ARTICLE IN PRESS
Journal of Econometrics xxx (xxxx) xxx
Contents lists available at ScienceDirect
Journal of Econometrics
journal homepage: www.elsevier.com/locate/jeconom

Who wins, who loses? Identification of conditional causal effects, and the welfare impact of changing wages[☆]

Maximilian Kasy
Department of Economics, University of Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 22 February 2019
Received in revised form 22 February 2019
Accepted 1 February 2021
Available online xxxx

JEL classification:
Codes
C21
H22
J3

Keywords:
Incidence
Identification
Marginal effects
Local average derivatives

ABSTRACT

The incidence of tax and other policy changes depends on their impact on equilibrium wages. In a standard model of labor supply, the impact of wage changes on a worker's welfare equals current labor supply times the induced wage change. Worker heterogeneity implies that wage changes vary across workers. In this context, in order to identify welfare effects one needs to identify the causal effect of policy changes on wages conditional on baseline labor supply and wages.

This paper characterizes identification of such outcome-conditioned causal effects for general vectors of endogenous outcomes. Even with exogenous policy variation, outcome-conditioned causal effects are only partially identified for outcome vectors of dimension larger than one. We provide assumptions restricting heterogeneity of effects just enough for point-identification and propose corresponding estimators.

This paper then applies the proposed approach to analyze the distributional welfare impact (i) of the expansion of the Earned Income Tax Credit (EITC) in the 1990s, using variation in state supplements in order to identify causal effects, and (ii) of historical changes of the wage distribution in the US in the 1990s. For the EITC, we find negative welfare effects of depressed wages as a consequence of increased labor supply, in particular for individuals earning around $20,000 per year. Looking at historical changes, we find modest welfare gains rising linearly with earnings.

1. Introduction and economic motivation

Reforms of the tax system have a direct impact on the welfare of tax payers and transfer recipients. They also affect incentives and thereby possibly labor supply and demand, so that they might shift the equilibrium in the labor market, and the wage distribution. This paper proposes methods for an ex-post evaluation of the welfare impact of such equilibrium effects in the labor market.

Two features make this a hard problem. First, welfare is not directly observable, in contrast to wages or earnings. By the envelope theorem, behavioral responses to marginal wage changes do not affect welfare, so that the welfare effect of a marginal wage change is equal to the wage change times a worker's baseline labor supply. Second, workers are heterogeneous, and thus so is the impact of policy changes on their wages. Suppose we are interested in the average welfare impact for workers of a given level of initial earnings; this is a key ingredient for a social welfare evaluation

☆ I would like to thank David Autor, Bence Bardoczy, Gary Chamberlain, Raj Chetty, Ellora Derenoncourt, Arindrajit Dube, Pirmin Fessler, Sergio Firpo, Bryan Graham, James Heckman, Nathaniel Hendren, Stefan Hoderlein, Larry Katz, Susanne Kimm, Nathan Lane, Suresh Naidu, Jesse Rothstein, Emmanuel Saez, and Jann Spieß for valuable discussions and comments. László Sándor provided helpful research assistance. This work was supported by NSF, USA grant SES-1354144 "Statistical decisions and policy choice".
E-mail address: maximilian.kasy@economics.ox.ac.uk.

https://doi.org/10.1016/j.jeconom.2021.02.001
0304-4076/

Please cite this article as: M. Kasy, Who wins, who loses? Identification of conditional causal effects, and the welfare impact of changing wages. Journal of Econometrics (2021), https://doi.org/10.1016/j.jeconom.2021.02.001.

when welfare weights vary across earnings levels. Then we need to identify the average causal effect of the policy change on wages conditional on the initial wage and labor supply of a worker, both of which are endogenous to the policy.

Abstractly, we need to identify the causal effect of a marginal policy change on a vector of endogenous outcomes, such as wage and labor supply, conditional on the initial value of this vector. Without conditioning, this is the standard problem of causal inference, to be solved with (quasi-)experimental variation of the policy. The case with conditioning on a one-dimensional endogenous outcome was solved by Hoderlein and Mammen (2007). The key technical contribution of this paper is a characterization of the case with conditioning on a vector of dimension larger than one, as needed for welfare evaluations. Our characterization is based on insights from continuum mechanics (fluid dynamics), relating changes of a density to the underlying flows.

Evaluating marginal changes to the tax system. Let us now spell out the economic motivation for our subsequent analysis of identification in more detail. Consider a local change to some tax or transfer policy, affecting net income. An example would be the expansion of the Earned Income Tax Credit (EITC) in the 1990s in the United States, which we will consider in Section 3.

Standard approaches for evaluating such (local) policy changes in public finance and optimal tax theory have two components. The first component considers the effect of the change in taxes on government revenue. This effect is due to both the mechanical change in taxes owed (holding behavior constant), and due to behavioral responses affecting the tax base. The total impact on the tax base of behavioral responses is the key causal parameter ("sufficient statistic") that many studies in public finance seek to estimate (Chetty, 2009).

The second component of such an evaluation considers the mechanical effect of the change in taxes on individual net income, holding pre-tax income constant. A key insight in optimal tax theory is that behavioral responses to tax changes do not affect individual welfare, so that the mechanical effect of policy changes is indeed equal to their welfare effect (equivalent variation) for individuals. This holds because of the envelope theorem, cf. Milgrom and Segal (2002). Individual welfare effects are then aggregated into an effect on social welfare using welfare weights (Saez and Stantcheva, 2016), which measure the relative value assigned to a marginal Dollar for each individual. Welfare weights typically depend on baseline income.

Equilibrium effects. An aspect that is often ignored in such evaluations, however, is the equilibrium effect of behavioral responses. Equilibrium effects are for instance assumed to not exist in the classic Mirrlees model of optimal income taxation (Mirrlees, 1971; Saez, 2001). Equilibrium effects imply that policy changes impact individual welfare not only through mechanical effects, but also via changes to prices and wages. The EITC, for instance, subsidizes work for low wage workers. Standard economic theory suggests that such subsidies are absorbed at least in part by employers, through reduced wages. And this is indeed what we will find below.

More generally, let Y_1 denote an individual's hours worked, and let Y_2 denote their hourly wage. Let furthermore $\dot{Y}_2$ be the causal effect of a marginal change of policy on the individual's wage; this notation will be introduced more formally below. Then the equivalent variation of the policy change, as mediated through equilibrium effects, is given by $\dot{Y}_2 \cdot Y_1$, by the envelope theorem.[1] Aggregating these individual level welfare effects using welfare weights Ω, we obtain an impact on social welfare that is equal to $E\left[\left(\dot{Y}_2 \cdot Y_1\right) \cdot \Omega \middle| X\right]$, where the expectation averages over the population of interest, and we condition on the baseline policy parameter X. The welfare weights Ω measure the relative value assigned to a marginal lump-sum Dollar for different individuals; these welfare weights reflect distributional preferences. Typically, the welfare weights Ω are a function of earnings $Y_1 \cdot Y_2$, or more generally of hours Y_1 and wages Y_2, so that $\Omega = \omega(Y)$ where $Y = (Y_1, Y_2)$ (Saez and Stantcheva, 2016). If that is the case, the law of iterated expectations allows us to write the marginal impact of a policy change on social welfare, as mediated by equilibrium effects, as

$$E\left[E\left[\dot{Y}_2 \middle| Y, X\right] \cdot Y_1 \cdot \omega(Y) \middle| X\right]. \tag{1}$$

This object is identified for arbitrary welfare weights of the form $\omega(Y)$ if and only if conditional causal effects of the form $E\left[\dot{Y}|Y, X\right]$ are identified. Pinning down the latter, then, is the main objective of the present paper.

In memory of Gary chamberlain. This paper is dedicated to the memory of Gary Chamberlain. The work presented here (and elsewhere) has benefited from numerous personal conversations with Gary, and builds on his many contributions including (Chamberlain, 1984; Chamberlain and Imbens, 2003; Chamberlain, 2011). This paper is particularly indebted to Gary's openness toward exploring ideas from physics and other fields and their relevance for econometrics.

Roadmap. The rest of this paper is structured as follows. Section 2 discusses identification of conditional causal effects in a general setting. Section 3 applies these identification results to evaluate the expansion of EITC transfers in the 1990s, and historical wage changes over the same period. Appendix A contains all proofs, and Appendix B contains additional tables and figures.

[1] For the case of a static model of labor supply, this follows immediately from the welfare analysis in Mas-Colell et al. (1995), chapter 3. This result holds much more generally, however, including for settings with discrete choices, fixed costs, dynamics, and incomplete information; see the review in Chetty (2009).

