Federal Reserve Bank of Minneapolis
Research Department Staff Report

November 2024

Improving Trade Policy Modeling with Best-Fit Elasticities

Timothy J. Kehoe

University of Minnesota,
Federal Reserve Bank of Minneapolis,
and National Bureau of Economic Research

Pau S. Pujolas

McMaster University

Jack Rossbach

Georgetown University Qatar

Kim J. Ruhl

University of Wisconsin Madison
and National Bureau of Economic Research

ABSTRACT

Applied General Equilibrium (AGE) models are the dominant tool for the analysis of trade policy. Unfortunately, AGE models have historically performed poorly when used to predict the industry-level impact of trade liberalization. We develop a flexible methodology for estimating industry-level trade elasticities based on finding the elasticities that yield model predictions that best fit previous liberalizations, which we identify using a genetic algorithm. When applied to future liberalization episodes, these Best-Fit Elasticities produce model predictions that significantly outperform elasticities from the literature, which are estimated on cross-sectional or year-to-year fluctuations.

JEL Codes: C58, F10, F17
Keywords: applied general equilibrium; trade liberalization; model validation; least traded products

*The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

Applied General Equilibrium (AGE) models, which feature multiple industries, multiple regions, and input-output linkages across industries, are the primary tool that policymakers employ when predicting the impact a trade liberalization will have on the industrial composition of the economy. Ideally, AGE models should offer reasonably accurate predictions, foreseeing which industries will grow and which will decline following liberalization. Unfortunately, they have often fallen short in this regard. In a series of papers, Kehoe (2005), Kehoe, Rossbach and Ruhl (2015, henceforth KRR), and Kehoe, Pujolas and Rossbach (2017, henceforth KPR), we have highlighted this inadequacy.[1] In particular, AGE model predictions often have zero (or even negative) correlation with the industry-level changes in trade flows that actually occur following a liberalization. Despite these shortcomings, the conclusion is not to discard AGE models altogether. Instead, the predictive accuracy of AGE models can and should be improved in order to best analyze the effects of trade policy.

In this paper, we demonstrate a novel methodology for calibrating sector-specific trade elasticities that are consistent with past liberalization episodes. We refer to these elasticities as "Best-Fit Elasticities" (BFE) to distinguish them from elasticities calibrated to be consistent with cross-sectional dispersions or year-to-year fluctuations. These elasticities deliver the AGE model predictions most aligned with actual changes in the data. Then, we show that BFE predict reasonably well other trade liberalizations not used in their computation. For example, to assess the impact of the Korea-U.S. Free Trade Agreement (KORUS), we calculate the elasticities for the North American Free Trade Agreement (NAFTA) that yield the highest weighted correlation between the predictions of our selected AGE model and the observed industry-level changes in Canada, Mexico, and the United States following NAFTA. We demonstrate that employing these "NAFTA-BFE" substantially enhances the predictive accuracy of the AGE model. Whereas the weighted correlation between the observed changes post-KORUS and the predictions generated by our standard AGE model is -0.31 when utilizing industry-level elasticities from Caliendo and Parro (2015, hereafter referred to as C-P), this correlation increases to 0.34 when employing our NAFTA-Best Fit elasticities.

[1] Overall, our analysis encompasses several Applied General Equilibrium (AGE) models and instances of trade liberalization. These include a prominent model developed for NAFTA around the time of its enactment (Brown et al., 1995), a significant recent model tailored for NAFTA long after its establishment (Caliendo and Parro, 2015), and the widely used AGE model employed by policymakers globally (GTAP, as outlined by Hertel, 2013).

We demonstrate that this enhanced performance remains consistent across various episodes of liberalization, encompassing both the liberalization events used to calibrate our elasticities and those to which we apply our AGE models. Our methodology tackles several well-known limitations associated with cross-sectional and other common forms of elasticity estimation. Specifically, while pre-existing elasticities are often employed to forecast the impacts of liberalization, there is scant justification for assuming that cross-sectional elasticity estimates accurately reflect the elasticities observed during specific liberalization episodes, as highlighted in Ruhl (2008).

Our approach to calibrating industry-level trade elasticities is guided by the principle that elasticities used for policy analysis should be consistent with the policy being evaluated and that the calibration method should allow for flexibility in modeling. When using an AGE model to conduct policy analysis, researchers will always have two things in hand: a model and a policy. Our suggestion is that researchers can simply apply their model to find the best fit elasticities for a similar policy, which is what we do in this paper applied to trade liberalizations, although we expect our approach to apply broadly. We are not arguing that this approach necessarily outperforms all alternative calibration methods, and the appropriate measure for assessing fit may differ across contexts. For example, to evaluate a potential liberalization between low income countries, it might be necessary to fit elasticities to match a similar policy involving low income countries, rather than for high income countries. Instead, what we contend is that our method is a straightforward and reasonable approach, which addresses two major issues with using off-the-shelf elasticities: they are not necessarily consistent with the policy being evaluated and they are not necessarily consistent with the model being employed. It is important to clarify this later point, since many papers estimate elasticities in a manner that is fully consistent with their model and we do not contend otherwise. Our point is that there is a common practice of borrowing elasticity estimates from other papers, and although these papers provide estimates that are consistent with their own models, these models may differ in meaningful ways from the models of researchers or policy makers borrowing those elasticities.

In KPR, we discuss the importance of calibrating Armington elasticities to be consistent with trade liberalizations if the goal is to evaluate the impact of trade liberalizations. Most applications of AGE models use elasticities that were calibrated to be consistent with either cross-sectional dispersion in trade costs and trade volumes or calibrated to match year-to-year

fluctuations in trade costs and trade volumes. There is good reason to doubt that elasticities consistent with those features of the data should also be consistent with trade liberalizations, in part because these elasticity estimates are not consistent with each other. The correlation between the industry-level elasticities in , which are calibrated to cross-sectional data, and the industry-level elasticities in Imbs and Mejean (2017, henceforth I-M), which are calibrated to annual fluctuations, is –0.01.

Our methodology offers several strengths beyond enhancing the predictive capabilities of AGE models. Firstly, when applied to the same trade liberalization under study, it establishes a theoretical upper limit on the performance of our AGE model with any set of elasticities—providing both a benchmark for assessing model efficacy and a metric for determining when extensions or modifications to an AGE model might be necessary to match the data. Secondly, our approach is very flexible. This flexibility is significant because different AGE models imply distinct optimal elasticities, yet it is common practice to employ the same set of elasticities across different models, irrespective of their specific features such as input-output linkages. Thirdly, our methodology allows for easy customization of the best fit criteria employed. While we primarily focus on weighted correlations in this paper due to their simplicity and consistency with metrics utilized in prior studies, such as Kehoe (2005), KRR, and KPR, other fit criteria may be more suitable depending on the objectives of the AGE model. This adaptability ensures that the methodology can be tailored to meet specific modeling requirements.

2. The Standard AGE Model

In this section we develop a standard multi-sector, multi-country, Applied General Equilibrium (AGE) model, which we will use to calibrate elasticity parameters. For simplicity of notation, we suppress time subscripts until we discuss using our AGE model to conduct policy experiments in section 4.

The household in country $i \in I$ derives utility from consuming c_i^s units of goods from industry $s \in S$, whose price is p_i^s. Her income is made of payments from market activities, w_i, plus rebated tariffs T_i. Households maximize utility subject to their budget constraint,

$$\max \sum_{s=1}^{S} a_i^s \log\left(c_i^s\right) \tag{1}$$

$$\text{s.t. } \sum_{s=1}^{S} p_i^s c_i^s = w_i L_i + T_i + D_i,$$

where a_i^s are the (country-specific) industry weights in the utility function and D_i is the trade balance. Let $Income_i = w_i L_i + T_i + D_i$, then consumption of c_i^s is equal to

$$c_i^s = a_i^s \frac{Income_i}{p_i^s}. \tag{2}$$

Each industry produces final industry output y_i^s combining domestic, y_{ii}^s, and foreign, y_{ij}^s (with i different that j), industry output according to the following Constant Elasticity of Substitution Armington specification,

$$y_i^s = \left(\sum_{j=1}^{J} \gamma_{ij}^s \left(y_{ij}^s \right)^{\frac{\sigma_s - 1}{\sigma_s}} \right)^{\frac{\sigma_s}{\sigma_s - 1}}. \tag{3}$$

Here γ_{ij}^s is the industry weights for consumption in country i, originating from country j, in industry s, and σ_s is the industry elasticity. The producers of this bundles are perfectly competitive and cost-minimize expenses according to

$$\min \sum_{j=1}^{J} p_{ij}^s y_{ij}^s, \tag{4}$$

subject to the production of output in equation (3). This problem has the following demand as solution

$$y_{ij}^s = y_i^s \left(\gamma_{ij}^s \frac{p_i^s}{p_{ij}^s} \right)^{\sigma_s}, \tag{5}$$

where

$$p_i^s = \left(\sum_{j=1}^{J} \left(\gamma_{ij}^s \right)^{\sigma_s} \left(p_{ij}^s \right)^{1-\sigma_s} \right)^{\frac{1}{1-\sigma_s}}. \tag{6}$$

Producers operate according to a constant returns to scale production function where intermediate inputs are utilized from other sectors in fixed proportions. This specification allows us to specify the firm's profit maximization profit separately for each destination. Specifically, firms from country j in sector s exporting to country i solve the profit maximization problem given by

$$\max \frac{p_{ij}^s y_{ij}^s}{\tau_{ij}^s} - w_j \ell_{ij}^s - \sum_{k=1}^{S} p_j^k x_{ij}^{s_k}, \tag{7}$$

subject to their production function

$$y_{ij}^{s} = \min\left\{ z_j^s \ell_{ij}^{s,m}, \frac{x_{ij}^{s_1}}{\alpha_j^{s_1}}, ..., \frac{x_{ij}^{s_k}}{\alpha_j^{s_k}}, ..., \frac{x_{ij}^{s_S}}{\alpha_j^{s_S}} \right\}, \tag{8}$$

where $x_{ij}^{s_k,m}$ is the quantity of composite intermediate inputs from industry k used in the production of industry s in country j for consumption in country i, and $\alpha_j^{s_k}$ is the unit input requirement of inputs from industry k used in the production of industry s, and τ_{ij}^{s} is the ad-valorem tariff rate paid by consumers. Note that the unit input requirements do not depend on the destination, only on where the output is produced and in which industry.

