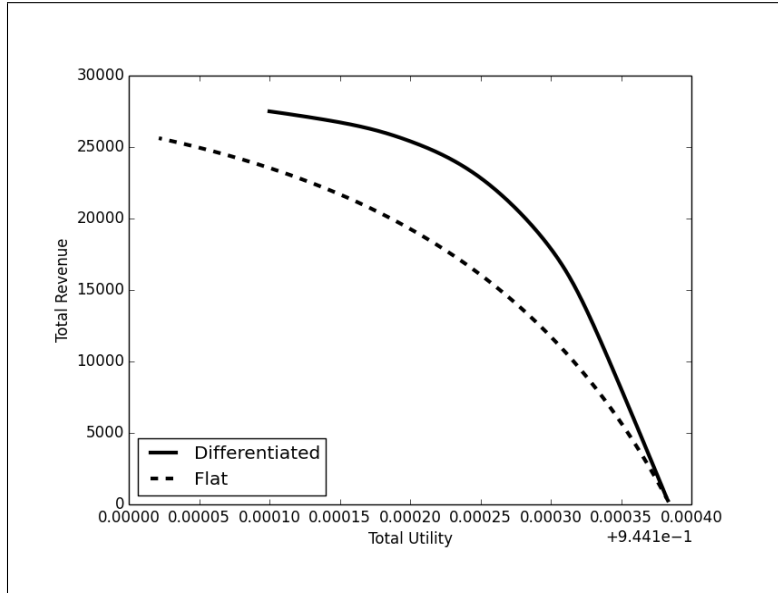
Our first experiment is to study the differences in economic outcomes between an unconstrained optimal sales tax schedule $\hat{\tau}$ that is allowed to be differentiated across goods i versus a flat tax schedule in which the tax rates across all goods are constrained to be equal $\bar{\tau}$ such that $\tau_i = \tau_j$. Figure 4 shows the difference in total utility versus total revenue frontiers associated with the two policy types. Each point on the two curves in Figure 4 represents a tax policy that is not strictly dominated in terms of total revenue and total utility given the underlying parameters of the model.⁸

Figure 5 shows how the optimal tax rate for each good i changes along the frontier from Figure 4 in terms of total revenue.⁹ The first goods categories on which the op-

⁸Note that total revenue-total utility frontier for the optimally differentiated tax policy does not have any revenues for the low levels of total utility that exist for the flat tax. This is because our tax policy space is bounded above somewhat arbitrarily by tax rates of 100 percent. If we had allowed the tax policy space to have a higher upper bound we could have populated that frontier in that extreme portion of Figure 4. However, we think commodity tax rates beyond 100 percent are not interesting for our broad categories of consumption goods.

⁹The smooth lines in Figure 5 are actually fitted logistic functions that minimize the L^1 norm of the points in tax policy space. This is because more fuzziness exists in the exact location of the

Figure 4: Total utility-revenue frontiers for optimal differentiated tax versus optimal flat tax



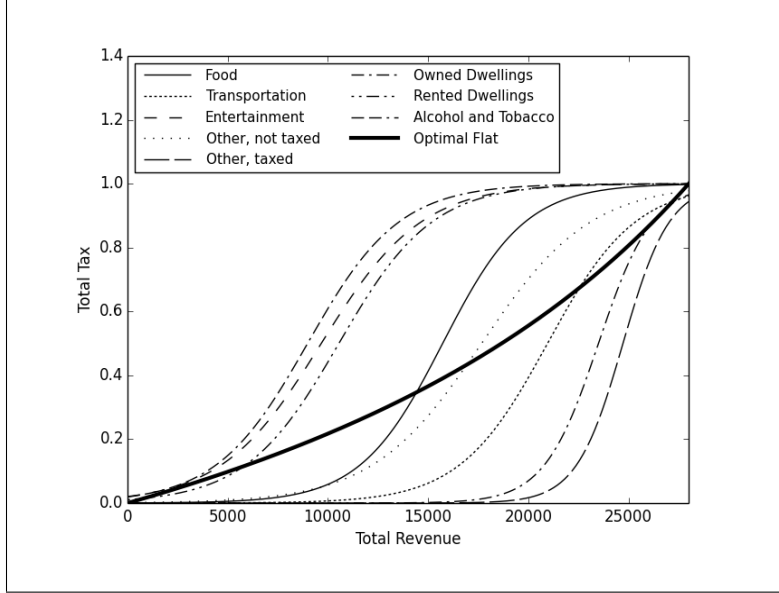
timal tax rates increase as the revenue requirements increase are alcohol and tobacco, entertainment, and rented dwellings. These goods have some of the lowest average elasticities of demand as shown in the last column of Table 2. The optimal tax rates that are the last to go up—other (taxed), owned dwellings, and transportation—have the highest average elasticities of demand.¹⁰

A broad initial interpretation of our results here confirms the Ramsey (1927) finding that optimal tax rates should vary with the elasticity of demand for a given good. With the underlying heterogeneity in elasticities of substitution, and thereby elasticities of demand, an optimally differentiated tax policy allows the policy maker to maximize total revenue collection given a particular total utility level. The optimal policies for the flat tax regime are everywhere interior to the optimally differentiated

optimal tax rates than in the resulting total utility and total revenue. Figure 18 in the Technical Appendix, which is available upon request, shows an example.

¹⁰The elasticity of demand of individual j on good consumption good i is a function of individual j 's elasticity of substitution among consumption goods η_j , his total consumption C_j (which is a function of income w_j), the minimum consumption of that particular good \bar{c}_i , and the expenditure share of the wealthy on that good α_i . Although the elasticity varies for each individual based on η_j and w_j , each individual's elasticity is inversely related to the product $\bar{c}_i \alpha_i^{-\eta_j}$. So the elasticity increases with α_i and decreases with \bar{c}_i . See the Technical Appendix for a derivation.

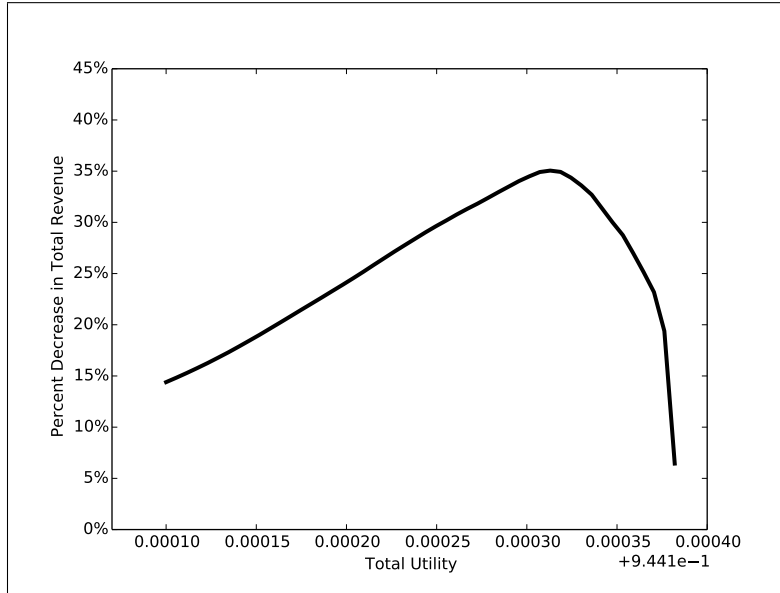
Figure 5: Optimal tax rates for good i for different levels of total revenue



sales tax regime.

However, we find that the revenue loss of an optimal flat tax regime relative to the optimally differentiated tax regime might not be that large. Figure 6 shows the percent loss in total revenue for a given total utility level for each of our experiments. The solid line in Figure 6 shows the revenue loss percent for the flat tax versus the optimally differentiated tax. The most revenue lost by the policy is just over 30 percent. This loss might not be that large if one considers the information and enforcement requirements and costs associated with an optimally differentiated sales tax system. We interpret this as evidence that a broad-based flat tax system might be a reasonable option for fundamental tax reform. Performing this experiment merely entailed re-computing the total revenue-total utility frontier on a portion of the database, which only took 1.4 seconds of computation time in contrast to the multiple hundred CPU hours to compute the database as shown in Table 3.