2. Identification

Notation. Throughout this paper, random variables are denoted by upper case letters, while values of these variables, as well as functions, are denoted by lower case letters. $Y \in \mathbb{R}^k$ is a k-dimensional random vector of outcomes, $X \in \mathbb{R}$ is a continuous policy or treatment variable, and U is a random element of arbitrary dimension capturing unobserved heterogeneity. We assume that Y is determined by the structural function

$$Y = y(X, U). \tag{2}$$

In the context of our motivating economic example, X indexes a tax policy, such as the magnitude of maximum EITC benefits. Y is a vector of individual outcomes, where Y_2 denotes an individual's hourly wage, and Y_1 denotes their hours worked. In this context, $y(\cdot)$ describes the reduced form (equilibrium) relationship between the policy variable X and the individual outcomes Y.

Subscripts j index components of y. Derivatives with respect to y_j are denoted ∂_j, and the gradient with respect to y is denoted $\nabla = (\partial_1, \ldots, \partial_k)$. Derivatives with respect to the policy parameter x are written $\dot{y}(x, u) = \partial_x y(x, u)$ etc. $f(y|x)$ is the conditional density of Y given X. The letter q denotes (conditional) quantiles. We define

$$\begin{aligned} g(y, x) &= E[\dot{y}(X, U)|Y = y, X = x], && \text{flow} \\ h(y, x) &= g(y, x) \cdot f(y|x), \text{ and} && \text{flow density} \\ \nabla \cdot h &= \sum_{j=1}^{k} \partial_j h_j. && \text{divergence} \end{aligned}$$

Setup. We next introduce our general setup, which is maintained throughout this section. Assumption 1 states that Y is an endogenous outcome vector that is determined by a policy X and unobserved heterogeneity U. The dimension of U is left unrestricted in order to allow for arbitrary heterogeneity. We assume that X varies randomly across observations, so that the standard problem of causal inference poses no problem — the distribution of Y given X is equal to the distribution of potential outcomes. What makes our problem non-standard is the assumption that $\dim(Y) > 1$, and the fact that we are interested in marginal causal effects conditional on Y and X.

Assumption 1.

1. **Observables:** We observe a random sample from the joint distribution of (X, Y), where $Y = y(X, U)$, $Y \in \mathbb{R}^k$, $X \in \mathbb{R}$, and U is a random element in a space of unrestricted dimension.
2. **Exogeneity and support of** X: X is statistically independent of U. The support of X contains an open neighborhood of 0, so that the observed data identify the conditional density $f(y|x)$ for $x \in (-\delta, \delta)$.
3. **Continuity and differentiability:** Y is continuously distributed given X, and $y(x, u)$ is differentiable in x. The function $h(y, x) = E[\dot{y}(X, U)|Y = y, X = x] \cdot f(y|x)$ is continuously differentiable in y, and equal to 0 outside the compact and convex set $\mathscr{Y}$.

The result of Hoderlein and Mammen (2007). The identification results in this paper can be seen as a generalization of the result of Hoderlein and Mammen (2007) to the case of multidimensional outcomes. Hoderlein and Mammen (2007) consider the case $k = \dim(Y) = 1$. Using our notation, they show (under some regularity conditions and exogeneity of X) that

$$\dot{q}_{Y|X}(v|x) = g(y, x). \tag{3}$$

for $y = q_{Y|X}(v, x)$. Recall that we defined $g(y, x) = E[\dot{y}(X, U)|Y = y, X = x]$. In words, the slope of the conditional v-quantile of Y given $X = x$, with respect to x, is equal to the average of the marginal causal effect $\dot{y}(X, U)$ of X on Y, conditional on $X = x$ and conditional on $Y = q_{Y|X}(v, x)$.

Roadmap. We will now develop a series of results characterizing the problem of identifying g (equivalently, h) based on knowledge of f, for the general case $k \geq 1$. Theorem 1 first shows that the divergence of h is identified given f via the identity $\dot{f} = -\nabla \cdot h$. Theorem 1 then shows that the reverse is also true: any flow density h that satisfies this equation is in the identified set, absent further restrictions. Theorem 2 characterizes the identified set when the dimension of y is equal to 1, 2, or 3. Theorem 3 imposes the additional restriction $\partial_j g_{j'}$ for $j > j'$ (a restriction on the heterogeneity of conditional causal effects), and shows that under this restriction h and g are just-identified by nonparametric quantile regressions with control functions.

Partial identification. The following theorem shows that knowledge of f identifies the divergence of h under Assumption 1. It also 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 1. Eq. (5) explicitly provides one particular function h^0 which satisfies the equation $\dot{f} = -\nabla \cdot h^0$. The theorem further shows that the difference $\widetilde{h}$ between any other function h in the identified set and h^0 is in the set (linear subspace) $\mathscr{H} = \{\widetilde{h} : \nabla \cdot \widetilde{h} \equiv 0\}$.

Theorem 1 (*Partial Identification, General Case*)**.** *Suppose Assumption 1 holds. Then*

$$\dot{f}(y|x) = -\nabla \cdot h(y,x). \tag{4}$$

Let V_j be the random variable $V_j = F_{Y_j|Y_1,\ldots,Y_{j-1},X}(Y_j|Y_1,\ldots,Y_{j-1},X)$, where F denotes the conditional cumulative distribution function. Define

$$h_j^0(y,x) = f(y|x) \cdot \dot{q}_{Y_j|V_1,\ldots,V_{j-1},X}(v_j|v_1,\ldots,v_{j-1},x), \tag{5}$$

and let

$$\mathscr{H} = \{\widetilde{h} : \ \nabla \cdot \widetilde{h} \equiv 0, \ \widetilde{h}(y,x) = 0 \text{ for } y \notin \mathscr{Y}\}. \tag{6}$$

Then the identified set for h is given by

$$h^0 + \mathscr{H}. \tag{7}$$

Characterizing the identified set. Since 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 2, 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 $\mathscr{H} = \{\widetilde{h} : \ \nabla \cdot \widetilde{h} \equiv 0\}$. Theorem 2 uses Poincaré's Lemma (cf. Rudin, 1991, chapter 10) to characterize the set $\mathscr{H}$ for dimensions $k = 1, 2$, and 3.[2] For dimension 1 we recover the point identification result of Hoderlein and Mammen (2007) as a special case (and thus provide an alternative proof for their result), while for higher dimensions point-identification breaks down; see also Hoderlein and Mammen (2009).

The case $k = 2$ is of special interest in the context of our motivating empirical example; recall that Y_1 denotes hours worked and Y_2 denotes wages for this example. For the case $k = 2$, the characterization takes on a particularly elegant form. The functions $\widetilde{h}$ in the kernel are exactly those functions which can be written as the gradient of some function $\boldsymbol{h}$, rotated by 90 degrees. $\widetilde{h}$ is thus a vector field pointing along the lines of constant height of $\boldsymbol{h}$.

Theorem 2 (*Partial Identification, Dimensions 1, 2, and 3*)**.** *Let $\mathscr{H} = \{\widetilde{h} : \ \nabla \cdot \widetilde{h} \equiv 0, \ \widetilde{h}(y,x) = 0 \text{ for } y \notin \mathscr{Y}\}$.*

1. *Suppose $k = 1$. Then $\mathscr{H} = \{\widetilde{h} \equiv 0\}$.*
2. *Suppose $k = 2$ and let $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then*

$$\mathscr{H} = \{\widetilde{h} : \widetilde{h} = A \cdot \nabla \boldsymbol{h}, \quad \boldsymbol{h} : \ \mathbb{R}^2 \to \mathbb{R}, \ \boldsymbol{h}(y,x) = 0 \text{ for } y \notin \mathscr{Y}\}.$$

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

$$\mathscr{H} = \{\widetilde{h} : \widetilde{h} = \nabla \times \boldsymbol{h}, \quad \boldsymbol{h} : \ \mathbb{R}^3 \to \mathbb{R}^3, \ \boldsymbol{h}(y,x) = 0 \text{ for } y \notin \mathscr{Y}\}.$$

where $\nabla \times \boldsymbol{h} = \left(\partial_2 \boldsymbol{h}_3 - \partial_3 \boldsymbol{h}_2, \partial_3 \boldsymbol{h}_1 - \partial_1 \boldsymbol{h}_3, \partial_1 \boldsymbol{h}_2 - \partial_2 \boldsymbol{h}^1\right)$.

Just-identification. Theorems 1 and 2 characterize the identified set for h absent any further identifying assumptions, that is if only Assumption 1 is imposed. The following theorem shows that the additional assumption of a "triangular" structure for $\nabla g(y,x)$ (derivatives above the diagonal are 0) yields just-identification of h. Note that the ordering of the components of y matters if we assume such a triangular structure.

