Economic Systems Research

ISSN: 0953-5314 (Print) 1469-5758 (Online) Journal homepage: http://www.tandfonline.com/loi/cesr20

CONSTRUCTION, STABILITY AND PREDICTABILITY OF AN INPUT–OUTPUT TIME-SERIES FOR AUSTRALIA

Richard Wood

To cite this article: Richard Wood (2011) CONSTRUCTION, STABILITY AND PREDICTABILITY OF AN INPUT–OUTPUT TIME-SERIES FOR AUSTRALIA, Economic Systems Research, 23:2, 175-211, DOI: 10.1080/09535314.2011.564156

To link to this article: http://dx.doi.org/10.1080/09535314.2011.564156

Published online: 15 Jul 2011.

Economic Systems Research, 2011, Vol. 23(2), June, pp. 175–211

CONSTRUCTION, STABILITY AND PREDICTABILITY OF AN INPUT–OUTPUT TIME-SERIES FOR AUSTRALIA

RICHARD WOOD*

Centre for Integrated Sustainability Analysis, A28, The University of Sydney NSW; and Industrial Ecology Program, NTNU, Norway

(*Received 19 February 2010; In final form 15 February 2011*)

This paper documents the development of a time series of Australian input–output tables. It describes the construction techniques employed in order to overcome the major issues encountered. Environmentally important processes were delineated using a range of detailed commodity data, thus expanding the original tables from roughly 100 industries into a temporally consistent 344 industries. Data confidentiality and inconsistency were overcome using an iterative constrained optimisation method called KRAS – a recent modification of RAS (Lenzen et al. 2006; 2007; 2009). The article concludes by analysing the stability of input–output coefficients over time similar to work in Dietzenbacher and Hoen (2006). The issue of stability of coefficients and multipliers was investigated under the Leontief and Ghosh models of supply/demand. Finally, the predictability of the models was examined under updated final demand or primary inputs and over varying time scales.

Keywords: Input–output analysis; Time series; Constrained optimisation; Stability; Predictability

1 INTRODUCTION

Data requirements have greatly increased with the increasingly complex problems addressed by environmentally extended input–output analysis. Two particular cases that are now becoming popular concern multi-regional input–output models and temporal studies. Both cases provide important insights into technology and society as well as technological and societal change, but require vast amounts of data compared with a single impact or attribution analysis.

Statistical agencies originally relied wholly on survey techniques to update input–output tables. However, the expense and time for such surveys has resulted in a move to more mathematical approaches, especially since Stone's exposition of the RAS technique (Stone and Brown, 1962). Today national statistical offices rely on a mix of survey and mathematical techniques for publishing updated input–output tables. Originally, the RAS technique was a procedure for balancing row and column totals, but it was subsequently modified to incorporate superior data on internal elements of a new table. It has undergone significant development, most of which is outlined recently elsewhere (Jackson and Murray, 2004; Lahr and de Mesnard, 2004; Lenzen et al., 2006, 2009; Mínguez et al., 2009).

*E-mail: richard.wood@ntnu.no

ISSN 0953-5314 print; ISSN 1469-5758 online
DOI: 10.1080/09535314.2011.564156

Many analysts are now taking on the work of updating input−output tables themselves by employing mathematical methods. Major issues include dealing with differing levels of data availability, changing classifications and changes in methodologies underlying the published input−output tables, such that when time series analysis or multi-regional analysis are performed, there is often no consistency across tables. Analysts are hence left with a considerable task of estimating input−output tables across time and/or regions, in a common classification, and, possibly, in different prices.

The key aim of this work is to construct such a time-series of the Australian economy, with an auxiliary aim to disaggregate input−output tables to a useful level of detail for environmentally important analysis. The function of the paper is to outline the techniques employed and issues encountered in estimating a time series of input−output tables for Australia at a high level of detail. In the first part of this work, a RAS variant is employed that allows for the consideration of any type of superior data that can be expressed as a linear combination of the table elements, and that act as constraints of the table to be balanced (Lenzen et al., 2006). The most innovative feature of the technique is that it can handle incompatibilities among these constraints (Lenzen et al., 2009), as well as any negative constraint or table value. The main considerations in this method are – creation of an initial estimate of the system, incorporation of constraints at different levels of classification on the system and balancing or optimisation of the system.

The application of these balancing techniques reduces the user's knowledge of uncertainties associated the source data. Error analysis can be undertaken by investigating distributions of uncertainty in model variables. Recent examples are Percoco et al. (2006) and Lenzen et al. (2010), who both estimate uncertainty using an assumed log-normal distribution of stochastic errors in input–output coefficients and final demand. However, whilst such a method allows for identifying the sectors most prone to error, it does not address non-stochastic change over time. In addition, it can be expected that, in real terms, large variations can occur from year to year, especially for industries affected by natural events. Instead, in this paper, the reliability, or predictability, of dated and mathematically updated tables is investigated.

Numerous papers have been written on analysing the change in structure over time. Bon (Bon 1984, 1986, 1988; Bon and Bing, 1993) has concentrated particularly on the predictability of supply and demand multipliers. Whilst Bon's work does not include updating tables, his work is of interest in comparing the accuracy of using older transactions matrices with updated data. Dietzenbacher and Hoen (2006), subsequently referred to as D&H, couple an analysis of the stability of model coefficients, with the predictability analysis of Bon.

This paper focuses on Australian input–output data. The situation of table availability for Australia is such that a semi-survey set of input–output tables (referred to as 'benchmark' tables in this paper) is published by the Australian Bureau of Statistics circa every two–four years (with 17 publications in total since 1975). In addition, more aggregated supply/use tables in purchaser prices were published for a time series of 10 years (1994/5–2003/4). Finally, the National Accounts, published annually, provide data for aggregate economic components. In this work, maximum information is incorporated in the estimation of the tables, and the accuracy of the estimated tables is determined through comparison with the published tables.

This paper begins with an outline of the construction method in Section 2. The context of the sector and temporal detail desired in this work is discussed at length (Sections 2.1.1

and 2.1.2), followed by a description of the system structure (Section 2.1.3); an overview of data (2.1.4) and processes (2.1.5); the creation of the initial estimate (Sections 2.2, 2.3 and 2.4 for constant price cables); and the application of constraint data (Section 2.5). The actual balancing is discussed in Section 2.6. The predictability and stability of the estimated time series is then analysed in Section 3. Conclusions are drawn in Section 4.

2 TABLE CONSTRUCTION

2.1 Context

2.1.1 Aggregation Level

An aggregated dataset is easier to construct than a disaggregated dataset, but is inherently more imprecise. A reduction in industry (or product) resolution can significantly limit a model for impact analysis and can potentially misinform subsequent decision making, especially when non-economic indicators are investigated (Lenzen, 2011). It is thus generally desirable to use the maximum level of detail when compiling integrated systems of disparate data, and aggregate results after impact analysis, if necessary.

Physical data exogenous to the input–output system is generally available at varying levels of detail for different industry sectors. The maintenance of the resolution of this data is one of the principal reasons for working at a high level of detail. For example, data on material requirements are detailed across mining and similar industries, energy data are detailed across manufacturing industries and employment data are detailed across service industries. When performing an energy impact analysis of a good, it is then important to have precise data on energy requirements of manufacturing processes; whilst potentially crude assumptions on the energy requirements of the service sector have limited impact on results. In essence, we seek holistic accuracy rather than the accuracy of individual elements (Jensen and West, 1980; Hewings et al., 1988; Gallego and Lenzen, 2008). The arguments follow the debate in the life-cycle assessment literature on precision versus accuracy (Suh et al., 2003).

To further highlight this issue (compare Gallego and Lenzen, 2008), the greenhouse gas emissions allocated to beef cattle are considerably different to the greenhouse gas emissions allocated to sheep farming. If, as is common, the beef cattle industry is aggregated with sheep farming in a ‘livestock’ industry, there is no way to discern the difference in the greenhouse gas intensity of the two industries. The consumption of livestock products will embody a large quantity of greenhouse gases, even if in reality all consumption was of the sheep products and none from beef cattle. This is important in Australia’s situation, where large changes in the output of sheep and cattle farming are occurring in opposing directions and where sheep and cattle sales are often destined for different levels of demand domestically and for export. In an aggregated framework, the differences in the production practices are not captured, whereas aggregation after analysis will include these differences. An empirical study demonstrating this effect is found in Lenzen et al. (2004).

In this work, the task of creating a standardised procedure to generate a detailed generalised input–output framework was undertaken. As a first step, choices of detail with respect to the product and industry classification had to be made. In Australia, the product and industry classifications in published “benchmark’ IO tables (in supply and use format) are both shown by the input–output industry classification (IOIC), which is generally around 100 product/industry sectors. In a second dataset, the product details are described in more detail with the input–output product classification (IOPC) of around 1000 products, as defined by the ABS (Australian Bureau of Statistics, 2006). These data only have more detail in the product classification (not the industry classification), hence, the tables are rectangular Supply and Use matrices measuring about 1000 × 100.

In order to have a temporally consistent set of input–output data, a subset of 344 commodity groups was selected from the IOPC. This aggregation level was detailed enough for most available auxiliary data on primary and environmental inputs, whilst enabling tractability of classification changes and handling of confidential cells.

When estimating such level of detail, there is always a balance between providing highly detailed models, and avoiding the manufacture of 'artificial data'. Whilst the precision of highly detailed models is something to strive for, the interpretation of results must reflect the source data. For example, if we know total mining production is increasing over time, an information gain is obtained if this datum is included in the model; even if no knowledge is available on the change in the relative amounts of iron ore mining to precious metal mining. Results can then reflect on the relative aggregate environmental impact of mining over time, but cannot give a reason for the change due to the mining of specific ores (even though these data are estimated). The detailed model becomes more powerful, providing higher quality results, but runs the risk of allowing estimates to become fact.

Returning to this work, the main data estimation across sectors occurs in changing from a rectangular matrix to a square matrix when disaggregating the 100 or so industries to 344 industries. Section 2.2 explains the procedure undertaken.

2.1.2 Temporal Considerations

In addition to sectoral estimation, a full time series of IO tables was produced from intermittently available tables in various levels of detail. This represents the second main form of data estimation – interpolation of disaggregate data between benchmark years (where full detail input–output and high detail commodity tables are available). This choice was based on the basic purpose of historical analysis in general – this being to (a) analyse general trends over time, whilst also (b) investigating responses to specific events. Obviously, the method of interpolation should not skew a trend that one would otherwise conclude from the data. The cross-benefit is that structural shocks, such as due to the oil crisis in the 1970s or other recession events are not overlooked if three or five year intervals are taken. It must also be remembered that in the investigation of temporal change without continuous data an analysis either explicitly or implicitly uses indexing methods (e.g. Laspeyres, Fischer, Paasche, or many more) to inform about the derivative of change (whether the change occurred consistently, or more rapidly at the start or end of the time period, etc). These methods are purely mathematical, and can have a reasonable effect on results in temporal analysis (Dietzenbacher and Los, 1998; Ang, 2000; Hoekstra and van den Bergh, 2002; Wood and Lenzen, 2006). It is thus logical to reduce the reliance on the particular choice of indexing method by estimating development paths across an annual time series. In the work here, as a first estimate, value added is used as a proxy for interpolation of industry developments from the prior and later benchmark data. This estimate is then balanced so that it complies with more aggregate data so that the interpolation gives added insight into the path of change.

A second important temporal consideration is the stability of classifications used in data provision over a long time-series, and the effect of changes in methodology on historical data. Changes in classification are inevitable as new products become available on the marketplace, and the relative importance of different products and industries grows and diminishes. The construction of concordance matrices deals with these changes, as explained further in Section 2.5.2.

Corrections for methodological changes provide a greater challenge. In Australia, 1974–75 began the first recording of intra-industry transactions, and also, the first use of mathematical techniques to estimate input–output tables (see the comprehensive coverage of changes in Gretton, 2005). Gretton identifies four major series of input–output tables. The first, ending in 1975, was superseded by a series covering the 1978 to 1993 tables. The major changes between 1975 and 1978 were the adoption of a new industry classification (ASIC 1978), and in methodology, changes to the way imputed finance, social services consumed by households and some indirect taxes were calculated (Gretton, 2005). The third series of tables from 1994 to 1999 includes some changes reversed in the 2002 tables. These included the handling of transport margins for when product suppliers pay transport invoices.

For this work, consistency in handling transport margins was considered important and was thus addressed explicitly (Section 2.5.3). For the changes on finance, social services and indirect taxes, as well as other minor changes, the available data were either included and changes were assumed negligible on structural results, or the data were excluded, and tables were balanced to aggregate totals. Being able to address all these changes explicitly would be ideal, but such a process would require data only (if at all) available in a statistical bureau, and require a lengthy process that was beyond the scope of this work. Furthermore, it is expected that such detail would have a negligible impact on the type of analysis under consideration here.

