Schürenberg-Frosch, Hannah

Working Paper

We could not care less about Armington elasticities - but should we? A meta-sensitivity analysis of the influence of Armington elasticity misspecification on simulation results

Ruhr Economic Papers, No. 594

Provided in Cooperation with:
RWI - Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Schürenberg-Frosch, Hannah (2015) : We could not care less about Armington elasticities - but should we? A meta-sensitivity analysis of the influence of Armington elasticity misspecification on simulation results, Ruhr Economic Papers, No. 594, ISBN 978-3-86788-689-5, http://dx.doi.org/10.4419/86788689

RUHR

ECONOMIC PAPERS

Hannah Schürenberg-Frosch

We Could not Care Less about Armington Elasticities – But Should We?

A Meta-Sensitivity Analysis of the Influence of Armington Elasticity Misspecification on Simulation Results

UNIVERSITÄT DUISBURG ESSEN

#594

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #594

Responsible Editor: Volker Clausen

ISSN 1864-4872 (online) – ISBN 978-3-86788-689-5

Ruhr Economic Papers #594

Hannah Schürenberg-Frosch

We Could not Care Less about Armington Elasticities – But Should We?

A Meta-Sensitivity Analysis of the Influence of Armington Elasticity Misspecification on Simulation Results

UNIVERSITÄT DUISBURG ESSEN

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788689
ISSN 1864-4872 (online)
ISBN 978-3-86788-689-5

Hannah Schürenberg-Frosch[1]

We Could not Care Less about Armington Elasticities – But Should We?

A Meta-Sensitivity Analysis of the Influence of Armington Elasticity Misspecification on Simulation Results

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. We resimulate 50 published CGE policy simulations each a 1000 times and randomize the elasticities. The results of this experiment clearly indicate that a change in the elasticities has noteworthy quantitative and qualitative effects on the results in more than half of the models. We thus conclude that the choice of the elasticities should get more attention and robustness of models with respect to elasticities should be tested by modellers.

JEL Classification: F14, C68, F17

Keywords: Armington; trade elasticities; computable general equilibrium; meta-study

November 2015

1 Hannah Schürenberg-Frosch, UDE. – The author thanks Zoryana Olekseyuk, Volker Clausen and Edward J. Balistreri as well as seminar and conference participants in Essen, Melbourne and Paris for helpful discussions and suggestions. Mehdi Belayet Lincon provided the best possible research assistance. The usual disclaimer applies.– All correspondence to: Hannah Schürenberg-Frosch, University of Duisburg-Essen, Chair of International Economics, Institute for Economics and Business Administration, Universitätsstr. 12, 45117 Essen, Germany, 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 from different countries even though the price for the domestic and foreign varieties is not equal. 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 country of origin. 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 offer 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.[1]

Given the large dissemination of Armington-type models, it is interesting to investigate the influence of this particular trade specification 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.[2] The correct choice of this elasticity will decide on the accuracy of model results. However, until recently the Armington elasticities have not received much attention in the literature. 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 by changing the Armington elasticities even a qualitative switch in the overall welfare result is possible. Schürenberg-Frosch (2015) 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 chosen elasticity set. A similar approach is used by Frey and Olekseyuk (2014) and Jensen and Tarr (2012) with comparable results. Apart from this, 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 misspecification in the un-

[1] See Balistreri et al. (2011) for a detailed discussion of the different trade specifications and their relation to each other.

[2] 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.

derlying econometric method and that the elasticities should be much higher.

Given these recent concerns about the appropriate choice of 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 worrisome 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 areas of application. 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.[3] Subsequently, we compare the range 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.[4] Each of these variables is investigated with respect to these measures for robustness: spread between maximum and minimum (in % relative to the mean), qualitative switches (yes/no), deviation of original model results from robustness simulation mean (in % relative to the mean). 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 qualitative switches are found in less than 1% of the cases.

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 360% below the mean result, the maximum result up to 400% above the mean.[5] The original model results deviate in many models by substantially more than 10%, often more than 50% from the mean obtained in our random-

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

[4]We also report household- and sector-specific results but these are not included in the paper for brevity.

[5]One model shows no convergence in a high number of scenarios and a deviation of up to 99800%

ized simulations. The most extreme deviation is above 400% for the change in real welfare and also 400% for the change in real GDP. In many cases the original simulation results lie 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 as 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 worrisome. 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.[6] 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 chosen elasticities and respective 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 counterintuitive. 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 different possible methodologies and includes as many countries as possible would be very helpful to improve the reliability of model results.

2 Related research

The *correct* size of Armington elasticities has been disputed ever since they 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 their estimation.

Since the 1970s, several studies with estimated Armington elasticities have been published.[7] 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 long run com-

[6] Though, the right strategy to estimate the elasticities is also subject to discussion in the community

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

pared to the short run. This last result is found both by altering the frequency 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 et al. (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 those for the US (further examples are Gibson, 2003; Kapuscinski and Warr, 1999; Olekseyuk and Schürenberg-Frosch, 2014; Welsch, 2006). 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.[8] Only very few studies use a nested approach and estimate both at the same time.[9] This will be 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 than the macro elasticities.

Third, the estimated elasticities increase with the level of disaggregation 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 similarity 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

[8]This distinction and denomination is adopted from Feenstra et al. (2014).

[9]It is thus rather worrying if these two approaches find comparable results even though the underlying structural model parameter in question is a different one.

estimated elasticities are to be used in a CGE model, the problem is somewhat more complex. The level of aggregation in the data used for 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.

