We couldn't care less about Armington elasticities – but should we?

A systematic analysis of the influence of Armington elasticity misspecification on model results

This version: June 11, 2015

Hannah Schürenberg-Frosch

Abstract

This paper investigates the robustness of CGE models with respect to the elasticities of substitution in demand between domestically produced goods and foreign goods – the so-called Armington elasticities. The Armington-type modeling of trade is still one of the most extensively used specifications in CGE modeling. For a long time the choice of the respective elasticities of substitution has not been given much attention. The most frequently used procedure was to adopt the elasticities from the literature, which meant using elasticities that had been estimated (or guessed) for a different country and often also for a different degree of data aggregation. However, recently, some authors have shown that the elasticities 1) vary more substantially over countries than had been expected and 2) are higher in more recent estimations than in those which have been published in the 1980s and 1990s.

JEL classification: F14, C68, F17

Keywords: Armington, trade elasticities, Computable General Equilibrium, Meta-Study

Acknowledgements The author thanks Zoryana Olekseyuk, Volker Clausen and Edward J. Balistreri for helpful discussions and suggestions. Mehdi Belayet Lincon provided the best possible research assistance. The usual disclaimer applies.

Preliminary version, comments are most welcome.

University of Duisburg-Essen, Chair of International Economics, Institute for Economics and Business Administration, Universitätsstr. 12, D-45117 Essen. hannah.schuerenberg-frosch@uni-due.de

1 Introduction

In his seminal paper *A Theory of Demand for Products Distinguished by Place of Production* Paul Armington has provided a theoretical basis to explain the stylized fact that consumers in a distinct country demand the same good from domestic and foreign suppliers and even not only from one foreign supplier but from many even though the price is not equal for the domestic and foreign varieties. With this work Armington has paved the path for applied economic modelling to incorporate consumer preferences for different varieties of the same good depending on their origins. Armington's approach is easily adaptable in Computable General Equilibrium (CGE) models and at the same time explains trade patterns in a surprisingly accurate manner. Hence, the Armington trade specification still prevails as trade specification in applied models until today. There exist, of course, more modern and detailed trade theories which come up with more complex explanations for the same stylized facts, it depends, however, on the model application whether a more complex trade specification is necessary and feasible. See Balistreri, R. H. Hillberry, and Rutherford (2011) for a detailed discussion of the different trade specifications and their relation to each other.

Given the large dissemination of Armington-type models, it is interesting to investigate the influence this trade specification has on model results. The crucial parameter in the Armington setting is the so-called Armington elasticity - or more precisely - the constant elasticity of substitution between domestic and foreign varieties in domestic demand.[1] The correct choice of this elasticity will decide over the accuracy of model results. However, the Armington elasticities have not got much attention in the literature until recently. McDaniel and Balistreri (2003) show in a simulation exercise that the choice of the elasticity might be crucial in determining welfare gains or losses from a given policy reform. They find that even a qualitative switch in the overall welfare result is possible by changing the Armington elasticities. Schuerenberg-Frosch (2014) shows by drawing elasticities randomly from a uniform distribution that even though the quantity variables are robust, price results are quite sensitive with respect to the elasticity set. A similar approach is used by Frey and Olekseyuk (2014) and Jensen and Tarr (2012) with comparable results. R. Hillberry and Hummels (2013) argue that the frequently used elasticities in the literature, which stem from time series estimations of CES-functions are subject to a mis-

[1] The respective models often use a comparable approach in modelling the decision between domestic market sales and export market sales on the producer side, but this is not in the scope of the present paper.

specification in the underlying econometric procedure and that the elasticities should be much higher.

Given these recent concerns that we might have been mistaken in choosing the elasticities, we perform a systematic assessment of the influence of a misspecification of Armington elasticities on model results. Possible errata in specifying the Armington elasticities are only worrying if they impact crucially on model results. Thus, this paper contributes a meta-study of existing CGE models which have been investigated with respect to their sensitivity to changes in Armington elasticities.

For this purpose, we apply a robustness testing procedure on published CGE models for a wide range of countries and application areas. We replicate the original model simulations for each model and afterwards rerun the models 1000 times with randomly drawn elasticities. We draw the elasticities from the interval given by all existing elasticity estimations which is between 0 and 18.[2] Subsequently, we compare the span of results from the 1000 simulations with the original model results. We analyze the sensitivity of the following key variables: real GDP, change in private household welfare, change in aggregate imports and exports, change in consumer prices, change in factor prices, sectoral production. Each of these variables is investigated with respect to these measures for robustness: spread between maximum(minimum) result and mean (in %), spread between maximum and minimum (in %), qualitative switches (yes/no), deviation of original model results from robustness simulation mean (in %). We consider a model robust if the spread between the maximum and minimum is not more than 50%, the original model result does not deviate from the mean by more than 25% and no qualitative switches are found.

Our results show a rather clear picture concerning the robustness of CGE model results when using substantially different Armington elasticities: They are not robust. For the vast majority of models we find quite noteworthy variation of simulation results. In general, GDP results are more sensitive than welfare results and sectoral results are more sensitive than aggregate results. In the most extreme case, the minimum result for relative GDP changes (in %) lies 300% below the mean result, the maximum result up to 500% above the mean. The original model results deviate in most of the models by substantially more than 10% from the mean obtained in our randomized simulations,

[2]So far, we have only investigated the so-called 'macro-elasticity', a follow-up study on 'micro-elasticities' is planned but not yet implemented.

the most extreme deviation is 130% for change in real welfare and 3000% for change in real GDP. In many cases the original simulation results lay rather at the margin of the span of our results. For most of the models at least some specifications lead to qualitative changes in the key variables i.e. a policy which has been found to increase welfare (or GDP) may in some specifications be considered harmful for private welfare (or GDP) and vice versa. Even though qualitative changes occur in the minority of cases, the large quantitative influence is still quite worrying. We expect that an extension of our pool of models will still reproduce the same result and thus argue strongly in favour of three key advices for future modelling: 1.) Modellers should, if possible, use estimated elasticities for the country in question (even though the right strategy to estimate the elasticities is also subject to discussion in the community). Adopting this would reduce the ambiguity of model results. 2.) Modellers should always test their model results for robustness concerning the elasticities and report the results transparently in their papers. 3.) The interpretation of model results should take into account this source for uncertainty and thus very small changes in key variables should be taken with some caution, especially if they are unintuitive. In addition we see a case for much more research in this area. It would be very helpful to know more about the determinants of the Armington elasticities and about their evolution over time. Thus, a systematic estimation of Armington elasticities which accounts for the different possible methodologies and includes as many countries as possible would be very helpful to improve the reliability of model results.

2 Related work

The *correct* size of Armington elasticities is disputed ever since they have entered CGE modelling as an important parameter. While the - perceived - majority of modellers is rather agnostic about the specification of the elasticities and simply adopts them from other studies with roughly comparable characteristics, there exists, indeed, a broad literature about the estimation of these.

Since the 1970s several studies with estimated Armington elasticities have been published.[3] At first glance the studies seem to come to rather comparable results. Elasticities found in this earlier branch of studies are often around or slightly smaller than unity with higher elasticities found for the short term compared to the long term. This last result is found both by altering the fre-

[3] A very complete overview of the literature can be found in McDaniel and Balistreri (2003) and Welsch (2008)

quency of the data from monthly to yearly as well as by using error correction models. This perceived homogeneity of econometric results might have led to the widespread practice to adopt elasticities from other studies or "guestimation" (See Welsch, 2008). However, this seeming agreement among studies has to be taken with a pinch of salt.

