

Piecewise Linear Taxation and Top Incomes

Yuri Andrienko
Sydney University Law School

Patricia Apps
Sydney University Law School and IZA

Ray Rees
University of Munich, University of Oslo, University of Warwick
and CESifo

30 April 2014

Abstract

This paper applies the model of optimal piecewise linear taxation to the issue of the taxation of top incomes. In a number of high-income countries there has been a large growth in inequality due to rising top incomes and a shift in the burden of taxation from the top to the middle of the distribution. Our results suggest that the appropriate response to rising inequality is a shift towards a more progressive multi-bracket income tax system, with relatively low rates in the lower half of the distribution and a highly progressive structure of rates towards the top percentiles.

Acknowledgements: The research was supported by the Australian Research Council's Discovery Project funding scheme (DP1094021).

Keywords Optimal taxation, income distribution, top incomes, inequality

JEL Classification H21, H24, D31, D63

Corresponding Author:

Patricia Apps, Faculty of Law, University of Sydney
NSW 2006, Australia

E: patricia.apps@sydney.edu.au

1 Introduction

The substantial growth in wage and income inequality in high-income countries from the early 1980's to the late 2000's, in particular the large increase in the income shares of the top 1% and 0.1% of income earners in these countries,¹ together with the fact that over the same period top tax rates in virtually all middle and high income countries have been falling rapidly,² has led to increased interest in the question of the appropriate levels of taxation of top incomes.

There is a wide divergence of opinion. In the UK for example, the Mirrlees Review of the income tax system³ accepted the argument that the top tax rate should not be raised, and indeed went further in proposing that the "normal rate of return to saving" should be tax-exempt.⁴ Since wealth in forms of assets other than home ownership and pension rights (taxation of which would be essentially left unchanged under the Review's proposals) is largely held by higher income households,⁵ this should also be construed as advocating a reduction in the relative tax burden on top incomes.

The contribution by Piketty, Saez and Stantcheva (2011) on the other hand comes to the conclusion that top tax rates should be significantly increased. They base their argument on a three-part decomposition of the supposed high degree of responsiveness of top incomes to taxation, which on conventional tax-theoretic grounds would be construed as arguing for low tax rates. They argue that a reduction in reported high incomes following an increase in the top tax rate would be composed of: a fall in labour supply or effort of the type usually considered in economic models; an increase in tax avoidance and evasion as income is underreported or diverted to forms which are subject to lower tax rates; and a fall in top incomes due to weakened bargaining power and consequently a lower share of rents, for example of senior executives in diverting rents from company shareholders to themselves. They then argue that the first of these components should be the main determinant of the level of the top tax rate, and that empirically this is very small. The second should be dealt with not by low tax rates but by dealing directly with the issues of tax avoidance and evasion. The third is actually an argument for higher tax rates, because of the inefficiencies involved in conflict over rents.⁶ Overall there is a good case for an increase in top tax rates, thus reserving, at least in part, the recent trends. These arguments may prove controversial, but do serve to move the debate in an important new direction.

In this paper we take the view that the top tax rate should be analysed in a theoretical framework of optimal taxation that addresses the entire income distribution, and so produces an optimal structure of marginal tax rates for this

¹See Atkinson et al. (2011) for a survey of the recent literature.

²Peter, Buttrick and Duncan (2010) document this for a large sample of countries.

³See Mirrlees et al., (2011).

⁴This reflects the position taken in a large body of theoretical literature. For a concise but comprehensive survey of this see Boadway (2012) Ch. 3.

⁵See Mirrlees et al (2011) for comprehensive UK data on this.

⁶It could also be an argument for higher profits taxes in rent-generating sectors such as financial markets. See also Bivens and Michel (2013).

distribution. However, the standard theoretical approaches to optimal taxation, based respectively on the mechanism design approach and the optimal linear tax model, do not provide the most empirically relevant or, in the case of the former, tractable way of doing this. Instead, we develop the approach introduced by Sheshinski (1989), which analyses the two-bracket optimal piecewise linear tax system.⁷ In this paper we extend the theoretical analysis to an arbitrary number of brackets and provide numerical calculations of the results for up to a 4-bracket tax system for the US, UK and Australia.

Given distributions of wages and earned incomes that are a reasonable approximation to the current empirical distributions, our results support the argument not only for a high tax rate at the very top, but also for a high degree of progressivity in the tax system overall, achieved by a multi-bracket system with relatively low tax rates in the lower half of the income distribution and multiple tax brackets, with a highly progressive rate structure, in the upper half. When we go on to introduce a further growth in inequality due to rising top wage rates, we find these results are strengthened. This represents a sharp reversal of the current trend towards a tax system in which the top rate is reached at incomes in the middle deciles of the distribution and then remains at a flat rate for all higher incomes.

The paper is set out as follows. In the next section we introduce the household model and analyse the choices of the individual income earner under a given m -bracket piecewise linear tax system, with $m \geq 2$. This forms the basis for the optimal tax analysis in Section 3. In Section 4 we describe the way in which we calibrate our numerical model and report the results of the numerical calculations of the welfare-optimal 2-, 3- and 4-bracket tax systems. Section 5 concludes.

2 Individual Choice Problems

Consumers have identical quasilinear utility functions⁸

$$u = x - c(l) \quad c' > 0, \quad c'' > 0 \quad (1)$$

where x is consumption and l is labour supply. Gross income is $y = wl$, with the wage rate $w \in [w_0, w_1] \subset \mathbf{R}_{++}$. Given an m -bracket tax system with parameters $(a, t_1, \dots, t_m, \hat{y}_1, \dots, \hat{y}_{m-1})$, with a the lump sum payment to all households, t_j the marginal tax rate in the j 'th bracket, $j = 1, \dots, m$, and \hat{y}_j the income level determining the upper limit of the j 'th bracket, $j = 1, \dots, m - 1$, the consumer faces the piecewise linear budget constraint defined by:

$$x \leq a + (1 - t_1)y \quad 0 < y \leq \hat{y}_1 \quad (2)$$

$$x \leq a + (1 - t_2)y + (t_2 - t_1)\hat{y}_1 \quad \hat{y}_1 < y \leq \hat{y}_2 \quad (3)$$

⁷See also Dahlby (1998), (2008), and Apps, Long and Rees (2011) and the literature cited there.