The Leontieff structure of the production function implies that

$$z_j^s \ell_{ij}^s = \frac{x_{ij}^{s_1}}{\alpha_j^{s_1}} = ... = \frac{x_{ij}^{s_k}}{\alpha_j^{s_k}} = ... = \frac{x_{ij}^{s_S}}{\alpha_j^{s_S}}, \tag{9}$$

which means that the demand for intermediate goods is given by

$$x_{ij}^{s_k} = \alpha_j^{s_k} z_j^s \ell_{ij}^s. \tag{10}$$

implying equilibrium prices given by

$$p_{ij}^{s} = \tau_{ij}^{s}\left(\frac{w_j}{z_{ij}^{s}} + \sum\nolimits_{k=1}^{S} p_j^k \alpha_j^{s_k} \right). \tag{11}$$

Tariff revenue in country i is collected by the government and rebated back to consumers,

$$T_i = \sum\nolimits_{s=1}^{S} \sum\nolimits_{j=1}^{J} \left(\tau_{ij}^{s} - 1\right)\left(p_{ij}^{s} y_{ij}^{s}\right). \tag{12}$$

We also allow for aggregate trade deficits, where the trade deficit for country i is given by,

$$D_i = \sum\nolimits_{s=1}^{S} \left(\sum\nolimits_{j=1}^{J} \frac{p_{ij}^{s} y_{ij}^{s}}{\tau_{ij}^{s}} - \sum\nolimits_{j=1}^{J} \frac{p_{ji}^{s} y_{ji}^{s}}{\tau_{ji}^{s}} \right). \tag{13}$$

Note that the aggregate trade deficit still allows for bilateral trade deficits to be endogenously determined. As an alternative to including aggregate trade deficits, we can also rebalance each countries trade with ROW so that trade is balanced on aggregate. This alternative specification has no impact on the results in our paper or the performance of AGE models for the liberalizations we study.

The final part of the AGE model are the market clearing conditions. The labor market clearing condition for country j is equal to the sum of labor used in production aggregated across all destination markets and industries,

$$L_j = \sum_{s=1}^{S} \sum_{i=1}^{J} \left(\ell_{ij}^{s} \right). \tag{14}$$

Similarly, the market condition for intermediate inputs is similarly given by

$$x_j^k = \sum_{s=1}^{S} \sum_{i=1}^{J} \left(x_{ij}^{s_k} \right), \tag{15}$$

where x_j^k is the total amount of final output from industry k that is used as intermediate inputs in country j, and the market clearing condition for final output in country j is given by

$$y_j^k = x_j^k + c_j^k \ . \tag{16}$$

3. Data and Calibration of Standard Model Parameters

In this section we explain our calibration approach for all model parameters other than the industry-level trade elasticities, σ_s. We develop a novel calibration method for those parameters, which we discuss in section 4, and what follows in this section takes σ_s as given.

To calibrate the model, we match the data exactly in the base period. Calibrating our model requires data on bilateral industry-level trade flows, production, and tariffs, along with industry-level input-output flows by country. Our primary data source for trade, production, and IO flows is the 2013 release of the World Input Output Database (WIOD), as described in Timmer et al. (2015). The WIOD covers 1995 to 2011 and contains industry-level trade (including self-trade) and IO flows for 40 countries, including a Rest of the World (ROW) aggregate, and 35 industries. We further aggregate the countries into 13 regions, including a European Union Customs Union aggregate. We focus on the 16 industries in the WIOD encompassing agriculture, forestry, fishing, mining, and manufacturing industries, where each WIOD industry is an aggregate of 3-digit International Standard Industrial Classification (ISIC) Rev 3 Codes.

The WIOD also includes information on industry-level output, value added, and employment for the period 1995-2009, which we use to compute industry-level productivity measures. We extend these series using data from United Nation's UNIDO Industrial Statistics (INDSTAT) Database, which contains information on employment, output, and value added for 3-digit ISIC Rev 3 manufacturing industry codes, which we supplement with OECD and country-specific KLEMS data — referring to data on capital (K), labor (L), energy (E), materials (M), and services (S) — for employment and productivity information for the agricultural and mining industries. This data, along with the list of industry and country aggregates, is described in more

detail in Appendix A.1. This additional information allows us to account for shocks that occurred post liberalization when calibrating tariffs and when evaluating model performance.

We combine the WIOD trade flows with bilateral product-level trade flows for each country from the United Nation's Comtrade Database at the 5-digit Standard International Trade Classification (SITC) Rev 2 level. There are 1,836 5-digit SITC product codes, which we are able to aggregate into 3-digit ISIC industries using a concordance provided by the World Integrated Trade Solution (WITS) Database, and then into WIOD industry codes using the concordance provided by Timmer et al. (2015). The Comtrade data allows us to extend our industry-level trade flows to 1989 to 2015.

We use data on bilateral industry-level tariffs at the 3-digit ISIC level from WITS, which calculates industry-level tariff equivalents using disaggregated tariff data from the UN's Trade Analysis Information System (TRAINS) database. When tariff data is not available in TRAINS, we use WITS industry-level trade flows based on the World Trade Organization's (WTO) Integrated Database. The strength of the TRAINS database is that it estimates ad-valorem tariff rate equivalents for non-tariff trade barriers such as quotas. Tariff agreements typically set caps on the maximum tariff rate that can be applied, however, in practice countries typically impose tariff rates below the maximum caps. Therefore, when possible we use effectively applied tariff rates, otherwise we use the minimum tariff rate cap that applies to the particular trade flow.

Teti (2023) notes that preferential tariff rates are not consistently reported by all countries across partners and years in the WITS TRAINS database, and WITS deals with the issue by interpolating missing preferential tariffs with most favored nation tariffs, which creates false spikes in the data that are not consistent across partners or years. We observe the same issue in the data we use; for instance, the simple average tariff rate for exports from the United States to Mexico in the Agricultural industry is 1.84 percent 2002 and 0 percent in 2004, but 18.39 percent in 2003, which is an example of a false spike.[2] We interpolate missing tariff rates with previous year values. To minimize the impact of false spikes we sequentially replace values that are more than double the rate of the previous year with a greater than 2 percentage point difference or that are 10 percentage points greater than the previous year with their prior year value for countries that have signed a regional trade agreement.

[2] Although false spikes are present for NAFTA countries, in data they do not occur in any of 1999, 2005, 2007, or 2009; which are the end years examined in our previous work.

Many of the parameters in our model are straightforward to calibrate. Since we are interested only in changes, initial wages and productivities have no effect on our results. Therefore, we begin our calibration by normalizing both wages and productivities to one. We are able to take trade deficits directly from the data, and we calculate implied tariff revenues from trade flows and industry-level tariff rates. We use the structure of the model to back out labor allocations using value added by industry and destination. With this information we can compute sectoral prices and outputs, as well as the Armington weight parameters, which govern home bias. Lastly, we calibrate input-output share parameters, consumption, intermediate inputs, and utility shares for each industry. In the remainder of this section we present the specific equations we use to calibrate each object.

In our model, all expenditures are from the consumer side and therefore inclusive of tariffs. From our trade flow data, we have expenditures exclusive of trade costs, $p_{ij}^s y_{ij}^s / \tau_{ij}^s$, which we combine with data on tariff rates, τ_{ij}^s, to get industry-level bilateral expenditures, $p_{ij}^s y_{ij}^s$.

With wages and productivities normalized to 1, $w_j = 1$ and $z_j^s = 1$, we calibrate the trade balance directly from equation (13), initial tariff revenues from equation (12), and income as

$$Income_i = \sum_{s=1}^{S} \sum_{i=1}^{J} p_{ij}^s y_{ij}^s - \sum_{s=1}^{S} p_i^s x_i^s \,. \tag{17}$$

Then aggregate labor supply comes directly from the budget constraint as

$$L_i = \frac{Income_i - T_i - D_i}{w_i}. \tag{18}$$

Furthermore, noting that an industry's value added is given by

$$VA_{ij}^s = \frac{p_{ij}^{s,non} y_{ij}^{s,non} + p_{ij}^{s,LTP} y_{ij}^{s,LTP}}{\tau_{ij}^s} - \sum_{k=1}^{S} \alpha_j^{s_k} p_j^k y_{ij}^{s,non} - t_{ij}^{s,m} \sum_{k=1}^{S} \alpha_j^{s_k} p_j^k y_{ij}^{s,LTP}, \tag{19}$$

which we combine using equation (11) to get p_{ij}^s. In turn, we can combine data on expenditure per industry with our computed prices to get output using

$$y_{ij}^s = \frac{p_{ij}^s y_{ij}^s}{p_{ij}^s}. \tag{20}$$

With information on prices, output, and elasticity, we calibrate country shares as

$$\gamma_{ij}^{s} = \frac{p_{ij}^{s}\left(y_{ij}^{s}\right)^{\frac{1}{\sigma_s}}}{\sum_{k=1}^{S} p_{ik}^{s}\left(y_{ik}^{s}\right)^{\frac{1}{\sigma_s}}}. \tag{21}$$

Lastly, we calibrate industry shares as

$$a_{i}^{s} = \frac{p_{i}^{s} c_{i}^{s}}{Income_{i}}. \tag{22}$$

With this last parameter, we have calibrated the entire model, save for the elasticity parameters. We explain how we do this step in the next section.

4. Estimating Best Fit Elasticities

We now discuss our methodology for estimating industry-level trade elasticities using previous liberalization episodes, the result of which we call the Best-Fit Elasticities (BFE). Export shares are defined as trade values (exclusive of tariffs) to a particular destination as a fraction of total industry output for a given source country. Given some base period t and final period t', changes in export shares are given by

$$\Delta\left(X_{ij}^{s}\right)_{t,t'} \equiv \log\left(\frac{X_{ij,t'}^{s}}{X_{j,t'}^{s}}\right) - \log\left(\frac{X_{ij,t}^{s}}{X_{j,t}^{s}}\right), \tag{23}$$

where the export shares from j to i in industry s is given by

$$X_{ij,t}^{s} = \frac{p_{ij,t}^{s} y_{ij,t}^{s}}{\tau_{ij,t}^{s}}, \tag{24}$$

and total industry output for industry s in country j is given by.

$$X_{j,t}^{s} = \sum_{i} X_{ij,t}^{s}.$$

Likewise, let

$$\Delta\left(\hat{X}_{ij}^{s}\right)_{t,t'} \equiv \log\left(\frac{\hat{X}_{ij,t'}^{s}}{\hat{X}_{j,t'}}\right) - \log\left(\frac{X_{ij,t}^{s}}{X_{j,t}}\right), \tag{25}$$

be the change in industry-level export shares in our AGE model for a given policy experiment, $\Omega_{ij,t}^{s} \to \hat{\Omega}_{ij,t'}^{s}$, where $\Omega_{ij,t}^{s}$ is the set of exogenous parameters subject to change and $i, j \in J$ are the

set of countries involved in the policy.[3] For example, if the only parameters that change between t and t' in the model are the set of bilateral industry-level tariffs then $\Omega_{ij,t}^{s} = \left\{\tau_{ij,t}^{s}\right\}_{J}$. Policy experiments can be much broader and include changes in multiple sets of exogenous parameters; such as industry-level expenditure shares or productivities

For changes in exports from country j to country i, we use as weights each industry's share of exports weighted across the base and final period,

$$\omega_{ij} = \frac{X_{ij,t}^{s} + X_{ij,t'}^{s}}{\sum_{s}\left(X_{ij,t}^{s} + X_{ij,t'}^{s}\right)}, \tag{26}$$

and compute the weighted correlation between the predicted changes in the AGE model and the changes in the data for, $\rho\left(\Delta\left(X_{ij}^{s}\right)_{t,t'} \Delta\left(\hat{X}_{ij}^{s}\right)_{t,t'} \omega_{ij}\right)$.[4] Formally, we solve the following problem to find the BFE,

$$\left\{\sigma_1,\ldots,\sigma_K\right\} = \arg\max \rho\left(\Delta\left(X_{ij}^{s}\right)_{t,t'} \Delta\left(\hat{X}_{ij}^{s}\right)_{t,t'} \omega_{ij}\right). \tag{27}$$

where we use weighted correlation as our fit metric and industries with zero production in the base or final period are excluded from the sum. Our methodology is easily adaptable to alternative fit criteria and distance metrics. Maximizing correlations is useful compared to alternative metrics for fit, like minimizing the distance between changes in trade shares in the model and in the data, in part because it can be used to evaluate elasticities when trade data is available, but production data is not. The WIOD 2013 database only contains production data through 2011, which does offer a long enough time period to evaluate many of the post-NAFTA trade liberalizations we study in our paper. Weighted correlations are a useful fit criteria when the data is noisy, but alternative fit criteria may be preferred for applications of our methodology to the calibration of parameters other than industry-level trade elasticities for AGE models.

