Statistics New Zealand
TATAURANGA AOTEAROA

An optimisation approach for updating product data in supply and use tables

Paper presented at the18th International Input-Output Conference,
Sydney, Australia,
25 June 2010

Jeroen Kole

Senior Research Statistician, National Accounts, Statistics New Zealand
P O Box 2922
Wellington, New Zealand
info@stats.govt.nz

www.stats.govt.nz

Abstract

This paper presents an optimisation approach for updating product breakdowns in a supply and use balancing process. Statistics New Zealand applies a supply and use framework in the National Accounts system to balance the annual current price production and expenditure estimates of gross domestic product. A new range of specific data collections has provided an update of input and output product breakdowns by industry. A linear programming approach has successfully supported the comprehensive and controlled introduction of the new data in the balancing process.

1 Introduction

This paper presents a linear optimisation model for supporting the large-scale update of product breakdowns in the supply and use framework in the New Zealand System of National Accounts (NZSNA).

The NZSNA is a national accounting framework based on an international standard, the System of National Accounts (Inter-Secretariat Working Group on National Accounts, 1993). According to this international standard, Statistics New Zealand applies a supply and use framework to confront and reconcile the annual current price production and expenditure estimates of gross domestic product (GDP). This framework is a powerful statistical and analytical tool used to balance the flow of goods and services (products) in the New Zealand economy. The accounts are balanced when for all industries total inputs equal total outputs, and for products when total supply equals total demand. As a result, the statistical discrepancy between the current price measures of GDP is zero in the years for which balancing has been carried out.

The supply and use framework improves the quality of both the current and constant price GDP series through data confrontation, and also provides a wealth of additional information. The balanced accounts provide a benchmark for the annual national accounts in following years, for quarterly estimates of GDP, and for other publications like the Tourism Satellite Account.

The framework requires detailed product breakdowns for inputs and outputs of industries. In the past in New Zealand, these were based on periodic industry censuses, and other sources. A change in the data collection strategy has resulted in detailed product data by industry being collected over the period 2003–2008. Some of the newly available product data has been used in the manual balancing process for the annual accounts of the 2003–2005 March years. The full set of new product data has been introduced in the balancing process for the accounts of the 2006 and 2007 March years.[1] Given time and resource constraints, the challenge was to find an effective and efficient approach to make the data fit-for-purpose for the 2006 and 2007 March years and remove inconsistencies.

A variety of techniques for updating and balancing input-output tables are discussed in the input-output literature. Two prominent numerical approaches are so-called RAS bi-proportional scaling methods and mathematical optimisation methods.[2] Constrained optimisation methods can handle inconsistent and conflicting external data.

---

[1] Annual accounts for the 2006 and 2007 March years have been published in November 2009. Detailed supply and use tables are not published annually.

[2] The RAS procedure (RAS) is named after the typical sequence of matrices. The matrix to be updated (A) is pre-multiplied with a diagonal matrix of row factors of correction (R) and post-multiplied with a diagonal matrix of column factors of correction (S) (European Commission, 2008, p451).

.

Recently a RAS-type method (KRAS) has been developed which is also able to find compromise solutions for conflicting external data (Lenzen, Gallego, & Wood, 2009).

Statistics NZ applies a modified RAS procedure in the end phase of the supply and use balancing process. However, to support the large-scale introduction of new and potentially conflicting product data in an earlier phase of the balancing process, we have developed and implemented a specific optimisation approach. The resulting balancing process is a combination of automated procedures and manual interventions.

In section 2 we introduce a simplified supply and use framework, followed by an overview of the balancing process in the NZSNA. The change in the data collection strategy is described in section 3. We then present the optimisation model (section 4), the integration in the balancing process including some results (section 5), and our conclusion (section 6).

2 The supply and use framework and the balancing process

For the purpose of this paper it is sufficient to introduce a simplified supply and use framework (section 2.1). We then give a general overview of the supply and use balancing process in the NZSNA (section 2.2).

