Institute for Food and Resource Economics

University of Bonn

Discussion Paper 2020:2

Maintaining plausible calorie intakes, crop yields and crop land expansion in long-run simulations with Computable General Equilibrium Models

Wolfgang Britz

Editor: Thomas Heckelei

Institute for Food and Resource Economics

University of Bonn

Nußallee 21

53115 Bonn, Germany

Phone: +49-228-732332

Fax: +49-228-734693

E-mail: thomas.heckelei@ilr.uni-bonn.de

Maintaining plausible calorie intakes, crop yields and crop land expansion in long-run simulations with Computable General Equilibrium Models

Wolfgang Britz

Abstract

We demonstrate how a combination of different elements can jointly provide plausible long-term trends for calorie intakes, crop yields and land use in Computable General Equilibrium (CGE) analysis. Specifically, we depict household demand based on a MAIDADS demand system estimated based on cross-sectional data. In order to control for calorie intake we first regress calorie intake on per capita income and construct a Leontief inverse to derive implicit calorie intakes from the final consumption of processed food. This allows jointly shifting preferences of the MAIDADS system by updating commitment terms and marginal budget shares, to arrive at plausible per capita calorie intakes during baseline construction. We control yields based on exogenous projections which we also use to parameterize our land supply functions. The contribution of the different elements is evaluated by comprising key developments in baselines up to 2050 constructed with different model variants

Keywords: long run baselines, food demand, calories, Computable General Equilibrium analysis

JEL classification: D12, C33, C68

1 Background

With still around 800 Mio people undernourished today, daily calorie intake will remain for the common decades a key indicator of human well-being (FAO et al. 2019). Abating hunger by provision of food is closely linked to future crop productivity changes and crop-land expansion (FAO 2018), the later a driver of natural eco-system loss and related carbon stock changes as other key challenges for a sustainable development. But improved access to food for the poor is at least as strongly linked to their future purchasing power which depends on macro-economic developments and related structural changes processes. Understanding and projecting likely developments in this nexus requires hence a framework extending beyond the agri-food and land use sectors. This renders it inviting to apply global dynamic CGE (Computable General Equilibrium) models as they combine sectoral detail including bi-lateral trade with a consistent description of income generation and use in the overall economy.

But depicting food supply and demand and its link to land use has proven challenging in long-run simulation with CGE models for a number of reasons. First, as underlined by Ho et al. 2020, most global CGE (Computable General Equilibrium) models used for long-run analysis do not incorporate directly empirically estimated demand systems with flexible Engel curves, but rely rather on more ad-hoc updates of parameters of the usual functional forms used in CGE models, such as of the Constant-Elasticity of Transformation demand system, the Linear Expenditure System or the Constant Difference in Elasticities (Hanoch 1975) type. Second, even a reasonable development of Engel curves for food in total or even sub-categories thereof, accomplished either by estimated rank-3 demand system or parameter updates of a simper one, does not guarantee plausible shifts in physical demands for calories, primary agricultural products and related to this, land use. As shown in the following, even the G-RDEM model (Britz and Roson 2019) as a CGE specifically designed for long-run analysis combined with a land use component where, for instance, yield shifts and crop land expansion reflect projections from a specialized study by FAO 2018 fails to provide defendable calorie intake estimates at the level of larger regional aggregates when simulating over decades into the futures. This reflects that calorie intakes are quite sensitive both to the parameterization of the demand system and the endogenous price developments simulated during baseline construction. A further challenge is to control simultaneously developments for the whole bundle of food products which jointly determine calorie intakes.

In the following, we discuss two elements which we consider as key elements to jointly contribute to more plausible and controlled outcomes with regard to food consumption and land use: integrating and empirically estimated demand system with detail for food and controlling explicitly for total calorie intake during baseline construction. These elements are integrated in an updated version of the G-RDEM model (Britz and Roson 2019). We evaluate in here their contribution with regard to outcomes in the focus, i.e. calorie intakes. Doing so requires to link calorie intake to the food categories depicted in the GTAP data base for which we propose a novel approach.

This paper is organized as follows. We first discuss the methodology comprising of four elements: (1) the integration of an empirically estimated MAIADS demand, (2) the estimation of the relation between calorie intake and income levels, (3) establishing the link calorie intake to final demand for processed food commodities in the GTTAP SAM and (4) controlling for calorie intakes during benchmarking and simulations as discussed next. This provides the basis for an evaluation based on comparing outcomes in a baseline construction exercise up to the year 2050 before we conclude and summarize.

2 Developing a demand system for a CGE which controls for calorie intakes

2.1 *Integration of an empirically estimated MAIDADS demand system*

Roson and Van der Mensbrugghe 2018, Britz and Roson 2019 and finally Britz 2020 stepwise improve the representation of final demand in CGE modelling based on estimating and demand systems of the AIDADS family from cross-sectional data provided by the International Comparison program (ICP) and integrating them into CGE models. These approaches built on previous work using an earlier release of the same data such as by Seale and Regmi 2006 and by Preckel et al. 2010, studies which did however not integrate their estimates into global models.

The study by Roson and Van der Mensbrugghe 2018 is based on the publicly available data set from the ICP with its ten broader commodity groups of which just one relates to food. The typical implementation of such more aggregated results for commodity groups in a CGE would consist of a nested approach with homothetic sub-systems where only price effects matter. In such a framework, the overall spent for food and thus the income effect is determined by the top level system. One aggregated food category in the top nest allows depicting Engel’s Law of falling budget shares for food in total. But Barnett’s law (Barnett 1941) could not be incorporated which states that the budget share of meats increases relative to staples. Britz and Roson 2019 therefore integrate detailed data on the composition of food only available on request from ICP. In their estimation, food is depicted by ten categories.

Britz 2020 improves on this by moving from an AIDADS to a MAIDADS presentation where also the commitment terms change with the utility and thus income level, following Preckel et al. 2010. He also introduces additional explanatory variables, namely the share of two age classes on total population, the share of Islamic population and men temperature. Additionally, he proposes an improved way to treat the error terms when using the parameter estimates for benchmarking. We employ his estimates and benchmarking approach in here, focusing on extensions to ensure plausible calorie developments. The three key equations of the MAIDADS demand system are shown below. The Marshallian per capita demands x at given consumer prices *p* and per capita income *Y* are identical to a LES as seen from (1) but the commitment terms γ and the marginal budget shares δ are a function of utility *u* and thus endogenous variables during simulation (we leave out indices for the region, the household and time in the following where not needed to increase readability):

$$x_i = \gamma_i + \frac{\delta_i}{p_i}\left[Y - \sum_j \gamma_j p_j\right] \tag{1}$$

The equations for the γ and δ are structurally identical. They determine a linear combination between estimated parameter values at low δ_i^{lo}, γ_i^{lo} and high utility δ_i^{hi}, γ_i^{hi} where the combination is determined by a logistic function depending on utility *u* and the estimated parameters ω_δ, ω_γ and κ_∂, κ_γ:

$$\delta_i = \frac{\delta_i^{lo} + \delta_i^{hi}\exp(\omega_\delta u - \kappa_\partial)}{1 + \exp(\omega_\delta u - \kappa_\partial)} \tag{2}$$

And

$$\gamma_i = \frac{\gamma_i^{lo} + \gamma_i^{hi}\exp(\omega_\gamma u - \kappa_\gamma)}{1 + \exp(\omega_\gamma u - \kappa_\gamma)} \tag{3}$$

The utility is defined similar as in case of the LES demand system but depends on the endogenous parameters γ and δ:

$$u = \sum_i \delta_i \ln\left(x_i - \gamma_i\right) \tag{4}$$

During simulation with the CGE, these four equations jointly determine at given prices and income the demands due to the implicit nature of the demand system: the simulated *x* determine the utility *u* which in turns defines based on (3) and (4) the parameters γ and δ which drives the demands *x*. Britz 2020 proposes a smooth quadratic exponential function as provided by GAMS in (2) and (3). This will yield identical results to e^z as long as z<10 as the chosen point where the smoothing starts but prevents mathematical overflows beyond this limit which can help to improve the numerical stability during estimation and, in here, simulation.