Theorem 3 (*Point-identification*)**.** *Suppose Assumption 1 holds. Assume additionally that*

$$\partial_j g_{j'}(y,x) = 0 \text{ for } j > j'. \tag{8}$$

Then g and h are point identified, and

$$g(y,x) = \dot{q}_{Y_j|V_1,\ldots,V_{j-1},X}(v_j|v_1,\ldots,v_{j-1},x), \tag{9}$$

where $V_j = F_{Y_j|Y_1,\ldots,Y_{j-1},X}(Y_j|Y_1,\ldots,Y_{j-1},X)$. The flow density h is equal to h^0 as defined in Theorem 1. There are no over-identifying restrictions implied by Eq. (8).

2.1. Discussion

Intuition and analogy to fluid dynamics. Fig. 1 provides some intuition for the first result of Theorem 1, that $\dot{f} = -\nabla \cdot h$. Consider the density of Y given X, in the shaded square. This density changes, as X changes, by (i) the difference between

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

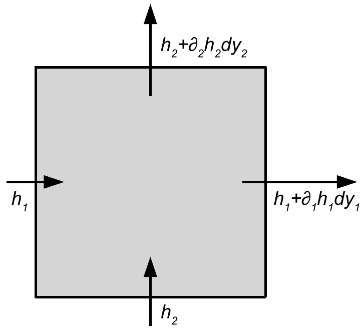

Fig. 1. This figure illustrates Theorem 1. It relates the change of density f (mass in the square) to the divergence of h (difference in flow on different sides).

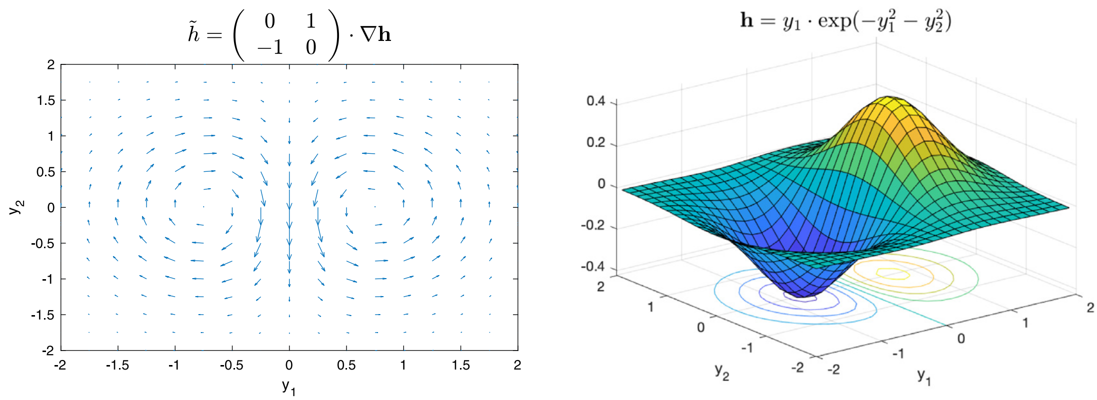

Fig. 2. This figure illustrates the characterization of $\mathscr{H}$ in Theorem 2 for the case $k = 2$. The vector field $\widetilde{h}$ (first graph) is given by a 90 degree rotation of the gradient of some function $\boldsymbol{h}$ (second graph), and thus points along the lines of equal level of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

the outflow to the right and the inflow to the left, $\partial_1 h_1 \cdot dy_1$, and (ii) the difference between the outflow on the top and the inflow on the bottom, $\partial_2 h_2 \cdot dy_2$. The sum of these changes is equal to $-\sum_{j=1}^{k} \partial_j h_j \cdot dy_j$. 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 1 has various analogies in physics, for instance in fluid dynamics. We can think of X as time, U indexing individual particles, and Y the position of a particle in space. The function $y(X, U)$ describes the trajectory of a particle over time. Then $f(y|x)$ is the density of the gas or liquid at location y and time x. 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 for instance approximately true for water.

The source of the identification problem we face is thus accurately illustrated by the following analogy: By stirring your coffee (or other beverage of choice), you can create a variety of different flows $h(y, x)$ which are all consistent with the same constant density $f(y|x)$ of the beverage being stirred. Fig. 2 illustrates this point. This figure plots a function $\boldsymbol{h}$ with a local maximum and a local minimum, and the corresponding $\tilde{h}$, a vector field that points along the lines of constant $\boldsymbol{h}$ circling these extrema.

Parametric example. We next consider a parametric example to provide further intuition for Theorems 1, 2, and 3. Suppose that $k = 2$, $U \sim Uniform(\{u : \ \|u\| \leq 1\})$, and $X \perp U$. Suppose further that $Y = y(X, U)$, where

$$y(x, u) = \begin{pmatrix} \cos(\alpha x) & \sin(\alpha x) \\ -\sin(\alpha x) & \cos(\alpha x) \end{pmatrix} \cdot U,$$

and α is an unknown parameter. For this setting, we have

$$f(y|x) = \frac{1}{\pi} \cdot \mathbf{1}(\|y\| \leq 1), \qquad g(y, x) = \alpha \cdot \begin{pmatrix} -\sin(\alpha x) & \cos(\alpha x) \\ -\cos(\alpha x) & -\sin(\alpha x) \end{pmatrix} \cdot u,$$

and thus $g(y,0) = \alpha \cdot \mathbf{1}(\|y\| \leq 1) \cdot (y_2, \ -y_1)$, as well as $h(y,0) = \frac{1}{\pi} \cdot g(y,0)$. Note in particular that for any value of α and x, $Y|X = x \sim Uniform(\{u : \ \|u\| \leq 1\})$, so that α is not identified from observation of the joint distribution of X and Y, without further restrictions. In this example, X is independent of unobserved heterogeneity, and the dimension of heterogeneity is the same as the dimension of Y, illustrating that the identification problem does not hinge on either of these assumptions. As implied by Theorem 1, non-identification is reflected in the fact that $\nabla \cdot h \equiv 0$ for any value of α. Corresponding to the characterization of Theorem 2, we can choose $\boldsymbol{h}(y,0) = \frac{1}{2\pi} \cdot \mathbf{1}(\|y\| \leq 1) \cdot \|y\|^2$, for which choice we get $h(y,0) = A \cdot \nabla \boldsymbol{h}(y,0)$. Lastly, suppose that we impose the additional assumption that $\partial_2 g_1(y,x) = 0$, as in Theorem 3. The only value of α consistent with this assumption is $\alpha = 0$, yielding point-identification.

Suppose now that we were to observe data generated according to our parametric example, with local variation of X around 0. We might wish to fit a linear random coefficient model to these data in order to estimate $g(y,0)$. This random coefficient model could take the form

$$\hat{y}(x,u) = \beta \cdot [u, \ x \cdot u],$$

where β is a 2×4 coefficient matrix. The empirical model estimated in Section 3 below is a generalization of this linear model, where we allow for individual level controls, state and time fixed effects, and variation of slopes by quantiles. For local variation of X around 0, we get in our parametric model that

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

By the preceding argument, different values of α are observationally equivalent. The assumption that $\partial_2 g_1(y,x) = 0$ corresponds to the assumption that the $(1,4)$ entry of β equals 0, thus pinning down α and β.

The identifying assumption of Theorem 3. The key assumption that yields point identification is given by Eq. (8). This assumption restricts the heterogeneity of causal effects. Consider the case $k = 2$ as in our labor market application, where Y_2 corresponds to hourly wages and Y_1 to hours worked. Eq. (8) then can be spelled out as $\partial_2 E[\dot{y}_1(X,U)|Y_1 = y_1, Y_2 = y_2, X = x] = 0$. This assumption implies that the average effect of a policy change on hours worked, conditional on labor supply and wages, does not depend on wages. The effect of a policy change on wages is still allowed to vary across values of the baseline hours worked, and to be heterogeneous given baseline wage and labor supply. In our application this assumption is imposed conditional on observed covariates, so that the policy effect may also vary across observed demographic groups. In Section 3 we discuss empirical results based on this assumption and results based on the analogous assumption with the role of wages and hours worked reversed, where the restriction $\partial_1 g_2(y,x) = 0$ is imposed. In both cases, we additionally allow for heterogeneity based on a rich set of exogenous covariates.

Eq. (8) restricts heterogeneity less than popular assumptions imposed in the labor economics literature. Papers in the literature on the effect of labor supply on wages often estimate parametric demand systems, see for instance Card (2009) or Autor et al. (2008). For such demand systems, shifts of labor supply only affect the relative wages of a finite number of observable types of workers. Relative wages of different workers belonging to the same type remain unchanged as labor supply varies. Types of workers are defined for instance in terms of education. In our notation, this assumption implies in particular that $g(y,x)$ does not depend on y at all. The literature on the distribution of treatment effects for binary treatments also needs to impose stronger restrictions than Eq. (8) to achieve point identification. Assumptions which achieve point identification include (conditional) independence of potential outcomes for different treatment values, or (conditional) perfect dependence; see for instance Abbring and Heckman (2007). Relative to the binary treatment case, we can impose less stringent assumptions because we have a continuous treatment x and assume smoothness of the response functions $y(x,u)$, and because we do not aim to identify the full distribution of treatment effects. Our setting finally allows for more heterogeneity than models designed to identify structural functions themselves; such models need to restrict the dimension of heterogeneity to be no larger than the dimension of observables, see e.g. Altonji and Matzkin (2005) and Kasy (2011).

