Journal of Global Economic Analysis, Volume 7 (2022), No. 1, pp. 44-75.

Disaggregating Agro-Food Sectors in the GTAP Data Base

BY WOLFGANG BRITZ[a]

We discuss the construction of the GTAP-AGROFOOD database for CGE modelling which is consistent with the standard GTAP Data Base Version 10a (Aguiar et al. 2019). GTAP-AGROFOOD departs from the full detail of 141 regions, 65 products and 8 factors in GTAP, by introducing 51 additional agro-food products and sectors in lieu of 11 original GTAP agro-food products, thereby yielding a database with 105 products. Such additional detail can improve and enrich, for instance, analysis of trade, bio-economy or climate change mitigation or adaptation issues. It also eases linkage to other, more detailed data sets, such as for nutrition accounting or irrigation water use. The main data sources used are the FABIO Multi-Regional Input-Output Database (Bruckner et al. 2019) which reports on production, land, seed, feed, and food use, mainly for primary agricultural products, and market balances for dairy products from FAOSTAT. They are combined with TASTE V10a (Pelikan et al. 2020) which provides bilateral trade and tariffs revenues at the level of tariff lines. The balancing methodology which ensures consistency with the GTAP Data Base is based on the linear loss based split utility of CGEBox (Britz 2021).

JEL codes: C67, C63.

Keywords: Computable general equilibrium analysis; Agro-food analysis.

1. Introduction

Despite the small weight of primary agriculture and agro-food industries in the global economy, the GTAP Data Base Version 10 (Aguiar et al., 2019) continues to provide quite some detail for agro-food sectors, with 14 primary agricultural sectors, 8 sectors related to food processing and 5 other sectors which can be linked to the wider bio-economy. But GTAP provides less detail compared to what specialized modelling systems for agro-food markets or land-use can offer, such as CAPRI (Britz and Witzke, 2014), GLOBIOM (Krey et al., 2020) or AGLINK-COSIMO (OECD, 2015). CAPRI, to give an example, comprises around 60 products in its global market model, partly produced by multiple activities in its supply module for Europe. Such specialized partial-equilibrium models however miss the economy-wide perspective of Computable General Equilibrium (CGE) models. There are several reasons why additional detail for agro-food products and sectors can be useful in quantitative analysis with CGE models. First, details for agro-food products might be crucial for a specific analysis, for instance, with regard to nutritional aspects, to environmental impacts linked to irrigation water or chemicals' use, or due to different climate sensitivities. Focus on specific bio-economy value chains can also warrant specific detail, such as in Nong et al. (2020), Escobar et al. (2018), and van Meijl et al. (2018). Second, product aggregation can also suggest substitution possibilities and thus competition in demand and trade which does not exist or only exists to a limited degree at the more detailed product level. For example, the product category "v_f" (vegetables and fruits) in GTAP comprises cash crops for exports as well as certain root and tubers such as Yams which are hardly traded and fall in the home

[a] Institute for Food and Resource Economics, University Bonn, Nussallee 21, D-53229 Bonn, Germany (wolfgang.britz@ilr.uni-bonn.de)

production/consumption category, and covers both temperate and (sub)tropical zone products. Third, agro-food products contain tariff lines with the highest protection rates and considerable tariff dispersion across products inside the same aggregate GTAP product. Thus, the additional detail provided by the database discussed here can improve the analysis and also help to apply approaches such as GTAP-HS (Chepeliev et al., 2021).

Several scholars and projects have therefore developed Social Accounting Matrices (SAMs) with higher agro-food detail. Mainar-Causapé et al. (2018), for instance, offer detailed SAMs for 28 European Member States drawing mainly on the database of the CAPRI model and EUROSTAT data. Their database distinguishes 45 sectors for primary agriculture and food processing. They introduce additional detail for bio-chemicals, four types of biofuels, energy crops, pellets and bioelectricity from the database of a special version of the MAGNET model from Philippidis et al. (2018). Philippidis et al. (2018) introduce this detail by splitting the GTAP Data Base based on multiple sources, and add fertilizers as a separate product and activity. Existing applications of CGEBox such as Escobar and Britz (2021) have introduced additional agro-food detail based on the same methodology and software as discussed here. In single country analysis, additional detail for agro-food sectors is also not uncommon such as in Nechifor et al. (2021).

But these databases with additional agro-food detail are either not directly compatible with the GTAP Data Base or not publicly available. The aim of the exercise documented here is to provide a data-set in the same format as the GTAP Data Base which can be used by other scholars. It maintains the full regional and factor detail of the original GTAP V10 database, but with increased agro-food detail, mainly for primary agricultural production and dairy. Users can then aggregate the database to their specific needs, including the possibility to disaggregate it further, for instance to introduce a differentiation between irrigated and rainfed crop activities. This paper documents key aspects of the process to arrive at this database, provides some statistics and discusses a comparison of key results compared to the GTAP V10 Data Base under a multi-lateral trade liberalisation experiment.

The process which disaggregates the GTAP Data Base to the additional detail ensures that the original aggregated data (intermediate, final and factor demands, tax and subsidy revenues, bilateral trade and related export and import revenues etc.) are consistently disaggregated. However, as detailed later, a filtered version of the GTAP Data Base where small cost and use shares have been removed is subjected to this disaggregation. This implies some quite small differences to the original, unfiltered GTAP Data Base when the detailed the database is re-aggregated to GTAP product level. Thanks to filtering, the GTAP-AGROFOOD database is about the same size as the GTAP Data Base in terms of non-zero items despite its higher detail.

2. Methodology

Figure 1[2] informs on the additional detail provided by the GTAP-AGROFOOD database. Each column depicts the more detailed products belonging to a product in the GTAP Data Base, shown in black above. Processed food products which are derived from primary ones share the same colour. The level of additional detail clearly differs across the 11 disaggregated products in the GTAP Data Base.

[2] See also Table A1 in the Appendix which reports on labels used for the new products and the detailed descriptions of the GTAP products.

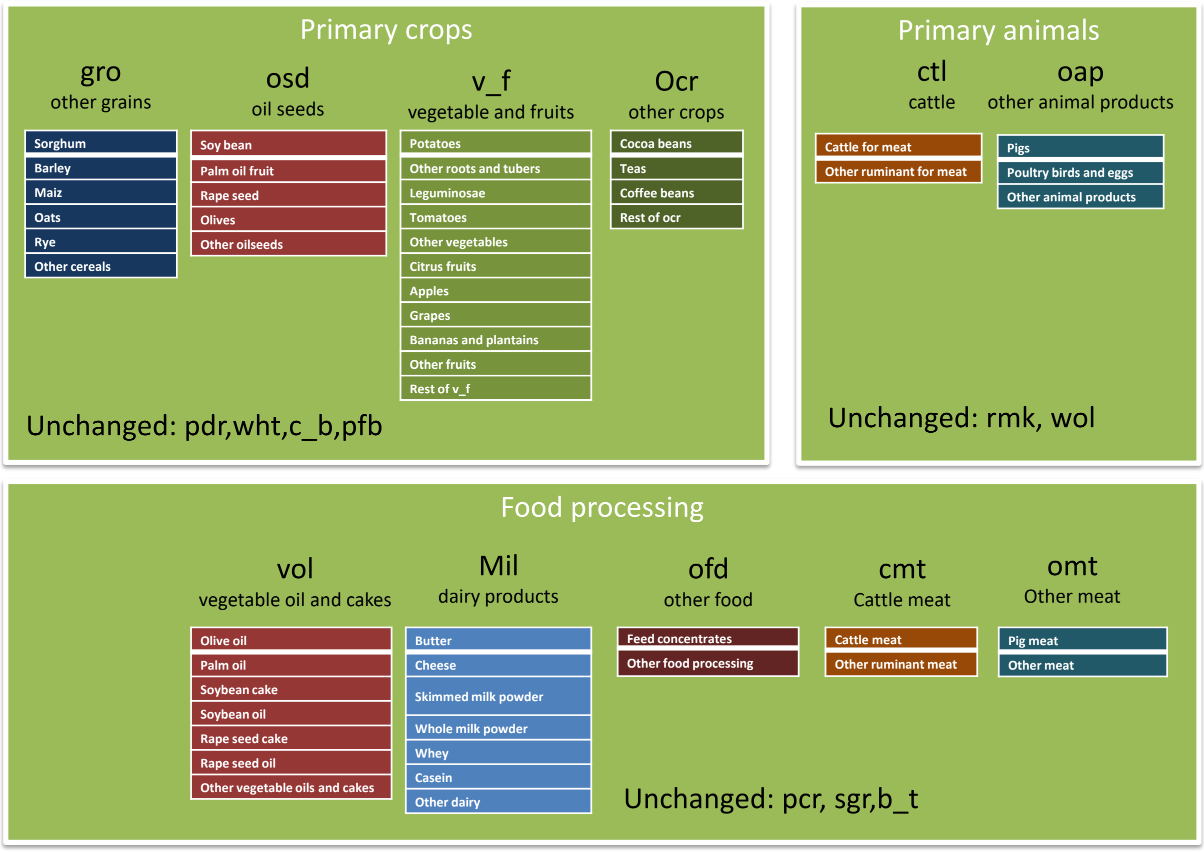

Figure 1. Product detail of the GTAP-AGROFOOD database

Source: Author

FABIO offers typically even more detail such that an even more disaggregated database could be derived. To give an example, the "Root and tubers" category in GTAP-AGRIFOOD is composed of four products in FABIO (Cassava; Yams; Sweet potatoes; Roots, others). Such aggregation over FABIO products and sectors reflects agronomic knowledge on product similarities and avoids disaggregated products and sectors with quite small shares relative to the related GTAP product and activity (see also Table 6 and Table 7). Introducing even more products provides also challenges for the numerical stability of the data processing chain. How such detail is introduced is discussed in the following.

The list of new activities is identical to the list of products shown below. For models which support a non-diagonal make structure, it is recommended to aggregate the two separate activities which output rape oil and cake to one rape crushing activity which produces both the cake and the oil, similarly for soybean crushing. Equally, all dairy processing activities can be combined into one dairy industry with multiple dairy outputs.

2.1 Overview of process and filtering

The split procedure is integrated in the database driver of CGEBox (Britz and Van der Mensbrugghe, 2018)[3], a GAMS based modular and flexible platform for CGE modelling. Its software code, including the scripts used to produce GTAP-AGROFOOD database are open-source and open-access. Figure 2 gives an overview of the different steps discussed in this paper. We depart from the GTAP Data Base with its full regional, product/sector and

[3] The code and data not licensed from the GTAP Center are available on the CGEBox repository https://svn1.agp.uni-bonn.de/svn/cgebox/. The status of the scripts used here is found under revision number 4150.

primary factor detail and filter out tiny cost and expenditure shares, to arrive at a sparse and better conditioned data set to disaggregate. This step is not specific to this application and therefore discussed in Appendix A1, along with statistics on its outcome. Next, a first disaggregation step takes place which requires information at a more detailed level with regard to production, demand, cost structures and bilateral trade. It introduces all detail except for dairy products. The resulting data are stored back in the same format as the original GTAP Data Base. This provides the input for a second round which adds detail for dairy and delivers a final database which can be used for benchmarking.

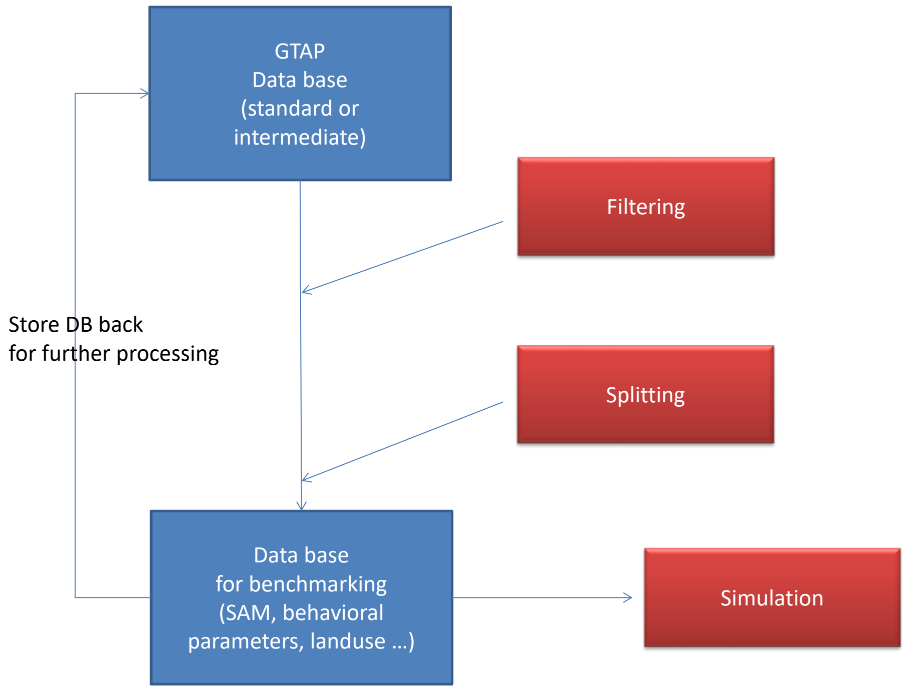

Figure 2. Overview on database construction

Source: Author

The additional information needed draws on different data sources, reported partly in quantities, not values, and differing in definitions from the GTAP Data Base. For instance, most data on demand are reported in contrast to the GTAP Data Base in so-called primary product equivalents. Any information is therefore converted into shares on totals to disaggregate, as discussed in sections 2.3 and 2.4. The resulting shares are not necessarily mutually compatible with each other. A Linear Programming framework therefore corrects these shares and ensures the consistent disaggregation while maintaining as much original information as possible, as discussed in section 2.6.