2.2 *Estimating calorie contents*

While the empirical estimates from Britz 2020 underlying the demand system should guarantee an overall plausible development of income spent on different food item groups over a large range of per capita incomes, this by itself does not yet guarantee that calorie intakes evolve plausible as well. The reason is first that the calorie content per dollar spent between these categories varies tremendously as discussed below. The total calorie intakes therefore react sensitive to changes of the budget shares inside the food bundle. Additionally, if price relations between food and non-food commodities and between different food categories diverge over the simulation horizon, developments of per capita demands will deviate from those of the budget shares, depending on the (implicit) price elasticities, an effect which is quite hard to control during benchmarking where the later simulated price changes are not yet known.

In partial equilibrium models of the agri-food sector, food demand is typically depicted based on the food balance sheets of the FAO using physical units where final demand of primary products and of derived products thereof such as flour or bread from wheat is expressed jointly in primary product equivalents. When constructing the data base, the calorie supply to final demand from, for instance, wheat is calculated by FAOSTAT based on a weighted aggregation of the physical demands of the different processed food products derived from wheat plus direct food consumption of wheat, where calorie contents act as aggregation weight. Adding up in parallel their primary product equivalents allows calculating an average calorie content of the (implicitly) consumed primary product. The resulting calorie contents per unit of primary product are hence not identical across countries but depend on the regionally specific bundle of primary and derived products consumed. This concept has its clear limitations when it comes to more complex food products such as, for instance, a pizza. One reason is that calculating each balance sheet requires data on imports and exports of the related processed products which are only available in a standardized format up to HS6. Second, for computational reasons, each derived product is only linked to one single primary product, such that for a pizza, one need to allocate its total calorie content to one main primary product ingredient, for instance, to wheat.

Gouel and Guimbard 2019 estimate an AIDADS demand system for six food product groups directly based on these data. Their food demands are therefore expressed in physical units of primary agricultural products. They use farm-gate prices and where not available import unit values in their estimation and not consumer prices. This gives insights on how total calorie supply per capita changes depending on income. Their demand system can clearly not be used to depict final household demand in a CGE as their estimation considers a kind of functional goods of which the final demand is actually not observed and to which also no consumer prices are attached. Finally, and that holds for our approach as well, the FAO data rather measures calorie supply to the food value chain which is clearly not equal to actual calories consumed by the final consumer due to food waste and potential losses of nutritional content along the supply chain and in the household itself.

Moreover, the work of Gouel and Guimbard 2019 underlines again that the concept of the FAO balance sheets as a source for nutritional accounting is hard to reconcile with the concept underlying a SAM where the spent on larger aggregates such as “other processed food” is reported. The composition of such aggregates is certainly different across countries and likely depending on income levels. The “other processed food” in the GTAP data base, to give an example, comprises staples such as bread or pasta, moderately processed foods such as breakfast cereals and completely precooked meals such as a frozen pizza. Britz and Roson 2019 partly address this problem by estimating changes in input-output coefficients for all sectors including the food processing industry as a function of per capita income and use their estimates to update them during baseline consumption.

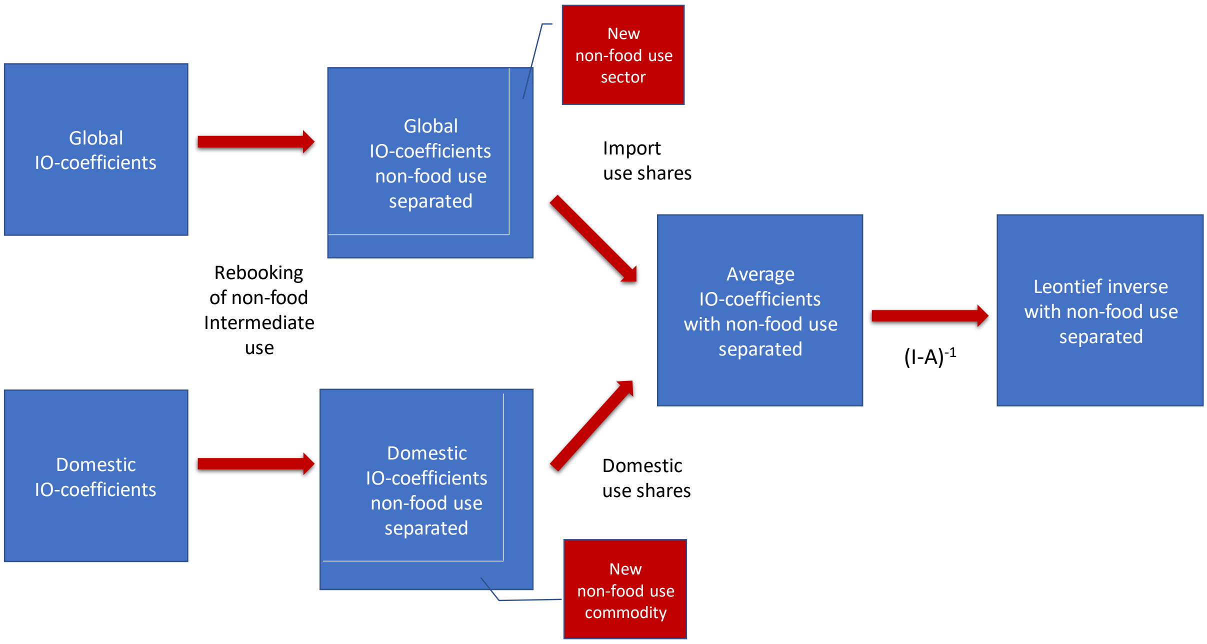
While calorie accounting is relatively common in partial equilibrium of the agri-food sector (cf. Britz and Witzke 2012, Rosegrant et al. 2008), the literature body on linking nutrition into CGE applications is limited. Thauw and Thurlow 2011 apply micro-simulation to update the food demand of each household according to the relative changes simulated with the CGE and attach nutritional values to the demand in a single country CGE for Tanzania. Their model features more agri-food detail compared to the GTAP data base (Aguiar et al. 2016), depicting 26 agricultural and 10 downstream agro-processing sectors. These details allow it to more easily attach (constant) calorie contents directly to final consumption. Moreover, in low income countries, the average processing level of food is still relatively low.

We want to make a step further and explicitly control for calories during baseline construction. This requires attaching calorie contents to the food commodities in the SAM used for final consumption. Rutten et al. 2013 propose here an iterative procedure in the context of the GTAP based CGE Model Magnet (Woltjer et al. 2014). They consider in a first round of their computations the input coefficients in a food processing sector for primary agricultural products to allocate the calories provided by the primary products to the processed one. In a second round, they consider the input coefficients of food processing sectors in the production of food processing and so forth. We see a Leontief-inverse as the more established way to track such linkages. However, as also stated by Rutten et al. 2013, we need to make sure that non-food demand of primary agricultural products, e.g. for feed, fibre or in the production of bio-fuels and of bio-chemicals is discarded during the calculations.

