THE UNIVERSITY OF CHICAGO

TOWARD A SYSTEMATIC APPROACH TO EX-ANTE FTA IMPACT ESTIMATION ANALYSIS

A PAPER SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE MASTER OF ARTS DEGREE IN THE MASTER OF ARTS IN COMPUTATIONAL SOCIAL SCIENCES

BY
ELLIOT DELAHAYE

PRECEPTOR: GINA PIETERS

CHICAGO, ILLINOIS
AUGUST, 2023

Acknowledgements

I am grateful for all those who have supported me through my journey at Chicago. In particular, I'd like to thank my family, who have always shown great faith in my ambitions, and who have always gone the extra mile to support me in them.

My time here would not have been the same without my closest friends. Special gratitude goes to Maria, who's constant support throughout the last year has always encouraged me to believe in my abilities but also take a break from studying. In addition to my housemates, Avery, Lynette, and Silvan, I'd like to thank my fellow economists Awaid Yasin and Eliot Weinstein; through the stress of many, many assignments, they have kept me sane and in good humour.

Finally, I'd like to thank my academic mentors at UChicago, in particular Gina Pieters. Her motivation in hard times has been much needed, and her mentorship has given me the confidence to excel as an economist.

Contents

Abstract

A standard approach to the analysis of free trade agreements (FTAs) involves simulating a general equilibrium model of the global economy in response to shocks that reflect trade cost liberalisation. A few methods are common to estimate the scale of these essential trade shocks, but little formal research has gone into evaluating the success of these different approaches. Developing a systematic approach to the evaluation of FTA trade shock estimates is essential to informed policy making in international trade, as well as to the academic literature on the evaluation of FTAs.

This paper contributes to the existing small literature focused on estimating trade impacts for use in economic modelling of trade. It extends a framework first used in a paper by [Baier et al., 2019]. This method presented a novel way to evaluate ex-ante trade shock estimation methodologies by posing the issue as a prediction problem to predict ex-post FTA shocks identified through a first stage gravity model. In this paper I extend the original paper's ex-ante analysis by making use of a wider dataset, as well as adopting recent methodological innovations from the machine learning literature. I find that amongst the most important considerations in making ex-ante FTA impact estimates are whether a model is trained on data excluding outliers and the type of machine learning model chosen. Furthermore, the broader set of data used is found to be important in improving success.

1 Introduction

A standard approach to the analysis of free trade agreements (FTAs) involves simulating a general equilibrium model of the global economy in response to shocks that reflect trade cost liberalisation. A few methods are common to estimate the scale of these essential trade shocks, but little formal research has gone into evaluating the success of these different approaches in generating accurate results ex-ante. Developing a systematic approach to the evaluation of FTA trade shock estimates has the promise to improve FTA evaluation, both in policy and the academic literature, and in enabling better informed policy making in international trade.

As an illustrative example, the guidance given by UNCTAD to member nations to create their official estimates of the value generated by signing an FTA between two or more countries is to use econometric estimates of trade barrier changes advising, among other tools, that gravity models allow the estimation of a wide range of trade policies. Specific implementations of this can be found in impact assessments published by the European Commission, the United States, and the United Kingdom, amongst others. Initially this estimate tended to be a high-level average treatment effect across FTAs but this left much unexplained variability in the estimates. The approach has evolved over recent decades as improved data on the contents of FTAs have allowed heterogenous treatment effects to be estimated. However, these estimates have primarily been ex-post estimates of the impact of an FTA or sub-type of FTAs, which are then repurposed to serve as the ex-ante estimate to be used in the analysis of the new FTA. This approach would seem inadequate for two reasons. Firstly, the improved data on FTA contents has shown that the provisions made within FTAs have changed dramatically over the past few decades, especially in terms of including non-tariff provisions [Hofmann et al., 2017]. This leaves a great scope for error if relying solely on ex-post estimates which will, by their nature, reflect a sample of FTAs different to those likely to be signed today. Secondly, research has shown that as much variation in FTA impact size is attributable to within-agreement variation as to between-agreement variation [Baier et al., 2019]. This shows the importance of other factors outside of FTA-specific variables and the consequent inadequacy of ex-post estimates which have traditionally reflected simple dummy variables for FTA presence, regardless of contents[1].