Figure 6: Percent loss in total revenue from optimal flat tax relative to optimal differentiated tax



4.2 Effect of exempting services

Our final experiment might have the most relevance to the U.S. economy. We compare the total revenue-total utility frontier from the optimally differentiated sales tax policy on all eight consumption goods from Section 4.1 to an optimally differentiated tax policy on only seven goods with the eighth good being exempted. For this experiment, we use the consumption category from our calibration in Section 2.3 of “Other (not taxed)” to be our exempted category.

As is described in Table 4 of Appendix A-2, the “Other (not taxed)” consumption category includes goods such as household services, vehicle finance charges, vehicle insurance, health care, and education. These categories are not taxed in the United States and represent approximately 14 percent of all U.S. consumption expenditures.

Figure 7 shows the difference in total revenue-total utility frontiers between the optimally differentiated unconstrained sales tax policy versus the optimal policy with the nontaxed services category exempted. The loss in total revenue for a given total

Figure 7: Total utility-revenue frontiers for optimal differentiated tax versus optimal differentiated tax with services exempted

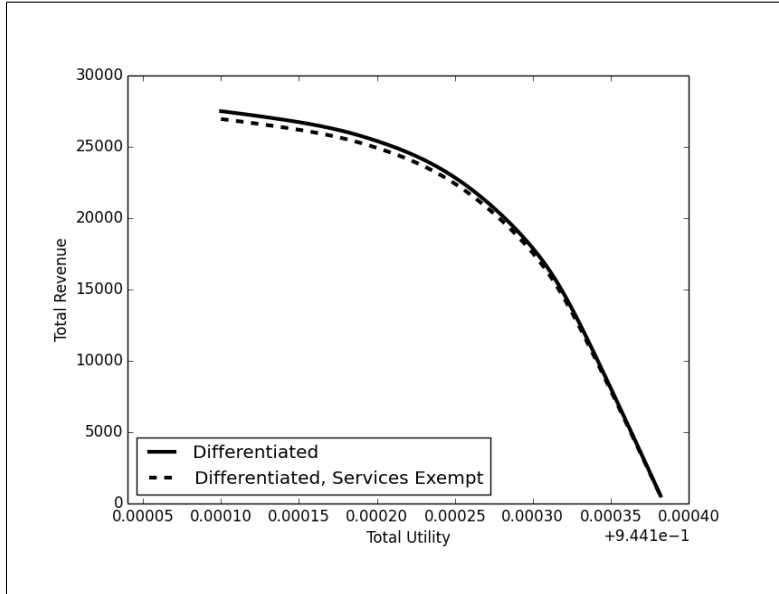


Figure 8: Difference in other seven optimally differentiated tax rates including non-taxed services minus tax rates exempting services

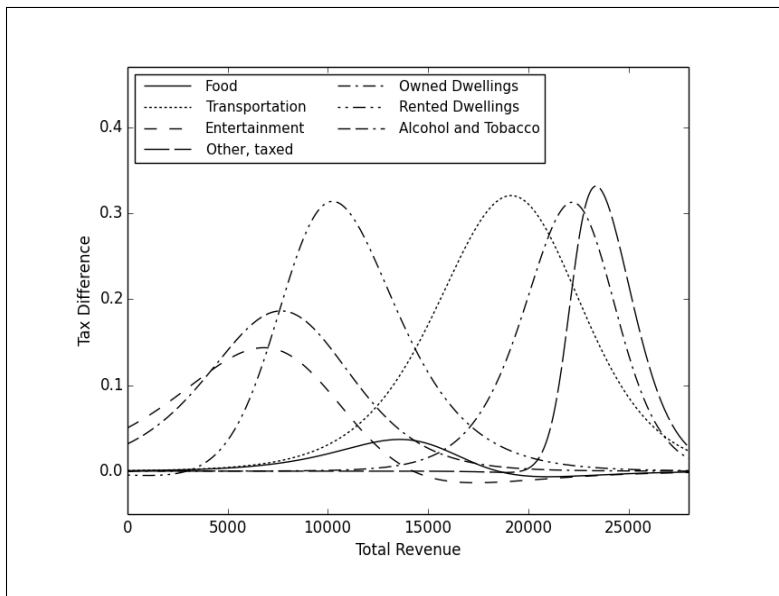
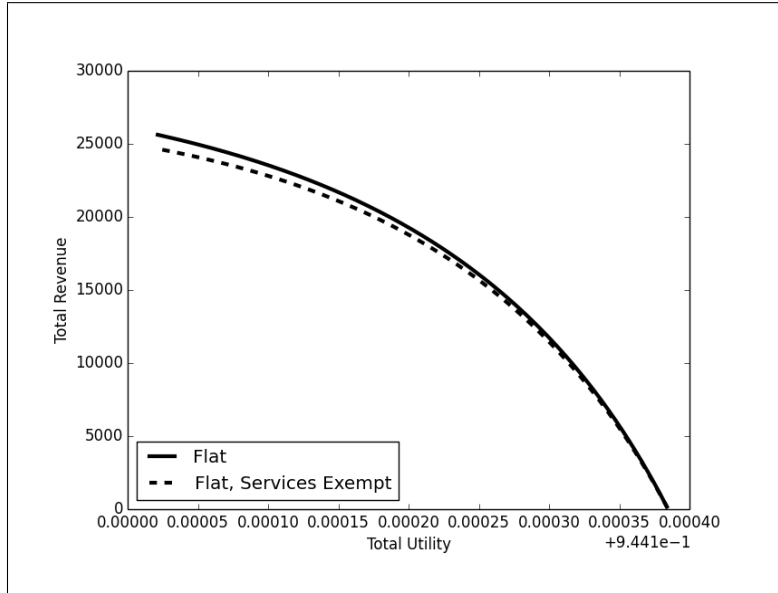


Figure 9: Total utility-revenue frontiers for optimal flat tax versus optimal flat tax with services exempted



utility level is roughly constant at about 2 percent for all levels of total utility. This is because the estimated share parameter α_i from the utility function is estimated to be a mid-to-low range value of $\alpha_i = 0.14$, as shown in Table 2. In addition, the estimated minimum consumption for this category is very low at $\bar{c}_i = \$29$. This implies a relatively high elasticity of demand for nontaxed services, and exempting them only slightly impairs the ability of the policy maker to raise revenue. The elasticities of demand for most of the other goods are lower relative to that of nontaxed services.

Figure 8 shows the difference in taxes for each good i , which is the optimal differentiated tax exempting services minus the optimal differentiated tax including an optimal tax on services. For low levels of total revenue, the revenue lost from exempting services is made up primarily by increases in the low elasticity of demand categories of entertainment, alcohol and tobacco, and rented dwellings. The optimal tax on rented dwellings increases by a maximum of 30 percent. However, at the high end of total revenues, it is primarily through tax increases on the high elasticity of demand categories of other (taxed), owned dwellings, and transportation that the difference is made up.

We test whether this result is specific to the optimally differentiated tax experiment by repeating it in a flat tax environment. That is, we compare the total revenue-total utility frontier from an optimal flat tax regime to that of an optimal flat tax in which the services category tax rate is constrained to be zero. Figure 9 shows the difference in total utility-revenue frontiers. The loss in total revenue for a given utility level is still small. The loss from the exempted flat tax relative to the all inclusive flat tax is roughly constant at nearly 3 percent for all levels of total utility, very similar to the loss of the exempted differentiated tax regime relative to the all inclusive differentiated tax regime.

For those interested in fundamental tax reform that broadens the tax base, this experiment suggests that economy-wide commodity taxation could maintain the exemption for services that currently exists in the U.S. tax system without overly impairing the ability of the policymaker to raise revenue. These two experiments with a services exemption and recomputing the total revenue-total utility frontier only took 3.4 seconds of computation time in contrast to the multiple hundred CPU hours to compute the database as shown in Table 3.

5 Conclusion

This paper proposes and describes a solution method for optimal policy problems that involve large dimensions of heterogeneity, nonconvex policy-maker optimization problems, and potential policy objective uncertainty. We use a big data approach to generate a large database of agent behaviors across the space of agent types and potential policies. Once this database is generated, we can very quickly compute frontiers in terms of the components of the policy objectives and perform experiments on how changes in the underlying distributions and other characteristics of the model affect the optimal outcomes.

