Distributional Analyses in Discrete Hours Behavioural Microsimulation Models

John Creedy, Guyonne Kalb and Rosanna Scutella[1]
The University of Melbourne
Final report prepared for the Department of Family and Community Services

July 2003

[1]We should like to thank the Department of Family and Community Services for funding this research. The views expressed in this paper are those of the authors and do not represent the views of the Minister for Family and Community Services, the Department of Family and Community Services or the Commonwealth Government.

Abstract

This paper provides a guide through additions made to the available output on the distribution of income in the Melbourne Institute Tax and Transfer Simulator (Mitts). The range of measures of inequality, poverty and redistribution have been expanded in the non-behaviouaral component, while a method has been proposed and incorporated into the behavioural component to take account of the probabilistic nature of the labour supply estimates.

Recent studies have examined tax policy issues using labour supply models characterised by a discretised budget set. Microsimulation modelling using a discrete hours approach is probabilistic; that is, it does not identify a particular level of hours worked for each individual after a policy change, but generates a probability distribution over the discrete hours levels used. This makes analysis of the distribution of income difficult as even for a small sample with a modest range of labour supply points the range of possible labour supply combinations over the sample becomes quite large. This paper proposes a method of approximating measures of income distribution and compares the performance of this method to alternative approaches in a microsimulation context. In this approach a pseudo income distribution is constructed, which uses the probability of a particular labour supply value occurring (standardised by the population size) to refer to a particular position in the pseudo income distribution. This approach is compared to using an expected income level for each individual and to a simulated approach, in which labour supply values are drawn from each individual's hours distribution and summary statistics of the distribution of income are calculated by taking the average over each set of draws. The paper shows that the outcomes of various distributional measures using the pseudo method converge quickly to their true values as the sample size increases. The expected income approach results in a less accurate approximation and the computational burden of the simulated approach is larger when a comparable degree of precision as in the pseudo method is required. As an illustration of the pseudo approach, we simulate the labour supply responses and associated implications for the distribution of income of a basic income-flat tax system using the Melbourne Institute Tax and Transfer Simulator.

1 Introduction

The current range of distributional measures available in the Melbourne Institute Tax and Transfer Simulator (MITTS) is limited to the non-behavioural component, MITTS-A. Even then, the range of measures is restricted to a small set of poverty and inequality estimates, with only an absolute poverty line available the level of which needs to be manually entered by the user. A relative poverty line can be estimated outside of the model and then entered in by the user but ideally the option to use various relative poverty lines within the model should be directly available to the user of Mitts. Other distributional measures to facilitate policy debate are also introduced such as jobless household measures.

In the behavioural part of the Melbourne Institute Tax and Transfer Simulator, MITTS-B, it is currently not possible to calculate the poverty and inequality measures that are available in the static part, MITTS-A. The reason for this is that in the behavioural part individuals have a distribution over a set of discrete labour supply points rather than being assigned one particular labour supply point. As a result each individual is associated with a distribution of incomes rather than one income level, which demands a different approach in calculating poverty and inequality measures. In this paper a method is proposed and the results of the implementation of this method in MITTS are presented.

The objective of this project is to extend the range of distributional measures in MITTS to facilitate the analysis of policy reforms. The availability of distributional analyses in MITTS-B would enable us to compare changes in poverty and inequality in the static environment (MITTS-A) with the expected changes when accounting for behavioural changes in labour supply (MITTS-B). A simulation comparing the effects on poverty and inequality of a policy change with and without accounting for behavioural changes will illustrate the differences in outcomes using MITTS-A and MITTS-B.

A description of the distributional measures available in MITTS is presented in Section 2. Section 3 outlines the method proposed to introduce distributional measures into MITTS-B, with the results of Monte Carlo sim-

ulations performed to test the adequacy of this method. Illustrations of the options now available in MITTS with examples of the front end menus are presented in Section 4 while as an application to highlight the additions to MITTS Section 5 presents the simulation results of a hypothetical reform of a basic income - flat tax system in Australia. Section 6 concludes.

2 Distributional measures

This section provides an overview of the range of distributional measures now available in MITTS. In the following, N is the weighted population estimate, n is the sample size, w_i are the frequency weights for each observation, y_i is income for each observation and $\bar{y}_w = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$ is weighted average income over the population.

2.1 Measures of inequality

90/10 ratio

$$I_{90/10} = y_{90pt}/y_{10pt} \tag{1}$$

where y_{90pt} & y_{10pt} refer to the income levels at the 90th and 10th percentiles of the population weighted net income distribution.

Atkinson measure of inequality If ϵ refers to the Atkinson inequality aversion parameter the Atkinson index of inequality can be calculated by:

$$I_A = 1 - \left(\frac{1}{N} \sum_{i=1}^{n} w_i (y_i/\bar{y}_w)^{1-\epsilon} \right)^{1/(1-\epsilon)} \qquad if \quad \epsilon \neq 1 \tag{2}$$

$$I_A = 1 - \exp\left(\sum_{i=1}^{n} \frac{w_i \ln(y_i)}{N})/\bar{y}_w \right) \qquad if \quad \epsilon = 1 \tag{3}$$

Gini coefficient

$$I_G = \frac{N+1}{N-1} - \frac{2}{N(N-1)\bar{y}_w} \sum_{i=1}^{n} \rho_i w_i y_i \tag{4}$$

where ρ_i refers to the ranking of each observation in the population from richest to poorest and is calculated recursively to take account of the population weights: $\rho_{i+1} = \rho_i + w_i$

Theil's entropy measure

$$I_E = \frac{1}{N} \sum_{i=1}^{n} w_i \frac{y_i}{\bar{y}_w} \ln(w_i \frac{y_i}{\bar{y}_w}) \tag{5}$$

Coefficient of variation

$$I_{CoV} = \hat{s}_w / \bar{y}_w \tag{6}$$

where $\hat{s}_w$ is the weighted estimate of the population standard deviation:

$$\hat{s}_w^2 = \frac{n}{n-1} \left(\sum_{i=1}^{n} w_i (y_i - \bar{y}_w)^2 \right) /N \tag{7}$$

Lorenz curve The Lorenz curve is a plot of the cumulative fraction of the population - from poorest to richest -on the x-axis against the cumulative fraction of income on the y-axis.

2.2 Measures of the redistributive impact of the tax-transfer system

Reynolds-Smolensky measure of redistribution

$$L = I_{G(x)} - I_{G(y)} \tag{8}$$

where x refers to income before taxes and benefits and y refers to post tax/benefit or net income.

Concentration curve and index A Concentration curve is similar to a Lorenz curve, differing in that individuals/families are ranked according to pre-tax/benefit incomes (x) and the proportion of people is related to the corresponding proportion of total net income these individuals have. Thus, as the Gini coefficient is the index summarising the information in a Lorenz curve, a similar index can be estimated known as the Concentration Index:

$$C_Y = \frac{N+1}{N-1} - \frac{2}{N(N-1)\bar{y}_w} \sum_{i=1}^{n} \rho_i w_i y_i \tag{9}$$

where ρ_i now refers to the ranking of each observation in the population from richest to poorest before taxes/benefits.

Tax concentration By plotting these individuals (ranking by pre-tax/benefit income) against the proportion of total taxes paid we obtain a measure of tax concentration. This can also be represented by an index by substituting taxes for income when calculating the Gini coefficient:

$$C_t = \frac{N+1}{N-1} - \frac{2}{N(N-1)\bar{t}_w} \sum_{i=1}^{n} \rho_i w_i t_i \tag{10}$$

where like in the concentration curve ρ_i now refers to the ranking of each observation in the population from richest to poorest before taxes/benefits. Benefits (negative taxes) are problematic in this measure.

Kakwani progressivity index

$$K = C_t - I_{G(x)} \tag{11}$$

Atkinson-Plotnick reranking index

$$R = I_{G(y)} - C_y \tag{12}$$

2.3 Measures of poverty

Headcount If y_{pov} is the poverty line (here is a relative poverty line) and 1( ) indicates an indicator variable returning 1 if the statement is true and zero otherwise, the poverty headcount can be found by:

$$p_0 = \frac{1}{N}\sum_{i=1}^{n} w_i 1(y_i \leq y_{pov}) \tag{13}$$

Poverty gap

$$p_1 = \frac{1}{N}\sum_{i=1}^{n}(1 - \frac{y_i}{y_{pov}})w_i 1(y_i \leq y_{pov}) \tag{14}$$

Foster, Greer and Thorbecke poverty index

$$p_\alpha = \frac{1}{N}\sum_{i=1}^{n}(1 - \frac{y_i}{y_{pov}})^\alpha w_i 1(y_i \leq y_{pov}) \tag{15}$$

where α is a user specified parameter, where the larger the parameter the more the measure penalises poverty gaps (so it is kind of an 'intensity aversion parameter'). α=0 corresponds to the headcount measure and alpha=1 the poverty gap measure. The most commonly used level of α is 2.

Sen index

$$p_s = p_0\,(1 - (1 - I_G^p)\,\frac{\bar{y}_w^p}{y_{pov}}) \tag{16}$$

where I_G^p is the Gini coefficient of inequality within the 'poor' population and $\bar{y}_w^p$ refers to weighted mean income of the poor.

TIP curve A TIP curve refers to the three I's of poverty: incidence, intensity and inequality. A TIP curve is obtained by plotting the proportion of the population on the x-axis against the cumulative proportion of the poverty gap on the y-axis. Thus the length of the curve along the x-axis from the origin to the point where the curve is flat represents the incidence of poverty and is equivalent to the headcount measure, the height represents intensity and is equivalent to the poverty gap measure and the curvature represents inequality.

3 Distributional analysis in behavioural microsimulation

In this section we propose a method of approximating the expected inequality and poverty measures when using discrete hours labour supply modelling in tax policy microsimulation studies. It examines the performance of this approximation against alternative approaches and against a benchmark which is known to be close to the true value but is impractical to use with real-life data.

The traditional approach to the modelling of labour supply assumes that the decision variable, hours of work, is continuous and unconstrained. However, a number of recent studies have examined tax policy issues using labour supply models characterised by a discretised budget set.[1] There are several reasons for this. Firstly, analysts increasingly question whether a model which allows continuous substitution of hours for leisure constitutes a realistic representation of supply choices. For many socio-demographic groups labour market participation takes the form of fixed wage and hours contracts, with individuals choosing from among a discrete set of hours combinations (most often at part-time levels of around 20 hours, and at full-time levels of between 38 and 40 hours per week). Secondly, there are statistical and practical reasons to favour a discrete approach to the modelling of labour supply in preference to continuous models. These largely stem from the difficulties associated with the treatment of nonlinear budget constraints in continuous estimation. The strategy adopted in the discrete approach is to replace the budget set with a finite number of points, and optimise over only those discrete points.

Microsimulation modelling using a discrete hours approach is essentially probabilistic. That is, it does not identify a particular level of hours worked for each individual after the policy change, but generates a probability dis-

[1] See, for example, Van Soest (1995) and Keane and Moffitt (1998); alternative approaches are surveyed by Creedy and Duncan (2002). The approach allows both for random preference heterogeneity and state-specific errors in perception, and can incorporate either directly estimated or indirectly imputed fixed costs in estimation.

tribution over the discrete hours levels used. As a result, individuals have a set of probabilities of being at different income levels and the usual formulae for poverty and inequality measures cannot be applied.

Subsection 3.1 introduces the outcomes available after simulating policy reforms using a discrete hours labour supply model, and discusses three alternative approaches to inequality measurement. Special attention is given to the case of the variance; the formulae for the other measures are far less tractable; they are discussed briefly in the Appendix. In view of the difficulty of establishing analytical results, subsection 3.2 presents a Monte Carlo study for these other more complex measures. In Section 5, as an illustration, the method is applied to a hypothetical reform of the Australian tax and transfer system using the Melbourne Institute Tax and Transfer Simulator (MITTS).

3.1 Some Analytical Results

When simulating labour supply effects of policy reforms using microsimulation models the outcome is essentially probabilistic. That is, it does not identify a particular level of hours worked for each individual after the policy change, but generates a probability distribution over the discrete hours levels used. As a result, individuals have a set of probabilities of being at different income levels and the usual formulae for poverty and inequality measures cannot be applied. This section using analytical results to explore the performance of three alternative approaches to estimating summary measures of the income distribution after simulating changes in labour supply to various policy reforms. The three measures are described in subsection 3.1.1. In subsection 3.1.2, some properties of the methods are discussed and compared for a simple example of two individuals with two discrete hours points. This provides an insight into the nature of the general case, which is presented in subsection 3.1.3. These analytical results are for the variance, other inequality measures are too intractable to examine in this way. As an illustration, formulae for the other measures are presented for the three alternative approaches in the Appendix.

3.1.1 Alternative Approaches

The outcome of simulated policy changes using a discrete hours labour supply model to calculate behavioural changes is a set of probabilities for labour supply at a number of different hours points, for each individual or income unit. For convenience the following discussion is in term of individuals. Suppose there are n individuals and k possible outcomes of labour supply (hours levels) for each individual. This would result in k^n possible combinations of labour supply, and thus income distributions. Each outcome results in a different value for poverty and inequality measures.