Fourth, more recent studies tend to find much higher elasticities compared to the first generation of studies – which we term in this paper as *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, underlying model, econometric procedure, underlying data and interpretation, from those 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 et al. (2003) and Blonigen and Wilson (1999) for the US, Kapuscinski and Warr (1999) for the Philippines, Gibson (2003) for South Africa and Welsch (2006) for France. Recent studies such as Saito (2004), Welsch (2008) and Németh et al. (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 among 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 et al. (2011) again identify variation over time and importing country but follow a nested procedure for the 'macro' and 'micro' elasticity.

Fifth, and most important: Hillberry and Hummels (2013) in line with Hertel et al. (2004) and Valenzuela et al. (2008) 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 therefore not appropriate. They propose a different strategy of estimation using a cross-section of trading partners and identifying a variation in prices across trading partners instead of over time (or over importing country as other cross-sectional and panel studies in the field) 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. The results in this branch of the literature lie substantially 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. They argue, in addition, that time series studies based on the use of unit prices are biased towards small elasticities. It has been demonstrated by Valenzuela et al. (2008) that elasticities around unity mistakenly lead to the identification of optimal tariffs in CGE model applications. As the estimations by Hillberry and Hummels (2013) and Hertel et al. (2004) reach a dimension being between 4 and 18 times higher than 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.[10]

3 Methodology

Our analysis investigates 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 substantial change in trade preferences regarding the substitutability of goods 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 and the *true* size of the elasticities is disputed. Results of simulations based on uncertain elasticity sets would have the same degree of uncertainty as the estimations. In addition, CGE models are not only used for trade policy analysis. Thus, if Armington

[10]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.

elasticities even impact on model results if trade policy is unchanged, the elasticity effect might distort the results of the actual policy experiment in case that the chosen elasticities are wrong whatever reason. We do not argue that models should in general not 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 in the model.

3.1 Data

This study comprises 50 original simulations within 16 models for 21 different countries.[11] 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 **??** in the appendix.

Most 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 sector-specific elasticities. This usage of one elasticity for all sectors may stem from computational limitations in older models. The degree of 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.[12] 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 aid 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.

[11] We are open to investigate further new models and are very grateful for any model provided by fellow researchers. If you are interested in contributing to this project please contact the author for details.

[12] 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 investigate whether the included models are robust with regard to the choice of elasticities or whether the influence of the Armington elasticities is noteworthy, we resimulate the original policy experiments 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[13] 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 Hillberry and Hummels, 2013).[14] As most of the models have their original elasticities around unity, which has traditionally been the case, our robustness check means for most of them a strong upward shift in the elasticity set. This is comparable to changing a model with older traditional elasticities to a newer elasticity set.

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.[15]

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 in our study
2. The deviation of the original result from the mean of our robustness test results in % relative to the mean in our study
3. The share of model runs in % in which 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 orig-

[13] 0.5 for NLP models due to the mathematical limits in the CES function in NLP-syntax.

[14] Note, the interval is indeed a large one but we did not want to exclude any of the elasticities present in the literature. The expected value of the elasticity in our draw is 9, which is in line with the average elasticity in widespread models such as the GTAP model. The broad interval has been chosen for the meta study, authors who want to test their own models might adjust the interval to their specific case, i.e. to the spread of elasticity estimations relevant for the specific application.

[15] Available on request.

inal result deviates by more than 25% from the mean in our simulations or if a qualitative switch occurs in our simulations.[16]

3.3 Implementation

As the models 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 differ and hence using the exactly same code for 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.[17]

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 a 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 **??** – **??** in the appendix.[18] In general, we find that most models are not robust by our criteria. Welfare

[16]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*.

[17]The testing code is made available on request.

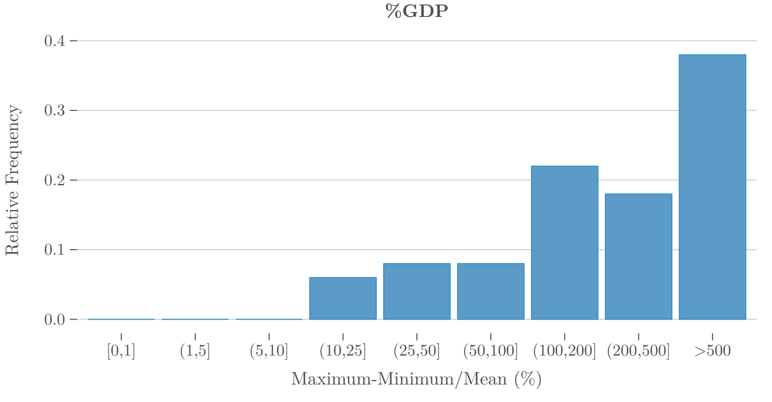
[18]We do not link the results to specific models here as we leave it free to the model authors to publish the results for their respective models. The randomly attributed scenario numbers in the graphs shown in this section correspond to those in the result tables, but not to those in the model overview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

results are more robust than GDP results and trade is even less robust - which is intuitive as trade effects are direct whereas GDP and welfare effects are indirect effects.[19]

4.1 Maximum-Minimum Spread

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

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



This criterion has emerged to be the strictest one which is only met by seven simulations for GDP and by 13 simulations (out of 50) for welfare. Even if we would define our criterion even less strong and allow for a variation of up to 100% around the mean, 39 simulations for GDP and 32 simulations for welfare would not fulfil this criterion.

