THE UNIVERSITY OF MELBOURNE

Department of Economics

Working Paper Series

Income Tax Revenue: Some Simple Analytics

John Creedy

July 2010

Research Paper Number 1101

ISSN: 0819-2642
ISBN: 978-0-7340-4454-9

Department of Economics
The University of Melbourne
Parkville VIC 3010
www.economics.unimelb.edu.au

Income Tax Revenue: Some Simple Analytics*

John Creedy
The University of Melbourne

Abstract

This paper takes familiar analytical expressions for total income tax revenue, expressed in terms of summary measures of the distribution of taxable income, and relates them to distances and slopes in the popular Lorenz curve diagram.

*In preparing this paper I have benefited from discussions with Jose Sanz.

1 Introduction

In obtaining simple analytical expressions for the total revenue obtained from an income tax structure having a set of rates and income thresholds, it is known that the form of the income distribution plays a crucial role. Thus the proportions of people within tax brackets, and the corresponding proportions of total income obtained by those people, provide sufficient information (along with the tax rates, and the value of average income) to calculate total revenue per person. The same proportions of people and income can clearly be identified with points on the Lorenz curve of taxable income, which formally traces the relationship between the (incomplete) first moment distribution function (average income of those below a given level divided by arithmetic mean over the whole population) and the distribution function of income (proportion of people below the given income level). While the analytics are well established, the present paper explores some of the diagrammatic properties relating tax revenue to the Lorenz curve.

First Section 2 considers the simplest case of a single tax rate applying to income measured above a tax-free threshold. Tax revenue is related to vertical distances in the Lorenz curve diagram, and the effects of a mean-preserving spread in the income distribution is examined diagrammatically. The approach is extended to the multi-step tax function in Section 3.

2 A Tax-Free Threshold

Consider a simple income tax structure having a single rate of t applying to taxable income, x, measured above a threshold of a. Thus for $x > a$:

$$T(x) = t(x - a) \tag{1}$$

For a population of size N where incomes are continuously distributed with distribution function, $F(x)$, for $0 < x < \infty$, total revenue is equal to:

$$R = Nt \int_a^\infty (x - a)\, dF(x) \tag{2}$$

and:

$$\begin{aligned} R &= Nt\left[\int_a^\infty xdF(x) - a\int_a^\infty dF(x)\right] \\ &= Nt\bar{x}\left[\{1-F_1(a)\} - \frac{a}{\bar{x}}\{1-F(a)\}\right] \end{aligned} \tag{3}$$

where $F_1(x)$ denotes the first (incomplete) moment distribution function, equal to the proportion of total income obtained by those with incomes less than x. Therefore $F_1(x) = \int_0^x udF(u) / \int_0^\infty udF(u)$ or $F_1(x) = \int_0^x udF(u)/\bar{x}$.

It is useful to define the function, $G(a)$ as the term in square brackets in (3), so that:[1]

$$G(a) = \{1-F_1(a)\} - \frac{a}{\bar{x}}\{1-F(a)\} \tag{4}$$

Revenue can thus be written as:

$$R = Nt\bar{x}G(a) \tag{5}$$

Hence changes in revenue arising from changes in the tax parameters are given by:

$$dR = \bar{x}G(a)\,dt + t\bar{x}\left(\frac{\partial G(a)}{\partial a}\right)da \tag{6}$$

where:

$$\frac{\partial G(a)}{\partial a} = -\frac{\partial F_1(a)}{\partial a} - \frac{1-F(a)}{\bar{x}} + \frac{a}{\bar{x}}\frac{\partial F(a)}{\partial a} \tag{7}$$

and since $\partial F(a)/\partial a = f(a)$ and $\partial F_1(a)/\partial a = af(a)/\bar{x}$, the first and last terms cancel and:

$$\frac{\partial G(a)}{\partial a} = -\frac{1-F(a)}{\bar{x}} \tag{8}$$

These results are not new, but the remainder of this section relates the various components of total tax revenue to slopes and distances in the Lorenz curve diagram.