In order to construct this Leontief-inverse, we need to distinguish between non-food use and food use multipliers as only the latter match the calorie accounting approach by the FAO. We introduce therefore during the computations temporarily a new sector and commodity called “non-food use” in the matrix of IO-coefficients. All intermediate input use of agri-food commodities which is not food use is rebooked to this commodity. In order to maintain total intermediate input use in the SAM, the share of the rebooked use of the agri-food sectors on their production serves as a distribution key to assign intermediate input use to the new “non-food use” sector. Let’s give an example for this procedure. The cattle sector, “ctl” in the GTAP data base, uses wheat “wht” as a feedstock; its output is mainly used by sector producing cattle meat “cmt”. Buying meat from the “cmt” sector therefore indirectly requires wheat to feed the cattle which in the FAO balance sheets is booked as feed use and does hence not enter the calorie accounting. We therefore remove this input coefficient depicting intermediate use of wheat by the cattle sector (wht -> ctl) and report it instead under the aggregate sector “non-food use” (non food use -> ctl). The intermediate use of wheat by the cattle sector relative to total wheat output serves as the distribution key to assign a part of the intermediate input use of wheat to the new “non-food use” sector. The same rebooking occurs for all IO entries which are considered as non-food use.

The computation needs also to reflect that final demand and intermediate input use are partly served by imports. In order to ease the computational load, we therefore first generate global average IO-coefficients including the new non-food use sector as described above. For each country, we construct next IO coefficients which are a weighted average of the domestic and the global one, using domestic and imported use as weights. This is only an approximation as we would need to construct for each country a different rest-of-the-world aggregate. From there, we construct a multiplier matrix which allows deriving the agricultural output required to produce on unit of food demand. The process is depicted in Figure 1.

Figure 1: Construction of the IO-matrix to derive calorie contents



Source: Authors

Table 1 below reports the reciprocal of the resulting factors, expressed as the price of a 1.000 daily calorie in USD 2011, for the relevant agri-food commodities of the GTAP data base and ten aggregate world regions. The underlying factors are scaled such that when multiplied with the final demand at the benchmark reported in the SAM – including government and investment demand – they exhaust the calories reported by the FAO. The reader should first note that direct final consumption of cereals – the first three lines – as well as for oilseeds is hardly reported in the data which also renders the multipliers, resulting factors and calorie prices somewhat dubious. As one might have guessed, processed rice provides the cheapest calories in the group of processed food, followed generally by vegetable oils, raw sugar and fruits and vegetables. Differences in prices especially for the residual category "Other food processing" are quite large and span a 1:10 range. We did some back-of-the-envelope checks using the prices reported below against the expenditure data from the ICP from which they are not derived and found quite reasonable results.

Table 1: Price for 1.000 calorie per day in USD

	Oceania	East Asia	South-East Asia	South Asia	North America	Latin America	EU_28	MENA	SSA
Paddy rice	14	31	128	39		60		148	167
Wheat	73	10	96	29		76	67	84	91
Coarse grains		266	133	57	55	139	264	218	91
Fruit & vegs	477	400	415	241	580	245	500	435	250
Oilseeds	7	754	421	825	601	880	2238	1029	536
Fish	2099	674	1109	1310	953	1104	1433	658	938
Cattle meat	580	358	443	140	917	687	848	793	876
Other meat	359	172	257	179	425	397	814	475	322
Vegetable oils	181	140	189	148	110	207	262	128	70
Dairy	878	665	833	397	559	799	1103	657	518
Processed rice	177	43	50	53	108	91	229	79	41
Sugar	333	88	143	122	69	156	165	124	96
Other food	1300	353	444	126	580	430	1228	366	239
Beverages and tobacco	2218	690	930	525	2264	1067	2015	1188	365

Source: Calculations based on FABIO and GTAP 9 Data bases

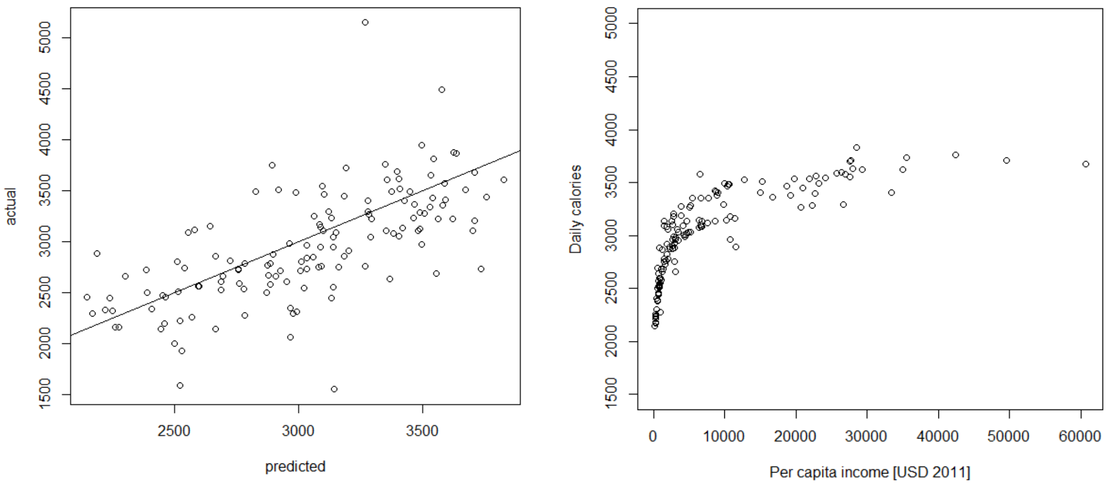
3 An empirical relation between calorie intake and income levels

In order to link per capita income and its changes to calorie intake, we need an empirical relation between the two, here provided by a simple regression from cross-sectional global data. Besides per capita income, we used mean temperatures (difference to mean in sample) as a control during estimation. The resulting relation (5) shows that the elasticity of the demand for calories demands decreases with larger incomes (see also Figure 2) as suggested by Engel's law:

$$\begin{aligned} calTot^* = &-947,275 - 16,257\,T \\ &+738,910\ln(Y) - 30,083\ln(Y)^2 \end{aligned} \tag{5}$$

The regression uses as dependents the reported calorie intakes from the FABIO data set (Bruckner et al. 2019) aggregated to the GTAP regions, while the per capita income is defined as the private consumption spent at agent prices as reported in the GTAP data base 9 divided by population size. This definition of income matches the variable which drives the demands during simulation. The adjusted R-squared of the regression is 65%, the p-values for the two income terms are at 0.3% and 0.04%, while the mean temperature has an even smaller p-value. It is reassuring to see that the estimated coefficients for the income dependency are not changing much (923.53, -38.20) when the mean temperature is removed from the model. It improves however the R-squared by 11%. In order to account for the fact that the GTAP regions are of quite different size, we used population size as weights in the regression.

