Econometric framework for NTM assessment: comparing methods to map NTMs into trade costs and welfare *

Eddy Bekkers
Joseph Francois
World Trade Institute, CEPR (London)
Hugo Rojas-Romagosa
CPB Netherlands Bureau for Economic Policy Analysis

January 2018

Abstract

This report provides an overview of the ways in which changes in non-tariff measures (NTMs) are mapped into welfare effects through changes in trade costs. This is done in two steps. First, methods to calculate the trade cost changes associated with changes in NTMs are discussed, concentrating on changes in NTMs as a result of the conclusion of FTAs. Second, the different ways to calculate the welfare effects of changes in trade costs are presented, compared, and assessed. In the first part the two main approaches to estimate NTM reductions associated with the implementation of FTAs are introduced and examined, the bottom-up and top-down approach. We then take the case study of the Transatlantic Trade and Investment Partnership (TTIP) to evaluate how expected reductions in NTMs as a result of a conclusion of an FTA are estimated in different impact assessment studies. We compare and analyse the main differences in these estimations and examine how they affect the overall economic impact of TTIP. We find that accounting for differences in the expected NTM reductions can explain a large share of the discrepancies regarding the overall potential economic effects between different impact assessments of TTIP. In the second part we compare the solution methods and baseline calibrations of three different quantitative trade models (QTMs) used to evaluate the effects of counterfactual experiments on reductions in NTMs: computable general equilibrium (CGE) models, structural gravity (SG) models and models employing exact hat algebra (EHA). The different solution methods generate identical results on counterfactual experiments with changes in NTMs if baseline trade shares or baseline trade costs are identical. SG models, calibrating the baseline to gravity-predicted shares, potentially suffer from bias in the predicted welfare effects as a result of misspecification of the gravity equation, whereas the other methods, calibrating to actual shares, potentially suffer from bias as a result of random variation and measurement error of trade flows. Simulations show that fitted shares calibration can generate large biases in predicted welfare effects if the gravity equation does not contain pairwise fixed effects or is estimated without domestic trade flows. Calibration to actual shares and to pairwise fixed effects based fitted shares display similar performance in terms of robustness to the different sources of bias.

Keywords: Quantitative trade models, baseline calibration, non-tariff measures, free trade agreements, gravity estimation

JEL codes: F13, F14, F15

*We want to thank participants in the research seminar at the World Trade Institute as well as Michael Pfaffermayr and Octavio Fernandez-Amador for useful comments. Parts of this report have been used in articles and book chapters under review: "Bekkers, Eddy and Hugo Rojas-Romagosa. Non-Tariff Measure Estimations in Different Impact Assessments." and "Bekkers, Eddy. The Welfare Effects of Trade Policy Experiments in Quantitative Trade Models: The Role of Solution Methods and Baseline Calibration." Address for correspondence: Joseph Francois, World Trade Institute, Hallerstrasse 6, 3012 Bern, Switzerland. joseph.francois@wti.org

1 Introduction

To evaluate the welfare effects of non-tariff measures (NTMs) trade economists typically proceed in two steps. First, the trade cost equivalent of the NTMs is calculated and second, welfare effects of the trade cost reductions are calculated with a general equilibrium model. In this report we provide a detailed overview of the different approaches to calculate the trade cost reductions of NTMs and to map these trade cost reductions into the welfare effects. To explain the different approaches we use as case study the different impact assessment studies on the Transatlantic Trade and Investment Partnership (TTIP), a free trade agreement (FTA) between the US and the EU, focusing on the expected changes in NTMs as a result of TTIP.

We distinguish between the two main approaches to calculate the reductions in NTMs as a result of TTIP: the bottom-up approach and the top-down approach. The bottom-up approach identifies the NTMs already in place between FTA partners and specifies scenarios for the expected trade cost reductions associated with this specific FTA. The top-down approach estimates the trade cost reductions from past FTAs, imposing that the expected trade cost reductions of the specific FTA in case, will be similar as in comparable agreements from the past. In general, most studies use the concept of ad valorem equivalent (AVE) of an NTM, which is the equivalent ad valorem trade cost for a particular NTM and we will discuss how AVEs associated with NTMs are calculated in the two approaches.

Quantitative trade models (QTMs) are employed to map trade costs associated with changes in NTMs into welfare effects. QTMs are employed mostly in ex ante studies on the expected effects of changes in NTMs, for example as a result of the introduction of free trade agreements like TTIP, CETA, and TPP (Felbermayr et al., 2013b; Fontagne et al., 2013; Aichele et al., 2014; Egger et al., 2015; Felbermayr et al., 2015; Petri et al., 2011; Ciuriak et al., 2016) or the WTO-agreement on trade facilitation (WTO, 2015). They are also used in ex post studies on for example NAFTA (Caliendo and Parro, 2015). In this report we distinguish between three types of QTMs in the literature: computable general equilibrium (CGE) models (for example Hertel (2013a)); structural gravity (SG) models (for example Anderson and van Wincoop (2003)); and models employing exact hat algebra, EHA-models (for example Dekle et al. (2008)).

Although the three approaches are microfounded, there are several differences. The first difference concerns the solution method of a counterfactual exercise. In the SG-approach and one of the approaches in the CGE-literature (dubbed CGE-in-levels) endogenous variables are solved for with baseline and counterfactual trade costs (in a different way) and the outcomes are compared to determine the change in welfare. In the other CGE-approach (dubbed CGE-in-relative-changes) and in the approach using EHA the relative change in welfare is calculated in one step (again in different ways).

The second difference is that SG-models and models applying EHA intend to estimate all behavioral parameters structurally, i.e. the estimating equations are derived from the same model as is used for the counterfactual exercises. Moreover, estimation and counterfactual exercise are based on the same dataset. CGE-models instead oftentimes also contain parameters taken from the literature. Third, in general the SG- and EHA-models tend to be more compact and parsimonious models, whereas CGE-models are more extensive including features like endogenous capital accumulation, non-homothetic preferences, and multiple factors of production. A fourth difference is the calibration of baseline trade costs. In the SG-approach the trade costs are structurally estimated based on gravity regressors. Baseline trade costs are equal to the fitted or predicted values of the estimated gravity equation and baseline import shares are thus equal to the fitted import shares. In the CGE- and EHA-approach instead baseline import shares are equal to the actual import shares in the data. Calibration to the fitted shares can be defended based on the argument that random variation in trade flows are filtered out in this way. Furthermore, it has been argued that this approach is robust to measurement error in observed trade flows (Yotov et al., 2016). On the other hand, calibration to actual shares does not suffer from potential misspecification of the gravity equation.

The choice to structurally estimate behavioral parameters will not systematically affect the predicted welfare effects. The impact of the scope of the models has been discussed in the literature (see for example Costinot and Rodríguez-Clare (2013) or Bekkers and Rojas-Romagosa (2017) for elements not addressed in Costinot and Rodríguez-Clare (2013) like endogenous factor accumulation and non-homothetic preferences).[1] Therefore in the second part of this report we concentrate on the other two differences. More precisely, we explore the impact of the different solution methods and baseline calibrations methods in QTMs on the predicted welfare effects of counterfactual trade policy experiments. To compare the different methods, a single-sector trade model with Armington preferences as in Anderson and van Wincoop (2003) is set up. Equilibrium equations, solution methods, and baseline calibration are mapped out in the three approaches: SG, CGE (both CGE-in-levels and CGE-in-relative-changes), and EHA. It is shown analytically and verified numerically that the different solution methods generate identical welfare effects if either the baseline import and export shares or the baseline iceberg trade costs, output and expenditure are identical. This implies that different solution methods available in the literature like solving for the baseline 'effective labor' as for example in Alvarez and Lucas (2007) and Levchenko and Zhang (2016) or the procedure to 'estimate' multilateral resistance terms as in GE PPML (Anderson et al., 2015) do not affect the results of a counterfactual trade cost experiment. The values of baseline trade costs, output and expenditures completely nail down the welfare effects of counterfactual experiments.

The impact of baseline calibration is examined with numerical simulations with the single sector model with both 121 regions from the GTAP9 data (120 countries and one rest-of-the-world) corresponding with cross-section gravity estimation and with 43 regions from WIOD corresponding with panel gravity estimation. The three sources of bias mentioned above, misspecification of the gravity equation, random variation in trade flows, and measurement error in observed trade flows, are explored for the two ways to calibrate baseline trade costs (to actual shares and to gravity-fitted shares). Before exploring these biases, various counterfactual trade experiments are implemented numerically to study the determinants of the bias. It is shown that the welfare effects will display large biases if baseline import and export shares are not correct. In particular, the welfare effects are biased upward if baseline import and export shares vis-a-vis the liberalizing partner are biased upward. So if for example baseline trade shares between the EU and the US are upward biased, a trade cost reduction experiment between these two regions such as the TTIP-experiment, will generate upward biased welfare effects for these regions.

The three sources of bias are evaluated with the counterfactual experiment of a reduction in trade costs between the USA and the EU (TTIP) with four main results. First, misspecification of the gravity equation can generate large biases in predicted welfare effects. Misspecification means in the context of this report that predicted values of the fitted gravity equation deviate systematically from the actual values, because the gravity equation is not well-specified. Misspecification is in particular severe when domestic trade flows are not included and predicted domestic trade flows are generated based on the estimated gravity equation employing only international trade flows. Second, the misspecification bias becomes relatively small, once pairwise fixed effects are included. In a panel data setting this implies that fitted import shares will be an average of the actual shares. In a cross-section setting fitted shares would become exactly equal to actual shares and the two calibration methods would then thus be identical.

Third, random variation in trade flows is not a reason to prefer calibration to fitted shares based on a gravity equation with pairwise fixed effects over calibration to actual shares. The bias generated by employing actual shares, and thus erroneously taking random variation in trade

[1] Costinot and Rodríguez-Clare (2013) point out that a disadvantage of more extensive and complex models is that results are more difficult to interpret (and become like a black box). CGE-proponents argue that by adding more features to a model, it is possible to explore effects at more detailed sector- and factor-level for example.

flows into account, is of similar size as the misspecification bias with fitted shares. To compare the two types of errors, hypothetical trade data are generated by adding random variation in trade flows to the actual trade flows. This setting is chosen instead of a setting employing data generated entirely randomly based on a data generating process, since the data-generating process would have to account for misspecification to be able to compare the sources of the bias. And the degree of misspecification would be based in turn on the actual data. Instead of using this roundabout way, random variation is added to actual trade data with random variation based on a data generating process with the variance disciplined by random variation in the estimated gravity equation including pairwise fixed effects.

Fourth, measurement error in trade flows is neither a reason to prefer calibration to fitted shares over calibration to actual shares. To draw this conclusion measurement error is added to the actual data with the data generating process for measurement error based on the difference between reported exports and imports in COMTRADE data.[2] This gap reflects oftentimes measurement error (besides the cif-fob margin) and is considered the best way to get a measure for the size of measurement error. The time-varying share of measurement error in the data is determined and based on this share, both time-varying and time-invariant measurement errors are generated. The simulations show that calibration to fitted shares with pairwise fixed effects does not perform better than calibration to actual shares, which seems to be due to the fact that the former calibration also picks up time-invariant measurement error and moreover still suffers from misspecification bias. To get rid of the time-invariant measurement error, calibration to fitted values using pairwise gravity variables and omitting pairwise fixed effects could perform better. However, the simulations show that this specification performs far worse, since the misspecification bias dominates the measurement error bias.

The report is related to three topics in the literature. First, various scholars study scope and solution methods of quantitative trade models, albeit in different literatures. Costinot and Rodríguez-Clare (2013) examine the impact of the scope of quantitative trade models employing exact hat algebra on the predicted welfare effects. Horridge et al. (2013) outline and compare numerical solution methods and softwares employed in the CGE-literature. Yotov et al. (2016) give an in-depth overview of the use of structural gravity models to conduct counterfactual trade policy experiments. However, none of these scholars compare the methods used in the different literatures.

Second, various researchers discuss size, implications, and fixes of measurement errors in trade data. Egger and Wolfmayr (2014) provide a detailed overview of different trade statistics available and their differences.[3] Gehlhar (1996) maps out the methodology employed by GTAP to reconcile trade and production data based on the reliability of reporting sources. And Yotov et al. (2016) argue that the methodology followed in SG-models is robust to measurement error and is henceforth an alternative remedy for measurement error in trade data.

Third, the report is related to work by Egger and Nigai (2015) who study the impact of unobserved trade costs on gravity estimation. They show that unobserved trade costs are large and that estimated technology parameters and coefficients on bilateral gravity variables like distance are biased in the presence of unobserved trade costs. This report instead studies the impact of unobserved trade costs on the bias in predicted welfare effects of counterfactual trade policy experiments through its impact on the baseline calibration of trade costs.

The report makes three important contributions to the literature. First, it compares the baseline calibration and solution methods of the different QTMs and thus sheds important light on the on-going discussion of the merits of the different approaches in QTMs. Second, it shows that baseline calibration can have a large impact on the predicted welfare effects of counterfactual trade policy experiments. Third, it evaluates arguments favoring the two ways of

[2] In this paper we are not interested in classical measurement error in the regressors, causing attenuation bias of the estimated coefficients. The focus is measurement error in the regressand, which is the value of trade.

[3] Jones et al. (2014) conduct a similar exercise concentrating on global input-output data.

baseline calibration, to actual and to fitted shares, concluding that calibration to actual shares and calibration to fitted shares based on a gravity estimation including pairwise fixed effects generate similar and accurate predictions. Calibration to fitted shares based on bilateral gravity regressors only instead can generate large biases in the predicted welfare effects.

We start this report with an outline of the different approaches to the estimation of trade cost reductions associated with NTM provisions in FTAs in Section 2. Then we introduce the different TTIP impact assessment studies and the expected reductions in NTMs used as an example to illustrate the calculation of NTMs in Section 3, also discussing so-called spillover effects from the introduction of FTAs on NTMs between FTA-partners and non-FTA-partners. In Section 4 we compare the predicted trade cost reductions for a set of studies on TTIP. Then we move to the second part of the report, the mapping of trade cost reductions into expected welfare effects. Section 5 compares the different methods in theory by outlining the employed trade model and gravity equation, introducing the four approaches used in the literature and presenting the differences in terms of baseline calibration and solution method. Section 6 explores the differences between baseline calibration and solution methods of the four approaches numerically. Section 7 explores the three sources of bias of calibration to fitted and actual shares, respectively gravity misspecification and random variation and measurement error. Section 8 concludes.

2 Calculating NTM reductions associated with FTAs

Since tariff levels are low for many countries negotiating deep FTAs like TPP, TTIP and CETA, the largest impact in these agreements will come from changes in NTMs. Thus, the proper estimation of expected trade cost reductions associated with provisions on NTMs has become a critical element for assessing the potential economic impacts of modern FTAs. To determine the welfare effects of deep FTAs with provisions on NTMs, researchers first have to calculate the associated reductions in trade costs. In this section we explain the two approaches to map the NTM provisions into associated reductions in trade costs. Before discussing the two approaches to calculating the trade cost reductions, we first point out how NTMs are modelled as trade costs.

2.1 Modelling trade costs caused by NTMs

The typical approach in the literature is to assume that NTMs generate resource-dissipating costs. Firms have to spend resources -i.e. time, working hours, financial resources– to comply with differences in national regulations and with barriers to trade. Such resource-dissipating costs are modelled in international trade using the concept of iceberg trade costs, where a fraction of goods "melts away" in transit from one country to another, and this fraction corresponds with the resources used or dissipated to deal with the particular NTM.

More formally, denoting iceberg trade costs for goods exported from i to j as τ_{ij}, means that τ_{ij} units have to be exported by country i to deliver one unit in country of destination j. Hence, $\tau_{ij} - 1$ units are lost or "melted away" when transporting the good. The implication is that the cost of selling a good from i to j is given by $\tau_{ij} c_i$, where c_i is the marginal cost of production in exporter i.

We define iceberg trade costs without any NTMs in place as $\tau_{ij}^{noNTM} = 1 + C_{ij}$, where C_{ij} encompasses all trade costs except for tariffs and non-tariff measures, such as international transport costs and other costs related to distance, language, and other barriers to trade. Iceberg trade costs inclusive of an NTM are defined as $\tau_{ij}^{NTM} = (1 + C_{ij})(1 + AVE_{ij}^{NTM})$, where AVE_{ij}^{NTM} is the ad valorem equivalent of an NTM, a concept often used to express the size of trade costs associated with an NTM. The AVE of an NTM is defined as the percentage increase in trade costs because of the NTM being present:

$$AVE_{ij}^{NTM} = \frac{\tau_{ij}^{NTM}}{\tau_{ij}^{noNTM}} - 1 \tag{1}$$

To calculate the AVE of an NTM, the empirical literature has employed two methods (Berden and Francois, 2015). First, the price-based method relates the price difference between countries to NTMs, controlling for other observable variables and fixed effects. Second, the value-based method estimates a gravity equation relating the value of trade between countries to NTMs, controlling again for other observable variables and fixed effects. Since all studies on TTIP discussed in this chapter employ value-based methods, we explain how the iceberg trade costs associated with an NTM, τ_{ij}^{NTM}, can be calculated with this method.

Starting with a generic theoretical gravity equation, the bilateral trade value between country i and j inclusive of bilateral tariffs, V_{ij}, can be expressed as follows:[4]

$$V_{ij} = \frac{\left((1+t_{ij})\,\tau_{ij}\right)^{-\theta} Y_i E_j}{\Pi_i \Omega_j} \tag{2}$$

where t_{ij} is the ad-valorem tariff rate, Y_i the total output of exporter i, E_j is total expenditures in country j, θ is the trade elasticity, and Π_i and Ω_j are the outward and inward multilateral resistance terms, respectively. The outward term denotes the attractiveness for exporter i to export to other destinations, and the inward term expresses the attractiveness for importer j to import from other sources. Failure to control for multilateral resistance leads to omitted variable bias (cf. Yotov et al., 2016).

From the theoretical formulation in Equation (2), the empirical gravity specification can be written as:

$$V_{ij} = \exp\left(\beta' C_{ij} - \theta \ln(1+t_{ij}) + \gamma_c \ln NTM_{ij}^c + \gamma_d NTM_{ij}^d + \eta_i + \nu_j\right)\varepsilon_{ij} \tag{3}$$

NTMs can consist of both continuous variables, NTM_{ij}^c, which are included in logs in the estimation and discrete variables, NTM_{ij}^d, which are included in levels. C_{ij} is captured by the standard gravity bilateral variables affecting the value of trade (e.g. distance, tariffs, common border, common language and colonial past). η_i and ν_j are the exporter and importer fixed effects and ε_{ij} is an error term.

Comparing Equations (2) and (3), we see that iceberg trade costs can be written as $\tau_{ij} = \exp\left(-\frac{\beta' C_{ij} + \gamma_c \ln NTM_{ij} + \gamma_d NTM_{ij}}{\theta}\right)$.[5] We can now define the AVE of an NTM –i.e. the percentage change in iceberg trade costs as a result of the presence of this NTM– for both continuous and discrete variables. For continuous variables the AVE of an NTM is defined as the percentage change in iceberg trade costs as a result of increasing the NTM-variable by 1%:

$$AVE_{ij}^{NTM^c} = \frac{d \ln \tau_{ij}}{d \ln NTM_{ij}^c} \tag{4}$$

If the NTM_{ij} is a dummy variable, the associated AVE is defined as the change in iceberg trade costs by switching the value of the NTM-dummy (NTM_{ij}) from 0 to 1:

$$AVE_{ij}^{NTM^d} = \frac{\tau_{ij}\left|NTM_{ij}^d = 1 - \tau_{ij}\right|NTM_{ij}^d = 0}{\tau_{ij}\left|NTM_{ij}^d = 0\right.} \tag{5}$$

[4]As shown by Head and Mayer (2014) many theoretical trade models can be expresses using this generic gravity equation.

[5]When running simulations the importer and exporter fixed effects in the theoretical gravity equations will change as a result of changing the values of the NTM variable, but this does not affect the calculation of the change in iceberg trade costs as a result of the NTM.

