WHO WINS, WHO LOSES? TOOLS FOR DISTRIBUTIONAL POLICY EVALUATION (PRELIMINARY AND INCOMPLETE)

Maximilian Kasy*
May 1, 2014

Most policy changes generate winners and losers. Political economy and optimal policy suggest questions such as: Who wins, who loses? How much? Given a choice of welfare weights, what is the impact of the policy change on social welfare? This paper proposes a framework to empirically answer such questions. The framework is grounded in welfare economics and allows for arbitrary heterogeneity across individuals as well as for endogenous prices and wages (general equilibrium effects). The proposed methods are based on imputation of money-metric welfare impacts for every individual in the data.

The key contribution of this paper are new identification results for marginal causal effects conditional on a vector of endogenous outcomes. These identification results are required for imputation of individual welfare effects. Based on these identification results, we propose methods for estimation and inference on disaggregated welfare effects, sets of winners and losers, and social welfare effects. We furthermore provide results relating aggregation with social welfare weights to the distributional decomposition literature. We apply our methods to analyze the distributional impact of the introduction of the Earned Income Tax Credit (EITC), using variation in state supplements to the federal EITC and the CPS-IPUMS data.

Keywords: Social welfare, distributional decomposition, nonparametric identification.

1. INTRODUCTION

Economists usually evaluate the welfare impact of policy changes based on their impact on individuals. To evaluate a policy change based on its impact on individuals, we need to (i) define how we measure individual gains and losses, (ii) estimate them, and (iii) take a stance on how to aggregate them. To understand the political economy of a policy change (who would oppose it and who would support it, based on economic self-interest), we need to characterize the sets of winners and losers of this policy change.

The answers to these questions are important to the extent that few changes of economic policy result in Pareto improvements; most policies, in particular controversial ones, do generate winners and losers. Some examples help to illustrate. Trade liberalization opposes net producers and net consumers of goods with rising / declining prices subsequent to liberalization. Progressive income tax reform opposes high and low income earners. Skill biased technical change

Assistant professor, Department of Economics, Harvard University, and junior associate faculty, IHS Vienna. Address: Littauer Center 200, 1805 Cambridge Street, Cambridge, MA 02138. email: maximiliankasy@fas.harvard.edu.

*I'd like to thank David Autor, Gary Chamberlain, Raj Chetty, Ellora Derenoncourt, Arindrajit Dube, Pirmin Fessler, Sergio Firpo, Nathaniel Hendren, Stefan Hoderlein, Larry Katz, Susanne Kimm, Nathan Lane, Suresh Naidu, and Emmanuel Saez for valuable discussions and comments. László Sándor provided valuable research assistance.

opposes suppliers of substitutes and complements to new technologies. A decrease of barriers to migration opposes would-be migrants as well as suppliers of complements to migrant labor to suppliers of substitutes to migrant labor.

The goal of this paper is to provide a general set of tools for empirical researchers who wish to analyze the distributional impact of policy changes or historical changes in settings such as these. The framework we propose is characterized by the following features: (i) We consider individual welfare as measured by utility. (ii) We allow for endogenous prices and wages. (iii) We allow for unrestricted heterogeneity across individuals in terms of preferences and in terms of policy impacts on wages, labor supply, etc. Within this framework, we devise procedures to answer various questions regarding the distributional impact of policy changes: What is the expected welfare impact on individuals conditional on their initial income and exogenous covariates? In particular, which income groups win or lose as a consequence of the policy change, and by how much? Given a choice of welfare weights, what is the impact of the policy change on social welfare? Should we support or oppose the policy change?

The procedures proposed here impute a money-metric expected welfare impact of the policy change under consideration to each individual. Based on this imputed impact, and given a choice of welfare weights, we can estimate aggregate welfare impacts. We can also estimate sets of winners and losers and their characteristics. The central econometric difficulty is identification of expected welfare impacts of the form $\dot{w} \cdot l$ (change in wage times baseline labor supply) *conditional* on baseline income $w \cdot l$. Welfare impacts differ from impacts on income by the behavioral effect $w \cdot \dot{l}$ (wage times change in labor supply). More generally, we need to identify the expected causal impact $\dot{x}$ of a policy change on x, conditional on initial x and policy level α, $E[\dot{x}|x, \alpha]$. We provide conditions under which such conditional causal effects, and in particular expected welfare impacts, are identified by the slopes of nonparametric quantile regressions with control functions, generalizing insights of Hoderlein and Mammen (2007) and Imbens and Newey (2009). Based on these identification results, we propose to estimate individual welfare impacts using local linear quantile regressions. These estimated expected welfare effects are then used to derive estimators for a variety of objects, in particular (i) average expected welfare impacts as a function of initial income, and (ii) descriptive statistics, such as covariate means and population shares, for the sets of winners and losers.

We furthermore provide results relating social welfare evaluations (as in optimal tax theory) to distributional decompositions (as in labor economics). We show that welfare weights in social welfare analysis are formally analogous to the derivatives of influence functions as introduced to the decomposition literature by Firpo et al. (2009). We further show that, given welfare weights, policy impacts on social welfare differ from impacts on distributional statistics by a "behavioral correction" term.

There are several literatures in economics aiming to empirically evaluate the distributional impact of policies or historical changes, including the empirical optimal tax literature in public finance (eg. Saez, 2001; Chetty, 2009), the labor economics literature on determinants of the wage distribution (eg. Autor et al., 2008; Card, 2009), and the distributional decomposition literature (eg. DiNardo et al., 1996; Firpo et al., 2009). Our proposed methods build on these literatures and generalize them in the following ways: (i) In contrast to most of the empirical (income) taxation literature, we allow for endogenous prices and in particular wages. (ii) In contrast to the wage distribution and decomposition literatures, we are interested in (unobserved) realized utility rather than observed wages or incomes. (iii) In contrast to more structural approaches estimating demand systems for the labor market, we allow for arbitrary heterogeneity across individuals in terms of policy impacts on their wages and on their labor supply.

The tools developed in this paper build on the insights of various literatures in- and outside economics. Several literatures in (empirical) economics consider distributional impacts of policies or historical changes. This includes the optimal taxation literature in public finance, where utilitarian social welfare functions[1] were introduced by Samuelson (1947) and the canonical model of redistributive income taxation was proposed by Mirrlees (1971). More recent references that this paper draws on include Saez (2001), Chetty (2009), Hendren (2013), and Saez and Stantcheva (2013). A large literature in labor economics analyzes the role of various determinants of the wage distribution (technology, migration, minimum wages, ...) in causing historical changes in wage inequality; partial reviews can be found in Autor et al. (2008) and Card (2009). An important and popular empirical tool for analyzing distributional impacts on observed outcomes are distributional decompositions. These originate in the work of Oaxaca (1973); standard references are DiNardo et al. (1996) and Firpo et al. (2009). Recent contributions to the econometrics of such decompositions are Rothe (2010) and Chernozhukov et al. (2013). The objects of interest we consider are inspired by questions central to the sociological analysis of social classes (cf. Wright, 2005). Disaggregated distributional analysis, in particular, allows to study both impacts of policies on inequality and antagonisms of interest. These are two of the main consequences of the class structure underlying the economy emphasized by class analysis. Dis-aggregated impacts allow us to study questions of political economy, following the research agenda proposed by Acemoğlu and Robinson (2013). They also allow readers to reach aggregate conclusions based on their own choice of welfare weights. They finally allow to recognize when policies generate both winners and losers. Deaton (1989) conducted a disaggregated analysis similar to the one proposed here for the case of a homogenous good (rice).

The main econometric challenge we face is the identification of policy effects conditional on multidimensional outcomes. The one-dimensional case has been elegantly characterized by Hoderlein and Mammen (2007); we derive identified sets in the multidimensional case and discuss conditions sufficient for point iden-

[1]The term "utilitarian" is used in this paper to describe methods evaluating welfare based on individual realized utilities. It is *not* used here to imply a comparison across individuals based on some notion of cardinal utility.

tification, drawing on tools from continuum mechanics (fluid dynamics) and the theory of differential forms (cf. Rudin, 1991, chapter 10). The estimators we propose build on the large literatures on quantile regression and nonparametric regression; important references include Koenker (2005), Newey (1994a), Matzkin (2003), Altonji and Matzkin (2005), and Chernozhukov et al. (2013).

The rest of this paper is structured as follows. Section 2 presents our assumptions and objects of interest and characterizes the effect of policy changes on individual welfare. Section 3 presents our results on identification; section 3.1 provides results on the identification of marginal causal effects conditional on outcomes, and section 3.2 discusses the use of instruments and controls as well as of panel data for identification of welfare effects in a nonparametric setting. Section 4 discusses aggregation and the relation between distributional decompositions and social welfare effects. Section 5 proposes estimators and inference procedures based on these identification results. Section 6 applies our results to analyze the distributional impact of the expansion of the Earned Income Tax Credit (EITC) using CPS-IPUMS data and identifying variation from state-level top ups of the EITC which vary over time. Section 7 concludes. Appendix A contains all proofs.

2. SETUP

This section presents the setup studied in this paper. We first discuss notation, then state the individual's consumption and labor supply problem, and finally introduce several empirical objects of interest which we will analyze. The setup considered is a static labor supply model with nonlinear income taxation and arbitrary heterogeneity of preferences and wages across individuals. Policies in this setup might affect prices, wages, and taxes.

2.1. *Notation*

Throughout, we consider a set of counterfactual policies indexed by $\alpha \in \mathbb{R}$, and a population of individuals i. *Potential outcomes* under policy α are denoted by superscripts, so that w^α is for instance the potential wage of an individual under policy α.[2] Letters without superscripts denote *random variables*, so that w is the wage of an individual as determined by the *realized* policy α. When we consider a sample of observations $i = 1, \dots, N$ in section 5 (a random subset of all individuals i), corresponding *draws* of random variables are denoted by a subscript i.

We use several short-hands for derivatives. *Partial derivatives* are denoted ∂ with a subscript, so that ∂_w is the derivative with respect to w. *Derivatives of*

[2]Potential outcomes in this paper are "reduced form" objects in the sense that they incorporate the impact of any general equilibrium effects of policy changes.

potential outcomes with respect to α will be denoted by a superscript dot, so that

$$\dot{w} := \partial_\alpha w^\alpha$$

denotes the marginal effect of a policy change on the wage of a given individual. Our identification results in section 3 will use the notation $\nabla H(x) := (\partial_{x^1} H, \ldots, \partial_{x^k} H)$ for the *gradient* of a real valued function H of a k dimensional vector x, and $\nabla \cdot h(x) := \sum_{j=1}^k \partial_{x^j} h^j$ for the *divergence* with respect to x of a vector field h.

Probability density functions, conditional or unconditional, are denoted by the letter f, *probabilities* and probability distributions by the letter P, *cumulative distribution functions* by the letter F, and *quantiles* by the letter Q. If it is clear from context which (conditional) distribution an expression refers to, subscripts will be omitted, so that for instance $f(w|l)$ denotes the density of w given l.

2.2. *Individual problem*

We discuss distributional policy evaluation in the context of labor markets, the wage distribution, and earnings taxes.[3] All variables depend on the policy α, as well as of unobserved individual heterogeneity ϵ, unless otherwise stated. We denote an individual's labor supply by l, her pre-tax market wage by w, and her pre-tax earnings by $z = l \cdot w$. She pays earnings tax $t = t(z)$ and receives unearned income y_0, so that her net income is $y = z - t(z) + y_0$. Using this notation, the individual's problem is given as follows.

