Università Politecnica delle Marche

Dipartimento di Scienze Economiche e Sociali

A Welfare Evaluation of the 1986 Tax Reform for Married Couples in the United States

Matteo Picchio and Giacomo Valletta

January 2016

Indirizzo: Matteo Picchio. Corresponding author. Department of
                Economics and Social Sciences, Marche Polytechnic Uni-
                versity, Piazzale Martelli 8, 60121 Ancona, Italy; Cen-
                tER, Tilburg University, The Netherlands; Sherppa, Ghent
                University, Belgium; IZA, Bonn, Germany. E-mail:
                `m.picchio@univpm.it`. Tel.: +39 071 220 7176.
                Giacomo Valletta. CORE - Université catholique
                de Louvain, Louvain-la-Neuve, Belgium. E-mail:
                `giacomovalletta@gmail.com`.

A Welfare Evaluation of the 1986 Tax Reform for Married Couples in the United States[*]

Matteo Picchio and Giacomo Valletta

1   Introduction

Since the beginning of the 1980s, the US went through several reforms of the federal income tax rules. As described in Eissa et al. (2008), a series of tax acts passed in 1981, 1986, 1990, 1993, 2001 and 2003. Although these tax reforms were designed with different targets and policy objectives, they defined a common trend in the change of the US tax system consisting in: i) the simplification of the tax code; ii) the decrease in the marginal tax rates; iii) the increasing importance of the Earned Income Tax Credit (EITC) for low-income taxpayers.

So far, the literature on the evaluation of the US tax reforms has mostly dealt with labor supply effects, especially those due to the increased role of the EITC as a public support for poor families and to the concurrent reduction of welfare benefits[1].The most recurring question concerns the labor market participation effects induced by the changes in the EITC.[2] The reason for replacing welfare benefits with an in-work benefit like the EITC has some theoretical foundations: the EITC should generate less distortions on the labor

[1]The 1996 reform of the welfare system abolished the Aid to Families with Dependent Children (AFDC) and replaced it with Temporary Assistance to Needy Families (TANF). See Schoeni and Blank (2000) for a discussion about the accomplishments of the 1996 welfare reform.

[2]The EITC was introduced in 1975 with the dual aim of supporting low income families with children and of stimulating work. The eligibility to the EITC depends indeed on the presence of children and on having positive earned income. The amount of EITC is a non-linear function of the family earned income. Three regions determine the credit schedule: i) the phase-in region in which the credit is increasing in earned income at a certain rate (the phase-in rate); ii) the flat region in which the credit is constant in earned income; iii) the phase-out region in which the credit is decreasing in income at a certain rate (the phase-out rate).

supply than welfare benefits and stimulate labor force participation. As the EITC depends on family income, it is expected to stimulate the labor supply at the extensive margin but, in the phase-out region, it should reduce incentives at the intensive margins. As a matter of fact, empirical studies find that the EITC expansions had heterogeneous effects on the labor supply, depending on the family composition. On the one hand, Dickert et al. (1995), Eissa and Liebman (1996), Ellwood (2000) and Meier and Rosenbaum (2000, 2001) show that the EITC reforms increased the workforce participation of single women with children, since it made work more attractive. On the other hand, Ellwood (2000) and Eissa and Hoynes (2004) find that the EITC expansion between 1984 and 1996 reduced the total family labor supply of married couples. As the EITC is based on family income, it generates adverse effects on participation decisions of the secondary family earners (Dickert et al., 1995) in the phase-out region. Overall, the positive effect on labor force participation seems to be large enough to offset the negative impact at the intensive margin in working hours (Scholz, 1996).

Some researchers focused also on other dimensions. Eissa et al. (2008) evaluated the efficiency effects of the reform. As an efficiency measure they use the excess burden of taxation, i.e. the sum of how much each individual would be willing to pay to get rid of all taxes and transfers. They found that the 1986, 1990, 1993 and 2001 tax reforms each generated relevant efficiency gains, especially through the increased participation of single mothers. Altig and Carlstrom (1999) studied how the change in the marginal tax rates introduced by TRA86 affected income inequality. They found that income inequality became greater with the introduction of TRA86.

Reforms cannot however be evaluated only from these limited perspectives. Most of the applied literature either completely neglects the normative consequences of the reforms or uses conventional measures of individual well-being based on a "reduced" version of the underlying model. For instance, many applied papers report only aggregate labor supply changes or, when it comes to distributional issues, they use disposable income as a proxy of individual well-being. Other papers rely on more sophisticated measures of well-being, such as the equivalent income, but they just consider the underlying preferences of a reference household so that heterogeneity in preferences is assumed away. See, for example, King (1983) and, more recently, Aaberge et al. (2004) and Aaberge and Colombino (2013).

In this paper we rely on indexes that fully account for different individual consumption-leisure preferences. Such indexes have been first introduced by Fleurbaey and Maniquet (2011). They propose an axiomatic construction so that one can clearly see the norma-

tive considerations and the ethical priors behind the ways interpersonal comparisons are performed, while still retaining preference heterogeneity. Fleurbaey and Maniquet (2011) also use the same ideas of fairness to single out ways to aggregate such indexes. Here we are mostly interested in the overall distribution of individual measures of well-being and how this is affected by a change in the fiscal system.

Different ethical priors yield different ways of measuring individual well-being. We intend to use an array of different notions of individual well-being in order to evaluate the effects of TRA86 under different ethical priors. More in detail, for any of the measures of individual well-being we will employ, we will identify the households who gained and who lost from TRA86. To the best of our knowledge, this is the first attempt to evaluate a fiscal reform using different ideas of fairness embedded in the particular measure of individual well-being that is employed. Bargain et al. (2013) and Decoster and Haan (2015) also use well-being indexes that are similar to those which we employ here. They do it however with different purposes: the former paper focuses on a cross-country comparison of well-being, the latter one to studies the sensitivity of different well-being rankings (in Germany) to different normative principles.

In our empirical analysis we use the 1984-1991 Panel Study of Income Dynamics (PSID) and focus on married couples. We find that TRA86 increased the well-being of 99.1% of them. This is essentially due to the fact that TRA86 enlarged the budget constraint for virtually everybody. However, the fraction of households that are better off after TRA86 is slightly smaller among the households that are at the bottom of the well-being distribution: approximatively 97% in the bottom decile and 99.9% in the top decile (this figures are more or less the same across different well-being measures). We therefore provide evidence that TRA86 was close to the goal of being distributionally neutral not only in terms of net income (Feldstein, 2011), but also in terms of well-being, at least for married couples. When analyzing what kind of households were able to climb the social ladder thanks to TRA86, we find that the well-being measure used (and hence the underlying notion of fairness) is important in identifying gainers and losers in terms of the relative position in the welfare ranking. Conclusions about the impact of TRA86 on households' relative position in terms of well-being and, thereby, on inequality cannot be separated from the normative principles inspiring the policy maker takes into account when performing interpersonal comparisons.

Our study is organized as follows. Section 2 describes the tax-benefit system in the US before and after TRA86. Section 3 presents the well-being indexes used to evaluate the impact of TRA86 on welfare. Section 4 deals with the estimation of household prefer-

ences over net income and leisure time, the first step to build indexes of welfare. Section 5 presents the dataset used in the empirical analysis and reports the estimation results of household preferences. Section 6 derives the indexes of welfare metrics under counterfactual scenarios and evaluate the impact of TRA86 on households' well-being. Section 7 concludes.

2   The 1986 Tax Reform Act

One of TRA86's main features was the simplification of the tax code through the drastic reduction in the number of the marginal tax rates' brackets. Their number was reduced from 15 to 5 in 1987 and then to 2 from 1988 until 1990.

Table 1 reports the main features of the federal income tax structure from 1984 until 1990. Before TRA86, the bottom bracket, \$0 - \$3,670, was characterized (for married couples filing jointly) by a 0% marginal tax rate. TRA86 removed this zero tax rate bracket by introducing a marginal tax rate of 15% for couples earning between \$0 and \$30,000. However, it compensated the removal of the zero tax rate bracket by increasing the standard deductions and, above all, the personal exemptions, which doubled in nominal terms between 1984 and 1990.

Another important feature of TRA86 is the important decrease of the highest marginal tax rate. As shown in Table 1, when TRA1986 was fully implemented, the highest marginal tax rate was 28% for earnings larger than \$29,750 in 1988, whereas it was 50% for earnings bigger than \$175,250 in 1986.

Table 1: Federal income tax parameters from 1984 until 1990 (nominal dollars)

	Marginal tax rates (%)		Number of	Bottom	Top	Standard deduction	Personal exemptions
Tax year	Minimum	Maximum	brackets	bracket (\$)	bracket (\$)	married couples (\$)	for the couple (\$)
1984	0.0	50.0	15	$0 - 3,400$	>162, 400	3,400	2,000
1985	0.0	50.0	15	$0 - 3,540$	>169, 020	3,550	2,080
1986	0.0	50.0	15	$0 - 3,670$	>175, 250	3,670	2,160
TRA86							
1987	11.0	38.5	5	$0 - 3,000$	>90, 000	3,760	3,800
1988	15.0	28.0	2	$0 - 29,750$	>29, 750	5,000	3,900
1989	15.0	28.0	2	$0 - 30,950$	>30, 950	5,200	4,000
1990	15.0	28.0	2	$0 - 32,450$	>32, 450	5,450	4,100

Source: Tax Foundation and authors' calculations from the Tax Simulation Model (TAXSIM) of the National Bureau of Economic Research (NBER), available online at http://www.nber.org/taxsim/.

Finally, TRA86 made the EITC more generous. It increased the phase-in rate, de-

creased the phase-out rate and enlarged the maximum tax credit. Table 2, which is extracted from Hotz and Scholz (2003), reports the parameters describing the structure of the EITC from 1984 until 1990. In this time span, the maximum credit and the maximum earnings still entitling to the tax credit almost doubled (in nominal terms). The tax reforms subsequent to TRA86 further consolidated the EITC as an instrument to reduce income tax liabilities of low-income taxpayers and to stimulate workforce participation: it went from \$3.9 billion in 1975 (in 1999 dollars) to \$31.9 billion in 1999 (Hotz and Scholz, 2003) and nowadays it is the largest poverty reduction program in the US.[3]

Table 2: Earned income tax credit from 1984 until 1990 (nominal dollars)

Tax year	Phase-in rate (%)	Phase-in range (\$)	Maximum credit (\$)	Phase-out rate (%)	Phase-out range (\$)
1984	10.00	$0-5,000$	500	12.50	$6,000-10,000$
1985	11.00	$0-5,000$	500	12.22	$6,500-11,000$
1986	11.00	$0-5,000$	500	12.22	$6,500-11,000$
TRA86					
1987	14.00	$0-6,080$	851	10.00	$6,920-15,432$
1988	14.00	$0-6,240$	874	10.00	$9,850-18,576$
1989	14.00	$0-6,500$	910	10.00	$10,240-19,340$
1990	14.00	$0-6,810$	953	10.00	$10,730-20,264$

Source: Figures in this table are extracted from Table 3.1 in Hotz and Scholz (2003).

Figure 1 summarizes, on the left panel, the relation between real earnings and federal income taxes, for married couples, before and after TRA86. Similarly, the right panel summarizes the relation between real earnings and total taxes, federal income, state income and Federal Insurance Contributions Act (FICA) taxes, in California (the most populated state in the US). It is clear that TRA86 reduced the average tax rate of everybody, both couples with kids and without kids. However, not all the types of households are impacted with the same intensity: lower earnings couples did not see their situation changing very much, especially if without kids. The biggest change in taxation involves gross earnings of about \$10,000. If we focus on federal income taxation, married couples enjoy the largest improvement, essentially induced by the larger generosity of the EITC. The change in the number of brackets, the marginal tax rates, the standard deductions and the personal exemptions reduced the average tax rate for couples without kids. However, relatively to couples with kids, their improvement is visibly smaller, both at the level of the federal income taxation and if we consider all the income taxes.

---

[3]Details about the evolution of the US tax system and of the EITC can be found, among others, in Eissa et al. (2008), Eissa and Hoynes (1998, 2004), Hotz and Scholz (2003), Meier and Rosenbaum (2001) and Blundell (2001).

Figure 1: Tax schedules before and after TRA86 by the presence of kids under age 17
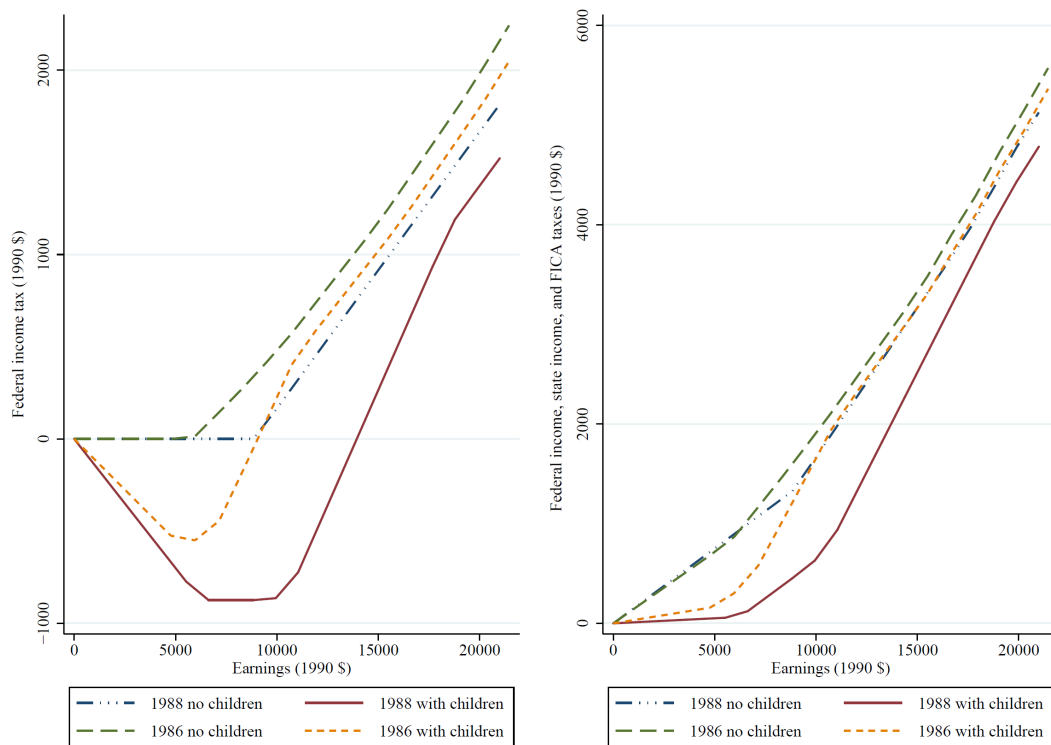
3 Well-being indexes and their normative framework

The study of the effect of a tax reform on the well-being of individuals with heterogeneous preferences and skills (hence different wage rates) is not easy. Usually, in the public economics literature, it implies the use of weighted utilitarian social welfare functions (Fleurbaey, 2008). However this typically relies on the arbitrary cardinalization of the agents' utility functions (Stiglitz, 1987), while the weights cannot be transparently connected to explicit ethical values (Fleurbaey and Maniquet, 2015).