The lowest maximum-minimum spread lies at 12.3% of the mean. 10% of our simulations have a spread below or equal to 36% of the mean, 20% lie below 89% of the mean. However, three quarters of our simulations show a spread between maximum and minimum result of more than 100% of the mean, more

[19]Trade results are not shown in the following graphs for brevity. Trade reacts always as expected.

than 50% lie above 300%. Even after excluding the simulations which did not show convergence in a noteworthy share of the simulations, we end up with more than half of the models being not considered as robust by this criterion.

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

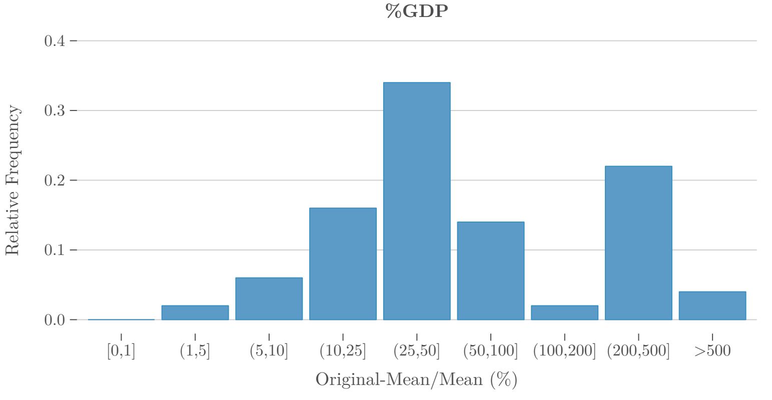
Welfare results vary far less than GDP results. Even though there exist, also here, some very high spreads between maximum and minimum in some models, the majority of models produces more robust welfare results compared to GDP. 25% of all simulations show a maximum-minimum spread for welfare below 50% of the mean, 40% of the results lie below 100%. However, there exist also some simulations with a spread above 200%. Even excluding an outlier model which does not converge in many of our simulations, still in quite a number of simulations in other models the Maximum-Minimum spread is more than 1000% relative to the mean. We still consider this worrying even though it is – of course – much less extreme than the GDP results.

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 12 out of the 50 simulations for the change in real GDP and by 19 out of the 50 simulations for the 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 (especially concerning GDP) if elasticities are changed substantially.

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



The highest deviation between the original results and the mean in the robustness tests, once again excluding non-convergent models from the results, is at 716% for GDP and at 460% for welfare, meaning that the GDP effect in the mean simulation in our study is more than 7 times as high as the result in the original study and 460% for welfare respectively. Roughly 30% of our simulations show deviation of the original result from the mean by not more than 25%, 55% of the simulations lie below 50% original-mean deviation. This is, indeed, a more positive picture compared to the first criterion. Nonetheless, a majority or at least half of the models are still to be considered not robust, also by this criterion. Again, this is worrisome as a strong misjudgement of the GDP effect of the simulated policy is possible.

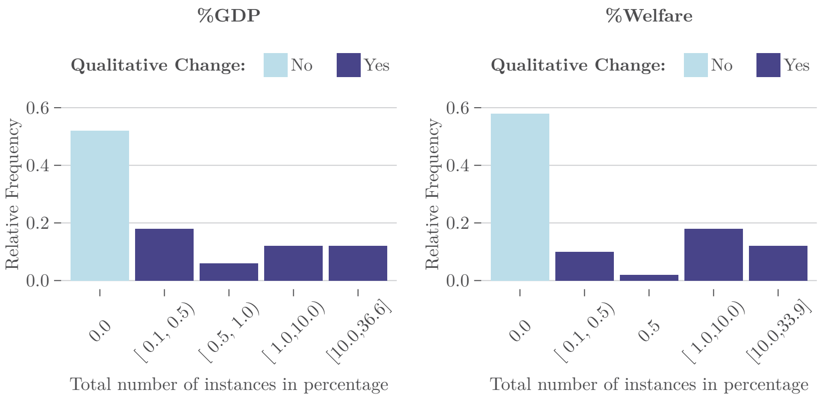
Figure 4: Hicks equivalent change in private welfare: Deviation of original result from mean %

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 the elasticity specification. It will consider a model as not robust if in more than 1% of the 1000 robustness simulations the result for either GDP or welfare has the opposite mathematical sign from the majority of the simulations. In other words, if the 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.

Figure 5: Qualitative switches in GDP and in welfare results



This criterion is not met by 12 of the 50 simulations for GDP and 15 for welfare which means that about one quarter of the simulations are clearly to be considered as not robust if elasticities are changed within a wide range. As figure **??** shows, there is a considerable number of models where qualitative switches occur in more than 10% of the cases. 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 models included are highly affected. However, the good news is that for this very lax criterion the majority of the models is robust, at least.

