DISCUSSION PAPER

February 2015 ■ RFF DP 14-40

A Microsimulation Model of the Distributional Impacts of Climate Policies

Hal G. Gordon, Dallas Burtraw, and
Roberton C. Williams III

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES FOR THE FUTURE

A Microsimulation Model of the Distributional Impacts of Climate Policies

Hal G. Gordon, Dallas Burtraw, and Roberton C. Williams III

Abstract

Carbon policies introduce potentially uneven cost burdens. Anticipating these outcomes is important for policymakers seeking to achieve an equitable outcome and can be politically important as well. This paper describes the details of a microsimulation model that utilizes the price and quantity changes predicted by economic models of carbon policies to make an estimation of economic incidence by income quintile or state, and potentially across other dimensions. After taking as inputs the aggregate output from partial or general equilibrium economic modeling, the microsimulation model uses data from the Consumer Expenditure Survey (CE), the State Energy Data System (SEDS), the National Income and Product Accounts (NIPA), estimations from the Congressional Budget Office (CBO), and the Haiku electricty model. These data sources are used to estimate the share of consumer and producer surplus changes that accrue to households in each income quintile and state. The model is unique among existing incidence models in its ability to drill down to the level of state incidence and to plug into a wide range of economic models.

Key Words: carbon price, carbon tax, emissions tax, cap and trade, distributional effects, equity, efficiency, incidence

JEL Classification Numbers: H22, H23, Q52, Q54

Contents

A Microsimulation Model of the Distributional Impacts of Climate Policies

Hal G. Gordon,[a] Dallas Burtraw,[a] and Roberton C. Williams III [a,b,c,*]

1. Introduction

When a new tax on greenhouse gas emissions is imposed, a new price is introduced through a cap-and-trade program, or prices are affected by the introduction of regulation, the economic welfare of households is affected through changes in product prices, tax obligations, and changes in income. The measure of the distribution of the changes in economic welfare on households is the incidence. Measures of incidence are always important to policymakers because of their implications for equity across income, geography, age, or other characteristics.

The policy we focus on in this paper is a tax on carbon emissions, but the methods and model we describe are applicable to other policies. Some taxes are designed, with varying levels of success, so that their incidence falls on those who gain from the use of the government revenue (e.g., gasoline taxes that pay for highway improvements), while others are designed to limit incidence on the poor (e.g., sales taxes that omit clothing and food). Pigouvian taxes on activities with a negative externality are often designed without the tax incidence in mind. For example, cigarette taxes have successfully reduced smoking, but their incidence has largely fallen on the poor.

Like cigarette taxes, carbon taxes are likely to be regressive, since expenditures on energy usually represent a larger share of income for low-income households. More broadly, while the environmental benefits of carbon taxes are diffuse across incomes, borders, age, and even time, its costs are likely to be concentrated in certain regions or segments of society. If the incidence is strongly concentrated, this creates strong incentives for those segments to oppose the tax politically, and the perceived regressivity of a carbon tax is one of its largest political hurdles (Hsu 2011).

Exempting certain protected classes from a carbon tax would present an enormous administrative challenge, but policymakers have one significant way to alter the incidence of a

* [a]Resources for the Future, [b]University of Maryland, [c]National Bureau of Economic Research. Model development was funded by RFF’s Center for Energy and Climate Economics and Mistra’s INDIGO program. The authors appreciate the excellent assistance of Samuel Grausz and Daniel Velez-Lopez in model development.

carbon tax—how they choose to use the revenue. In addition to providing environmental benefits, carbon taxes generate large amounts of revenue that can be returned to households in a lump sum, used to offset other taxes, pay for other government programs, or reduce the government's deficit. Each use will greatly alter the incidence and efficiency of the tax, which increases the importance of correctly estimating tax incidence.

Tax models typically are equipped to investigate economic efficiency, but often they cannot make an incidence estimate. While incidence is closely related to efficiency, analyses that examine tax proposals only for efficiency can produce misleading results because the aggregate effect of the policy may fall disproportionately across the population. For example, some research suggests that if the revenue from a carbon tax displaces a distortionary enough tax, a double dividend can occur, meaning that it has not only environmental benefits but also efficiency advantage, because the welfare loss from the carbon tax is less than the efficiency gain from using the tax revenues to offset other taxes. Yet only a good measure of the distribution of incidence can tell us whether this double dividend occurs for most households or the gains are concentrated in a minority of households. An economically efficient policy that has a wide range of incidence outcomes may not be as politically feasible as a less efficient policy with a more equitable outcome.

The model described in this paper, referred to as the microsimulation model, is designed to use the outputs, in the form of changes in consumer expenditures and incomes, from general equilibrium or partial equilibrium models (which we refer to as the economic model) of a new price or regulation to create an estimate of the incidence by income quintiles and states. The microsimulation model is designed to be interchangeable with many economic models, but we illustrate how it works with a dynamic general-equilibrium overlapping-generations (OLG) model of the US economy from Carbone et al. (2012), which we refer to as the OLG model.

2. Literature

The simplest approaches to assign the incidence of a carbon tax have used input-output tables to calculate the carbon intensity of goods, and then used estimates from the Consumer Expenditure Survey (CE) to map carbon emissions into the consumption portfolio of households to determine the carbon footprint of households by income group, geography, or age. These analyses have found that a carbon tax is regressive, bu, when emissions from all goods in the economy are measured, it is less regressive than the simplest estimates that are limited to energy goods (Hassett et al. 2009; Grainger and Kolstad 2010). In their detailed review of this literature, Morris and Munnings (2013) find that even before considering tax revenue redistribution, most recent studies, using more appropriate methods and measurements of incidence, have estimated a much lower rate of regressivity. For instance, some models take into account the fact that most transfer payments are indexed to inflation and changes in relative prices show less of an impact on low-income households (Blonz et al. 2012).

Research has sought to reduce the remaining estimated regressive outcomes through proposals for the uses of tax revenue. A small proportion of the revenue (11 to 12 percent) could be used to reduce the incidence on the lowest quintile to nothing through targeted direct rebates or tax reductions (Dinan 2012; Mathur and Morris 2012). Targeted programs that concentrate revenue recycling with low-income households have lower economic efficiency outcomes than lump-sum rebate policies that spread revenues more evenly, however (Rausch et al. 2009). Still, lump-sum rebates can have neutral or even progressive outcomes (Boyce and Riddle 2007).

A lump-sum divided ignores the substantial efficiency advantages that computable general equilibrium (CGE) modeling has estimated when the revenue from a carbon tax is used to reduce preexisting distortionary taxes (Parry et al. 1999). With this result in mind, many simpler models have estimated the effects of recycling revenues via tax cuts by allocating efficiency gains according to prepolicy distributions of tax burden. These models find that there are often trade-offs between the equity of the outcome and efficiency, but that middle-ground options exist (Dinan and Rogers 2002; Metcalf 2009; Parry and Williams 2010). For instance, while reducing capital taxes is likely to be the most economically efficient option, it primarily benefits the owners of capital, and even after residual benefits to wage earners materialize over time, the poorest households see little benefits (Mathur and Morris 2012). The literature also has applied this methodology to regional impacts by using regional electricity price models to simulate geographic diversity in carbon intensity (Burtraw et al. 2009). Most studies find that while energy use patterns vary widely by region, the outcomes for households vary only slightly (Mathur and Morris 2012; Hassett et al. 2009).

These previous models mostly focus only on how the tax affects the price of consumption goods; only some include changes to consumption quantity due to price changes. Typically, the partial equilibrium models do not account for how income is affected by the carbon price, the distribution of revenues from a tax, or how consumption changes with changes in income. Introducing a price on emissions will have indirect effects on income that vary across households depending on their sources of income (Fullerton and Karney 2009), and using the revenue to change other taxes also will have effects on income, consumption, and investment decisions. Finally, consumers will switch away from higher-priced energy-intensive goods, changing the consumption pattern of the entire economy. CGE models are designed to capture these changes.

Rausch et al. (2011) use a CGE model to estimate the incidence of various carbon policies on income groups and regions. This model is static, however, so it can model only a long-run equilibrium. A handful of dynamic models exist that include multiple agents (Jorgenson et al. 2012) and thus can estimate distributional incidence, but this requires strong assumptions about agents' preferences in order to make the problem computationally tractable. Lanz and Rausch's general equilibrium model (2014) calculates the difference in regional and income decile incidence when carbon permits are either auctioned or allocated for free to emitters.

The microsimulation model described in this paper presents a number of advances over the incidence methods previously developed. Because it can be used with CGE models, it can estimate the incidence from changes to income that is identified by a CGE model. We develop a new method to refine the data from the Consumer Expenditure Survey, using state-level direct energy consumption data from EIA and output from the Haiku electricity model, to improve the geographic accuracy of incidence at the state level. The microsimulation model can be attached relatively easily to any economy-wide or sector-specific model of a carbon tax that predicts changes in price and income. When new models are built that propose new policy solutions or improve on our current estimations, an incidence calculation can be easily derived with minimal alterations to the new model.

3. Inputs from the Economic Model

The microsimulation model uses the output from an economic model, which simulates the effects of a policy on price and quantities of one, many, or all goods in the US economy. For purposes of linking to the microsimulation model, economic models vary in two ways. First, partial equilibrium models report prices and quantities and their changes as exact values, while general equilibrium models use a numeraire good to index prices and report percent changes in price and quantity. Second, models vary in how much of the economy is covered (with partial equilibrium models possibly covering only some goods) and how the consumption and income goods are categorized (with partial equilibrium models typically providing greater sector-specific detail).