We demonstrate this method by performing two experiments in an environment calibrated to the United States looking at optimal commodity taxation. We assume the objective of the policy maker includes total revenue and societal utility as com-

ponents, but we make no assumptions as to the weighting of those components. We first test how the total revenue-total utility frontier changes from an environment in which the commodity taxes for each good can be set independently and optimally verses a flat tax environment in which all commodity tax rates are equal and are set optimally. We find that the percent loss of restricting policy to a flat tax regime incurs a loss no higher than 32 percent for a given level of total utility, but often much less than 30 percent. If one thinks that the information and enforcement costs of an optimally differentiated commodity tax are sufficiently high, this result could be evidence that a broad-based flat tax might be a good practical tax policy.

We also measure the effect of exempting a particular class of consumption good. In the United States, as well as in many other countries, many services are not taxed. Because the elasticity of demand is relatively high for the services that are exempted in the U.S. economy, we find a very small loss in terms of total revenue for a given level of total utility. This relationship is robust to the level of total utility and to whether the tax regime is an optimally differentiated sales tax or an optimal flat tax. In both cases, the loss in total revenue is everywhere less than 3 percent. This suggests that any tax reform that maintains the tax exemption for services might not overly impair the ability of the policy maker to raise revenue.

APPENDIX

A-1 Hybrid function to deal with minimum consumption requirement

Because some household's income is not enough to purchase the minimum consumption of each good, as described in (9), the total consumption aggregator (1) will be undefined for consumption levels less than the minimum $c_i < \bar{c}_i$ and household utility (2) will be negative infinity if any one of the household consumption levels is equal to the minimum $c_i = \bar{c}_i$.

$$C \equiv \left(\sum_{i=1}^I \alpha_i (c_i - \bar{c}_i)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \quad \forall \eta \geq 1 \quad (1)$$

$$u(C) = \frac{C^{1-\gamma} - 1}{1-\gamma} \quad \forall \gamma \geq 1 \quad (2)$$

The problem is that, given a tax schedule $\boldsymbol{\tau} = \{\tau_i\}_{i=1}^I$, some households income w is not sufficient to be able to afford the minimum amount of consumption.

$$\sum_{i=1}^I (1 + \tau_i) \bar{c}_i > w \quad (A.1)$$

This is a problem because the exponent inside the large parentheses in (1) is undefined for $c_i < \bar{c}_i$. The numerical methods used require that the utility function is everywhere defined. We will construct functions that extend the utility function and consumption aggregator so that they are globally valid.

Define the function $g_1(c)$ in the following way.

$$g_1(c) = c^{\frac{\eta-1}{\eta}} \quad \forall c \geq 0, \eta > 1 \quad (A.2)$$

Figure 10 shows a graph of $g_1(c)$ for $\eta = 4.2$. Note that $g_1(c)$ is not defined for $c < 0$.

We need a function equal to $g_1(c)$ above some cutoff $\varepsilon_0 > 0$ close to zero, but have it be defined for $c < 0$. We will construct a function $g(c)$ which has two pieces $g_1(c)$ and $g_2(c)$ and join them with a smooth transition. The piece for low values of consumption $g_2(c)$ should be a concave, monotone increasing function. The quadratic function is a good candidate. We want to stitch the function $g_2(c)$ to $g_1(c)$ for all $c < \varepsilon_0$. This quadratic function $g_2(c)$ must equal $g_1(c)$ at ε_0 and have the same first and second derivatives at that point.

$$\begin{aligned} g_2(c) &= a_0 + a_1 c + \frac{1}{2} a_2 c^2 \quad \forall a_0, a_1, a_2, c \\ s.t. \quad g_2(\varepsilon_0) &= g_1(\varepsilon_0), \\ g_2'(\varepsilon_0) &= g_1'(\varepsilon_0), \\ g_2''(\varepsilon_0) &= g_1''(\varepsilon_0), \end{aligned} \quad (A.3)$$

Figure 10: Function $g_1(c)$ for $c \geq 0$ and $\eta = 4.2$

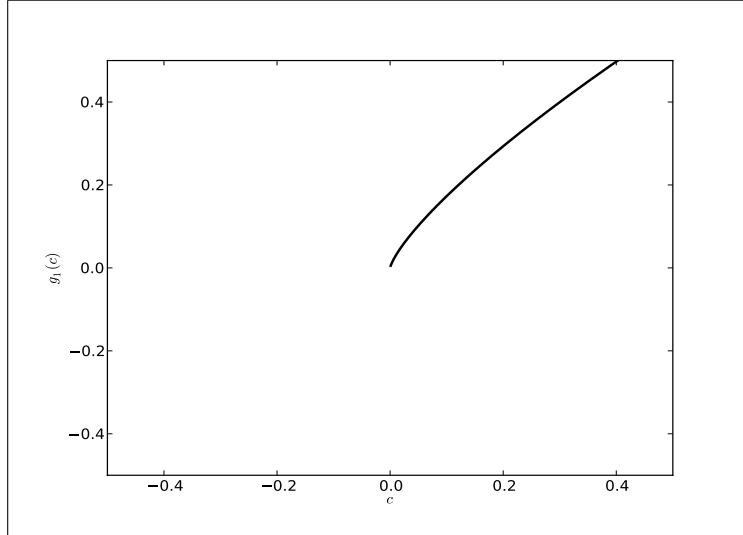
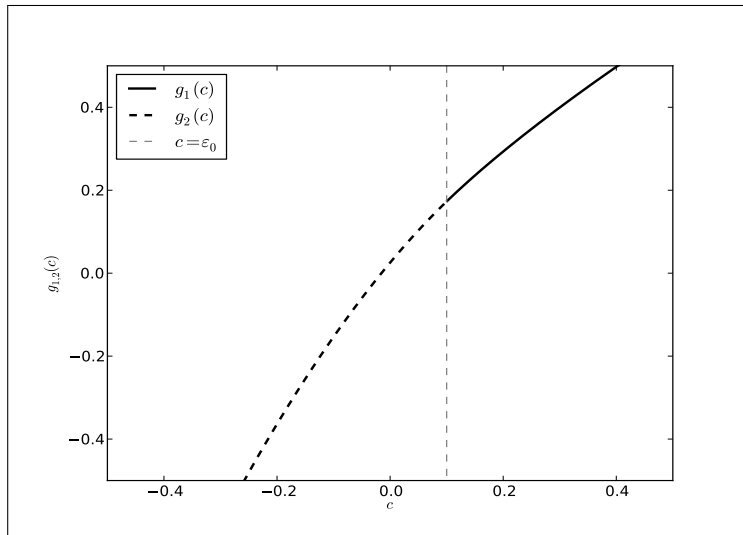


Figure 11: Function $g(c)$ for $\eta = 4.2$ and $\varepsilon_0 = 0.1$



We can solve for the three unknown coefficients a_0 , a_1 , and a_2 in (A.3) from the three conditions about connecting, first derivative, and second derivative. The solution is the following:

$$\begin{aligned} a_0 &= \frac{(1 + \eta)\varepsilon_0^{\frac{\eta-1}{\eta}}}{2\eta^2} \\ a_1 &= \frac{(\eta^2 - 1)\varepsilon_0^{-\frac{1}{\eta}}}{\eta^2} \\ a_2 &= -\frac{(\eta - 1)\varepsilon_0^{-\frac{\eta+1}{\eta}}}{\eta^2} \end{aligned} \tag{A.4}$$

Define the new hybrid function $g(c)$ in the following way.

$$g(c) = \begin{cases} g_1(c) & \text{for } c \geq \varepsilon_0 \\ g_2(c) & \text{for } c < \varepsilon_0 \end{cases} \tag{A.5}$$

Figure 11 shows the hybrid function $g(c)$.

In order to finish constructing our aggregator, we now need to construct the inverse function. To construct the inverse aggregator we choose a cutoff of $\varepsilon_1 = g(\varepsilon_0)$. Note that the value of ε_1 is dependent on η , and as such will be different for each household. We define the inverse of $g(c)$ to be the following.

$$g^{-1}(c) = \begin{cases} g_1^{-1}(c) & \text{for } c \geq \varepsilon_1 \\ g_2^{-1}(c) & \text{for } c < \varepsilon_1 \end{cases} \tag{A.6}$$

This again provides a smooth transition. Figure 12 shows the hybrid function $g^{-1}(c)$.

