



TAMPERE ECONOMIC WORKING PAPERS
NET SERIES

GROUPINGS AND THE GAINS FROM TAGGING

Ravi Kanbur
Matti Tuomala

Working Paper 93
February 2014

SCHOOL OF MANAGEMENT
FI-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191
ISBN 978-951-44-9402-4

Groupings and the Gains from Tagging¹

Ravi Kanbur and Matti Tuomala*

Cornell University, USA and University of Tampere, Finland

8 December, 2013

Contents

1. Introduction
2. A Simple Framework
3. Application to Finnish Data
4. Choice of Groups and Optimal Non-Linear income Taxation
5. Conclusion

Abstract:

The large literature on “tagging” shows that group specific tax and transfer schedules improve welfare over the case where the government is restricted to a single schedule over the whole population. The central assumption, however, is that the groupings available to the government are given and fixed. But how many and which types of groups should the government choose to tag? This is the question addressed in this paper. Starting with a simple framework and ending with numerical simulations based on data from Finland, we show how groupings should be formed for tagging, and provide a quantitative assessment of how group differences affect the gains from tagging, and of the marginal welfare gains from increasing the number of groups being tagged. We hope that these results are the first steps in a richer analysis of tagging which expands the question of design to the arena of choice over groups being tagged.

*E-Mail: Matti.Tuomala@Kolumbus.fi

¹ We are grateful to Spencer Bastani, Sören Blomquist, Bas Jacobs and the seminar participants at Uppsala University for useful comments and suggestions on an earlier version of this paper.

1. Introduction

It is widely recognized that there are potentially severe incentive and other costs of administering income-related transfers. One way of overcoming these costs is to differentiate the population by easily observable indicators that are correlated with the unobservable characteristic of interest. An individual's labour market status or demographic attributes, for instance, may convey information on underlying productivity.² Transfers can then be made contingent upon such characteristics. Akerlof (1978)³ was among the first to recognise that the use of contingent information to implement several tax/transfer schedules, one for each group, was bound to be superior to being restricted to a single schedule for the whole population. However, he did not say much about the quantitative gain from such differentiation, nor about the shapes of the schedules for the different groups.

The two decades following Akerlof's (1978) seminal publication saw the application and extension of the idea in a number of different directions and settings. Kanbur (1987) and Besley and Kanbur (1988) applied the idea to the targeting of anti-poverty transfers in developing countries. Kanbur and Keen (1989) provide some characterizations of linear group specific tax/transfer schedules with incentive effects. The design of distinct nonlinear income tax/transfer schemes for sub-groups of the population linked by intergroup transfers was provided by Immonen, Kanbur, Keen and Tuomala (1998) (hereafter IKKT), with a focus on two key issues: what are the shapes of optimal tax/transfer schedules when categorical information can be used to apply different schedules to different groups, and how substantial are the potential welfare gains from applying distinct schedules to distinct groups? The interplay between income-relation and categorical benefits is also examined by Stern (1982). A number of other papers have considered optimal taxes with tagging. For example, Bennett (1987) explores lump-sum transfers between different types of individuals, and Parsons (1996) studies the optimal benefit structure of an earnings insurance program when "eligibility requirements" are used as a tag to (imperfectly) identify those who are out of work.

The continuing power of the tagging idea is shown by a burgeoning literature post-2000, which has become more specific and considers tagging across different types of groupings. Viard (2001a, b) studies tagging in an optimal linear income tax framework allowing the demigrants to differ across

² Mirrlees (1971) noticed: "One might obtain information about a man's income-earning potential from his apparent I.Q., the number of his degrees, his address, age or colour..."

³ In fact the two-tier social dividend system in Meade report (1978) p.271-276 is a very similar idea.

groups but not the income tax rates; Alesina et al. (2007) advocate tagging based on gender; Blumkin, Margalioth, and Sadka (2009) examine the redistributive role of affirmative action policy, asking whether, supplementing the tax-transfer system with an affirmative action policy would enhance social welfare; Mankiw and Weinzierl (2010) study a model with many skill types who can be tagged on the basis of height; Jacquet and Van der Linden consider stigma in the tagging model; Cremer et al (2010) study the properties of tagging in an optimal income tax framework assuming quasi-linear preferences and a Rawlsian social welfare function; and Boadway and Pestieau (2006) have studied the issue of tagging with optimal income taxation in a two-group-two-skill-level setting.⁴

Following Kremer (2001), age based taxation, in particular, has received especially close attention in the last decade. Banks and Diamond (2010) argued that tagging based on age may be socially acceptable because everyone can reach a given age at some time during their life. The Mirrlees Review (2011) found this argument to be persuasive in advocating some age-related tax reforms to influence labour market participation decisions by older workers and parents with school-age children. Blomquist and Micheletto (2003, 2008) consider age-dependent nonlinear taxation in a dynamic Mirrleesian setting with heterogeneous agents and private savings using an overlapping generations (OLG) model where individuals face a stochastic wage process. Bastani, Blomquist and Micheletto (2013) examine the quantitative implications of implementing an optimal age-dependent nonlinear labor income tax and Weinzierl (2011) similarly provides a quantitative assessment of the welfare gains from age-dependent nonlinear income taxes.⁵

The tagging literature has thus grown, and is growing, by leaps and bounds. But its central assumption is still that the groupings available to the government are given and fixed. The government cannot rearrange these groupings—it cannot increase or decrease the number of groups at the margin, nor can it choose one type of grouping over another. Thus on the one hand the assumption is that the groupings are available to the government without cost, yet on the other hand that it is too costly for the government to deviate from the groupings specified by the analyst. However, if the implementation of tagging is itself costly, and if the costs are a function of the

⁴ Further, Kanbur-Tuomala (2005) analyze optimal aid allocation when the donor is faced with two potential recipient countries with their own specific characteristics. Each recipient government chooses its policies in light of its technology, preferences, and aid allocation. The donor has the task of choosing the aid allocation from a fixed pool of aid resources, to optimize the donor's welfare function. Bastani (2012) explores the optimal tax implications in a model with both singles and couples and inequality across as well as within households.

