

Technical Paper Series



Technical Paper 2004:2

**The Organisation of Trade Data
for inclusion in a Social
Accounting Matrix**

*Elsenburg
December 2004*

PROVIDE

PROJECT

The Provincial Decision-making Enabling Project

Overview


The Provincial Decision-Making Enabling (PROVIDE) Project aims to facilitate policy design by supplying policymakers with provincial and national level quantitative policy information. The project entails the development of a series of databases (in the format of Social Accounting Matrices) for use in Computable General Equilibrium models.

The National and Provincial Departments of Agriculture are the stakeholders and funders of the PROVIDE Project. The research team is located at Elsenburg in the Western Cape.

PROVIDE Research Team

Project Leader:	Cecilia Punt
Senior Researchers:	Kalie Pauw Melt van Schoor
Junior Researchers:	Lillian Rantho Benedict Gilimani
Technical Expert:	Scott McDonald
Associate Researchers:	Lindsay Chant Christine Valente

PROVIDE Contact Details

 Private Bag X1
Elsenburg, 7607
South Africa

 ceciliap@elsenburg.com
 +27-21-8085191
 +27-21-8085210

For the original project proposal and a more detailed description of the project, please visit www.elsenburg.com/provide

The Organisation of Trade Data for inclusion in a Social Accounting Matrix¹

Abstract

This paper describes the procedure followed to organise trade data obtained from the South African Revenue Service for inclusion in the PROVIDE Project Social Accounting Matrices. The data is aggregated in order to determine the custom's value of imports and exports, respectively, for each of the commodity groups identified in the Social Accounting Matrices. With regard to trading partners the trade data for South Africa is organized in such a way that it allows for the inclusion of multiple trading partners in the Social Accounting Matrices and that international trade at provincial level can be distinguished. Each data entry was mapped to a relevant commodity, province, and trading partner in three separate stages using three different mappings. General Algebraic Modelling Systems (GAMS) software was used to sort the trade data. Limitations of the data from the South African Revenue Service were addressed using additional information from Statistics South Africa, Global Insight and the South African Reserve Bank.

¹ The main author of this paper is Cecilia Punt, Project Leader of the PROVIDE Project.

Table of contents

1.	INTRODUCTION	3
2.	DESCRIPTION OF TRADE DATA FROM THE SOUTH AFRICAN REVENUE SERVICE.....	4
3.	IMPORTING DATA TO GDX FOR AGGREGATION IN GAMS	4
3.1.	<i>Record numbers and other identifiers</i>	4
3.2.	<i>Preparing the data files for loading to GDX.....</i>	5
3.3.	<i>Deriving unique elements for the GAMS set file.....</i>	7
3.4.	<i>Loading the data to GDX using a GAMS command (gms) file.....</i>	8
4.	CLASSIFICATION OF TRADE DATA TO MATCH SAM ACCOUNTS.....	8
4.1.	<i>Harmonised System codes to commodity accounts.....</i>	8
4.2.	<i>Country codes to trading partners (multiple trade accounts).....</i>	10
4.3.	<i>Postal codes to regions.....</i>	10
5.	RUNNING THE AGGREGATION PROCEDURE IN GAMS	11
5.1.	<i>GAMS gms files</i>	11
5.2.	<i>GAMS include files.....</i>	11
6.	ADDITIONAL INFORMATION SOURCES	11
7.	CONCLUSIONS	12
8.	REFERENCES	13
9.	APPENDIX	14
9.1.	<i>ASCIIICON specification file.....</i>	14
9.2.	<i>GAMS code for deriving unique elements</i>	14
9.3.	<i>GAMS code for loading data to GDX.....</i>	15
9.4.	<i>Mappings.....</i>	16
9.5.	<i>GAMS code for aggregating trade data</i>	23

List of tables

TABLE 1:	INFORMATION ON SARS TRADE DATASETS FOR DIFFERENT YEARS	4
TABLE 2:	UNDELIMITED TEXT SAMPLE FROM ORIGINAL EXPORT DATA FILE	6
TABLE 3:	COMMA SEPARATED VALUES AFTER CONVERSION USING ASCIIICON.....	6
TABLE 4:	DATA FOR LOADING TO GAMS.....	7
TABLE 5:	KEY BETWEEN INDUSTRY DESCRIPTIONS AND SIC CATEGORIES.....	17
TABLE 6:	LIST OF AGRICULTURAL COMMODITIES IN THE SAM	19
TABLE 7:	ORIGINAL LIST OF COUNTRY CODES	19
TABLE 8:	ADDITIONAL LIST OF COUNTRY CODES.....	21
TABLE 9:	SOUTH AFRICAN POSTAL CODES	22

1. Introduction

This paper describes the procedure followed to organize trade data obtained from the South African Revenue Service for inclusion in the PROVIDE Project Social Accounting Matrices. The aim of organizing the data is to create a submatrix for the SAM that presents the custom's value of imports and exports respectively by commodity for multiple trade (Rest of the World) accounts. National Social Accounting Matrices generally have at least one 'rest of the world' account that records trade between the relevant economy and the rest of the world. Exports are recorded in the column of the trade account, while imports are recorded in the row of the trade account. It is however possible to increase the number of trade accounts to allow for the distinction of trade with different trading partners. The trade data for South Africa is organized in such a way that it allows for multiple trading partners in the SAM for South Africa. Regional SAMs that are developed as part of the PROVIDE Project requires that the custom's value of international trade is determined at regional level as well. Information on trade is an important requirement for the construction of a SAM for South Africa as well as provincial/regional SAMs. The inclusion of detailed information on trade in the Social Accounting Matrices allows for subsequent analyses of trade related issues using Computable General Equilibrium models.