Relaxing random variation of X. Thus far, we have considered the problem of identifying $g(y,x)$ under the assumption that X is randomly assigned, so that $U \perp X$. We now discuss how to extend our results to quasi-experimental settings where this assumption does not hold, including (i) settings with conditional independence, (ii) instrumental variables, and (iii) panel data. Throughout the following we continue to assume that Assumption 1 holds, except for exogeneity of X. Suppose, first, that X is *conditionally exogenous* given a vector of observed covariates W, $X \perp U|W$. In this case, Theorems 1, 2, and 3 continue to hold verbatim after conditioning all expressions on W.

Suppose, next, that X itself is not exogenous, but an *instrument* Z is. Suppose additionally, as in Imbens and Newey (2009), that X is determined by a first-stage relationship of the form $X = k(Z,V)$, where k is monotonic in the one-dimensional source of heterogeneity V and $Z \perp (U,V)$. Then conditional independence of X and U holds given the control function $\tilde{V} = F_{X|Z}(X|Z)$. Again, Theorems 1, 2, and 3 hold verbatim after conditioning all expressions on $\tilde{V}$.

Suppose, finally, that X is not exogenous, but we have *panel data* at our disposition, where T denotes the time period of a given observation, and S denotes its cross-sectional unit (e.g., labor market or state). Assume that the policy X is a function of S and T, $X = x(S,T)$, and that the distribution of heterogeneity U does not vary over time within units

S, $U|T, S \sim U|S$. This assumption is known as "marginal stationarity", cf. Graham and Powell (2012) and Chernozhukov et al. (2013). Under this assumption, we can think of time T as an instrument for the policy X, conditional on S. Assume the restrictions of Eq. (8) apply given S, and define V accordingly. Then we get the following variation on the result of Theorem 3,

$$\frac{\partial_t Q_{Y_j|V_1,\ldots,V_{j-1},S,T}(v_j|v_1,\ldots,v_{j-1},s,t)}{\partial_t x(s,t)} = E[\dot{y}_j(X,U)|Y=y,X=x,S=s].$$

In panel data settings it is often assumed that other factors affecting outcomes change over time, in a manner that is shared across units s. To maintain identification of causal effects, it is necessary to restrict the way that time enters the outcome equation. The most common approach, the "difference in differences" approach, is to assume that time fixed effects enter additively. We impose a variant of this assumption in Section 3, where we implement a *quantile difference in differences* estimator, allowing for fixed effects that vary across quantiles.

3. Application

This section applies the theoretical results of Sections 2 and 3 to study the welfare impact of changing wages in the United States over the 14-year period 1989–2002. Both the welfare impact of the EITC expansion and the welfare impact of historical changes of the wage distribution are discussed.

To identify the impact of the EITC expansion, variation across states and time in state-level supplements to the federal EITC is used. This identification approach and the data preparation closely follow Leigh (2010), who estimated the impact of the EITC expansion on average log wages and labor supply by educational group. The results presented here differ from those of Leigh (2010) in that we focus on (i) disaggregated, distributional effects rather than averages, and (ii) on welfare rather than observable outcomes.

This section is structured as follows. It starts with a description of our data and the Earned Income Tax Credit, followed by a replication of some of the results of Leigh (2010) and a description of our estimation procedure. Then the empirical findings for both the impact of the EITC and of historical changes of the wage distribution are discussed, as is the robustness of our results to alternative specifications.

3.1. Data and estimation approach

Data. Following Leigh (2010), we use a subsample of the Current Population Survey Merged Outgoing Rotation Group. 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 1,346,058 observations. Welfare effects are imputed and described for the baseline population corresponding to the subsample of 157,737 observations from the year 2002.

Hourly wage, for those not reporting it directly, is calculated by dividing weekly earnings by usual weekly hours. Wages are converted to wages in 2000 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, and the same variables for a spouse if present. Unless otherwise noted, all earnings and welfare effects reported are on an annual scale.

The earned income tax credit. The EITC is a refundable tax credit for low to moderate income working individuals and couples. The amount of EITC benefits depends on a recipient's income and number of children. Fig. 3 plots the schedule of federal EITC payments in 2002 as a function of earnings and of the number of children in a household. For most of our analysis, we pool EITC schedules across numbers of children, and consider the weighted average maximum EITC payment (corresponding to the high-plateaus in Fig. 3), with weights 0.4 for 1 child, 0.4 for 2 children, and 0.2 for 3 children. These weights correspond roughly to population proportions. Fig. 4 plots the evolution of this weighted average of federal maximum EITC payments, measured in year 2000 US$, based on the CPI. As this figure shows, maximum payments increased about threefold in the 1990s. The literature documents that this EITC expansion increased labor supply, cf. Meyer and Rosenbaum (2001) and Chetty et al. (2013), and possibly depressed wages, cf. Rothstein (2010) and Leigh (2010).

Some states supplement federal payments using proportional top-ups. Table B.1 in the Appendix 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. Both the expansion of federal payments and state supplements take essentially the form of a proportional increase of payments, rather than varying the other parameters of the EITC schedule that govern the phase-in and phase-out of payments. This setting is therefore well described by a one-dimensional policy parameter X. Our definition of treatment X, indexing the generosity of EITC payments, is

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

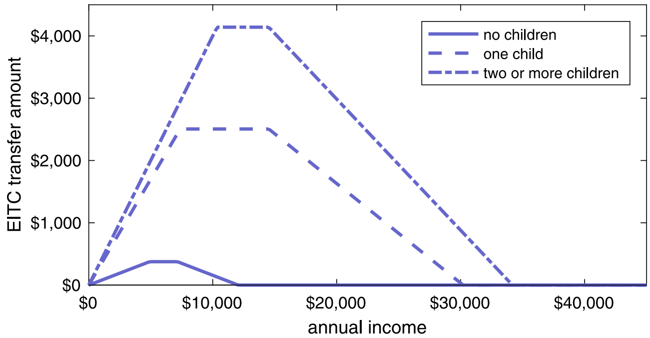

Fig. 3. This figure plots federal EITC payments in 2002 as a function of household income and number of children.

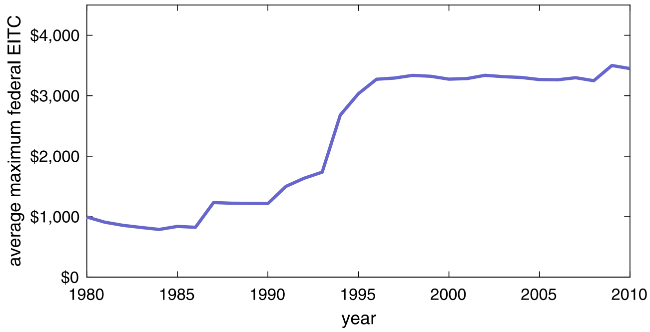

Fig. 4. This figure plots weighted average maximum federal EITC payments, measured in year 2000 US$.

where payments are measured in year 2000 US$. If, for example, a state provides a supplement of 10% to the federal EITC, this implies that the value of treatment X is increased by $0.095 \approx 10\%$ relative to what it would be in the absence of a state supplement.

Replicating (Leigh, 2010) Table B.2 in the Appendix reproduces the main estimates from tables 4 and 5 of Leigh (2010). These estimates imply that the expansion of the EITC increased the labor supply and depressed the 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 not necessarily out of line with estimates of labor demand using macro variation, 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 given that the reduction in wages appears to effectively cancel the subsidy of work provided by the EITC.

For our main estimates we slightly modify Leigh's specifications as follows. We recalculate the definition of treatment X, which yields slightly different values in some cases. For parsimony, we also drop state-level policy controls and drop weighting for the regressions. None of these modifications has a large effect; Table 1 reports the resulting estimates. The conclusions remain qualitatively the same, as can be seen from comparison of Tables B.2 and 1. Our main results build on quantile regression analogs of the specifications in Table 1.

Estimation. To estimate the equilibrium effects of the EITC expansion on welfare, we employ an estimation approach that is a parametric sample analog of the nonparametric identification result in Theorem 3. We estimate the quantile functions $q_{Y_j|V_1,\dots,V_{j-1},x}(v_j|v_1,\dots,v_{j-1},x)$ using linear quantile regressions.[3] We additionally allow for heterogeneity based on observables W. Our general estimation approach is summarized in Fig. 5.

