Staff Working Paper ERSD-2019-02 **8 March 2019**

World Trade Organization

Economic Research and Statistics Division

The welfare effects of trade policy experiments in quantitative trade models: the role of solution methods and baseline calibration*

Eddy Bekkers

Manuscript date: 20 February 2019

The welfare effects of trade policy experiments in quantitative trade models: the role of solution methods and baseline calibration[*]

Eddy Bekkers
World Trade Organization

ABSTRACT: This paper compares the solution methods and baseline calibration of three different quantitative trade models (QTMs): 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 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 predicted 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 fitted shares based on gravity estimation including pairwise fixed effects display similar performance in terms of robustness to the different sources of bias.

Keywords: Quantitative trade models, baseline calibration, free trade agreements, gravity estimation

JEL codes: F13, F14, F15

Disclaimer: The opinions expressed in this article should be attributed to its author. They are not meant to represent the positions or opinions of the WTO and its members and are without prejudice to Members' rights and obligations under the WTO. Any errors are attributable to the authors. Declarations of interest: none.

printdate: February 20, 2019

---

[*] I want to thank participants in th research seminar at the World Trade Institute as well as Michael Pfaffermayr, Robert Teh, Octavio Fernandez-Amador, Joseph Francois, and Hugo Rojas-Romagosa for useful comments. Address for correspondence: World Trade Organization, Rue de Lausanne 154, 1211 Geneve, Switzerland. email: eddy.bekkers@wto.org

The welfare effects of trade policy experiments in quantitative trade models: the role of solution methods and baseline calibration

ABSTRACT: This paper compares the solution methods and baseline calibration of three different quantitative trade models (QTMs): 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 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 predicted 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 fitted shares based on gravity estimation including pairwise fixed effects display similar performance in terms of robustness to the different sources of bias.

Keywords: Quantitative trade models, baseline calibration, FTAs, gravity estimation

JEL codes: F13, F14, F15.

1 Introduction

Quantitative trade models (QTMs) are employed frequently to determine the welfare effects of trade policy experiments. QTMs are employed both in ex ante studies on the expected effects of free trade agreements like TTIP, CETA, and TPP (Felbermayr et al., 2013; 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 of the breakup of free trade agreements like Brexit) and ex post studies on for example NAFTA ((Caliendo and Parro, 2015), but also in studies evaluating the expected effects of the WTO-agreement on trade facilitation (WTO, 2015). This paper distinguishes between three types of QTMs in the literature: computable general equilibrium (CGE) models (for example Hertel (2013)); 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 other 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 this paper concentrates on the other two differences. More precisely, the aim of this paper is to explore the impact of different baseline calibrations methods in QTMs on the predicted welfare effects of counterfactual trade policy experiments and to describe the different solution methods in one framework in a transparent way. As such the paper contributes to a better understanding of the differences between and similarities of the various approaches to quantitative trade modelling.

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

---

[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.

CGE-in-relative-changes), and EHA. 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 by conducting a typical counterfactual experiment in the literature (a reduction in trade costs between the USA and the EU, i.e. TTIP) based on Monte Carlo simulations with generated data disciplined by actual data. This exercise yields four main results. First, misspecification of the gravity equation can generate large biases in predicted welfare effects. Misspecification means in the context of this paper 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 flows into account, is of similar size as the misspecification bias with fitted shares, which is the result of using an average of trade flows over the entire estimation period based on 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 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.

