FISCAL STUDIES, vol. 29, no. 4, pp. 499–522 (2008) 0143-5671

Analysing Tax–Benefit Reforms Using Non-Parametric Methods*

CARLO V. FIORIO†

†*University of Milan; Econpubblica*
(carlo.fiorio@unimi.it)

Abstract

Static tax–benefit microsimulation models (MSMs) are widely used and well-regarded tools for public policy analysis, but it is essential to use them very carefully. This paper focuses on the analysis of MSM output, suggesting the use of non-parametric methods as a useful, informative and relatively straightforward complement to detect effects not always captured by measures often used to present MSM results.

Non-parametric methods are used here to analyse the output of an MSM applied to the 1998 Italian personal income tax reform, the main change in which concerned the tax schedule: the first tax rate was increased from 10 per cent to 18.5 per cent and the top one was reduced by 4.5 percentage points. Non-parametric methods highlight that the effects of this reform were very different for different types of households, with low-income pensioner households among the main losers. Results are checked for robustness by standard statistical methods and compared with empirical results obtainable using quintile histograms.

*Submitted May 2008.

The author is grateful to Frank Cowell, Conchita D'Ambrosio, Emmanuel Flachaire, Massimo Florio, Fabrizio Iacone and Daniela Mantovani for fruitful discussions and helpful comments. All remaining errors are his responsibility.

Keywords: tax–benefit microsimulation model, tax reform, losers and gainers, kernel density on bounded support, non-parametric regression.

JEL classification numbers: H24, D31, C14.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

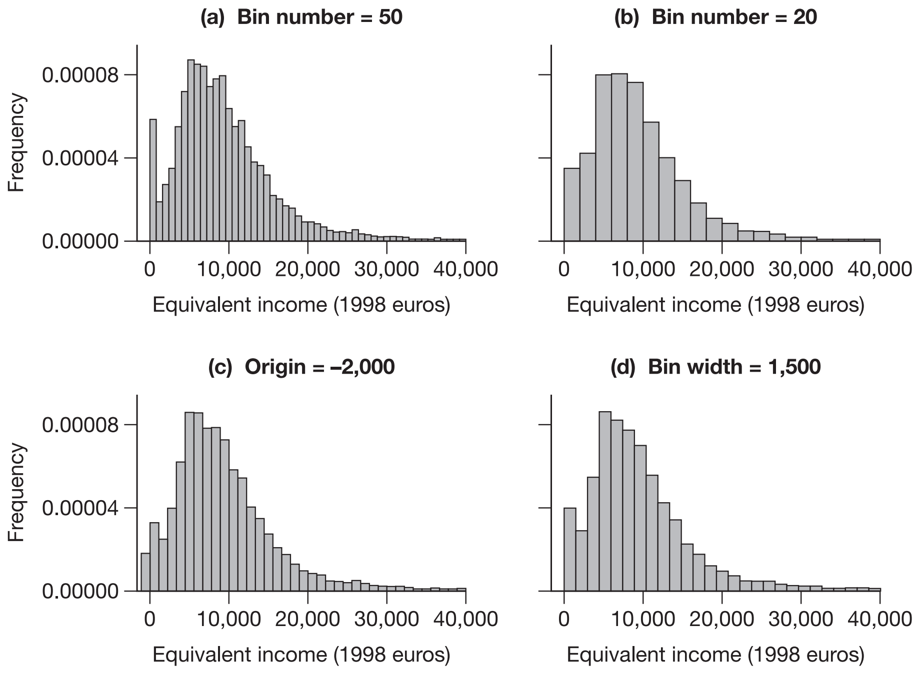
I. Introduction

Arithmetic tax–benefit microsimulation models (MSMs) analyse the 'morning-after' impact of tax–benefit reforms on the distribution of income and on poverty, and allow one to assess who are the gainers and losers. MSMs are mostly developed from household survey databases, which provide a picture of the population much closer to reality than any database using representative households. Nowadays, most developed countries have at least one MSM and multi-country models have also been developed (for example, EUROMOD, a Europe-wide tax–benefit model – see Sutherland (2001)). MSMs are important tools for orienting and evaluating tax and benefit policies, but they have to be handled with care.

The output of MSMs is often presented using distributions by deciles or by histograms. However, as discussed in Silverman (1986, ch. 2), although very intuitive, these methods often provide a biased and non-robust picture of reality. As a simple illustration, Figure 1 shows four histograms using the

FIGURE 1

A comparison of different histograms of after-tax 1998 equivalent income densities

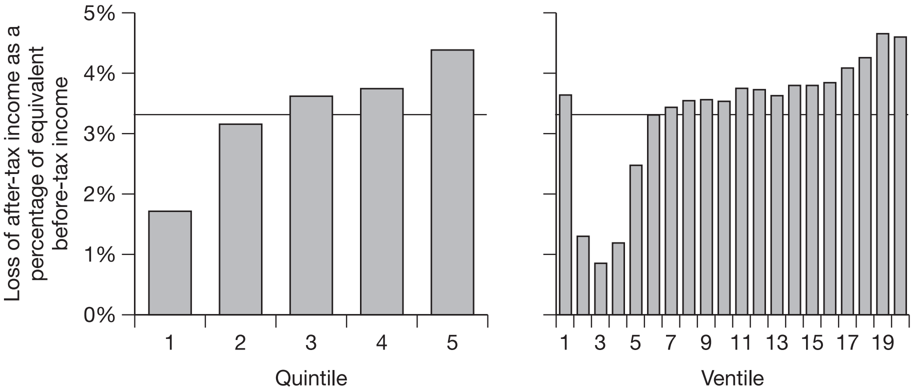


Notes: Histograms differ in bin number, origin and bin width choice. Data are in 1998 euros (€1 = Lit 1,936.27).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

FIGURE 2

An analysis of the 1998 reform using histograms, by quintiles and ventiles of equivalent before-tax income: households with pensioner head[a]



[a]For more details of the simulation, see Section IV.
Notes: The horizontal line shows the overall average loss for the whole sample. Data are in 1998 euros (€1 = Lit 1,936.27).

same data,[1] differing only in their choice of number of bins, origin and bin width. It suggests that each histogram might lead to different conclusions regarding the effects of the tax reform on the equivalent income distribution. For instance, while a histogram with 50 bins and origin set at zero shows a clear spike close to zero (panel a), the spike disappears using 20 bins (panel b) or setting the origin at –€2,000 (panel c), and only vaguely appears if the bin width is set at €1,500 (panel d). Figure 2 shows some histograms for the distribution of losses due to the 1998 Italian personal income tax reform[2] as opposed to the 1991 personal income tax system for families with a pensioner head. The average household in the whole sample lost 3.2 per cent of its equivalent before-tax income – as shown by the horizontal line – but the distribution of losses varies across income levels. It can be seen that histograms would have led to different conclusions depending on the histogram chosen: using a quintile distribution (left-hand panel) low-income pensioner households would have been relative gainers from the reform, but using a ventile distribution (right-hand panel) it emerges that pensioner households in the bottom 5 per cent of the equivalent income distribution would have lost more than the average.

Alternatively, non-parametric methods allow one to describe data distributions and relationships simply by letting the data speak by themselves. In contrast to histograms, non-parametric density estimation

[1]The data used here are the 1998 after-tax equivalent incomes of a representative sample of the Italian household population. They will be described further in Section III.

[2]This reform will be described in Section III.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

does not suffer major limitations such as choice of origin, limited robustness of estimates, ragged picture or absence of derivative. The interpolation of pointwise estimates of density provides a smooth picture, useful in detecting unusual behaviour of the distribution such as bimodality. Non-parametric regression methods help in detecting non-linearities in the distribution of average gains and losses at different levels of income and highlight distributional effects of reforms within any quantile interval.