Applying this approach to our setting, let W_1 be a vector of covariates including a constant, age, and a gender dummy, and let W_2 include gender by race dummies, age and age squared, and the same variables for the individual's spouse (if present), as well as state dummies (for states of the United States), and year by number of children dummies. We apply

3 For the quantile regressions, we follow Koenker (2005) and use the code provided by Roger Koenker at http://www.econ.uiuc.edu/~roger/research/rq/rq.html (accessed May 18 2017).

Table 1
Effect of EITC expansion. This table replicates the results of Leigh (2010), as in Table B.2 in the Appendix, using our definition of treatment X, without state-level controls for other policies, and without weighting observations.

	All adults	High school dropouts	High school diploma only	College graduates
	Dependent variable: Log real hourly wage			
Log max EITC	−0.105 (0.0768)	−0.376 (0.134)	−0.223 (0.0884)	0.0137 (0.0548)
	Dependent variable: whether employed			
Log max EITC	0.0182 (0.0156)	0.0807 (0.0413)	0.0342 (0.0224)	−0.00531 (0.0255)
	Dependent variable: Log hours per week			
Log max EITC	0.0307 (0.0178)	−0.0289 (0.0393)	0.0124 (0.0120)	0.0699 (0.0203)

Input: A sample of observations of (W_1, W_2, X, Y), including covariate vectors W_1 and W_2, the policy level X, and the outcome vector Y.

Output: Estimates of $g(y, x, w)$ for any (y, x, w).

Iterate for $j = 1, \ldots, k$:

1. Let $\tilde{W}_1 = \left(W_1, \hat{V}_1, \ldots, \hat{V}_{j-1}\right)$, and $\tilde{W}_2 = \left(W_2, \hat{V}_1, \ldots, \hat{V}_{j-1}\right)$.
2. For each $v \in \{.05, .10, \ldots, .95\}$, estimate the linear quantile regression
$$Y_2 = \left(\tilde{W}_1 \times X\right) \cdot \beta_1 + \tilde{W}_2 \cdot \beta_2 + \tilde{U} \tag{10}$$
by solving $\widehat{\beta} = \operatorname{argmin}_{\beta} E_n \left[\rho_v \left(Y_j - \left(\tilde{W}_1 \times X\right) \cdot \beta_1 + \tilde{W}_2 \cdot \beta_2\right)\right]$,
where $\rho_v(y) = y \cdot (v - \mathbf{1}(y < 0))$, and E_n denotes the sample average.
3. For each observation in the sample, impute $\hat{V}_j \in \{.05, .10, \ldots, .95\}$ corresponding the closest predicted value for Y_j, between these regressions for each value of v.

Impute: For any value of (y, x, w), impute $(\hat{v}_1, \ldots, \hat{v}_k)$ as in the iteration, and for each j also impute $\hat{g}_j(y, x, w) = \tilde{w}_1 \cdot \hat{\beta}_1$ corresponding to the same regressions.

Fig. 5. Estimation procedure.

this estimation approach separately for each educational group (high school dropout, high school, and some college or more). We set Y_1 equal to weekly hours worked, and Y_2 equal to hourly wage.[4]

Having run the estimation procedure described in Fig. 5, we then restrict attention to a baseline sample of observations from the year 2002. For each observation in this sample we impute a conditional quantile $\hat{V}_2$ based upon the closest predicted value of the wage regressions from the second step. We then impute a wage effect of the form $\hat{g}_2$ to each observation, using the coefficients $\hat{\beta}_1$ corresponding to the imputed $\hat{V}_2$.

Heterogeneity in these imputed wage effects is driven by age, gender, education, as well as wage and labor supply via the conditional quantiles $\hat{V}_1$ and $\hat{V}_2$. We finally impute welfare effects of the form $Y_1 \cdot \hat{g}_2$, and run kernel regressions of imputed wage effects and welfare effects on actual earnings, using an Epanechnikov kernel and a bandwidth based on Silverman's rule of thumb.

Inference. For inference, we use the Bayesian bootstrap, as introduced by Rubin (1981) and discussed by Chamberlain and Imbens (2003). It is implemented as follows: For each iteration, draw n i.i.d. exponentially distributed random variables B. Re-weight each observation by B, divided by the sum of B across all observations, and estimate the object of interest for the re-weighted distribution. Repeat the entire procedure, to obtain a set of replicate estimates for each object of interest.

The estimates obtained by this procedure are draws from the posterior corresponding to an (improper) Dirichlet process prior (with parameter 0) over the joint distribution of all observables. This procedure allows, in particular, to construct Bayesian credible sets for the object of interest, using quantiles of the re-sampling distribution. Additionally,

4 More precisely, in the context of this application, we estimate the quantile regression for wages using the log hourly wage as the dependent variable, and then impute $\hat{g}_2(y, x, w) = y_2 \cdot (w_1, \hat{v}_1) \cdot \hat{\beta}_1$, using the appropriate $\hat{\beta}_1$.

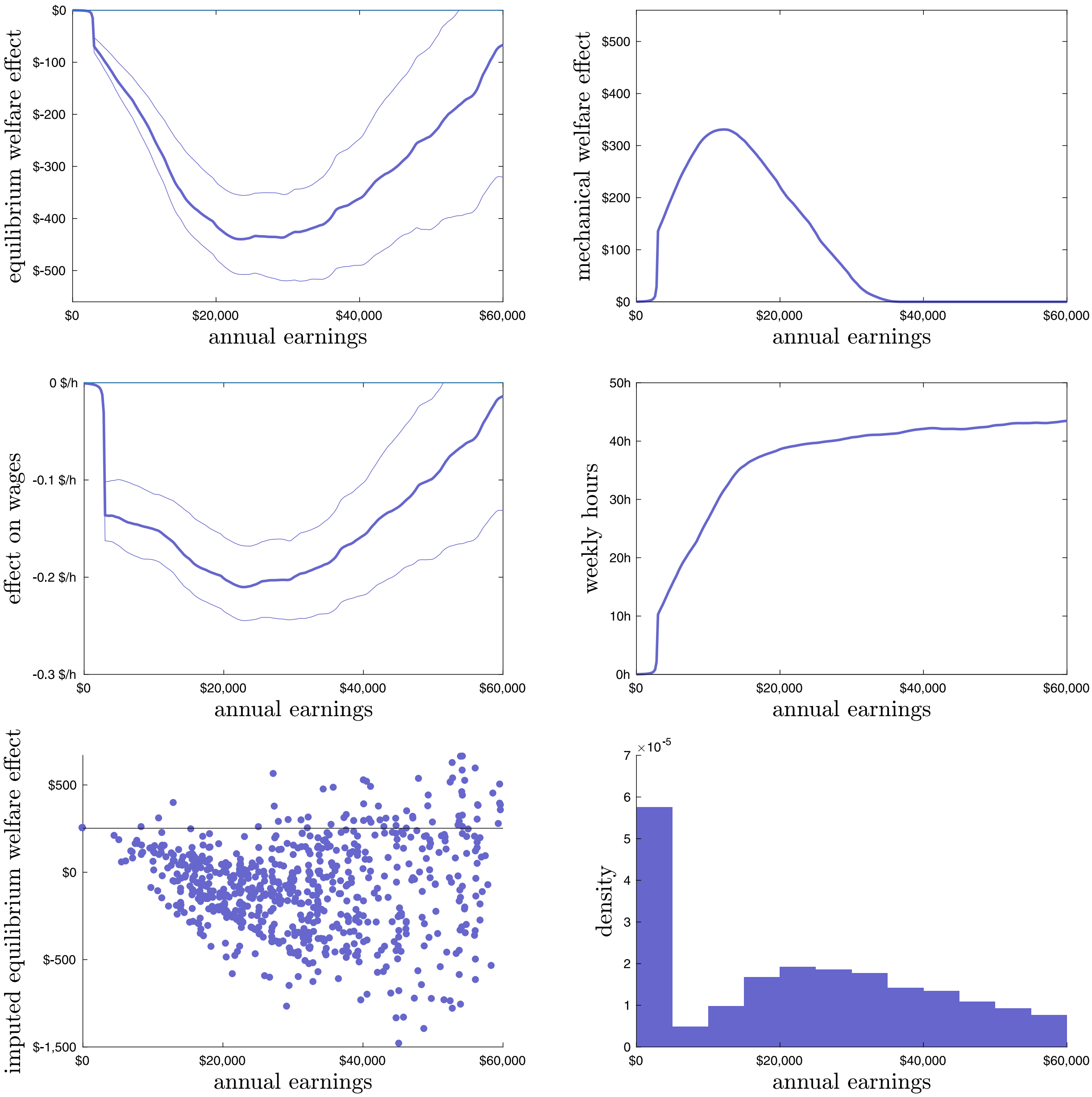

Fig. 6. These figures plot the following for our 2002 baseline sample: Kernel regressions on earnings of (i) the estimated pre-tax welfare effect $Y_1 \cdot \hat{g}_2$ for a 10% EITC expansion, (ii) the imputed mechanical effect, (iii) the imputed wage effect $\hat{g}_2$, and (iv) usual weekly hours worked Y_1. Furthermore (v) a scatter plot of imputed welfare effects for a random sample of 1000 observations from the baseline, and (vi) a histogram of the distribution of earnings. Imputations are based on the three step approach controlling for $\hat{V}_1$ as described in Section 3. Figures (i) and (iii) additionally show pointwise 95% credible bands based on the Bayesian bootstrap.

and similarly to the standard bootstrap, the re-sampling distribution also provides an asymptotically valid estimate of the frequentist sampling distribution for objects such as sample moments and smooth functions thereof. This allows to interpret the Bayesian credible sets as frequentist confidence sets.