A user of the database will perform (some of) these steps again, starting from the GTAP-AGROFOOD database, aggregating to the desired regional, product and factor detail, and potentially filtering again and splitting further, for instance, to disaggregate to rain-fed and irrigated crop activities. Processing of the GTAP-AGROFOOD database can also be performed with some other software, for instance using the GEMPACK programs ViewHar to convert the GDX containers to HAR files and aggregate them to desired detail with GTAPAgg(2).

2.2 Corrections to the GTAP Data Base

Before filtering, some corrections to the GTAP Data Base are introduced. This is possible as the filtering process will adjust transactions to fulfil the necessary balancing conditions, such as absorption of given imports, production of given exports, domestic market clearing, production value equal costs for all activities and absorption of income by expenditures. The balancing step is based on a Quadratic Programming set-up which penalizes squared relative deviations for data items with a non-zero a priori value and an additional penalty for the introduction of estimates with a zero a priori value to favour sparsity. It considers both changes from removing small transactions or from corrections based on plausibility considerations.

Specifically, the following corrections[4] to the original GTAP Data Base are introduced. Crop by crop inter-industry use should depict seed use. Therefore, use of primary crop products by other crop activities - such as potatoes being used in the production of wheat - is deleted from the database, independent of the size of the reported transaction. Excluded from deletion is the use of other crops "ocr" which comprises the production of seeds. These corrections will remove a larger number of usually quite small intermediate demand positions which are hard to interpret from an agronomic viewpoint.

Equally, intermediate demands of the remaining seed costs at agent prices (*vifa, vdfa*) cost are downward corrected with a uniform correction factor to ensure that total seed cost does not exceed 50% of the total production value of the crop. The related transactions at market prices (*vifm*, *vdfm*) are scaled with the same factor to maintain indirect tax rates.

The filtering step will also ensure that any production activity has a minimum cost share of 10% of value added at agent prices in total production costs. This is guaranteed by constraints in the quadratic framework. Very small or even zero cost shares of value added can provoke problems in simulations with CGE models, especially when intermediate input coefficients are fixed and no substitution between the value-added and intermediate composites is possible. In this case, the supply response is solely rooted in substitution inside the value-added nests. Without some minimal cost share of the value-added composite, shocks might not be absorbed in simulations and infeasibilities may result.

Equally, we delete all tax rates (taxes on imported or domestic demand, factor taxes and subsidies, production taxes, import taxes, export taxes and subsidies) smaller than 0.0001%. We take an indirect tax rate on domestic demand *dintx* as an example. It is derived from the value of indirect demand at agent and market prices. For instance, for imported demand by households, the related tax rate *rtpi* is derived from the value at agent prices *vipa* and at market prices *vipm* reported in the GTAP Data Base as:

$$rtpi = vipa/vipm - 1 \tag{1}$$

A value of 1.E-6 would mean that the agent would need to pay a tax of 1 cent on a purchase of 10,000 USD. Such small tax rates are unlikely to be part of any law book (related transaction costs most probably exceed tax revenues) and therefore are rather the outcome of database processing. Small tax rates are frequently observed if the GTAP Data Base is aggregated over regions, products and factors which results in averages over non-tax and taxed transactions. But this is not the case here, as we disaggregate this database in full detail. Removing such small tax rates should not affect simulation outcomes, but makes the database smaller.

We also correct implausibly high relative subsidization of factor use. To do so, we relate the sum of factor taxes paid and subsidies received for each factor by an activity to its costs at market prices. If the resulting ad-valorem rate is smaller than zero, the use of the factor in the activity is subsidized. If this ad-valorem subsidy rate is below minus 90%, we reset the rate

[4] Users of CGEBox have control over these different corrections by its Graphical User Interface and can construct their own variants of the database where different thresholds are used or some of the corrections are discarded.

to minus 90% and recalculate the value at market prices. This ensures that each production activity covers at least 10% of the factor costs at market prices. Leaving even higher subsidization rates in the database has shown to provoke problems in simulations as they almost remove allocative responses to changes in market prices of factors. This is especially the case if high subsidization rates are found for the use of multiple factors in a production activity.

Tests were conducted with an extremely low filtering threshold (1.E-8%) where basically no small data were removed. Their outcomes suggest that the corrections discussed above provoke some differences between the unfiltered and filtered (and corrected) database. The split process detailed below will ensure that the resulting filtered and corrected database is consistently disaggregated. This had been tested by re-aggregating the resulting, more detailed database back to the product and activity detail of the original GTAP Data Base and comparing the SAM and the domestic and import demand matrices to the filtered version. But as stated, this test cannot hold fully for the uncorrected and unfiltered GTAP Data Base.

2.3 Derivation of split factors

Split factors for production, land use and intra-agro-food intermediate use are mainly derived from the FABIO MRIO (Bruckner et al., 2019[5]) which covers 130 agro-food products, 121 related activities and 192 countries, drawing largely on FAOSTAT data. FABIO is defined in physical units while the GTAP Data Base reports economic transactions in USD. Using physical production or use shares to disaggregate the GTAP Data Base provides inconsistent split factors if the prices of the detailed products per physical unit differ. For instance, take the quite different prices per metric ton for yams and tropical fruits produced for exports as examples, both part of the vegetable and fruit aggregate (v_f) in the GTAP Data Base. We therefore convert the physical quantities reported in FABIO into values using producer prices in USD as reported by FAOSTAT. Let x denote data reported in physical units in FABIO and p the attached prices, and *spliti, splitj*[6] the newly introduced, more disaggregated products, and i the original, more aggregate ones found in the GTAP Data Base. Split factors *splitFac* are then defined as:

$$splitFac_{spliti \in i} = \frac{x_{spliti}\, p_{spliti}}{\sum_{splitj \in i} x_{splitj}\, p_{splitj}} \tag{2}$$

Equation (2) shows that only price relations inside a product aggregate matter. Price notifications can show larger inter-annual fluctuations. Equally, the relationships between the producer prices reported by FAO and the prices underlying national accounts defining the SAM are not clear. We use therefore 2013-2015 average prices and not prices for the benchmark year 2014 of the GTAP Data Base. In many cases, FAOSTAT reports prices for multiple detailed products for one newly introduced disaggregated product *spliti*. In such cases, unweighted averages across these more detailed products in FAOSTAT are used. In some cases, no producer prices are reported for a country and product combination. We use world averages in such cases, however correcting the world average price for this product with the average relation across all products between world average and regional prices. If no prices at all are reported for a country, world averages are used. FAOSTAT does not report producer prices for all FABIO products, such as for live animals, and oils and cakes. In these cases, we use average world f.o.b. prices derived from FABIO instead. An analysis of

[5] A version of the FABIO database converted in GAMS format can be found in “xxx/data/fabio.gdx” where xxx is the root folder into which CGEBox was installed by the user, usually, this folder is called “cgebox”.

[6] The notation for indices such as *spliti* and names of variables or parameters follow closely the GAMS syntax of the related script.

the f.o.b. prices resulting from the bilateral trade data in FABIO showed hard to interpret differences at the global level for one product across exporters, or even for different import destinations from one exporter. To keep things transparent and to avoid unwarranted impacts on split factors from using regionally (or even bilaterally) differentiated f.o.b. prices, one global average f.o.b. price was derived for each FABIO product.

All market balance positions in FABIO are expressed in so-called primary product equivalents, as are the FAOSTAT data underlying FABIO. This representation of intermediate and final demand differs fundamentally from a standard IO-Table, not only by using quantities. Feed or food use of rye in FAOSTAT and FABIO, for instance, comprises both the direct demand for the primary product rye, but also of products derived thereof (flour, bread, feed concentrates etc. made from rye). The use of derived products is expressed as the amount of rye needed directly or indirectly (such as rye flour used in bread production) in their production, and summed up with its direct use after using detailed physical conversion rates. The GTAP Data Base will report derived use (bread, feed concentrates etc.) separately from the primary products in product categories such as "Other Food Processing". Related processing activities report intermediate demand of the primary products, which in FABIO would be shown instead in the market balances of the primary product as food or feed use. Processed food products and related production activities are therefore mostly absent in FABIO; exemptions are, for instance, vegetable cakes and oils.

Despite this conceptual difference, it is assumed that the FABIO market balance positions multiplied with producer prices provide valuable information for the disaggregation at the GTAP product level. To give an example, the food use reported for maize, rye, oats etc. in FABIO, multiplied with producer prices, for the cereals mapped to the aggregate GTAP product "gro" (Other Grains: maize (corn), sorghum, barley, rye, oats, millets, other cereals) defines shares to split up the final household and government demand in GTAP. This assumes that the share of rye in direct final household and government demand of other grains is similar to its combined direct and indirect (flour, bread …) use for human consumption. What is helpful here is that FABIO considers explicitly some important food processing activities (beer, wine, oil seed and cakes), such that, for instance, the use of barley for beer brewing is not considered when deriving final demand shares. Similarly, feed or food use of oilseeds does not comprise its use in processing industries which produce oil seeds and cakes. The available differentiation in FABIO between the domestic and bilateral origin for the use positions is discarded.

The physical land use data from FABIO in hectares provide the information to split up the factor use of land. This implies the assumption of uniform returns to land across the newly introduced activities relating to the same original activity in the GTAP Data Base.

FABIO data on intermediate input use are aggregated over importer regions and the domestic origin to derive at input demands, again multiplied with producer prices (and if missing, with global f.o.b. prices) as weights. These data are available for around 130 agro-food products and activities covered by the FABIO database, but not for other sectors (extraction, non-agro-food manufacturing, services). For these other intermediate demands, proportionality assumptions are employed (see section 2.5.6). Identical indirect and factor tax/subsidies rates are used for the split-up activities and products belonging to the same aggregated activity or product in the GTAP Data Base. The same holds for export taxes/subsidies.

Split factors for processed dairy products are based on the recently released market balances for processed animal products by FAOSTAT, using the average of the years 2014 and 2015. Related f.o.b. prices are world prices for the year 2015 taken from the OECD agricultural outlook, release 2014. For the price of the "Other dairy" product category, a guestimate is used.

FABIO provides also bilateral trade data in primary product equivalents. Here, the TASTE database (Pelikan et al., 2020) is used instead[7] which disaggregates consistently bilateral trade at c.i.f. prices in the GTAP Data Base to the tariff line. TASTE delivers also information on bilateral tariff revenues which are not available in FABIO. This seems important as tariff protection in agro-food sectors is high while ad-valorem rates of disaggregated products belonging to the same aggregate GTAP product can vary substantially (see, for instance, Chepeliev et al. 2021).

The split into domestic and bilateral import demand in FABIO draws largely on proportionality assumptions, as usually in MRIO analysis. FABIO is internally consistent, for instance, the reported bilateral exports and domestic demands by agents exhaust domestic output, all defined in primary product equivalent. Defining bilateral and total imports and exports shares from TASTE and use and production shares from FABIO will distort this internal consistency, even if the differentiation between domestic and bilateral import demand in FABIO is removed here before split factors are defined. This is one reason why the split factors for different positions might not be mutually compatible and an overall balancing step is necessary.

The trade data at the HS6 level in TASTE were extracted for the agro-food sectors to disaggregate and stored in a GDX container. Using TASTE requires a mapping from HS6 tariff line classifications to the FABIO products[8]. TASTE does not show imports for the XTW (rest-of-the-world) region, here, the FABIO trade data are used instead. The same holds for a few further exemptions where no imports or exports of a detailed product are reported in TASTE, but some in FABIO.

It is important to note that except for the GTAP Data Base and TASTE, all other data used are open-access. Accordingly, only a GTAP license is required to use the resulting database. Inconsistencies between FABIO and TASTE are a key reason for mutually incompatible a priori information. For example, import use of a product might be found in the cleansed a priori information without exports for the same product being reported to this country by any other one. Such cases pose some challenges for the balancing methodology to arrive at a balanced database. One possible solution consists of introducing a small a priori value where no data are reported. This is avoided here to maintain a sparse data set of reasonable size and to stick close to the original data, including reported zeros. How such cases with conflicting a priori information are treated and further detail on the data processing are discussed next.

2.4 Preparation for the split balancing routine

The assignments from FABIO, FAOSTAT and TASTE provide values in million of USD for the newly introduced products and activities. Sums of these estimates for the related aggregates are calculated according to the mapping between the new more disaggregated products and the related GTAP product (see Figure 1). Afterwards, the individual positions are divided by their calculated related totals (see equation (2) above). This leads to the usual representation as so-called split factors, i.e. shares on the totals to disaggregate. To avoid problems with tiny transactions resulting from applying these factors, any split factor below 1.E-5[9] is deleted, with the exemption of bilateral trade where tiny shares are quite common. The remaining split factors are then scaled to unity again.

