Getting the best of two worlds? Combing single country data and the GTAP Data Base for work with MANAGE-WB

- DRAFT: NOT FOR PUBLICATION-

Wolfgang Britz[1,2], Alhassane Camara[1], Martin Aaroee Christensen[1], Mei Mei Aileen Lam[1], Lulit Mitik Beyene[1], Ragchaasuren Galindev[1], Noe Nicolas Reidt[1], Jan Ignacy Witajewski-Baltvilks[1,3]

[1]World Bank, Macroeconomics, Trade, and Investment Global Practice

[2]University of Bonn, Institute for Food and Resource Economics

[3] University of Warsaw, Faculty of Economic Sciences

Abstract

This paper details how GTAP Data are used to construct a SAM or complement SAMs for use with the World Bank's MANAGE-WB model. It discusses pros and cons of using a harmonized, ready-to-use global database versus building a single country SAM bottom up. The main advantage of the GTAP database for single country CGE work is that a SAM can be made available instantaneously for now around 140 countries, combined with relatively high commodity and factor detail, and the harmonized link to satellite data on emissions and land use. Dis-advantages compared to a bottom-up approach are a potential larger statistical time gap, corrections to observed trade and other statistics at country level to achieve global consistency and potentially missing detail for products in the focus of a specific project. The team around MANAGE-WB therefore uses a mixed approach, where GTAP data might be corrected or updated based on newer macro data if construction of a SAM bottom-up is not possible. GTAP data also play an important role in case a SAM is constructed from other data. GTAP data can provide information to split up certain transactions, for instance to introduce detail on power generation, and delivers emission inventories or data on land use.

Background

Building a Social Accounting Matrix (SAM) for a single country Computable General Equilibrium (CGE) or a data base for multi-regional CGE analysis differs in many aspects. For constructing a global database such as the GTAP database, bi-lateral trade related exchanges are usually determined first and are taken as fixed and given as constraints for further work on individual countries and regions. This implies that the Balance of Trade (B.o.T.) at single product level and in total cannot be changed for any country once a consistent set of trade flows is determined. Equally, trade margin demands, and import and export tax revenues for each product need to be kept unchanged. Multi-regional work will also try to use a "one build process and data source fits all" approach, to exploit returns-to-scale in designing software code for data processing and ease harmonization of definitions and documentation of the work. For

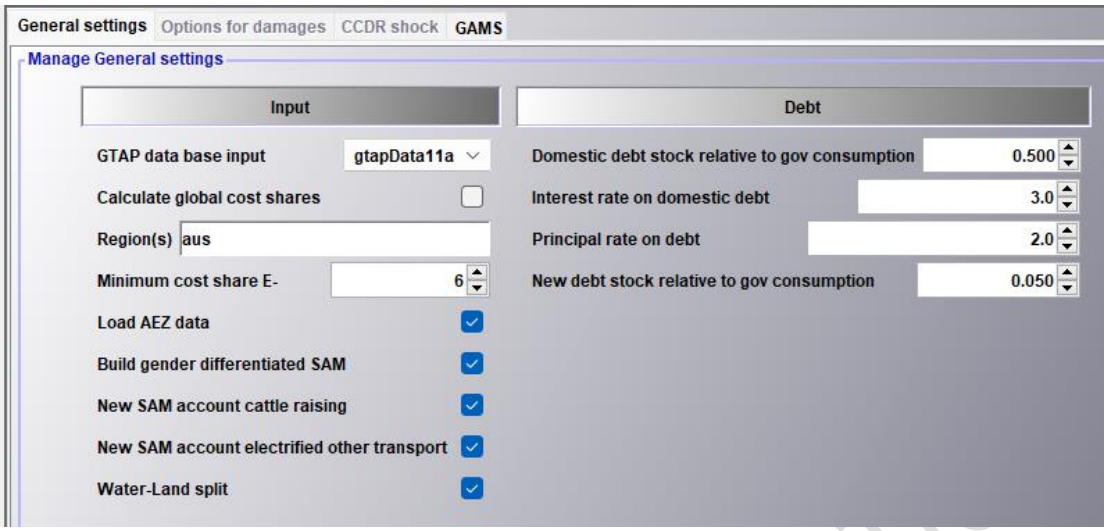
instance, the GTAP center uses wherever available the World Bank's World Development Indicators (WDI) database to obtain Gross Domestic Product (GDP) in total and its composition (private and government consumption, gross fixed capital formation). While total GDP is maintained in further processing, its composition is subject to change to match trade data. Penn World Tables provide data on capital stock, subject to outlier correction in case of capital stock ratios deemed implausible, while depreciation is estimated based on a depreciation rate of 4% (Aguiar et al., 2022). Such uniform sourcing and data treatment is key to compare results consistently across countries and model regions.