2.1 Diagrammatic Illustration

An alternative expression for total revenue is obtained from (3) by writing $N_T = N\{1-F(a)\}$ as the number of individuals who pay positive tax, that

[1] This function, with its further modifications, is discussed further in Creedy (1996).

is those who have $x > a$, so that:

$$R = t\bar{x}N_T \left\{ \frac{1 - F_1(a)}{1 - F(a)} - \frac{a}{\bar{x}} \right\} \tag{9}$$

The ratio $\{1 - F_1(a)\} / \{1 - F(a)\}$ has a simple interpretation in terms of the Lorenz curve of taxable income. Figure 1 shows the Lorenz curve in the top section, and the distribution function in the lower section. The ratio $\{1 - F_1(a)\} / \{1 - F(a)\}$ is the slope of the line AB, which must exceed 45 degrees. The point C corresponds to the arithmetic mean income, where the Lorenz curve has a slope of 45 degrees. Hence the vertical distance from point C to the Lorenz curve is the maximum distance from the curve to the line of equality, and is referred to as the Schutz index, S, given by:[2]

$$S = F(\bar{x}) - F_1(\bar{x}) \tag{10}$$

The effect of a mean-preserving spread in the distribution of taxable income is illustrated in Figure 2. The change in the distribution function, as shown in the lower section of the figure, involves a fixed arithmetic mean income. The mean is associated with the new Lorenz curve (at the point where that curve has a slope of 45 degrees), having a lower value of $F(\bar{x})$ than with the initial Lorenz curve. With a fixed value of the threshold, a, the value of $F(a)$ must be higher after the mean-preserving spread. Hence the slope of the line drawn from the point on the Lorenz curve, $(F_1(a), F(a))$, to the top right hand corner must be steeper than with the original distribution.

At the same time, the number of taxpayers, $N_T = N\{1 - F(a)\}$, must fall. Hence Figure 2 needs augmenting. Suppose new values are denoted by a prime. Using the function, G, it can be seen that revenue increases if:

$$\frac{F_1(a) - F_1'(a)}{\{1 - F'(a)\} - \{1 - F(a)\}} > \frac{a}{\bar{x}} \tag{11}$$

Consider Figure 3, which shows a further quadrant in the bottom left hand corner where, using a 45 degree line, $\bar{x}$ is translated to the new horizontal axis. Hence the right hand side of (11) is represented by the slope of the line

[2] On this measure and its interpretation, see for example, Lambert (1993, pp. 35-45).