3.2. Findings

Results: The effect of the EITC expansion. We now turn to our first set of empirical results, summarized by the plots in Fig. 6. These plots show estimated effects on the 2002 baseline population of a 10% expansion of the EITC. Estimates are based on the three step estimation procedure outlined above, and control in particular for the conditional quantile of labor supply.

The key plot, shown on the top left of Fig. 6, plots the estimated pre-tax conditional average welfare effect $E[Y_1 \cdot \hat{g}_2 | Y_1 \cdot Y_2]$ as a function of earnings $Y_1 \cdot Y_2$.[5] Subject to our identifying assumptions, this plot suggests that the general equilibrium effect of the EITC expansion led to a decrease the welfare of low income individuals, with the most negative effect for individuals earning between \$20,000 and \$30,000 per year. Note how this finding is not something that could have been deduced from the more conventional regression estimates of Tables B.2 and 1, since these regression estimates of the effect on observables neither allow us to obtain welfare effects, nor do they allow us to disentangle heterogeneity across income levels.

The next plot shows the mechanical effect of an EITC expansion by 10%. This plot corresponds to Fig. 3, averaging over number of children given earnings. A comparison between estimated equilibrium effects and mechanical effects shows that, strikingly, the two effects are of opposite sign but similar magnitude and shape. It seems that across the income distribution, the incidence of the EITC expansion was primarily on the employers' side, with low income recipients' welfare essentially unaffected on net.

To better understand the shape of equilibrium welfare effects, it is useful to separately plot kernel regressions of imputed wage effects $\hat{g}_2$ and of labor supply (usual weekly hours) Y_1 on earnings $Y_1 \cdot Y_2$. If the conditional correlation of $\hat{g}_2$ and Y_1 given $Y_1 \cdot Y_2$ were equal to zero, then the product of these two plotted functions would equal the conditional average welfare effects. These plots reveal that wages were depressed throughout the earnings distribution, but negative effects fade as earnings increase. Labor supply, on the other hand, is larger for individuals with higher earnings, but essentially flat and equal to around 40 h per week for individuals earning more than \$20,000 per year. The combination of these two shapes – rising hours per week for low income individuals, declining wage effects for higher income individuals – explains the U-shape of welfare effects.

The kernel regressions themselves obscure considerable conditional heterogeneity of welfare effects given earnings. This is revealed by the bottom left scatter plot, which shows earnings and imputed welfare effects for a random subsample of 1000 observations from the baseline population. The final plot shows a histogram of the distribution of earnings in the baseline sample; a significant number of individuals report zero earnings, while the modal non-zero level of earnings is slightly above \$20,000 per year.

Alternative restrictions on heterogeneity. The imputed welfare and wage effects reported thus far rely on the restriction of heterogeneity implied by the assumption $\partial_1 g_2(X, U) = 0$, as in Theorem 3. We can also construct an estimator based on the alternative assumption $\partial_2 g_1(X, U) = 0$. This assumption suggests a two-step procedure which proceeds exactly like the one discussed above, except that (i) the first step is omitted, and (ii) in the second step we do not control for $\hat{V}_1$ in the quantile regressions predicting Y_2. Fig. B.1 in the Appendix shows the resulting estimates from this two-step procedure. All the qualitative conclusions remain unchanged. This empirical finding provides comfort that our results appear to be robust to alternative assumptions about heterogeneity.

Historical changes. Thus far we have considered the causal impact of the EITC expansion on welfare and wages, under difference-in-difference type assumptions. We can also study the actual historical evolution of market wages, and the welfare impact of changing wages. Note that in this case, the problem of causal inference via exogenous treatment variation (which we solved using a difference-in-differences approach in the case of the EITC) disappears since we just consider changes over time, but we still face the problem of identifying conditional average effects. To do so, consider the same estimation approach as before, but drop all time fixed effects and replace the definition of treatment by $X = (T - 1988)/13$. The resulting estimates correspond to the welfare impact of historical changes over the period 1989–2002. The results of the three-step procedure are shown in the Fig. B.2 in Appendix. The conditional average welfare effects, shown on the top left, reveal welfare gains that are essentially proportional to earnings. This picture corresponds to modest annual welfare gains of about 0.6% of earnings throughout the earnings distribution. The corresponding regression of wage changes on earnings (bottom left) reveals some relative catch-up of wages for very low earnings. This, in conjunction with labor supply rising over the earnings distribution, generates the observed linear pattern of welfare effects. We again consider the results of a two step procedure based on the alternative restriction on heterogeneity in Fig. B.3 in the Appendix. As before, the results are similar to those using the three-step procedure, except that welfare gains appear both smaller and more heterogeneous given earnings, in this case.

Marginal versus discrete changes. We have imputed welfare effects equal to wage changes times baseline labor supply. For marginal changes, these welfare effects are equal to equivalent variation, compensating variation, and change in consumer surplus, see Mas-Colell et al. (1995) chapter 3. It is common practice in public finance to use this assumption of marginal changes to evaluate welfare effects, cf. Chetty (2009).

For discrete changes, such as the EITC expansion and historical shifts considered here, equivalent variation, compensating variation, and change in consumer surplus are generally not the same. If labor supply is monotonically increasing in X, however, then the imputed marginal effects provide a lower bound on all three of these. If we were to use a different baseline sample, we would obtain a corresponding upper bound.

5 We focus on pre-tax effects to avoid the need to impute taxes based on the limited information available in the CPS.

Appendix A. Proofs

Proof of Theorem 1. Let b be an arbitrary differentiable function defined on $\mathscr{Y}$ with bounded support $\mathscr{B}$. The function b will serve as a "test function" in the following proof. Let P denote the (marginal) distribution of U, and define

$$\begin{aligned}\bar{b}(x) = E[b(y(X,U))|X=x] &= \int b(y(x,u))dP(u)\\ &= \int b(y)f(y|x)dy.\end{aligned}$$

Corresponding to the two integral representations of $\bar{b}(x)$, there are two representations for $\dot{\bar{b}}(x)$. Using the first representation we get

$$\begin{aligned}
\dot{\bar{b}}(x) &= \int \partial_x\,(b(y(x,u)))\,dP(u) && \text{Exchanging integration, differentiation}\\
&= \int \nabla b(y(x,u))\cdot \dot{y}(x,u)dP(u) && \text{Chain rule}\\
&= \int \nabla b(y)\cdot E[\dot{y}(X,U)|Y=y, X=x]f(y|x)dy && \text{Law of iterated expectations}\\
&= \int \nabla b(y)\cdot h(y,x)dy && \text{Definition of } h\\
&= \sum_{j=1}^{k}\int \partial_j b(y)\cdot h_j(y,x)dy && \text{Inner product as sum}\\
&= -\sum_{j=1}^{k}\int b(y)\cdot \partial_j h_j(y,x)dy && \text{Partial integration}\\
&= -\int b(y)\cdot\sum_{j=1}^{k}\partial_j h_j(y,x) = -\int b(y)\cdot(\nabla\cdot h)(y,x)dy. && \text{Definition of divergence}
\end{aligned}$$

The two key steps of this derivation are (i) exchange of the order of differentiation and integration, and (ii) partial integration, from the third to the fourth line. (i) is justified since by assumption the functions involved are differentiable and have compact support $\mathscr{B}$. (ii) does not involve any boundary terms, again by the compact support assumption for b.

Using the second integral representation for $\bar{b}(x)$, we can alternatively write

$$\begin{aligned}\dot{\bar{b}}(x) &= \partial_x \int b(y)f(y|x)dy\\ &= \int b(y)\dot{f}(y|x)dy.\end{aligned}$$

We get that

$$\int b(y)\cdot\dot{f}(y|x)dy = -\int b(y)\cdot(\nabla\cdot h(y,x))dy$$

for any differentiable b with bounded support. Since $\nabla\cdot h$ is continuous by assumption, it follows that $\dot{f}(y|x) = -\nabla\cdot h(y,x)$.

We next show that the identified set for h is equal to $h^0 + \mathscr{H}$.

