# Who wins, who loses? Tools for distributional policy evaluation*

Maximilian Kasy†

November 4, 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 technical 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 expansion of the Earned Income Tax Credit (EITC), using variation in state supplements to the federal EITC and the CPS-IPUMS data. We find large negative effects of depressed wages as a consequence of increased labor supply. The estimated effects are largest for those earning around 20.000 US$ per year, as well as for high school dropouts.

JEL Codes: C14, D3, H22, J3
Keywords: Incidence, distributional decomposition, nonparametric identification

*I'd like to thank David Autor, Bence Bardoczy, Gary Chamberlain, Raj Chetty, Ellora Derenoncourt, Arindrajit Dube, Pirmin Fessler, Sergio Firpo, James Heckman, Nathaniel Hendren, Stefan Hoderlein, Larry Katz, Susanne Kimm, Nathan Lane, Suresh Naidu, Jesse Rothstein, and Emmanuel Saez for valuable discussions and comments. László Sándor provided valuable research assistance. This work was supported by NSF grant SES-1354144 "Statistical decisions and policy choice."

†Assistant professor, Department of Economics, Harvard University. Address: Littauer Center 200, 1805 Cambridge Street, Cambridge, MA 02138. email: maximiliankasy@fas.harvard.edu.

# I 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 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, in particular the impact of changing wages, but also of changing prices, taxes, and transfers. 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 (almost) unrestricted heterogeneity across individuals in terms of preferences and in terms of policy impacts on wages, labor supply, etc.[1] Within this framework, we devise procedures to answer various questions regarding the distributional impact of marginal 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 im-

1. This point is important and will be discussed in detail below. We do not restrict the dimension of heterogeneity in any way, in contrast to "structural" approaches, but we do need to impose some exclusion restrictions to achieve point identification.

pact of the policy change under consideration to each individual. Based on this imputed impact, we can estimate a number of objects of interest: We can plot expected welfare impacts given initial earnings and demographic covariates. We can estimate sets of winners and losers and their characteristics, such as means of covariates within each group. Given a choice of welfare weights, finally, we can estimate aggregate welfare impacts.

The central econometric difficulty is the first step, imputation of money-metric expected welfare impacts to each individual. Welfare impacts have components of the form $\dot{w} \cdot l$ (change in wage times baseline labor supply), and we need to estimate expected welfare impacts *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 can reframe the problem of estimating $E[\dot{w} \cdot l | w \cdot l, \alpha]$ as a special case of the problem of identifying the expected causal impact $\dot{x}$ of a policy change on the vector $x$, conditional on initial $x$ and policy level $\alpha$, $E[\dot{x}|x, \alpha]$. The special case we are interested in corresponds to considering $x = (w, l)$. We provide conditions involving exclusion restrictions 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 for the sets of winners and losers, for instance covariate means and population shares.

The results discussed so far concern the problem of inferring individual welfare effects from data. A second set of issues arises in normative policy analysis when we want to aggregate, that is, infer social welfare effects from individual welfare effects. Issues of aggregation arise both in optimal tax theory and in the distributional decomposition literature. We 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.

The rest of this paper is structured as follows. We conclude this introduction by reviewing some related literature. Section II presents our assumptions and objects of interest and characterizes the effect of policy changes on individual welfare. Our main theoretical contributions can be found in sections III and IV. Section III presents our results on identification; section IIIA provides results on the identification of marginal causal effects conditional on outcomes, and section IIIB discusses the use of instruments and controls as well as of panel data for identification of welfare effects in a nonparametric setting. Section IV discusses aggregation and the relation between distributional decompositions and social welfare effects. Section V proposes estimators and inference procedures based on these identification results. Section VI 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, following the analysis of Leigh (2010). Section VII concludes. Appendix A contains all proofs. Tables and figures are to be found in appendix B.

## *IA Related literature*

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 optimal taxation literature in public finance usually considers utilitarian social welfare functions,[2] which were introduced by Samuelson (1947); the

2. 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.

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); a standard reference is DiNardo et al. (1996). Recent contributions to the econometrics of such decompositions are Firpo et al. (2009), 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. Disaggregated impacts allow us to study questions of political economy, following the research agenda proposed by Acemoğlu and Robinson (2013). Dis-aggregated impacts 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).

Abbring and Heckman (2007) also review issues closely related to those discussed in this paper, in particular the distribution of treatment effects, and general equilibrium effects of policy changes. Our analysis differs from Abbring and Heckman (2007) as follows: (i) We are only interested in the conditional expectation of marginal causal effects of a continuous treatment, conditional on outcomes, rather than the full distribution of treatment effects of a discrete treatment. (ii) We analyze welfare effects rather than effects on observed outcomes. (iii) We only propose methods for the ex-post evaluation of realized wage and price changes, rather than predicting the general equilibrium effects of counterfactual policies.

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 identification, 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).

Our results might have applications in the estimation of consumer demand, which similarly features multidimensional outcomes, namely households' consumption bundles. In that case, it might be of interest to exploit restrictions imposed by revealed preference theory both for the purpose of improving finite sample performance and for the purpose of out-of-sample extrapolation, as in Blundell et al. (2014). Such restrictions are not available in our setting, since the main causal effect of interest will be on wages, and this causal effect is not restricted by general equilibrium theory in any way.

# II 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 introduce several empirical objects of interest which we will analyze. The section concludes with a characterization of the effect of policy changes on individual welfare, using standard envelope condition arguments. 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.

## *IIA 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 $\alpha$ are denoted by superscripts, so that $w^\alpha$ is for instance the potential wage of an individual under policy $\alpha$. We use the potential outcome notation as a short-hand for structural functions, as in

$$w^\alpha = w(\alpha, \epsilon),$$

where both the function $w(.,.)$ and individual heterogeneity $\epsilon$ are assumed to be invariant as policy $\alpha$ changes, and where $\epsilon$ is of unrestricted dimension.[3]

3. Potential outcomes and structural functions in this paper are "reduced form" objects in the sense that they incorporate the impact of any general equilibrium effects of policy changes.

Letters without superscripts denote *random variables* in this paper, so that $w$ is the wage of an individual as determined by the *realized* policy $\alpha$. When we consider a sample of observations $i = 1, \ldots, N$ in section V (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 $\partial$ with a subscript, so that $\partial_w$ is the derivative with respect to $w$. *Derivatives of potential outcomes with respect to* $\alpha$ will be denoted by a superscript dot, so that

$$\dot{w} := \partial_\alpha w^\alpha = \partial_\alpha w(\alpha, \epsilon)$$

denotes the marginal effect of a policy change on the wage of a given individual. Our identification results in section III 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 $h : \mathbb{R}^k \to \mathbb{R}^k$.[4]

*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 $l$.

## *IIB Individual problem*

We discuss distributional policy evaluation in the context of labor markets, the wage distribution, and taxes on earnings.[5] All variables depend on the policy $\alpha$, as well as on unobserved individual heterogeneity $\epsilon$, 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

4. As we will discuss below, the divergence measures the net out-flow from a point $x$, when $h$ describes a flow.

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

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 $c \in \mathbb{R}^{d_c}$ and $l \in \mathbb{R}$ 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 $\upsilon$.
- 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 $\alpha$, and as a consequence so do $c$, $l$, and $\upsilon$.
- For all individuals, $u$ is differentiable, increasing in the components of $c$ and decreasing in $l$, quasiconcave, and does not depend on $\alpha$.

**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 markets with heterogenous goods other than the labor market, 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.

- Even though our methods generalize to more general models allowing for additional constraints on individual choices, they crucially rely on 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 $w^\alpha$, counterfactual consumption $c^\alpha$ etc., as well as a realized policy $\alpha$ and corresponding realized wage $w$, realized consumption $c$ etc.

- We have assumed that policies are indexed by a one-dimensional parameter $\alpha$. This is best thought of as indexing a path within the space of feasible policies, which is in general of much larger dimension. We will be concerned with ex-post evaluation of actually implemented policy changes, and the path indexed by $\alpha$ corresponds to these policy changes.

- 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. This linearized version will allow us to restate some of our results in a more easily interpretable manner. 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 $u$, and if taxes $t(.)$ are progressive (convex), this problem in turn has the same solution as[6]

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

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

## *IIC 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). Since we are considering marginal policy changes, it is also equal to the equivalent variation, the compensating variation, and the change in consumer surplus corresponding to these policy changes.

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 $\alpha$, we can define our main objects of interest.

Definition 1 (Objects of interest – utility).

1. Expected conditional policy effect on welfare:

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

6. Quasiconcavity ensures that the optimum in the linearized budget set is the same as in the smaller original budget set.

2. Sets of winners and losers:

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

3. Policy effect on social welfare:

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

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

**Remarks:**

- The expected conditional policy effect $\gamma$ 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 $\gamma$ or objects such as $E[\gamma|y]$, the expected welfare impact given initial income. Figures VI, VII, and IX provide examples of such plots in the context of the application discussed in section VI. This allows to immediately visually assess the welfare impact of a policy change across the income distribution.

- The conditional expectation defining $\gamma$ conditions on income, covariates, and the policy level. This expectation averages over the population distribution of potential outcomes given observables. The expectation defining $\gamma$ corresponds to subjective expectations if (i) subjective expectations are rational, and (ii) individuals have no (unobserved) information which is predictive of welfare effects beyond the information captured by $l, w, W$.