Non-parametric estimation methods have proved to be an effective research tool in economics. They have been used in various fields of economics such as income inequality (e.g. Cowell, Jenkins and Litchfield, 1996; Jenkins, 1995; Pudney, 1993), economic growth (e.g. Quah, 1997) and labour economics literature (e.g. DiNardo, Fortin and Lemieux, 1996). Oddly, non-parametric estimation methods are not commonly considered in the analysis of tax and benefit reforms. This paper claims that they should be, as they are a useful, informative and relatively straightforward complement to detect effects not always captured by measures often used to present MSM results. Moreover, they allow straightforward inference analysis and are very effective for assessing the robustness of results, which is particularly important as MSMs are often used for public policy design.

Non-parametric methods are applied here to the 1998 Italian personal income tax (Irpef) reform, which included a new design for the tax bracket structure and a restructuring of tax credits to counterbalance the increased tax liability at low income levels.

The road map of the paper is as follows. Section II describes kernel density estimation on a bounded support and non-parametric regression methods to analyse the distribution of losses and gains. Section III describes the main elements of the 1998 reform and the data set. Section IV presents the microsimulation model used for the analysis of the reform and the main simulation exercise. Section V describes the results of the simulations performed, and finally Section VI concludes.

II. Non-parametric methods and tax reforms

Personal income tax reforms typically affect only those having positive income and the non-parametric density estimation based on a size-n sample, $\hat{f}(x)$, is not necessarily a consistent estimator of the true density, $f(x)$, for a point sufficiently close to the support boundary. There is an extensive literature on how to correct the so-called boundary effect, although there is no single dominating solution that corrects the boundary problem for all shapes of density.[3] The method suggested by Zhang, Karunamuni and Jones

[3]For an introductory discussion of density estimation on bounded support, see Silverman (1986, p. 29). Methods to correct for the boundary problem include the reflection method (Cline and Hart, 1991;

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

(1999) (henceforth, ZKJ), which is a combination of methods of pseudo-data, transformation and reflection, is non-negative everywhere (differently from most boundary kernel methods) and performs well compared with the existing methods for almost all shapes of densities and especially for densities with substantial mass near the boundary, which seems to be the case for the data used here (recall panel a of Figure 1).

The ZKJ method on the $[0,\infty)$ support can be described in three steps: (a) generate pseudo-data beyond the left endpoint of the support by transforming the original data $X_1,\ldots,X_n$ to $g(X_1),\ldots,g(X_n)$ while keeping the original data, where g is a non-negative, continuous and monotonically increasing function from $[0,\infty)$ to $[0,\infty)$; (b) reflect $g(X_1),\ldots,g(X_n)$ around the origin, resulting in $-g(X_1),\ldots,-g(X_n)$; (c) based on the enlarged data sample $-g(X_1),\ldots,-g(X_n),X_1,\ldots,X_n$, the new estimator is defined as

$$(1) \qquad \hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left(K\left(\tfrac{x-X_i}{h}\right)+K\left(\tfrac{x+g(X_i)}{h}\right)\right)$$

where h is the bandwidth ($h\to 0$ as $n\to\infty$), K is a non-negative symmetric kernel function with support $[-1,1]$ satisfying $\int_{-1}^{1}K(t)dt=1$, $\int_{-1}^{1}tK(t)dt=0$ and $0\neq\int_{-1}^{1}t^2K(t)dt<\infty$. The ZKJ estimator is obtained by estimating the function $g(x)$ as a cubic relationship of the data, using parameters set as suggested in Karunamuni and Alberts (2005) and Zhang and Karunamuni (1998).

As the kernel density estimation is more sensitive to the bandwidth parameter than to the kernel function, robustness checks can be performed by using different 'optimal' bandwidths, among which the Sheather and Jones (1991) (SJ) plug-in estimator, the direct plug-in (dpi) estimator (Wand and Jones, 1995) and the Silverman (1986) (S) estimator are the most frequently used. To assess the reliability of density estimates, the 90 per cent confidence bands can be computed as 1.645 standard errors around $\hat{f}(x)$.[4]

The analysis of average losses and gains due to a tax reform is performed using non-linear regression analysis as the distribution of losses and gains is a non-linear function of before-tax income. Let X be the before-tax income

Silverman, 1986), the boundary kernel method (Cheng, Fan and Marron, 1997; Jones, 1993; Zhang and Karunamuni, 1998), the transformation method (Marron and Ruppert, 1994) and the pseudo-data method (Cowling and Hall, 1996).

[4]For the bias and variance of (1), see Zhang, Karunamuni and Jones (1999). Note that confidence bands are different from confidence intervals since confidence bands do not consider the bias. Hall (1992) deals thoroughly with bias removal in confidence interval density estimation either by under-smoothing or by explicit bias removal in the context of fixed bandwidth kernels. However, these issues would greatly increase the complexity of the estimation without real benefit for the present analysis.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

and Y the relative loss of income caused by a tax reform. Given the size-n sample of observations (x,y), and assuming additive errors, the relationship between Y and X can be estimated as

$$(2) \qquad y = m(x) + \varepsilon$$

where ε is a random error with mean 0 and variance σ^2. While histograms of losers and gainers show the average loss or gain for households whose incomes belong to the same quantile interval, the estimated non-parametric regression, $\hat{m}(x)$, defines the average loss or gain at any possible value of X, showing whether the distribution of losses or gains is indeed uniformly distributed within a quantile interval.

Similarly to density estimation, which averages counts of the data locally, non-parametric regression averages the values of the variable y locally. The Nadaraya–Watson estimator is based on the function

$$(3) \qquad \hat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}$$

where the function K is the kernel function, as defined above. This means that most weight is given to the observations whose covariate values, X_i, are close to the point of interest, x. The smoothing parameter, h, similarly to non-parametric density estimation, controls for the width of the kernel function and hence the degree of smoothing applied to the data.[5]

The variance of (3) increases with smaller smoothing parameters and at values of x where neighbouring points are scarce. As h goes to zero, $\hat{m}(x)$ converges to Y_i, i.e. an interpolation of the data is obtained. On the other hand, if h goes to infinity, the estimator is a constant function that assigns the sample mean of Y to each x.[6] Often in non-parametric regressions, the cross-validation bandwidth (CV) is used, which is obtained by constructing an estimate of the mean integrated squared error and minimising it over h. The 90 per cent confidence bands are estimated as 1.645 standard errors around $\hat{m}(x)$.

III. The 1998 Italian personal income tax reform and the data set

The non-parametric methods outlined in the previous section are applied here to the 1998 Italian personal income tax (Irpef) reform. Irpef is a progressive income tax, defined on individual taxpayers. Progressivity is

[5]For details on the bias and variance of (3), see, among others, Bowman and Azzalini (1997).

[6]See, among others, Bowman and Azzalini (1997) and Härdle et al. (2004).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

obtained by an increasing tax rate schedule and by deductions and tax credits which depend on individual characteristics and the taxpayer's family burdens. Although Irpef was introduced as a comprehensive income tax, successive modifications occurred between its introduction in 1974 and the end of the 1980s. These made it a tax mainly on labour and pension income, as financial income is excluded from the tax base, agricultural and building and estate incomes are imputed using cadastral rather than market measures, and other exemptions exist. Moreover, taxable self-employment income is net of income-producing expenditures, differently from employment income, and tax evasion opportunities are larger for self-employment than for employment and pension income. The first significant reform of Irpef since its introduction occurred in 1998, ending a period of several small changes without an overall design that followed the 1992 financial and currency crisis (De Vincenti and Paladini, 2008).

The two main features of the 1998 Irpef reform concern the modifications of the tax brackets and of the tax credits structure, while there was no significant change in the definition of the income base. Compared with 1991, the year before the crisis, the number of fiscal brackets was reduced, from seven to five, while the highest tax rate was decreased (from 50 per cent to 45.5 per cent), the first tax rate was increased (from 10 per cent to 18.5 per cent) and the others were substantially changed (see Table 1). Tax credits for employment and self-employment were increased in amount and in number, tax credits for 'family burdens' were increased, and a new tax credit for pension recipients was introduced depending on income and a few other attributes.[7] The government that passed the reform claimed that the increase in tax credits would be sufficient to compensate low-income taxpayers for the increased tax rate of the first bracket.