⁵ Yet other recent analyses of age-dependent taxes include Erosa and Gervais (2002), Gervais (2003), Fennell and Stark (2005), and Lozachmeur (2006).

number and type of grouping available, the question arises—how many and which types of groups should the government choose to tag? This is the question addressed in this paper.

It should be intuitively obvious, and it is clear from the literature, that there are gains of moving from no grouping to some grouping, unless of course the groups chosen are identical to each other. But how do these gains depend on the nature of the groups? How do they depend on the differences between groups? And how do they depend on the number of groups? Answers to these questions are the building blocks for a deeper analysis of the design of tagging, where the groupings can also be chosen by the government. This paper takes the first steps in such an analysis. Starting with a simple framework and ending with numerical simulations based on data from Finland, we show how groupings should be formed for tagging, and provide a quantitative assessment of how group differences affect the gains from tagging, and of the marginal welfare gains from increasing the number of groups being tagged.

The plan of the paper is as follows. Section 2 of the paper sets out a starting framework, with two groups, simple transfers, and no behavioural responses. It derives results for special cases in order to sharpen intuition on the determinants of the gains from grouping. Section 3 introduces Finnish data on the age structure of income distribution, and provides illustrations of the simple results in the previous section. Section 4 moves to a more general framework of optimal non-linear income taxation with labor supply responses, where the optimal grouping problem can only be addressed through numerical simulations based on Finnish data, albeit guided by the intuitions developed in the previous section. Section 4 also takes up the case of more than two groups and, again using Finnish data for application, provide a quantitative assessment of the gains from increasing the number of groups to be tagged. Section 5 concludes.

2. A Simple Framework

In this section we develop a simple framework for assessing the gains from different types of groupings. We assume that there are no behavioural responses and we restrict attention to very simple tax and transfer regimes. The government's objective is to maximize a utilitarian social welfare function. Only two groups are allowed. The question of course is: which two groups? Because of its simplicity, the analytical framework allows us to derive closed form solutions, which in turn help to develop intuitions on what sorts of group differences are relevant for tagging. After

an illustration of the simple results with Finnish data in Section 3, Section 4 presents a more general model which relaxes many of these assumptions.

We focus attention on the case of two mutually exclusive and exhaustive groups, indexed 1 and 2. Let income be denoted z and let density function of income in the groups be $f_1(z)$ and $f_2(z)$ with means \bar{z}_1 and \bar{z}_2 respectively. Let the population shares of the groups be α_1 and α_2 , with $\alpha_1 + \alpha_2 = 1$. The overall density is then

$$f(z) = \alpha_1 f_1(z) + \alpha_2 f_2(z) \quad (1)$$

The government's objective function is given by

$$W = \int u(z) f(z) dz = \alpha_1 \int u(z) f_1(z) dz + \alpha_2 \int u(z) f_2(z) dz \quad (2)$$

where $u(z)$ is an individual level valuation function with $u' > 0$ and $u'' < 0$ in the usual way.

Consider now the simplest case of a group specific tax-transfer regime. A lump sum tax a_1 is imposed on each member of group 1 and the proceeds are used to finance a lump sum payment of a_2 to each member of group 2. The self-financing constraint implies that

$$a_2 = \frac{\alpha_1}{\alpha_2} a_1 \quad (3)$$

Social welfare after the transfer is

$$W = \alpha_1 \int u(z - a_1) f_1(z) dz + \alpha_2 \int u\left(z + \frac{\alpha_1}{\alpha_2} a_1\right) f_2(z) dz \quad (4)$$

and the impact of increasing a_1 on welfare is

$$\frac{dW}{da_1} = -\alpha_1 \left\{ \int u'(z - a_1) f_1(z) dz + \int u'\left(z + \frac{\alpha_1}{\alpha_2} a_1\right) f_2(z) dz \right\} \quad (5)$$

The optimal value of a_1 can be found by setting $\frac{dW}{da_1}$ equal to zero. This solves for a_1 implicitly and we can then find the maximized value of W . Although simple, the structure of the model still does not yield a closed form solution. We can, however, focus attention on small taxes and transfers. Evaluation $\frac{dW}{da_1}$ at $a_1 = 0$ gives us

$$\left. \frac{dW}{da_1} \right|_{a_1=0} = \alpha_1 \left\{ \int u'(z) f_2(z) dz - \int u'(z) f_1(z) dz \right\} \quad (6)$$

This depends solely on α_1 and on the properties of u' , $f_1(z)$ and $f_2(z)$, and can be used to sharpen our intuitions on what types of differences between $f_1(z)$ and $f_2(z)$ will maximize the welfare gain from the introduction of a tagged tax-transfer regime.

The two terms in curly brackets in (6) can be interpreted as the “distributional characteristic” of each group (Feldstein, 1972). The term in curly brackets as a whole is thus a measure of how different the two groups are along this metric. Equation (6) tells us that there are two features which determine groupings which will give the biggest impact on welfare with tagging—how different the groups are in terms of their population shares, and how different the groups are in terms of their distributional characteristic. Now, it might seem from the first feature that it is best to choose one very small and one very large group in terms of population share. But notice that in the limit, as one group comes closer and closer to becoming the whole population the difference in the curly brackets will disappear. There thus appear to be subtle tradeoffs in group choice, which will depend also on the exact form of the valuation function $u(\cdot)$. We now develop a number of special cases to investigate this further.