---

[3] Because our calibration strategy fits the data exactly in the base period, the denominator, $X_{ij,t}^{s} \big/ X_{j,t}$ or base period exports deflated by exporter GDP, is the same across both $\Delta\left(X_{ij}^{s}\right)_{t,t'}$ and $\Delta\left(\hat{X}_{ij}^{s}\right)_{t,t'}$.

[4] Given a vector of weights, ω, the weighted correlation between two vectors x and y is given by $\rho\left(x,y,\omega\right) = \dfrac{cov\left(x,y,\omega\right)}{cov\left(x,x,\omega\right)cov\left(y,y,\omega\right)}$ where $cov\left(x,y,\omega\right) = \sum_{i}\omega_i\left(x_i - \overline{x}\left(\omega\right)\right)\left(y_i - \overline{y}\left(\omega\right)\right)$ where ω_i is the ith element of ω and $\overline{x}\left(\omega\right) = \sum_{i}\omega_i x_i$ is the weighted mean of x.

There is no analytical expression for the solution to (27), and the multidimensionality of the problem and presence of local optima make classical numerical approaches such as example Newton's method unsuitable. Instead, we employ a genetic algorithm to solve this problem, which is a heuristic approach to global optimization, although other global optimization heuristics, such as grid search, particle swarm optimization, and differential evolution algorithms, have the potential to be similarly effective. Within the machine learning literature, our approach is comparable to that of hyperparameter optimization. The Armington elasticities are hyper-parameters for our AGE model, meaning they do not take their values from the data but they affect the values of the other parameters and the ability of the model to match the data and generate predictions. Within the hyper-parameter optimization literature, Alibrahim and Ludwig (2021) and Liashchynskyi and Liashchynskyi (2019) compare global optimization algorithms and find genetic algorithms to outperform alternatives such as grid search in both run time and accuracy. We utilize a genetic algorithm approach in this paper and show that it performs well in our context, however, we expect that alternative global optimization heuristics might be configured to perform similarly well.

Genetic algorithms work by creating an initial population of guesses for the set of industry-level elasticities $\{\sigma_1,\ldots,\sigma_K\}$. For each guess, which is a set of industry-level elasticities, we recalibrate our AGE model to perfectly fit the base period, conduct the policy experiment in the model, $\Omega^s_{ij,t} \rightarrow \hat{\Omega}^s_{ij,t'}$, and then evaluate the performance of the model according to our fit criterion. The genetic algorithm randomly selects sets of elasticities to serve as parents for the next iteration of guesses where the probability of selection is higher for sets of elasticities that yield a better fit criterion, in our case weighted correlation with the data. After selecting these sets as parents, the algorithm generates a new population of elasticity sets by pairing parent sets and randomly swapping the elasticities for subsets of industries, which is termed crossover, as well as randomly changing the elasticities for individual industries, which is termed mutation. The process continues iteratively until the algorithm converges to the set of elasticities yielding the highest fit criteria for the given policy. In the following section we provide a sketch of the algorithm, apply our methodology to find the BFE for NAFTA, and show our elasticities outperform alternatives when these BFE are used to predict the effects of other trade liberalization episodes.

4.1 Algorithm for Computing Best-Fit Elasticities

For clarity, we present the steps of the algorithm discussed above. The approach is identical when applied to our standard AGE model where $\theta = 1$, and the extended AGE model that allows for the LTP margin from KR, with the exception that our genetic algorithm will find θ at the same time as the other industry-level elasticities, $\{\sigma_1,\ldots,\sigma_K,\theta\}$.[5]

Step 1: Start with or update a population, N, of industry-level elasticity guesses. $\{\sigma_1,\ldots,\sigma_K\}_{n\in N}$. *Steps 2–4 are done separately for each guess,* $n \in N$. *These calculations can be done in parallel to improve run time.*

Step 2: Given guess $\{\sigma_1,\ldots,\sigma_K\}_n$, calibrate the remaining model parameters according to section 3.

Step 3: Using the parameters from Step 2, conduct the policy analysis $\Omega^s_{ij,t} \to \hat{\Omega}^s_{ij,t'}$ in the model to calculate $\Delta\left(\hat{X}^s_{ij}\right)_{t,t'}$.

Step 4: Compare the model results to the changes in the data according to the Fit Criteria in (27).

Step 5: Store the set of elasticities that yield the best Fit Criteria. If the Fit Criteria, $\rho\left(\Delta\left(X^s_{ij}\right)_{t,t'} \Delta\left(\hat{X}^s_{ij}\right)_{t,t'} \omega_{ij}\right)$, falls below a tolerance, or if the BFE, $\{\sigma_1,\ldots,\sigma_K\}_n$, did not change or changed by less than a tolerance, add to the Stall Counter. If there is change in either is greater than the tolerance, reset the Stall Counter.

Step 6: Repeat Steps 1–5, where the population of guesses is updated according to the genetic algorithm as described in the appendix, until the Stall Counter hits the pre-set limit for terminating the algorithm.

[5] Our approach can easily allow for industry-level LTP amplifier terms, i.e. θ_s. We do not do so in this paper due to concerns of overfitting, as different industries already differ in their proportion of LTPs, and to be consistent with prior work.

As the genetic algorithm is a heuristic, it is not possible to guarantee the algorithm will necessarily converge to the optimal set of elasticities. The genetic algorithm can be modified for both speed and performance; for example, increasing the Stall Limit will increase run time and decrease the likelihood that the algorithm terminates at a local optimum. Our practice in this paper was to take a conservative approach where we prioritized performance over speed, and we ran the genetic algorithm multiple times and ensured it converged to the same set of BFE each time as a verification step. Although researchers may have concerns about relying on a heuristic approach to optimization, we emphasize that many applications will have natural external validity checks that can be used to gain confidence in the results. In the next section we show that our NAFTA-BFE perform well in predicting the impact of other trade liberalizations, and we find that our elasticities perform well with alternative measures of fit, both of which are strong indicators that the BFE recovered by our approach are genuine and not artifacts of our heuristic. In the appendix, we show our methodology successfully recovers elasticities when applied to simulations where the true elasticities are known.

5. Trade Liberalizations and Predictive Performance

We apply the methodology from the previous section to find the Best-Fit Elasticities for several liberalization episodes, compare them to industry-level elasticities estimated by other recent papers, and evaluate the predictive performance of these elasticities when used to evaluate the impact of other liberalization episodes.

5.1. Comparison of Elasticity Estimates

As highlighted in KPR, previous research has shown that AGE models have historically done poorly in predicting the impacts of trade liberalizations. In this section, we revisit the performance of AGE models when elasticities are calibrated using our methodology, which ensures our elasticity estimates are consistent with prior trade liberalization episodes. In particular, we apply our methodology to find the Best-Fit Elasticities (BFE) for our AGE model for NAFTA, show that these elasticities differ meaningfully from the elasticity estimates in the literature, and show that these elasticities perform well in predicting the impact of future liberalization episodes.

Following KRR, we select 1989 as our base year for NAFTA and 2005 as our final year. [6] As with most free trade agreements, only 50 percent of initial tariffs were abolished immediately when NAFTA took effect, with the rest set to phase tariffs out over 15 years (1994–2009), however most tariffs were eliminated within 10 years, and this earlier base year allows us to avoid overlapping time periods when we apply our NAFTA elasticities to later trade liberalizations. Our results are similar with alternative base and end years, both for NAFTA and for the other liberalizations. Along with applying our methodology to NAFTA, we apply it to two separate post-NAFTA liberalization, the Korea-US Free Trade Agreement (KORUS) signed in 2007, and the India and Indonesia as part of the ASEAN-India Free Trade Agreement (AIFTA, with Indonesia being the only Association of Southeast Asian Nations member in our dataset) implemented in 2010. For both KORUS and AIFTA, we include in our fit criteria other major liberalization episodes signed by the key countries. For AIFTA we include also Indonesia and Australia as part of the ASEAN-Australia-New Zealand FTA, and India and Korea's liberalization, both of which were also implemented in 2010 and featured declines in tariff rates of more than 70 percent. Similarly, when we fit the elasticities for KORUS, we include information on Korea's free trade agreements with the European Union, India, and Indonesia (ASEAN). [7]

Table 1 reports the NAFTA-BFE for our standard AGE model, and compares them to those from Caliendo and Parro (2015), which are based on cross-sectional dispersion in trade volumes and tariff rates, and those from Imbs and Mejean (2017), which are reported in the C-P column I-M columns, respectively. The elasticities for the standard AGE model are reported in column σ^S. We also consider the elasticities reported in Bas, Mayer and Thoenig (2017; henceforth BMT), who construct two sets of elasticities, one using firm-level data (BMT Micro) and another one using aggregate trade data (BMT Macro).[8] BMT and I-M utilize panel data to be consistent with

[6] Although NAFTA was signed in 1994, Canada and the United States signed a trade agreement in 1988, however, WITS does not report tariff rate data prior to 1988. Further, as we show in KR, the adoption of the Harmonized System in 1989 made product-level comparisons before and after adoption implausible. Later base years as well as alternative end years yield largely similar results.

[7] KORUS was signed in 2007 and enforced starting in 2012. Korea-EU was signed in 2007, provisionally enforced in 2011, and fully enforced in 2015.

[8] We use the estimates reported in Table 6 of BMT. In particular, we use the estimates in column 1 for firms and in column 2 for aggregate. Note that in our model the elasticity parameters govern sectoral exports. Therefore, using elasticities on the number of exporters (column 3) or average exporters (column 4) would not result in an appropriate mapping.

annual data (1995–2004 for I-M, and 2000–2006 for BMT), while C-P utilizes cross-sectional data.