2.1 A simplified supply and use framework

The supply and use tables are a matrix representation of industry production accounts, goods and services accounts and generation of income accounts. A simplified supply and use framework is presented in the figure below.[3]

Figure 1: A simplified supply and use framework (*)

<u>Supply</u>

	Output by industries	*Imports*
Output by products	Goods and services produced by each domestic industry	Product composition of imports

<u>Use</u>

	Industry purchases	*Final demands*
Products consumed	Intermediate consumption by each industry to produce their own output	Value of products going to final demand categories
Primary inputs	Value added components of each industry	

(*) Matrix representation with row and column headers in *italics font*.

Total supply consists of domestic supply and imports. In figure 1, the supply table shows product groups produced by each domestic industry. The diversity of outputs in many

[3] For a more detailed description of a supply and use framework, see for example (European Commission, 2008).

industries (the industry product mix) is expressed in the columns. The rows represent the range of industries producing each product. The imports column in the supply table contains the product composition of imported goods and services.

The use table rows present the disposition of locally produced and imported products between industries' use as intermediate consumption and by final demand categories (e.g. final consumption, capital formation, exports). The columns show the composition of inputs by each industry, that is intermediate consumption and primary inputs (value added components like compensation of employees, operating surplus etc). Also included in the use table is the product composition of final demand categories.

In a balanced system, total inputs equal total outputs for each industry. Value added for each industry is the difference between domestic output and intermediate use of that industry. The supply and use of a product are balanced when domestic supply plus imports equal intermediate use plus final use of this product.

The working level in the Statistics NZ supply and use framework contains about 115 industries and 320 products.[4]

2.2 The balancing process

In the NZSNA, the process of reconciling the current price annual accounts has four stages:

- Industry and product analysis
- Manual supply and use balancing
- Automated balancing
- Final balancing

In preparing the initial industry production accounts, survey and other collected data are analysed and evaluated against alternative data sources, previous years' data, and real world events. Adjustments are made if the relevant industry totals are inconsistent with other information considered to be more reliable. An industry production account must always balance such that the sum of intermediate consumption and value added equals output. At this stage in the balancing process, the focus is on production account totals.

Similar to the production accounts for industries, totals for final demand categories are estimated and product analysis then compiled.

Itemisations of industry and final demand totals are reordered into product accounts. For the initial industry product breakdowns of domestic supply and intermediate use, proportions from the latest balanced year are applied.[5] The product accounts are unbalanced at this stage, as supply and use of products are not equal.

The imbalance may be the result of four possible factors:

- Inherent statistical differences in the data source relating to coverage, valuation, and timing. The differences occur because of the independence of estimates for production and final demand, using a variety of sources across different time periods in some cases.
- Inconsistent coding of products because of imprecise descriptions in data sources.
- Use of incorrect margins. The values in the product accounts are re-calculated in producers' prices, by removing Goods and Services Tax (GST) and retail and wholesale

[4] Industries are classified according to the Australian New Zealand Standard Industry Classification 1996 (ANZSIC 1996). Industries are further split into a sector of ownership (private, local or central government) and into market or non-market production group. For products, the classification being used is based on the Australian New Zealand Standard Commodity Classification 1996.

[5] Values reflecting these proportions are replaced with 'optimised' values from the linear program which we present in section 4.

margins. The sources for the margin proportions used in this re-calculation are distribution industry surveys and information from previous years' balancing process.[6]

- Sampling errors at all stages of data collection.

The manual balancing phase involves a variety of tasks and techniques: (i) analysts investigate data sources of supply and use for unbalanced products and their related products; (ii) new proportions for the input or output product mix of industries are calculated after investigating the balances of related products; and (iii) industry totals or final demand values are revisited and adjusted where necessary.

Manual balancing requires knowledge of the strengths and weaknesses of the data sources that feed the supply and use system. It is also an interactive (team) process, because balancing adjustments for one product can impact on the work done by other analysts. Priorities for the balancing work are based on targeted tolerances.

When all product imbalances are within acceptable limits the product and industry data undergo an automated balancing process, whereby all products are completely balanced. Although industry accounts may change marginally, they remain balanced throughout the process. The automated balancing process that is used is a modified RAS iteration procedure. Data that is considered of better quality can be 'locked'. The balancing process only affects the other data.

In the final balancing phase, analysis and checks on the full set of current price annual accounts could reveal a few remaining (consistency) problems, which are manually corrected.