⁸Thus we are ruling out income effects. This considerably clarifies the results of the analysis.

$$x \leq a + (1 - t_m)y + \sum_{k=2}^m (t_k - t_{k-1})\hat{y}_{k-1} \quad \hat{y}_{m-1} < y \quad (4)$$

We can write this in the general form

$$x \leq a + (1 - t_j)y + b_j \quad \hat{y}_{j-1} < y \leq \hat{y}_j \quad j = 1, \dots, m \quad (5)$$

where

$$b_j \equiv \sum_{k=1}^j (t_k - t_{k-1})\hat{y}_{k-1} \quad j = 1, \dots, m \quad (6)$$

and we adopt the notational conventions $t_0 = \hat{y}_0 = 0$, $\hat{y}_m \equiv \infty$. Note therefore that $b_1 \equiv 0$, and that we also have

$$\frac{\partial b_j}{\partial t_j} = \hat{y}_{j-1}; \quad j = 1, \dots, m \quad (7)$$

$$\frac{\partial b_j}{\partial t_k} = -(\hat{y}_k - \hat{y}_{k-1}); \quad \frac{\partial b_j}{\partial \hat{y}_k} = (t_{k+1} - t_k); \quad j = 2, \dots, m, \quad k = 1, \dots, j-1 \quad (8)$$

We assume a differentiable wage distribution function, $F(w)$, with continuous density $f(w) > 0$, strictly positive for all $w \in [w_0, w_1]$.

An important further assumption we make is:

Under any piecewise linear tax system under discussion the consumer's budget set in the (x, y) -plane is convex

That is, $t_j > t_{j-1}$, $j = 2, \dots, m$, so that we rule out the case in which marginal tax rates fall across tax brackets. As well as having considerable analytical advantages, the results in Apps, Long and Rees (2011) suggest that this assumption is reasonable in the light of the empirical distributions of wage rates and earned income which currently prevail in developed economies.

Given the consumer's problem of choosing optimal consumption and earnings (labour supply) under a given tax system there are two types of solution possibility:⁹

(i) Optimal income $y^* \in (\hat{y}_{j-1}, \hat{y}_j)$, $j = 1, \dots, m$

In that case we have the first order condition

$$1 - t_j - c'(\frac{y^*}{w})\frac{1}{w} = 0 \quad (9)$$

which yields the solution

$$y^* = \phi(t_j, w) \quad (10)$$

⁹It is assumed throughout that all consumers have positive labour supply in equilibrium. It could of course be the case that for some lowest sub interval of wage rates consumers have zero labour supply. We do not explicitly consider this case but it is not difficult to extend the discussion to take it into account.

giving in turn the indirect utility function

$$v(a, t_1, \dots, t_j, \hat{y}_1, \dots, \hat{y}_{j-1}, w) = a + (1 - t_j)\phi(t_j, w) + b_j - c\left(\frac{\phi(t_j, w)}{w}\right) \quad j = 1, \dots, m \quad (11)$$

Applying the Envelope Theorem to (11) yields the derivatives

$$\frac{\partial v}{\partial a} = 1; \quad \frac{\partial v}{\partial t_j} = -[\phi(t_j, w) - \hat{y}_{j-1}]; \quad \frac{\partial v}{\partial \hat{y}_j} = 0, \quad j = 1, \dots, m \quad (12)$$

$$\frac{\partial v}{\partial t_k} = -(\hat{y}_k - \hat{y}_{k-1}); \quad \frac{\partial v}{\partial \hat{y}_k} = (t_{k+1} - t_k), \quad k = 1, \dots, j - 1 \quad (13)$$

and note also that equilibrium utility is increasing with wage type

$$\frac{dv}{dw} = c'\left(\frac{y}{w}\right)\frac{\phi(t_j, w)}{w^2} > 0 \quad (14)$$

We define the unique values of the wage types \tilde{w}_j, \bar{w}_j by

$$\hat{y}_j = \phi(t_j, \tilde{w}_j) = \phi(t_{j+1}, \bar{w}_j), \quad j = 1, \dots, m - 1 \quad (15)$$

(ii) Optimal income $y^* = \hat{y}_j, j = 1, \dots, m - 1$. In that case the consumer's indirect utility is

$$v(a, t_1, \dots, t_j, \hat{y}_1, \dots, \hat{y}_{j-1}, w) = a + (1 - t_j)\hat{y}_j + b_j - c\left(\frac{\hat{y}_j}{w}\right) \quad (16)$$

and the derivatives of the indirect utility function are as in (12) and (13) above, except that:

$$\frac{\partial v}{\partial t_j} = -(\hat{y}_j - \hat{y}_{j-1}); \quad \frac{\partial v}{\partial \hat{y}_j} = (1 - t_j) - c'\left(\frac{\hat{y}_j}{w}\right)\frac{1}{w} \geq 0 \quad (17)$$

The last inequality, $\partial v / \partial \hat{y}_j \geq 0$, necessarily holds because these consumers, with the exception of types \tilde{w}_j , are effectively constrained at \hat{y}_j , in the sense that they would strictly prefer to earn extra gross income if it could be taxed at the rate t_j , since $c'(\hat{y}/w) < (1 - t_j)w$, but since it would in fact be taxed at the higher rate t_{j+1} , they prefer to stay at \hat{y}_j . A small relaxation of this constraint increases net income by more than the value of the marginal disutility of effort at this point. In what follows we denote this term more compactly by $v_{\hat{y}_j}$.

Note then that $w < \tilde{w}_j \Leftrightarrow y^* = \phi(t_j, w) < \hat{y}_j$, and $\partial \tilde{w}_j / \partial \hat{y}_j > 0, \partial \bar{w}_j / \partial \hat{y}_j > 0$. Also, since $\hat{y}_{j+1} > \hat{y}_j$ and $t_{j+1} > t_j$ we have $\tilde{w}_{j+1} > \bar{w}_j > \tilde{w}_j$, while $\phi(t_j, w) = \hat{y}_j \Leftrightarrow \tilde{w}_j \leq w \leq \bar{w}_j, j = 1, \dots, m - 1$.