Assumption 1 (Individual utility maximization)
- *There is a population of individuals indexed by* $i \in \mathscr{I}$, *and a schedule of counterfactual policies indexed by* $\alpha \in \mathbb{R}$.
- *Every individual* i *chooses* c *and* l *to solve*

$$\max_{c,l} u(c,l) \quad s.t. \quad c \cdot p \leq l \cdot w - t(l \cdot w) + y_0, \tag{1}$$

  taking w, p *and* $t(.)$ *as given. The value of* u *at the maximizing* (c,l) *is denoted* v.
- *The utility function* $u(.)$, *wage* w, *the consumption bundle* c, *and labor supply* l *may vary arbitrarily across individuals.*
- *Prices* p, *wage* w, *unearned income* y_0, *and taxes* $t(.)$ *may depend on* α, *and as a consequence so do* c, l, *and* v.
- *For all individuals,* u *is differentiable, increasing in the components of* c *and decreasing in* l, *quasiconcave, and does not depend on* α.

[3]Our arguments apply equally to other markets with heterogeneous goods, however, for instance to the housing market.

Remarks:

- Assumption 1 states a simple static model of labor supply subject to a budget constraint. We focus on this case for simplicity and specificity, and since it is similar to the settings considered in the wage distribution literature and in the income taxation literature. Labor markets are furthermore of central importance in determining the relative welfare of individuals. They are also of particular conceptual interest: Heterogeneity in wages and in wage responses to policy changes poses econometric challenges which are absent from the analysis of markets with more homogeneous goods.
- Our arguments do generalize to models with dynamics and additional constraints, and to other markets with heterogenous goods, by arguments similar to those discussed in Chetty (2009). Of particular interest is the housing market, since it is also characterized by very heterogeneous supply, and since most individuals are consumers of housing and many are owners of houses.
- A crucial limitation of the setting we discuss is the assumption that prices, wages, and taxes are the only constraints of the individual *which change* as a function of the policy change. If other constraints are binding *and* change as a function of the policy change, this would need to be incorporated in estimates of welfare effects. An example of such additional constraints would be involuntary unemployment.
- To each individual in the setup of assumption 1 there corresponds a schedule of counterfactual wages w^α, counterfactual consumption c^α etc., as well as a realized policy α and corresponding realized wage w, realized consumption c etc.
- In order to relate our model to the canonical model of consumer choice subject to a linear budget constraint, consider the following linearized version of the individual's problem. Define marginal net wage as

  $$n := \partial_l y = w \cdot (1 - \partial_z t),$$

  and virtual lump sump taxes as

  $$t_0 := t - \partial_z t \cdot z.$$

  Denote leisure $L = \overline{L} - l$ and total endowment with time $\overline{L}$. We can rewrite the individual' s utility maximization problem as

  $$\max_{c,L} u(c, \overline{L} - L) \quad s.t. \quad c \cdot p + L \cdot w \leq \overline{L} \cdot w - t(l \cdot w) + y_0.$$

  By quasiconcavity of u, this problem in turn has the same solution as

  $$\max_{c,L} u(c, \overline{L} - L) \quad s.t. \quad c \cdot p + L \cdot n \leq \overline{L} \cdot n - t_0 + y_0, \tag{2}$$

  where n and t_0 are treated as constants by the individual. This problem has the form of a standard linear consumer problem.

2.3. *Objects of interest*

The basic object of interest in this paper is the welfare impact of a policy change on individuals. All other objects we consider are functions of the individual-level welfare impact. This welfare impact is given by the impact $\dot{v}$ on realized utility v. We shall re-normalize this impact to get the impact on money-metric utility: Rescaling $\dot{v}$ by the marginal utility impact of a lump-sum transfer of money, $\partial_{y_0} v$, yields

$$\dot{e} := \dot{v}/\partial_{y_0} v. \tag{3}$$

$\dot{e}$ is the impact of the policy change on the expenditure function e (at baseline prices), as defined in (Mas-Colell et al., 1995, section 3.E).

We are also interested in aggregate welfare functionals which depend on individuals' realized utility v. For a finite population of N individuals, aggregate welfare is simply a function of the vector $(v_1, \ldots, v_N)$. More generally, welfare is a functional of $(v_i : i \in \mathscr{I})$.With these preliminaries, and denoting by W a vector of covariates which are not affected by α, we can define our main objects of interest.

Definition 1 (Objects of interest – utility)

1. *Expected conditional policy effect on welfare:*

$$\gamma(y, W) := E[\dot{e}|y, W, \alpha] \tag{4}$$

2. *Sets of winners and losers:*

$$\begin{aligned} \mathscr{W} &:= \{(y, W) : \gamma(y, W) \geq 0\} \\ \mathscr{L} &:= \{(y, W) : \gamma(y, W) \leq 0\} \end{aligned} \tag{5}$$

3. *Policy effect on social welfare:*

$$S\dot{W}F \tag{6}$$

where social welfare SWF maps $(v_i : i \in \mathscr{I})$ into $\mathbb{R}$.

Remarks:

- The expected conditional policy effect γ is the fundamental object of interest; it maps into all other objects we consider. Our proposed methods are based on imputing an estimate of $\gamma(y_i, W_i)$ to every observation $i = 1, \ldots, N$ in the baseline sample. We propose to plot γ or objects such as $E[\gamma|y]$, the expected welfare impact given initial income. This allows to immediately visually assess the welfare impact of a policy change across the income distribution.

- The sets of winners and losers $\mathscr{W}$ and $\mathscr{L}$ are central objects of interest for political economy considerations. To the extent that individuals' political actions reflect their economic self-interest, these sets correspond to potential coalitions supporting and opposing the policy change under consideration.
- The policy effect on social welfare SWF is the relevant object from an optimal policy perspective (cf. Saez, 2001; Chetty, 2009). If this effect is positive, the policy change should be implemented. To calculate this effect, we need to take a stance on the relative weight assigned to the welfare of different individuals. In fact, we show below that (under certain differentiability conditions) $S\dot{W}F$ can be written as $S\dot{W}F = E[\omega \cdot \dot{e}]$ for welfare weights ω and which measure the relative value assigned to a marginal dollar for each individual.
- The expected conditional welfare effect pins down aggregate welfare effects if either (i) the welfare weights implied by SWF (and discussed in section 4 below) are functions of y, W, or (ii) the policies considered have welfare effects which are functions of y, W. Both conditions are satisfied in standard models of optimal taxation such as the Mirrlees (1971) model.
- It is worth noting that any aggregate welfare evaluation corresponds to an implicit or explicit choice of welfare weights; we will elaborate on this point in section 4. Aggregation by summing up money metric utility across individuals, in particular, corresponds to a particular choice of welfare weights. The implied weights in that case are proportional to the inverse of marginal utility of income and are thus presumably larger for richer individuals.
- In this paper, we are mainly interested in welfare evaluations based on individual realized utility. It is however quite instructive, and provides useful connections to the distributional decomposition literature (cf. DiNardo et al., 1996; Firpo et al., 2009), to consider analogous objects for realized incomes rather than realized utility. Section 4 below considers these.

2.4. *Welfare effects of marginal policy changes on individuals*

We consider the effects of a marginal change in α on individual welfare. In the context of the model specified by assumption 1, such a change might affect individuals through (i) taxes t, (ii) wages w, (iii) unearned income y_0, and (iv) prices p. Indirectly, such a change might affect individuals' labor supply l and consumption vector c. We first derive the welfare effect on individuals, and compare it to the effect on net income y. Section 3 discusses identification of expected individual effects. Section 4 then considers aggregate effects, on social welfare SWF as well as on statistics of the income distribution θ.

The following lemma characterizes the effect of a marginal policy change on net income and on money-metric utility. We then discuss the difference between these two effects.

Lemma 1 (The effect of marginal policy changes on individuals)
Consider a marginal change in α. The effect of such a marginal change on net income equals

$$\dot{y} = (\dot{l} \cdot w + l \cdot \dot{w}) \cdot (1 - \partial_z t) - \dot{t} + \dot{y_0}. \tag{7}$$

The effect of such a marginal change on welfare (money metric utility) equals

$$\dot{e} = l \cdot \dot{w} \cdot (1 - \partial_z t) - \dot{t} + \dot{y_0} - c \cdot \dot{p}. \tag{8}$$

We can decompose the effect $\dot{y}$ of a marginal policy change on net income into four components,

1. the behavioral effect $b := \dot{l} \cdot w \cdot (1 - \partial_z t) = \dot{l} \cdot n$,
2. the wage effect $l \cdot \dot{w} \cdot (1 - \partial_z t)$,
3. the effect on unearned income $\dot{y_0}$,
4. and the mechanical effect of changing taxes $-\dot{t}$.

The effect $\dot{e}$ on money metric utility is given by the sum of

1. the wage effect,
2. the effect on unearned income,
3. the mechanical effect of changing taxes
4. and the price effect $-c \cdot \dot{p}$.

The difference between $\dot{y}$ and $\dot{e}$ is given by the sum of the behavioral effect and the price effect,

$$\dot{y} - \dot{e} = \dot{l} \cdot n + c \cdot \dot{p}. \tag{9}$$

The empirical application in section 6 assumes $\dot{p} = 0$ and $\dot{y_0} = 0$, that is, we ignore the effects of changing prices and of changes in unearned income. In this simplified case, we get

$$\dot{e} = l \cdot (1 - \partial_z t) \cdot \dot{w} - \dot{t}. \tag{10}$$

Using the linearized form of the consumer problem we can alternatively write this as

$$\begin{aligned} \dot{e} &= \dot{y} - \dot{l} \cdot n \\ &= l \cdot \dot{n} - \dot{t}_0. \end{aligned}$$

We can, in particular, obtain the welfare effect by subtracting the "behavioral correction" $b = \dot{l} \cdot n$ from the effect on realized net income.

Remark:

- Lemma 1 illustrates the main implication of a utilitarian framework for welfare economics: whatever choices people make are best for themselves – by assumption. As a consequence behavioral responses to policy changes have to be ignored when calculating the marginal impact of policy changes on individuals' welfare. This holds true regardless of the specific model under consideration. Note that behavioral responses can *not* be ignored when calculating the effect on other individuals; behavioral responses might affect other individuals through channels such as their effect on prices, the effect on the tax base, and externalities.
- The welfare effect $\dot{e}$ corresponds to the effect of changing prices and wages holding behavior constant. Defining an empirical counterpart of $\dot{e}$ requires us to specify the behavioral margins and associated prices which might be affected by the policy change, in contrast to the "sufficient statistic" literature reviewed in Chetty (2009). Sufficient statistic arguments rely on either fixed prices or known price-responses.[4] In particular, we *do* need to observe the relevant labor supply margins if wages are allowed to be endogenous.
- If $\dot{w} = \dot{y} = \dot{y}_0 = 0$, then equation (8) reduces to Roy's identity, $\dot{e} = -c \cdot \dot{p}$. In a precursor to the analysis proposed here and based on this identity, Deaton (1989) considers the distributional welfare impact of changing rice prices in Thailand.