Under the reasonable assumption that individuals' distributions of hours are independent, the probability of each income distribution, or outcome P_q, is given by the product of the relevant probabilities. Hence, if $p_{i,j}$ is the probability that individual j is at hours level i, the joint probability P_q is equal to $p_{i,1}p_{j,2}...p_{r,n}$, where q runs from 1 to k^n; and i, j, r can attain values between 1 and k (indicating the labour supply points chosen in combination q for each individual). In principle, each inequality or poverty measure can be calculated as the weighted sum of the measures over all possible outcomes, with weights equal to the probabilities P_q,where the sum over all $P_q$equals one. However, for any realistic sample size, even for few discrete labour supply points, the large number of possible combinations makes it computationally impractical to calculate all k^n distributions and associated probabilities P_q.[2]

Instead of examining all combinations of income levels, it would be possible to adopt a sampling approach. A large number of distributions could be obtained by taking random draws from each individual's hours distribution.[3] Each choice of discrete hours is drawn with the probability of it occurring for the relevant individual, so no weighting is required in averaging inequality

[2] It would be impossible to store all the information needed, but of course this would not be necessary as the appropriate weighted average could be obtained cumulatively using an algorithm for systematically working through all the k^n combinations. However, the computing time needed would be extremely long.

[3] This can be achieved by using a random number generator which produces random uniform variates between 0 and 1, which are then compared with the cumulative distribution at the hours points for each individual in order to select the relevant income for each sample drawn.

measures over the samples. With a sufficiently large number of randomly selected samples, the proportion of each hours combination would replicate the precise probabilities discussed above. This approach still requires a large computational effort, depending on the number of draws needed to obtain a good approximation. However, it provides a valuable way of examining the performance of alternative, less computer-intensive, approaches against this benchmark in a Monte Carlo experiment.

In considering alternative approaches which offer more practical solutions, the most obvious is perhaps a simple approach where the expected income is calculated for each individual and is used as if it were a single 'representative' level of income for that individual.[4] In addition, we explore an approach where all possible outcomes for every individual are used as if they are separate observations. The outcomes are weighted by the individual probabilities of labour supply to produce a pseudo distribution.

3.1.2 A Two-Person Example

Suppose there are just two individuals and two discrete hours levels. Details of associated incomes and probabilities are given in Table 1. There are therefore 4 possible alternative combinations, with associated probabilities, on the assumption that the probability distributions for different individuals are independent, as shown in Table 2.

Table 1: Income Distribution: A Two-Person Example

	Probabilities		Incomes	
Hours	person 1	person 2	person 1	person 2
h_1	p_1	p_2	$y_{1,1}$	$y_{1,2}$
h_2	$1-p_1$	$1-p_2$	$y_{2,1}$	$y_{2,2}$

Consider the arithmetic means of each combination, denoted $\overline{Y}_i$. The arithmetic mean of all these means, $\overline{Y}$, is given by:

$$\overline{Y} = p_1 p_2 \overline{Y}_1 + p_1 (1-p_2) \overline{Y}_2 + (1-p_1) p_2 \overline{Y}_3 + (1-p_1)(1-p_2) \overline{Y}_4 \quad (17)$$

[4] See for example, Gerfin and Leu (2003).

Table 2: Alternative Possible Income Distributions

Combination q	Incomes Person 1	Person 2	distributional measure g_q	Probability P_q
1	$y_{1,1}$	$y_{1,2}$	g_1	$p_1 p_2$
2	$y_{1,1}$	$y_{2,2}$	g_2	$p_1 (1-p_2)$
3	$y_{2,1}$	$y_{1,2}$	g_3	$(1-p_1) p_2$
4	$y_{2,1}$	$y_{2,2}$	g_4	$(1-p_1)(1-p_2)$
exact expected distributional measure:			$P_1 g_1 + P_2 g_2 + P_3 g_3 + P_4 g_4$	

Substituting for $\overline{Y}_1 = (y_{1,1} + y_{1,2})/2$, and so on, gives:

$$\overline{Y} = y_{1,1}\frac{p_1}{2} + y_{1,2}\frac{p_2}{2} + y_{2,1}\frac{1-p_1}{2} + y_{2,2}\frac{1-p_2}{2} \tag{18}$$

The arithmetic mean, $\overline{S}^2$, of variances for each possible combination, S_i^2, is given by:

$$\overline{S}^2 = p_1 p_2 S_1^2 + p_1 (1-p_2) S_2^2 + (1-p_1) p_2 S_3^2 + (1-p_1)(1-p_2) S_4^2 \tag{19}$$

Where $S_1^2 = \left(y_{1,1}^2 + y_{1,2}^2\right)/2 - \overline{Y}_1^2$, and so on. Hence:

$$\begin{aligned}
\overline{S}^2 \;=\; & p_1 p_2 \left\{ \frac{\left(y_{1,1}^2 + y_{1,2}^2\right)}{2} - \left(\frac{y_{1,1} + y_{1,2}}{2}\right)^2 \right\} \\
& + p_1 (1-p_2) \left\{ \frac{\left(y_{1,1}^2 + y_{2,2}^2\right)}{2} - \left(\frac{y_{1,1} + y_{2,2}}{2}\right)^2 \right\} \\
& + (1-p_1) p_2 \left\{ \frac{\left(y_{2,1}^2 + y_{1,2}^2\right)}{2} - \left(\frac{y_{2,1} + y_{1,2}}{2}\right)^2 \right\} \\
& + (1-p_1)(1-p_2) \left\{ \frac{\left(y_{2,1}^2 + y_{2,2}^2\right)}{2} - \left(\frac{y_{2,1} + y_{2,2}}{2}\right)^2 \right\}
\end{aligned} \tag{20}$$

Collecting terms, this can be written as:

$$\begin{aligned}
\overline{S}^2 \;=\; & p_1 \frac{y_{1,1}^2}{2} + p_2 \frac{y_{1,2}^2}{2} + (1-p_1)\frac{y_{2,1}^2}{2} + (1-p_2)\frac{y_{2,2}^2}{2} \\
& - \frac{1}{4}[p_1 y_{1,1}^2 + p_2 y_{1,2}^2 + (1-p_1) y_{2,1}^2 + (1-p_2) y_{2,2}^2 \\
& + 2 p_1 p_2 y_{1,1} y_{1,2} + 2 p_1 (1-p_2) y_{1,1} y_{2,2} \\
& + 2 p_2 (1-p_1) y_{2,1} y_{1,2} + 2(1-p_1)(1-p_2) y_{2,1} y_{2,2}]
\end{aligned} \tag{21}$$

The Expected Income Method Consider the use of the arithmetic mean income for each individual as a representative income.[5] These means are $\overline{y}_1 = p_1 y_{1,1} + (1-p_1)\, y_{2,1}$ and $\overline{y}_2 = p_2 y_{1,2} + (1-p_2)\, y_{2,2}$. The overall mean of the two individual expected values, $\overline{Y}_m$, is thus identical with (18). The variance of the individual expected incomes, S_m^2 is:

$$S_m^2 = \frac{\overline{y}_1^2 + \overline{y}_2^2}{2} - \overline{Y}_m^2 \tag{22}$$

The terms in (22) contain powers of the various probabilities. Hence $S_m^2 \neq \overline{S}^2$. The arithmetic means, as linear functions, are identical, but the variances, involving nonlinear functions of the various terms, are unequal. A similar feature is therefore expected for any inequality measure that is expressed as a nonlinear function of incomes.

Expanding (22) gives:

$$\begin{aligned} S_m^2 \;=\; & \frac{1}{4}[p_1^2 y_{1,1}^2 + p_2^2 y_{1,2}^2 + (1-p_1)^2 y_{2,1}^2 + (1-p_2)^2 y_{2,2}^2 \\ & +2p_1\,(1-p_1)\, y_{1,1} y_{2,1} + 2p_2\,(1-p_2)\, y_{1,2} y_{2,2}] - \\ & \frac{1}{2}[p_1 p_2 y_{1,1} y_{1,2} + p_1(1-p_2) y_{1,1} y_{2,2} + (1-p_1) p_2 y_{2,1} y_{1,2} + \\ & (1-p_1)(1-p_2) y_{2,1} y_{2,2}] \end{aligned} \tag{23}$$

The difference between this and $\overline{S}^2$ is thus given by:

$$\begin{aligned} 4\left(S_m^2 - \overline{S}^2\right) \;=\; & p_1^2 y_{1,1}^2 - p_1 y_{1,1}^2 + p_2^2 y_{1,2}^2 - p_2 y_{1,2}^2 + (1-p_1)^2 y_{2,1}^2 \\ & -(1-p_1) y_{2,1}^2 + (1-p_2)^2 y_{2,2}^2 - (1-p_2) y_{2,2}^2 \\ & +2p_1\,(1-p_1)\, y_{1,1} y_{2,1} + 2p_2\,(1-p_2)\, y_{1,2} y_{2,2} \\ =\; & p_1^2 y_{1,1}^2 + 2p_1\,(1-p_1)\, y_{1,1} y_{2,1} + (1-p_1)^2 y_{2,1}^2 \\ & +(1-p_2)^2 y_{2,2}^2 + 2p_2\,(1-p_2)\, y_{1,2} y_{2,2} + (1-p_2)^2 y_{2,2}^2 \\ & -p_1 y_{1,1}^2 - p_2 y_{1,2}^2 - (1-p_1) y_{2,1}^2 - (1-p_2) y_{2,2}^2 \\ =\; & (p_1 y_{1,1} + (1-p_1)\, y_{2,1})^2 + (p_2 y_{1,2} + (1-p_2) y_{2,2})^2 \\ & -(p_1 y_{1,1}^2 + (1-p_1) y_{2,1}^2) - (p_2 y_{1,2}^2 + (1-p_2) y_{2,2}^2) \\ =\; & (\overline{y_1})^2 - \overline{y_1^2} + (\overline{y_2})^2 - \overline{y_2^2} \end{aligned} \tag{24}$$

[5] Other possible candidates are the median and the mode of the hours distribution for each individual. These are rejected here on the grounds that they ignore potentially important information and the arithmetic means of resulting income distributions do not correspond to $\overline{Y}$.

Hence:

$$S_m^2 - \overline{S}^2 = \frac{1}{4}\left\{(\overline{y_1})^2 - \overline{y_1^2} + (\overline{y_2})^2 - \overline{y_2^2}\right\} \tag{25}$$

This confirms the trivial case where the two approaches give identical results for the variance if either the hours distributions are concentrated on a single hours level for each individual or the incomes are the same irrespective of the hours worked.

The Pseudo Income Distribution Method Consider an alternative approach involving the construction of a pseudo income distribution with four income levels, each associated with a corresponding probability. The distribution would have the four possible income levels $y_{1,1}$, $y_{1,2}$, $y_{2,1}$, $y_{2,2}$, with associated probabilities of $p_1/2$, $p_2/2$, $(1-p_1)/2$, and $(1-p_2)/2$. The division by 2 is required in order to ensure that the sum of the probabilities adds to 1.