Individuals choose bundles of net income y and leisure l (which we normalize to 1), from their budget constraint (modified by the tax function) so as to maximize their utility. If we observe that individual $i = 1, ..., n$ has chosen the bundle (y_i, l_i), then this bundle is

$$(y_i, l_i) = \arg\max \left\{ U_i(y, l) \mid y \leq w_i(1 - l) - T[w_i(1 - l); \mathbf{z}_i], 0 \leq l \leq 1, y \geq 0 \right\}, \quad (1)$$

where U_i is the utility function that represents her preferences over leisure and consumption and $T(\cdot)$ is the tax function, which transforms income into taxes, given the wage rate w_i[4] and other individual characteristics $\mathbf{z}_i$.[5] Individuals differ from each other because of: i) different preferences; ii) different personal productivity, measured by the wage rate w_i; iii) different initial endowments of personal characteristics $\mathbf{z}_i$.

TRA86 modified the tax function $T(\cdot)$ and hence the budget set of individuals. Our objective is to evaluate how the well-being of individuals was affected, both in absolute and relative terms. In order to do so, we need to define a way (or, different possible ways) to measure individual well-being. We wish to use indexes that fully respect individual (heterogeneous) preferences. Relying on heterogeneous preferences brings about the problem of interpersonal comparisons of utilities. This problem is often solved just by choosing an arbitrary calibration of the utility functions, whereas in our paper the interpersonal comparisons of individual situations relies on well defined value judgements. In the remaining part of this section we provide a formal definition of the family of well-being indexes we use. We also briefly describe the ethical considerations upon which the interpersonal comparison of individual situations rely. This discussion draws heavily from Decoster and Haan (2015) and Fleurbaey and Maniquet (2015). A deeper analysis of the notions we present here (and the corresponding social welfare functions) can be

---

[4]It is worth mentioning that, usually, the policy maker cannot observe w_i but only the earned income $w_i(1 - l)$, this is typically the main argument of the tax function.

[5]The actual tax paid by individual i may also depend on other characteristics such us family composition and capital income. These elements are embodied by the variable z_i.

found in Fleurbaey (2008), Fleurbaey and Maniquet (2011) and Fleurbaey and Blanchet (2013).

The well-being indexes we use are essentially money-metric utility functions. This concept, due originally to Samuelson (1974), can be defined in our framework as the value of the expenditure function for a reference wage rate w and the utility level $U_i(y_i, l_i)$:

$$m_i(w; y_i, l_i) = \min_{\{y,l\}}(y - w(1 - l)) \text{ s.t. } U_i(y, l) \geq U_i(y_i, l_i),$$

Intuitively, this is the amount of money that makes an individual indifferent between the bundle she has chosen and being able to choose from the *implicit* budget set with slope equal to w and intercept equal to $y - w(1 - l)$. A basic normative characteristic of money-metric utilities is that they only rely on ordinal information about individual preferences and they disregard any information about subjective utility (Fleurbaey and Blanchet, 2013). To put it differently, $m_i(w, y_i, l_i)$ is a particular cardinalization of individual preferences that does not rely on the initial cardinalization U_i of individuals preferences R_i, as the same result would be obtained thanks to any increasing transformation of U_i.

More specifically we rely on a family of money-metric utility functions which is based on two parameters. The first one is the reference wage $\widetilde{w}_i$ that is a linear combination of a reference wage common to everybody, $\widetilde{w}$, and the personal wage rate, i.e. $\widetilde{w}_i \equiv \lambda w_i + (1 - \lambda)\widetilde{w}$ with $\lambda \in [0, 1]$. The second one is $\widetilde{h} \in [0, 1]$, a reference value for the fraction of time devoted to work. The family so obtained looks as follows:

$$m_i(\widetilde{w}_i, y_i, l_i) + \widetilde{w}_i\widetilde{h}. \tag{2}$$

The construction of the value of the index in Eq. (2) for an individual $i = 1, ..., n$ for an optimal bundle $(y_i, 1 - l_i)$ is shown in Figure 2 in the space (h, y), with $h \equiv 1 - l$. Given the optimal bundle $(y_i, 1 - l_i)$ the well-being index for individual i is computed by: identifying the bundle $(y_i', 1 - l_i')$ which is the point of tangency between the indifference curve IC_i and the implicit budget line with slope $\widetilde{w}_i$ and then by looking at the value on the ordinate along the implicit budget line in correspondence of the reference working time $\widetilde{h}$. This family of well-being indexes is used in Fleurbaey and Maniquet (2015). They explain in detail how, by properly choosing the values of $\widetilde{w}, \lambda$ and $\widetilde{h}$ one can accommodate different value judgements, all addressed by the literature on responsibility-sensitive egalitarianism, underlying the interpersonal comparisons of individual situations.

Personal outcomes not only depend on endowed circumstances, but also on individual

choices. The basic idea of responsibility-sensitive egalitarianism is that individuals should be considered responsible responsible for the latter, but not for the former. In practice this entails, on the one hand, that inequalities due to endowed circumstances should be removed, this is the the so-called *principle of compensation*. On the other hand, inequalities due only to individual preferences are legitimate as individuals should be held responsible for their goals. This is the so-called principle of *neutrality/responsibility*. Both principles are logically independent and to some extent in conflict with each other (see, among others, Roemer, 1998; Fleurbaey, 2008; Fleurbaey and Maniquet, 2011, 2015).

Figure 2: The construction of well-being indexes
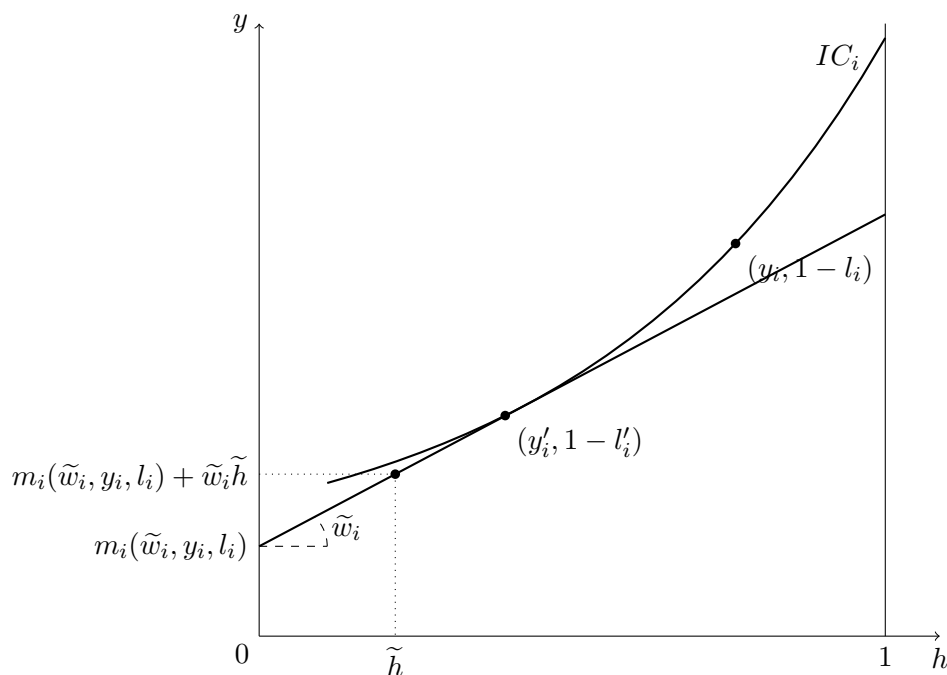
The parameter λ determines the main ethical choice. Indeed, if $\lambda = 1$ then the policy maker gives absolute priority to the principle of compensation: the index so obtained, when comparing two individuals with equal preferences, always considers that the individual on a dominated indifference curve is worse-off and, thereby, gives clear priority to the compensation goal of eliminating inequalities due to skills.

The choice of $\lambda = 0$ means that the social planner gives absolute priority to the principle of neutrality/responsibility: the well-being of any two agents with equal skills

will be the same only if they receive the same lump sum transfer (t) regardless of the shape of their indifference curves (their tastes). This ultimately accommodates the idea that there should not be redistribution between agents differing only in terms of their preference and one should not favor any kind of preferences in particular. To sum up,

1. when $\lambda = 0$ (and $\widetilde{h} = 0$) the index becomes $m_i(\widetilde{w}, y_i, l_i)$, generating compensation-minded indexes;

2. when $\lambda = 1$ the index is equal to $m_i(w_i, y_i, l_i) + w_i\widetilde{h}$, generating responsibility-minded indexes.

Of course, when neutrality is lost ($\lambda = 0$), one has to decide what preferences to favor. This is where the choice of the parameter $\widetilde{w}$ becomes relevant. If $\widetilde{w}$ is low, individuals who are more work-averse tend to obtain lower implicit budgets than individuals who are less averse (the contrary happens if $\widetilde{w}$ is high). This is illustrated in Figure 3b. At a low value of $\widetilde{w}$, individuals that are more work-averse (individual a) tend to be assigned a lower implicit budget. When $\widetilde{w} = 0$ in Figure 3b a is worse off than individual b. When the reference wage is higher, work-lovers tend instead to be assigned higher implicit budget lines: with $\widetilde{w} = \widetilde{w}_1$, individual b is ranked as worse off.

Figure 3: Responsibility and compensation criteria for two individuals, a and b

(a) Responsibility criteria, $U_a = U_b$ (b) Compensation criteria, $U_a \neq U_b$
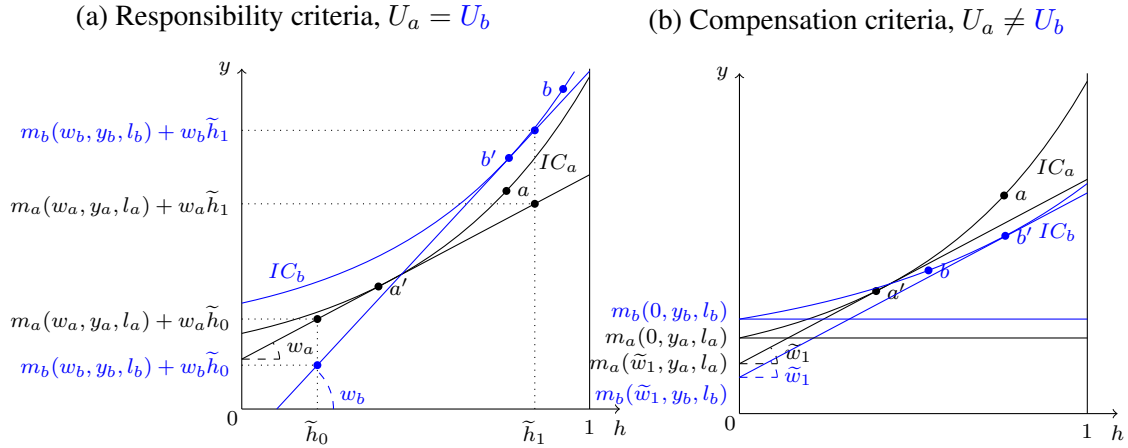
Notes: This figure draws from Figures 4 and 5 in Fleurbaey and Maniquet (2015).

As we have seen earlier, the responsibility-minded indexes are characterized by different slopes of the implicit budget lines. Hence, the individual productivity contributes in determining the well-being level: the higher the wage rate of individual i, the steeper her implicit budget line and the lower the value of $m_i(w_i, y_i, l_i)$. However, the larger is

$\widetilde{h}$, the more the preservation of inequalities due to unequal skills is softened. This means that one can reintroduce some concern for compensation. This can be seen in Figure 3a. Indeed, for low values of $\widetilde{h}$, individual a is better off, although the two individuals are assumed to have the same preferences and individual a reaches a lower indifference curve. The contrary happens when $\widetilde{h}$ is large enough.

4 The econometrics of a structural labor supply model

In order to empirically evaluate the impact of the 1986 tax reform on the well-being of married couples, we proceed in steps. Firstly, we estimate households' preferences for consumption and leisure. Such preferences are heterogeneous with respect to observed and unobserved characteristics. Secondly, we exploit the estimated preferences and the chosen bundles of leisure and consumption to compute several welfare metrics as proposed by Fleurbaey and Maniquet (2011, 2015). Finally, we simulate choices and compute welfare measures before and after the 1986 tax reform so as to study the impact of the change in the tax schedule on the variation of the well-being of married couples.

4.1 Specification of household preferences

In order to estimate the utility functions of agents with heterogeneous preferences (for consumption and leisure) we use a discrete choice structural labor supply model (van Soest, 1995; Hoynes, 1996; van Soest et al., 2002; Blundell and Shephard, 2012). In the 1980s women were typically secondary earners in the US. We focus on married or cohabiting women whose partners work at least 1,500 hours per year (30 hours per week in average). Partners' labor supply is assumed to be exogenously determined. Empirical evidence shows that secondary earners are mostly affected by changes in the tax system, whilst primary earners are subject to a different set of incentives (Triest, 1990; Blundell and Macurdy, 1999). A similar sample selection can be found in the methodological contribution of van Soest et al. (2002), in Eissa and Hoynes (2004) to study the labor force participation response of married couples to changes in the tax reforms in the US between 1984 and 1996, and in Bargain et al. (2013) and Decoster and Haan (2015) to construct welfare orderings of households according to several welfare criteria.

