

Confidence Intervals for Policy Reforms in Behavioural Tax Microsimulation Modelling*

John Creedy[‡], Guyonne Kalb^{} and Hsein Kew[‡]**

[‡]Department of Economics

The University of Melbourne

and

^{}Melbourne Institute of Applied Economic and Social Research**

The University of Melbourne

Melbourne Institute Working Paper No. 32/04

ISSN 1328-4991 (Print)

ISSN 1447-5863 (Online)

ISBN 0 7340 3174 2

December 2004

*We are grateful to the Department of Family and Community Services, who funded an earlier research project, from which the extension in this paper followed. 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.

Melbourne Institute of Applied Economic and Social Research

The University of Melbourne

Victoria 3010 Australia

***Telephone* (03) 8344 2100**

***Fax* (03) 8344 2111**

***Email* melb-inst@unimelb.edu.au**

***WWW Address* <http://www.melbourneinstitute.com>**

Abstract

This paper addresses the need for a measure of the uncertainty that is associated with the results calculated through tax policy behavioural microsimulation modelling. Deriving the analytical measure would be extremely complicated. Therefore, a simulated approach is proposed which generates a pseudo sampling distribution of aggregate measures based on the sampling distribution of the estimated labour supply parameters. This approach, which is very computer intensive, is compared to a more time-efficient approach where the functional form of the sampling distribution is assumed to be normal. The results show that in many instances the results from the two approaches are quite similar. The exception is when aggregate measures for minor types of payments, involving relatively small groups of the population, are examined.

1 Introduction

Behavioural tax microsimulation models, which are based on large scale cross-sectional surveys and use estimates of preference functions allowing for many individual characteristics, can provide valuable policy advice about the likely labour supply, revenue and expenditure implications of a wide range of policy reforms. However, these implications are usually in the form of point estimates and microsimulation models typically provide no information about the sampling distribution of labour supply or expenditure changes. Such information would be of much interest to those involved in designing policy reforms with specific objectives and constraints in mind.

Uncertainty regarding a model's projections can arise for a variety of reasons.¹ These include, for example, the fact that estimates of wage rates are used (this is particularly true for non-workers in the dataset where wages are obtained from estimated wage functions) and preference parameters are estimated. In addition, sampling variations arise from the fact that the database is a sample of the population and household weights must be used for aggregation purposes. This is particularly problematic for groups where only a small number of observations are available.² Errors can of course also arise from incorrect modelling of the complex administrative rules governing taxes and benefits, reporting errors and problems relating to the (possibly endogenous) incomplete take-up of benefits. The present paper focusses on just one of these sources of error. Its aim is to investigate methods of obtaining confidence intervals where the appropriate distribution of values arises from the sampling distribution of parameter estimates of preference functions. Maximum likelihood estimation gives rise to a variance/covariance matrix for the estimated parameters, which forms the basis of the approaches discussed here. Analytical derivation of the sampling distribution of the various microsimulation outcomes is ruled out in view of the complexity of the transformations

¹For a general discussion, see Klevmarken (1998).

²This source of error was examined by Pudney and Sutherland (1994). They concentrated on the use of weights to deal with differential responses by different population strata, assuming that response rates are known. Sample weights are now often based on calibration methods, following Deville and Särndal (1992), where extraneous information about selected population totals is used.

involved.³ The decision to concentrate on uncertainties associated with preference functions does not arise from any judgement that the other sources are not important. Rather, those sources exist with both arithmetic and behavioural modelling, and interest here is on the additional uncertainty which results from the estimation requirements of behavioural models.

The approach is discussed in the context of the Melbourne Institute Tax and Transfer Simulator (MITTS). This is a behavioural microsimulation model for Australia. It uses as a database the Survey of Income and Housing Costs (SIHC), made available by the Australian Bureau of Statistics in confidentialised unit record file (CURF) form. The labour supply parameters included in MITTS are based on estimates using the SIHC following a similar approach to Van Soest (1995). Labour supply is modelled using a discrete hours approach where individuals select from a number of discrete hours points, and estimated preference functions allow for considerable population heterogeneity.⁴ The discrete hours approach is described briefly in Section 2. A basic simulation approach to the construction of confidence intervals is described in Section 3. This involves the numerical production of what may be called a pseudo sampling distribution of the variable of interest. Numerical results for a simple policy reform, using this approach in MITTS, are presented in Section 4. This reform involves a flattening of budget constraints resulting from a reduction in benefit abatement, or taper, rates. The basic simulation method is highly computer intensive and can take several weeks or even months, depending on the number of households in the sample and the speed of the computer used. A much more time-efficient approximate method is suggested in Section 5. Conclusions are in Section 6.

³However, Pudney and Sutherland (1996) derived explicit results for the sampling variance of variables, such as total government expenditure for example, which are obtained by summing values over individuals in the sample. They estimated a three-state multinomial logit model for females, but did not consider hours variations (so that only earnings enter utility functions). Understandably, given the computation burden at the time, they rejected any kind of simulation approach.

⁴For details of the MITTS model, see Creedy *et al.* (2002).

2 Discrete Hours Labour Supply

This section describes the ingredients of the behavioural component of a tax microsimulation model when using an approach to labour supply modelling in which individuals are constrained to choose from a number of discrete hours levels. It is assumed here that the required parameters of the direct utility functions involved have been estimated and are now available for simulation.⁵ First, Subsection 2.1 describes the specification of utility functions. A microsimulation model must be able to capture a great deal of the heterogeneity in preferences usually found in a population of households. This subsection shows how personal characteristics are included in the direct utility functions. Utility maximisation is outlined in Subsection 2.2, where it is shown how a probability distribution over the available hours levels is generated for each individual, instead of a deterministic level of hours when a discrete hours labour supply model is used. A tax policy reform has the effect of changing individuals' net incomes corresponding to each discrete hours point.⁶ Subsection 2.3 describes the numerical procedure used to generate the distribution of hours worked following a tax policy reform, conditional on each individual being at the observed position before the reform.

2.1 Personal Characteristics' Effects on Utility

The MITTS model uses quadratic direct utility functions defined in terms of hours worked, h , and net income, y to evaluate combinations of net income and leisure/home production time. This can be expressed for individuals as:⁷

$$U(h, y) = \beta_y y + \beta_h h + \alpha_y y^2 + \alpha_h h^2 + \alpha_{yh} y h \quad (1)$$

For couples, this quadratic expression can be extended as follows:

$$\begin{aligned} U(h_m, h_f, y) = & \beta_y y + \beta_m h_m + \beta_f h_f + \alpha_y y^2 + \alpha_m h_m^2 \\ & + \alpha_f h_f^2 + \alpha_{ym} y h_m + \alpha_{yf} y h_f + \alpha_{fm} h_f h_m \end{aligned} \quad (2)$$

⁵For an introductory survey on the specification, estimation and microsimulation with discrete hours models, see Creedy and Kalb (2004).

⁶Some reforms modify eligibility conditions, and the enforcement of rules, in ways which cannot be modelled using current methods, but which may affect the take-up of benefits.

⁷An early example of its use can be found in Keane and Moffitt (1998).

where the subscript m denotes hours and parameters of the male, f denotes corresponding values of the female, and y represents net household income. The parameter α_{fm} indicates whether the male and female labour supplies are complements or substitutes.

This approach can be extended to make the preference parameters dependent on personal and household characteristics, such as age of children or an individual's own age, which are likely to influence the preference for work and income. For convenience, consider the simple linear utility function:

$$U = \beta_h h + \beta_y y \quad (3)$$

After including characteristics in the preference for work parameter, the utility function becomes:

$$U = \{\beta_{h1} + \beta_{h2}age + \beta_{h3}d\} h + \beta_y y \quad (4)$$

where, say, $d = 1$ if the age of the youngest child is 0 to 4, and $d = 0$ otherwise. This approach can help to explain differences in behaviour between individuals with similar wages but different personal characteristics. The utility functions discussed here represent 'deterministic utility' levels but, as discussed in the next subsection, the model allows for an additional additive stochastic term. This term could for example account for optimisation errors by the household.

The MITTS model allows for a wide range of characteristics affecting the utility function, including age, education level, number of children and age of the youngest child in the household. In addition, for sole parents gender is included because this is the only group for which labour supply is not estimated separately for men and women. For details of the estimation method, see Creedy and Kalb (2004). The estimates used in the current version of MITTS are reported in Kalb (2002), where estimation is carried out for separate demographic groups; these are single males and females, sole parents, and couples.

2.2 Utility Maximisation

Consider an individual with measured characteristics, X , who maximises utility by selecting the number of hours worked, h , where only a discrete number of hours levels, h_i , ($i = 1, \dots, n$) are available. As explained in Section 2.1, utility is determined by leisure and net income, but here it may be written with hours of work as the sole argument, because conditional on wage rates, other income and other characteristics, the hours of work determine net income so there is a one-to-one relationship between h_i and the level of y for each household.

Utility associated with each hours level is denoted U_i^* and is a function of deterministic utility $U(h_i|X)$ plus a random error term, v_i , so that:

$$\begin{aligned} U_i^* &= U(h_i|X) + v_i \\ &= U_i + v_i \end{aligned} \tag{5}$$

The term v_i arises from factors such as measurement errors concerning the variables in X , optimisation errors of the individual or the existence of unobserved preference characteristics. There is therefore a probability distribution over the available discrete hours levels, which depends on the distribution of v_i . The v s are assumed to be independent and any observation on h is associated with a set of possible draws of the random variables v_1 to v_n from their respective distributions.

Hours level i is chosen if:

$$U_i^* \geq U_j^* \quad \text{for all } j \tag{6}$$

Substituting for U_i^* , using (5), and rearranging, this condition is equivalent to:

$$v_j \leq v_i + U_i - U_j \quad \text{for all } j \tag{7}$$

If $f(\cdot)$ and $F(\cdot)$ are the density and distribution functions respectively for all v_i and all v_i are independent of each other, the probability of hours level i is given by:

$$p_i = \int_{-\infty}^{+\infty} \left\{ \prod_{j \neq i} F(v_i + U_i - U_j) \right\} f(v_i) dv_i \tag{8}$$

This expression takes all possible conditional probabilities $P(U_i^* \geq U_j^* | v_i)$, represented by $\prod_{j \neq i} F(v_i + U_i - U_j)$, and integrates v_i out to obtain the required probability p_i .

Suppose the distribution of v is described by the Extreme (maximum) Value Type I distribution:

$$f(v) = e^{-v} e^{-e^{-v}} \quad (9)$$

for which the distribution function is:

$$F(v) = e^{-e^{-v}} \quad (10)$$

It can be shown that substitution into (8) gives:⁸

$$p_i = \frac{e^{U_i}}{\sum_{j=1}^n e^{U_j}} \quad (11)$$

One approach to simulation would be to use (11) to generate probability distributions of hours worked for each person before and after a tax reform. Predicted labour supply changes could then be based on the difference between expected (average) hours. The MITTS model does not use this explicit form for the distribution of hours of work. Instead, a numerical method of repeatedly drawing from the error distribution is adopted, which generates a distribution of post-reform hours worked, conditional on the individual being at the observed pre-reform hours reported in the dataset.⁹ This so-called calibration process is described in the following subsection.

2.3 Individual Calibration

Given estimated parameters of specified preference functions, and a set of net incomes for each individual (or couple) arising from every possible discrete hours level, the effects on labour supply of tax policy changes can be

⁸See Madalla (1983) and Creedy and Kalb (2004) for details of the derivation of this result.

⁹The former method would be much less intensive from a computational point of view, as draws from error distributions would not be needed. However, following this approach the post-reform results would not be conditional on the observed pre-reform hours, which could be a disadvantage when aiming for realistic policy simulations.

examined. The simulation method applied by MITTS is to use the labour supply observed in the data set to obtain a starting point for each individual. This ensures that the optimal labour supply of each individual corresponds precisely to the observed labour supply under the particular tax and benefit system at the time of the survey. This process, referred to as ‘calibration’, is described in the present subsection: a more formal description is given in Appendix A.

The first step is to discretise the observed hours, so that the individual’s hours are set equal to the nearest discrete level available. Suppose there are n hours levels. In the calibration process, a set of n independent error terms is drawn from the relevant distribution (in this case, the extreme value distribution) and these are added to the deterministic utility levels at each of the hours points. If this results in the utility at the observed labour supply being the optimal choice for the individual, the set of random values is stored. It is useful to distinguish between ‘draws’ and ‘tries’: the former describes a set of values which are stored for further use, while the latter describes any set of random values from the extreme value distribution. If a ‘try’ does not result in optimal hours that correspond to observed hours, another set of error terms is obtained and another check is made. If, after say K attempts, no set of random values has been found that places the individual at actual pre-reform hours, instead of storing a ‘draw’, the individual’s hours of work are assumed to be constant for this particular ‘draw’ when examining a policy reform.

This process is repeated a specified number of times, say k , each time using K tries if necessary. For most individuals, the draws are usually obtained with few tries. However, for some individuals, a proportion of attempts to produce a draw may not succeed after K tries. In a small number of cases, no successful draws may be obtained even after a total of K times k sets of n random values have been taken from the error distribution.