To compare the three potential sources of bias, simulations are conducted with a data generating process disciplined by actual trade flows. In particular, hypothetical data are generated by adding random variation and measurement error to actual trade flows. This means that the bias as a result of misspecification is captured by starting with the actual data. Hence, this means that the data generating process is not entirely random. In such a setting with entirely randomly data, the data-generating process would have to account for misspecification to be able to compare the sources of bias across the different methodologies. And the degree of misspecification would have to be disciplined in turn by the degree of misspecification based on actual data. As pointed out into more detail in Subsection 4.2, this approach entails the risk that the degree of misspecification observed in actual data would not be correctly captured

---

[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.

and either under- or overestimated. As a result, this would bias the comparison between the different approaches. Therefore, in a first set of simulations random variation is added to actual trade data based on a data generating process with the variance disciplined by random variation in the error terms of a gravity equation including pairwise fixed effects estimated with real data. In a second set of simulations, measurement error is added to the actual data, based on measurement error observed in the data as measured by cif-fob margins distinguishing between time-varying and time-invariant measurement errors.

This paper 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 paper 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 paper 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 paper 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, which are employed extensively both in academic research and the evaluation of important policy decisions like the

---

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

creation or break-up of free trade agreements. 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 baseline calibration (to actual or 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.

The paper is organized as follows. Section 2 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 3 explores the differences between baseline calibration and solution methods of the four approaches numerically. Section 4 explores the three sources of bias of calibration to fitted and actual shares, respectively gravity misspecification and random variation and measurement error. Section 5 concludes.

2 Comparison of Methods: Theory

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

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

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:

$$P_j = \left(\sum_{i=1}^{J} \left(t_{ij} w_i\right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{2}$$

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

$$w_i L_i = \sum_{j=1}^{J} \left(t_{ij} w_i\right)^{1-\sigma} P_j^{\sigma-1} \left(1 + D_j\right) w_j L_j \tag{3}$$

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 (2)-(3) 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:

$$P_j = \left(\sum_{i=1}^{J} t_{ij}^{1-\sigma} \Pi_i^{\sigma-1} Y_i / Y_W \right)^{\frac{1}{1-\sigma}} \tag{4}$$

$$\Pi_i = \left(\sum_{j=1}^{J} t_{ij}^{1-\sigma} P_j^{\sigma-1} \left(1 + D_j\right) Y_j / Y_W \right)^{\frac{1}{1-\sigma}} \tag{5}$$

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:

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

2.2 Gravity estimation

The theoretical gravity equation in (1) corresponds with the following empirical gravity equation:

$$V_{ij}^{obs} = V_{ij} u_{ij} = \left(T_{ij} \eta_{ij}\right) X_i M_j \varepsilon_{ij} u_{ij} \tag{7}$$

V_{ij}^{obs} are the observed, actual trade flows, equal to the trade value V_{ij} times a measurement error, u_{ij}.[4] 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).[5] 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 in the value of trade. Observationally, η_{ij}, u_{ij}, and ε_{ij} can not be distinguished.[6] Rearranging (7) gives the following estimating equation:

$$V_{ij}^{obs} = T_{ij} X_i M_j \omega_{ij} \tag{8}$$

With:

$$\omega_{ij} = \eta_{ij} \varepsilon_{ij} u_{ij} \tag{9}$$

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.[7] Comparing equations (6)-(7) 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:[8]

$$t_{ij}^{1-\sigma} = T_{ij} \eta_{ij} \varepsilon_{ij} \tag{10}$$

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:

$$t_{ij}^{correct} = \left(E\left[t_{ij}^{1-\sigma} | T_{ij}, \eta_{ij} \right] \right)^{\frac{1}{1-\sigma}} = \left(T_{ij} \eta_{ij} \right)^{\frac{1}{1-\sigma}} \tag{11}$$

---

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

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

[6] 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.

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

[8] 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.

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} instead is a a random component in the value of trade, 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.

2.3 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:

$$\widetilde{W_i} = \frac{W_{c,i}}{W_i} - 1 = \frac{w_{c,i}}{w_i}\frac{P_i}{P_{c,i}} - 1 = \frac{Y_{c,i}}{Y_i}\frac{P_i}{P_{c,i}} - 1 \tag{12}$$

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).