First, the overwhelming majority of - older - time series estimations with disaggregated industries are for the US (e.g., Reinert and Roland-Holst (1992), Shiells and Reinert (1993), Blonigen and Wilson (1999) and Gallaway, McDaniel, and Rivera (2003)). Thus, it is not surprising that results differ only slightly as basically the same technique is applied to the same country. Changes hence only stem from a switch in preferences over time or from a different level of aggregation. Only very few time series analyses exist for other countries as also Welsch (2008) points out. These studies find, indeed, differing results compared to the US ones. Generally speaking, most estimations for non-US-countries find higher elasticities.

Second, while some studies estimate the so-called 'macro'-elasticity, i.e., the elasticity of substitution between domestic and foreign goods others estimate the 'micro'-elasticity which is the elasticity of substitution between different countries of origin.[4] Only a very limited number of studies use a nested approach and estimate both at the same time. This is a problem whenever the model which employs these estimated elasticities does not follow the same structure, i.e. does not make a distinction between different trading partners but this has been done in the underlying estimation or vice versa. Not surprisingly, the estimated micro elasticities are much higher compared to the macro elasticities.

Third, the estimated elasticities are higher if the data used is more disaggregate in terms of the number of sectors included. Again, a very plausible finding, as more disaggregate data contains sectors that are more homogeneous in the produced goods and thus also higher in their international substitutability. This phenomenon is generally considered as an "aggregation bias". The homogeneity of the results in first generation estimations for the U.S. thus somewhat lies in the fact that the studies use different waves of the same data set which is a rather disaggregate dataset for U.S. industries containing 192 sectors. If the estimated elasticities are to be used for a CGE model, the problem is somewhat more complex. The aggregation in the data used for

[4] This distinction and denomination is adopted from Feenstra, Obstfeld, and Russ (2012)

estimation should, in our view, match the aggregation that will be used in the respective CGE model. Hence, while estimated elasticities at the 2-digit-level might be too low for the use in a very disaggregate trade model, they might however be more convenient for a rather aggregated CGE model - a point also made by Welsch (2006). Given that this aggregation problem has been confirmed by many studies, one should, as McDaniel and Balistreri (2003) point out, be cautious in using elasticities from a very aggregate estimation in a more disaggregate setup or vice versa. However, this is a common practice.

Forth, younger studies tend to find much higher elasticities compared to the first generation of studies – that we will term in this paper as the traditional estimations – cited above. This could, misleadingly, be interpreted as a change in consumer preferences into the direction of higher integration in international trade. Unfortunately, this interpretation cannot be made unbiasedly as the techniques employed in more recent studies differ in terms of aggregation, econometric procedure and interpretation, from the ones used in the older ones: Most traditional time series studies, especially those for the US, use 3-digit-level data (i.e., between 150 and 200 sectors) and employ either a simple OLS, an OLS with lagged endogenous variables or, more recently, error correction approaches as the variables are typically integrated. Examples for time-series approaches are Reinert and Roland-Holst (1992), Shiells and Reinert (1993), Gallaway, McDaniel, and Rivera (2003) and Blonigen and Wilson (1999) for the US, Kapuscinski and Warr (1996) for the Philippines, Gibson (2003) for South Africa and Welsch (2006) for France. Recent studies such as Saito (2004), Welsch (2008) and Németh, Szabó, and Ciscar (2011) provide panel data results. The – younger – panel studies typically use a much higher aggregation with only 6-15 sectors. The elasticities found in panel studies are slightly smaller than those found in time-series studies thus contradicting the often formulated argument that cross-sectional studies per se obtain higher results. These studies are, even amoing this subgroup, not completely comparable as the dimensions of the panels differ. While Saito (2004) estimates a panel with variation over time and importing country for the OECD, Welsch (2008) estimates a panel over time and sectors and Németh, Szabó, and Ciscar (2011) again identifies variation over time and importing country but follows a nested procedure for the 'macro' and 'micro' elasticity.

Fifth, and most important: R. Hillberry and Hummels (2013) in line with Hertel et al. (2004) and Valenzuela, Anderson, and Hertel (2007) state that the still prevailing – traditional – time series approach to estimate the elasticities based on the price differential over time neglects important information and is not appropriate. They propose a different strategy of estimation identifying a variation in prices over trading partners instead of over time (or over importing country as other cross-sectional and panel studies in the field do) proxied by trading partner specific transport costs. Hence, they provide a cross-sectional approach which still produces country-and sector-specific estimates for the elasticity of substitution. Their results lie much above the results from most published studies in the field with elasticities of up to 18 compared to elasticities around unity in the time-series literature. In addition they show that the use of unit prices which is common in traditional estimations of the CES-function leads to biased results around 1. It can also be shown and has been demonstrated by Valenzuela, Anderson, and Hertel (2007) that elasticities around unity mistakenly lead to the identification of optimal tariffs in CGE model applications. As the estimations by R. Hillberry and Hummels (2013) and Hertel et al. (2004) reach a dimension which is between 4 and 18 times as high as the elasticities most frequently used in CGE modelling and as these new elasticities have entered the very widely used GTAP database, one should ask the question by how much this shift in the dimension of Armington elasticities impacts on the results.[5]

3 Approach and Data

With the presented analysis we aim at investigating systematically the influence of a noteworthy change in the Armington elasticities on model results. Noteworthy is here to be understood as changing from what we call the traditional elasticity interval (0.1–3) to the interval proposed by Hertel, Hummels and different co-authors (9–18) and vice versa. It is not surprising – and even desired – that such a change in trade preferences has an influence on the trade pattern of a country. However, the dimension of this effect is important if model results are used for policy simulation because, as mentioned in section 2, the *true* size of the elasticities is disputed. In addition, CGE models are not only used for trade policy analysis, thus, if Armington elasticities even impact on model results if trade policy is unchanged, the elasticity effect might distort the actual policy experiment in case that the chosen elasticities are wrong for one or another reason. We do not argue that models should in general not

[5]Please note: The strategy proposed by Hilberry, Hummels, Hertel and different co-authors and also employed by Saito is in fact a strategy of estimation of the micro elasticity as it distinguishes between trading partners. The elasticity found and employed must thus be considered as an average over micro elasticities for the respective country - it is hence quite obvious that it must be higher compared to the macro elasticity estimated in traditional studies. The author of this paper does not make any judgement over what strategy is the right one, especially as this depends strongly on the model setup in question. This paper only demonstrates the effects of the adoption of a different elasticity set.

react to changes in Armington elasticities, however, given the rather "hit-or-miss"-way many modellers chose the elasticity set, we consider it worrying if a change in the elasticities leads to an effect as high as or even higher than the original simulation experiment which had been done in the model.

3.1 Data

The study comprises (at the moment) 28 original simulations within 10 models for 11 different countries.[6] All models included are formulated in GAMS and use either MPSGE, MCP or NLP as syntax. We only included general equilibrium models. The models have been simulated with the respective original datasets provided by the authors. An overview of the models, scenarios and respective original elasticities is given in table 1 in the appendix.

Most of the models included in this study have elasticities around unity and have thus to be considered as belonging to the *traditional* branch of Armington-type models. Many also do not use different elasticities for the different sectors. This usage of one elasticity for all sectors may stem from computational limitations in older models. The sectoral aggregation is rather heterogeneous with one model having only three sectors while another one has 40. Most models are single country models but there are also multi-regional models included.[7] The areas of application differ as well, there are models for national fiscal policies like tax reforms, for external shocks like an inflow of foreign policy and for trade policy such as tariff elimination or joining trade unions.

As results of different policy scenarios within the same model differ and our robustness measures are based on median and mean results, we treat each scenario separately in our robustness tests. The same applies to multi-regional models where, of course, each country is treated separately when it comes to result comparison. Hence, each data point in the results shown below corresponds to one country and one policy scenario.