The economic model we describe in this paper is the Carbone et al. (2012) OLG model. The model produces a solution in 5-year time steps over a 100-year horizon (producing 21 solutions). Three different revenue recycling policies are simulated: carbon tax reduction, labor tax reduction, and a lump-sum rebate. For each policy and time period, we use the following outputs from the model:

- Baseline expenditures (x_i) and income (y_j) in all time periods for all 17 commodity goods (i) and 8 income sources (j) in the economy.
- Percent changes in price and quantity of the 17 commodities ($\frac{\Delta p_i}{p_i}, \frac{\Delta q_i}{q_i}$) and 6 of 8 income sources ($\frac{\Delta p_j}{p_j}, \frac{\Delta q_j}{q_j}$) due to the policy.
- Absolute changes in government transfer income (Δy_g) and carbon tax dividend (Δy_d) due to the policy. (By construction, the level of transfer income is explicitly controlled in the model so the percent changes in price and quantity of transfer income are not a model solution.)

Through the paper, we use this model as an example of an economic model that can be linked to the incidence model. Note that this economic model reports percent changes in prices and quantities, so calculations of changes in consumer and producer surpluses are made using those variables. Economic models producing numeric values of exact changes use a slightly different equation. In addition, this economic model reports outcomes for a relatively large amount of goods and income categories. Alternate models may report at a higher or lower level of detail, which requires use (or creation) of a different data crosswalk to link the economic model to the microsimulation incidence model and provides different levels of detail in the final results.

4. Economy-Wide Welfare Change

The goal is to measure incidence (distribution of burden) across households from specific policies that introduce prices or regulations that affect consumer expenditures and incomes. The output from the economic model is used to calculate incidence from the policy for each good and source of income included in the economic model. Incidence is estimated by changes to welfare due to the policy and is approximated by two components: changes to consumer surplus stemming from changes in the price and quantity of commodities; and changes to producer surplus stemming from changes in sources of income, which are represented by changes in the price and quantity of income goods. Summing the changes in consumer and producer surplus to calculate welfare closely approximates the measure of welfare change that is computed as equivalent variation in the OLG model.

In order to calculate changes in consumer surplus (Δcs) and producer surplus (Δps), we assume that the demand and supply curves are linear over the relevant range where changes are observed (the curves are not necessarily linear elsewhere) and calculate the change in the area of the surplus due to the change in price and quantity (see Figure 1, Panel A).

The change in consumer surplus (Δcs) is represented by the area $\triangle ADE - \triangle ABC$ in Panel A of Figure 1. The points C and E might be characterized as points along a partial equilibrium demand curve, in which case the slope in Figure 1 can be interpreted as own-price elasticity. Alternatively, a system of demand curves might be solved to provide new equilibrium outcomes, or a general equilibrium model might be solved, yielding equilibrium outcomes that incorporate changes in the prices of other commodities as well as income.

When the model of the economy provides exact values for the changes in price and quantity and the initial quantity, the change in consumer surplus (and producer surplus) can be directly calculated (see the second line in the derivation directly below). A general equilibrium model, like the one we use as an example, will define changes in prices and quantities relative to normalized values, providing estimates of the percent change in quantity and price. This can be mapped into the areas in Panel A of Figure 1 in the following way:

$$\Delta cs = \triangle ADE - \triangle ABC$$

$$\Delta cs = -(\Delta p(q+\Delta q) + \frac{1}{2}\cdot \Delta p \cdot -\Delta q)$$

$$\Delta cs = -(\Delta p \cdot \Delta q + \Delta p \cdot q - \frac{\Delta p \cdot \Delta q}{2})$$

$$\frac{\Delta cs}{p\cdot q} = -\frac{\Delta p \cdot q + \frac{\Delta p \cdot \Delta q}{2}}{p \cdot q}$$

$$\frac{\Delta cs}{p\cdot q} = -(\frac{\Delta p}{p} + \frac{\Delta p \cdot \Delta q}{2(p\cdot q)})$$

$$\Delta cs = -(1+\frac{\Delta q}{2q})\cdot \frac{\Delta p}{p}\cdot (p\cdot q)$$

The change in consumer surplus based on percent change in quantity and price for each commodity good is therefore written as follows, with x_i representing baseline expenditure on commodity i :

(1) $$\forall \textit{ commodity goods, i: } \Delta cs_i = -(1+\frac{\Delta q_i}{q_i}\cdot\frac{1}{2})\cdot\frac{\Delta p_i}{p_i}\cdot x_i$$

For the income goods, we calculate a change in producer surplus (Δps), which is the difference in area between the supply curve and the price line, or the area $\triangle ADE - \triangle ABC$ in Panel B of Figure 1. Just as for demand goods, where the model of the economy provides exact values for prices and quantities, the area can be calculated directly. If the model provides only percent change in prices and quantities, this can be mapped into the calculation of change in producer surplus as follows:

$$\Delta ps = \triangle ADE - \triangle ABC$$

$$\Delta ps = \Delta p \cdot q + \frac{1}{2} \cdot \Delta p \cdot \Delta q$$

$$\frac{\Delta ps}{p \cdot q} = \frac{\Delta p \cdot q + \frac{\Delta p \cdot \Delta q}{2}}{p \cdot q}$$

$$\frac{\Delta ps}{p \cdot q} = \frac{\Delta p}{p} + \frac{\Delta p \cdot \Delta q}{2(p \cdot q)}$$

$$\Delta ps = (1 + \frac{\Delta q}{2q}) \cdot \frac{\Delta p}{p} \cdot (p \cdot q)$$

The change in producer surplus based on percent change in quantity and price for each commodity good is therefore written as follows, with y_j representing baseline income from source j:

(2) $$\forall \text{ income goods } j \neq g,d: \ \Delta ps_j = (1 + \frac{\Delta q_j}{q_j} \cdot \frac{1}{2}) \cdot \frac{\Delta p_j}{p_j} \cdot y_j$$

Notice that equation 2 has the same form as equation 1. For transfer income (y_g) and income that might come from climate policy dividends (y_d),[1] there is no relevant definition for quantity. Because of this, the economic model provides exact values of changes to transfer or dividend income, which are used to directly predict changes in producer surplus.

(3) $$\Delta ps_g = \Delta y_g$$

(4) $$\Delta ps_d = \Delta y_d$$

The incidence of the policy is expressed as the change in welfare for the entire economy and is equal to the change in consumer surplus plus the change in producer surplus (a negative result is bad for households):

---

[1] Dividend income is the value of tax revenue returned to households from the income generated from the policy. Baseline income from dividends without the policy is always zero.

(5) $$\Delta W = \Delta CS + \Delta PS$$

where

(6) $$\Delta CS = \sum \Delta cs_i$$

(7) $$\Delta PS = \sum \Delta ps_j$$

Thus the policy incidence is

(8) $$\Delta W = \sum \Delta cs_i + \sum \Delta ps_j$$

5. Calculating Incidence across States and Income Quintiles

Solving for the change in welfare for the entire economy is possible using output from the chosen economic model, but additional sources of data are required to solve for the distribution of changes in welfare from the policy by state and income quintile.[2] We denote changes in welfare across these indexes as ΔW^k, where k is an index of either states or income quintiles, taken one at a time. For all goods except electricity, we assume that the percent changes to prices and quantities due to the policy are not different across states or income quintiles. However, the initial values for the levels of expenditures for each commodity (x_i^k) and source of income (y_j^k) differ across these groups. Using the changes to consumer and producer surplus indicated previously, we apply the following equations to proportionately assign changes across states and income groups according to the baseline level of consumption for each good and for each source of income:

(9) $$\forall \textit{commodity goods, } i \neq \textit{electricity and } \forall \textit{states or income quintiles, k: } \Delta cs_i^k = \Delta cs_i \cdot \frac{x_i^k}{x_i}$$

(10) $$\forall \textit{income goods, } j \neq \textit{dividend and } \forall \textit{states or income quintiles, k: } \Delta ps_j^k = \Delta ps_j \cdot \frac{y_j^k}{y_j}$$

[2] Income categories are defined by ranking all people by income, adjusted for household size. Following the practice used by the Congressional Budget Office, that adjustment is implemented by dividing income by the square root of the number of people in a household, which consists of the people who share a housing unit, regardless of their relationships (CBO 2013). Quintiles, or fifths, contain equal numbers of people. Households that have negative income (that is, if their business or investment losses are larger than other income) are excluded from the sample.

The important outside data one needs for this operation is the percentage of expenditures and percentage of income by state and quintile ($\frac{\Delta x_i^k}{x_i}$ and $\frac{\Delta y_j^k}{y_j}$).