The household's utility depends on leisure l, defined as the time which is not used for work, and on the total net income of the family y. The latter is determined by individuals' labor earnings plus non labor earnings (e.g. social security benefits, welfare

benefits, property income and pensions) less tax paid. Work decisions of married women are determined by their preferences over family net income y and leisure time l. Preferences are heterogeneous across households and they depend on observed characteristics $\mathbf{x}$ and unobserved characteristics $\mathbf{v}$ and ε. The direct utility function is specified as $U(y, l; \mathbf{x}, \mathbf{v}, \varepsilon) = u(y, l; \mathbf{x}, \mathbf{v}) + \varepsilon_l$ and, like in Blundell and Shephard (2012), the vector ε is independent on $(\mathbf{x}, \mathbf{v})$ and is the additive hours-specific error.

The deterministic part of the direct utility function is assumed to have the following Box-Cox form

$$u(y, l; \mathbf{x}, \mathbf{v}) = \phi_y(\mathbf{x}_y, v_y) \frac{y^{\alpha_y} - 1}{\alpha_y} + \phi_l(\mathbf{x}_l, v_l) \frac{l^{\alpha_l} - 1}{\alpha_l}, \tag{3}$$

with $l \equiv 1 - h \in [0, 1]$, where h is wife's fraction of time devoted to work. The Box-Cox specification is often assumed in welfare analysis when estimating utility functions. See, among others, Decoster and Haan (2015), Bargain et al. (2013) and Blundell and Shephard (2012). The functions $\phi_y(\mathbf{x}_y, v_y)$ and $\phi_l(\mathbf{x}_l, v_l)$ determine the impact of observed and unobserved heterogeneity on preferences.[6] They also determine the sign of the marginal utility and, jointly with the parameters α_y and α_l, the concavity of the deterministic part of the utility function. We restrict the power terms α_y and α_l to be between 0 and 1. We can thereby test the monotonicity and concavity conditions on net earnings and leisure by checking that the functions $\phi_y(\mathbf{x}_y, v_y)$ and $\phi_l(\mathbf{x}_l, v_l)$ have positive sign for each households in our sample. The marginal utilities with respect to k, with $k = y, l$, are indeed equal to $\frac{\partial u}{\partial k} = \phi_k(\mathbf{x}_k, v_k) k^{\alpha_k - 1}$, and the second derivatives are $\frac{\partial^2 u}{\partial k \partial k} = \phi_k(\mathbf{x}_k, v_k)(\alpha_k - 1) k^{\alpha_k - 2}$. Cross derivatives are nil. Differently from Blundell and Shephard (2012), who imposed ϕ_y and ϕ_l to be positive, we parametrize them as linear functions in observed and unobserved characteristics:

$$\phi_y(\mathbf{x}_y, v_y) = \mathbf{x}_y' \boldsymbol{\beta}_y + v_y \tag{4}$$

$$\phi_l(\mathbf{x}_l, v_l) = \mathbf{x}_l' \boldsymbol{\beta}_l + v_l. \tag{5}$$

As proposed by van Soest (1995), labor supply is treated as a discrete choice problem.[7] We assume that each family can choose bundles of income and leisure combinations

---

[6]$\mathbf{x}_y$ and v_y are observed and unobserved characteristics, respectively, affecting preferences towards net earnings; similarly, $\mathbf{x}_l$ and v_l are observed and unobserved characteristics, respectively, affecting preferences towards leisure.

[7]See van Soest et al. (2002) fore a discussion on the advantages and limits of the discrete labor supply model with respect to the continuous one.

among a finite set of alternatives $j = 1, \cdots, J$. In our benchmark empirical specification, we set J equal to 6, with wives' yearly discrete working hours points, denoted by H, taking values on $(0, 390, 1020, 1536, 1976, 2500)$.[8] Each of these working hours points corresponds to the empirical median in the following working hours ranges: 0, 1–750, 751–1250, 1251–1750, 1751–2250 and more than 2250.[9] Knowing the wage rate of each wife in our sample one can derive the yearly gross income corresponding to each working hours choice and, after applying the tax-benefit function, we can derive the after tax family income for each H_j. To sum up, each family $i = 1, \cdots, N$, when choosing a certain amount of leisure time l_{ij} with $j = 1, \cdots, J$, obtains the net income y_{ij} and the utility

$$U_{ij}(y_{ij}, l_{ij}; \mathbf{x}, \mathbf{v}, \boldsymbol{\varepsilon}) = u(y_{ij}, l_{ij}; \mathbf{x}, \mathbf{v}) + \varepsilon_j. \tag{6}$$

We assume that the ε_j's are independent and follow an extreme value type I (EV-I) distribution. They are random taste shifters capturing the fact that for a given family tastes may vary across bundles j because of unobservable attributes of the bundles. As pointed out by van Soest (1995) and Bargain et al. (2013), a natural way to interpret the ε_j's is unobserved alternative specific utility components, family's errors in perception of the alternative utilities, optimizations errors, or transitory situations.[10] Because of the independence assumption of the random terms, they cannot be interpreted as reflecting heterogeneous preferences due to unobserved family characteristics (van Soest, 1995). It is worthy to remark that the parametric assumption of the random taste shifters $\boldsymbol{\varepsilon}$ is not restrictive. McFadden and Train (2000) showed indeed that under mild regularity conditions any discrete choice model derived from random utility maximization can be approximated by EV-I errors in combination with a non-parametric specification of unobserved individual characteristics determining preferences.

Each family is supposed to choose the bundle (y_{ij}, l_{ij}) for which U_{ij} is the largest. Under the type I extreme value distribution of the random terms $\boldsymbol{\varepsilon}$, the probability that family i chooses bundle j has a logistic form:

$$\Pr[U_{ij} > U_{ik}, \ \forall \ k \neq j | \mathbf{x}_i, \mathbf{v}_i] = \frac{\exp\left[u(y_{ij}, l_{ij}; \mathbf{x}_i, \mathbf{v}_i)\right]}{\sum_{k=1}^{J} \exp\left[u(y_{ik}, l_{ik}; \mathbf{x}_i, \mathbf{v}_i)\right]}. \tag{7}$$

The probability in (7) is the base to build the sample likelihood function and estimate

---

[8]Leisure time enters the utility function through l, which is wife's fraction of time not devoted to work. More precisely, $l \equiv 1 - h = 1 - H/2500$. Hence, l is equal to 1 if $H = 0$ and is equal to 0 if $H = 2500$.

[9]See Table 4 for the empirical medians of the working hours in each interval.

[10]See Aaberge et al. (1999) for an economic model with this kind of random taste shifters.

the model by maximum likelihood (ML). However, we cannot use this density directly for ML estimation for two reasons. First, we need to know how y_{ij} is determined by l_{ij} for each $j = 1, \cdots, J$ and $i = 1, \cdots, N$. Second, we do not observe $\mathbf{v}_i \equiv (v_y, v_l)$.

The discussion about the determination of y_{ij} for each $j = 1, \cdots, J$ and $i = 1, \cdots, N$ will be addressed in the next subsection. As far as the latter problem is concerned, we integrate out $\mathbf{v}_i$ after imposing a particular parametric assumption on its distribution. This is approximated by a bivariate discrete distribution (Heckman and Singer, 1984) using a fixed number of support points with unknown locations and probability masses (this avoids too strict parametric assumptions on the joint distribution of v_y and v_c). We assume that v_y and v_l have both two points of support with corresponding probability masses defined as follows:

$$
\begin{aligned}
p_1 &\equiv \Pr(v_y = v_{1y}, v_l = v_{1l}) & p_2 &\equiv \Pr(v_y = v_{2y}, v_l = v_{2l}) \\
p_3 &\equiv \Pr(v_y = v_{2y}, v_l = v_{1l}) & p_4 &\equiv \Pr(v_y = v_{1y}, v_l = v_{2l}) = 1 - p_1 - p_2 - p_3.
\end{aligned}
$$

In this case, four points of support and three probability masses are to be estimated. The probabilities associated to the mass points are specified as logistic transforms:

$$
p_m = \frac{\exp(\lambda_m)}{\sum_{r=1}^4 \exp(\lambda_r)} \quad \text{with} \quad \lambda_4 = 0. \tag{8}
$$

Note that v_y and v_l are independent if and only if $p_1 p_2 = p_3 p_4$ (van den Berg et al., 1994; van den Berg and Lindeboom, 1998), making it easy to test for the independence of the unobserved components.

By integrating $\mathbf{v}_i$ out of density (7), we obtain the marginal density

$$
\Pr[U_{ij} > U_{ik}, \forall\, k \neq j | \mathbf{x}_i] = \sum_{m=1}^4 p_m \frac{\exp\left[u(y_{ij}, l_j; \mathbf{x}_i, \mathbf{v}_{im})\right]}{\sum_{k=1}^J \exp\left[u(y_{ik}, l_k; \mathbf{x}_i, \mathbf{v}_{im})\right]}. \tag{9}
$$

4.2 Budget constraint and tax simulation

As said, we need to know how y_{ij} is determined by l_{ij} for each $j = 1, \cdots, J$ and $i = 1, \cdots, N$. Different choices of the number of wives' working hours will lead to different before tax labor incomes and therefore different after tax net incomes. In the empirical analysis we will use data from PSID from 1984 to 1990, which contains information on working time, income (labor income, social security benefits, welfare benefits, property income and pensions) and spouses and family characteristics. In order to understand how

wife's choice of working hours affect after tax family income, we need to:

1. Calculate wives' before tax hourly wage, so that we can compute yearly labor incomes for each discrete working hour point H. Using the information in the PSID dataset, we can compute the average hourly wage of each workers. However, if the wife is not working, the information is missing. For non-workers, we need then to predict the before tax wage rate. In a panel data setting, we estimate log-wage equations for wives accounting for sample selection. Estimates of the wage equation are presented in Appendix A. The variables used to predict wives' wages are age, race, education, parents' education, time dummies and the unemployment rate in the state of residence.[11] As suggested by Hoynes (1996), to maintain a consistent stochastic specification, predicted hourly wages from the estimated equation are used both for worker and non-workers.[12]

2. Compute the family budget constraint per each discrete working hour point. The budget constraint of family i at each discrete $j = 1, \cdots, J$ is

$$y_{ij} = w_i H_j + E_i + V_i - T(w_i H_j, E_i, V_i, \mathbf{z}_i, t), \quad (10)$$

where w_i is wife's predicted hourly wage, E_i is husband's labor earnings, V_i is non-labor income (e.g. property income, interests, pensions, veteran benefits, social security benefits).[13] $T(w_i H_j, E_i, V_i, \mathbf{z}_i, t)$ is the tax function which determines the tax (subsidy) that the family has to pay (receive) for the fiscal year t at both federal and state level. The tax-benefit schedule is a function of labor and non-labor income but also of $\mathbf{z}_i$ which is a vector of family characteristics. It includes the state of residence and the number of dependent children. We assume that all married couples file jointly the tax declaration. The function $T(w_i H_j, E_i, V_i, \mathbf{z}_i, t)$ is numerically

---

[11]In the selection equation we include also the number of children younger than 17, an indicator for the presence of children younger than 6 and the number of family members.

[12]In order to properly take into account of measurement error in wages, we should simultaneously estimate the log wage equation and the parameters of the utility function (van Soest, 1995; van Soest et al., 2002; Blundell, 2001). However, we can derive the net family income from the gross income through an external tax simulator. Hence, we cannot integrate the tax-benefit simulation in each iteration of the ML estimation and we are forced, as in Hoynes (1996), Bargain et al. (2013) and Decoster and Haan (2015), to stick to a reduced form method for imputing wages to non-workers which is distinct from the structural model of labor supply.

[13]We consider social security benefits as independent on wives' working hours choice. In each year of our longitudinal dataset, less than 1% of the families received social security benefits. From 1984 until 1990, only 2.4% of the families (61 out of 2,529) received social security benefits in at least one year in which they were in the sample.

calculated using the Tax Simulation Model (TAXSIM) of the National Bureau of Economic Research (NBER).[14]

4.3 Maximum likelihood estimation

Equation (9) represents the contribution to the likelihood function of family i when it chooses H_j. In our empirical analysis, we will operate in a panel data framework. We assume that the bundle choices of each family are independent over time once we control for observed and unobserved characteristics determining preference heterogeneity. We also assume that families' preferences are time constant. So far, we have not indexed variables by the time indicator for the sake of clarity. We start now using the time subscript. Denote by g_{it} the variable indicating the bundle chosen by family i at time t and taking values on $\{1, \cdots, J\}$. The contribution to the likelihood function of family i is given by

$$
\begin{aligned}
\mathscr{L}_i &= \sum_{m=1}^{4} p_m \left\{ \prod_{t=1}^{T} \left[\prod_{j=1}^{J} \frac{\exp\left[u(y_{itj}, l_j; \mathbf{x}_{it}, \mathbf{v}_{im})\right]}{\sum_{k=1}^{J} \exp\left[u(y_{itk}, l_k; \mathbf{x}_{it}, \mathbf{v}_{im})\right]} \right]^{\mathbb{1}[g_{it}=j]} \right\} \\
&= \sum_{m=1}^{4} p_m \left\{ \prod_{t=1}^{T} \left[\prod_{j=1}^{J} \frac{\exp\left[\phi_y(\mathbf{x}_{it,y}, v_{m,y})\frac{y_{itj}^{\alpha_y}-1}{\alpha_y} + \phi_l(\mathbf{x}_{it,l}, v_{m,l})\frac{l_j^{\alpha_l}-1}{\alpha_l}\right]}{\sum_{k=1}^{J}\exp\left[\phi_y(\mathbf{x}_{it,y}, v_{m,y})\frac{y_{itk}^{\alpha_y}-1}{\alpha_y} + \phi_l(\mathbf{x}_{it,l}, v_{m,l})\frac{l_k^{\alpha_l}-1}{\alpha_l}\right]} \right]^{\mathbb{1}[g_{it}=j]} \right\} (11)
\end{aligned}
$$

where $\mathbb{1}[\cdot]$ is the indicator function, equal to 1 when its argument is true. The log-likelihood function sums the logarithm of expression (11) over all the families and its parameters $\boldsymbol{\beta}_y$, $\boldsymbol{\beta}_l$, α_y, α_l, $\mathbf{v}_y$, $\mathbf{v}_l$ and $(\lambda_1, \lambda_2, \lambda_3)$ are estimated by ML.

The set of observed characteristics in both $\mathbf{x}_{it,y}$ and $\mathbf{x}_{it,l}$ are wife's and husband's age, a non-white indicator, wife's and husband's education (dummies for up to grade 8, from grade 9 to 11, grade 12, tertiary education without diploma and tertiary education with diploma) and indicators for the health condition of both spouses. The marginal utility with respect to leisure is supposed to be affected also by the presence of dependent children, so we included into $\mathbf{x}_l$ also the number of dependent children below 18 and a dummy indicator for the presence of children younger than 6. The variables for the presence of children, health and age are time-varying. Variables for education and race are time constant.