Figure 2: Estimated versus observed calorie intakes and versus per capita income



Source: Own statistical analysis

The information on the calorie contents and the regression results are used twofold. First, in case where multiple households are depicted by the model and their consumption bundle is not provided, we estimate the calorie intake for each household based on the observed per capita income. This estimate for total calorie intake is complemented by the simulated per capita demands based on the MAIDADS system. Jointly, they are used in a Highest Posterior Density Estimator (Heckelei et al. 2008) to generate consumption bundles for each household with exhaust both their given income and the reported demands for the aggregate household in the SAM. Using the information on the calories has typically a limited impact only for poorer households in middle income and rich countries but helps to avoid implausibly high calorie intakes for richer households in these countries. The opposite holds for poor countries where partially very low per capita income can result in very low per capita calorie intakes. As the reported total calories intakes at country level need to be exhausted, implausible low estimates for one household groups result in higher ones for others and vice-versa. Considering the calorie intake resulting from an estimated consumption bundle during benchmarking is clearly not possible in case of a model layout with one aggregate household, only, where the per capita consumption is defined directly from the SAM.

4 Benchmarking the MAIDADS system while considering calorie intakes

During estimation and later simulation, the utility is implicitly driven by the demands which depend on the marginal budget shares and commitment levels which are functions of utility. In order to ease benchmarking, we proceed as Britz and Roson 2019 and use the outcome of a regression of the estimated utility levels on per capita income considering the additional factors driving the parameters of the MAIDADS systems in Britz 2020. This estimate of the utility level gives country and sector specific $\delta_{c,i}$ and $\gamma_{c,i}$ from (2) and (3) which need to be corrected to match the observed per capita demands from the SAM. In order to do so, we calculate the Marshallian per capita demands from (1) and from there for each product a scaling factor to the observed demand in the SAM, with a security threshold of 10%. With this scaling factor, we update δ_i^{lo}, γ_i^{lo}, δ_i^{hi}, γ_i^{hi} which at given utility also means that $\delta_{c,i}$ and $\gamma_{c,i}$ are scaled accordingly. Afterwards we scale the δ_i^{lo} and the δ_i^{hi} to unity to ensure adding up and repeat the scaling procedure again. That gives usually already a quite good match. Correcting both the commit terms and the marginal budget shares seems the better choice as larger changes to $\gamma_{c,i}$ or $\delta_{c,i}$ alone while keeping the other parameter unchanged can provoke quite curious elasticities.

But this alone could still provoke quite implausible calorie intakes during simulation. Assume, for instance, that at a country spends at the benchmark more than what the estimations would suggest for food which goes along with a higher than estimated calorie intake. The scaling procedure would increase the strictly positive commitment terms and the marginal budget share to match the benchmark observation such that this above average consumption share is also maintained when the income level increases which could provoke quite unrealistic developments. Here, it is important to keep in mind that the Engel curves are exponential and react sensitive to changes in the parameterization and also to which point on the curve they are calibrated which renders it hard to judge the impact on simulated outcomes.

As we have four parameters δ_i^{lo}, γ_i^{lo}, δ_i^{hi}, γ_i^{hi} for each product and can adjust the κ_γ and κ_δ as well, there is some room to impose plausible calorie intake developments with increasing income. Therefore, we introduce three additional income levels in the benchmarking procedure: (1) one at 10% reduced benchmark per capita income, (2) the income shifted with the per capita GDP growth to the final simulation year, (3) the mean between per capita income at the benchmark and (2). For each of these income levels, we use the regression function (5) to define the related calorie intake, corrected for the relation between the estimation from (5) at the benchmark and the calorie intake at the benchmark. Moreover, we introduce an equation into the benchmarking procedure which calculates the per capita calorie intake *calTot* from the demands *x* at different income levels, here denoted by t as they depict future simulated income level, and their calorie contents *cal*:

$$calTot_t = \sum_i x_{i,t} cal_i \tag{6}$$

The benchmarking step is defined as a constrained optimization problem. The constraints comprise: (I) adding up conditions for the vectors δ^{lo} and δ^{hi}, (II) the equations of the demand system, i.e. (1) to (4), however with δ_i^{lo}, γ_i^{lo}, δ_i^{hi}, γ_i^{hi}, κ_γ and κ_δ as endogenous variables and (III) the calculation of calorie intakes from (6). The objective function used considers (I) squared relative differences between the parameters δ_i^{lo}, γ_i^{lo}, δ_i^{hi}, γ_i^{hi} as estimated, scaled with the relation of the resulting demand estimate and the benchmark demand, (II) squared differences between calories from (6) and the estimates from the regression, corrected for benchmark level differences, (III) squared differences for κ_γ and κ_δ which are adjusted as well during the benchmarking process and, (IV) squared differences in utility levels as estimated from the utility level and derived at current parameters from (4). The demands for the benchmark point are fixed while for the three income levels, they are endogenous. The problem is solved independent for each household and region based on CONOPT4 on a grid. In order to improve plausible developments of calories, bounds on calorie intakes are introduced as an additional safeguard. The procedure can be interpreted as a Highest Posterior Density Estimator (Heckelei et al. 2008).

Unfortunately, tests of the resulting parameterization of the demand systems have shown that depending on income dynamics, calorie intakes can still look implausible (see also the result section below). We therefore introduce during baseline construction with the dynamic CGE additional equations which control for a reasonable calorie intake defined as based on regression function (5) which links calorie intakes to income levels. However, this relation between income and calories cannot be used directly in the simulation model as the calories determined from (6) are fixed and given at the benchmark based on the SAM values and will deviate from what the regression function suggest. We combine two options to deal with these error terms at the benchmark in (7). First, we drive in the long run the calorie intake towards the estimates suggested by (5) such that after simulating over two hundredth simulation years all households in all regions would end up on the regression estimate. Second, we add in each year the income depending change in the calorie intake to last year's intake, considering the relation between the intake at the benchmark and estimate at the benchmark. That gives the following equation for the desired calorie intake:

$$\begin{aligned} calTot_t &= \left(1-\frac{201-t}{200}\right) calTot_t^* \\ &\quad + \frac{201-t}{200}\left(calTot_{t-1}^* + \frac{calTot_0}{calTot_0^*}\left[calTot_t^* - calTot_{t-1}^* \right]\right) \end{aligned} \tag{7}$$

Adding (7) to the CGE requires an additional variable to maintain a square equation system while adjusting the preference structure of each household. We opt to shift both the commitment terms and marginal budget shares for food items endogenously after experimenting with each separately. Using the commitment alone runs the risk to drive the demands in the negative domain if larger downward corrections of the calorie demands are necessary. The frequently resulting negative commitments can also substantially change the price response. Equally, a downward correction of the commitment terms will increase the non-committed income which at unchanged marginal budget shares leads to a countervailing effect. Using a multiplicative update of the marginal budget share parameters is less likely to provoke negative outcome but requires an additional adding up condition for the updates of the δ. Specifically, if we downward correct all marginal budget shares of food items to reduce the calorie intake, we need to upward correct all non-food related ones and vice versa. This additional condition requires yet another variable to be determined endogenously. Furthermore, if the committed part of the income is high, the updates of the marginal budget shares must be quite substantial. Combining the two approaches has proven to work well even in case of multiple households in developing countries where estimated calorie intakes at the benchmark deviate substantially from the regression function and required corrections during baseline construction are large.

The correction factors for the marginal budget shares are integrated as follows where the endogenous correction factor *cor* differs for food *f* and non-food items *nf* and are defined relative to last year's marginal budget share, i.e. the corrections are defined in relative terms:

$$\delta^*_{i,t} = \delta_{i,t} + \delta^*_{i,t-1}\left(cor_f \wedge i \in f + cor_{nf} \wedge i \in nf\right) \tag{8}$$

For food items, cor_f it is implicitly determined by the calorie constraints (7). For non-food items, cor_{nf} is driven by the adding up conditions for marginal budget shares:

$$\sum_i \delta^*_{i,t-1}\left(cor_f \wedge i \in f + cor_{nf} \wedge i \in nf\right) = 0 \tag{9}$$

The correction for the commitment terms uses last year's per capita demands as weights. Using large year's marginal budget shares in (8) and last year's per capita demand in (9) should ensure that the provoked changes do not fundamentally change the composition inside of the food and inside of the non-food commodity bundle from what the benchmark step suggests.

