Developing a microsimulation model to forecast Australian company tax revenue.

Jun Bai and Michael Zeitlhofer
Modelling Development Unit, Australian Taxation Office

Abstract The Australian Taxation Office (ATO) is developing a new modelling approach to forecast company tax revenue which utilises microsimulation techniques. This involves predicting key variables for individual companies, calculating their resulting tax liability and aggregating the outcomes for the full population. These forecasts are required for a period of up to 5 years out for Federal government budgetary purposes. Although the model derives forecasts for individual company records, these are based on economic and other forecast parameters derived from macroeconomic models at the broad industry group level. This ensures the model conforms to overall economic conditions as forecast by the Treasury department. Adopting a microsimulation approach for volatile forecasts such as company profits and taxation involves a great deal of complexity. This paper gives an overview of some of the approaches taken in our attempt to develop a microsimulation model which produces reliable forecasts in a timely and user friendly manner.

1. INTRODUCTION

1.1 What is a microsimulation model?

Microsimulation involves simulating changes in individual unit records and aggregating the results for the full population. In our model, this involves simulating changes in a company's tax liability by forecasting their profits (or losses) and taxation reconciliation adjustments. The results are then aggregated across the population to arrive at an overall revenue forecast. As mentioned in the abstract, the model is designed to conform to a number of macro based targets, ensuring the model produces reliable simulations.

1.2 Why adopt a microsimulation approach?

At present, the ATO relies mainly on a macro based approach to forecast company tax revenue. This approach has the advantage of not being overly complicated and can yield good results given reliable input parameters and a fairly stable environment. The problem using a macro approach for companies is that often the environment is not particularly stable, with profitability being quite volatile and the presence of a number of very large companies that have a significant impact on overall revenue. It is difficult to isolate errors when they occur, and there are only minimal means of segmenting the population into industry based or other profiles. In addition, it is difficult to model the impact on revenue of losses recouped from previous years or loss transfers between entities within a corporate group.

Microsimulation attempts to overcome some of these difficulties by more realistically simulating this volatility at the company level. Some companies can be given large profits, some large losses, some small profits and some small losses etc. In addition, the results can be aggregated and profiled into any number of groups (provided that a critical mass is present and the group can be identified from the data).

Our colleagues in Inland Revenue, United Kingdom, have been leaders in the use of microsimulation for company tax forecasting for many years. The accuracy of their forecasting has improved with the use and refinement of microsimulation modelling.

There are a number of disadvantages in using a microsimulation approach. The development and maintenance of these models is resource intensive, both in terms of people and computing power. The models are reliant on the availability of relevant and accurate unit record data and forecast parameters. Realistic simulations often require complex algorithms, a range of user inputs and an experienced operator who is able to digest and validate model results. All of which present risks to the production of accurate and reliable forecasts.

2. MODEL METHODOLOGY

At present, the model has 3 main modules. The first involves forecasting net profits (and losses) for each company, before taking into account tax related adjustments. This is referred to as the Gross Surplus (GSP) Module. The second module forecasts tax specific adjustments that can be made to a company's net profit. These adjustments include loss transfers within a corporate group, prior year losses recouped and other adjustments relating to items such as accelerated depreciation, capital gains and research and development incentives. This module is referred to as the Tax Reconciliation Module. The final component involves profiling the forecast tax liability into an expected cash payment schedule based on the company instalment system. This is called the Cash Collections Module. Each of these modules is discussed below, with the major emphasis being on the first two modules.

2.1 Module 1 – The Gross Surplus (GSP) Module

The first step in the modelling process involves forecasting the level of profit (or loss) for each company. This profit (or loss) is the economic or accounting profit (or loss) generated by a company before any taxation specific adjustments are made (these adjustments are discussed later under Module 2). We refer to this as the Gross Surplus or GSP.

For the model to produce realistic results, the aggregated profit information has to conform to overall economic conditions as forecast by the Treasury department. Therefore, a number of macro inputs are required, including a parameter for the estimated growth in company profits. The model's task is to take this growth parameter and distribute it to individual records based on a range of other inputs determined by the user.

It is not appropriate to simply apply the same growth parameter to all companies, as this would yield similar results as a macro model, and is not a realistic simulation of company behavior. Rather, we need to have some means of simulating the volatility in company profits, while still conforming to our overall macro growth target.