2.3.1 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 (4)-(5) employing the actual values for Y_i and E_j and the fitted values for iceberg trade costs, t_{ij}^{sg}:

$$t_{ij}^{sg} = (T_{ij})^{\frac{1}{1-\sigma}} \tag{13}$$

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.[9] 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:

$$\frac{Y_{c,i}}{Y_i} = \left(\frac{\Pi_i}{\Pi_{c,i}}\right)^{\frac{\sigma-1}{\sigma}} \tag{14}$$

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 (4)-(5) and then calculate counterfactual values for $P_{c,i}$, $\Pi_{c,i}$ and income $Y_{c,i}$ using equations (4)-(5) and (14).[10] $t_{c,ij}^{sg}$ is calculated by changing the value of the policy variable, which is part of the fitted trade costs $\widehat{T_{ij}}$. 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:[11]

$$t_{c,ij}^{sg} = t_{ij}^{sg}\left(\frac{T_{c,ij}}{T_{ij}}\right)^{\frac{1}{1-\sigma}} = t_{ij}^{sg}\exp\left(\frac{\beta_{FTA}}{1-\sigma}\right) \tag{15}$$

---

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

[10] 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.

[11] Observable trade costs T_{ij} can be written as $T_{ij} = \exp\left\{\beta_{FTA}FTA_{ij} + \beta_{other}other_{ij}\right\}$, 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 (15).

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 following equations hold with Y_i and E_j actual output and expenditures:[12]

$$X_i = \Pi_i^{\sigma-1} \frac{Y_i}{Y_W} \tag{16}$$

$$M_j = P_j^{\sigma-1} E_j \tag{17}$$

Anderson et al. (2015) propose to use the finding by Fally to run counterfactual experiments by estimation in STATA instead of simulation.[13] 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 (16)-(17). Then the gravity equation (8) 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:

$$Y_{c,i} = \left(\frac{X_{c,i}}{X_i}\right)^{\frac{1}{1-\sigma}} Y_i \tag{18}$$

Then the gravity equation is re-estimated imposing new values for trade V_{ij} based on the theoretical gravity equation in (6):

$$V_{ij}^c = \left(\frac{t_{c,ij}^{sg}}{t_{ij}^{sg}}\right)^{1-\sigma} \frac{Y_i^c E_j^c}{Y_i E_j} \frac{\Pi_i^{1-\sigma} P_j^{1-\sigma}}{\Pi_{i,c}^{1-\sigma} P_{j,c}^{1-\sigma}} V_{ij} \tag{19}$$

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.

2.3.2 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).

---

[12] 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 (6) and (8), leads then to equations (16)-(17).

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

[14] Equations (2)-(3) 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 (15):

$$t_{c,ij}^{cge} = t_{ij}^{cge} \exp\left(\frac{\beta_{FTA}}{1-\sigma}\right) \tag{20}$$

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 (1)):[15]

$$t_{ij}^{cge} = \left(\frac{V_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}} \tag{21}$$

It can easily be shown that $w_i = P_j = 1$ is a solution of the equilibrium equations (2)-(3), given equation (21). This is therefore a convenient and harmless normalization.

2.3.3 CGE-in-relative-changes

CGE models such as the GTAP-model (Hertel (2013)) write the equilibrium equations in relative changes. Log differentiating the equilibrium equations (2)-(3) leads to:

$$\widetilde{P_j} = \sum_{i=1}^{J} impsh_{ij} \left(\widetilde{t_{ij}} + \widetilde{w_i}\right) \tag{22}$$

$$\widetilde{w_i} = \sum_{j=1}^{J} expsh_{ij} \left((1-\sigma)\left(\widetilde{t_{ij}} + \widetilde{w_i} - \widetilde{P_j}\right) + \widetilde{w_j}\right) \tag{23}$$

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 (22)-(23) as such would obviously lead to inexact solutions in case of larger shocks, as a first-order approximation is employed. However, the software employed

---

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

[15] 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 (29) this is equivalent to the CGE-in-levels calibration in equation (21) and implies that baseline import shares are equal to actual import shares in the data.

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. [16].