With some split factors relating to original non-zero data being deleted, the remaining ones are scaled upwards. Assume, for instance, that the production of sorghum relative to

[7] In Britz 2021 which details the split algorithm and compares it to alternatives, the TASTE data were not used.

[8] The mapping can be found in "build/split/split_based_on_fabio.gms".

[9] The factor is defined to match the filter tolerance used, as discussed below, and can be smaller in other applications of the CGEBox split utility.

the total other grains production (gro) is below the chosen 0.001% threshold (a split factor below 1.E-5). In this case, in order to maintain the total reported production of other grains in the GTAP Data Base, the production of maize, barley etc. must be slightly increased instead. Given the low threshold, the advantage of avoiding related small shares in use and production is favoured over maintaining the original data.[10] Split factors are used as reported in Table 1.

The motivation of the reported rules is to use the most detailed information available, while having fallbacks if it is missing. Accordingly, explicit split factors reported for a specific SAM cell (intermediate input or factor demand by an activity, output coefficients, demand positions, import flows) are given precedence. If no split factor for an individual intermediate input demand is found, an aggregate split factor for the sum of intermediate demand ("int") might be used. For intermediate input demands of split-up products by split-up activities, such as seed use of a crop, the product of the "int" and "prod" split factors is used if no explicit factor is present. If afterwards split factors for intermediate input demands are still missing, the split factor for production ("prod") of the activity demanding the input is used instead. For outputs of a new activity, if no explicit split factor is given, the split factor for production "prod" is used as well. The same holds if factor demands of new activities are missing. During these assignments, split factors representing original zeros in the underlying database are not overwritten.

After these assignments, a check tests for each SAM cell to disaggregate if related split factors are present. Missing split factors can result from removing small shares or from "true" zeros in the original data. If no spit-factors are found for a SAM cell to be disaggregated, the aggregated value is distributed in equal shares to the disaggregated positions. Afterwards, split factors for position disaggregate are guaranteed to be present. A priori estimates of new SAM cells to introduce are generated by multiplying their split factor with the related aggregate SAM cell.

[10] Due to large differences in the economic size of the countries in the GTAP Data Base combined with quite different shares of the detailed agro-food products on total production, demand and trade, the linear program used to balance the data faces an extremely badly conditioned problem where quite small and large coefficients occur. For any larger transaction, the chosen absolute feasibility tolerance of 1E-7 when solving the split problem is clearly irrelevant as the resulting absolute error is 1E-7 times 1.E6 USD = 0.1 USD. But in cases of extremely small estimates of production values, for instance, the feasibility tolerance could imply already hard to defend deviations from targeted cost shares. This is one key reason to remove estimates which result from split factor below 1.E-5.

Table 1. Overview on the derivation of split factors

Split-factors	Derivation / Treatment	Data Source
Production	Output quantity as reported in market balance multiplied with producer or f.o.b. price	FABIO market balances, FAOSTAT processed dairy market balances, FAO producer prices, f.o.b. price derived from FABIO
Human and government consumption	Food demand as reported in market balance multiplied with producer or f.o.b. price	FABIO market balances, FAOSTAT processed dairy market balances, FAO producer prices, f.o.b. price derived from FABIO
Intermediate demand for disaggregated agro-food products by disaggregated activities	As reported in FABIO, aggregated over delivering regions and multiplied with producer or f.o.b. price	FABIO I/O Data set, FAO producer prices, f.o.b. price derived from FABIO
Intermediate demand of new sector "Feed concentrates" for (disaggregated) agro-food products	Derived from FABIO feed use, multiplied with producer price	FABIO I/O Data set, FAO producer prices, f.o.b. price derived from FABIO
Intermediate demand of new sector "Other feed processing" for (disaggregated) agro-food products	Derived from FABIO food and other processing demand, multiplied with producer price	FABIO market balances, FAO producer prices, f.o.b. price derived from FABIO
Intermediate demand of disaggregated activities for other non-agro-food products	Link to known split factors based on proportionality assumptions, see section 2.5.6	-
Factor demand for land	-	FABIO land use data
Other factor demands	Link to known split factors based on proportionality assumptions, see section 2.5.6	-
Import flows	-	TASTE
Tariffs	No differentiation if large differences between TASTE and the corrected/filtered GTAP Data Base are found	TASTE

Source: Author

A challenge provides the decomposition of the individual demand positions to the domestic and imported origin. The original FABIO data provide (partly) data on this decomposition, but they are derived from trade data defined in primary product equivalents, information which is discarded as TASTE is used instead for bilateral trade. Using the same shares for the domestic and imported origin for the disaggregated product as reported in the GTAP Data Base for the related aggregate can lead to inconsistent a priori information. For instance, import demand of a disaggregated product in a region might be created using the aggregate import share where no imports are reported in TASTE; or a domestic demand share might be proposed where no production is reported in FABIO. To address this, the aggregate domestic (*domShare*) and import (*impshare*) shares for all demanders for the aggregate products are calculated first. This step produces import and domestic demand shares in sum over all Armington agents[11] *aa* (production sectors, final demand categories) for each national market and the aggregated products to split *i*:

$$impShare_{r,i} = \frac{\sum_{aa} xm0_{r,i,aa}}{\sum_{aa} sam0_{r,i,aa}} \tag{3}$$

$$domShare_{r,i} = 1 - impShare_{r,i} \tag{4}$$

Symbols with a *0* at the end, such as *xm0*, refer to the benchmark or a priori values. *xm0* refers to import demand, and *sam0* to a SAM constructed from the GTAP Data Base. Endogenous variables under control of the split balancing framework start with a *v_*. Accordingly, *v_sam* refers to the newly introduced, disaggregated SAM cells.

Next, imports for disaggregated products into region *r* from all exporter regions *r0* are set in relation to their total demands to arrive at import share estimates for the disaggregated products *spliti* in this region:

$$impShare_{r,spliti} = \frac{\sum_{r0} v\_sam_{r,r0,spliti}\left(1 + p\_imptx_{r0,r,spliti}\right)}{\sum_{aa} v\_sam_{r,spliti,aa}} \tag{5}$$

where *spliti* are the newly, more disaggregated products and *p_imptx* the bilateral ad-valorem import tax rates applied on the import flow at c.i.f. value, derived from TASTE.

If domestic output (*output*) is observed, the domestic share is determined as the residual, with a security threshold of 1%. This threshold is important for cases where the a priori estimates for the import demand exceed the reported total demand:

$$domShare_{r,spliti} = max(0 + 0{,}01 \ni output_{r,spliti}\ ,1 - impShare_{r,spliti}) \tag{6}$$

Due to the maximum operator and the fact that domestic shares must become zero if no domestic output is reported, the import share has to be re-calculated afterwards:

$$impShare_{r,spliti} = 1 - domShare_{r,spliti} \tag{7}$$

Based on these shares, relations (*impRel*, *domRel*) between the aggregate and disaggregated import and domestic use shares are defined for the disaggregated products:

$$impRel_{r,spliti} = \frac{impShare_{r,spliti}}{\sum_{i \ni spliti} impShare_{r,i}} \tag{8}$$

$$domRel_{r,spliti} = \frac{domShare_{r,spliti}}{\sum_{i \ni spliti} domShare_{r,i}} \tag{9}$$

This allows calculating scaling factors *scaleArm* which ensure that the adding up condition for SAM entries to disaggregate will hold after the distribution factors for the domestic and imported origin calculated above are applied:

[11] The set *aa* refers to the "Armington agents" which represent the individual production sectors and final demand categories for which the GTAP Data Base reports separate domestic and import shares. In the GTAP Standard model, each of these agents thus can show specific preferences for the domestic and imported origin as stated by the Armington assumption (even if the related substitution elasticities are identical), while their composition of bilateral trade is identical.

$$scaleArm_{r,spliti,aa} = \sum_{i\ni spliti} \frac{xdm0(rnat,i,aa)\ domRel_{r,spliti} + xmm0(rnat,i,aa)\ impRel_{r,spliti}}{sam0_{r,i,aa}} \quad (10)$$

where *xdm0* refers to domestic demand. Afterwards, the reported aggregated import and domestic demand shares for a product and an Armington agent are multiplied with the demand for the split-up product, as derived from the split factors ($v\_sam_{r,spliti,aa}$), and with the relation between the disaggregated and aggregated import respectively domestic shares, as well as with the scaling factor:

$$v\_mm_{r,spliti,aa} = v\_sam_{r,spliti,aa} \sum_{i\ni spliti} \frac{xmm0_{r,i,aa}}{sam0_{r,i,aa}} impRel_{r,spliti}\ scaleArm_{r,spliti,aa} \quad (11)$$

$$v\_dm_{r,spliti,aa} = v\_sam_{r,spliti,aa} \sum_{i\ni spliti} \frac{xdm0_{r,i,aa}}{sam0_{r,i,aa}} domRel_{r,spliti}\ scaleArm_{r,spliti,aa} \quad (12)$$

v_mm refers to the disaggregated import demands and *v_dm* to the disaggregated domestic ones, while $xdm0$ and $xmm0$ depict the demands at market prices from the filtered GTAP Data Base.

Afterwards, further checks and potential changes to the a priori data are introduced:

- If no domestic $v\_dm_{r,spliti,aa}$ or import $v\_mm_{r,spliti,aa}$ demand for any disaggregated product *spliti* for an related aggregate demand $sam0_{r,i,aa}$ is found, the given totals $xmm0_{r,i,aa}$, $xdm0_{r,i,aa}$ are assigned to each related disaggregated product *spliti.* The so-constructed disaggregated values are scaled to exhaust the given total, i.e. divided by the number of disaggregated products mapped to the aggregate product category in GTAP.
- In case where exports are reported, but no import demand (from aggregated import demand over all Armington agents), the export flow is deleted if there are other export flows of disaggregated products relating to the same GTAP product on the same trade link. If export flows cannot be deleted as they are the only ones reported, each Armington agent showing import demand in each import country where such missing export flows are observed gets assigned a very tiny import demand.
- In cases of missing production, export flows and domestic demand are deleted.

Afterwards, indirect, export and factor tax rates are defined from the aggregate transactions and assigned to the disaggregated ones. Bilateral import tariffs are taken from TASTE as long as the necessary scaling factor to align the detailed tariff rate to given tariff revenues is not larger than five or smaller than one fifth. If the ad-valorem rate for one of the disaggregated products exceeds that for the related GTAP product by more than 100 percentage points, all disaggregated information for the related GTAP product is removed.

After such corrections, the disaggregated data are re-scaled to exhaust the related totals, such as reported values for a SAM cell to split. The resulting data provide the a priori information for the linear-loss estimator framework discussed below (for a technical discussion see Britz, 2021). The main equations found in the framework are documented in the next section. During the solution of the split problem, solely weights for deviations from the "secure" data (TASTE, production, intermediate demand and land use from FABIO) enter the objective function for the linear programming problem, with higher weights for production, bilateral trade and land use. A priori values derived from proportionality assumptions are not considered in the objective, they are handled by specific constraints as discussed in section 2.5.6.

The framework also introduces penalty terms in its objective function for very small entries of new SAM cells to favour sparsity and comprises constraints to control for reasonable factor cost shares and tax rates.

2.5 Split balancing routine

This section reports the main balancing equations used in the split routine. The CGEBox documentation[12] comprises additionally all equations as coded in GAMS with comments. For the split routine, it is not necessary to balance all positions of the SAM and related detailed additional matrices, such as the Armington decomposition, or detailed tax revenues. It is sufficient to focus on the parts of the SAM and additional data subject to the split, see Figure 3. The split balancing routine is defined as a Linear Program where the objective function minimizes the weighted sum of relative absolute deviations between the consistent final estimates and the a priori ones.