The sets of error terms that result in the observed labour supply, the k draws, are then used to compute a distribution of hours worked for each individual after a specified reform. Given the individual’s characteristics and draws from the error term, utility at each hours level after the change can

be determined using the new net incomes implied by the tax reform. The hours level at which utility is maximised after the reform is determined for each draw. In this way, a frequency distribution, based on k values, of post-reform hours worked can be obtained and the probability of being at each of the discrete hours points, conditional on the pre-reform labour supply, is evaluated for each individual.¹⁰ As explained, some of the k draws, for a selection of individuals, simply fix labour supply at the pre-reform reported level.

3 Construction of Confidence Intervals

This section describes the simulation approach to generate a pseudo sampling distribution of aggregate measures, allowing for the estimated joint sampling distribution of the parameter estimates for the quadratic preference functions. These parameters are augmented, as explained in Section 2.1, to allow for a range of individuals' characteristics, and are estimated separately for different demographic groups. Essentially, the distributions for the aggregate measures are assembled by repeatedly drawing from the estimated sampling distribution of parameters. For each set of parameters, the calibration process and policy reform exercise as described in the previous section are carried out.

In view of the discrete choice nature of the hours decision, revenue and expenditure items for each individual are based on expected values, using the computed probability distribution over hours for the post-reform tax structure. For the pre-reform structure the revenue and costs are based on the discretised value of observed hours worked. Population aggregates are then obtained using the household weights provided with the SIHC by the Australian Bureau of Statistics.

The empirical confidence intervals are created using the point estimates

¹⁰This calibration method differs slightly from earlier versions of MITTS, in which a search for k draws was made using a specified total number of tries. If k draws were not obtained (even if the number was close to k), the individual's hours were fixed. The present method, however, makes use of any successful draws, even if not all k draws could be attained.

of the parameters in the labour supply models (that is, the estimated parameters of the direct utility functions) and the estimated variance/covariance matrix.¹¹ Suppose it is required to calculate the confidence intervals for the changes in expenditure and revenue resulting from a particular policy reform. These can be constructed through the use of the following simulation method. First, a vector of labour supply parameters is drawn randomly from a multivariate normal distribution with the mean and covariance matrix given by the point estimates of the labour supply parameters and the variance/covariance matrix for the parameters respectively.¹² This vector of labour supply parameters is used to carry out the calibration exercise described in the previous section; that is, the required number of sets of draws from the error distributions (the extreme value distributions) are generated such that optimal labour supply corresponds to actual discretised hours in the pre-reform situation. The process of these two steps taken together is described as a ‘replication’. Thus, each vector of preference parameters, drawn from the distribution of preference parameters, gives rise to a ‘replication’ of the preference function, to distinguish it from the set of ‘draws’ from the error distribution for each discrete hours level, which gives optimal hours as observed hours in the pre-reform situation, and the ‘tries’ allowed in collecting the set of draws.¹³

Combining the randomly drawn vector of preference parameters with all the sets of draws from the error distributions obtained in the calibration stage, it is possible to predict the proportion at each of the discrete hours points after the reform. Given these proportions, the expected changes in expenditure and revenue after the reform can be computed for each individual and for the sample as a whole. This procedure is repeated a large number of times. After each replication, the corresponding estimate for the expected change in, say, expenditure is stored in a vector. After completing

¹¹As mentioned earlier, the utility functions in MITTS are quadratic, with parameters that are functions of individual characteristics, as in equation (4).

¹²The method used to obtain random values from a multivariate normal distribution is described in Appendix B.

¹³The ‘tries’ resulting in successful ‘draws’ in the calibration exercise for one particular ‘replication’ may not place the individual at observed hours when a different replication is used. For this reason the calibration stage needs to be repeated for each replication.

all replications, this vector, the pseudo sampling distribution, is ranked in ascending order. A 90 percent confidence interval can be constructed using this vector by taking the 5th percentile and the 95th percentile as the lower and upper bounds of the interval. In a similar fashion, confidence intervals can be constructed for other measures of interest, such as the probability of participation or changes in revenue.

These calculations are computationally intensive. Indeed, it could take weeks or even months - depending on the number of households, the number of specified draws from the error distribution, the number of tries allowed for each draw and the number of replications on which the intervals are based - before confidence intervals for the results from the behavioural component could be provided for a particular policy change. For example, on a Pentium-4 machine with 512 Mb of RAM and 1.90 GHz, it takes 200 minutes to run one behavioural simulation for 7170 households in the 1997/98 SIHC dataset if, for each household, the total number of draws is set to 100 and the maximum number of tries allowed to calibrate the observed hours for each draw is set to 1000. If the simulation is repeated 1000 times it would take approximately 200,000 minutes (around 4.5 months) in total to calculate the confidence intervals for this sample.¹⁴ It is clear that this is impractical.

An obvious influence on the computation time is the number of individuals or households examined. In order to keep the time to a minimum, it is therefore best (where possible) to calculate the confidence intervals for each demographic group (for which separate preference parameter estimates are obtained) separately rather than using all households in the dataset. This allows the user to use more than one computer to calculate the intervals, which will reduce the total amount of (calendar) time needed for calculations for the entire population.

¹⁴The time required for each replication may not be the same, as different numbers of tries may be needed to obtain the requisite draws.

4 Numerical Examples: a Policy Change

This section uses the approach described in Section 3 to produce confidence intervals for the policy reform of reducing all taper rates in the March 1998 tax system, from 50 and 70 per cent to 30 per cent.¹⁵ This simulation uses the 1997/98 SIHC dataset. Results are reported here for couples with children and sole parents. Appendix C contains the results for the other groups for comparison, but these are not discussed separately in this section. As mentioned in Section 3, the number of draws as well as the maximum number of tries can be varied. Instead of taking an arbitrary number, it is useful to experiment to find the best combination; that is, the smallest set of draws and tries which produces stable results. This is discussed in Subsection 4.1. Confidence intervals are then reported in Subsection 4.2 for changes in labour supply, and in Subsection 4.3 for changes in government revenue and expenditure.

4.1 Sampling From the Error Distributions

Computation of confidence intervals is complicated by the large number of random variables generated in the calibration process, and the associated utility maximisation analysis, which has to be repeated for each set of preference parameters. This subsection considers how this part of the process can be reduced to a minimum.

Tables 1 and 2, for couples and singles respectively, provide an idea of how sensitive the results of one replication are to different combinations of draws and tries. The first two columns of these tables give the specified combination of k , the number of draws, and K , the number of tries for each draw. The third column gives the time in minutes to do a single simulation for all households, using the point estimates of preference functions (using a Pentium-4 machine with 512 Mb of Ram and 1.90 GHz). In each case the variable r indicates the number of draws for which labour supply is held

¹⁵This is the reform considered in Creedy, Kalb and Kew (2003) using point estimates of preference functions. However, the wage and preference functions used in MITTS have since been re-estimated.

constant at the base discretised hours (that is, after K tries, it was not possible to find a set of random variables from the error distribution giving optimal hours equal to observed hours). For all groups, 11 hours points are used in the labour supply model, except for married men, who are less likely to work part time than the other groups. Therefore, only 6 hours points are distinguished to model labour supply for this group.

Table 1: Average Hours Change by Number of Draws and Tries: All Couples

| k | K | Time (mins) | Average Hours Change | | $r < k$ | $r = k$ |
|-----|------|----------------|----------------------|------------------|---------|---------|
| | | | Married Men | Married Women | | |
| 100 | 1000 | 200 | 0.46 | -0.47 | 164 | 37 |
| 150 | 1000 | 280 | 0.46 | -0.47 | 206 | 27 |
| 100 | 500 | 151 | 0.32 | -0.50 | 256 | 64 |
| 100 | 100 | 80 | 0.13 | -0.45 | 1002 | 99 |
| 50 | 1000 | 115 | 0.46 | -0.48 | 93 | 64 |
| 40 | 1000 | 94 | 0.46 | -0.50 | 78 | 73 |
| 30 | 1000 | 80 | 0.45 | -0.48 | 57 | 81 |
| 20 | 1000 | 60 | 0.42 | -0.48 | 42 | 88 |
| 10 | 1000 | 40 | 0.38 | -0.49 | 21 | 99 |
| 10 | 700 | 38 | 0.35 | -0.49 | 21 | 125 |
| 10 | 500 | 35 | 0.26 | -0.51 | 37 | 155 |
| 10 | 100 | 30 | 0.01 | -0.45 | 286 | 289 |

Note: The total number of couples with and without children is 2474.

k = number of draws; K = number of tries per draw;

r = no of draws which involve fixed labour supply at base discretised hours.

When $r < k$, at least one, but less than k , draws are fixed at base discretised hours.

When $r = k$, all k draws are fixed at the base discretised hours.

The first row of each table, where the number of draws and tries are both relatively large, may be used as a benchmark. The results concerning the average change in hours worked remain almost identical when the number of draws is increased from 100 to 150, suggesting that the results become stable as the number of draws and/or tries is increased above a certain number. Choosing a smaller number of draws means that substantially less

Table 2: Average Hours Change by Draws and Tries: Singles

| k | K | Time (mins) | Single Men | | | Single Women | | | Sole Parents | | |
|-----|------|----------------|------------|---------|---------|--------------|---------|---------|--------------|---------|---------|
| | | | ΔH | $r < k$ | $r = k$ | ΔH | $r < k$ | $r = k$ | ΔH | $r < k$ | $r = k$ |
| 100 | 1000 | 200 | 0.33 | 53 | 6 | 0.14 | 32 | 2 | 2.28 | 7 | 0 |
| 150 | 1000 | 280 | 0.33 | 63 | 6 | 0.13 | 43 | 1 | 2.26 | 9 | 0 |
| 100 | 500 | 151 | 0.28 | 134 | 7 | 0.08 | 80 | 5 | 2.27 | 27 | 0 |
| 100 | 100 | 80 | 0.13 | 214 | 18 | -0.06 | 231 | 12 | 2.20 | 132 | 0 |
| 50 | 1000 | 115 | 0.30 | 32 | 7 | 0.13 | 22 | 5 | 2.40 | 3 | 0 |
| 40 | 1000 | 94 | 0.29 | 28 | 9 | 0.12 | 19 | 6 | 2.49 | 2 | 0 |
| 30 | 1000 | 80 | 0.28 | 24 | 11 | 0.13 | 18 | 6 | 2.55 | 1 | 0 |
| 20 | 1000 | 60 | 0.27 | 14 | 16 | 0.13 | 11 | 11 | 2.64 | 0 | 0 |
| 10 | 1000 | 40 | 0.29 | 10 | 18 | 0.13 | 6 | 12 | 2.71 | 0 | 0 |
| 10 | 700 | 38 | 0.26 | 17 | 33 | 0.06 | 10 | 23 | 2.71 | 1 | 0 |
| 10 | 500 | 35 | 0.24 | 23 | 52 | 0.01 | 9 | 31 | 2.71 | 2 | 2 |
| 10 | 100 | 30 | 0.10 | 63 | 104 | -0.10 | 58 | 64 | 2.47 | 24 | 16 |

Note: The total number of single men, single women and sole parents is 1392, 1038 and 503 respectively.

k = number of draws; K = number of tries per draw; ΔH = average hours change;

r = no of draws which involve fixed labour supply at base discretised hours.

When $r < k$, at least one, but less than k , draws are fixed at base discretised hours.

When $r = k$, all k draws are fixed at the base discretised hours.

time is needed to run one simulation. However, as the number of draws is decreased, the results become less accurate and more variable across simulations (depending more on the particular random draws that are taken from the error distribution). Decreasing the number of tries means labour supply is fixed at observed hours for more draws. That is, the average number of hours for the different subgroups are further from the benchmark results. The last row, which has only 10 draws and 100 tries, presents changes in the average numbers of hours worked that are very different from the benchmark figures.

The optimal number of draws and tries may be different for different subgroups and different policy simulations. Choosing 100 draws and 1000 tries is a safe choice for all groups, but it can most likely be reduced. It appears that married men, single men, and single women are less sensitive to the number of draws, but more sensitive to the number of tries. Sole parents, on the other hand, are less sensitive to the number of tries, but the number of draws plays an important role in achieving a stable result. The results for married women are the least sensitive of all groups. Before starting repeated calculations to generate confidence intervals, it is useful to experiment with the number of draws and tries in a simulation, using the point estimates of the labour supply model, to find a combination that takes as little time to run as possible while providing stable results. Investing some time at this stage can reduce the time needed to compute the confidence intervals considerably.

It was recommended in Section 3 that the confidence intervals should be calculated for one demographic group at a time rather than for all households together, so that more than one computer could be used for the calculations. A further advantage of separating the groups is that one demographic group may require more draws to obtain a stable result than another demographic group. By calculating the intervals separately for each subgroup the minimum number of draws sufficient for that subgroup can be chosen.