For most of the policies that one is likely to investigate, these equations would be sufficient (see section 7.2.2 for changes in consumer surplus by state from electricity). However, one approach that has emerged in the climate policy debate is the return of revenues raised by introducing a price on carbon returned to households as a dividend. One straightforward way this might be implemented is as a per capita (lump sum) dividend distributed proportionally across the population, although in practice various approaches have been proposed or observed. A prominent example of a dividend is the Alaska Permanent Fund, which provides every legal resident of Alaska, regardless of age or citizenship, an equal share of dedicated funds collected by the state from the extraction of mineral resources.[3] This means that dividend income is distributed by population:

$$(11) \qquad \forall \textit{states or income quintiles, k:} \quad \Delta ps_d^k = \Delta ps_d \cdot \frac{\pi^k}{\pi}$$

where $\frac{\pi^k}{\pi}$ is the proportion of the dividend to be distributed to households index k,[4] and Δps_d is the size of the dividend in the economic model (see equation 4). Tax incidence on welfare by index is calculated by summing all consumer and producer surplus values:

$$(12) \qquad \forall \textit{states or income quintiles, k:} \quad \Delta W^k = \sum_i \Delta cs_i^k + \sum_j \Delta ps_j^k$$

In order to calculate changes in consumer and producer surplus for states or income quintiles, represented by equations 9 and 10, and to calculate changes in consumer and producer surplus by our index, k (states or income quintiles), the proportion of each commodity category (i) and each income source (j) for each group ($\frac{\Delta x_i^k}{x_i}$ and $\frac{\Delta y_j^k}{y_j}$) must be identified. Note that all baseline economy-wide expenditure and income data (x_i, y_j) are given by the economic model solution. In the case of the OLG model, the source of those values is the Global Trade Analysis

[3] In California and some other states, similar payments have been made on the basis of electricity customer accounts. Some proposals call for equal payments per adult citizen.

[4] Williams et al. (2014) examine a case where each legal resident received an equal share, so $\frac{\pi^k}{\pi}$ equaled the proportion of US population in each index

Project (GTAP). Other sources that provide greater resolution over index *k* are unlikely to have the same economy-wide expenditure and income levels as the economic model. In fact, other data could provide biased totals for any particular commodity or income source. However, the only assumption necessary is that the shares of the total ($\frac{\Delta x_i^k}{x_i}$ and $\frac{\Delta y_j^k}{y_j}$) are not biased. For example, a data source could overestimate the national level of food expenditures, but as long as this bias is not correlated to an index (states or income quintiles), the estimations can be used in the microsimulation model without propagating the bias. Table 1 summarizes all of the data sources and notation in the microsimulation model. Each column corresponds to one of the sections below (with the first column, covering the economic model, already having been discussed in an earlier section).

6. Changes in Producer Surplus

Two sources of data are used for income shares $\frac{\Delta y_j^k}{y_j}$, one for states and one for income quintiles. Income shares are estimated for three sources: transfers, labor, and capital. These adequately represent the income sources in the OLG economic model and will do so for many economic models. If an economic model that differentiates income from separate labor sources (for example, oil and gas extraction labor) is used, the microsimulation model would need to be amended with a data source that provides categorical labor income by state and quintile. (NIPA Table SQ5 would work for states). It is assumed that the shares of national capital income and shares of resource income are equal for each subgroup of the population, as most of these assets are held as securities in diverse portfolios. Thus, even if an economic model splits resource income from capital income (as the OLG model does), we attribute the same proportion of resource income as we attribute to capital income.

The measures of total income coming from the OLG model (y_j) are after-tax income figures. However, before-tax data are used to calculate income shares. That is, it is assumed that costs and benefits from policies will impact owners in each state or income quintile in proportion to how much labor, capital, or transfers are earned prior to taxes. A reduction in the capital or income tax could be implemented in a variety of ways that would affect how producer surplus gains from capital or income would be distributed to states or, especially, income quintiles. Since most economic models (including the OLG model) do not specify how progressive or regressive these tax reductions would be, we feel the most logical way of distributing the gains would be proportional to the initial state of the world before any taxes. Alternative assumptions and approaches are possible, such as assuming that only marginal income tax rates are reduced, thus limiting the benefit to the higher-income households that pay that tax. Care would have to be taken in reconciling such alternatives with the results coming from the economic model.

6.1. Changes in Producer Surplus by Quintile

Changes in producer surplus to each quintile are assigned (Table 1, column 2) by using estimates produced by the Congressional Budget Office in an annual report (CBO 2013) that uses one administrative record, the Internal Revenue Service's Statistics of Income (SOI), and one survey, the Census Bureau's Current Population Survey (CPS). The SOI does not have reliable information from low-income households (which do not pay federal income taxes), and the CPS underreports income from capital (which is often the case in self-reported surveys) and lacks information on very high-income households, so the CBO estimates attempt to correct for each source's bias. The CBO provides before-tax income estimates by quintile for labor income and transfer income. We sum together CBO estimates of capital gains, business income, capital income, and other income (which is mostly retirement income) by quintile to create capital income estimates in the microsimulation model. CBO estimates for 2010 are used. Data from CBO estimates are identified with an accent mark, so equation 10 can be updated for income quintiles:

$$\forall \textit{ income sources, } j \textit{ and } \forall \textit{ income quintiles, } k\text{:}\ \Delta ps_j^k = \Delta ps_j \cdot \frac{y_j'^k}{y_j'} \tag{13}$$

6.2. Changes in Producer Surplus by State

The National Income and Product Accounts (NIPA) are the national accounts for the US economy, produced by the Bureau of Economic Analysis (BEA 2013), and are used to assign the changes in producer surplus by state (Table 1, column 3). BEA estimates personal income by source for each state on an annual basis primarily by using administrative records data from federal and state government social insurance programs and tax codes, including state unemployment insurance, state Medicaid, federal Medicare, social security, federal veterans' programs, and state and federal income tax data. The estimates for "net earnings by place of residence," "dividends, interest, and rent," and "personal current transfer receipts" are used for labor, capital, and transfer income, respectively. Since the NIPA estimates are considered to be the authoritative source on the US economy and use actual administrative records, rather than survey data, these estimates are unlikely to be biased by state. The annual before-taxes estimates for 2013 are used. NIPA estimates are identified with two bars, so equation 10 can be updated for states:

$$(14) \qquad \forall \textit{ income sources, } j \textit{ and } \forall \textit{ states, } k\text{: } \Delta ps_j^k = \Delta ps_j \cdot \frac{\bar{\bar{y}}_j^k}{\bar{\bar{y}}_j}$$

NIPA income is a measure of personal income, which is a more complete measure of income than monetary income (census estimates of household income are based on the latter). Personal income includes not only monetary income but also income that does not come in the form of money, such as imputed rent (i.e., the value from living in a house one owns, rather than having to pay rent), the value of employer-provided health insurance and retirement contributions, and a range of other nonmonetary income sources (see Ruser et al. 2004 for more details). Because most economic models use personal income, the mean income per household in each quintile and each state is likely to seem high to readers who are more used to the estimates of household income in census reports.

7. Changes in Consumer Surplus

Comprehensive administrative data sources for expenditures by states or income quintiles do not exist, so we rely on the Consumer Expenditure Survey (CE), a quarterly survey of randomly sampled households from 91 geographically contiguous primary sampling areas (PSUs) across the United States conducted by the Bureau of Labor Statistics (BLS 2012b). The CE is actually a composite of two surveys, a quarterly interview survey and a two-week-long diary survey, which seek to record everything a household consumes during the survey period along with demographic and socioeconomic information. In the interview survey, a BLS survey administer visits a household once every quarter for five consecutive quarters. The diary survey involves each household self-reporting expenditures over a two-week time period by recording every purchase as it takes place. Each survey includes about 7,000 households a year, and a survey weight (determined by demographic attributes and the likelihood of each home being surveyed) is assigned to each respondent to make the survey as representative of the nation as possible. While each survey attempts to cover all expenditures, BLS has determined which survey should be used for each category of goods (for instance, the diary survey is better at recording daily consumption, such as food purchases, and the interview survey is better at reporting large purchases, such as automobiles). The CE categorizes consumption using the Universal Classification Code (UCC), and after weights are applied to each observation in both surveys, expenditures for each UCC category are summed from either the interview survey or the diary survey, depending on which source is preferred by BLS for each good. After dropping irrelevant consumption categories, 777 UCC categories remain. This high level of granularity of consumption means that almost any categorization used by an economic model can be matched with the right crosswalk.

Unlike the NIPA and CBO estimates used for our income source estimation, CE data are strictly survey data and are not based on government administrative records, so the data potentially have systematic bias. BEA produces the Personal Consumption Expenditures (PCE) data, which are a component of NIPA and are nationwide estimates of household expenditures that are considered to be more accurate than CE data. However, PCE data are not broken out by states or income quintiles. For those UCC codes (CE data) and PCE codes (NIPA data) that are most comparable, some goods are consistently underreported by the CE (Passero et al. 2013). Some of these differences are clearly due to underreporting of morally questionable consumption (e.g., only $1 of gambling expenditures is reported in the CE for every $10 estimated in the PCE), but some also can be attributed to the large mental and time burden involved in recalling all consumption—it is hard to remember everything, and some people may limit their responses to make the survey end more quickly (Shields and To 2005). The fact that specific goods are underreported does not necessarily bias the proportion of expenditures that occurs in each index, however, as long as people across states and income quintiles underreport expenditures of each good at similar rates, which is our assumption.

7.1. Changes in Consumer Surplus by Quintile

To characterize consumption by income quintiles, we make no attempt to correct for potential inherent bias in the CE data because we assert that the systemic bias in total consumption by UCC code that may be present is not correlated to income. We use the publicly available microdata from 2004 to 2011 (we use many years' worth of data to increase the survey sample size) and create equally sized quintiles by dividing income by the square of the number of people in each household. CE data are identified with a tilde, and the index of UCC categories is represented by u , so we notate the proportion of expenditures of good u occurring in quintile k as $\frac{\tilde{x}_u^k}{\tilde{x}_u}$. These ratios cannot yet be substituted into equation 9 to estimate the consumer surplus change by quintile because UCC codes (u) do not align with the commodity goods (i) that are used in the OLG economic model, so a crosswalk will be required (see Table 1, column 4). If the economic model being used does not require a data crosswalk to connect to UCC codes, equation 9 could be updated at this point for income quintiles.