If $f_2(z)$ is a mean preserving spread of $f_1(z)$, then

$$\left. \frac{dW}{da_1} \right|_{a_1=0} > (<) 0 \Leftrightarrow u''' > (<) 0 \quad (7)$$

Furthermore, for any sign of u''' , the absolute magnitude of the impact of introducing the regime depends on the difference in inequality between the two groups. Among pairs of groupings with the same population share and the same mean, therefore the government should choose the pair with the maximum difference in inequality. It should also be clear that, more generally, a related statement can be made for second order dominance between f_1 and f_2 .

If f_1 and f_2 are not in the relation of a mean preserving spread, then further specification of the functional forms of either f_1 and f_2 , or of $u(\cdot)$, or of both, will be needed to get clear results. Let

$$u(z) = -\left(\frac{z^p - z}{z^p}\right)^\gamma; z \leq z^p \quad (8)$$

$$0; z > z^p$$

for $\gamma \geq 0$. Then W will be recognized to be nothing other than the negative of the famous FGT family of poverty indices (Foster, Greer and Thorbecke, 1984):

$$P_\gamma = \int_0^{z^p} \left(\frac{z^p - z}{z^p}\right)^\gamma f(z) dz \quad (9)$$

Here z^p is the poverty line and γ is interpreted as the degree of poverty aversion. When $\gamma = 0$, P_γ is simply the head count ratio of poverty, the fraction of individuals below the poverty line. When $\gamma = 1$ and $\gamma = 2$, the depth of poverty is also emphasized to different degrees.

Noting

$$u'(z) = \frac{\gamma}{z^p} \left(\frac{z^p - z}{z^p}\right)^{\gamma-1}; z \leq z^p \quad (10)$$

$$0; z > z^p$$

the expression (6) now becomes:

$$\left. \frac{dW}{da_1} \right|_{a_1=0} = \frac{\alpha_1 \gamma}{z^p} \{P_{2,\gamma-1} - P_{1,\gamma-1}\} \quad (11)$$

where the subscript 1 and 2 on P indicates group specific poverty, and the subscript $\gamma-1$ indicates a poverty aversion of $\gamma-1$.

Expressions such as (11) are to be found in the literature on anti-poverty targeting (Kanbur,1987, Besley-Kanbur, 1988). For our purposes what it shows is that if its objective is to minimize poverty P_γ , then for given population shares the government should choose groups with the biggest difference in $P_{\gamma-1}$. Thus if the objective is to minimize the poverty gap measure P_1 , the transfer should be across groups with the biggest difference in P_0 -in other words the biggest differences in the head count ratio.

Further simplification of the form of $u(\cdot)$ to a quadratic provides a particularly simple result and interpretation. Let

$$u(z) = -(z - \bar{z})^2 \quad (12)$$

where

$$\bar{z} = \alpha_1 \bar{z}_1 + \alpha_2 \bar{z}_2 \quad (13)$$

is the overall mean income. Thus the government's objective function is to minimize national variance through the mean preserving transfers across groups. In this case

$$\left. \frac{dW}{da_1} \right|_{a_1=0} = 2\alpha_1 (\bar{z}_1 - \bar{z}_2) \quad (14)$$

Thus if the government is restricted to only two groups, then it should choose groups with the largest differences in means holding population shares constant, or largest difference in population shares holding differences in mean constant.

A similar focus on group means arises if the densities $f_1(z)$ and $f_2(z)$ are assumed to be lognormal densities:

$$f_i(z) \sim \Lambda(m_i, \sigma_i^2), i = 1, 2 \quad (15)$$

where m_i and σ_i^2 are the mean and variance respectively of $\log z$ in the two groups. We further assume that

$$u(z) = \frac{1}{1-r} z^{1-r} \quad (16)$$

in other words, a utility function with constant relative inequality aversion r . In this case it can be shown that

$$\left. \frac{dW}{da_1} \right|_{a_1=0} = \alpha_1 [e^{-rm_1 + 0.5r^2\sigma_1^2}] [e^{r(m_1 - m_2) - 0.5r^2(\sigma_1^2 - \sigma_2^2)} - 1] \quad (17)$$

Thus once again the distributional difference between the groups matters, but the key metric is now

$$r(m_1 - m_2) - 0.5r^2(\sigma_1^2 - \sigma_2^2) \quad (18)$$

Notice that when relative inequality aversion is unity, in other words the utility function is logarithmic, then the metric collapses to

$$(m_1 + 0.5\sigma_1^2) - (m_2 + 0.5\sigma_2^2) \quad (19)$$

which again says that the government should use groupings with the largest difference in mean.

The final special case we consider is that of the utility function with constant absolute inequality aversion

$$u(z) = -\frac{1}{g} e^{-gz} \quad (20)$$

where g is the absolute inequality aversion parameter. Substituting this in Equation 6 still does not give a closed form solution. While this can be calculated for empirical distributions, as it will be in the next sub-section, further specification is needed for an analytical closed form. With this in mind, let the densities be exponential:

$$f_i(z) = h_i e^{-h_i z}, i = 1, 2 \quad (21)$$

Then

$$\left. \frac{dW}{da_1} \right|_{a_1=0} = \alpha_1 \left\{ \frac{h_1}{g+h_1} - \frac{h_2}{g+h_2} \right\} \quad (22)$$

Thus the impact is greatest when the two densities are most different from each other, as measured by the difference between their exponential parameters h_1 and h_2 .

3. Application to Finnish Data on Age Structure of Income Distribution

What do the expressions developed in the previous section look like for actual data? In this section we present an application to Finnish data, focusing on age based groupings. This will also allow us to introduce the data we will use in the rest of the paper.

Estimates on age structure of income distribution from 1990 to 2007 are calculated from the Income Distribution Statistics (IDS) data source for Finland. The IDS data is based on representative national sample survey of around 9000-11000 households drawn from households in Finland. The IDS contains information on incomes, taxes and benefits together with various socio-economic characteristics of the Finnish households. Most of the information contained in the IDS has been

collected from various administrative registers. Auxiliary information is collected through interviews. Examples of how this data has been used previously in other contexts are in Riihelä, Sullström and Tuomala (2008, 2012). These papers also provide further detail on the specific properties of the data and its sources.