Large commitment terms relative to the demand lead to a very low price elasticity and can provoke problems with solving the model. We fudge therefore the updated commitments with a Veelken-Ulbrich smooth min operator at 90% of last year's simulated per capita demands, see (10). This operator is an inbuilt function by GAMS such that it provides derivative information to the solver which might be better compared to, for instance, programming directly the otherwise somewhat simpler Huber max approximation. While we don't allow for negative commitment terms during benchmarking, the correction process could actually introduce negative ones which is another argument to control for calorie intakes at higher income level already during benchmarking to keep the changes from the correction small.

$$\gamma^*_{it} = -VU\begin{bmatrix} -\gamma_{i,t} - \dfrac{xa_{i,t-1}}{pop_{t-1}}\left(cor_f \wedge i \in f + cor_{nf} \wedge i \in nf\right), \\ -\dfrac{9}{10}\dfrac{xa_{i,t-1}}{pop_{t-1}} \end{bmatrix} \tag{10}$$

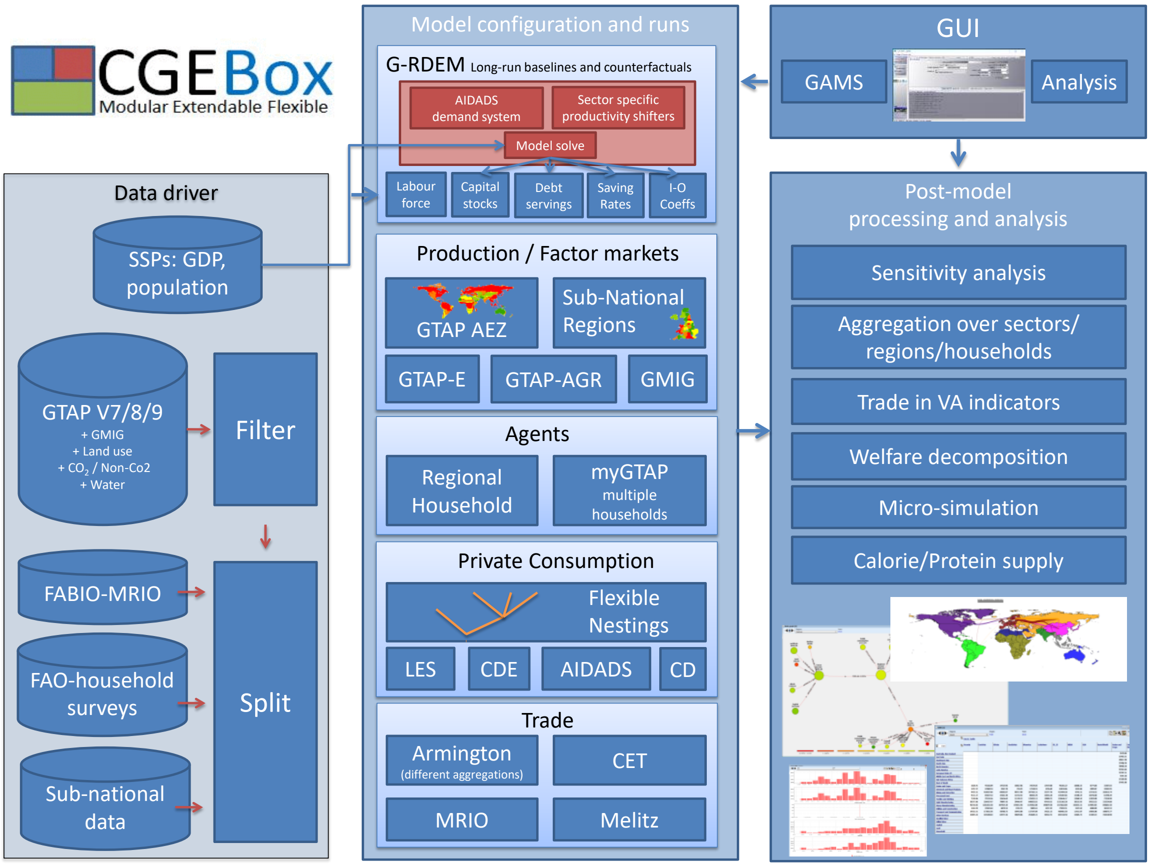
5 Testing the calorie module

We employ CGEBox (Britz and Van der Mensbrugghe 2018) as a flexible platform for CGE modelling, see Figure 3, to add the newly developed calorie module. The core of CGEBox consists of the widely employed GTAP Standard model (Hertel and Tsigas 1997) in its latest version seven as encoded in GAMS by van der Mensbrugghe 2018. CGEBox hosts in our application besides G-RDEM also implementations of GTAP-AGR (Keeney and Hertel 2005) and GTAP-E (McDougal and Golub 2007) to introduce more plausible cross-price relations into the CGE model. In the same vain, we also introduced CES sub-nests for cereals and meat for final demand and food processing sectors. We use all elements of G-RDEM, i.e. (1) capital accumulation, (2) macro-savings rate which depend on income levels and demographics, (3) sector differentiated productivity growth depending on the income level and its growth rate and (4) income dependent cost shares.

We touch here only briefly on how crop yields are controlled and land supply is depicted as this is discussed in detail in Britz and Escobar 2020. In CGEs without food and agricultural detail, it is not uncommon to treat land as part of the capital stock. In this case, capital accumulation would (implicitly) also increase the stock of land. Using the GTAP data base allows to single out land as a separate primary factor to avoid this implausibility. But it then requires assumptions on how the land stock develops over the long-run as the usual assumption of a fixed stock in comparative-static analysis makes limited sense, at least in regions which are still subject to considerable land-use dynamics as such as parts of Africa and Latin America. Multiple models have therefore introduced land supply functions which are typically driven as in Britz and Escobar 2020 by the land rent. The module by Britz and Escobar 2020 used in here is based on data on the remaining available crop land buffer in conjunction with supply elasticities which fit long-run projections of agri-food markets by FAO 2018. This allows expanding the GTAP-AEZ model to depict endogenous transition from natural vegetation to land in economic use. In order to maintain physical land balancing, the module employs the additive Constant-Elasticity-Of-Transformation function proposed by Van der Mensbrugge and Peters 2016. FAO 2018 also provides yield projections which are taken as exogenous drivers, based on a region and crop specific land productivity shifter.

We compare in here three model variants: (I) the proposed full model where parameters of the MAIDADS demand system are dynamically updated to let calorie intake follow a relation to income per capita, (II) a version with the MAIDADS demand system parameters as resulting from benchmarking and (III) one where the CDE demand system of the GTAP standard model is used at unchanged parameters. We are aware that (III) is not an easy to defend choice given the rather inflexible Engel curves of the CDE. But it allows us to better judge how important controlling for calories during simulation is compared to differences across functional forms. All three model variants are otherwise identically parameterised and structured.