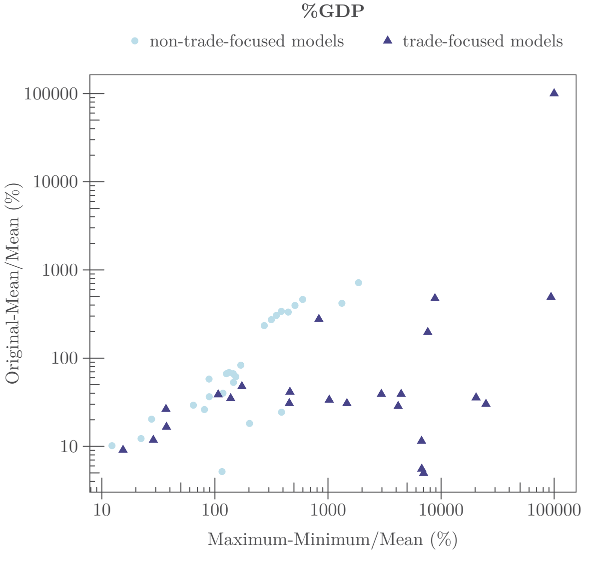
4.4 Relationship between robustness and application scenario

One might argue that it is completely intuitive and even desirable that a crucial change in trade preferences creates noteworthy effects. However, we 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 expect that trade policy scenarios might in general be more sensitive to changes in the elasticities compared to scenarios which only include national policies like tax reform or public spending scenarios. Hence, in figure **??** we present the results for GDP separately for trade and non-trade simulations. Figure **??** shows the same distinction for welfare effects.

It is 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.

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 set. This applies both to GDP as well as to welfare effects. Hence, in simulations in which 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.

Figure 6: GDP: Max-Min-Criterion and Original-Mean-Criterion distinguished by application

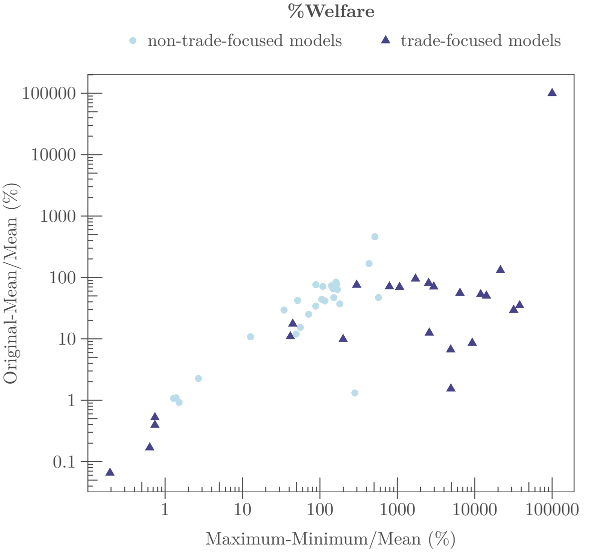


The same conclusion also applies to the *Qualitative-Switch-Criterion* where most of the models with qualitative switches are, indeed, trade models.

4.5 Relationship between robustness and base year

As noted above, the rather low elasticities found in the traditional time series estimations have mostly been published in the 1980s and 1990s. Hence, one might suspect that more recent models use higher elasticities and - as it is

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



argued that the low elasticities lead to a bias in model results - and are more robust. However, we cannot clearly confirm this reasoning. While we clearly see a trend towards more disaggregate and more complex models with more sectors and sector-specific elasticities, still, also in more recent models low elasticities around unity prevail. We have disaggregated our results by base year and initial elasticity and the results in figures **??** to **??** do not show a clear picture.

For GDP we see that the older models show a somewhat higher spread between maximum and minimum. However, the deviation between original result and mean is highest in younger models. For welfare, the picture is somewhat clearer with a tendency of older models to be less robust. This, however, is not completely due to lower elasticities. Figures **??** and **??** show that while models with lower elasticities produce a larger spread between maximum and minimum for GDP, the original-mean deviation is higher in models with higher elasticities. Furthermore, for welfare the picture is somewhat mixed. Hence, even though we find a tendency of older models, or models with lower elasticities, to be less robust, there is a considerable number of observations contradicting this.

Figure 8: GDP: Max-Min-Criterion and Original-Mean-Criterion distinguished by base year

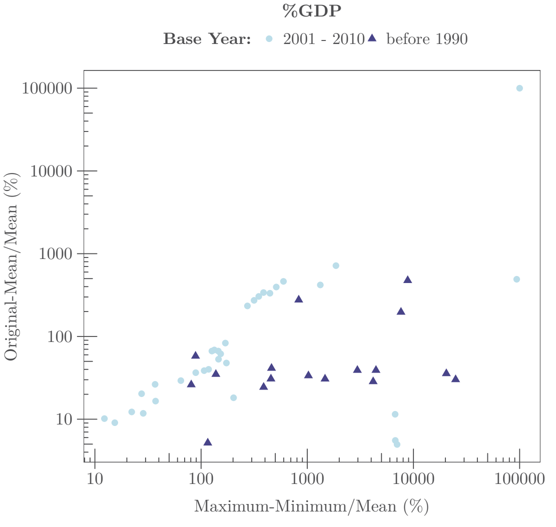


Figure 9: Welfare: Max-Min-Criterion and Original-Mean-Criterion distinguished by base year

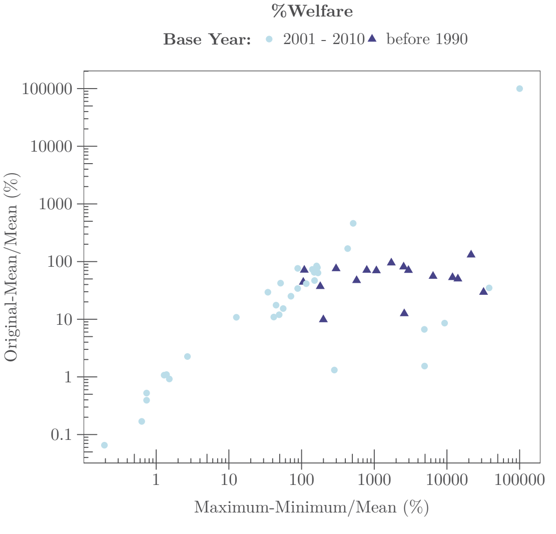


Figure 10: GDP: Max-Min-Criterion and Original-Mean-Criterion distinguished by initial elasticity