1. *h satisfies $\dot{f} = -\nabla\cdot h$ if and only if it is in the identified set:*
   The "if" part follows from the argument above. To show the "only if" part, taking h as given we need to construct a distribution of U and structural functions y consistent with (i) h, (ii) the observed data distribution, and (iii) Assumption 1. Such a data generating process can be constructed as follows: Let $y(0,u) = u$. This implies $f(u) = f(y|x=0)$. Let $y(.,u)$ be a solution to the ordinary differential equation

   $$\dot{y} = h(y,x),\;\; y(0,u) = u.$$

   Such a solution exists by Peano's theorem. It is easy to check that this solution satisfies all required conditions. This solution corresponds to the case where there is no heterogeneity of $\dot{y}$ conditional on X, Y.
2. *h^0 satisfies $\dot{f} = -\nabla\cdot h^0$:*
   Consider the model $y_j(x,u) = q_{Y_j|V^1,\dots,V^{j-1},X}(v^j|v^1,\dots,v^{j-1},x)$ where $u = (v^1,\dots,v^k)$. Then this model implies $E[\dot{y}(X,U)|Y=y,X=x]\cdot f(y|x) = h^0(y,x)$, 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 the first part of our theorem.

3. *h satisfies $\dot{f} = -\nabla \cdot h$ if and only if $h \in h^0 + \mathscr{H}$:*
   For any h in $h^0 + \mathscr{H}$, we have $\nabla \cdot h = \nabla \cdot h^0 + \nabla \cdot \widetilde{h} = -\dot{f} + 0$.
   Reversely, for any h such that $\dot{f} = -\nabla \cdot h$, let $\widetilde{h} := h - h^0$. Then $\widetilde{h} \in \mathscr{H}$. □

Proof of Theorem 2. Throughout this proof, assume that h is such that $\nabla \cdot h \equiv 0$.

1. $k = 1$: In this case, by definition, $\nabla \cdot h = \partial_y h = 0$. Since h has its support contained in $\mathscr{Y}$, 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 the differential form
   $$\omega = h_1 dy_2 - h_2 dy_1.$$
   Then $d\omega = (\partial_1 h_1 + \partial_2 h_2) dy_1 \wedge dy_2 = 0$ if and only if $\omega = d\boldsymbol{h} = (\partial_1 \boldsymbol{h}) dy_1 + (\partial_2 \boldsymbol{h}) dy_2$ for some $\boldsymbol{h}$.
3. This follows again from Poincaré's Lemma, applied to
   $$\omega = h_1 dy_2 \wedge dy_3 + h_2 dy_3 \wedge dy_1 + h_3 dy_1 \wedge dy_2$$
   and $\lambda = \sum_j \boldsymbol{h}_j dy_j$. □

Proof of Theorem 3.

1. *h^0 is consistent with this assumption:*
   Consider again the model $y_j(x, u) = Q_{Y_j|V_1,\ldots,V_{j-1},X}(v_j|v_1, \ldots, v_{j-1}, x)$ where $u = (v_1, \ldots, v_k)$, as in the proof of Theorem 1. This model implies $\partial_j h_{j'}(y, x) = 0$ for $j > j'$. This model is furthermore consistent with the observed data distribution and satisfies $E[\dot{y}(X, Y)|Y = y, X = x] \cdot f(y|x) = h^0(y, x)$.
2. *The only $\widetilde{h} \in \mathscr{H}$ consistent with this assumption is $\widetilde{h} \equiv 0$:*
   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 $y_1, \ldots, y_j$, and define $\widetilde{h}$ accordingly. Suppose we have shown $(\widetilde{h}_1, \ldots, \widetilde{h}_{j-1}) \equiv (0, \ldots, 0)$. Applying Theorem 1 to the j dimensional model immediately implies $\partial_j \widetilde{h}_j = 0$. Integrating with respect to y_j, and using the fact that the support of $\widetilde{h}$ is contained in the support $\mathscr{Y}$ of Y imply $\widetilde{h}_j \equiv 0$.
   Eq. (8) implies
   $$E[\dot{y}_{j'}(X, U)|Y_1, \ldots, Y_k, X] = E[\dot{y}_{j'}(X, U)|Y_1, \ldots, Y_j, X]$$
   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. □

Appendix B. Additional tables and figures

See Tables B.1, B.2 and Figs. B.1–B.3.

Table B.1
This table shows the percentage amounts of state supplements to the federal EITC for the years 1984–2002. Based on Table 2 from Leigh (2010).

State # of children	CO	DC	IA	IL	KS	MA	MD 1+	ME	MN 0	MN 1+	NJ	NY 1+	OK	OR	RI	VT	WI 1	WI 2	WI 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

Table B.2
Effect of EITC expansion on wages and labor supply. Estimates from Tables 4 and 5 of Leigh (2010), for workers with and without children.

	All adults	High school dropouts	High school diploma only	College graduates
	Dependent variable: Log real hourly wage			
Log max 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 max 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 max EITC	0.037	0.042	0.011	0.095
	(0.019)	(0.040)	(0.014)	(0.027)
Fraction EITC-eligible	9%	25%	12%	3%

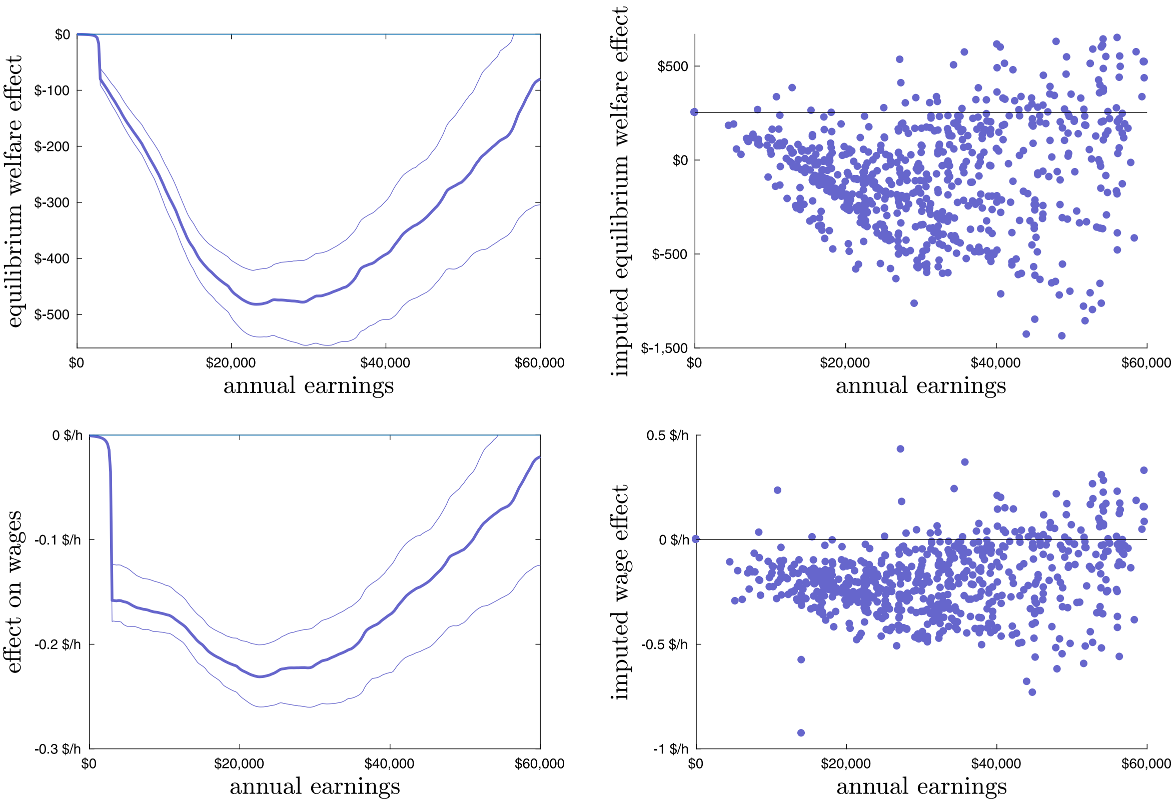

Fig. B.1. These figures show plots analogous to Fig. 6, where imputations are based on a two-step approach and do not control for V_1.