The arithmetic mean of this pseudo distribution, $\overline{Y}_p$, clearly takes exactly the same form as $\overline{Y}$ in (18). The variance of this pseudo distribution, S_p^2, is given by:

$$\begin{aligned} S_p^2 &= \frac{p_1}{2}y_{1,1}^2 + \frac{p_2}{2}y_{1,2}^2 + \frac{1-p_1}{2}y_{2,1}^2 + \frac{1-p_2}{2}y_{2,2}^2 \\ &\quad - \left[y_{1,1}\frac{p_1}{2} + y_{1,2}\frac{p_2}{2} + y_{2,1}\frac{1-p_1}{2} + y_{2,2}\frac{1-p_2}{2}\right]^2 \end{aligned} \tag{26}$$

Again, this expression depends on the powers of the various probabilities, which appear in the term in square brackets, so it cannot be expected to equal the arithmetic mean of the individual sample variances given above. However the first set of terms in the two equations are the same. Letting $2\Phi = p_1y_{1,1}^2 + p_2y_{1,2}^2 + (1-p_1)y_{2,1}^2 + (1-p_2)y_{2,2}^2$, equation (26) becomes:

$$\begin{aligned} S_p^2 &= \Phi - \frac{1}{4}[p_1^2y_{1,1}^2 + p_2^2y_{1,2}^2 + (1-p_1)^2y_{2,1}^2 + (1-p_2)^2y_{2,2}^2 \\ &\quad +2p_1p_2y_{1,1}y_{1,2} + 2p_1(1-p_2)y_{1,1}y_{2,2} \\ &\quad +2p_2(1-p_1)y_{2,1}y_{1,2} + 2(1-p_1)(1-p_2)y_{2,1}y_{2,2} \\ &\quad +2p_1(1-p_1)y_{1,1}y_{2,1} + 2p_2(1-p_2)y_{1,2}y_{2,2}] \end{aligned} \tag{27}$$

Hence the difference between this and $\overline{S}^2$ is thus given by:

$$\begin{aligned}
4\left(S_p^2-\overline{S}^2\right) &= -p_1^2y_{1,1}^2+p_1y_{1,1}^2-p_2^2y_{1,2}^2+p_2y_{1,2}^2-(1-p_1)^2y_{2,1}^2 \\
&\quad +(1-p_1)y_{2,1}^2-(1-p_2)^2y_{2,2}^2+(1-p_2)y_{2,2}^2 \\
&\quad -2p_1\left(1-p_1\right)y_{1,1}y_{2,1}-2p_2\left(1-p_2\right)y_{1,2}y_{2,2} \\
&= -p_1^2y_{1,1}^2-2p_1\left(1-p_1\right)y_{1,1}y_{2,1}-\left(1-p_1\right)^2y_{2,1}^2 \\
&\quad -\left(1-p_2\right)^2y_{2,2}^2-2p_2\left(1-p_2\right)y_{1,2}y_{2,2}-\left(1-p_2\right)^2y_{2,2}^2 \\
&\quad +p_1y_{1,1}^2+p_2y_{1,2}^2+(1-p_1)y_{2,1}^2+(1-p_2)y_{2,2}^2 \\
&= -\left(p_1y_{1,1}+\left(1-p_1\right)y_{2,1}\right)^2-\left(p_2y_{1,2}+\left(1-p_2\right)y_{2,2}\right)^2 \\
&\quad +(p_1y_{1,1}^2+(1-p_1)y_{2,1}^2)+(p_2y_{1,2}^2+(1-p_2)y_{2,2}^2) \qquad (28) \\
&= -\left\{\left(\overline{y_1}\right)^2-\overline{y_1^2}+\left(\overline{y_2}\right)^2-\overline{y_2^2}\right\}
\end{aligned}$$

and:

$$S_p^2-\overline{S}^2=-\frac{1}{4}\left\{\left(\left(\overline{y_1}\right)^2-\overline{y_1^2}\right)+\left(\left(\overline{y_2}\right)^2-\overline{y_2^2}\right)\right\} \qquad (29)$$

This shows that in the simple two by two case, the pseudo distribution overstates the value of $\overline{S}^2$ to precisely the same extent that the use of expected incomes understates the value.

3.1.3 The General Case

Suppose there are n individuals and k discrete hours levels, $h_1, ..., h_k$. Let $p_{j,i}$ and $y_{j,i}$ denote respectively the ith person's income at hours level j, and the probability of hours level j. On the assumption that the probability distributions for different individuals are independent, the joint probabilities of the k^n possible alternative combinations are as shown in Table 3. An arbitrary combination m from the set of all possible combinations consists of the set of hours points $h_{1m}, h_{2m}, ..., h_{nm}$ for persons 1 to n respectively, where h_{im} runs from 1 to k, indicating one of the k possible hours points for each person.

Consider the arithmetic means of each possible combination, denoted $\overline{Y}_i$. The arithmetic mean of all these means, $\overline{Y}$, is given by:

$$\overline{Y}=p_{1,1}p_{1,2}...p_{1,n}\overline{Y}_1+...+p_{h_{1m},1}p_{h_{2m},2}...p_{h_{nm},n}\overline{Y}_m+...+p_{k,1}p_{k,2}...p_{k,n}\overline{Y}_{k^n} \quad (30)$$

Table 3: Alternative Possible Income Distributions

Combination q	Incomes of persons: 1	2	...	n	distributional measure g_q	Probability P_q
1	$y_{1,1}$	$y_{1,2}$	$\cdots$	$y_{1,n}$	g_1	$p_{1,1}p_{1,2}\cdots p_{1,n}$
m	$y_{h_{1m},1}$	$y_{h_{2m},2}$	$\cdots$	$y_{h_{nm},n}$	g_m	$p_{h_{1m},1}p_{h_{2m},2}\cdots p_{h_{nm},n}$
k^n	$y_{k,1}$	$y_{k,2}$	$\cdots$	$y_{k,n}$	g_{k^n}	$p_{k,1}p_{k,2}\cdots p_{k,n}$
exact expected distributional measure:					$\sum_{q=1}^{k^n} P_q g_q$	

Substituting for $\overline{Y}_m = \frac{1}{n}\sum_{i=1}^{n} y_{h_{im},i}$, gives:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k} p_{h,i}y_{h,i} \tag{31}$$

The arithmetic mean, $\overline{S}^2$, of the variances for each possible combination, S_i^2, is:

$$\overline{S}^2 = p_{1,1}p_{1,2}\cdots p_{1,n}S_1^2 + \ldots + p_{h_{1m},1}p_{h_{2m},2}\cdots p_{h_{nm},n}S_m^2 + \ldots + p_{k,1}p_{k,2}\cdots p_{k,n}S_{k^n}^2 \tag{32}$$

Where $S_m^2 = \frac{1}{n}\sum_{i=1}^{n} y_{h_{im},i}^2 - \overline{Y}_m^2$.

Hence:

$$
\begin{aligned}
\overline{S}^2 &= \sum_{m=1}^{k^n} p_{h_{1m},1} p_{h_{2m},2} \cdots p_{h_{nm},n} \frac{1}{n} \sum_{i=1}^{n} y_{h_{im},i}^2 \\
&\quad - \sum_{m=1}^{k^n} p_{h_{1m},1} p_{h_{2m},2} \cdots p_{h_{nm},n} \left(\frac{1}{n} \sum_{i=1}^{n} y_{h_{im},i} \right)^2 \\
&= \frac{1}{n} \sum_{h=1}^{k} \sum_{i=1}^{n} p_{h,i} y_{h,i}^2 \\
&\quad - \sum_{m=1}^{k^n} p_{h_{1m},1} p_{h_{2m},2} \cdots p_{h_{nm},n} \left(\frac{1}{n} \sum_{i=1}^{n} y_{h_{im},i} \right)^2 \\
&= \frac{1}{n} \sum_{h=1}^{k} \sum_{i=1}^{n} p_{h,i} y_{h,i}^2 - \frac{1}{n^2} \sum_{h=1}^{k} \sum_{i=1}^{n} p_{h,i} y_{h,i}^2 \\
&\quad - \frac{1}{n^2} \sum_{m=1}^{k^n} p_{h_{1m},1} p_{h_{2m},2} \cdots p_{h_{nm},n} \left(2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} y_{h_{im},i} y_{h_{jm},j} \right) \\
&= \frac{1}{n} \sum_{h=1}^{k} \sum_{i=1}^{n} p_{h,i} y_{h,i}^2 - \frac{1}{n^2} \sum_{h=1}^{k} \sum_{i=1}^{n} p_{h,i} y_{h,i}^2 \\
&\quad - \frac{1}{n^2} \left(2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} \sum_{h=1}^{k} \sum_{l=1}^{k} p_{h,i} y_{h,i} p_{l,j} y_{l,j} \right) \qquad (33)
\end{aligned}
$$

The Use of Expected Incomes Using expected incomes, $\overline{y}_1 = \sum_{h=1}^{k} p_{h,1} y_{h,1}$ to $\overline{y}_n = \sum_{h=1}^{k} p_{h,n} y_{h,n}$, the overall mean of the n individual means, $\overline{Y}_m$, is

identical with (31). The variance of the individual mean incomes, S_m^2 is:

$$\begin{aligned}
\overline{S}_m^2 &= \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i^2 - \overline{Y}_m^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{h=1}^{k} p_{h,i}y_{h,i}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k} p_{h,i}y_{h,i}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{h=1}^{k} p_{h,i}^2 y_{h,i}^2 + 2\sum_{h=1}^{k}\sum_{j=1}^{h-1} p_{h,i}y_{h,i}p_{j,i}y_{j,i}\right) \\
&\quad -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{h=1}^{k} p_{h,i}^2 y_{h,i}^2 - \frac{2}{n^2}\sum_{i=1}^{n}\sum_{h=1}^{k}\sum_{j=1}^{h-1} p_{h,i}y_{h,i}p_{j,i}y_{j,i} \\
&\quad -\frac{2}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{h=1}^{k}\sum_{l=1}^{k} p_{h,i}y_{h,i}p_{l,j}y_{l,j}\right) \qquad (34)
\end{aligned}$$

The difference between the method using all combinations and the method using the expected income is:

$$\begin{aligned}
\overline{S}_m^2 - \overline{S}^2 &= \left(\frac{1}{n} - \frac{1}{n^2}\right)\sum_{i=1}^{n}\left(\sum_{h=1}^{k} p_{h,i}y_{h,i}\right)^2 - \left(\frac{1}{n} - \frac{1}{n^2}\right)\sum_{i=1}^{n}\sum_{h=1}^{k} p_{h,i}y_{h,i}^2 \\
&= \left(\frac{n-1}{n^2}\right)\sum_{i=1}^{n}\left(\left(\overline{y}_i\right)^2 - \overline{y_i^2}\right)
\end{aligned}$$

The Use of a Pseudo Income Distribution Consider the pseudo income distribution with nk income levels, each associated with a corresponding probability. The incomes are $y_{h,i}$, where h ranges from 1 to k and i ranges from 1 to n, and associated probabilities are $p_{h,i}/n$. The division by n ensures that the sum of the probabilities adds to 1. The $y_{h,i}$ values are placed in a single vector, $z = \{y_{h,i}\}$ with nk elements, with the associated probabilities given by $p' = \{p_{h,i}/n\}$. Hence:

$$\sum_{j=1}^{nk} p'_j = \sum_{i=1}^{n}\sum_{h=1}^{k} p_{h,i}/n = 1 \qquad (35)$$

The arithmetic mean of this pseudo distribution, $\overline{Y}_p$, must take the same form as $\overline{Y}$ in (31) above. The variance of this pseudo distribution, S_p^2, is

given by:

$$
\begin{aligned}
\overline{S}_p^2 &= \frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k} p_{h,i}y_{h,i}^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k} p_{h,i}y_{h,i}\right]^2 \\
&= \frac{1}{n}\sum_{h=1}^{k}\sum_{i=1}^{n} p_{h,i}y_{h,i}^2 - \frac{1}{n^2}\sum_{h=1}^{k}\sum_{i=1}^{n} p_{h,i}^2 y_{h,i}^2 \\
&\quad -\frac{2}{n^2}\sum_{i=1}^{n}\sum_{h=1}^{k}\sum_{j=1}^{h-1} p_{h,i}y_{h,i}p_{j,i}y_{j,i} \\
&\quad -\frac{2}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{h=1}^{k}\sum_{l=1}^{k} p_{h,i}y_{h,i}p_{l,j}y_{l,j}\right) \qquad (36)
\end{aligned}
$$

The difference between the method using all combinations and that using the pseudo method is:

$$
\begin{aligned}
\overline{S}_p^2 - \overline{S}^2 &= -\frac{1}{n^2}\left(\sum_{h=1}^{k}\sum_{i=1}^{n}\left(p_{h,i}^2 y_{h,i}^2 - p_{h,i}y_{h,i}^2\right) + 2\sum_{i=1}^{n}\sum_{h=1}^{k}\sum_{j=1}^{h-1} p_{h,i}y_{h,i}p_{j,i}y_{j,i}\right) \\
&= -\frac{1}{n^2}\left(\sum_{h=1}^{k}\sum_{i=1}^{n}\left(p_{h,i}^2 y_{h,i}^2 + 2\sum_{j=1}^{h-1} p_{h,i}y_{h,i}p_{j,i}y_{j,i}\right) - \sum_{h=1}^{k}\sum_{i=1}^{n} p_{h,i}y_{h,i}^2\right) \\
&= -\frac{1}{n^2}\left(\sum_{i=1}^{n}\left(\sum_{h=1}^{k} p_{h,i}y_{h,i}\right)^2 - \sum_{h=1}^{k}\sum_{i=1}^{n} p_{h,i}y_{h,i}^2\right) \\
&= -\frac{1}{n^2}\sum_{i=1}^{n}\left(\left(\overline{y}_i\right)^2 - \overline{y_i^2}\right) \qquad (37)
\end{aligned}
$$

Comparing this difference to the difference between the method using expected income demonstrates that for samples of more than two persons estimates of the variance using the pseudo method is closer to the method using all combinations, compared with estimates using the expected income method. Furthermore, the true outcome lies in between the pseudo method and the expected income method. For larger samples the approximation to the true value is expected to improve for the pseudo method.

3.2 Monte Carlo Experiments

The previous section compares differences in average income and the variance using three various approaches analytically. As mentioned earlier, to compare other summary measures of the income distribution analytically would be cumbersome and intractable, particularly with measures of inequality and poverty. Thus this section uses Monte Carlo simulation methods to examine the relative performance of the alternative approaches discussed in the previous section.