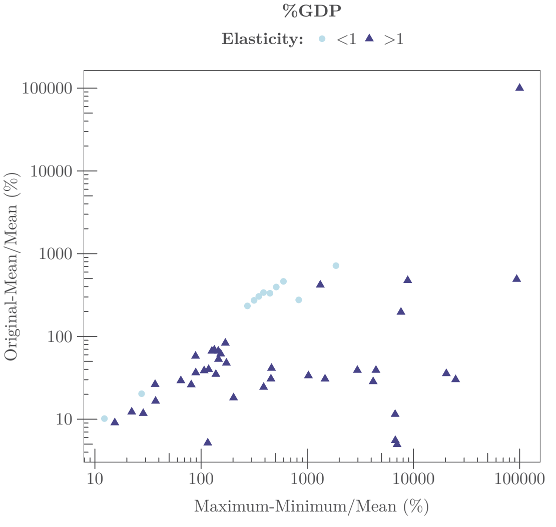
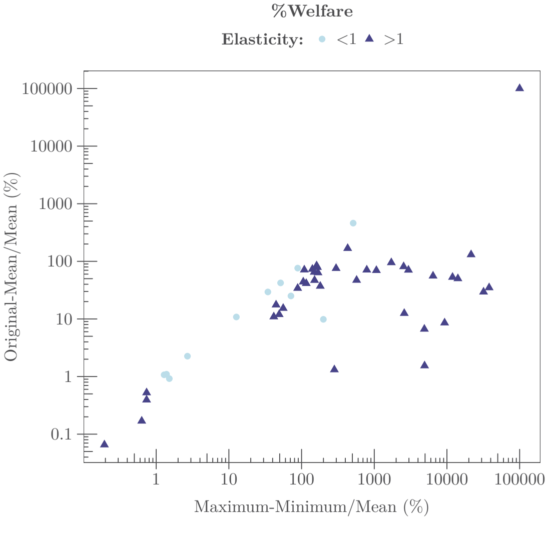


Figure 11: Welfare: Max-Min-Criterion and Original-Mean-Criterion distinguished by initial elasticity



5 Implications and future research

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 50 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. Older models with lower elasticities seem to be slightly more affected. The results in our robustness checks showed such a high variance that even qualitative switches in the results occurred in a noteworthy number of simulations. Hence, the same policy would be judged as either welfare increasing or decreasing depending on the elasticity set. This finding is in line with e.g. Valenzuela et al. (2008) who show that choosing a small elasticity may lead to an optimum-tariff-result in models which would find a tariff to be non-optimal in case of higher elasticities.

Our findings have strong 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 do 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 the role of 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 credibility 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 would have 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 the size and role of 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 about the correct way to estimate the elasticities and a considerable part of the CGE modelling community is also completely ignorant of the problem per se.

As an extension of this study we will have a closer look at more disaggregate results, which have not been included in this paper but are, as far as it 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, a point which is in line with the influential work of Arkolakis et al. (2012). We have also completely focused on real variables here even though also price variables are, of course, affected by the elasticity choice, as has been shown by Schürenberg-Frosch (2015). We will also investigate further whether there are critical values for the elasticities, i.e., alter the interval 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 the key determinants of Armington elasticities. This research program 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

Arkolakis, Costas, Arnaud Costinot, and Andres Rodriguez-Clare (2012). “New Trade Models, Same Old Gains?” In: *American Economic Review* 102.1, pp. 94–130.

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.

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.

Feenstra, Robert C. et al. (2014). *In Search of the Armington Elasticity*. NBER Working Paper 20063. National Bureau of Economic Research, Inc.

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.

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”. In: *Economics: The Open-Access, Open-Assessment E-Journal* 6.2012-1, p. 1.

Kapuscinski, Cezary A. and Peter G. Warr (1999). “Estimation of Armington elasticities: an application to the Philippines”. In: *Economic Modelling* 16.2, pp. 257–278.

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.

Olekseyuk, Zoryana and Hannah Schürenberg-Frosch (2014). *Are Armington Elasticities Different Across Countries and Sectors?* Ruhr Economic Papers 513. Essen: RWI Essen.

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.

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.

Schürenberg-Frosch, Hannah (2015). "How to Model a Child in School? A Dynamic Macrosimulation Study for Tanzania". In: *South African Journal of Economics* 83.1, pp. 117–139.

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.

Valenzuela, Ernesto, Kym Anderson, and Thomas Hertel (2008). "Impacts of trade reform: sensitivity of model results to key assumptions". In: *International Economics and Economic Policy* 4.4, pp. 395–420.

Welsch, Heinz (2006). "Armington elasticities and induced intra-industry specialization: The case of France, 1970–1997". In: *Economic Modelling* 23.3, pp. 556–567.