3. IDENTIFICATION

In this section, we discuss identification of $\gamma(y, W) = E[\dot{e}|y, W, \alpha]$. We first assume random ("experimental") variation of α and consider the case of no covariates W in section 3.1. The crucial challenge which section 3.1 addresses is identification of marginal causal effects conditional on a *vector* of endogenous outcomes, that is, identification of $E[\dot{x}|x, \alpha]$. In our context, this is necessary because $\gamma(y, W)$ involves terms of the form $E[l \cdot \dot{w}|l \cdot w, \alpha]$, or more general versions thereof. We provide conditions under which

$$E[\dot{x}^j|x, \alpha] = \partial_\alpha Q(v^j|v^1, \ldots, v^{j-1}, \alpha),$$

where $v^j = F(x^j|x^1, \ldots, x^{j-1}, \alpha)$. In section 3.2, we then generalize to quasi-experimental settings, assuming the availability of suitable controls and / or exogenous instruments.

The results in this section generalize those of the literature on nonparametric identification for the case $\dim(x) = 1$, in particular Hoderlein and Mammen (2007).

[4]I thank Nathaniel Hendren for discussions on this point.

3.1. *Effects conditional on outcomes*

Consider a simplified version of the setting of assumption 1, where we assume $\dot{p} = \dot{t} = \dot{y}_0 = 0$, so that welfare effects are driven solely by changing wages. In this case, equation (8) of lemma 1 implies that $\dot{e} = l \cdot (1 - \partial_z t) \cdot \dot{w}$. Let us additionally assume for notational simplicity that $\partial_z t = 0$, so that $\dot{e} = l \cdot \dot{w}$ Then the conditional expected welfare effect $\gamma(y, W)$ is equal to the wage effect $E[l \cdot \dot{w} | l \cdot w, \alpha]$. The latter is identified if $E[(\dot{l}, \dot{w}) | l, w, \alpha]$ is identified. Let $x = (l, w)$.[5] Our problem is to identify $E[\dot{x}|x, \alpha]$.

Suppose the distribution of x^α is known for a continuum of values of α. This is the case in an experimental setting, where α is independent of unobserved heterogeneity ϵ. The following series of results explores identification of $E[\dot{x}|x, \alpha]$ in this case.

Assumption 2 (Abstract setup)

- $x = x(\alpha, \epsilon)$
- $x \in \mathbb{R}^k$, $\alpha \in \mathbb{R}$, *ϵ has support of unrestricted dimension.*
- $\alpha \perp \epsilon$
- *The observed data identify $f(x|\alpha)$ for $\alpha \in (-\delta, \delta)$.*
- *x is continuously distributed given α.*
- *$x(\alpha, \epsilon)$ is differentiable in α.*
- *$E[\dot{x}|x, \alpha] \cdot f(x|\alpha)$ is continuously differentiable in x.*
- *The support of x given α is contained in a compact and convex set $\mathbf{X}$ which is independent of α.*

Recall that we are using the following notation: $f(x|\alpha)$ is the conditional density of x given α. The letter Q denotes (conditional) quantiles. Derivatives with respect to the policy parameter α are written $\dot{f} = \partial_\alpha f(x|\alpha)$, $\dot{x} = \partial_\alpha x(\alpha, \epsilon)$ etc. We further define

$$h(x, \alpha) := E[\dot{x}|x, \alpha] \cdot f(x|\alpha), \tag{11}$$

and denote the divergence of h by

$$\nabla \cdot h := \sum_{j=1}^{k} \partial_{x^j} h^j.$$

If h is identified, then so is our object of interest $E[\dot{x}|x, \alpha]$ by $E[\dot{x}|x, \alpha] = h(x, \alpha)/f(x|\alpha)$ for values of (x, α) in their joint support.

Remark:

- The setting of assumption 2 has various analogies in physics, most notably in fluid dynamics. We can think of α as time, ϵ indexing individual particles,

[5] In more general settings, x has to include other endogenous, heterogeneous variables such as y_0, and the conditioning arguments might include exogenous covariates W.

and x the position of a particle in space. The function $x(\alpha, \epsilon)$ describes the trajectory of a particle over time. Then $f(x|\alpha)$ is the density of the gas or liquid at location x and time α. As shown in theorem 1 below, the change of this density over time is given by the divergence of the net flow h. The case $\nabla \cdot h \equiv 0$ corresponds to the flow of an incompressible fluid, the density of which is constant over time, which is approximately true for water. The equation $\nabla \cdot h \equiv 0$ characterizes the kernel of the identified set for h in theorem 2 below.
- The source of the identification problem we face is accurately illustrated by the following analogy: By stirring your coffee (or other beverage of choice), you can create a variety of different flows $g(x, \alpha)$ which are all consistent with the same constant density $f(x|\alpha)$ of the beverage being stirred.

We will now develop a series of results characterizing this identification problem. Theorem 1 shows that the divergence of h is identified from the data via the identity $\dot{f} = -\nabla \cdot h$. Theorem 2 shows that the reverse is also true: any flow density h that satisfies this equation is in the identified set, absent any further restrictions. Theorem 3 characterizes the identified set, formalizing the "coffee stirring" analogy. Theorem 4 imposes the additional exclusion restrictions $\partial_{x^j} E[\dot{x}^i|x, \alpha] = 0$ for $j > i$, and shows that under these restrictions h and g are just-identified by nonparametric quantile regressions with control functions. Theorem 5, finally, restricts heterogeneity further and obtains just-identification of the structural functions $x(\alpha, \epsilon)$.

The following theorem shows that the data identify the divergence of h under assumption 2.

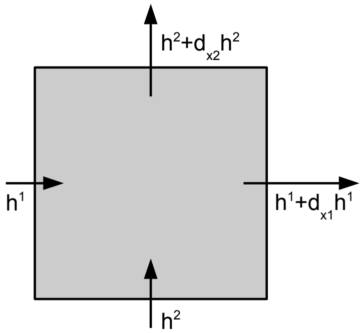
Theorem 1 *Suppose assumption 2 holds. Then*

$$\dot{f} = -\nabla \cdot h \tag{12}$$

Figure 1 provides some intuition for this result: Consider the density of observations in the shaded square. This density changes, as α changes, by (i) the difference between the outflow to the right and the inflow to the left, $\partial_{x^1} h^1 \cdot dx^1$, and (ii) the difference between the outflow on the top and the inflow on the bottom, $\partial_{x^2} h^2 \cdot dx^2$. The sum of these changes is equal to $-\sum_{j=1}^k \partial_{x^j} h^j \cdot dx^j$.

Our next result, theorem 2, shows that the data *only* identify the divergence of h. Any h such that $\dot{f} = -\nabla \cdot h$ is consistent with the observed data and assumption 2. Theorem 2 explicitly constructs one particular function h^0 which satisfies the equation $\dot{f} = -\nabla \cdot h$. It further shows that the difference $\tilde{h}$ between this function and any other function h in the identified set is in the set $\{\tilde{h} : \nabla \cdot \tilde{h} \equiv 0\}$.

FIGURE 1.— Divergence of flow and change of density



Notes: This figure illustrates theorem 1. It relates the change of density f (mass in the square) to the divergence of h (difference in flow on different sides).

Theorem 2 *Suppose assumption 2 holds.*
Let v_j be the random variable $v^j = F(x^j|x^1, \ldots, x^{j-1}, \alpha)$, define

$$h^{0j}(x, \alpha) = f(x|\alpha) \cdot \partial_\alpha Q(v^j|v^1, \ldots, v^{j-1}, \alpha), \tag{13}$$

and let

$$\mathscr{H} = \{\tilde{h} : \nabla \cdot \tilde{h} \equiv 0, \ \tilde{h}(x, \alpha) = 0 \text{ for } x \notin \mathbf{X}\}. \tag{14}$$

Then the identified set for h is given by

$$h^0 + \mathscr{H}. \tag{15}$$

Theorem 2 shows that the identified set for h is equal to $h^0 + \mathscr{H}$. Point identification fails if $\mathscr{H}$ has more than one element. Our next result, theorem 3, characterizes the nature of non-identification if this is the case. This theorem provides alternative representations of the "kernel" of the identified set which is given by $\mathscr{H} = \{\tilde{h} : \nabla \cdot \tilde{h} \equiv 0\}$. This is the set of flows that can be generated by "stirring the coffee," leaving the density of x invariant. Theorem 3 uses Poincaré's Lemma to characterize the set $\mathscr{H}$ for dimensions $k = 1, 2$, and 3.[6]

The case $k = 2$ is of special interest in the context of this paper – recall that $x = (w, l)$ in the simplified version of assumption 1 considered at the outset of this section. For the case $k = 2$, the characterization takes on a particularly elegant form. In this case, the functions $\tilde{h}$ in the kernel are exactly those functions which can be written as the gradient of some function H, rotated by 90 degrees. $\tilde{h}$ is thus a vector field pointing along the lines of constant height of H. Figure

[6]Similar results can be stated for higher dimensions, but require increasingly cumbersome notation.

2 illustrates.

Theorem 3 *Suppose assumption 2 holds.*

1. *Suppose $k = 1$. Then*[7]

$$\mathscr{H} = \{\tilde{h} \equiv 0\}. \tag{16}$$

2. *Suppose $k = 2$. Then*

$$\mathscr{H} = \{\tilde{h} : \ \tilde{h} = A \cdot \nabla H, \ H(x, \alpha) = 0 \textit{ for } x \notin \mathbf{X}\}. \tag{17}$$

where

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

3. *Suppose $k = 3$. Then*

$$\mathscr{H} = \{\tilde{h} : \ \tilde{h} = \nabla \times G, \ G(x, \alpha) = 0 \textit{ for } x \notin \mathbf{X}\}.\}. \tag{18}$$

where

$$\nabla \times G = \begin{pmatrix} \partial_{x^2} G^3 - \partial_{x^3} G^2 \\ \partial_{x^3} G^1 - \partial_{x^1} G^3 \\ \partial_{x^1} G^2 - \partial_{x^2} G^1 \end{pmatrix}.$$

Theorems 2 and 3 characterize the identified set for h if only assumption 2 is imposed. The following theorem shows that the additional assumption of a "triangular" structure for $\nabla E[\dot{x}^i | x, \alpha]$ (derivatives above the diagonal are 0) yields just-identification of h. Note that the ordering of the components of x matters if we assume such a triangular structure! The identified h differs depending on which ordering the triangular structure is imposed for.

Theorem 4 *Suppose assumption 2 holds. Assume additionally that*

$$\partial_{x^j} E[\dot{x}^i | x, \alpha] = 0 \textit{ for } j > i. \tag{19}$$

Then g and h are point identified, and

$$g(x, \alpha) = \partial_\alpha Q(v^j | v^1, \ldots, v^{j-1}, \alpha), \tag{20}$$

where $v^j = F(x^j | x^1, \ldots, x^{j-1}, \alpha)$. The flow density h is equal to h^0 as defined in theorem 2.

---

[7] This can be interpreted as a version of the result shown by Hoderlein and Mammen (2007). Non-identification for the case $k = 2$ was recognized by Hoderlein and Mammen (2009).

We conclude this section by discussing conditions which yield just identification of the structural functions x themselves. Such conditions have been explored by Rosa Matzkin, in particular Matzkin (2003). Point identification of structural functions follows under the rather restrictive conditions that (i) the dimensionality of unobserved heterogeneity is no larger than the dimensionality of endogenous outcomes y, and (ii) a triangular structure as in theorem 4 is imposed. Point identification of structural functions is useful in the context of discrete changes of α.