The methodology to simulate this volatility relies on the use of transitional probabilities. This concept is demonstrated with reference to *Figure 1*. Year 1 and Year 2 relate to historical information, while Year 3 is the year being forecast. In Year 1, a company could have had either a profit ($Profit_1$) or a loss ($Loss_1$). If a company was in profit in Year 1, in Year 2 that company could have *a)* increased their profit (P_1P_2+), *b)* decreased their profit (P_1P_2-) or *c)* moved to a loss (P_1L_2). Probabilities are then used to determine in which direction the profit (or loss) will move in the forecast year. Based on the probabilities in *Figure 1*, a company which was in profit in year 1, and had a bigger profit in year 2 (P_1P_2+) has 42% chance of having a bigger profit in year 3, a 46% chance of having a smaller profit in year 3 and a 12% chance of making a loss in year 3.

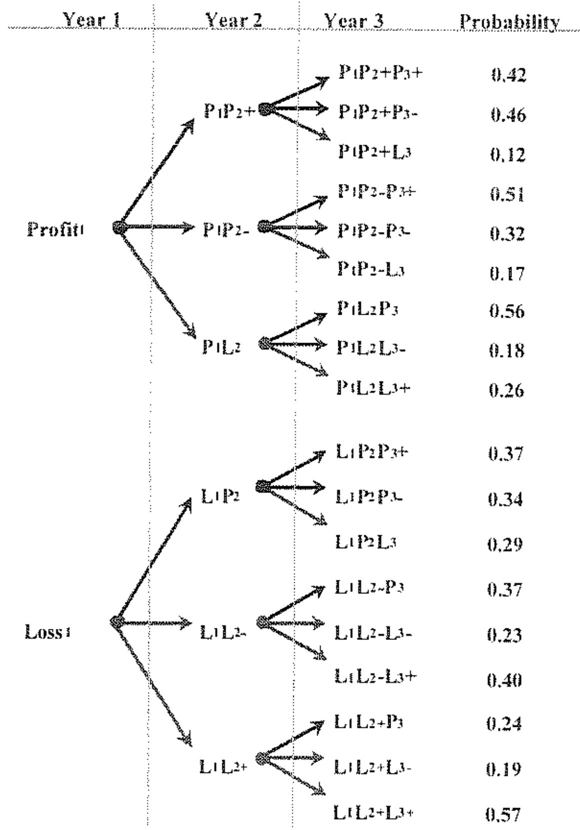
The user has responsibility for determining probabilities which are in line with forecast economic conditions and which are consistent with the base data and other inputs. For example, during periods of strong economic growth, one might expect more companies to increase their profits and fewer companies to increase their losses or move from profit to loss. To assist with determining appropriate probabilities, we have created parameter files that contain probabilities based on historical data dating back to 1990.

The change in profits (losses) from year 2 to year 3 can be summarised into 6 different Profit Groups. These groups relate to the profit transition in the base year (Year 2 on *Figure 1*) to the forecast year (Year 3 on *Figure 1*).

The 6 Profit Groups are:

Group 1: Profit in Year 2 to bigger profit in Year 3 (P_2P_3+)
Group 2: Profit in Year 2 to smaller profit in Year 3 (P_2P_3-)
Group 3: Profit in Year 2 to loss in Year 3 (P_2L_3)
Group 4: Loss in Year 2 to smaller loss in Year 3 (L_2L_3-)
Group 5: Loss in Year 2 to bigger loss in Year 3 (L_2L_3+)
Group 6: Loss in Year 2 to profit in Year 3 (L_2P_3)

Figure 1 Transitional Probability combinations



Random numbers are used to determine which companies are allocated to a particular group. Therefore, it is important for the user to note the random number seed of a particular model run if they want to rerun the model with the same allocation of records.

The user also needs to be aware of the effect of the random allocations on some of the largest companies. In certain industries, a small number of large companies make up the majority of profits. Under certain modelling conditions, some of these companies may end up with an unrealistic profit transition. This can have the effect of distorting the growth rates across the industry, leaving the user with poor results. In this situation, the user has the option of removing the random allocation of selected large companies and selecting the appropriate profit group for the company manually.