Table 1: Comparison of Elasticity Estimates for Standard AGE Model

sector	BFE: KORUS	BFE: AIFTA	BFE: NAFTA	C-P	I-M	BMT Micro	BMT Macro
1	4.6	1.3	1.1	9.1	-	4.9	4.7
2	23.0	13.6	1.1	13.5	-	-	-
3	8.0	23.6	1.1	2.6	5.1	2.1	3.0
4	2.3	1.2	23.9	8.1	12.6	2.5	2.6
5	8.0	1.2	1.1	8.1	9.5	4.1	6.8
6	2.2	11.8	17.3	11.5	6.0	9.4	9.6
7	4.0	2.7	11	16.5	5.4	-5.0	-6.1
8	11.8	9.6	79	64.9	7.9	0.0	-0.9
9	15.4	19.1	69.7	3.1	6.0	5.0	5.6
10	76.0	17.8	32.7	1.7	6.5	-5.1	-12.2
11	3.4	8.6	2.5	2.4	6.2	-	-
12	8.9	26.6	10.9	5.1	7.0	3.7	2.1
13	39.5	72.3	79	1.5	9.0	3.2	2.9
14	11.3	69.2	12.5	9.6	10.5	3.3	7.3
15	31.0	55.2	56.9	1.1	11.4	7.7	10.8
16	13.1	78.7	8.1	4.0	7.2	1.4	-0.3
Average	**16.4**	**25.8**	**25.5**	**10.2**	**7.9**	**2.7**	**2.6**

A few features stand out from the elasticity estimates in Table 1. First, the BFE tend to be higher than their counterparts from the literature which are not estimated on liberalization episodes. Further, the BFE are higher for the more recent trade liberalization episodes, which we hypothesize is due to our base year for NAFTA excluding some of the early responses to FTA between Canada and the United States; and the end year being less than full implementation. If we were to use an end year of 2007, the average elasticity is 22.5, which is closer to the average estimate for the other trade liberalizations. Second, there is more consistency between the different BFE estimates than there is with those in the literature.

Table 2 shows that the average Pearson and Spearman correlation coefficients and Kendall's Tau for each elasticity estimation method with the other methods, where the average excludes itself. The average Pearson correlation coefficient is 0.40 across the BFE estimates compared to 0.01 across non-BFE estimates. The average Spearman correlation coefficient and Kendall's Tau across BFE estimates similarly are significantly higher at 0.67 and 0.48

respectively, compared to 0.04 and 0.02 for non-BFE estimates.[9] The BMT estimates have the highest correlation with the other elasticity estimates, however, as Table 1 shows the BMT elasticities are much lower than the other estimates and this remains true even when negative elasticities are ignored. Taken together, our methodology produced elasticity estimates that are relatively similar to each other even when applied to alternative and non-overlapping time periods and countries. Our elasticity estimates differ significantly from those in the literature, which also differ amongst themselves.

Table 2: Correlations between Elasticity Estimate Methods

Estimation Method	Pearson		Spearman		Kendall Tau	
	BFE	**non-BFE**	**BFE**	**non-BFE**	**BFE**	**non-BFE**
BFE:NAFTA	0.53	-0.04	0.73	-0.05	0.54	-0.05
BFE:KORUS	0.34	0.00	0.65	-0.01	0.46	-0.02
BFE:AIFTA	0.32	0.06	0.62	-0.02	0.44	0.01
Average for BFE	**0.40**	**0.01**	**0.67**	**-0.03**	**0.48**	**-0.02**
C-P	-0.37	0.16	-0.50	0.15	-0.34	0.11
I-M	0.11	0.02	0.09	0.01	0.06	-0.01
BMT Micro	0.32	-0.19	0.30	-0.15	0.19	-0.12
BMT Macro	0.15	0.06	0.03	0.13	0.02	0.08
Average for non-BFE	**0.05**	**0.01**	**-0.02**	**0.04**	**-0.02**	**0.02**

5.2. Improved Performance of Liberalization Consistent Elasticities

Given the substantially differences between our elasticity estimates for each liberalization and those from the literature, it is to be expected that AGE models may produce substantially different predictions and vary in performance. Following KRR and KPR, we judge the predictive performance of a model by computing the correlation between predicted and actual changes after the liberalization. Our focus in this paper is on evaluating external performance, whereby the elasticity estimates from one trade liberalization or estimation procedure are used to generate predictions for an entirely separate liberalization. In particular, we will use the elasticity estimates from NAFTA to generate predictions for KORUS and AIFTA. We will also evaluate performance in the other direction, where we examine how elasticity estimates from AIFTA or KORUS would have performed in predicting NAFTA.

[9] The pairwise correlations in Table 2 do not include the elasticity estimates for sectors that are missing elasticity estimates for any of the methods (Agriculture, Mining, and Petroleum). If all sectors are included, the average Pearson correlation coefficient across LCE estimates falls to 0.15, however, the average Spearman correlation coefficient and Kendall's Tau across LCE estimates remain significantly higher at 0.49 and 0.35 respectively.

For changes in exports from country j to country i, we use as weights each industry's share of exports weighted across the base and final period, and compute the weighted correlation between the predicted changes in the AGE model and the changes in the data for, $\rho\left(\Delta\left(X_{ij}^{s}\right)_{t,t'}\Delta\left(\hat{X}_{ij}^{s}\right)_{t,t'}\omega_{ij}\right)$.[10]

Table 3 presents the out-of-sample performance of our model for two trade liberalizations, KORUS (inclusive of KOR-EUC) and AIFTA. Note that, although our elasticities were chosen to maximize fit for NAFTA, they were not chosen to maximize fit for these other trade liberalizations. We benchmark our results to those achieved by using off the shelf elasticities from the literature, reported under the columns C-P, I-M, and BMT (for BMT Macro based elasticities). The BFE elasticities fit to NAFTA do well in predicting the performance of these other liberalizations, particularly when benchmarked against results from the literature which achieve low or negative correlations.

Table 3: Correlations between Elasticity Estimate Methods

importer	exporter	NAFTA-BFE	C-P	I-M	BMT
KOR	USA	0.15	0.23	-0.33	-0.26
USA	KOR	0.16	-0.69	-0.10	0.11
KOR	EUC	0.41	-0.28	0.52	0.20
EUC	KOR	0.21	-0.46	0.20	0.20
Average: KORUS		**0.26**	**-0.30**	**0.07**	**0.06**
IDN	AUS	0.73	-0.29	-0.11	-0.31
AUS	IDN	-0.07	-0.42	-0.47	-0.21
IND	IDN	0.40	-0.09	-0.04	-0.03
IDN	IND	0.25	0.02	-0.14	-0.17
JPN	IDN	0.27	0.23	0.22	0.21
JPN	IDN	0.40	-0.06	0.15	-0.06
CHN	IDN	0.24	0.91	0.81	0.97
IDN	CHN	0.13	-0.25	-0.13	0.07
Average: AIFTA		**0.29**	**0.01**	**0.04**	**0.06**

[10] Given a vector of weights, ω, the weighted correlation between two vectors x and y is given by $\rho(x,y,\omega)=\frac{cov(x,y,\omega)}{cov(x,x,\omega)cov(y,y,\omega)}$ where $cov(x,y,\omega)=\sum_i \omega_i\left(x_i-\overline{x}(\omega)\right)\left(y_i-\overline{y}(\omega)\right)$ where ω_i is the ith element of ω and $\overline{x}(\omega)=\sum_i \omega_i x_i$ is the weighted mean of x.

Table 4 reports the results for the reverse exercise, where elasticities picked to be consistent with AIFTA and KORUS are shown to perform well for NAFTA. Table 4 shows that the model struggles to capture Mexican imports (average correlation of 0.02), relative to trade flows (average correlation of 0.56). One reason for this shortcoming may be the absence of data on applied tariffs from Mexico data prior to NAFTA.

Table 4: Alternatives sources for BFE

importer	exporter	AIFTA-BFE	KORUS-BFE
MEX	CAN	0.01	0.34
USA	CAN	0.46	-0.09
CAN	MEX	0.77	0.54
USA	MEX	0.59	0.07
CAN	USA	0.43	0.72
MEX	USA	0.03	-0.19
Average: NAFTA		**0.38**	**0.23**

Table 5 reports the weighted correlations between data and the model predictions for each set of trade elasticities, combining both the NAFTA liberalizations and the USA-KOR (not inclusive of KOR-EUC), and MEX-JPN liberalizations. Column 1 reports the C-P correlations, column 2 reports the correlations using the I-M elasticities, which is the best performing of the off-the-shelf elasticities with an average weighted correlation of 0.23, with three of the ten cases correlating negatively, and with the two correlations involving Mexican exports to NAFTA countries going as high as 0.84 and 0.86. Overall, while on average the correlations give a better fit than other sets, the performance is significantly lower than the performance of the Best-Fit Elasticities as reported in Tables 3 and 4. The last two columns use the estimates from BMT, first the one for firms, then the one for aggregate, which have average correlations of 0.12 and 0.08, respectively.

An important element regarding the predictions is qualitative: what percentage of predictions positively correlate with the data. For the case of C-P elasticities, we find that only 40% of the predictions correlate positively. I-M elasticities bring this number up to 70%, while BMT's are 60% (firms) and 50% (aggregate). These numbers contrast with the amount of cases that the model with model-consistent elasticities generate a positive correlation: 100% of the times (both the model with LTP and the model without LTP).

Table 5: Using external elasticities

importer	exporter	C-P	I-M	BMT firms	BMT aggregate
CAN	MEX	−0.46	0.86	0.62	0.59
CAN	USA	0.36	-0.27	-0.15	-0.13
MEX	CAN	−0.68	0.38	0.39	0.40
MEX	USA	−0.17	0.41	0.31	0.31
USA	CAN	0.35	0.05	0.08	0.07
USA	MEX	0.54	0.84	0.72	0.72
KOR	USA	-0.31	-0.26	-0.24	-0.31
USA	KOR	-0.32	0.19	-0.20	-0.35
JPN	MEX	-0.42	0.16	-0.34	-0.51
MEX	JPN	-0.02	-0.02	0.00	-0.02
simple average		−0.11	0.23	0.12	0.08
simple average [no NAFTA]		−0.27	0.02	−0.20	−0.30

5.3. Imperfect Insights: The Utility of Flawed Model Assessments

In a recent paper, Adão, Costinot, and Donaldson (2023), henceforth ACD, develop a novel method for evaluating the predictions of quantitative trade models based on constructing an instrumental variable from exogenous policy shifters. They formalize a statistical hypothesis test based on this IV that can be used to evaluate the null hypothesis that a given model's prediction are correct. Their approach has many strengths, in particular, that it can be deployed even in the presence of large shocks outside of the policy of interest. Due to the absence of this strength, ACD are critical of approaches, such as the one in this paper, that rely upon assessing models based on comparing the model predictions directly against changes in the data.

To be clear, we do not disagree with the central point of their criticism, indeed we acknowledged as such in Kehoe et al. (2017), writing "*While there are complications in evaluating models this way, particularly because many changes occur postreform that were not considered in the model, it can still be informative*". Certainly, we agree it is possible to construct a set of exogenous shocks such that the correlation coefficient and similar measures may inaccurately summarize the underlying accuracy of a model, both for the better or for worse. We believe this concern is minimized in our application due to several factors. First, the results of our assessment methodology have yielded consistent results across a range of different trade liberalizations, even after accounting for other major non-liberalization shocks, such as changes in tariffs between non-

liberalization countries and changes in industry-level productivities as shown in appendix A.3. ACD use simulations to show how the correlation of model predictions and changes in the data mechanically goes to zero as the variance of exogenous shocks not accounted for by the model goes to infinity, but they do not show what happens if these shocks are inputted into the model, as we do in this paper. Second, here as in earlier work (Kehoe et al. (2015), Kehoe et al. (2017)), we've provided an alternative methodology that we assess alongside outperforms the predictions of the standard model, which would not be possible if non-liberalization shocks were mechanically driving correlation coefficients to zero. Third, and most importantly, we provide economic interpretations for why we believe the models fall short in their performance; and we show that addressing these concerns improves model performance as expected.