2.3.4 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 (2)-(3) in the counterfactual and baseline and rearranging gives:

$$\widehat{P_j} = \left(\sum_{i=1}^{J} impsh_{ij} \left(\widehat{t_{ij}} \widehat{w_i} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{24}$$

$$\widehat{w_i} = \sum_{j=1}^{J} expsh_{ij} \left(\widehat{t_{ij}} \widehat{w_i} \right)^{1-\sigma} \widehat{P_j}^{\sigma-1} \widehat{w_j} \tag{25}$$

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. [17]

2.4 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.[18] Substituting equation (13) into (6), and applying equations (16)-(17), gives:

$$impsh_{ij}^{sg} = \frac{V_{ij}}{E_j} = \left(t_{ij}^{sg} \right)^{1-\sigma} \Pi_i^{\sigma-1} P_j^{\sigma-1} \frac{Y_i}{Y_W} = \frac{T_{ij} X_i M_j}{E_j} = \frac{V_{ij}^{fitted}}{E_j} \tag{26}$$

---

[16] Further details on solution methods in GEMPACK can be found for example in Horridge et al. (2013)

[17] Scholars applying exact hat algebra typically work with MATLAB.

[18] In the exposition below we focus on import shares. Expressions for export shares employed under the different methodologies would be similar.

Using the empirical gravity equation in (8), 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):

$$impsh_{ij}^{sg} = \frac{T_{ij} X_i M_j}{E_j} \tag{27}$$

$$impsh_{ij}^{eha,cge} = \frac{T_{ij} X_i M_j \eta_{ij} u_{ij} \varepsilon_{ij}}{E_j} \tag{28}$$

Alternatively, the baseline trade costs can be compared. Substituting the empirical expression for the actual value of trade V_{ij} in equation (8) into the expression for t_{ij} under CGE-calibration, equation (21), gives:

$$t_{ij}^{cge} = \left(\frac{V_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}} = \left(\frac{T_{ij}\eta_{ij} X_i M_j u_{ij}\varepsilon_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}} = \left(T_{ij}\eta_{ij}u_{ij}\varepsilon_{ij}\right)^{\frac{1}{1-\sigma}} \tag{29}$$

Under SG-calibration the error terms are neglected. The expression is given in equation (13).

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 (27)-(28) and the baseline trade costs in (13) and (29). However, based on equation (11) 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 4 the robustness of the different methods to misspecification, to random variation, and to measurement error is examined with simulations.

Comparing the solution methods, the SG-approach 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 is expected to lead to the same solution as CGE-in-levels or exact hat algebra. 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 (12) the percentage change in welfare, $\widetilde{W}_i$, is determined by $\widehat{w_i}$ and $\widehat{P_i}$. From equations (24)-(25), $\widehat{w_i}$ 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 (1), $impsh_{ij} = \frac{V_{ij}}{(1+D_j)Y_j}$ and $expsh_{ij} = \frac{V_{ij}}{Y_i}$ are determined by t_{ij}, D_j, w_j, P_j, and Y_i. Substituting equation (2) into equation (3) leads to:

$$Y_i = \sum_{j=1}^{J} \left(\frac{t_{ij} w_i}{\sum\limits_{k=1}^{J} \left(t_{kj} w_k \right)^{1-\sigma}} \right)^{1-\sigma} \left(1 + D_j\right) Y_j \tag{30}$$

It can be shown that equation (30) 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 (24), 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 three 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 Appendix B.2 it is shown that the solution procedure with t_{ij} based on the bilateral terms in the fitted gravity equation as in equation (13) 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.

Third, Appendix C verifies the claims in Result 1 numerically by showing that the different solution methods lead to the same welfare effects of a counterfactual experiment if (implied) baseline import shares are identical. This also means that there is 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 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.

3 Comparison of methods: simulations

In this section the impact of baseline calibration on the welfare effects of a counterfactual experiment is assessed with simulations. The model described in Subsection 2.1 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 (15) a shock of about 14%.