Theorem 5 *Suppose assumption 2 holds. Assume additionally that* $\epsilon \in \mathbb{R}^k$ *and*

$$x^j(\alpha, \epsilon) = x(\alpha, \epsilon^1, \ldots, \epsilon^j) \tag{21}$$

is strictly monotonically increasing in ϵ^j *and does not depend on* $\epsilon^{j+1}, \ldots, \epsilon^k$.

Then $(\epsilon^1, \ldots, \epsilon^j)$ *is a one-to-one transformation of* $(v^1, \ldots, v^j)$ *for any* $j \leq k$, *where*

$$v^j = F(x^j | x^1, \ldots, x^{j-1}, \alpha) = F(\epsilon^j | \epsilon^1, \ldots, \epsilon^{j-1}) \tag{22}$$

and

$$x^j(\alpha', \epsilon) = Q^{x^j}(v^j | v^1, \ldots, v^{j-1}, \alpha') \tag{23}$$

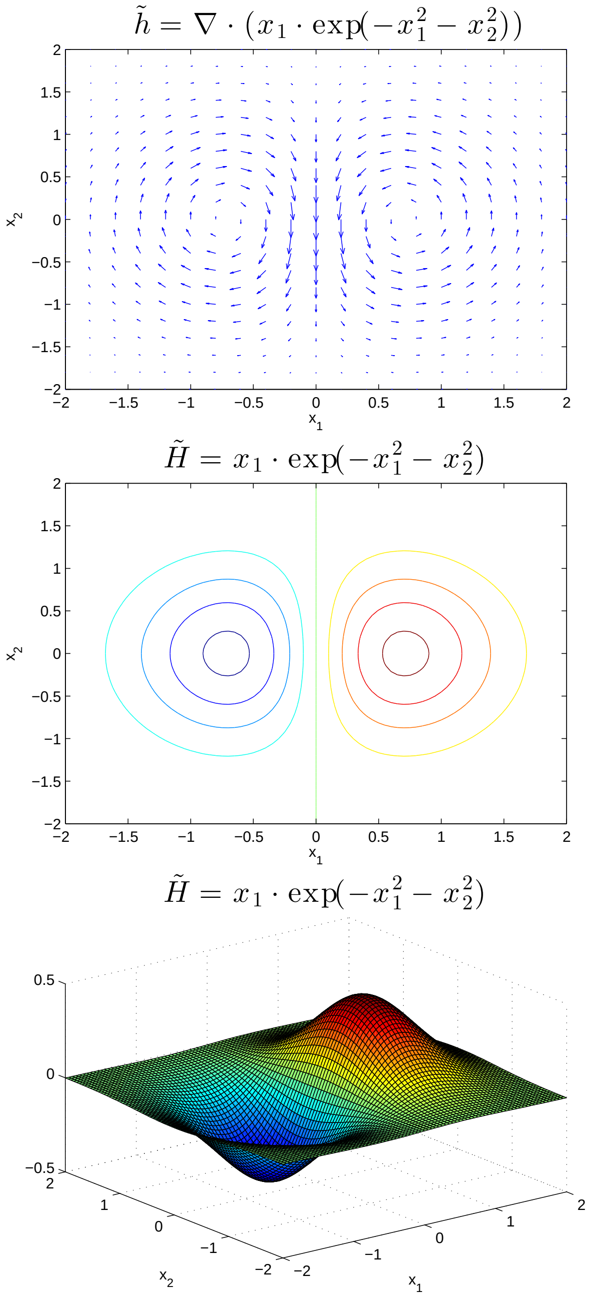
for any α, α'.

3.2. *Controls, instruments, and panel data*

In section 3.1, we considered the problem of identifying $E[\dot{x}|x, \alpha]$ under the assumption that α is randomly assigned. Most distributional evaluations have to rely on observational data in settings where this assumption can not plausibly be maintained. In this section we discuss identification of $E[\dot{x}|x, W, \alpha]$ if either (i) α is conditionally random, or (ii) there is a valid instrument Z, or (iii) we have panel data where changes of α over time are independent of changes of other factors affecting outcomes. Proposition 1 through 3 are generalizations of theorem 4 to these cases.

The following proposition 1 considers the approach taken by most of the distributional decomposition literature whenever decompositions are given a causal interpretation: It is assumed that treatment α is independent of unobserved heterogeneity ϵ once we condition on a set of available covariates W. This assumption might be a reasonable approximation to the truth when a rich set of covariates is available. Proposition 1 shows that under this condition policy effects on x (labor supply l and wage w) can be imputed using a quantile regression with the appropriate controls W and $v^{j'}$.

FIGURE 2.— Incompressible flow and rotated gradient of potential



Notes: This figure illustrates the characterization of $\mathscr{H}$ in theorem 3 for the case $k = 2$. The vector field $\tilde{h}$ (first graph) is given by a 90 degree rotation of the gradient of some function H (third graph), and thus points along the lines of equal level of the function H (second graph).

Proposition 1 (Controls) *Suppose assumption 2 holds, except that instead of* $\alpha \perp \epsilon$ *we have* $\alpha \perp \epsilon | W$. *Assume additionally that*

$$\partial_{x^j} E[\dot{x}^i | x, W, \alpha] = 0 \text{ for } j > i.$$

Then $E[\dot{x}^j | x, W\alpha]$ *is point identified for* (x, W, α) *in the interior of the support of the data, and equal to*

$$E[\dot{x}^j | x, W, \alpha] = \partial_\alpha Q(v^j | v^1, \ldots, v^{j-1}, W, \alpha), \tag{24}$$

where $v^j = F(x^j | x^1, \ldots, x^{j-1}, W, \alpha)$.

In settings where conditional independence of α can not plausibly be maintained, we might still have an instrument Z for which conditional independence holds, and which affects outcomes only through its effects on α. In the spirit of nonparametric identification, we would like identification not to depend on restrictions of functional form or the dimensionality of unobserved heterogeneity. Kasy (2014) shows identification of potential outcome distributions for the fully nonparametric case, assuming monotonicity of the first stage in the instrument and sufficient support of the data; the following proposition 2 reviews this result.

Proposition 2 (Instruments) *Suppose assumption 2 holds, except that instead of* $\alpha \perp \epsilon$ *we have* $Z \perp (\epsilon, \eta) | W$. *Assume additionally that* $\alpha = \alpha(Z, \eta)$, *where* $\alpha(., \eta)$ *is continuous and strictly increasing in* Z *for all* η. *Define the weighting function*

$$\varphi(\alpha, z, W) := -\frac{\partial_z F(\alpha | z, W)}{\partial_z F(z | \alpha, W)}, \tag{25}$$

assuming all derivatives and the ratio are well defined. Assume finally that $F(\alpha | z, W)$ *has full support* $[0, 1]$ *given* α *and* W. *Then*

$$f^{x^\alpha}(x | W) = f(x | \alpha, W) \cdot \varphi(\alpha, z, W), \tag{26}$$

and proposition 1 applies to the observed data distribution reweighted by φ.

In practice, the support requirement that $F(\alpha | z, W)$ has full support $[0, 1]$ given α and W might be fairly restrictive. If support is insufficient, we might proceed using the control function approach (Imbens and Newey, 2009), using $v^z := F(\alpha | z, W)$ as additional control in the quantile regression $Q(v^j | v^1, \ldots, v^{j-1}, v^z, W, \alpha)$, and relying on linearity assumptions to extrapolate outside the support of the data. For the case of sufficient support, it is shown in Kasy (2014) that the control function approach yields the same estimates as the reweighting approach of proposition 2.

The following proposition considers a panel data setup, where α varies as a function of time within groups s (states or metropolitan areas, for instance).

Similar to many approaches in the "Difference-in-differences" mould, such as Chamberlain (1984), Athey and Imbens (2006), and Graham and Powell (2012), we assume that the distribution of heterogeneity ϵ does not vary over time within states s. Time τ is allowed to have a causal impact on outcomes x, which is however assumed to not interact with the level of α.

Proposition 3 (Panel data) *Suppose assumption 2 holds, except that* $x = x(\alpha, \tau, \epsilon)$*, and* $\tau \perp \epsilon | s, W$*, where* $\alpha = \alpha(s, \tau)$*. Assume additionally that*

$$\begin{aligned} E[\dot{x}^j | x, s, W, \tau, \alpha] &= \quad E[\dot{x}^j | x^1, \ldots, x^j, W, \alpha] \\ E[\partial_\tau x^j | x, s, W, \tau, \alpha] &= \quad E[\partial_\tau x^j | x^1, \ldots, x^j, W, \tau] \end{aligned}$$

for $j = 1, \ldots, k$*. Then, for* $v^j = F(x^j | x^1, \ldots, x^{j-1}, s, W, \tau)$*,*

$$\begin{aligned} \partial_\tau Q(v^j | v^1, \ldots, v^{j-1}, s, W, \tau) = & \quad \partial_\tau \alpha(s, \tau) \cdot E[\dot{x}^j | x, W, \alpha] \\ & + E[\partial_\tau x^j | x, W, \tau] \end{aligned}$$

If, in particular, $\partial_\tau \alpha(s, \tau)$ *varies across* s *given* t *and* α*, then* $E[\dot{x}^j | x, W, \alpha]$ *is identified.*

The crucial identifying assumptions of this proposition are:

1. Heterogeneity is constant over time within states and given covariates. This assumption is known as "marginal stationarity" in the nonparametric panel literature.
2. The conditional average causal effects $E[(\dot{x}^i, \partial_\tau x^i) | x, s, W, \tau, \alpha]$ are the same for every state s. This is strictly weaker than the "common trends" assumption of Difference-in-difference models. This is also strictly weaker than the "changes-in-changes" assumption of Athey and Imbens (2006).

4. AGGREGATION

Lemma 1 characterizes the effect of a policy change on individual net income y and on money-metric utility e. In this section we discuss the corresponding effect on aggregate statistics θ of the income distribution and on social welfare SWF. Section 3 provided conditions sufficient for identification of $\gamma(y, W) = E[\dot{e} | y, W, \alpha]$. Under some restrictions on welfare weights to be discussed below, policy effects on social welfare can be written as $S\dot{W}F = E[\omega \cdot \gamma]$, so that identification of γ implies identification of $S\dot{W}F$.

We prove the following claims: (i) As long as we consider marginal policy changes, we can restrict our attention to social welfare functions which are linear in money metric utility. Policy effects on social welfare are a weighted average of their effect on individual welfare, $S\dot{W}F = E[\omega^{SWF} \cdot \dot{e}]$. Here ω^{SWF} is the welfare weight, or marginal value of an additional dollar, assigned to each individual.

(ii) Effects on social welfare relate to effects on statistics of the income distribution in that a) welfare effects are effects on income net of behavioral effects, and b) welfare weights correspond to the derivative of the influence function for distributional statistics. (iii) There are various equivalent ways of calculating $\dot{SWF}$ which are based on imputing conditional expected welfare effects γ or some counterfactual income to each individual. These equivalent representations can be used for alternative estimation approaches.

In addition to effects on social welfare, we discuss in this section effects on aggregate statistics θ of the income distribution. Typical examples of such distributional statistics are mean and variance, quantiles, and measures of inequality such as the Gini coefficient.

Definition 2 (Objects of interest – net income)

1. *Expected conditional policy effect on net income:* $\beta(y, W) := E[\dot{y}|y, W, \alpha]$
2. *Policy effect on a distributional statistic:* $\dot{\theta}$,
   where the distributional statistic θ *maps* P_y *into* $\mathbb{R}$.