Overall, we laude ACD as a step forward in the assessment of trade models. We have two minor cautions with their approach. First, the construction of the IV requires the existence of shocks that are orthogonal to the policy experiment of interest, and such shocks may be difficult to identify in practice. In simulations, ACD are able to show the IV-test works well in a hypothetical environment in which tariffs are drawn from an independent and identically distributed distribution; however, the assumption that tariffs are randomly generated is at odds with the literature on endogenous tariffs and goes against the popular understanding of how countries might respond to a trade war. Just as it is possible to imagine a set of exogenous shocks such that correlation coefficients are misleading, it is similarly possible to imagine a set of shocks such that an IV-based approach might be misleading. We expect this to be the case with any approach, so the caution here is to remind researchers that just as there is no universally best model of international trade, there cannot be a universally best method of model assessment and there will be times in which any method, no matter how sophisticated, falls short.

Second, although ACD marks a step forward in the sophistication of the assessment of international trade models, it offers more limited insight into how particular models might go astray or be improved. In the application of their methodology to the predictions of the model from Fajgelbaum et al. (2020), they find that some predictions are rejected by their IV-based test, and others are not. ACD largely leave it at that, and it's not clear their methodology offers much insight about what researchers should do, if anything, about these conflicting results. In contrast, in Kehoe et al. (2015) and Kehoe et al. (2017) we used our assessments to directly offer economic insights into why we believed the models were falling short and how they might be improved upon,

which we have attempted to deliver upon in this paper. In doing so, we were aided by the ease of applicability of our metric of model assessment, and in particular by the fact it could be equally applied to quantitative trade models and reduce form methods. This is not to say ACD cannot do similar comparisons with their approach, but it's less obvious how their approach can be used to directly compare two competing frameworks and offer insights form this comparison. Therefore, while there are clear limitations to the simplicity of our approach, we argue there are also notable strengths.

Overall, ACD correctly point out some of the limitations of an approach based on simple metrics of fit between data and model predictions, and offer a compelling alternative if a researcher is concerned that the presence of exogenous shocks might limit the reasonability of directly comparing a model to the data and these shocks cannot easily be inputted into the model. Given the simplicity and ease of implementing our approach, we see the two approaches as complimentary, rather than competing. Likewise, we anticipate there is room for researchers to build upon their insights to advance upon our methodology and develop alternative metrics of model fit that might be useful for calibrating elasticities from past liberalization episodes.[11]

6. Final remarks

In this paper we demonstrate that Applied General Equilibrium models can be used to make accurate predictions regarding the impact of a trade reform. The key insight from our paper is that using elasticities consistent with previous liberalization episodes allows the model to generate changes in the industry composition that positively correlate with the changes that actually occur after a trade reform. Our methodology can easily be applied to additional parameters and margins that we expect will further improve the performance of AGE modeling.

[11] For example, ACD's IV-based fit criteria might be adaptable as a metric of model fit that could be employed with our methodology. We do not attempt to do so in this paper, since the statistical properties behind their test relies on the law of large numbers as the number of products increases towards infinity. ACD focus on an application with 10228 products with a single common trade elasticity, whereas in our model we have only 16 industries, each with its own elasticity. If we were to separately model each product in an industry, rather than aggregating them together, we could more easily apply ACD's test; however, as our genetic algorithm requires solving our model a large number of times, increasing the size of our model by such a significant magnitude is not computational feasible without adaptations elsewhere in our implementation.

References

Alessandria, G., J. P. Kaboski and V. Midrigan (2010): "Inventories, Lumpy Trade, and Large Devaluations," *American Economic Review*, 100(5), 2304–39.

Alibrahim, H. and S. A. Ludwig (2021): "Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization," 2021 IEEE Congress Evolutionary Computation (CEC), 1551–59.

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

Arkolakis, C. (2016): "A Unified Theory of Firm Selection and Growth," *Quarterly Journal of Economics*, 131, 89–155.

Bas, M., T. Mayer and M. Thoenig (2017): "From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity," *Journal of International Economics*, 108, 1–19.

Brown, D. K., A. V. Deardorff and R. M. Stern (1995): "Estimates of a North American Free Trade Agreement," in P. J. Kehoe and T. J. Kehoe (Eds.), *Modeling North American Economic Integration*. Kluwer Academic Publishers, 59–74.

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

Feodora, T. (2023): "Missing Tariffs, False Imputation, and the Trade Elasticity," working paper.

Hertel, T. (2013): "Global Applied General Equilibrium Analysis Using the Global Trade Analysis Project Framework," *Handbook of Computable General Equilibrium Modeling*, vol. 1, ed. P. B. Dixon, D. Jorgenson, 815–76. Amsterdam: Elsevier.

Imbs, J. and I. Mejean (2017): "Trade Elasticities," *Review of International Economics*, 25(2), 383–402.

Kehoe, T. J. (2005): "An Evaluation of the Performance of Applied General Equilibrium Models on the Impact of NAFTA." In Frontiers in Applied General Equilibrium Modeling, ed. T.

J. Kehoe, T. N. Srinivasan, J. Whalley, pp. 341–77. Cambridge, UK: Cambridge Univ. Press.

Kehoe, T. J., C. Polo and F. Sancho (1995): "An Evaluation of the Performance of an Applied General Equilibrium Model of the Spanish Economy," *Economic Theory*, 6, 115–41.

Kehoe, T. J., P. S. Pujolas and J. Rossbach (2017): "Quantitative Trade Models: Developments and Challenges," *Annual Review of Economics*, 9(1), 295–325.

Kehoe, T. J., J. Rossbach and K. J. Ruhl (2015): "Using the New Products Margin to Predict the Industry-level Impact of Trade Reform," *Journal of International Economics*, 96, 289–97.

Kehoe, T. J. and K. J. Ruhl (2013): "How Important is the New Goods Margin in International Trade?" *Journal of Political Economy*, 121, 358–92.

Liashchynskyi, P. and P. Liashchynskyi (2019): "Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS," working paper.

Romalis, J. (2007): "NAFTA's and CUSFTA's Impact on International Trade," *Review of Economics and Statistics*, 89, 416–35.

Rubini, L. (2014): "Innovation and the Trade Elasticity," *Journal of Monetary Economics*, 66, 32–46.

Ruhl, K.J. (2008): "The International Elasticity Puzzle," working paper.

Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer and G. J De Vries (2015): "An Illustrated User Guide to the World Input-Output Database: the Case of Global Automotive Production," *Review of International Economics*, 23, 575–605.

A.1. Description of Data Sources

We rely on the 2013 revision of the World-Input Output Database (WIOD) described by Timmer et al. (2015) as our basis for the regions and industries included in our analysis. Our analysis focuses on reductions in tariff rates due to trade liberalizations, therefore we combine the European Union Customs Union countries, including the United Kingdom and Turkey[12], into a single aggregated region. This has the benefit of reducing the number of equations in our model, and thereby reducing code run time. We also combine Taiwan into the Rest of World aggregate, as the United Nations does not report product level trade flows for Taiwan, and rather attributes these trade flows to the region code for Other Asia, not elsewhere specified. Table A.1.1 details the 13 regions remaining for our analysis. Throughout the paper we often refer to countries rather than regions, however, there is no distinction between the two terms in our model and analysis.

Table A.1.1: Region Aggregates and Productivity Sources

Region Code	Region Name	INDSTAT Availability	Agg+Mining Data Source	Agg+Mining Availability
AUS	Australia	1988–2020	OECD	1988–2017
BRA	Brazil	1990–2020	OECD	2000–2020
CAN	Canada	1988–2020	OECD	1997–2020
CHN	China	1988–2020	REITI CIP 4.0	1988–2017
IDN	Indonesia	1988–2020	OECD	2011–2017
IND	India	1988–2020	RBI India Klems	1988–2020
JPN	Japan	1988–2020	OECD	1994–2020
KOR	Korea	1988–2020	OECD	2004–2020
MEX	Mexico	1988–2020	OECD	2005–2020
RUS	Russia	1993–2020	HSE Russia Klems	1995–2016
USA	United States of America	1988–2020	OECD	2000–2020
EUC	European Union Customs Union	1988–2020	OECD	1995–2019
ROW	Rest of World	1988–2020	Imputed	N/A

The WIOD comes with a Socio Economic Accounts supplement, which provides information on employment and value added per worker for each industry and region from 1995-2009. This time frame is acceptable when analyzing NAFTA, however, the timeframe is not suitable for accounting

[12] The UK exited the EU in 2020, which falls outside of the time period we analyze. Turkey and the European Union entered into a customs union agreement in 1995, meaning that there are no customs applied to goods transported between the two regions and equivalent tariff rates are applied by both the EU and Turkey to external trading partners.

for productivity changes that occurred during later trade liberalization episodes. We use the UNIDO Indstat 2-digit ISIC Rev 3 Database for information on value added per worker and employment for each industry. The dataset reports value added according to four different measures — producer prices, factor prices, basic prices, and other valuation — however, each country reports information according to only one measure. Labor per industry is reported as either the number of full-time employees or the number of person's engaged. Indstat contains only information on manufacturing sectors, therefore, we supplement this data with the OECD Productivity by Industry Database (PDBI), which reports data on value added per worker and employment for OECD countries to get data on the agriculture and mining industries.

We do not rely on the OECD PDBI database as our primary source of data both due to not lacking data for several important non-OECD countries, most notably China, and also the data is reported for ISIC Revision 4 industry codes, and there is not a 1-to-1 correspondence between ISCI Rev 4 industry codes and ISIC Rev 3 industry codes in general across manufacturing industries. For value added per worker and employment for the agriculture and mining industries for India, Russia, and China, we rely on country-specific KLEMS databases. The data source and annual availability for each country's value added per worker and employment data is reported in Table A.1.1. When a source reports value added in national currency in current prices, we convert to USD using the period average annual exchange rate and then deflate by the US GDP Deflator in each year, both from the World Bank's World Development Indicators (WDI) Database.

When a source reports value added in national currency in constant prices, we convert to USD using the base year period average annual exchange rate. Because we want to have data in each year, we sequentially fill in missing data for country-industry pairs with missing data after 1995 with the previous non-missing value. If there is no previous data available, or if the year is before 1995, we sequentially fill in missing data with the closest future year of data. Afterwards we smooth the value added per worker and employment numbers using a 3-year moving average, and then index them for each country-industry pair so that there is a base value of 1 in 1995. Productivity numbers are not available for the agricultural and mining sectors for the Rest of World (ROW) aggregate, therefore we impute them by predicting them from a linear regression with year fixed effects between ROW's value added per worker index and the median value added per

worker index for manufacturing industries. The imputed value added per worker index for the ROW ranges from 0.456 in 2006 for mining to 1.986 in 2020 for agriculture.