In recent years until 2008, two years were balanced at a time, and every year was balanced twice. For example in the November 2008 National Accounts release, Statistics NZ published the final balanced accounts for the March year 2004, the (first) balanced accounts for 2005, and provisional (non-balanced and less detailed) results for the 2006–2008 March years. From 2009, timeliness was improved by balancing three years. In November 2009 we have published balanced current price accounts for 2005 (final balanced), 2006, and 2007.

3 Change in the data collection strategy

Since 1990/91, Statistics NZ has been using a supply and use framework annually. In essence, the levels and detailed product structure of the industry and expenditure accounts in the supply and use model were based on the latest inter-industry (input-output) study, updated where possible using more up-to-date data to reflect changes in the economy over time. This resulted in an annual series of balanced GDP accounts, based on the latest inter-industry benchmark. The data for product breakdowns was mainly collected via periodic industry censuses, the Annual Enterprise Survey, tax data, the Producers Price Index, and company reports. For general or composite items of expenditure, standard breakdowns were used. Although new product information was used when available, this was largely restricted to key industries when product changes were known to be occurring. By and large, the product breakdown for most industries was strongly influenced by the historical benchmark proportions.

After publication of the 1995/96 inter-industry study in 2001, a new input-output strategy was introduced. New and additional data normally incorporated in the periodic benchmark inter-industry study is now progressively introduced directly in the annual supply and use

[6] In the core system of NZSNA, balancing takes place on a producers' price basis. Published supply and use tables have a basic price valuation. The difference is formed by taxes less subsidies on products. For definitions of these valuation bases, see for example European Commission (2008, pp91–92).

framework. As a consequence, the annual supply and use framework will have a similar quality status as an existing (five-yearly) inter-industry benchmark.

The new approach drives, and is supported by, a revised economic data collection strategy. Periodic industry censuses have ceased. These have been replaced by an expanded Annual Enterprise Survey, which collects similar financial data, but in less detail. The product data previously collected in the industry censuses is now collected via a new range of specific collections. For this, the Commodity Data Collection (CDC) programme was established. The strategy has changed from data collection for periodical benchmarks about every five years to a rolling process of collecting and incorporating updated information.

The primary vehicle for collecting income and expenditure details is a CDC survey, tailored to the specifications of each individual industry. The collection period for the first CDC programme was from 2003 to 2008.[7] The approved total sample size for the combined CDC surveys was 3,500 enterprises. Data for about 110 industries has been collected.

The product information collected in the CDC programme meets the needs of two key uses: (i) the supply and use tables in the national accounts; and (ii) the suite of industry producer price indexes (PPI's for inputs and outputs) produced by Statistics NZ. The PPI's (or their sub-indexes) are also used as deflators in the calculation of chain-volume measures of GDP.

The National Accounts team was involved in the CDC programme by providing specifications and priorities and reviewing samples, survey forms, and final results.

4 An optimisation approach to support the introduction of the new product mix

In this section we describe the basic idea for the optimisation approach and present the mathematical formulation of the linear program.

4.1 The basic idea for the optimisation model

The product breakdowns from the CDC programme, as they became available, were used as an additional data source in the manual supply and use balancing process for the 2003–2005 March years. However, the comprehensive introduction of CDC results, while manually balancing the 2006 and 2007 March years, would have involved a very complex and resource intensive process to remove data inconsistencies. Therefore, a (semi-)automated approach was developed and implemented.

Section 2.2 described that the manual balancing process starts with the input and output product structure from the latest balanced year. The idea for the optimisation approach presented here is to replace the product structure for the latest balanced year with a product mix better reflecting the new information from the comprehensive CDC results. The improved starting point for manual balancing comes from a re-allocation of detailed sales and intermediate expenditure values in the latest balanced year data. This re-allocation takes place via constrained optimisation.

As mentioned earlier, the CDC programme has mainly collected detailed information on sales of goods and services and expenditure on intermediate inputs. The CDC programme does not cover all domestic supply and intermediate use in the simplified supply and use framework presented in section 2.1. For example, production for own final use is part of domestic supply and changes in the value of raw material inventories have impact on intermediate use. However, for the presentation of the optimisation model we neglect these items. Also, the value added components don't play a role in our specific optimisation model.