The impact of variations in baseline calibration on predicted welfare effects of a set of counterfactual exercises is examined 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 2.1 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 in Appendix C. Hence, equations (2)-(3) 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 (12).

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 $t_{ij}^{1-\sigma} = \frac{V_{ij}^{data}}{E_j^{data}} \frac{E_j^{data}}{E_j} \upsilon_{ij}^{1-\sigma}$.[19] So in terms of equation (10) the setup corresponds with $T_{ij}\eta_{ij} = \frac{V_{ij}^{data}}{E_j^{data}}$ and $\varepsilon_{ij} = \frac{E_j^{data}}{E_j} \left(\upsilon_{ij}\right)^{1-\sigma}$.[20] Hence, applying equation (11), 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 conducted both with the correct baseline data and with the 100 randomly simulated baseline data.

The EHA-equilibrium equations, (24)-(25), 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:

$$\begin{aligned} diff\_welfare_{r,j} &= \beta_0 + \beta_1 diff\_impsh_{r,j,partner} + \beta_2 diff\_expsh_{r,j,partner} \\ &\quad + \beta_3 \left(diff\_domsh_{r,j}\right) + \eta_j + \zeta_r + \varepsilon_{r,j} \end{aligned} \tag{31}$$

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 1 displays the results of this regression for the EU countries. In column 1 all observations are included

---

[19] This expression for baseline trade costs corresponds with equilibrium values $w_i = P_j = 1$ as discussed in Subsection 2.3.2.

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

Table 1: 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)	(2)	(3)	(4)	(5)
	diff_welfare	diff_welfare	diff_welfare	diff_welfare (average across replications)	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

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

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 1 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 1. A larger import share from the US leads to a larger welfare effect.