Table A.1.2: WIOD Industry Aggregates

WIOD Industry Code	Industry Description	2-digit ISIC Rev. 3 Codes
1	Agriculture, Hunting, Forestry and Fishing	1,2,5
2	Mining and Quarrying	10-14
3	Food, Beverages and Tobacco	15,16
4	Textiles and Textile Products	17,18
5	Leather, Leather and Footwear	19
6	Wood and Products of Wood and Cork	20
7	Pulp, Paper, Paper , Printing and Publishing	21,22
8	Coke, Refined Petroleum and Nuclear Fuel	23
9	Chemicals and Chemical Products	24
10	Rubber and Plastics	25
11	Other Non-Metallic Mineral	26
12	Basic Metals and Fabricated Metal	27,28
13	Machinery, Nec	29
14	Electrical and Optical Equipment	30-33
15	Transport Equipment	34,35
16	Manufacturing, Nec; Recycling	36,37

We use product-level information trade flows reported at the 5-digit Standard International Trade Classification (SITC) Rev 2 level from the World Bank's World Integrated Trade Solution (WITS) database, which distributes data on tariffs from the United Nations' Trade Analysis Information Systems (TRAINS) databases and data on trade flows from the United Nations' Commodity Trade (UN Comtrade) database. WITS provides native support for the aggregation of countries into regions and products into groups and industries, therefore we use reported product-level trade flows for the European Union Customs Union and Rest of World aggregates reported directly from WITS. We compute the share of the share of least traded products in each industry using the concordance provided by WITS to assign each 5-digit SITC Rev.2 product code to a 2-digit ISIC Rev.3 industry codes using the provided concordance, and then to the WIOD industries as reported in Table A.1.2. We also collect trade flows and tariffs reported for each industry aggregate and region from WITS, to extend the industry trade flows in WIOD beyond the original

1995-2011 timeframe. After dropping industries with zero reported trade flows in either database, the correlation between trade flows normalized by exporter GDP in the WIOD database and Comtrade is 0.9625, and a linear regression between the two datasets yields a slope coefficient of 1.01.

We identify trade liberalization episodes from the Centre d'Etudes Prospectives et d'Informations Internationales' (CEPII) Gravity Database, which provides data from the World Trade Organization on regional trade agreements and can be used to identify the date of implementation. We select as trade liberalization episodes, all trade agreements between two or more regions in in Table A.1.1 that were signed between 1989 and 2012 and experienced at least a 75 percent decline in the average across WIOD industries of effectively applied weighted average industry tariffs. We do not consider liberalization episodes after 2012 due to wanting at least 7 years post-implementation for analysis, based on the results from Kehoe and Ruhl showing the LTP margin takes around 7 years to level out, as more recent data becomes available, it would be reasonable to extend our analysis to cover more recent liberalization episodes. The liberalization episodes identified according to these criteria are listed in Table A.1.3, where before refers to the average tariff rate 3 years prior to liberalization for non-NAFTA liberalizations and in 1989 for NAFTA, and after refers to 10 years following liberalization.

Table A.1.3: Liberalization Episodes with >75% Reduction in Tariffs

FTA Countries	FTA Year	Avg Tariff Before	Avg Tariff After
CAN-MEX-USA	1994	6.1	0.2
EUC-MEX	2001	9.7	2.6
AUS-USA	2005	3.2	0.5
CHN-IDN	2005	6.2	1.2
IDN-JPN	2008	4.0	0.9
AUS-IDN	2010	3.0	0.3
IDN-IND	2010	25.1	3.2
IND-KOR	2010	9.7	2.4
EUC-KOR	2012	5.7	0.5
KOR-USA	2012	15.1	1.2

A.2. An AGE Model featuring the Least Traded Products Margin

In this section we adapt our multi-sector, multi-country, applied general equilibrium model and nest within it the ability to capture the least traded products (LTP) extensive margin from KR. The

LTP margin is the observation that products that were previously traded in small, but positive, amounts, account for a disproportionate share of the total trade growth following a trade liberalization.[13] As noted in KRR and KPR, a simple reduced form approach based solely off of the share of least traded products in each industry performs better in predicting industry-level changes in trade flows for a wide array of trade liberalizations compared to both historically used and newly developed AGE models. The model we develop below allows us to account for the LTP margin explicitly while also allowing us to turn off this margin so that we can isolate how much of our increased model performance is due to our method for calibrating elasticities, which can be applied to a wide range of AGE models, and how much is due to the accounting for the LTP margin in the model. For simplicity of notation, we suppress time subscripts until we discuss using our AGE model to conduct policy experiments in section 4.

We start with the same standard AGE model described in section 3, where equations (1)–(6) are unchanged. To include the LTP margin, we assume the exports from j to i in industry s are themselves an Armington aggregate of non-LTPs and LTPs. This specification captures that the LTP margin is bilaterally defined at the industry-level. The elasticity across these two types of goods is the same as the elasticity across countries. Namely, the aggregate is given by

$$y_{ij}^{s} = \left(\left(1-\nu_{ij}^{s}\right)\left(y_{ij}^{s,non}\right)^{\frac{\sigma_s-1}{\sigma_s}} + \nu_{ij}^{s}\left(y_{ij}^{s,LTP}\right)^{\frac{\sigma_s-1}{\sigma_s}} \right)^{\frac{\sigma_s}{\sigma_s-1}}, \tag{28}$$

where ν_{ij}^{s} is the non-LTP weight in country i's industry s goods from country j. The producer of this bundle cost-minimizes, which again implies the demand functions

$$y_{ij}^{s,non} = y_{ij}^{s}\left(\left(1-\nu_{ij}^{s}\right)\frac{p_{ij}^{s}}{p_{ij}^{s,non}} \right)^{\sigma_s}, \tag{29}$$

and

$$y_{ij}^{s,LTP} = y_{ij}^{s}\left(\nu_{ij}^{s}\frac{p_{ij}^{s}}{p_{ij}^{s,LTP}} \right)^{\sigma_s}, \tag{30}$$

where

[13] Formally, KR define LTPs as the set of products with the least amount of trade such that they collectively account for 10 percent of pre-liberalization bilateral trade. They show, for example, that this set of products grew from 10 percent of exports pre-NAFTA to 30 percent of exports from Canada to Mexico following the implementation of NAFTA.

$$p_{ij}^{s}=\left(\left(1-\nu_{ij}^{s}\right)^{\sigma_s}\left(p_{ij}^{s,non}\right)^{1-\sigma_s}+\left(\nu_{ij}^{s}\right)^{\sigma_s}\left(p_{ij}^{s,LTP}\right)^{1-\sigma_s}\right)^{\frac{1}{1-\sigma_s}}. \tag{31}$$

Producers from country j of product $m=\{non,LTP\}$ (for non-LTPs and LTPs, respectively) operate the following production function with fixed proportions

$$y_{ij}^{s,m}=\frac{1}{t_{ij}^{s,m}}\min\left\{z_j^s\ell_{ij}^{s,m},\frac{x_{ij}^{s_1,m}}{\alpha_j^{s_1}},...,\frac{x_{ij}^{s_k,m}}{\alpha_j^{s_k}},...,\frac{x_{ij}^{s_S,m}}{\alpha_j^{s_S}}\right\}, \tag{32}$$

where $x_{ij}^{s_k,m}$ is the quantity of composite intermediate inputs from industry k used in the production of industry s in country j for consumption in country i, and $\alpha_j^{s_k}$ is the unit input requirement of inputs from industry k used in the production of industry s. Note that the unit input requirements does not depend on the destination, only on where the output is produced and in which industry.

The last object, $t_{ij}^{s,m}$ is the element we use to capture the increased growth of LTPs following trade liberalization. In particular, $t_{ij}^{s,m}$ is an additional iceberg cost that is equal to 1 if $m=non$, indicating no additional iceberg trade cost, and it is greater than 1 if $m=LTP$. In particular, we assume that

$$t_{ij}^{s,\mathrm{LTP}}=\left(\tau_{ij}^{s}\right)^{\theta-1}, \tag{33}$$

where $\tau_{ij}^{s,m}$ is the tariff for industry s in country i on imports from country j and θ is an amplifier that captures the degree to which LTP are more responsive to reductions in tariffs than non-LTP. The basic idea behind this trade cost is that we want our AGE model to capture the fact that LTPs are more responsive than non-LTPs for a given reduction in tariffs following a trade liberalization. Instead of giving these products different elasticities explicitly, we instead assume that they have the same elasticity, but that LTPs face amplified pre-liberalization trade costs, which lead LTPs to grow disproportionately following liberalization once these trade costs are eliminated. This choice of specification has two primary strengths. First, it allows us to capture the observation from KRR that LTPs do not grow more than LTPs in the absence of trade liberalization, which is consistent with there being additional trade costs that are removed during liberalization, but is not consistent with LTPs being more elastic in general. Second, it allows our model with LTPs to nest the

standard AGE model without the LTP margin by setting $\nu_{ij}^{s} = 0$ for some or all sectors. We expect that including a more rigorous mechanism for the growth in LTPs, for example based on Arkolakis (2016) or Rubini (2014), may yield further improved performance for AGE models.

The remainder of the model is unchanged compared to the standard AGE model, with the exception that instead of each equation being specified for only one product in each industry, it is specified separately for LTPs and non-LTPS. The Leontieff structure of the production function implies that

$$z_j^s \ell_{ij}^{s,m} = \frac{x_{ij}^{s_1,m}}{\alpha_j^{s_1}} = ... = \frac{x_{ij}^{s_k,m}}{\alpha_j^{s_k}} = ... = \frac{x_{ij}^{s_S,m}}{\alpha_j^{s_S}}, \tag{34}$$

which means that the demand for intermediate goods is given by

$$x_{ij}^{s_k,m} = \alpha_j^{s_k} z_j^s \ell_{ij}^{s,m}. \tag{35}$$

Consumers in country i pay tariffs equal to τ_{ij}^{s}. Therefore, the firm's profit maximization problem is given by

$$\max \frac{p_{ij}^{s,m} y_{ij}^{s,m}}{\tau_{ij}^{s}} - w_j \ell_{ij}^{s,m} - \sum\nolimits_{k=1}^{S} p_j^k x_{ij}^{s_k,m} \ , \tag{36}$$

implying equilibrium prices

$$p_{ij}^{s,m} = \tau_{ij}^{s} t_{ij}^{s,m} \left(\frac{w_j}{z_{ij}^{s}} + \sum\nolimits_{k=1}^{S} p_j^k \alpha_j^{s_k} \right). \tag{37}$$

In turn, LTP and non-LTP prices are related according to $p_{ij}^{s,LTP} = t_{ij}^{s,LTP} p_{ij}^{s,non}$.

Tariff revenue in country i is collected by the government and rebated back to consumers,

$$T_i = \sum\nolimits_{s=1}^{S} \sum\nolimits_{j=1}^{J} \left(\tau_{ij}^{s} - 1\right)\left(p_{ij}^{s,LTP} y_{ij}^{s,LTP} + p_{ij}^{s,non} y_{ij}^{s,non} \right) . \tag{38}$$

We also allow for aggregate trade deficits, where the trade deficit for country i is given by,

$$D_i = \sum\nolimits_{s=1}^{S} \left(\sum\nolimits_{j=1}^{J} \frac{p_{ij}^{s} y_{ij}^{s}}{\tau_{ij}^{s}} - \sum\nolimits_{j=1}^{J} \frac{p_{ji}^{s} y_{ji}^{s}}{\tau_{ji}^{s}} \right). \tag{39}$$

Note that the aggregate trade deficit still allows for bilateral trade deficits to be endogenously determined. As an alternative to including aggregate trade deficits, we can also rebalance each countries trade with ROW so that trade is balanced on aggregate. This alternative specification

has no impact on the results in our paper or the performance of AGE models for the liberalizations we study.