We begin with a preliminary look at basic patterns by age. Figure 1 shows pre-tax mean incomes by age in Finland in 1990, 2000 and 2007. 1990 and 2000 display the conventional inverse U-shaped pattern—the turning point is less clear in 2007. Figure 2 in turn displays Gini coefficients within age groups. Within age-group inequality is higher in younger and older groups compared to in the middle age groups.

Figure 1 Pre tax income (mean) in age groups (excl. pensioners, unemployed and students etc.) in Finland 1990,2000, 2007

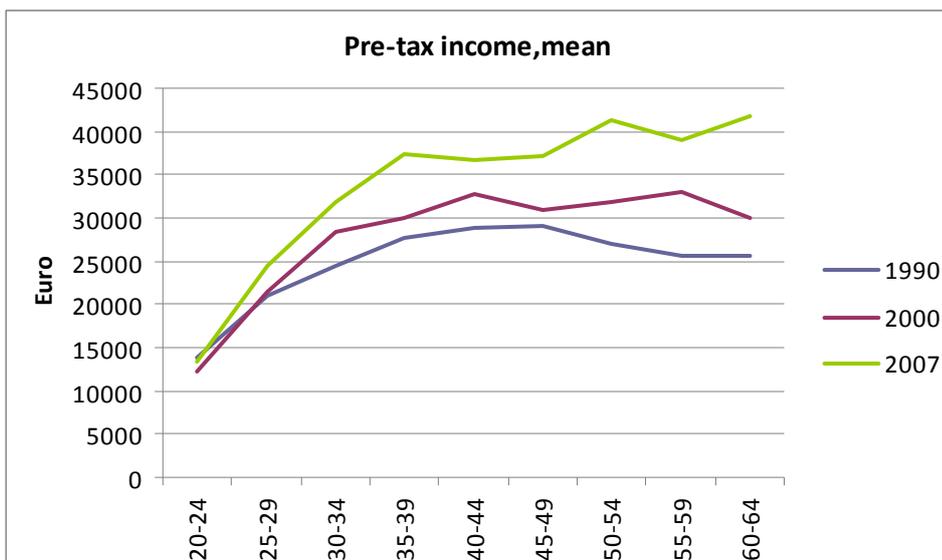


Figure 2 Gini coefficient for pre tax income in age groups (excl. pensioners, unemployed and students etc.) in Finland 1990,2000,2007



In what follows it will prove useful to represent within-cohort income distribution, as well as the overall income distribution, through a particular functional form. Extensive empirical work has shown that Finnish data are well represented by a two parameter Champernowne distribution:

$$f(z) = \theta \left(\frac{m^\theta z^{\theta-1}}{(m^\theta + z^\theta)^2} \right) \text{ and } F(z) = 1 - \frac{m^\theta}{(m^\theta + z^\theta)} \quad (23)$$

where m is a parameter of central tendency and θ is a parameter of spread or inequality. Among two parameter distributions the Champernowne distribution is the best fitting for pre-tax income distribution in Finland (1990-2010). The θ -parameter varies from 2.78 to 2.34. Over the period from the latter part of 1990's to 2010 the θ -parameter was almost constant being around 2.5 (see Figure 3). Hence $\theta = 2$ reflects a low range estimate (high inequality) and $\theta = 3$ in turn a high range estimate (low inequality). The Gini coefficients estimated by this distribution ($\text{Gini} = 1/\theta$) are quite close to those calculated from the data. The location parameter m (median) in the Champernowne distribution is also quite close to that calculated from the data. The Champernowne distribution also fits well for income distribution within age groups (Riihelä, Sullström, Tuomala,

2013). For this reason in the rest of the paper we will use the estimated Champernowne distribution to represent income distribution within age groups and nationally.

Figure 3 The shape parameter θ (with confidence interval): Champernowne distribution (Riihelä,Sullström,Tuomala,2013)