  Under the stronger condition that the dimensionality of unobservables is no larger than the dimensionality of observables, we will show that $\gamma = \dot{e}$. In this case our approach will recover actual rather than expected welfare effects. Such a dimensionality restriction is imposed by most structural estimation approaches, see for instance Matzkin (2003).

  To the extent that $\gamma$ corresponds to the subjective expectations of an individual assessing whether to advocate or oppose a proposed policy change, the sign of $\gamma$ should determine her decision. The sets of winners and losers $\mathscr{W}$ and $\mathscr{L}$ are then 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: 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 $\omega$ 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 $\omega$ implied by $SWF$ (and discussed in section IV 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. Under either condition $E[\omega \cdot \dot{e}|y, W, \alpha] = E[\omega|y, W, \alpha] \cdot \gamma(y, W)$.

  An example where these conditions might be violated is as follows: (i) Conditional on earnings, we assign higher welfare weights to sick people than to healthy people. (ii) The policy under consideration redistributes to sick people. (iii) Health status is not included among the covariates $W$.

- 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 IV. 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 IV below considers these.

### IID Welfare effects of marginal policy changes on individuals

We consider the effects of a marginal change in $\alpha$ 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 III discusses identification of expected individual effects. Section IV then considers aggregate effects, on social welfare $SWF$ as well as on statistics of the income distribution $\theta$.

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 $\alpha$. The effect of such a marginal change on net income $y$ and on welfare (money metric utility) $e$ equals

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

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

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 VI 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 and wages, their 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. This contrasts with the "sufficient statistic" literature reviewed in Chetty (2009). Sufficient statistic arguments rely on either fixed prices or known price-responses.[7] In particular, we *do* need

7. I thank Nathaniel Hendren for discussions on this point.

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.

# III Identification

In this section, we discuss identification of individual expected welfare effects $\gamma(y, W) = E[\dot{e}|y, W, \alpha]$. Once these are identified, we can estimate and visually plot them, use them to characterize sets of winners and losers, or aggregate them to obtain social welfare effects.

From the point of view of identification, the interesting (and difficult) component of $\gamma$ is the welfare effect of wage changes $\dot{w}$ caused by a policy change. To focus on this interesting (and difficult) component, consider a simplified version of the setting of assumption 1, where we assume $\dot{p} = \dot{t} = \dot{y}_0 = \partial_z t = 0$, so that welfare effects are driven solely by changing wages; we ignore welfare effects of changing prices, taxes, or unearned income for now, and we ignore attenuation of wage effects through redistributive taxation. Under these assumptions equation (8) of lemma 1 implies that $\dot{e} = l \cdot \dot{w}$. Assume further for the moment that there are no covariates $W$. The conditional expected welfare effect $\gamma(y, W)$ is then equal to the wage effect,

$$\gamma(y) = E[l \cdot \dot{w} | l \cdot w, \alpha].$$

The latter is identified if $E[(\dot{l}, \dot{w})|l, w, \alpha]$ is identified. Let $x = (l, w)$. Our problem is to identify

$$E[\dot{x}|x, \alpha],$$

for a vector of endogenous outcomes $x = x(\alpha, \epsilon)$; that is to identify marginal causal effects *conditional on a vector of endogenous outcomes.* This is the problem that we address in this section.

In section IIIA, we assume that we have random ("experimental") variation of $\alpha$. Given random variation of $\alpha$, the marginal distribution of potential outcomes $x^\alpha$ is identified for a range of $\alpha$-values, and we can focus on the question of what these marginal distributions tell us about conditional causal effects. We provide conditions under which

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

where $v^j$ is the control function $v^j = F(x^j|x^1, \ldots, x^{j-1}, \alpha)$, $F$ denotes the conditional cumulative distribution function, and $Q$ denotes the conditional quantile function. The conditions required for this identification result to hold include some restrictions on the heterogeneity of conditional causal effects; this will be discussed in detail below. In section IIIB, we then generalize to quasi-experimental settings, assuming the availability of suitable controls, exogenous instruments, or panel data.

In the special case where the dimension of the outcome $x$ is one, our identification results reduce to those of Hoderlein and Mammen (2007). In this paper, we are mainly concerned with the case where the dimension of $x$ is larger than one.

## *IIIA Effects conditional on outcomes*

Suppose the distribution of $x^\alpha$ is known for a continuum of values of $\alpha$ around 0. This is the case in an experimental setting, where $\alpha$ is independent of unobserved heterogeneity $\epsilon$ and the support of $\alpha$ contains a neighborhood $(-\delta, \delta)$ of 0. The following series of results explores identification of $E[\dot{x}|x, \alpha]$ in this case.

Assumption 2 (Abstract setup).

- $x = x(\alpha, \epsilon)$
- $x \in \mathbb{R}^k$, $\alpha \in \mathbb{R}$, $\epsilon$ 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 $\alpha$.
- $x(\alpha, \epsilon)$ is differentiable in $\alpha$.
- $E[\dot{x}|x, \alpha] \cdot f(x|\alpha)$ is continuously differentiable in $x$.
- The support of $x$ given $\alpha$ is contained in a compact and convex set $\mathbf{X}$ which is independent of $\alpha$.

Assumption 2 is stated in terms of structural functions $x(\alpha, \epsilon)$. It could be stated equivalently, though less transparently, in terms of potential outcomes, requiring $x^\alpha \perp \alpha$ etc.

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

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

$$h(x, \alpha) := g(x, \alpha) \cdot f(x|\alpha). \tag{12}$$

We denote the divergence of $h$ by

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

The "flow" $g$ is identified (on the support of $f$) if and only if the "flow density" $h$ is identified, since $h = g \cdot f$ and the density $f$ is known.

We will now develop a series of results characterizing the problem of identifying $g$ (equivalently, $h$) based on knowledge of $f$. 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. 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 knowledge of $f$ identifies the divergence of $h$ under assumption 2.

Theorem 1. Suppose assumption 2 holds. Then

$$\dot{f} = -\nabla \cdot h. \tag{13}$$

[Figure I here]

**Remark:**

- Figure I provides some intuition for this result: Consider the density of observations in the shaded square. This density changes, as $\alpha$ 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 $-\sum_{j=1}^{k} \partial_{x^j} h^j \cdot dx^j$. The divergence $\nabla \cdot h$ thus measures the net outflow at a point corresponding to a flow density $h$. If the density does not change, or equivalently the net outflow equals 0, then the divergence of $h$ has to be 0.

- The setting of assumption 2 has various analogies in physics, most notably in fluid dynamics. We can think of $\alpha$ as time, $\epsilon$ indexing individual particles, 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 $\alpha$. As shown in theorem 1, the change of this density over time is given by the divergence of the flow density (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.

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 $\widetilde{h}$ between this function and any other function $h$ in the identified set is in the set $\{\widetilde{h} : \nabla \cdot \widetilde{h} \equiv 0\}$.

THEOREM 2. Suppose assumption 2 holds.

Let $v^j$ be the random variable $v^j = F(x^j|x^1, \dots, x^{j-1}, \alpha)$,[8] define

$$h^{0j}(x,\alpha) = f(x|\alpha) \cdot \partial_\alpha Q(v^j|v^1, \dots, v^{j-1}, \alpha), \tag{14}$$

and let

$$\mathscr{H} = \{\widetilde{h} : \ \nabla \cdot \widetilde{h} \equiv 0, \ \widetilde{h}(x,\alpha) = 0 \text{ for } x \notin \mathbf{X}\}. \tag{15}$$

Then the identified set for $h$ is given by

$$h^0 + \mathscr{H}. \tag{16}$$

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 $\mathscr{H} = \{\widetilde{h} : \ \nabla \cdot \widetilde{h} \equiv 0\}$. 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.[9]

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 $\widetilde{h}$ in the kernel are exactly those functions which can be written as the gradient of some function $H$, rotated by 90 degrees. $\widetilde{h}$ is thus a vector field pointing along the lines of constant height of $H$. Figure II illustrates.

[Figure II here]

8. $v$ is a random variable since it is a function of $x$, which is a random variable.

9. Similar results can be stated for higher dimensions, but require increasingly cumbersome notation.

Theorem 3. Suppose assumption 2 holds.

1. Suppose $k = 1$. Then[10]
$$\mathscr{H} = \{\widetilde{h} \equiv 0\}. \tag{17}$$

2. Suppose $k = 2$. Then
$$\begin{aligned}\mathscr{H} = \{\widetilde{h} : \widetilde{h} = A \cdot \nabla H,& \\ H:\ \mathbb{R}^2 \to \mathbb{R},\ H(x, \alpha) = 0 \text{ for } x \notin \mathbf{X}\}.&\end{aligned} \tag{18}$$
where
$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

3. Suppose $k = 3$. Then
$$\begin{aligned}\mathscr{H} = \{\widetilde{h} : \widetilde{h} = \nabla \times G,& \\ G:\ \mathbb{R}^3 \to \mathbb{R}^3,\ G(x, \alpha) = 0 \text{ for } x \notin \mathbf{X}\}.&\end{aligned} \tag{19}$$
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 $h$ absent any further identifying assumptions, that is 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
$$\partial_{x^j} E[\dot{x}^{j'}|x, \alpha] = 0 \text{ for } j > j'. \tag{20}$$

10. 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).