We can determine the two AVEs by combining the theoretical gravity equation in (2) with the empirical gravity equation in (3). This gives:

$$\frac{d\ln V_{ij}}{d\ln NTM^c_{ij}} = \frac{d\ln V_{ij}}{d\ln \tau_{ij}}\frac{d\ln \tau_{ij}}{d\ln NTM^c_{ij}} \tag{6}$$

$$\frac{V_{ij}\left|NTM^d_{ij}=1\right.}{V_{ij}\left|NTM^d_{ij}=0\right.} - 1 = \left(\frac{\tau_{ij}\left|NTM^d_{ij}=1\right.}{\tau_{ij}\left|NTM^d_{ij}=0\right.}\right)^{-\theta} = \exp\left(\gamma_d\right) \tag{7}$$

The expressions for the AVEs as defined in equations ()-() can now easily be written as follows:

$$AVE^{NTM^c}_{ij} = -\frac{\gamma_c}{\theta} \tag{8}$$

$$AVE^{NTM^d}_{ij} = \exp\left(-\frac{\gamma_d}{\theta}\right) - 1 \tag{9}$$

Equation (3) shows that the coefficient on tariffs can be employed to identify the trade elasticity θ, used in the counterfactual experiments. This procedure is followed in some of the TTIP studies (Francois et al., 2013; Egger et al., 2015; Aichele et al., 2014).

Of course equation (8) is an appoximation which becomes increasingly poor when the change in the continuous NTM-measure (entered in logs in the gravity equation) becomes large. In that case we can better use the exact expression for the change in the NTM-measure of a continuous measure, from the old to the new level, so from $NTM^{c,0}_{ij}$ to $NTM^{c,1}_{ij}$ and denoted by $AVE^{\Delta NTM^c}_{ij}$. For completeness we also define the AVE associated with the change in a discrete variable (entered in levels in the gravity equation) from $NTM^{d,0}_{ij}$ to $NTM^{d,1}_{ij}$, $AVE^{\Delta NTM^d}_{ij}$:

$$AVE^{\Delta NTM^c}_{ij} = \frac{\tau_{ij}\left|NTM^c_{ij}=NTM^{c,1}_{ij}\right.}{\tau_{ij}\left|NTM^c_{ij}=NTM^{c,0}_{ij}\right.} - 1 = \exp\left(-\frac{\gamma^c \ln\frac{NTM^{c,1}_{ij}}{NTM^{c,0}_{ij}}}{\theta}\right) - 1 = \left(\frac{NTM^{c,1}_{ij}}{NTM^{c,0}_{ij}}\right)^{\frac{\gamma_c}{\theta}} - 1 \tag{10}$$

$$AVE^{\Delta NTM^d}_{ij} = \frac{\tau_{ij}\left|NTM^d_{ij}=NTM^{d,1}_{ij}\right.}{\tau_{ij}\left|NTM^d_{ij}=NTM^{d,0}_{ij}\right.} - 1 = \frac{\exp\left(-\frac{\gamma^d NTM^{d,1}_{ij}}{\theta}\right)}{\exp\left(-\frac{\gamma^d NTM^{d,0}_{ij}}{\theta}\right)} - 1 \tag{11}$$

Some studies calculate the percentage change in iceberg trade costs using an approximation. The relative change in the value of trade as a result of changing a zero/one variable like NTM is written as follows:[6]

$$\frac{\partial\ln V_{ij}}{\partial NTM^d_{ij}} = \frac{\partial\ln V_{ij}}{\partial\ln \tau_{ij}}\frac{\partial\ln \tau_{ij}}{\partial NTM^d_{ij}} \tag{12}$$

From the gravity equation in (3) we have:

$$\frac{\partial\ln V_{ij}}{\partial NTM_{ij}} = \exp\left(\gamma_d\right) - 1 \tag{13}$$

Combining equations (12)-(13) gives for the percentage change in iceberg trade costs:

$$\frac{\partial\ln \tau_{ij}}{\partial NTM_{ij}} = -\frac{\exp\left(\gamma_d\right) - 1}{\theta} \tag{14}$$

[6]Our notation is somewhat loose here, since NTM^d_{ij} is a discrete variable. Formally, we should write $AVE^{NTM^d,appr}_{ij} = \frac{\tau_{ij}\left|NTM^d_{ij}=1\right. - \tau_{ij}\left|NTM^d_{ij}=0\right.}{\tau_{ij}\left|NTM^d_{ij}=0\right.}$.

The exact change in iceberg trade costs instead is given in equation (9). For larger values of γ, equation (14) becomes an increasingly poor approximation of the exact expression in equation (17). For example with $\gamma = 1.21$ and $\theta = 7$, we would find a percentage change in τ_{ij} of -33.6% instead of the -15.9% based on equation (9). Kee et al. (2009) use equation (14) to calculate the AVE of NTMs. The AVE of an NTM is defined as the percentage change in iceberg trade costs, thus implicitly assuming that iceberg trade costs without the NTMs are equal to one. Since the level of trade costs without NTMs is unknown, this is a natural normalisation. But as we have shown with our numerical example, equation (14) is an approximation and it seems better to use equation (9) to calculate the gravity-inferred AVE.

2.2 Two approaches to calculate NTM-related trade cost reductions

The previous subsection has equipped us with the necessary tools to understand the differences between the two approaches to calculate the reduction in trade costs as a result of NTM provisions in FTAs: the bottom-up approach and top-down approach. To compare the expected trade cost reductions in the two approaches, we have to calculate the ad valorem equivalents of the introduction of an FTA.

2.2.1 Bottom-up approach

The bottom-up approach first calculates the size of NTMs already in place between FTA partners and then makes assumptions, or specifies scenarios, to infer the potential reductions in NTMs that can be expected from the implementation of that particular FTA. So, these studies start with initial levels of AVEs associated with current NTMs in place, and then impose scenarios for the percentage reductions in these NTMs or the AVEs of the NTMs as a result of the conclusion of an FTA.

Oftentimes the size of NTMs is estimated at the sector-level using detailed micro data –i.e. firm-level surveys and product-level data, and sometimes in combination with expert opinion. For instance, Francois et al. (2013) infer the size of NTMs based on business surveys with about 5,500 data points. These data are mapped into AVEs by estimating sector-level gravity equations. Fontagne et al. (2013) estimate the size of NTMs in goods based on product level data from the UNCTAD-TRAINS database for NTMs in goods, and estimate the size of NTMs in services at the sector level based on importer fixed effects in each country relative to a benchmark country with the lowest level of NTMs judged by the amount of trade (Fontagne et al., 2011). In Section 3.2 we provide a detailed explanation of the specific estimations used in both papers.

To compare the bottom-up studies with the top-down studies we have to calculate the implied AVE of the FTA based on the AVEs of NTMs in place and scenarios for the reductions in these NTMs. We can do this in two ways, using either assumed percentage reductions in NTMs or assumed percentage reductions in the AVEs of NTMs. The first approach to calculate the AVEs of an FTA in bottom up studies can be chosen when NTM-levels have been calculated based on gravity estimation and all estimation details are available. We can then apply equations (10)-(11) imposing a percentage reduction in the NTM from NTM_{ij}^{d0} to NTM_{ij}^{d1}. We see that for calculation of the associated AVE, $AVE_{ij}^{\Delta NTM^d}$ in case of a discrete NTM measure, we need the estimated coefficients and trade elasticities, respectively γ^d and θ.

The second approach to infer the AVEs of an FTA based on bottom-up studies can be followed when only initial AVEs of NTMs are available and scenarios for the percentages by which these AVEs are reduced. Defining the initial level of AVE in the bottom-up studies as $AVE_{ij}^{NTM,NO-FTA}$ and a percentage change in these AVEs as $pc\_AVE_{ij}^{FTA} = \frac{AVE_{ij}^{NTM,FTA} - AVE_{ij}^{NTM,NO-FTA}}{AVE_{ij}^{NTM,NO-FTA}}$, (in the CEPR and CEPII studies for example typically 25%), we can calculate AVE of the FTA,

AVE_{ij}^{FTA}, (so the percentage change in iceberg trade costs) as:

$$AVE_{ij}^{FTA} = \frac{\tau_{ij}^{FTA} - \tau_{ij}^{NO-FTA}}{\tau_{ij}^{NO-FTA}} = \frac{1 + AVE_{ij}^{NTM,FTA} - 1 + AVE_{ij}^{NTM,NO-FTA}}{1 + AVE_{ij}^{NTM,NO-FTA}}$$
$$= \frac{AVE_{ij}^{NTM,NO-FTA}}{1 + AVE_{ij}^{NTM,NO-FTA}} pc\_AVE_{ij}^{FTA} \tag{15}$$

For example, an initial AVE of 9.3% and a reduction by 25% as in the CEPR study for services in the EU leads to a percentage reduction in iceberg trade costs of 2.1%, $\frac{9.3}{109.3}25\% = 2.1\%$.

A somewhat different version of the bottom-up approach can be found in the study on the effects of Trans-Pacific Partnership (TPP) by Ciuriak et al. (2016). This is an ex-ante study of the expected effects of TPP that is based on the negotiated treaty text to infer expected trade cost reductions.[7] In particular, this study maps the TPP treaty text into trade cost indicators such as the OECD's services trade restrictiveness index (STRI). Hence, they identify provisions in the text that lead to changes in trade cost indicators, controlling for what has been dubbed as "legal inflation" – i.e. provisions in treaties that do not effectively reduce trade costs (Horn et al., 2009). These changes in the trade cost indicators can then be mapped into associated AVEs (so percentage reductions in iceberg trade costs) based on gravity estimations with these trade cost indicators.

2.2.2 Top-down approach

The top-down approach infers the expected reductions in NTM trade costs exclusively based on gravity estimation. Therefore, this method does not calculate initial levels of NTMs but determines the trade cost reduction associated with an FTA directly from a similar specification as in Equation (3). In particular, calculating the iceberg trade costs reductions as a result of an FTA ($\widetilde{\tau_{ij}^{FTA}}$) using the top-down approach is straightforward: the variable NTM_{ij} has to be replaced by FTA_{ij} in Equation (3), to estimate:

$$V_{ij} = \exp\left(\beta' C_{ij} - \theta \ln(1 + t_{ij}) + \delta FTA_{ij} + \lambda T_{ij} + \eta_i + \nu_j\right) \varepsilon_{ij} \tag{16}$$

where FTA_{ij} can be either a dummy variable that indicates the presence of an FTA between i and j or an index that determines the "depth" of the FTA between both countries. T_{ij} are the tariffs from Comparing equations (2) and (16), we observe that iceberg trade costs can be written as $\tau_{ij} = \exp\left(-\frac{\beta' C_{ij} + \delta FTA_{ij}}{\theta}\right)$. The associateed AVE of an FTA can then be calculated based on the change of the FTA-dummy, FTA_{ij}, from 0 to 1:[8]

$$AVE_{ij}^{FTA} = \frac{\tau_{ij}^{FTA} - \tau_{ij}^{NO-FTA}}{\tau_{ij}^{NO-FTA}} = \exp\left(-\frac{\delta}{\theta}\right) - 1 \tag{17}$$

The coefficient on the FTA variable (δ) assesses by how much FTAs have increased trade in the past and can thus be used to determine the percentage change reduction in iceberg trade costs as a result of the FTA. Note that in this specification both the tariffs and non-tariff measures associated with an FTA are grouped. Some studies, on the other hand, explicitly include tariff levels in equation (16), which allows to separately identify the NTM components associated with the implementation of the FTA. In this case, equation (17) will estimate AVE_{ij}^{NTM}

[7] In contrast, since negotiations are still ongoing on TTIP, there is still no final treaty text that can be used.

[8] Again, the importer and exporter specific terms will change with different FTA values, but this does not affect the iceberg trade costs estimations.

–i.e. the AVE reduction associated only with the NTM provisions in the FTA, independent of any tariff reductions in the agreement.

As discussed into detail in Section 3.2, the studies using the top-down approach differ with respect to the level of aggregation of trade data used (sectoral versus aggregate), whether endogeneity of FTAs is accounted for or not, and whether the trade elasticities are estimated structurally (from the same gravity equation) or not. For instance, some studies (Felbermayr et al., 2015, 2013b; Carrere et al., 2015) employ a simple 0-1 FTA dummy, implying that TTIP would generate the same trade cost reduction between EU countries and the US as FTAs have done on average in the past. On the other hand, Egger et al. (2015) and Aichele et al. (2014) instead consider the depth of an FTA and model TTIP as the move from no FTA to a deep FTA. Another important difference is that in the first set of studies the trade elasticity is taken from the literature, whereas in the studies employing the depth of FTAs the trade elasticities are estimated structurally with the gravity equation.

2.3 Comparing both approaches

When comparing both methodological approaches, there are advantages and drawbacks from each. On the one hand, the bottom-up approach requires a data-intensive process (i.e. the use of micro-data and surveys) that provides detailed information on different trade costs associated with NTMs. This process also provides asymmetric bilateral trade cost data, which reflects initial country- and sector-specific trade cost levels, and thus, the potential for reductions when negotiating FTAs.

On the other hand, the top-down approach can be applied using standard gravity datasets and econometric specifications, which makes it a less data and analytical-intensive process. In addition, when properly specified, this approach can also inform about non-observable trade costs that are overlooked by the bottom-up approach. However, this more straightforward approach also comes at the cost of assuming symmetric trade cost reductions, which disregard different initial NTM values by country and sector.

3 Overview of studies on TTIP

This section provides a brief introduction to the different studies that will be compared in this chapter. In general, the studies can be divided between CGE-based models and structural gravity (SG) models.[9] After this overview, we conduct a more systematic comparison of the way trade cost reductions are modelled in each TTIP study.

3.1 General characteristics of the TTIP studies

3.1.1 CEPR Study (Francois et al., 2013)

The most influential and mostly cited economic analysis of TTIP has been the CEPR study (Francois et al., 2013) using a CGE-based analysis.[10] This study employs a CGE model to simulate the expected economic impact of TTIP.[11] The model's main features are intermediate linkages, multiple sectors and production factors, endogenous capital accumulation, monopolistic competition and the inclusion of international transport margins, export subsidies, import

[9] See Bekkers and Rojas-Romagosa (2017) for a detailed discussion of the differences between both modelling frameworks, and the related TTIP economic effects.

[10] This is the reference study by the European Commission and DG Trade (cf. European Commission, 2013) and the study discussed by most commentators (see for example The Economist, 2013; Rodrik, 2015; Wolf, 2015; The Guardian, 2015; Mustilli, 2015). The relevance of this study is highlighted by the request of the European Parliament to conduct an independent evaluation of this study, which was done by Pelkmans et al. (2014).

[11] The study uses a variant of the GTAP CGE-model. The main characteristics and references to the standard GTAP model are detailed in Hertel and Tsigas (1997) and Hertel (2013b).

tariffs and other taxes. The CEPR study uses the GTAP-8 database (base year 2007), with 20 sectors and 11 regions, where the EU is treated as a single region rather than 28 disaggregated countries. A baseline scenario for 2027 is then projected by endogenising productivity growth such that macroeconomic aggregates in the baseline are equal to long-term GDP projections from the OECD and UN population projections.

The study explores the economy-wide impacts of several TTIP scenarios, of which reducing tariffs and NTMs are the most important. As an outcome of the CGE analysis, detailed simulation results are provided for expected changes in GDP, household disposable income, overall aggregate and bilateral export and import flows, trade diversion effects (from/to intra-EU, US and third-countries), terms-of-trade, tariff revenues, sectoral output and sectoral trade flows. The sustainability impact also includes detailed results on changes in wages for high- and low-skill workers, sectoral employment by skill level, labour displacement measures, changes in CO2 emissions and land use. Finally, the study provides GDP and trade effects for third-countries.

3.1.2 CEPII study (Fontagne et al., 2013)

The CEPII study starts with a description of the current trade and investment relations between the EU and the US. Given the limited average level of the import tariffs between both regions (2% in the US and 3% in the EU), they predict that these tariffs will not be the most important topics in the TTIP negotiations.[12] As with the CEPR study, this study finds that the corresponding levels of protection provided by the non-tariff measures are much higher on average than those provided by the tariffs. They also find that these differ significantly across sectors. Thus, they state that the significant negotiation topics at the macroeconomic level are on non-tariff measures, regulation in services, public procurement, geographical indication of origin and investment. They argue that these topics are contentious and provide an overview of each topic. The sector-specific trade barriers, together with NTMs and other contentious topics explain the overall sensitivity of the TTIP negotiations.[13] Finally, to assess the macroeconomic effects of TTIP they use the MIRAGE CGE model (Bchir et al., 2002) that similar to the CGE model in the CEPR study, provides a broad set of economic outputs.[14]

3.1.3 Egger et al. (2015)

Egger et al. (2015) examine the potential impact of TTIP with a hybrid approach that combines a CGE economic model with structural estimation of the trade elasticities and the expected trade cost reductions to generate estimates of the welfare effects for the EU, United States and third countries. The study follows a two-step approach. In the first step a gravity model is employed to yield estimates of reductions in trade costs. These values are then used as inputs in the second step where a CGE model simulates the economy-wide effects. This study focuses on and reports largely the same outcome measures as Francois et al. (2013). The CGE model employed in Egger et al. (2015) is also very similar to the model in Francois et al. (2013). The main difference with respect to the CGE application is that Egger et al. (2015) use the more recent GTAP-9 database with base year 2011 instead of the older GTAP-8 database with base year 2007.

[12]However, they do acknowledge that tariffs will be important for some sensitive sectors: for both the EU and US dairy products, clothing and footwear are sensitive. Furthermore, steel items are sensitive for the US and meat products for the EU.

[13]As in other studies reviewed here, they acknowledge that not all of the different aspects of the negotiations can be incorporated into a model.

[14]This CGE model is also related to the GTAP-class models.

3.1.4 Felbermayr et al. (2015)

The study by Felbermayr et al. (2015) is a typical structural gravity (SG) application based on a one-sector model with both a gravity equation to determine the trade costs associated with the TTIP experiment and a general equilibrium equation to calculate the welfare effects, where both equations follow directly from the theoretical gravity model. Baseline trade costs are calibrated to the fitted/predicted values of the estimated gravity equation. The paper does not follow the structural gravity literature in the determination of the parameters of the model, since the trade elasticity –a crucial parameter in the calculation of the welfare effects– is not estimated structurally, but is taken from the literature (i.e. Egger et al., 2011; Egger and Larch, 2011a).

3.1.5 Felbermayr et al. (2013a)

Felbermayr et al. (2013a) explore the effects of TTIP on both trade and welfare and on the labour market. This is done with two distinct SG models. The first model explores the impact on trade, GDP and welfare for 126 countries, whereas the second model is limited to 28 countries due to the lack of available labour market data for more countries. The first model is identical to the model employed by Felbermayr et al. (2015) except for the variables included in the gravity equation.

3.1.6 Aichele et al. (2014)

To date, the study by Aichele et al. (2014) is the most sophisticated SG-based study on TTIP. These researchers use a multi-country and multi-sector Ricardian trade model with national and international input-output linkages identical to the model employed by Caliendo and Parro (2015) to study the impact of tariff reductions as a result of NAFTA. Counterfactual outcomes are calculated employing so-called exact hat-algebra with baseline import shares calibrated to actual import shares as in the CGE-approach.

3.1.7 Carrere et al. (2015)

Carrere et al. (2015) use a multi-sector SG approach to determine the real wage and unemployment effects of TTIP by including labour market frictions and equilibrium unemployment in an Eaton-Kortum-type trade framework. The authors also present welfare effects based on a weighted average of real wage and unemployment effects. A salient feature of this study is that the negative unemployment effects dominate the positive real wage effects in their setup, which generate overall negative welfare effects in some countries. The negative unemployment effects stem from a reallocation of workers from sectors with less labour market frictions and smaller equilibrium unemployment to sectors with more labour market frictions and higher equilibrium unemployment.

3.2 NTM reductions in each TTIP study

In this section we explain how each particular impact assessment study estimates the associated trade cost reductions from implementing TTIP.

3.2.1 Francois et al. (2013)

Francois et al. (2013) take into account the three types of trade cost reductions associated with FTAs: tariffs, NTMs and spillover effects. To determine the size of reductions in NTMs as a result of TTIP, this study follows a bottom-up approach, taking four steps. First, the size of NTMs is inferred from business surveys (with about 5,500 data points). Firms from a particular country i are asked to rank the overall restrictiveness of an export market j between 0 and 100.

The bilateral indexes are aggregated per sector to importer j specific indexes. This aggregated number defines the size of NTMs for imports into country j. The import-specific indexes are then included in a standard gravity regression –as in Equation (3)– by interacting the NTM measures with dummies for intra-EU, intra-NAFTA and transatlantic trade. Actionable NTMs between the EU and the US are then defined as the difference in the ad valorem equivalent trade costs of the NTMs (using estimated tariff elasticities in goods, while for services a trade elasticity of 4 is used) for US-EU trade, intra-EU trade (NTMs into the EU), for US-EU trade and intra-NAFTA trade (NTMs into the US). Comparing these trade cost estimates, the actionable NTMs to be negotiated in TTIP consist of the difference in costs perceived by businesses surveyed of importing into an EU country from the US in comparison to importing into the same EU country from another EU country.