Thus, to summarise these results: the consumers can be partitioned into subsets according to their wage type, determined by where they choose to be on the given budget constraint facing all consumers: A consumer is either at a kink point or at a tangency point, or, for consumers of wage types $\tilde{w}_j, \bar{w}_j, j = 1, \dots, m - 1$, at both. We denote the subsets of wage types *not* positioned at kink points by

$$C_1 = [w_0, \tilde{w}_1), C_2 = (\bar{w}_1, \tilde{w}_2), \dots, C_m = (\bar{w}_{m-1}, w_1] \quad (18)$$

and the subsets at kink points by

$$\hat{C}_1 = [\tilde{w}_1, \bar{w}_1], \hat{C}_2 = [\tilde{w}_2, \bar{w}_2], \dots, \hat{C}_{m-1} = [\tilde{w}_{m-1}, \bar{w}_{m-1}] \quad (19)$$

with $C \equiv \{\cup C_j\}_{j=1}^m \cup \{\cup \hat{C}_j\}_{j=1}^{m-1} = [w_0, w_1]$.¹⁰ Given the continuity of $F(w)$, consumers are continuously distributed around the budget constraint, with both maximised utility v and gross income y continuous functions of w . Utility v is strictly increasing in w for all w , and y^* is also strictly increasing in w except over the intervals $[\tilde{w}_j, \bar{w}_j]$, where it is constant at \hat{y}_j .

Consider now the tax paid by a consumer of a given wage type. This can be written as

$$T(t_1, \dots, t_j, \hat{y}_1, \dots, \hat{y}_{j-1}, w) = t_j \phi(t_j, w) - b_j \quad w \in C_j \quad j = 1, \dots, m \quad (20)$$

and

$$\hat{T}(t_1, \dots, t_j, \hat{y}_1, \dots, \hat{y}_j, w) = t_j \hat{y}_j - b_j \quad w \in \hat{C}_j \quad j = 1, \dots, m-1 \quad (21)$$

The derivatives of the tax function are, for $w \in C_j$, $j = 1, \dots, m$:

$$\frac{\partial T}{\partial t_j} = \phi(t_j, w) + t_j \frac{\partial \phi(t_j, w)}{\partial t_j} - \frac{\partial b_j}{\partial t_j} \quad j = 1, \dots, m \quad (22)$$

$$\frac{\partial T}{\partial t_k} = \hat{y}_k - \hat{y}_{k-1}; \quad \frac{\partial T}{\partial \hat{y}_k} = -(t_{k+1} - t_k), \quad k = 1, \dots, j-1 \quad (23)$$

and for $w \in \hat{C}_j$, $j = 1, \dots, m-1$:

$$\frac{\partial \hat{T}}{\partial t_j} = \hat{y}_j - \hat{y}_{j-1}; \quad \frac{\partial \hat{T}}{\partial \hat{y}_j} = t_j; \quad j = 1, \dots, m \quad (24)$$

$$\frac{\partial \hat{T}}{\partial t_k} = \hat{y}_k - \hat{y}_{k-1}; \quad \frac{\partial \hat{T}}{\partial \hat{y}_k} = -(t_{k+1} - t_k), \quad k = 1, \dots, j-1 \quad (25)$$

We now turn to the optimal tax analysis.

3 Optimal Taxation

3.1 The optimal piecewise linear tax system

The planner chooses the parameters of the tax system to maximise a generalised utilitarian social welfare function (SWF) defined as

$$\Omega = \sum_{j=1}^m \int_{C_j} S[v(t_1, \dots, t_j, \hat{y}_1, \dots, \hat{y}_{j-1}, w)] dF + \sum_{j=1}^{m-1} \int_{\hat{C}_j} S[v(t_1, \dots, t_j, \hat{y}_1, \dots, \hat{y}_j, w)] dF \quad (26)$$

¹⁰In all that follows we assume that the tax parameters and wage distribution are such that none of these subsets is empty.

where $S(\cdot)$ is a continuously differentiable, strictly concave¹¹ and increasing function which expresses the planner's preferences over consumer utilities. The government budget constraint is

$$\Upsilon = \sum_{j=1}^m \int_{C_j} T(t_1, \dots, t_j, \hat{y}_1, \dots, \hat{y}_{j-1}, w) dF + \sum_{j=1}^{m-1} \int_{\hat{C}_j} \hat{T}(t_1, \dots, t_j, \hat{y}_1, \dots, \hat{y}_j, w) dF - a - G \geq 0 \quad (27)$$

where $G \geq 0$ is a per capita revenue requirement.

We can, on the assumption that the solution is an interior global optimum, characterise the optimal tax rates and bracket limits by first order conditions¹² given by:

Proposition 1: The optimal values of the tax parameters $a^*, t_1^*, \dots, t_m^*, \hat{y}_1^*, \dots, \hat{y}_{m-1}^*$, satisfy the conditions

$$\int_C \left(\frac{S'(v(w))}{\lambda} - 1 \right) dF = 0 \quad (28)$$

where λ is the shadow price of tax revenue;

$$t_j^* = \frac{\int_{C_j} [(S'/\lambda) - 1][\phi(t_j^*, w) - \hat{y}_{j-1}^*] dF + (\hat{y}_j^* - \hat{y}_{j-1}^*) \int_{C/\Gamma_j} [(S'/\lambda) - 1] dF}{\int_{C_j} \partial \phi(t_j^*, w) / \partial t_j dF} \quad j = 1, \dots, m-1 \quad (29)$$

where $\Gamma_j \equiv C_1 \cup \hat{C}_1 \cup C_2 \cup \hat{C}_2 \cup \dots \cup \hat{C}_{j-1} \cup C_j \quad j = 1, \dots, m$.

Since $C/\Gamma_m = \emptyset$, we have

$$t_m^* = \frac{\int_{C_m} [(S'/\lambda) - 1][\phi(t_m^*, w) - \hat{y}_{m-1}^*] dF}{\int_{C_m} \partial \phi(t_m^*, w) / \partial t_m dF} \quad (30)$$

Finally, the condition characterising each bracket limit is

$$\int_{\hat{C}_j} \left\{ \frac{S'}{\lambda} v_{\hat{y}_j} + t_j^* \right\} dF = -(t_{j+1}^* - t_j^*) \int_{C/(\Gamma_j \cup \hat{C}_j)} \left(\frac{S'}{\lambda} - 1 \right) dF \quad j = 1, \dots, m-1 \quad (31)$$

Proof: By differentiation of the Lagrange function $\Omega + \lambda \Upsilon$ with respect to $a, t_1, \dots, t_m, \hat{y}_1, \dots, \hat{y}_{m-1}$, then using the results from Section 2 and rearranging the resulting first order conditions.

3.2 Discussion

The first condition shows that the optimal payment a equalises the population average of the marginal social utility of income in terms of the numeraire, consumption, with the marginal cost of the transfer, which is 1. This is a familiar

¹¹This therefore excludes the utilitarian case, which can however be arbitrarily closely approximated. As is well known, the strict utilitarian case, with $S' = 1$, presents technical problems when a quasilinear utility function with consumption as numeraire is also assumed.