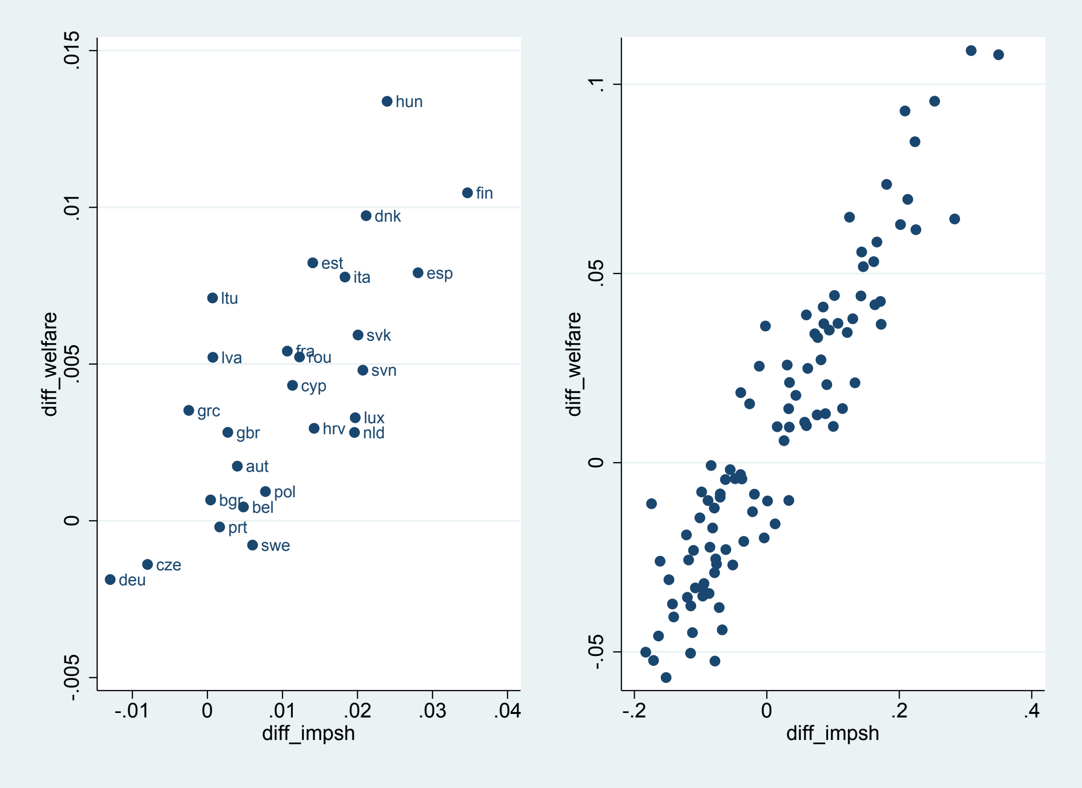

Figure 1: The impact of deviations in import shares from the US from their mean on deviations of welfare effects of the EU countries from their mean after a reduction in trade costs between the EU and the US (TTIP)

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.

4 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 4.1 only allows for bias as a result misspecification, Subsection 4.2 allows for bias as a result of both misspecification and random variation, and Subsection 4.3 allows for bias as a result of both misspecifaction and measurement error.

4.1 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.

4.1.1 Cross-section gravity estimation

Table 2 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. (2013) 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 2) 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 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.[21] Based on gravity specification (1) in column seven the welfare effects are

---

[21] The GTAP data contain domestic flows, but for example the COMTRADE data with a larger set of countries

Table 2: 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.

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.

4.1.2 Panel gravity estimation

Table 3 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 2 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

---

(up to 180) contain only international trade flows.

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, 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.

4.2 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 (7). 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 (8) 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.[22] Working with one variance

---

[22] 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.

Table 3: 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 Gravity specification	Actual shares	Fitted shares (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.

for all trade flows, would overstate the variance in trade flows for larger observations.[23]

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 Section 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. The randomly generated data would have to be a function of bilateral explanatory variables like distance to capture misspecification in cross-section gravity estimation without pairwise fixed effects, contain variation over time in the pairwise component to capture misspecification of gravity estimation with pairwise fixed effects, and finally contain random variation in the pairwise component to capture the bias involved in employing actual shares. Developing a data-generating process with all the three elements involves the risk of over- or underestimating the degree of misspecification or random variation with the implication that one method would erroneously be judged as superior over another method. Therefore, a data generating process was created by adding random variation to the actual data with the random variation disciplined by random variation in the actual data.

Table 4 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

---

[23] 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 4: The mean squared error of the predicted welfare effect under random variation of trade flows employing actual shares and fitted shares

Calibration to	Actual shares	Fitted shares		
Gravity specification		(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.

different groups, either population- or GDP-weighted:

$$MSE_{group}^{rv} = \sum_{i \in group} \frac{weight_i \frac{1}{100} \sum_{rep=1}^{100} \left(\widetilde{W}_{i,rep}^{rv} - \widetilde{W}_{i,rep}^{correct} \right)^2}{weight_i} ; weight = GDP, population \tag{32}$$

$\widetilde{W}_{i,rep}^{correct}$ 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 3 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 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.[24]

4.3 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 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.

---

[24] 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 5: 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 Gravity specification	Actual shares	Fitted shares (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.

Measurement errors are split up into time-varying and time-invariant measurement errors by regressing the measurement errors $\ln v_{ijt}$ 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:

$$\begin{aligned}\ln v_{ijt} &= \alpha_1 \ln distance_{ij} + \alpha_2 contiguity_{ij} + \alpha_3 common\_colony_{ij} + \varphi\sigma^2_{ijt} + \iota_{ijt} \\ \sigma^2_{ijt} &= \exp\{\lambda_0 + \lambda_1 \ln distance_{ij} + \lambda_2 contiguity_{ij} + \lambda_3 common\_colony_{ij} \\ &\quad + \lambda_4 gdp_{it} + \lambda_5 gdp\_pc_{it} + \lambda_6 gdp_{jt} + \lambda_7 gdp\_pc_{jt}\}\end{aligned} \tag{33}$$

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

Table 6: 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.