Second, NTMs are divided into two categories: cost-increasing barriers and rent-creating barriers.[15] On the basis of the Ecorys (2009) survey and expert judgement, on average (across sectors) 60% of the NTMs are cost-increasing. The other 40% of the claimed increase in prices as a result of NTMs, can be attributed to rents. The implication of these rents is that with existing NTMs firms have more market power and thus set higher prices.[16]

Third, the study assumes that only a fraction of NTMs can be reduced, as it is not possible to remove them because of legal, institutional or political constraints.[17] It is then assumed that 50% of the ad valorem equivalent of estimated NTMs can be reduced, or as they term it, are actionable.[18]

Fourth, as part of the TTIP experiment only a fraction of the actionable share of NTMs is assumed to be lowered as a result of signing TTIP. Different scenarios are employed, but the baseline scenario has a 50% reduction in the actionable NTMs. The result of combining the actionability share with the share of actionable NTMs reduced (both 50%) is that NTMs are expected to fall by 25% with TTIP.

3.2.2 Fontagne et al. (2013)

The CEPII study uses the estimates for NTMs in goods from the study by Kee et al. (2009) and their own CEPII estimates (Fontagne et al., 2011) for NTM values in services. Kee et al. (2009) estimate AVE NTMs using a 0-1 dummy variable on NTMs at a very detailed product (tariff line) level (HS6) using the UNCTAD's TRAINS database. To consolidate this information, they then regress import values per country and tariff line (corrected for endogenous tariffs) on the (instrumented) NTM-dummies, constraining the coefficients to be positive. Thus, their calculations are based on product-level import equations that consolidate information from several sources, in particular, the NTMs already contained in the UNCTAD-TRAINS database. Finally, trade elasticities are estimated based on the Feenstra (1994) approach –using identification by heteroskedasticity– and these are then used to obtain the NTM AVE values.

Fontagne et al. (2011) estimate NTMs in services using a residuals method based on the importer fixed effect in each importer country relative to the importer fixed effect in a benchmark country with the highest value of predicted trade relative to actual trade. The importer fixed effect of the benchmark country defines trade costs in the most liberalised country and thus, in the country with the largest scope for trade liberalisation. These estimates generate NTM

[15]See Box 1 in Francois et al. (2013) for more details.

[16]It is not clear from the description of the simulations whether prices are really reduced by 40% of the ad valorem equivalent of the actionable NTMs as a result of the TTIP experiment or whether the 60%/40% cut only serves to determine the share of NTMs that is cost-increasing.

[17]These limitations in reducing NTMs stem from legal and even constitutional restrictions, political and consumer sensibilities and technical limitations, among others. See Egger et al. (2015) for a more detailed discussion on this topic.

[18]The share is based on expert opinions, cross-checks with regulators, legislators and businesses opinions taken from the business survey in Ecorys (2009). More details are in Box 1 in Francois et al. (2013).

AVEs of cross-border trade flows in services for nine services sectors and 65 countries from the GTAP database.

The CEPII study then employ a 25% reduction in their estimated NTM AVE values, based on the same arguments used in Francois et al. (2013) –i.e. the level of actionability (50%) times the expected level of NTMs that can be actually negotiated in TTIP (also 50%).

3.2.3 Felbermayr et al. (2013a) and Felbermayr et al. (2015)

Both these studies follow a top-down approach as in Equation (16) separately accounting for tariffs and using a dummy variable for FTAs. So instead of calculating the ad valorem equivalent TTIP with the depth of FTAs, they use a simple zero-one dummy for all FTAs. The two studies only differ with respect to the number of countries included in the gravity estimation and the control variables employed.[19] The gravity equation is estimated with PPML accounting for endogeneity of FTAs, based on the approach in Egger et al. (2011). The two studies do not report the AVEs of their TTIP experiment, since they directly calculate the change in income, multilateral resistances and welfare as a result of the change in the FTA dummy from 0 to 1. However, the NTM AVE values can be easily obtained from Equation (17), using their estimated FTA coefficient (δ) and the employed trade elasticity (θ).

3.2.4 Egger et al. (2015)

This study calculates the reduction of NTMs in goods employing a top-down approach, while the reduction in NTMs in services is calculated with a bottom-up approach. To calculate the reductions in NTMs on manufacturing goods, a gravity equation as in Equation (16) is estimated including the depth of FTAs, while controlling for FTA endogeneity following the approach in Egger and Larch (2011b). To measure the depth of FTAs the authors use the index of the FTA-depth proposed by Dür et al. (2014) which ranges from 0 to 7, where TTIP is assumed to be a deep FTA with the maximum value of a FTA-depth of 7.[20] Based on the estimated trade elasticities (δ), the NTM AVEs are calculated as moving from no FTA (index=0) to a deep FTA (index=7).

The AVEs of trade restrictions in services are taken from Jafari and Tarr (2015) based on the World Bank's STRI database (Borchert et al., 2014). Both for manufacturing goods and services the remaining steps are the same as in Francois et al. (2013). Hence, NTMs are split up into cost-increasing and rent-generating NTMs and only the share of actionable NTMs are assumed to be reduced under TTIP. In this study the share of cost-increasing NTMs is again 50%, but in contrast to the CEPR study, the share of actionable NTMs is assumed to be now 100% (instead of 50%), which is consistent with the presumption that TTIP will be a deep FTA. On the other hand, the share of actionable NTMs in services varies by scenario. In one scenario there is no reduction in services NTMs and in the other scenario NTMs for non-financial services fall by 50%.

3.2.5 Aichele et al. (2014)

NTM reductions in this study are defined as a reduction in iceberg trade costs calculated using a top-down approach and separately accounting for tariff effects. A gravity model following Equation (16) is estimated including dummies for shallow and deep FTAs. A shallow FTAs is defined as an FTA with a score between 0 and 3 in the FTA depth index from Dür et al. (2014), and a deep FTAs is defined as an FTA with an index between 4 and 7. The FTA dummies are

[19] The exact approach in Felbermayr et al. (2013a) can be found by consulting the companion study (Felbermayr et al., 2013b).

[20] In terms of the coding of FTA-depth this means that the FTA should contain provisions on all seven topics identified by Dür et al. (2014).

instrumented by employing a measure for trade contagion proposed by Baldwin and Jaimovich (2012). To determine the reduction in overall iceberg trade costs for goods this study uses the estimated trade elasticities, whereas for services a trade elasticity of 5.9 is used, based on estimates from Egger et al. (2012). Then they assume that the AVE NTM reductions associated with TTIP can be calculated by shifting their FTA-dummy definition from 0 (shallow FTA) to 1 (deep FTA).

3.2.6 Carrere et al. (2015)

This study accounts for reductions in NTMs following the top-down approach in Equation (16) with a 0-1 FTA dummy and separately estimating the effect of tariffs. Although they use data on 35 sectors for their SG model, the gravity equation is estimated by pooling all sectors. Thus, this study estimate a single trade elasticity of 3.17 and a coefficient on the FTA-dummy of 0.52. Hence, despite having data at the sectoral level, only an aggregate NTM trade cost reduction is calculated.

3.3 Spillover effects

Another importance source of divergencies in estimated trade cost reductions comes from the so-called spillover effects –i.e. changes in trade costs in third countries that are related to the NTM reductions between the TTIP partners. These spillover effects were first introduced in the discussion on TTIP by Francois et al. (2013), distinguishing between direct and indirect spillovers. Direct spillovers occur when third countries find it less-costly to export to the EU and the US as a result of TTIP. Indirect spillovers take place when third countries partially take over the harmonised standards in the EU and US, resulting in lower trade costs between third-countries and for exports from the EU and US to third countries. The spillover effects stem from the expectation that TTIP will lead to improved regulatory cooperation between the EU and the US and that this new regulatory framework will become a global standard. Harmonisation of regulations will make it less costly to comply with the fixed costs of exporting to the EU and US market: these costs have to be incurred only once if regulations are harmonised.

The obvious effect of modelling spillovers is that negative trade diversion effects on third countries become smaller. The assumption of positive spillover effects is not uncontroversial. When two countries harmonise standards, they will also replace old standards, possibly agreed upon with third countries. This might make it more difficult for third countries to comply with the new standards, thus generating cost increases instead of cost reductions. The empirical literature on the scope for spillover effects is summarised in Baldwin (2014), although he organises the discussion around the concept of "negative trade diversion". When a country signs a deep FTA it might improve the functioning of its services sector, implement stricter rules on competition policy and streamline its government procurement, for example. To a large extent these measures are non-discriminatory in nature, thus also generating benefits for non-members. The studies cited in Baldwin (2014) indicate that the scope for negative trade diversion is very limited: in most cases trade with non-FTA partners also increases when an FTA is signed. This does not provide, however, conclusive evidence for the presence of direct spillover effects: trade with non-FTA partners might also increase after an FTA has been signed because countries signing an FTA might be implementing other types of reforms together with signing an FTA.

Only Francois et al. (2013) works with direct and indirect spillover effects in their main simulations, whereas most other studies report the effects of spillovers in the robustness checks. All studies, however, follow Francois et al. (2013) by assuming that direct spillovers (third countries exporting to the EU) are 20% of the trade cost reductions between the EU and the US and indirect spillovers are 10% (third countries exporting to each other).[21]

[21]For example, if there is a 5% total trade cost reduction between the EU and US, the direct spillover (i.e. 20% over total trade costs) will represent an additional 1% total trade cost decrease for third countries exporting

When we analyse the spillover effects, Table 1 shows that the economic effects of including spillovers varies considerably within studies. Whereas Fontagne et al. (2013) and Felbermayr et al. (2015) find that the spillover effects raise welfare by about two thirds, the other studies find effects in the range of 20% to 25%.[22] We can explain the large effect in Felbermayr et al. (2015) by the way baseline trade costs are modelled in their single-sector model. As discussed below, this approach biases the effects of trade cost reductions upwards and thus, also the effect of spillovers. Thus, by excluding both these studies as outliers, we conclude that including 20% direct and 10% indirect spillovers is expected to have additional welfare increases on the TTIP-partners of about 25%.

Table 1: Estimated TTIP welfare effects, with and without spillover effects

	Without spillovers		With spillovers		Relative difference
	EU	USA	EU	USA	
CEPR study	0.4	0.3	0.5	0.4	20.8%
CEPII study	0.3	0.3	0.5	0.5	66.7%
Egger et al. 2015	2.3	1.0	3.0	1.1	26.5%
Aichele et al. 2014	2.1	2.7	2.7	3.4	25.4%
Felbermayr et al. 2015	3.9	4.9	7.8	7.1	67.7%

Source: Own estimations based on reported results from cited studies.

4 Comparison of trade cost reductions between studies

In this section we do quantitative comparisons of the explicit and implicit trade cost values employed in each TTIP study and explain why there are significant differences in these estimations between studies. First, we explain differences in trade cost estimations when using the gravity model, and then we compare the iceberg trade cost reductions in each study.

4.1 Comparing trade cost estimations using the gravity estimations

The TTIP studies can be compared quantitatively based on the AVEs of the introduction of TTIP (or phrased differently the percentage reduction in iceberg trade costs as a result of implementing TTIP), $AVE_{ij}^{FTA} = \frac{\tau_{ij}^{FTA}}{\tau_{ij}^{NO-FTA}} - 1$. For example, from Felbermayr et al. (2013a) and Felbermayr et al. (2015) we can directly calculate the percentage change in iceberg trade costs using equation (17) with δ the coefficient on the FTA dummy and θ the assumed trade elasticity. A trade elasticity of 7 in the baseline and estimates of δ of respectively 1.24 and 1.21 generate the numbers displayed in Table 2.[23][24] For the bottom-up studies we employ the initial levels of AVEs of NTMs reported in the studies and apply percentage reductions in these AVEs. For services in the EU in the Egger et al. study for example, we apply the formula in equation (15) on an initial AVE of 17.6% and a 25% in the AVE (see Table 2) would lead to an AVE of

to the US or EU and an additional 0.5% indirect spillover reduction (i.e. half the size of the direct spillover decrease) for EU and US export costs to third countries and for trade between third countries.

[22]The large effects in Fontagne et al. (2013), however, could be explained from the fact that this study only reports effects rounded to one decimal. While the absolute differences are less significant.

[23]See Table 3 in Felbermayr et al. (2015) and Table II.2 in Felbermayr et al. (2013a).

[24]To calculate the average AVEs in Aichele et al. (2014) we can use the same methodology but we need to do some additional calculations, since the authors report different estimated trade elasticities and deep FTA dummies for more than 30 sectors. In particular, we first calculated the percentage changes in iceberg trade costs for the different sectors based on the estimated tariff elasticities and the coefficients for shallow and deep FTAs (in Tables 1 and 2 in Aichele et al. (2014)) and then calculated the weighted average AVEs for agriculture, manufacturing and services. We calculate AVEs per sector using the following formula, $\frac{\tau^{FTA}}{\tau^{NO-FTA}} - 1 = \exp\left(-\frac{\beta_{shallow}+\beta_{deep}}{\theta}\right) - 1$.

TTIP of $AVE^{ser,Egger}_{TTIP} = \frac{0.176}{1.176} * -25\% = -7.5\%$.[25]

4.2 Ad valorem equivalents of the introduction of TTIP in different studies

To make the TTIP experiment comparable across studies we report the AVE as a result of TTIP in Table 2. The bottom-up studies (Francois et al., 2013; Fontagne et al., 2013) report initial levels of NTMs and assume that the NTMs will be reduced by a fraction of the initial level. The top-down studies report the effect of (deep) FTAs on trade flows in gravity estimations. In Subsection 4.1 we have discussed how the AVEs (percentage reductions in iceberg trade costs) of TTIP can be calculated with the two approaches. For the studies reporting trade cost reductions at sectoral levels, we calculate a weighted average for the three main sectors -agriculture, manufactures and services- based on more disaggregated subsectors. The weights are given by the amount of trade from the EU to the US for US NTMs and from the US to the EU for EU NTMs (using trade data from the GTAP-9 database).

Table 2: Trade-weighted average percentage iceberg trade cost reductions

		AVE NTM initial levels			AVE TTIP: percentage change iceberg trade costs		
	Weighted averages	EU	USA	both regions	EU	USA	both regions
CEPR study	**Overall**	**10.6**	**14.3**	**12.5**	**-2.3**	**-2.8**	**-2.5**
	Agriculture/Primary	-	-	-	-	-	-
	Manufacturing	11.9	16.3	14.3	-2.5	-3.1	-2.8
	Services	9.3	10.0	9.7	-2.1	-2.1	-2.1
Egger et al., 2015	**Overall**	**16.5**	**16.5**	**16.5**	**-12.4**	**-13.6**	**-13.0**
	Agriculture/Primary	15.9	15.9	15.9	-15.9	-15.9	-15.9
	Manufacturing	15.9	17.0	16.5	-15.9	-17.0	-16.5
	Services	17.6	15.4	16.6	-6.7	-6.0	-6.4
CEPII study	**Overall**	**38.8**	**37.0**	**37.9**	**-6.5**	**-7.7**	**-7.1**
	Agriculture/Primary	48.2	51.3	48.8	-8.1	-8.5	-8.3
	Manufacturing	42.8	32.3	37.0	-7.5	-6.1	-6.9
	Services	32.0	47.3	39.0	-6.1	-8.0	-7.1
Aichele et al., 2014	**Overall**				**-12.7**	**-13.8**	**-13.2**
	Agriculture/Primary				-52.5	-52.5	-52.5
	Manufacturing				-16.5	-17.6	-17.1
	Services				-5.2	-4.6	-4.9
Felbermayr et al., 2013	**Overall**				**-16.2**	**-16.2**	**-16.2**
Felbermayr et al., 2015	**Overall**				**-15.9**	**-15.9**	**-15.9**
Carrere, 2015	**Overall**				**-15.1**	**-15.1**	**-15.1**

Sources: Own estimations based on NTMs and trade cost reductions estimates from referred studies and bilateral sector-specific US-EU trade data from the GTAP-9 database.

The weighted iceberg trade cost reductions reported in Table 2 show that the differences are large, ranging from 2.5% reduction in trade costs in Francois et al. (2013) to 16.2% in Felbermayr et al. (2013a). In addition, Table 2 shows that services trade contributes relatively little to the overall reduction in trade costs. Given that services trade is only about 6% of total bilateral trade between the two regions, its contribution to the total trade cost reduction is also small. In general we find that the studies working with a bottom-up approach based on micro-data on NTMs (Fontagne et al., 2013; Francois et al., 2013) come to smaller trade cost reductions than the top-down approaches based on average FTA effects (the rest of the studies). Both approaches are subject to criticism. In the bottom up approach the share of

[25] 7.5% deviates from the 6.7% reported in Table 2 because the numbers in the table are based on trade-weighted average AVEs of TTIP.

NTMs that will be reduced as a result of TTIP is hard to motivate and a percentage like 25% or 50% always seems somewhat arbitrary. Furthermore, NTM data are often of poor quality with negative estimated AVEs for many products. The top-down approach, on the other hand, can be criticised for the fact that it is debatable whether TTIP will create similar NTM effects as in past deep-FTAs. Moreover, modellers need to pay attention to estimation details of choosing the proper instrument, the correct FTA measure and the adequate functional form of the gravity equation (PPML).[26]

Comparing the studies within the two groups reveals that the CEPII study generates much larger NTM levels than the CEPR study. Furthermore, it is encouraging that the average trade cost reduction in the study by Aichele et al. (2014) is similar to the reduction in Egger et al. (2015), given that a similar methodology was used. Both studies calculate the impact of a move from no FTA to a deep FTA based on the same depth of FTA data. The difference is the operationalisation. Egger et al. (2015) maintain the 0-7 scale, whereas Aichele et al. (2014) convert this scale into two dummies, one for shallow FTAs and one for deep FTAs. For the agriculture sectors, nonetheless, there are notable differences, where Aichele et al. (2014) find much larger trade cost reductions than Egger et al. (2015). There are three important differences between the two gravity estimates. First the instruments used are different. Second, the estimation method differs: PPML in Egger et al. (2015) versus IV in Aichele et al. (2014). Finally, Egger et al. (2015) include a separate dummy for trade between EU members, thereby driving down the coefficient on the FTA-depth variable.

To summarise, it is difficult to give a value judgement on the expected trade cost reductions corresponding with TTIP. However, given the limitations on their estimation techniques, we consider that the approximately 16% NTM reductions reported in Felbermayr et al. (2013a) and Felbermayr et al. (2015) seem over-estimated. In particular, these studies do not account for the depth of FTAs and assume that TTIP will be similar to any average FTA in the past. This is an unrealistic assumption since most FTAs have strongly focused on relatively large tariff reductions in the past, but these are not possible within TTIP given the relatively low existing tariff levels. Furthermore, the effect of TTIP seems to rise significantly by not including a separate dummy for intra-EU trade. Carrere et al. (2015), moreover, do not account for the depth of FTAs and fail to correct for the endogeneity of FTAs. Therefore, we conclude that a reasonable lower bound for the average trade cost reduction as a result of TTIP is the 3% values in Francois et al. (2013) and a reasonable upper bound is the 13% values in Egger et al. (2015) and Aichele et al. (2014).