---

[14]We used the Internet Taxsim (v9). See Feenberg and Coutts (1993) for an introduction to TAXSIM. Further information can be found in Internet at http://www.nber.org/taxsim/. TAXSIM has already been used in the empirical studies of Eissa et al. (2008), Bargain et al. (2011) and Bargain et al. (2013).

4.4 Identification

In our labor supply model we impose a particular shape on the household utility function by assuming a Box-Cox form. This direct parametrization of the utility function certainly contributes to model identification. It is not however a necessary assumption for model identification. van Soest et al. (2002) have indeed shown how a structural labor supply model with a flexible non-parametric specification (series expansion in net income and hours) of the direct utility function can be used for the analysis of all sorts of tax-benefit changes. Moreover, in our framework, we exploit exogenous variation at different levels as further sources of model identification, leading us to a quasi-experimental set-up:

- Nonlinearities and discontinuities related to some family characteristics like the state of residence and the number of dependent children. This means that, at any point in time, families with the same before tax income but different $\mathbf{z}_{it}$ or different levels of nonlabor income will face different tax rates and will thereby have different after tax income. Variation across states in income tax and EITC is particularly important, as pointed out by Hoynes (1996) and Eissa and Hoynes (2004).

- Variations over time of the tax schedule with policy reforms both at state and federal level. Our longitudinal data cover the period from 1984 to 1990 and we can also take advantage of the important change in the tax schedule at federal level generated by TRA86.

- In our panel data setting, we observe families multiple times in making bundle choices. This is of help in the nonparametric identification of the joint distribution of unobserved characteristics.

5 Estimation results of the structural labor supply model

5.1 Data and sample

The data we use come from the 1984-1991 PSID.[15] At that time interviews were conducted annually on about 6,900–9,400 households each year. The PSID collects data on a wide set of social, demographic, economic and health topics. Information on income

---

[15]The PSID is produced and distributed by the Institute for Social Research, Survey Research Center, University of Michigan, Ann Arbor, MI.

from several sources and working hours is retrospectively collected: information on income at time t is collected at time $t+1$. Hence, our final dataset covers demographic, labor market and income information for the tax-years 1984–1990 and contains only families interviewed for at least two consecutive waves.

The sample includes married couples with the husband working at least 1,500 hours in a year and the wife between 20 and 55 years of age. We dropped families with one spouse working more 4,000 hours in a year, with missing state of residence, with composition changes over time in the husband or wife position, with negative non-labor income and with one spouse retired, disabled, in jail or studying. The resulting sample is made up of 2,542 different households with more than 1,900 couples interviewed per year. Column (1) of Table 3 displays the number of household observations in the original PSID dataset, whilst column (2) reports the number of observations after applying these selection criteria. The sample in column (2) is the one used to predict wives' hourly wages. After predicting wives' hourly wages and simulating the net income which corresponds to the discrete hour point of the interval actually chosen, we removed from the sample households in the last percentile of the pooled net income distribution (9 households removed). We also deleted households with nil net income when the wife is supposed not to work (4 households removed). The final sample used for the estimation of the utility function is made up of 2,529 different families. After pooling all 7 years from 1984 until 1990, the resulting sample size is 13,275 households. Detailed information on the size of the sample used for the estimation of the utility function is reported in column (3) of Table 3.

Table 3: Number of household observations from the original PSID dataset to the final selected sample

Year	Original PSID sample (1)	Sample used for predicting wives' hourly wages (2)	Final sample after removing outliers in predicted net consumption (3)
1984	6,918	1,953	1,934
1985	7,032	1,938	1,918
1986	7,018	1,945	1,929
1987	7,061	1,917	1,897
1988	7,114	1,874	1,849
1989	7,114	1,896	1,864
1990	9,371	1,912	1,884
# of different families (N)	11,774	2,542	2,529
# of pooled observations (NT)	51,628	13,435	13,275

Table 4 reports the distribution of wives' working hours. About 21% of wives in our sample do not work. The remaining married women are almost equally split into those working full-time (more than 1,750 hours per year, in average 35 hours per week) and

those working part-time (less than 1,750 hours per year). The top-left panel of Figure 4 displays the working hours distribution. The top-right panel is instead the distribution of married women's predicted hourly wages. The bottom panels of Figure 4 focus on wives' gross labor income and family yearly after tax income.[16] Table 5 reports the descriptive statistics of the variables used in the estimation of the utility function. Married women are on average 35.5 years old, 2.5 years younger than their husbands. There are 1.5 dependent children (17 years old or younger) in each married couple and in 68% of the couples there is a child younger than 6. One quarter of the couples are non-white. Almost 9% of the married women declare to be in bad health, whilst the same figure for men is about 7.4%. Married men are more likely to have a diploma of tertiary diploma (24.7%) than their wives (18.7%). However, married women are less likely than married men to have a low educational attainment: while 17.9% of the husbands stopped at grade 11 or less, only 14.7% of married women stopped at grade 11 or less. In comparing the distribution of these demographic characteristics between husbands and wives, it must be reminded that we selected families with the husband working at least 1,500 hours per year.

Table 4: Working hours distribution of married women

Yearly working hours	Absolute frequency	Relative frequency	Mean	Median
0 hours	$2,746$	0.207	0.000	0.000
(0, 750] hours	$1,732$	0.131	386.912	390.000
(750, 1250] hours	$1,428$	0.108	$1,013.692$	$1,020.000$
(1250, 1750] hours	$2,050$	0.154	$1,528.397$	$1,536.000$
(1750, 2250] hours	$4,354$	0.328	$1,978.260$	$1,976.000$
(2250, 4000] hours	965	0.073	$2,589.285$	$2,500.000$
Total	$13,275$	1.000	$1,232.610$	$1,476.000$

5.2   Estimation results

Table 6 reports the estimation results of the utility function in Eq. (3). We find that it is important to account for the presence of unobserved heterogeneity:[17] the log-likelihood function improves from $-22,291.6$ without unobserved heterogeneity to $-18,052.4$ with unobserved heterogeneity specified as explained in Subsection 4.1. We also find that the unobserved components determining preferences in leisure and the unobserved ones determining preferences in income are positively and significantly correlated.

Since we constrained α_l and α_y to be between 0 and 1, the marginal utilities are positive and the concavity conditions are satisfied if and only if $\phi_y(\mathbf{x}_y, v_y) > 0$ and

---

[16]The net income is deflated to 1990 prices using the consumer price index by the US Bureau of Labor

Figure 4: Distribution of wives' working hours, their wage rates, yearly gross income and family net income
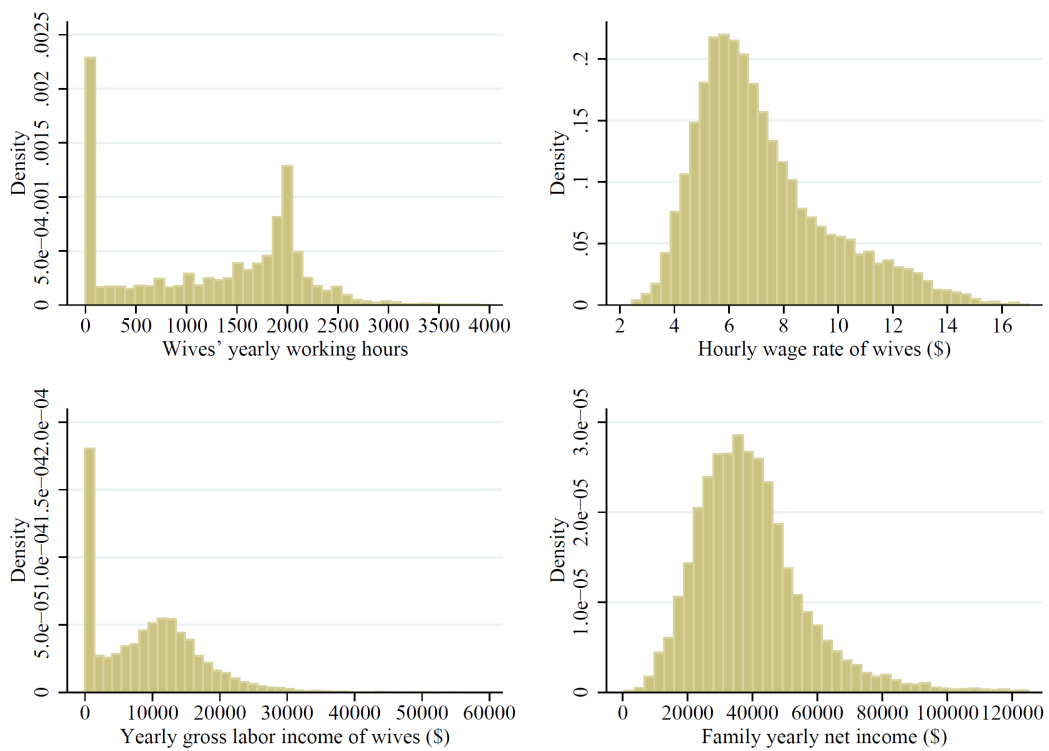
Table 5: Summary statistics

	Mean	Std. Dev.	Minimum	Maximum
Predicted family net income ($)	39,416.84	16,737.61	673.3	125,114.0
Age of wife	35.538	8.031	20	55
Age of husband	37.968	8.690	20	67
Number of kids below 18 years	1.473	1.132	0	4
Presence of kids below 6 years	0.680	0.864	0	4
Non-white	0.246	0.430	0	1
Education of wife				
Less than grade 9	0.032	0.177	0	1
Grades [9, 11]	0.115	0.319	0	1
Grade 12	0.437	0.496	0	1
Tertiary education without diploma	0.228	0.420	0	1
Tertiary education with diploma	0.187	0.390	0	1
Education of husband				
Less than grade 9	0.048	0.214	0	1
Grades [9, 11]	0.131	0.337	0	1
Grade 12	0.346	0.476	0	1
Tertiary education without diploma	0.228	0.420	0	1
Tertiary education with diploma	0.247	0.431	0	1
Wife in bad health	0.089	0.285	0	1
Husband in bad health	0.074	0.261	0	1
# of pooled observations (NT)		13,275		

$\phi_l(\mathbf{x}_l, v_l) > 0$. The utility function and its first and second derivatives depend however not only on observed characteristics but also on unobservables. Since we have estimated the distribution function of the unobservables, we can derive the mixed utility function and check whether the corresponding first and second derivatives satisfy the monotonicity and the concavity conditions. The estimated mixed utility function is

$$\sum_{m=1}^{4} \widehat{p}_m u(y, l; \mathbf{x}, \widehat{\mathbf{v}}_m) = \sum_{m=1}^{4} \widehat{p}_m \widehat{\phi}_y(\mathbf{x}_y, \widehat{v}_{my}) \frac{y^{\widehat{\alpha}_y} - 1}{\widehat{\alpha}_y} + \sum_{m=1}^{4} \widehat{p}_m \widehat{\phi}_l(\mathbf{x}_l, \widehat{v}_{ml}) \frac{l^{\widehat{\alpha}_l} - 1}{\widehat{\alpha}_l}. \quad (12)$$

Hence, the monotonicity and concavity conditions are satisfied if, for each $i = 1, \cdots, N$ and $t = 1, \cdots, T$,

$$\sum_{m=1}^{4} \widehat{p}_m \widehat{\phi}_y(\mathbf{x}_{it,y}, \widehat{v}_{my}) > 0 \quad \text{and} \quad \sum_{m=1}^{4} \widehat{p}_m \widehat{\phi}_l(\mathbf{x}_{it,l}, \widehat{v}_l) > 0.$$

We find that the monotonicity and concavity conditions are satisfied for every $i = 1, \cdots, N$

Statistics and available at ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt.

[17]We do not report the estimation result of the model without unobserved heterogeneity for the sake of brevity. They are available from the authors upon request.

Table 6: Estimated parameters of the Box-Cox utility function with unobserved heterogeneity

Variables	Leisure Coeff.		Std. Err.	Net income Coeff.		Std. Err.
Age of wife/10	0.446	***	0.116	−0.007		0.025
Age of husband/10	0.023		0.107	0.051	**	0.024
Number of kids below 18 years/10	−0.049		0.230	–		–
Presence of kids below 6 years	0.865	***	0.054	–		–
Non-white	0.381	***	0.100	0.118	***	0.024
Education of wife - Reference: up to grade 8						
Grades [9, 11]	0.490	*	0.263	0.054		0.066
Grade 12	1.720	***	0.243	0.410	***	0.062
Tertiary education without diploma	2.458	***	0.257	0.516	***	0.066
Tertiary education with diploma	3.884	***	0.277	0.703	***	0.073
Education of husband - Reference: up to grade 8						
Grades [9, 11]	−0.023		0.251	0.079		0.059
Grade 12	−0.228		0.234	0.034		0.054
Tertiary education without diploma	−0.264		0.242	0.050		0.056
Tertiary education with diploma	−0.472	*	0.249	−0.049		0.056
Wife in bad health	−0.435	***	0.155	−0.195	***	0.037
Husband in bad health	−0.122		0.146	−0.040		0.034
α_l	0.487	***	0.006	–		–
α_y	–		–	0.859	***	0.017
Unobserved heterogeneity support points						
v_1	−1.388	***	0.396	−0.174	*	0.094
v_2	3.642	***	0.403	2.040	***	0.160
Unobserved heterogeneity probability weights						
λ_1	0.459	***	0.062			
λ_2	0.155	**	0.062			
λ_3	−2.135	***	0.137			
λ_4	0.000		–			
Resulting unobserved heterogeneity probability masses§						
p_1	0.409	***	0.007			
p_2	0.302	***	0.007			
p_3	0.031	***	0.001			
p_4	0.258	***	0.011			
Log-likelihood	−18,052.4					
# of parameters	37					
AIC/NT	14.306					
NT	13,275					
N	2,529					
Corr(v_y, v_l)	0.494					
LR test for independence of v_y and v_l, $H_0: v_y \perp v_l$	$\chi^2(1) = 498.6$, p-value $= 0.000$					

Notes: *** Significant at 1%; ** significant at 5%; * significant at 10%.
§ The standard errors of the unobserved heterogeneity probability masses are estimated by the delta method.

and $t = 1, \cdots, T$.