This reform has been analysed by various authors. Among others, Bosi, Mantovani and Matteuzzi (1999), CER (1998) and Birindelli et al. (1998)

TABLE 1

Actual 1998 and counterfactual 1991 structure of Irpef tax brackets

1991 Irpef		1998 Irpef	
Income bracket (€)	*Tax rate (%)*	*Income bracket (€)*	*Tax rate (%)*
0–5,188	10.0	0–7,747	18.5
5,188–10,299	22.0	7,747–15,494	26.5
10,299–25,710	26.0	15,494–30,987	33.5
25,710–51,572	33.0	30,987–69,722	39.5
51,572–128,778	40.0	Over 69,722	45.5
128,778–257,631	45.0		
Over 257,631	50.0		

Note: Data are in 1998 euros (€1 = Lit 1,936.27).

[7]For details, see the appendix.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

studied the 1998 reform compared with the previous year's tax system, while Giannini and Guerra (1999) compared the 1999 tax system with that of 1990. They all concluded that the reform caused an overall increase in Irpef liability on Italian households, with a slight increase in progressivity and a small reduction in inequality, but there is less agreement in identifying the most and least affected groups when the sample is divided into subgroups. Although different MSMs using slightly different simulation assumptions are likely to produce numerically different results, the exclusive use of discrete, non-robust descriptive tools blurs the main picture of the distributional effects of the 1998 reform even more.

For the simulation of the 1998 tax reform, the MSM used here uploads the 1998 Survey of Household Income and Wealth (SHIW) data.[8] The SHIW is a long-standing survey, based on face-to-face interviews and available free-of-charge from the Bank of Italy website. The SHIW is the most frequently used data set for Italian MSMs, as well as for any kind of household income analysis at the national level. The 1998 data set contains detailed micro-data about 7,147 households and 20,901 individuals on 1998 disposable income, consumption, labour market, monetary and financial variables. Data are released with sampling weights, which are used throughout the following analysis. The interviews include only recall questions and do not include information about people who do not have a registered dwelling or who are in hospital or other kinds of institution.[9]

The limitations of this data set for microsimulation modelling relate to the type of income recorded: only disposable income, excluding taxes and social contributions paid and benefit received, is recorded; hence, the first role of an MSM on SHIW data is to simulate before-tax income, prior to introducing any other policy simulation.

IV. Simulating a counterfactual tax system

The MSM used here is TABEITA98, a TAx–BEnefit microsimulation model on ITAlian 1998 SHIW data. TABEITA98 simulates 1998 personal income taxation[10] net of social contributions. TABEITA98 is a static model without behavioural response. It can be described as a deterministic transformation of a given sample into a new one. Let Y_A and Y_B be the $n\times 1$ vectors of after-tax and before-tax income, respectively: the former vector is obtained from the latter through a tax function, say τ_i for $i = 1,\ldots,n$, where n is the number

[8]Sutherland (1991) provides an illuminating discussion of the bias produced by using data sets of a previous year inflated using a price index, such as the consumer price index (CPI) or the retail price index (RPI).

[9]For more detail on the data set, see Banca d'Italia (2000).

[10]Irpef accounted for 75.8 per cent of total Italian revenues from direct taxation in 1998 (Banca d'Italia, 1999).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

of individuals in the sample. As only Y_A is known to the analyst, and the tax function τ_i is not the same for all individuals and is highly non-linear, Y_B has to be obtained numerically, by recursive approximations.[11]

In this model, the main assumption is that the tax and benefit legislation, τ_i, is perfectly known by the individual and applied without error. Only systematic errors leading to under- or over-reporting are considered and the final model is calibrated on actual aggregated data coming from the population of tax forms and published by the Italian Ministry of Finance. As data from the SHIW tend to simulate larger tax revenues than data from the Ministry of Finance, a calibration is performed by main types of income assuming that part of Y_B is hidden from the tax authorities (tax evasion) so that the aggregate error in tax revenues between the calibrated MSM output and Ministry of Finance data is less than 1 per cent. This calibration procedure is common also to other Italian MSMs (see, for example, Mantovani (1998) and Coromaldi and Toso (2004)). As Irpef does not allow for negative income tax (i.e. it is set equal to zero if net tax is negative because tax credits exceed the gross tax), only incomes greater than zero are considered in the analysis.

In this paper, TABEITA98 is used to compare the 1998 Irpef system with the 1991 one, in real values at 1998 prices,[12] which will also be referred to as the 'actual' and the 'counterfactual' Irpef respectively. In the first stage, using the after-tax income data contained in the n-dimensional 1998 SHIW sample (Y_{A98j}, $j = 1,\ldots,n$) and the 1998 Irpef legislation, the MSM is used to obtain the before-tax income (Y_{B98j}, $j = 1,\ldots,n$). In the second stage, the counterfactual estimation is performed starting from 1998 before-tax income (Y_{B98j}) and simulating the 1991 Irpef system. The comparison year was taken as 1991 because the public finance management changed greatly after the 1992 crisis. The counterfactual distribution can be described as the 'distribution of income that would have prevailed in 1998 if personal taxation had been replaced by 1991 Irpef and each income recipient had obtained exactly the same income, before personal taxation'. However, it is not claimed that this simulation is fully suitable for comparing 1998 and 1991 Italian personal income taxation because behavioural responses to taxation are not taken into account. Nor is it claimed that this simulation is

[11]For an introduction to TABEITA family models, see D'Amuri and Fiorio (2006) and Cavalli and Fiorio (2006).

[12]The 1991 real values at 1998 prices are obtained by inflating the 1991 nominal values by the GDP growth rate between 1991 and 1998. Although other inflating procedures might be used, this is a standard way to partly neutralise bracket-creeping effects more effectively than using the consumer price index (CPI). However, it should be noted that had the CPI been used, the results would not have changed dramatically (results are obtainable from the author upon request).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

informative on overall disposable income as indirect taxation and other transfers to households are ignored.[13]

Given the focus on household welfare, the Italian Poverty Commission equivalence scale, derived from the Engel methodology, is conventionally adopted, assuming equal distribution of income among members of the same household (De Santis, 1998). The equivalent income of each member of

TABLE 2

Mean incomes, poverty indices and inequality indices for different types of income

	1998 before-tax income		*1991 after-tax income*		*1998 after-tax income*	
Taxpayer incomes						
Mean (€)	15,869	(206.16)	13,554	(127.43)	12,890	(122.96)
No. of observations	11,895		11,895		11,895	
Equivalent incomes						
Mean (€)	12,648	(191.17)	10,419	(119.17)	10,081	(116.35)
No. of observations	20,901		20,901		20,901	
Deciles						
1	3,353.09	(34.64)	3,210.14	(53.83)	3,094.78	(44.30)
2	4,976.85	(40.77)	4,791.29	(35.27)	4,628.65	(28.13)
3	6,449.64	(39.33)	6,020.01	(44.33)	5,790.09	(42.73)
4	7,857.44	(57.67)	7,220.91	(44.98)	6,918.76	(34.25)
5	9,436.95	(55.83)	8,508.15	(37.84)	8,149.88	(35.45)
6	11,114.39	(60.34)	9,847.95	(53.14)	9,388.18	(40.57)
7	13,234.51	(92.89)	11,528.82	(66.30)	10,969.73	(66.41)
8	16,029.91	(117.11)	13,681.40	(67.95)	13,007.40	(71.75)
9	21,741.23	(183.45)	17,756.96	(168.61)	16,754.07	(146.62)
Poverty indices						
Headcount	0.203	(0.003)	0.180	(0.003)	0.177	(0.003)
Poverty-gap	0.175	(0.003)	0.153	(0.003)	0.151	(0.003)
$(\text{Poverty-gap})^2$	0.152	(0.002)	0.131	(0.002)	0.129	(0.002)
Inequality indices[a]						
Gini	0.418	(0.005)	0.377	(0.004)	0.373	(0.004)
GE(0)	0.370	(0.009)	0.326	(0.008)	0.319	(0.007)
GE(1)	0.350	(0.014)	0.271	(0.009)	0.267	(0.009)
GE(2)	0.698	(0.076)	0.424	(0.034)	0.427	(0.039)

[a]GE(*x*) are generalised entropy indices. GE(0) is also known as mean log deviation, GE(1) as Theil and GE(2) as half the squared coefficient of variation.