These results show that it is impractical to set a large number of draws and tries when calculating confidence intervals, because the simulation needs to be repeated so many times that the time needed with many draws becomes problematic from a computational point of view. A large number of draws

and tries may be unnecessary in many cases because the predicted values become stable for much lower numbers of draws and tries. In obtaining the results in the next subsections, in each case the decision was made to set the number of draws to 40 with 1000 tries per draw. The number of replications was set to 1000.

4.2 Labour Supply

Table 3 presents the simulation results regarding labour supply changes. All values refer to percentages of the demographic group, with the exception of the last row of each segment which reports the change in average hours worked. Within each section of the table, the first item, ‘all workers’ refers to the percentage of the demographic group working, including the self-employed along with wage and salary earners. The second item ‘wage/sal. workers’ refers to wage and salary earners, who are working and for whom hours of work are observed, as a percentage of the total number of people in the group. In the simulation, the behaviour of the self-employed is not modelled, so their gross income is assumed to remain constant. The final three columns of Table 3 show the 5th and 95th percentiles of the pseudo sampling distribution, along with the median value.

The effect of the flattening of the budget constraints has an ambiguous effect on labour supply, since marginal tax rates are reduced at lower working hours while they are increased at higher levels of labour supply. The average hours for women is expected to fall as a result of the reform. Although the percentage of women who work fewer hours after the reform is actually lower than the corresponding value for men, more married women move from work to non-work. For men and women in couples, the confidence intervals are not particularly wide and it seems that these labour supply responses are obtained with a reasonable amount of precision.

The reform is expected to have a larger effect on the labour supply of sole parents. A much larger proportion moves from non-work to work than for couples, though this result is less precise as indicated by the wider confidence interval. The increase in average hours is significantly larger than for people

Table 3: Simulated Labour Supply Responses to Reduction in Taper Rates

| | Confidence Intervals | | | |
|-------------------------------------|----------------------|-------|--------|-------|
| | mean | 5% | median | 95% |
| Couples with children: Men | | | | |
| All workers: base | 87.3 | | | |
| wage/sal. workers: base | 72.57 | | | |
| wage/sal. workers: post | 75.67 | 75.51 | 75.67 | 75.81 |
| non-work to work | 3.53 | 3.40 | 3.53 | 3.65 |
| work to non-work | 0.43 | 0.36 | 0.43 | 0.52 |
| workers working more | 0.67 | 0.57 | 0.67 | 0.75 |
| workers working less | 2.59 | 2.42 | 2.58 | 2.77 |
| average hours change | 0.99 | 0.91 | 0.99 | 1.06 |
| Couples with children: Women | | | | |
| all workers: base | 59.94 | | | |
| wage/sal. workers: base | 51.12 | | | |
| wage/sal. workers: post | 50.74 | 50.52 | 50.74 | 50.97 |
| non-work to work | 2.28 | 2.07 | 2.28 | 2.49 |
| work to non-work | 2.67 | 2.5 | 2.67 | 2.83 |
| workers working more | 0.2 | 0.16 | 0.2 | 0.23 |
| workers working less | 1.02 | 0.89 | 1.02 | 1.15 |
| average hours change | -0.63 | -0.7 | -0.63 | -0.55 |
| Sole Parents | | | | |
| all workers: base | 47.37 | | | |
| wage/sal. workers: base | 42.71 | | | |
| wage/sal. workers: post | 50.94 | 49.29 | 50.7 | 53.46 |
| non-work to work | 8.45 | 6.81 | 8.2 | 10.96 |
| work to non-work | 0.22 | 0.10 | 0.22 | 0.36 |
| workers working more | 1.89 | 1.28 | 1.87 | 2.54 |
| workers working less | 2.34 | 1.94 | 2.32 | 2.82 |
| average hours change | 2.63 | 1.98 | 2.56 | 3.49 |

in couples, though again the confidence interval is wider. Comparing the standard deviations for the point estimates of the labour supply parameters for sole parents and couples, not reported here, it is clear that those for couples are relatively smaller, indicating more precise estimates.¹⁶ Obviously, this has translated into more precise predictions for behavioural changes in the simulation for couples.

4.3 Revenue and Expenditure

Table 4 reports results for the demographic groups regarding a number of revenue and expenditure aggregates. This table reports the aggregates in millions of dollars for the base tax and transfer system under the pre-reform heading. These are evaluated at the discretised hours, the discrete hours level nearest to the actual hours reported. The effects of the reform, under the extreme assumption that labour supplies do not change, is given under the heading ‘fixed labour’, which shows the absolute change in the expenditure or revenue on each item. The final columns show details of the absolute change, while allowing for labour supply changes, except for the self-employed, the retired and disabled, whose labour supplies are fixed at their observed pre-reform hours.¹⁷

The comparisons show first that allowing for potential labour supply effects of tax reforms matters in the estimation of taxes and expenditures. Not allowing for such responses would, for example, lead to a significant overstatement of the income tax revenue change for couples. The assumption of fixed labour supply suggests that tax rebates for sole parents are expected to fall, yet the positive net labour supply responses imply that rebates increase. Furthermore, the increase in allowances for sole parents is substantially lower when labour supply is endogenous. The increase in net expenditure on this demographic group is significantly overstated when labour supply is held fixed.

¹⁶For details, see Kalb (2002).

¹⁷In addition, there is the small number of individuals who cannot be placed at their observed hours level in the base system, even after Kk sets of n random values are examined.

Table 4: Change in Tax and Transfer Costs Resulting from Reduced Taper Rates

| | Pre-Reform (\$m) | Changes from Pre- to Post-Reform | | | | |
|---------------------|---------------------|----------------------------------|------------------------|----------------------|--------|--------|
| | | Fixed Labour (\$m) | Variable Labour Supply | | | |
| | | | Mean (\$m) | Confidence Intervals | | |
| | | | 5% | median | 95% | |
| Couples | | | | | | |
| Revenue | | | | | | |
| Income Tax | 28929.8 | 527.1 | 392.6 | 349.4 | 393.0 | 431.6 |
| Medicare | 1736.5 | 74.6 | 70.2 | 67.1 | 70.2 | 73.0 |
| Total Revenue | 30666.3 | 601.7 | 462.8 | 417.1 | 463.0 | 504.0 |
| Expenditure | | | | | | |
| Tax Rebates | 1181 | -363.4 | -350.3 | -353.5 | -350.3 | -346.9 |
| Fam Payment | 4078.3 | 1686.9 | 2035.6 | 2017.7 | 2035.3 | 2055.4 |
| FTP/FTB | 402.4 | 179.2 | 214.7 | 211.6 | 214.6 | 217.7 |
| Allowances | 4271.7 | 3620.8 | 3468.0 | 3426.9 | 3467.1 | 3509.2 |
| Pensions | 737.7 | 42.5 | 27.0 | 23.5 | 27.0 | 33.5 |
| Pharm Allow | 8.9 | 0.6 | 0.1 | 0.0 | 0.1 | 0.2 |
| Rent Allow | 359.8 | 119.9 | 156.7 | 153.3 | 156.7 | 160.0 |
| Total Expenditure | 11039.9 | 5286.6 | 5551.6 | 5493.8 | 5551.9 | 5611.9 |
| Net Expenditure | -19626.4 | 4684.8 | 5088.8 | 4992.6 | 5088.2 | 5191.5 |
| Sole parents | | | | | | |
| Income Tax | 1900.4 | 75.6 | 158.7 | 123.9 | 156.4 | 199.7 |
| Medicare | 83.9 | 4.8 | 5.6 | 4.0 | 5.6 | 7.1 |
| Total Revenue | 1984.4 | 80.4 | 164.3 | 128.1 | 161.8 | 206.9 |
| Expenditure | | | | | | |
| Tax Rebates | 553.6 | -12.5 | 18.7 | 12.6 | 17.9 | 28.0 |
| Fam Payment | 2139.7 | 90.2 | 132.5 | 124.6 | 132.2 | 141.1 |
| FTP/FTB | 229.9 | 0 | 0 | 0 | 0 | 0 |
| Allowances | 3025.2 | 264.2 | 93.7 | 35.6 | 97.8 | 136.6 |
| Pensions | 155 | 1.1 | -4.1 | -5.6 | -4.0 | -2.6 |
| Pharm Allow | 49.8 | 6.0 | 7.5 | 7.2 | 7.4 | 7.8 |
| Rent Allow | 405.7 | 5.8 | 14.8 | 12.6 | 14.8 | 17.2 |
| Total Expenditure | 6558.9 | 354.7 | 263.1 | 220.2 | 265.4 | 299.3 |
| Net Expenditure | 4574.5 | 274.3 | 98.8 | 11.3 | 102.4 | 171.0 |

In addition, Table 4 shows that for couples, the confidence intervals for most items are relatively small. However, some of the sole parent confidence intervals are relatively large. For example, changes in tax rebates, allowances and especially net expenditure are estimated with a relatively low precision. The width of the confidence interval for the change in net expenditure is about 200 million dollars for couples with children and about 160 million dollars for sole parents. Given the much lower number of sole parents in the population, this means that per individual the uncertainty associated with the expected change after the reform is substantially larger for sole parents.

5 An Alternative Approach

The previous section has demonstrated the value of examining confidence intervals using the proposed method of computing a pseudo sampling distribution. It showed that all predicted changes were significantly different from zero. This indicates that there is some confidence in the direction of the predicted effects. The main drawback of the method is the length of time needed, given the large computational burden. Even with fast modern computers, the time involved is too large to make it viable for regular use, particularly when attempting to design policy reforms. The design of policy changes, with a range of objectives in mind, including a government budget constraint, typically involves the examination of a large number of experiments. Hence the question arises of whether a faster method of constructing confidence intervals can be found. This question is addressed in the present section.

The role of the sets of random draws from the error distributions relating to the direct utility function, with a maximum number of tries, has been discussed. Before obtaining confidence intervals, experiments were carried out to find the minimum combination of k and K (draws and tries respectively) that appeared to be viable. Hence there is little room to further reduce this aspect of the computational burden.

Examination of Tables 3 and 4 shows that the mean and median changes resulting from the tax change are in all cases close to each other, suggesting

that the relevant distributions are close to being symmetric.¹⁸ The potential for reducing the number of replications was explored by plotting histograms of the various empirical pseudo sampling distributions, having obtained a range of distributions, such as distributions for various expenditure and revenue items, as well as total expenditure and revenue, based on 1000 replications. Examples for sole parents are shown in Figure 1. The left hand set of distributions relate to labour supply changes and the right hand set is for total revenue and expenditure, and net expenditure. In each histogram, the continuous line shows the density function of a normal distribution having the same arithmetic mean and variance as the empirical frequencies.

While the distributions do not appear to be well-described by the normal distribution, and in fact normality is rejected for some of the distributions, it is useful to consider an approach that is based on the properties of the normal distribution given the reasonably symmetric shape of many of the distributions. Thus, instead of taking a large number of replications and finding the 5th and 95th percentiles, consider using a smaller number of replications to estimate the mean μ and standard deviation σ for the outcomes of interest. Assuming normality, the appropriate percentiles of the standard normal $N(0, 1)$ distribution can then be used to find the required confidence interval based on $N(\mu, \sigma^2)$. For example, the 5th and 95th percentiles are obtained as $\mu \pm 1.645\sigma$.

Results of this exercise are given in Figure 2, which shows, for each summary variable, the 90 per cent confidence limits obtained when using 1000 replications (represented by the solid horizontal lines in the graphs) and the limits produced when an assumption of normality is used, as the number of replications (N) is increased. It can be seen that the confidence limits produced by assuming normality are reasonably close to those obtained from 1000 replications, even for as few as 50 replications. However, the limits using 1000 replications take about 20 times longer to produce than those obtained using only 50 replications. This means that most results could be produced within one or two days with this more approximate approach.

¹⁸Furthermore, inspection of the (unreported) results using only the point estimates of preference functions shows that these are also quite close to the mean and median values.