By simply looking at the estimated parameters reported in Table 6, it is not easy to understand the impact of the covariates on the utility function and, above all, on the marginal rate of substitution of income with leisure. We therefore employ the estimated parameters to compute, by way of microsimulations, the income variation needed to compensate a marginal variation in working hours (or leisure). The simulated marginal rate of substitution is computed per each observation and then averaged across the pooled sample. We also estimate the average partial effect of selected characteristics on the marginal rate of substitution. The simulated marginal rate of substitution and the average partial effects are reported in Table 7. We describe the simulation algorithm to obtain the figures in Table 7 in Appendix B. Across our sample of married couples, a reduction of 50 hours of work (increase of 50 hours of leisure) of the wife must be compensated, in order to remain on the same indifference curve, by a decrease in the net income of $504. The estimated average partial effects show that the reduction in net income is significant and large if there are children in prescholar age in the household and if the wife is in bad health. *Ceteris paribus*, if in the family there is a kid younger than 6, the extra reduction in the net income to compensate a reduction of 50 hours of work amounts to $46.6. If the wife is in bad health, it amounts to $41.4. Hence, the presence of young children and wife's bad health conditions remarkably increase wife's distaste for work. Also husbands' bad health condition and wife's age have a significant and positive impact on distaste for work, but the effect is small in size. When the wife turns one (ten) year(s) older, the marginal rate of substitution becomes larger in absolute value by $2.4 ($24.2). Wives in non-white families, with an older husband and with a husband with a diploma of tertiary education are instead more work lovers and their indifference curves are flatter.

Another way to understand what the estimated parameters of the structural labor supply mean in terms of labor supply behaviour is to simulate labor supply elasticities. This is done by increasing the net income by 1%, re-predicting wives' labor supply and looking at its change. Table 8 reports simulated labor supply elasticities both at the intensive and the extensive margins for the whole sample, but also by selected characteristics.[18] Intensive elasticities, which measure the per cent change in working hours when net income increases by 1%, are large. Across the whole sample, a 1% increase in net income implies an overall increase in working hours of 0.55%, corresponding to an increase of 6.65 yearly working hours on average, and an increase in the employment rate by 0.21% (0.14 percentage points on average). These estimated elasticities are in line with those found by

─────────────────────

[18]The simulation algorithm with regard to labor supply elasticities is described in Appendix B.

Table 7: Marginal rates of substitution: Yearly income variation to compensate a yearly decrease of 50 hours of work

	Marginal rate of substitution (in $1,000)	95% confidence interval[§]	
Whole sample	−0.5043	−0.7721	−0.3458
Average partial effects of selected characteristics on the marginal rate of substitution			
Age of wife (+ 1 year)	−0.0024	−0.0015	−0.0032
Age of wife (+ 10 years)	−0.0242	−0.0150	−0.0369
Age of husband (+ 1 year)	0.0014	0.0005	0.0027
Age of husband (+ 10 years)	0.0136	0.0051	0.0263
Number of kids below 18 years (+ 1)	0.0002	−0.0000	0.0026
Presence of kids below 6 years	−0.0466	−0.0365	−0.0580
Non-white	0.0159	0.0044	0.0370
Wife with tertiary education with diploma	−0.0044	−0.0016	0.0779
Husband with tertiary education with diploma	0.0062	0.0005	0.0276
Wife in bad health	−0.0414	−0.0173	−0.0878
Husband in bad health	−0.0063	−0.0003	−0.0249

[§] The 95% confidence intervals are computed from the predicted marginal rates of substitution by sampling 999 times the vector of parameter estimates assuming that the estimator is Normally distributed around the point estimates with a variance-covariance matrix equal to the estimated one.

Eissa (1995), who analysed the responsiveness of married women's labor supply to the 1986 tax reform.[19]

5.3   Goodness-of-fit

In Section 6 we will exploit the estimated parameters of the utility function to simulate the choice of net income-leisure bundles of families under a counterfactual scenario in order to isolate the effect of TRA86. As the reliability of the simulations depends on the capability of our discrete choice structural labor supply model to predict the realized household choices between net income and leisure, we first report goodness-of-fit checks of the estimated model. The goodness-of-fit statistics are constructed on the basis of simulations of 9,999 bundle choices for each family in the sample and for each year the family remains in the sample. Since we replicate the simulations 9,999 times, we can construct 95% confidence intervals of the predicted frequencies of working hours and check whether the predicted frequencies are close enough to the empirical ones. The simulation algorithm with regard the goodness-of-fit is in Appendix B.

The first panel of Table 9 contrasts the empirical frequencies of the yearly working hours discrete categories with the simulated ones across the whole time window 1984–

---

[19]Eissa (1995) focused on labor supply elasticities at the top of the income distribution and found that, the higher the income, the more sensitive the labor supply behaviour. This might explain why our estimates of the labor supply elasticities are somewhat smaller than those in Eissa (1995).

Table 8: Simulated labor supply elasticities

Characteristics	Yearly variation in total hours	Intensive elasticity[§]	Variation in employment rate[†]	Extensive elasticity[‡]	# of pooled observations
Whole sample	6.649	0.549	0.141	0.207	13,275
Children between 0 and 6	6.922	0.604	0.165	0.245	5,166
Children between 6 and 17	6.690	0.547	0.142	0.210	4,765
No children	6.165	0.480	0.104	0.149	3,344
Wife in bad health	6.724	0.587	0.161	0.246	1,187
Wife in good health	6.640	0.546	0.139	0.204	12,088
Wife younger than 35	6.744	0.566	0.155	0.228	7,346
Wife older than 35	6.529	0.530	0.124	0.182	5,929
With tertiary education	5.876	0.438	0.085	0.117	5,515
Without tertiary education	7.261	0.643	0.182	0.279	7,760
White	6.487	0.537	0.144	0.212	10,015
Non-white	7.141	0.588	0.133	0.195	3,260

[§] Labor supply intensive elasticities measures the percentage point increase in labor supply when net income increases by 1%.
[†] The variation in employment rate measures the change in level of the employment rate when net income increases by 1%.
[‡] Labor supply extensive elasticities is the percentage point increase in employment rates when net income increases by 1%.

1990. The estimated model overpredict the number of married wife who are not working and underpredict the number working less than 750 hours per year. We instead obtain very good fit of the frequencies for each category of married women working more than 750 hours per year.

The bottom panel of Table 9 reports the model fit for two particular years, the first one (1984) and the last one (1990) of our time window. The aim is to check whether the estimated model is able to replicate the actual data over time although time enters the model specification in a very limited way (only through the wage equation and time-varying regressors).[20] The goodness-of-fit of the frequencies in 1984 and 1990 are in line with the model fit on the full sample.

6   Simulating the effects of TRA86

We use the estimated parameters of the utility function to simulate the well-being level of families in 1986, i.e., just before the application of TRA86, and compare it with the well-being level they would have reached in the counterfactual scenario where the new fiscal system is already operative. More in detail, for each family, we predict the 1986 net income-leisure choice using alternatively the actual 1986 fiscal rules and the new 1988

---

[20]We also tried to include time dummies in the linear preference terms of leisure and net income. Since their coefficients were not jointly significant at the usual 5% statistical level and the goodness-of-fit did not show any relevant improvement, we removed the time dummies from the set of covariates to save in degrees in freedom.

Table 9: Goodness-of-fit: Empirical and predicted frequencies, 1984–1990

Yearly working hours categories	Actual frequencies[†]	Predicted frequencies[‡]	95% confidence interval[§]	
Pooled from 1984 until 1990				
0 hours	**0.207**	0.280	0.194	0.431
(0, 750] hours	0.130	0.047	0.023	0.083
(750, 1250] hours	**0.108**	0.087	0.046	0.127
(1250, 1750] hours	**0.154**	0.177	0.096	0.217
(1750, 2250] hours	**0.328**	0.314	0.200	0.470
(2250, 4000] hours	**0.073**	0.096	0.070	0.129
1984				
0 hours	**0.215**	0.299	0.208	0.473
(0, 750] hours	0.147	0.058	0.029	0.095
(750, 1250] hours	**0.118**	0.105	0.055	0.145
(1250, 1750] hours	**0.161**	0.185	0.098	0.237
(1750, 2250] hours	**0.292**	0.272	0.185	0.432
(2250, 4000] hours	**0.067**	0.082	0.057	0.112
1990				
0 hours	0.193	0.288	0.202	0.451
(0, 750] hours	0.121	0.050	0.021	0.094
(750, 1250] hours	**0.102**	0.096	0.047	0.144
(1250, 1750] hours	**0.151**	0.189	0.097	0.241
(1750, 2250] hours	**0.342**	0.305	0.200	0.474
(2250, 4000] hours	**0.091**	0.090	0.063	0.123

[†] Empirical frequencies lying in the 95% confidence intervals of the simulated frequencies are in bold.

[‡] The predicted frequencies are calculated using the empirical dataset and the maximum likelihood estimates of the parameters of the labor supply structural model presented in Section 4.1.

[§] The 95% confidence intervals are computed from the predicted frequencies by sampling 9,999 times the vector of parameter estimates assuming that the estimator is Normally distributed around the point estimates with a variance-covariance matrix equal to the estimated one.

fiscal system to go from gross to net incomes. Then, we compute the indirect utilities and the well-being level, using some of the metrics described before, attained by each family under the actual and the counterfactual scenario. We can thereby compute the fraction of families which are better off with the new fiscal system. Finally, in order to understand whether the fiscal reform might have differently affected the *poorest* and the *richest*, depending on the different welfare metrics we consider, we compute the fraction of families which are better off by selected percentiles of the chosen welfare measure and we study what families gained or lost relatively to the other members of the population. Appendix C describes the simulation algorithm to calculate the ranking according to different welfare metrics under the actual 1986 fiscal law and the counterfactual scenario.

6.1 Compensation and responsibility criteria in 1986

In Table 10 we report the correlation matrix of the 1986 well-being rankings, computed using six different indexes. The first index is just the net income: the higher the net income the higher the position in the ranking. Then, we computed the distribution of individual well-being using two compensation minded indexes: we set $\lambda = 0$ and $\widetilde{h} = 0$, using as the reference wage $\widetilde{w}$ zero and the average wage rate in 1986, respectively. Finally, we computed three responsibility-minded indexes by setting $\lambda = 1$ and the reference value for the fraction of time devoted to work, $\widetilde{h}$, equal to 1, 0.5 and 0.

The two compensation criteria, which attribute importance to the wage rate in determining the ranking, are highly correlated to the net income. The responsibility criterion with $\widetilde{h} = 1$ also tends to rank first people with high wage rates. As a matter of fact, we find that it is highly correlated with the compensation criteria (linear correlation larger than 0.990). When $\widetilde{h}$ decreases and tends to 0, the correlation between the responsibility and the compensation criteria decreases. However, it remains fairly large: the correlation between the responsibility criterion with $\widetilde{h} = 0$ and the compensation criterion with $\widetilde{w} = 0$ is 0.894. This might be due to the fact that preferences over net income and working time were not so heterogeneous in the US in the 1980s. Although in Table 6 we noted that individual characteristics are statistically highly significant in determining the profile of the utility function, the resulting indifference maps for each individual might not be so different from each other. In what follows, we continue our analysis by sticking to the two compensation criteria and the responsibility criteria with $\widetilde{h} = 0$ and $\widetilde{h} = 0.5$.

The different indexes generate a ranking of households from the worst off to the better off. For each household and for each welfare metric, we compute the natural logarithm of the position in the 1986 ranking and, then, regress it on the wage rate and the same set

Table 10: Correlation matrix of welfare metrics

Criterion	Net income	Compensation criteria		Responsibility criteria		
		$\widetilde{w} = 0$	$\widetilde{w} = 1986$ mean wages	$\widetilde{h} = 1$	$\widetilde{h} = 0.5$	$\widetilde{h} = 0$
Net income	1.000					
Compensation $\widetilde{w} = 0$	0.996	1.000				
Compensation $\widetilde{w} = $ mean	0.996	0.998	1.000			
Responsibility $\widetilde{l} = 1$	0.998	0.997	0.999	1.000		
Responsibility $\widetilde{l} = 0.5$	0.965	0.983	0.981	0.973	1.000	
Responsibility $\widetilde{l} = 0$	0.854	0.894	0.886	0.869	0.960	1.000

Notes: The number of observations is equal to $1,929$, i.e. the number of households observed in our sample in 1986.

of characteristics entering the utility function. Table 11 report the Ordinary Least Squares (OLS) estimation results. They reflect some of the normative considerations expressed in Section 3. *Ceteris paribus*, the wage rate has the strongest effect on compensation criteria: one more dollar of wage rate increases the ranking position by about 19.3–19.7%. In the responsibility criteria, the wage rate plays a less important role, which is increasing in $\widetilde{h}$. With $\widetilde{h} = 0.5$, one more dollar of wage rate increases the ranking position by 9.8%. When $\widetilde{h} = 0$, the one dollar increase in the wage rate decreases the position in the ranking by 7.7%. Hence, with the compensation criteria, individuals with high wage rates are more likely to be considered as better off and, thereby, if the social planner has a redistributive attitude, more likely to pay more taxes. With the responsibility criteria, inequalities due to unequal wage rates are instead less of a priority: with the extreme case of $\widetilde{h} = 0$, individuals with higher wage rates are more likely to be considered as worse off, *ceteris paribus*, and therefore a social planner using this criterion for the redistributive purposes would have a kind of laissez-faire attitude with respect to workers' productivity.