2.1.3 System Structure

The estimated system was set up to be consistent over time and to incorporate all input–output flows in a supply-use format, and in various pricing layers (United Nations Department for Economic and Social Affairs Statistics Division, 1999). Variables are thus chosen for supply and use tables, imports, 11 margin layers and four tax layers (defined in Table 1).

The variable $\mathbf{Z}^t$ is used henceforth to represent the full input–output system at a single point in time t (over a time period $\mathbf{t}$ – bold capital indicating a matrix, and bold lowercase denoting a vector, italic denoting scalar). For convenience, rather than working in a matrix representation, the lowercase $\mathbf{z}^t$ is used to represent the vectorised form of the input–output system – that is, it is a one-dimensional representation of every flow of the input–output system. $\mathbf{z}^t$ includes the Supply matrix, Use matrix, Imports matrix, and all margins and taxes matrices. The size of $\mathbf{z}^t$ is hence dependent on the number and size of the variables of the input–output system. In this case, with 344 commodities, 344 industries, five primary inputs, seven levels of final demand and 17 margin and tax tables as well as the Supply, Use and Imports tables, the size of $\mathbf{z}^t$ is circa 2.5 million variables. $\mathbf{z}^t$ is defined as:

$$\mathbf{z} = \text{vectorisation}(\mathbf{V};\, \mathbf{U};\, \mathbf{M};\, \mathbf{P}^k;\, \mathbf{T}^k)$$

TABLE 1. Variables of open input–output system, *m* commodities, *n* industries, *a* types of primary inputs and *b* types of final demand.

Variable	*Dimension*	*Description*
$\mathbf{x}^I$	*m*	Output, industry.
$\mathbf{x}^c$	*n*	Output, commodity.
V	*m,n*	Supply table, showing sales of Industry *i* of commodity *j*, basic prices, domestic flows.
U	$m + a,n + b$	Use table, showing purchases of commodity *m* by industry *n*; final demand, showing purchases of commodity *m* by destination of final demand *b*; primary inputs into industry demand, showing inputs into production of input type *a* into industry *n*, basic prices, domestic flows.
M	$m,n + b$	Imports table, showing purchases of imported commodity *m*, by industry *n* or final demand destination *b*, c.i.f. valuation, imported flows.
$\mathbf{T}^k$	$m + 1,n + b$	Individual taxes or subsidies, *k*, paid on product *m* (plus re-exports) by industry *n* or final demand destination *b*. *k* tax tables include GST, Duties, Other taxes, Subsidies.
P	$m + 1,n + b$	Total margins (incl. redistribution) paid on product *m* (plus re-exports) by industry *n* or final demand destination *b*.
$\mathbf{P}^k$	$m + 1,n + b$	Individual margins, *k*, paid on product *m* (plus re-exports) by industry *n* or final demand destination *b*. *k* margins tables include Wholesale, Retail, Hospitality, Road, Rail, Pipeline, Water, Air, Port, Marine, Gas. Dimensions as above.
gdp	1×1	GDP measure.
$\mathbf{U}^{pp}$	*m,n*	Use table, showing purchases of commodity *m* by industry *n*; final demand, showing purchases of commodity *m* by destination of final demand *b*; primary inputs into industry demand, showing inputs into production of input type *a* into industry *n*, purchases prices, domestic flows.

Whilst slightly convoluted here, this distinction is made for practical purposes, as the matrix format is convenient when using concordance matrices between row/column classifications in constructing constraints (see Appendix), whilst a vector format is convenient for mathematical manipulation. An indexing matrix $\mathbf{Z}^{\mathbf{r}}$ is used to translate between the original square format **Z** and an individual flow in the vectorised format $\mathbf{z}_r$. It should also be noted that the term 'input–output tables' is used throughout this piece, referring to the full set of the supply, use, imports and margins tables, not the symmetric input–output table (United Nations Department for Economic and Social Affairs Statistics Division, 1999).

2.1.4 Available Data

Of course, the key to having a reasonably informed model is having reasonable data input. Here source data are not available with a stable classification because of confidentiality issues, classification changes, methodological changes and because of temporal gaps (see Section 2.1.2 and later, as well as Table 2). As the available data are in varying levels of aggregation and consistency, they cannot be used directly in the model.

Instead, the data must be used to refer to sub-aggregates of transactions – much as row/column totals constrain all production and uses of a product/industry.

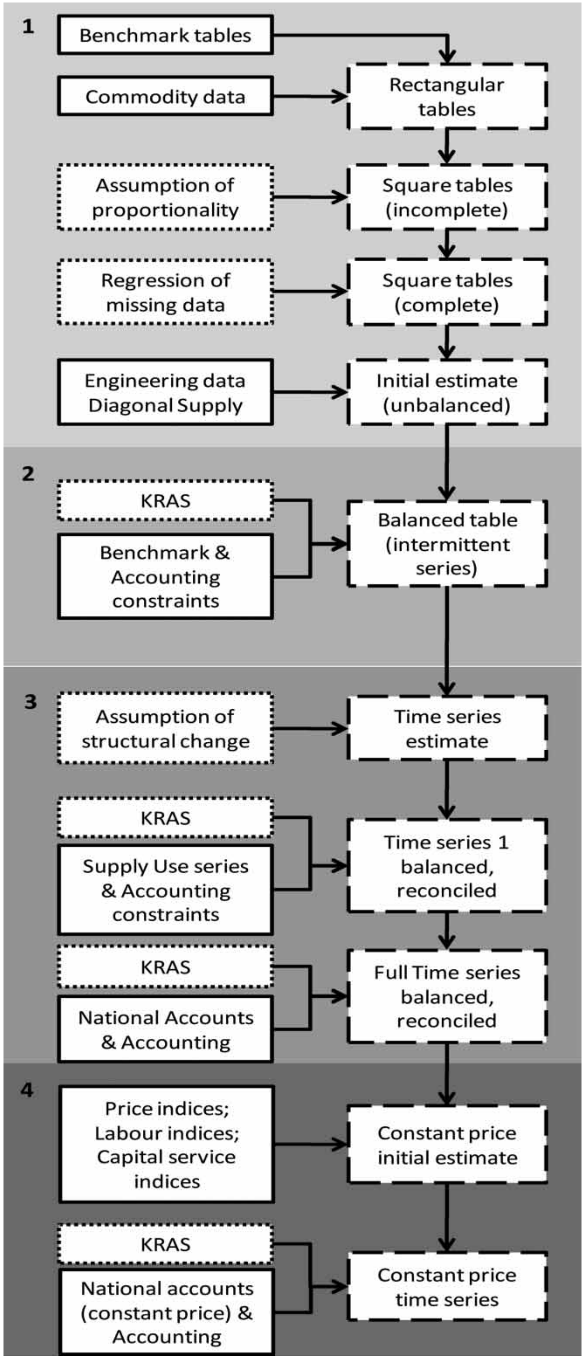
The source data availability (Table 2) in this work are 17 sets of 'benchmark' input–output tables in a supply use format including all margin layers and with a classification of approximately 100 products/industries. Eleven of the sets of benchmark tables are accompanied by detailed data on the roughly 1000 products by the 100 (or so) industries, again in supply/use format and in basic and purchaser prices, with all margin and tax tables (however, post 1999 only aggregate margins paid on each product were available). Note that these 'benchmark' and 'detailed' tables are under changing classification and methodology. An actual time series of aggregated supply/use table in purchaser prices was also available from the ABS, aggregated in 224 products by 52 industries. Whilst not consistent with the benchmark IO data, these data did provide a true time series under consistent methodology and classification, and thus were useful for extracting relative change (reflecting the combination of both volumetric and price changes). National accounts data are updated annually and back casted in methodology such that the datasets are consistent with the most recent benchmark table, and provide annual series of key national accounts aggregates over the full 30 years. Selective uses of the data were made in the estimation and balancing process (outlined in the next section). Finally, some 'engineering' data were also used in the creation of the initial estimate (Section 2.2) when changing from rectangular to square tables.

TABLE 2. Summary of main data used in the initial estimate and constraints employed in this work.

	Year	*Products*	*Industries*	*Primary Inputs*	*FD*	*Pricing*
IO-data (benchmark)	2005,2002,1999… (17 years)	103–118 (varies by year)	103–118 (varies by year)	7	7	ALL
IO-data (commodity details)	2005,2002,1999,… (11 years)	~1000	103–118	-	7	ALL
Purchaser's price Supply/Use margin totals	2004–1995 (10 years)	224	52	7	7	p.p. +
National Accounts (current price)	1975–2005 (30 years)	Varies	Varies	Varies	Varies	b.p., p.p
National Accounts (constant price)	1975–2005 (30 years)	Varies	Varies	Varies	Varies	b.p., p.p
Engineering data	n/a	Specific	Specific	0	Specific	n/a

Notes: All data are on Australian financial year (July-June) basis, but referred here by the final year (e.g. 2001–02 data referred to as 2002). b.p. = basic price; p.p. = purchaser price; margin totals = total margin on each product by type of margin; ALL = full detail on basic price, imports and each individual margin layer.

FIGURE 1. Flow chart of data (black boxes) and processes (dotted boxes) used in the construction of the stages (dashed boxes) of the time series.



2.1.5 Overview of Processes

Starting from the set of square 100 by 100 benchmark tables in supply and use format, the following processes were used in order to achieve a time series of tables in 344 sectors (Figure 1). In stage 1, an initial estimate was created using available benchmark data (Section 2.2). Aggregate data were first expanded by detailed data on additional product rows, before being made square under the assumption of proportionality of inputs – i.e. that disaggregated columns have the same coefficients as the aggregated column. Next, a regression process was undertaken in order to estimate missing data without introducing bias, before tables are adjusted according to additional assumptions on supply and input coefficients. In stage 2, the KRAS balancing of the initial estimate of the tables was performed using the benchmark and accounting (balancing) constraints (Sections 2.5.1, 2.5.2 and for the balancing, 2.6) to give an intermittent series of tables corresponding to when benchmark data were available. A full time series was constructed in stage 3 – firstly by creating an initial estimate for years without benchmark data (Section 2.3). This was followed by applying supply use data in purchaser prices (Section 2.5.3) for 1995–2004 (time series 1), and then balancing to the consistent series of national accounts data (Section 2.5.4) across the full time series. A fourth step was then performed to convert the time series in current prices to one in constant prices – again, firstly by creating an unbalanced initial estimate, this time with price indices (Section 2.4); followed by balancing to constant price national accounts data (Section 2.5.4).

The application of the KRAS balancing process (Section 2.6) is the same in each stage – only with a different initial estimate $\mathbf{z}^0$ and constraint set $\mathbf{c}$.

2.2 Creation of Initial Estimate for Benchmark Years

The Australian benchmark input–output tables have historically been subject to numerous classification changes. The first step in creating the full time series was to create an estimate of the tables in a classification stable across all years. This involved adding auxiliary information and assumptions to go from the temporally inconsistent 100 by 100 (or so) tables to the selected standard classification of 344 products/industries. As previously mentioned, more detailed commodity data are available in addition to the data available in the 100 by 100 tables. This detail on commodities only refers to the type of good, not where it was consumed, and thus only allows the expansion of the input–output system by the number of product rows (i.e. the data allow creation of rectangular supply/use tables). In order to move from rectangular to square tables, data on consumption by disaggregated industry are initially included by assuming proportionality between product output and input coefficients. That is – the relative output of an individual product rx_j^c is calculated compared with the total output of the aggregated product group $\sum_{j*} x_j^c$, according to $rx_j^c = x_j^c / \sum_{j*} x_j^c$, where j^* refers to the aggregate product group. Disaggregated inputs into supply V_{ij}, and use U_{ij} (and imports and margin, tax layers; $\mathbf{M}$, $\mathbf{P}$, $\mathbf{T}$) are calculated as (per use): $U_{ij} = rx_j^c \times U_{ij*}$. This gives an initial crude disaggregation of inputs that are to be refined later. At this stage, the coefficients of the disaggregated industries are identical to the original aggregated industry. This step results in data being organised in tables with a standard temporally consistent classification, thus allowing for inter-temporal adjustments to be made.

Due to both confidentiality issues and classification issues, data points for some commodities are only available for a select number of years. In addition to this, various elements of the original tables are also confidential. One basic requirement before balancing the tables is that there is an initial estimate of every flow for at least one point in time. Hence, in order to fill tables for years with confidential or otherwise missing data points, an initial estimate was required for this data. The options canvassed to create this initial estimate were: assume data points are constant from known years to unknown years; linear regression; scaling using aggregated data; or some other form of regression (such as log-linear).