¹²In deriving these conditions, it must of course be taken into account that the limits of integration \bar{w} and \underline{w} are functions of the tax parameters. Because of the continuity of utility, optimal gross income and tax revenue in w , these effects all cancel and the first order conditions reduce to those shown here.

condition from optimal linear taxation. Since $S(\cdot)$ is strictly concave and $v(\cdot)$ is increasing monotonically in w , this marginal social utility of income S'/λ is falling monotonically in w . Thus an initial subset of wage types will have above-average marginal social utilities.

The denominators of the expressions for the optimal marginal tax rates t_j^* are also familiar from optimal linear tax theory. They are the frequency-weighted sums over the wage types in the respective tax brackets of their compensated derivatives of earnings with respect to the tax rate, determined by the slopes of the individual labour supply functions with respect to the net of tax wage rate. They are a measure of the deadweight loss or labour supply distortion created at the margin by the tax rate. They are negative, and the greater their absolute value for a given tax bracket the lower, other things equal, must be the corresponding tax rate.

The numerators of the marginal tax rate expressions for $j = 1, \dots, m - 1$ represent the main departure from optimal linear tax theory. In place of the simple covariance between the marginal social utility of income and income, which defines the equity effect of the tax in the linear tax model, we have, for all tax brackets except the highest, two terms that represent respectively the equity effects of the tax within the given tax bracket and the sum of its equity effects across all higher tax brackets. The marginal effect of the tax rate t_j on the utility of a wage type in equilibrium in the interval C_j is given by the portion of her income falling within the corresponding bracket, $\phi(t_j^*, w) - \hat{y}_{j-1}^*$. To obtain the first term these are weighted by the deviation of the consumer's marginal social utility of income from the population average and summed across all consumers in that bracket. It could be the case that in the lower tax brackets, for example $j = 1$, this term could be positive, given the distribution of the terms $[(S'/\lambda) - 1]$. In the absence of the second term, this would imply a negative marginal tax rate.

The second term reflects the fact that the tax rate t_j is an intramarginal, nondistortionary tax on incomes in all brackets $j + 1, \dots, m$. The tax rate t_j has a marginal effect on the utilities of all the consumers in higher tax brackets proportional to $(\hat{y}_j^* - \hat{y}_{j-1}^*)$, and the equity effects of this are found by weighting this term by the frequency-weighted sum of the deviations of these consumers' marginal social utilities of income from the population average. Given the condition (28), this sum must be negative, and so this term overall always contributes positively to the value of the tax rate.

It can be shown¹³ that t_1^* is strictly positive, and given that optimal marginal tax rates are increasing and that income increases with wage type this will also apply to all higher tax rates. The intuition is straightforward. If t_1^* were zero, earnings choices of consumers in C_1 are undistorted, and so a marginal increase in t_1^* has zero first order effects on welfare in this subset. However, it has a strictly positive first order effect on tax revenue resulting from the positive lump sum nondistortionary tax on all higher brackets \hat{C}_1, C_2, \dots , allowing

¹³See Apps, Long and Rees (2011) for the proof of this in the two-bracket case, which readily extends to m brackets.

a transfer from consumers with lower to those with higher marginal social utilities of income. Thus with a given revenue requirement overall welfare can be increased.

This second term is missing from the expression for t_m^* since there are no higher tax brackets. Given the restriction to piecewise linear tax systems, we do not have the "no distortion at the top" or zero marginal tax rate result of the Mirrlees model, since, because of condition (31), this top interval of wage types must be of nonzero length, and so both numerator and denominator terms must be negative, giving a positive tax rate overall.

The final condition characterises the optimal bracket limits. The left hand side gives the marginal social benefit of a slight relaxation of the j 'th bracket limit. As shown in the previous section, this first of all gives a positive benefit $v_{\hat{y}_j}$ to almost all the wage types in \hat{C}_j , since the marginal increase in net income exceeds the marginal disutility of the increased effort. This is weighted by the marginal social utility of income to these wage types, S'/λ . The increase in gross income also increases tax revenue at the rate t_j^* . The marginal cost of the relaxation of the bracket limit, the right hand side of (31), reflects a worsening in the equity of the tax system. All consumers of wage type higher than \bar{w}_j receive a marginal benefit $(t_{j+1}^* - t_j^*)d\hat{y}_j > 0$, and this is weighted by the sum of deviations of their marginal social utilities of income from the population average, which must be negative, because of condition (28). So the optimal bracket limit equalises these marginal costs and benefits. The assumption that a piecewise linear tax system with increasing marginal tax rates is globally optimal implies that this condition holds for at least one bracket limit in the interior of the set of optimal gross incomes generated by the tax system.

The question of the optimal finite number of tax brackets is not addressed in this paper. This would require a specification of the costs associated with the number of brackets and associated complexity of the tax system and then the comparison of the increases in these as we go from m to $m + 1$ brackets, $m = 1, 2, \dots$, with the increase in maximised social welfare social resulting from this. The general form of the solution is quite obvious, and the real challenge would be to obtain the data that would allow the problem to be solved in practice.¹⁴ In the next section we contribute to this by examining in a relatively simple parameterised model the latter part of this calculation.

4 Numerical Results

We illustrate the general characteristics of the m -bracket model set by showing how the structure of optimal tax parameters for the 2-, 3- and 4-bracket cases depend on the shape of the wage distribution, labour supply elasticities and the degree of inequality aversion specified in the social welfare function. The analysis proceeds in two steps. We first solve for the optimal tax parameters for "reference" wage distributions constructed from survey data for, respectively, the US, UK and Australia. We then show how the optimal tax parameters

¹⁴The computational problems should also not be underestimated.

change when inequality in each reference distribution increases as a result of steeply rising wages across the top percentiles.

We assume throughout that tax revenue is equal to the amount required to pay to each household the optimal lump sum a^* , that is, we take for purposes of illustration the case of "pure redistribution" (in the government budget constraint (27) $G = 0$). As a result, all the cases we consider are in that sense revenue-equivalent.

The next subsection discusses data sources and the construction of the reference wage distributions. The subsection following presents the results for the structure of the optimal tax parameters first, as we increase the number of tax brackets across each reference wage distribution and secondly, as wage rates rise in the top percentiles of each distribution.