7.2. Changes in Consumer Surplus by State

More work is needed to create estimates of the proportion of expenditures by states because there are a number of clear sources of geographic bias in the CE survey. In fact, the BLS maintains that the CE microdata are not meant to be representative of state-level data, but are instead representative only of the region level (West, East, North, Midwest). In addition to this bias, it is likely that electricity prices will change differently in different states in response to a carbon policy, so a key assumption of equation 9 (that price changes are equivalent across the index of states or quintiles) does not hold for this good, and thus additional work is needed.

To accurately estimate consumer surplus changes by state, we rely on three sources of data. For the most goods that have the largest price and quantity changes, direct energy goods, we rely on the Energy Information Agency's (EIA) State Energy Data System (SEDS), which provides state estimates of expenditures from administrative reports (Table 1, column 5). For electricity consumer surplus changes, we use output from Haiku, an electricity industry model (Table 1, column 6). For all other consumption, we alter CE data to build plausible estimates for expenditures by state.

7.2.1. Changes in Consumer Surplus by State from Direct Energy Sources

Expenditures on electricity, natural gas, other fuel oils, and gasoline are of particular importance to the microsimulation model because they represent direct expenditure on energy goods and services and are directly subject to the carbon policy. These goods do exist in the CE and suffer from far less bias than some other goods. It is relatively easy to report a household's utility bills to a survey worker except when utilities are included in rent. Nonetheless, the EIA's SEDS estimates are superior. SEDS estimates are highly accurate over the general population and provide statewide estimates of retail electricity, natural gas, and other fuel oils (fuel oil, kerosene, liquefied petroleum gas) residential expenditures (EIA 2012). SEDS provides statewide gasoline expenditures for all transportation uses (residential expenditures are not reported separately). The primary data sources for these estimates are compulsory surveys of utilities and petroleum refiners that are collected by the EIA, and thus the estimates are likely to closely reflect reality.

The main potential for state-level bias in SEDS is in the gasoline estimates. Since SEDS does not provide an estimate of gasoline expenditures for residential customers, we must assume that the proportion of all motor gasoline that is purchased by households is the same in each state. We are unaware of any existing data on this question, but it stands to reason that the factors that drive commercial and government expenditures on motor gasoline (economic activity, density, state fuel taxes) affect household purchases in the same direction. We thus assume this potential bias is minimal.

Because the 2011 SEDS estimates are likely to be less biased than CE estimates and have complete data for all states, we use these estimates (signified with an upside-down hat) to calculate the proportion of expenditures of direct energy goods by state.

We use 2011 SEDS estimates to replace CE data for UCC direct energy categories. Unlike with CE data, we cannot find any systematic bias correlated to geography with the SEDS estimates, so no more data manipulation is necessary. We cannot yet substitute an estimate of the ratio of expenditures on direct energy goods across states ($\frac{\check{x}_u^k}{\check{x}_u}$, where the index k denotes states and u denotes UCC direct energy goods categories). Because UCC codes (u) do not align with the commodity goods (i) that are used in the OLG economic model, a crosswalk is required before changes in consumer surplus are calculated. SEDS provides no estimates of energy expenditures by income quintile, so SEDS data are used only for our state-level estimates.

7.2.2. Changes in Consumer Surplus by State from Electricity

In equation 9, we assume that the percent change in price from a carbon policy and the behavioral response to a percent change in price would be the same across the country and income spectrum. In one case, this assumption clearly does not hold. The carbon content of electricity generation and consumption is very different across states, so price and quantity changes due to the policy vary by geography. A separate electricity industry model, Haiku, is used to calculate a change in consumer surplus from electricity by state (Δcs_e^k). The Haiku electricity market model identifies equilibrium in electricity, fuel, and environmental markets in 22 linked regions of the country over a 25-year horizon (Paul et al. 2013).

Dependent variables are time dependent and anticipate the path of compliance with other environmental regulations and changing market conditions. The Haiku model solves for a specific set of simulation years. For illustration, in a recent exercise, the model solved for 2015, 2017, 2020, 2027, and 2035. The model provides absolute and percent changes in electricity price and quantity consumed due to a social cost of carbon policy by state ($\frac{\Delta\bar{p}_e^{k,t}}{\bar{p}_e^{k,t}}, \frac{\Delta\bar{q}_e^{k,t}}{\bar{q}_e^{k,t}}$) as well as electricity expenditures for each time period in a baseline forecast without the policy ($\bar{x}_e^{k,t}$), where k is the index of states, e refers to electricity, and t is the index of time periods for which Haiku calculates estimates. Data coming from the Haiku model are identified with a single bar.

Modeling results from Haiku may not correspond directly to the output from the economic model because it may solve for different years or have a different specification of the climate policy. For example, the carbon tax in the OLG economic model is $30/ton and does not change over time. In Haiku, the carbon price increases over time to improve the equilibrium calculation in that model but averages $30/ton between 2015 and 2035. Moreover, in the Haiku model, the geographic location of investments in abatement are partly dependent on the price of carbon and the variation of the marginal cost of abatement in each state, but also on the path of compliance, so no one year from the Haiku model provides the best approximation of the policy from the example economic model. Thus, in this example, we interpolate the change in price and quantity for each state between 2015 and 2035 and use these values for calculating the distribution of consumer surplus changes across states for every time period of the economic model. If an economic model is used that has a different social cost of carbon, different Haiku model results that reflect this price should be sought or further adjustments would be necessary.

The total loss in consumer surplus due to the policy in the electricity sector (Δcs_e) is given by the economic model using equation 1. The national percent changes in price and quantity of electricity in the Haiku model are not likely to be equivalent to those coming from the economic model, so the resulting national change in consumer surplus will be different ($\Delta cs_e \neq \Delta \overline{cs}_e$).[5] Losses to consumer surplus in each year over the 20 years of the Haiku model are calculated and summed, and the proportion of consumer surplus change from each state is used to assign the proportion of the economic model's consumer surplus change to each state.

To illustrate, first Haiku consumer surplus changes by state and year are calculated:[6]

$$(15)\quad \forall \textit{ states, } k \neq AK \textit{ or } HI \textit{ and } \forall \textit{ years, } 0 \leq t \leq 20\text{:}\ \ \Delta \overline{cs}_e^{k,t} = -(1 + \frac{\Delta \overline{q}_e^{k,t}}{\overline{q}_e^{k,t}} \cdot \frac{1}{2}) \cdot \frac{\Delta \overline{p}_e^{k,t}}{\overline{p}_e^{k,t}} \cdot \overline{x}_e^{k,t}$$

Note that the subscript e reminds us that this equation is applicable only to electricity. As noted in the previous section, the best estimates for the initial statewide residential electricity

5 It is possible to scale the quantity percent changes so the national quantity percent change in Haiku matches the economic model. However, because changes to the national price are affected by both changes in state price *and* state quantity (national price is a weighted average price over all electricity consumed, not the average of state prices), it is impossible to evenly scale state prices in such a way that the national price change is equal to a predetermined level (assuming all quantity percent changes are not equivalent).

6 Alaska and Hawaii are not included in Haiku, so we use SEDS expenditure proportions and the national price changes from the economic model to calculate CS losses for these states.

expenditures are SEDS estimates because they are based on administrative data. SEDS expenditure estimates from 2012 (identified with an upside-down hat) are substituted for the Haiku expenditure estimates for the first year (2015, $t=0$) of Haiku results ($\breve{x}_e^{k,0}$ replaces $\overline{x}_e^{k,0}$). In order to solve equation 15 for years after t=0, SEDS expenditures by states estimates need to be updated for each year from 2016 to 2035. The initial SEDS estimates are scaled up and down by the pattern over time, as reflected in the changes projected in the Haiku expenditures estimates:

$$(16) \qquad \forall\ \textit{states, } k \neq AK \textit{ or } HI \textit{ and } \forall\ \textit{years, } 0 \leq \mathrm{t} \leq 20\textit{:}\ \breve{x}_e^{k,t} = \breve{x}_e^{k,0} \cdot \left(1 + \frac{\overline{x}_e^{k,t} - \overline{x}_e^{k,0}}{\overline{x}_e^{k,0}}\right)$$

Inserting the SEDS-based expenditure estimates, we update equation 15:

$$(17) \quad \forall\ \textit{states, } k \neq AK \textit{ or } HI \textit{ and } \forall\ \textit{years, } 0 \leq \mathrm{t} \leq 20\textit{:}\ \Delta\overline{cs}_e^{k,t} = -\left(1 + \frac{\Delta\overline{q}_e^{k,t}}{\overline{q}_e^{k,t}} \cdot \frac{1}{2}\right) \cdot \frac{\Delta\overline{p}_e^{k,t}}{\overline{p}_e^{k,t}} \cdot \breve{x}_e^{k,t}$$

Then Haiku consumer surplus changes for each state for the whole time period for which Haiku reports results are calculated:

$$(18) \qquad \forall\ \textit{states, } k \neq AK \textit{ or } HI\text{:}\ \Delta\overline{cs}_e^{k} = \sum_t \Delta\overline{cs}_e^{k,t}$$