The final part of the AGE model are the market clearing conditions. The labor market clearing condition for country j is equal to the sum of labor used in production of LTPs and non-LTPs aggregated across all destination markets and industries,

$$L_j = \sum_{s=1}^{S} \sum_{i=1}^{J} \left(\ell_{ij}^{s,non} + \ell_{ij}^{s,LTP} \right). \tag{40}$$

Similarly, the market condition for intermediate inputs is similarly given by

$$x_j^k = \sum_{s=1}^{S} \sum_{i=1}^{J} \left(x_{ij}^{s_k,non} + x_{ij}^{s_k,LTP} \right), \tag{41}$$

where x_j^k is the total amount of final output from industry k that is used as intermediate inputs in country j, and the market clearing condition for final output in country j remains unchanged.

When calibrating the model, we use product level trade flows to calculate the share of LTP in each industry for each bilateral trade flow according to KRR. For a given base year, we sort products by trade flows over the first three years and then define as LTPs the set of products with the smallest trade flows such that they cumulatively account for 10 percent of trade in the base period. [14] [15] By definition, LTPs account for 10 percent of trade in the base period for a given bilateral trade flow, however, in each industry the share of LTPs may range from between 0 percent and 100 percent.

A.3. Robustness: Productivity Shocks and Extended AGE Model with LTP Margin

The analysis in the main paper shows that our benchmark model (and also the one without LTP) are already able to make sectoral predictions that correlate well with the observed changes. Kehoe et al. (1995) show that the performance of the AGE for Spain can improve further when they introduce shocks that were unknown at the time of the prediction, specifically, a draught in the agricultural industry and changing oil prices.

[14] This prevents products that are traded in large amounts, but irregularly, from being classified as least traded products. See Alessandria et al. (2010) for a discussion of trade lumpiness.

[15] KRR note that the exact cutoff does not matter as long as it is not so small as to exclude a significant portion of the extensive margin or so large that it fully dilutes it. In particular, their results are robust to cutoffs of 5 percent or 20 percent instead of 10 percent.

The WIOD includes information on industry-level output, value added, and employment for the period 1995-2009, which we use to compute industry-level productivity measures. The primary difficulty, is that the liberalizations we study exceed this time range. We extend these series using data from United Nation's UNIDO Industrial Statistics (INDSTAT) Database, which contains information on employment, output, and value added for 3-digit ISIC Rev 3 manufacturing industry codes, which we supplement with OECD and country-specific KLEMS data — referring to data on capital (K), labor (L), energy (E), materials (M), and services (S) — for employment and productivity information for the agricultural and mining industries. This data, along with the list of industry and country aggregates, is described in more detail in Appendix A.1. This additional information allows us to account for shocks that occurred post liberalization when calibrating tariffs and when evaluating model performance.

In Appendix A.2, we developed an extended version of our AGE model that incorporates the least traded products margin from KR, and show how both the standard version of our AGE model and the extended model perform when we incorporate shocks beyond just changes in tariffs between liberalizing countries. In particular, we include changes in industry-level productivity, changes in industry final consumption shares, and changes in tariffs between non-liberalizing countries.

Table A.3.1 reports the correlation between model and data for our Standard AGE model and our extended AGE model that incorporates an explicit LTP margin where both the industry-level elasticities and the LTP amplifier, θ, are estimated to fit observed changes during NAFTA. Columns 1 and 3 report the correlations of the model with and without the LTP margin when only tariffs between liberalizing countries change. Columns 2 and 4 report the correlations when these additional shocks are included, the most impactful of which is industry-level productivity changes.

Table A.3.1: Performance of BFE: NAFTA Elasticities for other FTAs

FTA	Standard AGE	LTP AGE	Standard + Info	LTP + Info
AUSUS	0.02	0.13	0.22	0.02
INDJPN	-0.11	0.31	0.31	0.40
MEXEU	0.69	0.51	-0.02	0.64
KORUS	0.26	0.17	0.15	0.15
AIFTA	0.59	0.55	0.37	0.57
Simple Average	0.29	0.33	0.21	0.35

We can also conduct the same exercise with elasticities and an LTP amplifier estimated to be consistent with AIFTA. Table 2.2. reports these elasticities.

Table A.3.2: Performance of BFE: AIFTA Elasticities for other FTAs

FTA	Standard AGE	LTP AGE	Standard + Info	LTP + Info
AUSUS	0.36	0.23	-0.27	-0.26
INDJPN	0.23	0.46	0.62	0.35
MEXEU	0.47	0.30	0.07	0.70
NAFTA	0.38	0.20	0.37	0.57
Simple Average	0.29	0.30	0.21	0.35

For both AIFTA and NAFTA, the model with an LTP margin performs better than the standard model, reinforcing the importance of the LTP margin from KR, and showing how our methodology can be easily adapted to calibrate model parameters outside of simply elasticities. This increase in performance is not uniform. For example, the standard model does the best at predicting the impact of the free trade agreement between Mexico and the European Union. More surprisingly, including productivity shocks sometimes leads to decreases in model performance; particularly for the standard model. We believe this is due to a mismatch between the concept of productivity in the model and productivity in the data. If a sector shrinks due to firms with lower labor productivities shutting down, this will mechanically cause an increase in labor productivity in the data. If we insert this mechanical increase in labor productivity from the data into our AGE model, our model predicts the country should allocate more labor towards this more productive sector; which goes in the exact opposite direction of the model. We anticipate that models that feature a robust extensive margin will more uniformly benefit from the inclusion of productivity data; however, our results show that our elasticity estimates still perform well in capturing observed changes when additional shocks are included, even with this caveat.

A.4. Updating Parameter Guesses with Genetic Algorithm

In this section we describe how the genetic algorithm updates the population of guesses for the set of trade elasticities in the algorithm described in section 4.1.[16] Genetic algorithms work through a combination of exploration, in which the algorithm searches over a diverse set of possible

[16] In this paper we employ and adapt the genetic algorithm implementation from the MATLAB Global Optimization Toolkit, however, other implementations, such as PyGAD for Python and GA for R, work similarly. The initialization, mutation, and adaptive restart features we discuss are not available natively in Matlab's implementation, but can be implemented through user written functions and code provided in our data appendix.

parameter combinations to avoid local optima, and refinement, in which the algorithm retains and searches near the best performing prior guesses. In our application of the genetic algorithm we constrain the elasticity estimates to be between 1.1 and 30, which envelops the range of trade elasticity estimates found in Imbs and Mejean (2017), and in the extended version of our model we constrain θ to be between 1 and 20. We also discretize the problem implement an integer constraint for each parameter value multiplied by 100, that is, we constrain our search over trade elasticity estimates to two decimal places. In essence, we create a grid of elasticity values, but instead of implementing a standard grid search approach, we search over the grid of possible elasticity combinations using a genetic algorithm.

Each iteration of the genetic algorithm corresponds to a given generation. The algorithm initializes the first generation by randomly generating $N = 200$ [17] sets of trade elasticities from a uniform distribution. Independently for each set of trade elasticities, we calibrate the remaining AGE model parameters, simulate NAFTA or an alternative past trade liberalization of choice in the calibrated AGE model, and calculate our measure of model fit by comparing the changes in trade shares in our model to the changes in the data. To generate the next generation of guesses, we rank each set of trade elasticities by its fit value. The top ten $(0.05N)$ ranked sets of elasticity estimates are termed elites and included in the next generation without modification. The remaining 190 sets of elasticities in the next are called children and are generated from parents in the current generation, where elites are included as parent candidates, and then modified according to either crossover or mutation. We implement nonlinear rank scaling, where parents for the next generation are always selected randomly with probability proportional to $1/\sqrt{rank}$.

Crossover accounts for 80 percent of the children, and involves first selecting two parents. We implement scattered crossover, which means that each trade elasticity is independently randomly selected from one of the two parents with equal probability. Let $\{\sigma_s\}_{n'}$ represent the trade elasticity estimate for sector s for the guess n' in a given generation. If n' is a child generated by crossover via parents m and n, then $\{\sigma_s\}_{n'}$ takes value $\{\sigma_s\}_m$ with probability 0.5 and takes value $\{\sigma_s\}_n$ with probability 0.5.

Mutation accounts for the remaining 20 percent of children. Mutation involves selecting one parent from a child, and then randomly modifying each component. We employ an adaptive power mutation, which involves two stages. The first stage determines whether the elasticity for a given sector will increase or decrease by comparing a draw from random uniform distribution to the scaled distance between the elasticity and the upper and lower bounds. For example for a child n' of mutation from parent n, if $\{\sigma_s\}_n = 4$, then the scaled distance is $0.1003 = (4-1.1)/(30-1.1)$, which means if the uniform random variable is less than 0.1003 the algorithm will choose a value less than 4 for $\{\sigma_s\}_{n'}$ and greater than 4 otherwise. In the second stage, a random uniform distribution raised to a power (4) is used to determine how far along the distance between the value and the bound the child guess will be, and then the final value is rounded to conform to our integer

[17].

constraints. If the second random variable is equal to 0.5, then the child estimate will be $\{\sigma_s\}_{n'} = 3.82 \approx 4 - 0.5^4(4 - 1.1)$.

Each of the different stages of a genetic algorithm can be modified from the methodology presented here. In fact, the genetic algorithm could be considered to contain its own set of hyperparameters, such as population size, that affect the genetic algorithm's performance and run time. Rather than attempting optimization ad infinitum, we select these values through a combination of trial and error, ease of implementation, and the borrowing of general practices from the machine learning literature We informally tested a range of different initialization and seeding, fitness scaling, crossover, and mutation schemes with simulated data, along with both discretized and continuous functions. The features described above outperformed alternatives in our particular environment. The feature that was most important for accelerating runtime was discretization or considering elasticities only to two decimal places. The features that were most important for ensuring convergence to the correct solution were seeding and adaptive restart. Small population sizes tend to perform well the majority of the time and the code executes significantly faster with small population sizes; however, occasionally the algorithm will fail to converge to the globally optimal solutions and this occurs more frequently than when there is a very large population size. We took a conservative approach in our population size to ensure accuracy, but smaller population sizes are likely appropriate when there is more emphasis on runtime. The choice of mutation scheme had less impact compared to other features, but perhaps because we restricted ourselves to mutation schemes that have been shown to perform well in the literature.

The steps of our implementation of the genetic algorithm can be summarized as follows.

Algorithm A.1: Genetic Algorithm

Step 0: Randomly generate $(N - \#Seeds)/2$ sets of industry-level trade elasticity guesses from a discrete random uniform distribution with bounds (1.1,30). For each set of elasticities, including the initial seed guesses, create generate the opposite number using the provided bounds ($\{30 + 1.1 - \sigma_s\}_n$ is the opposite elasticities for guess n); so that the total initial population is N.

Step 1: Calculate the fit criteria, $\rho\left(\Delta\left(X_{ij}^s\right)_{t,t'} \Delta\left(\hat{X}_{ij}^s\right)_{t,t'} \omega_{ij}\right)$, for each set of trade elasticities. Rank the sets by their fit.

Step 2: Keep 10=#Elites sets of elasticity guesses with the best fit as elites. Generate 80 percent of children, or 152=(0.8(N-#Elites)) children, using Laplace crossover from two randomly selected parents. Generate 28= (0.2(N-#Elites)) children according to power mutation from a single parent. For both crossover and mutation, each child's parents are randomly selected with probability proportional to $1/\sqrt{rank}$.