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

$$MSE^{me}_{group} = \sum_{i \in group} \frac{weight_i \frac{1}{100} \sum_{rep=1}^{100} \left(\widetilde{W}^{me}_{i,rep} - \widetilde{W}^{correct}_{i,rep} \right)^2}{weight_i}; weight = GDP, population \qquad (34)$$

$\widetilde{W}^{me}_{i,rep}$ is the percentage change in welfare based on the data with measurement error. The table displays the same patterns as Table 4 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 6 displays the MSE based on trade flows including measurement errors generated with variances varying by the gravity regressors.[25] 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.

5 Concluding remarks

This paper classified and compared different quantitative trade models, in terms of model exposition, solution method and calibration of baseline trade costs. As such the paper 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 paper 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 estima-

---

[25] The regression outcomes of the model in equation (33) to determine the variance of the measurement error are given in Appendix Appendix F

tion 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.[26] 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.[27]

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.[28] 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. 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)

---

[26] 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.

[27] 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.

[28] 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.

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.

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.

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.

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.

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

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

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.

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

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., M. Larch, L. Flach, and E. Y. und Sebastian Benz (2013). “Dimensionen und Auswirkungen eines Freihandelsabkommens zwischen der EU und den USA,” Tech. rep., CES-IFO. Studie im Auftrag des Bundesministeriums fuer 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).

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.

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

Hertel, T. (2013). “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.

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.

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.

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.

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.

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:[29]

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

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}:

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

---

[29] 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.

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 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 (4)-(5) in the baseline employing the actual values for Y_i and the fitted values for iceberg trade costs, t_{ij}^{sg}, in equation (13). 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 (15) and adding equation (14) 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 (16)-(17):

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

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

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 (2)-(3), 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:

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

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

Appendix B.3 CGE-in-levels

Equations (2)-(3) 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 (21). In the baseline w_i and P_j are both set at 1, which are solutions for equations (2)-(3) and obviously give baseline import and export shares equal to the actual shares.

Appendix B.4 CGE-in-relative-changes

The equilibrium equations (22)-(23) 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:

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

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

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

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.

Table 7: 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

Appendix B.5 Exact hat algebra

Equations (24)-(25) 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:

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

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

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

Appendix C The impact of solution methods

Table 7 displays the welfare effects of the solution methods described in 2.4.[30] 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

---

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

Table 8: 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 (16)-(17).

import shares (called estibration by Yotov et al. (2016), see footnote 9) all lead to exactly the same solution.[31] 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 8 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.

---

[31] 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.

Supplementary appendices to be distributed as online appendix

Appendix D Derivation equations

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

Equations (22)-(23)

Hat-differentiating equations (2)-(3) gives:

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

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

It is easy to show that the coefficients in the summations of equations (D.1)-(D.2) are respectively equal to the import shares $impsh_{ij}$ and export shares $expsh_{ij}$ and thus lead to equations (22)-(23).

Equations (4)-(6)

To start equation (3) can be rearranged as follows, imposing $Y_i = w_i L_i$:

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

Hence, Π_i can be defined as follows, equation (5) in the main text:

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

Equation (5) implies the following expression for w_i:

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

Substituting equation (D.4) into equations (1)-(2) leads to equations (4) and (6):

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

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

Equation (14)

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 (14) will generate an approximate solution.