Then $g$ and $h$ are point identified, and

$$g(x, \alpha) = \partial_\alpha Q(v^j | v^1, \ldots, v^{j-1}, \alpha), \tag{21}$$

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.

There are no over-identifying restrictions implied by equation (20).

**Remark:**

- It is useful to discuss the triangularity assumption of equation (20) and its implications in the context of our labor market setting. In that setting, let $x = (w, l)$. Equation (20) can then be rewritten as

  $$\partial_l E[\dot{w} | w, l, \alpha] = 0,$$

  that is, the average effect of a policy change on wages, conditional on labor supply and wages, does not depend on labor supply. If that is the case, then theorem 4 tells us that

  $$E[\dot{w} | l, w, \alpha] = \partial_\alpha Q(w | \alpha).$$

  Recall that $E[\dot{w} | w, \alpha] = \partial_\alpha Q(w | \alpha)$ holds without any exclusion restrictions – this was the result of Hoderlein and Mammen (2007).

- If, alternatively, we set $x = (l, w)$, so that the order of $w$ and $l$ is reversed, then equation (20) can be rewritten as

  $$\partial_w E[\dot{l} | w, l, \alpha] = 0,$$

  that is, the average effect of a policy change on labor supply, conditional on labor supply and wages, does not depend on wages. If that is the case, then theorem 4 tells us that

  $$E[\dot{w} | l, w, \alpha] = \partial_\alpha Q(w | v_1, \alpha),$$
  $$\text{where } v_1 = F(l | \alpha).$$

  This looks deceivingly similar to the control-function result of Imbens and Newey (2009). The results differ both in terms of assumptions and in terms of interpretation, however. Imbens and Newey (2009) would propose to use $v_1 = F(l|\alpha)$ as a control when $\alpha$ is an (excluded, random) instrument, $l$ is an endogenous treatment, $w$ is an outcome of interest, and first stage heterogeneity is assumed to be one-dimensional. Their result involves slopes of the form $\partial_l Q(w|v_1, l)$, interpreted as causal effect of $l$ on $w$. Our result, in contrast, suggests to interpret slopes of the form $\partial_\alpha Q(w|v_1, \alpha)$ as the causal effect of $\alpha$ on $w$. We furthermore do not impose restrictions on the dimensionality of heterogeneity.

- Is either exclusion restriction reasonable? In the context of actual applications, these exclusion restrictions will be imposed *conditional* on a set of predetermined covariates $W$. Arguably, maintaining that the causal effect of a policy change on wages is not predicted by the level of labor supply, conditional on the level of wages and covariates such as age, gender, education, ethnicity, geographic location, etc., is a quite reasonable approximation; and at least as much so as is the assumption of exogeneity of $\alpha$ conditional on $W$. Exogeneity of $\alpha$ conditional on $W$, or a similar assumption, needs to be imposed for identification of $f(x^\alpha)$.

- Note also that either exclusion restriction is considerably weaker than what is maintained in the literature on determinants of the wage distribution, such as Card (2009) or Autor et al. (2008). These papers, and all structural approaches that I am aware of, require that there is no systematic heterogeneity in causal effects on log wages conditional on covariates:

$$E[\dot{w}/w|l, w, W, \alpha] = E[\dot{w}/w|W, \alpha].$$

  This assumption implies that $\alpha$ does not affect inequality systematically conditional on $W$. This might be problematic if our object of interest is the impact of $\alpha$ on inequality.

  Fully structural approaches tend to impose the additional and even stronger restriction that there is no variation in unobserved heterogeneity conditional on outcomes.

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.

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

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

is strictly monotonically increasing in $\epsilon^j$ 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

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

and

$$x^j(\alpha', \epsilon) = Q^{x^j}(v^j | v^1, \ldots, v^{j-1}, \alpha') \tag{24}$$

for any $\alpha$, $\alpha'$.

Equation (24) allows to predict counterfactual outcomes under alternative $\alpha$, for any given unit of observation, once we know her realized $x$ and $\alpha$. This equation tells us that we can predict counterfactual outcomes based on the conditional $v^j$th quantile given $\alpha$ and the controls $v^1, \ldots, v^{j-1}$, which in turn are functions of $x$ and $\alpha$ – if we are willing to maintain the assumptions of theorem 5.

Under the dimensionality restriction on unobserved heterogeneity imposed by theorem 5, there is no heterogeneity in causal effects conditional on outcomes, so that

$$\dot{x} = E[\dot{x} | x, \alpha].$$

It follows that the result of theorem 4, $E[\dot{x}|x, \alpha] = \partial_\alpha Q(v^j | v^1, \ldots, v^{j-1}, \alpha)$, can be strengthened to

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

This equation is a differentiated version of the result of theorem 5.

## IIIB Controls, instruments, and panel data

In section IIIA, we considered the problem of identifying $E[\dot{x}|x, \alpha]$ under the assumption that $\alpha$ 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) $\alpha$ is conditionally random, or (ii) there is a valid instrument $Z$, or (iii) we have panel data where changes of $\alpha$ over time are independent of changes of other factors affecting outcomes. This corresponds to estimation of causal effects using controls, instrumental variables, or differences in differences. 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 $\alpha$ is independent of unobserved heterogeneity $\epsilon$ 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 quantile regressions with the appropriate controls $W$ and $v^{j'}$.

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

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

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

$$E[\dot{x}^j|x, W, \alpha] = \partial_\alpha Q(v^j|v^1, \ldots, v^{j-1}, W, \alpha), \tag{25}$$

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

In settings where conditional independence of $\alpha$ can not plausibly be maintained, we might instead have an instrument $Z$ for which conditional independence holds, and which affects outcomes only through its effects on $\alpha$. 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 $\eta$. Define the weighting function

$$\varphi(\alpha, z, W) := -\frac{\partial_z F(\alpha|z, W)}{\partial_z F(z|\alpha, W)}, \tag{26}$$

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

$$f^{x^\alpha}(x|W) = f(x|\alpha, W) \cdot \varphi(\alpha, z, W), \tag{27}$$

and proposition 1 applies to the observed data distribution reweighted by $\varphi$.

In practice, the support requirement that $F(\alpha|z, W)$ has full support $[0, 1]$ given $\alpha$ 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 $\alpha$ varies as a function of time $\tau$ 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 $\epsilon$ does not vary over time within states $s$. Time $\tau$ is allowed to have a causal impact on outcomes $x$, which is however assumed to not interact with the level of $\alpha$.

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 $\tau$ and $\alpha$, 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).

# IV Aggregation

Recall that one of the main goals of our analysis is to make principled normative statements about whether policy changes increase or decrease social welfare. To do so, it is necessary to first characterize the effect of policy changes on individual welfare, and then to identify and estimate these individual welfare effects. We finally need to take individual welfare effects and aggregate them into social welfare statements, trading off the welfare of different individuals.

We have already taken care of the effect of policy changes on individuals: Lemma 1 characterizes the effect of a policy change on individual net income $y$ and on money-metric utility $e$. Section III then provided conditions which are sufficient for identification of expected individual welfare effects $\gamma(y, W) = E[\dot{e} | y, W, \alpha]$.

In this section we take policy effects $\dot{y}$ on individual realized income and $\dot{e}$ on individual welfare as given, and we discuss aggregation of these individual effects. That is, we discuss the effect of policy changes on aggregate statistics $\theta$ of the income distribution, and on social welfare $SWF$, which is a function of individual welfare $v_i$ for all individuals $i \in \mathscr{I}$. Under some restrictions on welfare weights to be discussed below, policy effects on social welfare can be written as $S\dot{W}F = E[\omega \cdot \gamma]$ for welfare weights $\omega$, so that identification of $\gamma$ implies identification of $S\dot{W}F$.

The purpose of this section is twofold. First, we suggest various alternative approaches for estimating aggregate social welfare effects. These estimation approaches are based on the identification results for individual welfare effects discussed in section III. Second, we provide a conceptual discussion which relates two literatures, the literature on social welfare evaluations in normative public finance, and the statistical literature on distributional decompositions.

We prove the following claims:

1. For marginal policy changes, it is without loss of generality to only consider social welfare functions which are can be written as weighted averages of individual (money metric) utility. Policy effects on social welfare are therefore given by a weighted average of policy effects on individual welfare, $S\dot{W}F = E[\omega^{SWF} \cdot \dot{e}]$. Here $\omega^{SWF}$ is the welfare weight, or marginal value of an additional dollar, assigned to each individual.

2. 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.

3. There are various equivalent ways of calculating $S\dot{W}F$ which are based on imputing either conditional expected welfare effects $\gamma$ or some counterfactual income to each individual. These equivalent representations can be used for alternative estimation approaches.

## *IVA Preliminaries*

In addition to effects on social welfare, we discuss in this section effects on aggregate statistics $\theta$ 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 – 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 $\theta$ maps $P_y$ into $\mathbb{R}$.