In order to elegantly characterize and relate $\dot{\theta}$ and $\dot{SWF}$, we need to impose additional differentiability conditions on either functional; the following assumption 3 does so. Definition 2 assumes θ is a statistic of the income distribution P_y. We can also, however, think of it as a functional of the random variable $(y_i : i \in \mathscr{I})$.[8] The random variable y has a probability distribution P_y, where the latter "forgets" about the index i – who earns how much. The following assumption imposes differentiability of θ for either representation.

Assumption 3 (Differentiability)

1. SWF *is Gateaux-differentiable*[9] *on the set of random variables* v*, equipped with the* L^2 *norm.*
   θ *is Gateaux-differentiable on the set of random variables* y*, equipped with the* L^2 *norm.*
2. θ *is Gateaux-differentiable on the set of probability distributions* P_y*, equipped with some norm, so that the influence function* $IF(y)$ *of* θ *exists.*
3. *The influence function* $IF(y)$ *of* θ *is differentiable in* y*.*

Theorem 6 (Welfare weights and influence functions)
Suppose that assumption 1 holds. Let $\dot{y}$ *and* $\dot{e}$ *be the impact of a marginal policy change on individuals' income and welfare at* $\alpha = 0$*, and consider the corresponding impact on* θ *and* SWF*.*

[8]The random variables y and v map the underlying probability space $\mathscr{I}$ of individuals i, endowed with the uniform distribution, into $\mathbb{R}$.

[9]A functional is "Gateaux-differentiable" if it is differentiable along paths in the spaces of random variables or probability measures. For finite populations $i = 1, \ldots, N$, "Gateaux-differentiability" corresponds to the usual notion of differentiability.

1. ***Welfare weights:***
   Suppose that assumption 3.1 holds. Then there exist random variables ω^{SWF} and ω^{θ}[10] *such that*

$$S\dot{W}F = E[\omega^{SWF} \cdot \dot{e}] \tag{27}$$

$$\dot{\theta} = E[\omega^{\theta} \cdot \dot{y}]. \tag{28}$$

2. ***Influence function:***
   Suppose that assumption 3.2 holds. Then

$$\dot{\theta} = \partial_{\alpha} E\left[IF(y^{\alpha})\right] = \partial_{\alpha} \int IF(y) dF_{y^{\alpha}}(y). \tag{29}$$

3. ***Relating the two:***
   Suppose that assumptions 3.1-3.3 hold. Then

$$\omega^{\theta} = \partial_y IF(y). \tag{30}$$

Remarks:

- It is instructive to consider the case of a finite population. In that case, the welfare weights of equation (27) are equal to

$$\omega_i = \partial_{v_i} SWF(v) \cdot \partial_{y_0} v_i. \tag{31}$$

  This is the relative value attached to a marginal dollar for a given individual. Saez and Stantcheva (2013) argue for a direct specification of such weights (without the detour over some social welfare function), in order to reflect distributional preferences. In the majority of public finance applications, ω^{SWF} is a function of y.
- Differentiability of the influence function of θ, as required for the identity $\omega^{\theta} = \partial_y IF(y)$ is violated for some distributional statistics of interest, most notably quantiles. We can think of quantiles as assigning "infinite weight" to the welfare (income) of individuals right at the quantile. Differentiability holds for moments of the form $\nu = E[G(y)]$ for differentiable G, and for statistics which are locally well approximated by such moments.
- Theorem 6 provides two representations of $\dot{\theta}$, the first in terms of welfare weights and the second in terms of the influence function. These two representations correspond to the two ways of thinking about θ, as a functional of the random variable y and as a functional of the distribution P_y.
- There are two ways for estimating $\dot{\theta}$ proposed in the distributional decomposition literature, reweighting DiNardo et al. (1996) and RIF regression Firpo et al. (2009). Reweighting corresponds to directly estimating

[10] More precisely, $(\omega_i^{SWF} : i \in \mathscr{I})$ and $(\omega_i^{\theta} : i \in \mathscr{I})$.

$\partial_\alpha \theta\left(P_{y^\alpha}\right)$ after constructing the counterfactual distributions P_{y^α}. RIF regression corresponds to estimating $E\left[IF(y^\alpha)\right]$ by suitable regressions of $IF(y)$ on α and controls.

The following theorem provides alternative representations of $\dot{SWF}$ under the assumption that the welfare weights ω^{SWF} are a function of y, and that $\omega^{SWF} = \omega^\theta$, which allows to relate $\dot{SWF}$ to $\dot{\theta}$.

Theorem 7 (Counterfactual income and behavioral correction) *Suppose that assumptions 1 and 3 hold. Assume further that $\omega^{SWF} = \omega^\theta = \omega$ and that $\dot{p} = 0$. Define the counterfactual income $\tilde{y}^\alpha = l^0 \cdot w^\alpha - t^\alpha(l^0 \cdot w^\alpha) + y_0^\alpha$. and the behavioral effect $b = \dot{l} \cdot n$.*

Then

$$\dot{e} = \dot{\tilde{y}} = \dot{y} - b, \tag{32}$$

and $\dot{SWF}$ can be rewritten in the following ways.

1. ***Welfare weights:***

$$\begin{aligned} \dot{SWF} &= E[\omega \cdot \dot{\tilde{y}}] \\ &= E[\omega \cdot \gamma] \end{aligned} \tag{33}$$

   with γ as in definition 1.

2. ***Counterfactual income distribution:***

$$\dot{SWF} = \partial_\alpha \theta\left(P_{\tilde{y}^\alpha}\right). \tag{34}$$

3. ***Influence function:***

$$\dot{SWF} = \partial_\alpha E\left[IF(\tilde{y}^\alpha)\right] = \partial_\alpha \int IF(y) dF_{\tilde{y}^\alpha}(y). \tag{35}$$

4. ***Behavioral correction of distributional decomposition:***

$$\begin{aligned} \dot{\theta} - \dot{SWF} &= E[\omega \cdot b] \\ &= \partial_\alpha \theta\left(P_{\check{y}^\alpha}\right) \\ &= \partial_\alpha E\left[IF(\check{y}^\alpha)\right]. \end{aligned} \tag{36}$$

   where $\check{y}^\alpha = l^\alpha \cdot w^0 - t^0(l^\alpha \cdot w^0) + y_0^0$.

Remarks:

- Theorem 7 defines two counterfactual income variables, $\tilde{y}^\alpha$ and $\check{y}^\alpha$. $\tilde{y}^\alpha$ is the income an individual would receive given baseline ($\alpha = 0$) labor supply and policy α wages, taxes, and unearned income. The derivative of $\tilde{y}^\alpha$ with respect to α at $\alpha = 0$ gives the welfare effect $\dot{e}$.

$\check{y}^\alpha$ is the income an individual would receive given policy α labor supply and baseline ($\alpha = 0$) wages, taxes, and unearned income. The derivative of $\check{y}^\alpha$ with respect to α at $\alpha = 0$ gives the "behavioral correction" $b = \dot{l} \cdot n$.

- The equivalent representations of $S\dot{W}F$ in theorem 7 suggest several alternative ways of estimating $S\dot{W}F$:
  1. We can impute an estimate of $\gamma(y, W) = E[\dot{e}|y, W, \alpha]$ to every observation, and then use $S\dot{W}F = E[\omega \cdot \gamma]$, where welfare weights ω are directly specified. This is the route we will pursue.
  2. We can impute $\tilde{y}^\alpha$, based on counterfactual wages, taxes, and unearned income to individuals in the baseline sample. Or impute $\tilde{y}^\alpha$, based on counterfactual labor supply to individuals in the policy α sample. Either way, we can apply distributional decomposition methods such as reweighting or RIF regression for statistics of the distribution of $\tilde{y}^\alpha$.
  3. We can impute $\check{y}^\alpha$, similarly to imputing $\tilde{y}^\alpha$, and apply one of the decomposition methods to the distribution of $\check{y}^\alpha$. We can then use $S\dot{W}F = \dot{\theta} - \partial_\alpha \theta \left(P_{\check{y}^\alpha}\right)$ and thus obtain $S\dot{W}F$ by applying a "behavioral correction" to a standard decomposition.

  The first of these approaches has two important advantages. First, it is possible to identify γ under weaker conditions then necessary to identify counterfactual outcomes such as $\tilde{y}^\alpha$ and $\check{y}^\alpha$. Second, this approach allows to directly construct estimates of the sets of winners and losers, $\mathscr{W}$ and $\mathscr{L}$, and to plot the conditional expectation of $\dot{e}$ given baseline income or other variables.

5. ESTIMATION

This section discusses estimation based on the identification results of section 3 and the aggregation results of section 4. We first consider the baseline case as discussed in section 3.1, with random variation in α and no covariates W. We provide an estimator for $g(x, \alpha) = E[\dot{x}|x, \alpha]$ in this baseline case, using the identification-result of theorem 4. The proposed procedure estimates $\partial_\alpha Q$ by local linear quantile regression, and replaces the "control-functions" v_j by estimated versions thereof. We then generalize this estimation procedure to the settings considered in section 3.2, using controls, instrumental variables, or panel data.

No matter how g is estimated, we can impute estimated values of g for every individual in a baseline sample. These estimated values can in turn be used to construct estimates of γ, $\mathscr{W}$, $\mathscr{L}$, and $S\dot{W}F$. This is discussed in section 5.3. In section 5.4 we discuss estimation of the structural functions $x(\alpha, \epsilon)$ under the more restrictive identifying assumptions of theorem 5. The section concludes with a brief discussion of inference. Analytic standard errors are complicated to construct in our setting and require re-derivation of influence functions for every object of interest and every identification approach; we opt instead for a procedure based on the Bayesian bootstrap. This is described in section 5.5.

5.1. *Estimation of g in the baseline case*

Suppose that the assumptions of theorem 4 hold. Denote sample averages by E_N, so that for instance $E_N[x] = 1/N \sum_i x_i$. Then $g(x, \alpha) = E[\dot{x}|x, \alpha]$ can be estimated by iterating the following procedure over the components $j = 1, \ldots, k$ of g:

1. Fix a point (x, α) and take $(\widehat{v}^1, \ldots, \widehat{v}^{j-1})$ as given.
2. Define the following local weights around $(\alpha, \widehat{v}^1, \ldots \widehat{v}^{j-1})$.

$$K_i^j = \frac{1}{\rho^j} \cdot K\left(\frac{1}{\rho}\left\|\alpha_i - \alpha, \widehat{v}_i^1 - \widehat{v}^1, \ldots, \widehat{v}_i^{j-1} - \widehat{v}^{j-1}\right\|\right) \tag{37}$$

   for a kernel function K[11] and a suitably chosen bandwidth ρ.[12]
3. Let

$$\widehat{v}^j = \frac{E_N[K_i^j \cdot \mathbf{1}(x_i^j \leq x^j)]}{E_N[K_i^j]}. \tag{38}$$

4. Let, finally,

$$\widehat{g}^j = \underset{g^j}{\operatorname{argmin}}\ E_N\left[K_i^j \cdot U_i^j \cdot (\widehat{v}^j - \mathbf{1}(U_i^j \leq 0))\right], \text{ where} \tag{39}$$

$$U_i^j = x_i^j - x^j - \alpha \cdot g^j. \tag{40}$$

Then $\widehat{v}^j$ is an estimate of $v^j = F(x^j|x^1, \ldots, x^{j-1}, \alpha)$ and $\widehat{g}^j$ is an estimate of $\partial_\alpha Q(v^j|v^1, \ldots, v^{j-1}, \alpha)$. The latter is equal to $g^j(x, \alpha) = E[\dot{x}^j|x, \alpha]$ under the assumptions of theorem 4.