5 Comparison of methods to calculate the welfare effects of trade cost reductions: theory

5.1 Theoretical model

Since a more extensive model to show the differences between solution methods and baseline calibration is not needed, a simple single-sector Anderson and van Wincoop (2003) endowment economy without intermediate linkages is employed. The Anderson and Van Wincoop endowment economy is equivalent to an Eaton and Kortum (2002) economy in terms of the welfare effects of trade policy experiments upon reinterpretation of the trade elasticity. In a single-sector setting without endogenous factor accumulation it is also equivalent to a Krugman economy (See Arkolakis et al. (2012)). Each country i has endowments equal to L_i. Preferences in each importer j are characterized by Armington love-of-variety preferences across goods from

[26] All studies using the top down approach account for endogeneity of FTAs, although in a different way. All studies estimate the gravity equation using PPML, except for Aichele et al. (2014) who use Instrumental Variables (IV).

different sourcing countries. With this setup the value of trade from country i to j is given by:

$$V_{ij} = (t_{ij} w_i)^{1-\sigma} P_j^{\sigma-1} E_j \tag{18}$$

With t_{ij} iceberg trade costs, w_i the price of endowments in country i, E_j expenditure in country j, σ the substitution elasticity, and P_j the price index defined as:

$$P_j = \left(\sum_{i=1}^{J} (t_{ij} w_i)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{19}$$

Equilibrium requires that the value of sales from country i to all its trading partners, $\sum_{j=1}^{J} V_{ij}$, is equal to the value of endowments, $w_i L_i$:

$$w_i L_i = \sum_{j=1}^{J} (t_{ij} w_i)^{1-\sigma} P_j^{\sigma-1} (1 + D_j) w_j L_j \tag{20}$$

D_j is the trade deficit ratio, $D_j = \frac{E_j - w_j L_j}{w_j L_j}$, and it is assumed to be fixed. Equilibrium of the economy is given by a solution of equations (19)-(20) for w_i and P_j. These equations can easily be rewritten into the following two equations employed by Anderson and Van Wincoop, thus solving for inward and outward multilateral resistance (MR), respectively P_j and Π_i:

$$P_j = \left(\sum_{i=1}^{J} t_{ij}^{1-\sigma} \Pi_i^{\sigma-1} Y_i / Y_W \right)^{\frac{1}{1-\sigma}} \tag{21}$$

$$\Pi_i = \left(\sum_{j=1}^{J} t_{ij}^{1-\sigma} P_j^{\sigma-1} (1 + D_j) Y_j / Y_W \right)^{\frac{1}{1-\sigma}} \tag{22}$$

Y_i is income in country i, $Y_i = w_i L_i$, and Y_W is world income. The corresponding Anderson and Van Wincoop gravity equation is given by:

$$V_{ij} = t_{ij}^{1-\sigma} \Pi_i^{\sigma-1} P_j^{\sigma-1} \frac{Y_i E_j}{Y_W} \tag{23}$$

5.2 Gravity estimation

The theoretical gravity equation in (18) corresponds with the following empirical gravity equation:

$$V_{ij}^{obs} = V_{ij} u_{ij} = (T_{ij} \eta_{ij}) X_i M_j \varepsilon_{ij} u_{ij} \tag{24}$$

V_{ij}^{obs} are the observed, actual trade flows, equal to the trade value V_{ij} times a measurement error, u_{ij}.[27] X_i and M_j capture exporter and importer specific variation and T_{ij} is a function of bilateral observables like distance, and dummies for sharing a border, common colony, domestic trade flows and the presence of a free trade agreement (FTA).[28] η_{ij} represents unobserved variation in trade flows not captured by the observables in T_{ij} thus generating misspecification error. η_{ij} could consist of unobservable trade frictions for example. ε_{ij} is the random disturbance

[27] Time subscripts are omitted. In the simulations both cross-section and panel data settings are explored and time subscripts will be introduced when appropriate.

[28] As they would not affect the main conclusions of the analysis, tariffs are omitted from the analysis.

in the value of trade. Observationally, η_{ij}, u_{ij}, and ε_{ij} can not be distinguished.[29] Rearranging (24) gives the following estimating equation:

$$V_{ij}^{obs} = T_{ij} X_i M_j \omega_{ij} \tag{25}$$

With:

$$\omega_{ij} = \eta_{ij} \varepsilon_{ij} u_{ij} \tag{26}$$

The problem for baseline calibration is that it is unknown how large η_{ij}, ε_{ij}, and u_{ij} are and that they might generate errors in the baseline calibration of trade costs t_{ij} or import shares.[30] Comparing equations (23)-(24) shows that iceberg trade costs (raised to the power $1-\sigma$), $t_{ij}^{1-\sigma}$, consist of observable and unobservable trade costs and random variation:[31]

$$t_{ij}^{1-\sigma} = T_{ij} \eta_{ij} \varepsilon_{ij} \tag{27}$$

When calculating baseline trade costs or baseline import shares, unobserved trade costs η_{ij} should be taken into account, but ε_{ij} and u_{ij} not. Phrased differently, the correct expression to use for t_{ij} in counterfactual experiments is given by the expectation conditional on both observed and unobserved trade costs:

$$t_{ij}^{correct} = \left(E\left[t_{ij}^{1-\sigma} | T_{ij}, \eta_{ij} \right] \right)^{\frac{1}{1-\sigma}} = \left(T_{ij} \eta_{ij} \right)^{\frac{1}{1-\sigma}} \tag{28}$$

η_{ij} is the unobserved component of trade costs, so it is part of trade costs and should be taken into account in calibrating baseline trade costs. ε_{ij} is a random component in trade costs which is different each time data are drawn from the sample, so in practice, each year trade data are observed. So in a correct counterfactual experiment, the influence of random variation should be neglected. A researcher is interested in the predicted welfare effect based on average trade costs and not based on actual trade costs for a specific observation (so in a specific year). u_{ij} is measurement error in the value of trade and should henceforth also be neglected in calibration of the baseline trade costs or import shares. As shown below, the structural gravity approach neglects the error terms and can be expected to perform well relative to the other approaches if ε_{ij} and u_{ij} are large and η_{ij}, whereas the CGE- and EHA-approaches take into account the error term and thus can be expected to perform well relative to th SG-approach if η_{ij} is large and ε_{ij} and u_{ij} are small.

5.3 Four approaches to calculate the welfare effects of a counterfactual trade experiment

Four approaches to calculate the welfare effects of a counterfactual experiment can be distinguished in the literature. The four approaches differ in two ways, the baseline calibration and the solution method. The four approaches are first presented and then systematically compared based on differences in baseline calibration and solution method. The size of the shock to iceberg trade costs will be identical as well as the substitution elasticity σ. All the approaches solve for the relative change in welfare W_i as a result of a policy experiment to t_{ij}, $\widetilde{W}_i$. The relative change in welfare can be written as follows:

$$\widetilde{W}_i = \frac{W_{c,i}}{W_i} - 1 = \frac{w_{c,i}}{w_i} \frac{P_{c,i}}{P_i} - 1 = \frac{Y_{c,i}}{Y_i} \frac{P_{c,i}}{P_i} - 1 \tag{29}$$

[29] Writing measurement error u_{ij} as a separate term is in line with the way measurement error in dependent variables is usually examined (See for example chapter 4 of Wooldridge (2002)). Unobserved variation η_{ij} and random variation ε_{ij} are distinguished for the purpose of identifying the different sources of bias in the calculation of the welfare effects of counterfactual experiments.

[30] As pointed out below, solution of a counterfactual exercise requires either values for baseline trade costs or baseline import and export shares.

[31] It is assumed that $1-\sigma$ is a fixed, non-stochastic parameter, implying that the transformation from t_{ij} to $t_{ij}^{1-\sigma}$ is not affected by potential variation in σ.

The subscript c stands for counterfactual. Since endowments are fixed, the change in real wages is identical to the change in real GDP per capita. Moreover, since a constant trade deficit ratio is imposed, the ratio of expenditures E_j and output Y_i are identical and the relative change in real wages and real GDP are equal and also generate the relative change in welfare (the equivalent variation of a counterfactual experiment).

5.3.1 Structural gravity

The structural gravity approach emerging from Anderson and van Wincoop (2003) solves the model in levels, both without and with changes in trade costs, so with initial and counterfactual trade costs. So this approach first solves for a baseline and then a counterfactual and compares the two outcomes. The baseline trade costs come from the predicted trade costs of the estimated gravity equation.

The baseline values of Π_i and P_j are solved from equations (21)-(22) employing the actual values for Y_i and E_j and the fitted values for iceberg trade costs, t_{ij}^{sg}:

$$t_{ij}^{sg} = (T_{ij})^{\frac{1}{1-\sigma}} \tag{30}$$

Hence, this approach neglects the error terms with as advantage that measurement errors u_{ij} and random variation ε_{ij} in trade flows are not taken into account.[32] The disadvantage is that unobserved trade costs, η_{ij}, are neglected.

The counterfactual is solved from the same equations, replacing baseline trade costs, t_{ij}^{sg}, by counterfactual trade costs, $t_{c,ij}^{sg}$, and adding an equation to solve for counterfactual GDP following from the assumption that endowment L_i is fixed:

$$\frac{Y_{c,i}}{Y_i} = \left(\frac{\Pi_i}{\Pi_{c,i}}\right)^{\frac{\sigma-1}{\sigma}} \tag{31}$$

So the procedure is to first solve for baseline values of P_i and Π_i (baseline Y_i taken from the data) from equations (21)-(22) and then calculate counterfactual values for $P_{c,i}$, $\Pi_{c,i}$ and income $Y_{c,i}$ using equations (21)-(22) and (31).[33] $t_{c,ij}^{sg}$ is calculated by changing the value of the policy variable, which is part of the fitted trade costs $\widehat{T_{ijt}}$. The simulations will focus on the example of the welfare effects of the introduction of an FTA, corresponding with a change in the value of the FTA-dummy from 0 to 1 for the countries introducing an FTA. Counterfactual trade costs for countries i and j introducing the FTA are equal to:[34]

$$t_{c,ij}^{sg} = t_{ij}^{sg}\left(\frac{T_{c,ij}}{T_{ij}}\right)^{\frac{1}{1-\sigma}} = t_{ij}^{sg}\exp\left(\frac{\beta_{FTA}}{1-\sigma}\right) \tag{32}$$

Fally (2015) has shown that under estimation of the gravity equation with PPML, the multilateral resistance terms can be calculated based on the fixed effects. In particular, the

[32] Measurement error could affect the fitted iceberg trade costs as discussed further below.

[33] Derivations are in the webappendix, which also shows that solving for the price index and the price of input bundles/endowments generates the same results.

[34] Observable trade costs T_{ij} can be written as $T_{ij} = \exp\{\beta_{FTA}FTA_{ij} + \beta_{other}other_{ij}\}$, with FTA_{ij} a dummy for the presence of an FTA and $other_{ij}$ a vector of other gravity regressors implying the expression used for $T_{c,ij}/T_{ij}$ in equation (32).

following equations hold with Y_i and E_j actual output and expenditures:[35]

$$X_i = \Pi_i^{\sigma-1} \frac{Y_i}{Y_W} \tag{33}$$

$$M_j = P_j^{\sigma-1} E_j \tag{34}$$

Anderson et al. (2015) propose to use the finding by Fally to run counterfactual experiments by estimation in STATA instead of simulation.[36] In particular, they estimate the gravity equation, calculate the MR-terms Π_i and P_j from the fixed effects X_i and M_j using equations (33)-(34). Then the gravity equation (25) is re-estimated imposing counterfactual values for the policy variable to recompute the new values of the MR-terms, $\Pi_{c,i}$ and $P_{c,i}$ from the fixed effects $X_{c,i}$ and $M_{c,j}$. Counterfactual GDP $Y_{c,i}$ is calculated based on the counterfactual fixed effect $X_{c,i}$ using the following expression:

$$Y_{c,i} = \left(\frac{X_{c,i}}{X_i}\right)^{\frac{1}{1-\sigma}} Y_i \tag{35}$$

Then the gravity equation is re-estimated imposing new values for trade V_{ij} based on the theoretical gravity equation in (23):

$$V_{ij}^c = \left(\frac{t_{c,ij}^{sg}}{t_{ij}^{sg}}\right)^{1-\sigma} \frac{Y_i^c E_j^c}{Y_i E_j} \frac{\Pi_i^{1-\sigma} P_j^{1-\sigma}}{\Pi_{i,c}^{1-\sigma} P_{j,c}^{1-\sigma}} V_{ij} \tag{36}$$

The authors then iterate between calculation of counterfactual MRTs and GDP, respectively, $\Pi_{c,i}$, $P_{c,i}$, and $Y_{c,i}$ and gravity estimation until convergence.

5.3.2 CGE-in-levels

CGE-models such as GTAPinGAMS (Lanz and Rutherford (2016)) and a flexible framework of different CGE models (Britz and van der Mensbrugghe (2017)) solve like structural gravity for both the baseline and counterfactual values of the endogenous variables and compare the difference to calculate percentage changes in welfare (and possibly other outcome measures). [37] Equations (19)-(20) are solved for the endogenous variables w_i and P_j both with baseline and counterfactual trade costs. Counterfactual trade costs are calculated as a function of baseline iceberg trade costs employing an equation similar to (32):

$$t_{c,ij}^{cge} = t_{ij}^{cge} \exp\left(\frac{\beta_{FTA}}{1-\sigma}\right) \tag{37}$$

The crucial difference between the SG-approach and the CGE-in-levels approach is the calibration of baseline trade costs. Whereas the SG-approach calibrates baseline trade costs employing predicted trade costs from the gravity equation, the CGE-in-levels approach calibrates baseline trade costs such that baseline import shares are equal to actual import shares in the data. Normalizing baseline wages w_i and price levels P_j to 1, this corresponds with the following expression for t_{ij}^{cge} (from equation (18)):[38]

$$t_{ij}^{cge} = \left(\frac{V_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}} \tag{38}$$

[35] Fally (2015) has shown that fitted income and expenditures are equal to actual income and expenditures when estimating with PPML and upon inclusion of exporter and importer fixed effects, i.e. $Y_i = \widehat{Y_i}$, and $E_j = \widehat{E_j}$. Employing this information and combining the structural and empirical gravity equation, respectively equations (23) and (25), leads then to equations (33)-(34).

[36] In the SG-literature researchers typically employ MATLAB to solve the system of non-linear equations.

[37] In this approach scholars typically works with the software GAMS.

[38] Yotov et al. (2016) call this approach "estibration. " With estibration, baseline trade costs are given by $t_{ij}^{1-\sigma} = T_{ij}\omega_{ij}$ (see page 91 of Yotov et al. (2016)). As shown in equation (46) this is equivalent to the CGE-in-levels calibration in equation (38) and implies that baseline import shares are equal to actual import shares in the data.

It can easily be shown that $w_i = P_j = 1$ is a solution of the equilibrium equations (19)-(20), given equation (38). This is therefore a convenient and harmless normalization.

5.3.3 CGE-in-relative-changes

CGE models such as the GTAP-model (Hertel (2013a)) write the equilibrium equations in relative changes. Log differentiating the equilibrium equations (19)-(20) leads to:

$$\widetilde{P_j} = \sum_{i=1}^{J} impsh_{ij} \left(\widetilde{t_{ij}} + \widetilde{w_i} \right) \tag{39}$$

$$\widetilde{w_i} = \sum_{j=1}^{J} expsh_{ij} \left((1-\sigma) \left(\widetilde{t_{ij}} + \widetilde{w_i} - \widetilde{P_j} \right) + \widetilde{w_j} \right) \tag{40}$$

With $\widetilde{x} = \frac{x_c - x}{x}$. In this approach the import and export shares, $impsh_{ij}$ and $expsh_{ij}$, are equal to the shares in the data.

Solving equations (39)-(40) as such would obviously lead to inexact solutions in case of larger shocks, as a first-order approximation is employed. However, the software employed in the CGE-in-relative-changes literature, GEMPACK, calculates the solution of a counterfactual experiment using multiple steps. This means that the first-order approximation actually becomes a higher-order approximation leading to accurate solutions. [39].

5.3.4 Exact hat algebra (EHA)

Exact hat algebra, introduced in the trade literature by Dekle et al. (2008), solves exactly for the ratio of the counterfactual and baseline endogenous variables. Dividing equations (19)-(20) in the counterfactual and baseline and rearranging gives:

$$\widehat{P_j} = \left(\sum_{i=1}^{J} impsh_{ij} \left(\widehat{t_{ij}} \widehat{w_i} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{41}$$

$$\widehat{w_i} = \sum_{j=1}^{J} expsh_{ij} \left(\widehat{t_{ij}} \widehat{w_i} \right)^{1-\sigma} \widehat{P_j}^{\sigma-1} \widehat{w_j} \tag{42}$$

With $\widehat{x} = \frac{x_c}{x}$.

As in the CGE models, the import and export shares, $impsh_{ij}$ and $expsh_{ij}$, are set equal to the shares in the data in the EHA-literature. [40]

5.4 Differences between methods: baseline calibration and solution method

The exposition in the previous section shows that the four approaches differ with respect to baseline calibration and solution method. The second, third, and fourth approach calibrate the baseline such that baseline import and export shares are equal to actual shares in the data. CGE-in-relative-changes and exact hat algebra work explicitly with these shares and set them equal to the actual shares. CGE-in-levels calibrates the baseline trade costs such that the import and export shares are equal to the actual shares. The baseline import shares under structural gravity instead are equal to the fitted or predicted import shares following from the estimated gravity equation. Substituting equation (30) into (23), and applying equations (33)-(34), gives:

$$impsh_{ij}^{sg} = \frac{V_{ij}}{E_j} = \left(t_{ij}^{sg} \right)^{1-\sigma} \Pi_i^{\sigma-1} P_j^{\sigma-1} \frac{Y_i}{Y_W} = \frac{T_{ij} X_i M_j}{E_j} = \frac{V_{ij}^{fitted}}{E_j} \tag{43}$$

[39] Further details on solution methods in GEMPACK can be found for example in Horridge et al. (2013)
[40] Scholars applying exact hat algebra typically work with MATLAB.

Using the empirical gravity equation in (25), the baseline calibration of structural gravity (SG) can be compared with the other approaches, exact hat algebra (EHA) and CGE-in-levels and CGE-in-relative-changes (CGE):

$$impsh_{ij}^{sg} = \frac{T_{ij} X_i M_j}{E_j} \tag{44}$$

$$impsh_{ij}^{eha,cge} = \frac{T_{ij} X_i M_j \eta_{ij} u_{ij} \varepsilon_{ij}}{E_j} \tag{45}$$

Alternatively, the baseline trade costs can be compared. Substituting the empirical expression for the actual value of trade V_{ij} in equation (25) into the expression for t_{ij} under CGE-calibration, equation (38), gives:

$$t_{ij}^{cge} = \left(\frac{V_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}} = \left(\frac{T_{ij}\eta_{ij} X_i M_j u_{ij}\varepsilon_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}} = \left(T_{ij}\eta_{ij} u_{ij}\varepsilon_{ij}\right)^{\frac{1}{1-\sigma}} \tag{46}$$

Under SG-calibration the error terms are neglected. The expression is given in equation (30).

Hence, the difference between SG-calibration on the one hand and CGE- and EHA-calibration on the other hand is whether the terms η_{ij}, u_{ij}, and ε_{ij} are taken into account or not. This follows both from comparing the baseline import shares in equations (44)-(45) and the baseline trade costs in (30) and (46). However, based on equation (28) unobserved trade costs η_{ij} should be taken into account, whereas random variation ε_{ij} and measurement error u_{ij} should not be taken into account. This implies that it can be expected that CGE- and EHA-calibration will perform better if unobserved trade costs η_{ij} are large relative to measurement error u_{ij} and random variation ε_{ij}. SG instead is expected to perform better if unobserved trade costs are negligible and measurement error and random variation are large. In Section 7 the robustness of the different methods to misspecification, to random variation, and to measurement error is examined with simulations.

Comparing the solution methods, CGE-in-levels and EHA should generate exactly the same results. In the first approach the model has to be solved twice, whereas in the second approach only one simulation is needed. However, both methods are based on the same equilibrium equations and should thus give identical results for the same counterfactual experiment. In the SG-approach the same counterfactual experiment will obviously lead to different outcomes than with the other approaches, since baseline calibration is different. With the same baseline calibration the solution method employed in the SG-approach will, however, also lead to the same solution as CGE-in-levels, since the equilibrium equations are equivalent. More formally, the following holds:

Result 1 *Identical changes in counterfactual iceberg trade cost, $\widehat{t_{ij}}$, lead to identical changes in welfare, given either:*

(i) Identical import shares, $impsh_{ij}$, and export shares, $expsh_{ij}$, in the baseline

(ii) Identical trade costs, t_{ij}, output, Y_i, and trade deficit ratio, D_j, in the baseline

Proof: From equation (29) the percentage change in welfare, $\widetilde{W}_i$, is determined by $\widehat{w_i}$ and $\widehat{P_i}$. From equations (41)-(42), $\widehat{w_i}$ and $\widehat{P_i}$ are determined by the baseline import and export shares, $impsh_{ij}$ and $expsh_{ij}$, and the change in iceberg trade costs $\widehat{t_{ij}}$, thus establishing (i). Given equation (18), $impsh_{ij} = \frac{V_{ij}}{(1+D_j)Y_j}$ and $expsh_{ij} = \frac{V_{ij}}{Y_i}$ are determined by t_{ij}, D_j, w_j, P_j, and Y_i. Substituting equation (19) into equation (20) leads to:

$$Y_i = \sum_{j=1}^{J} \left(\frac{t_{ij} w_i}{\sum\limits_{k=1}^{J} \left(t_{kj} w_k\right)^{1-\sigma}}\right)^{1-\sigma} \left(1 + D_j\right) Y_j \tag{47}$$

Solution method: Baseline calibration to:	Compare levels with and without policy experiment	Equilibrium in relative changes	Calculate ratio of old and new equilibrium values
Fitted shares	Structural gravity (SG)		
Actual shares	CGE-in-levels	CGE-in -relative-changes	Exact hat algebra (EHA)

Table 3: Comparison of four methods based on choice of solution method and baseline calibration

It can be shown that equation (47) contains a unique solution for w_i for all i as a function of t_{ij}, D_j and Y_j (See for example Alvarez and Lucas (2007)). Given that w_i determines P_j from equation (41), this implies that t_{ij}, D_j, and Y_i determine $impsh_{ij}$ and $expsh_{ij}$ and thus the welfare effects of a change in trade costs, $\widehat{t_{ij}}$.