[6] We are still including new models and are very grateful for any model provided by fellow researchers. If you are interested in contributing to this study please contact the authors for details.

[7] The technique described is not limited to single-country models, static models or to small scale models, however it was difficult to get access to more complex models as authors are very reluctant to provide access to their model code and data.

3.2 Approach

In order to test whether the included models are robust with regard to the elasticities or if the influence of the Armington elasticities is noteworthy, we resimulate the original policy experiments done in the models 1000 times but we draw the elasticities randomly from a uniform distribution. Thus we run the same simulation with 1000 different elasticity constellations and report the results. The interval for the elasticities is between 0.0001 and 18. This large interval represents all estimations which have – as far as we know – been published on the size of Armington elasticities.(See also R. Hillberry and Hummels, 2013) As most of the models, we could access have their original elasticities around unity, which has traditionally been the case, our robustness check means for most of them a strong shift upwards in the elasticity set. This is comparable changing a model with older elasticities to a newer elasticity set, as has been mentioned in the literature survey in the previous section.

We report the results for some macroeconomic key variables like GDP, Hicks equivalent change in private welfare, aggregate exports and imports and consumer price index. In addition, we report sectoral results for production, exports and imports as well as welfare per household group. However, due to the differences in sector and household disaggregation as well as in order to keep this paper readable, we do not present these disaggregate results within this paper.

We define three different measures for model robustness:

1. The deviation between the maximum and minimum result in our robustness test in % relative to the mean
2. The deviation of the original result from the mean of our robustness test results in % relative to the mean
3. Whether a qualitative change occurs in our robustness test results

We regard a model as not robust if the Maximum-Minimum deviation is above 50% compared to the mean of all 1000 simulation results or if the original result deviates by more than 25% from the mean in our simulations or if a qualitative switch occurs in our simulations.[8]

[8] We initially considered much smaller deviations as indications for non-robustness - namely the more "‘traditional" confidence bands of 5 and 10% - but given the rather significant change in the elasticities of up to factor 18, we have opted for a less strict definition of *robust*.

3.3 Implementation

As the models included are quite heterogeneous in both their dimension as well as their application, automation of the robustness checks was difficult. Many models feature multiple scenarios within the code as well as their own reporting procedure, the definition of sectors, elasticity parameters and reporting variables differs and hence using the exactly same code on all models was not possible. The easiest way to run our robustness checks without having to rewrite the whole scenario definitions is, to run the model as it is within a loop over different elasticity sets where the first run is the original setup and up from run 2 elasticities are randomized.

The concrete implementation consists of the following steps:

1. Running the model without changes and checking for correct benchmark replication and consistency of results.
2. Identifying the elasticity-parameter or introducing it if elasticities were set in other ways than by specifying a parameter.
3. Introducing a loop set and programming a loop with 1000 elements.
4. Defining the elasticity as random draw from the interval 0.00001 – 18.
5. Including reporting of key results depending on the loop run and writing the value of the elasticities into the report, too.
6. Analysing the results, calculating descriptive statistics such as maximum, minimum, mean and standard deviation.

4 Results

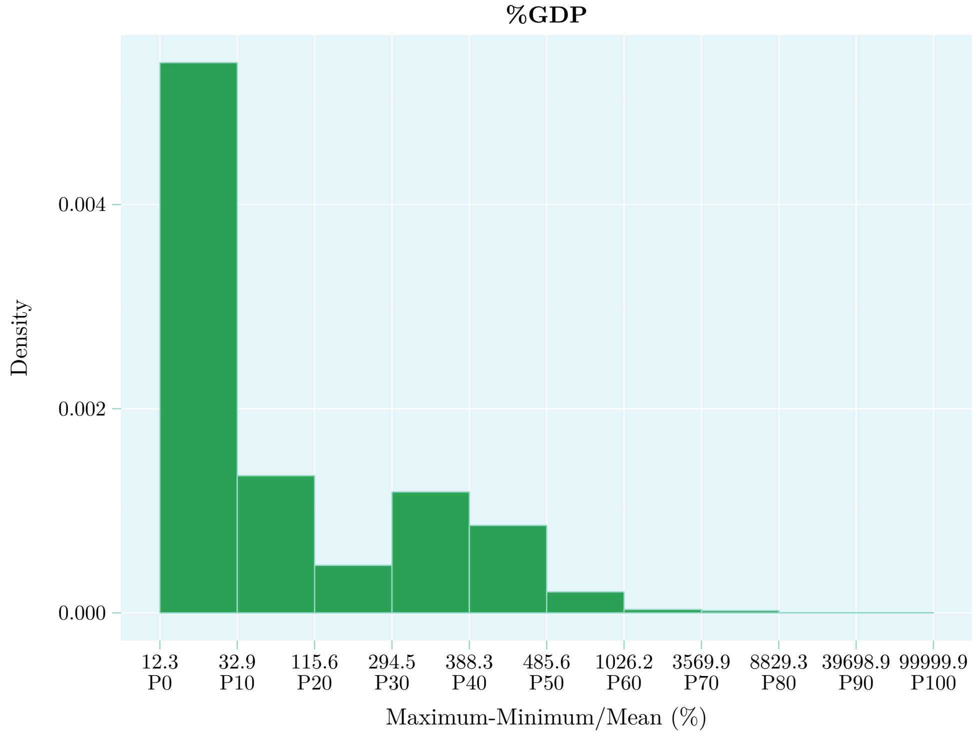
An overview of the results is given in tables 2 – 3 in the appendix.[9] In general it must be concluded that most of the models are not robust as of our results and criteria. Welfare results are more robust compared to GDP results and trade is even less robust - which is intuitive as trade effects are direct whereas GDP and Welfare effects are indirect effects.

[9] We do not link the results to specific models here as we leave it free to the model authors to publish the results for their respective models. The randomly attributed scenario numbers in the graphs shown in this section correspond to those in the result tables, but not to those in the model overview in 1.

4.1 Maximum-Minimum Spread

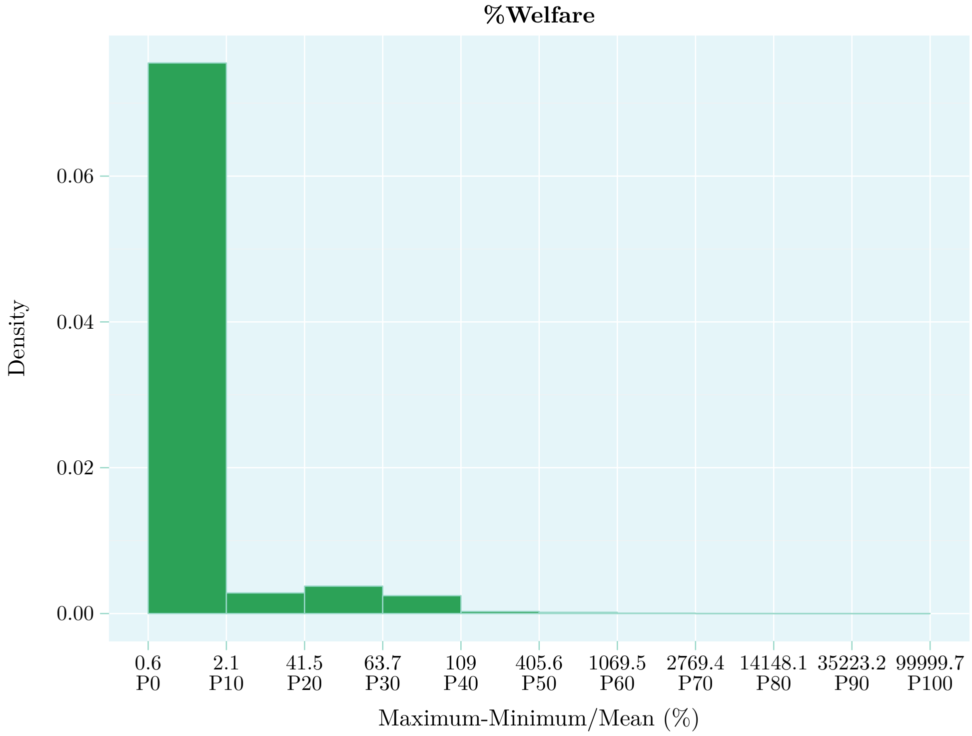
The maximum-minimum spread criterion measures whether the variance of our results is above 50% of the mean of our results. Hence, in the case of a normal distribution a model would be considered robust if the minimum lay not more than 25% below the mean and the maximum lay not above 25% above the mean. Our criterion however allows for a non-normal distribution where the mean lies closer to the minimum or maximum respectively if the span does not exceed 50% alltogether.