To fill the gap in research regarding the optimal approach to estimating FTA trade shocks ex-ante, this paper extends one of the first frameworks that attempts to make ex-ante estimates [Baier et al., 2019]. This method presented a novel two-step approach to make ex-ante trade shock estimation methodologies by posing the issue as a prediction problem to predict ex-post FTA shocks identified through a first stage gravity model. In this paper I extend the original paper's ex-ante analysis by making use of a wider dataset, as well as adopting

[1]Whilst technically these other factors could be included in ex-post estimates through interaction variables, including a large number of interactions is often impractical and the non-linearity imposed by this approach makes estimating the relative importance of each variable intractable.

recent methodological innovations from the machine learning literature. In this way, I evaluate which analytical components, in terms of data and methodology, are important to accurate out-of-sample estimation of FTA impacts. The paper proceeds as follows, in the section 2, I review the literature on trade shock estimation for ex-ante FTA analysis, both in academic research and in the practice applied by policy institutions; in section 3, I outline the methodology approach pursued in the analysis in two-steps; in section 4, I outline the results from each stage; in section 5, I offer concluding remarks.

2 Literature

The trade cost shock estimates used in practice often take the form of three broad approaches: bottom-up estimates, top-down estimates, and ex-post gravity estimates. It should be noted that the focus of the below is largely on trade cost changes arising from changes in non-tariff barriers. This reflects the relative ease with which tariff barrier changes can be quantified, given their explicit nature. The focus of this paper is trade costs in general but this could be easily changed to focus on NTB changes through the inclusion of an specific tariff variable.

Bottom-up estimates tend to work from datasets on the incidence of NTBs to impute the overall level of barriers using trade elasticities. A key reference for this approach uses UNCTAD TRAINS incidence data to impute barriers [Kee et al., 2009]. To obtain changes, it is then common practice to remove some proportion of the observed NTB incidences to calculate how the total barrier changes. This is the approach taken by the LSE in their modelling of the EU-Australia and EU-New Zealand FTAs, undertaken as a consultation for the European Commission [London School of Economics, 2017]. The bottom-up barriers of [Kee et al., 2009] are taken and reduced across manufacturing sectors by an assumed 10% of their original values. A weakness with these approaches in general is that they rely on estimating the impact of the "count" of NTBs reported in a country, where these are likely to be highly heterogeneous in reality; by essentially imposing that they have a uniform impact, estimates tend to have a high variance and also depend on the quality of reporting for each country.

Top-down approaches start from estimates of the initial level of NTBs in place in each partner which are inferred from unobservable distortions in trade to calculate barrier that would be required to explain the portion of trade that can't be explained through variation in other, observable factors[2].A certain portion of this initial level is then removed to reflect the agreement using methods that range from the simplistic to the more sophisticated. Top-down estimates are available from a variety of sources, but a common reference is [Fontagné et al., 2016]. Once the initial level is determined, researchers then decide a proportion of the barrier to remove to reflect the provisions specified in the agreement. A simple method for doing this can be found in the United

[2]For an overview of these methods, see [Abbyad and Herman, 2017]

Kingdom's preliminary reports on the value of an FTA prior to commencing trade negotiations; a simple assumption is made that 50% of goods barriers and 33% of services barriers could be removed in principle and then two scenarios are run where 25% and 50% of these remaining "actionable" barriers are removed [UK Department for International Trade, 2020]. More sophisticated approaches have also been used to try to integrate policy judgement in a more systematic way. In an assessment of the Trans-Pacific Partnership, a Peterson Institute report broke NTB provisions into 23 different weighted chapter areas and scored each chapter as to its depth. Along with a 67% actionability assumption on level estimates, this generates the final NTB AVE change [Petri et al., 2012]. A more analytical approach can be seen in the US government report on the impact of USMCA. Here the estimates from Fontagne et al. are used as the left-hand side variable in a regression where data on the OECD's Services Trade Restrictiveness Index is used as an explanatory variable to estimate the proportion that should be reduced by the agreement [United States International Trade Commission, 2019].