Once companies have been allocated to one of the profit groups, the next step is to calculate the growth rate to be applied to each company. Different growth rates are calculated for each of the 6 profit groups, with each company within a profit group receiving the same growth rate. In order for the model to calculate these growth rates, three macro inputs are required:

a) Macro growth forecast
b) Profit to loss ratio
c) Profit and loss distributions

These inputs are defined below by way of example.

a) *Macro growth forecast*

This is the macro growth target for net profit across the population being modelled.

Net Profit in Year 2	= $180m
Macro growth forecast (@ 11%)	= $ 20m
Forecast Net Profit in Year 3	= $200m

b) Profit to loss ratio

The profit figures used above are net figures. That is, they are the sum of all profits (positive values) and losses (negative values). The ratio of the sum of the profits to the sum of the losses is the profit to loss ratio. If we use a profit to loss ratio of 3:1 on the above example, we can determine the macro target for gross profits and gross losses.

Gross Profit in Year 3	= $300m
Gross Loss in Year 3	= -$100m
Forecast Net Profit in Year 3	= $200m

c) Profit and loss distributions

The gross profits and gross losses need to be split further to set the macro targets for each of the 6 profit groups. For example, the gross profit target of $300m needs to be distributed between the 3 profit groups P_2P_3+, P_2P_3- and L_2P_3 *(Figure 2)*. Similarly, the gross loss target of -$100m needs to be distributed between L_2L_3-, L_2L_3+ and P_2L_3 *(Figure 3)*. The distributions are calculated by applying a percentage to the gross profit and gross loss figures.

Figure 2 Profit Distributions

Profit Group	Distribution	Profit Amount
P_2P_3+	36.7%	110m
P_2P_3-	38.3%	115m
L_2P_3	25.0%	75m
	100%	300m

Figure 3 Loss Distributions

Profit Group	Distribution	Loss Amount
L_2L_3-	60.0%	-60m
L_2L_3+	15.0%	-15m
P_2L_3	25.0%	-25m
	100%	-100m

Given the macro inputs outlined above and the allocations performed through the use of transitional probabilities, the model then calculates the growth rates to be applied to each company record. An example of this is shown at *Figure 4*. It can be seen that by using the growth from Year 2 to Year 3, the model achieves the forecast macro growth rate of 11% (180m to 200m), the profit to loss ratio of 3:1 and the desired profit and loss distributions. Also, each profit group has a growth that is consistent. That is, P_2P_3+ shows an increase in profits, P_2P_3- shows a decrease in profits etc. The growth rate to be applied to each record is calculated by dividing Year 3 by Year 2. When developing a model run, the user needs to ensure the various inputs are consistent with the underlying data so that reliable results are achieved.

It should also be noted that the forecast growth rates would change if a different random number seed were used on different model runs. This is due to different companies being allocated to each of the 6 profit groups.

Figure 4 Forecast growth for each Profit Group

Profit Group	Year 2 $	Year 3 $	Growth $	Growth Rate
Gross Profit				
P_2P_3+	90m	110m	20m	1.22
P_2P_3-	150m	115m	-35m	0.77
L_2P_3	-10m	75m	85m	-7.50
Gross Loss				
L_2L_3-	-80m	-60m	20m	0.75
L_2L_3+	-5m	-15m	-10m	3.00
P_2L_3	35m	-25m	-60m	-0.71
TOTAL	**180m**	**200m**	**20m**	**1.11**

Although the methodology discussed above only refers to 'one' population of records, the process is undertaken for 7 separate industry groupings. Therefore, a separate set of macro growth targets, profit to loss ratios, profit and loss distributions, and transitional probabilities are required for each industry group. The model calculates growth rates for each of the 6 profit groups within each industry, resulting in 42 different growth rates being used. The 7 industry groups are:

1. Banking & Finance
2. Manufacturing
3. Mining
4. Services & Other
5. Insurance & Superannuation
6. Media & Communications
7. Property & Construction

2.2 Module 2 – The Tax Reconciliation Module

Once a company's profit (or loss) has been forecast in module 1, further adjustments are made to certain records to reflect tax effect accounting. These adjustments are made before the final tax liability can be calculated and are made to take account of special treatments of certain types of income and expenses available under taxation laws. Examples of these include prior year loss recoupment, loss transfers within corporate groups, research & development concessions, accelerated depreciation and capital gains.