The trade data obtained from the South African Revenue Service is described in more detail in the next section. Data can be read into GAMS in various ways. Due to the magnitude of the trade data files, the trade data was read in as comma separated value (CSV) files and then written to GDX. The process of loading data to GDX for sorting the data using GAMS software is discussed in section 3. The mapping of trade data to the accounts included in national and regional SAMs is discussed in section 4. The mappings for the national SAM include both the mapping to relevant commodity groups and the mapping used to distinguish multiple trading partners. An additional mapping has to be carried out to derive international trade by region. Therefore for the regional SAMs trade data was mapped to each region using postal code information. The process of sorting the data using GAMS is explained in section 5. Due to limitations of the raw SARS data additional information from three different sources were used in order to derive the value of imports and exports for inclusion in the SAMs. Data from additional sources, Statistics South Africa, Global Insight and the South African Reserve Bank, are discussed in section 6. Summary and conclusions follow in the final section.

2. Description of trade data from the South African Revenue Service

Annual (calendar year) trade data for South Africa are available from the South African Revenue Service (SARS). The trade data are either received as two files, one for imports and one for exports, or as one file, containing both import and export data. If the data is included in one file, the data are sorted by postal code (ascending order) of the receiver or sender of the traded goods. The import data are listed separately from the export data, therefore the two sets can be distinguished by reference to the postal codes. The traded goods are classified according to the harmonised system used internationally for the classification of commodities. The data are received in undelimited ASCII format, with a fixed width file type: postal code (4 digits), country of origin for imports, or destination for exports (2 digits), tariff or Harmonized System code (9 digits), quantity imported or exported (15 digits, including two decimal places) and custom's value (15 digits). The layout of data files for different years could vary therefore it remains the responsibility of the researcher to ensure that the data is consistent with this layout before extracting the data.

Table 1 shows some key statistics and characteristics about the datasets for 1996, 1998 and 2000 that can be used as first check on accuracy. The dataset for 2000 contains 182 003 rows of data on exports and 432 096 rows of data on imports. The export value in 2000 was R196 377 million and the import value was R199 433 million.

Table 1: Information on SARS trade datasets for different years

	1996		1998		2000	
	Exports	Imports	Exports	Imports	Exports	Imports
Rows of data (number)	155 506	405 997	166 438	417 394	182 003	432 096
Custom's value (R million)	83 331	117 241	117 452	151 850	196 377	199 433

SARS trade data reports only trade in merchandise and gold, with very limited information on service trade. For purposes of a SAM trade data on services are also needed. Additional data from sources other than SARS were therefore consulted. The data from other sources are discussed in section 6.

3. Importing data to GDX for aggregation in GAMS

3.1. Record numbers and other identifiers

In order to read data into GAMS a set file that contains a list of all the identifiers of the data is needed. The postal codes, Harmonised System (HS) codes and country codes are regarded as the identifiers for each data entry. The identifiers are used to determine the respective town of the local trader, the type of commodity that is traded and the trading partner. Calculations are

performed only on the custom's value and quantity of trade and not on identifiers. The data are sorted by postal codes therefore most of the postal codes appear more than once in the data. Each data record however must have a unique identifier in order for GAMS to recognize it as a unique entry. Hence a record number was assigned to each data row in order to create a unique identifier for each record. The record number that is assigned is the actual row number of a particular data record in the original data set. The numbering of rows has the added advantage of providing a check to ensure that no data rows are excluded when loading the data.

Some of the postal codes as well as some of the HS codes start with a zero. For numeric values any zeros at the front of a number will disappear as soon as the numbers are used in computations or exported to Excel. Dropping zeros that form part of a code effectively changes the code and should therefore be avoided. Furthermore, the record numbers, postal codes and HS codes are numeric values, but typical calculations such as summing are never performed on these codes since the result would not have any economic meaning. In order to avoid potential errors each of the numeric identifiers, i.e. the record number, postal code and HS code were initialised with a letter to ensure that GAMS reads the identifiers as text instead of numeric values. All record numbers were changed to start with R, postal codes with P and HS codes with H. These letters were added to the data before reading it into GAMS.

3.2. Preparing the data files for loading to GDX²

Data were imported via comma separated value (CSV) files. The original trade data text file was divided and saved as two text files, one for imports and one for exports. ASCIICON software³ was used to convert the original file containing undelimited text (Table 2) to a text file containing comma separated values (Table 3). The conversion process places commas between data as specified. A specification file (see section 9.1) is used to give commands to ASCIICON. The specification file indicates the digit from which a string (series of values or text) should start, as well as how many digits the string should consist of. When an "n" is placed at the end of the command the string is numeric as opposed to text. This is desirable since ASCIICON places text strings in inverted commas, which will then later have to be removed manually in the GAMS include file. Since the user specifies the placement of the commas, it is possible to exclude some of the digits from the final file. For example, the first four digits in the original data file appeared to be all zeros and of no significance. These were dropped from the CSV file. Also the tariff reported in the original file is a nine-digit code, but the HS code used in the HS-SIC mapping uses only the first eight digits, hence the ninth digit

² GDX is an interface between Excel (or other file types) and GAMS.

³ ASCIICON is software that converts ASCII files into a number of different formats. Trial versions can be downloaded from the web at www.ASCIICON.com.

of the tariff was dropped. The first row of export data in Table 2 can be interpreted as showing that the address of the exporter is indicated by postal code 0117 (Pretoria). The exports are destined for Angola (country code AO). The tariff code 24220000 indicates that it is a product classified as ‘tobacco’. One unit was exported to the value of R38.

Table 2: Undelimited text sample from original export data file

0000000117AO240220000000000000000010000000000000038
0000000117CD250100905000000003570000000000000019570

Table 3: Comma separated values after conversion using ASCIICON

0117,AO,24022000,000000000000100,000000000000038
0117,CD,25010090,000000003570000,000000000019570

The comma separated value (CSV) text file created in ASCIICON was opened in GAMS as a text file and saved as an include file⁴, which is a type of file that can be called into GAMS through the main gms (command) file. The next step was to include the record numbers. First commas to separate the postal codes and the record numbers were placed in front of every postal code. This was done by selecting a column of commas using the shift-alt option in GAMS⁵ and to copy the commas to the front of the postal codes. A list of record numbers starting with “R” were created in Excel and then copied into a clean GAMS file. The record numbers were then selected using the shift-alt option to select a column. This was copied to the include file containing the data, in front of the added column of commas. Similarly columns of “H” and “P” were created in a GAMS file and copied in front of the tariff and postal codes respectively.