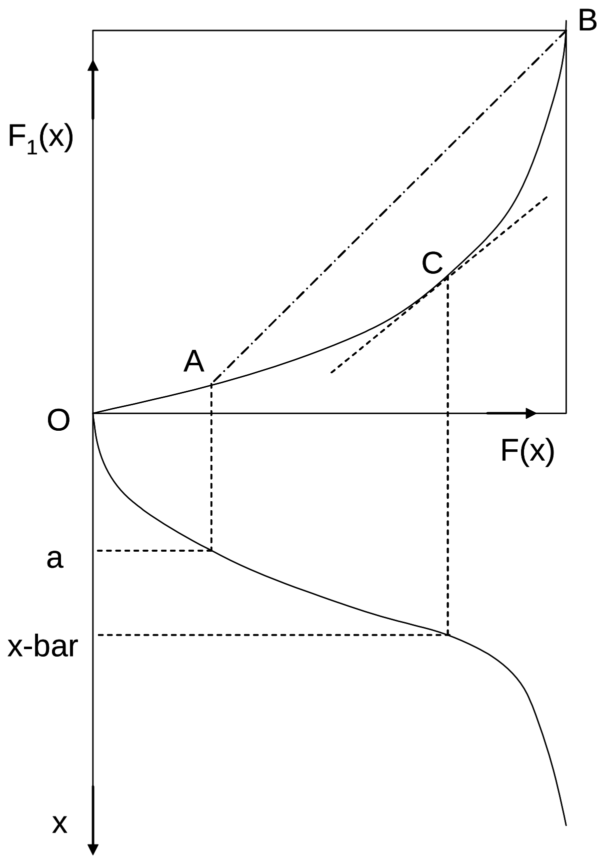

Figure 1: Tax Revenue and the Lorenz Curve

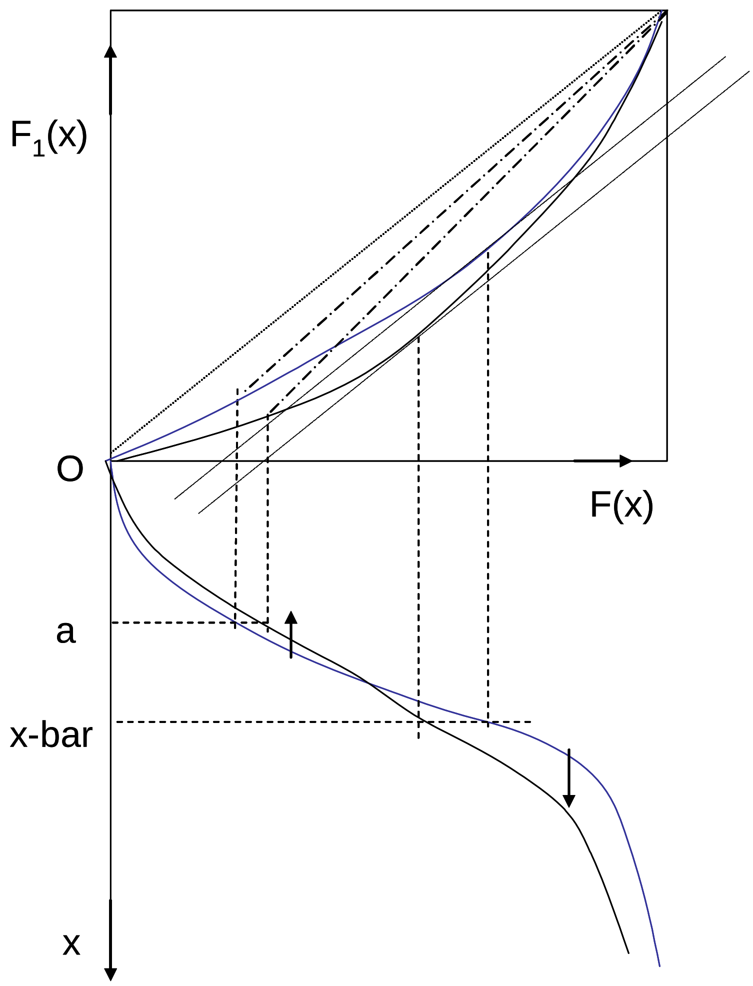

Figure 2: An increase in Inequality: Fixed Mean Income

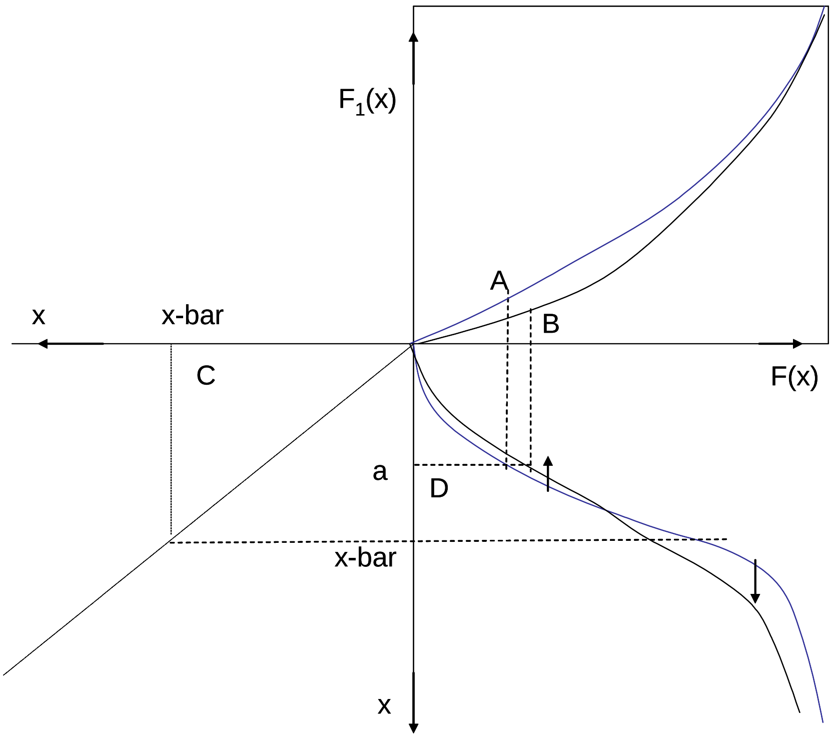

Figure 3: Tax Revenue and Mean-preserving Spread

CD. The left hand side is represented by the slope AB. Comparison of total revenue thus requires a comparison of the two slopes.