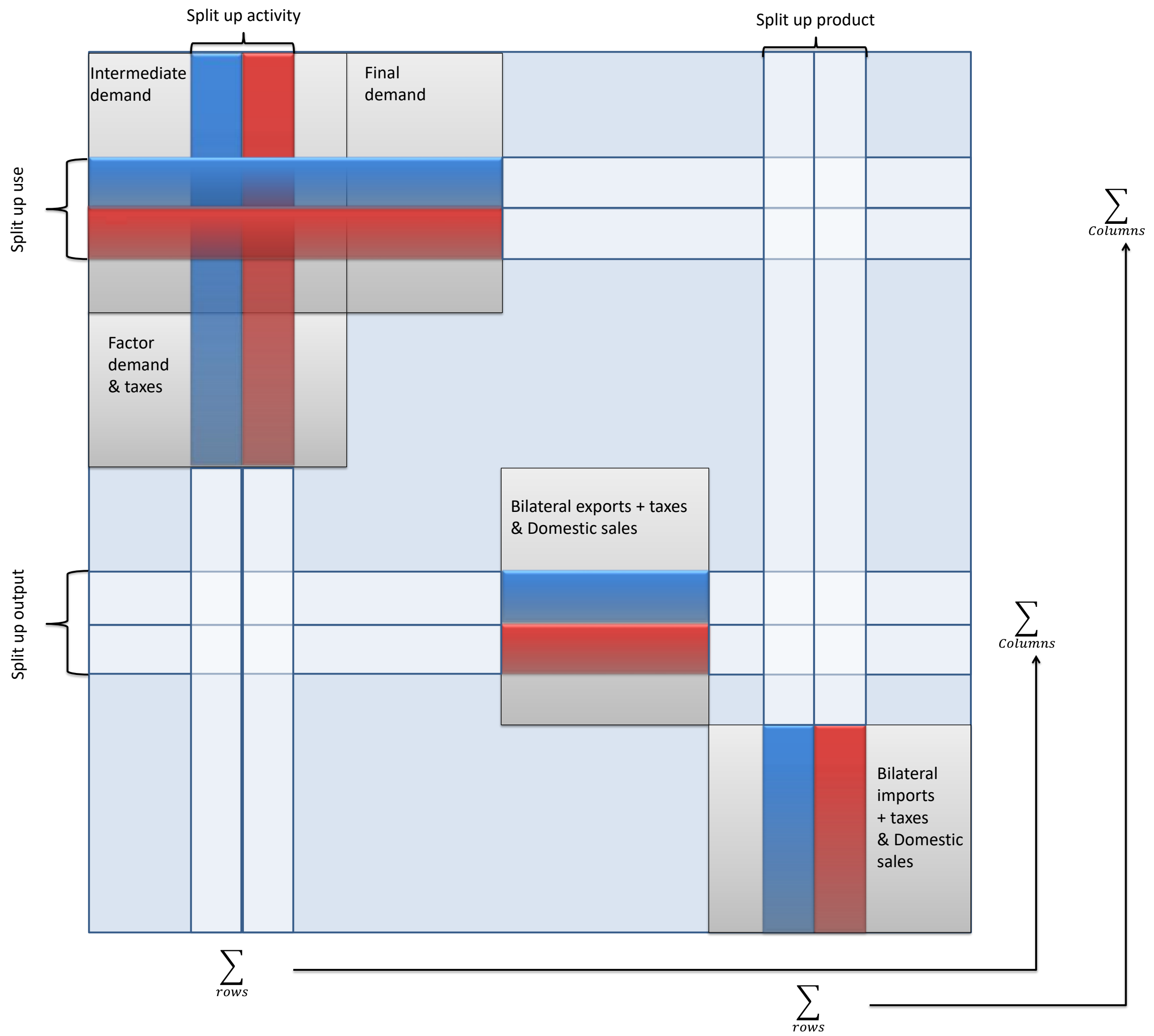

Figure 3. Schematic overview of a split problem with two additional products and sectors

Source: Author

[12] https://www.ilr.uni-bonn.de/em/rsrch/cgebox/cgebox_GUI.pdf

Note: The areas in red and blue depict the disaggregated SAM cells for the two new products and related disaggregated sectors. Areas in light blue are zero by definition. Areas in dark grey comprise unchanged non-zero SAM cells.

2.5.1 *Exhaustion of aggregate SAM cells*

In order to maintain the overall balance of the SAM, the given transactions in a SAM cell to disaggregate must be exhausted by the newly created, more detailed SAM cells.

For input and factor demands as well as taxes relating to aggregated activities to split *a*, this is ensured by the following constraint:

$$\overline{sam0_{r,\iota s,a}} = \sum_{splita \epsilon a} v\_sam_{r,is,splita} + \sum_{splita \in a, spliti \in is} v\_sam_{r,spliti,splita} \tag{13}$$

The first term on the RHS depicts SAM rows *is* which do not refer to products subject to the current split (for instance, demands for services in our example); transactions referring to such disaggregated SAM rows are added with the second expression. The set of new and original SAM rows is indexed with *is; splita* and *spliti* are the newly introduced activities and products, respectively. Variables which are zero by definition (such as demand for products to disaggregate in the matrix *v_sam*) are fixed exogenously to zero.

The exhaustion condition for the SAM columns relating to activities in the rows (outputs, production taxes) is defined as follows:

$$\overline{sam0_{r,a,\iota s}} = \sum_{splita \epsilon a} v\_sam_{r,splita,is} + \sum_{splita \in a, spliti \in is} v\_sam_{r,splita,spliti} \tag{14}$$

The exhaustion condition of SAM cells referring to products to split found in the rows of the SAM, such as exports (excluding input demand of split-up activities, captured in the equation above), is defined as:

$$\overline{sam0_{r,\iota,\iota s}} = \sum_{spliti \epsilon i} v_{sam_{r,spliti,is}} \tag{15}$$

The exhaustion of SAM cells referring to products to split in columns, such as imports (excluding the ones referring to split-up activities), is given by the following constraint:

$$\overline{sam0_{r,\iota s,\iota}} = \sum_{spliti \epsilon i} v_{sam_{r,is,spliti}} \tag{16}$$

Jointly, these four equations ensure that the original balance of the SAM is maintained.

2.5.2 *Balancing conditions for new SAM cells*

The next two equations ensure that newly introduced rows and columns in the SAM, representing the disaggregated products and activities, have matching sums:

$$\sum_{is} v\_sam_{r,spliti,is} = \sum_{is} v\_sam_{r,is,spliti} \tag{17}$$

$$\sum_{is} v\_sam_{r,splita,is} = \sum_{is} v\_sam_{r,is,splita} \tag{18}$$

2.5.3 *Exhaustion of tax revenues*

Four additional equations are needed which relate to factor taxes. The first two refer to exhaustion of the given totals, where $fxttxY0_{r,f,a}$ are given factor tax revenues for the aggregate activities to split, $p\_fcttx_{r,f,splita}$ are the a priori tax rates for the disaggregated activities and $v\_fcctx_{r,f,splita}$ are penalized tax revenues deviations, i.e. their expected value is zero. These variables which allow deviations from the uniform a priori tax rates are fixed to zero for the application discussed here, as no information on differentiated tax rates is available in FABIO or FAOSTAT. Accordingly, factor taxes for new activities all default to the taxes rates of the original aggregate activity.

$$fxttxY0_{r,f,a} = \sum_{splita \epsilon a} v\_sam_{r,f,splita}\, p\_fcttx_{r,f,splita} + v\_fcctx_{r,f,splita} \tag{19}$$

An identical equation is introduced for factor subsidies:

$$fxttsY0_{r,f,a} = \sum_{splita \epsilon a} v\_sam_{r,f,splita}\, p\_fctts_{r,f,splita} + v\_fccts_{r,f,splita} \tag{20}$$

The estimated tax and subsidy revenues add up to SAM entries for the new activities as follows:

$$v\_sam_{r,\text{"fcctx"},splita} = \sum_{f} v\_sam_{r,f,splita}\, p\_fcttx_{r,f,splita} + v\_fcctx_{r,f,splita} \tag{21}$$

$$v\_sam_{r,\text{"fccts"},splita} = \sum_{f} v\_sam_{r,f,splita}\, p\_fctts_{r,f,splita} + v\_fccts_{r,f,splita} \tag{22}$$

As tax revenue correction terms *v_fccts* and *v_fcctx* are fixed to zero in the application discussed here we do not document additional constraints in the SAM balancing framework which ensure that plausible tax rates are maintained.

2.5.4 *Decomposition of demands to the domestic and imported origin*

The decomposition of product demands $v\_sam_{r,i,aa}$ to the domestic $v\_dm_{r,i,aa}$ and imported origin $v\_mm_{r,i,aa}$ requires additional constraints. Aggregated demand at market prices for product *i* and all Armington agents *aa* as found in the SAM are defined as:

$$v\_sam_{r,i,aa} = v\_dm_{r,i,aa} + v\_mm_{r,i,aa} \tag{23}$$

In these constraints, all combinations of product *i* and Armington agents *aa* are considered where *i* refers to split-up products and/or *aa* to split-up activities. The import and domestic demand at market prices must be exhausted, as ensured by the two following constraints:

$$x\_mm0_{r,i,aa} = \sum_{spliti \epsilon i} v\_mm_{r,spliti,aa} + \sum_{splita \epsilon aa} v\_mm_{r,i,splita} \tag{24}$$

$$x\_dm0_{r,i,aa} = \sum_{spliti \epsilon i} v\_dm_{r,spliti,aa} + \sum_{splita \epsilon aa} v\_dm_{r,i,splita} \tag{25}$$

The exhaustion of the aggregated demands for the imported origin at agent prices is based on the following equations:

$$xma0_{r,i,aa} = \sum_{j\epsilon i} v\_mm_{r,j,aa}\left(1 + p\_mintx_{r,j,aa}\right) + v\_mintx_{r,j,aa} + \sum_{splita\epsilon aa} v\_mm_{r,j,splita}\left(1 + p\_mintx_{r,j,aa}\right) + v\_mintx_{r,j,aa} \quad (26)$$

Similar for the domestic origin:

$$xda0_{r,i,aa} = \sum_{j\epsilon i} v_{dm_{r,j,aa}}\left(1 + p_{dintx_{r,j,aa}}\right) + v_{dintx_{r,j,aa}} + \sum_{splita\epsilon aa} v_{dm_{r,j,splita}}\left(1 + p_{dintx_{r,j,aa}}\right) + v_{dintx_{r,j,aa}} \quad (27)$$

(28)

Where *p_mintx* and *p_dintx* are indirect tax rates identical to the aggregate product to split, and *v_mintx* and *v_dintx* are correction terms.

Based on these equations, indirect taxes revenues for each agent are defined:

$$sam0_{r,\text{"indtx"},aa} = \sum_{j\epsilon i} v_{mm_{r,j,aa}}\left(1 + p_{mintx_{r,j,aa}}\right) + v_{mintx_{r,j,aa}} \quad (29)$$

As for factor taxes, the indirect tax revenue correction terms *v_mintx* and *v_dindtx* are fixed to zero for this application as the a priori information assumes uniform tax rates. We therefore do not document additional constraints in the SAM balancing framework introduced to yield plausible tax rates.

2.5.5 *Market clearance conditions*

The following market clearance condition for the output of newly introduced products must hold. It ensures that total domestic use plus exports minus export taxes is equal to the domestic output:

$$\sum_{aa} v\_dm_{r,spliti,aa} + \sum_{r0} v\_sam_{r,spliti,r0} - v_{sam_{r,\text{"exptx"},spliti}} = \sum_{splita\epsilon aa} v\_sam_{r,splita,spliti} \quad (30)$$

Imports of split up products must be exhausted by the import demands of the Armington agents at c.i.f. prices, taking tariffs into account:

$$\sum_{aa} v\_mm_{r,spliti,aa} = \sum_{r0} v_{sam_{r,r0,spliti}}(1 + imptxyY0_{r0,r,spliti} + v\_imptx_{r0,r,spliti}) \quad (31)$$

Note there that *v_imptx* is an endogenous correction of the tariff revenues and $imptxyY0_{r0,r,spliti}$ are the priori ad-valorem tariff rates derived from TASTE.

The link between bilateral imports at c.i.f. prices (LHS) and exports at f.o.b. prices (RHS) is given by the following equation, assuming identical per unit trade margins across disaggregated products related to the same GTAP product:

$$v\_sam_{r,spliti,r0} = v\_sam_{r0,r,spliti}\ (1 + \sum_{spliti\epsilon i,m} \frac{xmarg0_{m,r0,i,r}}{sam0_{r0,i,r}} \quad (32)$$

Where $xmarg0_{m,r0,i,r}$ is the value of the transport margins demand by transport mode m as reported in the GTAP Data Base.

2.5.6 *Proportionality assumptions for cost structures*

The problem discussed here is quite typical for disaggregation exercises of SAMs where most often a priori information is available for selected intermediate input and factor demands, only. Split factors for the remaining cost items are then derived based on proportionality assumptions. Take a crop production activity as an example. FABIO will report data on its output value, seed and land use. From the given information, split factors are derived as discussed above. Multiplied with the given data in the GTAP Data Base for the related aggregated production activity, we can derive a priori estimates for the output value, seed and land use of the disaggregated activity, but not for other intermediate or factor demands. What we know however for these so far missing costs is that their sum should exhaust the difference between output value minus seed and land use cost for each disaggregated activity. We can calculate the sum of these missing costs for each newly introduced disaggregated activity and the related aggregate one in the GTAP Data Base. The next step then uses the reported cost relations inside the residual sum from the aggregate to derive estimates for so-far missing intermediate and primary factor demands.