GAMS has certain reserved words which correspond with five of the country codes. As a precaution these codes can be renamed to avoid errors. However it was found that the use of these five country codes do not create problems. If these codes are changed to different two letter codes potential error can arise if confused with different countries’ codes and if it is changed to a three letter code, the fixed width characteristic of the data file is lost. The country codes were therefore used as they appear in the data. It must be noted that there exists a very limited number of cases in which a row of data from the original text file has a country code consisting of only one letter. This creates misalignment of data, but GAMS will provide

⁴ Because of the size of these files it is quicker to change the type of file in GAMS than to copy the data from a text file and to paste it into an include file.

⁵ The option of selecting columns of text in GAMS with the shift-alt option is especially useful when working with trade data because 1) the large number of data rows would render the use of CSV files inefficient if every row had to be altered one by one and 2) the fixed width characteristic of CSV files makes it especially useful to make consistent changes down a “column” of data.

an error message in this regard to guide the researcher to the error when compiling the set include file (section 3.3). The code can then be corrected if known, or reclassified by using the two letter code to indicate that the country is unknown.

GAMS reads a CSV file as a table and as such the table of data needs a heading for each column of data and these heading should also be comma delimited. A sample of the final data will therefore appear as indicated in Table 4.

Table 4: Data for loading to GAMS

Record,	post,	country,	hs,	quant,	cvalue
R5550,	P0117,	AO,	H24022000,	000000000000100,	000000000000038
R5551,	P0117,	CD,	H25010090,	000000003570000,	000000000019570

The main advantage of the CSV data files as opposed to Excel data files is that all exports can be included in a single file and all imports in one other file. In Excel the number of rows in a sheet are limited to 65 536 rows, therefore multiple sheets are needed for all the export and import data. GAMS handles these large data include files with relative ease.

3.3. Deriving unique elements for the GAMS set file

Besides the data files GAMS also needs set files in order to load data to GDX. Set files give a list of the different sets over which each parameter is defined, as well as the elements that are included in each of the sets. For import and export data, the parameter is defined over the following sets: record number, postal codes, country code, HS code and category. “Category” is a set that includes two elements, i.e. the quantity traded and the custom’s value of the traded goods.

As expected, most of the data entries have either a postal code or HS code that correspond to that of another data entry. The set files require a list of unique elements, i.e. every entry must appear only once. GAMS has a facility that allows the reporting of unique element lists, either in order of appearance or sorted. To the knowledge of the author GAMS does not allow the exporting of such a list to GDX or Excel. Also the reporting facility only allows for the elements to be reported in rows, five entries per row. It is therefore necessary to copy the list to Excel, to reorganise it into one column prior to sorting it. A sample of the include file to read all the codes into GAMS as well as the GAMS code for deriving the unique element list from the global list is included in section 9.2. The list of unique elements that is derived this way can be copied from Excel to an include file in order to read the sets into GAMS (see the next section).

3.4. Loading the data to GDY using a GAMS command (gms) file

The data is loaded to GDY using a gms file that calls the CSV data files (section 3.2) and the include file that defines the sets (section 3.3). Any inconsistencies between the codes in the data and those in the set file are indicated by an error message in GAMS. Loading data therefore serves as a useful check on the data and the sets that will be used in the aggregation procedure. Errors in the data, such as single digit country codes will be brought to the attention of the researcher. This allows for the data to be cleaned of errors in the HS, postal or country codes prior to sorting the data. The GAMS code for loading data to GDY and extracts of the data and set include files are included in section 9.3.

4. **Classification of trade data to match SAM accounts**

In order to include the trade data obtained from SARS into the SAM, the SARS classifications had to be mapped to the SAM classifications or appropriate SAM accounts. The unique element lists described in section 3.3 provide a list of all the harmonized system (HS) codes, country codes and postal codes that appear in the trade data obtained from SARS. Three different classification mappings / keys were used to categorise the data for inclusion in the SAM. The first mapping was used to map each of the harmonized system codes that appeared in the unique element list for HS codes to an appropriate SAM commodity account. The second mapping was used to map each country code that appeared in the unique element list for country codes to a relevant trading partner according to the multiple trade accounts in the SAM. The third mapping was needed specifically for the regional SAM and entails linking every postal code that appear in the unique element list for postal codes to one of the four regions for purposes of the PROVIDE regional SAMs.

4.1. Harmonised System codes to commodity accounts

The mapping of the HS codes to the SAM accounts take place in two stages. The HS codes are first mapped to the Standard Industrial Classification (SIC) codes (CSS, 1993). As a second stage each of the SIC codes are mapped to the commodity accounts that are desired for the particular SAM. Although the SIC codes pertain to industries and HS codes classify commodities, the commodity accounts in the SAMs follow the Standard Industrial Classification (SIC) codes based on the assumption of principal product production by each industry classified according to SIC.

A mapping between HS codes and SIC codes were obtained from the Industrial Development Corporation (IDC, 2000)⁶. The mapping is based on an eight-digit code, hence

⁶ The authors gratefully acknowledge the initial HS-SIC mapping received from Gerhard Kuhn (Industrial Development Corporation).

only the first eight digits of the nine-digit code in the SARS data were used. The process of reading in trade data to GDX requires a set file with a complete list of all the HS codes found in the data. The mapping from IDC contained 21485 mappings. Using this complete list of HS codes and the mapping from IDC a “match” function in Excel was used to find the HS codes that were missing from the original list of mappings. If an HS code was missing, the result was indicated as #N/A (not available). The whole column of the export data and of the HS codes from the trade data were highlighted and the auto filter function was used to select only the #N/A HS codes. For the 2000 trade data 2083 HS codes did not form part of the original mappings. An attempt was made to map these HS codes to SIC codes, but descriptions for these codes do not exist in the available HS code list. Hence it was decided to retain these unmapped codes as unclassified trade. In 2000 the unclassified trade accounted for 0.0008% of the total value of imports and 0.69% of the total value of exports.