National Haiku consumer surplus change is the sum of all states:

$$(19) \qquad \Delta\overline{cs}_e = \sum_k \Delta\overline{cs}_e^{k}$$

Because Haiku does not provide estimates for Hawaii and Alaska, one might assume that the price and quantity changes are equivalent to the national averages, or one could adjust the national average based on the carbon intensity of electricity generation in those states. Only SEDS expenditures are used to determine the proportion of consumer surplus change that occurs in each of these states:

$$(20) \qquad \forall\ \textit{states, } k = AK \textit{ or } HI\text{:}\ \Delta cs_e^{k} = \Delta cs_e \cdot \frac{\breve{x}_e^{k}}{\breve{x}_e}$$

Finally, the loss in consumer surplus from electricity by state is

$$(21) \qquad \forall\ \textit{states, } k \neq AK \textit{ or } HI\text{:}\ \Delta cs_e^{k} = (\Delta cs_e - \Delta cs_e^{AK} - \Delta cs_e^{HI}) \cdot \frac{\Delta\overline{cs}_e^{k}}{\Delta\overline{cs}_e}$$

Unlike all other goods, the electricity commodity category in the OLG economic model is equivalent to the consumption category in our state data, so no crosswalk is needed (commodity electricity is equivalent to consumption electricity).[7]

7.2.3. Changes in Consumer Surplus by State from All Other Goods

We alter CE data to estimate all other consumption by state. Raw CE data do not accurately estimate consumption by state due to nonrandom sampling; that is, respondents are not drawn from geographically random areas. Instead, each year, the sample is drawn from 91 geographically contiguous primary sampling units (PSUs) that have remained the same from year to year since 2006 (the last year the PSUs were changed). If a certain part of a state is not in a PSU, its residents are never included in the CE. Seven states (IA, NM, ND, MS, OK, RI, and VT) have not been included in any PSU since 2006. Of the states that are included in the CE, some are either over- or undersampled. Since the BLS weights are meant only to achieve accurate expenditure levels at regional and national levels, a state that is oversampled will also be oversampled after those weights are applied. While this does not bias the expenditures per household ($\frac{\tilde{X}^k}{\tilde{h}^k}$), the total expenditures per state ($\tilde{X}^k$) and total households per state ($\tilde{h}^k$) are easily biased up or down by over- or undersampling (recall that data from the CE survey are identified with a tilde).

In addition to nonrandom geographic sampling, the ability to use publicly available microdata to make accurate state-level estimates suffers from geographic top coding, or veiling of data. If a household can be identified as coming from a geographic area of 100,000 people or fewer, the state and region indicators are not reported. Of the original sample from 2004 to 2011, 13 percent (or 14.6 percent after weights are applied) of households have top-coded state identifying variables. In some cases, entire states are top-coded (all households from AR, MS, MT, NC, and SD are top-coded). Of the 14.6 percent of the weighted sample that was top-coded, 85.5 percent did not live in a metropolitan core-based statistical area (CBSA; these are defined by the Office of Management and Budget). Of the 14.8 percent of the weighted sample that do not live in a metropolitan CBSA, 84.8 percent were top-coded. Not only does top coding reduce

[7] For example, crosswalks are needed for gasoline because SEDS data refer to consumption of refined gasoline, the manufacture of which involves a number of GTAP commodities (the categories in which the OLG economic model reports price changes), including refined petroleum, chemical products, and transportation.

the number of states with attributable data, but it also adds more geographically correlated bias because it is not randomly done.

Because of these problems, the CE estimates of the proportion of expenditures across states ($\frac{\tilde{x}_u^k}{\tilde{x}_u}$, where k is an index for states and u is an index for UCC consumption goods) are potentially biased. We first correct for under- or oversampling of states, next create an estimate for top-coded areas, and then construct estimates for states that are completely missing using regional data. Finally, we correct for under- or oversampling of populations within states by scaling up or down each state's CE total consumption data so that they are more representative of the state's income.

7.2.4. First Source of Potential Bias: Under- or Over-sampling of States

Since some states are over- or undersampled in the CE data, there are too many or too few households (and thus too many or too few expenditures) per state. In order to correct over- or underrepresentation of the number of households per state in the CE, we use the count of households per state in the 2010 census (h^k) to scale up or down expenditures:

(22) $$\forall \textit{ UCC consumption goods, } u \neq \textit{direct energy and } \forall \textit{ states, } k\text{: } \ddot{\tilde{x}}_u^k = h^k \cdot \frac{\tilde{x}_u^k}{\tilde{h}^k}$$

where $\tilde{h}^k$ represents the households per state in CE, h^k is the census households per state, and $\ddot{\tilde{x}}_u^k$ is the adjusted CE estimates (since values taken from the CE are notated with a tilde, we notate adjusted CE data with a tilde and two dots).

7.2.5. Second Source of Potential Bias: Top-Coded Data

As previously reported, 14.8 percent of households represented in the CE publicly available microdata are stripped of their state and region identifying variables. In order to assign the top-coded expenditures to geographic regions, we must estimate the proportion of households in each region that are top-coded. Understanding exactly why each observation is or is not top-coded is impossible with the publicly available microdata, but some assumptions can be made that will reduce state-by-state bias. As indicated previously, a large portion, 85.5 percent of all top-coded observation,s did not live in a metropolitan CBSA, while 84.8 percent of all observations that did not live a metropolitan CBSA were top-coded (Figure 2 displays the overlap of all observations). No other variable in the microdata is so closely related with top coding, so in order to create actionable rules for assigning top-coded data, we assume that all top-coded households are not living in metropolitan CBSAs and that all households not living in metropolitan CBSAs are top-coded. The result is an assumption that all state attributable data (that is, the data that are not top-coded) represent only the metropolitan areas of each state, while the top-coded data are representative of nonmetropolitan areas.

While it is not possible to identify which state or region any specific top-coded household belongs to, it is possible to split the total top-coded expenditures of each UCC category by the four census regions (East, West, Midwest, and South). The BLS publishes total expenditures by region each year, so subtracting the geographically attributable regional totals from the microdata from these published regional totals allows for a calculation of total expenditures from top-coded households by region and UCC category.

To infer the amount of top-coded expenditures by region:

(23) $\forall$ *UCC consumption goods, u* $\neq$ *direct energy and* $\forall$ *regions, r:* $\tilde{\xi}_u^r = \bar{\tilde{x}}_u^r - \tilde{x}_u^r$

And to infer the number of top-coded households by region:

(24) $\forall$ *regions, r:* $\tilde{\eta}^r = \bar{\tilde{h}}^r - \tilde{h}^r$

where $\bar{\tilde{x}}_u^r$ and $\bar{\tilde{h}}^r$ are total regional expenditures and households in BLS summary tables (we notate data coming from the summary tables with a tilde and a bar), $\tilde{x}_u^r$ and $\tilde{h}^r$ are non-top-coded regional expenditures and households in the microdata, and $\tilde{\xi}_u^r$ and $\tilde{\eta}^r$ are regional expenditures and households for nonmetropolitan CBSAs (top-coded observations).

For the 38 states and 1 district that have attributable households in the CE microdata, we combine the state expenditures (which we assume are representative of only metropolitan areas) with regional top-coded data in a way that is proportional to the urban-rural makeup of each state to create a less biased estimate of state expenditures. In other words, we assume the nonmetropolitan population throughout a region has similar expenditures, so equation 22 is updated:

(25) $\forall$ *consumption goods, u* $\neq$ *direct energy and* $\forall$ *non top-coded states and regions, k and r:*

$$\ddot{\tilde{x}}_u^k = h^k\left(\mu^k \cdot \frac{\tilde{x}_u^k}{\tilde{h}^k} + (1-\mu^k)\frac{\tilde{\xi}_u^r}{\tilde{\eta}^r}\right)$$

where μ^k is the proportion of each state that lives in a metropolitan CBSA according to the 2010 census and $\ddot{\tilde{x}}_u^k$ represents the adjusted CE estimates. To review, h^k is the number of households per state according to census, while $\frac{\tilde{x}_u^k}{\tilde{h}^k}$ are the expenditures per household on good u in state k

’s metropolitan areas, and $\frac{\tilde{x}_u^k}{\tilde{x}_u}$ are the expenditures per household on good u in state k ’s nonmetropolitan areas.

7.2.6. Third Source of Potential Bias: States That Are Missing from the CE Survey

Seven states have no primary sampling units (PSUs) in the CE survey, while five more states are fully top-coded and thus have no identifiable observations in the microdata. In order to create expenditure estimates for these missing states, we use expenditure data from other states in the same region that are not top-coded in the microdata and our regional top-coded data. Equation 25 is used, but with regional metropolitan CE expenditures per household replacing the missing state metropolitan expenditures:

(26) $\forall$ *consumption goods, u* $\neq$ *direct energy and* $\forall$ *top-coded states and regions, k and r:*

$$\ddot{\tilde{x}}_u^k = h^k(\mu^k \cdot \frac{\tilde{x}_u^r}{\tilde{h}^r} + (1-\mu^k)\frac{\tilde{\xi}_u^r}{\tilde{\eta}^r})$$

where $\ddot{\tilde{x}}_u^k$ represents the adjusted CE estimates.

7.2.7. Fourth Source of Potential Bias: Within-State Nonrandom Sampling

At this point, we have two types of state estimates. To review, for states that are not top-coded, we have created an estimate based on state-specific metropolitan data and regionally specific rural data. For top-coded states, our estimate is composed of regional metropolitan and regional rural data. Both types of state estimates have been corrected so that total expenditures reflect average expenditures in CE multiplied by the number of households in the 2010 estimates (equation 22), and SEDS expenditures completely replace direct energy estimates.