5.2. *Estimation of g using controls, instruments, or panel data*

In the context of the applications of interest for the procedures considered in this paper, experimental variation of α will rarely be available. The estimator just sketched immediately generalizes, however, to the more general settings considered in 3.2. The estimator has to be modified as follows to be used in these settings.

1. **Controls**
   Suppose that the assumptions of proposition 1 hold. Then the estimator

[11] For instance the Epanechnikov-kernel $K(a) = \max(0, 1 - a^2)$.

[12] We use a common bandwidth ρ for all variables for simplicity of notation; in general different bandwidth for different variables might be desirable.

of section 5.1 can be used to estimate $g(x, W, \alpha) = E[\dot{x}^j | x, W, \alpha]$ once we replace the local weights by

$$K_i^j = \frac{1}{\rho^{\dim(W)+j}} \cdot K\left(\frac{1}{\rho}\left\|\alpha_i - \alpha, W_i - W, \widehat{v}_i^1 - \widehat{v}^1, \ldots, \widehat{v}_i^{j-1} - \widehat{v}^{j-1}\right\|\right). \tag{41}$$

2. **Instruments**
Suppose that the assumptions of proposition 2 hold. Then the estimator for the case of controls can be used to estimate $g(x, W, \alpha)$ after reweighting the data by $\widehat{\varphi}(\alpha, z, W)$, where

$$\widehat{\varphi}(\alpha, z, W) := -\frac{\partial_z \widehat{F}(\alpha|z, W)}{\widehat{f}(z|\alpha, W)}. \tag{42}$$

We can use a kernel density estimator for the denominator,

$$\widehat{f}(z|\alpha, W) = \frac{1}{\rho} \frac{\sum_i K\left(\frac{1}{\rho}(\alpha_i - \alpha)\right) \cdot K\left(\frac{1}{\rho}\|W_i - W, Z_i - z\|\right)}{\sum_i K\left(\frac{1}{\rho}\|\alpha_i - \alpha, W_i - W\|\right)},$$

and a local linear regression estimator for the numerator,

$$\partial_z \widehat{F}(\alpha|z, W) = \underset{b}{\operatorname{argmin}} \min_a \sum_i \left(L\left((\alpha - \alpha_i)/\rho_\alpha\right) - a - b \cdot (Z_i - z)\right)^2 \cdot K\left(\frac{1}{\rho}\|W_i - W, Z_i - z\|\right). \tag{43}$$

In the latter expression, L is the cumulative distribution function of a smooth symmetric distribution with support $[-1, 1]$, and ρ_α is a further bandwidth parameter. The "dependent variable" $L\left((\alpha - \alpha_i)/\rho_\alpha\right)$ in this regression is a smoothed version of the indicator $\mathbf{1}(\alpha - \alpha_i \leq 0)$.

3. **Panel data**
Suppose that the assumptions of proposition 3 hold. Then $g(x, W, \alpha)$ can be estimated using a two-stage approach:
   (a) Estimate $\partial_\tau Q(v^j | v^1, \ldots, v^{j-1}, s, W, \tau)$ using the exact same estimator as for the case of estimation with controls, with τ taking the place of α.
   (b) Then regress $\partial_\tau Q(v^j | v^1, \ldots, v^{j-1}, s, W, \tau)$ on $\partial_\tau \alpha(s, \tau)$ across values of s and τ. The slope of this regression provides an estimator of $g(x, W, \alpha)$.

5.3. *Estimation of γ, $\mathscr{W}$, $\mathscr{L}$, and $S\dot{W}F$*

Ultimately, we are not interested in $g(x, W, \alpha) = E[\dot{x}|x, W, \alpha]$ itself, but rather in derived objects such as γ, $\mathscr{W}$, $\mathscr{L}$, and $S\dot{W}F$ as introduced in definition 1. To estimate γ, we need to take a suitable average of the estimated conditional effects of the policy change on w, y^0, and t. Assume for simplicity that $\dot{y}^0 = \dot{t} = \dot{p} = 0$, and that $w = x^j$. Then $\dot{e} = l \cdot \dot{w}$, and we can estimate γ by

$$\widehat{\gamma}(y, W) = E_N\left[K_i \cdot l \cdot (1 - \partial_z t) \cdot \widehat{\dot{w}}\right] / E_N[K_i] \tag{44}$$

where $\widehat{\dot{w}} = \widehat{g}^j$ is estimated using any of the approaches we discussed (experimental variation, controls, instruments, panel data), and

$$K_i = \frac{1}{\rho^{2+\dim(W)}} \cdot K\left(\frac{1}{\rho}\left\|\alpha_i - 0, y_i - y, W_i - W\right\|\right). \tag{45}$$

We can finally plug our estimate of γ into the definitions of $\mathscr{W}$ and $\mathscr{L}$, and into the first characterization of $S\dot{W}F$ in theorem 7 to obtain

$$\begin{aligned} \widehat{\mathscr{W}} &= \{(y, W) : \widehat{\gamma}(y, W) \geq 0\} \\ \widehat{\mathscr{L}} &= \{(y, W) : \widehat{\gamma}(y, W) \leq 0\} \\ \widehat{S\dot{W}F} &= E_N[\omega_i \cdot \widehat{\gamma}(y_i, W_i)]. \end{aligned} \tag{46}$$

We can furthermore obtain estimates of objects characterizing the sets of winners and losers, for instance the moments of covariates for each of these sets,

$$\widehat{E}[W|\mathscr{W}] = \frac{E_N[W \cdot \mathbf{1}(\widehat{\gamma}(y_i, W_i) > 0)]}{E_N[\mathbf{1}(\widehat{\gamma}(y_i, W_i) > 0)]}. \tag{47}$$

5.4. *Estimation of $x(., \epsilon)$ under stronger restrictions of heterogeneity*

So far we have discussed estimation of $g(x, \alpha) = E[\dot{x}|x, \alpha]$, and of objects which are functions of g. If we are willing to put stronger restrictions on heterogeneity, as in theorem 5, we can identify and estimate the structural functions $x(\alpha, \epsilon)$ themselves, using nonparametric quantile regressions. Assume that the assumptions of theorem 5 hold, and that w.l.o.g. $\epsilon^j = v^j$; this is just a normalization of scale for ϵ. Under the assumptions of theorem 5, this normalization implies

$$\epsilon^j = F(x^j | x^1, \ldots, x^{j-1}, \alpha) = F(\epsilon^j | \epsilon^1, \ldots, \epsilon^{j-1}).$$

We can then estimate $x(\alpha, \epsilon)$ by

$$\begin{aligned} \widehat{x}^j(\alpha', \epsilon) &= \widehat{Q}^{x^j}(\epsilon^j | \epsilon^1, \ldots, \epsilon^{j-1}, \alpha') \\ &= \underset{x^j}{\operatorname{argmin}}\; E_N\left[K_i^j \cdot (x_i^j - x^j) \cdot (\epsilon^j - \mathbf{1}(x_i^j - x^j \leq 0))\right] \end{aligned} \tag{48}$$

where

$$K_i^j = \frac{1}{\rho^j} \cdot K\left(\frac{1}{\rho}\left\|\alpha_i - \alpha, \widehat{v}_i^1 - \epsilon^1, \ldots, \widehat{v}_i^{j-1} - \epsilon^{j-1}\right\|\right) \tag{49}$$

and $\widehat{v}$ is as in section 5.1.

5.5. *Standard errors and confidence sets*

Inference on all parameters of interest ϑ we consider could proceed using the standard approach of deriving a linear (first-order) approximation to the statistic of interest, and estimating the variance of the corresponding "influence-function," plugging in estimators of any relevant nuisance-parameters; see for instance Newey (1994a). The asymptotic variance of $c_n \cdot (\widehat{\vartheta} - E(\widehat{\vartheta}))$ (rescaled by an appropriate diverging sequence c_n), in particular, can be consistently estimated by c_n^2/n times the sample variance of the influence function of $\widehat{\vartheta}$, so that

$$\mathrm{Var}\left(\widehat{\vartheta}\right) \approx \frac{1}{n^2}\sum_i \widehat{\psi}_i^2,$$

where $\widehat{\psi}_i = \frac{\partial\widehat{\vartheta}}{\partial p_n(w_i,l_i,W_i,\ldots)}$. The derivative in the last expression is to be understood as the derivative of $\widehat{\vartheta}$ with respect to the mass p_n put by the empirical distribution on the i^{th} observation. Details and background can be found in (van der Vaart, 2000, chapter 20) and Newey (1994b).

While possible in principle, such an approach requires a separate derivation of influence functions for each object of interest and each identification approach. This is rendered cumbersome, in particular, by the presence of the generated regressors $\widehat{v}_i^j$; cf. Hahn and Ridder (2013).

We opt for an alternative approach, the Bayesian bootstrap introduced by Rubin (1981), and discussed by Chamberlain and Imbens (2003). This approach proceeds as follows:

1. Draw i.i.d. exponentially distributed random variables V_i.
2. Reweight each observation by $V_i/\sum_{i'} V_{i'}$.
3. Estimate the object of interest for the reweighted distribution.
4. Iterate the entire procedure, to obtain a set of R replicate estimates $(\widehat{\vartheta}_r)_{r=1}^R$ for the object of interest.

The estimates $\widehat{\vartheta}_r$ obtained by this procedure are draws from the posterior distribution for the object of interest when the prior over the joint distribution of all observables is a Dirichlet process with parameter 0.[13] This allows, in particular,

[13]Strictly speaking, this is an improper prior which is the limit of a sequence of proper Dirichlet processes.

to construct Bayesian credible sets for the object of interest, using quantiles of the sampling distribution $(\widehat{\vartheta}_r)_{r=1}^R$ as boundary values of the credible sets.

The re-sampling distribution of the object of interest can also be considered as an approximation to the frequentist asymptotic distribution for objects satisfying certain regularity conditions, in particular sample moments and smooth functions thereof. This allows to interpret the Bayesian credible sets as frequentist confidence sets. All our objects of interest are functions of sample moments, for given (fixed) bandwidth parameters. In appendix B we provide evidence on the frequentist accuracy of this inference procedure using calibrated Monte Carlo simulations.

6. APPLICATION

We shall now turn to an application of the proposed methods. This section reevaluates the welfare impact of the extension of the Earned Income Tax Credit (EITC) during the 1990s. A large literature documents that the EITC expansion increased labor supply, see for instance Meyer and Rosenbaum (2001) and Chetty et al. (2013). Rothstein (2010) and Leigh (2010) note that these increases in labor supply are likely to depress wages in the labor markets affected. If this is so, the effective incidence of the EITC might be quite different from the nominal incidence.

Following up on this argument, this section provides a disaggregated welfare-evaluation of the EITC expansion using the framework introduced in section 2. We estimate the impact of the EITC expansion using variation across states and time in state-level supplements to the federal EITC, as in Leigh (2010).