Take the following simple example where only the production value and the land demand for two new crops is known. We also show the sum of intermediates costs, only:

Table 2. Didactic example of using proportionality assumption, given data

	Agg. crop	New crop 1	New crop 2
Intermediates	600		
Capital	100		
Labour	100		
Land	**200**	**50**	**150**
Production value	**1000**	**300**	**700**
Unknown cost sum	*800*	*250*	*550*

Source: Author

Estimates for so-car unknown costs (not shown in bold above) are now derived by using cost relations for the aggregate. For instance, the relation between total intermediate use and the unknown cost sum for the aggregate (600 / 800 = 75%) allows estimation of the intermediate use for the first disaggregated crop as 75% of 250 = 187.5. Identical calculations for all missing cost positions of the two crops result in the following internally consistent estimates:

Table 3. Didactic example of using proportionality assumption, estimated costs

	Agg. Crop	New crop 1	New crop 2
Intermediates	600	187.50	412.50
Capital	100	31.25	68.75
Labour	100	31.25	68.75
Land	**200**	**50**	**150**
Production value	**1000**	**300**	**700**
Unknown cost sum	*800*	*250*	*550*

Source: Author

We make these calculations transparent in our estimation framework by explicit constraints which control the proportionality assumptions. The aim is to stick to the cost relations of the aggregate where no other data are available. This allows us to steer the derived costs estimates solely based on the available a priori information (mostly production, feed, seed, land use etc.) and the cost relations of the related aggregate, without introducing penalty terms for these estimates in the objective function. If for instance, the balancing framework has to deviate from a priori estimate for the production value (for instance, to ensure consistency to trade), the introduced costs estimates will still follow the cost relations of the aggregate.

Let $propAssmpt$ define a three-dimensional indicator set of region, input or factor demand and activity combinations derived from proportionality assumptions. For these cases, the following conditions are enforced:

$$v\_sam_{r,is,splita} = v\_propAssmpt_{r,splita} \sum_{splita \in a} sam0_{r,is,a} \quad \bigwedge propAssmpt_{rnat,is,splita} \tag{33}$$

$$v\_sam_{r,spliti,a} = v\_propAssmpt_{r,spliti} \sum_{spliti \in i} sam0_{r,i,a} \quad \bigwedge propAssmpt_{rnat,spliti,a} \tag{34}$$

Where $v\_propAsspmt$ are endogenously determined distribution factors which ensure that all cost shares included in $propAssmpt$ receive the same share on the total. These factors are not entering the objective function. Equation (33) above will ensure that relations between cost items based on proportionality assumptions are kept fixed. Identical equations use the same distribution factor $v\_propAssmpt$ to define the imported and domestic demands.

Accordingly, the relation of the cost shares of capital and a labour category will be identical to the aggregate activity (see Table 4 for an example taken from the final data base), reflecting the constraint above. A lower yield, such as in the case of rye in the example below, will imply a higher cost share of land and lower cost shares of other factors (and intermediate demand, not shown in the table). This point is again picked up in the discussion section. Equation (34) relates to the distribution of split up products to non-split up activities where proportionality assumptions are used.

Table 4. Example of factor cost shares resulting from the split framework (Germany, other grains)

	Sorghum	Barley	Maize	Oats	Rye	Other cereals
Total factor demand	0.5112	0.5280	0.4666	0.5505	0.6095	0.5208
Land	0.1704	0.1841	0.1003	0.1979	0.2664	0.1634
Capital	0.1324	0.1337	0.1423	0.137	0.1333	0.1389
UnSkLab	0.1380	0.1393	0.1483	0.1428	0.1389	0.1448
SkLab	0.0703	0.0710	0.0756	0.0728	0.0708	0.0738
Capital/land	0.780	0.730	1.420	0.690	0.500	0.850
Capital/Unsklab	0.959	0.960	0.960	0.959	0.960	0.959

Note: Small differences in cost share relations stem from rounding when copy-pasting the results.

2.5.7 *Emission factors and land use*

In order to allow the use of the more detailed database in model extensions in CGEBox or other models drawing on GTAP, the split driver calculates CO2, non-CO2, energy use in mTOE and air emission factors per unit for the aggregated economic transactions to disaggregate, and multiplies them with the disaggregated ones to arrive at detailed emission inventories. The disaggregation of the land use database (AEZ) for the disaggregated crop and animal activities is based on the national shares on factor demand of land of the related GTAP activity.

3. Some statistics on the database

The new database features 105 products (plus the investment demands "cdgs"), 105 activities, 141 regions and 8 factors. For the bilateral trade transactions, this implies up to 141x141x105=1,908,576 transactions. Due to sparsity, around 1.3 million non-zero transactions are found in the final database (see Table 5). Another block of data with matrices of a similar dimensionality relates to domestic and imported intermediate demands by firms with potentially 141x106x105=1,272,384 non-zero entries. Here, the sparsity is higher; around 0.5 respectively 0.4 million non-zero transactions are reported. The expected use of the database is clearly that some regional aggregation (and potentially an aggregation of products and/or activities, and factors) will be applied to produce a benchmarking data set of suitable size and detail for the specific CGE analysis.

Table 5. Number of non-zero elements of largest parameters and density [%]

Parameter	Description	GTAP-V10 (unfiltered)	GTAP-AGRIFOOD
vtwr	Bilateral trade margin demand	1,011,171	1,308,192
vims	Bilateral trade, tariffs inclusive	1,272,397 *98.46%*	1,174,510 *56.26%*
vdfm	Domestic demand of firms, at market prices	604,044 *99.86%*	505,849 *32.23%*
vifm	Import demand of firms, at market prices	604,044 *99.86%*	403,179 *25,69%*
vfm	Factor demand of firms, at market prices	57,415 78.30%	80,035 *66.93%*

Note: Related transactions (such as at agent prices) not reported

Table 6 reports some key metrics for the newly introduced activities based on global averages. Of special interest when deciding how to potentially re-aggregate the detailed activities for analysis might be their shares relative to the related aggregated activity in the GTAP Data Base. The lines in bold report the given aggregated values and report their shares on the sum over all activities in the database. For other cereals, both oats and rye have relatively small shares globally and might be aggregated with the "other cereals" category. For oilseed, the economic importance of olives is quite small at the global level; this is also mirrored for olive oil. In the vegetable and fruit category, multiple disaggregated products show output shares on the related aggregate at five percent or below. This also reflects the relatively high product detail in this category with nine products.

The share of output relative to intermediate and factor demand, especially land, gives some indication of differences in cost shares. It should be noted that in the case of crops, FABIO will report data for output, seed use and land use in hectares, only, all other cost shares are based on proportionality assumptions. The differences for cost share shown in the table below additionally reflect the geographically dispersed production of the disaggregated crops as they also reflect cost share differences of the related GTAP activity across regions. In the case of animal activities, FABIO reports feed use by product and use of young animals, information which is also used in the construction of the disaggregated database.

Table 6. Value of global transactions for split-up activities and shares on aggregate activity

	Output	Interm. demand	Factor demand	Land	Output	Interm. demand	Factor demand	Land
ALL sectors	155,512.64	81,476.22	63,392.55	678.95				
GRO	**336.45**	**116.78**	**216.30**	**55.29**	**0.22%**	**0.14%**	**0.34%**	**8.14%**
Barley	30.81	11.93	17.65	5.48	9.16%	10.22%	8.16%	9.91%
Maize	259.86	94.64	163.93	40.73	77.24%	81.04%	75.79%	73.67%
Sorghum	22.04	3.99	17.60	3.96	6.55%	3.42%	8.14%	7.16%
Oats	2.77	1.09	1.72	0.86	0.82%	0.93%	0.80%	1.56%
Rye	1.81	0.82	1.01	0.51	0.54%	0.70%	0.47%	0.92%
Other cereals	19.16	4.31	14.39	3.75	5.69%	3.69%	6.65%	6.78%
OSD	**284.78**	**94.57**	**189.1**	**55.01**	**0.18%**	**0.12%**	**0.30%**	**8.10%**
Soy bean	98.95	37.89	60.29	18.76	34.75%	40.07%	31.88%	34.10%
Palm oil fruit	29.96	8.16	21.05	10.72	10.52%	8.63%	11.13%	19.49%
Rape seed	25.9	10.85	16.04	7.52	9.09%	11.47%	8.48%	13.67%
Olives	6.8	2.7	3.8	0.47	2.39%	2.86%	2.01%	0.85%
Other oilseeds	123.17	34.97	87.92	17.54	43.25%	36.98%	46.49%	31.89%
V_F	**779.07**	**211.15**	**546.62**	**133.37**	**0.50%**	**0.26%**	**0.86%**	**19.64%**
Potatoes	64.02	22.24	39.98	11.89	8.22%	10.53%	7.31%	8.92%
Other roots and tubers	147.55	23.09	120.26	17.62	18.94%	10.94%	22.00%	13.21%
Leguminosae	30.39	4.81	25.77	18.51	3.90%	2.28%	4.71%	13.88%
Other vegetables	229.04	62.43	159.95	38.07	29.40%	29.57%	29.26%	28.54%
Tomatoes	60.61	21.05	36.98	2.92	7.78%	9.97%	6.77%	2.19%
Apples	16	6.18	9.51	3.09	2.05%	2.93%	1.74%	2.32%
Grapes	33.66	11.92	20.98	5.71	4.32%	5.65%	3.84%	4.28%
Citrus fruits	37.91	12.89	24.01	4.85	4.87%	6.10%	4.39%	3.64%
Bananas and plantains	31.86	11.99	19.01	4.01	4.09%	5.68%	3.48%	3.01%
Other fruits	109.41	30.84	75.37	17.47	14.04%	14.61%	13.79%	13.10%
Rest of v_f	18.62	3.71	14.8	9.23	2.39%	1.76%	2.71%	6.92%
OCR	**279.3**	**111.91**	**163.54**	**49.77**	**0.18%**	**0.14%**	**0.26%**	**7.33%**
Cocoa beans	16.88	4.53	11.91	4.84	6.04%	4.05%	7.28%	9.72%
Teas	23.92	9.84	14.53	7.25	8.56%	8.79%	8.88%	14.57%
Coffee beans	39.47	12.58	25.59	9.17	14.13%	11.24%	15.65%	18.42%
Other crops	199.03	84.96	111.51	28.51	71.26%	75.92%	68.19%	57.28%
CTL	**383.18**	**194.35**	**185.61**	**46.7**	**0.25%**	**0.24%**	**0.29%**	**6.88%**
Cattle for meat	181.2	104.61	75.36	18.57	47.29%	53.83%	40.60%	39.76%
Other ruminant for meat	201.98	89.74	110.25	28.13	52.71%	46.17%	59.40%	60.24%
OAP	**664.37**	**376.55**	**280.82**	**82.15**	**0.43%**	**0.46%**	**0.44%**	**12.10%**
Pigs	153.98	78.94	75.28	44.83	23.18%	20.96%	26.81%	54.57%
Poultry birds and eggs	466.92	279.61	181.41	34.59	70.28%	74.26%	64.60%	42.11%
Other animal products	43.47	18	24.13	2.73	6.54%	4.78%	8.59%	3.32%

(Continued)

Table 6. Value of global transactions for split-up activities and shares on aggregate activity (Continued)

	Output	Interm. demand	Factor demand	Land	Output	Interm. demand	Factor demand	Land
VOL	**358.84**	**275.95**	**68.7**		**0.23%**	**0.34%**	**0.11%**	
Olive oil	19.84	16.62	2.34		5.53%	6.02%	3.41%	
Palm oil	65.1	50.24	13.84		18.14%	18.21%	20.15%	
Soybean cake	128.47	91.18	31.46		35.80%	33.04%	45.79%	
Soybean oil	78.18	61.71	12.35		21.79%	22.36%	17.98%	
Rape seed cake	16.22	13.71	1.87		4.52%	4.97%	2.72%	
Rape seed oil	51.03	42.49	6.84		14.22%	15.40%	9.96%	
CMT	**709.1**	**522.39**	**155.32**		**0.46%**	**0.64%**	**0.25%**	
Cattle meat	597.62	435.73	133.46		84.28%	83.41%	85.93%	
Other ruminant meat	111.48	86.66	21.86		18.65%	19.89%	16.38%	
OMT	**614.44**	**464.01**	**124.65**		**0.40%**	**0.57%**	**0.20%**	
Pig meat	309.38	235.29	60.92		50.35%	50.71%	48.87%	
Other meat	305.06	228.72	63.73		49.65%	49.29%	51.13%	
OFD	**2932.28**	**2030.67**	**743.11**		**1.89%**	**2.49%**	**1.17%**	
Feed concentrates	517.54	434.13	67.99		17.65%	21.38%	9.15%	
Other food processing	2414.74	1596.54	675.12		82.35%	78.62%	90.85%	

Source: Author

Another perspective gives Table 7 by showing differences in output shares across ten world regions, compared to the global average or sum. This underlines that the regional distribution of the disaggregated activities is quite disperse. Soy beans, to give an example, have a share of about 83% of aggregated oilseed production in Latin America, but only of about 3% in the European Union.