The first option, assuming data points are constant, increases the error when some commodity data are known and some commodity data are unknown. In essence, price and volume changes would be applied to the known commodity data, but no estimates of changes would be applied to the unknown commodity data. As some specific commodity data are confidential for most of the 30 year time period, the lack of consideration of these changes would cause considerable error. The second option of using linear regression was found to be lacking, as any rapid economic changes would be reflected in the regression by negative or very large estimates due to endpoints not always being available. Also, in reality, very little economic growth is linear anyway. The third option, using aggregated data such as the more aggregated level industry data can only occur if the industry data are non-confidential, which is not always the case. Further, if the industry data are not confidential, it will already be applied in the constraints of the balancing process (see next section).

In order to overcome these issues, the fourth option of applying some other form of regression was taken. Due to the issues previously identified, a stable sign-preserving form of regression was desired. Assuming the economy has undergone consistent growth, the exponential form was taken, with a limiting factor to prevent very large estimates. Using the vectorised definition of all flows at time t $\mathbf{z}^t$, and for a subset $\mathbf{s}$ of time $\mathbf{t}$ $(\mathbf{s} \supset \mathbf{t})$ where a flow z_r^s is known, the growth coefficient a is estimated, such that $a < \alpha$, for some α – (a limiting factor, chosen to reflects estimated limits to growth rates and to maintain stability over the time series), according to $\ln(z_r^{(\mathbf{t}=\mathbf{s})}) = a_r(\mathbf{t} = \mathbf{s})$, solved using ordinary least squares. The missing elements of the initial estimate are then estimated for known a_r from the regression $\log(z_r^{(\mathbf{t}\neq\mathbf{s})}) = a_r(\mathbf{t} \neq \mathbf{s})$.

For example, for $\mathbf{s} = \{1999, 1997$ and prior$\}$, the flow z_r^s that refers to any type of final consumption of agricultural services is known. However, for $t = \{2002, 2005\}$, this flow is confidential. Thus, the set of years 1999 and earlier is used to estimate a_r which is then used to estimate z_r^t for $t = \{2002, 2005\}$.

This method should not be considered ideal – volume and price changes are not accounted for separately due to data availability; the regression is not fully tested through subsequent statistical analysis. However, it was considered adequate for the current work because the actual number of data points estimated in this procedure is small (only confidential data, or missing data due to classification changes – probably less than 10% of data). The importance of these data points is also generally small (most important flows in the input–output tables are non-confidential and represented in stable classification). All flows are also subsequently balanced to available industry data, row totals and column totals; and the method is not applied to years where no input–output data are available. The advantages are that this method provides a reasonable first estimate of otherwise unknown flows without introducing an obvious bias.

Following this estimation of missing data, tables are now available in a standard classification (first step), and without missing elements (second step). As a third step, the rather crude splitting of inputs into disaggregate industries (and supply of disaggregate industries) is improved by means of ‘engineering’ data, following a session convened to utilise expert knowledge (Foran et al., 2003). These data improved the initial splitting of the aggregate industries by firstly assuming diagonal supply (compare Mastoris et al., 2002) within the main diagonal block of the aggregate industry. Secondly, data on inputs were adjusted according to expectations of zero flows (e.g. no inputs of bauxite into any metal ore processing industries apart from alumina). Secondary data sources from the expert session (Foran et al., 2003) were then used to give rough approximations of inputs for products such as electricity – with, for example slightly less than two-thirds of electricity use by all types of non-ferrous metal coming from aluminium production, 15% coming from alumina, and the remainder coming from other non-ferrous metals. The disaggregation of inputs from other sectors such as services and agriculture was left from the original proportionate distribution at this stage (prior to balancing). When analysing specific questions such as electricity use by different disaggregated industries, this type of allocation performed here would possibly not be adequate. An example of extending this basic system to allow physical allocation of such commodities as electricity in order to address differences in non-ferrous metals is shown in Wood and Dey (2009).

To re-establish the context to this method, a further example is given: in this work, the industry grains are disaggregated into wheat, barley, rice, oilseeds, and other grains. For primary environmental inputs, considerable differences could be expected, but for economic inputs, the coefficients of the wheat industry would be expected to be reasonably similar to that of the barley industry, and somewhat similar to the rice industry. The importance of having the industry disaggregated comes when the environmental data are appended to the IO tables. We know rice has a much larger direct extraction of water than the other industries – hence both the water intensity and multiplier of rice will be much larger than the intensity/multiplier of the original ‘grains’ industry. Beer is made from barley and other grains (no rice), and some spirits in Australia are made from rice. Hence, in an aggregated study, beer would obtain part of the water footprint of rice. In a disaggregated study, we could make that distinction.

2.3 Method for Creating Initial Estimate for Non-Benchmark Years

For years with no benchmark input–output data, a weighted average calculation is made from available data for preceding and subsequent benchmark years to arrive at an initial estimate. The initial estimate for years with no IO data was only constructed once a complete balancing had been performed for years with IO data (refer Figure 1).

Weights were constructed according to:

$$\mathbf{w}^1 = \left(1 - \frac{|t^1 - t^t|}{(t^1 - t^0)}\right) * (\mathbf{W} * (\widehat{\mathbf{q}^1}./\mathbf{q}^t))$$

$$\mathbf{w}^0 = \left(1 - \frac{|t^0 - t^t|}{(t^1 - t^0)}\right) * (\mathbf{W} * (\widehat{\mathbf{q}}^0./\mathbf{q}^t))$$

where t^t is the year to be estimated, $\mathbf{w^1}$ is the weighted ratio applied from the year t^1 of subsequent IO data. The scalar (or element-wise) division ./ of $\mathbf{q^1}$ the vector of value added in the year t^1 and $\mathbf{q^t}$ the vector of value added in the year t^t is multiplied by a concordance matrix **W** showing the relationship of the value added classification to the input–output classification. The diagonal (here denoted by a hat) of the expanded value added ratios is then taken for later matrix multiplication. The first term in the weighting calculation provides a linear weighting of temporal distance, where the distance from the years t^1 or t^0 to t^t provides the relative weight. The calculation is repeated for the preceding year t^0 of IO data to give $\mathbf{w^0}$. Data on value added is taken from the National Accounts data set (Australian Bureau of Statistics 2007b), with 18 distinct industries.

The initial estimate is then calculated as:

$$\mathbf{Z}^t = \mathbf{W}^1 * \mathbf{Z}^1 + \mathbf{W}^0 * \mathbf{Z}^0$$

Once the initial estimate is created, it is balanced against available data. From 1995 onwards, this includes purchaser price tables (Section 2.5.3) and National Accounts data of value added and final demand (see Section 2.5.4) and according to row column balances (see Section 2.5.1).

2.4 Constant Price Tables

An additional step was taken after the establishment of the time series of current price tables in order to estimate constant price tables (see Figure 1 and Section 2.6). Price indices were used on the current price input–output tables to give a first estimate of constant price tables. These tables are not balanced – the application of product grouped price indices, and specific indices for value added, creates non-additivity over rows compared with columns, and due to chaining of indices within National Accounts data, the classic problem of aggregate non-additivity presents itself (see Reich, 2008, for a more complete overview). A choice can be made on whether to leave or reconcile these inconsistencies. Reconciliation has been chosen in this work, for the following reasons: as the non-additivity problem is a mathematical construct, a mathematical reconciliation seems plausible; inconsistencies are generally well within the expected accuracy of the data (Reich, 2008; Lenzen et al., 2010); from an end user perspective, having balanced row/column sums, as well as consistency with aggregate economic variables, is desirable – especially given the relatively large uncertainty of the applied price indices. This work differs from the standard double deflation method (United Nations Department for Economic and Social Affairs Statistics Division 1999), in that specific deflators are used for components of value added, rather than calculating value added as a residual. The choice reflects the desire to obtain a greater amount of direct information on quantity changes in value added. The importance can be seen when analysing aspects such as labour rates – utilising direct physical information should give much more stable compensation of employee rates ($ per hour worked), than if compensation of employees is calculated as a residual as is done in the double deflation method.

As per the tables in nominal values, constant price tables are balanced in three steps. First, the application of appropriate price indices to the current price tables to create an initial estimate; second the selection of appropriate aggregate data in real values; third, the balancing of the initial estimate to both accounting constraints and aggregate data.

2.4.1. Price Indices

In order to deflate input–output tables from nominal to real values, price indices were required. Ideally, price indices would be available for every input–output transaction. However, in reality, the statistics bureaux collects price indices for only selected products and for an even more select group of purchasers. It is hence assumed, that without better available data, each individual purchaser is subject to the price index of its closest related purchaser. In terms of products, price indices of individual products are assumed to be that of its overarching product group where no individual indices exist. The price indices used within this work are given in Table 3.

TABLE 3. Applicable price indices used in this work.

Name	No of products	Catalogue no	Years available
Consumer price index	144	6401	1972; 1980; 1989–2007
Producer price index – Manufacturing use	189	6427	1970; 1971; 1973; 1981–2007
Producer price index – Manufacturing demand	53	6427	1970;1996-2007
Producer price index – Construction output	7	6427	1996;1998-2007
Producer price index – Transport and storage	36	6427	1996;1998-2007
Producer price index – Property and Business	40	6427	1996;1998-2007

Two sets of price indices were constructed, one with emphasis on producer price indices (PPI) (modified from Australian Bureau of Statistics, 2008d), and one with emphasis on consumer price index (CPI) (modified from Australian Bureau of Statistics, 2008c). It was assumed that producer price indices were most closely related to intermediate demand and that the consumer price index was related most closely to final consumption. Hence, producer price indices were used when available for the intermediate transactions for a particular product group, and, if not available, consumer price indices were used; and consumer price indices were prioritised for products of final demand when available.

2.4.2 Labour Index

The compensation of employees was deflated using an index of hours worked. As a first step, employee numbers were estimated for one year by industry using data from the Australian Business register (Australian Bureau of Statistics, 1998), which provides a very high level of detail. These data were balanced to time series industry level data (228 industry subdivisions) of the ABS from 1984–2008 (Australian Bureau of Statistics, 2008a). Employment prior to 1984 was recast from the 1984 data using aggregate industry numbers for each year 1966–84 (Table 7 in Australian Bureau of Statistics, 2007a), scaled to consistent total employment level figures from 1978–2008.

Labour data in terms of hours worked were estimated as a second step using the relative amount of employment by industry data and the total hours worked by aggregated industry (19 industries) (Table 11 in Australian Bureau of Statistics, 2008), since 1984. Prior to 1984, average weekly hours worked (Table 8 in Australian Bureau of Statistics 2007a) were multiplied by total employment (Table 7 in Australian Bureau of Statistics 2007a) for the same 19 industries. Consistent time series were available from 1978–2008 at the aggregate level, and data were scaled to match these aggregates. Data from 1975–77 were scaled at the same rate as 1978, as it is assumed that the differences are due to different methodologies employed, and that this difference is consistent across each data set.

Indices were then calculated with respect to the 2005 base year. Constant price data were constructed based on the current price data for 2005 and the labour indices. Price indices for compensation of employees to convert between current and constant prices are finally constructed as the ratio between the two measures.

2.4.3 Capital Index

Gross operating surplus was deflated using an index of capital services. A capital service index is available for the market sector (Table 58 in Australian Bureau of Statistics, 2007b), and an index was estimated from the quantity of net capital stock in chain volume measures for the non-market sector (Table 58 in Australian Bureau of Statistics, 2007b). Constant price data were then constructed based on the current price data for 2005 and the estimated capital indices. Price indices for gross operating surplus to convert between current and constant prices are finally constructed as the ratio between the two measures.

2.4.4 Deflation

Indices were applied to all elements of the IO table. Price indices were applied to the product purchased in the case of intermediate and final transactions (shown by pre-multiplication), whilst value added was deflated using the capital and labour indices where applicable. Price indices (PIs) have a 2005 base year.

PIs were applied to all elements of the basic price supply and use and transport margins. As a first estimate, and subject to subsequent balancing with national accounts data, PPIs were applied to all intermediate transactions, and CPIs were applied to final demand.

The form of the deflation follows the basic price table:

$$\mathbf{U}^{real}_{1:m,1:n} = diag(\mathbf{PPI}) * \mathbf{U}_{\mathrm{1:m,1:n}}$$

$$\mathbf{y}^{real}_{1:n,b} = diag(\mathbf{CPI}) * \mathbf{y}_{\mathrm{1:n,b}}$$

subject to the above considerations. It should be noted, that whilst export and import price indices were available, these were only available in aggregate, and hence the application of actual constant price import and export data points as constraints provides a more accurate deflation method.