The mapping between the SIC codes and the SAM accounts is kept consistent with that used by Statistics SA in their commodity classification of the 2000 Use Table (Statistics SA, 2003). The mapping is included in Table 5 of section 9.4.

4.1.1. *Agricultural data*

Typically national SAMs have only a single account for agriculture. In order to have more information on agriculture the commodity and activity accounts for agriculture are further disaggregated. Since imports and exports refer to only commodities and not activities, the agricultural commodity accounts are of interest when sorting the trade data. There are 20 different agricultural commodities defined in the first versions of the SAMs for the PROVIDE Project. The list of agricultural commodities included in the SAM can be viewed in Table 6 of section 9.4. Changes in the agricultural commodity account specification in the SAM will have implications for the mapping of the trade data to each of these accounts. The disaggregation of the agricultural accounts is carried out as a second stage.

After all the HS codes from the trade data were mapped to the SIC codes and subsequently to the 95 SAM commodities. The sort function in Excel was used to select only those HS codes that fall under the commodity category, Agriculture, Forestry and Fishing. That is commodity number 1 in the SAM. Data can alternatively be sorted by SIC code. The HS-SIC mapping obtained from IDC already distinguishes between four agricultural categories, i.e. 1110 (Growing of Crops; Market Gardening; Horticulture), 1120 (Farming of Animals), 1220 (Forestry; Logging; Related Services) and 1310 (Fishing; Operation of Fish Hatcheries and Fish Farms). In the 2000 trade data 356 HS codes related to agriculture. Each of these codes were further mapped to one of the 20 agricultural commodities included in the SAM.

4.2. Country codes to trading partners (multiple trade accounts)

The data for the multiple trade (Rest of the World) accounts in the SAM were obtained by mapping the country code of each data entry to a relevant country or trading block. A list with 234 country names and respective country codes was obtained from SARS (Table 7 of section 9.4). The data for 2000 contained 310 country codes. After conversations with officials from SARS it is still uncertain to what extent country codes were changed or updated over time, or whether the unknown country codes could be ascribed to incorrect use of the codes. Descriptions for an additional 28 codes were obtained from the International Organization for Standardization website (ISO, 2003), COMTRADE website of the United Nation (United Nations, 2003) and from SARS Officials (Table 8 of section 9.4). The information suggests that country codes were changed from time to time, which leads to one country having more than one code. In the 2000 data set 65 codes still remain unnamed. In the 2000 data set exports to unknown countries amount to 1.3 percent of the total value of exports reported by SARS. Imports from unknown countries amount to 0.96 percent of the total value of imports reported by SARS.

4.3. Postal codes to regions

For a national SAM with a single trade account only the HS-SIC mapping has to be used to organize the data. However, in order to derive trade data for regional SAMs the trade data has to be linked to each of the four regions identified for purposes of the PROVIDE project. Similar to the national SAM, each of the regional SAMs can still have either a single trade account or multiple trade accounts.

The postal code of each data entry allowed the mapping of trade data to each of the four regions. The South Africa Post Office's list of postal codes (Post Office, 2003) was used to identify from which regions exports were sent, or for which province the imports were destined. The mapping of postal codes to each of the provinces is included in Table 9 of section 9.4.

The postal code for each data entry is that of the importer or exporter of the goods. Ideally the postal code of the producer or user of the goods should be used to derive exports and imports at provincial level, however this information is not included in the trade data or otherwise available to the knowledge of the authors. Researchers from Global Insight Consultancy Firm indicated that they use their local knowledge to try and rectify some of the main discrepancies to gain a better representation of regional trade. Hence it was decided to make use of their data as well as discussed in section 6.

5. Running the aggregation procedure in GAMS

The aggregation procedure in GAMS entails giving GAMS the correct commands and classification information to enable it to derive the desired aggregated/sorted set of trade data that can be included in the SAM. The aggregation command is programmed in a GAMS gms (command) file (see section 9.5.1). The classification mappings discussed in section 4 are incorporated in a GAMS include file in order for GAMS to read the mappings. A sample of the include file is presented in section 9.5.2.

5.1. GAMS gms files

The first command in the gms file reads in the data from the GDX file (section 3.4). However in order for GAMS to recognise the data, the set include file with the unique elements (section 3.3) needs to be called up prior to loading the data. The include file with the mapping information (section 5.2) is read in prior to the aggregation command. During the aggregation process GAMS adds together the quantity and custom's value data of the original groups to derive the quantity and custom's value information for the new categories as specified in the mapping file. The results, i.e. data according to SAM accounts, are written out to another GDX file. The results can be exported to Excel from the GDX viewer either as a pivot table or as an ordinary table of results.

Additional commands can be included in GAMS in the gms file in order to check various totals or subtotals. The display statement in GAMS makes it possible to view these checks for accuracy in the GAMS list file.

5.2. GAMS include files

The include file firstly defines new categories for each identifier and the desired elements of the new categories or sets, e.g. the new elements of the set for postal codes will be regions 1 to 4. Secondly a mapping statement links all the unique elements that appear in the original data to one of the elements of new sets, i.e. each of the postal codes will be mapped to one of the four regions. This process is carried out for postal codes, HS codes and country codes.

6. Additional information sources

Limitations of SARS data necessitated turning to other sources for additional information. One of the main limitations is that SARS data does not fully record services trade. For both imports and exports the pattern of trade in services is derived from the published data in the Supply and Use Tables for 2000 (Statistics SA, 2003).

Another limitation of SARS data is that monetary gold exports are not distinguished by the country of destination, but as one total value. The COMTRADE data on the United Nations website (United Nations, 2003) was consulted in this regard but it contains data on trading partners for non-monetary gold imports and exports only. Destination countries for 99 percent of total gold exports are therefore unknown.