Table 7. Comparison of output share across world regions

	World	Australia, New Zealand	East Asia	Southeast Asia	South Asia	North America	Latin America	European Union 25	Middle East and North Africa	Sub-Saharan Africa	Rest of World
GRO	**336**	**3**	**86**	**10**	**7**	**60**	**33**	**25**	**24**	**69**	**20**
Barley	9.16%	66.26%	1.06%	0.10%	5.25%	2.56%	1.94%	30.98%	40.93%	0.98%	36.36%
Maize	77.24%	7.29%	96.85%	98.86%	64.64%	93.34%	91.92%	53.21%	51.76%	59.85%	44.42%
Sorghum	6.55%	15.20%	0.71%	0.29%	8.56%	3.47%	5.04%	0.00%	5.14%	22.27%	0.10%
Oats	0.82%	6.69%	0.00%	0.00%	0.00%	0.51%	0.74%	2.85%	0.51%	0.01%	5.92%
Rye	0.54%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.09%	0.47%	0.00%	4.74%
Other cereals	5.69%	4.56%	1.38%	0.76%	21.55%	0.12%	0.37%	9.87%	1.19%	16.89%	8.47%
OSD	**285**	**2**	**56**	**35**	**16**	**49**	**60**	**15**	**10**	**35**	**8**
Soy bean	34.75%	3.81%	17.65%	1.46%	9.02%	69.17%	83.01%	2.59%	3.79%	5.27%	13.30%
Palm oil fruit	10.52%	0.42%	0.00%	69.42%	0.00%	0.00%	3.06%	0.00%	0.00%	11.17%	0.13%
Rape seed	9.09%	34.32%	12.07%	0.00%	7.54%	15.92%	0.15%	52.48%	5.23%	0.12%	13.43%
Olives	2.39%	2.97%	0.00%	0.00%	0.06%	0.00%	0.68%	12.93%	44.47%	0.00%	0.88%
Other oilseeds	43.25%	58.47%	70.28%	29.12%	83.37%	14.91%	13.09%	31.99%	46.52%	83.45%	72.27%

(Continued)

Table 7. Comparison of output share across world regions (Continued).

	World	Australia, New Zealand	East Asia	Southeast Asia	South Asia	North America	Latin America	European Union 25	Middle East and North Africa	Sub-Saharan Africa	Rest of World
V_F	**779**	**11**	**100**	**42**	**69**	**84**	**80**	**74**	**96**	**176**	**48**
Potatoes	8.22%	6.76%	5.40%	0.57%	11.53%	10.56%	5.69%	17.47%	6.20%	1.94%	29.51%
Other roots and tubers	18.94%	1.52%	6.55%	26.81%	2.44%	0.90%	13.32%	0.01%	0.21%	66.11%	0.00%
Leguminosae	3.90%	24.57%	0.59%	4.27%	8.10%	6.67%	4.80%	2.16%	1.79%	3.37%	2.43%
Other vegetables	29.40%	23.24%	61.94%	21.28%	35.80%	29.36%	15.51%	30.52%	35.49%	11.09%	37.41%
Tomatoes	7.78%	3.05%	5.80%	0.80%	5.94%	13.49%	5.57%	10.45%	19.11%	2.03%	9.79%
Apples	2.05%	4.29%	2.95%	0.00%	1.15%	3.23%	1.78%	4.47%	2.27%	0.17%	3.98%
Grapes	4.32%	18.10%	1.28%	0.14%	0.96%	7.05%	5.09%	14.82%	5.24%	0.51%	6.08%
Citrus fruits	4.87%	2.95%	3.45%	2.31%	3.66%	10.27%	12.40%	5.15%	7.11%	0.75%	0.23%
Bananas and plantains	4.09%	1.52%	0.87%	7.29%	6.32%	0.76%	15.68%	0.20%	0.98%	5.15%	0.00%
Other fruits	14.04%	11.33%	10.11%	34.29%	22.42%	11.57%	19.33%	12.92%	18.69%	6.31%	9.24%
Rest of v_f	2.39%	2.67%	1.07%	2.24%	1.68%	6.14%	0.83%	1.83%	2.93%	2.57%	1.34%
OCR	**279**	**1**	**36**	**19**	**58**	**14**	**41**	**58**	**7**	**37**	**8**
Cocoa beans	6.04%	3.57%	0.00%	13.56%	0.19%	0.84%	7.25%	0.88%	3.03%	25.26%	12.62%
Teas	8.56%	1.79%	17.23%	5.85%	10.60%	0.00%	8.41%	0.88%	10.95%	12.58%	12.86%
Coffee beans	14.13%	26.79%	0.98%	31.17%	3.13%	9.42%	55.51%	0.90%	8.93%	12.55%	12.62%
Other crops	71.26%	67.86%	81.80%	49.41%	86.09%	89.74%	28.83%	97.33%	77.09%	49.61%	61.90%
CTL	**383**	**18**	**95**	**8**	**16**	**73**	**53**	**38**	**24**	**35**	**24**
Cattle for meat	47.29%	52.15%	21.10%	44.42%	23.84%	89.86%	75.36%	43.02%	19.02%	20.06%	46.77%
Other ruminant for meat	52.71%	47.85%	78.90%	55.58%	76.16%	10.14%	24.64%	56.98%	80.98%	79.94%	53.23%
OAP	**664**	**6**	**283**	**48**	**21**	**93**	**62**	**72**	**25**	**17**	**37**
Pigs	23.18%	10.71%	23.80%	24.25%	8.61%	27.21%	18.29%	38.26%	3.29%	10.92%	14.94%
Poultry birds and eggs	70.28%	75.32%	71.90%	72.22%	89.36%	65.63%	76.60%	47.23%	87.07%	70.50%	78.48%
Other animal products	6.54%	13.97%	4.31%	3.53%	2.03%	7.16%	5.11%	14.51%	9.64%	18.58%	6.58%
VOL	**359**	**2**	**135**	**56**	**13**	**33**	**59**	**46**	**9**	**2**	**4**
Olive oil	5.53%	5.95%	0.00%	0.00%	0.08%	0.00%	0.45%	32.93%	46.71%	0.00%	0.27%
Palm oil	18.14%	22.62%	0.56%	94.29%	0.53%	0.42%	10.39%	6.38%	4.61%	77.67%	2.39%
Soybean cake	35.80%	0.00%	48.04%	3.41%	31.02%	47.42%	51.65%	17.46%	16.23%	10.19%	43.35%
Soybean oil	21.79%	0.60%	27.25%	2.19%	18.11%	28.01%	36.71%	8.46%	18.20%	9.71%	27.39%
Rape seed cake	4.52%	17.86%	5.80%	0.00%	14.44%	5.03%	0.17%	8.33%	3.73%	0.49%	6.38%
Rape seed oil	14.22%	52.98%	18.35%	0.11%	35.83%	19.11%	0.62%	26.44%	10.53%	1.94%	20.21%

(Continued)

Table 7. Comparison of output share across world regions (Continued).

	World	Australia, New Zealand	East Asia	Southeast Asia	South Asia	North America	Latin America	European Union 25	Middle East and North Africa	Sub-Saharan Africa	Rest of World
CMT	**709**	**23**	**150**	**11**	**12**	**158**	**85**	**97**	**26**	**59**	**89**
Cattle meat	84.28%	79.75%	76.72%	74.57%	88.04%	91.90%	94.80%	85.65%	73.71%	66.15%	88.73%
Other ruminant meat	15.72%	20.25%	23.28%	25.43%	11.96%	8.10%	5.20%	14.35%	26.29%	33.85%	11.27%
OMT	**614**	**4**	**130**	**36**	**3**	**129**	**74**	**178**	**17**	**18**	**25**
Pig meat	50.35%	30.18%	70.00%	50.37%	15.18%	41.04%	27.48%	59.98%	3.63%	25.07%	52.58%
Other meat	49.65%	69.82%	30.00%	49.63%	84.82%	58.96%	72.52%	40.02%	96.37%	74.93%	47.42%
OFD	**2,932**	**50**	**815**	**172**	**95**	**557**	**228**	**639**	**164**	**85**	**129**
Feed concentrates	17.65%	5.03%	30.27%	18.58%	6.69%	11.52%	17.71%	13.45%	8.70%	9.46%	13.53%
Other food processing	82.35%	94.97%	69.73%	81.42%	93.31%	88.48%	82.29%	86.55%	91.30%	90.54%	86.47%

Source: Author

4. An example application

We compare under a full multilateral trade liberalization scenario the results for two databases, both using an aggregation to ten world regions. The first data set corresponds to the GTAP Data Base V10 at full sectoral resolution with its 65 products from which we depart, the second is the derived GTAP-AGROFOOD data set with its 105 products. We use the GTAP Standard model as encoded in CGEBox under default settings, i.e. labour and capital are fully mobile while land is sluggish with a substitution elasticity of unity; trade balances are flexible based on the global bank approach; no separate accounts for private household and government under the regional household approach; international trade is depicted by a two-stage Armington approach. The parametrisation is based on the values found in the GTAP Data Base V10, with parameters of more detailed products in GTAP-AGROFOOD being equal to the ones of the related product in the GTAP Data Base V10. Armington elasticities are an exemption, they are in both cases aggregated using trade weights from HS6 level using the recent estimates by Fontagné et al., 2019. No additional nestings in the production function or for factor or final demand are used.

Obviously, all the results provided by the GTAP standard model (output, intermediate and final demand, factor use, bi-lateral trade and related prices etc.) become available at the more detailed product and sector level when the more disaggregated database is used. But besides this additional detail, we also find relevant differences for economy-wide global effects. The money metric utility per capita, for instance, decreases globally from around 22.5 to 21.7 USD per capita when using the more detailed GTAP-AGROFOOD database. This might come as a surprise given the relatively low weights of the agro-food sector in the global economy. Differences at regional level can deviate even more as shown in Table 8 below. These changes reflect on the hand the effect of tariff dispersion across products for the original GTAP sectors; similar results might be expected when using models working at the tariff line such as GTAP-HS (Chepeliev et al., 2021) for agro-food sectors. Higher tariff dispersion when using a more detailed database will increase the impacts of trade liberalisation, this was also found by Britz and van der Mensbrugghe (2016) when increasing the regional detail of the GTAP Data Base. However, trade relations in the GTAP-AGROFOOD database are likely more specialized which dampens the effect of trade liberalization, related to what Chepeliev et al. (2021) call the "false competition" effect from aggregation. For instance, major exporters and importers of minor coarse grains differ considerably from the average of all other grains; soybean cakes are imported mostly by regions with intensive livestock production, while soy oil is most imported for cooking purposes into Asian countries. Horridge (2018) suggests also aggregation of substitution elasticities using benchmark values as weights as a potential source of aggregation bias if cost shares differ. This might be relevant here for the Armington elasticities where larger differences in bilateral trade shares for disaggregated products belonging to the same aggregate are not uncommon.

Table 8. Selected outcomes under the same full trade liberalization shock and the same regional aggregation, full sectoral detail, GTAP-AGROFOOD compared to GTAP V10 Data Base

		Real GDP [delta %]	Money metric utility per capita [USD per head]	Foreign savings [delta m.USD]	Calories [delta %]
World	GTAP	0.26%	22.48		1.03%
	AGROFOOD	0.23%	21.70		0.80%
Australia, New Zealand	GTAP	0.10%	87.21	2.10	0.29%
	AGRFOOD	0.11%	99.06	2.55	0.25%
East Asia	GTAP	0.74%	91.00	58.48	2.44%
	AGRFOOD	0.67%	86.79	50.44	0.00%
Southeast Asia	GTAP	0.30%	6.66	2.02	0.38%
	AGRFOOD	0.31%	5.63	1.84	0.12%
South Asia	GTAP	0.43%	1.01	12.41	0.77%
	AGRFOOD	0.41%	2.66	11.03	0.58%
North America	GTAP	0.03%	-4.96	-55.09	0.31%
	AGRFOOD	0.03%	-1.31	-49.40	0.24%
Latin America	GTAP	0.06%	-20.31	3.58	-0.18%
	AGRFOOD	0.06%	-20.97	3.42	-0.21%
European Union 28	GTAP	0.15%	24.29	-49.19	1.18%
	AGRFOOD	0.12%	19.08	-46.50	1.00%
Middle East and North Africa	GTAP	0.30%	19.89	10.51	0.91%
	AGRFOOD	0.28%	19.44	10.49	0.82%
Sub-Saharan Africa	GTAP	0.39%	-1.64	11.20	0.30%
	AGRFOOD	0.38%	-1.96	11.46	0.27%
Rest of World	GTAP	-0.10%	-0.08	3.99	0.77%
	AGRFOOD	-0.10%	0.55	4.67	0.79%

Source: Author

A look at the prices changes for product aggregates (see Table 9) reveals that the more detailed model simulates considerably larger price drops for agro-food sectors, but shows, as expected, almost no differences for non-agro-food aggregates. Interesting to note are also far larger price effects for paddy rice, wheat and processed rice (not shown in the table), three sectors which are not subject to disaggregation in the GTAP-AGROFOOD database.