Figure 1: GDP: Maximum-minimum spread relative to mean %



This criterion has evolved to be the toughest one which is only met by one single simulation for GDP and by six simulations (out of 28) for welfare. Even if we would define our criterion even less strong and allow a variation of up to 100% around the mean, 25 simulations for GDP and 17 simulations for welfare would not fulfil this criterion. Hence, from the point of view of spread of the results most models are to be considered not robust.

Figure 2: Hicks equivalent change in private welfare: Maximum-Minimum spread relative to mean %



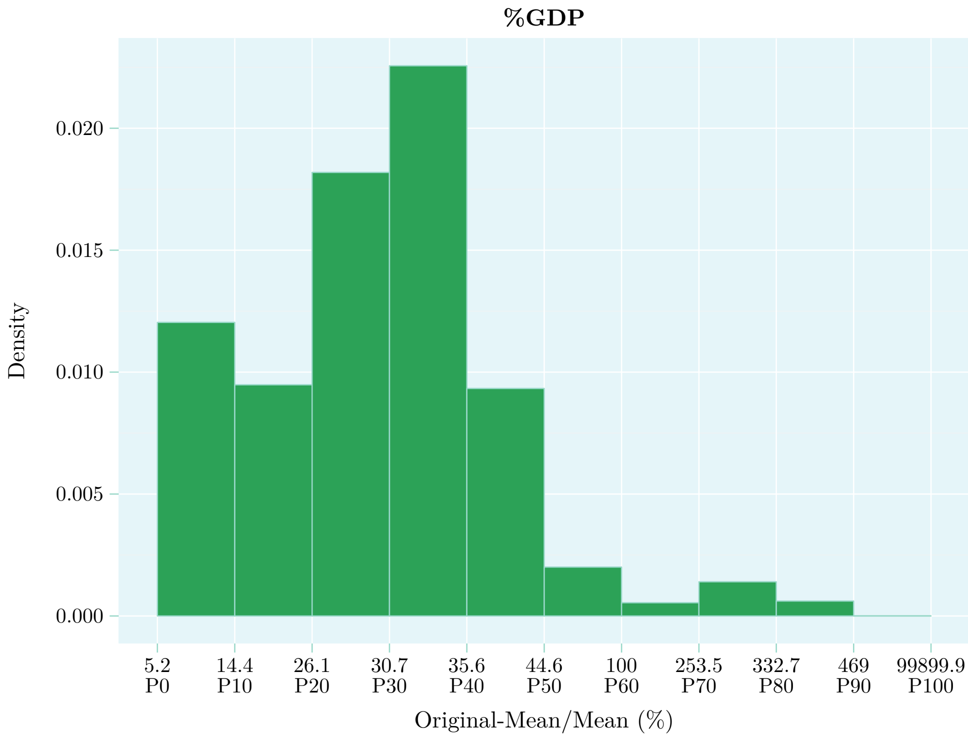
The highest spread between maximum and minimum result is more than 3000% compared to the mean which means that the real GDP effect might lie by up to 3000% higher or lower than the result obtained in one single simulation with one specific set of elasticities. This is of course quite a noteworthy span of results which leads to a high uncertainty about the "real" effect of the simulated policy.

4.2 Mean-Original Spread

This is the second strongest criterion. A model is regarded as robust if the original simulation result does not deviate by more than 25% from the mean in our simulations.

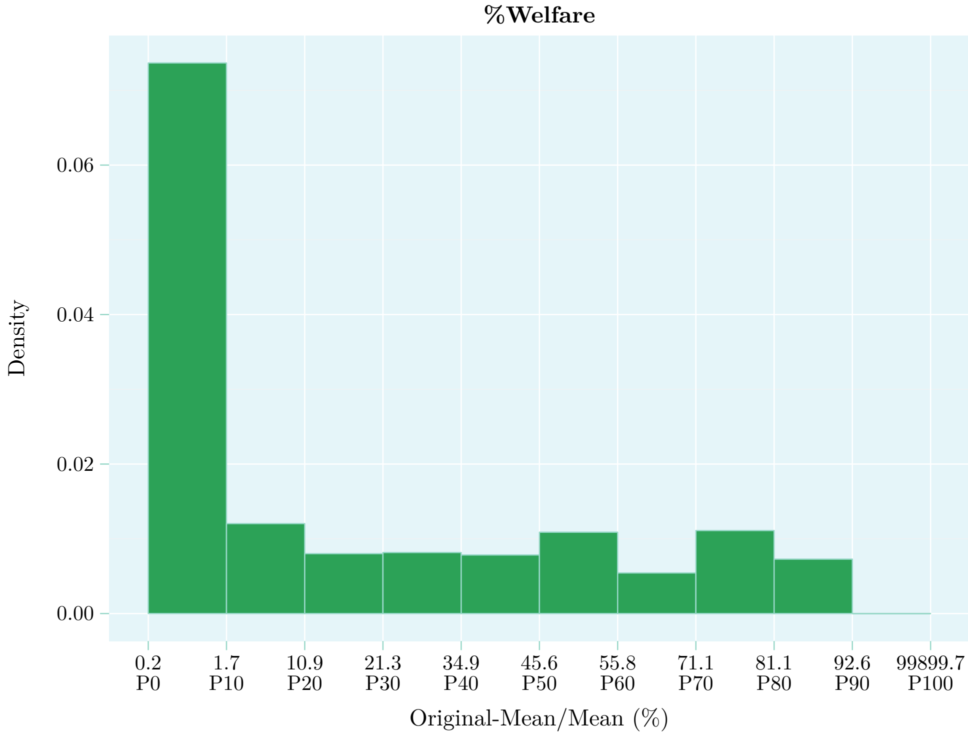
This criterion is met by 2 out of the 28 simulations for change in real GDP and by 12 out of the 28 simulations for change in real welfare. Hence, some further models are judged as robust if this criterion is used, still, most of the models have to be regarded as not robust if elasticities are changed largely.

Figure 3: GDP: Deviation of original result from mean %



The highest deviation between the original results and the mean in the robustness tests is at about 700% for GDP and at about 130% for welfare, meaning that the GDP effect in the mean simulation in our study lies 700% away from the result in the original study and 130% for welfare respectively. Again, this is worrying as a strong misjudgement of the GDP effect of the simulated policy is possible.