These adjustments are difficult to forecast as they involve a range of issues and often depend on the particular circumstances of an individual company. For example, the decision to sell or purchase a capital asset which is subject to capital gains or accelerated depreciation concessions may be made purely on company cashflows rather than any general economic indicator.

2.3 Module 3 – The Cash Collections Module

The third module takes the tax liability information calculated in Module 2 and profiles it into estimated monthly cash collections. At present, this is based on a quarterly taxation instalment regime that is applicable in Australia. However, the instalment system used for companies in Australia is being changed to a Pay As You Go System, which will mean changes need to be made to this module.

At present, the module produces a summary sheet detailing due payments for each month based on the aggregated source of the payments. The source information includes the summary size group (large, medium, and small) and the accounting balance period (e.g. June balancing, March balancing). It is not proposed to provide any more detail on this module in this paper.

3. COMPUTER SYSTEM AND DATABASE

The model is being developed using PC SAS in a Windows NT LAN environment. Graphical user interfacing is being used, giving the model the look and feel of a professional Windows application. The current server is a Compaq 5500 with quad Pentium II 400 Xeon processors, 1gb RAM and approx. 200gb of disk capacity. The Workstations are Pentium based running as part of the ATO LAN environment. To make use of the processing power of the server, model runs are developed at the workstation and submitted for processing on the server. Client server products from Argent are used which allow full batching and scheduling facilities from the workstation to the server, including priority queues, overnight queues and job dependencies. Our modelling activities earned the SAS Institute Australia's '1998 Innovation Award'.

Any microsimulation model is reliant on unit record data. Our model uses company tax return form data that is captured on the ATO corporate systems. All companies in Australia are required to lodge annual tax returns that provide a range of income, expense and statistical data items. We have historical databases dating back to the 1989/90 income year through to the 1997/98 income year, with each database holding 450,000 to 550,000 records. As a new year dataset becomes available (i.e. end of processing year), this dataset is downloaded from our mainframe system onto our LAN server environment. As with a number of statistical datasets, a deal of effort is undertaken to ensure the data is of the highest quality possible.

Although the full amount of information on these datasets is used for analysis, not all of it is used for the model. New datasets are created with less information, and with records matched over a 2 year period. That is, where information for a particular company record is not held for two consecutive years, that record is not used in a model run. This is due to the profit forecasting methodology using transition probabilities, which requires historical profit data for 2 consecutive years. No further sampling is done of the datasets, so a typical model dataset holds approximately 500,000 records. A 'grossing up' algorithm is being developed which takes account of any shortfall in the number of companies in our model database due to the 2 year matching of records.

4. CONCLUSION

In Australia, companies contribute approximately $20 billion dollars of revenue each year in income tax. This is a significant amount of revenue, so the government needs the best possible forecasts to facilitate budgetary planning. The development of a microsimulation model for this purpose is consistent with the ATO's commitment to world best practice in taxation administration.

This development however, has not been without its problems. Work commenced on the model in 1996, and the model is yet to be used as a production forecasting tool. A number of delays have been experienced, in particular around difficulties with the methodology, data quality and infrastructure (both IT and staffing). However, significant progress has been made, and we are now well placed to have one of the leading company tax forecasting models in the world.

Our efforts over the coming year will focus on testing the sensitivities and reliabilities of the model results, deriving reliable input parameters, user interfacing, data quality and output reporting. The complex nature of microsimulation modelling, and the profit and tax volatility of the corporate population make the process a slow and difficult one.

At the time of writing, Modules 1 & 3 have been coded into the model, and Module 2 is nearing completion. As such, minimal information has been provided in this paper about the forecasting accuracy of the model methodology being used. However, results from preliminary testing using historical data have proven positive. On presentation of this paper, more information should be available regarding a number of these issues.

The authors would be happy to provide further information about the Australian Company Income Tax Model should people be interested. They would also welcome any comments or suggestions. They can be contacted at:

Modelling Development Unit
Revenue Analysis Branch
Australian Taxation Office
PO Box 900
Civic Square
ACT 2608
AUSTRALIA

Email: jun.bai@ato.gov.au
michael.zeitlhofer@ato.gov.au