Table 9. Output changes under the same full trade liberalization shock and the same regional aggregation, full sectoral detail, GTAP-AGROFOOD compared to GTAP V10 Data Base

		World	Austr. New Zealand	East Asia	South-East Asia	South Asia	North America	Latin America	European Union 28	Middle East and North Africa	Sub-Saharan Africa	Rest of World
ALL sectors	GTAP	0.05%	0.07%	-0.01%	0.17%	0.15%	0.15%	-0.05%	0.07%	0.15%	-0.52%	-0.11%
	AGROFOOD	0.04%	0.05%	0.01%	0.05%	0.03%	0.14%	-0.05%	0.04%	0.13%	-0.50%	-0.11%
Grains and Crops	GTAP	-2.23%	-7.01%	-4.41%	-3.84%	-0.04%	-2.76%	-0.47%	-2.03%	-3.82%	-0.62%	2.35%
	AGROFOOD	-2.25%	-2.75%	-5.26%	-3.21%	-0.74%	-2.65%	-1.46%	-1.48%	-2.05%	-0.53%	4.65%
Livestock and Meat Products	GTAP	0.49%	20.08%	-1.39%	46.07%	0.47%	4.53%	10.37%	-17.90%	-1.36%	-2.04%	-4.47%
	AGROFOOD	0.27%	22.13%	-2.54%	27.26%	0.35%	4.89%	9.68%	-13.54%	-1.32%	-1.63%	-4.21%
Mining and Extraction	GTAP	0.13%	-0.60%	-1.46%	-0.50%	-0.84%	0.17%	1.67%	0.86%	0.56%	1.40%	0.57%
	AGROFOOD	0.13%	-0.84%	-1.37%	-0.34%	-0.93%	0.16%	1.76%	0.76%	0.53%	1.41%	0.55%
Processed Food	GTAP	0.31%	5.91%	2.89%	6.72%	-9.19%	1.88%	1.20%	-1.47%	-3.45%	-6.47%	-5.33%
	AGROFOOD	0.08%	6.43%	1.25%	4.28%	-7.36%	2.27%	1.42%	-1.20%	-2.88%	-5.54%	-4.47%
Textiles and Clothing	GTAP	1.75%	-4.47%	8.56%	5.02%	2.80%	-7.54%	-8.20%	-5.78%	-2.50%	-17.81%	-8.05%
	AGROFOOD	1.77%	-4.80%	8.76%	5.74%	1.74%	-7.53%	-8.12%	-5.80%	-2.66%	-17.86%	-8.11%
Light Manufacturing	GTAP	0.02%	-5.81%	-0.15%	-1.48%	2.02%	0.98%	-4.04%	1.67%	-1.17%	-8.29%	-7.43%
	AGROFOOD	0.02%	-6.41%	0.18%	-0.86%	1.63%	0.86%	-3.90%	1.39%	-1.48%	-8.51%	-7.73%
Heavy Manufacturing	GTAP	-0.14%	-2.88%	-1.37%	-3.80%	0.95%	0.92%	-1.68%	1.69%	1.50%	2.69%	3.30%
	AGROFOOD	-0.12%	-3.88%	-1.09%	-2.88%	0.47%	0.72%	-1.47%	1.36%	1.18%	2.29%	3.05%
Utilities and Construction	GTAP	0.23%	0.40%	1.17%	-0.01%	1.26%	-0.69%	0.67%	-0.69%	0.77%	3.82%	-0.28%
	AGROFOOD	0.23%	0.45%	1.08%	0.07%	1.18%	-0.61%	0.68%	-0.66%	0.76%	3.83%	-0.28%
Transport and Communication	GTAP	0.28%	-0.01%	0.46%	0.44%	0.19%	0.08%	0.21%	0.30%	0.79%	0.00%	0.14%
	AGROFOOD	0.26%	-0.01%	0.42%	0.53%	0.19%	0.08%	0.20%	0.27%	0.72%	-0.02%	0.11%
Other Services	GTAP	-0.06%	-0.11%	0.04%	-0.59%	-0.47%	-0.08%	-0.05%	-0.03%	-0.02%	0.14%	-0.12%
	AGROFOOD	-0.06%	-0.13%	0.05%	-0.42%	-0.33%	-0.07%	-0.03%	-0.05%	-0.06%	0.11%	-0.16%

Source: Author

The reasons for the price (equal cost) differences are mainly considerably different changes in land rents. At the global level (Table 10), the more detailed database predicts an increase of 3% in the average land rent, opposed to a drop of close to 5% when using the original GTAP Data Base. These differences in results seem sufficient to underline that more details can matter, here the likely main reason being differences in ad-valorem equivalents of trade policy instruments.

Table 10. Changes in land rents, GTAP-AGROFOOD compared to GTAP V10 Data Base

	GTAP %	AGROFOD %	Difference [percentage points]
World	-4.82%	3.09%	-7.91%
Australia, New Zealand	25.12%	24.43%	-0.55%
East Asia	-14.98%	1.94%	19.90%
Southeast Asia	5.51%	9.77%	4.04%
South Asia	-2.76%	8.97%	12.07%
North America	-6.14%	-3.29%	3.04%
Latin America	11.98%	12.70%	0.64%
European Union 28	-17.01%	-12.68%	5.22%
Middle East and North Africa	-14.01%	-7.06%	8.08%
Sub-Saharan Africa	-4.29%	-5.04%	-0.78%
Rest of World	5.81%	1.92%	-3.68%

Source: Author

5. Discussion

As discussed above, cost shares for capital, labour categories and non-agrofood intermediate use are based on proportionality assumptions. These neglect possible differences in cost shares across detailed activities which can be relevant in specific applications. A higher economic return to land for a specific crop compared to the average of the related aggregate one, which is equivalent to overall higher costs per ha, implies under proportionality assumptions that all its intermediate demands as well capital and labour use per unit of output are scaled upwards by the same factor compared to what the GTAP Data Base reports for the related aggregate product. But depending on the production activity, higher cost per ha might stem mostly from changes in specific inputs, for instance, from the need of irrigation, from high seed costs, from the necessity to harvest manually, to use heated glass houses, or to grow and maintain trees in the case of perennials.

The AGROSAM data base (Mainar-Causapé et al., 2018) which covers the EU with high detail therefore derives detailed intermediate costs shares from primary agricultural production activities from the CAPRI data base (Britz and Witzke, 2014), which in turn draws on data from the Farm Accountancy Data Network and other sources to allocate costs reported in Economic Accounts for Agriculture to production activities. National statistical offices might also provide detailed information, for instance, via supply and use tables. Statistics Canada, to give an example, reports "Greenhouse, nursery and floriculture production (except cannabis)" as a separate activity in its supply-and-use table (SUT)[13] while the US Bureau of Economic Analysis offers SUTs,[14] which include, for instance, "Fruit and tree nut farming", "Vegetable and melon farming" and "Greenhouse, nursery, and floriculture production" as separate activities.

But cost shares, for instance, for irrigation or heating of glasshouses, might depend on climatic conditions, others, such as for labour and capital, reflect wages and prevalent technologies, while costs for plant protection or fertilization, for instance, might depend on national agri-environmental standards or the share of organic production. It is therefore challenging to use cost share observations from single countries as a basis to disaggregate a global database. But users are free to improve the database for specific studies by extending

[13] https://www150.statcan.gc.ca/n1/pub/15-602-x/15-602-x2017001-eng.htm
[14] https://www.bea.gov/industry/input-output-accounts-data

the code defining the split factors[15] by adding their own data or by using tools such as SplitCom or TPMH0104 in GEMPACK.[16]

Another point of caution relates to the parameterization of models drawing on GTAP-AGROFOOD. Behavioural and other parameters (expansion and substitution elasticities for the CDE demand system, substitution elasticities between primary factors) are borrowed from the aggregate products respectively activities. For Armington elasticities, only, we can draw on the recent detailed estimates by Fontagné et al. (2019) at the HS6 level.

More generally, higher product detail might require a partial re-structuring of the CGE model. Take the input coefficients for other food processing (ofd), vegetables cakes and oils (vol) and dairy production (mil) into animal production activities (ctl, oap, rmk, wol) as an example. They capture jointly a large share of feed costs, especially of concentrates. As multiple cakes, oils and dairy products are present in the GTAP-AGROFOOD database, it might be necessary to allow for substitution of products in same category. This avoids unrealistic rigidity in the technology and potentially implausible price developments of close substitutes; the same might hold for other cases, such as for multiple grains. Similarly, once multiple meats, vegetables and fruits or dairy products are used in a model, separate CES nests under the final demand system might be necessary to depict cross-price effects in an appropriate manner. The G-RDEM model (Britz and Roson, 2018) comprised as a module in CGEBox adds automatically such nests if product detail is found in the database used in a simulation, see also the paper by Ho et al. (2020) on final demand in construction of long-term baselines. Equally, factor mobility assumptions might need to be adjusted, for instance, to reflect that the land allocation between annual crops is more flexible compared to perennials. Whether such changes are necessary depends on whether the related model mechanism is important for the study at hand.

An alternative to the full disaggregation of a CGE model based on the GTAP-AGROFOOD database is the GTAP-HS model (Chepeliev et al. 2021) which disaggregates the trade components of the GTAP Standard model to the tariff line, and adds as Constant-Elasticity-of-Transformation functions to distribute production output to multiple tariff lines. Even less data demanding is the approach by Jafari et al. (2021) which disaggregates solely trade on selected bilateral links to multiple tariff lines. Both approaches require less data compared to what is discussed here, but cannot assess cases where (global) value chains involving more detailed agro-food products are the focus. This motivated similar split exercises as underlying the GTAP-AGROFOOD database for some selected agro-food products, only, in Escobar et al. (2018) and Escobar and Britz (2021) for bio-plastics or Nong et al. (2020) for bio-chemicals.

Some further words of caution might be needed. Both the data from GTAP and FABIO data refer to single years, 2014 and 2013, respectively. Stochastic elements, such as weather, international conflicts, pests or other disasters can lead to larger inter-annual changes in outputs, use, trade and related prices. The higher the regional, activity and product detail, the more likely it is that the observed data for a specific year are not representative for larger time periods which also challenges benchmarking. For instance, ex-post observed cost shares might not have been the planned ones. For future updates, if possible, multi-year averages should be used to derive split factors. This was not possible here as the version of the FABIO database has 2013 as the most recent year.

[15] gams/build/split/fabio_agr.gms, gams/build/split/split_dairy.gms

[16] https://www.copsmodels.com/splitcom.htm and https://www.copsmodels.com/archivep.htm

5. Summary and conclusions

This document reports some key aspects of developing the GTAP-AGROFOOD database for CGE modelling, which adds 51 additional agro-food products and sectors to the GTAP Data Base Version 10, maintaining its full regional and factor detail. This more detailed database is available as supplementary material. It is organized as a zip container with the data in the same structure, definitions and units as the original GTAP Data Base. The data are comprised in GDX files from which they can be also easily converted to GEMPACK har files, for instance with ViewHAR.

Data on bilateral trade and related tariff revenues are derived from the TASTE V10 database. The disaggregation of production, land use and intra-agro-food intermediate use (seed and feed use, input of young animals in livestock processes) draws on the MRIO database FABIO (Bruckner et al., 2019), detail for dairy stems from market balances by FAOSTAT. Remaining data are based on proportionality assumption. The resulting a priori data are balanced such as to maintain consistency with the GTAP Data Base. This is based on a linear-loss framework (Britz, 2021) as part of the data driver of CGEBox, a platform for CGE modelling in GAMS (Britz and Van der Mensbrugghe, 2018). The process involves removing cost, use and trade shares <1.E-3% from the GTAP Data Base before the disaggregation step, in order to reduce numerical problems with small shares, while maintaining macro totals. This implies some quite limited differences to the GTAP Data Base when re-aggregating to the original product resolution. It also leads to non-dense matrices, for instance for bilateral trade and intermediate demand by firms. In parallel, some corrections are introduced, such as deleting any tax rates below 1.E-6. Shares on land use at country level are used to split-up the AEZ database to the additional details for crop production, while uniform emission factors for CO2, non-CO2 and air pollution emissions across disaggregated products and activities deliver estimated detailed emission inventories such that user can also use these auxiliary data sets and related model equations. All data used besides the data bases from GTAP are open-source and open-access such that only a GTAP license is necessary to use the GTAP-AGROFOOD database.

Users with a specific regional or sector focus might improve the database further by introducing more detailed split factors to reflect further cost share differences. Equally, changes to the default model parametrization should be considered where parameter values for detailed products are currently equal to the related aggregate value. Users will also need to reflect if changes to their model layout are warranted, such as with regard to the nestings in production and demand function or to the nestings in factor markets, for instance, to address the fact that many of the highly detailed agro-food products are close substitutes.

Acknowledgements

The author would like to thank two anonymous reviewers for their detailed and helpful comments which improved the paper and the underlying methodology. This research was supported by the German Federal Ministry of Education and Research (funding code 031B0792A).