Figure 4: Hicks equivalent change in private welfare: Deviation of original result from mean %

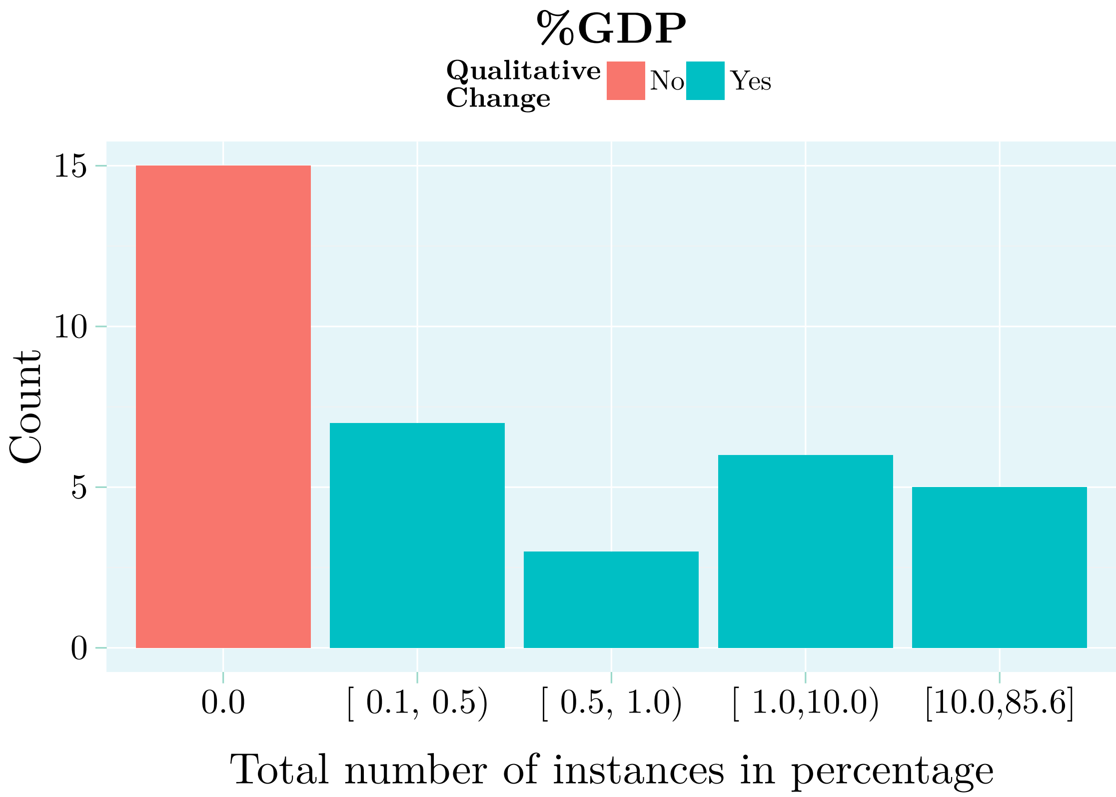


Interestingly, the models which deviate largely in the case of this criterion are not necessarily the same as those which do not meet the maximum-minimum criterion and this is not only explicable by a generally higher elasticity in the original simulation. It seems that while some models produce a strong variation (close to random in some cases) in results if elasticities are altered other ones have a rather small spread of results but the original simulation is closer to a corner solution than to the mean or median of all simulations. Both these findings are worrying even if implications differ. A large spread of possible results reduces the reliability of any possible result produced by the respective model while a corner solution in the original simulation gives reason to question the concrete effects presented in the respective papers while not necessarily reducing the reliability of the model per se.

4.3 Qualitative switches

This criterion is the least strict one and on the other hand the one which points most strikingly at problems in elasticity specification. It will consider a model not robust if in any of the 1000 robustness check simulations the result for either GDP or welfare effect has the opposite mathematical sign from the ma-

Figure 5: Qualitative switches in GDP results



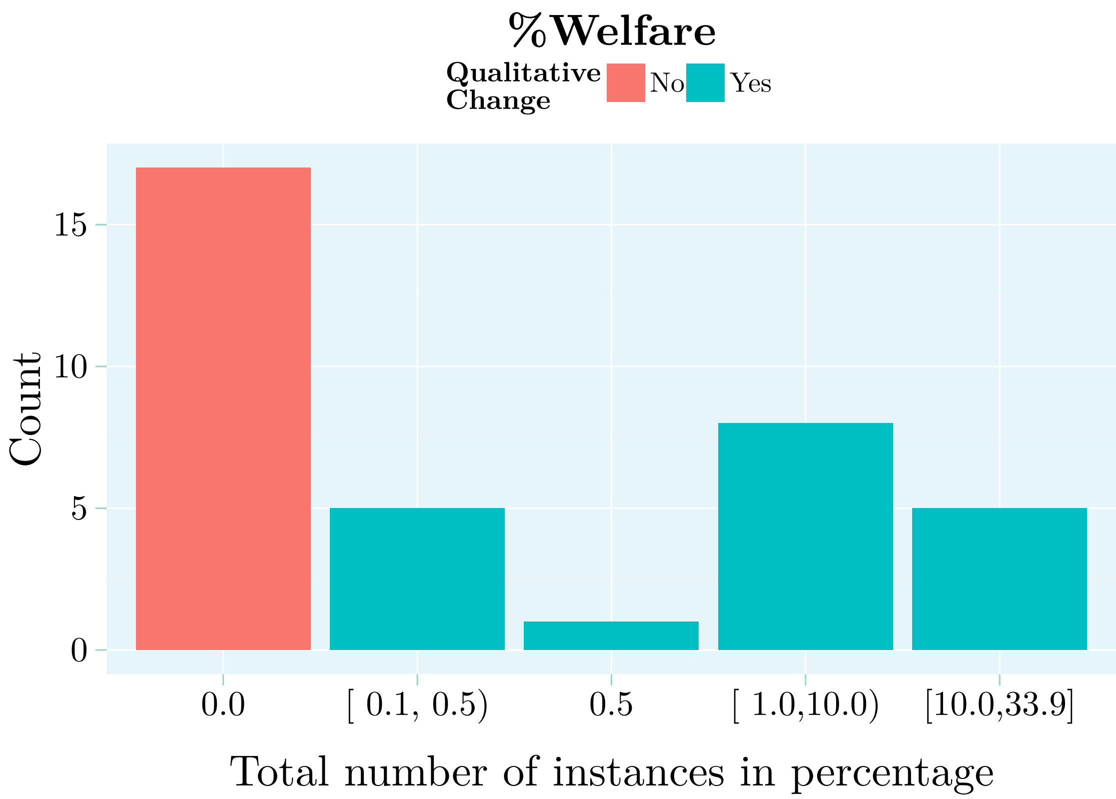
jority of the simulations i.e. if any simulation produces qualitatively different results. This is obviously a severe problem. If a policy scenario is considered GDP increasing under one elasticity specification and GDP decreasing under another or welfare respectively, the model results are completely ambiguous.

This criterion is not met by 14 of the 28 simulations for GDP and 12 for welfare which means that about half of the simulations are clearly to be considered not robust if elasticities are changed within a wide range. It might well be that changing the elasticities within the range of traditional time series estimations (0.5–5) would not cause any noteworthy problems but with the span simulated here (0.00001–18) the results of many of the models included are highly affected.

4.4 Relationship between robustness and application scenario

One might argue that it is completely intuitive and even desirable that a crucial change in the trade preferences creates noteworthy effects. However, we

Figure 6: Qualitative switches in welfare results



do not present the direct effects on trade itself here, but the indirect effect on GDP and welfare. Moreover, only part of the simulations included are trade-related policy scenarios. One would guess that trade policy scenarios might in general be more sensitive to changes in the elasticities compared to scenarios which only include national policy like tax reform scenarios or public spending scenarios. Hence, in 7 we present the results for GDP separately for trade and non-trade simulations, 8 shows the same distinction for welfare effects.