[7] A next (five-to-six-year) cycle of a product data collection programme has not started yet.

4.2 Notation
The matrices

$$S = \begin{bmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \vdots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{bmatrix} \text{ and } U = \begin{bmatrix} u_{11} & \cdots & u_{1n} \\ \vdots & \vdots & \vdots \\ u_{m1} & \cdots & u_{mn} \end{bmatrix}$$

contain the model's decision variables with the following notation:

S = domestic supply

U = intermediate use

m = number of products

n = number of industries

S^0 = $m \times n$ matrix containing the domestic supply values reflecting the (old) product structure from the latest balanced year

U^0 = $m \times n$ matrix containing the intermediate use values reflecting the (old) product structure from the latest balanced year

S^1 = $m \times n$ matrix containing the domestic supply values reflecting the (new) product structure based on CDC data

U^1 = $m \times n$ matrix containing the intermediate use values reflecting the (new) product structure based on CDC data

$\overline{S} = \left[\overline{S}_1, \ldots, \overline{S}_n\right]$ = row vector of column totals of S^0 and S^1

$\overline{U} = \left[\overline{U}_1, \ldots, \overline{U}_n\right]$ = row vector of column totals of U^0 and U^1

$\overline{M} = \left[\overline{M}_1, \ldots, \overline{M}_m\right]^T$ = column vector of imports for the *m* products

$\overline{F} = \left[\overline{F}_1, \ldots, \overline{F}_m\right]^T$ = column vector of final expenditure values for the *m* products

Figure 2 shows the matrix notation for the optimisation model in the context of the simplified supply and use framework as presented in figure 1.

Figure 2: Matrix notation for the optimisation model in the supply and use framework

	Output by n industries	*Imports*
Output by m products	S	$\overline{M}$

	Industry purchases	*Final demands*
Products consumed	U	$\overline{F}$

4.3 Model constraints and variable boundaries

The optimisation model is constrained by column and row totals and boundaries on the variables.

When re-allocating sales and expenditure values, industry (column) totals for supply and intermediate use must stay the same:

$$\sum_{i=1}^{m} s_{ij} = \overline{S}_j \text{ for all } j = 1,\ldots,n \text{ and } \sum_{i=1}^{m} u_{ij} = \overline{U}_j \text{ for all } j = 1,\ldots,n$$

Also, for each product (row) total supply must equal total use:

$$\sum_{j=1}^{n}\left(s_{ij} - u_{ij}\right) = \overline{F}_i - \overline{M}_i \text{ for all } i = 1,\ldots,m$$

To control the results of the optimisation procedure we introduce specific boundaries for the variables. For all variables, ‘old’ and ‘new’ values are available. The old values reflect information from the latest balanced year. Although they may not be fully up to date, they have gone through a process of data collection, analysis, and balancing, and are consistent with other information in the supply and use framework. The new values have been collected relatively recently, and we assume they are up to date. They have been analysed, but not yet in the supply and use framework (see section 5 for more details). So there is potential for inconsistencies when using the CDC information. For example, for a specific product, total use in the CDC data doesn’t match total supply.

Both the previous and the new values for the product mix have margins of error. However, if a compromise solution is needed we choose the previous and new values as variable boundaries:

$$\min\left\{s_{ij}^0, s_{ij}^1\right\} \le s_{ij} \le \max\left\{s_{ij}^0, s_{ij}^1\right\} \text{ and}$$

$$\min\left\{u_{ij}^0, u_{ij}^1\right\} \le u_{ij} \le \max\left\{u_{ij}^0, u_{ij}^1\right\} \text{ for all } i = 1,\ldots,m \text{ and } j = 1,\ldots,n$$

We assume all values in S^0, S^1, U^0 and U^1 are non-negative. Note that the values in S^0 and U^0 for the decision variables in S and U satisfy all constraints and boundaries. The optimisation problem, for which the objective function will be introduced in section 4.4, always has a feasible and optimal solution.