6.1. *Data and background*

Table I, which reproduces table 2 from Leigh (2010), shows the variation of state supplements to federal EITC payments across states and time for those states that do provide supplements. Effective EITC payments are equal to (federal EITC payments)(1+state EITC supplement). We use variation of these supplements, interacted with the federal expansion of EITC payments over the period considered, in order to identify the impact of the EITC expansion on wages and welfare conditional on initial incomes.

Table II reports the main estimates from table 4 and 5 of Leigh (2010). These estimates imply that the expansion of the EITC increased labor supply and depressed wages of those without high school diplomas, while only having a smaller effect on the rest of the population.

[to be continued]

TABLE I.— State EITC supplements 1984-2002

State:	**CO**	**DC**	**IA**	**IL**	**KS**	**MA**	**MD**	**ME**	**MN**	**MN**	**NJ**	**NY**	**OK**	**OR**	**RI**	**VT**	**WI**	**WI**	**WI**
# of children:							1+		0	1+		1+					1	2	3+
1984																	30	30	30
1985																	30	30	30
1986															22.21				
1987															23.46				
1988															22.96	23			
1989															22.96	25	5	25	75
1990			5												22.96	28	5	25	75
1991			6.5						10	10					27.5	28	5	25	75
1992			6.5						10	10					27.5	28	5	25	75
1993			6.5						15	15					27.5	28	5	25	75
1994			6.5						15	15		7.5			27.5	25	4.4	20.8	62.5
1995			6.5						15	15		10			27.5	25	4	16	50
1996			6.5						15	15		20			27.5	25	4	14	43
1997			6.5			10			15	15		20		5	27.5	25	4	14	43
1998			6.5		10	10	10		15	25		20		5	27	25	4	14	43
1999	8.5		6.5		10	10	10		25	25		20		5	26.5	25	4	14	43
2000	10	10	6.5	5	10	10	15	5	25	25	10	22.5		5	26	32	4	14	43
2001	10	25	6.5	5	10	15	16	5	33	33	15	25		5	25.5	32	4	14	43
2002	0	25	6.5	5	15	15	16	5	33	33	17.5	27.5	5	5	25	32	4	14	43

Notes: This table reproduces table 2 from Leigh (2010). It shows the percentage amounts of state supplements to the federal EITC. These supplements are the source of variation we use for identification of welfare effects.

TABLE II

Effect of EITC expansion on average wages and labor supply

	All adults	**High school dropouts**	**High school diploma only**	**College graduates**
	dependent variable: Log real hourly wage			
Log maximum EITC	-0.121	-0.488	-0.221	0.008
	[0.064]	[0.128]	[0.073]	[0.056]
Fraction EITC-eligible	9%	25%	12%	3%
	dependent variable: whether employed			
Log maximum EITC	0.033	0.09	0.042	0.008
	[0.012]	[0.046]	[0.019]	[0.022]
Fraction EITC-eligible	14%	34%	17%	4%
	dependent variable: Log hours per week			
Log maximum EITC	0.037	0.042	0.011	0.095
	[0.019]	[0.040]	[0.014]	[0.027]
Fraction EITC-eligible	9%	25%	12%	3%

Notes: Estimates from Table 4 and 5 of Leigh (2010), for workers with and without children.

6.2. *Results*

6.3. *Discussion*

We conclude this section by discussing, first, the relationship of our approach to alternative measures of the individual-level welfare impact of the EITC. We then mention some potential shortcomings of our analysis.

Various measures of the individual-level welfare impact of the EITC expansion seem possible:

1. **Evaluation based on income:**
   A perspective which is interested in the realized incomes of the poor might consider the EITC to be desirable both (i) because it provides transfer income, and (ii) because increased labor supply implies increased earnings. In our notation, both $-\dot{t} > 0$ and $\dot{l} \cdot n > 0$.
2. **Evaluation based on utility, assuming fixed wages:**
   A perspective in the tradition of the Mirrlees model of income taxation might consider the EITC as less desirable than transfers to the unemployed poor (i) because it does not reach those most in need, and (ii) the increase in labor supply has a zero first order effect on private welfare, but a negative effect on public revenues. The induced increases in labor supply cause "deadweight loss." This is because marginal taxes are negative for low incomes due to the EITC.
   In our notation, $\dot{l} \cdot n$ figures in the expression for $\dot{y}$ given in lemma 1, but not in the expression for $\dot{e}$.
3. **Evaluation based on utility, with endogenous wages:**
   Work in the Mirrleesian tradition has assumed wages to be exogenously given and unaffected by policy changes. This contrasts with much of the literature in labor economics, as noted by Rothstein (2010) and Leigh (2010). Our analysis considers endogenous wages to be an important channel for the distributional welfare impact of the EITC expansion, while otherwise adopting the utility-based perspective of optimal tax theory.

In the terminology introduced in our discussion of lemma 1, we expect the expansion of the EITC to have a positive mechanical effect for the working poor, a positive labor supply effect for those eligible, and a negative wage effect for both those eligible, and for those ineligible but competing in the same labor markets. The three approaches to evaluating the EITC expansion correspond to

1. mechanical + wage + labor supply effect, $-\dot{t} + l \cdot \dot{w} \cdot (1 - \partial_z t) + \dot{l} \cdot n$
2. mechanical effect, $-\dot{t}$
3. mechanical + wage effect, $-\dot{t} + l \cdot \dot{w} \cdot (1 - \partial_z t)$.

Case 3, which corresponds to our analysis, makes the EITC look worse than both the income based, and the utility based, fixed-wage evaluation.

Our analysis (like any "sufficient-statistic" type analysis) has an important limitation – we do not account for the welfare effects of involuntary unemployment. As mentioned before, the result of lemma 1 relies on the assumption that only monetary constraints of individuals are affected by the policy change. Non-monetary constraints of various kinds can be allowed for, but they may not be affected by the policy change. The non-monetary constraint which is the biggest concern in the context of labor markets is involuntary unemployment. Policies that shift labor supply or demand likely not only impact the wage distribution but also the degree of involuntary unemployment. Empirical results such as the ones discussed in this section should be interpreted as only capturing welfare effects mediated through transfers as well as wages. Effects through involuntary unemployment, in our context, are likely to make the EITC look less desirable.

7. CONCLUSION

APPENDIX A: PROOFS

Proof of lemma 1 The expression for $\dot{y}$ follows by simple differentiation of

$$y = w \cdot l - t(w \cdot l) + y_0.$$

The expression for $\dot{e} = \dot{v}/\partial_{y_0} v$ follows from a variant of Roy's identity (cf. Mas-Colell et al., 1995, p73); see also Chetty (2009): Assuming w.l.o.g. an interior solution, we can obtain it using

1. $\dot{v} = \partial_{(c,l)} u \cdot (\dot{c}, \dot{l})$ and $\partial_{y_0} v = \partial_{(c,l)} u \cdot (\partial_{y_0} c, \partial_{y_0} l)$,
2. the individual's first order condition $\partial_{(c,l)} u = \lambda \cdot (p, -n)$ for some Lagrange multiplier λ, and
3. Walras' law $(c \cdot p = l \cdot w - t(l \cdot w) + y_0))$, which implies
$$(p, -n) \cdot (\dot{c}, \dot{l}) = l \cdot \dot{w} \cdot (1 - \partial_z t) - \dot{t} + \dot{y_0} - c \cdot \dot{p},$$
and $(p, -n) \cdot (\partial_{y_0} c, \partial_{y_0} l) = 1$.

□

Proof of theorem 1: Let

$$\begin{aligned} A(\alpha) &:= E[a(x(\alpha, \epsilon))|\alpha] = \int a(x(\alpha, \epsilon)) dP(\epsilon) \\ &= \int a(x) f(x|\alpha) dx. \end{aligned}$$

for any differentiable a with bounded support. Corresponding to the last two representations of $A(\alpha)$, there are two representations for $\dot{A}(\alpha)$. Using the first representation and partial integration, we get

$$\begin{aligned} \dot{A}(\alpha) &= E[\partial_x a \cdot \dot{x}|\alpha] = \sum_{j=1}^{k} E[\partial_{x^j} a \cdot \dot{x}^j|\alpha] \\ &= \sum_{j=1}^{k} E[\partial_{x^j} a \cdot h^j/f|\alpha] = \sum_{j=1}^{k} \int \partial_{x^j} a \cdot h^j dx \\ &= -\int a \cdot \sum_{j=1}^{k} \partial_{x^j} h^j dx = -\int a \cdot (\nabla \cdot h) dx. \end{aligned}$$

We can alternatively write

$$\begin{aligned} \dot{A}(\alpha) &= \partial_\alpha \int a(x) f(x|\alpha) dx \\ &= \int a(x) \dot{f}(x|\alpha) dx. \end{aligned}$$

As these equations hold for any differentiable g with bounded support and h is continuous by assumption, we get $\dot{f} = -\nabla \cdot h$. □

Proof of theorem 2:

1. h satisfies $\dot{f} = -\nabla \cdot h$ if and only if it is in the identified set:
The "if" part follows from theorem 1. To show the "only if" part, taking h as given we need to construct a distribution of ϵ and structural functions x consistent with h, the

observed data distribution, and assumption 2: Let $\epsilon = x(0,\epsilon)$, and thus $f(\epsilon) = f(x|\alpha = 0)$, and let $x(.,\epsilon)$ be a solution to the ordinary differential equation

$$\dot{x} = h(x,\alpha),\ x(0,\epsilon) = \epsilon.$$

Such a solution exists by Peano's theorem. It is easy to check that this solution satisfies all required conditions.

2. h^0 satisfies $\dot{f} = -\nabla \cdot h^0$:
   Consider the model $x^j(\alpha,\epsilon) = Q^{x^j|v^i,\ldots,v^{j-1},\alpha}(v^j|v^1,\ldots,v^{j-1},\alpha)$ where $\epsilon = (v^1,\ldots,v^k)$. Then this model implies $E[\dot{x}|x,\alpha]\cdot f(x) = h^0(x)$, where h^0 is defined as in the statement of the theorem. This model is furthermore consistent with the observed data distribution, and thus in particular satisfies $\dot{f} = -\nabla \cdot h^0$ by theorem 1.
3. h satisfies $\dot{f} = -\nabla \cdot h$ if and only if $h \in h^0 + \mathscr{H}$:
   For any h in $h^0 + \mathscr{H}$, we have $\nabla \cdot h = \nabla \cdot h^0 + \nabla \cdot \tilde{h} = -\dot{f} + 0$.
   Reversely, for any h such that $\dot{f} = -\nabla \cdot h$, let $\tilde{h} := h - h^0$. Then $\tilde{h} \in \mathscr{H}$.