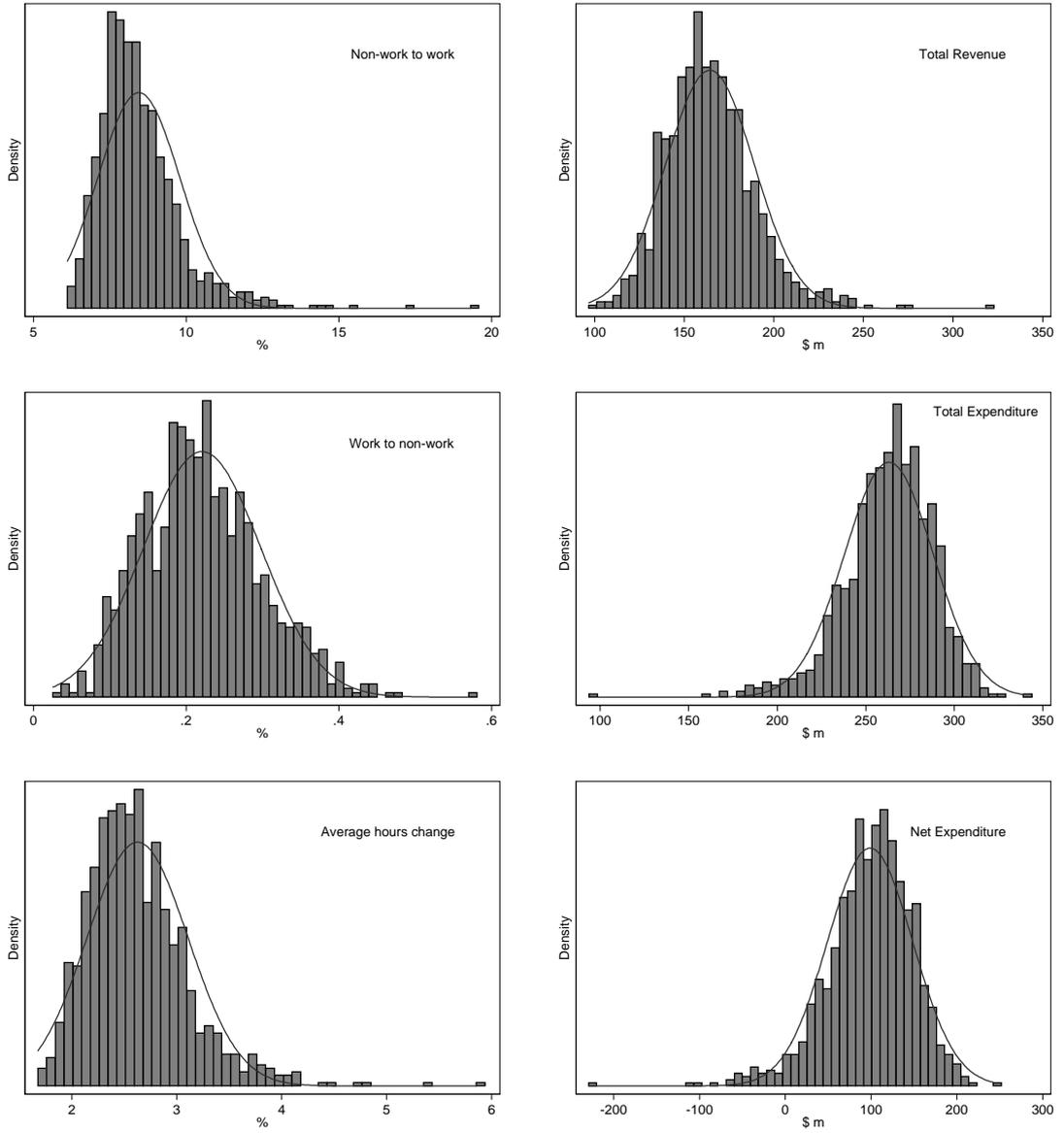


Figure 1: Sole Parents' Histograms of Simulation Outcomes

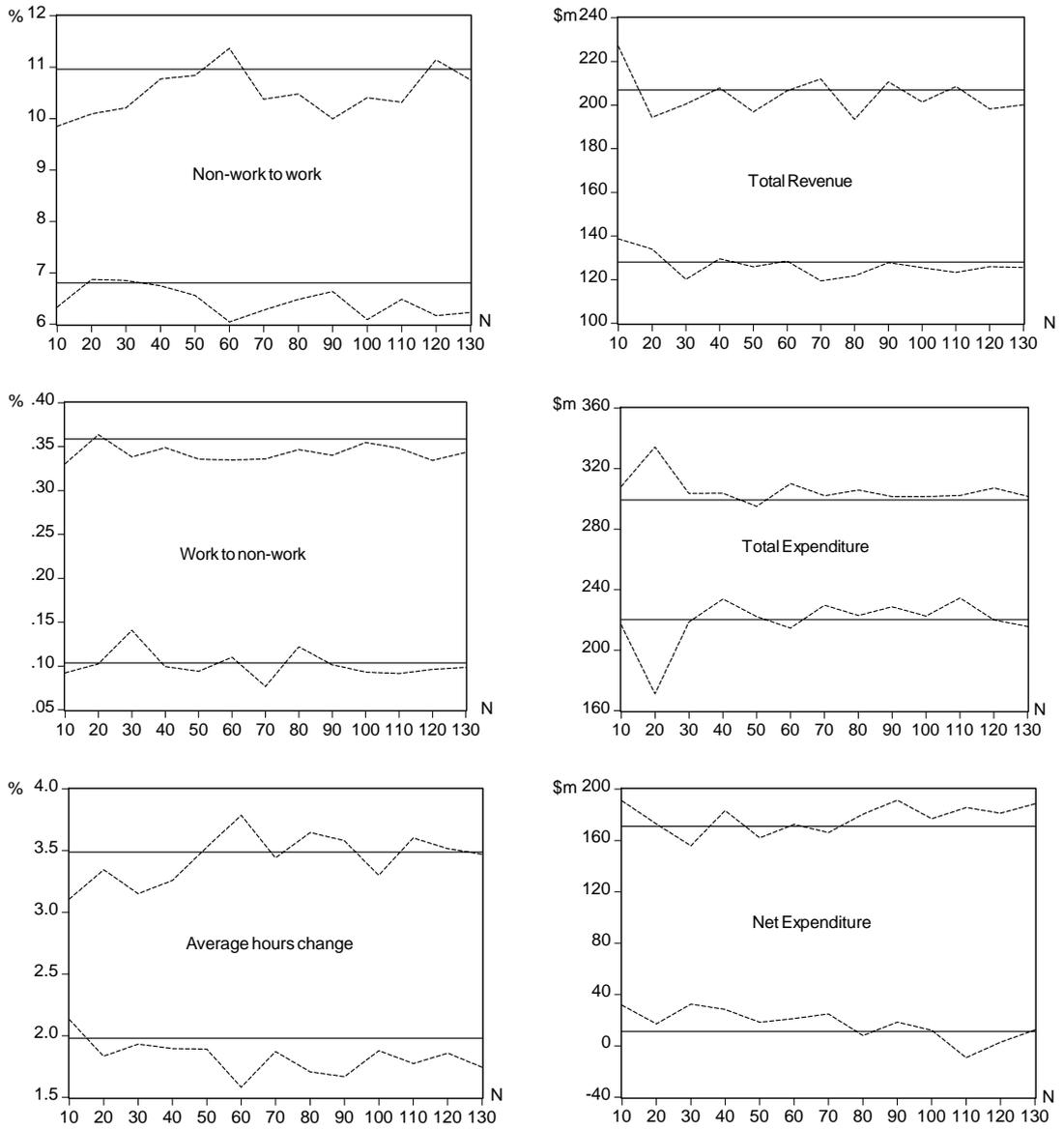


Figure 2: Number of Replications (N) and Confidence Intervals: Sole Parents

Figures 3 and 4 show the histograms for the labour supply summary changes for married men and married women respectively. For this group, the pseudo sampling distributions are closer to normality (as also shown by a range of distributions for other variables, not presented here). The right hand segments of these figures compare the confidence limits obtained from 1000 replications compared with the use of small samples and an assumption of normality. Again, good approximations are produced with as few as 50 replications. Similar results apply to tax and expenditure totals for couples as shown in Figure 5.

Similar results were found to apply to a wide range of expenditure and revenue categories. The same is true for other demographic groups, and brief results for single men and women, and couples without children are reported in Appendix C. Even where the plotted histograms are not particularly close to being normal, the use of a smaller number of replications combined with an assumption of normality provides a good guide to the confidence intervals that are produced by the use of a large number of replications to generate a pseudo sampling distribution. Exceptions to this general rule were found to arise in the case of the pharmaceutical allowance for couples and sole parents. Where very few individuals in the sample are in receipt of a particular benefit, as in this case, caution must be exercised. However, the aggregate amount of this payment compared to those of other payments is negligible, so the uncertainty associated with it tends to have only a small effect on the total expenditure and revenue.

6 Conclusions

This paper has examined one important source of uncertainty surrounding tax policy simulations of labour supply, taxation and revenue changes, when using a behavioural tax microsimulation model. This arises from the fact that the preference functions, on which predicted labour supply behaviour is based, are estimated using econometric methods, and typically only the point estimates are used in simulations. While the uncertainty associated with the estimated parameters is just one source of uncertainty regarding

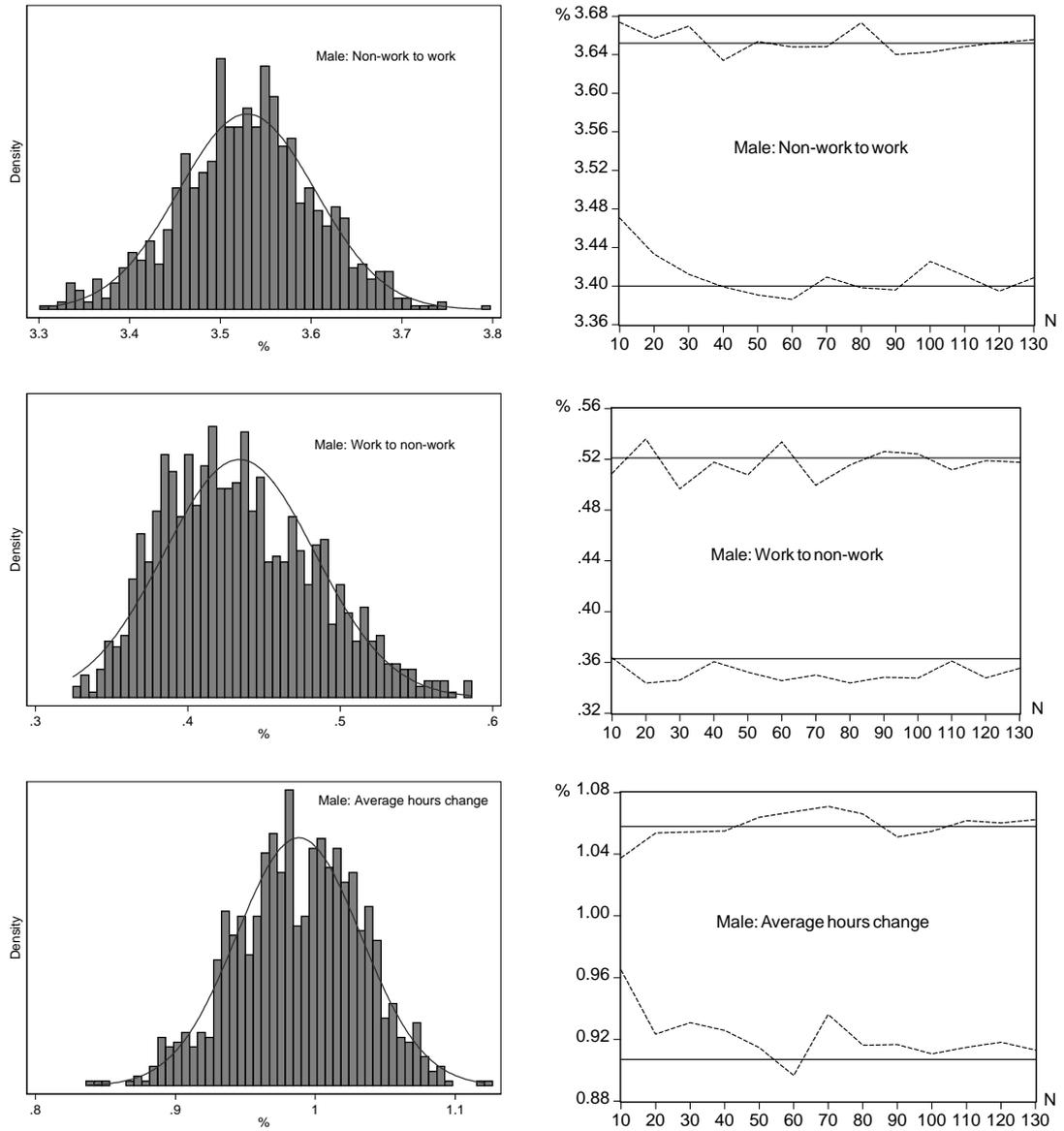


Figure 3: Married Men: Histograms of Simulation Outcomes and Confidence Intervals by Number of Replications (N)