When for a specific industry j no CDC information is available, we set:

$$s_{ij}^1 = s_{ij}^0 \text{ and } u_{ij}^1 = u_{ij}^0 \text{ for all } i = 1,\ldots,m$$

With the variable boundaries introduced above, this means that the optimisation variables for industry j are fixed:

$$s_{ij} = s_{ij}^0 \text{ and } u_{ij} = u_{ij}^0 \text{ for all } i = 1,\ldots,m$$

So the matrices S^1 and U^1 contain the most up-to-date available information about the product structure.

4.4 The objective function

The aim of the optimisation is to get as close as possible to the product structure as reflected in the CDC information, given the constraints and boundaries introduced in the previous section. We want to move the variables in the right direction.

There are various options for the choice of an objective function. The sum of weighted squared differences results in a quadratic program and is often used as a distance metric. For reasons of computational efficiency and integration of the optimisation procedure in the current supply and use systems, we have used a linear objective function similar to weighted absolute differences.

To shorten the notation, we combine and order the supply and use variables in S and U (and the values in S^0, U^0, S^1 and U^1) in vectors:

$x = (s_{11}, s_{21}, \ldots, s_{mn}, u_{11}, u_{21}, \ldots, u_{mn}) = (x_1, \cdots, x_{2mn})$ = the vector of decision variables

x^0 = the corresponding vector of values reflecting the old product structure

x^1 = the vector of values reflecting the most up-to-date information for the product structure

The distance between the vector of optimisation variables x and the target vector x^1, measured as the sum of weighted absolute differences
(with weight factors w_i for variables x_i, which will be explained later), can be written as:

$$\sum_{i=1}^{2mn} w_i \left| x_i - x_i^1 \right|$$

The minimisation of this non-linear function is equivalent to a linear program. For this, additional variables to distinguish positive and negative variations need to be included in the linear program (see for example Gonin & Money (1989)).

However, making use of the variable boundaries introduced in section 4.3 ($\min\{x_{ij}^0, x_{ij}^1\} \le x_{ij} \le \max\{x_{ij}^0, x_{ij}^1\}$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$), an alternative objective function without additional variables can be derived:

$$\min_x \sum_{i=1}^{2mn} w_i \left| x_i - x_i^1 \right| = \min_x \left(\sum_{\{i: x_i^0 \ge x_i^1\}} w_i \left(x_i - x_i^1 \right) + \sum_{\{i: x_i^0 < x_i^1\}} w_i \left(x_i^1 - x_i \right) \right) = \min_x \left(\sum_{\{i: x_i^0 \ge x_i^1\}} w_i x_i + \sum_{\{i: x_i^0 < x_i^1\}} \left(-w_i x_i \right) + C \right)$$

with a constant term $C = \sum_{\{i: x_i^0 \ge x_i^1\}} \left(-w_i x_i^1 \right) + \sum_{\{i: x_i^0 < x_i^1\}} \left(w_i x_i^1 \right)$.

By choosing weights $w_i = \left| x_i^0 - x_i^1 \right|$ for all $i = 1, \ldots, 2mn$, the linear program gives higher priority to elements of the product mix which require a large update in absolute ($) values. The constant term C has no impact on the optimal solution, so the objective function can then be written as follows:

$$\sum_{i=1}^{2mn} \left(x_i^0 - x_i^1 \right) \cdot x_i$$

We have not systematically compared this specific objective function against alternatives. This function served our purpose of moving the variables in the right direction to improve the start situation for manual balancing. Higher weights are attached to variables which need a larger update, as suggested by the (reviewed) external data.[8]

[8] An initial experiment with an additional large weight factor in the objective function with the aim to 'lock' certain values proved unsuccessful, because other variables changed in an undesired manner.