We also look at the impact of the other individual characteristics on the ranking position. We note that wife's age, presence of kids and husband's education are positively associated to the ranking position of all the welfare metrics. The presence of kids younger than 6 does not affect the position in the ranking of the compensation criterion with $\widetilde{w} = 0$, whilst it increases the position in the ranking from all the other metrics by about 12–13%. When $\widetilde{w}$ is low, individuals that are more work averse tend to obtain lower value of the corresponding welfare metric (Fleurbaey and Maniquet, 2015). We have seen in Subsection 5.2 that the presence of kids in prescholar age increases the distaste for work. This explains why when we move from a welfare metric that better treat work averse individuals to welfare metrics that pay better attention to individuals with a low work aversion, the presence of kids younger than 6 increases the position in the welfare ordering. Fi-

Table 11: OLS estimates of the log ranking equation in the welfare metrics in 1986

Variables	Compensation criterion $\tilde{w} = 0$ Coeff.		S.E.[§]	Compensation criterion $\tilde{w} = $ mean Coeff.		S.E.[§]	Responsibility criterion $\tilde{h} = 0.5$ Coeff.		S.E.[§]	Responsibility criterion $\tilde{h} = 0$ Coeff.		S.E.[§]
Wage rate	0.193	***	0.025	0.197	***	0.025	0.098	***	0.030	−0.077	**	0.035
Age of wife/10	−0.009		0.057	0.089		0.055	0.064		0.063	0.058		0.071
Age of husband/10	0.244	***	0.050	0.204	***	0.049	0.224	***	0.056	0.226	***	0.066
Number of kids below 18	0.110	***	0.021	0.109	***	0.022	0.127	***	0.022	0.133	***	0.022
Presence of kids below 6	−0.038		0.051	0.138	***	0.050	0.123	**	0.056	0.132	**	0.057
Non-white	−0.319	***	0.047	−0.338	***	0.048	−0.356	***	0.050	−0.350	***	0.051
Education of wife - Reference: up to grade 8												
Grades [9, 11]	−0.106		0.162	0.023		0.160	0.052		0.166	0.087		0.143
Grade 12	0.089		0.151	0.110		0.152	0.083		0.165	0.063		0.149
Tertiary without diploma	0.103		0.159	0.168		0.158	0.111		0.183	0.047		0.175
Tertiary with diploma	−0.233		0.202	−0.180		0.196	−0.061		0.245	−0.027		0.253
Education of husband - Reference: up to grade 8												
Grades [9, 11]	0.760	***	0.164	0.621	***	0.155	0.695	***	0.167	0.592	***	0.146
Grade 12	1.068	***	0.160	0.929	***	0.153	1.019	***	0.164	0.900	***	0.141
Tertiary without diploma	1.268	***	0.159	1.114	***	0.153	1.210	***	0.165	1.084	***	0.144
Tertiary with diploma	1.345	***	0.158	1.209	***	0.151	1.373	***	0.165	1.390	***	0.150
Wife in bad health	−0.046		0.069	0.052		0.063	0.051		0.068	0.037		0.072
Husband in bad health	−0.280	***	0.092	−0.240	***	0.087	−0.218	**	0.091	−0.153	*	0.086
Constant	3.249	***	0.266	3.036	***	0.288	3.597	***	0.267	4.859	***	0.239
R-squared	0.375			0.375			0.269			0.178		
# of households (N)				1,929								
# of parameters				17								

Notes: *** Significant at 1%; ** significant at 5%; * significant at 10%.
[§] Standard errors are robust to heteroskedasticity.

nally, non-white individuals are worse off according to all the metrics and their rank is about 30% lower than white individuals. Since they are ranked worst in the responsibility criteria, especially when $\widetilde{h} = 0$, non-white households would be better treated by a social planner seeking for redistribution inspired by a laissez-faire attitude.

6.2 Absolute gainers from TRA86

In this subsection we actually compare, at household level, the well-being level attained in 1986 under the old fiscal system with the counterfactual well-being level attained if the 1988 fiscal set-up was in place. By doing so, we can understand the fraction of households that experienced an increase in well-being because of the change of the fiscal system. Table 12 reports the fraction of families who would have been strictly better under the 1988 fiscal system, both in terms of net income and in terms of well-being. We can see that almost all the households experienced an increase in both dimensions. This is due to the fact that TRA86 enlarged the budget constraint almost for everybody. However, a small fraction of households (1.3%) experienced an increase in well-being jointly with a reduction in their net income: the EITC expansion introduced by TRA86 reduced the labor supply of these married women so much that their households experienced a decrease in net income but an increase in well-being. In terms of labor supply, we find that TRA86 had a positive effect on labor force participation: on average the fraction of time devoted to work increased by about 2 percentage points, which amounts to a yearly increase of 80 hours of work.

Table 12: Gainers from TRA86 according to net income and utility levels

	Fraction of families who are strictly better off	95% confidence interval§		Observations
Net income	0.995	0.992	0.998	$1,929$
Utility	0.991	0.986	0.995	$1,929$
Fraction of those not experiencing an income increase among those experiencing a utility improvement	0.013	0.008	0.019	$1,913$

§ The 95% confidence intervals are computed from the predicted frequencies by sampling 999 times the vector of parameter estimates assuming that the estimator is Normally distributed around the point estimates with a variance-covariance matrix equal to the estimated one.

Although almost all the households experienced an increase in well-being (99.1%), it might be interesting to understand what kind of families were less likely to experience it. We simulated the income-leisure choices and the indirect utility 999 times per each

household, each time by sampling the vector of parameter estimates assuming that the estimator is Normally distributed around the point estimates with a variance-covariance matrix equal to the estimated one, so as to take into account the variability of the estimated parameters. Across the 999 simulations, we computed the average of a dummy indicator equal to one if the household experienced an increase in utility, which can be interpreted as the probability of experiencing an increase in utility. Then, we regress it on the wage rate and the set of characteristics determining observed preference heterogeneity. Table 13 reports the OLS estimation results. The higher the wage rate, the lower the probability of an increase in well-being. This might be due to the fact that the expansion of the EITC favored especially low income and low skilled individuals. Older households, households with kids both below and above 6 years of age, households with the wife in bad health and households with a better educated husband are more likely to experience an increase in utility. Non-white households are less likely to experience an increase in well-being.

Table 13: OLS regression of the probability of gaining in terms of utility from TRA86 on family characteristics

Variables	Coeff.		Std. Err.§
Wage rate ($)	−0.0014	***	0.0004
Age of wife/10	0.0024	**	0.0011
Age of husband/10	0.0017	*	0.0010
Number of kids below 18 years	0.0009	**	0.0004
Presence of kids below 6 years	0.0033	***	0.0011
Non-white	−0.0038	***	0.0011
Education of wife - Reference: up to grade 8			
Grades [9, 11]	−0.0029		0.0020
Grade 12	−0.0042	**	0.0018
Tertiary without diploma	−0.0012		0.0021
Tertiary with diploma	−0.0049	*	0.0028
Education of husband - Reference: up to grade 8			
Grades [9, 11]	0.0038		0.0028
Grade 12	0.0068	**	0.0027
Tertiary without diploma	0.0073	***	0.0028
Tertiary with diploma	0.0109	***	0.0027
Wife in bad health	0.0022	**	0.0011
Husband in bad health	−0.0008		0.0016
Constant	0.9783	***	0.0037
R^2		0.0757	
Breusch-Pagan test		$\chi^2(1) = 500.8$	
for heteroskedasticity		p-value $= 0.000$	
# of households (N)		1,929	
# of parameters		17	

Notes: *** Significant at 1%; ** significant at 5%; * significant at 10%.
§ Standard errors are robust to heteroskedasticity.

Finally, we checked whether the probability of an increase in well-being differs across

the different well-being distributions generated by the different indexes we use. Table 14 reports the fraction of families who are better off by selected percentiles of the ranking of each welfare metric. The figures are quite stable across the percentiles of different welfare metrics: families in the first decile of the distribution of each welfare metric are those who are less likely to experience an increase in well-being (about 97%). Moving across the percentiles, the probability of an increase in well-being monotonically increases and, in the last decile, it amounts to 99.9%. We therefore provide evidence that TRA86 was close to the goal of being distributionally neutral in terms of well-being, at least for married couples, beside being designed to be revenue neutral (Feldstein, 2011).

Table 14: Gainers from TRA86 by percentiles according to selected compensation and responsibility criteria

	Fraction of families who are strictly better off	95% confidence interval[§]		Observations
		Compensation criteria		
		$\widetilde{w} = 0$		
Between 0th and 10th percentiles	0.973	0.948	0.995	193
Between 10 and 50th percentiles	0.988	0.978	0.995	772
Between 50 and 90th percentiles	0.997	0.992	1.000	771
Between 90 and 100th percentiles	0.999	0.995	1.000	193
		$\widetilde{w} = 1986$ mean wages		
Between 0th and 10th percentiles	0.971	0.943	0.990	193
Between 10 and 50th percentiles	0.988	0.978	0.995	772
Between 50 and 90th percentiles	0.997	0.992	1.000	771
Between 90 and 100th percentiles	0.999	0.995	1.000	193
		Responsibility criteria		
		$\widetilde{h} = 0.5$		
Between 0th and 10th percentiles	0.969	0.943	0.990	193
Between 10 and 50th percentiles	0.988	0.979	0.995	772
Between 50 and 90th percentiles	0.997	0.992	1.000	771
Between 90 and 100th percentiles	0.999	0.995	1.000	193
		$\widetilde{h} = 0$		
Between 0th and 10th percentiles	0.969	0.943	0.990	193
Between 10 and 50th percentiles	0.988	0.979	0.995	772
Between 50 and 90th percentiles	0.997	0.992	1.000	771
Between 90 and 100th percentiles	0.999	0.994	1.000	193

[§] The 95% confidence intervals are computed from the predicted frequencies by sampling 999 times the vector of parameter estimates assuming that the estimator is Normally distributed around the point estimates with a variance-covariance matrix equal to the estimated one.

6.3   Relative gainers from TRA86

In this subsection we analyze what kind of households were able to climb the social ladder after the introduction of TRA86. In other words, we will shed light on what type of households gained relatively to the others. Table 15 reports descriptive statistics about

the shuffle of households' positions in the (different) well-being rankings when switching from the 1986 fiscal set-up to the 1988 tax system. Panel (a) of Table 15 displays the fraction of households climbing the social ladder thanks to TRA86. About 40% of the married couples have gained positions in the welfare ranking. This is roughly the same for the four welfare metrics we consider.

If we disaggregate these statistics by selected percentiles for each different welfare metric distribution, we do not find noticeable differences. The compensation minded indexes show a somewhat smaller fraction of relative gainers in the top decile. Panel (b) of Table 15 focuses instead on the average number of positions gained by households in selected percentiles of each welfare metric distribution. Both for compensation and responsibility minded indexes, the worst off (below the median of the corresponding welfare metric distribution) gained positions on average, whilst the better off lost positions. When moving from the compensation to the responsibility criteria, this feature becomes more pronounced and reaches its peak with the responsibility criterion with $\widetilde{h} = 0$. In this respect we could say that TRA86 had strongest distributive effects in the responsibility-minded perspective as compared to the compensation-minded perspective. Finally, if we compare the two compensation-minded indexes and focus on the difference between the average position gain across percentiles, we realize that TRA86 generated a stronger well-being redistribution if the society opts for a welfare metric treating better the work averse individuals ($\widetilde{w} = 0$).

Table 15: Gainers from TRA86 in terms of relative welfare ranking

	Compensation criteria		Responsibility criteria		
	$\widetilde{w} = 0$	$\widetilde{w} = 1986$ mean wages	$\widetilde{h} = 0.5$	$\widetilde{h} = 0$	Observations
(a) Fraction of households gaining positions in the welfare ranking					
Overall	0.396	0.413	0.430	0.399	$1,929$
Selected percentiles of the welfare metric					
Between 0th and 10th percentiles	0.451	0.466	0.456	0.435	193
Between 10 and 50th percentiles	0.382	0.403	0.421	0.389	772
Between 50 and 90th percentiles	0.393	0.406	0.418	0.396	771
Between 90 and 100th percentiles	0.409	0.430	0.487	0.420	193
(b) Average position gain in the welfare ranking					
Overall min. and max.[§]	$[-67, 117]$	$[-50, 102]$	$[-69, 110]$	$[-97, 245]$	$1,929$
Selected percentiles of the welfare metric					
Between 0th and 10th percentiles	0.503	0.140	0.513	0.984	193
Between 10 and 50th percentiles	0.415	0.306	0.426	1.611	772
Between 50 and 90th percentiles	-0.472	-0.326	-0.459	-1.647	771
Between 90 and 100th percentiles	-0.275	-0.062	-0.383	-0.850	193

[§] We do not report the average of the overall average gain in positions in the welfare ranking since it is 0 by construction.

What are the characteristics of the households climbing the social ladder thanks to TRA86? To answer this question, we computed for each household, the ranking position in each welfare metric before and after TRA86 and calculated its variation. We regressed the ranking position variation on the wage rate and the characteristics determining observed preference heterogeneity. Table 16 reports the OLS estimation results of the equation for the variation in the welfare ranking position.

Two points are worth mentioning. First, TRA86 favored people with higher wage rates regardless of which index we consider. However the magnitude of this effect changes considerably across the different indexes. If society gives priority to the reduction of inequalities due to unequal skills favoring work averse individuals (compensation criterion with $\widetilde{w} = 0$), one more dollar in hourly wage translates into 0.97 positions gained in the welfare ranking (out of $1,929$ observations) with the introduction of TRA86. If the society instead opts for a responsibility-minded attitude, one more dollar in the wage rate would imply a gain of 16.8 positions. Hence, depending on the ethical priors, we get different conclusions about the impact of TRA86 on the well-being of high skilled married women.

Second, whilst according to all the welfare metrics the presence of kids below 18 has a positive effect on the ranking position, essentially because of the larger generosity of the EITC, the presence of kids in pre-scholar age differently affects the variation in the position in the welfare ranking depending on the welfare metric. For the responsibility minded indexes and the compensation minded index with $\widetilde{w}$ equal to the mean, the presence of a kid below 6 years of age reduces the ranking position by about 5 positions after TRA86 is introduced. If one considers the compensation minded index with $\widetilde{w} = 0$, the impact of the presence of kids in scholar age is instead positive and equal to about 0.8 positions. Hence, the responsibility minded indexes and the compensation minded index tend not to favor too much work averse individuals. We thereby conclude that TRA86 worsened the relative well-being of households with kids in pre-scholar age.

Finally, if one considers the compensation minded indexes, families with younger wives and non-white gained less than families with older wives and white. This age and racial gap in the effect of TRA86 disappears if one uses the responsibility minded indexes.