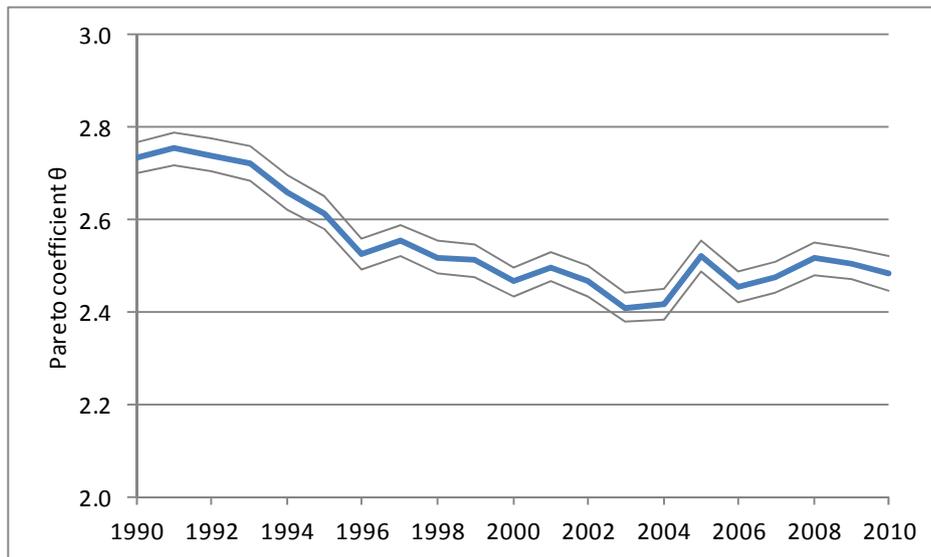
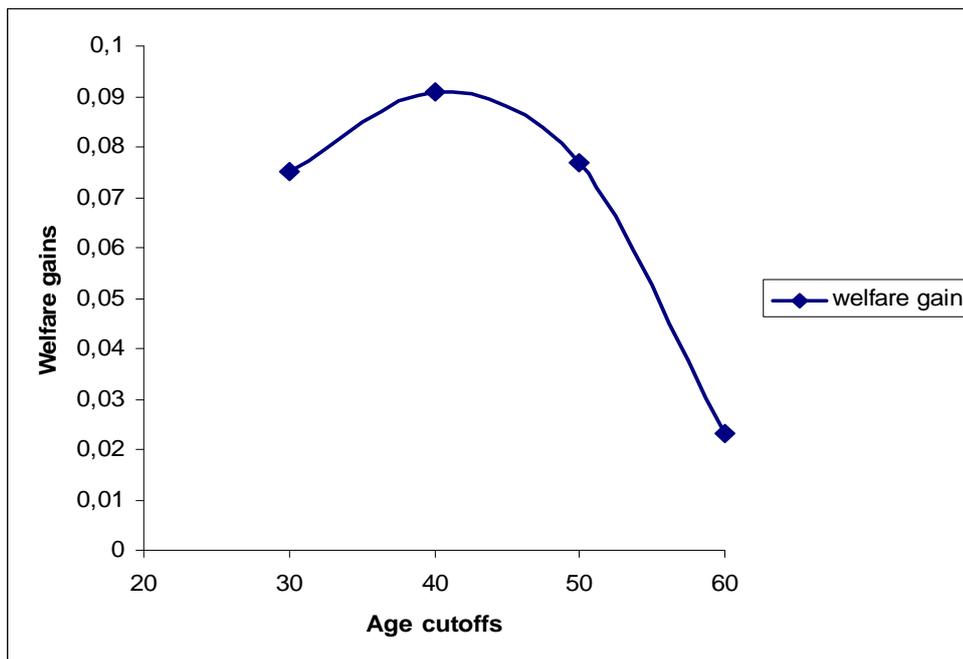


Figure 2 shows that within cohort inequality tends to have an inverse-U shape, with inequality being lowest in the middle age cohorts. Thus if we want to form two groups with disparate inequality, we would combine the very young and the very old into one group and keep the middle age cohorts in another group. However, policy typically works through dividing the population into groups ranked by age. Thus if we wanted to form two groups of young versus old with most disparate inequalities, it is not a priori obvious from Figure 2 where the cut-off should be drawn. The same is true if our focus was difference in group means (Figure 1). In any event, as we know from (6) it is not just these differences between groups which matters; the relative population in the groups which will also matter in determining the gains from tagging. In what follows we present a series of quantitative assessments for an age cut-off of 30, 40, 50 and 60 years.

Let us start then with the case where the government's objective is to minimize national variance through mean preserving transfers between two groups, which leads to the criterion given by (14). Which groups should these be—in other words, which age cut-off maximizes the value of (14) in Finnish data? Figure 4 provides the answer—the cut off which maximizes the gain to tagging across the groups it creates occurs at 40 years. What if the objective is to minimize the poverty gap, which

leads to criterion (11) with $\gamma = 1$? Then Figure 5 tells us that the cut-off to use for tagging is 30 years. Finally, consider the case where the objective is a utilitarian objective function with exponential utility function (20) with $g = 0$ or $g = 1$. Figure 6 shows the value of the criterion (6) for the different age cut-offs⁶. It is seen that for both $g = 0$ and for $g = 1$, the optimum cut-off to create groups for tagging is 40 years.

Figure 4 The government's objective function is to minimize national variance through the mean preserving transfers across groups (estimates are based on year 2007).



⁶ Here age cut-offs are 30, 40 and 50 years

Figure 5 The government's objective is to minimize poverty gap (estimates are based on year 2002)