The constraints applied to the constant price table were only available from National Accounts data, and included Total Final demand by destination in purchaser's prices including data on exports; household final demand by product group; value added by industry group; total taxes on products; and competing imports. See the Appendix for more details.

2.5. Constraints

Once the initial estimate was fully populated, constraints were applied across the full time series. Two types of constraints were applied – accounting or balancing constraints (to ensure relationships such as gross output equals gross input); and data constraints (such as from industry, commodity and national accounts data). A list of constraints $\mathbf{c}_l$ was constructed, where the l refers to a subset of constraints rather than individual constraints. $\mathbf{c}_l$ is constructed alongside a concordance matrix $\mathbf{G}$ that links the constraint values $\mathbf{c}_l$ with the initial estimate $\mathbf{z}^0$. $\mathbf{G}$ is simply a sparse binary matrix showing the existence or non-existence of a relationship between $\mathbf{c}_l$ and each element of the vector $\mathbf{z}^0$.

2.5.1 Accounting Constraints

The full list of constraint data are included in the appendix. To briefly summarise, accounting constraints were included to ensure the balance between the expenditure and income approaches to GDP; the column or industry output of supply and use tables; the row or total product output of supply and use tables; the balance of imports including re-exports; the re-distribution of margin tables to margin rows; and the balance of margin tables, including re-exports.

For data constraints, three main data sets are utilised.

2.5.2 Benchmark Data

In years with available benchmark data, balanced input–output tables (in supply and use format) as well as detailed commodity data for a subset of years were available (see Table 2). There is a higher level of confidentiality in the product details data, as well as a greater uncertainty in the data. The comparison (aggregation) to the square IO tables is not always fully consistent. The commodity data and the industry level data of the input output tables are used as constraints on the initial estimates of the IO system, subject also to the accounting constraints above.

Classification Changes. Changes in classification have occurred frequently over the time series in the benchmark IO tables. These changes are mainly due to the changing importance of different products in the economy, as well as the introduction of new products (e.g. computers) non-existent in early years. As such, concordance matrices vary from year to year. These classification changes confound temporal comparability at the level at which the changes occurred. As such, when a classification change occurred, concordance matrices were adjusted such that data were aggregated upwards. That is, if a classification change made it impossible to have temporal consistency between two (or more) products separately, then these products would be aggregated together, and the sum of the products would be used as a constraint (provided there was then temporal consistency between the sum of the products and the final classification, otherwise, further aggregation of the source data was undertaken until consistency was ensured).

Methodological Changes. Numerous methodological changes have occurred over time (Gretton, 2005). These have been dealt with more recently (from 1995) by estimating relative rather than absolute change based on the consistent set of supply-use tables (see below). Prior to 1995, most major changes occurred with the classification, and these were dealt with as above. Other methodological changes were considered relatively minor in terms of overall economic evolution and subject to mathematical balancing.

2.5.3 Purchaser price Supply Use Data

A time series of Supply Use data in purchaser prices is available for Australia, updated according to current methodology, at the conclusion of each year.[1] The most recent data available at the time were from 1994–95 to 2003–2004. The data are published as a Supply table in basic prices and a Use table in purchases prices for 225 commodities and 52 industries. In addition, there are commodity totals for imports, subsidies, taxes, and for a breakdown into 10 margins. The time series of these source tables are all under consistent classification and methodology.

The purchaser price data provided by the ABS were of sufficient deviation from the benchmark input–output tables that a secondary treatment step was deemed beneficial. Rather than using the absolute values depicted within the supply/use tables, growth rates were constructed for each year relative to 2001–2002. 2001–2002 provides a benchmark input–output table under current national accounts methodology. By using ratios to the 2001–2002 values, and because the purchaser price Supply/Use tables are all produced under consistent methodology, it is hence possible to circumnavigate some of the methodological changes introduced during the 1990s in the benchmark input–output data (handling of transport margins in particular). As the purchaser price tables provide a slightly more aggregate dataset, but still for all products under all margins, only if methodological differences were significant between the application of the margins to receiving industries would errors remain. This relates principally to the transport margins, where the 1994–99 benchmark input–output tables were produced under a different method of invoice allocation. Gretton et al. (2004) discuss this issue and provide some examples of the differences to be expected under these years.

This approach can be summarised mathematically as follows. The growth rates are calculated from the raw Supply/Use data $\mathbf{Z}_{SU}^{t}$ at time $t = \{1995,\ldots,2004\}$, relative to the raw Supply/Use data $\mathbf{Z}_{SU}^{2002}$ at $t = 2002$. This ratio is calculated element-wise for each row i and each column j. The 2002 benchmark table $\mathbf{Z}_{Bm}^{2002}$ is collapsed to the same classification of the Supply/Use tables by a row or commodity concordance matrix $\mathrm{W}_{SU,Bm}^{comm}$ between the Supply/Use tables and the classification of the 2002 Benchmark table, and a concordance matrix $\mathrm{W}_{SU,Bm}^{ind}$ relating the columns of the Supply/Use tables to the column classification of the 2002 benchmark table. The Hadamard or element-wise product (denoted $\circ$) is then applied between the collapsed 2002 benchmark table $(\mathbf{W}_{SU,Bm}^{comm})' \times \mathbf{Z}_{Bm}^{2002} \times \mathbf{W}_{SU,Bm}^{ind}$ and the growth rates.

$$\mathbf{Z}_{SU*}^{t} = (\mathbf{W}_{SU,Bm}^{comm})' \times \mathbf{Z}_{Bm}^{2002} \times \mathbf{W}_{SU,Bm}^{ind} \circ \left(\forall i, j \frac{\mathbf{Z}_{SUi,j}^{t}}{\mathbf{Z}_{SU,i,j}^{2002}} \right)$$

The adjusted Supply/Use table $\mathbf{Z}_{SU}^{t}$ is then used as a constraint directly. This process of creating adjusted tables avoids non-unitary coefficients in the balancing process compared with using growth rates or some other form of ratio – with the same effect.

[1] Currently under review by the ABS

2.5.4 National Accounts Data

Finally, National Accounts aggregates are available across the whole time series in current and constant prices. Constraints are hence constructed from all available data. The applicable National Accounts data from 5204.0 includes data on value added by industry, GDP, final demand by destination, competing imports, inputs to production by type, household final demand by product, compensation of employees and gross operating surplus by industry, and finally a dataset of agricultural inputs by industry and primary or intermediate delivery. As previously mentioned, a separate section of the National Accounts provides data in constant prices.

These 38 subsets of Constraints $\mathbf{c}_l$ make up the full list of constraints **c. c** is constructed for each year of the time series. These constraints act on the initial estimate as outlined in the next section.

2.6 System Balancing

In order to update the initial estimate $\mathbf{z}^{t,0}$ at time t to our final estimate, a large amount of information is used. Roughly half a million constraint values are used for the benchmark years. Standard updating procedures (RAS, etc) usually limit themselves to row and column totals, ignoring additional information on purchases, types of final demand and value added, etc. These procedures reduce variances in technological change, matching previous flows to total production only. As such, these procedures can produce flawed estimates when known changes in the economy have occurred. A lot of effort in the literature has been expended on comparing different updating procedures and different target functions when limiting data inputs only to row and column totals. The approach taken here is to leave fewer data to be estimated mathematically by including all applicable data in the balancing process. Such an approach is consistent with the modification to RAS (Paelinck and Walbroeck, 1963; Allen, 1974; Lecomber, 1975) to include constraints on known elements. It has been shown that this approach leads to superior outcomes (Gilchrist and St Louis, 1999, 2004; Lenzen et al., 2006), and considering the data situation in most countries, where data on some sub-groups of flows are easier to come by than updated aggregate production figures, it seems appropriate to include these data when possible (provided it does not bias the balancing process).

The basis of the balancing process is that the initial estimate $\mathbf{z}^{\mathbf{t,0}}$ will not necessarily (and is unlikely to) satisfy the list of constraints $\mathbf{c}^{\mathbf{t}}$. Hence, some type of balancing or optimisation process is required in order to obtain $\mathbf{z}^{t,*}$ such that:

$$\mathbf{z}^{\mathrm{t},*} = \mathbf{z}^{\mathrm{t,n}} \Leftrightarrow \mathbf{c}^{\mathrm{t}} = \mathbf{G}^{\mathrm{t}} \times \mathbf{z}^{\mathrm{t,n}}$$

where $\mathbf{z}^{\mathbf{t,n}}$ is solved iteratively through balancing $\mathbf{c}^{\mathbf{t}}$ and $\mathbf{G}^{\mathbf{t}} \times \mathbf{z}^{\mathbf{t,n}}$ (such that overall information gain in $\mathbf{z}^{\mathbf{t,n}}$ is reduced). As mentioned, the most commonly used technique is that of RAS (Stone and Brown, 1962). The modification from the basic bi-proportional approach into a method that operates on subsets of flows, handles negative values and handles constraint incompatibility is documented elsewhere (Lahr and de Mesnard, 2004; Lenzen et al., 2009), and its critical application in the form of KRAS to the Australian case is also shown elsewhere (Lenzen et al., 2006, 2007, 2009).

As is well known, RAS and variants are a form of mathematical optimisation process where a logarithmic form is used to minimise a target function f that represents the distance of $\mathbf{z}^{\mathbf{t,0}}$ from $\mathbf{z}^{\mathbf{t,n}}$ (Bacharach, 1970):

$$f(\mathbf{z}^{\mathrm{t,n}}, \mathbf{z}^{\mathrm{t,0}}) = \sum_r |z_r^{t,n}| s_r^t \ln(S_r^t)$$

with

$$S_r^t = \frac{z_i^{t,0}}{e * z_r^{t,n}}$$

See Lenzen et al. (2007) for the justification of the presence of the exponential, and de Mesnard (1994) for investigation of equivalency. As the data are from different sources, it is unlikely that the criteria $\mathbf{c}^{\mathbf{t}} = \mathbf{G}^{\mathbf{t}} \times \mathbf{z}^{\mathbf{t,n}}$ are met.

We thus need also to minimise differences between $\mathbf{c}^{\mathbf{t}}$ and $\mathbf{G}^{\mathbf{t}} \times \mathbf{z}^{\mathbf{t,n}}$ according to estimates of the reliability of the constraints, which is called $\mathbf{sc}^{\mathbf{t}}$, assessed in this case through a log-normal regression of known error values. The constraint balance is thus defined as:

$$\mathbf{c}^{\mathrm{t}} = \mathbf{G}^{\mathrm{t}} \times \mathbf{z}^{\mathrm{t,n}} + \alpha^{\mathrm{t}} \cdot \mathbf{sc}^{\mathrm{t}}$$

where α is a vector of the same size as $\mathbf{sc}^{\mathbf{t}}$ and the dot shows scalar multiplication. The target function is then modified to

$$f(\mathbf{z}^{\mathrm{t,n}}, \mathbf{z}^{\mathrm{t,0}}) = \sum_r |z_r^{t,n}| s_r^t \ln(s_r^t) + \sum_r |\alpha_r^t|$$

The implementation of the approach in an iterative fashion is described in an alternate paper (Lenzen et al., 2009) where convergence is said to be achieved once a threshold is reached. In the iterative approach, $\boldsymbol{\alpha}^{\mathbf{t}}$ is advanced stepwise, where each step should be chosen to be sufficiently small to approximate a linear function.

This balancing process was undertaken in stages for each year of the input–output system (1975–2005). The intention of the balancing process was to adjust the system to the hierarchy of data constraints – first to available data for the benchmark years, secondly to consistent time series data from the supply use tables, and thirdly to the full series of consistent national accounts data. This hierarchy is chosen so that temporal consistency is most paramount (i.e. the national accounts is the most binding constraint set), and also works from the most detailed constraint set upwards to the most aggregate constraint set.

Two options are available for establishing a hierarchy – either embedding the hierarchy ‘softly’ in the reliability term $\mathbf{sc}^{\mathbf{t}}$, or by enforcing the hierarchy through successive balancing. The first approach was taken here for within the three main data sets, and the second approach was taken for between the data sets. Within each dataset, a factor 10 preference was established such that more aggregate data were given preference over less aggregate data. This implies that in case of conflict, a constraint from a less aggregate data set would move 10 times that of a constraint from a more aggregate data set (e.g. the 100 sector tables versus the detailed commodity data at 1000 levels of detail). Obviously the more consistent the datasets, the more accurate the results, as the less balancing occurs between the constraint data. Where constraints are from the same level of aggregation, they are moved according to the error estimation. When implementing the balancing through the hierarchy, only three different main data sets were used, making consecutive runs relatively straightforward. The balancing process is called in exactly the same way for each constraint set, albeit restricted to the particular subset of constraints $\mathbf{c^t}$ for that dataset (and also inclusive of all accounting constraints).