The necessity to have data for all countries available for the reference year of the database can imply a considerably statistical time gap, for GTAP Version 11 with its reference year in 2017 released in 2023 amounting to seven years. While trade and macro-economic data all refer to the same reference year, the time points at which Input Output (IO)-Tables were submitted for a country in the database can differ widely. And clearly, often a larger time gap exists between the date of submission of an IO-Table to the GTAP center and the data year(s) of the underlying statistical data. A global database requires the same set of activities, products, factors, final demands etc. for each covered country and region, and a uniform structure to account for taxes, subsidies, and margins. Constructing a SAM for a single country and specific project gives instead freedom to reflect country specifics, for instance, in data availability, sector importance or the tax system.

Projects with MANAGE-WB operate mostly in close co-operation with World Bank's country desks and national stakeholders which scrutinize data for plausibility and actuality. The team therefore tries to exploit the most recent data as detailed below. This is discussed after we detail the process of constructing a SAM according to MANAGE-WB guidelines from GTAP data.

Constructing a SAM in MANAGE format based on GTAP data

There are three main different uses of the GTAP data for MANAGE-WB. The one coming close to multi-regional analysis by the GTAP community constructs a SAM according to MANAGE-WB guidelines directly from the GTAP data, at the resolution of the GTAP-Power Data Base. The process is steered by the Graphical User Interface of MANAGE-WB (GUI) as depicted in Graphic 1 below.



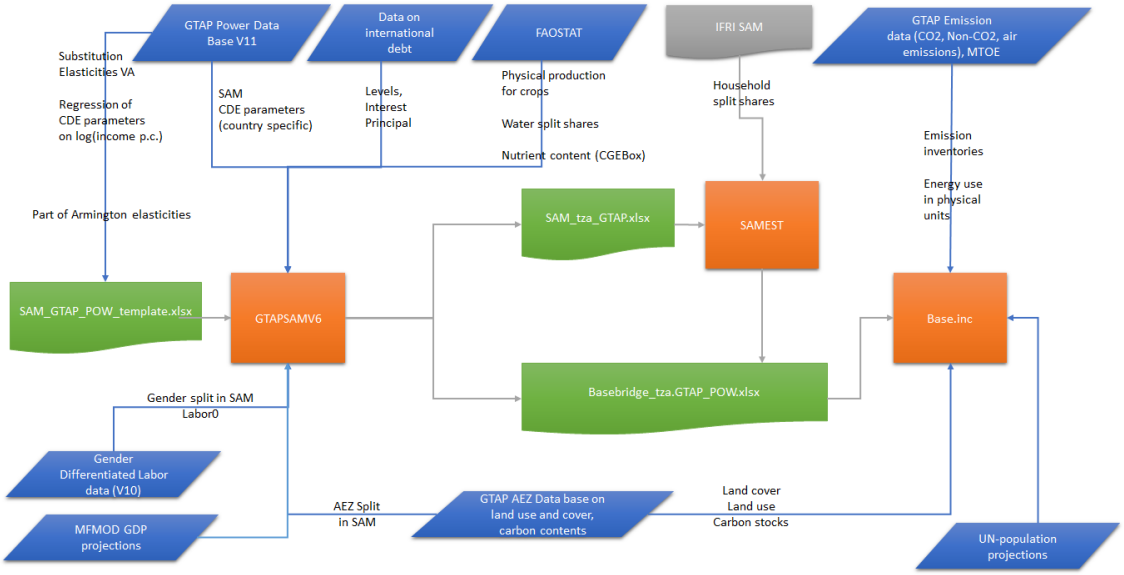
Graphic 1: Interface option for constructing a SAM from GTAP data

The process, coded in GAMS, comprises a simple filter possibility to set to zero small cost and expenditure shares, defined relative to column/row totals. A balancing routine is run afterwards. The objective function of this SAM balancing step minimizes squared relative differences to the raw data. The aim is to remove errors from filtering, rounding or adding detail on Agro-Ecological Zones (AEZ), water, gender or regarding additional activities and government debt.