Based on Result 1 two remarks can be made on approaches followed in the literature. First, there is no difference between solving for w_i and P_j as in CGE-in-levels and solving for Π_i and P_j as in SG, as long as the baseline t_{ij}, D_j, and Y_i or the import and export shares are identical. For example, different baseline trade costs t_{ij} can be used as long as the implied import and export shares are identical. In the appendix it is shown that the solution procedure with t_{ij} based on the bilateral terms in the fitted gravity equation as in equation (30) and Π_i and P_j determined by the estimated fixed effects (GEPPML) leads to the same baseline import and export shares as calibrating the baseline to the fitted import shares working with $w_i = P_j = 1$ in the baseline. This corresponds with different baseline values for t_{ij}, but since baseline import shares are identical the welfare effects of a counterfactual trade policy experiment will be the same.

Second, the welfare effects of counterfactual experiments in the calibration approach followed by for example Alvarez and Lucas (2007) and Levchenko and Zhang (2016) do not hinge on the chosen solution method for wages and "effective labor" in the baseline, but are determined by the way baseline trade costs are set. In this approach values for trade costs and trade shifters are set at specific values (Alvarez and Lucas (2007)) respectively based on gravity estimation (Levchenko and Zhang (2016)). A fixed point iteration procedure is then used to calculate baseline wages and baseline effective labor, imposing that wages times effective labor are equal to output in the data. However, the exact solution procedure and solution for effective labor do not affect the impact of counterfactual experiments, which are determined by the values for baseline trade costs and income only. Since fitted gravity-based trade costs are employed, this strand of literature thus follows the SG-approach to baseline calibration. As per the first remark a calibration procedure with $w_i = P_j = 1$ in the baseline would generate the same effects of counterfactual experiments as long as the baseline import and export shares are identical.

Table 3 provides a comparison of the four methodologies based on the solution method and the baseline calibration. As shown in the next section, other combinations of solution method and baseline calibration than currently made in the literature are possible. The exact hat algebra solution method can for example be combined with baseline calibration to fitted import shares.

6 Comparison of methods: simulations

6.1 Introduction

In this section the impact of the solution method and baseline calibration on the welfare effects of a counterfactual experiment is assessed with simulations. The model described in Subsection

Table 4: Wellfare effect of reducing trade costs between the EU and the USA (TTIP) with calibration to actual import shares employing different solution methods

Solution method	SG	CGE in levels	CGE rel. changes	EHA
Welfare effects (perc. change)				
Population weighted				
EU	.4472	.4472	.4472	.4472
USA	.5993	.5993	.5993	.5993
Third countries	-.0149	-.0149	-.0149	-.0149
All countries	.0464	.0464	.0464	.0464
GDP weighted				
EU	.5073	.5073	.5073	.5073
USA	.5993	.5993	.5993	.5993
Third countries	-.0155	-.0155	-.0156	-.0156
All countries	.2332	.2332	.2332	.2332

5.1 is used and calibrated to 121 countries using GTAP9 data. No intermediate linkages are included and a single-sector model is employed, as this is sufficient to compare the different approaches to baseline calibration. The 120 countries available in GTAP9 are employed and the rest of the regions are merged into one rest-of-the-world region. The counterfactual experiment to compare solution methods and baseline calibration is a reduction in trade costs between the European Union countries and the USA (TTIP-experiment). Without loss of generality the size of the shock comes from Felbermayr et al. (2015). $\beta_{FTA} = 1.21$ and $\sigma = 8$ as reported in their paper imply from equation (32) a shock of about 14%.

6.2 The impact of solution methods

Table 4 displays the welfare effects of the solution methods described in 5.4.[41] In this subsection random variation in trade flows is neglected, focusing instead on differences in solution methods, thus imposing $\omega_{ij} = 1$. The table shows that CGE-in-levels, EHA, and SG calibrated to actual import shares (called estibration by Yotov et al. (2016), see footnote 9) all lead to exactly the same solution.[42] The table also shows that CGE-in-relative-changes generates identical results as well. Horridge et al. (2013) have conducted a similar exercise, showing that GEMPACK and GAMS-versions of the same multi-sector multi-country model generate identical outcomes on counterfactual exercises. Hence, multi-step log-differentation of the model generates the same results. How the different solution methods and also softwares compare in larger-scale models with multiple sectors and factors, intermediate linkages, and endogenous factor accumulation is an open question left for future research.

Table 5 shows the results of employing the four solution methods based on calibration to fitted trade costs (and import shares). The table shows that also with a different baseline calibration the solution methods generate identical welfare effects. These results show that there is hence no one-to-one link between solution method and baseline calibration. For example, exact hat algebra or CGE-in-relative-changes can be combined with baseline calibration to fitted trade costs. Also in a more extensive CGE-model no rebalancing of the other data of the model

[41]The equilibrium equations, baseline calibration, and starting values are provided in Appendix B.1.

[42]CGE-in-levels, EHA and estibration are progammed in GAMS and CGE-in-relative-changes in GEMPACK. Although scholars applying EHA and SG typically use MATLAB this does not make a difference in this single-sector model.

Table 5: Wellfare effect of reducing trade costs between the EU and the USA (TTIP) with calibration to fitted import shares employing different solution methods

Solution method	SG1	SG2	CGE in levels	CGE rel. changes	EHA
Welfare effects (perc. change)					
Population weighted					
EU	.2579	.2578	.2578	.2578	.2578
USA	.2666	.2666	.2659	.2666	.2666
Third countries	-.0105	-.0105	-.0104	-.0106	-.0106
All countries	.0215	.0215	.0215	.0214	.0214
GDP weighted					
EU	.2269	.227	.2268	.2268	.2268
USA	.2666	.2666	.2659	.2666	.2666
Third countries	-.0087	-.0091	-.0087	-.0092	-.0092
All countries	.103	.1026	.1029	.1027	.1027

Notes: SG1 solves for Π_i and P_j using calibrated import shares to determine t_{ij} and normalizing P_j at 1 and Π_i at $(Y_j/Y_W)^{(1/(\sigma-1))}$, whereas SG2 solves for Π_i and P_j using fitted trade costs T_{ij} to determine t_{ij} and Π_i and P_j in the baseline determined by the fixed effects according to equations (33)-(34).

would be necessary with calibration to fitted trade costs. The reason is that PPML estimation of the gravity equation including exporter and importer fixed effects implies that the sum of fitted exports and the sum of fitted imports are equal to respectively the sum of actual exports and imports. Practically this means that if the simulations below would indicate that calibration to fitted shares outperforms calibration to actual shares, CGE models could be easily recalibrated only changing baseline international trade values.

6.3 The impact of baseline calibration

This subsection explores the impact of variations in baseline calibration on predicted welfare effects of a set of counterfactual exercises without going into the causes of those variations, which could be unobserved trade costs, measurement error in trade flows, or random variation. The model described in Section 5.1 is employed and only one solution method, CGE-in-levels, is used, since the solution method does not affect the predicted welfare effects as shown above. Hence, equations (19)-(20) are solved for the endogenous variables w_i and P_j with both baseline and counterfactual trade costs and the welfare change is calculated according to equation (29).

100 baseline trade values V_{ij} are generated according to $V_{ij} = V_{ij}^{data} \upsilon_{ij}^{1-\sigma}$ with V_{ij}^{data} based on the trade flows in the GTAP9-data and $\ln \upsilon_{ij}$ is drawn from a standard normal distribution (and υ_{ij} thus a log-normal distribution). This subsection does not focus on the size of the deviations of predicted welfare effects from the mean, but only on its determinants. Therefore, working with a variance of 1 for $\ln \upsilon_{ij}$ is inconsequential.

With this setup baseline trade costs are given by $t_{ij}^{1-\sigma} = \frac{V_{ij}^{data}}{E_j^{data}} \frac{E_j^{data}}{E_j} \upsilon_{ij}^{1-\sigma}$.[43] So in terms of equation (27) the setup corresponds with $T_{ij}\eta_{ij} = \frac{V_{ij}^{data}}{E_j^{data}}$ and $\varepsilon_{ij} = \frac{E_j^{data}}{E_j} (\upsilon_{ij})^{1-\sigma}$.[44] Hence, applying equation (28), the trade costs implied by the data are by assumption the correct trade costs (corresponding with the expectation of trade costs). The counterfactual experiment is

[43]This expression for baseline trade costs corresponds with equilibrium values $w_i = P_j = 1$ as discussed in Subsection 5.3.2.

[44]The term E_j^{data}/E_j will be negligible given that there are 121 countries.

conducted both with the correct baseline data and with the 100 randomly simulated baseline data.

Table 6: Effect of import and export share with FTA-partner on welfare effects of the EU of the introduction of an FTA with the USA (TTIP)

	(1) diff_welfare	(2) diff_welfare	(3) diff_welfare	(4) diff_welfare (average across replications)	(5) diff_welfare (average across countries)
diff_impsh	0.23***	0.23***	0.23***	0.20***	0.19***
	(0.0010)	(0.0010)	(0.00098)	(0.010)	(0.0047)
diff_expsh	0.18***	0.18***	0.18***	0.21***	0.17***
	(0.00094)	(0.00095)	(0.00090)	(0.0066)	(0.0051)
diff_domsh	-0.0021***	-0.0021***	-0.0017***	-0.00049	0.00010
	(0.00020)	(0.00020)	(0.00019)	(0.0011)	(0.00092)
Observations	2800	2800	2800	28	100
R^2	0.98	0.98	0.98	1.00	0.99
Adjusted R^2	0.98	0.98	0.98	0.99	0.99
Country fixed effects	No	Yes	No	No	No
Replication fixed effects	No	No	Yes	No	No

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The EHA-equilibrium equations, (41)-(42), show that the import and export shares of the liberalizing countries determine the impact of the counterfactual, since these shares pre-multiply the shocks to trade costs. To study how these shares affect welfare predictions more into detail, the difference between the calculated and correct welfare effect is regressed on differences in the calculated and correct import and export shares with the liberalizing trading partner, and on the domestic spending share, as specified in the following equation:

$$\begin{aligned} diff\_welfare_{r,j} = \beta_0 + \beta_1 diff\_impsh_{r,j,partner} + \beta_2 diff\_expsh_{r,j,partner} \\ + \beta_3 \left(diff\_domsh_{r,j}\right) + \eta_j + \zeta_r + \varepsilon_{r,j} \end{aligned} \tag{48}$$

With $diff\_var_{r,j} = var\_random_{r,j} - var\_correct_j$ for the variables $var = welfare$, $impsh$, $expsh$, $domsh$ and η_j and ζ_r country- and replication fixed effects, respectively. Table 6 displays the results of this regression for the EU countries. In column 1 all observations are included and fixed effects omitted. In column 2 country fixed effects are added and the effect is thus identified by variation within countries across replications. In column 3 replication fixed effects are added. In columns 4 and 5 country- and replication-averages are used. The tables show that observations with larger import shares and export shares with the liberalizing trading partner, and large domestic shares than the correct shares display too large welfare effects. The estimated coefficient of 0.18 on $diff\_impsh$ means that on average an excess baseline import share of a European country from its TTIP partner the US of 1 percentage point generates an excess welfare effect of TTIP of 0.18%. The size of the deviation of the welfare effect will of course also rise with the size of the trade cost shock and the trade elasticity. However, the exercise shows that small deviations in import shares can already lead to large changes in welfare effects for reasonable trade cost shocks.

Table 6 shows as well that the deviations from the mean import and export shares vis-a-vis the FTA partner and from the mean domestic spending shares explain almost all of the variation in the bias of the welfare effects. Result 1 suggests that it also the other trade shares matter, but the table shows that they play a minor role.

Figure 1 displays scatter plots of the averages of $diff\_welfare$ and $diff\_impsh$ across importers and replications, respectively. The figure illustrates the findings of Table 6. A larger import share from the US leads to a larger welfare effect.

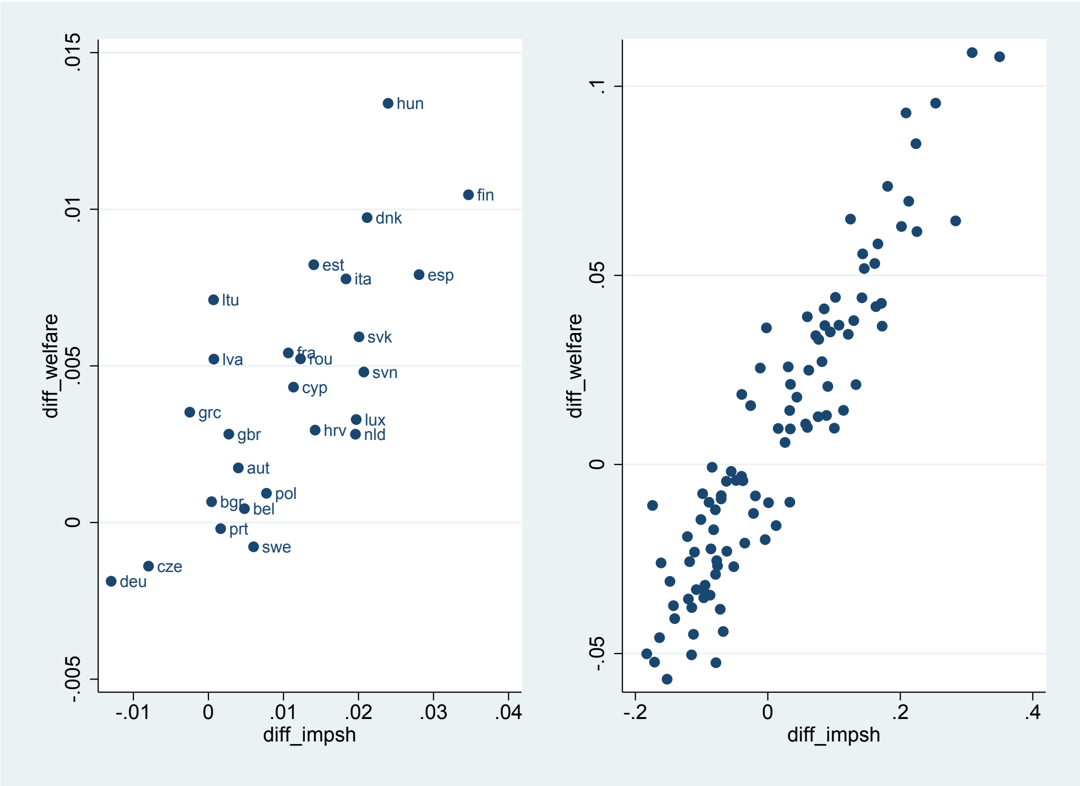

Figure 1: The impact of deviations in import shares from the US from their mean on deviations of welfare effects of the EU countries from their mean after a reduction in trade costs between the EU and the US (TTIP)

The online appendix shows similar patterns for the US under TTIP and for other counterfactual experiments. Larger import and export shares vis-a-vis the EU lead to an upward bias in the predicted welfare effect for the US. Three other counterfactual experiments are conducted: first, a bilateral trade cost reduction between only two countries, Mexico and the US; second, a unilateral trade cost reduction in one country, Mexico; and third, a multilateral reduction between trade costs in all countries. In each of the cases the trade cost reduction is identical, 14%. The additional experiments show that the upward bias in the welfare effects is consistently larger if the import share with the trading partners is larger. In particular the USA-Mexico FTA-experiment shows like the TTIP-experiment that larger import and export shares with the trading partner lead to an excess welfare effect. The unilateral trade liberalization experiment shows that a larger import share in Mexico leads to excess welfare effects, but also that larger import and export shares of the other countries with Mexico generate a too large effect as expected. The multilateral experiment shows as expected that a larger total import share of countries generates an upward bias. An import share of 10 percentage points more than the mean import share in the baseline calibration generates an increase in the welfare effect of 3 percentage points.

7 The importance of different sources of bias

The previous section has shown that deviations of baseline import and export shares from their mean can create large deviations in the predicted welfare effect of a counterfactual experiment. In this section the impact of three sources of error in baseline shares is explored: first, errors because of unobserved trade costs not accounted for because the gravity equation does not capture them and is thus misspecified; second, errors because of random variation in trade flows erroneously included in the baseline calibration; and third, errors driven by measurement error in observed trade flows. The three errors correspond respectively with variation in η_{ijt}, ε_{ijt}, and u_{ijt}. The first error generates a bias in calibration to fitted import shares, the second generates a bias in calibration to actual import shares, and the third generates a bias in both approaches. Subsection 7.1 only allows for bias as a result misspecification, Subsection 7.2 allows for bias as a result of both misspecification and random variation, and Subsection 7.3 allows for bias as a result of both misspecifaction and measurement error.

7.1 Bias as a result of misspecification

This section explores the potential bias of calibration to fitted trade costs as in structural gravity as a result of misspecification of the gravity equation. It is assumed that there is no random variation and no measurement error in the data, $u_{ij} = \varepsilon_{ij} = 1$. So welfare effects can be biased because the actual trade costs $t_{ij}^{cge} = (T_{ij}\eta_{ij})^{\frac{1}{1-\sigma}}$ are not equal to the fitted trade costs $t_{ij}^{sg} = (T_{ij})^{\frac{1}{1-\sigma}}$. Henceforth, by construction SG-calibration will be biased and CGE/EHA calibration is unbiased. The absence of random variation is a strong assumption, which makes it possible to focus on the impact of misspecification. Below random variation and measurement error are included, also generating bias in CGE- and EHA-calibration. The first subsection examines the predicted welfare effects in a setting with cross-section gravity estimation based on GTAP data and the second subsection explores these effects with panel gravity estimation based on WIOD data.

7.1.1 Cross-section gravity estimation

Table 7 displays the welfare effects of the same experiment as used before, so a shock of 14% to iceberg trade costs between the EU and the US, employing data on 121 countries from GTAP9 data for 2011. Column 1 shows the effects with calibration of baseline trade costs to actual import shares, which are assumed to be the correct welfare effects in the absence of random variation and measurement error in the trade flows. The remaining columns are based on structural gravity-type simulations, so with baseline trade costs calibrated to their fitted values from the gravity estimation. In columns two to six the gravity estimation is based on data not including domestic flows, as for example in Felbermayr et al. (2013b) and Felbermayr et al. (2015). Domestic trade costs are based on their fitted values. Since the bilateral explanatory variables are available for intra-country, out-of-sample observations, it is possible to generate fitted values. Columns two to six show that calibration to fitted values without using domestic flows generates a large upward bias in the welfare effects of the TTIP-experiment. This bias can be explained from the overestimated import shares of the EU and the USA vis-a-vis their FTA-partner (bottom rows of Table 7) in the baseline. The welfare effects in column two are close to the welfare effects reported in Felbermayr et al. (2015). With the same substitution elasticity, the same shock to iceberg trade costs and the same gravity variables in the regression, this suggests that Felbermayr et al. (2015) have worked with the described baseline calibration, omitting domestic trade flows and calibrating domestic trade costs to the out-of-sample fitted trade costs.