Total *exports* for South Africa as reported by the South Africa Reserve Bank (SARB) consist of three parts, i.e. merchandise, gold and services. Total *imports* for South Africa as reported by the SARB consist of two parts, i.e. merchandise and services. For each of the mentioned subcomponents of exports and imports the figures reported by SARB are used as control totals and the patterns (shares) of trade per commodity within these categories are derived from the data from the South African Revenue Service (SARS). There are discrepancies between the SARS and SARB data. The SARS data captures the trade information at the time that the goods are at port, whereas SARB makes some changes to the data to try and capture the trade within the year during which the payment took place and not the year during which the goods arrived at the port. The SARB control totals are therefore deemed to be more accurate.

As mentioned in section 4.3 the postal codes in the trade data represent the postal code of the trader and not necessarily the postal code of the producer. For this reason the data supplied by Global Insight (2004) is deemed to be more accurate than the raw SARS data. Global Insight uses SARS data as a starting point, but then uses local knowledge to make some adjustment to the data with regard to the origins of some of the exports. This is particularly relevant to mining exports. The trade data obtained from Global Insight however reports trade at a more aggregated level than what is required for the SAM. It can therefore not replace the PROVIDE Projects efforts in obtaining trade information from SARS, but only supplement the data. Global Insight reports trade for 34 commodity categories as opposed to the desired 95 categories used in the PROVIDE SAMs. The 95 SAM categories are therefore mapped to the 34 categories reported by Global Insight and then the Global Insight trade values are used as control totals for these subsets of SAM commodities.

7. Conclusions

Trade data were obtained from SARS to estimate commodity trade at national and regional level to multiple trading partners. Each data record contains a Harmonised System (HS) code, a country code and a postal to provide the necessary information on what is traded and by whom. Lack of complete lists of HS code descriptions and country code descriptions necessitated that some of the data remained unclassified. The unclassified data however represents less than two percent of the total trade value.

The use of GAMS software to handle trade data proved very useful because GAMS handles large datasets with relative ease. Using GAMS for loading and subsequently aggregating the trade data has the added benefit that it provides a check on accuracy with regard to the identifiers used, since the set include files necessitates a list of all identifiers and the aggregation include file necessitates that each of the identifiers are mapped to a new category. GAMS warns the user of any errors.

The main limitations of the raw SARS data include its omission of service trade, the fact that postal codes represents the location of traders instead of producers, and the omission of trading partners for gold exports. The latter remains an unresolved issue at present. However additional data sources were used to overcome some of these problems, i.e. service trade as reported in the supply and use table published by Statistics South Africa, the control totals of trade values for merchandise, services and gold by the South African Reserve Bank, and the regional trade data for 34 commodity groups released by Global Insight.

8. References

- CSS (1993). SIC Standard Industrial Classification of all Economic Activities. Report 09-90-02. Pretoria: CSS.
- ISO (2003). *International Organisation for Standardization*. <http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html>
- IDC (2003). *Unpublished HS-SIC Mapping*. Industrial Development Corporation: Johannesburg.
- Global Insight (2004). *Regional Economic Focus Data*. Centurion: Global Insight Southern Africa.
- Post Office (2003). *An informative Guide to Postcodes and Postal Services in South Africa*. Alphabetical List.
- SARS (2003). *Unpublished trade data for South Africa for 2000*. South African Revenue Service: Pretoria.
- Statistics SA (2003). *Final Supply and Use Tables, 2000*. Reports no. 04-04-01 (2000). Pretoria, Statistics South Africa.
- United Nations (2003). COMTRADE database, United Nations. <http://unstats.un.org/unsd/comtrade>.

9. Appendix

9.1. ASCIICON specification file

ASCIICON was used to convert undelimited text files to comma separated text files as part of the preparation of the data files for use in GAMS. The specification file looks as follows:

```
1,6n
7,4n
11,2n
13,8n
22,15n
37,15n
```

The inclusion of an “n” indicates that data should be treated as numeric values. The file indicates that the first value starts at digit 1 and is six digits long. The second value, which is the postal code, starts at digit seven and is four digits long, etc.

9.2. GAMS code for deriving unique elements

GAMS code to derive unique elements by selecting the include file for either country, HS or postal codes:

```
*### Declare sets ###
$onuellist
*$INCLUDE uecountry00.inc
*$INCLUDE uehs00.inc
$INCLUDE uepost00.inc
```

The extract from the include file “uepost00.inc” to derive unique element list for postal codes shows the postal codes from record number 217 813 to 217 823. From the extract below GAMS will create a list containing three elements: P1509, P1510, P1511, as apposed to the previous list of 11 codes.

```
SET
R217813 /P1509/
R217814 /P1509/
R217815 /P1509/
R217816 /P1509/
R217817 /P1510/
R217818 /P1510/
R217819 /P1510/
```

R217820 /P1510/
 R217821 /P1511/
 R217822 /P1511/
 R217823 /P1511/
 ;

9.3. GAMS code for loading data to GDX

```

### SET DECLARATION
### Sets are declared and defined in an include file
* read in sets for base data
$include 2000sets.inc

*Select either imports or exports
### IMPORTS
*Data are loaded from a comma separated file
*$ontext
TABLE imports2000(rec,post,w,hs,cat) trade data
$ondelim
$INCLUDE 2000imports.inc
$offdelim

#### Export results to.gdx
Execute_Unload 'imports2000.gdx', imports2000 ;
*$offtext

### EXPORTS
*Data are loaded from a comma separated file
*$ontext
TABLE exports2000(rec,post,w,hs,cat) trade data
$ondelim
$INCLUDE 2000exports.inc
$offdelim

#### Export results to.gdx
Execute_Unload 'exports2000.gdx', exports2000 ;
*$offtext

```

An extract from the **data** include file to show the format of the include file:

rec,	post,	w,	hs,	quant,	value
R155507,	P0001,	CM,	H01060000,	0000000000010000,	0000000000008000
R155508,	P0001,	GN,	H01060000,	0000000000001300,	0000000000007699
R155509,	P0001,	US,	H02074190,	0000000314912003,	0000000009202089

An extract from the **set** include file to show the format of the include file:

```
SETS
rec  record accounts /
R1*R600000 /
hs  hs code number /
H00281700
H00603100
H00700910
H00810000
H01000000 /
w  country /
AD
AE
AF
AG
AI
ZW /
cat categories /
quant  Quantity
value  Customs value / ;
```

9.4. Mappings

Table 5 shows the mapping between the SAM accounts and the Standard Industrial Classification codes. This is taken from Statistics SA's report on Supply and Use Tables (Stats SA, 2003) since the PROVIDE SAM categories are kept consistent with Statistics South Africa's activity classifications.