In order to elegantly characterize and relate $\dot{\theta}$ and $S\dot{W}F$, we need to impose additional differentiability conditions on either functional; the following assumption 3 does so. Definition 2 assumes $\theta$ is a statistic of the income distribution $P_y$. We can also, however, think of it as a functional of the random variable $(y_i : i \in \mathscr{I})$.[11] 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 $\theta$ for either representation. This assumption also refers to the influence function $IF(y)$ of $\theta$. The influence function allows to approximate $\theta$ by an expectation, $\theta(\alpha) \approx \theta(0) + E[IF(y^\alpha)]$, in a neighborhood of $\alpha = 0$.

Assumption 3 (Differentiability).

1. $SWF$ is Gateaux-differentiable[12] on the set of random variables $v$, equipped with the $L^2$ norm.
   $\theta$ is Gateaux-differentiable on the set of random variables $y$, equipped with the $L^2$ norm.
2. $\theta$ is Gateaux-differentiable on the set of probability distributions $P_y$, equipped with some norm, so that the influence function $IF(y)$ of $\theta$ exists.
3. The influence function $IF(y)$ of $\theta$ is differentiable in $y$.

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

12. 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.

## *IVB Characterizing aggregate policy effects*

The following theorem 6 characterizes marginal policy effects on $SWF$ and on $\theta$, using differentiability assumption 3. This theorem provides, in particular, 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 $\theta$, as a functional of the random variable $y$ and as a functional of the distribution $P_y$.

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 $\theta$ and $SWF$.

1. **Welfare weights:**
   Suppose that assumption 3.1 holds. Then there exist random variables $\omega^{SWF}$ and $\omega^{\theta}$ such that[13]
$$S\dot{W}F = E[\omega^{SWF} \cdot \dot{e}] \tag{28}$$
$$\dot{\theta} = E[\omega^{\theta} \cdot \dot{y}]. \tag{29}$$

2. **Influence function:**
   Suppose that assumption 3.2 holds. Then
$$\dot{\theta} = \partial_\alpha E\left[IF(y^\alpha)\right] = \partial_\alpha \int IF(y) dF_{y^\alpha}(y). \tag{30}$$

3. **Relating the two:**
   Suppose that assumptions 3.1-3.3 hold. Then
$$\omega^{\theta} = \partial_y IF(y). \tag{31}$$

**Remarks:**

- It is instructive to consider the case of a finite population. In that case,

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

the welfare weights of equation (28) are equal to

$$\omega_i = \partial_{v_i} SWF(v) \cdot \partial_{y_0} v_i. \tag{32}$$

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, $\omega^{SWF}$ is a function of $y$.

- Differentiability of the influence function of $\theta$, 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.

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

The following theorem provides alternative representations of $S\dot{W}F$ under the assumption that the welfare weights $\omega^{SWF}$ are a function of income $y$ and covariates $w$, and that $\omega^{SWF} = \omega^\theta$, which allows to relate $S\dot{W}F$ 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 $\widetilde{y}^\alpha = l^0 \cdot w^\alpha - t^\alpha(l^0 \cdot w^\alpha) + y_0^\alpha$. and the behavioral effect $b = \dot{l} \cdot n$.

Then

$$\dot{e} = \dot{\widetilde{y}} = \dot{y} - b, \tag{33}$$

and $S\dot{W}F$ can be rewritten in the following ways.

1. **Welfare weights:**

$$
\begin{aligned}
S\dot{W}F &= E[\omega \cdot \dot{\tilde{y}}] \\
&= E[\omega \cdot \gamma]
\end{aligned} \tag{34}
$$

with $\gamma$ as in definition 1.

2. **Counterfactual income distribution:**

$$
S\dot{W}F = \partial_\alpha \theta \left(P_{\widetilde{y}^\alpha}\right). \tag{35}
$$

3. **Influence function:**

$$
S\dot{W}F = \partial_\alpha E\left[IF(\widetilde{y}^\alpha)\right] = \partial_\alpha \int IF(y) dF_{\widetilde{y}^\alpha}(y). \tag{36}
$$

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

$$
\begin{aligned}
\dot{\theta} - S\dot{W}F &= E[\omega \cdot b] \\
&= \partial_\alpha \theta \left(P_{\breve{y}^\alpha}\right) \\
&= \partial_\alpha E\left[IF(\breve{y}^\alpha)\right].
\end{aligned} \tag{37}
$$

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

**Remarks:**

- Theorem 7 defines two counterfactual income variables, $\widetilde{y}^\alpha$ and $\breve{y}^\alpha$.

  $\widetilde{y}^\alpha$ is the income an individual would receive given baseline ($\alpha = 0$) labor supply and policy $\alpha$ wages , taxes, and unearned income. The derivative of $\widetilde{y}^\alpha$ with respect to $\alpha$ at $\alpha = 0$ gives the welfare effect $\dot{e}$.

  $\breve{y}^\alpha$ is the income an individual would receive given policy $\alpha$ labor supply and baseline ($\alpha = 0$) wages, taxes, and unearned income. The derivative of $\breve{y}^\alpha$ with respect to $\alpha$ 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 $\omega$ are directly specified. This is the route we will pursue.

2. We can impute $\widetilde{y}^\alpha$, based on counterfactual wages, taxes, and unearned income to individuals in the baseline sample. Or impute $\widetilde{y}^\alpha$, based on counterfactual labor supply to individuals in the policy $\alpha$ sample. Either way, we can apply distributional decomposition methods such as reweighting or RIF regression for statistics of the distribution of $\widetilde{y}^\alpha$.

3. We can impute $\breve{y}^\alpha$, similarly to imputing $\widetilde{y}^\alpha$, and apply one of the decomposition methods to the distribution of $\breve{y}^\alpha$. We can then use $S\dot{W}F = \dot{\theta} - \partial_\alpha \theta\left(P_{\breve{y}^\alpha}\right)$ and thus obtain $S\dot{W}F$ by applying a "behavioral correction" to a standard decomposition.

The first of these approaches has three important advantages. First, it is possible to identify $\gamma$ under weaker conditions than necessary to identify counterfactual outcomes such as $\widetilde{y}^\alpha$ and $\breve{y}^\alpha$. 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. Third, it allows the researcher to be agnostic about welfare weights, and lets readers aggregate reported welfare effects $\gamma$ using their own welfare weights.

# V Estimation

This section discusses estimation based on the identification results of section III and the aggregation results of section IV. We first consider the baseline case as discussed in section IIIA, with random variation in $\alpha$ 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 IIIB, 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 $\gamma$, $\mathscr{W}$, $\mathscr{L}$, and $S\dot{W}F$. This is discussed in section VC. In section VD 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 VE.

## VA Estimation of $g$ in the baseline case

Suppose that the assumptions of theorem 4 hold, so that $g^j(x,\alpha) = E[\dot{x}^j|x,\alpha] = \partial_\alpha Q(v^j|v^1,\ldots,v^{j-1},\alpha)$, where $v^j = F(x^j|x^1,\ldots,x^{j-1},\alpha)$. 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 $(x,\alpha)$ and take $(\widehat{v}^1,\ldots,\widehat{v}^{j-1})$ as given.
2. Define the following local weights around $(\alpha,\widehat{v}^1,\ldots\widehat{v}^{j-1})$.
$$K_i^j = \frac{1}{\rho^j}\cdot K\left(\frac{1}{\rho}\left\|\alpha_i-\alpha,\widehat{v}_i^1-\widehat{v}^1,\ldots,\widehat{v}_i^{j-1}-\widehat{v}^{j-1}\right\|\right) \tag{38}$$
for a kernel function $K$[14] and a suitably chosen bandwidth $\rho$.[15]
3. Let
$$\widehat{v}^j = \frac{E_N[K_i^j\cdot\mathbf{1}(x_i^j\le x^j)]}{E_N[K_i^j]}. \tag{39}$$
4. Let, finally,
$$\widehat{g}^j = \underset{g^j}{\operatorname{argmin}}\; E_N\left[K_i^j\cdot U_i^j\cdot(\widehat{v}^j-\mathbf{1}(U_i^j\le 0))\right],\ \text{where} \tag{40}$$
$$U_i^j = x_i^j - x^j - \alpha\cdot g^j. \tag{41}$$

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

14. For instance the Epanechnikov-kernel $K(a) = \max(0,1-a^2)$.
15. We use a common bandwidth $\rho$ for all variables for simplicity of notation; in general different bandwidth for different variables might be desirable.

## VB *Estimation of* $g$ *using controls, instruments, or panel data*

In the context of most applications of interest for distributional policy evaluation, experimental variation of $\alpha$ will not be available. The estimator just sketched immediately generalizes, however, to the more general settings considered in IIIB. 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 of section VA can be used to estimate $g(x, W, \alpha) = E[\dot{x}^j | x, W, \alpha]$ once we replace the local weights by

$$K_i^j = \frac{1}{\rho^{\dim(W)+j}} \cdot K\left(\frac{1}{\rho}\left\|\alpha_i - \alpha, W_i - W, \widehat{v}_i^1 - \widehat{v}^1, \ldots, \widehat{v}_i^{j-1} - \widehat{v}^{j-1}\right\|\right). \quad (42)$$

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

$$\widehat{\varphi}(\alpha, z, W) := -\frac{\partial_z \widehat{F}(\alpha|z, W)}{\widehat{f}(z|\alpha, W)}. \quad (43)$$

   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,

$$\partial_z \widehat{F}(\alpha|z, W) = \operatorname*{argmin}_b \min_a \sum_i \left(L\left((\alpha - \alpha_i)/\rho_\alpha\right) - a - b \cdot (Z_i - z)\right)^2 \cdot K\left(\frac{1}{\rho}\|W_i - W, Z_i - z\|\right). \quad (44)$$

   In the latter expression, $L$ is the cumulative distribution function of a smooth symmetric distribution with support $[-1, 1]$, and $\rho_\alpha$ 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 $\mathbf{1}\left(\alpha - \alpha_i \leq 0\right)$.