There are still sources of bias in the adjusted CE estimates that we are concerned about. As a check on how well our estimations did at improving CE estimates, we compare our estimates of total statewide expenditures to NIPA estimates of statewide income (with expenditures on direct energy from SEDS, which we assume is unbiased, subtracted from each side). In order to compare estimates of consumption to the NIPA administrative data on income, we assume that rates of taxation and savings are roughly equivalent across states (a state with 5 percent of national income also has 5 percent of national expenditures).

Differences that are in the same direction for all states are not indicative of bias, but the observed differences between the adjusted CE estimates we construct and the NIPA estimates are occasionally large and are in both directions. A comparison is then made between $\ddot{\tilde{X}}^k$ and $\bar{\bar{X}}^k$.

For states that are missing from the CE microdata, it is easy to imagine how regional data may not be a perfect proxy for state expenditures (for instance, using South regional data, which include wealthy states like Virginia and Maryland, overestimates the level of expenditures of a relatively poor top-coded state like Alabama). For states that have attributable observations, this bias is likely due to the fact that metropolitan observations are not randomly selected, since they are selected from predetermined regions within states (for example, a state with PSUs in only wealthy areas of the state will have too many total expenditures). This type of error in the total expenditures means that the proportion of adjusted CE expenditures for each good across states ($\frac{\ddot{x}_u^k}{\ddot{x}_u}$) is biased. However, it does not imply that there is bias in the proportion of expenditures for each good *within* each state ($\frac{\ddot{x}_u^k}{\ddot{X}^k}$), which we assume is relatively accurate (i.e., the proportion of *national* food expenditures that occurs in a given state may be biased, but the proportion of expenditures *within* that state that are on food is assumed to be accurate).

The NIPA personal income estimates by state from BEA are used to help correct for these clear systematic biases correlated to state geography in total expenditures in the CE data. NIPA’s estimates of state income are used as a proxy for expenditures by state. There is likely some bias in these estimates. For example, states with higher income possibly have higher savings rates, and states with older residents may have lower savings rates. However, this potential bias is much less severe than the irregularities that exist in the unadjusted CE data.

A composite estimate (identified by a hat) is calculated to correct for this bias by scaling adjusted CE total state expenditures to total state income from NIPA:

(27) $$\forall \textit{ consumption goods, } u \neq \textit{direct energy and } \forall \textit{ states, } k\text{: } \hat{x}_u^k = \ddot{x}_u^k \cdot \frac{\bar{\bar{Y}}^k - \sum_{\varepsilon} \breve{x}_{\varepsilon}^k}{\ddot{X}^k}$$

and

(28) $$\forall \textit{ consumption goods, } u \neq \textit{direct energy: } \hat{x}_u = \sum_k \hat{x}_u^k$$

where $\bar{\bar{Y}}^k$ is total statewide NIPA income (recall that NIPA values are notated with a double bar), $\sum_{\varepsilon} \breve{x}_{\varepsilon}^k$ is the sum of all SEDS direct energy expenditures by state (we are using an epsilon to denote all direct energy goods), and $\ddot{X}^k$ is the sum of all adjusted CE consumption (which does not include any direct energy expenditures). The composite measures of expenditures ($\hat{x}_u^k$) are a compound of the adjusted CE expenditure totals and the NIPA estimates for state income. After correcting for this fourth source of bias, we cannot find any more systematic bias in the share of expenditure by state of UCC category consumption ($\frac{\hat{x}_u^k}{\hat{x}_u}$).

8. Data Crosswalks

The most challenging task in linking the microsimulation incidence model to the economic model is in matching the 3 income goods and 777 consumption goods that exist in the microsimulation model to whatever set of goods exists in the economic model.

Matching income sources is relatively straightforward, and the process has already been described in the paper. Economic models are likely to closely match the three income sources (labor, capital, and transfer) in the microsimulation model. For economic models that report results in more detailed categories, either the results are summed to create the three categories or additional details would have to be added to the microsimulation model. In the example OLG economic model, only capital goods are provided in greater detail, but as previously noted, we assume the households within an index own the same proportion of each of the capital categories. As a result, the predicted changes to capital goods from the economic model are summed together, and the proportions of income are inserted into equation 10 (see equations 13 and 14).

Consumption goods are much less likely to match up. The microsimulation model reports proportion of consumption by 777 UCC goods, which are unlikely to relate directly to the categories used by an economic model. BLS publishes a concordance table (BLS 2012a) that matches many UCC codes to the 46 personal consumption expenditure (PCE) goods categories from BEA. Because of differences in definitions, the concordance table makes clear that some categories are not comparable, and in other cases, it divides a UCC category between two or more PCE categories. To minimize these inconsistencies and include as many UCC categories as possible, we summed the PCE categories into 23 more broadly defined groups, which we notate as α, and built a crosswalk between the two categories:

$$\text{(29)} \qquad \forall \textit{ PCE consumption good categories, } \alpha\text{: } \frac{\hat{x}_\alpha^k}{\hat{x}_\alpha} = \sum_u (\rho_u^\alpha \cdot \frac{\hat{x}_u^k}{\hat{x}_u})$$

where ρ_u^α is the proportion of expenditures in UCC consumption group u that is represented in PCE consumption group α. All direct energy PCE consumption good categories are made up only of direct energy UCC categories that have been replaced by SEDS estimates for states. As a result, there is no mixing of SEDS and CE expenditures to create any PCE consumption good estimates.

The OLG economic model does not report PCE consumption categories. Instead, it estimates price and quantity changes for 17 commodity goods (i), which are summations of the 57 commodity sectors in the Global Trade Analysis Project (GTAP) and do not represent final consumption goods. In order to relate commodity goods and consumption goods, an additional data crosswalk is necessary, which we borrow from Elliot and Fullerton (2014). The transformation matrix records the proportion of each GTAP commodity good (i) that went into producing 23 consumption goods (α) that are summations of BEA's PCE codes. For example, the commodity (i) ferrous metal corresponds to the consumption (α) goods furnishings, appliances, and autos. This allows us to map changes in consumer surplus by commodity into consumer surplus changes by consumption good:

$$\text{(30)} \qquad \forall \textit{ PCE consumption goods, } \alpha\text{: } \Delta cs_{\alpha} = \sum_{i} (\rho_i^{\alpha} \cdot \Delta cs_i)$$

where ρ_i^{α} is the proportion of expenditures in GTAP commodity group i that is represented in PCE consumption group α. As previously noted, the electricity commodity good and the electricity consumption good are the same ($\rho_e^e = 1$), so this crosswalk does not affect the Haiku-related equations (equations 15–21).

9. Final Consumer and Producer Surplus Equations

After all changes in consumer surplus and proportions of expenditures are both translated to the same group of consumption categories (in our example, PCE categories, or α), we can now calculate all consumer surplus changes.

For consumer surplus by quintile (Table 1, column 4), we use proportions from the CE survey (identified with a tilde):

$$\text{(31)} \qquad \forall \textit{ consumption goods, } \alpha \textit{ and } \forall \textit{ income quintiles, k: } \Delta cs_{\alpha}^{k} = \Delta cs_{\alpha} \cdot \frac{\tilde{x}_{\alpha}^{k}}{\tilde{x}_{\alpha}}$$

For consumer surplus by state from direct energy sources other than electricity (Table 1, column 5), we use proportions from SEDS (identified with an upside-down hat):

$$\text{(32)} \qquad \forall \textit{ consumption goods, } \alpha\textit{=energy goods except electricity and } \forall \textit{ states, k: } \Delta cs_{\alpha}^{k} = \Delta cs_{\alpha} \cdot \frac{\check{x}_{\alpha}^{k}}{\check{x}_{\alpha}}$$

State electricity equations already were given in section 7.2.2, but to review, for consumer surplus from electricity in Alaska and Hawaii (which are not included in the Haiku model), we use proportions from SEDS (identified with an upside-down hat; also note that these equations were already given above and are rewritten here for summarization purposes):

(20) $$\textit{For } \alpha = \textit{electricity and } \forall \textit{ states, } k=AK \textit{ or } HI\textit{:} \quad \Delta cs_{\alpha}^{k} = \Delta cs_{\alpha} \cdot \frac{\breve{x}_{\alpha}^{k}}{\breve{x}_{\alpha}}$$

For consumer surplus from electricity in other states (Table 1, column 6), we use proportions of CS change in the Haiku model (identified with a single bar):

(21) $$\textit{For } \alpha = \textit{electricity and } \forall \textit{ states, } k \neq AK \textit{ or } HI\textit{:} \quad \Delta cs_{\alpha}^{k} = (\Delta cs_{\alpha} - \Delta cs_{\alpha}^{AK} - \Delta cs_{\alpha}^{HI}) \cdot \frac{\Delta \overline{cs}_{\alpha}^{k}}{\Delta \overline{cs}_{\alpha}}$$

For consumer surplus by state from all other consumption (Table 1, column 7), we use the composition proportions we manufactured from CE survey data (identified with a hat):

(33) $$\forall \textit{ consumption goods, } \alpha \neq \textit{energy goods and } \forall \textit{ states, } k\textit{:} \quad \Delta cs_{\alpha}^{k} = \Delta cs_{\alpha} \cdot \frac{\hat{x}_{\alpha}^{k}}{\hat{x}_{\alpha}}$$

Producer surplus change equations appeared above, but to review, they are

(13) $$\forall \textit{ income sources, } j \textit{ and } \forall \textit{ income quintiles, } k\textit{:} \quad \Delta ps_{j}^{k} = \Delta ps_{j} \cdot \frac{y_{j}'^{k}}{y_{j}'}$$

(14) $$\forall \textit{ income sources, } j \textit{ and } \forall \textit{ states, } k\textit{:} \quad \Delta ps_{j}^{k} = \Delta ps_{j} \cdot \frac{\overline{\overline{y}}_{j}^{k}}{\overline{\overline{y}}_{j}}$$

where $\frac{y_{j}'^{k}}{y_{j}'}$ are from CBO estimates and $\frac{\overline{\overline{y}}_{j}^{k}}{\overline{\overline{y}}_{j}}$ are from NIPA estimates.