Figure 3: Overview on CGEBox modeling platform



Source: Authors

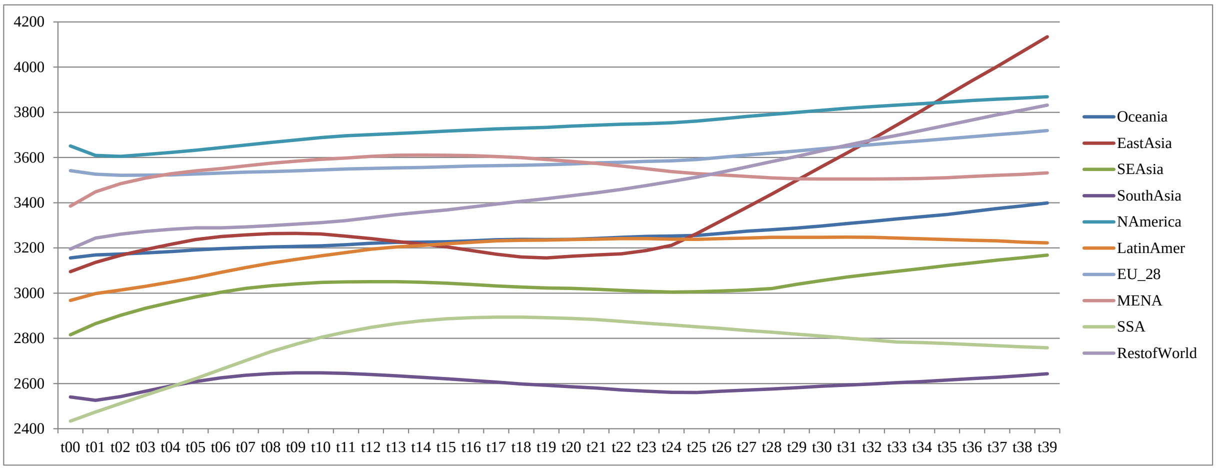
As similar comparisons of model variants where also reported in Britz and Escobar 2020 and Ho et al. 2020 we find it useful to use the same data base as used in these two studies for our comparison exercise. This data base maintains the full 57 sector resolution of GTAP 9 and aggregates to ten world regions. In order to shed light on differences across the model variants, we drive them all with the same GDP, demographics and education level projections under the Shared Socio-Economy Pathway 2 (SSP, Riahi et al. 2017) as published by the International Institute for Applied Systems Analysis, Shared Socio Economic Pathways Database, Version 1.1.). These SSPs were developed in the context of the IPPC reports on Climate Change. For each of these five SSPs, a single population and urbanization scenario jointly developed by IIASA and NCAR (National Center for Atmospheric Research) can be combined with three alternative SSPs GDP projections from the OECD, IASSA and PIK (Potsdam Institute for Climate Impact Research). These GDP and population projections are available in 5-year steps up to 2100 at single country basis. They are mapped in G-RDEM to the desired regional aggregation level and interpolated to yield a yearly time series.

6 Results from the different model variants

We start with a variant using the CDE demand system with benchmark parameters, see Figure 4. Note that the CDE is a rank-2 demand system where income elasticities are constant. The CDE system seems to exhibit relatively high demand elasticities in average for food items as calorie demands increase far stronger compared to what the regression function suggest. The graph highlights that the default parameters for the CDE denmand system provided with the GTAP data base comprises the plausible assumption that the income elasticity is higher for low income countries which is especially visible for East Asia. However, as these elasticities are kept unchanged, this can lead to implausible developments if income changes are high such as for East Asia where calorie intakes per capita are projected to increase from the current global averages of around 3,000 to over 8,000 calories per day. Therefore, most models used for long-run analysis reviewed in Ho et al. 2020 apply some updates to the parameters of their demand systems.

The high demand for food also requires a larger overall agricultural sector and, at given yield projections, far larger land expansion. Indeed, we were only able to solve the model by doubling the land supply elasticity for all model regions and for South Asia even by factor four plus increasing its land reserves by 250%. Otherwise, factor price relations diverge towards the end of the simulation periods to a point where the model becomes infeasible. In order to allow for comparison, we kept these changes to the default model parameterization active for further analysis. This implies that users will find different results when using the default parameters proposed for the GTAP-AEZ module of CGEBox which are aligned with land expansion rates from the FAO 2018 study.

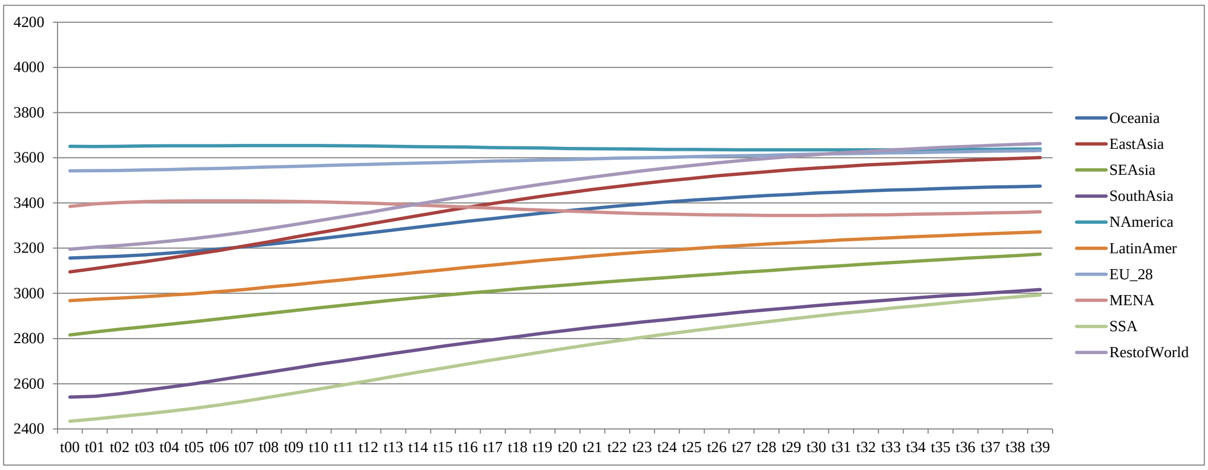
Figure 4: Simulated daily calories intake from 2011-2050, SSP2, CDE



Source: Model simulation

Figure 5 below shows the simulated calorie intakes for the model variant proposed as a future G-RDEM default, i.e. using the extended MAIADS systems in combination with the endogenous control for calories during baseline construction. Accordingly, these values are the outcome of adding equation (7) to the model such that in each period, the calorie intake is exogenously controlled. For the two poorest regions, SSA and South Asia which consume around 2,500 calories per day at the benchmark, the income growth projected under SSP2 would imply an increase to around 2,900 daily calories in 2050 which is about the current world average. North America even shows a slight decline in calorie intakes which reflects that its inhabitants consume currently with over 3,600 daily calories more than what the regression would suggest in the starting point given their income level and mean temperature. A similar effect can be seen for the MENA region which comprises oil exporting countries with a quite high per capita income. As to be expected due to the saturation effect with regard to calorie intakes and overall raising per-capita incomes in all regions, calorie intakes become more similar across regions over the simulation horizon.