Returning to the notation from section 4.2, the linear program which has been implemented can be summarised as follows:

$$\min_{s_{ij},u_{ij}} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\left(s_{ij}^{0} - s_{ij}^{1}\right) \cdot s_{ij} + \left(u_{ij}^{0} - u_{ij}^{1}\right) \cdot u_{ij} \right)$$

subject to (bounded optimisation variables targeting new information, and giving higher weight to larger shifts in the product mix)

$\sum_{i=1}^{m} s_{ij} = \overline{S}_j$ for all $j = 1,\ldots,n$ (industry totals for domestic supply are given)

$\sum_{i=1}^{m} u_{ij} = \overline{U}_j$ for all $j = 1,\ldots,n$ (industry totals for intermediate use are given)

$\sum_{j=1}^{n} \left(s_{ij} - u_{ij}\right) = \overline{F}_i - \overline{M}_i$ for all $i = 1,\ldots,m$ (supply and use balance for each product)

$0 \le \min\left\{s_{ij}^{0}, s_{ij}^{1}\right\} \le s_{ij} \le \max\left\{s_{ij}^{0}, s_{ij}^{1}\right\}$ for all $i = 1,\ldots,m$ and $j = 1,\ldots,n$ (variable boundaries)

$0 \le \min\left\{u_{ij}^{0}, u_{ij}^{1}\right\} \le u_{ij} \le \max\left\{u_{ij}^{0}, u_{ij}^{1}\right\}$ for all $i = 1,\ldots,m$ and $j = 1,\ldots,n$ (variable boundaries)

5 Integration of the optimisation model in the balancing process

In this section, we describe the steps added to the balancing process to integrate the optimisation model and present some process results.

5.1 Additional steps in balancing process

The linear program presented in the previous section has been integrated in the usual supply and use balancing process. As described earlier, the optimisation re-allocates the product mix in the latest balanced year (2005) towards a structure that is a better reflection of the recently collected product data (in 2003–2008). This is the starting point for manual balancing (2006 and 2007). In the balancing process as described in section 2.2, three steps were added between the usual industry and product analysis, and manual supply and use balancing:

1. Industry review: Adjusting CDC survey results to improve relevance and plausibility of the product mixes for each industry.
2. Product review: Removing large and obvious imbalances in the new product structure.
3. Optimisation review: Analysing the results from the linear program and further improving the plausibility of the product data in an iterative process.

Parallel supply and use systems to the production system were set up to support steps 2 and 3. The linear program was implemented in SAS version 8.2, using mainly PROC SQL and PROC IML to prepare and process data and PROC LP to solve the linear program itself. Manual interventions were loaded into the system via spreadsheets. The number of variables in the program was about 90,000 (127 industries and 356 products) and the number of constraints 610.[9] Only about 13,700 variables were relevant, so the linear program has been set up in a sparse data format. The run time for the optimisation was approximately 3-4 minutes on a PC (with 2.66 GHZ and 2GB RAM).

In step 1 above, National Accounts analysts reviewed the CDC results by industry and made adjustments where necessary to ensure the external data was representative for the years to be balanced (2006 and 2007). For example, statistical units (kind of activity units) may have ceased production in New Zealand recently. If the relevant CDC survey took place prior to 2006, the resulting industry product mix for outputs and inputs may need to be revised.

Secondly, a parallel supply and use system for 2006 and 2007 was filled with a product structure according to the (revised) data. An analysis by groups of products took place. The aim was to discover and remove obvious inconsistencies that could reduce large supply and use imbalances in the system. An example is the reporting of supply and use of a product across multiple industries under different classification codes.

The first two steps have improved the quality of the external data to be introduced in the supply and use framework. In the notation from section 4, vector x^1 has been adjusted, changing the variable boundaries for the optimisation procedure.

The third step was the interactive use of the linear program. Imbalances in the supply and use system based on the optimised product structure were analysed, leading to further adjustments to the variable boundaries. The linear program was very useful in instances where the use of minor input products was underreported in CDC surveys. In case of higher total supply values, and positive values for the relevant input products in the previous data, the linear program corrected the intermediate use side.[10]

Aim of the optimisation process was to achieve a situation that was similar to the starting point for manual balancing in past years.

5.2 Process results

A measure we use for the total imbalance is the sum of the absolute values of imbalances for all products in the supply and use system. The ratio of this measure and the value of total use gives a relative measure of the total imbalance. Although it varies across years, we usually started manual balancing with an imbalance of roughly 5 percent. The manual balancing process reduces this to about 1 percent. The remainder is balanced with the automated iteration procedure.