It is quite obvious that the most important outliers in the span of results measured by the Max-Min-Criterion are all trade-focused models. However, noteworthy deviations of original results from the mean also occur in non-trade focused models, especially in the case of GDP.

Still, taking all results and both criteria into account non-trade simulations are much more likely to be robust with respect to changes in the elasticity, this applies both to GDP as well as to welfare effects. Hence, in simulations were the trade effects are only indirect and not direct effects from the simulated policy, the choice of the elasticities is somewhat less crucial compared to models focused on trade.

The same conclusion also applies to the *Qualitative-Switch-Criterion* were most of the models with qualitative switches are, indeed, trade models.

Figure 7: GDP: Max-Min-Criterion and Original-Mean-Criterion distinguished by application

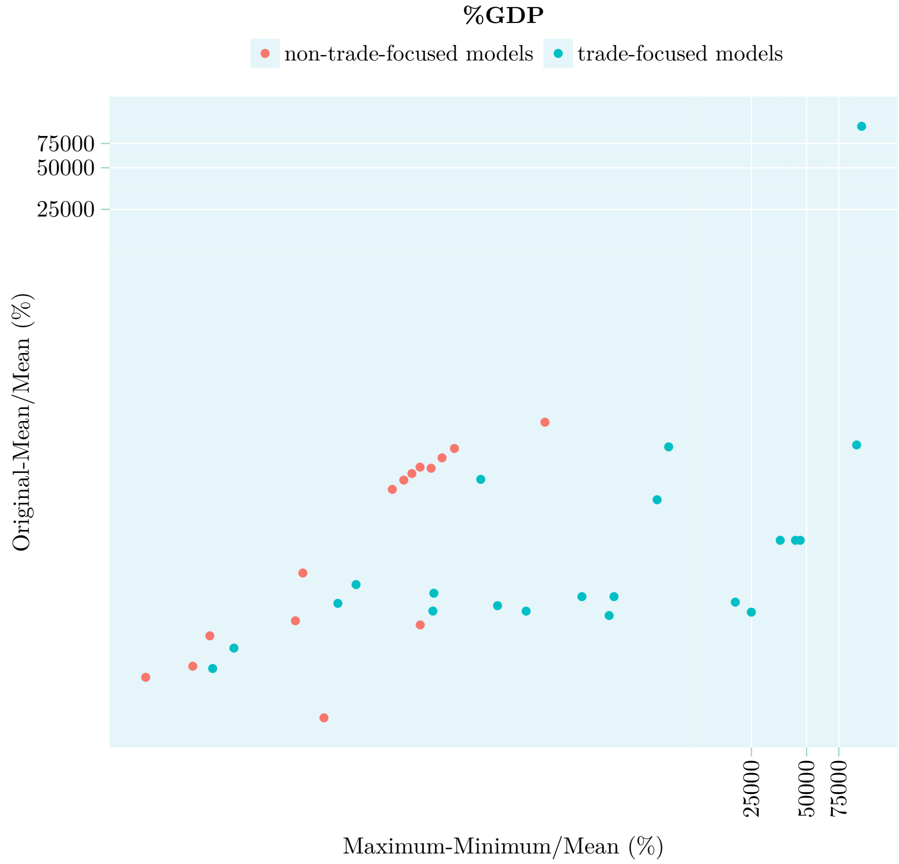
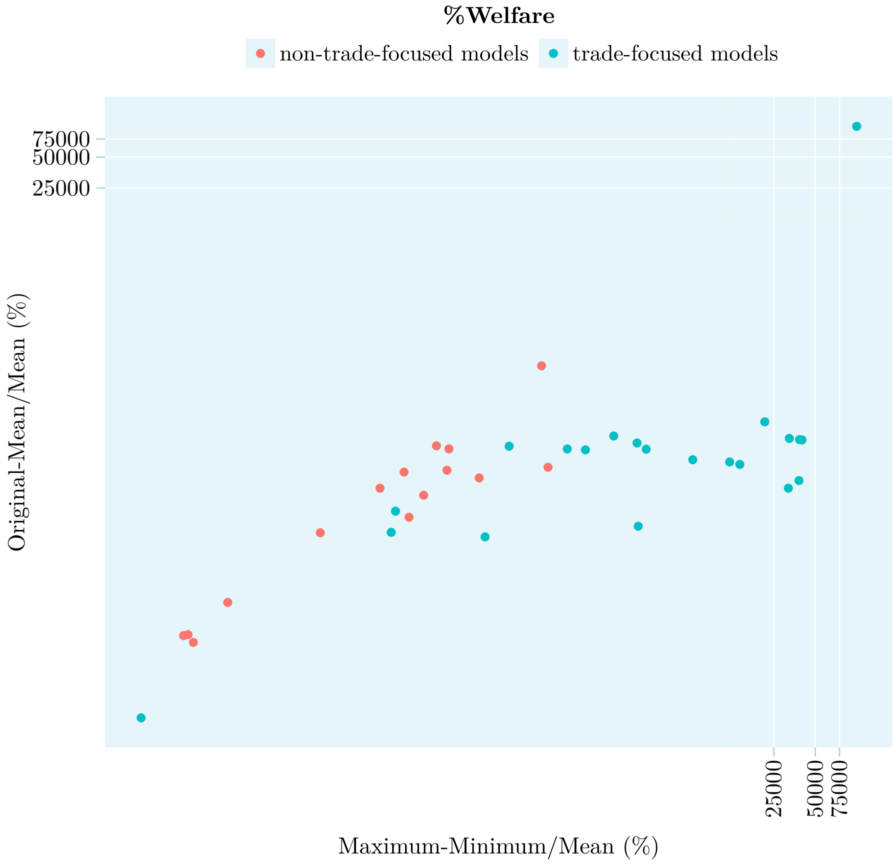


Figure 8: Welfare: Max-Min-criterion and Original-Mean-Criterion distinguished by application



5 Implications and future work

In brief, our answer to the question posed in the title is: "We should, indeed, care much more about Armington elasticities." We have shown in a small meta-study including 28 model simulations that altering the Armington elasticities by an important amount has a strong influence on model results in at least half of the cases. This conclusion applies to both increased elasticities in models with rather low initial ones and the opposite, lowering elasticities in models with initially high ones. The largest effect has been found in trade-focused models which initially used one single elasticity for all sectors. The results in our robustness checks showed such a high variance that even qualitative switches in the results occurred in many simulations. Hence, the same policy would be judged as either welfare increasing or decreasing depending on the policy set. This finding is in line with e.g. Valenzuela, Anderson, and Hertel (2007) who show that choosing a small elasticity may lead to an optimum-tariff-result in models which would find a tariff non-optimal in case of higher elasticities.

Our findings have severe implications for applied economic modelling especially in ex-ante simulations of planned trade reforms. We do explicitly not argue here that one or the other method of estimating the elasticities is the "right" one, however, we emphasize that elasticities matter. And modellers who use, for any reason, a specific set of elasticities should be aware that their results are only a local result conditional on this specific elasticity set. Hence, we call for more transparency concerning Armington elasticities. Model results, even for the same country, year and policy are not comparable if elasticities differ substantially. It is thus important for both the scientific community as well as for policymakers, who make choices based on model results, to know which elasticities have been used – and ideally also why these and not other ones.