Hence there is not an unambiguous result regarding total tax revenue. Hence for two distributions with the same arithmetic mean and population size, if distribution A 'Lorenz dominates' distribution B – that is, if the Lorenz curve of A lies inside that of B – the income tax revenue from a structure having a single rate above a tax-free threshold is not necessarily lower in A than in B.

From the same kind of diagram as Figure 2, it can also be seen that if the Lorenz curve of B were to display less inequality than that of A in the income ranges up to and including the threshold, a, but more inequality elsewhere (so that the Lorenz curve of B intersects that of A from above, when moving from left to right), the slope of the line from $(F_1(a), F(a))$ to the top right hand corner at $(1,1)$ is flatter in B. But, in addition, the number of taxpayers, N_T, must increase. Hence intersecting Lorenz curves, with fixed $\bar{x}$, also means that tax revenue comparisons are ambiguous.

A further rearrangment of the function, $G(a)$ gives:

$$G(a) = \{F(a) - F_1(a)\} - \left(1 - \frac{a}{\bar{x}}\right)\{1 - F(a)\} \tag{12}$$

The terms in this expression also have simple interpretations in terms of the Lorenz curve, as shown in Figure 4. This illustrates $\{F(a) - F_1(a)\}$ as the vertical distance from the Lorenz curve (corresponding to a) to the line of equality. The term $\{1 - F(a)\}$ is the horizontal distance to the right of the appropriate point on the Lorenz curve, and the term $1 - \frac{a}{\bar{x}}$ is the slope of the line from C to a in the bottom left hand quadrant of the diagram. Hence, it is possible to express the product $\left(1 - \frac{a}{\bar{x}}\right)\{1 - F(a)\}$ as a vertical distance, as in Figure 5 by drawing a line from the point $F(a)$ on the horizontal axis of the Lorenz curve which has the same gradient as the line from C to a. This has the effect of projecting the appropriate vertical distance on the right hand side of the box containing the Lorenz curve. Hence $G(a)$ is represented as the sum of two vertical distances.

Furthermore, noting that the second term in (12) can be written as $\{1 - F(a)\} - \frac{a}{\bar{x}}\{1 - F(a)\}$, and that the Gini inequality measure, $Gini$,