7 Conclusions

In this paper we evaluate the effect of TRA86, the US fiscal reform of 1986, on the well-being of married couples. First, we estimated a structural model of labor supply in which

Table 16: Estimation results of the equation for the variation in the positions in the welfare ranking from TRA86

Variables	Compensation criterion $\tilde{w} = 0$		Compensation criterion $\tilde{w} = 1986$ mean wage		Responsibility criterion $\tilde{h} = 0.5$		Responsibility criterion $\tilde{h} = 0$	
	Coeff.	S.E.§	Coeff.	S.E.§	Coeff.	S.E.§	Coeff.	S.E.§
Wage rate ($)	0.973 **	0.447	1.281 ***	0.443	6.513 ***	0.584	16.763 ***	0.977
Age of wife/10	1.960 **	0.908	−0.223	0.914	−1.174	1.211	−1.490	1.687
Age of husband/10	−1.357	0.845	−0.709	0.857	−2.037 *	1.150	−1.437	1.601
Number of kids below 18	3.745 ***	0.414	3.721 ***	0.402	2.986 ***	0.492	4.202 ***	0.628
Presence of kids below 6	0.873	0.879	−5.058 ***	0.870	−5.771 ***	1.088	−5.119 ***	1.463
Non-white	−2.222 ***	0.685	−1.384 **	0.690	0.275	0.914	0.229	1.362
Education of wife - Reference: up to grade 8								
Grades [9, 11]	0.785	1.735	0.578	1.820	0.421	2.535	1.045	3.477
Grade 12	0.413	1.769	1.487	1.844	1.532	2.429	0.453	3.351
Tertiary without diploma	1.027	2.059	0.266	2.150	1.684	2.728	−2.308	3.817
Tertiary with diploma	−3.142	3.025	−7.861 **	3.079	−3.624	3.850	−8.871	5.485
Education of husband - Reference: up to grade 8								
Grades [9, 11]	−0.660	1.299	1.730	1.319	−2.029	1.833	−2.618	2.527
Grade 12	−0.520	1.403	2.736 *	1.421	−2.626	1.857	−1.378	2.550
Tertiary without diploma	−2.148	1.463	1.793	1.467	−3.984 **	1.892	−2.350	2.541
Tertiary with diploma	0.059	1.605	2.982 *	1.591	−6.142 ***	2.042	−1.832	2.891
Wife in bad health	1.546	1.166	−1.636	1.255	−1.009	1.547	−0.416	2.280
Husband in bad health	0.370	1.350	−0.363	1.339	1.462	1.915	0.604	2.636
Constant	−12.842 ***	3.136	−9.615 ***	3.164	−30.989 ***	4.364	−102.669 ***	6.807
R-squared	0.070		0.069		0.260		0.597	
# of households (N)			1,929					
# of parameters			17					

Notes: *** Significant at 1%; ** significant at 5%; * significant at 10%.
§ Standard errors are robust to heteroskedasticity.

households have preferences over net income and leisure that are allowed to be heterogeneous in observed and unobserved characteristics. Second, we used the estimated preferences to predict the optimal bundles of leisure and net income for 1986 households before TRA86 and in the counterfactual scenario where TRA86 was in place. Third, on the basis of households' choices under the actual and counterfactual scenarios, we computed the well-being level of each household using several indexes, as proposed by Fleurbaey and Maniquet (2011, 2015), which are based on different ethical priors, i.e., they reflect different ethical preferences of the social planner. Finally, we analyzed how many and which type of households experienced an increase in well-being, both in absolute and relative terms.

Our approach to the evaluation of a tax reform is original. So far, the literature on the evaluation of the US tax reforms has mostly dealt with labor supply effects. Some researchers focused also on other dimensions. Eissa et al. (2008) evaluated the efficiency effects measured by the excess burden of taxation, i.e. the aggregation of how much the individual would be willing to pay to get rid of all taxes and transfers. Altig and Carlstrom (1999) studied how the change in the marginal tax rates introduced by TRA86 affected income inequality. Most of the applied literature typically overlooks normative aspects when analysing a reform. An explicit definition of normative criteria for interpersonal comparisons is necessary if one wish to meaningfully asses individual situations.

TRA86 changed the tax scheduled and increased the generosity of the EITC so that virtually everybody benefited of a reduced average tax rate and thereby of an expansion of the budget set. As a matter of fact, we found that more than 99% of married couples experienced an increase in well-being, with a higher net income and an increase in labor force participation of married women by 2 percentage points on average. When we looked at the probability of experiencing an increase in well-being by selected percentiles of the well-being distributions, we found that families in the first decile of the distribution, for each welfare metric, are less likely to experience an increase in well-being with the introduction of TRA86, although their probability of a well-being increase is fairly large and about 97%.

When it comes to the household characteristics determining the impact of TRA86 on the ability of households to climb the social ladder, according to different welfare metrics, we show that families with highly skilled wives gained more, especially if the social planner has a responsibility minded attitude. Moreover, the stronger is such attitude, or the lower the attention paid to work averse individuals, the bigger the loss in terms of positions in the well-being ranking for households with kids in pre-scholar age. Finally,

whilst using compensation minded indexes older and white families gained most, with responsibility minded indexes, these age and race gaps disappear.

We conclude that the introduction of TRA86 increased the well-being of almost every couple in the US, independently on the type of welfare metric. However, depending on the normative considerations behind the different ways of measuring well-being, we arrive at different verdicts about what type of households gained most, relatively to the other members of society. More precisely, if the social planner has a compensation minded attitude and tends to favor work averse individuals, then older and white families are those who gained most from the reform. If the social planner has a responsibility minded attitude, then families with higher skilled wives and without kids in pre-scholar age are those who gained most from the reform.

Appendix

A   Sample selection corrected estimation of log-wage equation

In order to account for nonworkers' missing wages, log wage equations for the wives are estimated taking into account sample selection bias. We first tested for sample selection following Wooldridge (1995).[21] The log wage equation is

$$\ln w_{it} = \mathbf{x}_{it,w}\boldsymbol{\beta}_w + c_i + u_{it}, \quad t = 1, \cdots, T, \tag{A.1}$$

where w_{it} is observed only if wife i is working, $\mathbf{x}_{it,w}$ is the set of exogenous regressors determining wages at time t, c_i is time-invariant unobserved characteristics and u_{it} is an idiosyncratic error term. Denote by s_{it} the binary indicator equal to 1 if woman i is working at time t and 0 otherwise. Assume that, for each t, s_{it} is determined by the following probit model

$$s_{it} = \mathbb{1}[\mathbf{x}_{it,s}\boldsymbol{\beta}_s + \bar{\mathbf{x}}_{i,s}\boldsymbol{\delta}_s + v_{it}], \quad v_{it}|\mathbf{x}_{it,s} \sim \text{Normal}(0, 1), \tag{A.2}$$

where $\mathbf{x}_{it,s}$ contains $\mathbf{x}_{it,w}$ and $\bar{\mathbf{x}}_{i,s}$ is the average of $\mathbf{x}_{it,s}$ over t and represents the Mundlak (1978) approach to allow the unobservables to be correlated with some elements of the observed characteristics in a discrete choice panel data model.

The parameters of the model in Equation (A.2) are estimated by pooled probit and reported in the first columns of Table A.1. We can test for sample selection by computing the inverse Mills ratio obtained from the pooled probit estimation and by plugging it into the log wage equation as a further regressor. Under the null assumption of no sample selection, the inverse Mills ratio should

---

[21]The notation in this section is inspired by the one in Wooldridge (2010, Subsection 19.9.2).

not be significant when the log wage equation is estimated by fixed effects (FE). The bottom line of Table A.1 displays a Wald test for the significance of the inverse Mills ratio after the FE estimation of the log wage equation. The test is made robust to heteroskedasticity and serial correlation. As the inverse Mills ratio is highly significant, a problem of sample selection is detected and should be corrected for.

In order to correct for sample selection, we run pooled ordinary least square regression on the log wage equation augmented by the inverse Mills ratio, its interaction with the time dummies and $\bar{\mathbf{x}}_{i,s}$. The estimation results of the coefficients of the log wage equation are reported in the last three columns of Table A.1. These estimated coefficients are used to predict wages for both working and non-working wives.

B The simulation algorithm with regard to marginal rates of substitution, labor supply elasticities and the goodness-of-fit

We start by describing the simulation algorithm for the goodness-of-fit, since it is the starting point of the algorithms for the estimation of the marginal rates of substitution and labor supply elasticities.

The simulation algorithm with regard to the goodness-of-fit

The simulation algorithm with regard to the goodness-of-fit proceeds according to the following steps:

1. Draw a vector of parameter estimates $\widehat{\Theta}$ assuming normality around the point estimates with a variance-covariance matrix equal to the estimated one. This ensures that the Monte Carlo confidence intervals encompass the parameter estimation precision.

2. Assign to each family the observed explanatory variables, observed family net income for each wife's leisure choice and a vector of unobserved characteristics drawn with probabilities given by Equation (8).

3. For each bundle $j = 1, \cdots, 6$ and each $i, \cdots, N$, compute the predicted deterministic part of the utility function $u(y_{ij}, l_j | \mathbf{x}_i, \widehat{\mathbf{v}}_i; \widehat{\Theta})$ and thereby the predicted cumulative distribution function from the theoretical one implied by the type I extreme value distribution of the

Table A.1: Estimated parameters of the employment selection equation and sample selection corrected log-wage equation

Variables	Probit selection equation Coeff.		Std. Err.	Corrected log-wage equation Coeff.		Std. Err.§
Constant	0.957	***	0.311	0.301		0.263
Age	0.051	***	0.015	0.085	***	0.013
Age square	−0.001	***	0.000	−0.001	***	0.000
Race – Reference: White						
Black	0.245	***	0.034	0.020		0.033
Other	0.253	***	0.082	0.032		0.071
Education of wife - Reference: up to grade 5						
Grades [6, 8]	−0.454	***	0.143	−0.276	**	0.134
Grades [9, 11]	−0.270	**	0.128	−0.247	**	0.103
Grade 12	0.100		0.126	−0.135		0.095
Tertiary education without diploma	0.271	**	0.128	0.019		0.099
Tertiary education with diploma	0.451	***	0.130	0.335	***	0.102
Education of father - Reference: up to grade 5						
Grades [6, 8]	−0.027		0.049	0.014		0.044
Grades [9, 11]	−0.025		0.059	0.054		0.052
Grade 12	0.022		0.054	0.088	*	0.046
Tertiary education without diploma	−0.026		0.070	0.112	*	0.066
Tertiary education with diploma	−0.092		0.068	0.133	**	0.063
Education of mother - Reference: up to grade 5						
Grades [6, 8]	0.239	***	0.061	0.045		0.066
Grades [9, 11]	0.126	**	0.063	−0.004		0.068
Grade 12	0.137	**	0.062	0.077		0.065
Tertiary education without diploma	0.032		0.077	0.055		0.076
Tertiary education with diploma	0.059		0.081	0.007		0.079
Time dummies – Reference: 1984						
1985	−0.051		0.049	0.039		0.044
1986	−0.061		0.054	0.097	*	0.050
1987	−0.029		0.064	0.124	**	0.055
1988	0.024		0.076	0.191	***	0.060
1989	0.080		0.087	0.281	***	0.066
1990	−0.005		0.099	0.310	***	0.074
State unemployment rate	2.647		1.670	−1.306	*	0.789
Number of kids below 18 years	−0.015		0.046	–		–
Presence of kids below 6 years	−0.281	***	0.050	–		–
Number of household members	−0.070		0.044	–		–
Time average of age square	−0.000		0.000	−0.000		0.000
Time average of state unemployment rate	−9.856	***	1.861	0.295		1.251
Time average of the number of kids below 18 years	−0.131	**	0.058	−0.100	***	0.033
Time average of presence of kids below 6 years	−0.430	***	0.066	0.121	**	0.061
Time average of the number of household members	0.107	*	0.055	−0.002		0.029
Inverse Mills ratio	–		–	−0.261		0.200
Inverse Mills ratio*1985	–		–	0.108		0.128
Inverse Mills ratio*1986	–		–	−0.032		0.140
Inverse Mills ratio*1987	–		–	0.010		0.151
Inverse Mills ratio*1988	–		–	−0.098		0.153
Inverse Mills ratio*1989	–		–	−0.351	**	0.172
Inverse Mills ratio*1990	–		–	−0.280	*	0.163
Log-likelihood	−6,429.3			−10,416.8		
# of parameters	35			39		
Pseudo R² / R²	0.072			0.195		
NT	13,435			10,597		
N	2,542			2,305		
F-test for sample selection	$F(1, 2304) = 17.42$, p-value $= 0.000$					

Notes: *** Significant at 1%; ** significant at 5%; * significant at 10%.

§ The standard errors are computed by bootstrapping the results 1,000 times (individual-cluster bootstrap with replacement).

random term ε and the structural model of labor supply:

$$
\begin{aligned}
F(U_{ij}|\mathbf{x}_i, \mathbf{v}_i) &= \sum_{r=1}^{j} \Pr[U_{ir} > U_{ik}, \ \forall \, k \neq r | \mathbf{x}_i, \mathbf{v}_i] \\
&= \sum_{r=1}^{j} \frac{\exp\left[u(y_{ir}, l_r; \mathbf{x}_i, \mathbf{v}_i)\right]}{\sum_{k=1}^{J} \exp\left[u(y_{ik}, l_k; \mathbf{x}_i, \mathbf{v}_i)\right]}.
\end{aligned}
\tag{B.3}
$$

For $i = 1, \cdots, N$, we draw κ_i from a standard uniform distribution. Define $\widehat{F}(U_{i0}|\mathbf{x}_i, \widehat{\mathbf{v}}_i) = 0$. If $\widehat{F}(U_{ij-1}|\mathbf{x}_i, \widehat{\mathbf{v}}_i) \leq \kappa_i < \widehat{F}(U_{ij}|\mathbf{x}_i, \widehat{\mathbf{v}}_i)$, family i is predicted to choose bundle j of net income and leisure.

4. Repeat steps 1 to 3 $R = 9,999$ times to get R independent realizations and build Monte Carlo confidence intervals.

The simulation algorithm with regard to the marginal rates of substitution

The simulation algorithm with regard to the marginal rates of substitution is very similar to the one for the goodness-of-fit. The difference is that after step 3, i.e. after simulating the bundle choice of family i, we computed the deterministic part of the utility which corresponds to that bundle choice. Finally, for each i we calculated the variation in the net income needed to reach the same level of predicted deterministic utility if leisure would increase by 50 hours per year (decrease of 50 hours of work). Finally, we averaged across the sample to get the average marginal rate of substitution and we repeated the procedure 999 times to build Monte Carlo confidence intervals.

The average partial effects in Table 7 of selected characteristics on the marginal rate of substitution are computed by taking the variation in the simulated marginal rate of substitution when we let one particular covariate vary. If the covariate is a dummy variable, we measure the variation in the marginal rate of substitution when the covariate changes from 0 to 1. When the covariate has a continuous support (wife's age, husband's age, or the number of children), we look at the variation when it increases by one.