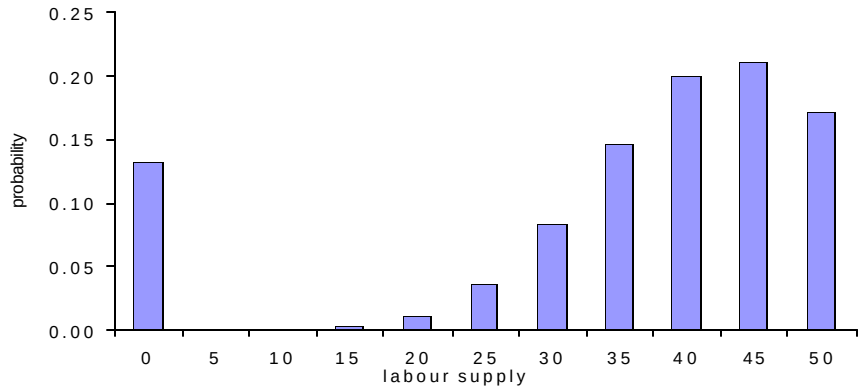
3.2.1 The Experimental Approach

To ensure that the simulations were based on realistic distributions, a base sample of 10,293 individuals was produced using the Melbourne Institute Tax and Transfer Simulator (MITTS).[6] This sample was generated from the confidentialised unit record files from the 1996/7 and 1997/8 Surveys of Income and Housing Costs made available by the Australian Bureau of Statistics. To avoid the additional complexity (and additional run time) introduced by examining the hours distribution of couples jointly and to keep the number of hours points to a minimum for the experiment, a sample of single persons was selected. The sample provides, for each individual, the incomes and probabilities associated with $k = 11$ discrete hours levels ranging from 0 to 50 hours in five-hourly bands.

In simulating the effects of reforms, MITTS produces a set of probabilities of being at each labour supply point for each income unit after the reform. These probabilities are estimated through a process of calibration, ensuring the pre-reform labour supply is equal to the observed value. As this generally produces relatively high probabilities for a particular hours point, which would improve the outcomes of the approximating methods examined here, the raw probabilities generated from the initial labour supply estimation are used for the Monte Carlo study. This generates a more diverse range of probabilities for each individual and tests the usefulness of the different methods more rigorously. Examples of three hours distributions are shown in Table 4.

[6] For details of this model, see Creedy *et al.* (2002).

Table 4: Examples of Hours Distributions



It is of interest to consider the properties of the measures as the sample size n varies, so values of n of 50, 100, 250 and 500 were examined. The use of the smaller sizes is relevant where samples are divided into particular socio-economic or demographic groups. In each case, 1000 subsamples of size n were drawn from this basic data set. Gini, G, and Atkinson, A (with relative inequality aversion of 0.4), inequality measures were computed along with the variance, V. In addition, three poverty measures were computed from the Foster, Greer and Thorbecke (1984) family. The measures chosen are the headcount measure, P_0; one that depends on the extent to which individuals fall below the poverty line, P_1; and finally a measure that also depends on the coefficient of variation of incomes of those in poverty, P_2. The poverty line was set in relative terms at 50 per cent of median income; hence the poverty line varies across samples.

First it is necessary to consider the behaviour of the method of sampling from the set of possible alternative distributions. Although the number of possible distributions is huge (even with samples sizes of 50, with 11 hours levels, there are 11^{50} possible combinations), it was found that the values for the inequality measures and the values for mean income converged by $50,000$ draws.[7] The convergence of mean income and the Gini inequality measure are shown in Figures 1 and 2 respectively, for the case of $n = 50$. Of course, this does not provide a practical approach that could be used in microsimulation models, particularly for larger samples.[8] However, the associated stable values can be regarded as the 'true' values against which the performance of the alternative approaches may be gauged.[9]

[7] Convergence takes place more quickly for the larger samples. However in many instances, the difference between the 'true' measure or mean and the measure or mean found through the sampling approach is already small using many fewer draws even for the smaller sample sizes.

[8] For $n = 500$, taking 1000 repetitions, each of which involves $50,000$ random draws from the 500 individuals' hours distributions in the sample, the computation time is substantial at around four days.

[9] There is one exception. It has been demonstrated above that the use of average incomes, and the pseudo distribution, provide exact measures of the (appropriately weighted) mean, $\overline{X}$. Where 10 draws are taken with the sampling approach, the mean from the pseudo distribution has been taken as the 'true' value, rather than that produced by 50,000 draws.

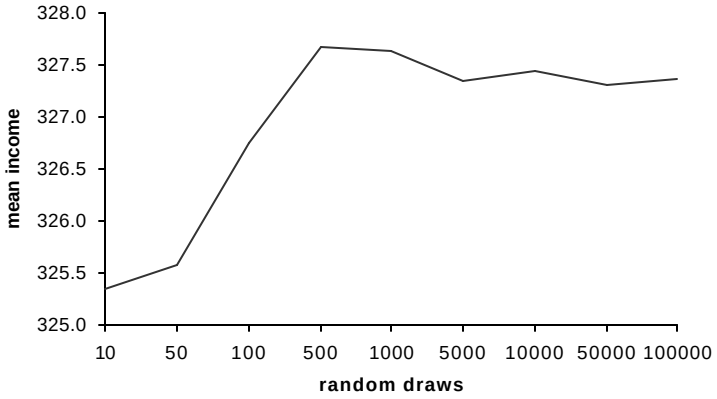


Figure 1: Convergence for Mean Income

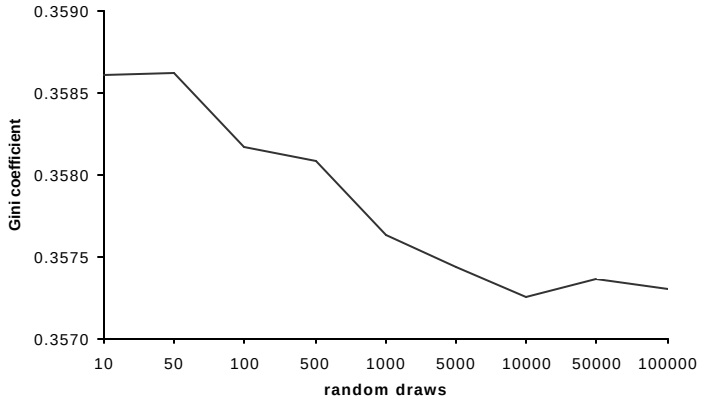


Figure 2: Convergence For Gini Inequality Measure

3.2.2 Experimental Results

Tables 5 to 8 present the Monte Carlo results for the three approaches comparing deviations of the various approaches from the 'true' values for 1000 replications.[10] The first two methods are the expected income method and the pseudo method. To explore the quality of the sampling approach when few draws are used, results are also produced for the sampling approach with 10 draws. The use of 10 draws would be feasible in practice and this number has been shown to be sufficient in other situations, such as for example, Simulated Maximum Likelihood estimations.[11]

Table 5: Distribution of Differences from 'True Value': n=50

	G	A	P_0	P_1	P_2	$\overline{X}$	V
Average "true" values							
	0.33391	0.084995	0.115037	0.062394	0.047425	318.27	42674.11
Use of each individual's expected income							
Mean	-0.0336	-0.01199	-0.00526	9.18E-06	0.000372		-7303.94
mse[a]	0.001174	0.000148	0.000404	2.78E-05	1.26E-05		55194286
Min	-0.06056	-0.01993	-0.0821	-0.01785	-0.01328		-11498.1
Max	-0.01467	-0.00621	0.08906	0.025669	0.017747		-3531.46
Use of pseudo distribution of income							
Mean	0.004543	0.000275	0.001585	0.000748	0.000459		149.8281
mse[a]	0.000021	7.87E-08	0.000083	2.52E-06	8.82E-07		23739.04
Min	0.001503	0.000079	-0.0253	-0.00373	-0.00267		4.246241
Max	0.006876	0.000499	0.041592	0.006866	0.004414		314.5633
Use of 10 random draws from possible distributions							
Mean	-3.5E-05	-6.3E-06	0.000419	9.77E-05	7.14E-05	0.093905	-41.5227
mse[a]	0.000017	3.24E-06	0.000057	5.67E-06	3.1E-06	14.72302	2736878
Min	-0.01408	-0.0078	-0.02878	-0.01001	-0.00698	-11.542	-11104.1
Max	0.011255	0.006216	0.049411	0.013506	0.012192	13.13847	9338.877
Note a: mse stands for the mean squared error.							

From all measures in the tables, it is clear that the pseudo method per-

[10] The experiment where a sample of 500 individuals is used took roughly four days to run. This involved drawing around 50,000,000 times (which is less than 10^8) from the 500 hours distributions and computing averages across the 50,000 draws in each of the 1000 replications. The exact measure for the same sample would involve 11^{500} different combinations which compared to the above drawing of 50,000 would require an enormous amount of time and be infeasible to carry out.

[11] In Van Soest (1995), 5 draws seemed sufficient in one of the models estimated using the Simulated Maximum Likelihood approach.

Table 6: Distribution of Differences from 'True Value': n=100

	G	A	P_0	P_1	P_2	$\overline{X}$	V
Average "true" values							
	0.337054	0.085744	0.112875	0.062622	0.047801	318.11	43145.16
Use of each individual's expected income							
Mean	-0.03405	-0.01218	-0.00248	4.47E-05	0.000303		-7369.36
mse[a]	0.00118	0.00015	0.000194	1.23E-05	5.97E-06		55064023
Min	-0.0494	-0.0167	-0.05305	-0.01051	-0.00849		-10896.5
Max	-0.0194	-0.00813	0.062672	0.014378	0.008512		-4648.41
Use of pseudo distribution of income							
Mean	0.002273	0.000137	0.001157	0.000364	0.000237		74.73766
mse[a]	5.23E-06	1.94E-08	0.000039	7.18E-07	2.51E-07		5944.869
Min	0.001236	0.000063	-0.02872	-0.00286	-0.0011		6.374102
Max	0.003116	0.000244	0.025762	0.003468	0.002167		155.9846
Use of 10 random draws from possible distributions							
Mean	1.91E-05	1.3E-05	0.000204	6.33E-05	3.97E-05	0.185203	31.26735
mse[a]	7.41E-06	1.55E-06	0.000022	2.54E-06	1.54E-06	7.814952	1479075
Min	-0.00835	-0.0045	-0.02395	-0.00516	-0.00408	-10.4604	-7314.95
Max	0.009678	0.005463	0.022216	0.005611	0.004703	10.86006	5933.581

Note a: mse stands for the mean squared error.

Table 7: Distribution of Differences from 'True Value': n=250

	G	A	P_0	P_1	P_2	$\overline{X}$	V
Average "true" values							
	0.338943	0.086112	0.114024	0.063022	0.048044	317.97	43322.27
Use of each individual's expected income							
Mean	-0.03437	-0.01227	-0.00156	0.000181	0.000301		-7417.44
mse[a]	0.001191	0.000151	0.000108	5.11E-06	2.53E-06		55352057
Min	-0.04348	-0.01506	-0.03549	-0.00592	-0.00429		-10211.7
Max	-0.02511	-0.01005	0.03473	0.008382	0.005616		-5802.55
Use of pseudo distribution of income							
Mean	0.000911	5.38E-05	0.00057	0.00012	8.41E-05		29.81149
mse[a]	8.35E-07	3.05E-09	2.05E-05	1.75E-07	5.52E-08		1006.12
Min	0.000678	0.000016	-0.01327	-0.00145	-0.0008		-10.9953
Max	0.001146	0.000104	0.014827	0.001469	0.000927		85.59714
Use of 10 random draws from possible distributions							
Mean	-8.5E-05	-3.4E-05	2.62E-05	8.59E-06	-7.9E-07	0.029564	-10.633
mse[a]	3.25E-06	6.53E-07	9.87E-06	9.77E-07	5.83E-07	2.999281	595096.4
Min	-0.00574	-0.00267	-0.01384	-0.00367	-0.00302	-5.36467	-3252.28
Max	0.004986	0.002338	0.016108	0.003251	0.002815	4.945107	2578.688

Note a: mse stands for the mean squared error.

Table 8: Distribution of Differences from 'True Value': n=500

	G	A	P_0	P_1	P_2	$\bar{X}$	V
Average "true" values							
	0.340422	0.086779	0.115116	0.063697	0.048700	317.89	43486.24
Use of each individual's expected income							
Mean	-0.03424	-0.01227	-0.00090	0.000426	0.000361		-7422.69
mse[a]	0.001177	0.000151	0.000077	2.96E-06	1.35E-06		55254183
Min	-0.04077	-0.01474	-0.02098	-0.00449	-0.00378		-8674.50
Max	-0.02718	-0.01000	0.02746	0.006148	0.004455		-6235.62
Use of pseudo distribution of income							
Mean	0.000454	0.000026	0.00021	0.000010	0.000020		14.33324
mse[a]	2.07E-07	7.68E-10	1.68E-05	5.30E-08	1.55E-08		260.649
Min	0.000374	-0.000001	-0.01341	-0.000817	-0.000471		-20.8162
Max	0.000543	0.000051	0.012726	0.000951	0.000596		42.10908
Use of 10 random draws from possible distributions							
Mean	0.000072	3.12E-05	-5.84E-05	-3.02E-05	-7.50E-06	-0.001639	24.1993
mse[a]	1.55E-06	3.05E-07	6.84E-06	5.18E-07	3.12E-07	1.522162	256246.7
Min	-0.003427	-0.001878	-0.01053	-0.00198	-0.00184	-4.19253	-1608.05
Max	0.004071	0.001896	0.010913	0.002368	0.001782	3.823399	1670.176

Note a: mse stands for the mean squared error.

forms much better than the expected income method. As shown analytically for the variance, the two inequality measures are also always underestimated by the expected income method and overestimated by the pseudo method for samples with 50 or 100 individuals. However, for samples with 250 or 500 individuals, some negative differences between the "true" variance and the variance from the pseudo method are found. This does not happen very often and it can be explained by the fact that our "true" values which are used as benchmarks in the experiment are approximations as well. As the pseudo method improves with sample size, for n=250 or 500 the pseudo method may sometimes get closer to the true value than the approximation by 50,000 draws, causing the difference to be negative rather than positive. This will happen more often for larger samples, making the pseudo method a computationally very efficient approach with an approximate value close to the true value.