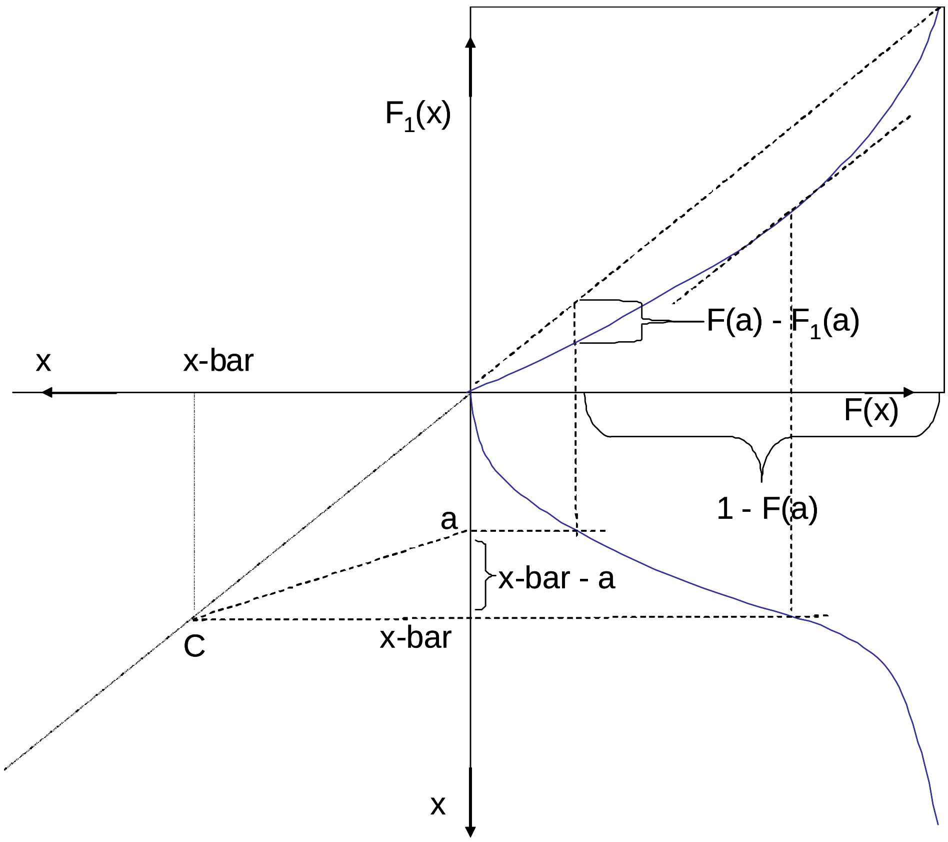

Figure 4: Components of $G(a)$

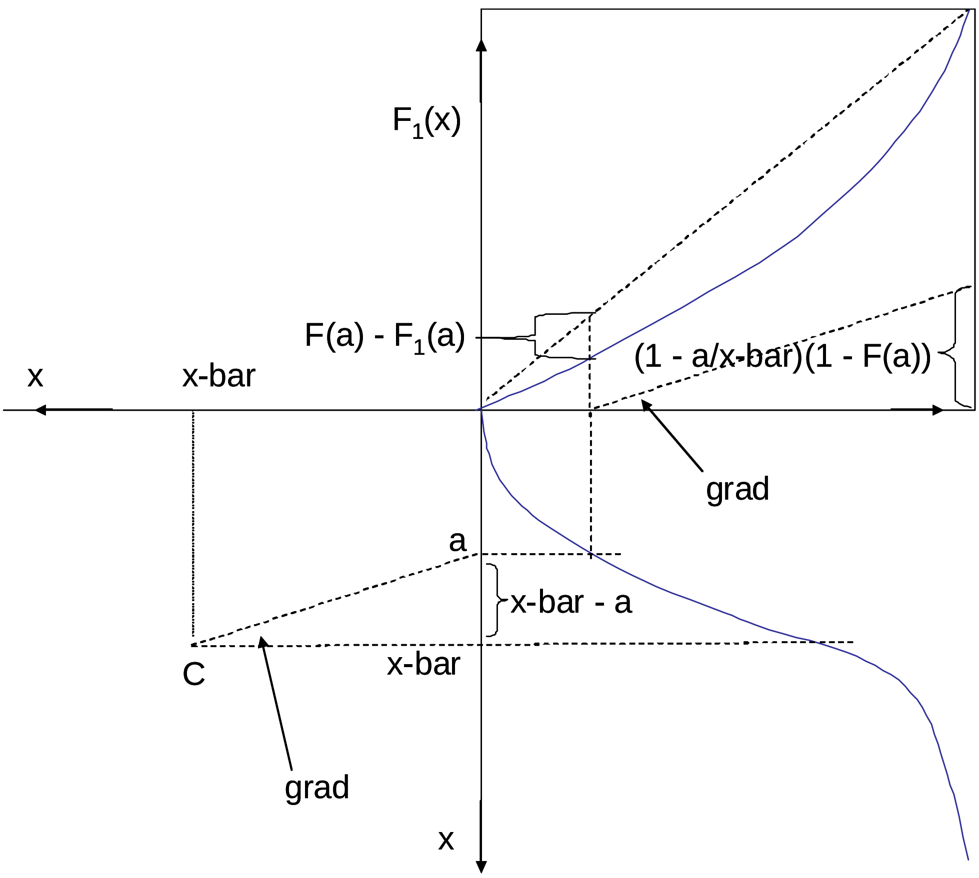

Figure 5: $G(a)$ as the Sum of Two Distances

can be written as:[3]

$$Gini = -\frac{2}{\bar{x}} Cov\left\{x, 1 - F(x)\right\} \tag{13}$$

the average value of the second term (using the fact that, for example, $Cov(x, y) = E(xy) - \bar{x}\bar{y}$) turns out to be simply half the Gini measure of inequality.

3 A Multi–step Tax Function

The most common form of income tax function used in practice is the multi–step function, which is described by a series of marginal tax rates and income thresholds over which the rates apply.[4] Formally, the multi–step income tax function can be written as:

$$\begin{aligned} T(x) &= 0 && 0 < x \leq a_1 \\ &= t_1(x - a_1) && a_1 < x \leq a_2 \\ &= t_1(a_2 - a_1) + t_2(x - a_2) && a_2 < x \leq a_3 \end{aligned} \tag{14}$$

and so on.