For computational purposes, we will use the following aggregator function,

$$C \equiv g^{-1} \left(\sum_{i=1}^I \alpha_i g(c_i - \bar{c}_i) \right) \tag{A.7}$$

where $g(\cdot)$ is given in (A.5) and $g^{-1}(\cdot)$ is given in (A.6). This function will behave just like (1) for most individuals. But the adjusted aggregator in (A.7) will also accomodate consumption of individual goods less than the minimum $c_i < \bar{c}_i$, which is necessary for some individuals with low enough incomes as well as for computation.

A similar problem arises with the utility function (2). The aggregated consumption from (A.7) may take on values where the utility function is not defined. We can apply the same extension technique to make the utility function everywhere defined.

Define the function $h_1(c)$ in the following way.

$$h_1(c) = \frac{c^{1-\gamma} - 1}{1 - \gamma} \quad \forall \gamma \geq 1 \tag{A.8}$$

Figure 12: Function $g^{-1}(c)$ for $\eta = 4.2$ and $\varepsilon_1 = 0.17$

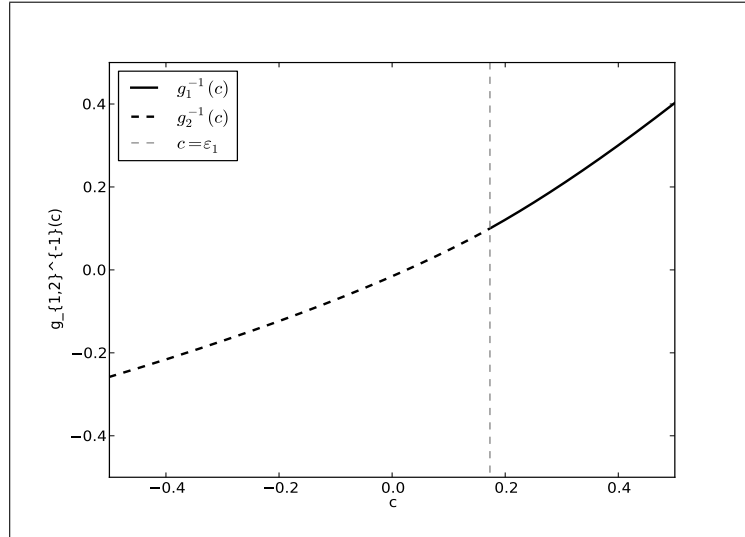
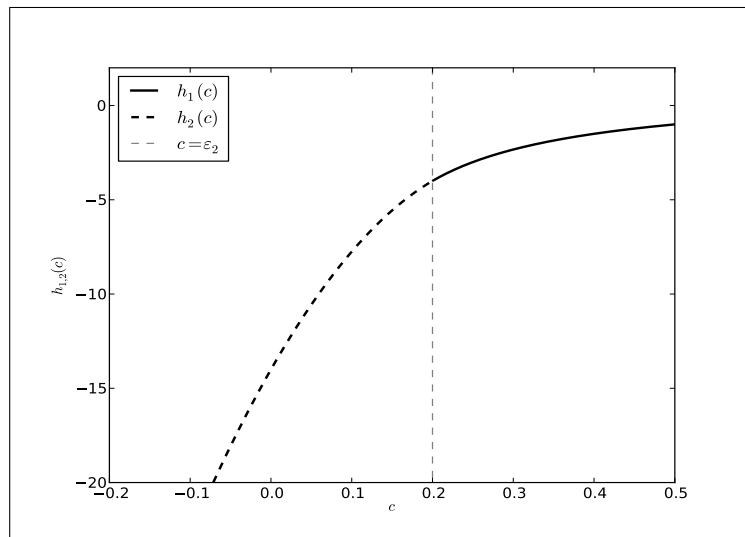


Figure 13: Function $h(c)$ for $\gamma = 2$ and $\varepsilon_2 = 0.2$



As before, we need a function equal to $h_1(c)$ above some cutoff $\varepsilon_2 > 0$ which is defined for $c \leq 0$. We will construct a globally valid $h(c)$ consisting of two pieces joined together with a smooth transition at ε_2 . We again use a quadratic form, and define the second piece as follows.

$$\begin{aligned}
h_2(c) &= b_0 + b_1c + \frac{1}{2}b_2c^2 \quad \forall b_0, b_1, b_2, c \\
s.t. \quad h_2(\varepsilon_2) &= h_1(\varepsilon_2), \\
h_2'(\varepsilon_2) &= h_1'(\varepsilon_2), \\
h_2''(\varepsilon_2) &= h_1''(\varepsilon_2),
\end{aligned} \tag{A.9}$$

Given these three conditions, we can solve for the coefficients b_0 , b_1 , and b_2 in (A.9). The solution is the following:

$$\begin{aligned}
b_0 &= \frac{\varepsilon_2^{-\gamma}(-\gamma\varepsilon_2 - \gamma^2\varepsilon_2 + 2\varepsilon_2^\gamma)}{2(\gamma - 1)} \\
b_1 &= \varepsilon_2^{-\gamma}(1 + \gamma) \\
b_2 &= -\gamma\varepsilon_2^{-1-\gamma}
\end{aligned} \tag{A.10}$$

We can now define the hybrid utility function as follows.

$$h(c) = \begin{cases} h_1 = (c) & \text{for } c \geq \varepsilon_2 \\ h_2 = (c) & \text{for } c < \varepsilon_2 \end{cases} \tag{A.11}$$

Figure 13 shows the hybrid function $h(c)$. For computational purposes we will use $h(\cdot)$ as given in (A.11) as the utility function. This function will behave just like (2) for most individuals, but will be everywhere defined.

A-2 Consumer Expenditure Survey (CEX) category description and estimation of α_i and \bar{c}_i

To calibrate the expenditure share of the wealthiest individuals α_i and the minimum consumption levels \bar{c}_i , we used U.S. consumption data from the broad category summary tables of the 2011 Consumer Expenditure Survey (CEX). We aggregate expenditures into eight categories ($I = 8$): food, transportation, entertainment, owned dwellings, rented dwellings, alcohol and tobacco, other (taxed), and other (not taxed). Table 4 details the individual consumption expenditure categories that make up each of the eight broad consumption categories that we use.

Figure 2 shows average total expenditure by household income category on each of the eight categories of consumption goods. One problem with the measure of income on the x -axis of Figure 2 is that it does not match up well with the concept of income in our model. Our static model includes no savings decisions. However, income in the CEX is used for both consumption and savings. For this reason, we compute total expenditure by households in a particular income category as our empirical analog of income in our model. Figure 3 shows the expenditures from Figure 2 as a percent of average total expenditure by households in a particular income category.

We estimate the minimum consumption of the poorest individuals \bar{c}_i and the consumption expenditure share of the wealthiest individuals α_i as the coefficients from the following regression in (11). Table 2 presents the estimated values for \bar{c}_i and α_i obtained by running the OLS regression in (11) for each of the eight consumption goods categories. Because negative values for minimum consumption levels \bar{c}_i and expenditure shares α_i do not make sense, we adjust our calibration of the minimum consumption of goods that are projected to be negative for $AVGTOTEXP_j = 0$ to be $\bar{c}_i = 0$. This adjustment was made for the four categories of Transportation, Entertainment, Other (taxed), and Owned dwellings. The adjustment of $\alpha_i = 0.01$ for any initially estimated $\alpha_i < 0$ only had to be implemented for the one consumption category of Alcohol and tobacco. For our calibration of \bar{c}_i and α_i , we use the adjusted values in the rightmost two columns of Table 2.

Note from Figure 2 that a linear specification as in (11) is likely not fully appropriate. However, a linear specification is the only one that allows us to estimate, in a consistent manner, the consumption minimum \bar{c}_i and the expenditure share α_i for each consumption category i . If we did a log specification for $AVGTOTEXP_j$, we would not be able to estimate \bar{c}_i by taking the predicted value of $c_{i,j}$ at $AVGTOTEXP_j = 0$. We would also not be able to estimate the consumption share of the wealthiest individuals α_i by letting $AVGTOTEXP_j$ go to infinity because the share would necessarily go to zero. The same problems exist for a quadratic specification of (11). For this reason, we use the linear specification to calibrate our model.