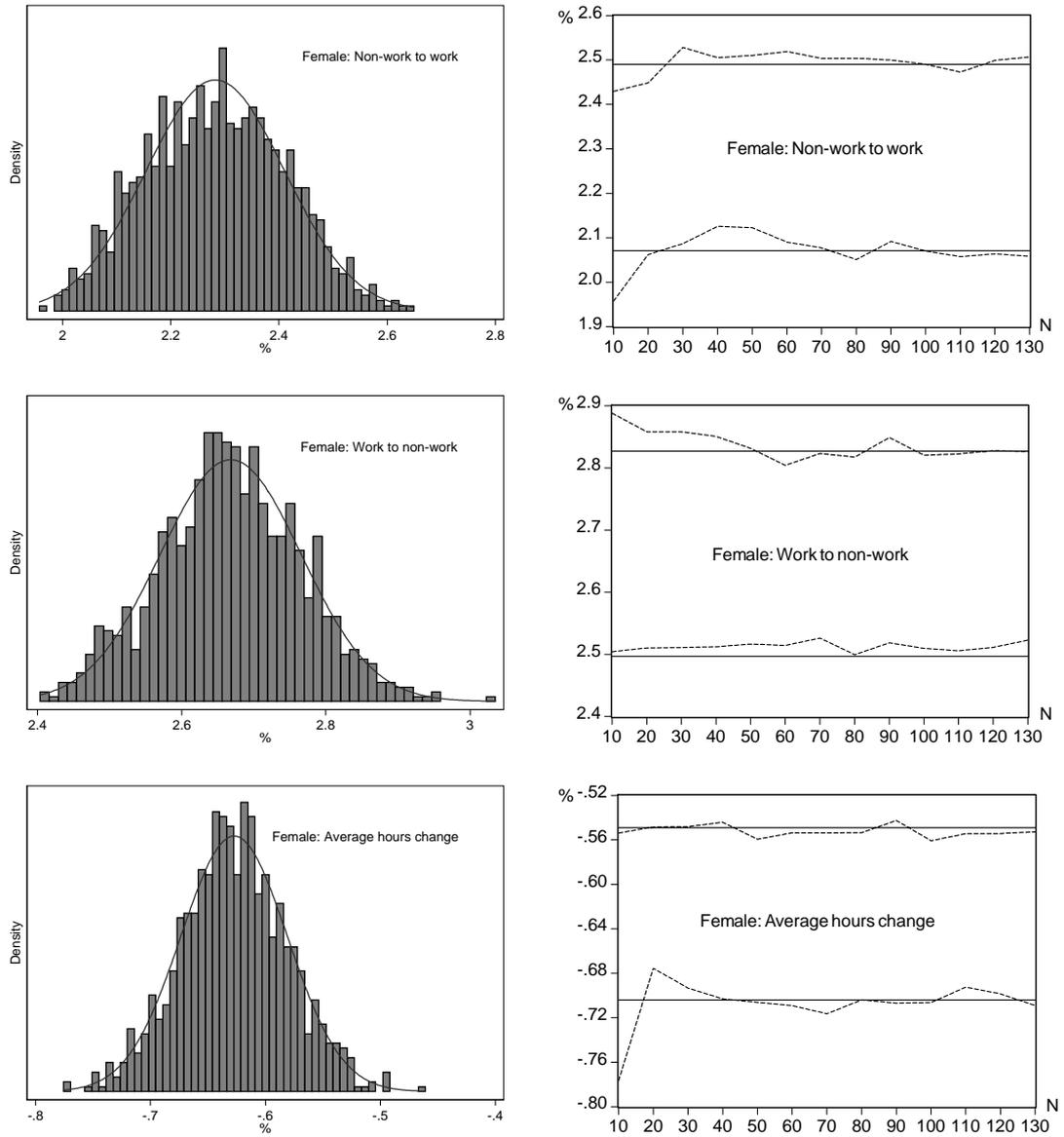


Figure 4: Married Women: Histograms of Simulation Outcomes and Confidence Intervals by Number of Replications (N)

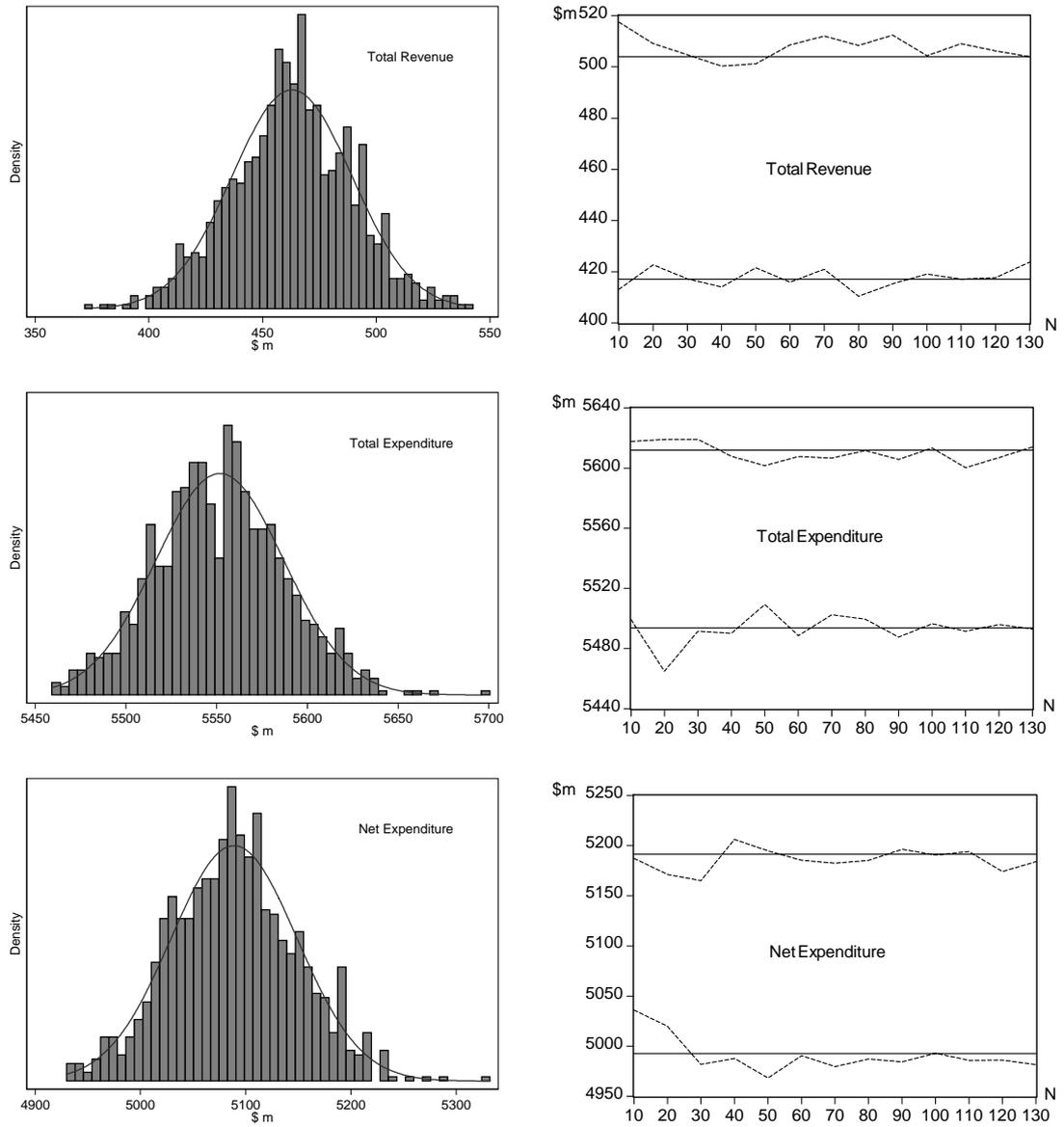


Figure 5: Couples' Revenue and Expenditure: Histograms of Simulation Outcomes and Confidence Intervals by Number of Replications (N)

simulations of policy reforms, it represents the major additional source generated by the extra requirements of behavioural microsimulation compared with arithmetic, or non-behavioural, models.

The analysis was carried out in the context of the MITTS model for the Australian direct tax and transfer system. MITTS uses a discrete hours structural approach to labour supply modelling, whereby a probability distribution over a set of discrete hours levels is generated after a policy reform, conditional on each individual being placed at the observed hours in the base dataset. For each variable of interest, a pseudo sampling distribution was generated by taking repeated random draws from the multivariate distribution of parameter values in the preference functions, obtained from maximum likelihood estimation. A broad finding is that, for most demographic groups and variables, the 90 per cent confidence intervals are relatively narrow. Hence the additional uncertainty introduced by behavioural modelling seems low. Exceptions were found for several variables in the case of sole parents. The large confidence intervals in those cases were associated with the higher estimated variances in the variance/covariance matrix for this demographic group.

A disadvantage of this approach, which requires a large number of replications, is that the computational requirements are great. Simulation results for a single demographic group can take several weeks or months to complete. Further investigation revealed that results close to the intervals based on 1000 replications can generally be obtained by taking a smaller number of replications, sometimes as low as 40 to 50, and using these to estimate the mean and variance of the outcome of interest. Under the assumption of the outcome being normally distributed, a confidence interval can then be constructed. This method appears to overcome the computational burden without sacrificing much accuracy.

Appendix A: Labour Supply Calibration and Simulation

This appendix provides a more formal description of the calibration and simulation processes used in MITTS. Assume an individual is observed to work at the discretised hours level, h_i . In the calibration stage (for a given set of preference function parameter values), suppose that it is required to obtain k sets of draws from the error distributions (for each hours level) for which the individual's optimal hours are equal to the observed hours. A total of up to K attempts, or 'tries' are allowed to generate each draw. There are n discrete hours levels.

Define the following terms and consider the generation of a single draw. Let

$$v^{(j)} = [v_1^{(j)}, \dots, v_n^{(j)}]' \quad (12)$$

denote the j th vector of n random variables (the j th 'try') drawn from the Extreme Value Type 1 distribution. Let

$$U^b = [U_1^b, \dots, U_n^b]' \quad (13)$$

denote the n -vector of utilities corresponding to the hours levels, calculated using the tax and transfer system to obtain the associated net incomes for each discrete hours level under the base (or pre-reform) system. Finally, let $U^{*b(j)}$ denote the vector of utilities:

$$\begin{bmatrix} U_1^{*b(j)} \\ \cdot \\ \cdot \\ U_n^{*b(j)} \end{bmatrix} = \begin{bmatrix} U_1^b \\ \cdot \\ \cdot \\ U_n^b \end{bmatrix} + \begin{bmatrix} v_1^{(j)} \\ \cdot \\ \cdot \\ v_n^{(j)} \end{bmatrix}$$

For each 'try' a check is made to see if $U_i^{*b(j)}$ is the maximum value among $U_\ell^{*b(j)}$ for all other $\ell \neq i$. If so, this means that the individual chooses to work at h_i , which corresponds to the observed discrete hours for this individual. In this case, the vector $v^{(j)}$ is stored for use in the simulation stage; this corresponds to a 'draw'. If $U_i^{*b(j)}$ is not the maximum value, $v^{(j)}$ is discarded and another set of n random values from the error distribution (another 'try')

is selected. This process is repeated until a successful ‘draw’ has been found. If, after K tries, no set is found for which optimal and observed discretised hours are equal, hours are held fixed at the observed hours for this draw.

This process is repeated k times. For most individuals, k draws are obtained using far less than the K tries for each.

The simulation of a policy reform proceeds as follows. Let $U^r = [U_1^r, \dots, U_n^r]'$ denote the n -vector of utilities after the tax reform. Then calculate $U^{*r(j)}$, based on the j th stored vector $v^{(j)}$, for $j = 1, \dots, k$:

$$\begin{bmatrix} U_1^{*r(j)} \\ \cdot \\ \cdot \\ U_n^{*r(j)} \end{bmatrix} = \begin{bmatrix} U_1^r \\ \cdot \\ \cdot \\ U_n^r \end{bmatrix} + \begin{bmatrix} v_1^{(j)} \\ \cdot \\ \cdot \\ v_n^{(j)} \end{bmatrix} \quad (14)$$

For each $j = 1, \dots, k$ a check is made to see which $U_\ell^{*r(j)}$, for $\ell = 1, \dots, n$, is the maximum value in the vector $U^{*r(j)}$. This determines the discrete hours level after the reform for the j th set of draws, given that the individual was at h_i in the base system. The k draws thus generate the probabilities of being at each of the discrete hours points after the reform by counting the number of times for which a draw results in a particular hours point, and dividing this number by the total number of draws k . As mentioned before, for some individuals, a number of the k draws may be replaced by an assumption that labour supply is unchanged as a result of the tax reform. In that case, the discrete hours level after the reform is the same as the observed discrete hours level before the reform.