Table 5: Key between industry descriptions and SIC categories

SAM account	Account description	Corresponding SIC (Standard Industrial Classification, fifth edition) categories
1	Agriculture	1110, 1120, 1130, 1140, 1150, 1160, 1210, 1220, 1310, 1320
2	Coal	2100
3	Gold	2300
4	Other mining	2210, 2410, 2420, 2510, 2520, 2530, 2900
5	Meat	3011
6	Fish	3012
7	Fruit	3013
8	Oils	3014
9	Dairy	3020
10	Grain mills	3031, 3032
11	Animal feeds	3033
12	Bakeries	3041
13	Sugar	3042
14	Confectionery	3043
15	Other food	3044, 3049
16	Beverages & tobacco	3051, 3052, 3053, 3060
17	Textiles	3111, 3112
18	Textile articles	3121
19	Carpets	3122
20	Other textiles	3123, 3129
21	Knitting mills	3130
22	Wearing apparel	3140, 3150
23	Leather	3161
24	Handbags	3162
25	Footwear	3170
26	Wood	3210, 3221, 3222, 3223, 3229
27	Paper	3231
28	Containers of paper	3232
29	Other paper	3239
30	Publishing	3241, 3242, 3249, 3251, 3252
31	Recorded media	3243, 3260
32	Petroleum	3310, 3321, 3322, 3323, 3324, 3325, 3329, 3330
33	Basic chemicals	3341
34	Fertilizers	3342
35	Primary plastics	3343, 3360
36	Pesticides	3351
37	Paints	3352
38	Pharmaceuticals	3353
39	Soap	3354
40	Other chemicals	3359
41	Tyres	3371
42	Other rubber	3379
43	Plastic	3380
44	Glass	3411
45	Non-structural ceramics	3421
46	Structural ceramics	3422, 3423
47	Cement	3424
48	Other non-metallic	3425, 3426, 3429
49	Iron and steel	3510, 3531
50	Non-ferrous metals	3520, 3532

SAM account	Account description	Corresponding SIC (Standard Industrial Classification, fifth edition) categories
51	Structural metal	3541, 3542, 3543
52	Treated metals	3551, 3552
53	General hardware	3553
54	Fabricated metal	3559
55	Engines	3561
56	Pumps	3562
57	Gears	3563
58	Lifting equipment	3565
59	General machinery	3564, 3569
60	Agricultural machinery	3571
61	Machine tools	3572
62	Mining machinery	3574
63	Food machinery	3575
64	Special machinery	3573, 3576, 3577, 3579
65	Household appliances	3580
66	Office machinery	3590
67	Electric motors	3610
68	Electricity apparatus	3620
69	Wire and cable	3630
70	Accumulators	3640
71	Lighting equipment	3650
72	Electrical equipment	3660
73	Radio and television	3710, 3720, 3730
74	Optical instruments	3741, 3742, 3743, 3750, 3760
75	Motor vehicles	3810, 3820
76	Motor vehicle parts	3830
77	Other transport	3841, 3842, 3850, 3860, 3871, 3872, 3879
78	Furniture	3910
79	Jewelery	3921
80	Other manufacturing	3922, 3923, 3924, 3929, 3951, 3952
81	Electricity	4110, 4120, 4130
82	Water	4200
83	Buildings	5021, 5024, 5031, 5032, 5033, 5039, 5041, 5049
84	Other construction	5010, 5022, 5023, 5050
85	Trade	6110, 6120, 6130, 6140, 6150, 6190, 6210, 6220, 6230, 6240, 6250, 6260, 6310, 6320, 6330, 6340, 6350
86	Hotels	6410, 6420
87	Transport services	7110, 7120, 7130, 7210, 7220, 7300, 7410
88	Communications	7510, 7520
89	Insurance	8110, 8190, 8210, 8310, 8320
90	Real estate	8410, 8420
91	Business activities	8510, 8520, 8530, 8610, 8620, 8630, 8640, 8650, 8690, 8710, 8720, 8810, 8820, 8830, 8890
92	General government	9110, 9120, 9130, 9400
93	Health and social work	9311, 9312, 9319, 9320, 9330
94	Activities/services	9200, 9500, 9600, 9900, 0200, 0900

Source: Statistics SA (2003)

Table 6 lists the accounts for agricultural commodities included in the PROVIDE SAMs.

Table 6: List of agricultural commodities in the SAM

Agricultural Commodity Accounts	
1 Summer Cereals	11 Other Horticulture
2 Winter Cereals	12 Livestock Sales
3 Oilseeds	13 Milk and Cream
4 Sugarcane	14 Animal Fibres
5 Other Field Crops	15 Poultry
6 Potatoes and Vegetables	16 Game
7 Wine grapes	17 Fish
8 Citrus	18 Other Animals
9 Subtropical	19 Forestry
10 Deciduous	20 Wild Flowers Compost and Firewood

Table 7 gives the country codes and descriptions that were obtained from SARS. The list is sorted by the name of the country.