Table 4: Detail of CEX categories included in eight broad model categories

Model category	Broad CEX category	Specific CEX category
Food	Food	all subcategories
Transportation	Transportation	all subcategories minus vehicle finance charges and vehicle insurance
Entertainment	Entertainment	all subcategories
Owned dwellings	Housing	Shelter: Owned dwellings
Rented dwellings	Housing	Shelter: Rented dwellings
Alcohol and Tobacco	Alcoholic Beverages Tobacco products and smoking supplies	n/a n/a
Other (taxed)	Housing Apparel and services Personal care products and services Reading Miscellaneous	Shelter: Other lodging Utilities, fuels, and public services Household operations: Other household expenses Housekeeping supplies Household furnishings and equipment all subcategories n/a n/a n/a
Other (not taxed)	Housing Transportation Health care Education	Household operations: Personal services Other vehicle expenses: Vehicle finance charges Vehicle insurance all subcategories n/a

The two expenditure categories of “Cash contributions” and “Personal insurance and pensions” were excluded because they could be classified more as savings than as consumption.

References

- Alm, James and Mikhail I. Melnik**, “Sales Taxes and the Decision to Purchase Online,” *Public Finance Review*, March 2005, 33 (2), 184–212.
- Armington, Paul S.**, “A Theory of Demand for Products Distinguished by Place of Production,” *IMF Staff Papers*, March 1969, 16 (1), 159–178.
- Balcer, Yves, Irwin Garfinkel, Kathy Krynski, and Efraim Sadka**, “Income Redistribution and the Structure of Indirect Taxation,” in Elhanan Helpman, Assaf Razin, and Efraim Sadka, eds., *Social Policy Evaluation: An Economic Perspective*, Academic Press, 1983, pp. 279–297.
- Ballard, Charles L. and Jaimin Lee**, “Internet Purchases, Cross-border Shopping, and Sales Taxes,” *National Tax Journal*, December 2007, 60 (4), 711–725.
- Blake, David**, “Efficiency, Risk Aversion and Portfolio Insurance: An Analysis of Financial Asset Portfolios Held by Investors in the United Kingdom,” *Economic Journal*, September 1996, 106 (438), 1175–1192.
- Bruce, Donald, William F. Fox, and LeAnn Luna**, “State and Local Government Sales Tax Revenue Losses from Electronic Commerce,” *State Tax Notes*, May 2009, 52 (7), 537–558.
- Campbell, John Y.**, “Understanding Risk and Return,” *Journal of Political Economy*, April 1996, 104 (2), 298–345.
- Christopoulou, Rebekka and Philip Vermeulen**, “Markups in the Euro Area and the US of the Period 1981-2004: A Comparison of 50 Sectors,” *Empirical Economics*, February 2012, 42 (1), 53–77.
- Collette, Andrew**, *Python and HDF5: Unlocking Scientific Data*, O’Reilly Media, 2013.
- Current Population Survey**, “2012 Annual Social and Economic (ASEC) Supplement,” Technical Report, Bureau of the Census and Bureau of Labor Statistics 2012.
- Deaton, Angus**, “Equity, Efficiency, and the Structure of Indirect Taxation,” *Journal of Public Economics*, December 1977, 8 (3), 299–312.
- Diamond, Peter A. and James A. Mirrlees**, “Optimal Taxation and Public Production I: Production Efficiency,” *American Economic Review*, March 1971, 61 (1), 8–27.
- and —, “Optimal Taxation and Public Production II: Tax Rules,” *American Economic Review*, June 1971, 61 (3, Part 1 of 2), 261–278.
- Dixit, Avinash K. and Joseph E. Stiglitz**, “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review*, June 1977, 67 (3), 297–308.

- Einav, Liran and Jonathan D. Levin**, “The Data Revolution and Economic Analysis,” NBER Working Paper No. 19035, National Bureau of Economic Research May 2013.
- , **Dan Knoepfle, Jonathan Levin, and Neel Sundareshan**, “Sales Taxes and Internet Commerce,” *American Economic Review*, January 2014, *104* (1), 1–26.
- Ellison, Glenn and Sara Fisher Ellison**, “Tax Sensitivity and Home State Preferences in Internet Purchasing,” *American Economic Journal: Economic Policy*, August 2009, *1* (2), 53–71.
- Garner, C. Alan**, “Consumption Taxes: Macroeconomic Effects and Policy Issues,” *Economic Review*, Federal Reserve Bank of Kansas City Second Quarter 2005.
- Goolsbee, Austan**, “In a World without Borders: The Impact of Taxes on Internet Commerce,” *Quarterly Journal of Economics*, May 2000, *115* (2), 561–576.
- , “Internet Commerce, Tax Sensitivity, and the Generation Gap,” in James M. Poterba, ed., *Tax Policy and the Economy*, Vol. 14 National Bureau of Economic Research MIT Press January 2000, pp. 45–66.
- Gruber, Jonathan**, *Public Finance and Public Policy*, 3rd ed., Worth Publishers, 2011.
- Hines, James R., Jr.**, “Taxing Consumption and Other Sins,” *Journal of Economic Perspectives*, Winter 2007, *21* (1), 49–68.
- Judd, Kenneth L.**, *Numerical Methods in Economics*, MIT Press, 1998.
- Kaplow, Louis**, “Taxing Labor Complements,” *Economic Inquiry*, October 2010, *48* (4), 1065–1071.
- Kocherlakota, Narayana R.**, “The Equity Premium: It’s Still a Puzzle,” *Journal of Economic Literature*, March 1996, *34* (1), 42–71.
- Mankiw, N. Gregory and Stephen P. Zeldes**, “The Consumption of Stockholders and Nonstockholders,” *Journal of Financial Economics*, March 1991, *29* (1), 97–112.
- McDonald, James B.**, “Some Generalized Functions for the Size Distribution of Income,” *Econometrica*, May 1984, *52* (3), 647–663.
- **and Yexiao Xu**, “A Generalization of the Beta Distribution with Applications,” *Journal of Econometrics*, March-April 1995, *66* (1-2), 133–152.
- , **Jeff Sorensen, and Patrick A. Turley**, “Skewness and Kurtosis Properties of Income Distribution Models,” *Review of Income and Wealth*, June 2013, *59* (2), 360–374.