The constraint data $\mathbf{c^t}$ contained positive, negative and zero constraints, as would be expected in national account and input–output databases, and the matrix of flows $\mathbf{Z^t}$ also contains positive, negative and zero entries, again, as to be expected in an input–output table. The generalisation of RAS is able to handle such variation; however, whilst conflicting constraints of opposite sign can converge, the balancing of $\mathbf{Z^t}$ is sign invariant. A search algorithm was run through all constraint data for each year to identify which elements of $\mathbf{Z^t}$ were negative. Such an algorithm is not straightforward when data are in various levels of aggregation, but the basic assumption employed was that all data were positive unless a disaggregate constraint showed the flows of $\mathbf{Z^t}$ to be negative. Where a constraint showed $\mathbf{Z^t}$ to be negative, all flows addressed by this constraint were assumed negative, unless an even more disaggregate constraint showed some flows to be positive, and so forth.

System convergence was helped greatly by correct sign definition of $\mathbf{Z^t}$. GDP had the strongest affect of the balancing constraints due to its simultaneous affect on row/column balances through the income and expenditure approaches of measuring GDP. Convergence (occurring when differences between $\mathbf{c^t}$ and $\mathbf{G^t} \times \mathbf{z^{t,n}}$ are less than a specified threshold) was achieved after several thousand iterations of the KRAS algorithm taking several hours of run time. For more general results on the performance of the KRAS algorithm in these problems, the reader is referred to Lenzen et al. (2009).

3 STABILITY, PREDICTABILITY AND ACCURACY

Following the construction effort of the time series, this section gives an investigation into how much change could be expected in input–output coefficients over time for Australia and the accuracy of using an old input–output table compared with a new input–output table. Three components are given – an analysis of stability of coefficients and multipliers of the time series (Section 3.2); an analysis of predictability given known final demand or value added (Section 3.3); and finally (Section 3.4), we return to the need for updated tables – i.e. the expectant error from using dated tables.

It is important to note that this section does not seek to validate the methods and data of Section 2. The temporal inconsistency of the source data makes a comparison between source data and the final time series infeasible.

3.1. Input–Output Model Choice

The symmetric input–output model used in this section is calculated from the derived supply and use tables under the industry technology assumption in industry by industry format (United Nations Department for Economic and Social Affairs Statistics Division

1999). In the Leontief (demand driven) model, gross output **x** is defined as the summation of intermediate transactions **T** and final demand **y. e** is defined as the summation vector (a vector of ones, size $n \times 1$):

$$\mathbf{x} = \mathbf{T}\mathbf{e} + \mathbf{y} \tag{1}$$

A is defined as the direct requirements or coefficients matrix, relating proportions of intermediate usage to gross output (the hat defining diagonalisation):

$$\mathbf{A} = \mathbf{T} * \widehat{\mathbf{x}}^{-1} \tag{2}$$

The Leontief model then redistributes to:

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{y} \tag{3}$$

$$\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{y} \tag{4}$$

where **I** is the identity matrix.
In the Ghosh (supply driven) model,

$$\mathbf{x}' = \mathbf{e}'\mathbf{T} + \mathbf{v}' \tag{5}$$

$$\mathbf{B} = \mathbf{T}' * \widehat{\mathbf{x}}^{-1} \tag{6}$$

$$\mathbf{x}' = \mathbf{x}'\mathbf{B} + \mathbf{v}' \tag{7}$$

$$\mathbf{x}' = \mathbf{v}'(\mathbf{I} - \mathbf{B})^{-1} \tag{8}$$

The dash represents transposition.

3.2 Stability

As a precursor to investigating the predictability of the Leontief and Ghosh models, following Dietzenbacher and Hoen (2006, hereafter D&H), the stability of the coefficients matrix is first investigated. Alternative methods to investigate stability include de Mesnard's work (de Mesnard 1997, 2004) on biproportional filters. As opposed to the use of data in current price in D&H, here the deflated IO tables in constant prices are used in order to exclude price-related fluctuations in the analysis.[2]

As per D&H, the coefficient of variation is utilised to see the change in model parameters over time. The coefficient of variation **cvar** is defined as the average square

[2] Note: the analysis was performed for both current and constant price data – with the conclusions drawn not being affected, but as expected, less variation shown in the constant price data set. Owing to space limitations, the results are not included here, but available from the author.

root of the variance of a variable **z** over a time period **t** = 1,. . ..*p*, i.e.:

$$cvar_r = \frac{\sqrt{var_r}}{\overline{z_r}} \tag{9}$$

where mean $\overline{z_r}$ is calculated according to

$$\overline{z_r} = \frac{\sum_{t=1:\,p} z_r^t}{p} \tag{10}$$

and variance is calculated according to

$$var_r = \sum t = 1 : p \frac{(z_r^t - \overline{z_r})^2}{p} \tag{11}$$

The coefficient of variation of the Leontief coefficients matrix **A** (Figure 2) shows the changing technological relationships. As is expected, the greater relative change has occurred within smaller coefficients. The magnitude of **cvar** is highly dependent on aggregation level, such that the results presented have a much larger **cvar** than those found in D&H with only 12 sectors. Comparing the Leontief model (Figure 2) with the Ghosh model (Figure 3), there is greater variation in the Ghosh compared with the Leontief. In fact, the average coefficient of the Ghosh model was some 25% larger than the Leontief model, implying smaller primary inputs compared with final consumption, whilst the median coefficient was 15% larger in the Ghosh model, implying a greater concentration of inputs than outputs on average.

FIGURE 2. Coefficient of variation **cvar** of the Leontief coefficients matrix *A*.

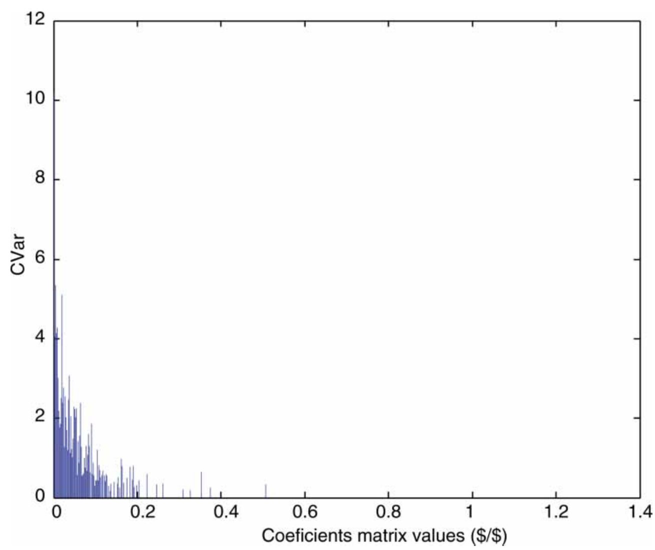
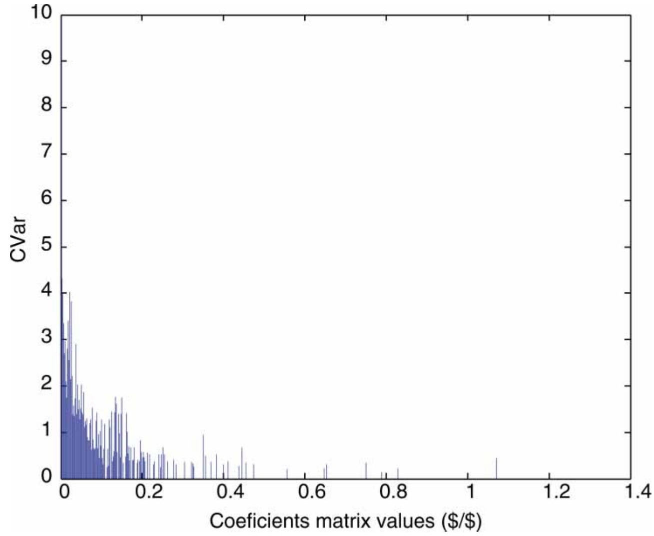


FIGURE 3. Coefficient of variation **cvar** of the Ghosh coefficients matrix *B*.



The coefficient of variation for both the Leontief and Ghosh models for large coefficients (coefficients greater than 0.1, or 10% of total inputs/outputs of the respective sectors) is generally in the range of 0.5 to 1.0, meaning the standard deviation of the coefficients across time are in the same order as the mean of the coefficients.

Of greater impact is the variation observed in multipliers rather than coefficients. Whilst the direct relationships of **A** and **B** may change depending on the presence or lack of presence of intermediate sectors, when mapping environmental flows, of

FIGURE 4. Coefficient of variation **cvar** of the Leontief multiplier matrix *L*.

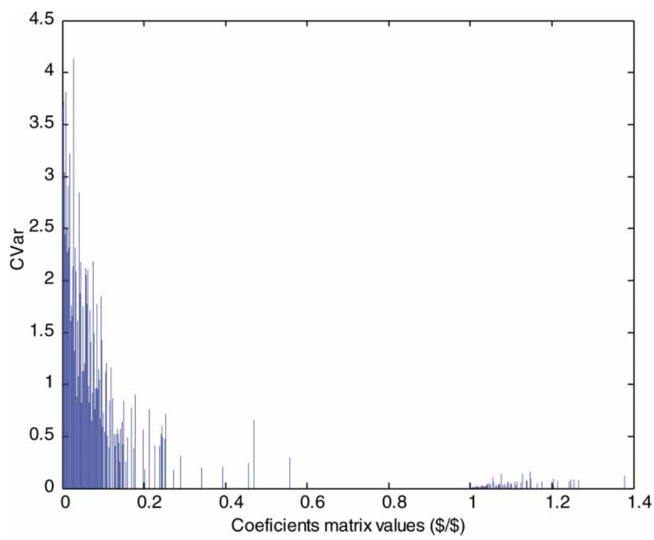
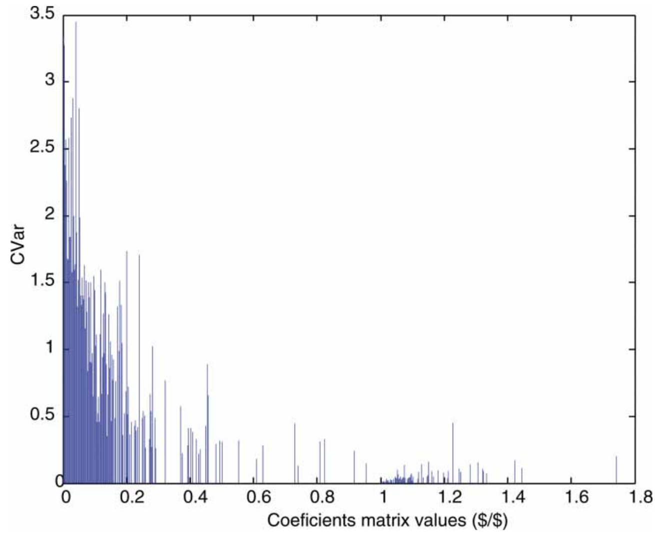


FIGURE 5. Coefficient of variation **cvar** of the Ghosh multiplier matrix G.



greatest interest is the variability in total (direct and indirect) relationships. Again, both Leontief (Figure 4) and Ghosh (Figure 5) multipliers show increased variation in small multipliers, with the Ghosh this time showing much greater variation in comparison with the Leontief matrix. The cluster of multiplier values greater than 1 represents the diagonals of the multiplier matrices, which interpret the direct linkage between final consumption and primary inputs. As would be expected, the coefficient of variation of these linkages is very minor. The relatively larger coefficients of variations of the Ghosh model imply this model is less accurate when predicting important interrelationships in the economy. The observed difference between the Ghosh and Leontief models can be explained by a greater variation in primary inputs compared with final consumption.

3.3 Predictability (Joint Stability of Supply and Demand Multipliers)

In this section, the predictability of multipliers of the Leontief and Ghosh production functions is analysed. By predictability, the lead of D&H is followed in referring to what Bon named ‘stability’. Again, in this and the ensuing section, tables in constant prices are used in order to compensate for price effects.

In this study, the predictability of the Leontief model is assessed under known final demand $\mathbf{y}_{t+1}$ by comparing known gross output $\mathbf{x}_t$ with estimated gross output $\widetilde{\mathrm{x}}_{t+1}$ calculated according to:

$$\widetilde{\mathbf{x}}_{t+1} = (\mathbf{I} - \mathbf{A}_t)^{-1}\mathbf{y}_{t+1} \tag{12}$$

The assumption being tested is that the coefficients matrix $\mathbf{A}$ is reasonably constant over time. For large variation in $\mathbf{A}$, a large difference in $\widetilde{\mathbf{x}}_{t+1} - \mathbf{x}_t$ will be evident.