Table 7: Original List of Country Codes

Code	Name of Country	Code	Name of Country	Code	Name of Country
AF	Afghanistan	GM	Gambia	NF	Norfolk Island
AL	Albania	GE	Georgia	MP	Northern Mariana Island
DZ	Algeria	DE	Germany	NO	Norway
AS	American Samoa	GH	Ghana	OM	Oman
AD	Andorra	TI	Gibraltar	PK	Pakistan
AO	Angola	GR	Greece	PW	Palau
AI	Anguilla	GL	Greenland	PA	Panama
AQ	Antarctica	GD	Grenada	PG	Papua New Guinea
AG	Antigua and Barbuda	GP	Guadelope	PY	Paraguay
AR	Argentina	GU	Guam	PE	Peru
AM	Armenia	GT	Guatemala	PH	Philippines
AW	Aruba	GN	Guinea	PN	Pitcairn
AU	Australia	GW	Guinea-Bissau	PL	Poland
AT	Austria	GY	Guyana	PT	Portugal
AZ	Azerbaijan	HT	Haiti	PR	Puerto Rico
BS	Bahamas	HM	Heard Island and McDonald Islands	QA	Qatar
BH	Bahrain	VA	Holy See	RE	Reunion
BD	Bangladesh	HU	Honduras	RO	Romania
BB	Barbados	IS	Iceland	RU	Russian Federation
BY	Belarus	IN	India	SH	Saint Helena
BE	Belgium	ID	Indonesia	KN	Saint Kitts and Nevis
BZ	Belize	IR	Iran, Islamic Republic of	LC	Saint Lucia
BJ	Benin	IQ	Iraq	VC	Saint Vincent and Grenadines
BM	Bermuda	IE	Ireland	WS	Samoa
BT	Bhutan	IL	Israel	SM	San Marino
BO	Bolivia	IT	Italy	SA	Saudi Arabia

Code	Name of Country	Code	Name of Country	Code	Name of Country
BA	Bosnia and Herzegovina	JM	Jamaica	SN	Senegal
BW	Botswana	JP	Japan	SC	Seychelles
BV	Bouvet Island	JO	Jordan	SL	Sierra Leone
BR	Brazil	KZ	Kazakhstan	SG	Singapore
IO	British Indian Ocean Terri	KE	Kenya	SK	Slovakia
BN	Brunei Darussalam	KI	Kiribati	SI	Slovenia
BG	Bulgaria	KP	Korea, Democratic Peoples Republic	ST	Soa Tome and Principe
BF	Burkina Faso	KR	Korea, Republic of	SB	Solomon Islands
BI	Burundi	KW	Kuwait	SO	Somalia
KH	Cambodia	KG	Kyrgyzstan	ZA	South Africa
CM	Cameroon	LA	Lao Peoples Democratic Republic	GS	South Georgia and the South Sandwich Island
CA	Canada	LV	Latvia	ES	Spain
CV	Cape Verde	LB	Lebanon	LK	Sri Lanka
KY	Cayman Islands	LS	Lesotho	PM	St Pierre and Miquelon
CF	Central African Republic	LR	Liberia	SD	Sudan
TD	Chad	LY	Libyan Arab Jamahiriya	SR	Suriname
CL	Chile	LI	Liechtenstein	SJ	Svalbard and Jan Mayen
CN	China	LT	Lithuania	SZ	Swaziland
CX	Christmas Island	LU	Luxembourg	SE	Sweden
CC	Cocos (Keeling) Islands	MO	Macau	CH	Switzerland
CO	Colombia	MK	Macedonia (former Yugoslav Republic)	SY	Syrian Arab Republic
KM	Comoros	MG	Madagascar	TW	Taiwan, Prov of China
CG	Congo	MW	Malawi	TJ	Tajikistan
CD	Congo, Democratic Republic of	MY	Malaysia	TZ	Tanzania
CK	Cook Islands	MV	Maldives	TH	Thailand
CR	Costa Rica	ML	Mali	TG	Togo
CI	Cote D'Ivoire	MT	Malta	TK	Tokelau
HR	Croatia	MH	Marshall Islands	TO	Tonga
CU	Cuba	MQ	Martinique	TT	Trinidad and Tobago
CY	Cyprus	MR	Mauritania	TN	Tunisia
CZ	Czech republic	MU	Mauritius	TR	Turkey
DK	Denmark	YT	Mayotte	TM	Turkmenistan
DJ	Djibouti	MX	Mexico	TC	Turks and Caicos Islands
DM	Dominica	FM	Micronesia, Fed States of	TV	Tuvalu
DO	Dominican Republic	MD	Moldova, Republic of	UG	Uganda
TP	East Timor	MC	Monaco	UA	Ukraine
EC	Ecuador	MN	Mongolia	AE	United Arab Emirates
EG	Egypt	MS	Montserrat	GB	United Kingdom
SV	El Salvador	MA	Morocco	US	United States
GQ	Equatorial Guinea	MZ	Mozambique	UM	United States Minor Outlying Islands
ER	Eritrea	NM	Myanmar	UY	Uruguay

Code	Name of Country	Code	Name of Country	Code	Name of Country
EE	Estonia	NA	Namibia	UZ	Uzbekistan
ET	Ethiopia	NR	Nauru	VU	Vanuatu
FK	Falkland Islands (Malvinas)	NP	Nepal	VE	Venezuela
FO	Faroe Islands	NL	Netherlands	VN	Vietnam
FJ	Fiji	AN	Netherlands Antilles	VI	Virgin Islands, US
FI	Finland	NC	New Caledonia	WF	Wallis and Futuna
FR	France	NZ	New Zealand	EH	Western Sahara
GF	French Guiana	NI	Nicaragua	YE	Yemen
PF	French Polynesia	NE	Niger	YU	Yugoslavia
TF	French Southern Territories	NG	Nigeria	ZM	Zambia
GA	Gabon	NU	Niue	ZW	Zimbabwe

Source: South African Revenue Service (2003)

Table 8 presents descriptions to country codes in addition to those appearing on the list obtained from SARS in Table 7. Information in the first and second columns were obtained from the Comtrade database on the website of the United Nations. The third and fourth column were obtained from personnel at SARS and the website of the International Organisation for Standardisation.