Appendix B: Random Draws From a Multivariate Normal Distribution

This appendix describes how random draws from a multivariate normal distribution may be obtained, given a vector of means and a variance/covariance matrix. In the univariate case, if η is a random draw from an $N(0, 1)$ distribution, a random observation, x , from a univariate normal, $N(\mu, \sigma^2)$ is given by:

$$x = \mu + \sigma\eta \quad (15)$$

The corresponding result for an n -element vector x from a multivariate normal, $MN(\mu, \Omega)$, is:

$$x = \mu + \Omega^{\frac{1}{2}}\eta \quad (16)$$

where μ and η are n -element vectors and Ω is an n by n variance-covariance (symmetric positive definite) matrix:

$$\begin{aligned} \Omega &= E(x - \mu)(x - \mu)' \\ &= \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \sigma_{n1} & \dots & \sigma_{nn} \end{bmatrix} \end{aligned} \quad (17)$$

To generate a random set of draws from the multivariate distribution requires the matrix $\Omega^{1/2}$, which is the square root of Ω , and η , which consists of n draws from the standard normal distribution $N(0, 1)$. This matrix can be obtained using the Cholesky factorisation, which states that a symmetric positive definite matrix Ω can be expressed as:

$$\Omega = LL' \quad (18)$$

where L is a lower triangular matrix. Given $\Omega^{1/2} = L$, (16) can be rewritten as:

$$x = \mu + L\eta \quad (19)$$

Confirmation of this result can be obtained by examining the mean and variance, as follows. The mean is:

$$\begin{aligned} E(x) &= E(\mu + L\eta) \\ &= \mu + LE(\eta) = \mu \end{aligned} \quad (20)$$

while the variance:

$$\begin{aligned}
 E [(x - \mu) (x - \mu)'] &= E [L\eta (L\eta)'] \\
 &= LE [\eta\eta'] L' \\
 &= LL' = \Omega
 \end{aligned} \tag{21}$$

The Cholesky factorisation, that is, the solution for L in (18), can be achieved using the following simple recursive algorithm:

$$\begin{aligned}
 \text{Start with:} & \quad L_{11} = \sqrt{\sigma_{11}} \\
 \text{Then calculate for } i = 2, \dots, n: & \quad L_{i1} = \sigma_{i1}/L_{11} \\
 \text{For all } j = 2, \dots, n: & \\
 \text{First calculate:} & \quad L_{jj} = \sqrt{\sigma_{jj} - \sum_{k=1}^{j-1} L_{jk}L_{jk}} \tag{22} \\
 \text{Then for all } i = j + 1, \dots, n: & \quad L_{ij} = \left(\sigma_{ij} - \sum_{k=1}^{j-1} L_{ik}L_{jk} \right) / L_{jj} \\
 \text{For all } i < j: & \quad L_{ij} = 0
 \end{aligned}$$

where L_{ij} denotes the element in the i th row and the j th column of L .

Appendix C: Results for Singles and Couples Without Children

Section 4 concentrated on simulation results for couples with children and sole parents. This appendix reports comparable results for single men and women, and couples without children. The tables show that in general the confidence intervals are quite narrow for these groups, and the use of a small number of replications combined with an assumption of normality can provide reliable estimates of those intervals.

Table 5: Simulated Labour Supply Responses to a Reduction in Taper Rates: Single Men and Women

| | Confidence Intervals | | | |
|-------------------------|----------------------|-------|--------|-------|
| | mean | 5% | median | 95% |
| Single Men | | | | |
| all workers: base | 67.30 | | | |
| wage/sal. workers: base | 59.57 | | | |
| wage/sal. workers: post | 61.31 | 60.89 | 61.30 | 61.78 |
| non-work to work | 1.86 | 1.45 | 1.85 | 2.33 |
| work to non-work | 0.11 | 0.06 | 0.12 | 0.16 |
| workers working more | 0.08 | 0.00 | 0.07 | 0.17 |
| workers working less | 1.24 | 1.04 | 1.21 | 1.57 |
| average hours change | 0.41 | 0.27 | 0.41 | 0.55 |
| Single Women | | | | |
| all workers: base | 48.01 | | | |
| wage/sal. workers: base | 44.90 | | | |
| wage/sal. workers: post | 46.92 | 46.59 | 46.91 | 47.28 |
| non-work to work | 2.16 | 1.83 | 2.14 | 2.52 |
| work to non-work | 0.13 | 0.10 | 0.13 | 0.17 |
| workers working more | 0.13 | 0.07 | 0.12 | 0.20 |
| workers working less | 2.79 | 2.39 | 2.77 | 3.22 |
| average hours change | 0.17 | 0.07 | 0.17 | 0.26 |

Table 6: Change in Tax and Transfer Costs Resulting from a Reduction in Taper Rates: Single Men and Women

| | Pre-Reform | Changes from Pre- to Post-Reform | | | | |
|---------------------|------------|----------------------------------|------------------------|------------|----------------------|--------|
| | | Fixed Labour | Variable Labour Supply | | | |
| | (\$m) | | (\$m) | Mean (\$m) | Confidence Intervals | |
| | | | 5% | median | 95% | |
| Single Men | | | | | | |
| Revenue | | | | | | |
| Income Tax | 12667.8 | 352.2 | 389.7 | 353.8 | 390.9 | 421.1 |
| Medicare | 863.1 | 35.0 | 43.6 | 40.2 | 43.6 | 47.1 |
| Total Revenue | 13530.9 | 387.1 | 433.3 | 394.2 | 434.4 | 468.1 |
| Expenditure | | | | | | |
| Tax Rebates | 436 | -8.7 | -24.9 | -29.9 | -24.9 | -20.5 |
| Allowances | 3564.5 | 1266.8 | 1108.2 | 1053.5 | 1107.8 | 1158.7 |
| Pensions | 3336.7 | 142.7 | 134.3 | 130.5 | 134.2 | 137.7 |
| Pharm Allow | 58.2 | 3.1 | 3.2 | 3.1 | 3.2 | 3.2 |
| Rent Allow | 356.3 | 333.2 | 344.0 | 342.0 | 343.7 | 346.9 |
| Total Expenditure | 7751.7 | 1737.1 | 1564.7 | 1506.1 | 1564.3 | 1619.7 |
| Net Expenditure | -5779.2 | 1349.9 | 1131.3 | 1036.3 | 1130.5 | 1224.3 |
| Single Women | | | | | | |
| Revenue | | | | | | |
| Income Tax | 7192.9 | 329.3 | 267.4 | 242.3 | 267.5 | 291.8 |
| Medicare | 473.9 | 32.0 | 36.0 | 34.0 | 36.1 | 38.1 |
| Total Revenue | 7666.7 | 361.3 | 303.4 | 276.8 | 303.8 | 329.8 |
| Expenditure | | | | | | |
| Tax Rebates | 707.8 | -23.4 | -36.0 | -38.5 | -36.0 | -33.3 |
| Allowances | 2264.8 | 1051.4 | 1023.7 | 995.3 | 1022.5 | 1053.6 |
| Pensions | 7456.4 | 236.6 | 203.9 | 196.4 | 204.5 | 209.5 |
| Pharm Allow | 117.2 | 4.6 | 4.6 | 4.6 | 4.6 | 4.6 |
| Rent Allow | 289.3 | 261.0 | 278.2 | 275.3 | 278.2 | 281.0 |
| Total Expenditure | 10835.4 | 1530.2 | 1474.5 | 1442.1 | 1473.8 | 1508.5 |
| Net Expenditure | 3168.7 | 1169.0 | 1171.1 | 1112.5 | 1168.3 | 1229.7 |

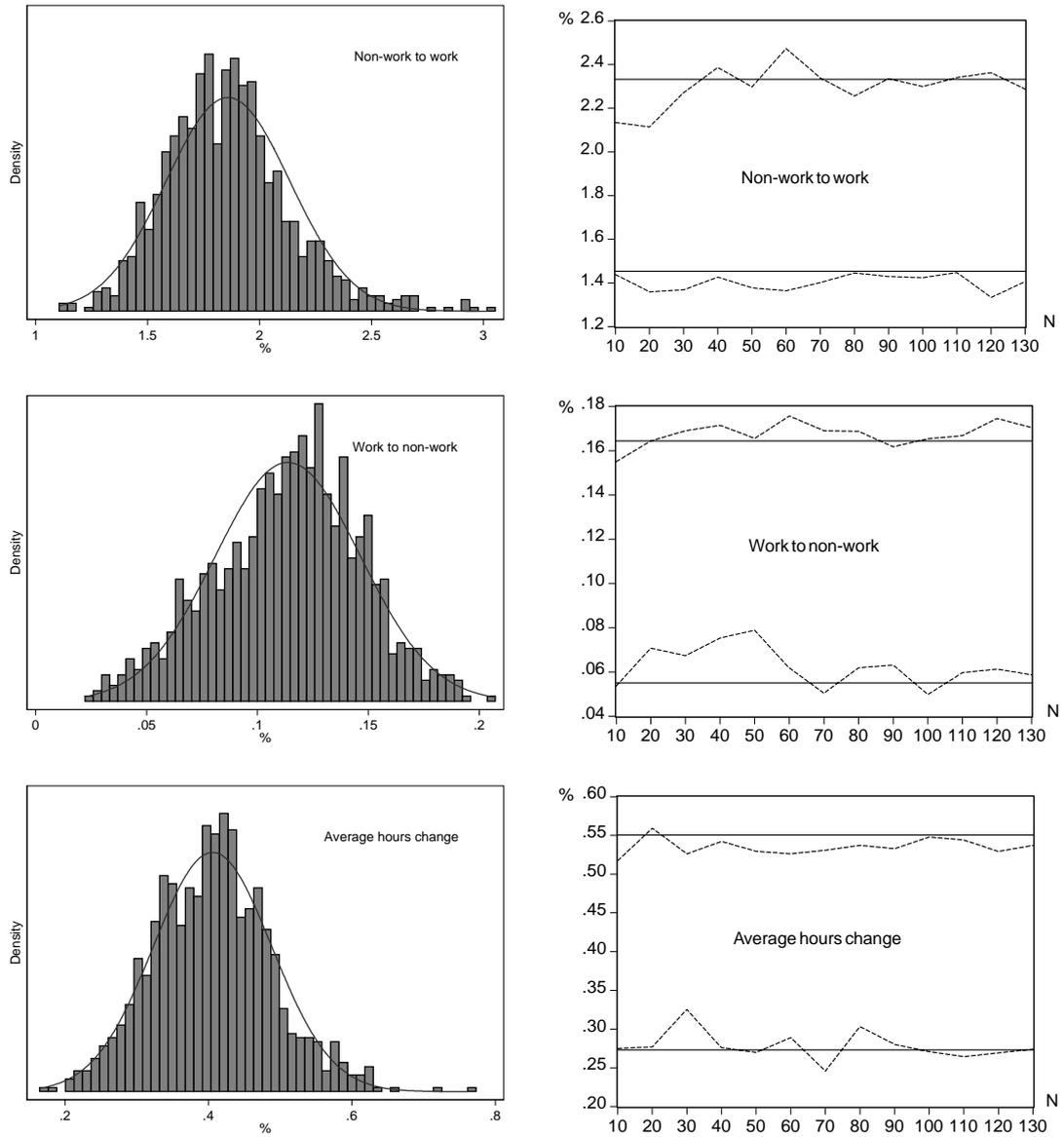


Figure 6: Single Men: Histograms of Labour Supply Simulation Outcomes and Confidence Intervals by Number of Replications (N)

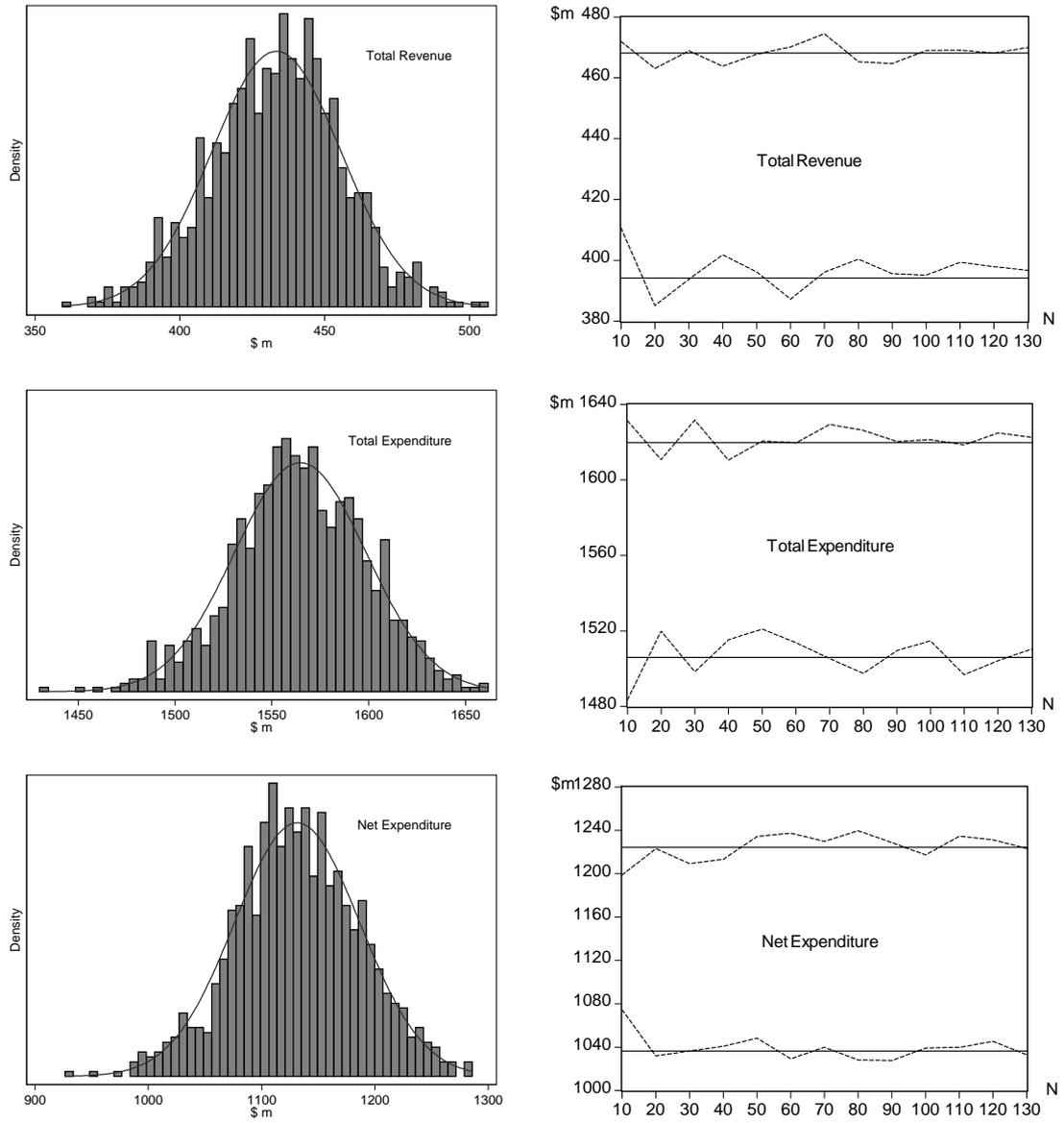


Figure 7: Single Men: Histograms of Revenue and Expenditure Simulation Outcomes and Confidence Intervals by Number of Replications (N)