Equation 12, which calculates welfare changes, remains unchanged, except for the subscript on consumer surplus:

(12) $$\forall \textit{ states or income quintiles, } k\textit{:} \quad \Delta W^{k} = \sum_{i} \Delta cs_{i}^{k} + \sum_{j} \Delta ps_{j}^{k}$$

10. Conclusion

The microsimulation model described in this paper contributes to the existing suite of carbon tax incidence models as the first model that we are aware of that estimates incidence by state instead of region. A more practical advantage of the model, however, is in the ease of application to different or new economic models.

The incidence estimation in the model could be extended beyond quintiles and states with the addition of further data. The consumer expenditure survey includes data on age, so the burden of a carbon tax across generations could be investigated. Measuring the effects of the policy by income group and region is also possible, but the same estimates for states are possible only for states that have large enough consumer expenditure samples. For those missing or lightly sampled states, examining effects by income quintile is impossible. This framework would work well for analyzing a state that both is well represented in the CE survey and is considering instituting its own statewide carbon tax or cap-and-trade regime. While states are the lowest geographic unit in the model currently, most states include a large and diverse set of households. Modeling the incidence by regional urban, suburban, and rural areas may be more informative, but the data challenges of identifying consumption and income patterns may be prohibitive.

Future economic models could expand on the precision of the income and geographic incidence estimations. Giving more detail to sources of labor income would result in increased variation between the incidence estimates, especially concerning quintiles or states with heavy employment in carbon intensive industries. The data exist to incorporate this level of detail by state, but detailed labor income by quintile may be challenging.

Adding greater detail to the microsimulation model could provide politicians with the information necessary to proceed with a plan likely to touch every part of the economy, but taken too far, this complexity could result in less reliable results.

Figures and Tables

Figure 1. Changes in Consumer and Producer Surplus

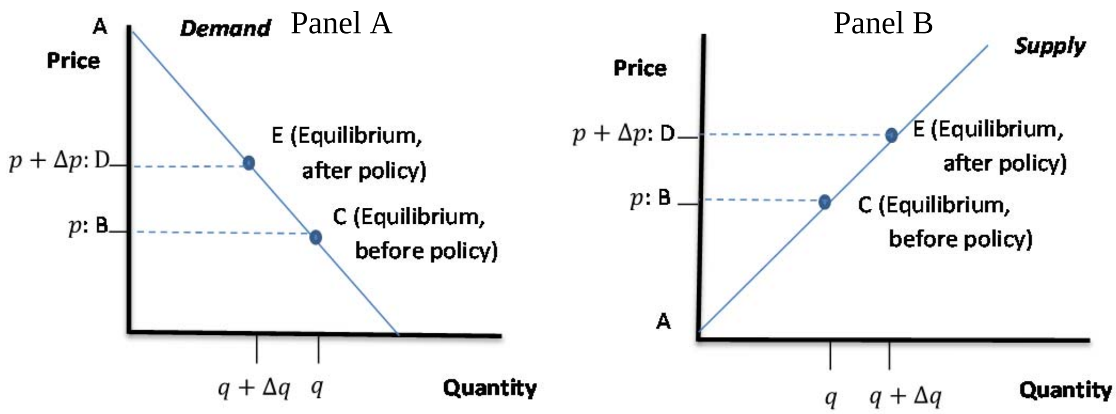

Figure 2. Weighted Population in CE 2004–2011

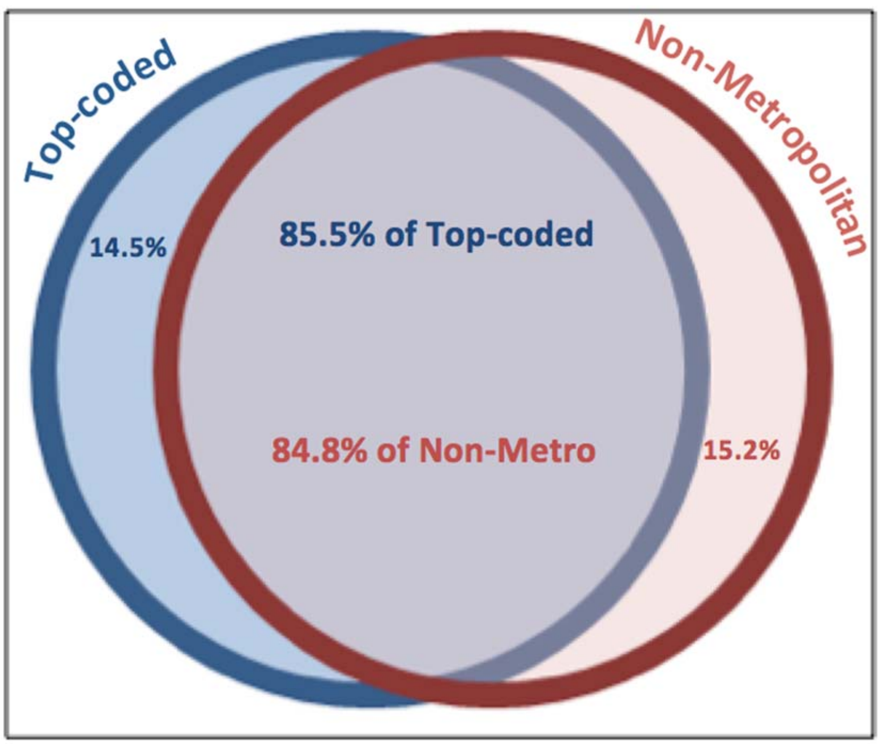

Table 1. Selected Estimation in the Microsimulation Model

	Economic model	Producer surplus		Consumer surplus			
	1. Model outputs	2. By quintile	3. By state	4. By quintile	5. Energy, state	6. Electricity, by state	7. All other consumption, state
Inputs	Initial expenditures and income: $x_i\ ,\ y_j$ Percent change in price and quantity: $\frac{\Delta p_i}{p_i},\frac{\Delta q_i}{q_i},\frac{\Delta p_j}{p_j},\frac{\Delta q_j}{q_j}$ Changes in government transfers, tax dividends: $\Delta y_g\ ,\ \Delta y_d$ Crosswalk between i and PCE codes: ρ_i^{α}	CBO estimates of proportion of income by quintile: $\frac{y_j'^k}{y_j'}$	NIPA estimates of proportion of income by state: $\frac{\overline{\overline{y}}_j^k}{\overline{\overline{y}}_j}$	Consumer expenditure estimates of proportions of expenditures by quintile: $\frac{\tilde{x}_u^k}{\tilde{x}_u}$ After crosswalk: $\frac{\tilde{x}_\alpha^k}{\tilde{x}_\alpha}$	SEDS estimate of proportion of expenditures by state: $\frac{\breve{x}_u^k}{\underset{\smile}{x}_u}$ After crosswalk: $\frac{\breve{x}_\alpha^k}{\underset{\smile}{x}_\alpha}$	Haiku model expenditures and changes in price and quantity: $\overline{x}_e^{k,t},\frac{\Delta\overline{p}_e^{k,t}}{\overline{p}_e^{k,t}},\frac{\Delta\overline{q}_e^{k,t}}{\overline{q}_e^{k,t}}$ SEDS electricity expenditures: $\breve{x}_e^{k,0}$	CE microdata household expenditure by state/region: $\frac{\tilde{x}_u^k}{\tilde{h}^k},\ \frac{\tilde{x}_u^r}{\tilde{h}^r}$ CE summary table regional expenditures, households: $\bar{\tilde{x}}_u^r,\bar{\tilde{h}}^r$ Census households per state and proportion of population in metro CBSA: $h^k\ ,\ \mu^k$ NIPA income by state: $\overline{\overline{Y}}^k$
Equations	(1) $\Delta cs_i=-(1+\frac{\Delta q_i}{q_i}\cdot\frac{1}{2})\cdot\frac{\Delta p_i}{p_i}\cdot x_i$ (2) $\Delta ps_j=(1+\frac{\Delta q_j}{q_j}\cdot\frac{1}{2})\cdot\frac{\Delta p_j}{p_j}\cdot y_j$ (30) $\Delta cs_\alpha=\sum(\rho_i^\alpha\cdot\Delta cs_i)$	(13) $\Delta ps_j^k=\Delta ps_j\cdot\frac{y_j'^k}{y_j'}$	(14) $\Delta ps_j^k=\Delta ps_j\cdot\frac{\overline{\overline{y}}_j^k}{\overline{\overline{y}}_j}$	(31) $\Delta cs_\alpha^k=\Delta cs_\alpha\cdot\frac{\tilde{x}_\alpha^k}{\tilde{x}_\alpha}$	(32) $\Delta cs_\alpha^k=\Delta cs_\alpha\cdot\frac{\breve{x}_\alpha^k}{\underset{\smile}{x}_\alpha}$	(16) $\breve{x}_e^{k,t}=\breve{x}_e^{k,0}\cdot(1+\frac{\overline{x}_e^{k,t}-\overline{x}_e^{k,0}}{\overline{x}_e^{k,0}})$ (17) $\Delta\overline{cs}_e^{k,t}=-(1+\frac{\Delta\overline{q}_e^{k,t}}{\overline{q}_e^{k,t}}\cdot\frac{1}{2})\cdot\frac{\Delta\overline{p}_e^{k,t}}{\overline{p}_e^{k,t}}\cdot\breve{x}_e^{k,t}$ (21) $\Delta cs_e^k=(\Delta cs_e-\Delta cs_e^{AK}-\Delta cs_e^{HI})\cdot\frac{\Delta\overline{cs}_e^k}{\Delta\overline{cs}_e}$	(22) $\ddot{\tilde{x}}_u^k=h^k\cdot\frac{\tilde{x}_u^k}{\tilde{h}^k}$ (23) $\tilde{\xi}_u^r=\bar{\tilde{x}}_u^r-\tilde{x}_u^r$ (24) $\tilde{\eta}^r=\bar{\tilde{h}}^r-\tilde{h}^r$ (25) $\ddot{\tilde{x}}_u^k=h^k(\mu^k\cdot\frac{\tilde{x}_u^k}{\tilde{h}^k}+(1-\mu^k)\frac{\tilde{\xi}_u^r}{\tilde{\eta}^r})$ (27) $\hat{x}_u^k=\ddot{\tilde{x}}_u^k\cdot\frac{\overline{\overline{Y}}^k-\sum\breve{x}_\varepsilon^k}{\ddot{\tilde{X}}^k}$ After crosswalk: (33) $\Delta cs_\alpha^k=\Delta cs_\alpha\cdot\frac{\hat{x}_\alpha^k}{\hat{x}_\alpha}$