Columns two to six convey three other messages. First, the negative welfare effects for third countries are overestimated in the calibration based on gravity estimation without domestic

Table 7: Effects of TTIP on real income with GTAP data, comparing actual import shares calibration with gravity-predicted import share calibration based on different gravity specifications

Calibration to	Actual shares	Fitted shares								
Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Domestic flows		No	No	No	No	No	Yes	Yes	Yes	Yes
Gravity specification		(1)	(1)	(2)	(3)	(4)	(1)	(4)	(5)	(6)
Welfare effects (perc. change)										
Population weighted										
EU	.45	4.45	3.98	3.77	3.92	1.95	.16	.26	.58	.49
USA	.6	5.15	6.31	5.96	6.21	2.29	.18	.27	.65	.59
Third countries	-.01	-.68	-.5	-.55	-.5	0	-.01	-.01	-.02	-.01
All countries	.05	-.05	.13	.06	.13	.25	.01	.02	.05	.05
GDP weighted										
EU	.51	4.44	3.89	3.83	3.66	1.83	.15	.23	.53	.49
USA	.6	5.15	6.31	6.21	5.96	2.29	.18	.27	.65	.59
Third countries	-.02	-.96	-1.2	-1.19	-1.2	-.41	-.01	-.01	-.02	-.01
All countries	.23	1.55	1.56	1.53	1.43	.67	.07	.1	.25	.23
Trade shares (scaled to 100)										
Domestic shares										
EU	77.29	1.16	1.15	.94	1.21	6.9	63.32	57.08	62.85	77.29
USA	90.7	32.37	30.73	29.39	31.17	69.51	90.54	93.47	89.04	90.7
Third ountries	71.22	.96	1.01	1	1.05	6.16	45.2	60.95	65.55	71.22
All countries	76	1.37	1.36	1.19	1.42	7.25	59.35	58.28	63.69	76
Import shares partner										
EU	1.5	12.14	10.85	10.15	10.65	4.82	.35	.96	2.67	1.8
USA	.06	.43	.48	.43	.47	.2	.02	.03	.07	.07

Notes: All gravity specifications include exporter and importer fixed effects and the following regressors. Specification (1): log-distance, contiguity, common language and history of common colonizer; Specification (2): as in (1) but history of common colonizer replaced by current colonial relation; Specification (3): as in (2) and moreover the difference in political competition score from PolityIV; Specification (4): as specification (1) but domestic trade costs normalized at 1. Specification (5): as in (1) and moreover a dummy for domestic trade flows; Specification (6): as in (1) and moreover a dummy for domestic trade flows and a dummy for domestic trade flows interacted with GDP and GDP per capita; Specification (7): as in (1) and moreover a country-specific dummy for domestic trade flows. Column (2) uses GDP, the other columns gross output for Y_i. Columns (2) to (6) do not include domestic flows in the gravity estimation. Columns (7) to (10) do. In specifications with and without domestic trade flows, trade imbalances are modelled respectively not modelled.

flows, which is related to the underestimation of domestic spending shares in the baseline for these countries. As a result the negative trade diversion effects operate on too large trade shares with the FTA-partners thus overestimating these effects. Second, columns three, four, and five show that the welfare effects are sensitive to the included bilateral gravity variables, the only difference between these columns. Third, setting domestic trade costs in all countries at 1 reduces the welfare effects considerably, related to the fact that this leads to smaller baseline import shares vis-a-vis the FTA-partners.

The results reported in columns seven to ten are based on gravity estimations including domestic trade flows.[45] Based on gravity specification (1) in column seven the welfare effects are hugely underestimated, related to the fact that trade shares vis-a-vis FTA-partners are largely underestimated. Including a dummy for domestic trade flows in column eight improves the welfare calculations, but still leaves a downward bias. Including interaction terms of domestic trade flows with GDP and GDP per capita brings the fitted domestic spending shares and import shares vis-a-vis FTA-partners closer to the actual values especially for the USA. Thereby the welfare effects also come closer to the correct welfare effects. Finally, column ten includes country-specific dummies for domestic trade flows, thus leading to the correct domestic spending shares and thereby also bringing the welfare effects relatively close to the correct welfare effects. This suggests that the last two specifications are interesting candidates to explore when including random variation and measurement error in trade flows. One step further would be to include pairwise fixed effects, which would effectively exhaust all degrees of freedom and make SG-calibration based on fitted shares identical to CGE/EHA-calibration based on actual shares.

7.1.2 Panel gravity estimation

Table 8 displays the welfare effects of the TTIP-experiment employing gravity estimations based on panel data from WIOD for 2000-2014. Since the model is static, the baseline has to be calibrated to a specific year, which is set at 2011 without loss of generality. Phrased differently, it is assumed that welfare in 2011 is the main outcome variable of interest. Column one displays the results of calibration to the actual shares in 2011, whereas the remaining columns are based on fitted shares from various gravity estimations. In the first gravity specification a dummy for domestic flows is included in the gravity estimation, leading like in Table 7 to a downward bias of the welfare effect of almost 100% for the USA. In the second specification interactions of the domestic dummy with GDP and GDP per capita are added reducing the bias. The third specification includes country-specific dummies for domestic flows, reducing the bias further for the USA but raising it for the EU-average. Finally, the fourth specification contains pairwise (time-invariant) fixed effects. This leads to domestic spending shares and trade shares vis-a-vis the FTA-partner close to the actual trade shares in 2011 and thereby also welfare effects close to the welfare effects based on actual-shares-calibration. However, the fitted import share of the EU-countries from the USA is somewhat below the share in 2011 indicating a growing share over the sample period, since the pairwise fixed effects generate fitted import shares equal to the average over the sample period. The underestimated fitted import share of the EU-countries from the USA corresponds with a small underestimation in the welfare effect for the EU. To address trends in trade flows over the sample period, specification five employs fitted shares from a gravity estimation including both pairwise fixed effects and interactions of the pairwise fixed effects with a linear time trend. The predicted welfare effects under specification five are slightly closer to the welfare effects employing actual shares although the difference with specification four is marginal. The predicted welfare effect for the USA rises from 0.34 to 0.35, slightly closer to the effect under actual shares, 0.36. This finding is in line with a somewhat larger import share from the EU in the USA under specification five than specification four,

[45]The GTAP data contain domestic flows, but for example the COMTRADE data with a larger set of countries (up to 180) contain only international trade flows.

Table 8: Effects of TTIP on real income with WIOD panel data, comparing actual import shares calibration with gravity-predicted import share calibration based on different gravity specifications

Calibration to	Actual shares	Fitted shares				
Gravity specification		(1)	(2)	(3)	(4)	(5)
Welfare effects (perc. change)						
Population weighted						
EU	.36	.25	.35	.44	.34	.35
USA	.5	.27	.43	.49	.48	.48
Countries not in TTIP	-.01	-.01	-.01	-.01	-.01	-.01
All countries	.04	.02	.03	.04	.03	.04
GDP weighted						
EU	.43	.22	.35	.39	.4	.41
USA	.5	.27	.43	.49	.48	.48
Countries not in TTIP	-.01	-.01	-.01	-.01	-.01	-.01
All countries	.19	.1	.16	.18	.18	.18
Trade shares (scaled to 100)						
Domestic shares						
EU	85.45	84.33	85.81	84.31	85.92	85.94
USA	91.85	93.06	91.93	90.97	92.02	92.04
Countries not in TTIP	73.61	63.18	75.11	66.08	75.1	74.18
All countries	78.06	71.07	79.14	72.86	79.17	78.59
Import shares partner						
EU	1.54	.92	1.19	1.9	1.39	1.46
USA	.05	.04	.05	.06	.05	.05s

Notes: All gravity specifications include exporter and importer fixed effects and the following regressors. Specification (1): log-distance, contiguity, common language, history of common colonizer and a dummy for domestic trade flows; Specification (2): as in (1) and moreover interaction terms of the domestic dummy with GDP and GDP per capita; Specification (3): as in (1) and moreover a country-specific dummy for domestic trade flows; Specification (4): pairwise (time-invariant) fixed effects; Specification (5): pairwise (time-invariant) fixed effects and interactions of pairwise (time-invariant) fixed effects with a linear time trend.

respectively 1.46 and 1.39. The findings suggest that the gravity-based calibration with pairwise fixed effects can be improved upon if there is a trend in trade flows.

The findings in this section imply both a potential advantage and disadvantage of the gravity-based calibration with pairwise fixed effects. If there are large swings in trade shares across years, calibration to the actual shares in a specific year might pick up undesired random variation in these shares, as will be explored in the next subsection. At the same time gravity-based calibration to a specific year might miss trends in trade shares, which could become particularly pressing in a longer panel. As shown above, this problem can be addressed by including interactions of pairwise fixed effects with a time trend, slightly improving the outcomes. However, this specification is computationally very intensive.

7.2 Bias as a result of random variation in trade flows

This section explores the potential bias of calibration to actual shares in case of random variation in trade flows, corresponding with $\varepsilon_{ijt} \neq 1$ in gravity equation (24). Actual shares might pick up this random variation, although it should be disregarded. In this section measurement error is still omitted, so $u_{ijt} = 1$. Of course, the bias of calibration to actual shares will rise in the variance of ε_{ij}, so this variance should be disciplined. To do so, the random variation in the gravity equation estimated with actual data and pairwise fixed effects is employed. More formally, equation (25) is estimated with $T_{ijt} = \mu_{ij}$ with μ_{ij} pairwise fixed effects, employing the WIOD data for 2000 to 2014. The estimated variance of the error terms is then employed as a proxy for the variance of ε_{ijt}. The estimated variance varies by the size of trade flows and therefore the variance is estimated for each decile of trade flows.[46] Working with one variance for all trade flows, would overstate the variance in trade flows for larger observations.[47]

100 baseline trade values are generated, according to $V_{ijt} = V_{ijt}^{data}\varepsilon_{ijt}$ with V_{ijt}^{data} equal to the trade flows in the WIOD-data and $\ln \varepsilon_{ijt}$ drawn from a normal distribution with mean zero and variance as described above. As in Subsection 6.3 this setup implies that baseline trade costs are given by $t_{ijt}^{1-\sigma} = \frac{V_{ijt}^{data}}{E_{jt}^{data}} \frac{E_{jt}^{data}}{E_{jt}} \varepsilon_{ijt}$ and so by assumption the trade costs in the data are the correct trade costs. The counterfactual experiment is conducted both with the correct baseline data and with the 100 randomly simulated baseline data. This setup makes it possible to examine the impact of both random variation and misspecification together.

Almost the same estimators as in the previous section are then compared, in particular calibration to actual shares and calibration to fitted shares based on gravity estimation with (i) country-specific domestic dummies; (ii) a domestic dummy and its interaction with GDP and GDP per capita; and (iii) pairwise fixed effects. Using the actual data as the true data makes it possible to compare the misspecification bias in the previous section with the bias as a result of random variation in trade flows. The alternative would be to set up a data generating process accounting for both misspecification and random variation. The misspecification would then have to be disciplined, based on the data. Instead of using this roundabout way, the actual data are taken as the true data and random variation is added.

Table 9 displays the mean squared error (MSE) of the predicted welfare effects over the 100 simulations employing the different methods indicated, averaged over the countries in the different groups, either population- or GDP-weighted:

$$MSE_{group}^{rv} = \sum_{i \in group} \frac{weight_i \frac{1}{100} \sum_{rep=1}^{100} \left(\widetilde{W}_{i,rep}^{rv} - \widetilde{W}_{i,rep}^{correct} \right)^2}{weight_i}; weight = GDP, population \quad (49)$$

$\widetilde{W}_{i,rep}^{correct}$ and $\widetilde{W}_{i,rep}^{rv}$ are the percentage changes in welfare based on respectively the actual data and the data with random variation added. The table shows that calibration to actual shares and calibration to fitted shares based on pairwise fixed effects perform an order better than the other two methods based on fitted shares. The reason for the poor performance with the other two methods for especially the EU countries is that the two gravity equations misspecify the baseline shares, in particular in some of the EU countries. This was not clear in Table 8 in the previous subsection evaluating misspecification, as this table displayed average effects, obscuring upward and downward biases within the groups. Only for the USA the two fitted shares approaches without pairwise fixed effects generate relatively accurate predictions, which seems to be merely accidental.

Comparing the results based on actual shares and parwise fixed effects based fitted shares shows that the differences are small between these two approaches. The MSE is smaller for the EU-countries employing fitted shares, whereas pairwise fixed effects based fitted shares display a lower MSE for the USA and countries not in TTIP. The average MSE for all countries in the

[46] This is done separately for domestic and international trade flows, since the variance of domestic trade flows is an order smaller than of international trade flows. The results are in the online appendix.

[47] Egger and Nigai (2015) find as well that the variance of the predicted trade flows in estimation with actual data varies by decile, but do not impose a varying variance in their Monte Carlo exercises, which are entirely based on generated data and thus do not contain differences inherent differences across observations.

Table 9: The mean squared error of the predicted welfare effect under random variation of trade flows employing actual shares and fitted shares

Calibration to Gravity specification	Actual shares	Fitted shares (1)	 (2)	 (3)
Welfare effects (perc. change)				
Population weighted				
EU	.00549	.276	.142	.0112
USA	.000449	.0000389	.00300	.000184
Countries not in TTIP	2.75e-06	.0000120	4.05e-06	1.89e-06
All countries	.000417	.0199	.0104	.000815
GDP weighted				
EU	.00731	.285	.192	.0164
USA	.000449	.0000389	.00300	.000184
Countries not in TTIP	4.82e-06	.0000438	.0000100	3.08e-06
All countries	.00185	.0682	.0466	.00395

Notes: All gravity specifications include exporter and importer fixed effects and the following regressors. Specification (1): log-distance, contiguity, common language, history of common colonizer and country-specific dummies for domestic trade flows; Specification (2): as in (1) but instead of country-specific dummies for domestic trade flows, one dummy for domestic flows and interaction terms of a domestic dummy with GDP and GDP per capita; Specification (3): pairwise (time-invariant) fixed effects.

sample is smaller with actual shares. These results show that the bias generated by random variation under actual shares is comparable in size with the bias generated by misspecifaction under pairwise fixed effects based fitted shares.[48]

7.3 Bias as a result of measurement error

This section examines the bias as a result of measurement error in trade flows of both calibration to actual shares and to gravity-fitted shares, corresponding with $u_{ijt} \neq 1$. Random variation in trade flows is omitted again, so $\varepsilon_{ijt} = 1$. Although measurement error in trade flows is listed as an advantage of gravity-based calibration relative to actual shares calibration, it is by no means clear that gravity-based calibration is robust to measurement error. The gravity-based calibrations could suffer from two potential problems. First, the gravity estimation with pairwise fixed effects might pick up measurement error not varying over time, and second the gravity estimations with pairwise variables might suffer from correlation between the measurement error and the gravity covariates like distance, leading to biased estimates of the gravity coefficients and therefore also of the fitted trade values.

Since measurement error is unobserved it is hard to generate data to explore the influence of measurement error on the different welfare-estimators. To generate measurement error, the log difference between reported export and import flows in UN-COMTRADE is employed, exp_{ijt}^{CT} and imp_{ijt}^{CT}, for the countries and years in the WIOD-database, $\ln \upsilon_{ijt} = \ln \frac{exp_{ijt}^{CT}}{imp_{ijt}^{CT}}$. The same trade flows are reported by both importer authorities and exporter authorities and they often display large differences. Although the difference between reported exports and imports partially picks up the cif-fob margin, there are many other reasons for a discrepancy between

[48]As shown in the previous subsection, the latter approach can be slightly improved upon by including an interaction of the pairwise fixed effects with a time trend. Estimation of the gravity equation with this specification requires about 2.5 hours on a desktop with 16GB of ram. The required computation time together with the very small improvement in the estimation with actual data in the previous section was the reason not to explore this specification in the simulations.

Table 10: The average mean squared error of the predicted welfare effect with trade flows under measurement error based on decile-specific variance employing actual shares and fitted shares

Calibration to	Actual shares	Fitted shares		
Gravity specification		(1)	(2)	(3)
Welfare effects (perc. change)				
Population weighted				
EU	.00810	.276	.142	.0120
USA	.000346	.000115	.00283	.000255
Countries not in TTIP	3.92e-06	.0000126	3.87e-06	2.60e-06
All countries	.000601	.0199	.0104	.000879
GDP weighted				
EU	.0113	.285	.192	.0174
USA	.000346	.000115	.00283	.000255
Countries not in TTIP	7.02e-06	.0000459	.0000100	4.38e-06
All countries	.00277	.0681	.0465	.00421

Notes: All gravity specifications include exporter and importer fixed effects and the following regressors. Specification (1): log-distance, contiguity, common language, history of common colonizer and country-specific dummies for domestic trade flows; Specification (2): as in (1) but instead of country-specific dummies for domestic trade flows, one dummy for domestic flows and interaction terms of a domestic dummy with GDP and GDP per capita; Specification (3): pairwise (time-invariant) fixed effects.

the two, which are related to measurement error. The difference can represent for example misreporting of country of origin and destination (related to transit trade), ambiguity about the exchange rate date used to convert values into a common currency, and ambiguity over timing of exports and imports to report (Gaulier and Zignago (2010)). Lacking better data on measurement error, the difference between trade flows reported by exporter and importer is used as the most informative proxy of measurement error.

Measurement errors are split up into time-varying and time-invariant measurement errors by regressing the measurement errors $\ln \upsilon_{ijt}$ in the sample on a set of pairwise fixed effects, showing that 36% of the variation of the measurement error is time-invariant. Only for international data measurement error is generated, since COMTRADE does not contain domestic flows. Furthermore, two different ways are employed to partition the sample when calculating the variance of the measurement error to generate the simulation data. First, the measurement error is split up in deciles by size of trade flows as with random variation in the previous subsection. Second, the variance of measurement error is expressed as a function of the gravity regressors to allow for correlation between the measurement error and the gravity regressors. In this setup, observation specific variances are used to generate the simulation data. In particular, the following equation was estimated with the variance a function of the gravity regressors:

$$\begin{aligned} \ln \upsilon_{ijt} &= \alpha_1 \ln distance_{ij} + \alpha_2 contiguity_{ij} + \alpha_3 common\_colony_{ij} + \varphi\sigma^2_{ijt} + \iota_{ijt} \\ \sigma^2_{ijt} &= \exp\{\lambda_0 + \lambda_1 \ln distance_{ij} + \lambda_2 contiguity_{ij} + \lambda_3 common\_colony_{ij} \\ &\quad + \lambda_4 gdp_{it} + \lambda_5 gdp\_pc_{it} + \lambda_6 gdp_{jt} + \lambda_7 gdp\_pc_{jt}\} \end{aligned} \tag{50}$$

As in the previous section 100 baseline trade values are generated, according to $V_{ijt} = V^{data}_{ijt} u_{ijt}$ with V^{data}_{ijt} equal to the WIOD trade flows and the measurement error $\ln u_{ijt}$ drawn from a normal distribution with mean zero and variance calculated in the two ways described above. The welfare effects of the TTIP counterfactual experiment are then calculated based on the actual shares and based on fitted shares using three gravity specifications. So the same four estimators as in the previous section are compared.

Table 11: The average mean squared error of the predicted welfare effect with trade flows under measurement error based on variance correlated with gravity regressors employing actual shares and fitted shares

Calibration to	Actual shares	Fitted shares		
Gravity specification		(1)	(2)	(3)
Welfare effects (perc. change)				
Population weighted				
EU	.000575	.276	.143	.0114
USA	3.03e-06	.0000139	.00381	.000365
Countries not in TTIP	1.98e-07	.0000101	4.20e-06	1.24e-06
All countries	.0000416	.0198	.0104	.000835
GDP weighted				
EU	.000862	.285	.194	.0167
USA	3.03e-06	.0000137	.00381	.000365
Countries not in TTIP	3.28e-07	.0000375	.0000105	2.24e-06
All countries	.000207	.0681	.0472	.00406

Notes: All gravity specifications include exporter and importer fixed effects and the following regressors. Specification (1): log-distance, contiguity, common language, history of common colonizer and country-specific dummies for domestic trade flows; Specification (2): as in (1) but instead of country-specific dummies for domestic trade flows, one dummy for domestic flows and interaction terms of a domestic dummy with GDP and GDP per capita; Specification (3): pairwise (time-invariant) fixed effects.

Table 10 displays the MSE based on measurement errors generated with decile-specific variances:

$$MSE^{me}_{group} = \sum_{i \in group} \frac{weight_i \frac{1}{100} \sum\limits_{rep=1}^{100} \left(\widetilde{W}^{me}_{i,rep} - \widetilde{W}^{correct}_{i,rep} \right)^2}{weight_i}; weight = GDP, population \quad (51)$$

$\widetilde{W}^{me}_{i,rep}$ is the percentage change in welfare based on the data with measurement error. The table displays the same patterns as Table 9 on random variation in trade flows. The approaches based on actual shares and pairwise fixed effects based fitted shares generate similar results, whereas the other two approaches based on fitted shares perform considerably worse. This shows that time-invariant measurement error is not a reason to employ fitted shares based on gravity regressors only without pairwise fixed effects. The fixed effects also pick up time-invariant measurement error and could thus potentially perform worse than fitted shares based on gravity regressors only.

Table 11 displays the MSE based on trade flows including measurement errors generated with variances varying by the gravity regressors.[49] The results of this exercise are somewhat different from the previous one, since calibration based on actual shares clearly outperforms calibration based on fitted shares. The MSEs based on actual shares in this simulation exercise are smaller than in the previous exercise (first column), whereas the MSEs based on fitted shares are similar in the two exercises. The variance is observation specific and determined mostly by time-invariant regressors in this exercise. This makes the performance of fitted shares calibration worse as the pairwise fixed effects pick up time invariant measurement errors.

Based on these simulations it can be concluded that neither measurement error nor random variation in trade flows provide a good reason to prefer working with fitted shares instead of actual shares.