In general, a weakness of the above two approaches is that, in estimating the existing level of barriers and removing a proportion of them, they neglect evidence on what trade changes have actually been achieved by FTAs historically. A more empirically grounded approach is to use gravity modelling to estimate ex-post impacts of FTAs. This approach is common in both academic and policy literatures and has developed over the past decade to obtain greater sophistication. Initially, the literature estimated average treatment effects of FTAs, but as data improved, heterogeneous treatment effects became more common. A notable improvement was the Design of Trade Agreements, "DESTA", dataset which classified FTAs by their type (i.e. customs unions, partial scope agreements, and standard FTAs) as well as whether they included 7 key chapter within their texts [Dür et al., 2014]. This was then extended significantly extend by a World Bank dataset, The Content of Deep Trade Agreements, which summarised whether FTAs included 52 different chapter types [Hofmann et al., 2017], which was then extended even further to provided detail on whether FTAs included over 1000 different types of provision [Mattoo et al., 2020]. An example of this approach being used for ex-ante evaluation is the UK's approach in recent impact assessments where DESTA data is combined with country NTB level estimates to obtain heterogenous treatments [UK Department for International Trade, 2021]. Whilst not focused on ex-ante estimation specifically, research work in the academic literature has made some initial use of machine learning methods to sort through this large quantity of data. Recent work has combined standard gravity model estimation using PPML with a LASSO penalty term to decide which provisions from the most recent World Bank dataset are relevant to trade [Breinlich et al., 2021]. However, this work is limited for ex-ante estimation for the same reasons as the earlier work on FTA evaluation; a focus on causality limits its usefulness for out-of-sample prediction and, to the extent that there is a discussion of predictive success, the use of LASSO in the first stage focuses this as prediction of trade generally, rather than measuring prediction of pre-identified FTA impacts.

Much of this work has used ex-post estimates without questioning their applicability for ex-ante estimation purposes. A small academic literature focusing specifically on how to conduct ex-ante estimation does exist. A recent contribution to this literature recognises the shortcomings of using ex-post gravity estimates, with the authors noting that even heterogeneous FTA treatment estimates are likely to be limited in the information they can give about a particular FTAs impact. Instead, the authors use heterogeneous estimates of the border variable obtained through gravity modelling to approximate the change in trade costs that occurs when switching between two policy regimes (i.e. within-EU trade costs vs non-EU countries trade costs with EU member states) [Larch et al., 2023]. This is an effective solution for estimating the specific trade cost change that will occur when acceding to a pre-existing trade policy agreement, such as the EU, where the border estimate for within-FTA trade can be obtained. However, this in essence still relies on ex-post estimates and so is limited in being dependent on the sample of currently observable FTAs; it cannot be used to imply the trade cost change that would occur for unseen FTAs and is therefore limited to use in cases such as joining plurilateral agreements. Furthermore, as noted above, prior research has shown that as much variation between FTA impacts is attributable to within-FTA heterogeneity as it is to between-FTA heterogeneity. By using a border estimate for an existing agreement, this approach assumes homogeneity in trade cost changes between previous FTA parties and the new country. Consequently, the existing work on ex-ante estimation misses much of the variability present across FTA impacts, and is limited to use on existing agreements.