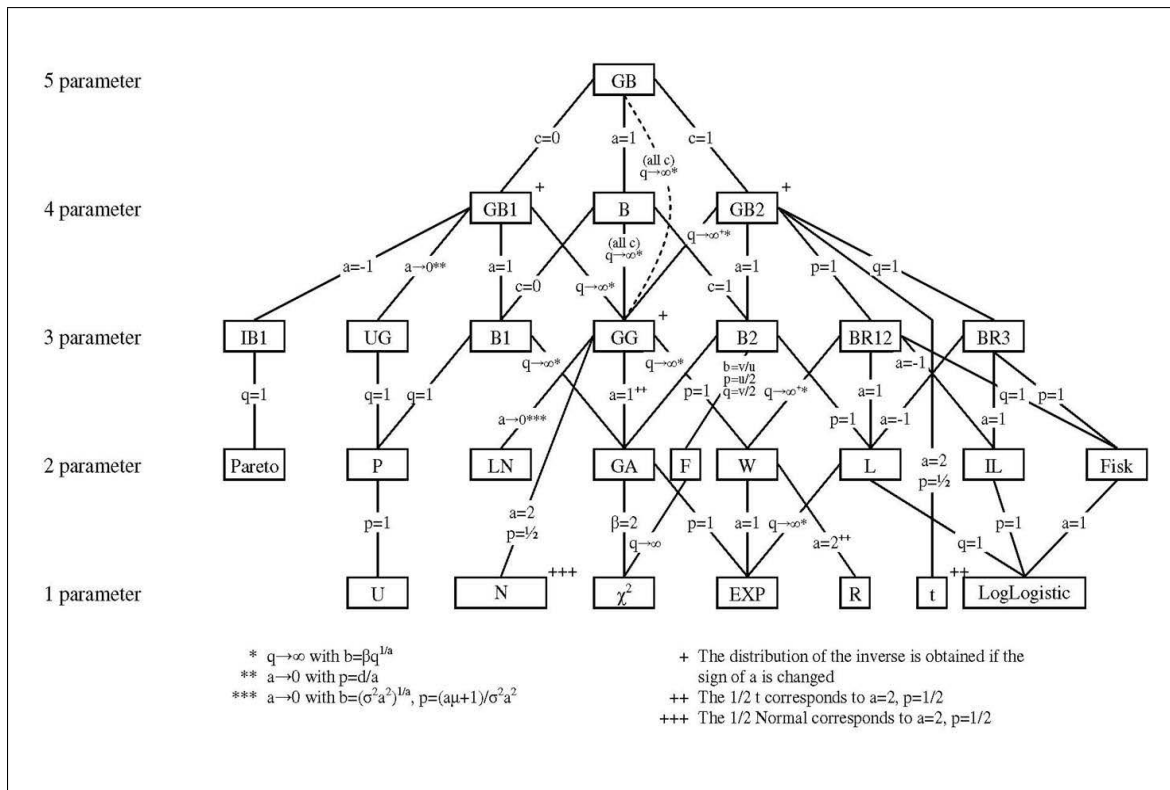
- Mehra, Rajnish and Edward C. Prescott**, “The Equity Premium: A Puzzle,” *Journal of Monetary Economics*, March 1985, *15* (2), 145–161.
- Naito, Hisahiro**, “Re-examination of Uniform Commodity Taxes under a Non-linear Income Tax System and Its Implication for Production Efficiency,” *Journal of Public Economics*, February 1999, *71* (2), 165–188.
- Ramsey, Frank P.**, “A Contribution to the Theory of Taxation,” *Economic Journal*, March 1927, *37* (145), 47–61.
- Saez, Emmanuel**, “The Desirability of Commodity Taxation under Non-linear Income Taxation and Heterogeneous Tastes,” *Journal of Public Economics*, February 2002, *83* (2), 217–230.
- **and Stefanie Stantcheva**, “Generalized Social Marginal Welfare Weights for Optimal Tax Theory,” NBER Working Paper No. 18835, National Bureau of Economic Research February 2013.
- Scanlan, Mark A.**, “Tax Sensitivity in Electronic Commerce,” *Fiscal Studies*, December 2007, *28* (4), 417–436.
- Slemrod, Joel and Jon Bakija**, *Taxing Ourselves: A Citizen’s Guide to the Debate over Taxes*, 4th ed., MIT Press, 2008.
- White, Tom**, *Hadoop: The Definitive Guide*, O’Reilly Media, 2012.

TECHNICAL APPENDIX

T-1 Fitting a generalized gamma (GG) distribution to the U.S. income distribution

McDonald (1984) and McDonald et al. (2013) have documented that the members of the generalized beta (GB) family of distributions with three parameters or more fit the U.S. distribution of income much better than the more commonly used two-parameter distributions in the GB family, such as the log normal distribution. Figure 14 shows the GB family of distributions. In our case, the generalized gamma (GG) distribution captures both the shape of the distribution where most of the mass is as well as the thick tails.

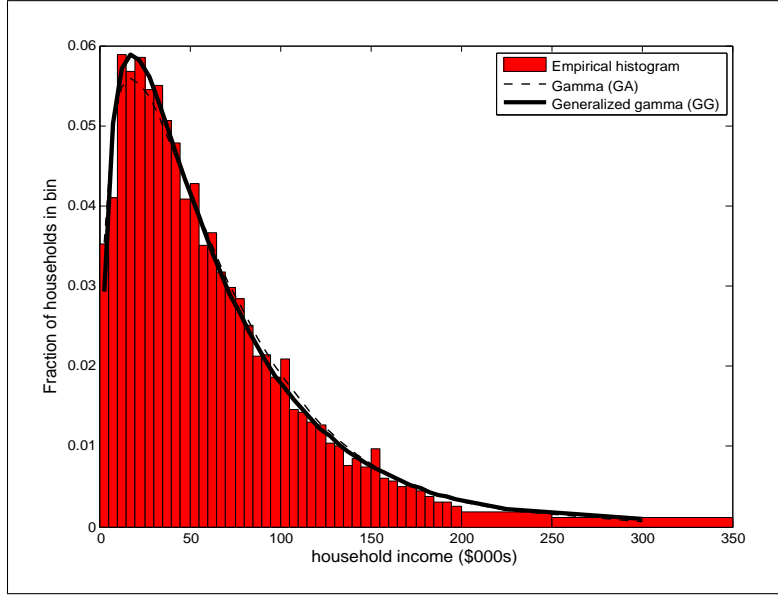
Figure 14: Generalized beta family of distributions [taken from McDonald and Xu (1995, Fig. 2)]



Our approach to fitting the best distribution from the GB family of distributions in Figure 14 is to start with a simple two-parameter distribution. Because of the shape of the empirical histogram in Figure 15 (which is just a reproduction of Figure 1 from the paper), we start with the gamma (GA) distribution shown in (T.1.1).

$$(GA): \quad f(y; a, b) = \frac{1}{b^a \Gamma(a)} y^{a-1} e^{-\left(\frac{y}{b}\right)} \quad y \in [0, \infty), \quad a, b > 0 \quad (T.1.1)$$

Figure 15: Histogram of U.S. household income and estimated generalized gamma (GG) distribution



The GMM estimated parameters of the gamma (GA) distribution are $a = 1.36$ and $b = 48,362$. The fit of the GA distribution is shown in Figure 15 and listed in the fourth column of Table 5. The fit of the GA distribution is pretty good, although it does not put enough density in the \$20,000 to \$50,000 and above \$200,000 income ranges and puts too much density in the \$80,000 to \$120,000 range.

We then use the estimates for a and b from the two-parameter gamma (GA) distribution and move one rung up the generalized beta (GB) distribution family tree in Figure 14 to the three-parameter generalized gamma (GG) distribution. The probability density function for the generalized gamma distribution is the following.

$$(GG) : f(y; a, b, m) = \frac{m}{b^a \Gamma\left(\frac{a}{m}\right)} y^{a-1} e^{-\left(\frac{y}{b}\right)^m} \quad y \in [0, \infty), \quad a, b, m > 0 \quad (6)$$

It is clear from comparing the generalized gamma (GG) distribution in (6) to the gamma (GA) distribution in (T.1.1) that the GA is a nested case of the GG.

$$GA(y; a, b) = GG(y; a, b, m = 1) \quad (T.1.2)$$

We use the GMM estimates of the two gamma (GA) parameters, as well as $m = 1$, as initial guesses for the generalized gamma (GG) parameters in the GMM estimation. The estimated GG parameters are reported in Table 4 and are $a = ?$, $b = ?$, and $m = ?$. Figure 15 shows that the GG distribution fits better than the GA distribution in the \$20,000 to \$50,000, \$80,000 to \$120,000, and above \$200,000 income ranges. The fifth column of Table 5 lists how well the GG moments match the empirical moments. In particular, note that the mean income from the estimated GG distribution comes closer to matching the empirical mean income. This signifies that the GG distribution

is doing a better job of matching the fat right tail of the income distribution than is the GA distribution.

We did go one more rung up the generalized beta (GB) distribution tree in Figure 14 and estimated a four-parameter generalized beta 2 (GB2) distribution. However, with the 42 empirical moments of the U.S. income distribution that we have, the fit of the GB2 is nearly identical to that of the generalized gamma (GG) distribution.