The balancing routine comprises the usual constraints which ensure equal column and row sums, plus several further conditions. They ensure firstly that exports cannot exceed production and secondly, that factor taxes cannot exceed factor cost at market prices. Thirdly, a constraint is added which ensures that B.o.T. is equal to Balance of Payment (B.o.P) net of transfers. In the standard GTAP model, the net trade position is the sole determinant of B.o.P. and the saving account is closed accordingly. The data driver for MANAGE-WB adds transactions related to foreign debt from an IMF database which could violate the B.o.P. condition. Data on domestic debt must be added by the analyst directly on the interface, as seen in the screenshot below.

An overview on the different data sources entering the step of constructing a SAM from GTAP and other data for the example of Tanzania, including some projections, are shown in Graphic 2 below. The step discussed above refers to the box GTAPSAMV6 as the start of the process. As seen from the Graphic, besides data from GTAP Power data, the GTAP-AEZ data are used to split up the land account. Moreover, data from a Food Agriculture Organization (FAO) study (FAO 2018) are used to split crops into an irrigated and rainfed variant, inspired by the work of Haqiqi et al. (2009). As these data are at country level, the same distribution shares of the crop variants are used for each AEZ. The data from the FAO study are also used to break up the land account into rents to land and to irrigation water. Some activity splits (electrified transport, cattle raising) can be activated via the GUI and are used for the Tanzania case.

The labor account is usually split up by gender. Here, two sources can be used depending on data availability. Using recent ILOSTAT data, wage sums differentiated by male and female at the level of the labor categories and activities of the GTAP V11 data base has been estimated for about seventy regions, including Tanzania as the chosen example. In case a SAM is constructed for a country not found in the data set, distribution shares are derived from the so-called Gender Differentiated Labor database by the World Bank (GDLD, ref.) is used which matches to the GTAP V10a data set. In this case, SAM will feature four labor categories (higher or lower skilled, male and female).



Graphic 2: Example of a data flow to populate data for a country based on the GTAP and other data

Substitution elasticities, expansion and substitution parameters for the Constant Difference of Elasticity (CDE) demand system are taken from GTAP, too. Armington elasticities stem from Fontagne et al. (2022), only missing ones are taken from GTAP.

The expansion and substitution parameters of the CDE demand system for all countries and model regions covered by GTAP V11 have been regressed on the logarithm of GDP per capita, after being aggregated to product groups which are also used in the simulation model. These coefficients can be used during baseline construction to shift the parameters of the demand system to consider income dynamics. To quantify supply of calories, fat and proteins, estimated nutrient contents based on Britz (2020) are stored along with the other parameters. To ease interaction with bio-physical models, physical production data are added for some clearly defined agri-food commodities.

Bottom-up construction of the SAM

If the SAM is not constructed from GTAP data, the process consists of three major steps: i) collecting reliable data, ii) building a macro-SAM and updating the proto-SAM, iii)splitting the updated SAM if needed.

First, a suitable SAM or data to construct one needs to be located. Suitability refers to actuality and sufficient detail, reflecting project aims as agreed upon with the client. This might imply commodity detail beyond what is found in the GTAP database, for instance, to cover products accounting for larger shares of export revenues. It regularly also involves incorporating household details or labor markets segments in the SAM, often in close cooperation with the World Bank's poverty team which manages the household survey data and related tools available at the Bank.

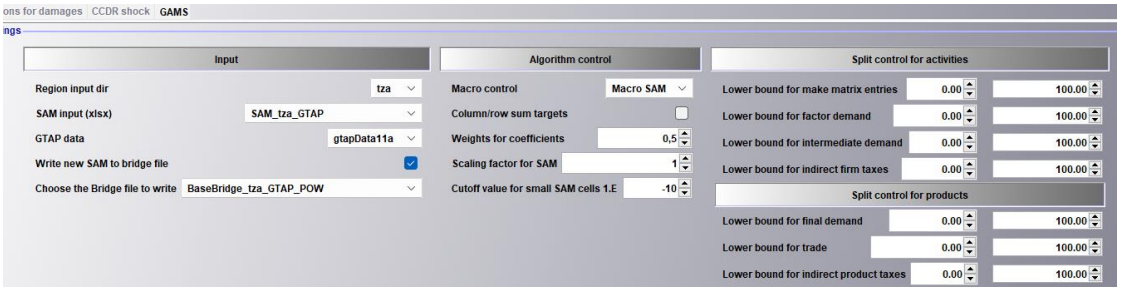
The process of finding appropriate data is typically supported by the bank's country desk in the respective country or directly by the client government. This also helps to bridge potential language barriers if statistical data are available in local languages, only. If the MANAGE-WB team can locate a newer IO-Table than the one currently used by the GTAP center, it follows a policy to share it with the center.