If x falls into the kth tax bracket, so that $a_k < x \leq a_{k+1}$, and $a_0 = t_0 = 0$, $T(x)$ can be written for $k \geq 1$ as:

$$T(x) = t_k(x - a_k) + \sum_{j=0}^{k-1} t_j(a_{j+1} - a_j) \tag{15}$$

The expression for $T(x)$ in (15) can be rewritten as:

$$T(x) = t_k x - \sum_{j=1}^{k} a_j(t_j - t_{j-1}) \tag{16}$$

Hence:

$$T(x) = t_k(x - a_k') \tag{17}$$

[3] See Creedy (1996, p. 20).

[4] For further discussion of the multi-step function, see Creedy (1996, pp. 54-6) and on the revenue elasticity properties, see Creedy and Gemmell (2006, pp. 25-32).

where:

$$
\begin{aligned}
a'_k &= a_k - \sum_{j=0}^{k-1} \left(\frac{t_j}{t_k} \right) (a_{j+1} - a_j) \\
&= \frac{1}{t_k} \sum_{j=1}^{k} a_j \left(t_j - t_{j-1} \right) \qquad (18)
\end{aligned}
$$

The implication of (17) and (18) is that the tax function facing any individual taxpayer is equivalent to a tax function with a single marginal tax rate, t_k, applied to income measured in excess of a single threshold, a'_k. The term, a'_k, is the effective threshold for individuals in the kth class, and is a weighted sum of the a_js, with weights, $\left(t_j - t_{j-1}\right)/t_k$, determined by the structure of marginal rate progression.

Rewrite (17) as:

$$
\begin{aligned}
T(x) &= t_k x - t_k a'_k \\
&= t_k x - \sum_{j=1}^{k} a_j \left(t_j - t_{j-1} \right) \qquad (19)
\end{aligned}
$$

The interpretation of the second term in (19) can be seen from Figure 6, which shows a section of the multi-step tax function. For threshold a_k, the value of $t_k a'_k$ is the sum of areas like A to the left of a_k. If all income were taxed at the rate t_k, too much tax would be paid, so the total area indicated represents this excess tax over the actual tax from the multi-rate structure. The artificial threshold, a'_k, thus ensures that the correct revenue is obtained.

Using (17), aggregate tax revenue can be written, for K tax brackets with $a_{K+1} = \infty$, as:

$$
R = N \sum_{k=1}^{K} \left[t_k \int_{a_k}^{a_{k+1}} \left(x - a'_k \right) dF(x) \right] \qquad (20)
$$

And using

$$
\int_{a_k}^{a_{k+1}} dF(x) = F(a_{k+1}) - F(a_k) \qquad (21)
$$

and

$$
\int_{a_k}^{a_{k+1}} x dF(x) = \bar{x} \left\{ F_1(a_{k+1}) - F_1(a_k) \right\} \qquad (22)
$$

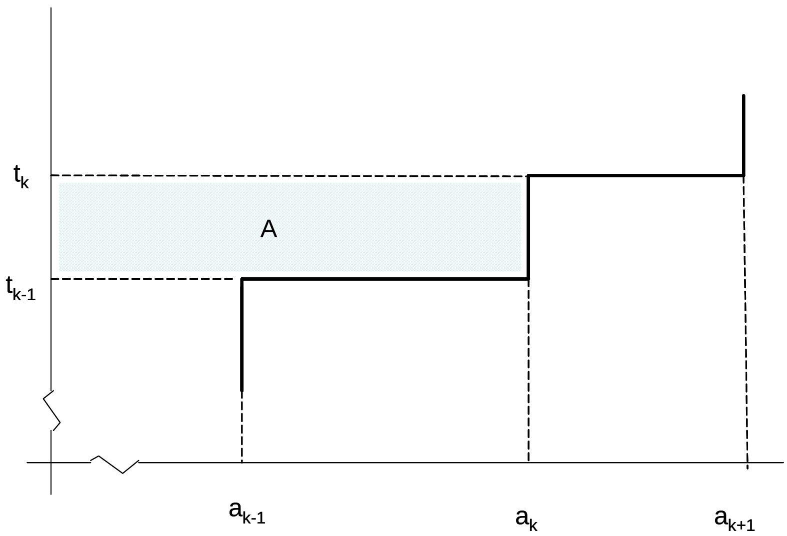

Figure 6: Mutli-rate Thresholds and Rates

it is found that:

$$R = N\bar{x}\sum_{k=1}^{K} t_k \left[\{F_1(a_{k+1}) - F_1(a_k)\} - \frac{a'_k}{x}\{F(a_{k+1}) - F(a_k)\} \right] \tag{23}$$