The sampling method, using 10 draws ranks in between the two other methods. Only the mean in this approach is always smaller, but that is because negative and positive values partly offset each other in this approach, whereas the other two approaches are either always over- or underestimating

the true values.

With an increase in sample size, the methods tend to perform better with the exception of the variance and inequality measures for the expected income method. Although the minimum and maximum values move towards each other in the expected income method, a bias remains. In the pseudo method and the sampling method with 10 draws the minimum and maximum value both move towards the true value. The pseudo method improves faster with sample size than the expected income method. The pseudo method clearly outperforms the sampling method with 10 draws except for the case of the headcount measure P_0. Nevertheless, on many occasions the accuracy of the sampling method using only 10 draws would be sufficient.

4 Additions to MITTS

This section provides a short visual guide through the distributional measures now available in MITTS with details of the various options available in the front end menus. All of the revisions made to MITTS have been fully documented in the MITTS manual.[12] First we provide a guide through the options now available in Mitts-A and then we examine the new options in Mitts-B. For information on any particular measure of inequality, poverty or redistribution see section 2.

4.1 Mitts-A

After a simulation run in Mitts-A there are a various number of options to generate distributional measures pre and post reform. The first step in generating these is to choose the option "Analyse results" in the array of options made available in Mitts-A as shown in figure 3 below.

Once "Analyse results" is selected a range of possible items are made available at the bottom of the screen: Winners/Losers, Income Changes, Inequality, Poverty and METRs and RRs (see figure 4).

[12] A link to the latest version of the MITTS manual can be found at http://www1.ecom.unimelb.edu.au/iaesrwww/lsfs/mitts.html.

Figure 3: Mitts-A menu items

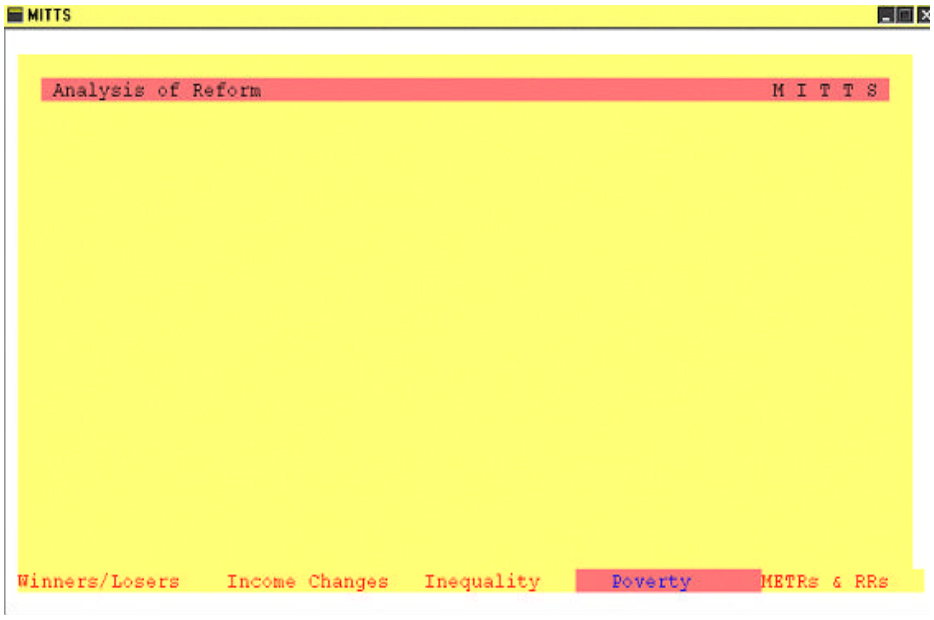


Figure 4: Menu items in "Analyse Results"

Figure 5: Various measures of income in inequality and poverty settings - Mitts-A

If the "inequality" option is selected, a menu pops up with a range of inequality measures listed as is shown in figure 6. The first option is "INEQUAL SETTINGS" and if selected allows the user to select the unit of measurement for income. There are now seven options to choose from as are shown in figure 5: 1) income unit income per adult equivalent, 2) income unit income per adult equivalent weighted for size of income unit, 3) per capita income unit income – where income is shared equally between each member of an income unit 4) household income per adult equivalent, 5) household income per adult equivalent weighted for household size, 6) and 7) Individual income. The default setting is option 1) income unit income per adult equivalent. Note that income always refers to net income.

The coefficient of variation and Theil's Entropy measure are the new additions to the range of inequality measures. We have also revised the output in relation to the Atkinson indices. Rather than allow the user to specify the inequality aversion parameter here, we now provide indices for a range of parameter values (0.1,0.5,1,2 and 3) so the user can see how sensitive

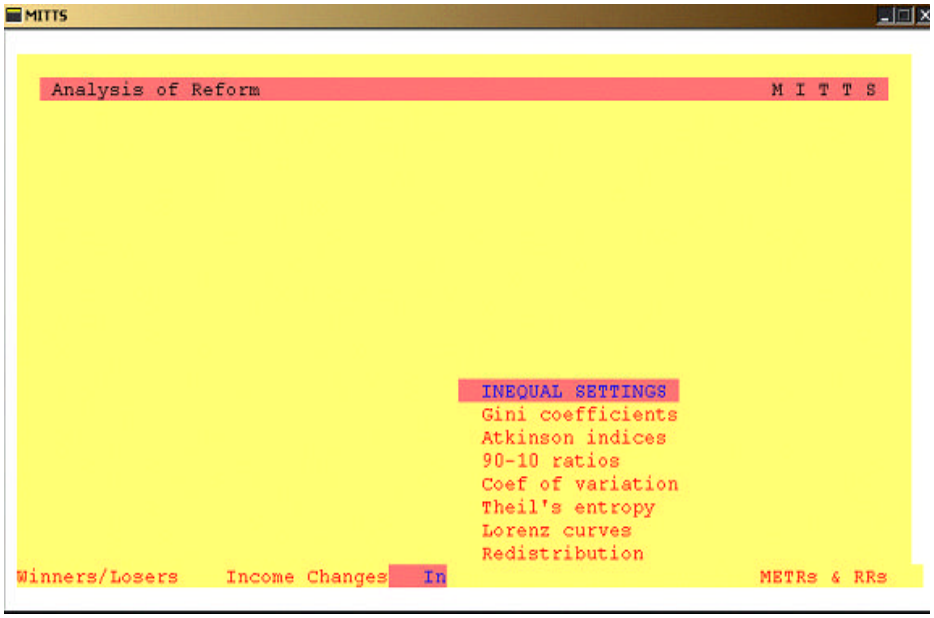


Figure 6: Inequality measures

the measure is to the various values.

Figure 7 provides an example of the output window if a particular inequality measure is required. The simulation reports the results of a hypothetical change to the tax-transfer system, which involves moving from the current system to a system with universal benefits and a constant marginal tax rate (more details of this simulation are provided in the section on the simulation example in section 5. The example shown in the figure is the gini coefficient, where the coefficient is presented by a broad measure of family type.

The final option available in the set of inequality measures, "Redistribution", if selected provides an array of options relating to the redistributive ability of the tax-transfer system, along with various measures of progressivity of the system. The options available are presented in figure 8.

Figure 9 shows an example screen of the Reynolds-Smolensky redistribution measure by broad family type.

Figure 10 presents the options available when “poverty” is selected. The

MITTS

Analysis of Reform M I T T S

MITTS Version 4.00 March 2003

Gini Coefficients by UNITTYPE
IU income per adult equivalent

Group	Before	After	change
COUPLE	0.3103	0.2758	-0.0345
CPL&DEP	0.2460	0.1963	-0.0497
SINFEM	0.2925	0.2616	-0.0309
SINMALE	0.3297	0.2825	-0.0472
SOLEPAR	0.2154	0.2067	-0.0087
all	0.3014	0.2654	-0.0361

Save: <Alt-S> Exit: <ESC>

Figure 7: Example of set of Gini coefficients by family type before and after reform

MITTS

Analysis of Reform M I T T S

INEQUAL SETTINGS
Gini coefficients

R-S redistribution
Concentration index
Tax concentration
Kakwani progressivity
A-P reranking index

Winners/Losers Income Changes In METRs & RRs

Figure 8: Range of measures of redistributive effect of tax-transfer system

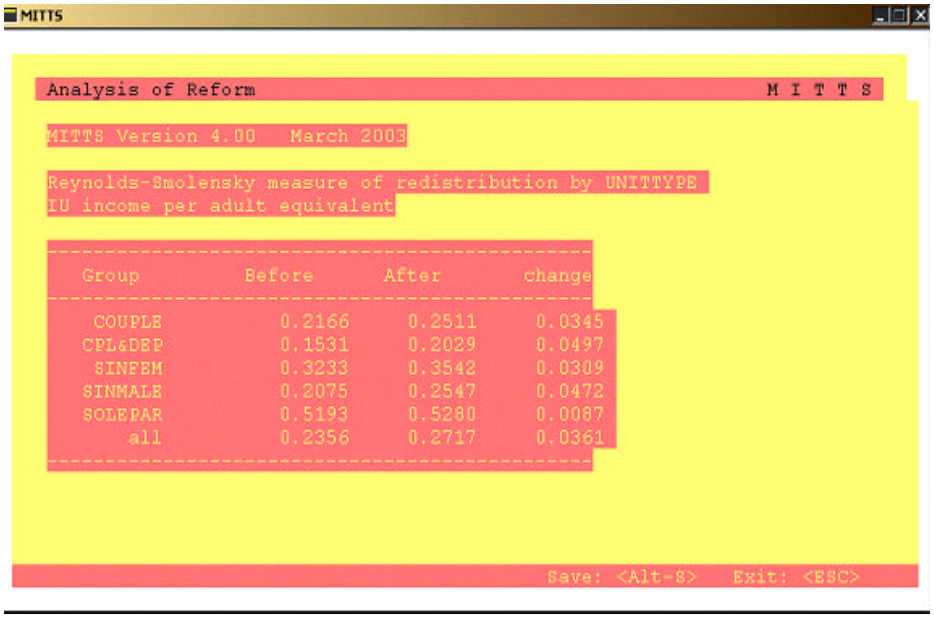


Figure 9: Example of Reynolds-Smolensky measure of redistribution by family type

'Poverty Settings', like the inequality settings, enable the user to select the measure of income to be used. The second option is to set the poverty line. If this option is selected a further menu providing a range of relative poverty lines appears as is shown in figure 11. The choices here are between 40%, 50% or 60% of either mean or median income. The default setting is 50% of median income, as this seems to be one of the more commonly used poverty lines in the literature. Note that the poverty line is calculated based on the particular measure of net income selected in "POVERTY SETTINGS". Five various poverty indices are then available (Headcount, Poverty Gap, Sen's indices, Foster, Greer and Thorbecke Indices, and the Johnson-Dixon indices). The final option calculates a TIP curve.