For further processing, the potentially not fully balanced SAM is copied into a sheet of an EXCEL-Workbook which serves an input in the GAMS based updating and splitting utility of MANAGE-WB (SAMEST in Graphic 2). Besides the SAM, the workbook hosts an aggregation mapping from the raw to the proto-SAM to construct, the non-GTAP derived split factors, for instance relating to household detail as well as the information on the structure of the macro-SAM and its link to the detailed SAM accounts.

The construction of this Macro-SAM is the second important step after a detailed SAM is made available. The Macro-SAM will close the statistical time gap further by incorporating the latest national accounts available which are usually more up to date than the detailed SAM. The accounts of the Macro-SAM will typically at least comprise the disaggregation of GDP to different final demands also used by GTAP as mentioned above, but preferrable also more detail on income generation.

The proto-SAM is next updated based on this Macro-SAM, which requires re-balancing. During updating, cost and expenditure are usually targeted as shares in column totals, whereas total expenditures for final demanders or production values of the sectors are targeted in absolute terms. The Macro-SAM accounts serve as constraints in the updating and balancing step. The Macro-SAM must be already perfectly balanced and is not subject to change. Formally, we assume that relative errors of the proto-SAM cells or coefficient are normally distributed to define the updating problem as a quadratically constrained problem with a quadratic objective using a Highest Posteriori Density estimator (see Britz, 2021). During updating, certain cells of the detailed SAM can be fixed. This is useful to recover specific transactions at a detail beyond the Macro-SAM, such as imports and exports of certain products. The updating process can be also applied to GTAP derived SAM.

The relevant interface options are shown below in Graphic 3.



Graphic 3: Interface option for balancing and splitting a SAM bottom-up

The updated SAM is then split to further detail if deemed necessary. During splitting, the cells of the balanced SAM are treated as fixed and given. As in the step above, it is possible to fix certain SAM dis-aggregated cells. If the project has a focus on GHG emissions, the power generation sector is usually disaggregated to further detail based on the GTAP-Power Data base. Frequently, also additional agricultural detail is provided by the GTAP database. Instead of using the GTAP Power Data Base to derive split factors, the analyst can also opt to use more detailed GTAP derived SAMs constructed as discussed in the first section, for instance, to add a dis-aggregation of land rent to AEZ level or gender differentiated detail for labor accounts.

The splitting step can be also applied to a GTAP derived SAM for a country. For instance, detail on multiple households can be added as in the example shown in Graphic 2 for Tanzania. The graphic also depicts the central role of the bridge-file, here named “basebridge_tza_GTAP_POW” as the inventory of all country specific data.

The team has now started for some applications to first construct a GTAP-SAM with AEZ detail as detailed above and use this SAM to split out agricultural detail including the AEZ differentiation for the land account. If the GTAP database or a GTAP derived SAM is used to construct splitting shares, the EXCEL workbook provides the necessary mappings between the SAM accounts on the GTAP products, sectors and factors. The calculation of the split shares is part of the common GAMS code of MANAGE-WB. The split problem uses linear constraints, only, and the same approach for the objective function.

Various statistics are produced during updating and splitting. As part of the Graphical User Interface, a reporting tool allows the analyst to easily compare targeted and final values and to inspect certain statistics.

Organizing the data in MANAGE-WB

The code of MANAGE-WB clearly separates code from data. Moreover, project specific data and code, the latter basically consisting solely of shock files, are hosted in project specific folders to clearly separate common code and data from project specific ones. Common code and data are all under version control. The common data comprise the global databases shared across projects (UN population projections, GDP projections by the Word Bank, GTAP data, FAO yield and acreage projections etc.).