The term in square brackets can be written more succinctly by modifying the function $G(a)$ defined earlier. Let $G_k(a_k)$ denote the term in square brackets, so that:

$$R = N\bar{x}\sum_{k=1}^{K} t_k G_k(a_k) \tag{24}$$

or, alternatively:

$$R = \bar{x}\sum_{k=1}^{K} t_k N_k \left[\frac{F_1(a_{k+1}) - F_1(a_k)}{F(a_{k+1}) - F(a_k)} - \frac{a'_k}{\bar{x}} \right] \tag{25}$$

where N_k is the number of taxpayers in the kth tax bracket. This result is thus a simple extension of equation (9), where there is just one threshold. The first term in square brackets in (25) is the slope of the Lorenz curve between the two adjacent income thresholds. This is illustrated in Figure 7, as the slope AB. The term to be subtracted is not the actual threshold, but the effective threshold, a'_k, as a ratio of the arithmetic mean income.

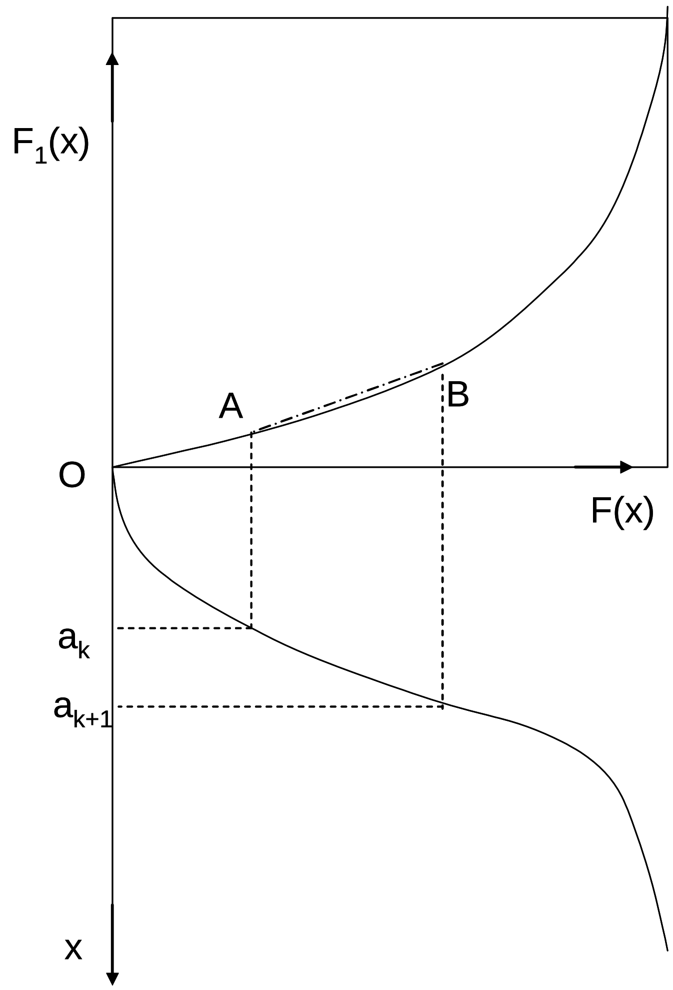

Figure 7: The Multi-step Tax Function

The effects of changes in population size, arithmetic mean income, the structure of marginal rates and thresholds, and of the dispersion of incomes (including a 'mean preserving spread' – a change in the dispersion which leaves arithmetic mean unchanged), can thus easily be examined, given the Lorenz curve and using the expression given above. It must be remembered that a change in the structure of marginal rates, with income thresholds unchanged, nevertheless changes the a'_k values.

4 Conclusions

This paper has shown how familiar expressions for total income tax revenue can be related to slopes and vertical distances in the popular Lorenz curve diagram, used to examine relative income inequality. This involves adding a quadrant showing the distribution function (the cumulative frequency density, or ogive curve). It is thereby possible to examine diagrammatically how income distribution changes can affect total tax revenue.

References

[1] Creedy, J. (1996) *Fiscal Policy and Social Welfare*. Cheltenham: Edward Elgar.

[2] Creedy, J. and Gemmell, N. (2006) *Modelling Tax Revenue Growth*. Cheltenham: Edward Elgar.

[3] Lambert, P. (1993) *The Distribution and Redistribution of Income*. Manchester: Manchester University Press.