Table 8: Additional List of Country Codes

Source: UN Comtrade database		Source: SARS and other	
Code	Name of Country	Code	Name of Country
AB	Aruba	BU	Burundi
IO	British Indian Ocean Territories	CO	Colombia
VG	British Virgin Islands	DN	Denmark
HK	China, Hong Kong	FL	Falkland Islands
CS	Czechoslovakia	FX	France Metropolitan
PC	Former Pacific Islands	HN	Honduras
PZ	Former Panama Canal Zone	LE	Lebanon
FQ	Former South Antarctic Territories	PU	Misc Pacific Islands
SU	Former USSR	NW	Norway
FQ	French South Antarctic Territories	PO	Poland
GI	Gibraltar	DR	Rep. Congo
HU	Hungary	XX	Ship's Stores
MM	Myanmar	SP	Spain
PS	Occupied Palestinian Territories	SW	Sweden
RW	Rwanda	TL	Tokelau
ZN	SACU	TU	Turkey
YU	Serbia and Montenegro & Yugoslavia	UR	Uruguay
TP	Timor-Leste	YG	Yugoslavia
WL	World	ZR	Zaire

Table 9: South African Postal Codes

PROVIDE SAM Regions	Hub/Area	Postal Code
Gauteng / Free State / North West	Pretoria 1	P0001 - P0204
	Germiston	P1400 - P1699
	Heidelberg	P1438 - P1444
	Krugersdorp	P1700 - P1799
	Soweto	P1800 - P1870
	Vanderbijlpark	P1871 - P1999
	Witspos (Jhb)	P2000 - P2199
	Krugersdorp	P2495 - P2519
	Potchefstroom	P2520 - P2709
	Mafikeng	P2710 - P2899
	Boemfontein 1	P9300 - P9409
	Welkom	P9410 - P9699
Boemfontein 2	P9700 - P9999	
Mpumalanga / Limpopo	Pretoria 2	P0205 - P0698
	Pretoria 3	P1000 - P1199
	Nelspruit	P1200 - P1399
	Pretoria 4	P2200 - P2494
	Pietersburg	P0699 - P0999
Northern Cape / Western Cape	Upington	P8180 - P8299
	Kimberly	P8300 - P8799
	Upington	P8800 - P8999
	George	P6500 - P6699
	Worcester	P6700 - P6899
	Beaufort West	P6900 - P7099
	Cape Mail	P7100 - P8179
KwaZulu-Natal / Eastern Cape	Ladysmith	P2900 - P3199
	Pietermaritzburg	P3200 - P3309
	Ladysmith	P3310 - P3599
	Durmail 2	P3600 - P3799
	Richards Bay	P3800 - P3999
	Durmail 1	P3991 - P4179
	Port Shepstone	P4180 - P4299
	Durmail 2	P4300 - P4641
	Port Shepstone	P4642 - P4730
	Port Shepstone	P4740 - P4799
	Umtata	P4800 - P4899
	Umtata	P4735 - P4739
	East London	P4920 - P5049
	Umtata	P5050 - P5199
	East London	P5200 - P5750
Port Elizabeth	P5751 - P6499	

Source: Post Office (2003)

9.5. GAMS code for aggregating trade data9.5.1. *GMS file for aggregating trade data*

```

*## Sets are declared and defined in an include file
* read in sets for base data
$include 00loadsets.inc
*### Procedures for Aggregation
$include tradagg00.inc
*## DATA ENTRY
PARAMETERS
    exports00(rec,post,w,hs,cat) 2000 export trade data
    imports00(rec,post,w,hs,cat) 2000 import trade data ;
* Data are read in from a *.gdx file – select imports or exports
$ontext
$GDXIN exports00.gdx
$LOAD exports00 = exports00
$GDXIN
$offtext
*$ontext
$GDXIN imports00.gdx
$LOAD imports00 = imports00
$GDXIN
*$offtext
*## AGGREGATION
PARAMETER
    NEWDAT(npost,nw,nhs,cat) Aggregated trade derived ;
    NEWDAT(npost,nw,nhs,cat) = SUM((rec,post,w,hs)
        $(MAPPOST(npost,post)
        $MAPW(nw,w)
        $MAPHS(nhs,hs)),
*        exports00(rec,post,w,hs,cat)) ;
        imports00(rec,post,w,hs,cat)) ;
*## EXPORT OF RESULT FILES
*Export new aggregation to GDX
*Execute_Unload 'sortedex00.gdx', NEWDAT ;
Execute_Unload 'sortedim00.gdx', NEWDAT ;

```

9.5.2. *Include file for aggregating trade data*

* Set include file for aggregating RSA Trade data

SET

npost postal code number /

npost1 Region 1

npost2 Region 2

npost3 Region 3

npost4 Region 4 /

nw country /

nw1 rest of world /

nhs hs code number /

nhs1 hs code group1

nhs2 hs code group 2

nhs3 hs code group 3 /

MAPPOST(npost,post) mapping postal districts to new districts /

```
npost1  .(  P0001
          P0002
          P0005
          P0201
          P0204  )
```

```
npost2  .(  P0212
          P0216
          P0219
          P0232
          P9886
          P9950
          P9980  ) /
```

MAPW(nw,w) mapping countries to new countries /

```
nw1  .(  AD
        AE
        AF  ) /
```

MAPHS(nhs,hs) mapping hs codes to new new hs groups /

```
nhs1  .(  H00080000
```

H01011900
H52032090
H52081260)
nhs2 .(H27011000
H27012000
H27021000
H27030000)
nhs3 .(H71082000)
nhs4 .(H25010000
H25010010
H25020000
H44039900) / ;

--oo0oo--

Technical Papers in this Series

Number	Title	Date
TP2003: 1	Creating a 1995 IES Database in STATA	September 2003
TP2003: 2	Creating a 1995 OHS and a combined OHS-IES Database in STATA	September 2003
TP2003: 3	A Standard Computable General Equilibrium Model Version 3: Technical Documentation	September 2003
TP2003: 4	Measures of Poverty and Inequality: A Reference paper	October 2003
TP2004: 1	SeeResults: A spreadsheet Application for the Analysis of CGE Model Results	November 2004
TP2004: 2	The Organisation of Trade Data for inclusion in Social Accounting Matrix	December 2004
TP2004: 3	Creating a 2000 IES Database in Stata	August 2004
TP2004: 4	Forming Representative Household Groups in a SAM	July 2004

Other PROVIDE Publications

Background Paper Series

Working Papers

Research Reports