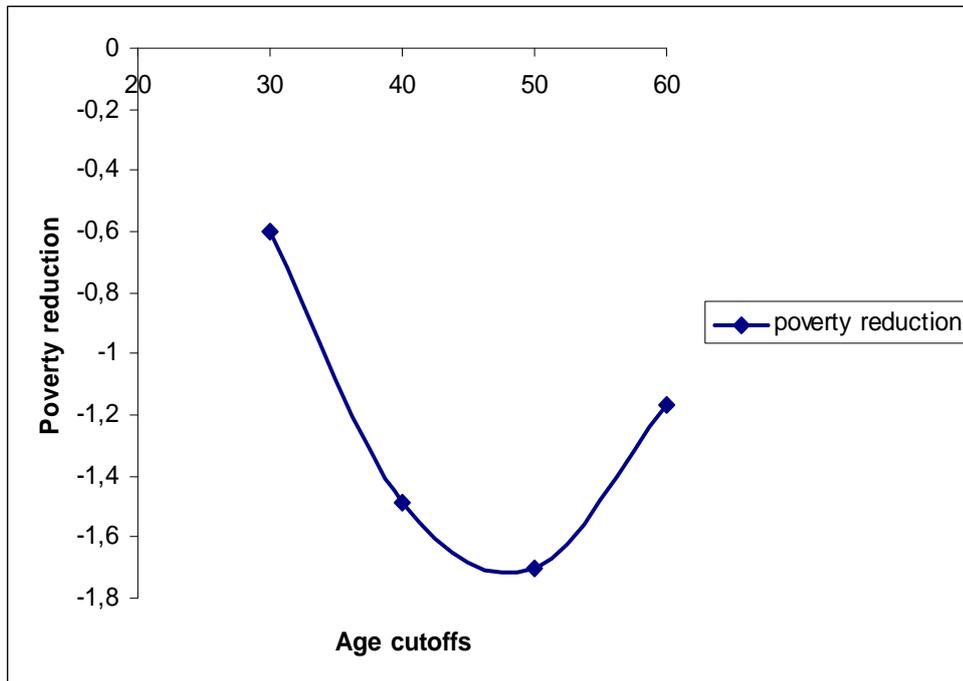
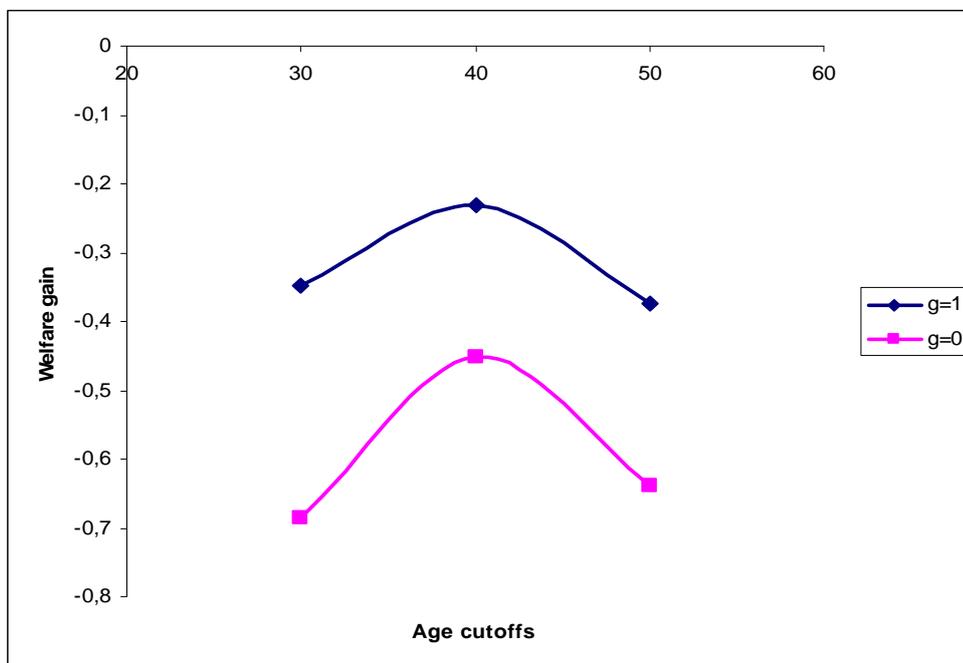


Figure 6 Equation (6) with exponential utility function (20) with $g=0$ and $g=1$



Thus, when we apply the simple framework of the previous section to actual data, we are able to give concrete form to the basic intuitions embodied in the expressions developed in the theoretical analysis. The application shows that optimal groupings for tagging can indeed be identified, and that they will change as the objective function of the government changes. The next section moves beyond the simple analysis by relaxing the many assumptions underlying it.

4. Choice of Groups and Optimal Non-linear Income Taxation

The simple analytical framework of the last section, and the special functional forms used there, are useful for developing and sharpening intuition. However, they are clearly special in (i) the form of the tax-transfer regime, (ii) the government's objective function, (iii) the distributional forms used and, perhaps most important, (iv) the assumption of no behavioural responses. In this section we turn to a more general formulation where these restrictions are relaxed. We do this by setting the problem of choosing groups in the Mirrlees (1971) framework of optimal non-linear income taxation.

Suppose as before that the population (the size of which is normalized to unity) can be divided into two mutually exclusive and exhaustive groups, labelled 1 and 2. Individuals are unable to alter or disguise the group to which they belong, which is observed costlessly by the government. Members of each group i ($i=1, 2$) have preferences $u_i = U(x) + V(1-y)$ defined over consumption x and labour supply y , but differ in their hourly gross wage (alternatively, their skill or ability), n , with $U_x > 0$ and $V_y < 0$ (subscripts indicating partial derivatives). Individuals differ only in the pre-tax wage n they can earn. Gross income is $z = ny$. The groups differ in the distribution of abilities, the latter being described for each group by a continuous density function f_i (with corresponding distribution $F_i(n)$) on support $[\underline{n}, \bar{n}]$. The within-group structure of the model is thus exactly as in Mirrlees (1971).

Suppose that the aim of policy is to design tax/benefit schedules $T_i(z)$ for two different groups $i=1, 2$ to maximize the following social welfare criterion

$$W = \int_{\underline{n}}^{\bar{n}} \sum \alpha_i (G(u_i(n)) f_i(n) dn, \quad (24)$$

where α_i denotes the proportion of the population in each group i and G is an increasing and concave function of utility. The government cannot observe individuals' productivities and thus is restricted to setting taxes and transfers as a function only of earnings, $T_i(z)$. The government maximizes W subject to the revenue constraint

$$\int_{\underline{n}}^{\bar{n}} \sum \alpha_i (z_i(n) - x_i(n)) f_i(n) dn = R \quad (25)$$

and the second constraint making use of workers utility maximization condition - in each group, a person with wage n chooses y to maximize u_i subject to $x = ny - T_i(ny)$, giving us the incentive compatibility constraint constraints,

$$\frac{du_i}{dn(i)} = \frac{y_i V_{y(i)}}{n(i)} \text{ for } i=1,2. \quad (26)$$

As shown in IKKT (1998), it is helpful to think of this problem as consisting of two steps. First we derive group specific optimal tax schedules, given a group specific revenue requirement R_i . This means solving the standard Mirrlees problem for each group. A number of treatments (for example, Tuomala, 1990) set out how this is done, and the implications for the tax schedule.

Our focus, however, is on the gains from having two tagged groups rather than being forced to apply a single schedule to the population as a whole. This takes us to the second step. Given the solution the first step, the government chooses the optimal allocation of the aggregate R over groups; in other words it chooses R_i to maximize overall W , having first maximized each W_i for given total R_i .

Let us now apply this framework to the specific case of the data on the age structure of income distribution in Finland. We begin by specifying social objectives further. Social welfare is taken to be utilitarian, so that

$$G(u) = u \quad (27)$$

We assume identical individual preferences of the form

$$u = -\frac{1}{x} - \frac{1}{(1-y)} \quad (28)$$

implying an elasticity of substitution between consumption and leisure in each groups of 0.5.