Table 5: Distribution of Household Money Income by Selected Income Class, 2011 and estimated GA and GG fit

Income class	# households (000s)	households %	GA est. %	GG est. %
All households	121,084	100.0	100.0	100.0
Less than \$5,000	4,261	3.5	3.5	2.9
\$5,000 to \$9,999	4,972	4.1	5.0	5.0
\$10,000 to \$14,999	7,127	5.9	5.5	5.7
\$15,000 to \$19,999	6,882	5.7	5.6	5.9
\$20,000 to \$24,999	7,095	5.9	5.5	5.8
\$25,000 to \$29,999	6,591	5.4	5.4	5.6
\$30,000 to \$34,999	6,667	5.5	5.1	5.3
\$35,000 to \$39,999	6,136	5.1	4.9	5.0
\$40,000 to \$44,999	5,795	4.8	4.6	4.7
\$45,000 to \$49,999	4,945	4.1	4.3	4.3
\$50,000 to \$54,999	5,170	4.3	4.0	4.0
\$55,000 to \$59,999	4,250	3.5	3.8	3.7
\$60,000 to \$64,999	4,432	3.7	3.5	3.4
\$65,000 to \$69,999	3,836	3.2	3.2	3.1
\$70,000 to \$74,999	3,606	3.0	3.0	2.9
\$75,000 to \$79,999	3,452	2.9	2.8	2.7
\$80,000 to \$84,999	3,036	2.5	2.6	2.4
\$85,000 to \$89,999	2,566	2.1	2.4	2.2
\$90,000 to \$94,999	2,594	2.1	2.2	2.1
\$95,000 to \$99,999	2,251	1.9	2.0	1.9
\$100,000 to \$104,999	2,527	2.1	1.8	1.7
\$105,000 to \$109,999	1,771	1.5	1.7	1.6
\$110,000 to \$114,999	1,723	1.4	1.5	1.5
\$115,000 to \$119,999	1,569	1.3	1.4	1.3
\$120,000 to \$124,999	1,540	1.3	1.3	1.2
\$125,000 to \$129,999	1,258	1.0	1.2	1.1
\$130,000 to \$134,999	1,211	1.0	1.1	1.0
\$135,000 to \$139,999	918	0.8	1.0	0.9
\$140,000 to \$144,999	1,031	0.9	0.9	0.9
\$145,000 to \$149,999	893	0.7	0.8	0.8
\$150,000 to \$154,999	1,166	1.0	0.8	0.7
\$155,000 to \$159,999	740	0.6	0.7	0.7
\$160,000 to \$164,999	697	0.6	0.6	0.6
\$165,000 to \$169,999	610	0.5	0.6	0.6
\$170,000 to \$174,999	617	0.5	0.5	0.5
\$175,000 to \$179,999	530	0.4	0.5	0.5
\$180,000 to \$184,999	460	0.4	0.4	0.4
\$185,000 to \$189,999	363	0.3	0.4	0.4
\$190,000 to \$194,999	380	0.3	0.4	0.4
\$195,000 to \$199,999	312	0.3	0.3	0.3
\$200,000 to \$249,999	2,297	1.9	2.0	2.2
\$250,000 and over	2,808	2.3	1.2	1.8
Mean income	\$69,677		\$65,860	\$67,108
Median income	\$50,054		\$50,625	\$49,625

Source: 2011 Current Population Survey household income count data **Current Population Survey (2012, Table HINC-01)**

T-2 Properties and advantages of equidistributed sequences

Equidistributed sequences are deterministic sequences of real numbers where the proportion of terms falling in any subinterval is proportional to the length of that interval. A sequence $\{x_j\}_{j=1}^{\infty} \subset D \subset \mathbb{R}^n$ is equidistributed over D if and only if

$$\lim_{n \rightarrow \infty} \frac{\mu(D)}{n} \sum_{j=1}^n f(x_j) = \int_D f(x) dx \quad (\text{T.2.3})$$

for all Riemann-integrable $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, where $\mu(D)$ is the Lebesgue measure of D .

There are a number of equidistributed sequences that possess this property. Let p_1, p_2, \dots denote the sequence of prime numbers $2, 3, 5, \dots$, and let $\{x\}$ represent the fractional part of x , that is $\{x\} = x - \lfloor x \rfloor$. Table 6 contains examples of a number of equidistributed sequences. Figure 16 shows the first 10 000 points for two-dimensional Weyl, Haber, Niederreiter, and Baker sequences.

Table 6: Equidistributed Sequences in \mathbb{R}^n

Name of Sequence	Formula for (x_1, x_2, \dots, x_n)
Weyl	$(\{np_1^{1/2}\}, \dots, \{np_n^{1/2}\})$
Haber	$(\{\frac{n(n+1)}{2}p_1^{1/2}\}, \dots, \{\frac{n(n+1)}{2}p_n^{1/2}\})$
Niederreiter	$(\{n2^{1/(n+1)}\}, \dots, \{n2^{n/(n+1)}\})$
Baker	$(\{ne^{r_1}\}, \dots, \{ne^{r_n}\})$, r_j rational and distinct

In generating our database we use a scaled Baker sequence. The n^{th} element of the type space is given by

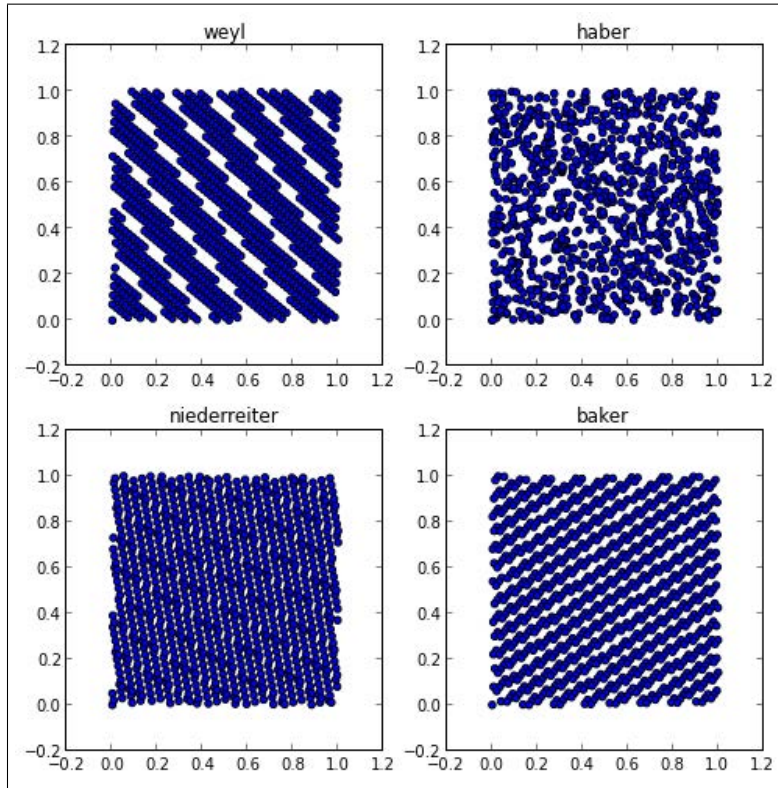
$$\theta_n = (\eta_n, w_n) = (0.8 (ne^2 - \lfloor ne^2 \rfloor) + 3.7, 492000 (ne^3 - \lfloor ne^3 \rfloor) + 8,000). \quad (\text{T.2.4})$$

The initial sample points in the policy space are similarly chosen using an eight-dimensional Baker sequence. Figure 17 shows the equidistributed points of a two-dimensional Baker sequence on $[0, 1]^2$.