Notes: Data are in 1998 euros (€1 = Lit 1,936.27). Standard errors are shown in parentheses. The poverty line is set at half the median income.

[13]The simulations performed in this paper look at Irpef reform only and do not aim to evaluate the overall Italian tax and benefit system, which also includes social assistance transfers and family benefits as well as other taxes that families may be liable to pay, such as the building and real estate assets tax (ICI) and the tax on income from productive activities (IRAP) (Coromaldi and Toso, 2004).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

household h is estimated as $X_h = Z_h/s^{\eta}$ where Z_h is the sum of all incomes in household h and s is the household size. Using the 1998 SHIW data, η is estimated to be 0.757.

Some summary statistics reported in Table 2 show that the Irpef reform caused an average equivalent after-tax income loss of 3.2 per cent, increasing tax liability across all income deciles. Looking at inequality and poverty indices,[14] it might be observed that personal income taxation significantly reduces poverty and inequality using a wide range of measures. There are a couple of issues worth noting: first, looking at differences in inequality and poverty indices between the actual and counterfactual Irpef systems, one cannot conclude that the poverty and inequality changes were statistically significant; second, as generalised entropy (GE) inequality indices have different sensitivity at different levels of income (Cowell, 1995) and GE(0) decreases while GE(2) increases between the 1991 and 1998 systems, changes might have been different at different levels of income. Both these remarks provide some stimuli for looking more deeply at the distributional effects of the 1998 Irpef reform.

V. Distributional effects of the reform

1. Effects on income distribution

Figure 3 presents density estimations on the whole sample for before-tax (BT) and actual and counterfactual after-tax (AT) income distributions with SJ, S and dpi bandwidths. This graph produces a clear picture of the concentration effect induced by personal income taxation: the 1998 BT income density presents a lower maximum and a higher mode than AT income. The AT density presents a thinner upper tail than the BT income density, showing that the 1998 Irpef system is effective in reducing the overall density at income levels over the median. The AT income density shows some bimodality around the mode, which was not clear in the BT income density.[15]

Panel d of Figure 3 depicts the difference between counterfactual and actual AT distributions (i.e. the density loss due to introducing the 1998 instead of the 1991 Irpef) for the three bandwidths considered. It shows that density at income levels below about €3,000 was higher in 1998 than with the counterfactual 1991 tax system. Another relevant issue is clearly evident from the kernel density estimation: the BT income density at zero equivalent income shows that there is a non-negligible probability of households with

[14]The poverty line was set at half the median income.

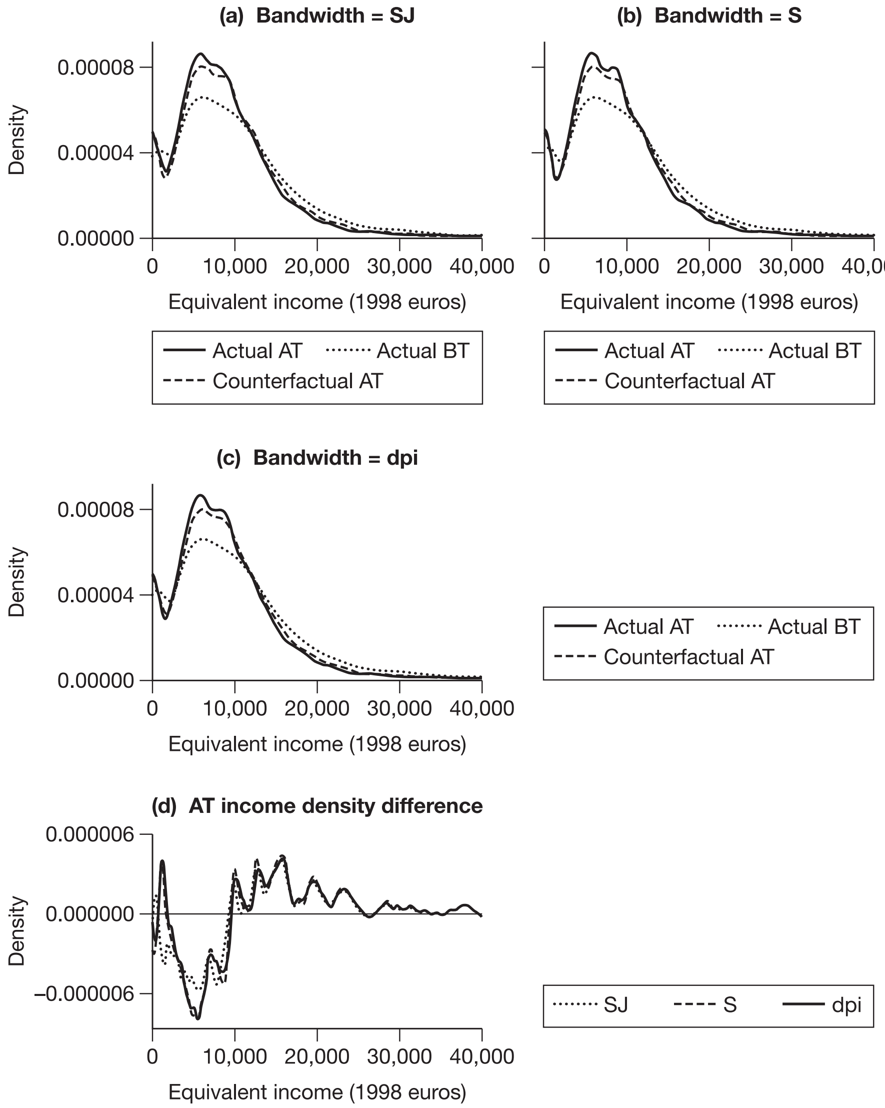
[15]This result adds something to findings of Pittau and Zelli (2001) and D'Ambrosio (2001): the bimodality of equivalent AT income is partly due to and magnified by personal income taxation.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

zero income. Hence, Figure 3 also shows how non-parametric density estimation can be a robust complement for methods usually employed to analyse MSM output, which are effective for a more policy-oriented audience. In particular, it suggests that out of the histograms in Figure 1, the

FIGURE 3

Analysis of density estimates using different bandwidths on the whole sample



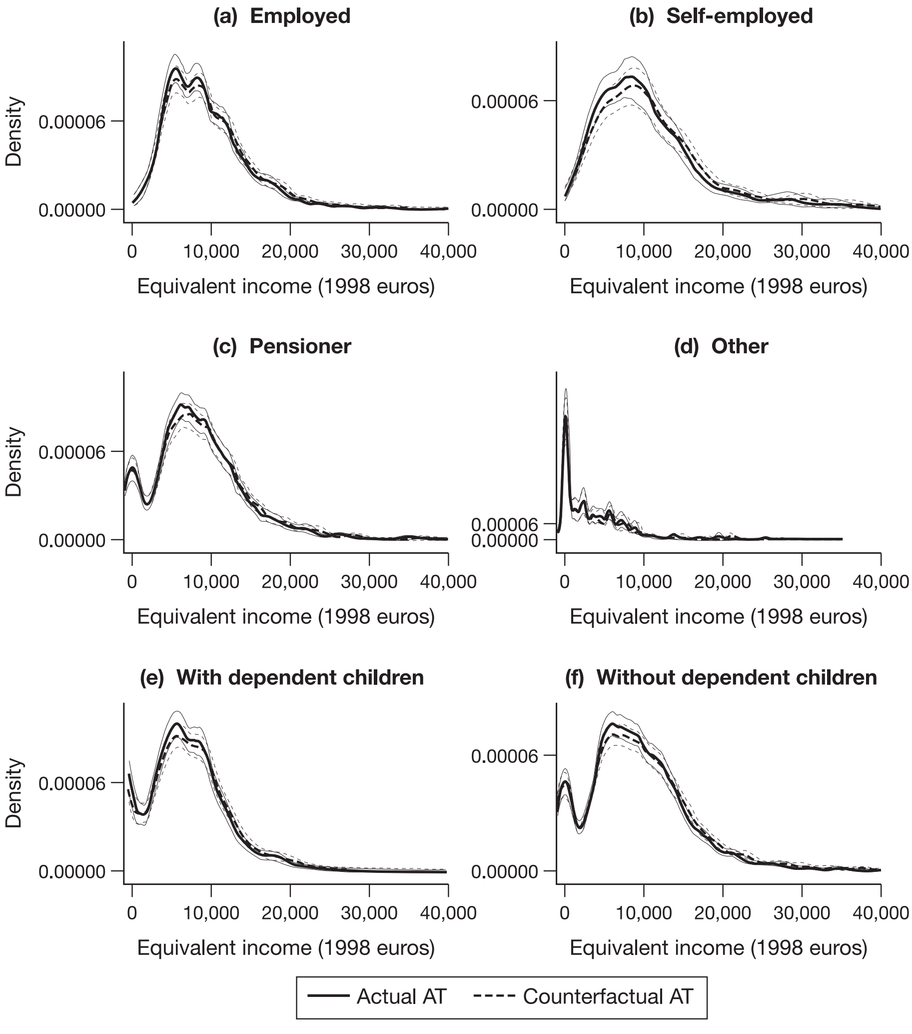
Notes: Panel d shows the difference between counterfactual AT and actual AT income densities. Data are in 1998 euros (€1 = Lit 1,936.27).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