Outputs	Changes in consumer surplus and producer surplus by good: $\Delta cs_i, \Delta ps_j$ After crosswalk: Δcs_α	Changes in producer surplus by quintile: Δps_j^k	Changes in producer surplus by state: Δps_j^k	Changes in consumer surplus by quintile: Δcs_α^k	Changes in nonelectricity direct energy consumer surplus by state: Δcs_α^k	Changes in electricity consumer surplus by state: Δcs_e^k	CE expenditures, households in nonmetro CBSAs: $\tilde{\xi}_u^r, \tilde{\eta}^r$ Changes in all other consumer surplus by state: Δcs_α^k

11. References

BEA (Bureau of Economic Analysis). 2013. Local Area Personal Methodology. Washington, DC: BEA. Retrieved 12.1.14 from http://www.bea.gov/regional/pdf/lapi2012.pdf.

Blonz, J., D. Burtraw, and M. Walls. 2012. Social Safety Nets and US Climate Policy Costs. *Climate Policy* 12: 1–17.

BLS (Bureau of Labor Statistics). 2012a. Concordance Mapping of CE UCC Codes to PCE Series from Table 2.4.5U: Expenditures by Type of Product. Washington, DC: BLS. Retrieved 12.1.14 from http://www.bls.gov/cex/cecomparison.htm.

———. 2012b. Consumer Expenditure Interview Survey Public Use Microdata; 2011 User's Documentation. Washington, DC: BLS. Retrieved 12.1.14 from http://www.bls.gov/cex/2011/csxintvw.pdf.

Boyce, J. K., and M. E. Riddle. 2007. Cap and Dividend: How to Curb Global Warming while Protecting the Incomes of American Families. Working paper no. 150. Amherst, MA: Political Economy Research Institute.

Burtraw, D., R. Sweeney, and M. Walls. 2009. The Incidence of US Climate Policy: Alternative Uses of Revenues from a Cap-and-Trade Auction. *National Tax Journal* 62(3): 497–518.

Carbone, J. C., R. D. Morgenstern, and R. C. Williams III. 2012. Carbon Taxes and Deficit Reduction. Forthcoming working paper.

CBO (Congressional Budget Office). 2013. The Distribution of Household Income and Federal Taxes, 2010. Washington, DC: CBO. Retrieved 12.1.14 from http://www.cbo.gov/publication/44604.

Dinan, T. 2012. Offsetting a Carbon Tax's Costs on Low-Income Households. Working paper 2012-16. Washington, DC: Congressional Budget Office.

Dinan, T. M., and D. L. Rogers. 2002. Distributional Effects of Carbon Allowance Trading: How Government Decisions Determine Winners and Losers. *National Tax Journal* 55(2): 199–221.

EIA (Energy Information Agency). 2012. State Energy Data System. Price and Expenditure Estimates Technical Notes. Washington, DC: EIA. Retrieved 12.1.14 from http://www.eia.gov/state/seds/sep_prices/notes/pr_technotes.pdf.

Elliott, J., and D. Fullerton. 2014. Can a Unilateral Carbon Tax Reduce Emissions Elsewhere? *Resource and Energy Economics* 36(1): 6–21.

Fullerton, D., and D. Karney. 2009. Does Environmental Protection Hurt Low-Income Families? Policy Forum. Urbana: Institute of Government and Public Affairs, University of Illinois. 21: 2.

Grainger, C. A., and C. D. Kolstad. 2010. Distribution and Climate Change Policies. In *Climate Change Policies: Global Challenges and Future Prospects*, edited by E. Cerda and X. Labandiera, 129–46. Cheltenham, UK: Edward Elgar.

Hassett, K. A., A. Mathur, and G. E. Metcalf. 2009. The Incidence of a U.S. Carbon Tax: A Lifetime and Regional Analysis. The Energy Journal 30(2): 157–79.

Hsu, S.-L. 2011. Arguments against a Carbon Tax. *The Case for a Carbon Tax: Getting Past Our Hang-Ups to Effective Climate Policy*. Washington, DC: Island Press.

Jorgenson, D. W., R. J. Goettle, M. S. Ho, and P. J. Wilcoxen. 2012. Energy, the Environment, and U.S. Economic Growth. In *Handbook of Computable General Equilibrium Modeling*, edited by P. B. Dixon and D. W. Jorgenson, 477–552. Amsterdam: Elsevier.

Lanz, B., and S. Rausch. 2014. Emissions Trading in the Presence of Price-Regulated Polluting Firms: How Costly Are Free Allowances? *Energy Journal*. Paper 14-087 (forthcoming).

Mathur, A., and A. C. Morris. 2012. Distributional Effects of a Carbon Tax in Broader U.S. Fiscal Reform. Climate and Energy Economics Discussion Paper. Washington, DC: Brookings Institution.

Metcalf, G. E. 2009. Designing a Carbon Tax to Reduce U.S. Greenhouse Gas Emissions. *Review of Environmental Economics and Policy* 3(1): 63–83.

Morris, D. F., and C. Munnings. 2013. Progressing to a Fair Carbon Tax: Policy Design Options and Impacts to Households. RFF issue brief 13-03. Washington, DC: Resources for the Future.

Parry, I., and R. C. Williams III. 2010. What Are the Costs of Meeting Distributional Objectives for Climate Policy? *B.E. Journal of Economic Analysis & Policy* 10(2): ISSN (Online) 1935–1682

Parry, I., R. C. Williams III, and L. Goulder. 1999. When Can Carbon Abatement Policies Increase Welfare? The Fundamental Role of Distorted Factor Markets. *Journal of Environmental Economics and Management* 37(1): 52–84.

Passero, W., T. I. Garner, and C. McCully. 2013. Understanding the Relationship: CE Survey and PCE. BLS working paper 462. Washington, DC: US Bureau of Labor Statistics.

Paul, A., B. Beasley, and K. Palmer. 2013. Taxing Electricity Sector Carbon Emissions at Social Cost. RFF discussion paper 13-23-REV. Washington, DC: Resources for the Future.

Rausch, S., G. E. Metcalf, and J. M. Reilly. 2011. Distributional Impacts of Carbon Pricing: A General Equilibrium Approach with Micro-Data for Households. *Energy Economics* 33: S20–S33.

Rausch, S., G. E. Metcalf, J. M. Reilly, and S. Paltsev. 2009. Distributional Impacts of a U.S. Greenhouse Gas Policy: A General Equilibrium Analysis of Carbon Pricing. MIT Joint Program on the Science and Policy of Global Change. Cambridge: Massachusetts Institute of Technology.

Ruser, J., A. Pilot, C. Nelson. 2004. Alternative Measures of Household Income: BEA Personal Income, CPS Money Income, and Beyond. BEA Report to the Federal Economics Statistics Advisory Committee. Washington, DC. Retrieved 12.1.14 from http://www.bls.gov/bls/fesacp1061104.pdf.

Shields, J., and N. To. 2005. Learning to Say No: Conditioned Underreporting in an Expenditure Survey. Paper presented at the annual meeting of the American Association for Public Opinion Research. Retrieved 12.1.14 from http://www.amstat.org/sections/srms/proceedings/y2005/Files/JSM2005-000432.pdf.

Williams, R. C., H. Gordon, D. Burtraw, J. C. Carbone, and R. D. Morgenstern. 2014. The Initial Incidence of a Carbon Tax across US States. *National Tax Journal* 67(4): 807–30.