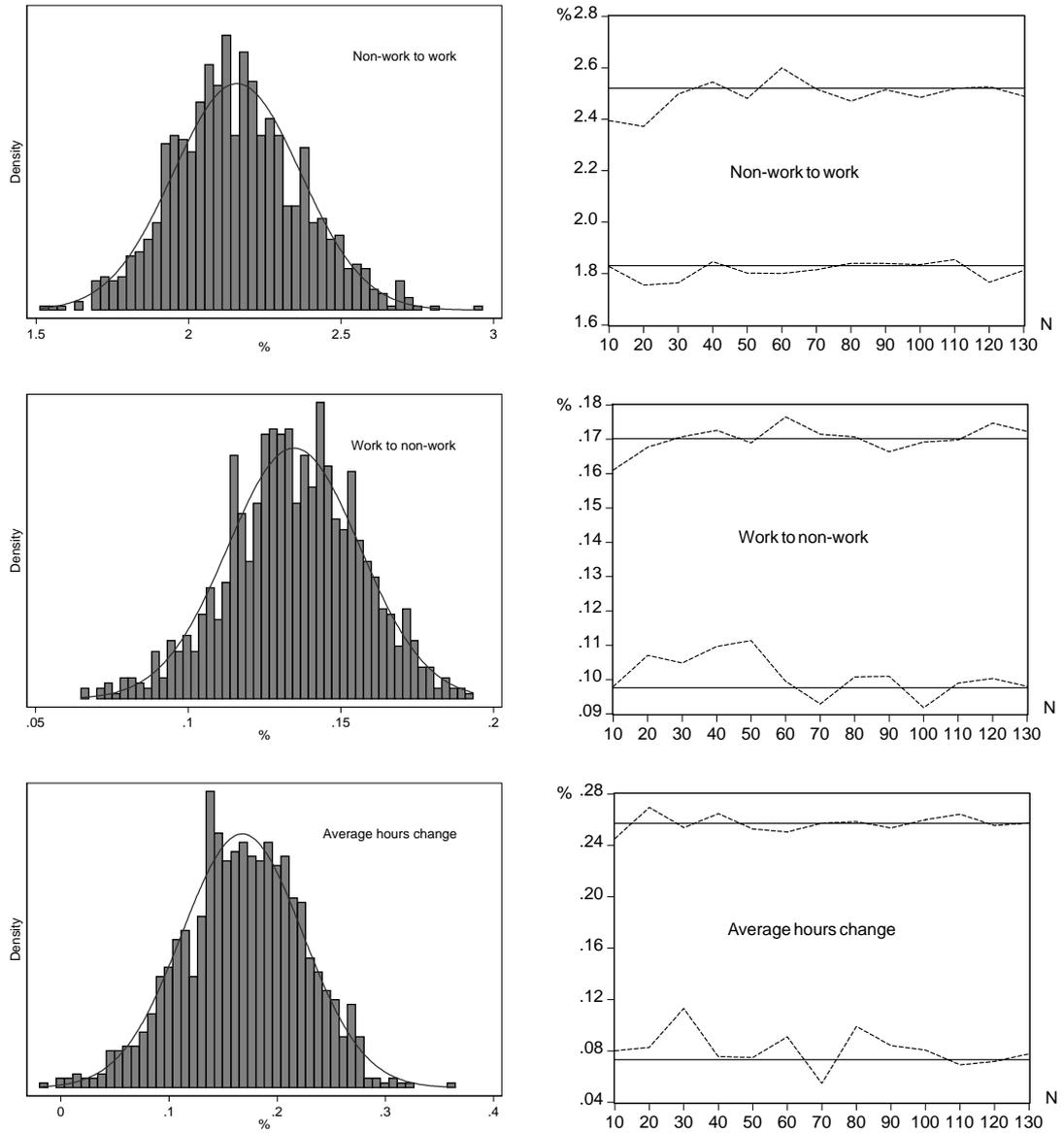


Figure 8: Single Women: Histograms of Labour Supply Simulation Outcomes and Confidence Intervals by Number of Replications (N)

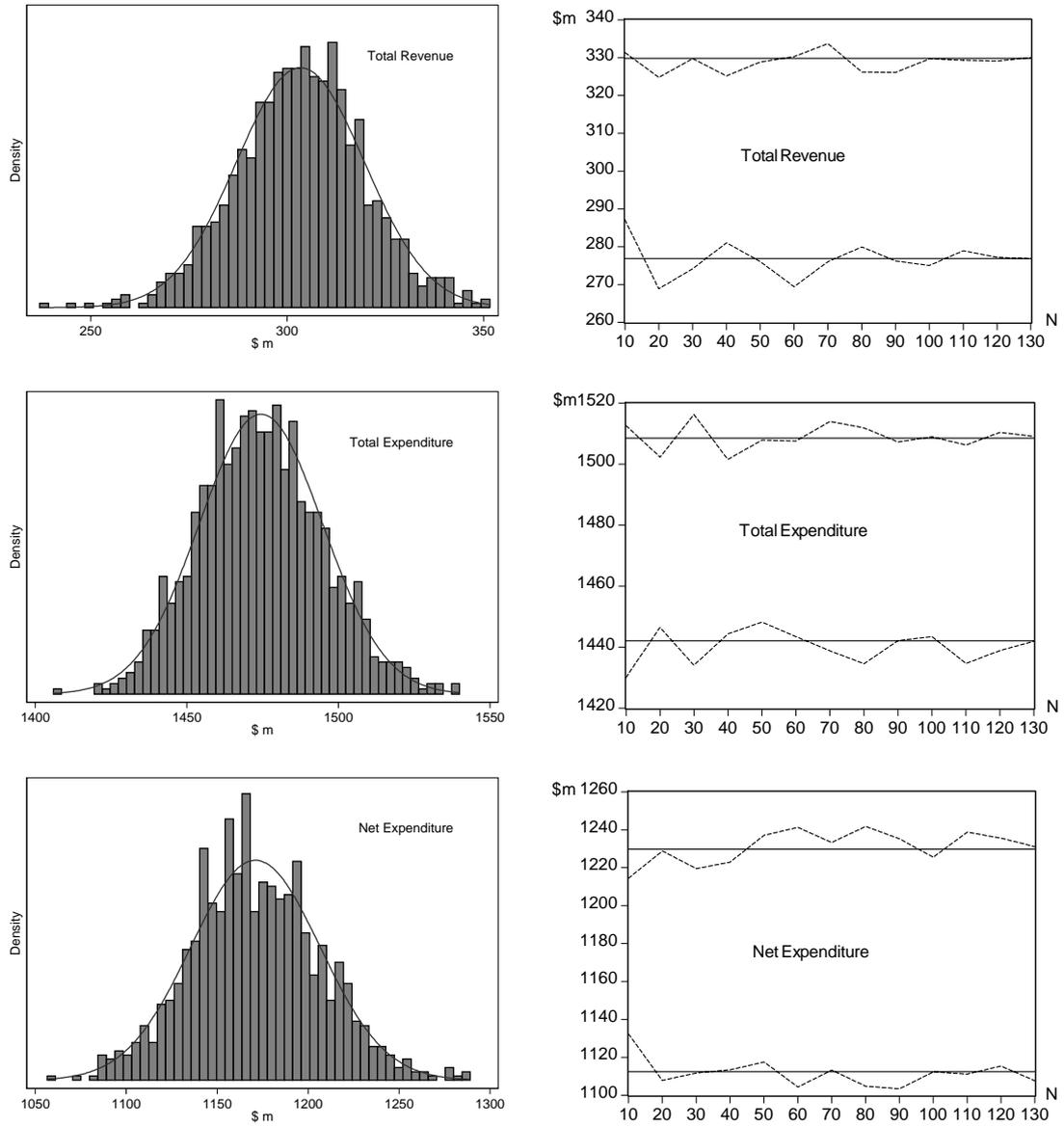


Figure 9: Single Women: Histograms of Revenue and Expenditure Simulation Outcomes and Confidence Intervals by Number of Replications (N)

Table 7: Simulated Labour Supply Responses to a Reduction in Taper Rates

| | Confidence Intervals | | | |
|---------------------------------------|----------------------|-------|--------|-------|
| | mean | 5% | median | 95% |
| Married men without children | | | | |
| all workers: base | 57.17 | | | |
| wage/sal. workers: base | 45.24 | | | |
| wage/sal. workers: post | 45.58 | 45.43 | 45.58 | 45.74 |
| non-work to work | 1.11 | 0.99 | 1.11 | 1.22 |
| work to non-work | 0.76 | 0.66 | 0.76 | 0.86 |
| workers working more | 0.22 | 0.15 | 0.21 | 0.28 |
| workers working less | 0.79 | 0.72 | 0.79 | 0.85 |
| average hours change | 0.03 | -0.03 | 0.03 | 0.10 |
| Married women without children | | | | |
| all workers: base | 46.63 | | | |
| wage/sal. workers: base | 39.79 | | | |
| wage/sal. workers: post | 39.22 | 39.05 | 39.22 | 39.41 |
| non-work to work | 0.57 | 0.46 | 0.56 | 0.68 |
| work to non-work | 1.13 | 0.98 | 1.13 | 1.27 |
| workers working more | 0.24 | 0.17 | 0.23 | 0.33 |
| workers working less | 0.73 | 0.66 | 0.73 | 0.79 |
| average hours change | -0.34 | -0.40 | -0.34 | -0.27 |

Table 8: Change in Tax and Transfer Costs Resulting from Reduced Tapers

| Couples without children | Pre-Reform (\$m) | Changes from Pre- to Post-Reform | | | | |
|--------------------------|------------------|----------------------------------|------------------------|----------------------|--------|--------|
| | | Fixed Labour (\$m) | Variable Labour Supply | | | |
| | | | Mean (\$m) | Confidence Intervals | | |
| | | | 5% | median | 95% | |
| Revenue | | | | | | |
| Income Tax | 21001.1 | 527.3 | 337.6 | 308.0 | 338.2 | 367.7 |
| Medicare | 1327.4 | 58.0 | 48.4 | 46.3 | 48.4 | 50.6 |
| Total Revenue | 22328.5 | 585.3 | 386 | 354.2 | 386.7 | 418.2 |
| Expenditure | | | | | | |
| Tax Rebates | 1327.8 | -194.0 | -193.9 | -196.3 | -193.9 | -191.7 |
| Allowances | 2612.2 | 2403.6 | 2495.5 | 2463.6 | 2496.4 | 2524.8 |
| Pensions | 10151.2 | 771.5 | 768.5 | 762.4 | 769.1 | 772.4 |
| Pharm Allow | 109.1 | 10.1 | 10.2 | 10.1 | 10.2 | 10.4 |
| Rent Allow | 170.8 | 127.0 | 141.7 | 139.7 | 141.8 | 143.4 |
| Total Expenditure | 14371.0 | 3118.3 | 3222.0 | 3187.7 | 3222.5 | 3254.7 |
| Net Expenditure | -7957.5 | 2533.0 | 2836.1 | 2770.8 | 2835.8 | 2897.9 |