An example of an output screen for a headcount measure of poverty is presented in figure 12. Again, the hypothetical reform was a move a basic income - flat tax system. The estimates reflect the proportion of the population (with the outcome weighted to reflect population estimates) with income below the poverty line. The measure of income used, the unit of analysis and

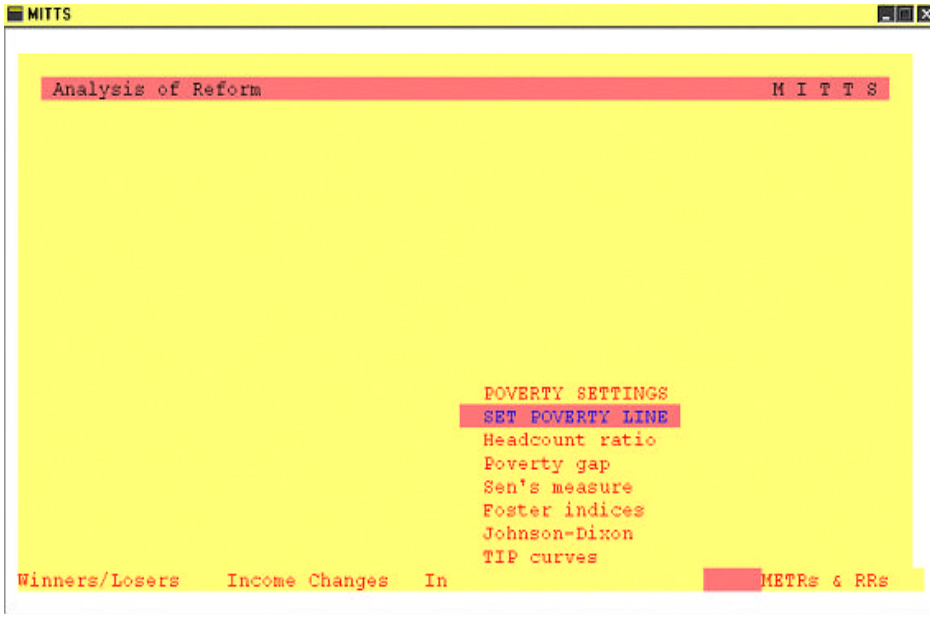


Figure 10: Range of poverty measures available

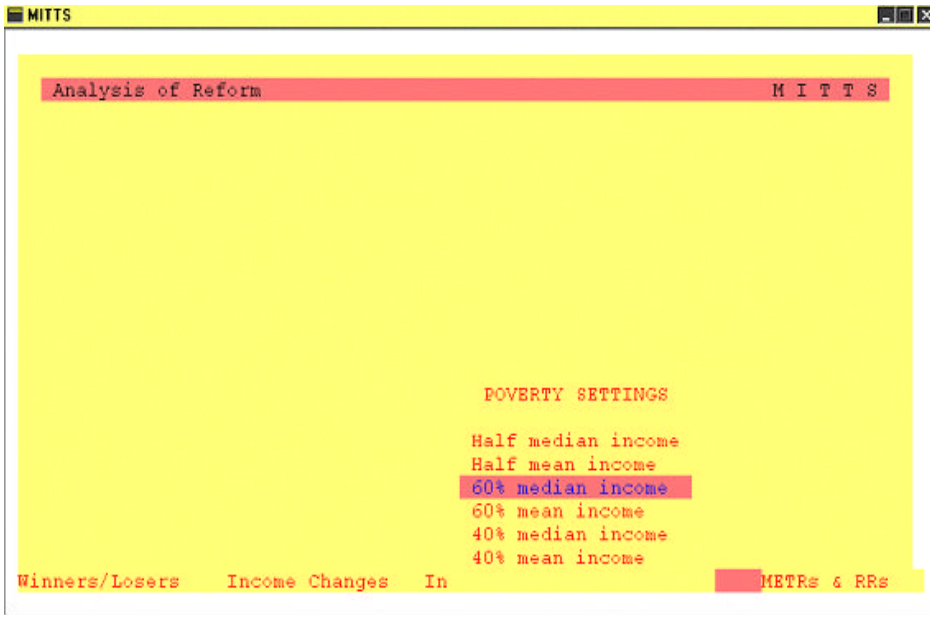


Figure 11: Relative poverty line options

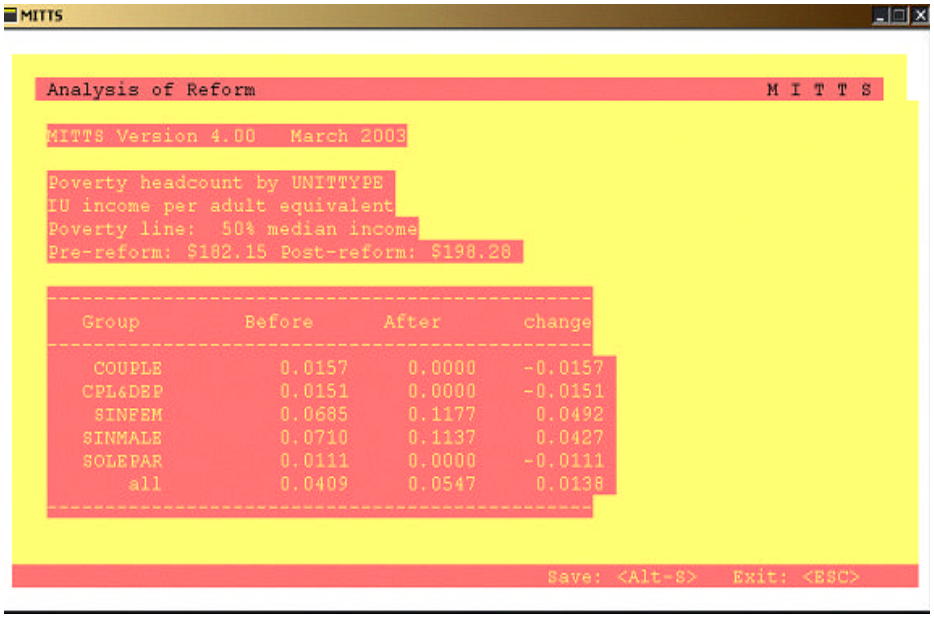


Figure 12: Example of poverty headcount by family type

the poverty line used is shown in the header with the actual estimated level of the poverty line pre and post-reform presented.

4.2 Mitts-B

After a simulation run in Mitts-B there are various options to generate distributional measures pre and post reform. The first step in generating these is to choose the option "Analyse results" in the array of options made available in Mitts-B as shown in figure 13 below.

Once "Analyse results" is selected a range of possible items are made available at the bottom of the screen: Summary, Transitions, Participation, Hours, and Ineq/Poverty (see figure 14). Implications of the reforms on the distribution of income, incorporating changes in labour supply, are found in the final option, 'Ineq/Poverty'. For more information on the method used see figure 3. Once selected, two options appear, Inequality and Poverty. The measures available are as those in Mitts-A with two exceptions, the

Figure 13: Mitts-B menu items

measures of redistribution are not available, being replaced by measures of the distribution of employment as the final category in the array of Inequality measures. The options available in the inequality and poverty settings also differ slightly in that the unit of analysis is constrained to that of the income unit, that is three options are available being i) income unit income per adult equivalent ii) income unit income per adult equivalent adjusted for household size and iii) income unit income shared equally between members of a household adjusted for household size.

An example of the output provided when selecting Gini coefficients and a headcount of poverty by family type in Mitts-B is shown in figures 16 and 17 respectively.

5 Simulation example

To illustrate the use of the pseudo method in generating summary measures for the distribution of income in behavioural microsimulation and some of the new poverty measures included in Mitts-A, a large-scale hypothetical

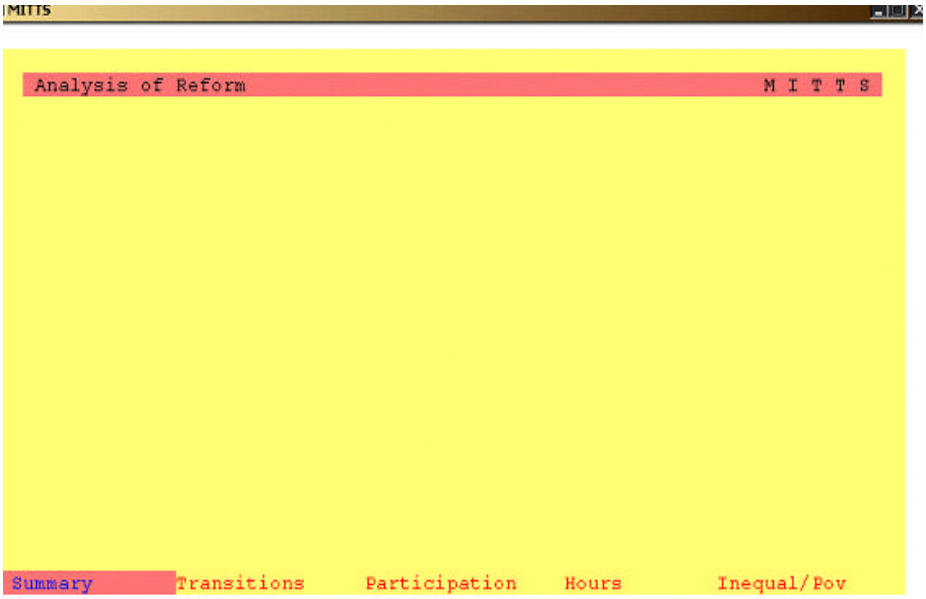


Figure 14: Analyse results options in Mitts-B

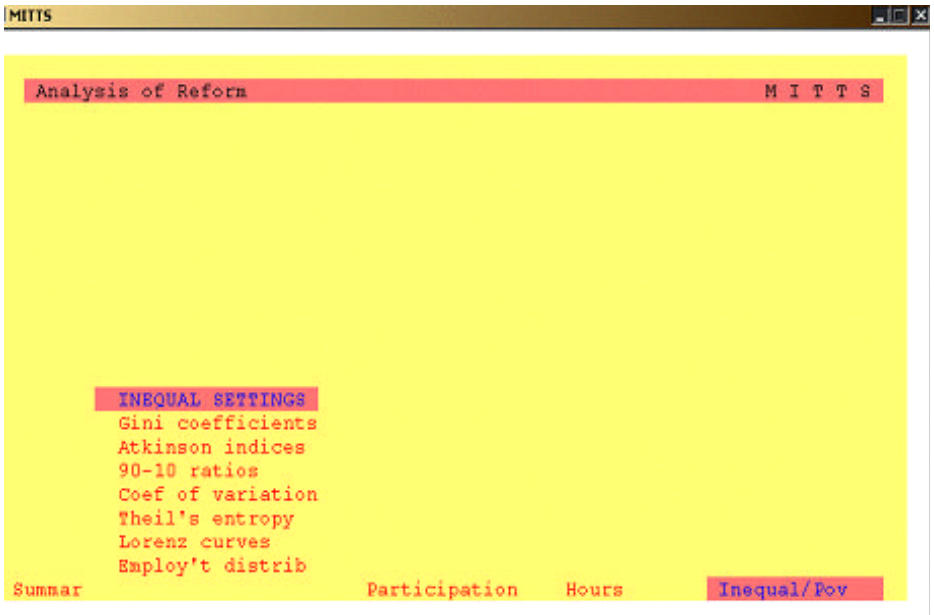


Figure 15: Inequality measures in Mitts-B

MITTS

Analysis of Reform M I T T S

MITTS Version 4.00 March 2003

Gini Coefficients by UNITTYPE
Income unit income using Mitts-B pseudo distribution

Group	Before	After	change
COUPLE	0.2984	0.2639	-0.0345
CPL&DEP	0.2338	0.1847	-0.0491
SINFEM	0.3036	0.2663	-0.0373
SINMALE	0.3545	0.3010	-0.0535
SOLEPAR	0.1982	0.1931	-0.0051
all	0.3040	0.2659	-0.0381

Save: <Alt-S> Exit: <ESC>

Figure 16: Gini coefficients by family type in Mitts-B

MITTS

Analysis of Reform M I T T S

MITTS Version 4.00 March 2003

Poverty headcount by UNITTYPE
Income unit income using Mitts-B pseudo distribution
Poverty line: 50% median income
Pre-reform: $175.60 Post-reform: $220.54

Group	Before	After	change
COUPLE	0.0175	0.0561	0.0386
CPL&DEP	0.0075	0.0000	-0.0075
SINFEM	0.0852	0.4121	0.3269
SINMALE	0.1074	0.3006	0.1933
SOLEPAR	0.0000	0.0285	0.0285
all	0.0523	0.1821	0.1299

Save: <Alt-S> Exit: <ESC>

Figure 17: Poverty headcount by family type in Mitts-B

reform of the Australian tax and transfer system is examined. The Australian transfer system is very complex, with many different types of benefit, each with its set of thresholds and taper (or withdrawal) rates, giving rise to highly nonlinear budget constraints. The simulations described in this section replace all existing basic social security benefits and additional payments such as rent assistance, pharmaceutical allowance and family payments by a basic non-taxable level of income. The existing direct tax structure (which includes the Medicare levy and all tax rebates) is replaced with a constant marginal tax rate on all taxable income (that is, all non-benefit forms of income). The reform is simulated using the Melbourne Institute Tax and Transfer Simulator (MITTS). The method of simulating labour supply responses to tax reforms in MITTS is briefly described in subsection 5.1. The results of the simulation are described in subsection 5.2.

5.1 The Melbourne Institute Tax and Transfer Simulator

The behavioural component of the Melbourne Institute Tax and Transfer Simulator takes into account labour supply changes resulting from policy changes to taxes and transfers. The behavioural responses are based on the use of quadratic preference functions whereby the parameters are allowed to vary with an individual's characteristics. These parameters have been estimated for five demographic groups, which include married or partnered men and women, single men and women, and sole parents. The framework is one in which individuals are considered to select from a discrete set of hours levels, rather than being able to vary labour supply continuously.