4.1 Wage distributions

The reference wage distributions are based on data for the earnings and hours of work of the primary earner of couples selected from the US Panel Study of Income Dynamics (PSID) 2009, the British Household Panel Survey (BHPS) 2009, and the Australian Bureau of Statistics Survey of Income and Housing (ABS SIH) 2009-10. The primary earner is defined as the partner with the higher labour income. A sample of couples from each survey is selected on the criteria that both partners are aged from 25 to 59 years and the primary earner works at least 30 hours per week. We drop the bottom 5 percentiles in order to exclude very low wage earners who are likely to be recipients of categorical welfare payments.¹⁵ The number of observations is, respectively, 2553, 2261 and 4053 in the US, UK and AU samples. The wage in each percentile is calculated as average gross hourly earnings with hours smoothed across the distribution.¹⁶ Figure 1 plots the profile for each country.¹⁷

Figure 1 about here

The most striking characteristic of the wage distributions is that they rise relatively slowly and are virtually linear over the initial seven to eight deciles, then turn sharply upward, reflecting the general inequality in income distribution in each country. Consistent with studies that track inequality over recent decades,¹⁸ of the three countries the US has far higher wage rates and earnings in the top percentiles.

We stress that our results for these reference distributions, though suggestive of the characteristics of optimal tax systems in general, cannot be strictly interpreted as empirical estimates of optimal tax systems for the three economies because we cannot realistically incorporate the actual structure of marginal tax

¹⁵An assumption of the optimal tax simulations we present is that the individual's wage type cannot be observed. However very low wage individuals may be tagged according to observable characteristics that attract additional payments.

¹⁶We use the Lowess method for smoothing the profile of the percentile distribution of hours as a function of earnings.

¹⁷For the purpose of comparison we use historically average exchange rates adjusted for differences in prices, 1.60 USD/GBP and 0.75 USD/AUD for the UK and Australia, respectively.

¹⁸See Atkinson et al. (2011) and Piketty and Saez (2003).

rates and lump sums making up the tax-transfer system of each country. The three countries have complex tax-transfer systems. While each applies a progressive formal rate scale to income,¹⁹ the *effective* rates on primary earnings may be far less progressive than those of the formal rate scale. All three countries provide income-tested credits and family payments that raise effective marginal rates across the lower and middle percentiles of the income distribution, and they offer exemptions and opportunities for avoidance towards the upper percentiles that lower the effective top rates of the scale. While formal marginal rates may vary dramatically across narrow income bands, when "smoothed" the overall "effective" scale may be relatively flat and close to the average rate profile for the "in-work" samples we have selected.²⁰

We therefore begin by selecting hypothetical smoothed marginal rates. We present results for a constant marginal tax rate of 0.2 and, as a robustness check, for a marginal rate that rises from 0.2 in the first percentile to 0.3 in the top percentile across each distribution. The results for the latter are reported in Table A of the Appendix.

With the net wage given by $\hat{w} = (1 - \tau)w$, where τ denotes the smoothed marginal tax rate, we derive the form of the utility function generating optimal labour supplies l^* that broadly match the data.²¹ Figure 2 plots the labour supply elasticity profiles.²² Given a quasilinear utility function, these are compensated elasticities and therefore contain the compensated derivatives entering the denominators of the expressions for the optimal marginal rates in (29) and (30).

Figure 2 about here

From Figure 2 we see that the elasticity profiles at first decline rapidly and then level off across the percentile wage distribution. This general pattern reflects the tendency for hours of work to rise linearly across wage distributions that are initially relatively flat and then rise steeply in the upper percentiles. Thus elasticities in the upper half of the distributions are lower for the US than for UK and Australia because US wage rates are much higher in the top percentiles and there is no matching increase in hours of work. In all three countries hours of work profiles across the wage distribution are broadly similar.

To show how optimal tax rates change with rising inequality, we construct a second set of distributions by introducing wage growth in the top decile of each reference distribution. We allow a growth in wage rates beginning at 3% in the

¹⁹ Additional complexity arises with variation in the tax base across countries. The US Federal Income Tax is based on joint income while the Australian and UK formal income tax systems are based on individual incomes.

²⁰ See Apps and Rees (2009, Ch. 6) for a detailed analysis of the effective rates scales of all three countries.

²¹ We fit a monotonic function to the pairs (l^*, \hat{w}) to derive the utility function in (1), which has the form $u(\hat{w}) = u(0) + b_j + \int_0^{\hat{w}} l^*(w)dw$, where j is the bracket for the net wage \hat{w} and b_j is given in (6). For all simulations we set $u(0) = 0$.

²² Elasticities are smoothed using the lowess method for the mid-point elasticity as a function of the wage.

91st percentile and rising uniformly to 30% in the top percentile.²³

4.2 Simulation results

For each wage distribution we solve for the optimal parameters of the tax system, $a^*, t_1^*, \dots, t_m^*, \hat{y}_1^*, \dots, \hat{y}_{m-1}^*$, which maximise a SWF of the form $[\sum_{i=1}^n v_i^{1-\rho}]^{1/(1-\rho)}$, where ρ is a measure of inequality aversion, $n = 100$ is the number of wage types, and v_i is the indirect utility function in (11). We find the global maximum of the SWF by applying a general grid search algorithm across marginal tax rates lying in the interval $[0, 1]$ with an increment of 0.01 and an integer bracket limit rising by one dollar increments in weekly earnings.

Panels US, UK and AU in Table 1 present the results for the three reference distributions with $\tau = 0.2$. Each panel reports the optimal tax parameters for a linear system and the 2-, 3-, and 4-bracket piecewise linear systems for $\rho = 0.1, 0.2$ and 0.3 . Thus these results show the effects on the optimal parameters a^*, t_j^*, \hat{y}_j^* of increasing the number of tax brackets for the given wage distributions, as well as the robustness of the results to variations in the degree of inequality aversion in the SWF.

Table 1 about here

The changes in optimal tax parameters as we move from the linear to the 2-, 3- and 4-bracket piecewise systems exhibit a number of consistent features for each distribution and across the values of ρ . The most striking are:

- the marginal tax rates in the lower brackets tend to fall as the number of brackets increases;
- the optimal lump sum payment typically declines;²⁴ and
- the effect of increasing ρ is to increase the optimal degree of progressivity by raising the whole structure of marginal rates and funding a larger lump sum in each system.