– (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	model type	Policy Simulation	Original Elasticities
1a	Tanzania	MPSGE, single-country, static	Increase in production and export of cotton and clothing	2 for all sect
1b	Tanzania	MPSGE, single-country, static	Increase in production and export of tea and coffee	2 for all sect
1c	Tanzania	MPSGE, single-country, static	Increase in production and export of tobacco	2 for all sect
2	Peru	MPSGE, single-country, static	Tariff abolition	5 for all sect
3	Zambia	MPSGE, single-country, static	Increase in public budget by 2% of GDP	around 1 based on lit.
4a	Zambia	MPSGE, single-country, static	Increase in public capital investment by 5-100%, no productivity effect	around 1 based on lit.
4b	Zambia	MPSGE, single-country, static	Increase in public capital investment by 5-100%, low productivity effect	around 1 based on lit.
4c	Zambia	MPSGE, single-country, static	Increase in public capital investment by 5-100%, high productivity effect	around 1 based on lit.
5	Zambia	MPSGE, single-country, static	Increase in public consumption by 5-100%	around 1 based on lit.
6	Zambia	MPSGE, single-country, static	Increase in remittances by 2% of GDP	around 1 based on lit.
7	Zambia	MPSGE, single-country, static	Increase in FDI by 2% of GDP	around 1 based on lit.
8	Zambia	MPSGE, single-country, recursive-dynamic	Investment in road infrastructure	around 1, based on lit.
9	Finland	MPSGE, single-country, static	Budget-neutral tax-reform	2 for all sect.

Table 1: Model and simulation overview

Model Index	Country	model type	Policy Simulation	Original Elasticities
10	Cameroon	MPSGE, single-country, static	Uniform tariff	slightly below or around 1
11a	UK, EC, CW	MCP, four-country, static	Economic effects of UK joining the EEC, Integration without transfer	3 for all sect.
11b	UK, EC, CW	MCP, four-country, static	Economic effects of UK joining the EEC, Integration with tax transfer	3 for all sect.
11c	UK, EC, CW	MCP, four-country, static	Global free trade	3 for all sect.
12	Uganda	NLP, single country, recursive-dynamic	Consequences of aid-funded public expenditure programs	0.5 for all sect.
13	Brazil, Germany, Ghana, India, South Korea, Netherlands, Poland, United States, Tanzania, Saudi Arabia	MPSGE, world model with 12 regions	Increase in European producer tax	around 5
14	Ukraine	MPSGE, single-country, static	Import tariff abolition for EU-goods	around 5
15	Ukraine	MPSGE, single-country, static	Import tariff reduction by 50%	around 5
16	Ukraine	MPSGE, single-country, static	Import tariff reduction by 50%	around 5

Table 2: Result summary GDP

Scen#	Original result	Max result	Min result	Mean	Median	Dev. Max-Min	Dev. Original-Mean	Std. Dev.	Qualitative change	Qualitative change(%)
1	0.7544	0.7544	0.1030	0.1865	0.1757	349.3562	304.5951	0.0515	No	0.0
2	3.0592	3.4778	3.0592	3.4056	3.4332	12.2898	10.1709	0.0743	No	0.0
3	4.0384	175.5323	-94.3497	3.8478	4.0733	7,013.9839	4.9537	11.2564	Yes	4.4
4	-0.3710	-0.0674	-0.3710	-0.1111	-0.1049	273.2809	233.9568	0.0278	No	0.0
5	10.7196	10.7196	9.2067	9.8278	9.8212	15.3938	9.0744	0.2238	No	0.0
6	-0.0197	0.3457	-6.6828	-0.0282	-0.0195	24,956.4664	30.1444	0.2151	Yes	0.5
7	-0.3527	-0.2156	-1.5510	-1.0564	-1.1228	126.4042	66.6128	0.3187	No	0.0
8	2.9069	7.1472	0.9918	4.4671	4.2369	137.7953	34.9264	1.0968	No	0.0
9	0.8484	1.1415	0.8484	1.0648	1.0871	27.5264	20.3214	0.0590	No	0.0
10	0.4705	0.5745	0.2117	0.3394	0.3280	106.8894	38.6287	0.0650	No	0.0
11	0.3771	55.0291	-35.5301	-0.0965	0.2371	93,858.5095	490.8310	5.2189	Yes	6.7
12	1.4253	2.3731	0.6355	1.5030	1.4770	115.6085	5.1724	0.3580	No	0.0
13	0.4423	1.0248	-1.0575	-0.2506	-0.2767	830.9353	276.4887	0.2544	Yes	14.4
14	-0.0298	-0.0162	-0.1458	-0.0892	-0.0983	145.2316	66.5919	0.0380	No	0.0
15	5.1248	5.1248	3.6284	4.0552	4.0340	36.9001	26.3756	0.1872	No	0.0
16	0.2254	12.2107	-0.9509	0.3153	0.2863	4,174.9356	28.5040	0.4066	Yes	0.2
17	-1.1233	1.0284	-27.8595	-0.3786	-0.4221	7,630.8671	196.7359	1.1343	Yes	30.2
18	0.1988	0.2068	0.1432	0.1706	0.1695	37.2813	16.5575	0.0122	No	0.0
19	0.1679	8.1187	-0.0525	0.2756	0.2645	2,964.8877	39.0656	0.2714	Yes	0.9
20	-0.0207	0.0214	-0.0316	-0.0040	-0.0028	1,326.8019	419.1689	0.0102	Yes	36.6
21	0.8769	0.8769	-0.0532	0.1558	0.1445	596.9567	462.7934	0.0863	Yes	1.9
22	-0.6859	-0.1055	-0.6859	-0.1839	-0.1730	315.6437	272.9971	0.0496	No	0.0
23	4.4952	176.7540	-94.3248	4.0325	4.5277	6,722.2968	11.4737	11.7395	Yes	5.2