[49]The regression outcomes of the model in equation (50) to determine the variance of the measurement error are given in Appendix Appendix E

8 Concluding remarks

This report classified and compared different quantitative trade models, in terms of model exposition, solution method and calibration of baseline trade costs. As such the report could contribute to a dialogue between CGE and SG practitioners. Three main conclusions can be drawn based on the analysis. First, biases in predicted welfare effects of counterfactual trade cost experiments are driven by biases in baseline import and export shares. More specifically, overestimated trade shares vis-a-vis countries with which trade costs are reduced overestimate predicted welfare effects. Second, calibration to fitted shares can generate large biases because of misspecification of the employed gravity equation if no pairwise fixed effects are included. Third, calibration to actual shares and calibration to pairwise-fixed-effects-based fitted shares display similar performance in terms of robustness to three potential sources of bias, misspecification of the gravity equation, random variation in trade flows, and measurement error in observed trade flows. The findings of the report show that robustness to measurement error is not an advantage of fitted over actual shares. Therefore, it is not useful to compare SG with other approaches in the literature to address measurement error in trade data like Gehlhar (1996), who use mirror trade statistics to confront this problem.

Based on these findings, the analysis conveys three messages for trade modellers conducting counterfactual experiments. First, it is not a good idea to use fitted shares based on an estimation with only international trade flows and an estimation with a cross-section of data. Either actual shares should be used or panel data including pairwise fixed effects based on trade data with both international and domestic trade flows.[50] If there is a strong need to include as many countries as possible in the sample (for example to identify parameters of observable trade cost measures), it seems better to be flexible and deviate from structural estimation which would impose the use of the same dataset for estimation and simulation. Furthermore, it seems good practice to report the fit of the gravity estimation. Only reporting an R2 is not sufficient in this respect since a large correlation between fitted and actual trade flows could still go along with large differences between actual and fitted shares for example for domestic flows.[51]

Second, CGE-models can be merged with the SG-approach to baseline calibration of trade flows. Under PPML estimation of the gravity equation, fitted and actual output and expenditure are identical if all exporter and importer fixed effects are identified. Therefore, fitted trade values can easily be incorporated into a balanced CGE-database. Third, differences between baseline calibration to actual and pairwise-fixed-effects-based fitted shares are small and seem to be a matter of taste given the current knowledge. However, future work lies ahead which can tip the balance in favor of using one of the two approaches. Two questions should therefore be explored.

First, which method offers the best way to conduct sensitivity analysis, generating confidence intervals for the predicted welfare effects? Anderson and Yotov (2016) and Pfaffermayr (2017) calculate confidence intervals for the estimated effects of the presence of FTAs between countries, their main counterfactual experiment. They do so by bootstrapping the gravity estimates 500 times.[52] Since the behavioral parameters in their study, the sectoral trade elasticities, are fixed, this exercise comes down to varying the estimated baseline trade costs, based on pairwise fixed effects, and the FTA-coefficients determining the size of the trade cost shocks.

[50] As pointed out before, including pairwise fixed effects in a cross-section setting with domestic flows will give equivalent baseline shares as calibration to actual shares.

[51] An interesting example of the need to compare fitted with actual shares is Heid and Larch (2016). They estimate two gravity equations, first a gravity equation with pairwise fixed effects to identify the coefficient on a time-varying FTA dummy, and then a gravity equation without pairwise fixed effects to identify the effects of time-invariant bilateral variables imposing a the FTA-coefficient from their first equation. They do not report which estimate is employed to generate baseline trade costs, which does make a big difference for the outcomes of counterfactual experiments as shown by the simulations in this paper.

[52] The authors provide only few details on their exercise. For example, it is not clear whether the bootstrapping consists of panel bootstrapping or normal bootstrapping.

In the CGE-literature scholars have instead conducted sensitivity analyses based on variation in behavioral parameters (the trade elasticities in the exercises of Anderson and Yotov (2016) and Pfaffermayr (2017)) applying Gaussian quadrature (Arndt (1996)). Future research could attempt to combine the two approaches so as to be able to account for random variation in both behavioral parameters, counterfactual experiments, and baseline trade costs.

Second, oftentimes policy-makers are interested in the effects of a trade policy experiment in the future. This requires baseline trade data lying in the future and the two approaches to baseline calibration suggest different ways to generate future baselines. CGE-modellers typically make projections for macroeconomic variables like productivity, population and human capital growth, and changes in the trade balance to generate baselines in the future based on a dynamic model with endogenous capital and labor (see for example Dixon and Rimmer (1998), Bekkers et al. (2017)). An alternative in the spirit of the SG-models would be to generate predictions for future trade flows based on a panel gravity estimation delivering out-of-sample predictions. Future work could compare the merits of the two approaches based on their out-of-sample performance.

References

Aichele, R., G. Felbermayr, and I. Heiland (2014). "Going Deep: The Trade and Welfare Effects of TTIP," CESifo Working Paper 5150, Center for Economic Studies and Ifo Institute.

Alvarez, F. and R. E. J. Lucas (2007). "General Equilibrium Analysis of The Eaton Kortum Model of International Trade," *Journal of Monetary Economics*, 54: 1726–1768.

Anderson, J. E., M. Larch, and Y. V. Yotov (2015). "Estimating General Equilibrium Trade Policy Effects: GE PPML," Working Paper 5592, CESIFO.

Anderson, J. E. and E. van Wincoop (2003). "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93(1): 170–192.

Anderson, J. E. and Y. V. Yotov (2016). "Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990-2002," *Journal of International Economics*, 99: 279–298.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1): 94–130.

Arndt, C. (1996). "An Introduction to Systematic Sensitivity Analysis via Gaussian Quadrature," Technical Paper 2, GTAP.

Baldwin, R. E. (2014). "Multilateralising 21st Century Regionalism," Global Forum on Trade, OECD.

Baldwin, R. E. and D. Jaimovich (2012). "Are Free Trade Agreements Contagious?" *Journal of International Economics*, 88(1): 1–16.

Bchir, M. H., Y. Decreux, J.-L. Guï¿½rin, and S. Jean (2002). "MIRAGE, a Computable General Equilibrium Model for Trade Policy Analysis," Tech. Rep. 2002-17, CEPII.

Bekkers, E., J. Francois, and H. Rojas-Romagosa (2017). "Melting Ice Caps and the Economic Impact of Opening the Northern Sea Route," *Economic Journal*, Forthcoming.

Bekkers, E. and H. Rojas-Romagosa (2017). "The Welfare Effects of FTAs in Quantitative Trade Models: A Comparison of Studies over TTIP," Working paper, University of Bern.

Berden, K. and J. Francois (2015). "Quantifying Non-Tariff Measures for TTIP," in *Rule-Makers or Rule-Takers? Exploring the Transatlantic Trade and Investment Partnership*, ed. by D. S. Hamilton and J. Pelkmans, Centre for European Policy Studies (CEPS) and Center for Transatlantic Relations (CTR).

Borchert, I., B. Gootiiz, and A. Mattoo (2014). "Policy Barriers to International Trade in Services: Evidence from a New Database," *World Bank Economic Review*, 28(1): 162–188.

Britz, W. and D. van der Mensbrugghe (2017). "A Flexible, Modular, and Extendable Framework for CGE Analysis in GAMS," Tech. rep., Institute for Food and Resouce Economics, University of Bonn.

Caliendo, L. and F. Parro (2015). "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, 82(1): 1–44.

Carrere, C., A. Grujovicz, and F. Robert-Nicoud (2015). "Trade and Frictional Unemployment in the Global Economy," Mimeo, Universite de Geneve.

Ciuriak, D., A. Dadkhah, and J. Xiao (2016). "Taking the Measure of the TPP as Negotiated," Tech. rep., Ciuriak Consulting Inc.

Costinot, A. and A. Rodríguez-Clare (2013). "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, North Holland, vol. 4.

Dekle, R., J. Eaton, and S. Kortum (2008). "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 55: 511–540.

Dixon, P. B. and M. T. Rimmer (1998). "Forecasting and Policy Analysis with a Dynamic CGE Model of Australia." Working Papers 90, Centre of Policy Studies/Impact Centrer.

Dür, A., L. Baccini, and M. Elsig (2014). "The Design of International Trade Agreements: Introducing a New Dataset," *Review of International Organizations*, 9(3): 353–375.

Eaton, J. and S. Kortum (2002). "Technology, geography, and trade," *Econometrica*, 70(5): 1741–1779.

Ecorys (2009). "Non-Tariff Measures in EU-US Trade and Investment: An Economic Analysis," Report prepared by Koen G. Berden, Joseph F. Francois, Saara Tamminen, Martin Thelle, and Paul Wymenga for the European Commission, Reference OJ 2007/S180-219493, ECORYS Nederland BV.

Egger, P., J. Francois, M. Manchin, and D. Nelson (2015). "Non-tariff Barriers, Integration, and the Transatlantic Economy," *Economic Policy*, forthcoming.

Egger, P. and M. Larch (2011a). "An assessment of the Europe agreements? effects on bilateral trade, GDP, and welfare," *European Economic Review*, 55: 263–279.

Egger, P. and M. Larch (2011b). "An Assessment of the Europe Agreements' Effects on Bilateral Trade, GDP, and Welfare," *European Economic Review*, 55(2): 263–279.

Egger, P., M. Larch, and K. E. Staub (2012). "Trade Preferences and Bilateral Trae in Goods and Services: A Structural Approach," *CEPR DIscussion Paper*, 9051.

Egger, P., M. Larch, K. E. Staub, and R. Winkelmann (2011). "The Trade Effects of Endogenous Preferential Trade Agreements," *American Economic Journal: Economic Policy*, 3(3): 113–143.

Egger, P. and S. Nigai (2015). “Structural gravity with dummies only: Constrained ANOVA-type estimation of gravity models,” *Journal of International Economics*, 97: 86–99.

Egger, P. and Y. Wolfmayr (2014). “What Economists Should Know about International Goods Trade Data,” Working Paper 475/2014, WIFO.

European Commission (2013). “Transatlantic Trade and Investment Partnership: The Economic Analysis Explained,” Technical report, European Commission.

Fally, T. (2015). “Structural Gravity and Fixed Effects,” *Journal of International Economics*, 97: 76–85.

Feenstra, R. C. (1994). “New Product Varieties and the Measurement of International Prices,” *American Economic Review*, 84(1): 157–177.

Felbermayr, G., B. Heid, M. Larch, and E. Yalcin (2015). “Macroeconomic Potentials of Transatlantic Free Trade: A High Resolution Perspective for Europe and the World,” *Economic Policy*, July 2015: 491–537.

Felbermayr, G., B. Heid, and S. Lehwald (2013a). “Transatlantic Trade and Investment Partnership (TTIP): Who benefits from a free trade deal?” Report, Global Economic Dynamics (GED), Bertelsmann Foundation.

Felbermayr, G., M. Larch, L. Flach, and E. Y. und Sebastian Benz (2013b). “Dimensionen und Auswirkungen eines Freihandelsabkommens zwischen der EU und den USA,” Tech. rep., CES-IFO. Studie im Auftrag des Bundesministeriums fï¿½r Wirtschaft und Technologie. Endbericht.

Fontagne, L., J. Gourdon, and S. Jean (2013). “Transatlantic Trade: Whither Partnership, which Economic Consequences?” CEPII Policy Brief No. 1, September, Centre d’Etudes Prospectives et d’Informations Internationales (CEPII).

Fontagne, L., A. Guillin, and C. Mitaritonna (2011). “Estimations of Tariff Equivalents for the Services Sectors,” CEPII Working Paper 2011-24.

Francois, J., M. Manchin, H. Norberg, O. Pindyuk, and P. Tomberger (2013). “Reducing Transatlantic Barriers to Trade and Investment: An Economic Assessment,” Final Project Report, prepared under implementing Framework Contract TRADE10/A2/A16, CEPR, London.

Gaulier, G. and S. Zignago (2010). “BACI: International trade database at the product-level. The 1994-2007 version,” Working Paper 2010-23, CEPII.

Gehlhar, M. (1996). “Reconciling Bilateral Trade Data for Use in GTAP,” Technical papers, GTAP.

Head, K. and T. Mayer (2014). “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Elsevier, vol. 4, 131–195.

Heid, B. and M. Larch (2016). “Gravity with Unemployment,” *Journal of International Economics*, forthcoming.

Hertel, T. (2013a). “Global Applied General Equilibrium Analysis Using the Global Trade Analysis Project Framework,” in *Handbook of Computable General Equilibrium Modeling*, ed. by P. Dixon and D. Jorgenson, Amsterdam: Elsevier: North Holland.

Hertel, T. (2013b). "Global Applied General Equilibrium Analysis Using the Global Trade Analysis Project Framework," in *Handbook of Computable General Equilibrium Modeling*, ed. by P. Dixon and D. Jorgenson, Amsterdam: Elsevier.

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

Horn, H., P. C. Mavroidis, and A. Sapir (2009). *Beyond the WTO? An Anatomy of EU and US Preferential Trade Agreements*, vol. VII of *Bruegel Blueprint Series*.

Horridge, M., A. Meeraus, K. Pearson, and T. F. Rutherford (2013). *Solution Software for Computable General Equilibrium Modeling*, Elsevier: North Holland, chap. 20: 1331–1381.

Jafari, Y. and D. G. Tarr (2015). "Estimates of Ad Valorem Equivalents of Barriers Against Foreign Suppliers of Services in Eleven Services Sectors and 103 Countries," *World Economy*, doi: 10.1111/twec.12329.

Jones, L., Z. Wang, L. Xin, and C. Degain (2014). "The Similarities and Differences among Three Major Inter-Country Input-Output Databases and their Implications for Trade in Value-Added Estimates," Working Paper 2014-12B, USITC.

Kee, H. L., A. Nicita, and M. Olarreaga (2009). "Estimating Trade Restrictiveness Indices," *Economic Journal*, 119(534): 172–199.

Lanz, B. and T. F. Rutherford (2016). "GTAPINGAMS: Multiregional and Small Open Economy Models," *Journal of Global Economic Analysis*, 1(2): 1–77.

Levchenko, A. and J. Zhang (2016). "The Evolution of Comparative Advantage:Measurement and Welfare Implications," *Journal of Monetary Economics*, 78: 96–111.

Mustilli, F. (2015). "Estimating the Economic Gains of TTIP," *Intereconomics*, 50(6): 321–327.

Pelkmans, J., A. Lejour, L. Schrefler, F. Mustilli, and J. Timini (2014). "EU-US Transatlantic Trade and Investment Partnership: Detailed Appraisal of the European Commission's Impact Assessment," Ex-Ante Impact Assessment Unit Study, European Parliamentary Research Service, European Parliament.

Petri, P. A., M. G. Plummer, and F. Zhai (2011). "The Trans-Pacific Partnership and Asia-Pacific Integration: A Quantitative Assessment," Working Paper 119, East-West Center.

Pfaffermayr, M. (2017). "Constrained Poisson Pseudo Maximum Likelihood Estimation of Structural Gravity Models," Tech. rep., University of Innsbruck.

Rodrik, D. (2015). "The War of Trade Models," Dani Rodrik's weblog (rodrik.typepad.com), May 4th.

The Economist (2013). "Transatlantic Trade Talks: Opening Shots," July 6th, Print Edition.

The Guardian (2015). "What is TTIP The Controversial Trade Deal Proposal Explained," July 3rd, Print Edition.

Wolf, M. (2015). "The Embattled Future of Global Trade Policy," May 12th, The Financial Times.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

WTO (2015). "World Trade Report. Speeding up Trade: Benefits and Challenges of Implementing The WTO Trade Facilitation Agreement," Tech. rep., World Trade Organization.

Yotov, Y. V., R. Piermartini, J.-A. Monteiro, and M. Larch (2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*, WTO Publications.

Appendix A Theoretical model

A representative consumer with Armington preferences in country j has utility U_j over varieties from different countries i:[53]

$$U_j = \left(\sum_{i=1}^{J} q_{ij}^{\frac{\sigma-1}{\sigma}} d\omega_{ij} \right)^{\frac{\sigma}{\sigma-1}} \tag{A.1}$$

q_{ij} is the demand for goods from i and is equal to:

$$q_{ij} = p_{ij}^{-\sigma} P_j^{\sigma-1} E_j \tag{A.2}$$

p_{ij} is the price of goods shipped from i to j, E_j is expenditure in country j, and P_j is the price index defined as:

$$P_j = \left(\sum_{i=1}^{J} p_{ij}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{A.3}$$

There is a fixed endowment of goods in country i denoted by L_i with price w_i. Shipping goods from country i and j comes with iceberg trade costs t_{ij}, implying $p_{ij} = t_{ij} w_i$. The value of trade from i to j, V_{ij}, is equal to the quantity of trade q_{ij} times the price p_{ij}:

$$V_{ij} = (\tau_{ij} w_i)^{1-\sigma} P_j^{\sigma-1} E_j \tag{A.4}$$

Income in country i, $Y_i = w_i L_i$, is equal to the value of sales to all destination countries j:

$$w_i L_i = \sum_{j=1}^{J} (\tau_{ij} w_i)^{1-\sigma} P_j^{\sigma-1} E_j \tag{A.5}$$

Expenditures E_j have been written as a function of income by defining the trade deficit ratio D_j as $D_j = \frac{E_j - w_j L_j}{w_j L_j}$.

Appendix B Equilibrium equations and baseline calibration in the different approaches

In this appendix the equilibrium equations, baseline calibrations of trade costs, and employed starting values of the endogenous variables of the four methods are listed. First, calibration of trade costs from and to the rest-of-the-world is discussed.

Appendix B.1 Calibration trade costs Rest-of-the-World (ROW)

In a balanced dataset of trade flows between countries there is also a residual region, ROW. The question is how trade costs should be calibrated for this region. With calibration to actual shares, trade costs are simply set based on the actual import and export shares. With calibration to fitted shares the problem is that ROW is not part of the gravity estimation, so

[53] Since we assume identical firms we can write the mass of firms n_{ij} immediately in the utility function instead of using an integral over the set of varieties from country i.

no fitted values can be calculated. Under estimation with PPML including domestic flows, however, fitted and actual total imports and exports of each country are identical. Therefore, trade costs from and to ROW can be calibrated to the actual import shares. Under estimation without domestic trade flows, however, total fitted imports are not equal to actual imports, since domestic sales (imports) are not part of the estimation. Therefore, in the simulations based on gravity estimation without domestic flows, fitted trade costs were calculated as an average of trade costs with all other countries.

Appendix B.2 Structural gravity

Π_i and P_j are solved from equations (21)-(22) in the baseline employing the actual values for Y_i and the fitted values for iceberg trade costs, t_{ij}^{sg}, in equation (30). The counterfactual is solved from the same equations with baseline trade costs, t_{ij}^{sg}, replaced by counterfactual trade costs, $t_{c,ij}^{sg}$, using equation (32) and adding equation (31) to solve for counterfactual GDP. The starting values for Π_i and P_j in the baseline can be set as follows based on the fixed effects, equations (33)-(34):

$$\Pi_i = \left(X_i / \frac{Y_i}{Y_W} \right)^{\frac{1}{\sigma-1}} \tag{B.1}$$

$$P_j = (M_j/E_j)^{\frac{1}{\sigma-1}} \tag{B.2}$$

If all fixed effects are identified in the estimation stage and the gravity equation is estimated with PPML, these expressions are exact solutions of the baseline model, as pointed out by Fally (2015).

As an alternative, baseline trade costs can be set based on the fitted import shares with corresponding solutions for P_j and Π_i:

$$t_{ij} = \left(impsh_{ij}^{sg} \right)^{\frac{1}{1-\sigma}} = \left(\frac{T_{ij} X_i M_j}{E_j} \right)^{\frac{1}{1-\sigma}} \tag{B.3}$$

$$P_j = 1 \tag{B.4}$$

$$\Pi_i = \left(\frac{Y_i}{Y_W} \right)^{\frac{1}{1-\sigma}} \tag{B.5}$$

A third way to solve the model is to employ the equilibrium equations (19)-(20), solving for w_i and P_j, setting t_{ij} as in (B.3) in the baseline and using as starting values $w_i = P_j = 1$.