one closest to the true 1998 AT distribution is in panel a, where there is a clear spike at incomes close to zero.[16]

FIGURE 4

AT income densities by occupation of householder and some family burdens



Notes: The SJ bandwidth is used, with 90 per cent confidence bands shown by the thinner lines. Data are in 1998 euros (€1 = Lit 1,936.27).

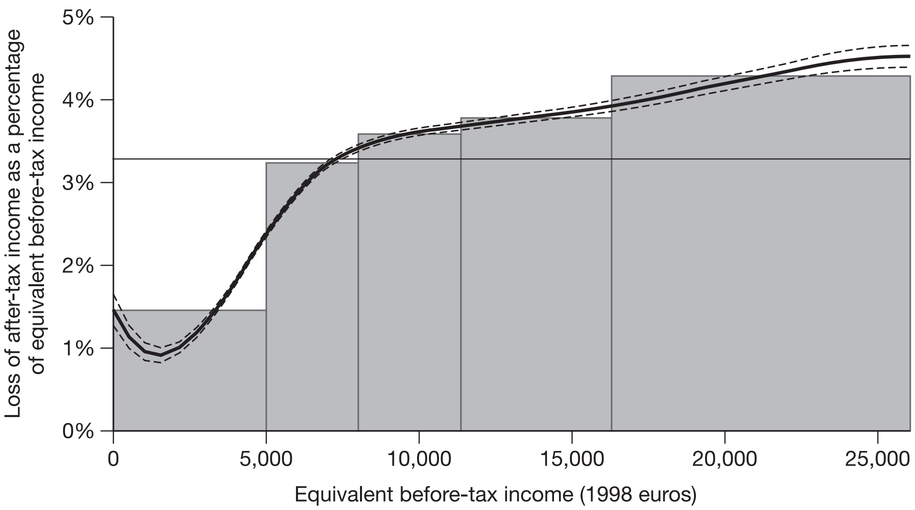
[16]This is mainly due to the fact that Irpef does not allow tax credits in the case of negative or zero tax liability. The possibility of negative taxation is an important issue in recent Italian tax debate. An extensive discussion of negative income taxation can be found in the recent 'White Book' on Irpef reform (De Vincenti and Paladini, 2008).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

Other interesting observations can be made when breaking the sample down by occupation of the householder and by households with and without dependent children. This is done since the householder's income comprises most of the family income[17] and Irpef allows for very different tax credits (and hence tax liability) depending on the taxpayer's type of income earned (employment, self-employment or pension) and family burdens, which implicitly define different groups in the population. Figure 4 depicts the actual and counterfactual AT density estimates with SJ bandwidth using continuous and dashed lines respectively.[18] It shows that density distributions are very different depending on the subgroup considered. Bimodality of AT income is a feature of the income distribution of employed and pensioner households and of households without dependent children, pointing to the fact that for these groups some degree of polarisation exists and should probably be monitored across time. The concentration of households at zero income is relatively large for pensioner households, the residual group (labelled 'other', mostly consisting of non-working householders) and households with no dependent children, suggesting that these should be the main target of public policies for low-income support.

FIGURE 5

Distribution of losses: whole sample



Notes: The horizontal solid line shows the average relative loss due to the 1998 reform. The heights of the bars represent the average loss within each quintile interval (first, second, third and fourth) of before-tax equivalent household income. Bandwidth is selected by cross-validation (CV), with 90 per cent confidence bands shown by the dashed lines. Data are in 1998 euros (€1 = Lit 1,936.27).

[17]On average, over 65 per cent of family income was produced by the householder in 1998.

[18]The thinner lines show the 90 per cent confidence bands of the two density estimates.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

2. The distribution of losses and gains

Figure 5 shows the non-parametric estimate of the average relative loss caused by the 1998 tax reform compared with the 1991 Irpef as a percentage of 1998 before-tax income, i.e. a non-parametric regression of Y against X as defined in Section II. The horizontal line shows the total average loss and the heights of the bars denote the average loss within each quintile of before-tax income.[19] The graph shows that all households experienced a positive increase in their tax liability and that households in the first quintile of income had a lower average increase in tax liability. However, the relative loss is the lowest for equivalent income of about €2,000 and quickly increases between €2,000 and the second quintile.

It should, however, be noted that some households have been affected by the reform more than others. In Figure 6, the non-parametric regressions by type of household are reported. It shows that while the 1998 Irpef reform caused no loss for very low-income households in the group of employed householders and in the group with dependent children, the loss for very low-income pensioner households was larger than the average. This is very likely to be due to the increased first tax rate of the 1998 reform, which was only partly compensated by the change in tax credits (see the appendix). A histogram by quintiles of before-tax income would have completely hidden the distributive differences among the households in the first quintile of income. The 90 per cent confidence bands show that estimates are reasonably reliable, with the only exception being incomes in the top quintile of the residual group due to the small sample size.

The use of histograms of losers and gainers would have hidden most of these distributional differences or even left some doubts about the average loss in each quantile group (recall Figure 2, where histograms were constructed using exactly the same data as in panel c of Figure 6).