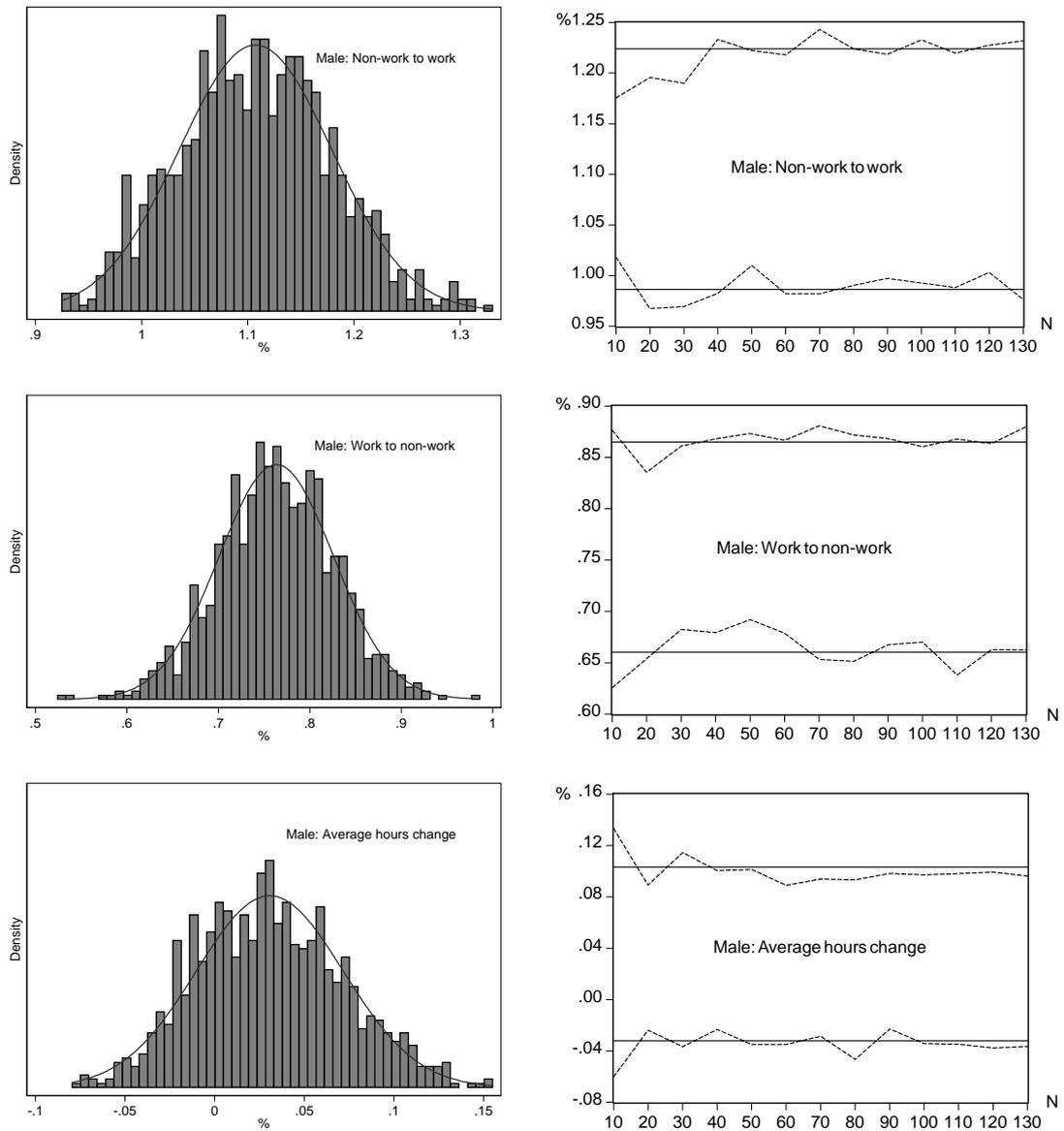


Figure 10: Couples without children: Histograms of Male Labour Supply Simulation Outcomes and Confidence Intervals by Number of Replications (N)

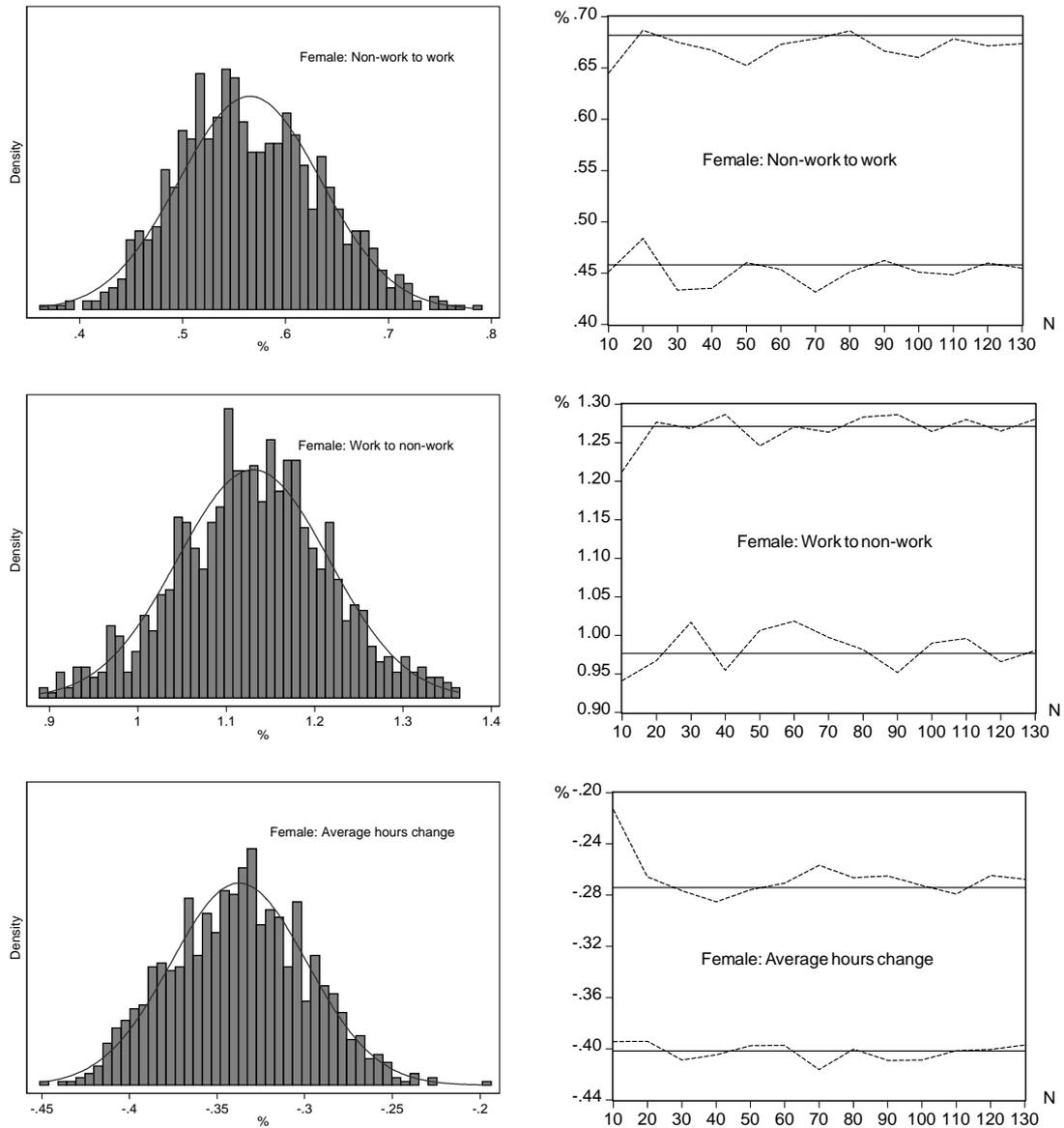


Figure 11: Couples without children: Histograms of Female Labour Supply Simulation Outcomes and Confidence Intervals by Number of Replications (N)

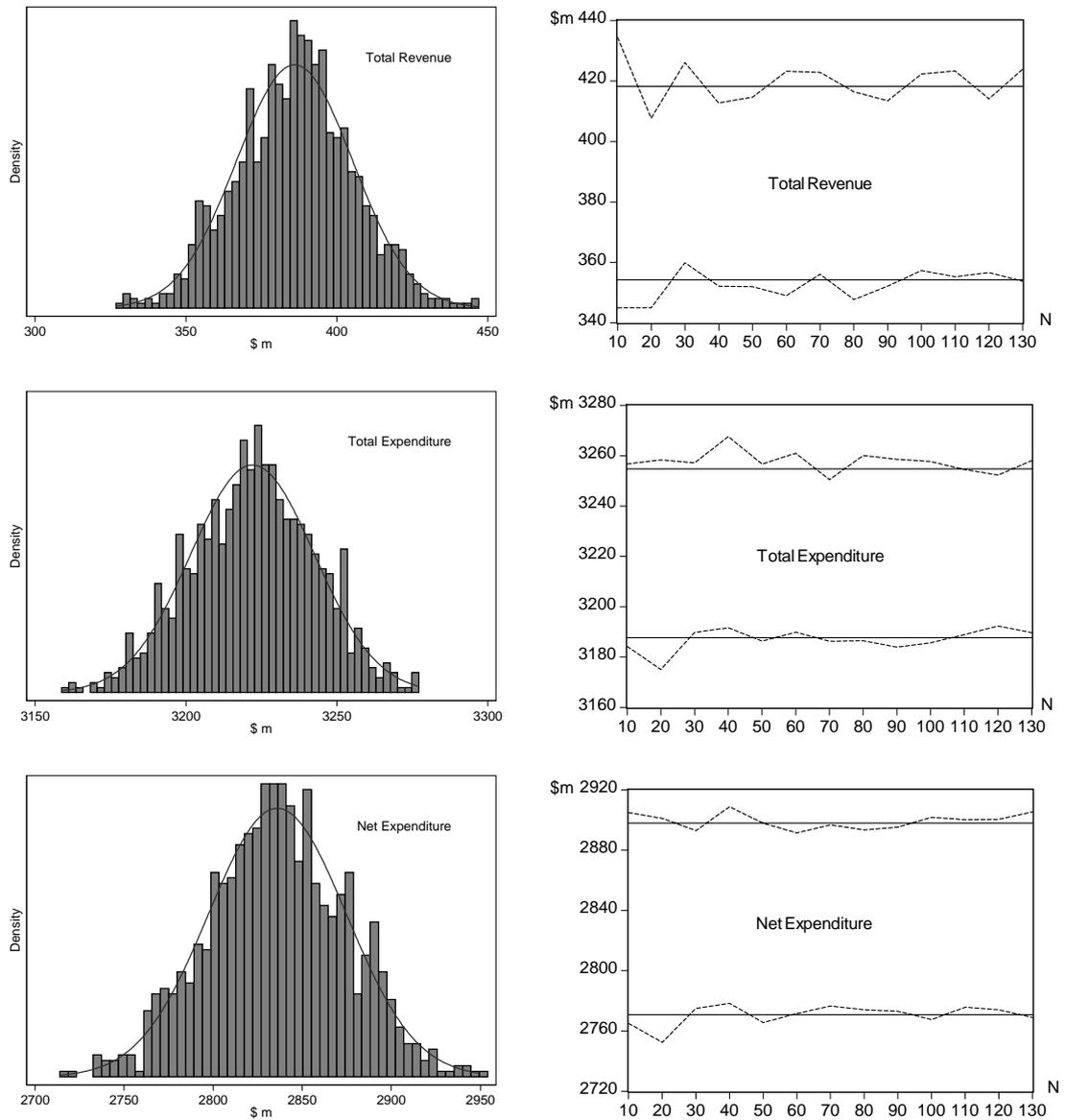


Figure 12: Couples Without Children: Histograms of Revenue and Expenditure Simulation Outcomes and Confidence Intervals by Number of Replications (N)

References

- [1] Creedy, J., A.S. Duncan, M. Harris, R. Scutella (2002) *Microsimulation Modelling of Taxation and The Labour Market: The Melbourne Institute Tax and Transfer Simulator*. Cheltenham: Edward Elgar.
- [2] Creedy, J. and Kalb, G. (2004) Discrete hours labour supply modelling: specification, estimation and simulation. *Journal of Economic Surveys* (forthcoming).
- [3] Creedy, J., Kalb, G. and H. Kew (2003) Flattening the effective marginal tax rate structure in Australia: policy simulations using the Melbourne Institute Tax and Transfer Simulator. *Australian Economic Review*, 36, pp. 156-172.
- [4] Deville, J.-F. and Särndal, C.-E. (1992) Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, pp. 376-382.
- [5] Kalb, G. (2002) Estimation of Labour Supply Models for Four Separate Groups in the Australian Population. Melbourne Institute Working Paper Series, WP No. 24/02, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.
- [6] Keane, M. and R. Moffitt (1998) A structural model of multiple welfare program participation and labor supply. *International Economic Review*, 39, pp. 553-589.
- [7] Klevmarcken (1998) Statistical inference in microsimulation models: Incorporating external information. Uppsala University Department of Economics Working paper, 1998:20.
- [8] Maddala, G.S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.
- [9] Pudney, S. and Sutherland, H. (1994) How reliable are microsimulation results? *Journal of Public Economics*, 53, pp. 327-365.

- [10] Pudney, S. and Sutherland, H. (1996) Statistical reliability in microsimulation models with econometrically-estimated behavioural responses. In *Microsimulation and Public Policy* (ed. by A. Harding), pp. 473-503. Amsterdam: North-Holland.
- [11] Van Soest, A. (1995) Structural models of family labor supply; a discrete choice approach. *Journal of Human Resources*, 30, pp. 63-88.