Step 3 (hybrid optimization): Run a local optimization algorithm with an initial guess set equal to the solution of the genetic algorithm with the best fit criteria. Include the solution of the local optimization algorithm as one of the elite children for future iterations, and select it as the final solution after the algorithm terminates if its fit criteria is better than those from step 2.

Step 4: Iterate Steps 1- 3 until the genetic algorithm terminates due to the weighted geometric mean of fitness tests is below the tolerance. Let F_n be the fit criteria from Step 1 of the top ranked set of elasticities. Then the algorithm stops when the geometric weighted average relative change in the best fit, $F_n - F_{n-1}$, where the weight for the change n steps ago is $\left(\frac{1}{2}\right)^n$, falls below the tolerance.

$$\left(\prod_{n=0}^{N-1}\left(\left|F_{N-n} - F_{(N-(n+1))}\right|\right)^{\left(\frac{1}{2}\right)^n}\right)^{1/\sum_{n=0}^{N-1}\left(\frac{1}{2}\right)^n} < tol$$

A.5. Performance of the Genetic Algorithm in Recovering Trade Elasticities

Here we explore the robustness of our genetic algorithm in recovering trade elasticities using simulations from an AGE model with known trade elasticities. For each simulation, we first set the number of sectors to match the number of industries in WIOD, $S = 16$, and the number of countries to the minimum size, $J = 3$. Let $U(0,1)$ represent random variables drawn from a continuous uniform distribution with lower bound 0 and upper bound 1. We initialize each simulation by generating a matrix of initial expenditures, $p_{ij,t}^s y_{ij,t}^s \sim U(0, 100S/J)$, initial tariffs, $\tau_{ij,t}^s \sim U(1,2)$, post-liberalization tariffs, $\tau_{ij,t'}^s \sim \left(\tau_{ij,t}^s - 1\right)U(0,1)$, and industry-level trade elasticities, $\sigma_s \sim U(1.1, 30)$, where elasticities are rounded to the nearest decimal place. For simplicity, we set IO flows and LTP shares to zero, but our algorithm performs similarly for our extended model and when we randomly generate those values as well.

Table A.5.1 shows the performance of our genetic algorithm with different population sizes (the number of distinct sets of elasticities evaluated at each iteration of the algorithm). Table A.2.1 highlights that larger population sizes both perform better, and have larger run times. It also shows that a hybrid optimization approach, whereby we run a local optimization algorithm following the genetic algorithm, yields improved performances and while also potentially reducing overall runtime.

Table A.5.1. Performance of Different Fit Criteria (Fit = Fisher's Transform)

Population Size	Hybrid Optimizer	Exact Match (%)	Mean Absolute Error	Mean Absolute Percent Error	Average Runtime (Seconds)
N=40	Yes	99.94	0.0002	0.00	267.9
N=160	Yes	100.00	0.0000	0.00	421.9
N=640	Yes	100.00	0.0000	0.00	1047.9
simple average	Yes	99.98	0.0001	0.00	579.2
N=40	No	58.44	0.1236	0.96	149.6
N=160	No	96.63	0.0074	0.04	472.6
N=640	No	100.00	0.0000	0.00	1217.2
simple average	No	85.02	0.0437	0.33	613.1

Figure A.5.2 shows the Weighted Correlation between the model and the data for different elasticity values for sector 1 and sector 2, for a simulation where the data is generated by a model where the true elasticity is set to 8 for each sector, and the other sectors are at their correct elasticity. This shows us, that when all the other sectors are set to the true elasticity, the weighted correlation between the model and data is locally stable around the true elasticities. This local stability around the solution when the other sectors are at the correct solution is what leads a hybrid approach to be so beneficial.

Figure A.5.2. Weighted Correlation between Model and Simulated Data

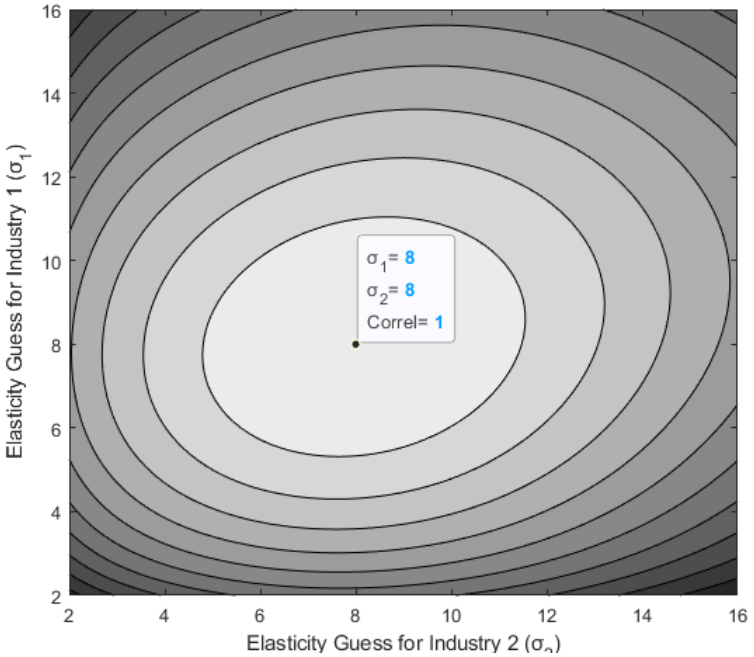


Figure A.5.3 offers a similar picture of how various metrics of model fit perform under three different scenarios. In each scenario, the simulated data is generated from a model where the elasticities are set to 8. The bold line shows how the metric of model fit varies with the elasticity guess for sector 1 when the elasticity for the other 15 sectors is correctly set to 8. For each of the different fit metrics, Weighted Correlation, Fisher's Transform of Weighted Correlation, Weighted Sum of Squared Errors, and Mean Absolute Error, have optima when sector 1's elasticity is also set to the correct value of 8. The dotted line shows what happens when the other elasticities are perturbed, so that the elasticity guess for sector $s = 8 + (1/3)((-1)^{(s-1)}(s-1))$. This means that none of the elasticities are set to the same value, many elasticities are far from their true value, however, the average bias across elasticities is relatively small. In this scenario, the overall fit according to each metric is much lower than in the first scenario, and depending on the metric, there may be some degree of bias in terms of whether the local optimum is achieved at an elasticity for sector 1 of 8 or a different value. The third scenario shows the fit metric for different elasticity values of sector 1, when the elasticity for each of the other sectors in the model is set to 10, a constant perturbation of +2 versus the true elasticity in these sectors. Figure A.2.3. highlights both the somewhat lower fit for each of the metrics, as well as the bias in the locally optimal solution when the other elasticities are not set to their true values. This figure highlights why a local optimization

algorithm ability to converge to the global optimum is highly dependent on its initial guess. Second, it highlights how the genetic algorithm is able to locate the global optima due to difference in Fit Criteria, and it highlights how these differences are more pronounced when the fit criteria is Fisher's Transform or Mean Absolute Error.

Figure A.5.3. Fit Metrics between Model and Simulated Data

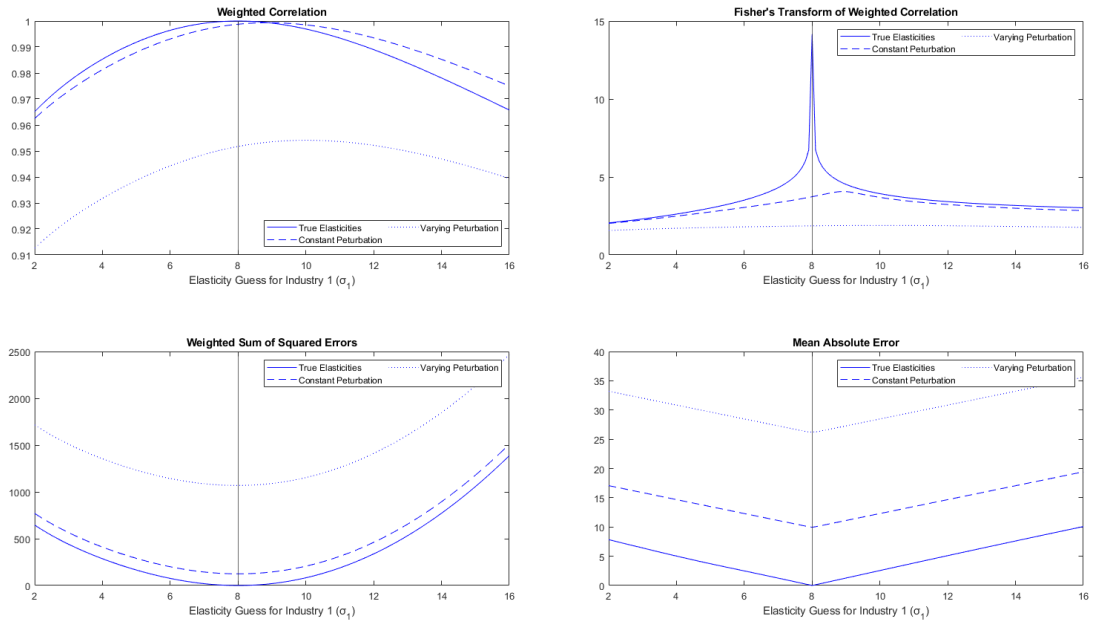


Table A.5.2. shows the performance of the genetic algorithm with each Fit Criteria with a population size of 40. This population size is intentionally set low, since with a large enough population size each of the algorithms would converge exactly to the globally optimal solution and the only variation would be in run-time. This table shows that several fit metrics perform well, with Mean Absolute Error performing the best, ignoring runtime, although Fisher's Transform of Weighted Correlation also performs well. These results suggest that one might prefer to use either the Weighted SSE or Mean Absolute Error instead of the Fisher's Transform we rely on the in paper. Our opinion is that either can perform satisfactorily, however, scaling is a concern for us given the WIOD does not cover out complete sample, and therefore we do not have production data for all sectors and all years. Incorrectly deflating or scaling changes in the data would necessarily lead to incorrect elasticity estimates when using a metric that is not scale invariant. However, for researchers that are working with data and are able to work with shares (industry exports to a destination divided by total industry production), rather than values, we would indeed recommend Mean Absolute Error or Weighted SSE as a reasonable choice that may be more intuitive in some contexts.

Table A.5.2. Performance of Different Fit Criteria (Population Size = 40)

Fit Criteria	Hybrid Optimizer	Exact Match (%)	Mean Absolute Error	Mean Absolute Percent Error	Average Runtime (Seconds)
Fisher's Transform	Yes	99.94	0.0002	0.00	276.4
Weighted Correl.	Yes	80.19	0.0413	0.24	199.8
Mean Absolute Error	Yes	100.00	0.0000	0.00	203.4
Weighted SSE	Yes	100.00	0.0000	0.00	138.0
Simple Average	Yes	95.03	0.0104	0.06	204.4
Fisher's Transform	No	54.06	0.1357	1.08	160.2
Weighted Correl.	No	42.44	0.1856	1.29	175.7
Mean Absolute Error	No	92.25	0.0219	0.22	122.6
Weighted SSE	No	84.88	0.0363	0.42	139.6
Simple Average	No	68.41	0.0949	0.75	149.5