3. A revenue-neutral reform simulation

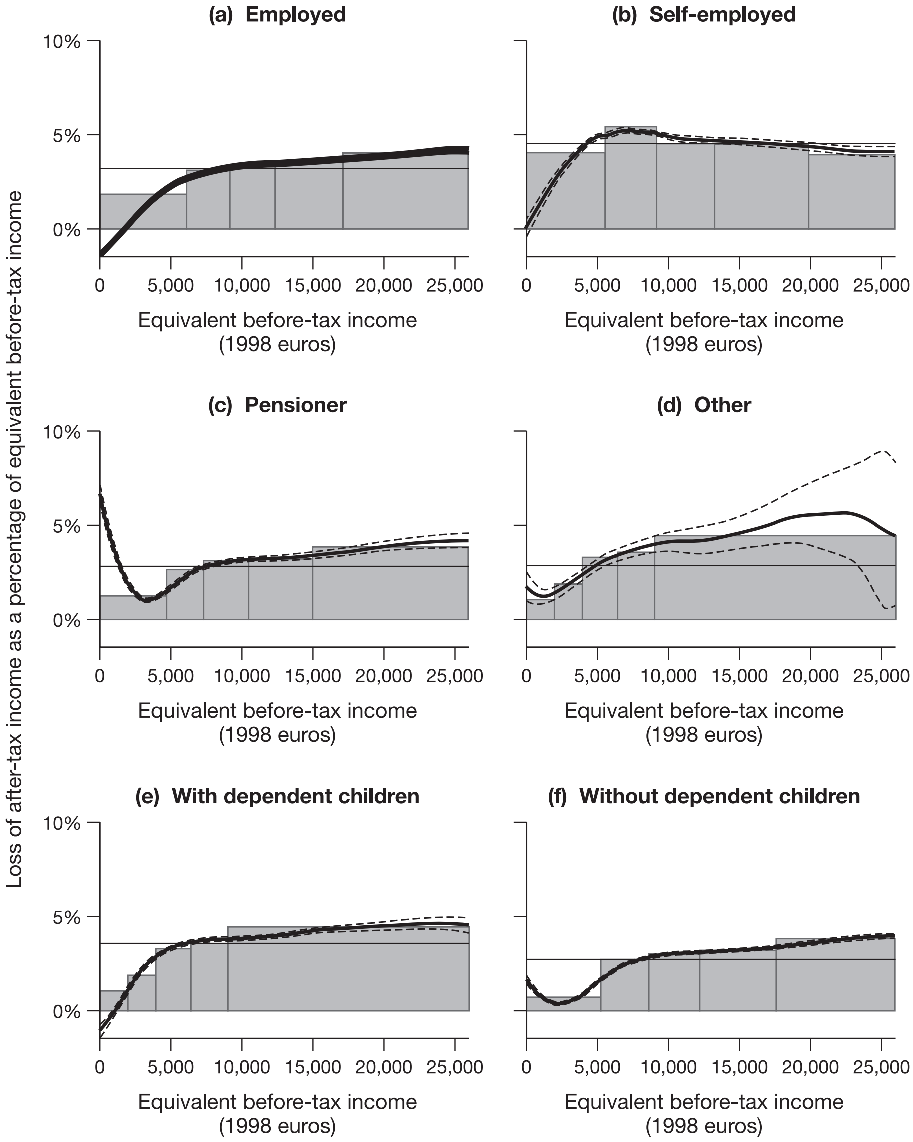
The 1998 Irpef reform induced a significant increase in revenues compared with the 1991 system. It is, of course, difficult to assess how the increased tax revenue was employed, partly because other taxes and welfare reforms were introduced in the same period and partly because the increased tax revenue was not constrained to be used for any particular policy. Here, two different revenue-neutral simulations are described: in the first, the excess revenue is added to the counterfactual income of each taxpayer in proportion to their before-tax income; in the second, the excess revenue is equally attributed to the counterfactual income of each taxpayer.

[19]These histogram bars differ from those of commonly used histograms only in presenting a non-constant width.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

FIGURE 6

Distribution of losses by occupation of the householder

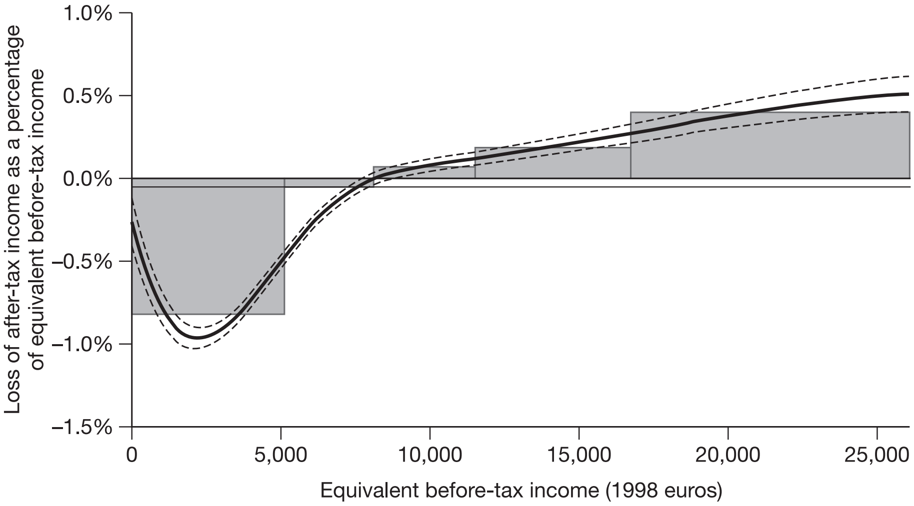


Notes: In each panel, the horizontal solid line shows the average relative loss for that group due to the 1998 reform. The heights of the bars represent the average loss within each quintile interval (first, second, third and fourth) of before-tax equivalent household income, within each group. Bandwidth is selected by cross-validation (CV), with 90 per cent confidence bands shown by the dashed lines. Data are in 1998 euros (€1 = Lit 1,936.27).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

FIGURE 7

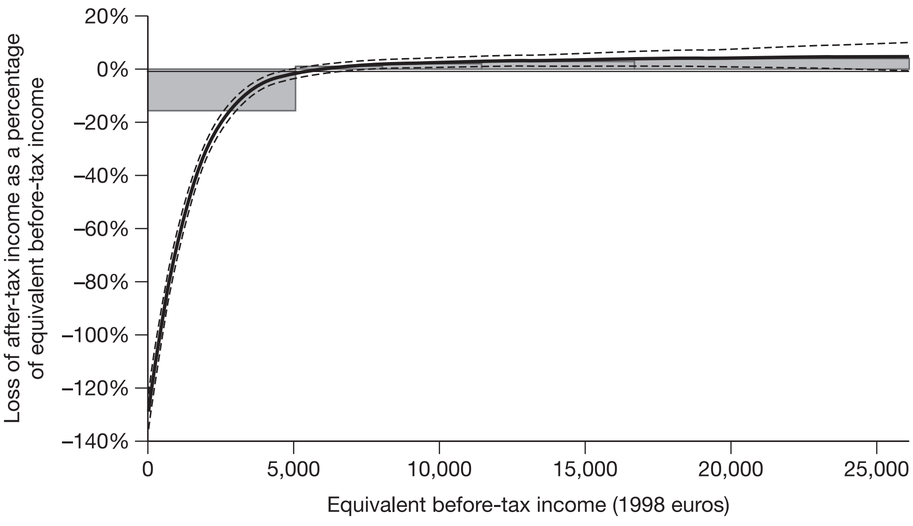
Distribution of losses for a revenue-neutral simulation with proportional redistribution: whole sample



Notes: The horizontal solid line shows the average relative loss due to the 1998 reform under this revenue-neutral simulation. The heights of the bars represent the average loss within each quintile interval (first, second, third and fourth) of before-tax equivalent household income. Bandwidth is selected by cross-validation (CV), with 90 per cent confidence bands shown by the dashed lines. Data are in 1998 euros (€1 = Lit 1,936.27).

FIGURE 8

Distribution of losses for a revenue-neutral simulation with lump-sum redistribution: whole sample



Notes: See Notes to Figure 7.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

Figure 7 shows average losses when the 1998 excess revenue has been redistributed proportionally to before-tax income; Figure 8 shows the case where the excess revenue has been redistributed as an equal lump sum to all taxpayers.[20] While the proportional redistribution would have reduced but not eliminated the distributional differences at different income levels, the lump-sum redistribution would have greatly increased the after-tax income of households with income in the first quintile, the poorest of which would have more than doubled their income with respect to the 1991 tax system. The strong distributional differences within the first quintile would not have been captured using quintile histograms, especially in the case of the lump-sum redistribution.

4. Decomposing the tax reform

The difference between counterfactual and actual 1998 income was decomposed to investigate the overall importance of changes to tax credits as opposed to changes to brackets and rates, which are the main features of the 1998 Irpef reform.

Two alternative scenarios are simulated and contrasted to the actual 1998 after-tax income. Scenario 1 is characterised by a tax credit system the same as the one in use in 1991 but with a tax bracket structure and tax rates like those of 1998. Scenario 2 simulates a tax credit system as in 1998 but with a tax bracket structure and tax rates as in 1991.

In Figure 9, the relative loss variable, Y, is defined as the difference between Scenario 1 and actual 1998 after-tax income as a percentage of actual before-tax income. It shows that had tax credits been as in 1991, the excess revenue of the 1998 reform would have been slightly negative, with households in the first income quintile enjoying the largest gains due to the increase in tax credits.