Table 2: Result summary GDP

Scen#	Original result	Max result	Min result	Mean	Median	Dev. Max-Min	Dev. Original-Mean	Std. Dev.	Qualitative change	Qualitative change(%)
24	-0.3275	0.0094	-0.3275	-0.0757	-0.0706	445.1770	332.7201	0.0352	Yes	0.6
25	-1.8512	-0.6649	-10.1693	-2.4478	-2.3008	388.2784	24.3728	0.9752	No	0.0
26	-0.0625	-0.0450	-0.1327	-0.0984	-0.1054	89.0566	36.5006	0.0225	No	0.0
27	-0.2092	0.1961	-0.2808	-0.0256	-0.0343	1,862.7358	716.9836	0.0652	Yes	28.7
28	-0.3336	0.5975	-3.1552	-0.2553	-0.2004	1,470.1859	30.6799	0.1823	Yes	0.2
29	-3.4140	-3.4140	-7.1469	-4.6211	-4.5762	80.7799	26.1222	0.5332	No	0.0
30	2.0171	6.7048	-24.4632	3.0373	3.0659	1,026.1744	33.5904	1.3999	Yes	0.4
31	-0.3319	6.6105	-3.9789	-0.2387	-0.2024	4,436.6244	39.0734	0.2776	Yes	0.4
32	-0.0188	0.2695	-5.6916	-0.0292	-0.0194	20,407.3556	35.6287	0.1819	Yes	0.4
33	-0.4548	0.0043	-0.7742	-0.3848	-0.4197	202.2880	18.1703	0.1870	Yes	0.3
34	-0.0420	-0.0234	-0.1907	-0.1095	-0.1180	152.9242	61.6230	0.0416	No	0.0
35	0.1299	0.1337	0.1006	0.1162	0.1160	28.5316	11.7749	0.0069	No	0.0
36	0.0068	0.0068	0.0000	0.0000	0.0000	99,999.8975	99,899.8961	0.0002	Yes	22.0
37	2.7237	20.6049	-0.7865	4.6438	4.5100	460.6414	41.3482	1.7053	Yes	0.1
38	-0.2008	0.0060	-0.2008	-0.0405	-0.0311	510.5488	395.7329	0.0367	Yes	6.9
39	-0.2347	-0.1626	-1.1629	-0.7503	-0.8153	133.3210	68.7134	0.3115	No	0.0
40	0.1718	0.5908	0.0210	0.3290	0.3278	173.1949	47.7750	0.1007	No	0.0
41	-0.8677	8.3898	-4.9293	-0.1509	-0.3064	8,829.2582	475.2130	0.6791	Yes	35.5
42	0.0849	0.8515	-0.0032	0.5048	0.5661	169.3209	83.1906	0.2646	Yes	0.2
43	3.7710	175.0551	-94.3639	3.9922	3.8573	6,748.6261	5.5404	11.7606	Yes	3.9
44	1.0088	4.3569	-2.2679	1.4563	1.4267	454.8972	30.7283	0.2930	Yes	0.1
45	-0.0144	-0.0041	-0.0488	-0.0306	-0.0334	145.9238	53.0452	0.0127	No	0.0
46	0.8854	0.8854	0.1043	0.2016	0.1858	387.3800	339.1000	0.0642	No	0.0

Table 2: Result summary GDP

Scen#	Original result	Max result	Min result	Mean	Median	Dev. Max-Min	Dev. Original-Mean	Std. Dev.	Qualitative change	Qualitative change(%)
47	-0.0789	-0.0380	-0.1930	-0.1316	-0.1420	117.7652	40.0533	0.0390	No	0.0
48	4.1588	12.9275	4.1588	9.8860	10.1254	88.6987	57.9324	1.7414	No	0.0
49	-0.0301	-0.0262	-0.0537	-0.0426	-0.0452	64.4954	29.2689	0.0076	No	0.0
50	31.7145	39.7431	31.7145	36.1403	36.4326	22.2151	12.2463	1.6852	No	0.0