The predictability of the Ghosh model under known value added $\mathbf{v}_{t+1}$ is assessed by comparing known gross output $\mathbf{x}_t$ with estimated gross output $\widetilde{\mathbf{x}}_{t+1}$ calculated according to:

$$\widetilde{\mathbf{x}}'_{t+1} = \mathbf{v}'_{t+1}(\mathbf{I} - \mathbf{B}_t)^{-1} \tag{13}$$

In assessing predictability, five tests are used to determine the differences between the estimated gross output and the known gross output. Each test behaves slightly differently, and no test is deemed conclusive (Butterfield and Mules, 1980). Refer to Lenzen et al. (2009) for the source of these measures:

The relative arithmetic mean of absolute differences

$$AMAD = \frac{\sum_i |x_{t\,i} - \tilde{x}_{t\,i}|}{\sum_i x_{t\,i}} \tag{14}$$

The relative geometric mean of absolute differences

$$GMAD = \frac{\sqrt{\sum_i (x_{t\,i} - \tilde{x})^2_{t\,i}}}{\sqrt{\sum_i x^2_{t\,i}}} \tag{15}$$

The Isard/Romanoff Similarity Index which has been used as a 'Dissimilarity Index' by Thissen and Löfgren (1998) and Lenzen et al. (2009)

$$SIM = \frac{\sum_i \frac{|x_{t\,i} - \tilde{x}_{t\,i}|}{x_{t\,i} + \tilde{x}_{t\,i}}}{N^2} \tag{16}$$

The arithmetic mean of relative differences

$$AMRD = \frac{\sum_i \frac{|x_{t\,i} - \tilde{x}_{t\,i}|}{x_{t\,i}}}{N^2} \tag{17}$$

The relative mean square error

$$RMSE = \sqrt{\frac{1}{N}\sum_i \left(\frac{(x_{t\,i} - \tilde{x}_{t\,i})}{x_{t\,i}}\right)^2} \tag{18}$$

One of the main differences is in the weighting of the differences, whether the difference is relative to the single variable or to the sum over the variables. The RMSE is relative to the single variable, and is more influenced by small flows. This is less than ideal in an input–output system when small flows have less importance. It is included here for comparability to D&H.

TABLE 4. Predictability of Leontief coefficients, 1976–2006.

	AMAD	GMAD	SIM	AMRD	RMSE
1977	6%	1%	4%	8%	10%
1978	6%	1%	4%	7%	9%
1979	9%	3%	5%	9%	13%
1980	9%	2%	5%	10%	20%
1981	6%	2%	3%	5%	9%
1982	6%	1%	4%	7%	12%
1983	4%	1%	2%	4%	7%
1984	11%	3%	6%	12%	19%
1985	3%	1%	2%	4%	7%
1986	3%	1%	2%	3%	6%
1987	10%	3%	4%	8%	11%
1988	4%	1%	2%	4%	6%
1989	3%	1%	2%	4%	5%
1990	10%	4%	5%	9%	15%
1991	10%	4%	4%	11%	37%
1992	3%	1%	2%	4%	5%
1993	6%	2%	2%	5%	8%
1994	9%	3%	5%	10%	19%
1995	19%	4%	12%	32%	33%
1996	6%	1%	4%	7%	10%
1997	5%	1%	4%	9%	21%
1998	5%	1%	3%	6%	8%
1999	5%	1%	3%	6%	8%
2000	6%	1%	4%	8%	11%
2001	3%	1%	2%	4%	6%
2002	3%	0%	2%	5%	8%
2003	3%	0%	3%	5%	9%
2004	2%	0%	2%	4%	5%
2005	7%	1%	5%	9%	15%
2006	5%	1%	3%	6%	6%
Average	6%	2%	4%	8%	12%

Average results (Tables 4 and 5) are in line with D&H and Bon, with the Leontief and Ghosh results being comparable in magnitude of error. A noticeable year is 1995, with larger errors evident in this year across both models. This is the first year the Time series of Supply/Use tables in purchaser prices was published by the ABS, and mark a significant change in methodology.

Figures 6 and 7 present the breakdown by industry, with sectors arranged from 1 to 109 (primary industries on the left and tertiary industries on the right). Like D&H, there appears to be less variation in the Leontief model in the more established primary sectors, and considerably more variation in the Leontief model in tertiary sectors. This result, whilst consistent with D&H is opposite to the findings of Bon. D&H's explanation was that they were studying post-war reconstruction. The same explanation does not apply to this study, although there has been considerable change away from agricultural industries and to knowledge services in Australia.

TABLE 5. Predictability of Ghosh coefficients, 1976−2006.

	AMAD	GMAD	SIM	AMRD	RMSE
1977	8%	1%	4%	8%	9%
1978	7%	1%	4%	8%	13%
1979	5%	3%	3%	6%	9%
1980	6%	2%	3%	6%	12%
1981	6%	2%	3%	5%	7%
1982	5%	1%	3%	5%	7%
1983	4%	1%	3%	5%	10%
1984	8%	3%	5%	9%	13%
1985	3%	1%	2%	3%	4%
1986	2%	1%	1%	3%	4%
1987	4%	3%	2%	5%	7%
1988	3%	1%	2%	4%	6%
1989	3%	1%	2%	4%	5%
1990	7%	4%	4%	9%	14%
1991	12%	4%	6%	14%	17%
1992	4%	1%	2%	4%	6%
1993	4%	2%	3%	5%	7%
1994	9%	3%	5%	11%	18%
1995	20%	4%	11%	26%	31%
1996	6%	1%	4%	8%	12%
1997	6%	1%	4%	8%	11%
1998	5%	1%	4%	7%	12%
1999	4%	1%	2%	5%	9%
2000	6%	1%	5%	8%	14%
2001	4%	1%	3%	5%	8%
2002	3%	0%	3%	6%	15%
2003	4%	0%	3%	7%	10%
2004	3%	0%	2%	5%	7%
2005	6%	1%	5%	10%	14%
2006	5%	1%	3%	7%	7%
Average	6%	2%	3%	7%	11%

3.4 Analysis of Temporal Predictability

In the previous section, the predictability of $\mathbf{x}_t$ was calculated from $\mathbf{L}_{t-1}$. In essence, the gross output was calculated given current final demand and the previous year's technological structure. A common occurrence, however, is that there is often a greater lag between available input–output tables and current final demand. Analysts often cite the stability of technological coefficients as justification for using dated structural data. In this section, this assumption is analysed for Australia for 2005 final demand (again using constant price tables). The series of equations investigated are represented by:

$$\forall_{t=1975:2005}\{\widetilde{\mathrm{x}}_t = (\mathrm{I} - \mathrm{A}_t)^{-1}\mathrm{y}_{t=2005}\} \tag{19}$$

FIGURE 6. RMSE of Leontief coefficients across time series.

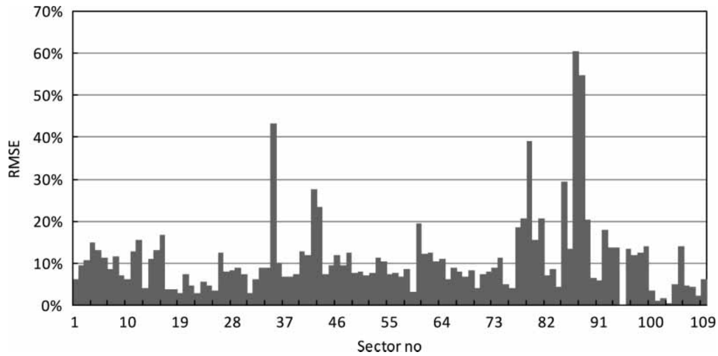
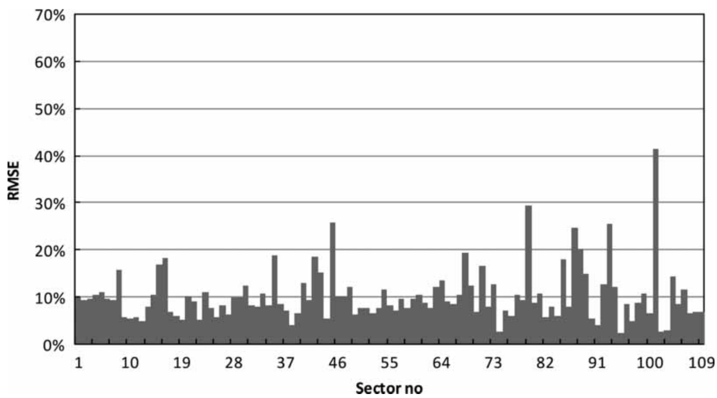


FIGURE 7. RMSE of Ghosh coefficients across time series.



The measure of AMAD is used which gave one of the better representations in Section 0 to map the evolution of $\tilde{x}_t$

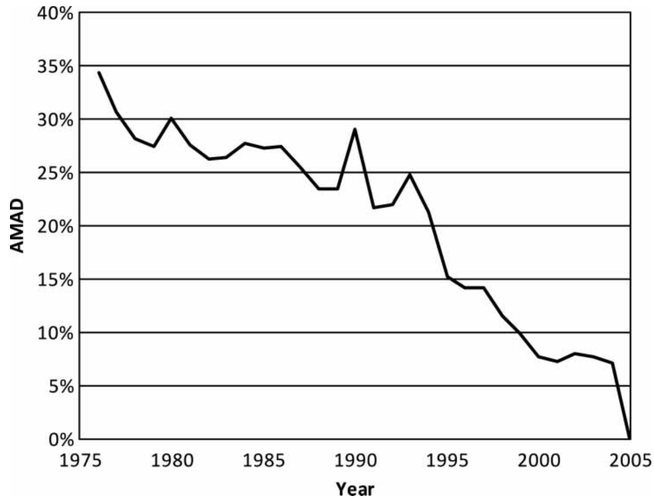
$$AMAD = \frac{\sum_i \left| x_{2005,i} - \tilde{x}_{t\ i} \right|}{\sum_i x_{2005,i}} \quad (20)$$

Figure 8 presents the AMAD between the estimated gross outputs $\tilde{x}_t$ to 2005 gross output $\mathbf{x}_{2005}$.

As expected, gross output estimates shown in Figure 8 are poor at a greater distance from the benchmark year (2005). A general trend ($R^2 = 0.89$) is found between the temporal distance from the benchmark and the measure of error. Using previous or subsequent year data gives an error in the range of 4–7%, as previously examined in Table 5.

Most statistical agencies have a lag time of three–four years in producing IO tables (Australia currently has a three year lag, the UK's last published tables are from 1995). From the results presented here, it would be expected that using the dated tables would

FIGURE 8. Evolution of gross output estimates $\widetilde{x}$ compared to real $\mathbf{x}_{2005}$.



give an error of absolute differences in the range of 5–15%. However, as the measures employed present only the absolute error in industrial output, the error may be reduced if the uncertainty is stochastic (only somewhat likely, as we are seeing economic evolution) and normally distributed.

4 CONCLUSIONS AND DISCUSSION

This paper describes the construction of a time series of Australian input–output tables. A range of detailed commodity data were utilised in order to estimate a temporally consistent set, containing 344 sectors. Initial estimates were generated using an exponential regression, and were subsequently balanced by applying a range of accounting constraints and data sources. Tables were constrained firstly by published IO data (17 tables over 30 years), secondly by adding in unpublished supply/use data (10 recent years), and finally by including National Accounts data (all years). This enabled a consistent time series to be constructed, whilst incorporating the highest level of detail from published and unpublished IO tables in order to delineate environmentally important processes.

The stability and predictability of input–output coefficients are important considerations when using input–output data. Often the analyst has available current final expenditure data, but older data on technical coefficients. Knowledge of the evolution of technical coefficients can help inform the validity of modelling exercises. Hence, in this paper, the characteristics of the temporal model developed were explored along the lines of Dietzenbacher and Hoen (2006) in order to give comparability for the Australian case. To begin with, the issue of stability of coefficients and multipliers was investigated for the Leontief and Ghosh models of supply/demand. As expected, there is considerably more variation in smaller coefficients than large coefficients. Multipliers were generally found to be more stable than coefficients. The interpretation from this finding is that adjustments to individual components of production chains are greater than adjustments to the corresponding full production chains, so some substitution of intermediate inputs is occurring.

A comparison between the Ghosh and Leontief models showed superior performance of the Leontief model, with the Ghosh being subject to increased variation in some large coefficients and multipliers. Hence, in Australia, there has been more stability in demand side components than supply side components.