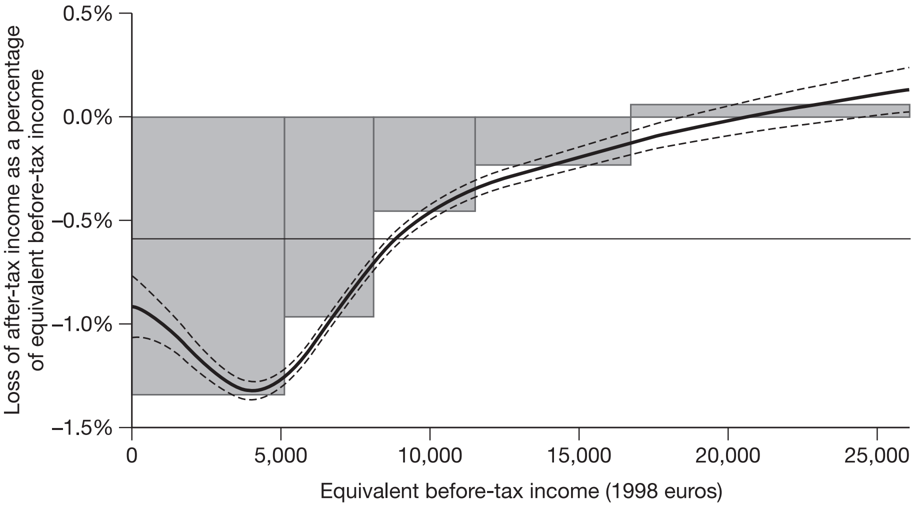
In Figure 10, the relative loss variable, Y, is defined as the difference between Scenario 2 and actual 1998 after-tax income as a percentage of actual before-tax income. It shows that had tax brackets and tax rates been set at their 1991 levels, more tax revenue would have been obtained, with losses increasing very quickly for equivalent incomes between the first and second quintiles. The use of histograms would have left a lot of uncertainty about the size of losses in the first two quintiles.

[20]The solid horizontal line, which as usual represents the average equivalent household loss, is not exactly at zero in either case, as redistribution in these simulations occurs at the individual level while distributional analysis is performed at the household level.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

FIGURE 9

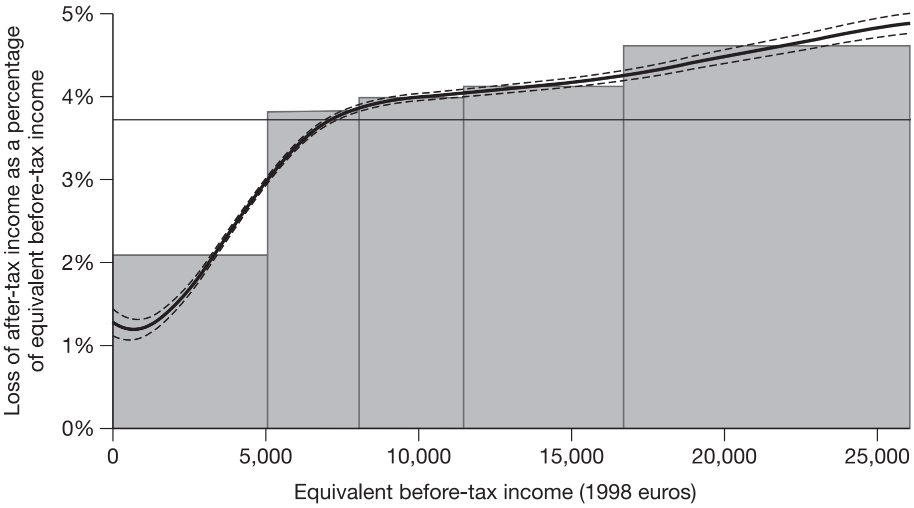
Distribution of losses for a simulation of Scenario 1: whole sample



Notes: The horizontal solid line shows the average relative loss due to the 1998 reform under Scenario 1. The heights of the bars represent the average loss within each quintile interval (first, second, third and fourth) of before-tax equivalent household income. Bandwidth is selected by cross-validation (CV), with 90 per cent confidence bands shown by the dashed lines. Data are in 1998 euros (€1 = Lit 1,936.27).

FIGURE 10

Distribution of losses for a simulation of Scenario 2: whole sample



Notes: The horizontal solid line shows the average relative loss due to the 1998 reform under Scenario 2. The heights of the bars represent the average loss within each quintile interval (first, second, third and fourth) of before-tax equivalent household income. Bandwidth is selected by cross-validation (CV), with 90 per cent confidence bands shown by the dashed lines. Data are in 1998 euros (€1 = Lit 1,936.27).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

VI. Conclusions

This paper suggests using non-parametric methods to analyse tax–benefit reforms, as they can increase the understanding of results, provide useful insights about the impact of tax–benefit reforms on income distribution and add to analysis of losers and gainers. They can be used as a useful, informative and relatively straightforward complement to detect effects not always captured by the measures often used to present results produced by microsimulation models, such as histograms and average gains and losses by quintiles or deciles.

Simulating the 1998 Italian personal income tax reform, the strong concentration effect induced by the progressive tax system is illustrated. Non-parametric density estimation on a bounded support shows that histograms not showing a clear spike close to zero income should not be presented.

Decomposing the sample into different subgroups highlights the fact that some households have been affected more than others. While low-income employed households and households with dependent children were basically unaffected by the reform, those with non-working, and especially pensioner, heads suffered major losses. In fact, for these groups of households, the increased tax credits were not enough to offset the increased tax rate of the first income bracket.

An analysis of the 1998 Italian personal income tax reform using non-parametric methods shows that even in typical reforms of developed economies, which do not tend to produce dramatic changes in the income distribution, one may find interesting effects that could not be clearly detected using histograms or distributions of gains and losses by quintiles or deciles.

Appendix: 1998 Irpef vs. 1991 Irpef

Italian personal income tax (Irpef) is a tax on individual income. The amount of tax due is obtained by applying the tax bracket structure to the taxable income (i.e. total individual income minus exempt incomes and tax deductions) and subtracting tax credits, which depend on the family and individual characteristics of the taxpayer. To keep the simulation analysis simple, in this paper only Irpef is considered and the local property tax (ICI), taxation of financial income, taxation of severance pay, the tax on productive activities (IRAP), social insurance contributions and other family benefits are not considered.

The tables in this appendix provide some detail of the difference between the 1991 and 1998 systems. In 1998, the number of tax brackets was reduced

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

TABLE A1

Actual 1998 and counterfactual 1991 tax credits for dependent spouse

1991 Irpef		1998 Irpef	
Income level	*Tax credit (€)*	*Income bracket (€)*	*Tax credit (€)*
Any income	515	0–15,494	546
		15,494–30,987	491
		30,987–51,646	459
		Over 51,646	422

Note: Data are in 1998 euros (€1 = Lit 1,936.27).

TABLE A2

Actual 1998 and counterfactual 1991 tax credits for dependent children and other relatives

	1991 tax credit (€)	*1998 tax credit (€)*
Dependent child	60	174
Other dependent relative	82	174

Note: Data are in 1998 euros (€1 = Lit 1,936.27).

TABLE A3

Actual 1998 and counterfactual 1991 tax credits for employment income

1991 Irpef		1998 Irpef	
Income bracket (€)	*Tax credit (€)*	*Income bracket (€)*	*Tax credit (€)*
0–9,460	649	0–4,700	868
9,460–9,658	572[a]	4,700–4,803	826
Over 9,658	494	4,803–7,747	775
		7,747–7,902	697
		7,902–8,057	646
		8,057–8,212	594
		8,212–15,494	542
		15,494–20,658	491
		20,658–25,823	439
		25,823–30,987	387
		30,987–31,142	336
		31,142–36,152	284
		36,152–41,317	232
		41,317–46,481	181
		46,481–46,688	129
		46,688–51,646	77
		Over 51,646	52

[a]This figure is an average. The actual tax credit (in euros) was computed as $851–[(y_B–12,400)\times0.78]$ where y_B is before-tax income.
Note: Data are in 1998 euros (€1 = Lit 1,936.27).