The process described in section 5.1 allowed some comparisons between different start situations. The previous product structure from the 2005 March year (vector x^0), as starting point for manual balancing the 2006 and 2007 March years, provided a relatively small imbalance of about 5 percent, but did of course not reflect the new product structure. With the initial new product structure as starting point, prior to any revisions to the new data for subsequent runs of the linear program (initial vector x^1), the imbalance was more than 20 percent for the 2006 and 2007 March years, indicating major consistency issues in the data.

We aimed to reduce this to about 5 percent with the optimisation approach before starting the usual manual balancing activities in the supply and use production system. This goal was achieved with a 4 percent imbalance for 2006 and 6 percent for 2007.

[9] Some industries were split for different sectors and market/non-market production. The full product classification has 356 groups, in practice we use around 320.

[10] The data sources for these minor input products in the previous data are probably the periodic industry censuses. When applying the optimisation approach in subsequent years for (small) CDC surveys, the variable boundaries for the relevant product could converge to a zero value. In future data collection and analysis, this needs to be taken into account.

Almost half (approximately 8 percent) of the reduction in the imbalances can be directly attributed to the linear program. The rest is achieved via manual interventions in the variable boundaries and other adjustments in the supply and use system. However, many decisions were also informed by analysis of the optimisation results.

How close did the final optimised variables (vector x) get to the projected product mix (that is the vector x^1 in the last run with the linear program)? About 52 percent of the projected values were achieved exactly. The optimal value was within a one million dollar (NZ$1m) range for 72 percent and within NZ$10m for 91 percent of all 13,700 relevant variables. Only 85 cells or 0.6 percent of the variables were more than NZ$100m away from the goal. These have been further investigated in manual balancing. To put these numbers in the context of the New Zealand national accounts: total industry output for the March year 2005 was about NZ$ 308 billion.

6 Conclusion and outlook

Our specific optimisation approach to introduce an updated product mix in a supply and use framework has been successful in practice. The optimisation model was integrated in the balancing process, allowing an effective and efficient combination of automated and manual balancing techniques.[11]

The linear program we have used could be extended with variables in final demand categories if new external data becomes available. Also, variable industry totals could be brought into the optimisation to further improve the start point for manual balancing, although first experiments have indicated that this comes at a cost of much longer run times. A newly available SAS version in the production environment of NZSNA, containing interior-point algorithms, and running the optimisation on a server instead of PC would probably make quadratic objective functions an interesting alternative to investigate. In future work, we could also look at weights reflecting reliability of the data, and variable boundaries taking error margins in the data into account.

Our focus at the moment is on implementing a new industry classification (Australia and New Zealand Standard Industrial Classification, ANZSIC) in the NZSNA. A new product classification will be introduced as well. The sample design in the CDC programme was based on the old industry classification ANZSIC 1996, but the new industry classification ANZSIC 2006 was taken into account. Currently the ANZSIC 1996 based product structure is being converted to an ANZSIC 2006 basis. The first National Accounts release on the new basis is scheduled for November 2011. After the new classifications have been implemented we will be looking at extending our framework with supply and use balancing in volume terms. When a new product data collection programme based on ANZSIC 2006 will start and in what form has not yet been confirmed. In future, we continue aiming to find the right ‘balance’ of interventions by skilled analysts and automated procedures.

Acknowledgements

The author is thankful to numerous people at Statistics New Zealand providing valuable comments on earlier versions of this paper.

[11] For our external users the updated product structure is unfortunately not visible yet, as we have yet to release an update of detailed supply and use tables (and input-output tables) in basic prices. Our current core systems do not facilitate an easy transformation from producers’ to basic prices.

References

European Commission (2008). *Eurostat manual of supply, use and input-output tables.* Luxemburg: Office for Official Publications of the European Communities. Available from http://epp.eurostat.ec.europa.eu.

Gonin, R, & Money, AH (1989). *Nonlinear Lp-norm estimation*. New York: Marcel Dekker.

Inter-Secretariat Working Group on National Accounts (1993). *System of national accounts 1993*. Commission of the European Communities – Eurostat, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, and World Bank: Brussels/Luxembourg, New York, Paris, Washington, DC. Available from http://unstats.un.org.

Lenzen, M, Gallego, B, & Wood, W (2009). Matrix balancing under conflicting information. *Economic Systems Research*, *21(1)*, 23–44.