The predictability of the models was then examined under updated final demand or primary input components. The model results showed errors in the 0 – 15% range depending on choice of metric. The average root mean square error of 11 – 12% was higher than found in Dietzenbacher and Hoen (2006), most likely due to the significantly higher disaggregation level used in this study. Using error measures that weigh out large changes in small values, the errors were a more reasonable 2–8%.

No significant difference between the two models could be discerned at the whole economy level. Whilst differences were observed in model performance at the industry level, no reliable trends were found for a maturing economy. This is in line with Dietzenbacher and Hoen (2006), but contradictory to Bon's findings.

The predictability of the input–output model over varying time scales was also investigated. The results showed the importance of using up-to-date coefficient matrices, with weighted error measurements subsiding from almost 35% on a 30 year time scale to nearer 5% for previous year tables.

Acknowledgements

Thanks go to the reviewers and editor (Manfred Lenzen) for the critiques and significant suggestions that have helped to improve the clarity and focus of this paper.

References

Allen, R.I.G. (1974) Some Experiments with the RAS Method of Updating Input–Output Coefficients. *Oxford Bulletin of Economics and Statistics*, 36, 217–228.

Ang, B.W. (2000) A Survey of Index Decomposition Analysis in Energy and Environmental Studies. *Energy*, 25, 1149–1176.

Australian Bureau of Statistics (1998) *Business Register Counts 1998*. Canberra, Australia, Australian Bureau of Statistics.

Australian Bureau of Statistics (2006) Australian National Accounts, Input–Output Tables (Product Details), 2001-02, Cat. no. 5215.0.55.001.

Australian Bureau of Statistics (2007a) 6204.0.55.001 – Labour Force, Australia, Detailed, Quarterly. Canberra, Australia, Australian Bureau of Statistics.

Australian Bureau of Statistics (2007b) Australian National Accounts, ABS Catalogue No. 5204.0. Canberra, Australia, Australian Bureau of Statistics.

Australian Bureau of Statistics (2008a) 6291.0.55.003 – Labour Force, Australia, Detailed, Quarterly. Canberra, Australia, Australian Bureau of Statistics.

Australian Bureau of Statistics (2008b) Australian National Accounts, Input – Output Tables, 2004–05, Cat. no. 5209.0.55.001. Canberra, Australia, Australian Bureau of Statistics.

Australian Bureau of Statistics (2008c) Consumer Price Index, Australia, Sep 2008, Cat. no. 6401.0. Canberra, Australia, Australian Bureau of Statistics.

Australian Bureau of Statistics (2008d) Producer Price Index, Australia, Sep 2008, Cat. no. 6427.0. Canberra, Australia, Australian Bureau of Statistics.

Bacharach, M. (1970) *Biproportional Matrices & Input–Output Change*. Cambridge, Cambridge University Press.

Bon, R. (1984) Comparative Stability Analysis of Multi-Regional Input–Output Models: Column, Row, and Leontief–Strout Gravity Coefficient Models. *Quarterly Journal of Economics*, 99, 79–815.

Bon, R. (1986) Comparative Stability Analysis of Demand-Side and Supply-Side Input–Output Models. *International Journal of Forecasting*, 2, 231–235.

Bon, R. (1988) Supply-Side Multiregional Input–Output Models. *Journal of Regional Science*, 28, 41–50.

Bon, R. and X. Bing (1993) Comparative Stability Analysis of Demand-Side and Supply-Side Input–Output Models in the UK. *Applied Economics*, 25, 75–79.

Butterfield, M. and T. Mules (1980) A Testing Routine for Evaluating Cell by Cell Accuracy in Short-Cut Regional Input–Output Tables. *Journal of Regional Science*, 20, 293–310.

de Mesnard, L. (1994) Unicity of Biproportion. *SIAM Journal of Matrix Analysis Applications*, 15, 490–495.

de Mesnard, L. (1997) A Biproportional Filter to Compare Technical and Allocation Coefficient Variations. *Journal of Regional Science*, 37, 541–564.

de Mesnard, L. (2004) Biproportional Methods of Structural Change Analysis: A Typological Survey. *Economic Systems Research*, 16, 205–230.

Dietzenbacher, E. and A.R. Hoen (2006) Coefficient Stability and Predictability in Input–Output Models: A Comparative Analysis for the Netherlands. *Construction Management and Economics*, 24, 671–680.

Dietzenbacher, E. and B. Los (1998) Structural Decomposition Techniques: Sense and Sensitivity. *Economic Systems Research*, 10, 307–323.

Foran, B., J. Lennox, M. Lenzen, C. Dey and R. Wood (2003) Meeting. Canberra, Australia, Resource Futures, CSIRO Sustainable Ecosystems.

Gallego, B. and M. Lenzen (2008) Estimating Generalised Regional Input–Output Systems: A Case Study of Australia. In: M. Ruth and B. Davidsdottir (eds.) *Dynamics of Industrial Ecosystems*. Boston, MA, MIT Press.

Gilchrist, D.A. and L.V. St Louis (1999) Completing Input–Output Tables using Partial Information, With an Application to Canadian Data. *Economic Systems Research*, 11, 185–193.

Gilchrist, D.A. and L.V. St Louis (2004) An Algorithm for the Consistent Inclusion of Partial Information in the Revision of Input–Output Tables. *Economic Systems Research*, 16, 149–156.

Gretton, P. (2005) Australian Input–Output Tables. *The Australian Economic Review*, 38, 319–332.

Gretton, P., I. Bobbin and M. Horridge (2004) A User's Perspective on the Treatment of Transport Activities in Input–Output Tables and Suggestions for the Future. *International Conference on Input–Output and General Equilibrium: Data, Modelling and Policy Analysis*, Brussels, Belguim.

Hewings, G.J.D., M. Sonis and R.C. Jensen (1988) Fields of Influence of Technological Change in Input–Output Models. *Papers of the Regional Science Association*, 64, 25–36.

Hoekstra, R. and J.C.J.M. van den Bergh. (2002) Structural Decomposition Analysis of Physical Flows in the Economy. *Environmental and Resource Economics*, 23, 357–378.

Jackson, R.W. and A.T. Murray (2004) Alternative Input–Output Matrix Updating Formulations. *Economic Systems Research*, 16, 135–148.

Jensen, R.C. and G.R. West (1980) The Effect of Relative Coefficient Size on Input–Output Multipliers. *Environment and Planning A*, 12, 659–670.

Lahr, M.L. and L. de Mesnard (2004) Biproportional Techniques in Input–Output Analysis: Table Updating and Structural Analysis. *Economic Systems Research*, 16, 115–134.

Lecomber, J.R.C. (1975) A Critique of Methods of Adjusting, Updating and Projecting Matrices. In: R.I.G. Allen and W.F. Gossling (eds.) *Estimating and Projecting Input–Output Coefficients*. London. Input–Output Publishing Company, 1–25.

Lenzen, M. (2011) Aggregation Versus Disaggregation in Input–Output Analysis of the Environment. *Economic Systems Research*, 23, 73–89.

Lenzen, M., L.-L. Pade and J. Munksgaard (2004) CO_2 Multipliers in Multi-Region Input–Output Models. *Economic Systems Research*, 16, 391–412.

Lenzen, M., B. Gallego and R. Wood (2006) A Flexible Approach to Matrix Balancing under Partial Information. *Journal of Applied Input–Output Analysis*, 11,12.

Lenzen, M., R. Wood and B. Gallego (2007) Some Comments on the GRAS Method. *Economic Systems Research*, 19, 461–465.

Lenzen, M., B. Gallego and R. Wood (2009) Matrix Balancing under Conflicting Information. *Economic Systems Research*, 21, 23–44.

Lenzen, M., R. Wood and T. Wiedmann (2010) Uncertainty Analysis for Multi-Region Input–Output Models – A Case Study of the UK'S Carbon Footprint. *Economic Systems Research*, 22, 43–63.
Mastoris, I., J. Travis and A. Welsh (2002) Regional Input–Output Tables – Australia. In: B.V. Dimaranan and R.A. McDougall (eds.) *Global Trade Assistance and Protection: The GTAP 5 Data Base*. Indianapolis, IN, Center for Global Trade Analysis, Purdue University.
Mínguez, R., J. Oosterhaven and F. Escobedo (2009) Cell-Corrected RAS Method (CRAS) for Updating or Regionalizing an Input–Output Matrix. *Journal of Regional Science*, 49, 329–348.
Paelinck, J. and J. Walbroeck (1963) Etude empirique sur l'évolution de coefficients 'input–output'. *Economie Appliquée*, 16, 81–111.
Percoco, M., G.J.D. Hewings and L. Senn (2006) Structural Change Decomposition through a Global Sensitivity Analysis of Input–Output Models. *Economic Systems Research*, 18, 115–131.
Reich, U.-P. (2008) Additivity of Deflated Input–Output Tables in National Accounts. *Economic Systems Research*, 20, 415–428.
Stone, R. and A. Brown (1962) *A Computable Model of Economic Growth*. London, Chapman and Hall.
Suh, S., M. Lenzen, G.J. Treloar, H. Hondo, A. Horvath, G. Huppes, O. Jolliet, U. Klann, W. Krewitt, Y. Moriguchi, J. Munksgaard and G. Norris (2003) System Boundary Selection in Life-Cycle Inventories Using Hybrid Approaches. *Environmental Science & Technology*, 38, 657–664.
Thissen, M. and H. Löfgren (1998) A New Approach to SAM Updating with an Application to Egypt. *Environment and Planning A*, 30, 1991–2003.
United Nations Department for Economic and Social Affairs Statistics Division. (1999) *Handbook of Input–Output Table Compilation and Analysis*. New York, United Nations.
Wood, R. and C.J. Dey (2009) Australia's Carbon Footprint. *Economic Systems Research*, 21, 243–266.
Wood, R. and M. Lenzen (2006) Zero-Value Problems of the Logarithmic Mean Divisia Index Decomposition Method. *Energy Policy*, 34, 1326–1331.

APPENDIX

The accounting or balancing and data constraints of the input–output system are defined in this appendix. The notation used is to help show the links between the data points and the final input–output system (see Section 2.1.3). It is the constraint list **c** and the concordance matrix **G** that are constructed (see Section 2.6) (using the indexing vector described in Section 2.1.3) in the preparation of these constraints.

Accounting Constraints

First, define GDP relationships, by income approach − summation of value added (gross operating surplus, compensation of employees, taxes on production) and taxes on products:

$$\mathbf{C}_1 \Rightarrow \sum_{i=m+1,m+a} \sum_{j=1,n+b} U_{ij} - gdp = 0$$

and expenditure approach – summation of domestic final demand and exports (including taxes less subsidies), minus imports (excluding re-exports):

$$\mathbf{C}_2 \Rightarrow \sum_{i=1,m+a} \sum_{j=n+1,n+b} \mathbf{U}_{ij} - \left(\sum_{i=1,m} \sum_{j=1,n} \mathbf{M}_{ij} \right) - gdp = 0$$

Match column totals of supply, use and imports matrices:

$$\mathbf{C}_3 \Rightarrow \sum_{i=1,m} \mathbf{V}_{i,1:n} - \left(\sum_{i=1,m+a} \mathbf{U}_{i,1:n} + \sum_{i=1,m} \mathbf{M}_{i,1:n} \right) = 0$$

and Match row totals of supply, and use matrices

$$\mathbf{C}_4 \Rightarrow \sum_{j=1,n} \mathbf{V}_{1:m,j} - \sum_{j=1,n+b} \mathbf{U}_{1:m,j} = 0$$

Supply of imports is equal to competing imports into production (PIIMP)

$$\mathbf{C}_5 \Rightarrow \sum_{j=1,n+b} \mathbf{M}_{1:m,j} - \mathbf{U}'_{PIIMP,1:n} = 0$$

And for re-exports (REX),

$$\mathbf{C}_6 \Rightarrow \sum_{i=1:m} \mathbf{M}_{i,n+1:n+b} - \mathbf{U}_{PIIMP,n+1:n+b} = 0$$

Net taxes (less subsidies) on products are the aggregation of the respective tables on taxes, including taxes on re-exports.

$$\mathbf{C}_7 \Rightarrow \sum_{k=1:4} \sum_{i=1:m+1} \mathbf{T}^k_{i,1:n+b} - \mathbf{U}_{Tax,1:n+b} = 0$$

Net margins are not simply the aggregation of the d margins tables, but the redistribution of the margins. Using a concordance vector W^k for $k = 1{:}d$ margins, which maps margin table totals to margin rows of **P**.

$$\mathbf{C}_8 \Rightarrow \mathbf{P}_{1:m,1:n+b} - \sum_{k=1:d} \mathbf{P}^k_{1:m,1:n+b} - \sum_{k=1:d} \mathbf{W}^k x \sum_{i=1:m} \mathbf{P}^k_{i,1:n+b} = 0$$