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 $\tau$ taking the place of $\alpha$.

   (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 $\tau$. The slope of this regression provides an estimator of $g(x, W, \alpha)$.

   As an alternative to this approach, and this is indeed the route we take in the context of the application discussed in section VI, one might estimate a (flexible) parametric model for $Q(v^j|v^1, \ldots, v^{j-1}, s, W, \alpha, \tau)$ of the form[16]

$$q^j(W, \alpha, v_1, \ldots, v^j) + \delta_s^j(v_j) + \delta_\tau^j(v_j).$$

   This model restricts state and time to enter in the form of additive fixed effects $\delta_s$ and $\delta_\tau$ that are not interacted with each other nor with $\alpha$, but which might vary across quantiles $v_j$. The slope of this regression with respect to $\alpha$, $\partial_\alpha q^j(W, \alpha, v_1, \ldots, v^j)$, provides an estimator of $g(x, W, \alpha)$.

## *VC Estimation of $\gamma$, $\mathscr{W}$, $\mathscr{L}$, 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 $\gamma$, $\mathscr{W}$, $\mathscr{L}$, and $S\dot{W}F$ as introduced in definition 1. To estimate $\gamma$, we need to calculate a conditional average, given income $y$ and covariates $W$, of the estimated effects of the policy change on wages $w$ (scaled by labor supply), unearned income $y^0$ and taxes and transfers $t$. Assume for simplicity, as in section III, that only wages are affected by policy changes, so that we can ignore unearned income and taxes. In that case welfare effects

16. Note that this is more restrictive than necessary from the point of view of identification.

simplify to $\dot{e} = l \cdot \dot{w}$, and we can estimate $\gamma$ by

$$\widehat{\gamma}(y, W) = E_N \left[ K_i \cdot l \cdot (1 - \partial_z t) \cdot \widehat{\dot{w}} \right] / E_N[K_i] \tag{45}$$

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

$$K_i = \frac{1}{\rho^{2+\dim(W)}} \cdot K\left( \frac{1}{\rho} \|\alpha_i - 0, y_i - y, W_i - W\| \right). \tag{46}$$

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

$$\begin{aligned} \widehat{\mathscr{W}} &= \{(y, W) : \widehat{\gamma}(y, W) \geq 0\} \\ \widehat{\mathscr{L}} &= \{(y, W) : \widehat{\gamma}(y, W) \leq 0\} \\ \widehat{S\dot{W}F} &= E_N[\omega_i \cdot \widehat{\gamma}(y_i, W_i)]. \end{aligned} \tag{47}$$

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,

$$\widehat{E}[W|\mathscr{W}] = \frac{E_N[W \cdot \mathbf{1}(\widehat{\gamma}(y_i, W_i) > 0)]}{E_N[\mathbf{1}(\widehat{\gamma}(y_i, W_i) > 0)]}. \tag{48}$$

## VD 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 $\epsilon$. 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

$$\begin{aligned}
\widehat{x}^j(\alpha', \epsilon) &= \widehat{Q}^{x^j}(\epsilon^j | \epsilon^1, \ldots, \epsilon^{j-1}, \alpha') \\
&= \underset{x^j}{\operatorname{argmin}} \; E_N \left[ K_i^j \cdot (x_i^j - x^j) \cdot (\epsilon^j - \mathbf{1}(x_i^j - x^j \leq 0)) \right] \qquad (49)
\end{aligned}$$

where

$$K_i^j = \frac{1}{\rho^j} \cdot K \left( \frac{1}{\rho} \left\| \alpha_i - \alpha, \widehat{v}_i^1 - \epsilon^1, \ldots, \widehat{v}_i^{j-1} - \epsilon^{j-1} \right\| \right) \qquad (50)$$

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

It is worth noting that under the assumptions of theorem 4 the slope of $x(., \epsilon)$, estimated in this way, identifies $g$. This is true even if the additional assumption of restricted heterogeneity imposed in theorem 5 is incorrect.

## VE Standard errors and confidence sets

Inference on all parameters of interest $\vartheta$ 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

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

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 $B_i$.

2. Reweight each observation by $B_i / \sum_{i'} B_{i'}$.

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.[17] This allows, in particular, 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.

# VI Application

We shall now turn to an application of the proposed methods. This section re-evaluates the welfare impact of the extension of the Earned Income Tax Credit (EITC) during the 1990s. The EITC is a refundable tax credit for low to moderate income working individuals and couples. The amount of EITC benefit depends on a recipient's income and number of children. Figure III plots the schedule of federal EITC payments in 2002 as a function of earnings and of the number of children in a household.

[Figure III here]

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

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

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 wage-effects of the EITC expansion using the framework introduced in section II. 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).

### *VIA Data and background*

For our analysis we use the same subsample of the Current Population Survey Merged Outgoing Rotation Group as Leigh (2010). We restrict the sample to the 14-year period 1989-2002, and to individuals aged 25-55 and not self-employed. Extreme observations with reported earnings less than half the federal minimum wage, or more than 100 times the federal minimum wage are excluded. This leaves us with 1346058 observations.

Hourly wage, for those not reporting it directly, is calculated by dividing weekly earnings by usual weekly hours. Wages are converted to wages in 2002 dollars using the CPI. Labor supply is set to "usual hours worked" for those working, and zero for those not employed. The individual-level controls we use include age as well as dummies for gender, whether the respondent identified as black or as hispanic, educational attainment, and number of children.

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 to any given household are equal to (federal EITC payments to this household) times (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 on welfare, conditional on initial incomes. Both the expansion of federal payments and state supplements take essentially the form of a proportional increase of payments, so that this setting is well described by a one-dimensional policy parameter $\alpha$, despite the fact that EITC schedules depend on several parameters governing phase in and phase out of payments. Our definition of treatment $\alpha$, indexing the generosity of EITC payments, is

$$\alpha := \log(\text{maximum attainable EITC payments in a state and year}).$$

If, for a example, a state provides a supplement of 10% to the federal EITC, this implies that our value for treatment $\alpha$ is increased by $0.095 \approx 10\%$ relative to what it would be in the absence of a state supplement.

[Table I here]

Table II reproduces 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 high school dropouts and of those with a high school diploma only, while only having a small effect on the rest of the population.

Notice the large magnitude of these effects. The reported coefficient for wages suggests that a 10% expansion of EITC payments results in a 5% drop of wages for high school dropouts, and in a 2% drop of wages for those with a high school diploma only. Wage effects of this size might more than cancel the increase in EITC payments. While large, these magnitudes are in line with standard estimates of labor demand elasticities of around $-0.3$, given the estimated labor supply effects of the EITC. What is somewhat surprising – but is in line with other studies estimating the effects of the EITC expansion – is the magnitude of labor supply effects. This magnitude is surprising to the extent that the reduction in wages effectively cancels the subsidy of work provided by the EITC. Table III reports analogous estimates, but drops the state-level policy controls used by Leigh. The results remain qualitatively the same.

[Tables II and III here]

We will next estimate disaggregated welfare effects of the EITC expansion, using the identification result for panel data of proposition 3. Our specifications will use the same variation and controls as the regressions used for estimating wage effects in table III; we control, in particular, for the same set of demographic covariates, and state and time fixed effects.

## *VIB Results*

We shall now turn to our estimates of disaggregated welfare impacts. Recall that our approach relies on two crucial identifying assumptions: (i) Quasi-random changes of $\alpha$ over time within states, as in difference-in-difference approaches to the identification of causal effects. This assumption is necessary to guarantee that we correctly identify the distribution of potential outcomes $(w^\alpha, l^\alpha)$ given covariates.

(ii) The restriction on heterogeneity imposed by theorem 4, that either causal effects of policy on wages $w$ do not vary systematically with labor supply *conditional* on wages and covariates, or reversely that causal effects on labor supply $l$ do not vary systematically with wages *conditional* on labor supply and covariates. This assumption is necessary to guarantee that we correctly identify conditional causal effects from the distribution of potential outcomes $(w^\alpha, l^\alpha)$ given covariates. It is worth emphasizing again that this assumption is restrictive, but (a) less restrictive than what is imposed in the existing literature on the impact of policies on wages, and (b) plausible given that we control for both the endogenous outcome itself as well as for a fairly rich set of covariates.

We estimate disaggregated welfare impacts by implementing a semi-parametric version of the 3-step estimation procedure proposed in section V. In the first step, we impute estimates of $v_1$ and $v_2$ to every observation, using nonparametric kernel regression. The control function $v_1$ is given by the conditional cdf of labor supply (hours worked; set to 0 for those without employment) given demographic covariates (gender, race, age, education, number of children), state, and year. The variable $v_2$ is given by the conditional cdf of hourly wages given $v_1$, demographic covariates, state, and year.