It would be in the interest of both liability and comparability as well as good scientific practice that modellers include robustness checks as the ones shown in this paper per default into their articles and reports. Such a practice had a number of positive effects: a) it would be visible within which span of elasticities the results remain reliable, b) results would be more comparable across models, c) the problem would be quantified further, d) such a transparent approach would increase the trust in CGE models and acceptance of these as an important tool for economic policy analysis given that CGEs are at the moment much too often perceived as "'black boxes".

A broadened and deepened scientific discussion about Armington elasticities is necessary. Consistent estimations of Armington elasticities are still the best way to deal with the demonstrated non-robustness of model results. If one was sure that he or her chose the Armington elasticities wisely, even a strong influence of these on model results would not be worrying. However, we are simply not sure which is the correct way to estimate the elasticities and a big part of the modelling community is also completely ignorant of the problem per se.

This study is still work in progress. We aim at including as many models as possible and explicitly encourage modellers to contribute their models and data. We will have a closer look on more disaggregate results, which have so far not been included in this paper but are, as far as is visible at the moment, even more sensitive to changed elasticities. Especially the influence of the elasticity choice on income distribution seems worth a closer look. We have also completely focused on real variables so far even though also price variables are, of course, affected by the elasticity choice. We will also investigate further whether there are critical values for the elasticities i.e. alter the span of elasticities from which we draw the elasticity sets randomly. The meta-study presented here will later be complemented by an econometric study which estimates the Armington elasticities for a large group of countries and investigates which are the key determinants of Armington elasticities. This work as a whole will hopefully shed more light on both the influence of Armington elasticities and the correct way to chose these even if they are highly influential.

References

Balistreri, Edward J., Russell H. Hillberry, and Thomas F. Rutherford (2011). "Structural estimation and solution of international trade models with heterogeneous firms". In: *Journal of International Economics* 83.2, pp. 95–108. ISSN: 0022-1996. DOI: 10.1016/j.jinteco.2011.01.001.

Blonigen, Bruce A. and Wesley W. Wilson (1999). "Explaining Armington: What Determines Substitutability between Home and Foreign Goods?" In: *The Canadian Journal of Economics / Revue canadienne d'Economique* 32.1, pp. 1–21. ISSN: 0008-4085. DOI: 10.2307/136392.

Feenstra, Robert C., Maurice Obstfeld, and Katheryn N. Russ (2012). *In Search of the Armington Elasticity*. http://www.econ.ucdavis.edu/faculty/knruss/FOR_6-1-2012.pdf.

Frey, Miriam and Zoryana Olekseyuk (2014). "A general equilibrium evaluation of the fiscal costs of trade liberalization in Ukraine". en. In: *Empirica*, pp. 1–36. ISSN: 0340-8744, 1573-6911. DOI: 10.1007/s10663-014-9249-z.

Gallaway, Michael P., Christine A. McDaniel, and Sandra A. Rivera (2003). "Short-Run and Long-Run Industry-Level Estimates of U.S. Armington Elasticities". In: *North American Journal of Economics and Finance* 14.1.

Gibson, Katherine Lee (2003). "Armington Elasticities for South Africa: Long- and Short-Run Industry Level Estimates". In: *TIPS Working Paper* 2003.12.

Hertel, Thomas et al. (2004). *How Confident Can We Be in CGE-Based Assessments of Free Trade Agreements?* NBER Working Paper 10477. National Bureau of Economic Research, Inc.

Hillberry, Russell and David Hummels (2013). "Chapter 18 - Trade Elasticity Parameters for a Computable General Equilibrium Model". In: *Handbook of Computable General Equilibrium Modeling*. Ed. by Peter B. Dixon and Dale W. Jorgenson. Vol. 1. Handbook of Computable General Equilibrium Modeling SET, Vols. 1A and 1B. Elsevier, pp. 1213–1269.

Jensen, Jesper and David G. Tarr (2012). "Deep Trade Policy Options for Armenia: The Importance of Trade Facilitation, Services and Standards Liberalization". en. In: *Economics: The Open-Access, Open-Assessment E-Journal* 6.2012-1, p. 1. ISSN: 1864-6042. DOI: 10.5018/economics-ejournal.ja.2012-1.

Kapuscinski, Cezary A. and Peter G. Warr (1996). *Estimation of Armington Elasticities: An Application to the Philippines*. Departmental Working Paper 1996-08. The Australian National University, Arndt-Corden Department of Economics.

McDaniel, Christine A. and Edward J. Balistreri (2003). "A review of Armington trade substitution elasticities". In: *Economie Internationale* 2003.2, pp. 301–313.

Németh, Gabriella, László Szabó, and Juan-Carlos Ciscar (2011). "Estimation of Armington elasticities in a CGE economy–energy–environment model for Europe".

In: *Economic Modelling* 28.4, pp. 1993–1999. ISSN: 0264-9993. DOI: 10.1016/j.econmod.2011.03.032.

Reinert, Kenneth A. and David W. Roland-Holst (1992). “Armington elasticities for United States manufacturing sectors”. In: *Journal of Policy Modeling* 14.5, pp. 631–639. ISSN: 0161-8938. DOI: 10.1016/0161-8938(92)90033-9.

Saito, Mika (2004). “Armington Elasticities in Intermediate Inputs Trade: A Problem in Using Multilateral Trade Data”. In: *The Canadian Journal of Economics / Revue canadienne d’Economique* 37.4, pp. 1097–1117. ISSN: 0008-4085.

Schuerenberg-Frosch, Hannah (2014). “How to Model a Child in School? A Dynamic Macrosimulation Study for Tanzania”. en. In: *South African Journal of Economics*, n/a–n/a. ISSN: 1813-6982. DOI: 10.1111/saje.12042.

Shiells, Clinton R. and Kenneth A. Reinert (1993). “Armington Models and Terms-of-Trade Effects: Some Econometric Evidence for North America”. In: *The Canadian Journal of Economics / Revue canadienne d’Economique* 26.2, pp. 299–316. ISSN: 0008-4085. DOI: 10.2307/135909.

Valenzuela, Ernesto, Kym Anderson, and Thomas Hertel (2007). *Impacts of Trade Reform: Sensitivity of Model Results to Key Assumptions*. Centre for International Economic Studies Working Paper 2007-09. University of Adelaide, Centre for International Economic Studies.

Welsch, Heinz (2006). “Armington elasticities and induced intra-industry specialization: The case of France, 1970–1997”. In: *Economic Modelling* 23.3, pp. 556–567. ISSN: 0264-9993. DOI: 10.1016/j.econmod.2006.02.008.

– (2008). “Armington elasticities for energy policy modeling: Evidence from four European countries”. In: *Energy Economics* 30.5, pp. 2252–2264.