References

Aguiar, A., Chepeliev, M., Corong, E., McDougall, R., & van der Mensbrugghe, D. (2019). "The GTAP Data Base: Version 10". *Journal of Global Economic Analysis, 4(1)*, 1-27

Britz W. (2021): "Comparing penalty functions in balancing and disaggregating Social Accounting Matrices", *Journal of Global Economic Analysis*, Analysis 6(1): 34-81

Britz, W., & Witzke, P. (2014). "CAPRI model documentation 2012". *Institute for Food and Resource Economics, University of Bonn, Bonn*, https://www.capri-model.org/docs/capri_documentation.pdf

Britz, W., van der Mensbrugghe, D. (2016): "Reducing unwanted consequences of aggregation in large-scale economic models - A systematic empirical evaluation with the GTAP model", *Economic Modelling* 59: 462-473

Britz, W., van der Mensbrugghe, D. (2018): "CGEBox: A Flexible, Modular and Extendable Framework for CGE Analysis in GAMS", *Journal of Global Economic Analysis* 3(2): 106-176

Chepeliev, M., Golub, A., Hertel, T., Saeed, W., & Beckman, J. (2021). Disaggregating the Vegetables, Fruits and Nuts Sector to the Tariff Line in the GTAP-HS Framework, *Journal of Global Economic Analysis* 6(1): 82-127

Bruckner, M., Wood, R., Moran, D., Kuschnig, N., Wieland, H., Maus, V., & Börner, J. (2019). "FABIO—the construction of the food and agriculture biomass input–output model". *Environmental Science & Technology*, *53*(19), 11302-11312

Escobar, N., Haddad, S., Börner, J., Britz, W. (2018): "Land use mediated GHG emissions and spillovers from increased consumption of bioplastic", *Environmental Research Letters* 13(12)

Escobar, N., Britz, W. (2021): "Metrics on the sustainability of region-specific bioplastics production, considering global land use change effects", *Resources, Conservation and Recycling* 167(April 2021): 105345

Fontagné, L., Guimbard, H., & Orefice, G. (2019). Product-Level Trade Elasticities. CEPII Working Paper No 2019-17. http://www.cepii.fr/PDF_PUB/wp/2019/wp2019-17.pdf

Ho, M., Britz, W., Delzeit, R., Leblanc, F., Roson, R., Schuenemann, F., Weitzel, M. (2020): Modelling Consumption and Constructing Long-Term Baselines in Final Demand, *Journal of Global Economic Analysis* 5(1): 63-80

Horrigde, M. (2018). Aggregating CES elasticities for CGE models, paper presented the 21st Annual Conference on Global Economic Analysis, Cartagena, Colombia. https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=5491

Jafari, Y., Himics, M., Britz, W., Beckman, J. (2021): "It is all in the details: A bilateral approach for modelling trade agreements at the tariff line", Canadian Journal of Agricultural Economics 1-28

Krey, V., Havlik, P., Kishimoto, P., Fricko, O., Zilliacus, J., Gidden, M., ... & Riahi, K. (2020). "MESSAGEix-GLOBIOM Documentation-2020 release".

Mainar-Causapé, A., Philippidis, G., & Caivano, A. (2018). BioSAMs for the EU Member States: Constructing Social Accounting Matrices with a detailed disaggregation of the bio-economy

Nechifor, V., Ramos, M. P., Ferrari, E., Laichena, J., Kihiu, E., Omanyo, D., ... & Kiriga, B. (2021). "Food security and welfare changes under COVID-19 in Sub-Saharan Africa: Impacts and responses in Kenya". *Global food security*, *28*, 100514

Nong, D., Escobar, N., Britz, W., Börner, J. (2020). "Long-term impacts of bio-based innovation in the chemical sector: a dynamic global perspective". *Journal of Cleaner Production* 272(November 2020): 122738

OECD, F. (2015). "Aglink-Cosimo Model Documentation. A partial equilibrium model of world agricultural markets".

Pelikan, J., Horridge, M., & Mustakinov, D. (2020). "TASTE for GTAP 10: A Tariff Analytical and Simulation Tool for Economists". https://www.gtap.agecon.purdue.edu/resources/download/10140.pdf

Philippidis, G., Bartelings, H., Helming, J., M'barek, R., Ronzon, T., Smeets, E., ... & Shutes, L. (2018). "The MAGNET model framework for assessing policy coherence and SDGs". https://core.ac.uk/download/pdf/157830375.pdf

van Meijl, H., Tsiropoulos, I., Bartelings, H., Hoefnagels, R., Smeets, E., Tabeau, A., & Faaij, A. (2018). "On the macro-economic impact of bioenergy and biochemicals–Introducing advanced bioeconomy sectors into an economic modelling framework with a case study for the Netherlands." *Biomass and Bioenergy*, 108, 381-397

Appendix

Table A1. Newly introduced products and their link to the GTAP Data Base

	Label	Description	Related GTAP product	
1	sorg	Sorghum	gro	Other Grains: maize (corn), sorghum, barley, rye, oats, millets, other cereals
2	barl	Barley		
3	maiz	Maiz		
4	oat	Oats		
5	rye	Rye		
6	ocer	Other cereals		
7	soy	Soy bean	osd	Oil Seeds: oil seeds and oleaginous fruit
8	palm	Palm oil fruit		
9	rape	Rape seed		
10	olv	Olives		
11	oso	Other oilseeds		
12	pota	Potatoes	v_f	Veg & Fruit: vegetables, fruit and nuts, edible roots and tubers, pulses
13	rttb	Other roots and tubers		
14	leg	Leguminosae		
15	toma	Tomatoes		
16	oveg	Other vegetables		
17	citr	Citrus fruits		
18	appl	Apples		
19	grap	Grapes		
20	banp	Bananas and plantains		
21	fruit	Other fruits		
22	v_fo	Rest of v_f		
23	coco	Cocoa beans	ocr	Other Crops: stimulant; spice and aromatic crops; forage products; plants and parts of plants used primarily in perfumery, pharmacy, or for insecticidal, fungicidal or similar purposes; beet seeds (excluding sugar beet seeds) and seeds of forage plants; natural rubber in primary forms or in plates, sheets or strip, living plants; cut flowers and flower buds; flower seeds, unmanufactured tobacco; other raw vegetable materials nec
24	teas	Teas		
25	coff	Coffee beans		
26	ocro	Rest of ocr		
27	olivOil	Olive oil	vol	Vegetable Oils: margarine and similar preparations; cotton linters; oil-cake and other residues resulting from the extraction of vegetable fats or oils; flours and meals of oil seeds or oleaginous fruits, except those of mustard; vegetable waxes, except triglycerides; degras; residues resulting from the treatment of fatty substances or animal or vegetable waxes; animal fats
28	palmOil	Palm oil		
29	soyCake	Soybean cake		
30	soyOil	Soybean oil		
31	rapCake	Rape seed cake		
32	rapOil	Rape seed oil		
33	volo	Other vegetable oils and cakes		
34	cattle	Cattle for meat	ctl	Cattle: bovine animals, live, other ruminants, horses and other equines, bovine semen
35	orum	Other ruminant for meat		
36	pig	Pigs	oap	Other Animal Products: swine; poultry; other live animals; eggs of hens or other birds in shell, fresh; reproductive materials of animals; natural honey; snails, fresh, chilled, frozen, dried, salted or in brine, except sea snails; edible products of animal origin n.e.c.; hides, skins and furskins, raw; insect waxes and spermaceti, whether or not refined or coloured
37	poul	Poultry birds and eggs		
38	oapo	Other animal products		

	Label	Description	Related GTAP product	
39	ctlMeat	Cattle meat	cmt	Cattle Meat: fresh or chilled; meat of buffalo, fresh or chilled; meat of sheep, fresh or chilled; meat of goat, fresh or chilled; meat of camels and camelids, fresh or chilled; meat of horses and other equines, fresh or chilled; other meat of mammals, fresh or chilled; meat of mammals, frozen; edible offal of mammals, fresh, chilled or frozen
40	orumMeat	Other ruminant meat		
41	pigMeat	Pig meat	omt	Other Meat: meat of pigs, fresh or chilled; meat of rabbits and hares, fresh or chilled; meat of poultry, fresh or chilled; meat of poultry, frozen; edible offal of poultry, fresh, chilled or frozen; other meat and edible offal, fresh, chilled or frozen; preserves and preparations of meat, meat offal or blood; flours, meals and pellets of meat or meat offal, inedible; greaves
42	othMeat	Other meat		
43	butt	Butter	mil	Milk: dairy products
44	ches	Cheese		
45	smp	Skimmed milk powder		
46	wmp	Whole milk powder		
47	whey	Whey		
48	case	Casein		
49	milo	Other dairy		
50	ofdAnim	Feed concentrates	ofd	Other food processing
51	ofdOther	Other food processing		

Note: The 51 new products refer to 11 original ones which are dropped from the database. Description for GTAP sectors is from the GTAP web site (www.gtap.org).

Appendix A1: Filtering

The first step in the process is to apply the filter routine. This implies some smaller deviations from the original GTAP Data Base. Filtering is necessary as the linear loss problems applied after filtering comprises already around 3.5 million constraints and 6 million variables. As expected, these problems are quite badly scaled, i.e. they comprise a mix of very small and large coefficients in the constraint matrix and objective, often in the very same (in)equality. Without filtering, many more small coefficients would be introduced in the estimation framework, challenging its numerical stability further. The filtering process, using a threshold of 0.001%[17], reduces the number of non-zero data items in the GTAP Data Base from around 3.6m to 2.8m, or by around 20%. This reflects that the original database comprises a large amount of quite tiny cost, expenditure and trade shares, frequently relating to transactions in the range of 1.E-6m USD or below, i.e. less than 1 USD. The filter program (see Britz and Van der Mensbrugghe 2016 for details) maintains important macro totals. This implies for the example here that global GDP is almost not corrected at all and that the value of world trade is kept constant. The changes related to small items in the original GTAP Data Base should have no relevance for any analysis. Macro-total such as GDP can be maintained only during filtering if not-deleted transactions are upward corrected. Accordingly, some limited differences between the original and the filtered version can be expected also for not-deleted items.

Table A2 reports some key outcomes of the filtering process. It underlines that the changes from filtering in GDP, total factor demand (*vfm*), production value (*vom*) and global trade (*vtwr*) as well as for many other totals are all <0.005%, in the table reported as zeros. Further penalties in the objective function of the filter step reflect deviations for total spent,

[17] The threshold is applied to related totals, for instance, for bilateral imports flows for a product and region, the total imports of this product and this region are used for comparison.

i.e. the expenditures for import and domestic demand by private households, government, investment and all firms for each region. Combined with the hard restriction that the value of global trade and import tax revenues of each country has to be maintained, this leads to some quite limited shifts between import and domestic demand for the different Armington agents.

Table A2. Changes in value of transactions in filtered compared to unfiltered GTAP Data Base

Items	Unfiltered [USD m]	Filtered [USD m]	Difference [USD m]	Difference [%]
GDP	74,483,137	74,483,074	-64	0
vfm	63,095,678	63,096,294	616	0
vom	155,517,294	155,518,923	1.630	0
vxmd	20,449,665	20,449,665	0	0
vtwr	620,260	620,260	1	0
vifm	14,543,991	14,543,361	-630	0
vdfm	67,173,397	67,174,101	704	0
vipm	4,311,123	4,311,605	482	0.01
vdpm	38,225,060	38,224,436	-624	0
vigm	252,566	252,696	130	0.05
vdgm	12,956,459	12,956,403	-56	0
viim	2,643,855	2,643,876	21	0
vdim	16,094,074	16,094,057	-16	0

Source: Author

Table A3 reports the number of small transactions removed during filtering. In relative terms, the largest amount of tiny use shares is found in government consumption (*vigm, vdgm*) where around 60% of the non-zero (including the 1.E-10) cases are removed. Even in this case, there are no differences in the resulting totals. For government consumption, the average absolute correction was 0.03m USD for the imported demand (*vigm*) and 0.33m USD for the domestic demand (*vdgm*). This underlines again that filtering should not have a discernible impact on simulations results while reducing model size and increasing numerical stability.

The split process documented above can also be applied to databases with some user defined regional, product and factor aggregation, instead of disaggregating the 1:1 version of the original GTAP Data Base which results in the database described here. The split process had been tested with different split problem and differently detailed database variants (see also Britz 2021). For more aggregated and thus less detailed input data, less aggressive filtering thresholds can be used such that differences between the filtered and unfiltered (aggregated) version of the GTAP Data Base become even smaller. If it is deemed important, the data driver of CGEBox gives the user the option to also refrain from the corrections and cost share safeguards described above. However, the resulting input data to disaggregate with potentially quite tiny cost/use/trade shares might provoke feasibility problems in the subsequent split step. Sparing the user the need to familiarize herself in detail with the

CGEBox data driver and to deal with potential feasibility problems is a key reason why instead a version with full product, factor and region detail of the original GTAP Data Base is released. This also gives users of the GTAP Data Base without the GAMS license required for CGEBox access to the data.

Table A3. Changes in non-zero transactions in filtered compared to unfiltered GTAP Data Base

Items	Unfiltered [# of non-zeros]	Filtered [# of non-zeros]	Difference [# of non-zeros]	Difference [%]
All	3,571,533	2,827,366	-744,167	-20.84
vfm	57,415	54,784	-2,631	-4.58
vom	9,165	8,783	-382	-4.17
vxmd	1,264,576	1,042,809	-221,767	-17.54
vtwr	1,011,171	990,492	-20,679	-2.05
vifm	588,784	321,133	-267,651	-45.46
vdfm	588,816	379,030	-209,786	-35.63
vipm	9,165	7,589	-1,576	-17.20
vdpm	9,165	7,984	-1,181	-12.89
vigm	8,319	3,302	-5,017	-60.31
vdgm	8,319	3,062	-5,257	-63.19
viim	8,319	3,820	-4,499	-54.08
vdim	8,319	4,578	-3,741	-44.97

Source: Author