In the second step, we implement a parametric version of our estimator for panel data. For every value $v_2$, we run linear quantile regressions of wages on the generosity of EITC benefits, the control variable $v_1$, demographic covariates, state and year fixed effects. These quantile regressions are analogous to the mean regression estimates reported in table III, except that we now allow policy effects to vary with education, gender, age, and $v_1$.

In the third step, we use the estimates from the second step in order to impute wage effects of a 10% expansion of EITC benefits to everyone in a baseline sample; we choose observations in 2002 (the last year in our sample) as the baseline. These imputed wage effects vary across individuals, since the coefficient on $\alpha$ varies by demographics, $v_1$, and $v_2$. Under our identifying assumptions, these imputed effects are consistent estimates of the true effects given $w, l$, and covariates. Multiplying imputed wage effects by labor supply yields estimates of $E[\dot{e}|w, l, W, \alpha]$, the estimated welfare impact of a 10% expansion of the EITC on a given individual through it's impact on wages.

Let us now turn to the results. Figure IV plots estimates of $E[\dot{e}|w, l, W, \alpha]$, the expected welfare impact of a 10% expansion of the EITC induced by changing wages, given annual pre-tax earnings, for a random subsample of 1000 households.[18] These estimated welfare impacts are "pre tax," as well. In order to obtain actual welfare impacts, they should be multiplied by 1 minus the marginal tax rate, which is negative in the phase in range of the EITC and positive in the phase-out range. Since household income, which is necessary for computation of marginal tax rates, is unobserved for most of our sample, we report only pre-tax effects.

Figure V shows estimates of $E[\dot{e}|w, l, W, \alpha]$ for the subgroups of high school dropouts and those with a high school diploma only. We can summarize these estimated welfare effects by plotting kernel regression estimates of expected conditional welfare effects for the various educational subgroups given earnings. This is done in figure VI. Figure VII shows the same estimates as figure VI in separate plots, including pointwise 95% confidence bands obtained using the Bayesian bootstrap as described in section VE.

[Figures IV, V, VI, and VII here]

From these figures we learn the following. Average welfare losses due to reduced wages are substantial and largest for high school dropouts earning 20.000 US$ or more, who lose the equivalent of 900 US$. These results are quite close to the values that a "back of the envelope" calculation based on Leigh's estimates would suggest: His estimates imply that a 10% expansion of the EITC would lead to a 4% reduction in wages.

Those with just a high-school diploma similarly experience large losses. Losses rise with earnings for earnings between 0 and 20.000 US$ per year, and then decline slightly for higher earnings. This reflects two counteracting effects: On the one hand, those with little or no initial earnings have "little left to lose," in US$ terms, even for large percentage drops in their wages. On the other hand, the negative impact on wages appears to be concentrated among those with low wages, as economic intuition would suggest. People with higher earnings experience, therefore, a smaller loss in percentage terms.

To put these estimated welfare effects into context, let us consider the effect

18. More draws would result in a plot that is visually too cluttered – there are almost 160.000 observations in our baseline sample.

of a 10% EITC expansion on realized earnings (as opposed to welfare), as well as the welfare effects of the observed historical changes of wages for the period 1989-2002.

The effect of a 10% increase in EITC payments on earnings is shown in figure VIII.[19] The results in this figure are consistent with the intuition about the difference between $\dot{y}$ and $\dot{e}$ provided by lemma 1, and with the empirical results of Leigh (2010). The effects of an EITC expansion on earnings are of the same sign, but attenuated magnitude relative to the effects on realized welfare. The reason is that the EITC increases labor supply, which affects earnings positively, that is $w \cdot \dot{l}$ is positive. This term enters $\dot{y}$ but does not enter $\dot{e}$, by the envelope theorem argument.

The welfare effects of historical changes shown in figure IX clearly show that high school dropouts did worse over this period than the rest of the population. These results do not reflect a clear pattern of "polarization" in the labor market, as discussed by Autor and Dorn (2013) and others. The latter would manifest itself in a decline in middle-incomes and simultaneous rise of incomes for those at the bottom, decreasing inequality within the poor, while increasing overall inequality.

[Figures VIII and IX here]

## *VIC 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 approaches to evaluate the individual-level 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

19. Note that this figure is on a different scale relative to figure VI.

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). We follow up on their argument. Our results suggest endogenous wages to be an important channel for the distributional welfare impact of the EITC expansion; we otherwise adopt the utility-based perspective of optimal tax theory.

In the terminology introduced in our discussion of lemma 1, our empirical findings suggest that the EITC has a positive mechanical effect for the working poor eligible to receive EITC payments, a positive labor supply effect for the same, 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. Evaluation based on income:
   mechanical + wage + labor supply effect, $-\dot{t} + l \cdot \dot{w} \cdot (1 - \partial_z t) + \dot{l} \cdot n$
2. Evaluation based on utility, assuming fixed wages:
   mechanical effect, $-\dot{t}$
3. Evaluation based on utility, with endogenous wages:
   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.

# VII Conclusion

In this paper, a set of tools is provided for estimation of individual expected welfare impacts (impacts on realized utility) of policy changes. We consider a framework which allows for a large degree of heterogeneity across individuals as well as for endogenous prices and wages.

Conditional expected welfare effects, given earnings and demographic covariates, can be plotted using the estimators we proposed. Such disaggregated welfare effects are relevant for both normative and positive statements. For normative statements, as in optimal policy theory, individual welfare effects can be aggregated to social welfare effects. Given a choice of social welfare weights that trade off the welfare of different individuals, such aggregation allows to make statements about the social desirability of a given policy change. For positive statements, as in the field of political economy, individual welfare effects allow to construct and characterize sets of winners and losers. To the extent that political actions reflect economic self-interest, the latter correspond to the potential advocates and opponents of a policy change.

The main technical contribution of this paper is to characterize identification of individual welfare effects. In a baseline case, these take the form $E[\dot{w} \cdot l | w \cdot l, \alpha]$. This requires identification of causal effects on a vector of endogenous outcomes $x = (w, l)$, conditional on $x$ itself, that is identification of $E[\dot{x}|x, \alpha]$. We characterize identification of such conditional causal effects borrowing tools from continuum mechanics (fluid dynamics), and derive estimators based on our identification results. These estimators take the form of local linear quantile regressions with control functions.

# 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 $\dot{e} = \dot{\upsilon}/\partial_{y_0}\upsilon$ 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. $\dot{\upsilon} = \partial_{(c,l)}u \cdot (\dot{c}, \dot{l})$ and $\partial_{y_0}\upsilon = \partial_{(c,l)}u \cdot (\partial_{y_0}c, \partial_{y_0}l)$,
2. the individual's first order condition $\partial_{(c,l)}u = \lambda \cdot (p, -n)$ for some Lagrange multiplier $\lambda$, 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 $a$ with bounded support and $h$ is continuous by assumption, we get $\dot{f} = -\nabla \cdot h$. $\square$

**Proof of theorem 2:**

1. <u>$h$ satisfies $\dot{f} = -\nabla \cdot h$ if and only if it is in the identified set:</u>
   The "if" part follows from theorem 1. To show the "only if" part, taking $h$ as given we need to construct a distribution of $\epsilon$ 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. <u>$h^0$ satisfies $\dot{f} = -\nabla \cdot h^0$:</u>
   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. <u>$h$ satisfies $\dot{f} = -\nabla \cdot h$ if and only if $h \in h^0 + \mathscr{H}$:</u>
   For any $h$ in $h^0 + \mathscr{H}$, we have $\nabla \cdot h = \nabla \cdot h^0 + \nabla \cdot \widetilde{h} = -\dot{f} + 0$.
   Reversely, for any $h$ such that $\dot{f} = -\nabla \cdot h$, let $\widetilde{h} := h - h^0$. Then $\widetilde{h} \in \mathscr{H}$.

$\square$

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

2. <u>The only $\widetilde{h} \in \mathscr{H}$ consistent with this assumption is $\widetilde{h} \equiv 0$:</u>
As we have already shown $h^0$ to be consistent with this assumption, it is enough to show that $\nabla \cdot \widetilde{h} = 0$ implies $\widetilde{h} \equiv 0$ if $\widetilde{h}$ 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 $\widetilde{h}$ accordingly. Suppose we have shown $(\widetilde{h}^1, \ldots, \widetilde{h}^{j-1}) = (0, \ldots, 0)$. Applying theorem 1 to the $j$ dimensional model immediately implies $\partial_{x^j} \widetilde{h}^j = 0$. Integrating with respect to $x^j$, and using the fact that the support of $\widetilde{h}$ is contained in the support $\mathbf{X}$ of $x$ implies $\widetilde{h}^j \equiv 0$.