Table 3: Result summary welfare

Scen#	Original result	Max result	Min result	Mean	Median	Dev. Max-Min	Dev. Original-Mean	Std. Dev.	Qualitative change	Qualitative change(%)
1	7.2530	8.2832	7.2530	8.1321	8.1468	12.6684	10.8108	0.0852	No	0.0
2	1.7971	3.1265	1.4133	2.3984	2.4834	71.4305	25.0721	0.3408	No	0.0
3	3.2007	65.1623	-94.8393	3.2508	3.2364	4,921.9275	1.5418	7.5916	Yes	2.5
4	46.8603	46.8603	46.2642	46.3633	46.3537	1.2856	1.0718	0.0536	No	0.0
5	1.0548	1.0548	1.0470	1.0507	1.0506	0.7374	0.3936	0.0013	No	0.0
6	-0.0233	0.3205	-6.2886	-0.0467	-0.0349	14,148.1372	50.1930	0.2024	Yes	0.5
7	-0.4696	-0.2093	-2.6760	-1.7504	-1.8589	140.9144	73.1717	0.5843	No	0.0
8	0.4920	4.6320	-1.3789	2.0148	1.7901	298.3309	75.5793	1.0711	Yes	2.5
9	2.2201	2.2201	1.1058	1.2598	1.2117	88.4518	76.2308	0.1459	No	0.0
10	1.0028	1.0034	1.0014	1.0022	1.0021	0.1938	0.0654	0.0004	No	0.0
11	0.1538	60.0771	-29.8591	0.2364	0.0452	38,047.5773	34.9279	5.0579	Yes	33.9
12	-0.3938	-0.2075	-1.3396	-0.6261	-0.6138	180.8072	37.1088	0.2569	No	0.0
13	1.5655	3.0147	-0.4474	1.7365	1.7925	199.3747	9.8463	0.4756	Yes	0.2
14	-0.0350	-0.0131	-0.1626	-0.1000	-0.1096	149.3723	65.0094	0.0424	No	0.0
15	1.0245	1.0245	1.0170	1.0191	1.0190	0.7378	0.5246	0.0009	No	0.0
16	0.4124	11.5093	-0.6793	0.4714	0.4467	2,585.4997	12.5230	0.3733	Yes	0.2
17	-1.6272	0.2674	-27.8884	-0.9534	-0.9788	2,953.3504	70.6796	1.0721	Yes	5.1
18	0.1722	0.1846	0.1193	0.1465	0.1455	44.5541	17.5599	0.0122	No	0.0
19	0.0537	7.6374	-0.1858	0.1215	0.1172	6,438.8501	55.7786	0.2556	Yes	9.7
20	-0.0198	0.0838	-0.0412	0.0290	0.0334	430.6473	168.0952	0.0263	Yes	17.0
21	1.5282	2.8896	1.5282	2.6462	2.6652	51.4479	42.2476	0.1159	No	0.0
22	45.0751	45.0751	43.8885	44.0798	44.0618	2.6917	2.2578	0.1063	No	0.0
23	3.2268	232.7994	-94.8378	3.5291	3.2635	9,283.8405	8.5663	13.7056	Yes	3.3

Table 3: Result summary welfare

Scen#	Original result	Max result	Min result	Mean	Median	Dev. Max-Min	Dev. Original-Mean	Std. Dev.	Qualitative change	Qualitative change(%)
24	46.9481	46.9481	46.3043	46.4387	46.4289	1.3864	1.0970	0.0605	No	0.0
25	0.3671	1.5586	-2.4079	0.6939	0.7465	571.6575	47.0919	0.4088	Yes	4.2
26	-0.0753	-0.0518	-0.1528	-0.1143	-0.1218	88.3582	34.1338	0.0239	No	0.0
27	46.8579	46.8579	46.1544	46.4304	46.4280	1.5153	0.9208	0.0834	No	0.0
28	0.0374	1.0017	-2.4815	0.0162	0.0321	21,484.4869	130.7157	0.1160	Yes	27.3
29	-0.7764	-0.6711	-2.1350	-1.3873	-1.3572	105.5159	44.0385	0.2712	No	0.0
30	0.2334	5.5904	-25.7832	1.2371	1.2645	2,536.0113	81.1348	1.3776	Yes	11.7
31	0.0397	6.5155	-3.2550	0.0306	0.0294	31,891.8583	29.4488	0.2389	Yes	24.2
32	-0.0222	0.2595	-5.3871	-0.0473	-0.0357	11,938.9195	53.0449	0.1720	Yes	0.4
33	-0.4901	0.2084	-1.1945	-0.4967	-0.5328	282.4649	1.3155	0.3267	Yes	7.7
34	-0.0561	-0.0258	-0.2851	-0.1541	-0.1658	168.2181	63.5991	0.0616	No	0.0
35	0.0794	0.0886	0.0589	0.0716	0.0715	41.4735	10.9132	0.0055	No	0.0
36	0.0027	0.0027	0.0000	0.0000	0.0000	99,999.7390	99,899.7353	0.0001	Yes	22.3
37	0.7709	18.3061	-2.6682	2.6591	2.5278	788.7644	71.0110	1.6728	Yes	3.1
38	1.3880	1.3880	0.1176	0.2477	0.1943	512.8861	460.3697	0.1672	No	0.0
39	-0.1596	-0.0429	-1.6492	-0.9976	-1.0912	161.0144	83.9973	0.5130	No	0.0
40	1.0028	1.0076	1.0012	1.0045	1.0045	0.6322	0.1691	0.0012	No	0.0
41	0.0353	8.8476	-3.5312	0.7212	0.6427	1,716.4028	95.1104	0.6116	Yes	0.4
42	0.4684	3.7048	0.0632	2.2037	2.4713	165.2522	78.7436	1.1031	No	0.0
43	3.0509	64.9639	-94.8449	3.2689	3.1533	4,888.7261	6.6712	7.6789	Yes	1.7
44	0.1930	3.5142	-3.2882	0.6360	0.6069	1,069.5065	69.6562	0.2937	Yes	0.1
45	-0.0165	-0.0029	-0.0499	-0.0311	-0.0337	150.8527	46.8900	0.0123	No	0.0
46	2.4966	3.7132	2.4966	3.5370	3.5516	34.3990	29.4151	0.0995	No	0.0

Table 3: Result summary welfare

Scen#	Original result	Max result	Min result	Mean	Median	Dev. Max-Min	Dev. Original-Mean	Std. Dev.	Qualitative change	Qualitative change(%)
47	-0.0769	-0.0367	-0.1888	-0.1312	-0.1414	115.9388	41.4173	0.0376	No	0.0
48	1.2218	5.8527	1.2218	4.2486	4.2670	108.9977	71.2435	0.7123	No	0.0
49	-0.0382	-0.0323	-0.0537	-0.0434	-0.0444	49.1758	11.9649	0.0046	No	0.0
50	38.1162	55.3104	30.1572	45.0162	45.3665	55.8759	15.3279	4.2141	No	0.0