The rising value of the SWF as the number of brackets increases for all values of ρ implies that there are gains in moving from a linear to a four bracket piecewise linear tax system.²⁵ Essentially, increasing the number of brackets

²³According to the recent Bivens and Mishel (2013) survey of evidence on changes in the distribution of income in the US during the period 1979-2005/07 the annual rate of growth of the bottom nine deciles was close to zero while that of the top decile was around 1.5 per cent and that of the top percentile, around 3 per cent per annum, implying an overall rate of growth of more than 30 per cent over a ten year period.

²⁴In some cases when we move from the 3- to the 4-bracket system and t_1 remains the same, the lump sum remains the same, as for example in the case of the UK for $\rho = 0.2$ and 0.3 where t_1 remains at 11 and 15 per cent, respectively. The result reflects the precision attainable in the grid search with increments of 0.01 in the marginal tax rate and one dollar in weekly earnings for each bracket limit. The models were run on a supercomputer.

²⁵The fact that the absolute differences in the SWF values are relatively small is a result of the simplifying assumption of quasilinear utilities. Introducing more concavity into the utility function would increase the measure of utility differences, but would introduce income effects and thus greatly complicate the analysis.

allows the marginal rate scale and therefore the intramarginal nondistortionary taxes on the higher brackets to be more finely-tuned to the shape of the wage distribution and variation in labour supply responses, to achieve an optimal tax system that is more progressive overall. This can be seen in condition (29), where reducing the bracket widths ($\hat{y}_j^* - \hat{y}_{j-1}^*$) in the numerator will, other things equal, reduce the marginal tax rates. The redistributive loss from a lower lump sum payment as the number of brackets rises is more than offset by lower tax rates on the lower wage brackets.²⁶

The results for the US distribution, with its far higher top wage rate, stand apart in a number of respects. The bracket points for the top rate of the 3- and 4-bracket systems, \hat{y}_2 and \hat{y}_3 , are consistently at the 99th percentile. The top rates range from 63 to 73 per cent. For the UK and AU distributions, the bracket point for the top rate does not exceed the 97th percentile and the highest tax rate is 56 per cent.

It is interesting to consider the results for a country and a ρ -value, for example that of the US with $\rho = 0.2$, as we move from a 2- to a 3- to a 4-bracket tax system. The tax rates in the 2-bracket system are respectively 32 and 62 per cent with the latter rate coming in at the 96th percentile. This is virtually a flat tax with the top 4 per cent of income earners paying a rate that is almost double the "standard rate". Referring to Figure 1, this is clearly due to the sudden very sharp increase in wage rates at around that percentile. This relatively high standard rate reflects its use as a non-distortionary tax on the top incomes. Moving to a 3-bracket system reduces the rate on the lowest bracket somewhat, to 28 per cent, but greatly increases the degree of progressivity within the upper part of the income distribution. The tax rate rises sharply at the 81st percentile, but is still around 25 per cent lower than the previous top rate, and the significantly higher top rate, rising from 62 per cent to 70 per cent, kicks in later, at the 99th percentile. Thus we are seeing a more differentiated, progressive structure between high and very high incomes. Finally, moving to a 4-bracket structure leads to further significant falls in the lowest two tax rates, with a fairly sharp increase in tax rate at the 53rd percentile, a similarly sharp increase at the 91st percentile, though the tax rate in this bracket is still well below the top rate, and then an even sharper increase to the previous top rate at the 99th percentile.

Thus the overall pattern as the structure of tax brackets becomes finer is a falling tax rate on the lower half of the distribution, accompanied by a lower lump sum, together with a more differentiated, highly progressive structure of tax rates on the upper half, with quite a sharp differentiation between the top 10% and the top 1%. A similar pattern is shown for the other two countries.

Table 2 presents the optimal tax parameters for the second set of wage distributions in which wage rates rise uniformly in increments of 3 per cent from the 90th percentile, thus increasing the degree of inequality in the wage distribution. All lump sum payments are larger, reflecting the optimally higher

²⁶We can also expect this to be desirable for reasons outside the framework of the present optimal tax analysis. Setting high disincentives for low wage individuals to work, while maintaining their living standard by high lump sum transfers, would be regarded as perpetuating the cycle of "welfare dependency" in a socially undesirable way.

degree of progressivity with rising inequality. The changes in tax rates and bracket limits indicate that the larger lump sums tend to be funded by higher top tax rates, lower bracket limits, or some combination of both, with the specific result in each case being highly sensitive to the point at which wage rates begins to rise. For example, the optimal top tax rates of the 3- and 4-bracket systems, t_3 and t_4 , are higher than in Table 1 for all three degrees of inequality aversion for the UK and US distributions, with the higher US rates continuing to apply at the 99th percentile. In the UK distribution the bracket limits for the top rate, t_4 , fall to the 90th percentile, the point at which wage rates begin to rise, for all values of ρ . For t_3 the rate falls for $\rho = 0.1$ and rises for $\rho = 0.2$ and 0.3 , with the bracket point consistently at the 90th percentile. In the AU results the bracket limit of the top tax rate falls to the 90th percentile for all values of ρ while the optimal taxes, t_3 and t_4 , tend to stay the same.

These results suggest that the optimal response to the significant increase in income inequality associated with growth in the share of the top 10 per cent, and even more markedly in the share of the top 1 per cent, is a shift towards a more progressive multi-bracket income tax system. In contrast to this direction of reform, recent decades have seen a number of OECD countries, such as the US, UK and Australia, move towards less progressive income tax systems. Australia, for example, has significantly reduced taxes on top incomes by combining lower top tax rates with upward shifts in the top bracket limits at which the rates apply. At the same time effective marginal tax rates on low to average incomes have risen with the introduction of income-tested tax offsets, credits and family payments.

Table A of the Appendix reports the results of simulations with τ increasing from 0.20 to 0.30 across each wage distribution. Very similar patterns of optimal parameters to those in Tables 1 are obtained. As we increase the number of brackets the optimal tax rates on the lower brackets tend to fall and the optimal lump sums typically fall.²⁷ The value of the SWF rises as the number of brackets increases for all values of ρ and all distributions. The top tax rate tends to be around the same despite the higher elasticities associated with the rising value of τ .²⁸ The consistency of the pattern of the rates and bracket points with those in Table 1 suggest that the qualitative results are quite robust to changing the initial smoothed marginal tax rate on which we base our calibrations of the labour supply and utility functions.

5 Conclusions

In this paper we have used the approach of optimal piecewise linear income taxation to address the issue of the taxation of top incomes. This focuses attention

²⁷ Again, where t_1 remains the same remain as the number of brackets increases, the lump sum can stay the same or actually rise slightly, as for example in the case of the UK for $\rho = 0.1$. As noted previously, these slight deviations from the overall pattern of the results reflect the precision attainable in the grid search.