Thus the two tagged groups differ only in their distribution of abilities. As noted in the previous section, we assume that pre-tax incomes follow a Champernowne distribution nationally as well as within each age group. Inference of parameters from observed empirical earnings distributions is a long-standing issue in the optimal income taxation literature. A number of methods have been proposed, each with its own strengths and weaknesses. Saez (2001) calibrates the exogenous ability distribution such that the actual $T(\cdot)$ yields empirical income distribution. To calculate the optimal tax schedule, Saez makes additional assumptions about the model's structure. He assumes that the labour elasticity is constant. Given this utility function he infers the ability distribution from the empirically observed distribution of incomes in the current tax regime (assuming a linear tax schedule). However, the strong assumptions required for structural identification of the model reduce the confidence of the optimal tax schedule calculations. Alternatively, Kanbur-Tuomala (1994) calibrate the skill distribution indirectly so that the income distribution inferred from the skill distribution matches the actual distribution. Using this procedure there is no need to narrow further the set of functional forms used in simulations. We follow this route in our illustrations on tagging with optimal non-linear income taxation.

We begin by dividing the Finnish working population into two groups on the basis of age with different age cutoffs. As in Sections 2 and 3, the cut-offs are 30, 40 and 50 years. For each cutoff, we calculate the welfare gain from using that grouping. The welfare gain reported in Table 1 is the proportional increase in equivalent consumption in moving from the optimal single schedule to the

optimal group-specific schedules. Table 1 immediately gives us the answer to question to which grouping is best for tagging—it is the one which uses the age cut-off of 40 years.

Table 1 Welfare effects of different age cutoffs (Estimates of θ and m are based on year 2007)

	utilitarian	
	x^0	change %
Single schedule	0.163	
two groups*	0.167	1.81
two groups**	0.170	4.10
two groups***	0.165	1.23

x^0 : Consumption equivalent: x^0 is that consumption which, if equally distributed with zero work hours, would give the same social welfare integral as the allocation $((x, y,))$ arising from a given tax schedule.

Single: $\theta=2.5$, $m=0.368$

Two groups *: group 1 [$\theta_1=2.3$, $m_1=0.202$, $\alpha_1=0.21$], group 2 [$\theta_2=2.6$, $m_2=0.407$, $\alpha_2=0.79$], cutoff at age 30

Two groups **: group 1 [$\theta_1=2.4$, $m_1=0.317$, $\alpha_1=0.41$], group 2 [$\theta_2=2.7$, $m_2=0.417$, $\alpha_2=0.59$], cutoff at age 40

Two groups ***: group 1 [$\theta_1=2.5$, $m_1=0.333$, $\alpha_1=0.64$], group 2 [$\theta_2=2.6$, $m_2=0.427$, $\alpha_2=0.36$], cutoff at age 50

The discussion so far has maintained the number of groups at two. But each of these groups could be further sub-divided, until there as many tax schedules as individuals. Of course if increasing the number of instruments in this way was costless, it would make sense to do so because welfare cannot decrease with more instruments available. However, what if instruments are costly—what if the costs of distinguishing between and monitoring across groups increases as a function of the number of groups? Then it would be optimal to limit the number of groups to well before the point where each individual is a group. But how many groups is optimal? The answer depends on the costs of administering each additional group and, crucially, the marginal welfare gain from increasing the number of groups. We now turn to quantifying the gains from additional groups, in the specific context of our Finnish data set.

We proceed as follows. We already know the welfare levels as a result of the optimal use of tagging for when there is only one group, and when there are two groups with an age cut off at 30. We will now calculate the welfare level with three groups (under 30, between 30 and 40, and over 40) and

four groups (under 30, between 30 and 40, between 40 and 50, and over 50). In each case we calculate the welfare when the government uses all the information available to tag groups and implements separate non-linear income tax schedules for each group to maximize overall social welfare. These welfare levels are given in Table 2.

Table 2 Welfare levels of increasing the number of groups (Estimates of θ and m are based on year 2007)

utilitarian		
	\mathbf{x}^0	change %
Single schedule	0.163	
2 groups*	0.167	2.4
3 groups**	0.168	3.0
4 groups***	0.1684	3.21

\mathbf{x}^0 : Consumption equivalent: \mathbf{x}^0 is that consumption which, if equally distributed with zero work hours, would give the same social welfare integral as the allocation $((x, y,))$ arising from a given tax schedule.

Two groups: group 1 [$\theta_1=2.3, m_1=0.202, \alpha_1=0.21$], group 2 [$\theta_2=2.6, m_2=0.407, \alpha_2=0.79$],

Three groups: group 1 [$\theta_1=2.3, m_1=0.202, \alpha_1=0.21$], group 2 [$\theta_2=2.4, m_2=0.317, \alpha_2=0.20$], group 3 [$\theta_3=2.7, m_3=0.427, \alpha_3=0.59$],

Four groups: group 1 [$\theta_1=2.3, m_1=0.202, \alpha_1=0.21$], group 2 [$\theta_2=2.4, m_2=0.317, \alpha_2=0.20$], group 3 [$\theta_3=2.6, m_3=0.407, \alpha_3=0.23$], group 4 [$\theta_4=2.5, m_4=0.387, \alpha_4=0.36$]

It should be clear from Table 2 that there are strong diminishing returns to increasing the number of groups. For this utilitarian case, welfare compared to the single group case increases by 2.4% with the introduction of two groups, but only a further 0.6% of the base welfare is added when the groupings are increased to three, and going from three groups to four groups only gives an additional 0.21%. Thus the gains from increasing the number of groupings fall off quite rapidly.