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

TABLE A4

Actual 1998 and counterfactual 1991 tax credits for self-employment income

1991 Irpef		1998 Irpef	
Income bracket (€)	*Tax credit (€)*	*Income bracket (€)*	*Tax credit (€)*
0–5,188	128	0–4,700	362
5,188–5,340	69[a]	4,700–4,803	310
		4,803–4,958	258
		4,958–5,113	207
		5,113–7,747	155
		7,747–15,494	103
		15,494–30,987	52

[a]This figure is an average. The actual tax credit (in euros) was computed as 168–[(y_B–6,800)×0.78] where y_B is before-tax income.

Note: Data are in 1998 euros (€1 = Lit 1,936.27).

from seven to five. For a fiscally dependent spouse, tax credit depends on before-tax income (Table A1). Tax credits for dependent children and other dependent relatives do not depend on before-tax income (Table A2). Tax credits for employment and self-employment income are different and depend on before-tax income (Tables A3 and A4). If the taxpayer receives only a pension income of less than €9,626.22, does not own any building except their own dwelling and does not receive any additional transfer, there is an additional tax credit equal to €36.15 in 1998.

References

Banca d'Italia (1999), *Relazione Annuale sul 1999* (*Annual Report for the Year 1999*), Roma: Assemblea Generale Ordinaria dei Partecipanti tenuta in Roma il giorno 31 maggio 1999 (Ordinary General Meeting of Shareholders, held in Rome on 31 May 1999), Banca d'Italia.

— (2000), 'I bilanci delle famiglie italiane nell'anno 1998' ('Survey on Household Income and Wealth 1998'), *Supplementi al Bollettino Statistico* (*Supplements to the Statistical Bulletin*), year X, no. 22, Roma: Banca d'Italia.

Birindelli, L., Inglese, L., Proto, G. and Ricci, L. (1998), 'Gli effetti redistributivi della politica economica e sociale' ('Redistributive effects of the economic and social policy'), in N. Rossi (ed.), *Il lavoro e la sovranità sociale, 1996–1997. Quarto rapporto CNEL sulla distribuzione e redistribuzione del reddito in Italia* (*Labour and Social Sovereignty, 1996–1997. Fourth CNEL Report on Distribution and Redistribution of Income in Italy*), Bologna: Il Mulino.

Bosi, P., Mantovani, D. and Matteuzzi, M. (1999), *Analisi degli effetti redistributivi della riforma IRAP-IRPEF* (*Analysis of Redistributive Effects of the IRAP-IRPEF Reform*), Nota di lavoro (Technical Report) no. 2, Bologna: Prometeia.

Bowman, A. W. and Azzalini, A. (1997), *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Oxford: Oxford University Press.

Cavalli, M. and Fiorio, C. V. (2006), 'Individual vs family taxation: an analysis using TABEITA04', Econpubblica (Centre for Research on the Public Sector, Milan), Working Paper no. 116.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

CER (1998), *La riforma fiscale: una simulazione con il modello macroeconomico del CER* (*The Fiscal Reform: A Simulation with the CER Macroeconomic Model*), Technical Report, Roma: Centro Europa Ricerche.

Cheng, M., Fan, J. and Marron, J. (1997), 'On automatic boundary corrections', *Annals of Statistics*, vol. 25, pp. 1691–708.

Cline, D. B. H. and Hart, J. D. (1991), 'Kernel estimation of densities with discontinuities or discontinuous derivatives', *Statistics*, vol. 22, pp. 69–84.

Coromaldi, M. and Toso, S. (2004), *EUROMOD Country Report: Italy (2001 Tax–Benefit System)*, Colchester: ESRC Research Centre on Micro-Social Change, University of Essex.

Cowell, F. A. (1995), *Measuring Inequality*, second edition, Hemel Hempstead: Harvester Wheatsheaf.

—, Jenkins, S. P. and Litchfield, J. (1996), 'The changing shape of the UK income distribution: kernel density estimates', in J. Hill (ed.), *New Inequalities*, Cambridge: Cambridge University Press.

Cowling, A. and Hall, P. (1996), 'On pseudodata methods for removing boundary effects in kernel density estimation', *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 551–63.

D'Ambrosio, C. (2001), 'Household characteristics and the distribution of income in Italy: an application of social distance measures', *Review of Income and Wealth*, vol. 47, pp. 43–64.

D'Amuri, F. and Fiorio, C. V. (2006), 'Grossing-up and validation issues in an Italian tax–benefit microsimulation model', Econpubblica (Centre for Research on the Public Sector, Milan), Working Paper no. 117.

De Santis, G. (1998), *Le misure della povertà in Italia: scale di equivalenza e aspetti demografici* (*Poverty Measures in Italy: Equivalence Scales and Demographic Aspects*), Roma: Presidenza del Consiglio dei Ministri, Dipartimento degli Affari Sociali.

De Vincenti, C. and Paladini, R. (eds) (2008), *Libro Bianco della Commissione di Studio sull'Imposta sui Redditi delle Persone Fisiche e sul Sostegno delle Famiglie* (*White Book of the Commitee for the Study of the Personal Income Taxation and Family Support*), Roma: Scuola Superiore dell'Economia e delle Finanze 'Ezio Vanoni', Ministero dell'Economia e delle Finanze.

DiNardo, J., Fortin, N. M. and Lemieux, T. (1996), 'Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach', *Econometrica*, vol. 64, pp. 1001–44.

Giannini, S. and Guerra, M. C. (1999), 'Il sistema tributario verso un modello di tassazione duale' ('The tax system towards a model of dual taxation'), in L. Bernardi (ed.), *La finanza pubblica italiana: rapporto 1999* (*Report on Italian Public Finance: 1999*), Bologna: Il Mulino.

Hall, P. (1992), 'Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density', *Annals of Statistics*, vol. 20, pp. 675–94.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004), *Nonparametric and Semiparametric Models*, Berlin: Springer.

Jenkins, S. P. (1995), 'Did the middle class shrink during the 1980s? UK evidence from kernel density estimates', *Economics Letters*, vol. 49, pp. 407–13.

Jones, M. (1993), 'Simple boundary correction for kernel density estimation', *Statistics and Computing*, vol. 3, pp. 135–46.

Karunamuni, R. and Alberts, T. (2005), 'On boundary correction in kernel density estimation', *Statistical Methodology*, vol. 2, pp. 191–212.

Mantovani, D. (1998), 'Manuale DIRIMOD95' ('DIRIMOD95 manual'), mimeo, Prometeia, Bologna.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008

Marron, J. and Ruppert, D. (1994), 'Transformations to reduce boundary bias in kernel density estimation', *Journal of the Royal Statistical Society, Series B*, vol. 56, pp. 653–71.
Pittau, M. G. and Zelli, R. (2001), 'Income distribution in Italy: a nonparametric analysis', *Statistical Methods and Applications*, vol. 10, pp. 175–89.
Pudney, S. (1993), 'Income and wealth inequality and the life cycle: a non-parametric analysis for China', *Journal of Applied Econometrics*, vol. 8, pp. 249–76.
Quah, D. (1997), 'Empirics for growth and distribution: stratification, polarization, and convergence clubs', *Journal of Economic Growth*, vol. 2, pp. 27–59.
Sheather, S. J. and Jones, M. C. (1991), 'A reliable data-based bandwidth selection method for kernel density estimation', *Journal of the Royal Statistical Society, Series B*, vol. 53, pp. 683–90.
Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.
Sutherland, H. (1991), 'Constructing a tax–benefit model: what advice can one give?', *Review of Income and Wealth*, vol. 37, pp. 199–219.
— (ed.) (2001), 'EUROMOD: an integrated European benefit–tax model – final report', University of Cambridge, Microsimulation Unit, EUROMOD Working Paper no. EM9/01.
Wand, M. P. and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman and Hall.
Zhang, S. and Karunamuni, R. (1998), 'On kernel density estimation near endpoints', *Journal of Statistical Planning and Inference*, vol. 70, pp. 301–16.
—, — and Jones, M. (1999), 'An improved estimator of the density function at the boundary', *Journal of the American Statistical Association*, vol. 94, pp. 1231–41.

© 2008 The Author
Journal compilation © Institute for Fiscal Studies, 2008