²⁸ Since the net wage falls as τ rises and hours are given by the data, the labour supply elasticity rises with τ .

not just on the "top tax rate" alone but rather on the entire rate structure and set of bracket limits of the overall tax system. Our numerical results suggest that the appropriate response to the large increase in wage and income inequality over the past few decades would have been a shift towards a more progressive multi-bracket income tax system with lower marginal rates on the lower half of the distribution and an increasing degree of differentiation and marginal rate progressivity in the upper half of the distribution. Further inequality growth strengthens the case for these features of an optimal tax system. Certainly, given the characteristics of the empirical wage and income distributions, the actual changes in tax systems that have taken place, with sharp reductions in the tax burden on top incomes and considerable shifting of this burden on to the middle deciles of income, cannot be rationalised in this model.

References

- [1] Apps, P., N. V. Long, and R. Rees, 2011. Optimal piecewise linear income taxation. IZA DP#6007. Forthcoming in *Journal of Public Economic Theory*.
- [2] Apps, P. and R. Rees, 2012. Optimal taxation, child care and models of the household. IZA, DP#6823.
- [3] Apps, P. and R. Rees, 2009. *Public economics and the household*. Cambridge, UK: Cambridge University Press.
- [4] Bivens, J. and M. Michel, 2013. The pay of corporate executives and financial professionals as evidence of rents in top 1 percent incomes. *Journal of Economic Perspectives*, 27(3), 57-78.
- [5] Atkinson A. B., T. Piketty and E. Saez, 2011. Top incomes in the long run of history. *Journal of Economic Literature*, 49(1), 3-71.
- [6] Boadway, R., 2012. *From optimal tax theory to tax policy*, Cambridge, Mass: MIT Press
- [7] Dahlby, B., 1998. Progressive taxation and the marginal social cost of public funds. *Journal of Public Economics*, 67, 105-122.
- [8] Dahlby, B., 2008. *The marginal cost of public funds*. Cambridge, Mass: MIT Press.
- [9] Mirrlees, J., S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba, (2011). *Tax by design*. Oxford, UK: Oxford University Press.
- [10] Peter, K. S., S. Buttrick and D. Duncan, 2010. Global reform of personal income taxation, 1981-2005: Evidence from 189 countries. *National Tax Journal*, 63(3), 447-478.

- [11] Piketty T., and E. Saez, 2003. Income Inequality in the United States, 1913-1998. *Quarterly Journal of Economics*, 118 (1), 1-39.
- [12] Piketty T., E. Saez and S. Stantcheva, 2011. Optimal taxation of top labor incomes: A tale of three elasticities. NBER WP#17616. Forthcoming in *American Economic Journal: Economic Policy*.
- [13] Sheshinski, E., 1989. Note on the shape of the optimum income tax schedule. *Journal of Public Economics* 40, 201-215.