Equation (20) implies

$$E[\dot{x}^{j'}|x^1, \dots, x^k, \alpha] = E[\dot{x}^{j'}|x^1, \dots, x^j, \alpha]$$

for $j \geq j'$. As a consequence, $\widetilde{h}^{j'} = 0$ in the $j$ dimensional model immediately implies $\widetilde{h}^{j'} = 0$ 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 (22) immediately implies that $x^j(\alpha, \epsilon) \leq x^j(\alpha, \epsilon')$ if and only if $\epsilon^j \leq \epsilon^{j\prime}$ when $(\epsilon^1, \dots, \epsilon^{j-1}) = (\epsilon^{1\prime}, \dots, \epsilon^{j-1\prime})$. By the induction assumption we get

$$\begin{aligned} v^j = F(x^j|x^1, \dots, x^{j-1}, \alpha) &= F(x^j|\epsilon^1, \dots, \epsilon^{j-1}, \alpha) \\ &= F(\epsilon^j|\epsilon^1, \dots, \epsilon^{j-1}, \alpha) = F(\epsilon^j|\epsilon^1, \dots, \epsilon^{j-1}) \end{aligned}$$

independent of $\alpha$. The claims regarding $v$ follow. As for the second claim, we get

$$\begin{aligned} Q^{x^j}(v^j|v^1, \dots, v^{j-1}, \alpha') &= Q^{x^j}(v^j|\epsilon^1, \dots, \epsilon^{j-1}, \alpha') \\ &= Q^{x^j}(F(x^j|x^1, \dots, x^{j-1}, \alpha')|\epsilon^1, \dots, \epsilon^{j-1}, \alpha') \\ &= x^j(\alpha', \epsilon). \end{aligned}$$

□

**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). $\square$

**Proof of proposition 3:**

This is again an immediate consequence of theorem 4. $\square$

**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 $\widetilde{\omega}$ such that $S\dot{W}F = E[\widetilde{\omega} \cdot \dot{\upsilon}]$. We can renormalize and define

   $$\omega^{SWF} := \widetilde{\omega} \cdot (-\partial_{t_0} \upsilon), \tag{51}$$

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

2. Assumption 3.2 immediately implies that

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

   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.

$\square$

**Proof of theorem 7:**

The equations $\dot{e} = \dot{\widetilde{y}} = \dot{y} - b$ follow from simple differentiation and lemma 1.

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

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

2. Note that $S\dot{W}F = E[\omega \cdot \dot{\widetilde{y}}]$. $S\dot{W}F = \partial_\alpha \theta \left(P_{\widetilde{y}^\alpha}\right)$ 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 $\widetilde{y}$ instead of $y$.

4. $\dot{\theta} - S\dot{W}F = E[\omega \cdot b]$ holds since $\dot{y} - \dot{e} = b$. The other claims follow analogously to item 2 and 3 of this theorem, once we note that $\dot{\check{y}} = b$.

□

# References


Abbring, J. H. and J. J. Heckman (2007): "Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation," *Handbook of econometrics*, 6, 5145–5303.

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. and D. Dorn (2013): "The growth of low-skill service jobs and the polarization of the US labor market," *The American Economic Review*, 103, 1553–1597.

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.

Blundell, R., D. Kristensen, and R. Matzkin (2014): "Bounding quantile demand functions using revealed preference inequalities," *Journal of Econometrics*, 179, 112–127.

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.

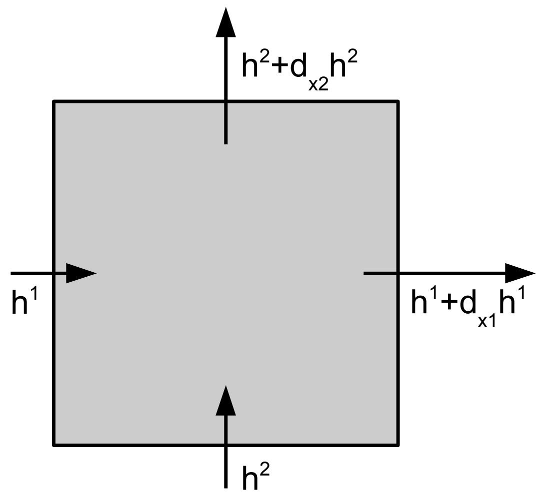
Samuelson, P. (1947): "Foundations of Economic Analysis," .

van der Vaart, A. (2000): *Asymptotic statistics*, Cambridge University Press.

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

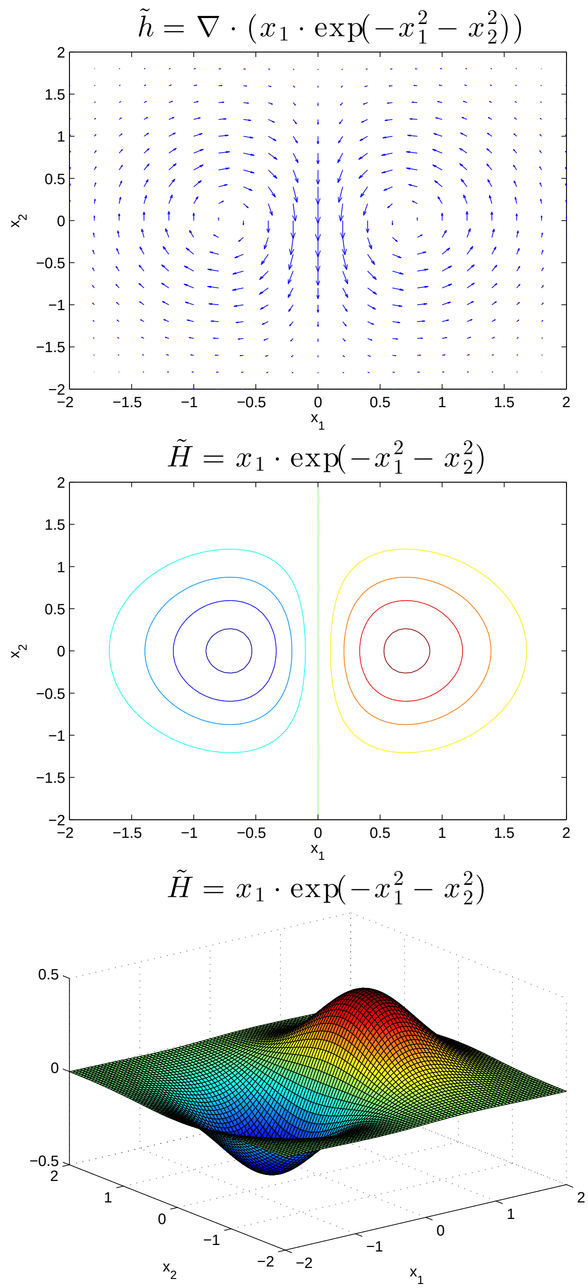
# B Tables and figures

Figure I: Divergence of flow and change of density



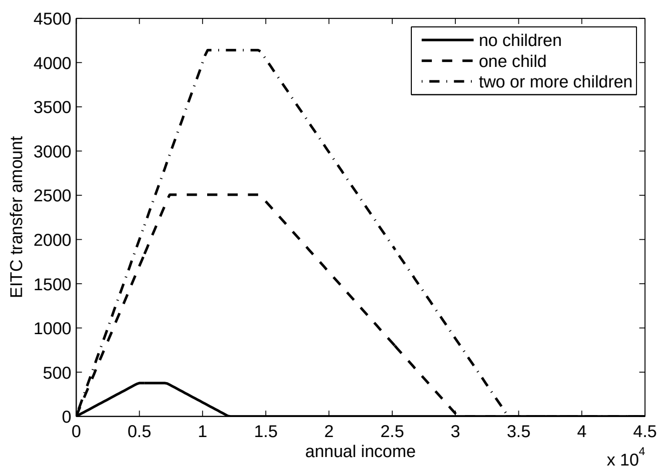

*Notes:* This figure illustrates theorem 1. It relates the change of density $f$ (mass in the square) to the divergence of $h$ (difference in flow on different sides).

Figure II: 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 $\widetilde{h}$ (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).

Figure III: Federal EITC schedule 2002




*Notes:* This figure plots federal EITC payments in 2002 as a function of household income and number of children.

TABLE I: State EITC supplements 1984-2002

| State: # of children: | CO | DC | IA | IL | KS | MA | MD | ME | MN | MN | NJ | NY | OK | OR | RI | VT | WI | WI | WI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 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% |

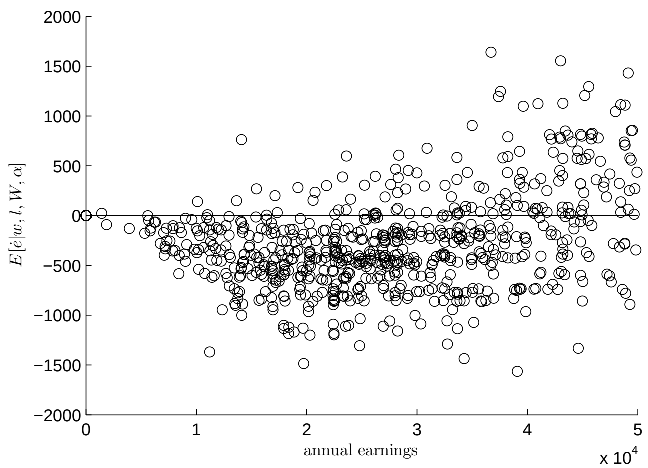
*Notes:* Estimates from Table 4 and 5 of Leigh (2010), for workers with and without children.