The simulation algorithm with regard to the labor supply elasticity

The labor supply elasticity predicted by the model is computed by looking at the change in the predicted discrete working hour points generated by a 1% increase in the net income corresponding to each discrete working hour point. We modified the simulation algorithm for the goodness-of-fit and replaced step 3 and 4 with the following steps:

3′ For each bundle $j = 1, \cdots, 6$ and each $i, \cdots, N$, compute the predicted deterministic part of the utility function $u(y_{ij}, l_j | \mathbf{x}_i, \widehat{\mathbf{v}}_i; \widehat{\Theta})$ and the counterfactual utility in case of a

1% increase in net income $u(1.01 \times y_{ij}, l_j | \mathbf{x}_i, \widehat{\mathbf{v}}_i; \widehat{\Theta})$. Using the theoretical cumulative distribution function in Equation (B.3), predict cumulative distribution function without the 1% increase in net income, $\widehat{F}(U_{ij} | \mathbf{x}_i, \widehat{\mathbf{v}}_i)$, and with the 1% increase in net income, $\widehat{F}(\tilde{U}_{ij} | \mathbf{x}_i, \widehat{\mathbf{v}}_i)$. For $i = 1, \cdots, N$, we draw κ_i from a standard uniform distribution. Define $\widehat{F}(U_{i0} | \mathbf{x}_i, \widehat{\mathbf{v}}_i) = \widehat{F}(\tilde{U}_{i0} | \mathbf{x}_i, \widehat{\mathbf{v}}_i) = 0$. If $\widehat{F}(U_{ij-1} | \mathbf{x}_i, \widehat{\mathbf{v}}_i) \leq \kappa_i < \widehat{F}(U_{ij} | \mathbf{x}_i, \widehat{\mathbf{v}}_i)$, family i is predicted to choose bundle j of net income and leisure. If $\widehat{F}(\tilde{U}_{ij-1} | \mathbf{x}_i, \widehat{\mathbf{v}}_i) \leq \kappa_i < \widehat{F}(\tilde{U}_{ij} | \mathbf{x}_i, \widehat{\mathbf{v}}_i)$, family i is predicted to choose bundle j when the net income increases by 1%. On the basis of the predicted working hours without and with the 1% increase in net income, we calculate different measures of labor supply elasticity both at the extensive and the intensive margins by averaging across the sample (or across subgroups conditional on selected observed characteristics).

4′ Repeat steps 1, 2 and 3′ 999 times to get 999 independent realizations and build Monte Carlo confidence intervals.

C Calculation of welfare metrics: compensation and responsibility criteria

In order to understand the effect of TRA86 on households' choices and, thereby, on the attained well being level and on the position in the well-being ranking, we predicted twice the choices of the households in our sample in 1986 using the first 3 steps of the simulation algorithm used for the goodness-of-fit in Subsection B: first, using the actual 1986 tax rule to go from gross to net incomes, getting the optimal bundle (l^*, y^*); secondly, using the 1988 tax rule to transform 1986 gross incomes into net incomes, getting the optimal bundle (l', y'). By substituting these bundles into the estimated deterministic utility functions, we get the indirect utilities attained in the actual and counterfactual scenarios.

Calculation of the compensation criterion with $\widetilde{w} = 0$

Once we get the indirect deterministic utilities from the first three steps of the simulation algorithm for the goodness-of-fit, the computation of the compensation criterion with $\widetilde{w} = 0$ for individual i boils down to the calculation of the vertical intercept in the space (h, y) of the indifference curve reached by individual i (see Figure 3b), i.e.

$$m_i^*(\widetilde{w} = 0, y_i^*, l_i^*) = \left[1 + \frac{\widehat{\alpha}_l}{\widehat{\phi}_y} \widehat{u}(y^*, l^*) \right]^{1/\widehat{\alpha}_l} \quad \text{for the actual tax scenario;} \quad (C.4)$$

$$m_i'(\widetilde{w} = 0, y_i', l_i') = \left[1 + \frac{\widehat{\alpha}_l}{\widehat{\phi}_y} \widehat{u}(y', l') \right]^{1/\widehat{\alpha}_l} \quad \text{for the counterfactual tax scenario.} \quad (C.5)$$

Calculation of the compensation criterion with $\widetilde{w} > 0$

To get the compensation criterion with $\widetilde{w} > 0$, we must first identify the bundle (l_0, y_0) located along the indifference curve attained by individual i with optimal choice (l^*, y^*) (or (l', y') in the counterfactual scenario), where the indifference curve has slope equal to $\widetilde{w}$. In other words, we have to solve a system of two equations for y_0 and l_0: the marginal rate of substitution between l and y equal to $\widetilde{w}$ and the indirect utility at the optimal bundle equal to the usual formula in Equation (3) for (l_0, y_0):

$$
\begin{cases}
\dfrac{\widehat{\phi}_l}{\widehat{\phi}_y} \dfrac{l_0^{\widehat{\alpha}_l - 1}}{y_0^{\widehat{\alpha}_y - 1}} = \widetilde{w} & \text{(C.6)} \\[2ex]
\widehat{u}(y^*, l^*) = \widehat{\phi}_y \dfrac{y_0^{\widehat{\alpha}_y} - 1}{\widehat{\alpha}_y} + \widehat{\phi}_l \dfrac{l_0^{\widehat{\alpha}_l} - 1}{\widehat{\alpha}_l}. & \text{(C.7)}
\end{cases}
$$

Solving Equation (C.6) for y_0 and substituting into Equation (C.7) yield

$$
\frac{\widehat{\phi}_y}{\widehat{\alpha}_y} \left[\left(\frac{\widehat{\phi}_l}{\widehat{\phi}_y} \frac{l_0^{\widehat{\alpha}_l - 1}}{\widetilde{w}} \right)^{\frac{\widehat{\alpha}_y}{\widehat{\alpha}_y - 1}} - 1 \right] + \widehat{\phi}_l \frac{l_0^{\widehat{\alpha}_l} - 1}{\widehat{\alpha}_l} - \widehat{u}(y^*, l^*) = 0, \tag{C.8}
$$

which cannot be solved analytically for l_0. Hence, we numerically minimize the absolute value of Equation (C.8) with respect to l_0 and, by substituting the solution into Equation (C.6), we get the solution for y_0. If $l_0 < 0$ or $l_0 > 1$, we are in the presence of a corner solution and l_0 is replaced by $l_0 = 0$ or $l_0 = 1$, respectively.

Once we know the location of (l_0, y_0), the compensation criterion with $\widetilde{w} > 0$ is the vertical intercept of the budget set passing through (l_0, y_0) with slope $\widetilde{w}$ (see Figure 3b). Hence, $m_i^*(\widetilde{w}, y_i^*, l_i^*) = y_0 - \widetilde{w}(1 - l_0)$.

Calculation of the responsibility criteria

The calculation of the responsibility criterion starts by identifying the bundle (l_0, y_0) located along the indifference curve attained by individual i with optimal choice (l^*, y^*) (or (l', y') in the counterfactual scenario) where the indifference curve has slope equal to the personal wage rate w_i. The strategy to find the bundle (l_0, y_0) is the same followed for the computation of the compensation criterion with $\widetilde{w} > 0$. The only difference is that we have to replace $\widetilde{w}$ with w_i in Equations (C.6)–(C.8). Once we have determined the bundle (l_0, y_0), the responsibility criterion is the height on the vertical axis of the budget line with slope w_i passing through (l_0, y_0) in correspondence of $\widetilde{l} = 1 - \widetilde{h}$ (see Figure 3a). Hence, the responsibility criterion with $\widetilde{h} \in [0, 1]$ is given by $y_0 - w_i(1 - l_0) + w_i \widetilde{h}$.

References

Aaberge, R. and U. Colombino, "Using a microeconometric model of household labour supply to design optimal income taxes," *Scandinavian Journal of Economics*, 2013, *115* (2), 449–475.

— , — , **and S. Strøm**, "Labour supply in Italy: An empirical analysis of joint household decisions, with taxes and quantity constraints," *Journal of Applied Econometrics*, 1999, *14* (4), 403–22.

— , — , **and** — , "Do more equal slices shrink the cake? An empirical investigation of tax-transfer reform proposals in Italy," *Journal of Population Economics*, 2004, *17* (4), 767–785.

Altig, D. and C.T. Carlstrom, "Marginal tax rates and income inequality in a life-cycle model," *American Economic Review*, 1999, *89* (5), 1197–1215.

Bargain, O., A. Decoster, M. Dolls, D. Neumann, A. Peichl, and S. Siegloch, "Welfare, labor supply and heterogeneous preferences: Evidence for Europe and the US," *Social Choice and Welfare*, 2013, *41* (4), 789–817.

— , **K. Orsini, and A. Peichl**, "Labor supply elasticities in Europe and the US," IZA Discussion Papers 5820, Institute for the Study of Labor (IZA) June 2011.

Blundell, R., "Welfare reforms for low income workers," *Oxford Economic Papers*, 2001, *53* (2), 189–214.

— **and A. Shephard**, "Employment, hours of work and the optimal taxation of low-income families," *Review of Economic Studies*, 2012, *79* (2), 481–510.

— **and T. Macurdy**, "Labor supply: A review of alternative approaches," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics, Volume 3*, Amsterdam: Elsevier, 1999, chapter 27, pp. 1559–1695.

Decoster, A.M.J. and P. Haan, "Empirical welfare analysis with preference heterogeneity," *International Tax and Public Finance*, 2015, *22* (2), 224 – 251.

Dickert, S., S. Houser, and J.K. Scholz, "The earned income tax credit and transfer programs: A study of labor market and program participation," in J.M. Poterba, ed., *Tax Policy and the Economy, Volume 9*, Cambridge: The MIT Press, 1995, chapter 1, pp. 1–50.

Eissa, N., "Taxation and labor supply of married women: The tax reform act of 1986 as a natural experiment," 1995. NBER Working Paper No. 5023.

— **and H.W. Hoynes**, "The earned income tax credit and the labor supply of married couples," 1998. NBER Working Paper No. 6856.

— **and** — , "Taxes and the labor market participation of married couples: the earned income tax credit," *Journal of Public Economics*, 2004, *88* (9-10), 1931–1958.

— **and J.B. Liebman**, "Labor supply response to the earned income tax credit," *Quarterly Journal of Economics*, 1996, *111* (2), 605–637.

— , **H.J. Kleven, and C.T. Kreiner**, "Evaluation of four tax reforms in the United States: Labor supply and welfare effects for single mothers," *Journal of Public Economics*, 2008, *92* (3-4), 795–816.

Ellwood, D., "The impact of the earned income tax credit and social policy reforms on work, marriage and living arrangements," *National Tax Journal*, December, Part 2 2000, *53* (4), 1063–1105.

Feenberg, D.R and E. Coutts, "An introduction to the TAXSIM model," *Journal of Policy Analysis and Management*, 1993, *12* (1), 189–194.

Feldstein, M.S., "The tax reform act of 1986: Comment on the 25th Anniversary," 2011. NBER Working Paper No. 17531.

Fleurbaey, M., *Fairness, responsibility and welfare*, New York: Oxford University Press, 2008.

— **and D. Blanchet**, *Beyond GDP. Measuring Welfare and Assessing Sustainability*, Oxford: Oxford University Press, 2013.

— **and F. Maniquet**, *A theory of fairness and social welfare*, Cambridge: Cambridge University Press, 2011.

— **and** — , "Optimal taxation theory and principles of fairness," 2015. CORE Discussion Paper No. 2015/5, Louvain-la-Neuve.

Heckman, J.J. and B. Singer, "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 1984, *52* (2), 271–320.

Hotz, V.K. and J.K. Scholz, "The earned income tax credit," in R.A. Moffitt, ed., *Means-Tested Transfer Programs in the United States*, Chicago: University of Chicago Press, 2003, chapter 3, pp. 141–197.

Hoynes, H.W., "Welfare transfers in two-parent families: Labor supply and welfare participation under AFDC-UP," *Econometrica*, 1996, *64* (2), 295–332.

King, M.A., "Welfare analysis of tax reforms using household data," *Journal of Public Economics*, 1983, *21* (2), 183–214.

McFadden, D. and K. Train, "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 2000, *15* (5), 447–470.

Meier, B.D. and D.T. Rosenbaum, "Making single mothers work: recent tax and welfare policy and its effects," *National Tax Journal*, December, Part 2 2000, *53* (4), 1027–1061.

— **and** — , "Welfare, the earned income tax credit, and the labor supply of single mothers," *Quarterly Journal of Economics*, 2001, *116* (3), 1063–1114.

Mundlak, Y., "On the pooling of time series and cross section data," *Econometrica*, 1978, *46* (1), 69–85.

Roemer, J. E., *Equality of Opportunity*, Cambridge: Harvard University Press, 1998.

Samuelson, P.A., "Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory," *Journal of Economic Literature*, 1974, *12* (4), 1255–1289.

Schoeni, R.F. and R.M. Blank, "What has the welfare reform accomplished? Impacts on welfare participation, employment, income, poverty, and family structure," 2000. NBER Working Paper No. 7627.

Scholz, J.K., "In-work benefit in the United States: The earned income tax credit," *Economic Journal*, 1996, *106* (434), 156–169.

Stiglitz, J.E., "Pareto efficient and optimal taxation and the new new welfare economics," 1987. NBER Working Paper No. 2189.

Triest, R.K., "The effect of income taxation on labor supply in the United States," *Journal of Human Resources*, 1990, *25* (3), 491–516.

van den Berg, G.J. and M. Lindeboom, "Attrition in panel survey data and the estimation of multi-state labor market models," *Journal of Human Resources*, 1998, *33* (2), 458–478.

— , — , **and G. Ridder**, "Attrition in longitudinal panel data and the empirical analysis of dynamic labour market behaviour," *Journal of Applied Econometrics*, 1994, *9* (4), 421–35.

van Soest, A., "Structural models of family labor supply: a discrete choice approach," *Journal of Human Resources*, 1995, *30* (1), 63–88.

— , **M. Das, and X. Gong**, "A structural labour supply model with flexible preferences," *Journal of Econometrics*, 2002, *107* (1-2), 345–374.

Wooldridge, J.M., "Selection corrections for panel data models under conditional mean independence assumptions," *Journal of Econometrics*, July 1995, *68* (1), 115–132.

— , *Econometric Analysis of Cross Section and Panel Data*, Cambridge: The MIT Press, 2010.