6 Appendix

Table 1: Model and simulation overview

Model Index	Country	Author	Policy Simulation	Original Elasticities
0002	Peru	T. Rutherford	Tariff abolition	5 for all sect
0005a	Zambia	H.Schuerenberg-Frosch	Increase in public budget by 2% of GDP	around 1 based on lit.
0005b	Zambia	H.Schuerenberg-Frosch	increase in public capital investment by 5-100%, no productivity effect	around 1 based on lit.
0005c	Zambia	H.Schuerenberg-Frosch	increase in public capital investment by 5-100%, low productivity effect	around 1 based on lit.
0005d	Zambia	H.Schuerenberg-Frosch	increase in public capital investment by 5-100%, high productivity effect	around 1 based on lit.
0005e	Zambia	H.Schuerenberg-Frosch	increase in public consumption by 5-100%	around 1 based on lit.
0005f	Zambia	H.Schuerenberg-Frosch	increase in remittances by 2% of GDP	around 1 based on lit.
0005g	Zambia	H.Schuerenberg-Frosch	increase in FDI by 2% of GDP	around 1 based on lit.
0006	Zambia	H.Schuerenberg-Frosch	Investment in road infrastructure	around 1, based on lit.
0007a	Finland	H. Torma, T. Rutherford	"Consistent Tax Reform" multiple ways to hold budget constant	2 for all sect.
0007b	Finland	H. Torma, T. Rutherford	"Consistent Tax Reform" multiple ways to hold budget constant	2 for all sect.
0007c	Finland	H. Torma, T. Rutherford	"Consistent Tax Reform" multiple ways to hold budget constant	2 for all sect.
0007d	Finland	H. Torma, T. Rutherford	"Consistent Tax Reform" multiple ways to hold budget constant	2 for all sect.
0008	Cameroon	T. Condon, H. Dahl, S. Devarajan	uniform tariff	slightly below or around 1
0011a	UK, EC, CW	M. H. Miller, J. E. Spencer	economic effects of UK joining the EEC, Integration without transfer	3 for all sect.
0011b	UK, EC, CW	M. H. Miller, J. E. Spencer	economic effects of UK joining the EEC, Integration with tax transfer	3 for all sect.
0011c	UK, EC, CW	M. H. Miller, J. E. Spencer	global free trade	3 for all sect.
0018	Ukraine	Z. Olekseyuk, M. Frey	Import tariff abolition for EU-goods	around 5 based on lit.
0028	Ukraine	F. Pavel, V. Movchan	Import tariff reduction by 50%	around 5 based on lit.

Table 2: Result summary GDP

Scen#	Original result	Max result	Min result	Mean	Median	Dev. Max-Min	Dev. Original-Mean	Std. Dev.	Qualitative change
1	-0.0197	0.3457	-66.828	-0.0282	-0.0195	249.564.664	301.444	0.2152	Yes
2	-0.6859	-0.1055	-0.6859	-0.1839	-0.173	3.156.437	2.729.971	0.0497	No
3	10.088	43.569	-22.679	14.563	14.267	4.548.972	307.283	0.2932	Yes
4	-18.512	-0.7194	-101.693	-2.547	-23.702	371.022	273.169	10.224	No
5	317.145	397.431	317.145	361.403	364.326	222.151	122.463	1.686	No
6	-0.3319	66.105	-39.789	-0.2387	-0.2024	44.366.244	390.734	0.2778	Yes
7	0.7544	0.7544	0.103	0.1865	0.1757	3.493.562	3.045.951	0.0516	No
8	-0.371	-0.0674	-0.371	-0.1111	-0.1049	2.732.809	2.339.568	0.0278	No
9	-0.3336	0.5975	-31.552	-0.2553	-0.2004	14.701.859	306.799	0.1824	Yes
10	14.253	23.896	0.6355	15.209	01.05	1.153.351	62.887	0.3605	No
11	-0.8677	83.898	-49.293	-0.1509	-0.3064	88.292.582	475.213	0.6795	Yes
12	-0.0188	0.2695	-56.916	-0.0292	-0.0194	204.073.556	356.287	0.182	Yes
13	0.1718	0.5908	0.021	0.329	0.3278	1.731.949	47.775	0.1008	No
14	0.2254	122.107	-0.9509	0.3153	0.2863	41.749.356	28.504	0.4068	Yes
15	27.237	206.049	-0.7865	46.438	0.451	4.606.414	413.482	17.062	Yes
16	-0.2092	0.1961	-0.2808	-0.0256	-0.0343	18.627.358	7.169.836	0.0652	Yes
17	0.8769	0.8769	-0.0532	0.1558	0.1445	5.969.567	4.627.934	0.0863	Yes
18	20.171	67.048	-244.632	30.373	30.659	10.261.744	335.904	14.006	Yes
19	0.8854	0.8854	0.1043	0.2016	0.1858	387.38	339.1	0.0643	No
20	0.1679	81.187	-0.0525	0.2756	0.2645	29.648.877	390.656	0.2715	Yes
21	29.069	71.472	0.9918	44.671	42.369	1.377.953	349.264	10.974	No
22	-0.3275	0.0094	-0.3275	-0.0757	-0.0706	445.177	3.327.201	0.0352	Yes
23	0.4423	10.248	-10.575	-0.2506	-0.2767	8.309.353	2.764.887	0.2546	Yes
24	41.588	131.814	41.588	99.412	10.195	907.597	581.663	1.753	No
25	-11.233	10.284	-278.595	-0.3786	-0.4221	76.308.671	1.967.359	11.349	Yes
26	-3.414	-3.414	-65.215	-4.646	-45.583	668.876	265.174	0.53	No

Table 3: Result summary welfare

Scen#	Original result	Max result	Min result	Mean	Median	Dev. Max-Min	Dev. Original-Mean	Std. Dev.	Qualitative change
1	-0.0233	0.3205	-6.2886	-0.0467	-0.0349	14148.1372	50.193	0.2025	Yes
2	45.0751	45.0751	43.8885	44.0798	44.0618	2.6917	2.2578	0.1063	No
3	0.193	3.5142	-3.2882	0.636	0.6069	1069.5065	69.6562	0.2938	Yes
4	0.3671	1.5946	-2.4079	0.6544	0.7124	611.6121	43.9042	0.4271	Yes
5	38.1162	55.3104	30.1572	45.0162	45.3665	55.8759	15.3279	4.2162	No
6	0.0397	6.5155	-3.255	0.0306	0.0294	31891.8583	29.4488	0.239	Yes
7	7.253	8.2832	7.253	8.1321	8.1468	12.6684	10.8108	0.0853	No
8	46.8603	46.8603	46.2642	46.3633	46.3537	1.2856	1.0718	0.0536	No
9	0.0374	1.0017	-2.4815	0.0162	0.0321	21484.4869	130.7157	0.1161	Yes
10	-0.3938	-0.2075	-1.3396	-0.6217	-0.6108	182.0875	36.6635	0.2417	No
11	0.0353	8.8476	-3.5312	0.7212	0.6427	1716.4028	95.1104	0.6119	Yes
12	-0.0222	0.2595	-5.3871	-0.0473	-0.0357	11938.9195	53.0449	0.1721	Yes
13	1.0028	1.0076	1.0012	1.0045	1.0045	0.6322	0.1691	0.0012	No
14	0.4124	11.5093	-0.6793	0.4714	0.4467	2585.4997	12.523	0.3735	Yes
15	0.7709	18.3061	-2.6682	2.6591	2.5278	788.7644	71.011	1.6736	Yes
16	46.8579	46.8579	46.1544	46.4304	46.428	1.5153	0.9208	0.0835	No
17	1.5282	2.8896	1.5282	2.6462	2.6652	51.4479	42.2476	0.116	No
18	0.2334	5.5904	-25.7832	1.2371	1.2645	2536.0113	81.1348	1.3783	Yes
19	2.4966	3.7132	2.4966	3.537	3.5516	34.399	29.4151	0.0996	No
20	0.0537	7.6374	-0.1858	0.1215	0.1172	6438.8501	55.7786	0.2557	Yes
21	0.492	4.632	-1.3789	2.0148	1.7901	298.3309	75.5793	1.0717	Yes
22	46.9481	46.9481	46.3043	46.4387	46.4289	1.3864	1.097	0.0606	No
23	1.5655	3.0147	-0.4474	1.7365	1.7925	199.3747	9.8463	0.4759	Yes
24	1.2218	5.9656	1.2218	4.258	4.3028	111.4117	71.3065	0.7038	No
25	-1.6272	0.2674	-27.8884	-0.9534	-0.9788	2953.3504	70.6796	1.0726	Yes
26	-0.7764	-0.7162	-2.135	-1.3952	-1.3791	101.6882	44.3563	0.2701	No