5. Conclusions

The large literature on “tagging” shows that group specific tax and transfer schedules improve welfare over the case where the government is restricted to a single schedule over the whole population. The central assumption, however, is that the groupings available to the government are

given and fixed. But how many and which types of groups should the government choose to tag? This is the question addressed in this paper. Starting with a simple framework and ending with numerical simulations based on data from Finland, we show how groupings should be formed for tagging, and provide a quantitative assessment of how group differences affect the gains from tagging. We also provide a quantitative assessment of the welfare gains from increasing the number of tagged groups. We hope that these results are the first steps in a richer analysis of tagging which expands the question of design to the arena of choice over groups being tagged

References

- Akerlof, G. A. (1978). The economics of 'tagging' as applied to the optimal income tax, welfare programs and manpower planning. *American Economic Review*, 68, 8-19.
- Alesina, A, A, Ichino, and L, Karabarbounis. 2007. "Gender Based Taxation and the Division of Family Chores." National Bureau of Economic Research Working Paper 13638.
- Bastani,S. Blomquist,S and L Micheletto (2013),The Welfare Gains of Age Related Optimal Income Taxation, *International Economic Review*.
- Bastani,S, (2013), Gender-Based and Couple-Based Taxation, *International Tax and Public Finance*, Volume 20, Issue 4, p. 653-686
- Bennett, J. 1987. "The Second-Best Lump-Sum Taxation of Observable Characteristics." *Public Finance*, 42(2): 227 - 35.
- Besley, T. (1990). Means testing versus universal provision in poverty alleviation programmes. *Economica*, 57, 119-29.
- Besley, T. and Kanbur, R. (1988) 'Food subsidies and poverty alleviation', *The Economic Journal* 98, 701-719.
- Blomquist, S, and L Micheletto. (2008). Age-related optimal income taxation. *The Scandinavian Journal of Economics*, 110(1):pp. 45–71.
- Blumkin, T, Y Margalioth, and E. Sadka (2009), Incorporating Affirmative Action into the Welfare State, *Journal of Public Economics* 93, pp. 1027-1035.
- Boadway, R. and Pestieau, P. (2006). Tagging and redistributive taxation. *Annales d'Economie et de Statistique*, 83/84:123–147.
- Cowell, F. (1986). Welfare benefits and the economics of take-up. Working paper no. 89, ESRC Programme on Taxation, Incentives and the Distribution of Income, London School of Economics.
- Cremer, H., Gahvari, F., and Lozachmeur, J. M. (2010). Tagging and income taxation: theory and an application. *American Economic Journal: Economic Policy*, 2(1):31–50.
- Feldstein, M. 1972 Distributional equity and the optimal structure of public prices, *American Economic Review* 62:32-36.
- Gervais, M.(2003), On the Optimality of Age-Dependent Taxes and the Progressive US Tax System. Unpublished. <http://aix1.uottawa.ca/~vbarham/PT03.pdf> .
- Hamilton, J, and P. Pestieau. 2005. "Optimal Income Taxation and the Ability Distribution:Implications for Migration Equilibria." *International Tax and Public Finance*, 12(1): 29 - 45.

- Immonen, R., Kanbur, R., Keen, M., and Tuomala, M. (1998). Tagging and taxing: The use of categorical and income information in designing tax/transfer schemes. *Economica*, 65:179–192.
- Kaplow, L. 1989. “Horizontal Equity - Measures in Search of a Principle.” *National Tax Journal*, 42(2): 139-54.
- Kanbur, R. (1987). Transfers, targeting and poverty. *Economic Policy*, 4, 112-36 and 141-7. and Kanbur, R. and Keen, M. J. (1989). Poverty, incentives and linear income taxation. In A. Dilnot and I. Walker (eds.), *The Economics of Social Security*. Oxford: Oxford University Press. and Kanbur, R. Keen, M. and Tuomala, M. (1994). Optimal non-linear income taxation for the alleviation of income poverty. *European Economic Review*, 38, 1613-32.
- Kanbur R- M Tuomala, (2006), Incentives, Inequality and the allocation of aid when conditionality doesn't work, p.331-352, in *Poverty, Inequality and Development; Essays in Honor of Eric Thorbecke*, edited by Alain de Janvry and Ravi Kanbur, Kluwer Academic Publishers, USA
- Keen, M. J. (1992). Needs and targeting. *Economic Journal*, 102, 670-9.
- Kremer, M. (2001) Should Taxes be Independent of Age?
<http://www.economics.harvard.edu/faculty/kremer/papers.html>.
- Lozachmeur, J. M. (2006), "Optimal Age Specific Income Taxation", *Journal of Public Economic Theory*, vol. 8, n. 4, p. 697-711.
- Lydall, H. F. (1968). *The Structure of Earnings*. Oxford: Clarendon Press.
- Mankiw, N. G. and Weinzierl, M. (2010). The optimal taxation of height: A case study of utilitarian income redistribution. *American Economic Journal: Economic Policy*, 2(1):155–76.
- Meade Report (1978), *The Structure and Reform of Direct Taxation*. George Allen&Unwin, IFX, London.
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38, 175-208.
- Riihelä M, Sullström R and M Tuomala, (2008), *Economic Poverty in Finland 1971-2004*, *Finnish Economic Papers* 21,57-77.
- Riihelä M, Sullström R and M Tuomala (2012) *Top Incomes and Top Tax Rates: Implications for Optimal Taxation of Top Incomes in Finland*, Tampere Economic Working Papers, Net Series , University of Tampere.
- Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68:205–229.
- Stern, N. (1976). On the specification of optimum income taxation. *Journal of Public Economics*, 6, 123-62.

Stern, N. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, 17, 181-211.

Tuomala, M. (1990). *Optimal Income Tax and Redistribution*. Oxford: Clarendon Press.

Weinzierl, M. C. (2011). The surprising power of age-dependent taxes. *Review of Economic Studies*, 78(4):1490–1518.