Equation (18)

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

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

Comparing the two expressions for P_j in (2) and (4) shows that w_i can be written as:

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

Using the expression for the estimated fixed effect, equation (16), w_i can be written as:

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

Substituting equation (D.9) into equation (D.7) and rearranging then leads to equation (18) in the main text:

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

Baseline shares equal to actual shares with CGE in levels approach

Calculating the import and export shares from equation (1) and substituting the CGE-in-levels calibration of baseline trade costs in equation (21) leads to:

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

Showing that equation (21) *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 (2)-(3), given equation (21):

$$
\begin{aligned}
P_j &= \left(\sum_{i=1}^{J} (t_{ij}w_i)^{1-\sigma}\right)^{\frac{1}{1-\sigma}} = \left(\sum_{i=1}^{J} \frac{V_{ij}}{E_j}\right)^{\frac{1}{1-\sigma}} = 1 \\
w_i L_i &= \sum_{j=1}^{J} (t_{ij}w_i)^{1-\sigma} P_j^{\sigma-1} (1+D_j)\, w_j L_j = \sum_{j=1}^{J} \frac{V_{ij}}{E_j} E_j
\end{aligned}
$$

Appendix E 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:

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

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

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

With:

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

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

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

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 F Additional counterfactual experiments and auxiliary regression results

Table 9: 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 10: 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 11: 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 12: 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

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

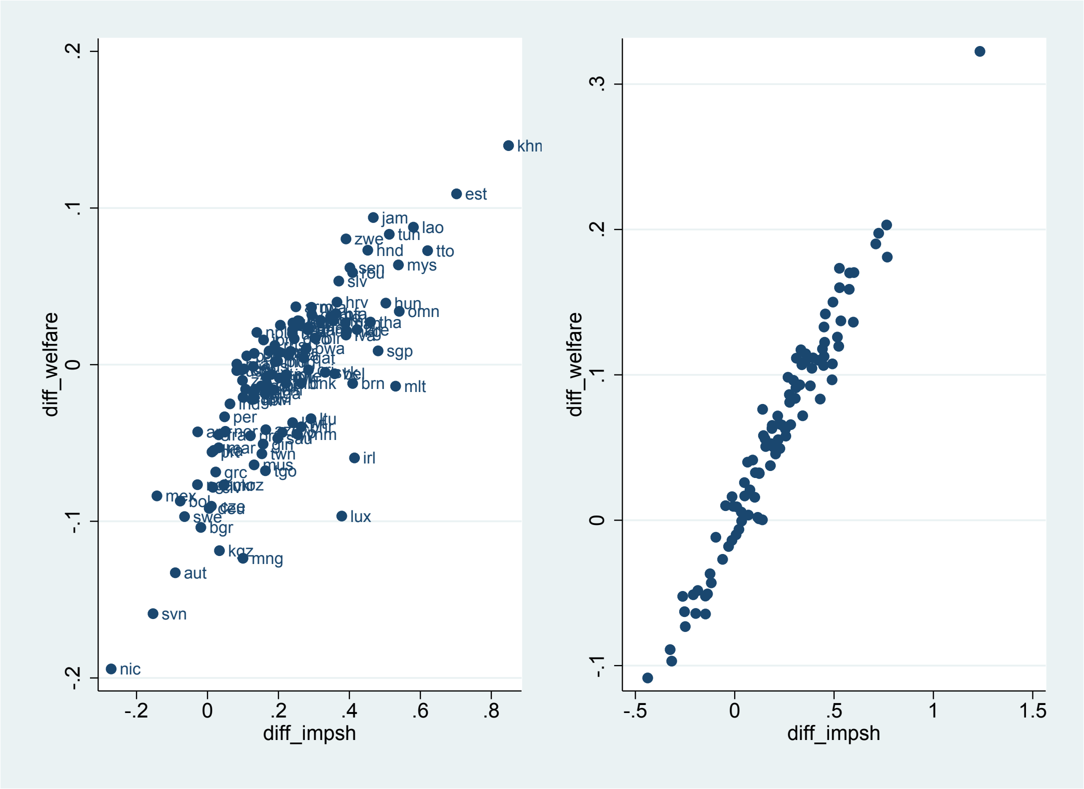

Figure 2: The impact of deviations in average import shares from their mean on deviations of welfare effects from their mean after a reduction in trade costs between all countries (multilateral liberalization)

Table 13: 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 14: 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

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