For the couples in the labour supply estimation, two sets of discrete labour supply points are used. Given that the female hours distribution covers a wider range of part-time and full-time hours than the male distribution, which is mostly divided between non-participation and full-time work, women's labour supply is divided into 11 discrete points, whereas men's labour supply is represented by just six points. The couple's joint labour supply is estimated simultaneously, unlike the popular approach in which

female labour supply is estimated with the spouse's labour supply taken as exogenous.[13]

The simulation is essentially probabilistic. That is, rather than identifying a particular level of hours worked for each individual after a policy change, a probability distribution is generated over the discrete hours levels used.[14] The behavioural simulations begin by taking the discrete hours level for each individual that is closest to the observed hours level. Then, given the parameter estimates of the quadratic preference function (which vary according to a range of characteristics), a random draw is taken from the distribution of the 'error' term. This draw is rejected if it results in an optimal hours level that differs from the discretised value observed. The accepted drawings are then used in the determination of the optimal hours level after the policy change. A user-specified total number of 'successful draws' (that is, drawings which generate the observed hours as the optimal value under the base system for the individual) are produced. This gives rise to a probability distribution over the set of discrete hours for each individual under the new tax and transfer structure. In computing the transition matrices, which show probabilities of movement between hours levels, the labour supply of each individual before the policy change is fixed at the discretised value and a number of transitions are produced for each individual, which is equal to the number of successful draws specified.[15]

[13] For those individuals in the data set who are not working, and who therefore do not report a wage rate, an imputed wage is obtained. This imputed wage is based on estimated wage functions, which allow for possible selectivity bias, by first estimating probit equations for labour market participation. However, some individuals are excluded from the database if their imputed wage or their observed wage (obtained by dividing total earnings by the number of hours worked) is unrealistic. The wage functions are reported in Kalb and Scutella (2002) and the preference functions are in Kalb (2002): these are updated versions of results reported in Creedy *et al.* (2002).

[14] Some individuals, such as the self employed, the disabled, students and those over 65 have their labour supply fixed at their observed hours.

[15] When examining average hours, the labour supply after the change for each individual is based on the average value over the successful draws, for which the error term leads to the correct predicted hours before the change. This is equivalent to calculating the expected hours of labour supply after the change, conditional on starting from the observed hours before the change. In computing the tax and revenue levels, an expected value is also obtained after the policy change. That is, the tax and revenue for each of the accepted draws are computed for each individual and an average over these is taken.

In some cases, the required number of successful random draws producing observed hours as the optimal hours cannot be generated from the model within a reasonable number of total drawings. The number of random draws tried, like the number of successful draws required, is specified by the user. If after the total draws from the error term distribution, the model fails to predict the observed labour supply, the individual is left at their observed hours in policy simulations. In the simulation example used here, the maximum number of random draws is set to 5000 with 100 successful draws required.

5.2 Simulation Results

In the simulations examined in this section, a basic non-taxable level of income replaces all existing basic social security benefits, and additional payments such as rent assistance, pharmaceutical allowance and family payments. The structure of the September 2001 tax and transfer system is used as the pre-reform system. In the reform system, the existing tax structure (which includes the Medicare levy and all tax rebates) is replaced with a constant marginal tax rate on all taxable income (all non-benefit forms of income). Basic income rates are set at the benefit levels as they were at September 2001 with basic income levels varying for different groups in the population reflecting the current situation. Characteristics that currently entitle individuals to a pension are used to determine whether an individual is entitled to a higher rate of basic income, which is referred to as the pension rate. This group includes those of age pension age, those with a disability, carers, veterans and sole parents. This payment is then differentiated by marital status to reflect the economies of scale of larger households. The remaining subset of the population receives a basic income level set at the current allowance payment rates. These payments also differ by age, such that single persons aged 16 or 17 years receive a lower basic income than older individuals, with youths still living at home receiving a lower rate again and those aged 60 years or over receive a higher level than 18 to 59 year olds. Members of a couple are entitled to a lower payment rate each than the single rate, reflecting economies of scale. To ensure revenue neutrality of the

new system compared to 2001 system, the marginal tax rate required in this system was 55 per cent.

The removal of means testing reduces the disincentives to work for those on low incomes as it reduces the effective marginal tax rates significantly. Lowering the marginal tax rate has two opposing effects. A lower tax rate would induce substitution out of leisure and into work, as the price of leisure is higher. However, at the same time net incomes are higher at lower levels of labour supply, so some individuals/families may decide to reduce their hours of work. The net result depends on relative preferences for leisure and income. Higher taxes imposed at the high end of the income distribution may have an adverse effect on the labour supply of higher income earners.

A summary of the estimated labour supply responses across demographic groups is presented in Table 9 with the expected effect on net expenditure reported in Table 10. The responses differ greatly across the groups. Singles without children exhibit labour supply responses similar to those of married men. However, single persons seem more likely to decrease participation and to reduce their work effort as a result of a heavier tax burden than married men. Married males are the least likely to decrease their labour supply overall. Married women and sole parents tend to have larger income effects thus having a greater tendency to move out of the labour force with an increase in household income. In addition, the current payments to sole parents are withdrawn very gradually and thus a tax rate of over fifty per cent is higher than the effective marginal tax rate currently faced by them at low income levels, reducing their incentive to work after the reform compared with the 2001 system.

The adverse labour supply responses associated with a relatively high marginal tax rate imposes a large increase in net government expenditure. A large share of each individual's income is paid in tax, so a reduction in labour supply considerably reduces the revenue collected by the government. The large reduction in the labour supply of married females increases net government expenditure on couples. The higher tax rates on middle to high income earners are not enough to offset the decrease in revenue resulting from the decrease in participation and average working hours for this group. With a

Table 9: Behavioural Responses

	Couples:		Single	Single	Single
	Men	Women	Men	Women	Parents
Workers (% base)	58.6	45.7	55.1	44.2	42.4
Workers (% reform)	58.5	41.2	53.6	43.3	37.2
Non-work –> work (%)	1.2	0.4	0.3	0.5	0.4
Work –> non-work (%)	1.3	4.8	1.8	1.4	5.6
Workers working more	1.3	0.3	0.0	0.1	2.0
Workers working less	2.4	1.9	1.4	2.3	1.1
Average hours change	-0.2	-1.9	-0.9	-0.8	-0.6

Table 10: Net Expenditure Summary (in millions of dollars)

	Couples	Single Men	Single Women	Single Parents	Total
Fixed hours	-4,024.5	1,735.8	2,042.8	-282.5	-528.4
Variable hours	2,044.3	2,859.5	2,895.4	-172.0	7,627.2

general reduction in labour supply across all groups, net government expenditure increases after the reform when labour supply responses are taken into account.

To examine the effects of this policy change on the distribution of income before and after labour supply changes, three summary measures are presented. First, two commonly used measures of inequality, the Gini coefficient and the Atkinson measure of inequality, are examined. As the reform system provides relatively generous basic income levels financed by heavily taxing the working population, it is not surprising that inequality is reduced quite significantly after the reform; see Tables 11 and 12. Adjusting for labour supply responses does not have a strong effect on the presented measures of inequality in this case, since those who move out of the labour force were on low wages.

For the analysis of poverty, values of the Foster, Greer and Thorbecke Index of poverty are presented in Table 13. The inclusion of labour supply responses in the simulation clearly affects the headcount measure, resulting from the larger shift in median income. Taking into account labour supply

Table 11: Gini Coefficient by Income Unit Type

	Fixed hours			Variable hours		
Group	Before	After	% change	Before	After	% change
Couple	0.3103	0.2758	-11.1	0.3030	0.2701	-10.9
Cple+dep	0.2460	0.1963	-20.2	0.2333	0.1857	-20.4
Single Females	0.2925	0.2616	-10.6	0.2950	0.2609	-11.6
Single Males	0.3297	0.2825	-14.3	0.3343	0.2839	-15.1
Sole Parents	0.2154	0.2067	-4.0	0.2101	0.2029	-3.4
all	0.3014	0.2654	-12.0	0.2980	0.2616	-12.2

responses results in more poverty than would otherwise be the case. This effect becomes less pronounced when the extent of the poverty is taken into account in the poverty gap and the Foster, Greer and Thorbecke measure of poverty with $\alpha = 2$.

6 Conclusions

In this report we have outlined approaches to the measurement of summary measures of the distribution of income in the context of behavioural microsimulation with specific reference to the Melbourne Institute Tax and Transfer Simulator (Mitts). Part of this has involved updating the range of measures in Mitts-A, the non-behavioural component of Mitts, and allowing all of the measures offered to be weighted. The other major part of this report involves comparing alternative approaches to the measurement of inequality and poverty indices in the context of behavioural microsimulation with discrete hours labour supply models. Special consideration is needed because microsimulation modelling using a discrete hours approach is probabilistic; that is, it does not identify a particular level of hours worked for each individual after a policy change, but generates a probability distribution over the discrete hours levels used. This makes analysis of the distribution of income difficult because, even for a small sample with a modest range of hours points, the range of possible labour supply combinations becomes too large to handle.

The approaches examined include the use of an expected income level for

Table 12: Atkinson Inequality Measures by Income Units

	Fixed hours			Variable hours		
Group	Before	After	% change	Before	After	% change
$\varepsilon = 0.1$						
Couple	0.0155	0.0123	-20.6	0.0147	0.0117	-20.4
Cple+dep	0.0106	0.007	-33.0	0.0095	0.0062	-34.7
Single Females	0.0141	0.0112	-20.6	0.0145	0.0112	-22.1
Single Males	0.0182	0.0136	-25.8	0.0188	0.0137	-27.1
Sole Parents	0.0083	0.0077	-7.2	0.008	0.0075	-7.5
all	0.0151	0.0117	-22.5	0.0148	0.0113	-23.0
$\varepsilon = 0.5$						
Couple	0.0749	0.0597	-20.3	0.0711	0.0568	-20.1
Cple+dep	0.0498	0.0332	-33.3	0.0448	0.0294	-34.4
Single Females	0.0680	0.0540	-20.6	0.0697	0.0539	-22.7
Single Males	0.0879	0.0646	-26.5	0.0906	0.0652	-28.0
Sole Parents	0.0387	0.0360	-7.0	0.0373	0.0348	-6.4
all	0.0724	0.0564	-22.1	0.0710	0.0546	-23.0
$\varepsilon = 1$						
Couple	0.1429	0.1145	-19.8	0.1355	0.1091	-19.5
Cple+dep	0.0934	0.0627	-32.9	0.0843	0.0560	-33.6
Single Females	0.1300	0.1028	-20.9	0.1332	0.1020	-23.4
Single Males	0.1686	0.1222	-27.5	0.1740	0.1232	-29.3
Sole Parents	0.0711	0.0661	-7.0	0.0684	0.0641	-6.3
all	0.1380	0.1078	-21.9	0.1355	0.1046	-22.8
$\varepsilon = 2$						
Couple	0.2563	0.208	-18.9	0.2429	0.1981	-18.4
Cple+dep	0.1684	0.1151	-31.7	0.1524	0.1037	-32.0
Single Females	0.2383	0.1840	-22.7	0.2451	0.1814	-26.0
Single Males	0.3114	0.2191	-29.7	0.3221	0.2202	-31.6
Sole Parents	0.1218	0.1127	-7.5	0.1170	0.1098	-6.2
all	0.2527	0.1968	-22.1	0.2501	0.1914	-23.5

Table 13: Alternative Poverty Measures

	Fixed hours			Variable hours		
	Before	After	%	Before	After	%
Poverty line						
50% median income	$182.15	$198.28		$176.28	$221.20	
$\alpha = 0$ (*headcount*)						
Couple	0.0157	0.0000	-100.0	0.0150	0.0490	227.3
Cple+dep	0.0151	0.0000	-100.0	0.0173	0.0027	-84.4
Single Females	0.0685	0.1177	71.8	0.0768	0.3980	418.1
Single Males	0.0710	0.1137	60.1	0.0784	0.2798	256.9
Sole Parents	0.0111	0.0000	-100.0	0.0035	0.0161	362.9
all	0.0409	0.0547	33.7	0.0444	0.1724	288.3
$\alpha = 1$ (*poverty gap*)						
Couple	0.0054	0.0000	-100.0	0.0053	0.0017	-67.9
Cple+dep	0.0031	0.0000	-100.0	0.0029	0.0001	-93.1
Single Females	0.0224	0.0157	-29.9	0.0258	0.0420	62.8
Single Males	0.0260	0.0103	-60.4	0.0299	0.0316	5.7
Sole Parents	0.0005	0.0000	-100.0	0.0001	0.0002	100.0
all	0.0135	0.0061	-54.8	0.0151	0.0177	17.2
$\alpha = 2$						
Couple	0.0021	0.0000	-100.0	0.0020	0.0001	-100.0
Cple+dep	0.0007	0.0000	-100.0	0.0006	0.0000	-100.0
Single Females	0.0096	0.0034	-64.6	0.0115	0.0080	-29.6
Single Males	0.0121	0.0015	-87.6	0.0140	0.0054	-61.4
Sole Parents	0.0000	0.0000	0.0	0.0000	0.0000	0.0
all	0.0058	0.0012	-81.0	0.0067	0.0032	-52.2

each individual. Alternatively, a simulated approach could be used in which labour supply values are drawn from each individual's hours distribution and summary statistics of the distribution of income are calculated by taking the average over each set of draws. Finally, the construction of a pseudo income distribution was proposed. This uses the probability of a particular labour supply value occurring (standardised by the population size) to refer to a particular position in the pseudo income distribution.