Figure 1 Reference wage distributions

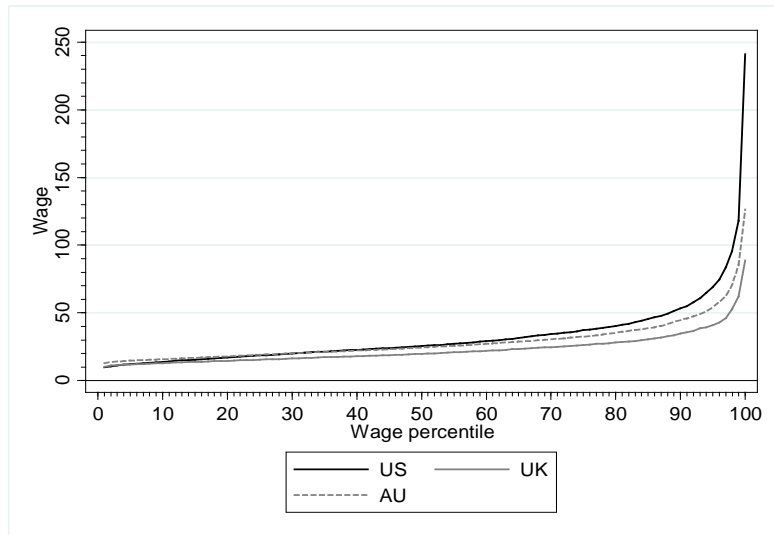


Figure 2 Labour supply elasticities

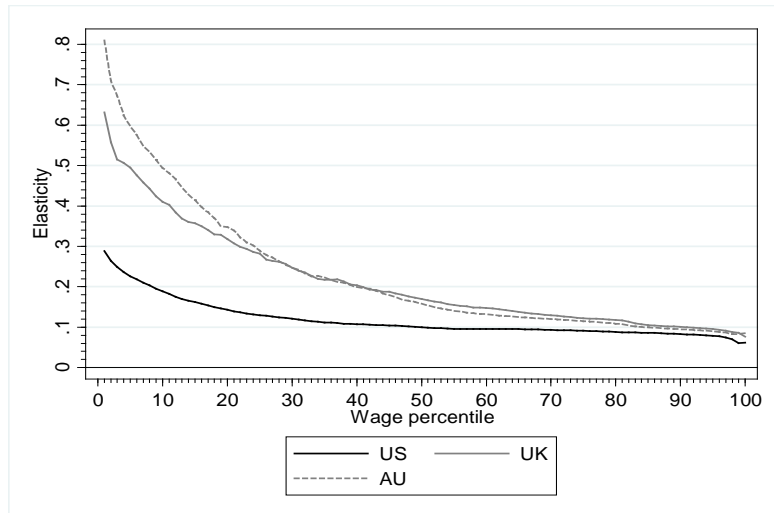


Table 1 Optimal tax parameters: reference distribution $\tau = 0.2$

Dist*	ρ	t_1^{**}	t_2^{**}	t_3^{**}	t_4^{**}	\hat{y}_1^{***}	\hat{y}_2^{***}	\hat{y}_3^{***}	a	SWF/ 10^3
US	0.1	28	-	-	-	-	-	-	402	200.29
		23	51	-	-	96	-	-	363	200.63
		20	36	63	-	87	99	-	339	200.74
		16	26	39	63	53	96	99	314	200.78
	0.2	38	-	-	-	-	-	-	538	374.44
		32	62	-	-	96	-	-	488	375.76
		28	46	70	-	81	99	-	460	376.16
		22	36	49	70	53	91	99	416	376.32
	0.3	43	-	-	-	-	-	-	603	843.17
		35	63	-	-	91	-	-	541	847.42
		28	48	73	-	56	99	-	500	848.63
		28	42	55	73	53	91	99	494	849.12
UK	0.1	20	-	-	-	-	-	-	202	121.49
		11	23	-	-	71	-	-	131	121.63
		8	19	34	-	45	97	-	117	121.66
		4	11	19	34	19	53	97	94	121.67
	0.2	22	-	-	-	-	-	-	222	227.66
		15	31	-	-	57	-	-	186	228.06
		11	23	37	-	38	83	-	163	228.19
		11	23	33	45	38	83	97	163	228.24
	0.3	27	-	-	-	-	-	-	269	512.83
		19	41	-	-	73	-	-	222	514.38
		15	27	43	-	38	83	-	204	514.85
		15	27	40	51	38	83	97	204	515.00
AU	0.1	20	-	-	-	-	-	-	259	152.56
		13	30	-	-	86	-	-	193	152.74
		8	21	40	-	43	97	-	165	152.81
		5	13	21	40	21	55	97	144	152.83
	0.2	24	-	-	-	-	-	-	309	285.24
		13	35	-	-	57	-	-	235	286.05
		12	30	52	-	46	97	-	232	286.27
		8	21	34	52	24	69	97	209	286.35
	0.3	28	-	-	-	-	-	-	357	641.54
		20	44	-	-	71	-	-	309	644.65
		18	38	56	-	57	97	-	299	645.34
		13	28	40	56	36	71	97	264	645.69

*Reference wage distribution

** MTR percentage

*** Bracket limit percentile

Table 2 Optimal tax parameters: $\tau = 0.2$ with rising top wage rates

Dist*	ρ	t_1^{**}	t_2^{**}	t_3^{**}	t_4^{**}	\hat{y}_1^{***}	\hat{y}_2^{***}	\hat{y}_3^{***}	a	SWF/10 ³
US	0.1	30	-	-	-	-	-	-	460	214.48
		20	49	-	-	90	-	-	387	214.99
		20	45	67	-	90	99	-	392	215.08
		16	26	45	67	53	90	99	372	215.13
	0.2	40	-	-	-	-	-	-	602	399.95
		28	62	-	-	90	-	-	516	402.35
		28	55	74	-	87	99	-	520	402.76
		22	36	56	74	53	90	99	480	402.96
	0.3	46	-	-	-	-	-	-	684	899.15
		34	68	-	-	90	-	-	599	907.24
		34	61	77	-	87	99	-	601	908.23
		26	42	62	76	53	90	99	544	908.94
UK	0.1	20	-	-	-	-	-	-	212	128.43
		11	34	-	-	83	-	-	152	128.66
		8	19	36	-	45	90	-	140	128.70
		6	17	23	36	37	83	90	128	128.71
	0.2	25	-	-	-	-	-	-	262	239.92
		17	45	-	-	83	-	-	221	240.94
		11	26	49	-	46	90	-	188	241.11
		11	23	35	48	38	83	90	190	241.16
	0.3	29	-	-	-	-	-	-	301	539.37
		22	53	-	-	88	-	-	267	543.02
		15	30	55	-	46	90	-	231	543.67
		15	27	40	54	38	83	90	233	543.79
AU	0.1	20	-	-	-	-	-	-	273	162.07
		13	39	-	-	87	-	-	227	162.48
		8	21	41	-	45	90	-	199	162.54
		5	13	21	41	21	55	90	180	162.56
	0.2	27	-	-	-	-	-	-	363	302.07
		18	51	-	-	87	-	-	305	303.93
		11	27	51	-	43	87	-	265	304.19
		6	19	30	52	21	56	90	235	304.25
	0.3	31	-	-	-	-	-	-	413	677.90
		22	56	-	-	86	-	-	360	684.43
		13	31	56	-	39	87	-	305	685.44
		9	23	35	56	22	57	87	285	685.65

*Reference wage distribution

** MTR percentage

*** Bracket limit percentile

Appendix

Table A Optimal tax parameters: reference distribution $\tau = 0.2$ to 0.3

Dist*	ρ	t_1^{**}	t_2^{**}	t_3^{**}	t_4^{**}	\hat{y}_1^{***}	\hat{y}_2^{***}	\hat{y}_3^{***}	a	SWF/ 10^3
US	0.1	30	-	-	-	-	-	-	432	202.38
		23	51	-	-	96	-	-	365	202.69
		21	39	64	-	91	99	-	351	202.79
		16	29	44	64	53	96	99	325	202.82
	0.2	37	-	-	-	-	-	-	527	378.37
		32	63	-	-	96	-	-	491	379.64
		27	47	71	-	81	99	-	454	380.04
		23	37	49	71	53	91	99	431	380.21
	0.3	43	-	-	-	-	-	-	606	852.03
		35	64	-	-	91	-	-	545	856.19
		27	48	75	-	54	99	-	498	857.46
		27	43	56	74	53	91	99	492	857.98
UK	0.1	30	-	-	-	-	-	-	299	123.46
		9	21	-	-	53	-	-	122	124.10
		6	17	32	-	37	96	-	103	124.13
		6	17	23	34	37	83	97	107	124.14
	0.2	30	-	-	-	-	-	-	299	231.93
		14	31	-	-	53	-	-	183	232.71
		11	26	44	-	38	96	-	166	232.84
		11	24	33	46	38	83	97	166	232.89
	0.3	30	-	-	-	-	-	-	299	523.21
		17	37	-	-	57	-	-	219	524.93
		14	29	44	-	38	83	-	203	525.37
		14	29	41	53	38	83	97	203	525.53
AU	0.1	30	-	-	-	-	-	-	383	154.80
		15	36	-	-	94	-	-	212	155.57
		7	19	36	-	35	94	-	158	155.64
		4	16	24	40	21	71	97	145	155.65
	0.2	30	-	-	-	-	-	-	383	290.22
		14	36	-	-	57	-	-	249	291.35
		11	30	50	-	43	97	-	227	291.58
		9	23	34	50	25	69	97	222	291.65
	0.3	30	-	-	-	-	-	-	383	653.44
		18	43	-	-	61	-	-	302	656.62
		17	37	58	-	50	97	-	297	657.33
		11	27	41	58	25	71	97	261	657.63

*Reference wage distribution

** MTR percentage

*** Bracket limit percentile