The country specific data (SAM, physical labor statistics, emission inventories etc.), parameters and projections used to construct a baseline or counterfactuals (GDP, on

demand its composition, power generation shares, Value Added (VA) composition by broader sectors, carbon taxes or emission ceilings etc.) are assembled in a so-called bridge-file, stored in the project specific folder. It also comprises all necessary sets and maps, including the nesting structures for the production function, final demand and factor supply. This EXCEL workbook acts as the unique entry point for baseline construction and counterfactuals. Mapping sets for activities, products, factors and tax streams are defined in the bridge file to link the GTAP data to a given SAM.

This GTAP data in the common folder also comprises the "satellite" data which are extensively used in MANAGE-WB. This includes CO2 and non-CO2 emissions, air emissions, data on land use and cover, and carbon stocks. How they are used is depicted in graphic 2 above. Where the benchmark year of the country SAM deviates from the GTAP data, GHG emission inventories can be provided. Along with mapping information how these inventories relate to the GTAP products and activities, scaling factors are derived to align the GTAP emission data with the given inventory. From there, emission factors are derived.

Discussion and summary

As discussed above, based on the example of MANAGE-WB model, the GTAP database is an interesting starting point or complement to analysis with single country CGE models. This includes the so-called satellite accounts on emissions and land use. Compared to a bottom-up construction of a single country SAM for a project, using the GTAP database is the faster approach but didacts the maximal resolution regarding product, activities and factors. Depending on the age of the data underlying an IO-Table sent by a data provider to the GTAP center, there can also be a considerable statistical time gap. Especially for fast growing economies, this can imply outdated cost and expenditures shares due to technological and income dynamics. Moreover, rendering bi-lateral trade data globally consistent as required for a global database can provoke unwanted corrections to import and export statistics from a single country perspective. If these potential disadvantages of using a harmonized and readily available global database such as GTAP outweigh the considerable efforts to build instead a country specific SAM bottom-up is case and project specific. For instance, reducing the statistical time gap requires access to a younger IOT or Supply and Use Table (SUT) for a country, which might not always be possible. The data utilities as part of MANAGE-WB offer multiple options to improve a single-country SAM and related data with the GTAP database. Ready-to-use code calculates split-shares where the GTAP database offers higher detail such as regularly for power generation.

As the GTAP Data base now covers many of the low and middle-income countries subject to studies of MANAGE-WB, one could compare findings based on using a SAM based on the GTAP data against those based on a bottom-up built SAM. This could inform on how sensitive results are to these different input data. There are multiple reasons why such systematic comparisons are unlikely. First, clients scrutinize the quality of a country-specific study often based on comparing their data to the ones used in the study. A larger time gap or other deviations from their data, such as from

corrections to render data consistent to a global database, are easily taken as an indicator of a poor quality of the study as a whole. Second, repeated findings from comparison studies that key results, such as changes in GDP, do not depend much on the database are not proof that the database chosen does not matter considerably for a new study. Finally, as the databases compared are likely to be not completely identical with regard to products, sectors, factors, tax instruments etc., part of the shocks would need to be redefined, which could be quite labor intensive. Therefore, building a SAM bottom-up will remain the dominant approach for MANAGE-WB for any larger study.

References

Aguiar, A., Chepeliev, M., Corong, E. & van der Mensbrugghe, D. (2022). The GTAP Data Base: Version 11. *Journal of Global Economic Analysis, 7(2)*, 1-37

Britz, W. (2021): Comparing Penalty Functions in Balancing and Dis-aggregating Social Accounting Matrices, *Journal of Global Economic Analysis* 6(1): 34-81

Britz, Wolfgang, 2020. "**Maintaining plausible calorie intakes, crop yields and crop land expansion in long-run simulations with Computable General Equilibrium Models**," Discussion Papers 302922, University of Bonn, Institute for Food and Resource Economics.

Fontagné, L., Guimbard, H., & Orefice, G. (2022). A new dataset on product-level trade elasticities. *Data in Brief*, *45*, 108668

FAO 2018, The future of food and agriculture – Alternative pathways to 2050. Rome. 224 pp. Licence: CC BY-NC-SA 3.0 IGO

Haqiqi, I., Taheripour, F., Liu, J., & van der Mensbrugghe, D. (2016). Introducing irrigation water into GTAP data base version 9. *Journal of Global Economic Analysis*, *1*(2), 116-155