It was shown that the outcomes of various distributional measures using the pseudo distribution method converge to their 'true' values as the sample size used in microsimulation increases. However, the method performs well for a sample as small as 50 individuals. The pseudo method performs better than the expected income method or the sampling approach with just 10 draws. The latter has a similar computational burden to the pseudo approach. The feasibility of the pseudo approach is demonstrated in an illustrative example where the labour supply responses and implications for the distribution of income were examined for a hypothetical reform of the Australian tax and transfer system.

7 Appendix: Inequality and Poverty Measures Using the Different Approaches

This appendix lists the expressions for the inequality and poverty measures used in this paper, for the alternative approaches. For notational reasons the weights are excluded below but it is straightforward to include them in the formulae. The measures in MITTS-B allow for weights just like in MITTS-A.

First, the *Atkinson measure* (for unit relative inequality aversion) is, for all combinations:

$$1-\sum_{m=1}^{k^n} p_{h_{1m},1}p_{h_{2m},2}\cdots p_{h_{nm},n}\frac{\exp\frac{1}{n}\sum_{i=1}^{n}\ln(y_{h_{im},i})}{\frac{1}{n}\sum_{i=1}^{n}y_{h_{im},i}}$$

The sampling approach is similar, where a sample is drawn from all possible combinations and the probability of drawing a particular combination m is equal to the probability $p_{h_{1m},1}p_{h_{2m},2}\cdots p_{h_{nm},n}$. Using expected incomes:

$$1-\frac{\exp\frac{1}{n}\sum_{i=1}^{n}\ln(\sum_{h=1}^{k}p_{hi}y_{hi})}{\frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k}p_{hi}y_{hi}}$$

Using the pseudo approach:

$$1-\frac{\exp\frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k}p_{hi}\ln(y_{hi})}{\frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k}p_{hi}y_{hi}}$$

Second, the *Gini Coefficient* for all combinations is expressed as:

$$\frac{n+1}{n-1}-2\sum_{m=1}^{k^n} p_{h_{1m},1}p_{h_{2m},2}\cdots p_{h_{nm},n}\frac{\sum_{i=1}^{n}\rho_{l_{im}}y_{h_{l_{im}}m l_{im}}}{n(n-1)\frac{1}{n}\sum_{i=1}^{n}y_{h_{im},i}}$$

with l_{im} indicating the index for the household with the ith ranked income in the mth combination and $\rho_{l_{im}}=\rho_{l_{i-1,m}}+1$. Using expected incomes it is:

$$\frac{n+1}{n-1}-2\frac{\sum_{i=1}^{n}\rho_{l_i^*}\sum_{h=1}^{k}p_{hl_i^*}y_{hl_i^*}}{n(n-1)\frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k}p_{hi}y_{hi}}$$

with l_i^* indicating the index for the household with the ith ranked expected income and $\rho_{l_i^*}=\rho_{l_{i-1}^*}+1$. Using the pseudo method the Gini measure is:

$$\frac{n+1}{n-1}-2\frac{\sum_{i=1}^{n}\sum_{h=1}^{k}\rho_{v_{h(i-1)+h},\widehat{l}_{h(i-1)+h}}p_{v_{h(i-1)+h},\widehat{l}_{h(i-1)+h}}y_{v_{h(i-1)+h},\widehat{l}_{h(i-1)+h}}}{n(n-1)\frac{1}{n}\sum_{i=1}^{n}\sum_{h=1}^{k}p_{hi}y_{hi}}$$

with $\widehat{l}_{h(i-1)+h}$ indicating the index for the household with the $(h(i-1)+h)^{th}$ ranked income within the nh possible incomes, $v_{h(i-1)+h}$ indicating the index for the discrete hours point with the $(h(i-1)+h)^{th}$ ranked income and $\rho_{v_{h(i-1)+h},\widehat{l}_{h(i-1)+h}} = \rho_{v_{h(i-1)+h-1},\widehat{l}_{h(i-1)+h-1}} + p_{v_{h(i-1)+h},\widehat{l}_{h(i-1)+h}}$.

Third, the *90/10 Ratio* for all combinations is:

$$\sum_{m=1}^{k^n} p_{h_{1m},1} p_{h_{2m},2} \cdots p_{h_{nm},n} \frac{y_{h_{l_{im}m} l_{im}}}{y_{h_{l_{jm}m} l_{jm}}}$$

with l_{im} indicating the index for the household with the ith ranked income in the mth combination, where $i/n = 0.9$ and l_{jm} indicating the index for the household with the jth ranked income in the mth combination, where $j/n = 0.1$. Using expected incomes it is:

$$\frac{\sum_{h=1}^{k} p_{hl_i^*} y_{hl_i^*}}{\sum_{h=1}^{k} p_{hl_j^*} y_{hl_j^*}}$$

with l_i^* indicating the index for the household with the ith ranked expected income, where $i/n = 0.9$ and l_j^* indicating the index for the household with the jth ranked expected income, where $j/n = 0.1$. Using the pseudo method it is:

$$\frac{y_{v_{h(i-1)+h},\widehat{l}_{h(i-1)+h}}}{y_{v_{h(j-1)+h},\widehat{l}_{h(j-1)+h}}}$$

with $\widehat{l}_{h(i-1)+h}$ indicating the index for the household with the $(h(i-1)+h)^{th}$ ranked income within the nh possible incomes, $v_{h(i-1)+h}$ indicating the index for the discrete hours point with the $(h(i-1)+h)^{th}$ ranked income, where $h(i-1)+h$ indicates the income where $\sum_{k=1}^{h(i-1)+h} p_{v_k,\widehat{l}_k} = 0.9$, and $\widehat{l}_{h(j-1)+h}$ indicating the index for the household with the $(h(j-1)+h)^{th}$ ranked income within the nh possible incomes, $v_{h(j-1)+h}$ indicating the index for the discrete hours point with the $(h(j-1)+h)^{th}$ ranked income, where $h(j-1)+h$ indicates the income where $\sum_{k=1}^{h(j-1)+h} p_{v_k,\widehat{l}_k} = 0.1$.

Fourth, the *headcount measure* for all combinations is:

$$\sum_{m=1}^{k^n} p_{h_{1m},1} p_{h_{2m},2} \cdots p_{h_{nm},n} \frac{1}{n} \sum_{i=1}^{n} 1(y_{h_{im},i} \leq y_{pov,m})$$

where $y_{pov,m} = 0.5 \times y_{h_{l_{im}m} l_{im}}$ and l_{im} indicates the index for the household with the ith ranked income in the mth combination, where $i/n = 0.5$. Using expected income it is:

$$\sum_{i=1}^{n} 1(\sum_{h=1}^{k} p_{hi} y_{hi} \leq y_{pov})$$

where $y_{pov} = 0.5 \times \sum_{h=1}^{k} p_{hl_i^*} y_{hl_i^*}$ with l_i^* indicating the index for the household with the ith ranked expected income, where $i/n = 0.5$. Using the pseudo method it is:

$$\sum_{i=1}^{n} \sum_{h=1}^{k} p_{hi} 1(y_{hi} \leq y_{pov})$$

where $y_{pov} = 0.5 \times y_{v_{h(i-1)+h}, \widehat{l}_{h(i-1)+h}}$ with $\widehat{l}_{h(i-1)+h}$ indicating the index for the household with the $(h(i-1)+h)^{th}$ ranked income within the nh possible incomes, $v_{h(i-1)+h}$ indicating the index for the discrete hours point with the $(h(i-1)+h)^{th}$ ranked income, where $h(i-1)+h$ indicates the income where $\sum_{k=1}^{h(i-1)+h} p_{v_k, \widehat{l}_k} = 0.5$.

Fifth, the *Poverty Gap measure* for all combinations is:

$$\sum_{m=1}^{k^n} p_{h_{1m},1} p_{h_{2m},2} \cdots p_{h_{nm},n} \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{y_{h_{im},i}}{y_{pov,m}}\right) 1(y_{h_{im},i} \leq y_{pov,m})$$

where $y_{pov,m} = 0.5 \times y_{h_{l_{im}m} l_{im}}$ with l_{im} indicating the index for the household with the ith ranked income in the mth combination, where $i/n = 0.5$. Using expected income it is:

$$\sum_{i=1}^{n} \left(1 - \frac{\sum_{h=1}^{k} p_{hi} y_{hi}}{y_{pov}}\right) 1(\sum_{h=1}^{k} p_{hi} y_{hi} \leq y_{pov})$$

where $y_{pov} = 0.5 \times \sum_{h=1}^{k} p_{hl_i^*} y_{hl_i^*}$ with l_i^* indicating the index for the household with the ith ranked expected income, where $i/n = 0.5$. Using the pseudo method it is:

$$\sum_{i=1}^{n} \sum_{h=1}^{k} p_{hi} \left(1 - \frac{y_{hi}}{y_{pov}}\right) 1(y_{hi} \leq y_{pov})$$

where $y_{pov} = 0.5 \times y_{v_{h(i-1)+h}, \widehat{l}_{h(i-1)+h}}$ with $\widehat{l}_{h(i-1)+h}$ indicating the index for the household with the $(h(i-1)+h)^{th}$ ranked income within the nh possible

incomes, $v_{h(i-1)+h}$ indicating the index for the discrete hours point with the $(h(i-1)+h)^{th}$ ranked income, where $h(i-1)+h$ indicates the income where $\sum_{k=1}^{h(i-1)+h} p_{v_k,\widehat{l}_k} = 0.5$.

Sixth, the *Foster, Greer and Thorbecke measure* with $\alpha = 2$, can be expressed, for all combinations, as:

$$\sum_{m=1}^{k^n} p_{h_{1m},1} p_{h_{2m},2} \cdots p_{h_{nm},n} \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{y_{h_{im},i}}{y_{pov,m}}\right)^{\alpha} 1(y_{h_{im},i} \leq y_{pov,m})$$

where $y_{pov,m} = 0.5 \times y_{h_{l_{im}m} l_{im}}$ with l_{im} indicating the index for the household with the ith ranked income in the mth combination, where $i/n = 0.5$. Using expected incomes it is:

$$\sum_{i=1}^{n} \left(1 - \frac{\sum_{h=1}^{k} p_{hi} y_{hi}}{y_{pov}}\right)^{\alpha} 1(\sum_{h=1}^{k} p_{hi} y_{hi} \leq y_{pov})$$

where $y_{pov} = 0.5 \times \sum_{h=1}^{k} p_{hl_i^*} y_{hl_i^*}$ with l_i^* indicating the index for the household with the ith ranked expected income, where $i/n = 0.5$. Using the pseudo method it is:

$$\sum_{i=1}^{n} \sum_{h=1}^{k} p_{hi} \left(1 - \frac{y_{hi}}{y_{pov}}\right)^{\alpha} 1(y_{hi} \leq y_{pov})$$

where $y_{pov} = 0.5 \times y_{v_{h(i-1)+h}, \widehat{l}_{h(i-1)+h}}$ with $\widehat{l}_{h(i-1)+h}$ indicating the index for the household with the $(h(i-1)+h)^{th}$ ranked income within the nh possible incomes, $v_{h(i-1)+h}$ indicating the index for the discrete hours point with the $(h(i-1)+h)^{th}$ ranked income, where $h(i-1)+h$ indicates the income where $\sum_{k=1}^{h(i-1)+h} p_{v_k,\widehat{l}_k} = 0.5$.

References

[1] Creedy, J. and Duncan, A. S. (2001) Behavioural microsimulation with labour supply responses. *Journal of Economic Surveys*, 16, pp. 1-39.

[2] Creedy, J., Duncan, A.S., Scutella, R. and Harris, M. (2002) *Microsimulation Modelling of Taxation and The Labour Market: The Melbourne Institute Tax and Transfer Simulator*. Cheltenham: Edward Elgar.

[3] Foster, J., Greer, J. and Thorbecke, E. (1984) A class of decomposable poverty measures. *Econometrica*, 52, pp. 761-762.

[4] Gerfin, M. and Leu, R.E. (2003) The impact of in-work benefits on poverty and household labour supply: a simulation study for Switzerland. Discussion Paper No. 762, IZA, Bonn.

[5] Kalb, G. (2002) Estimation of labour supply models for four separate groups in the Australian population. Melbourne Institute Working Paper No. 24/02, The University of Melbourne.

[6] Kalb, G. and Scutella. R. (2002) Estimation of wage equations in Australia: allowing for censored observations of labour supply. Melbourne Institute Working Paper No. 8/02, The University of Melbourne.

[7] Keane, M., and Moffitt, R. (1998) A structural model of multiple welfare program participation and labour supply. *International Economic Review,* 39, pp. 553-590.

[8] Van Soest, A. (1995) Discrete choice models of family labour supply. *Journal of Human Resources,* 30, pp. 63-88.