Further, margin commodities are 100% related to each margin (i.e. there are no non-margin components):

$$\mathbf{C}_9 \Rightarrow \underset{k=1:d}{\forall} \mathbf{W}^k \times \mathbf{P}_{1:m,1:n+b} = 0$$

According to Australian input–output tables (Australian Bureau of Statistics, 2008), margins on re-exports are excluded from each margin commodity and included in the total. i.e.

$$\mathbf{C}_{10} \Rightarrow \mathbf{P}_{REX,1:n+b} = \sum_{k=1:d} \mathbf{P}^{k}_{REX,1:n+b}$$

National accounts data are mainly available in purchases prices, hence, a Use table in purchases prices is constructed

$$\mathbf{C}_{11} \Rightarrow \mathbf{U}^{PP}_{1:m,1:n+b} - \left(\mathbf{U}_{1:m,1:n+b} + \sum_{k=1:4} \mathbf{T}^{k}_{1:m,1:n+b} + \mathbf{P}_{1:m,1:n+b} \right) = 0$$

Due to varying aggregation levels of data sources used, the Supply matrix is kept as homogeneous as possible. Hence, for the respective industry data, a concordance matrix $\mathbf{W}^{comm,ind}$ mapping commodities to industry units is used such that:

$$\mathbf{C}_{12} \Rightarrow \mathbf{V}_{1:m,1:m} - \sum_{j=1,n} \mathbf{V}_{1:m,j} * \mathbf{W}^{\mathrm{comm,ind}}_{\mathrm{1:m,j}} = 0$$

Benchmark Data

The 17 years of published IO data are referred to as the ‘benchmark’ years. In these years, the data available includes balanced input–output tables as well as commodity data. Some conflicts are evident in the commodity data and within the input–output tables such that the system whilst close, is not fully in equilibrium. The commodity data and the industry level data of the input output tables is used as constraints on the initial estimates of the IO system, subject also to the accounting constraints above. Hence, $\mathbf{Z}_{Bm}$ is used to represent a component of the IO system (such as **U, P, T, V**) in original format from the benchmark years, and $\mathbf{W}^{n,ind}_{Bm}$ to represent the concordance matrix from the benchmark data to the respective table in the estimated IO system ($\mathbf{Z}^{*}$). A dash ’ refers to the transpose.

For square systems, such as the transactions of the Supply or Use matrix, where the classification of the products m is equal to that of industries n:

$$\mathbf{C}_{13} \Rightarrow \mathbf{Z}^{*m,n}_{i,j} = \mathbf{W}^{n,ind}_{Bm} \times \mathbf{Z}^{ind,ind}_{Bm,i,j} \times \left(\mathbf{W}^{n,ind}_{Bm}\right)'$$

For rows (e.g. Value Added):

$$\mathbf{C}_{14} \Rightarrow \mathbf{Z}^{*a,n}_{i,j} = \mathbf{Z}^{a,ind}_{Bm,i,j} \times \left(\mathbf{W}^{n,ind}_{Bm}\right)'$$

For columns (e.g. Final Demand):

$$\mathbf{C}_{15} \Rightarrow \mathbf{Z}^{*m,b}_{i,j} = \mathbf{W}^{n,ind}_{Bm} \times \mathbf{Z}^{ind,b}_{Bm,i,j}$$

And for commodity data (Australian National Accounts, Input–output Tables (Product Details), Cat. no. 5215.0.55.001), using a concordance matrix $\mathbf{W}^{m,comm}_{Bm}$ to map between estimated product data and benchmark commodity data $\mathbf{Z}^{comm,b}_{Bmi,j}$. As benchmark commodity data are only available by row, with columns still in industry classification, the benchmark data $\mathbf{Z}_{Bmi,j}^{comm,b}$ is pre-multiplied by the commodity concordance, and post-multiplied by industry concordance:

$$C_{16} \Rightarrow Z_{i,j}^{*m,n} = W_{Bm}^{m,comm} \times Z_{Bm,i,j}^{comm,ind} \times (W_{Bm}^{n,ind})'$$

Categories of final demand (or other columns) are simply pre-multiplied:

$$\mathbf{C}_{17} \Rightarrow \mathbf{Z}_{i,j}^{*m,b} = \mathbf{W}_{Bm}^{n,comm} \times \mathbf{Z}_{Bmi,j}^{comm,b}$$

There is no additional row data for value added at the commodity level.

The constraints $C_{13} - C_{17}$ are calculated for each available table – i.e. for each of the Supply, Use, Imports, Margins and Taxes tables.

Purchaser Price Supply Use Data

Similar to the benchmark data, constraints are constructed using an industry $\mathbf{W}_{SU}^{n,ind}$ and commodity $\mathbf{W}_{SU}^{m,comm}$ concordance matrix on the adjusted supply use data in purchaser prices $\mathbf{Z}_{SU^*,i,j}^{comm,ind}$

$$\mathbf{C}_{18} \Rightarrow \mathbf{Z}_{i,j}^{*m,n} = \mathbf{W}_{SU}^{m,comm} \times \mathbf{Z}_{SU^*,i,j}^{comm,ind} \times \left(\mathbf{W}_{SU}^{n,ind}\right)'$$

$$\mathbf{C}_{19} \Rightarrow \mathbf{Z}_{i,j}^{*a,n} = \mathbf{Z}_{SU^*,i,j}^{a,ind} \times \left(\mathbf{W}_{SU}^{n,ind}\right)'$$

$$\mathbf{C}_{20} \Rightarrow \mathbf{Z}_{i,j}^{*m,b} = \mathbf{W}_{SU}^{m,comm} \times \mathbf{Z}_{SU^*,i,j}^{comm,b}$$

The Supply/Use constraints are constructed on the non-confidential points of the Supply and Use tables, the totals (row/column) of the Supply and Use tables and on the row totals of the imports, margins and taxes tables.

National Accounts Data

Finally, National Accounts aggregates are available across the whole time series. Constraints are hence constructed from all available data. The applicable National Accounts data from 5204.0 includes the following.

Value Added by 18 industries:

$$\mathbf{C}_{21} \Rightarrow \sum_{i=m+1,m+a-1} \mathbf{U}_{i,j} = \mathbf{Q}_{1:18}^{ec,NA} \times \left(\mathbf{W}_{NA}^{n,18}\right)'$$

where $\sum_{i=m+1,m+a-1} \mathbf{U}_{i,j}$ is the value added by industry (excluding taxes less products from classification of primary inputs a), $\mathbf{Q}_{1:18}^{ec,NA}$ is the value added by industry in the National Accounts classification (18 sectors), and $\left(\mathbf{W}_{NA}^{n,18}\right)$ is the concordance matrix between the 18 sectors of the National Accounts and the ISAPC.

GDP (1 data point)

$$\mathbf{C}_{22} \Rightarrow gdp = gdp^{NA}$$

Total Final demand by destination (seven data points), in purchases prices

$$\mathbf{C}_{23} \Rightarrow \sum_{i=1:n} \mathbf{U}^{pp}_{i,m+1:m+b} = \mathbf{y}^{T,NA}_{1:7}$$

where the sum over final demand rows $\sum_{i=1:n} \mathbf{U}^{pp}_{i,m+1:m+b}$ is equal to final demand category totals of the national accounts $\mathbf{y}^{T,NA}_{1:7}$.

Total competing imports of the National Accounts $M^{T,NA}$ (one data point) is the sum over all imports

$$\mathbf{C}_{24} \Rightarrow \sum_{i=1:m} \sum_{j=1,n+b} \mathbf{M}_{ij} = M^{T,NA}$$

Total inputs into production by type (Compensation of employees, Gross Operating Surplus, Taxes on products, Taxes on production) $\mathbf{Q}^{ec,NA}_{1:a}$, four data points.

$$\mathbf{C}_{25} \Rightarrow \sum_{j=1:n+b} \mathbf{U}_{n+1:n+a,j} = \mathbf{Q}^{ec,NA'}_{1:a}$$

Household final demand in purchases prices is provided by 38 commodities in the National Accounts $\mathbf{y}^{NA}_{1:38,1}$, and relates estimated final demand $\mathbf{U}^{pp}_{1:n,m+1}$ by the concordance matrix $\mathbf{W}^{n,38}_{y,NA}$:

$$\mathbf{C}_{26} \Rightarrow \mathbf{U}^{pp}_{1:m,n+1} = \left(\mathbf{W}^{n,38}_{y,NA}\right) \times \mathbf{y}^{NA}_{1:38,1}$$

Compensation of employees by industry (17 data points) is calculated similarly:

$$\mathbf{C}_{27} \Rightarrow \mathbf{U}_{m+1,1:n} = \mathbf{Q}^{ec,NA}_{1,1:17} \times \left(\mathbf{W}^{n,17}_{Q,NA}\right)'$$

And for gross operating surplus by industry (17 data points)

$$\mathbf{C}_{28} \Rightarrow \mathbf{U}_{m+2,1:n} = \mathbf{Q}^{ec,NA}_{2,1:17} x \left(\mathbf{W}^{n,17}_{Q,NA}\right)'$$

Agricultural income in the National Accounts (eight data points) is defined in more detail for gross production:

$$\mathbf{C}_{29} \Rightarrow \sum_{i=1:m+a} \mathbf{U}^{pp}_{i,1:n} = \mathbf{U}^{pp,NA}_{Ag,T,1:8} \times \left(\mathbf{W}^{m,8}_{Ag,NA}\right)'$$

intermediate inputs:

$$\mathbf{C}_{30} \Rightarrow \sum_{i=1:m} \mathbf{U}^{pp}_{i,1:n} = \mathbf{U}^{pp,NA}_{II,1:8} \times \left(\mathbf{W}^{n,8}_{Ag,NA}\right)'$$

value added:

$$\mathbf{C}_{31} \Rightarrow \sum_{i=n+1:n+a-1} \mathbf{U}_{i,1:n} = \mathbf{Q}^{ec,NA}_{VA,1:8} \times \left(\mathbf{W}^{n,8}_{Ag,NA}\right)'$$

taxes on products:

$$\mathbf{C}_{32} \Rightarrow \sum_{i=m+a} \mathbf{U}_{i,1:n} = \mathbf{Q}^{ec,NA}_{TAX,1:8} \times \left(\mathbf{W}^{n,8}_{Ag,NA}\right)'$$

compensation of employees:

$$\mathbf{C}_{33} \Rightarrow \mathbf{U}_{m+1,1:n} = \mathbf{Q}^{ec,NA}_{COMP,1:8} \times \left(\mathbf{W}^{n,8}_{Ag,NA}\right)'$$

Constant Price Data

The constraints applied to the constant price table were only available from National Accounts data, and included Total Final demand by destination (seven data points), in purchaser's prices and including data on exports:

$$\mathbf{C}^{\mathrm{Re}al}_{34} \Rightarrow \sum_{i=1:n} \mathbf{U}^{pp,real}_{i,n+1:n+b} = \mathbf{y}^{T,NA}_{1:7}$$

Constant price household final demand is available for 17 categories in the National Accounts $\mathbf{y}^{NA}_{1:17,1}$, and is related to estimated final demand $\mathbf{U}^{pp,real}_{1:n,m+1}$ by the concordance matrix $\mathbf{W}^{n,17}_{y,NA}$:

$$\mathbf{C}^{\mathrm{Re}al}_{35} \Rightarrow \mathbf{U}^{pp,real}_{1:n,m+1} = \left(\mathbf{W}^{n,17}_{y,NA}\right) \times \mathbf{y}^{NA}_{1:17,1}$$

Constant price value added of 36 categories is similarly implemented:

$$\mathbf{C}^{\mathrm{Re}al}_{36} \Rightarrow \sum_{i=m+1:m+a-1} \mathbf{U}^{real}_{i,1:n} = \left(\mathbf{W}^{n,36}_{y,NA}\right) \times \mathbf{Q}^{ec,NA}_{1:36VA'}$$

Total taxes on products:

$$\mathrm{C}^{\mathrm{Re}al}_{37} \Rightarrow \sum_{i=m+a} \mathrm{U}^{real}_{i,1:n} = \mathrm{Q}^{ec,NA}_{TAX}$$

Total competing imports of the National Accounts $M^{T,NA}$ (1 data point) is the sum over all imports

$$\mathbf{C}_{38}^{\mathrm{Re}al} \Rightarrow \sum_{i=1:m} \sum_{j=1,n+b} \mathbf{M}_{ij}^{real} = M^{T,NA}$$

It should be noted, that whilst export and import price indices were available, these were only available in aggregate, and hence, the application of actual constant price import and export data points as constraints provides a more accurate deflation method.