Figure 5: Simulated daily calories intake from 2011-2050, SSP2, MAIDADS plus calorie correction

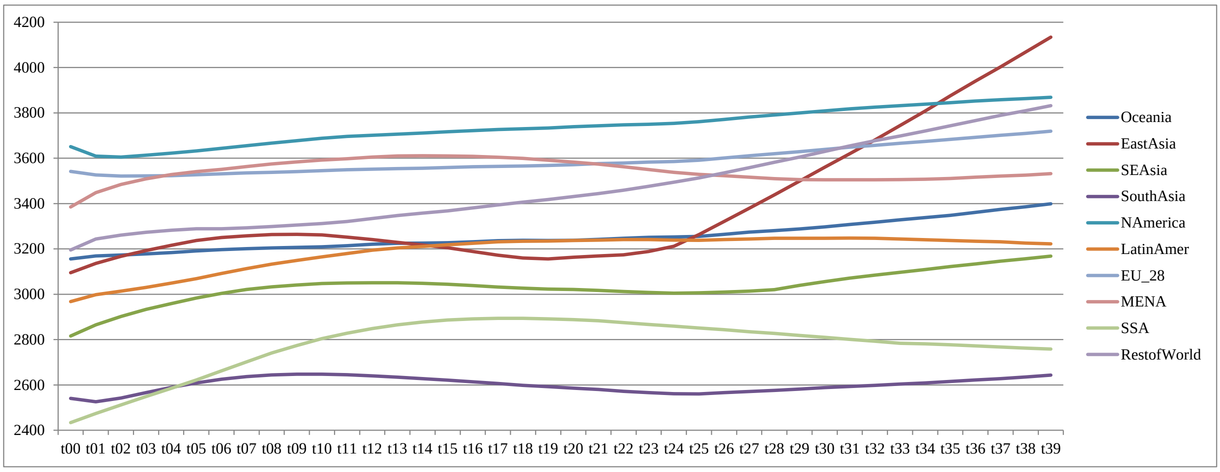


Source: Model simulation

Note: Axis scaled such is can be easily compared to the case without control for calorie intake as depicted in Figure 6.

That controlling for calories during baseline construction is recommended becomes visible if we consider the simulated calorie intake without using the calorie correction as depicted in Figure 6. As expected, the simulated developments of the empirically estimated rank-3 MAIDADS demand system are clearly more plausible than what we find under a CDE demand system with unchanged benchmark parameters as depicted in Figure 4. However, even the MAIDADS system can provoke curious developments as, for instance, for East Asia in our application. The regression function for the calories suggests an intake of around 3,600 calories per day at the end of simulation period (see Figure 5) but the model without the inbuilt correction for the calories would simulate levels of around 4,200 calories per day. For the Sub-Saharan Africa aggregates, Figure 6 shows up to the year 16 a plausible development of increasing calorie intakes given the real income development projected under SSP2, only slightly higher than what the regression function would suggest (compare to Figure 5). But afterwards, despite increasing incomes, calorie intake drops. This reflects that in average, price decreases from higher productivity of capital and labor are stronger for non-food items. This is especially visible in Sub-Saharan where high population growth jointly with higher income increases overall food demand and thus pressures on land. The higher land prices, combined with modest yield increases projected by the FAO, drive an increasing wedge between food and non-food prices, especially for crops. The real per capita spent on household consumption would suggest further increases in daily calorie intakes to around 3,000 calories a day, see Figure 5, the uncontrolled model suggest instead a value of around 2,750 calories in 2050, approximately the level reached in the uncontrolled simulation already in the year 2020.

Figure 6: Simulated daily calories intake from 2011-2050, SSP2, without calorie correction, MAIDADS



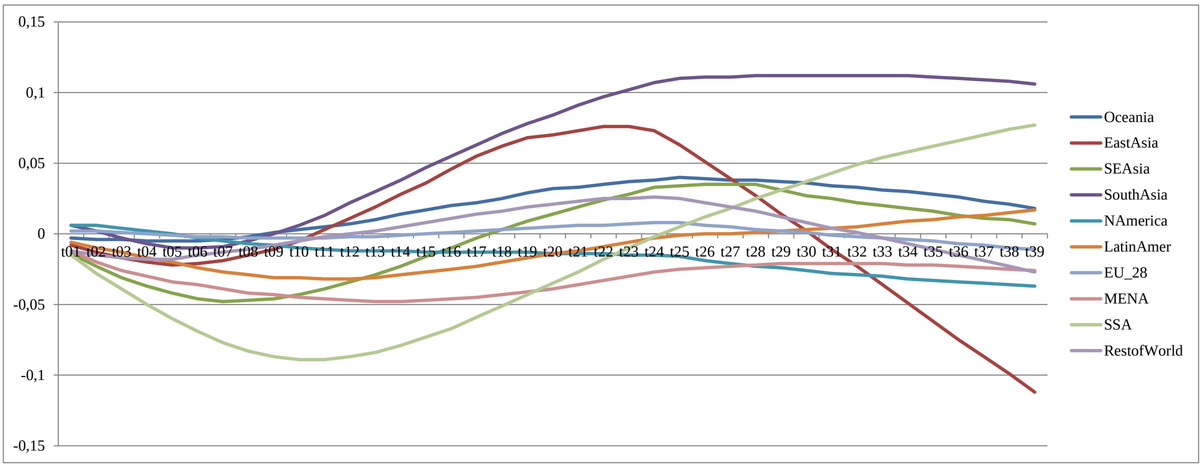
Source: Model simulation

The size of required parameters corrections of the MAIDADS demand system to shift the calorie intake developments from the unrestricted simulation shown in Figure 6 to the ones depicted in Figure 5 is visualized in Figure 7 for the food commodities. The highest positive correction is found for South Asia where from simulation year 22 onwards, marginal budget shares for food items are multiplied by a factor of around 1.1 which is reported as 10% in the figure. For East Asia, we see a similar upward correction up to the year 22. Afterwards, the correction factor drops to zero to the year 30 and afterwards turns negative, to offset the developments depicted in Figure 6. Given that budget shares for food are largely small already at the benchmark or fall considerably for countries with low incomes at the benchmark, these relative corrections are judged as relatively minor. The sweep observed for East Asia reflects that the parameters of the MAIDADS system depend exponentially on utility, i.e. income, see equations (2) and (3). This can lead to rapid changes especially if the expressions in the exponents of the logistic functions change their sign.

Why are such adjustments necessary if we control for future calorie intake developments already during benchmarking? First, during benchmarking, as mentioned already above, the simulated prices of the baseline are not yet known such that benchmark prices are used and only the income effect is considered.

Due to sector differentiated productivity shifts for capital and labour in G-RDEM, the development of the land prices driven by the land supply module, productivity shifts for land to match the exogenous yield projections and differences in cost shares, import shares etc. price developments over the baseline will differ among commodities and determine the consumption bundle besides the income effect. Second, we can't use formula (7) during benchmarking as it would require a full time series of calorie intakes.