Quasi-Monte Carlo integration is used to integrate over the type space for each point in policy space given the type space distribution. Quasi-Monte Carlo integration is similar to Monte Carlo integration, but chooses points using equidistributed sequences instead of pseudorandom numbers. This allows for a faster rate of convergence for a large number of points. With N points in s dimensions, quasi-Monte Carlo techniques converge in $O\left(\frac{(\log N)^s}{N}\right)$ as opposed to $O\left(\frac{1}{\sqrt{N}}\right)$ for Monte Carlo techniques. ¹¹

¹¹See Judd (1998, Ch. 9) on quasi-Monte Carlo methods for a more thorough discussion of the advantages of using equidistributed sequences to execute simulation-based methods

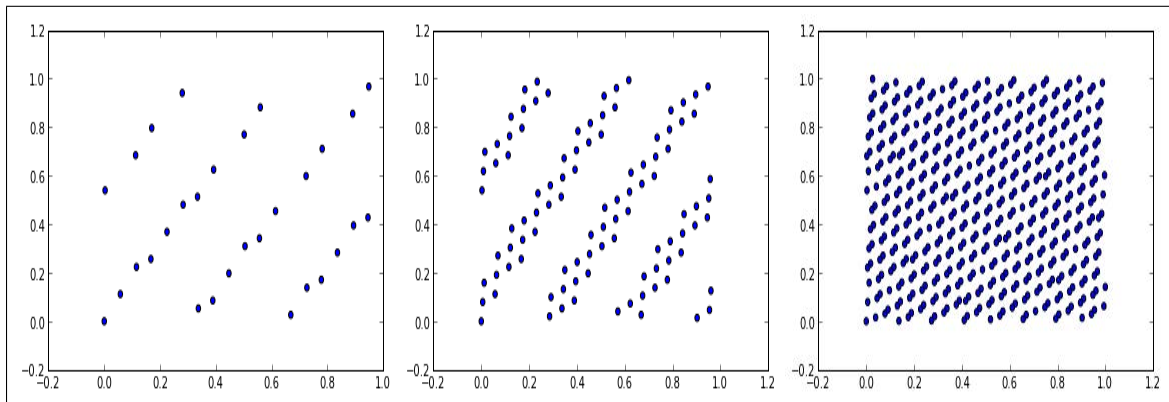
Figure 16: Four two-dimensional equidistributed sequences with $n = 10\,000$



A key distinction between equidistributed sequences and pseudorandom sequences is that equidistributed sequences do not “look like” random numbers. As can be seen in Figures 16 and 17, they generally display substantial serial correlation. From the outset, equidistributed sequences are chosen so as to perform accurate integration, and are not encumbered by any other requirements of random numbers

There are other practical advantages to using equidistributed sequences in this setting aside from the merits of quasi-Monte Carlo integration. Using equidistributed sequences enables us to represent the entire space of households and tax policies as a one-dimensional $M \times N$ list, which allows for easy partitioning across computing nodes. Additionally, using equidistributed sequences makes for easy expansion of the database. One has merely to append additional points to the end of the list.

Figure 17: Two-dimensional Baker sequence on $[0, 1]^2$ for $n = 30$, $n = 100$, and $n = 1000$



T-3 Fitting polynomials to individual good i tax policies on the frontier

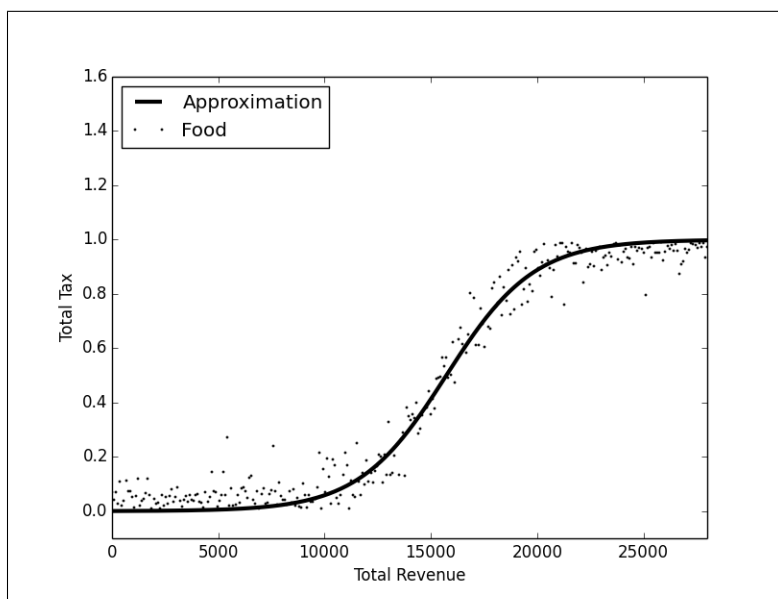
Each point in the total utility-total revenue frontiers shown in Figures 4 and 7 represents a tax policy $\tau = \{\tau_1, \tau_2, \dots, \tau_I\}$ that is not strictly dominated by any other policy in terms of both total utility and total revenue. Figure 5 shows how each of the optimal tax rates τ_i changes along the total utility-total revenue frontier. However, due to the higher number of dimensions in policy space, the computed path of taxes is relatively fuzzy compared to the smoothness of the corresponding total utility-total revenue frontier. For this reason, we use logistic function approximations that minimize the L^1 distance to the points on the tax path to produce the curves in Figure 5.

Suppose that there are a total of n points along the total utility-total revenue frontier. Then we have a list of tax policies $\tau_1, \tau_2, \dots, \tau_n$ and the corresponding values for total revenue r_1, r_2, \dots, r_n . We will write the tax on good k corresponding to revenue r_i as $\tau_{i,k}$. We use the following form of the generalized logistic function, parameterized by A , B , and C , to represent $\tau_{i,k}$ as a function of total revenue.

$$f_k(r_i) = \frac{1}{1 + A_k e^{B_k(r_i - C_k)}} \quad (\text{T.3.5})$$

For each good k , we calculate the values of A_k , B_k , and C_k that minimize the sum of the absolute values of the residuals, $\sum_{i=1}^n |\tau_{i,k} - f_k(r_i)|$. Figure 18 shows an example of the logistic function fit for the food category.

Figure 18: Logistic fit of optimal differentiated tax on food category for points on total revenue frontier



T-4 Derivation of price elasticity of demand

We derive the price elasticity of demand of individual household j for consumption good i by solving for the individual demand function c_i without normalizing prices p_i . Household j maximizes,

$$\max_{\{c_{ij}\}_{i=1}^I} \frac{\left(\left[\sum_{i=1}^I \alpha_i (c_{ij} - \bar{c}_i)^{\frac{\eta_j-1}{\eta_j}} \right]^{\frac{\eta_j}{\eta_j-1}} \right)^{1-\gamma} - 1}{1-\gamma} \quad \text{s.t.} \quad \sum_{i=1}^I p_i c_{ij} \leq w_j \quad (\text{T.4.1})$$

Particularly note in the budget constraint that we have not normalized p_i to 1 and that the tax rate is included in p_i . This means that the price elasticity of demand that we compute will be a function of price as influenced by both tax rates and by other factors that might influence price. Also note that we can ignore the boundary of \bar{c}_i because only a few individuals will have that bind and we have adjusted the utility function to be defined over all consumptions positive and negative as described in Appendix A-1. However, these critical values add more complexity to the actual calculation of the price elasticity of demand.

The first order condition for household j gives rise to the following demand function for consumption good i ,

$$c_{ij} = \left(\frac{p_i C_j^{\gamma - \frac{1}{\eta_j}}}{\alpha_i} \right)^{-\eta_j} + \bar{c}_{ij} \quad (\text{T.4.2})$$

where C_j is total consumption for individual j and is the expression in the first parentheses in the utility function in (T.4.1). The derivative of c_{ij} with respect to price p_i is,

$$\frac{\partial c_{ij}}{\partial p_i} = -\eta_j p_i^{\eta_j-1} \alpha_i^{\eta_j} C_j^{1-\gamma\eta_j}. \quad (\text{T.4.3})$$

The price elasticity of demand can then be written as,

$$\begin{aligned} \frac{\partial c_{ij}}{\partial p_i} \frac{p_i}{c_{ij}} &= -\eta_j \frac{p_i^{-\eta_j} \alpha_i^{\eta_j} C_j^{1-\gamma\eta}}{p_i^{-\eta_j} \alpha_i^{\eta_j} C_j^{1-\gamma\eta} + \bar{c}_i} \\ &= -\eta_j \frac{p_i^{-\eta_j} C_j^{1-\gamma\eta}}{p_i^{-\eta_j} C_j^{1-\gamma\eta} + \bar{c}_i \alpha_i^{-\eta_j}} \end{aligned} \quad (\text{T.4.4})$$

If we normalize prices $p_i = 1$, as we have done in our model, the final expression for individual j 's' price elasticity of demand for good i is the following,

$$\frac{\partial c_{ij}}{\partial p_i} \frac{p_i}{c_{ij}} = -\eta_j \frac{C_j^{1-\gamma\eta}}{C_j^{1-\gamma\eta} + \bar{c}_i \alpha_i^{-\eta_j}} \quad (\text{T.4.5})$$

One way to interpret this price elasticity of demand is that the average elasticity (given individual characteristics j) increases with α_i and decreases with \bar{c}_i .