Although the three calibrations seem different, they all give exactly the same results for counterfactual exercises since they correspond with the same baseline import and export shares. With the first calibration import and export shares are given by:

$$impsh_{ij}^{sg1} = \frac{V_{ij}}{E_j} = t_{ij}^{1-\sigma} \Pi_i^{\sigma-1} P_j^{\sigma-1} \frac{Y_i}{Y_W} = \frac{T_{ij} X_i M_j}{E_j} = \frac{V_{ij}^{grav}}{E_j} \tag{B.6}$$

$$expsh_{ij}^{sg1} = \frac{V_{ij}}{Y_i} = t_{ij}^{1-\sigma} \Pi_i^{\sigma-1} P_j^{\sigma-1} \frac{E_j}{Y_i} \frac{Y_i}{Y_W} = \frac{T_{ij} X_i M_j}{Y_i} = \frac{V_{ij}^{grav}}{Y_i} \tag{B.7}$$

The second calibration delivers the same baseline shares:

$$impsh_{ij}^{sg2} = \frac{V_{ij}}{E_j} = t_{ij}^{1-\sigma} \Pi_i^{\sigma-1} P_j^{\sigma-1} \frac{Y_i}{Y_W} = t_{ij}^{1-\sigma} = \frac{T_{ij} X_i M_j}{E_j} = \frac{V_{ij}^{grav}}{E_j} \tag{B.8}$$

$$expsh_{ij}^{sg2} = \frac{V_{ij}}{Y_i} = t_{ij}^{1-\sigma} \Pi_i^{\sigma-1} P_j^{\sigma-1} \frac{E_j}{Y_i} \frac{Y_i}{Y_W} = t_{ij}^{1-\sigma} \frac{E_j}{Y_i} = \frac{T_{ij} X_i M_j}{Y_i} = \frac{V_{ij}^{grav}}{Y_i} \tag{B.9}$$

And the third calibration generates also the same baseline shares:

$$impsh_{ij}^{sg3} = \frac{V_{ij}}{E_j} = (t_{ij}w_i)^{1-\sigma} P_j^{\sigma-1} = t_{ij}^{1-\sigma} = \frac{T_{ij}X_iM_j}{E_j} = \frac{V_{ij}^{grav}}{E_j} \tag{B.10}$$

$$expsh_{ij}^{sg3} = \frac{V_{ij}}{Y_i} = (t_{ij}w_i)^{1-\sigma} P_j^{\sigma-1}\frac{E_j}{Y_i} = t_{ij}^{1-\sigma}\frac{E_j}{Y_i} = \frac{T_{ij}X_iM_j}{Y_i} = \frac{V_{ij}^{grav}}{Y_i} \tag{B.11}$$

Appendix B.3 CGE-in-levels

Equations (19)-(20) are solved for the endogenous variables w_i and P_j both with baseline and counterfactual trade costs and baseline trade costs are given by equation (38). In the baseline w_i and P_j are both set at 1, which are solutions for equations (19)-(20) and obviously give baseline import and export shares equal to the actual shares.

Appendix B.4 CGE-in-relative-changes

The equilibrium equations (39)-(40) can be solved for $\widetilde{P_j}$ and $\widetilde{w_i}$ as a function of $\widetilde{t_{ij}}$ with the import and export shares taken from the data. But in GEMPACK, with the model coded in terms of quantities, the following equations are solved for $\widetilde{P_j}$, $\widetilde{w_i}$ and $\widetilde{q_{ij}}$, the quantity of trade from i to j:

$$\widetilde{P_j} = \sum_{i=1}^{J} \frac{(t_{ij}w_i)^{1-\sigma}}{\sum\limits_{k=1}^{J}(t_{kj}w_k)^{1-\sigma}}\left(\widetilde{t_{ij}} + \widetilde{w_i}\right) = \sum_{i=1}^{J}\frac{trade_{ij}}{\sum\limits_{k=1}^{J} trade_{kj}}\left(\widetilde{t_{ij}} + \widetilde{w_i}\right) \tag{B.12}$$

$$\widetilde{q_{ij}} = \widetilde{t_{ij}} - \sigma\left(\widetilde{t_{ij}} + \widetilde{w_i} - \widetilde{P_j}\right) - \widetilde{P_j} + \widetilde{w_j} + \widetilde{L_j} \tag{B.13}$$

$$\widetilde{L_i} = \sum_{j=1}^{J}\frac{trade_{ij}}{\sum\limits_{k=1}^{J} trade_{ik}}\widetilde{q_{ij}} \tag{B.14}$$

Since the number of endowments is fixed, $\widetilde{L_i}$ is set at zero or technically set as exogenous variable in the command file. The value of trade $trade_{ij}$ is updated in each step according to:

$$trade_{ij} = w_i q_{ij} \tag{B.15}$$

The online appendix outlines the GEMPACK code of the equilibrium equations.

Appendix B.5 Exact hat algebra

Equations (41)-(42) are solved for $\widehat{P_j}$ and $\widehat{w_i}$ as a function of $\widehat{t_{ij}}$. The initial values for $\widehat{P_j}$ and $\widehat{w_i}$ are set at respectively -0.1 and 0.1. If the model becomes more complicated, it might become more urgent to pick starting values closer to the expected solution and work with country-specific starting values.

As alternative one could solve for $\widehat{P_j}$ and $\widehat{Y_i}$ as often done in the EHA-approach. The two equilibrium equations become in this case:

$$P_j = \left(\sum_{i=1}^{J}\left(t_{ij}\frac{Y_i}{L_i}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \tag{B.16}$$

$$Y_i = \sum_{j=1}^{J}\left(t_{ij}\frac{Y_i}{L_i}\right)^{1-\sigma} P_j^{\sigma-1}(1+D_j)\,Y_j \tag{B.17}$$

Exact hat-differentiating equations (B.16)-(B.17) gives then the same equations as before, but with w_i replaced by Y_i.

Supplementary appendices to be distributed as online appendix

Appendix C Derivation equations

This section provides more detailed derivations of some of the equations in the main text.

Equations (39)-(40)

Hat-differentiating equations (19)-(20) gives:

$$\widehat{P_j} = \sum_{i=1}^{J} \frac{(t_{ij}w_i)^{1-\sigma}}{\sum\limits_{k=1}^{J} (t_{kj}w_k)^{1-\sigma}} \left(\widehat{t_{ij}} + \widehat{w_i}\right) \tag{C.1}$$

$$\widehat{w_i} = \sum_{j=1}^{J} \frac{(t_{ij}w_i)^{1-\sigma} P_j^{\sigma-1} (1+D_j) w_j L_j}{\sum\limits_{k=1}^{J} (t_{ik}w_i)^{1-\sigma} P_k^{\sigma-1} (1+D_k) w_k L_k} \left((1-\sigma)\left(\widehat{t_{ij}} + \widehat{w_i} - \widehat{P_j}\right) + \widehat{w_j}\right) \tag{C.2}$$

It is easy to show that the coefficients in the summations of equations (C.1)-(C.2) are respectively equal to the import shares $impsh_{ij}$ and export shares $expsh_{ij}$ and thus lead to equations (39)-(40).

Equations (21)-(23)

To start equation (20) can be rearranged as follows, imposing $Y_i = w_i L_i$:

$$Y_i = \sum_{j=1}^{J} (t_{ij}w_i)^{1-\sigma} P_j^{\sigma-1} (1+D_j) Y_j$$

$$w_i^{1-\sigma} = \frac{Y_i}{\sum\limits_{j=1}^{J} t_{ij}^{1-\sigma} P_j^{\sigma-1} (1+D_j) Y_j}$$

$$w_i = \left(\sum_{j=1}^{J} t_{ij}^{1-\sigma} P_j^{\sigma-1} \frac{(1+D_j) Y_j}{Y_W}\right)^{\frac{1}{\sigma-1}} \left(\frac{Y_i}{Y_W}\right)^{\frac{1}{1-\sigma}}$$

Hence, Π_i can be defined as follows, equation (22) in the main text:

$$\Pi_i = \left(\sum_{j=1}^{J} t_{ij}^{1-\sigma} P_j^{\sigma-1} \frac{(1+D_j) Y_j}{Y_W}\right)^{\frac{1}{1-\sigma}} \tag{C.3}$$

Equation (22) implies the following expression for w_i:

$$w_i = \frac{\left(\frac{Y_i}{Y_W}\right)^{\frac{1}{1-\sigma}}}{\Pi_i} \tag{C.4}$$

Substituting equation (C.4) into equations (18)-(19) leads to equations (21) and (23):

$$P_j = \left(\sum_{j=1}^{J} \left(t_{ij} \frac{\left(\frac{Y_i}{Y_W}\right)^{\frac{1}{1-\sigma}}}{\Pi_i}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}} = \left(\sum_{j=1}^{J} t_{ij}^{1-\sigma} \Pi_i^{\sigma-1} \frac{Y_i}{Y_W}\right)^{\frac{1}{1-\sigma}} \tag{C.5}$$

$$V_{ij} = (t_{ij}w_i)^{1-\sigma} P_j^{\sigma-1} E_j = (t_{ij})^{1-\sigma} \Pi_i^{\sigma-1} P_j^{\sigma-1} \frac{Y_i E_j}{Y_W} \tag{C.6}$$

Equation (31)

Next the expression for the ratio of counterfactual and baseline output can be derived, starting from equation () and imposing fixed endowments, $L_{c,i} = L_i$:

$$\frac{Y_{c,i}}{Y_i} = \frac{w_{c,i}L_i}{w_i L_i} = \frac{\frac{\left(\frac{Y_{c,i}}{Y_W}\right)^{\frac{1}{1-\sigma}}}{\Pi_{c,i}}}{\frac{\left(\frac{Y_i}{Y_W}\right)^{\frac{1}{1-\sigma}}}{\Pi_i}} = \frac{\Pi_i}{\Pi_{c,i}} \frac{\left(\frac{Y_{c,i}}{Y_W}\right)^{\frac{1}{1-\sigma}}}{\left(\frac{Y_i}{Y_W}\right)^{\frac{1}{1-\sigma}}}$$

$$\frac{Y_{c,i}}{Y_i} = \left(\frac{\Pi_i}{\Pi_{c,i}}\right)^{\frac{\sigma-1}{\sigma}}$$

It is assumed that $Y_W = Y_{c,W}$, implying that the solution of a counterfactual experiment based on equation (31) will generate an approximate solution.

Equation (35)

Given fixed endowments $Y_{c,i}$ can be written as a function of $w_{c,i}$:

$$Y_{c,i} = \frac{w_{c,i}}{w_i} Y_i \tag{C.7}$$

Comparing the two expressions for P_j in (19) and (21) shows that w_i can be written as:

$$w_i^{1-\sigma} = \Pi_i^{\sigma-1} \frac{Y_i}{Y_W} \tag{C.8}$$

Using the expression for the estimated fixed effect, equation (33), w_i can be written as:

$$w_i^{1-\sigma} = \Pi_i^{\sigma-1} \frac{Y_i}{Y_W} = X_i \tag{C.9}$$

Substituting equation (C.9) into equation (C.7) and rearranging then leads to equation (35) in the main text:

$$Y_{c,i} = \frac{w_{c,i}}{w_i} Y_i = \left(\frac{X_{c,i}}{X_i}\right)^{\frac{1}{1-\sigma}} Y_i \tag{C.10}$$

Baseline shares equal to actual shares with CGE in levels approach

Calculating the import and export shares from equation (18) and substituting the CGE-in-levels calibration of baseline trade costs in equation (38) leads to:

$$impsh_{ij} = (t_{ij} w_i)^{1-\sigma} P_j^{\sigma-1} = \left(\left(\frac{V_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}}\right)^{1-\sigma} = \frac{V_{ij}}{E_j}$$

$$expsh_{ij} = \frac{(t_{ij} w_i)^{1-\sigma} P_j^{\sigma-1} E_j}{Y_i} = \frac{\left(\left(\frac{V_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}}\right)^{1-\sigma} E_j}{Y_i} = \frac{V_{ij}}{Y_i}$$

Showing that equation (38) *corresponds with* $w_i = P_j = 1$ *in baseline*

It can be easily shown that $w_i = P_j = 1$ is a solution of the equilibrium equations (19)-(20), given equation (38):

$$P_j = \left(\sum_{i=1}^{J} (t_{ij} w_i)^{1-\sigma}\right)^{\frac{1}{1-\sigma}} = \left(\sum_{i=1}^{J} \frac{V_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}} = 1$$

$$w_i L_i = \sum_{j=1}^{J} (t_{ij} w_i)^{1-\sigma} P_j^{\sigma-1} (1 + D_j) w_j L_j = \sum_{j=1}^{J} \frac{V_{ij}}{E_j} E_j$$

Appendix D Code CGE-model in relative changes

In GEMPACK notation (using as much as possible the same symbols as in the GTAP-GEMPACK-code), equations (B.12)-(B.15) correspond with:

$$pim(s) = sum(k, REG, MSHRS(k,s) * [pm(k) + itc(k,s)]) \tag{B.1}$$

$$qxs\,(r,s) = itc\,(r,s) - ESBD * (itc\,(r,s) + pm\,(r) - pim\,(s)) - pim\,(s) + pm\,(s) + qo\,(s) \tag{B.2}$$

$$qo\,(r) = sum\,(s, SHRXMD(r,s) * qxs\,(r,s)) \tag{B.3}$$

With:

$$trade\,(r,s) = pm\,(r) * qxs\,(r,s) \tag{B.4}$$

$$MSHRS\,(r,s) = \frac{trade\,(r,s)}{sum\,(k, trade\,(k,s))} \tag{B.5}$$

$$SHRXMD\,(r,s) = \frac{trade\,(r,s)}{sum\,(k, trade\,(r,k))} \tag{B.6}$$

Instead of using the symbol *ams*, the technological change in trade in GTAP-GEMPACK, *itc*, iceberg trade costs, is used observing that $ams = -itc$. As data only $trade_{ij}$ is needed, the value of trade.

Appendix E Additional counterfactual experiments and auxiliary regression results

Table 12: Effect of import and export share with FTA-partner on welfare effects for the USA of the introduction of an FTA with the EU (TTIP)

	(1) diff_welfare	(2) diff_welfare	(3) diff_welfare
diff_impsh	9.40***	4.52***	4.45***
	(0.33)	(0.074)	(0.072)
diff_expsh		4.95***	4.76***
		(0.063)	(0.078)
diff_domsh			-0.0026***
			(0.00068)
Observations	100	100	100
R^2	0.89	1.00	1.00
Adjusted R^2	0.89	1.00	1.00

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Effect of import and export share with FTA-partner on welfare effects of the USA and Mexico for a Mexico-USA FTA

	Mexico			USA		
	(1)	(2)	(3)	(4)	(5)	(6)
	diff_welfare	diff_welfare	diff_welfare	diff_welfare	diff_welfare	diff_welfare
diff_impsh	0.17***	0.055***	0.0018***	0.071***	0.032***	0.032***
	(0.0078)	(0.0095)	(0.00039)	(0.0023)	(0.0015)	(0.0012)
diff_expsh		0.12***	-0.0024***		0.049***	0.044***
		(0.0087)	(0.00054)		(0.0016)	(0.0015)
diff_domsh			-0.15***			-0.0011***
			(0.00056)			(0.00015)
Observations	100	100	100	100	100	100
R^2	0.82	0.94	1.00	0.91	0.99	0.99
Adjusted R^2	0.82	0.94	1.00	0.91	0.99	0.99

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Effect of import and export share with FTA-partner on welfare effects of FTA for Mexico and the rest of the world of a unilateral liberalization in Mexico

	Mexico	Rest of the world				
	(1)	(2)	(3)	(4)	(5)	(6)
	diff_welfare	diff_welfare	diff_welfare	diff_welfare	diff_welfare	diff_welfare
diff_impsh	0.15***	0.046***	0.046***	0.046***	0.048***	0.042***
	(0.00016)	(0.00014)	(0.00014)	(0.00013)	(0.0012)	(0.0022)
diff_expsh		0.047***	0.046***	0.049***	0.051***	0.023***
		(0.00027)	(0.00027)	(0.00026)	(0.0020)	(0.0025)
diff_domsh		-0.000015*	-0.000015*	-0.0000062	-0.000061	0.00019
		(0.0000068)	(0.0000069)	(0.0000064)	(0.000053)	(0.00012)
Observations	100	12000	12000	12000	120	100
R^2	1.00	0.93	0.93	0.94	0.94	0.85
Adjusted R^2	1.00	0.93	0.93	0.94	0.94	0.85

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Effect of total import share on welfare effects of multilateral liberalization

	(1)	(2)	(3)	(4)	(5)
	diff_welfare	diff_welfare	diff_welfare	diff_welfare	diff_welfare
diff_impsh	0.32***	0.32***	0.33***	0.24***	0.27***
	(0.0043)	(0.0043)	(0.0018)	(0.016)	(0.0056)
Observations	12100	12100	12100	121	99
R^2	0.31	0.31	0.88	0.65	0.96
Adjusted R^2	0.31	0.31	0.88	0.65	0.96

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

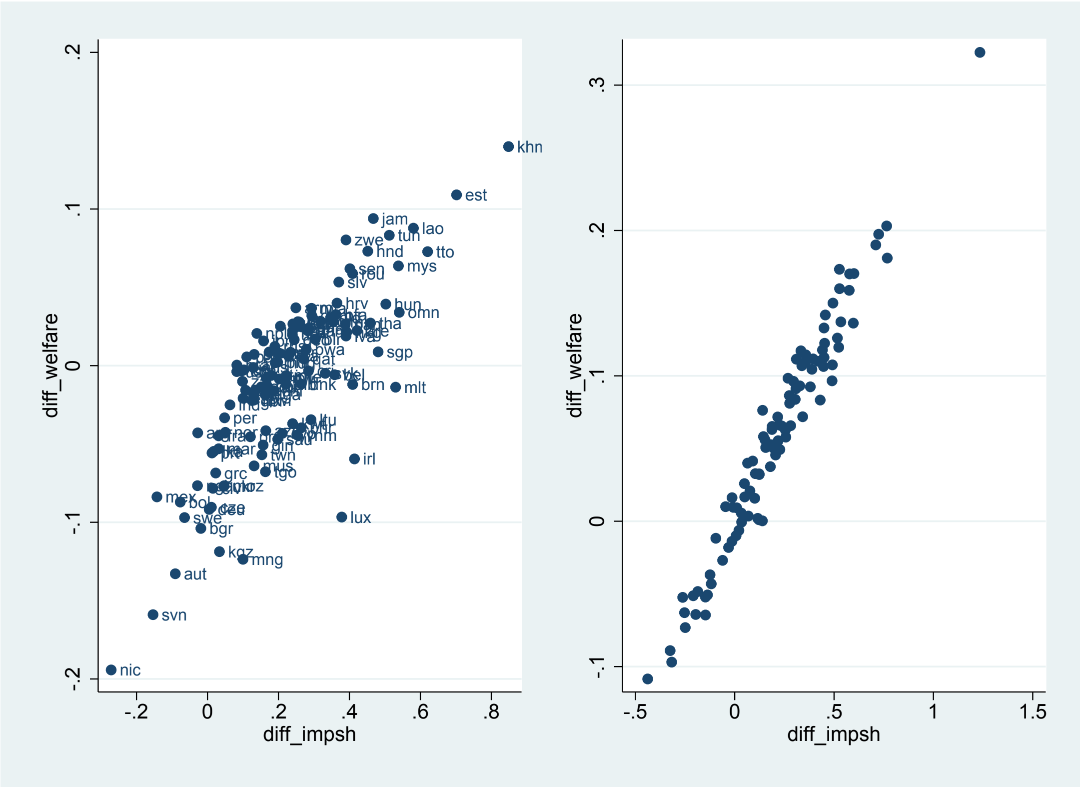

Figure 2: The impact of deviations in average import shares from their mean on deviations of welfare effects from their mean after a reduction in trade costs between all countries (multilateral liberalization)

Table 16: Standard deviations used to generate random variation and measurement error in the trade data

Decile trade flows	St.dev. random variation $ln(\epsilon_{ijt})$		St.dev. measurement error $ln(\upsilon_{ijt})$
	International flows	Domestic flows	All trade flows
1	0.838	0.116	0.891
2	0.562	0.058	0.605
3	0.471	0.054	0.52
4	0.394	0.031	0.454
5	0.353	0.019	0.381
6	0.33	0.022	0.369
7	0.282	0.017	0.31
8	0.268	0.009	0.297
9	0.244	0.01	0.274
10	0.18	0.004	0.274
N	27,090	645	23,220

Table 17: Explaining variance of measurement error as a function of gravity regressors and importer and exporter GDP and GDP per capita

	(1)	
	lomega	
Log(omega)		
Log(Distance)	-0.087***	(0.0031)
Contiguity	0.010	(0.011)
Common language	0.044***	(0.0092)
Common colony	0.11***	(0.031)
Constant	0.60***	(0.025)
HET		
Log(Distance)	0.38***	(0.0047)
Contiguity	-0.50***	(0.022)
Common language	-0.39***	(0.020)
Common colony	-0.63**	(0.21)
GDP exporter	-0.000000083***	(1.7e-09)
GDP PC exporter	-8.09***	(0.13)
GDP importer	-0.00000017***	(1.4e-09)
GDP PC importer	-0.65***	(0.14)
Constant	-3.78***	(0.036)
Observations	23220	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$