□

Proof of theorem 3:

1. $k = 1$: In this case, $\nabla \cdot h = \partial_x h = 0$. Since h has its support contained in $\mathbf{X}$, integration immediately yields $h \equiv 0$.
2. $k = 2$: This result is a special case of Poincaré's Lemma, which states that on convex domains differential forms are closed if and only if they are exact; cf. Theorem 10.39 in Rudin (1991). Apply this lemma to

   $$\omega = h^1 dx^2 - h^2 dx^1.$$

   Then

   $$d\omega = (\partial_{x^1} h^1 + \partial_{x^2} h^2) dx^1 \wedge dx^2 = 0$$

   if and only if

   $$\omega = dH = \partial H/\partial x^1 dx^1 + \partial H/\partial x^2 dx^2$$

   for some H.
3. This follows again from Poincaré's Lemma, applied to

   $$\omega = h^1 dx^2 \wedge dx^3 + h^2 dx^3 \wedge dx^1 + h^3 dx^1 \wedge dx^2$$

   and

   $$\lambda = \sum_j G^j dx^j.$$

□

Proof of theorem 4:

1. h^0 is consistent with this assumption:
   Consider again the model $x^j(\alpha,\epsilon) = Q^{x^j|v^i,\ldots,v^{j-1},\alpha}(v^j|v^1,\ldots,v^{j-1},\alpha)$ where $\epsilon = (v^1,\ldots,v^k)$, as in the proof of theorem 2. Then this model implies $\partial_{x^j} E[\dot{x}^i|x,\alpha] = 0$ for $j > i$. This model is furthermore consistent with the observed data distribution and satisfies $E[\dot{x}|x,\alpha]\cdot f(x) = h^0(x)$.

2. The only $\tilde{h} \in \mathscr{H}$ consistent with this assumption is $\tilde{h} \equiv 0$:
   As we have already shown h^0 to be consistent with this assumption, it is enough to show that $\nabla \cdot \tilde{h} = 0$ implies $\tilde{h} \equiv 0$ if $\tilde{h}$ is consistent with this assumption. We proceed by induction in j.
   Consider the model where we only observe $x^1, \ldots, x^j$, and define $\tilde{h}$ accordingly. Suppose we have shown $(\tilde{h}^1, \ldots, \tilde{h}^{j-1}) = (0, \ldots, 0)$. Applying theorem 1 to the j dimensional model immediately implies $\partial_{x^j} \tilde{h}^j = 0$. Integrating with respect to x^j, and using the fact that the support of $\tilde{h}$ is contained in the support $\mathbf{X}$ of x implies $\tilde{h}^j \equiv 0$.
   Equation (19) implies
   $$E[\dot{x}^i | x^1, \ldots, x^k, \alpha] = E[\dot{x}^i | x^1, \ldots, x^j, \alpha]$$
   for $j \geq i$. As a consequence, $\tilde{h}^i = 0$ in the j dimensional model immediately implies $\tilde{h}^i = 0$ in the $j+1$ dimensional model. The claim now follows by induction.

□

Proof of theorem 5:
We proceed by induction in j. Assume the statements of the theorem hold for $j-1$. For $j = 0$ the claims hold trivially. Equation (21) immediately implies that $x^j(\alpha, \epsilon) \leq x^j(\alpha, \epsilon')$ if and only if $\epsilon^j \leq \epsilon^{j\prime}$ when $(\epsilon^1, \ldots, \epsilon^{j-1}) = (\epsilon^{1\prime}, \ldots, \epsilon^{j-1\prime})$. By the induction assumption we get
$$\begin{aligned} v^j &= F(x^j | x^1, \ldots, x^{j-1}, \alpha) = F(x^j | \epsilon^1, \ldots, \epsilon^{j-1}, \alpha) \\ &= F(\epsilon^j | \epsilon^1, \ldots, \epsilon^{j-1}, \alpha) = F(\epsilon^j | \epsilon^1, \ldots, \epsilon^{j-1}) \end{aligned}$$
independent of α. The claims regarding v follow. As for the second claim, we get
$$\begin{aligned} Q^{x^j}(v^j | v^1, \ldots, v^{j-1}, \alpha') &= Q^{x^j}(v^j | \epsilon^1, \ldots, \epsilon^{j-1}, \alpha') \\ &= Q^{x^j}(F(x^j | x^1, \ldots, x^{j-1}, \alpha') | \epsilon^1, \ldots, \epsilon^{j-1}, \alpha') \\ &= x^j(\alpha', \epsilon). \end{aligned}$$

□

Proof of proposition 1:
This is an immediate consequence of theorem 4. □

Proof of proposition 2:
This is an immediate consequence of proposition 1 in Kasy (2014). □

Proof of proposition 3:
This is again an immediate consequence of theorem 4. □

Proof of theorem 6:

1. Assumption 3.1 implies that there exists a linear map $d\theta$ such that $\dot{\theta} = d\theta(\dot{y}(.))$. This map is furthermore continuous with respect to the L^2 norm of $\dot{y}$. Riesz' representation theorem then implies existence of $\omega^\theta \in L^2$, such that $\dot{\theta} = E[\omega^\theta \cdot \dot{y}]$.

   For SWF, assumption 3.1 similarly implies existence of $\tilde{\omega}$ such that $S\dot{W}F = E[\tilde{\omega} \cdot \dot{v}]$. We can renormalize and define
   $$\omega^{SWF} := \tilde{\omega} \cdot (-\partial_{t_0} v), \tag{50}$$

which immediately implies $\dot{SWF} = E[\omega^{SWF} \cdot \dot{e}]$.

2. Assumption 3.2 immediately implies that

$$\theta(\alpha) = E[IF(y^\alpha)] + o(\|P_{y^\alpha} - P_{y^0}\|),$$

where $\|.\|$ is the appropriate norm on the space of probability distributions; see also (cf. van der Vaart, 2000, p291ff). The claim follows.

3. By 2, we have $\theta^\alpha = \theta^0 + E[IF^\theta(y^\alpha)] + o(\alpha)$. Differentiating this expression yields $\dot{\theta} = E\left[\partial_y IF(y) \cdot \dot{y}\right]$. Comparing this expression to the first representation of $\dot{\theta}$ yields

$$E[\omega^\theta \cdot \dot{y}] = E\left[\partial_y IF(y) \cdot \dot{y}\right].$$

As this equation holds for any direction of change $\dot{y}(.)$, $\omega^\theta = \partial_y IF(y)$ follows.

□

Proof of theorem 7:
The equations $\dot{e} = \dot{\tilde{y}} = \dot{y} - b$ follow from simple differentiation and lemma 1.

1. $\dot{SWF} = E[\omega \cdot \dot{\tilde{y}}]$ follows from theorem 6 and the identity $\dot{e} = \dot{\tilde{y}}$. $E[\omega \cdot \dot{e}] = E[\omega \cdot \gamma]$ holds by the law of iterated expectations, since ω is a function of y by assumption:

$$E[\omega \cdot \dot{e}] = E[E[\omega \cdot \dot{e}|y, W]] = E[\omega \cdot E[\omega \cdot \dot{e}|y, W]] = E[\omega \cdot \gamma].$$

2. Note that $\dot{SWF} = E[\omega \cdot \dot{\tilde{y}}]$. $\dot{SWF} = \partial_\alpha \theta\left(P_{\tilde{y}^\alpha}\right)$ follows by analogy to $\dot{\theta} = E[\omega^\theta \cdot \dot{y}] = \partial_\alpha \theta\left(P_{y^\alpha}\right)$.
3. By theorem 6, $\dot{\theta} = \partial_\alpha E\left[IF(y^\alpha)\right]$. Apply this result to $\tilde{y}$ instead of y.
4. $\dot{\theta} - \dot{SWF} = E[\omega \cdot b]$ holds since $\dot{y} - \dot{e} = b$. The other claims follow analogously to item 2 and 3 of this theorem, once we note that $\dot{\check{y}} = b$.

□

APPENDIX B: MONTE CARLO SIMULATIONS

REFERENCES

Acemoğlu, D. and J. A. Robinson (2013): "Economics versus Politics: Pitfalls of Policy Advice," *Journal of Economic Perspectives*.

Altonji, J. and R. Matzkin (2005): "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica*, 73, 1053–1102.

Athey, S. and G. W. Imbens (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74, 431–497.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008): "Trends in US wage inequality: Revising the revisionists," *The Review of Economics and Statistics*, 90, 300–323.

Card, D. (2009): "Immigration and Inequality," *The American Economic Review*, 99, 1–21.

Chamberlain, G. (1984): "Panel Data," *Handbook of Econometrics*, 2.

Chamberlain, G. and G. W. Imbens (2003): "Nonparametric applications of Bayesian inference," *Journal of Business & Economic Statistics*, 21, 12–18.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013): "Inference on counterfactual distributions," *Econometrica*, 81, 2205–2269.

Chetty, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–488.

Chetty, R., J. N. Friedman, and E. Saez (2013): "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, 103, 2683–2721.

DEATON, A. (1989): "Rice prices and income distribution in Thailand: a non-parametric analysis," *The Economic Journal*, 1–37.

DINARDO, J., N. FORTIN, AND T. LEMIEUX (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64, 1001–1044.

FIRPO, S., N. FORTIN, AND T. LEMIEUX (2009): "Unconditional quantile regressions," *Econometrica*, 77, 953–973.

GRAHAM, B. S. AND J. L. POWELL (2012): "Identification and Estimation of Average Partial Effects in ìIrregularî Correlated Random Coefficient Panel Data Models," *Econometrica*, 80, 2105–2152.

HAHN, J. AND G. RIDDER (2013): "Asymptotic variance of semiparametric estimators with generated regressors," *Econometrica*, 81, 315–340.

HENDREN, N. (2013): "The Policy Elasticity," .

HODERLEIN, S. AND E. MAMMEN (2007): "Identification of marginal effects in nonseparable models without monotonicity," *Econometrica*, 75, 1513–1518.

——— (2009): "Identification and estimation of local average derivatives in non-separable models without monotonicity," *The Econometrics Journal*, 12, 1–25.

IMBENS, G. W. AND W. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

KASY, M. (2014): "Instrumental variables with unrestricted heterogeneity and continuous treatment," *Review of Economic Studies*, forthcoming.

KOENKER, R. (2005): *Quantile Regresssion*, Wiley Online Library.

LEIGH, A. (2010): "Who benefits from the earned income tax credit? Incidence among recipients, coworkers and firms," *The BE Journal of Economic Analysis & Policy*, 10.

MAS-COLELL, A., M. WHINSTON, AND J. GREEN (1995): *Microeconomic theory*, Oxford university press.

MATZKIN, R. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71, 1339–1375.

MEYER, B. D. AND D. T. ROSENBAUM (2001): "Welfare, the earned income tax credit, and the labor supply of single mothers," *The Quarterly Journal of Economics*, 116, 1063–1114.

MIRRLEES, J. (1971): "An exploration in the theory of optimum income taxation," *The Review of Economic Studies*, 175–208.

NEWEY, W. K. (1994a): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253.

——— (1994b): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253.

OAXACA, R. (1973): "Male-female wage differentials in urban labor markets," *International economic review*, 14, 693–709.

ROTHE, C. (2010): "Nonparametric estimation of distributional policy effects," *Journal of Econometrics*, 155, 56–70.

ROTHSTEIN, J. (2010): "Is the EITC as Good as an NIT? Conditional Cash Transfers and Tax Incidence," *American Economic Journal: Economic Policy*, 177–208.

RUBIN, D. B. (1981): "The bayesian bootstrap," *The annals of statistics*, 130–134.

RUDIN, W. (1991): *Principles of mathematical analysis*, McGraw-Hill.

SAEZ, E. (2001): "Using elasticities to derive optimal income tax rates," *The Review of Economic Studies*, 68, 205–229.

SAEZ, E. AND S. STANTCHEVA (2013): "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," Tech. rep., National Bureau of Economic Research.

SAMUELSON, P. (1947): "Foundations of Economic Analysis," .

VAN DER VAART, A. (2000): *Asymptotic statistics*, Cambridge University Press.

WRIGHT, E. O. (2005): *Approaches to class analysis*, Cambridge Univ Pr.