TABLE III: Effect of EITC expansion without policy controls

| | **All adults** | **High school dropouts** | **High school diploma only** | **College graduates** |
|---|---|---|---|---|
| | dependent variable: Log real hourly wage | | | |
| Log maximum EITC | -0.125 | -0.380 | -0.221 | -0.018 |
| | [0.081] | [(0.135] | [0.097] | [0.055] |
| Fraction EITC-eligible | 9% | 25% | 12% | 3% |
| | dependent variable: whether employed | | | |
| Log maximum EITC | 0.044 | 0.09 | 0.059 | 0.013 |
| | [0.014] | [0.047] | [0.022] | [0.027] |
| Fraction EITC-eligible | 14% | 34% | 17% | 4% |
| | dependent variable: Log hours per week | | | |
| Log maximum EITC | 0.035 | 0.035 | 0.015 | 0.083 |
| | [0.020] | [0.037] | [0.012] | [0.027] |
| Fraction EITC-eligible | 9% | 25% | 12% | 3% |

*Notes:* This table replicates the results of table II without state-level controls for changes in other policies.

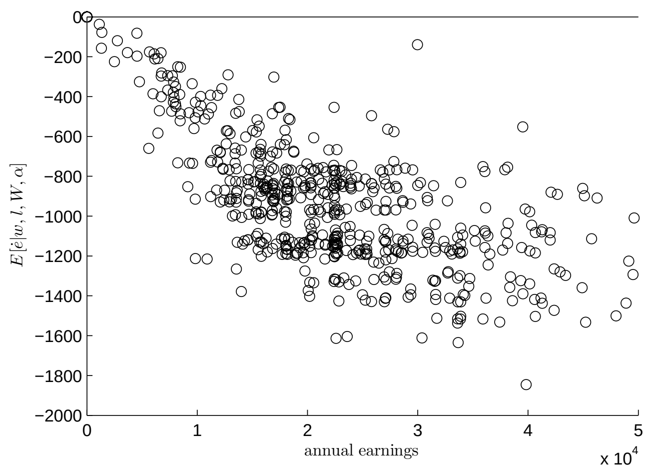
Figure IV: Welfare effects of wage changes induced by a 10% expansion of the EITC



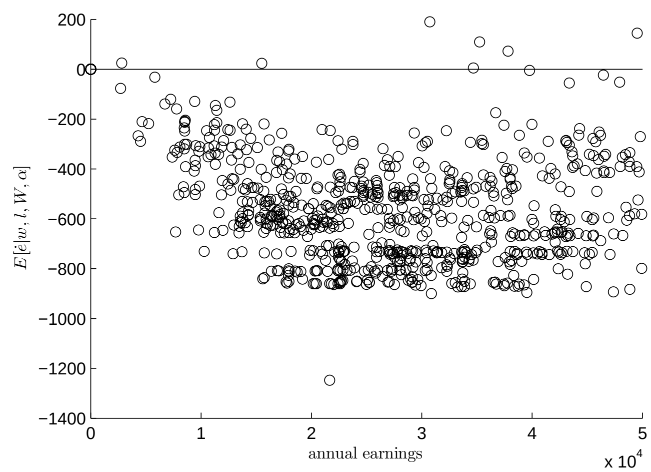

*Notes:* This figure shows the estimated welfare effect $l \cdot \dot{w}$ for a subsample of 1000 households, plotted against their earnings.

Figure V: Welfare effects of wage changes induced by a 10% expansion of the EITC

(a) high school dropouts

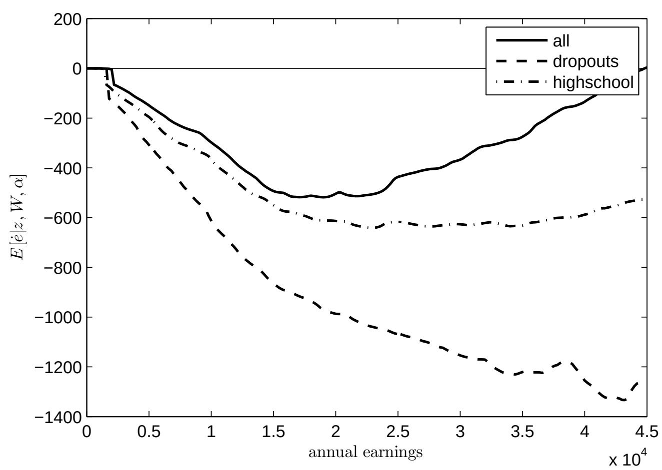



(b) high school diploma only




*Notes:* These figures, similarly to figure IV, show the estimated welfare effect $l \cdot \dot{w}$ for a subsample of 1000 households for the subgroups of high school dropouts and those with a high school diploma only. (Please note the different scales across figures!)

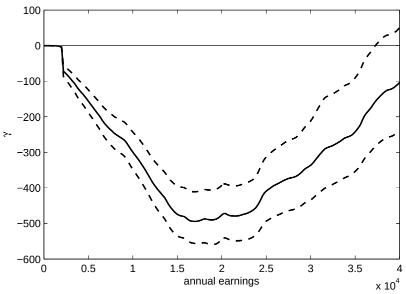
Figure VI: Welfare effects of wage changes induced by a 10% expansion of the EITC



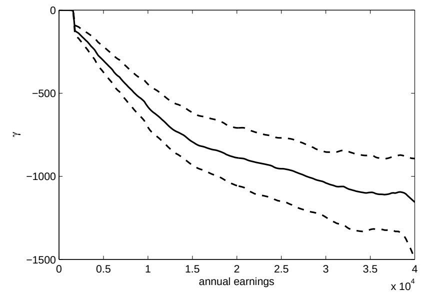

*Notes:* This figure shows kernel regressions of the estimated welfare effect $l \cdot \dot{w}$ on pre-tax earnings, for the entire baseline sample and for educational subgroups.

Figure VII: Welfare effects of wage changes induced by a 10% expansion of the EITC

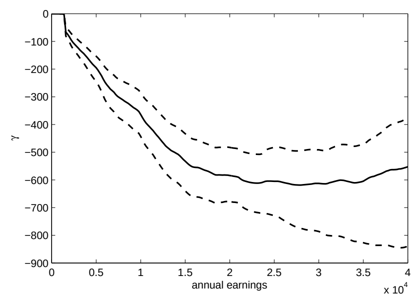
(a) all




(b) high school dropouts

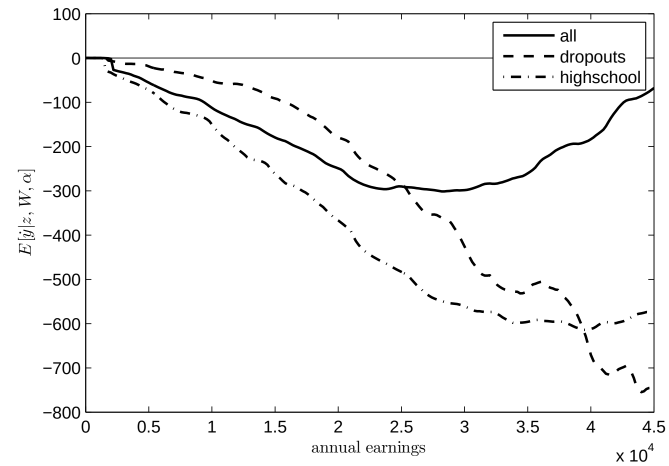



(c) high school diploma only



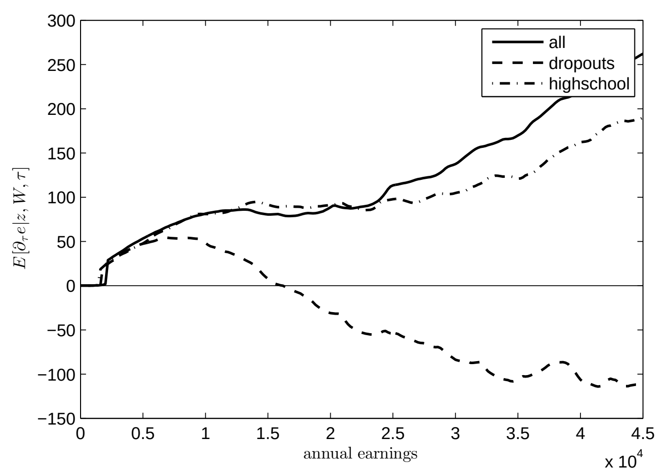

*Notes:* These figures show kernel regressions of the estimated welfare effect $l \cdot \dot{w}$ on pre-tax earnings, for the entire baseline sample and for educational subgroups, as in figure VI, including 95% confidence bands obtained using the Bayesian bootstrap, as described in section VE. (Please note the different scales across figures!)

Figure VIII: Earnings effects of wage and labor supply changes induced by a 10% expansion of the EITC




*Notes:* This figure shows kernel regressions of the estimated earnings effect $\dot{y} = l \cdot \dot{w} + \dot{l} \cdot w$ on pre-tax earnings, for the entire baseline sample and for educational subgroups.

Figure IX: Welfare effects of wage changes over the period 1989-2002




*Notes:* This figure shows kernel regressions of the estimated welfare effect for historical wage changes over the period 1989-2002 on pre-tax earnings, for the entire baseline sample and for educational subgroups.