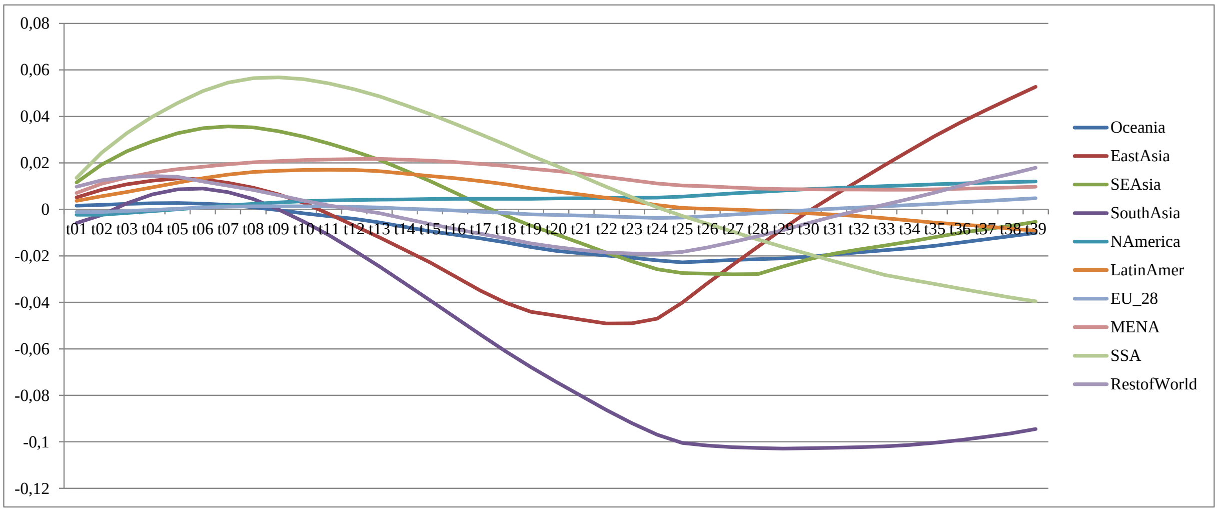
Figure 7: Correction factors to marginal budget shares and commitment to meet calorie intake estimates, food commodities



Source: Model simulation

The correction for marginal budget shares for food items must be offset by a correction of non-food items as depicted in Figure 8 according to equation (9). We see that maximal corrections are smaller than for food items, compare Figure 7. This reflects that the correction factor for non-food items receive the total marginal budget share of non-food as the weight in (9) which is considerably higher compared to food for most countries. Take Oceania as an example. Figure 7 suggest a maximal relative upward correction for food items of around 4% in the year 25 to meet the estimated calorie intake at simulated incomes. This requires a downward correction of the marginal budgets for non-food of just 2.3%, i.e. a correction factor of around 0.98. The relative small magnitude of the correction factors underlines that the total calorie intake reacts quite sensitive to relative small changes of parameters of the demand system which provides the motivation to introduce an endogenous correction factor and not to rely solely on the benchmarking process.

Figure 8: Correction factors to marginal budget shares to meet calorie intake estimates, non-food commodities



Source: Model simulation

7 Summary and conclusions

We present a methodology to derive nutritional indicators for processed food commodities in CGE analysis based on a Leontief inverse which separates food and on-food use. We use this indicator for calorie intakes during long-run baseline construction to shift preferences such that per capita calorie intakes follow an empirical determined relation to real per capita income. The preferences shifts are implemented by simultaneously updating the commitment terms and marginal budget shares of an empirically estimated MAIADS demand system. We assess this approach by constructing a baseline up to 2050 using the full 57 sector resolution of the GTAP data base 9 and ten world regions, comparing key results against two versions where calories are allowed to develop freely, one using the same MAIADS demand system and the second the more usual CDE demand system as part of the GTAP Standard model. The simulation results suggest that this approach indeed helps to yield plausible calorie intake developments also as a basis for long-run counterfactual analysis. Due to close link between calorie demand and agricultural land use, improving the model in this respect can also help to arrive at more plausible reactions at the intensive and extensive margin of land use and related carbon stock changes. Beyond a better control of results during baseline construction, the nutrition indicators are also an interesting aspect when analysing model outcomes in counterfactuals.

References

Aguiar, A., Narayanan, B., & McDougall, R. (2016). An overview of the GTAP 9 data base. *Journal of Global Economic Analysis*, 1(1), 181-208.

Britz W. (2020): A global panel estimation of a MAIDADS demand system with detail for agri-food considering demography, climate and norms, under review in Agricultural Economics

Britz W. and Escobar, N. (2020): Introducing carbon stock changes from natural land cover loss in Computable General Equilibrium analysis, under review in Journal of Global Economic Analysis

Britz, W. and Roson, R. (2019): G-RDEM: A GTAP-Based Recursive Dynamic CGE Model for Long-Term Baseline Generation and Analysis, *Journal of Global Economic Analysis* 4(1): 50-96

Britz, W. and Witzke, P., (2012). CAPRI model documentation 2012. Institute for Food and Resource Economics. Bonn: University of Bonn.

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

Bruckner, M., Wood, R., Moran, D., Kuschnig, N., Wieland, H., Maus, V., and Börner, J. (2019). FABIO - The Construction of the Food and Agriculture Biomass Input–Output Model. *Environmental Science &. Technology*, 53(19) 11302-11312

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

FAO, I., WFP, W. and UNICEF, (2019). The state of food security and nutrition in the world 2019: safeguarding against economic slowdowns and downturns.

Gouel, C., and Guimbard, H. (2019). Nutrition transition and the structure of global food demand. *American Journal of Agricultural Economics*, 101(2): 383-403.

Hanoch, G. (1975). Production and Demand Models with Direct or Indirect Implicit Additivity. *Econometrica* 43 (3): 395-419

Heckelei T., Jansson, T. and R. Mittelhammer (2008). A Bayesian alternative to generalized cross entropy solutions for underdetermined econometric models, Discussion Paper 2008: 2, http://purl.umn.edu/56973

Hertel, T.W. and Tsigas, M.E. (1997). Structure of GTAP, in: T.W. Hertel (ed.), *Global Trade Analysis: Modeling and Applications*, Cambridge University Press

Ho, M., Britz, W., Delzeit, R., Leblanc, F., Roson, R., Schuenemann, F. and Weitzel M. (2020). Modelling Consumption and Constructing Long-Term Baselines in Final Demand. Under second review in the *Journal of Global Economic Analysis*

Keeney, R. and Hertel, T., (2005). GTAP-AGR: A framework for assessing the implications of multilateral changes in agricultural policies. GTAP Technical Papers (no.25).

McDougall, R., and Golub, A., (2007). GTAP-E: A revised energy-environmental version of the GTAP model. GTAP Research Memorandum, 15.

Preckel P.V., Cranfield J.A.L., and Hertel T.W.A. (2010). Modified, Implicit, Directly Additive Demand System. *Applied Economic*s, 42(2):143–155

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'neill, B. C., Fujimori, S., ... & Lutz, W. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global Environmental Change*, 42, 153-168.

Rosegrant, M.W., Msangi, S., Ringler, C., Sulser, T.B., Zhu, T. and Cline, S.A., 2008. International model for policy analysis of agricultural commodities and trade (IMPACT): model description.

Roson, R. and van der Mensbrugghe, D., 2018. Demand-Driven Structural Change in Applied General Equilibrium Models. In *The New Generation of Computable General Equilibrium Models* (39-51). Springer, Cham

Rutten, M., Tabeau, A., and Godeschalk, F. (2013). A new methodology for incorporating nutrition indicators in economy-wide scenario analyses .FoodSecure Technical Paper No 1, December 2013

Seale, J.L., and Regmi A. (2006). Modeling International Consumption Patterns. *Review of Income and Wealth*, 52(4): 603-24.

Van der Mensbrugghe, D. (2018). The Standard GTAP Model in GAMS, Version 7. *Journal of Global Economic Analysis*: 3(1), 1-83

van der Mensbrugghe, D., and Peters, J. C. (2016). Volume preserving CES and CET formulations. Presented at the 19th Annual Conference on Global Economic Analysis, 15-17 June 2016 Washington DC (USA).

Woltjer, G. B., and Kuiper, M. H. (2014). The MAGNET Model: Module description. LEI Report 14-057,