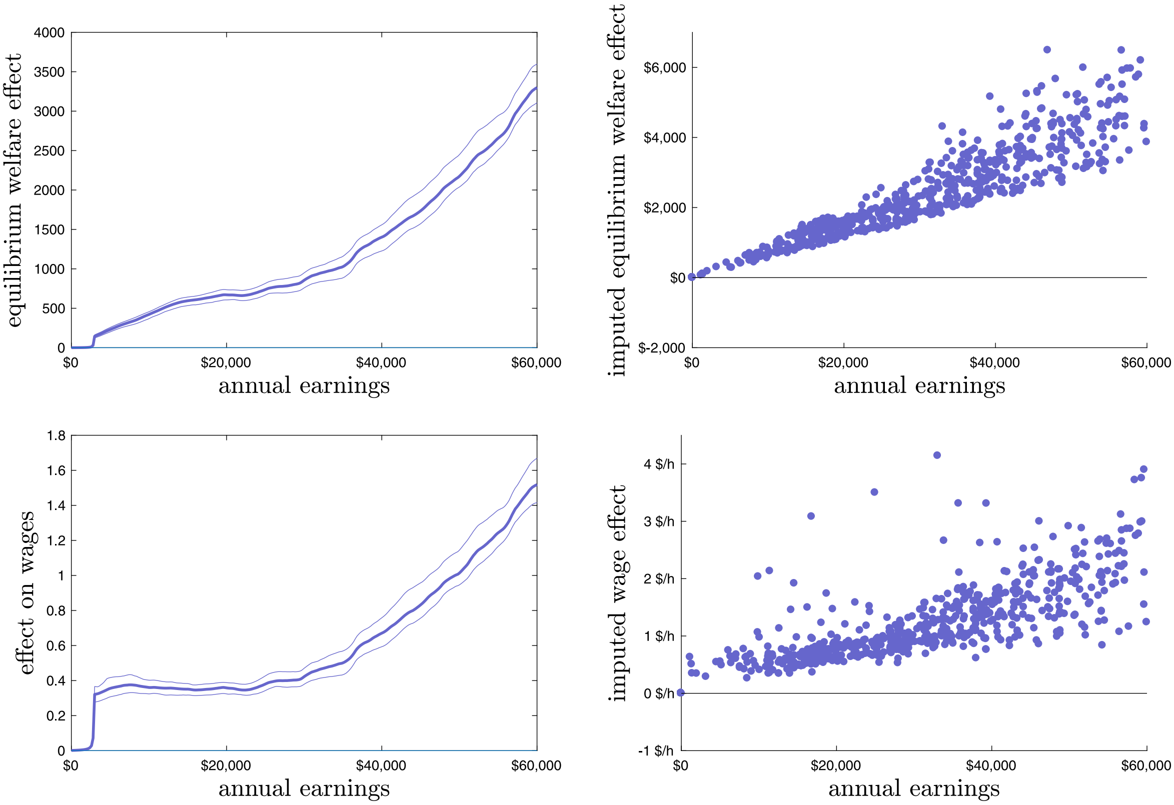

Fig. B.2. These figures show plots analogous to Fig. 6, where now treatment X has been replaced by normalized time $X = (T - 1988)/13$, and time fixed effects have been dropped from the controls. Effects correspond to actual historical changes over the time period 1989–2002.

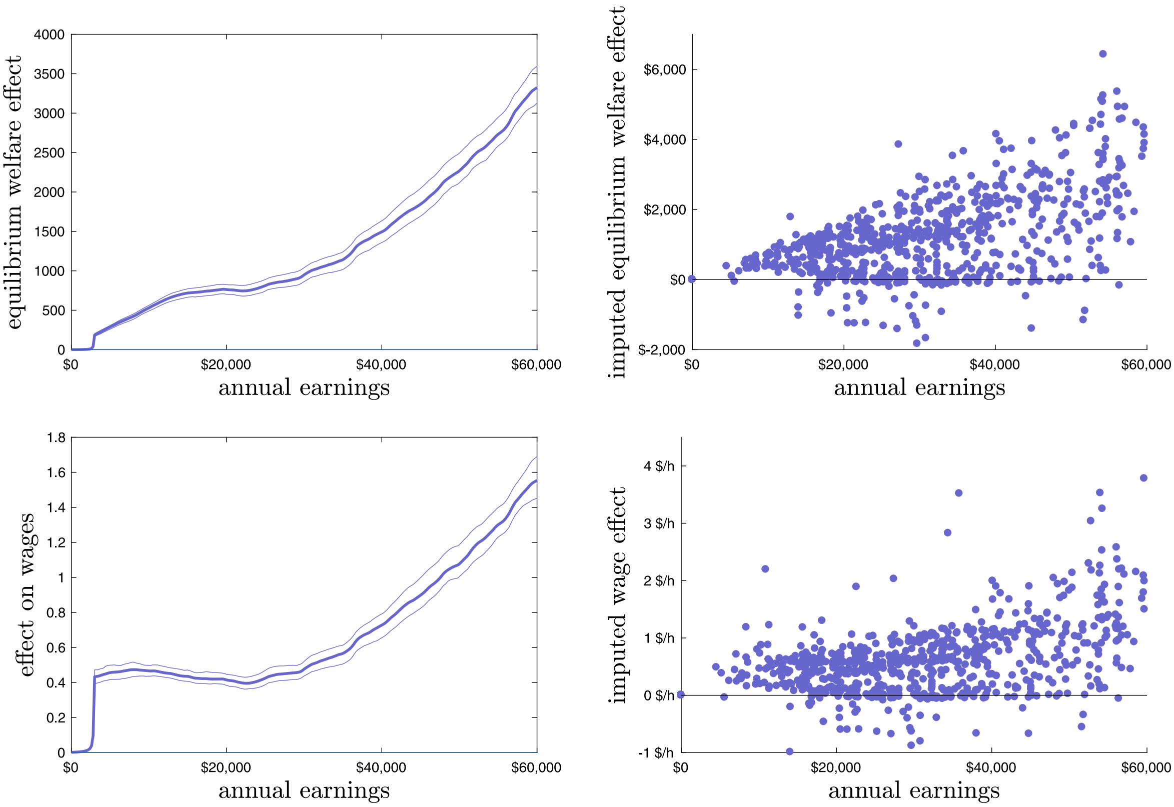

Fig. B.3. These figures show plots analogous to Fig. B.2, where imputations are based on a two-step approach and do not control for V_1.

References

Abbring, J.H., Heckman, J.J., 2007. Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. Handb. Econom. 6, 5145–5303.
Altonji, J., Matzkin, R., 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. Econometrica 73 (4), 1053–1102.
Autor, D.H., Katz, L.F., Kearney, M.S., 2008. Trends in US wage inequality: Revising the revisionists. Rev. Econ. Stat. 90 (2), 300–323.
Card, D., 2009. Immigration and inequality. Amer. Econ. Rev. 99 (2), 1–21.
Chamberlain, G., 1984. Panel data. Handb. Econom. 2.
Chamberlain, G., 2011. Bayesian aspects of treatment choice. In: Oxford Handbook of Bayesian Econometrics.
Chamberlain, G., Imbens, G.W., 2003. Nonparametric applications of Bayesian inference. J. Bus. Econom. Statist. 21 (1), 12–18.
Chernozhukov, V., Fernandez-Val, I., Hahn, J., Newey, W., 2013. Average and quantile effects in nonseparable panel models. Econometrica 81 (2), 535–580.
Chetty, R., 2009. Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. Annu. Rev. Econ. 1 (1), 451–488.
Chetty, R., Friedman, J.N., Saez, E., 2013. Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. AER 103 (7), 2683–2721.
Graham, B.S., Powell, J.L., 2012. Identification and estimation of average partial effects in irregular correlated random coefficient panel data models. Econometrica 80 (5), 2105–2152.
Hoderlein, S., Mammen, E., 2007. Identification of marginal effects in nonseparable models without monotonicity. Econometrica 75 (5), 1513–1518.
Hoderlein, S., Mammen, E., 2009. Identification and estimation of local average derivatives in non-separable models without monotonicity. Econom. J. 12 (1), 1–25.
Imbens, G.W., Newey, W.K., 2009. Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77, 1481–1512.
Kasy, M., 2011. Identification in triangular systems using control functions. Econometric Theory 27 (03), 663–671.
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. BE J. Econ. Anal. Policy 10 (1).
Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory. Oxford University Press.
Meyer, B.D., Rosenbaum, D.T., 2001. Welfare, the earned income tax credit, and the labor supply of single mothers. QJE 116 (3), 1063–1114.
Milgrom, P., Segal, I., 2002. Envelope theorems for arbitrary choice sets. Econometrica 70 (2), 583–601.
Mirrlees, J., 1971. An exploration in the theory of optimum income taxation. Rev. Econom. Stud. 175–208.
Rothstein, J., 2010. Is the EITC as good as an NIT? Conditional cash transfers and tax incidence. Amer. Econ. J. Econ. Policy 177–208.
Rubin, D.B., 1981. The Bayesian bootstrap. Ann. Statist. 130–134.
Rudin, W., 1991. Principles of Mathematical Analysis. McGraw-Hill.
Saez, E., 2001. Using elasticities to derive optimal income tax rates. Rev. Econom. Stud. 68 (1), 205–229.
Saez, E., Stantcheva, S., 2016. Generalized social welfare weights for optimal tax theory. AER 106 (1), 24–45.