Interestingly, there is much greater consensus on ex-post evaluation of FTAs. As documented by the guidance on methodology followed by USITC in a 2021 report, a standard approach is gravity modelling estimates of the trade agreements [United States International Trade Commission, 2021]. An example of this in practice is the European Commission's evaluation of the EU-Korea FTA. Here a gravity model is used to evaluate partner specific export impacts by sector with the fixed effects of the gravity modelling ensuring that these detailed heterogeneous impacts are identified [European Commission, 2019].

The work that I seek to extend in this paper proposed a method that can reconcile the agreed upon ex-post estimation methodology with ex-ante estimation [Baier et al., 2019]. By using ex-post estimates of FTAs from a first-stage gravity model, the authors obtain a set of identified impacts which they then use to train a second stage model to predict the impact of an unseen agreement. This ex-ante estimation portion of their analysis is a secondary concern compared to a primary focus on causal drivers of trade cost changes and the authors do not explore in depth the question of which analytical features lead to the best predictive performance. However, their methodology presents a potential framework to establish a single criterion for the evaluation of an ex-ante estimation methodology; by framing the problem as a prediction problem, standard evaluation criteria such as root mean square error (RMSE) and other metrics can be used to select amongst approaches. Indeed, a feature of the literature is that papers often discuss their estimates in isolation, not comparing to each other against a common standard of success.

3 Methodological Overview

3.1 First Stage Ex-Post Estimates of Historic FTAs

The method follows that of [Baier et al., 2019] who break down estimation into two parts; ex-post estimation of individual FTA impacts and then out-of-sample prediction of these ex-post estimates to obtain ex-ante estimates. This first stage can be conducted at three levels of aggregation; at the individual FTA level, the symmetric FTA-pair level, and the directional FTA-pair level. The latter two levels of disaggregation allow within-FTA heterogeneity which is often assumed away in many ex-ante estimates but has been shown to be important in ex-post assessments of FTA impacts [European Commission, 2019]. Unpacking the sources of this variability is key to policy relevance as it informs how individual countries are affected by the agreement.

The first stage estimates are made using a gravity model of comparable specification to that in the original paper:

$$y_{ijt} = \exp(FTA'_{ijt}\beta + \alpha_{it} + \delta_{jt} + \eta_{ij}) + \varepsilon_{ift}. \quad (1)$$

Here y_{ijt} is bilateral trade from country i to country j at time t, and α_{it}, δ_{jt}, and η_{ij} are exporter-time, importer-time, and pairwise fixed effects respectively. FTA_{ijt} is a vector containing the set of FTA treatment variables which in each successive regression refer to individual coefficients for each FTA, individual coefficients for symmetric FTA-pair combinations, and individual coefficients for directional FTA-pair combinations. ε_{ift} is the usual error term. The exponential functional form shown above represents the Poisson-distributed maximum likelihood estimation that is now accepted as standard in gravity modelling. The inclusion of the fixed effects is important for successful identification of our set of treatments in terms of structural variables and reverse causality.

3.2 Second Stage Ex-Ante Estimates

The ex-ante estimation portion [Baier et al., 2019] uses a simple OLS regression of the form

$$\hat{\beta}_a = x'_{ija}\kappa + \epsilon_{ijt}$$

where $\hat{\beta}_a$ is the set of estimated treatment effects from equation (1) and x'_{ija} is a vector of second-stage treatments whose effect on FTA impact are of interest[3]. It is estimated with a leave-one-out cross-validation approach whilst using the following manual selection procedure for their model specification; starting from the set of variables found to be significant in their causal analysis, the authors

[3]Here the a subscript refers to variation across the FTA dimension. The i and j subscripts are retained for the explanatory variables as some variables are importer and exporter specific.

chose their final prediction model by manually dropping individual variables and assessing the out-of-sample predictive success against their preferred measure of fit. The measure of fit used to evaluate the success of their models involve regressing the predicted values for the first-stage estimates against the original estimates. The R^2 for this model summarises the extent to which the predictions successfully explain the observed estimates (I refer to this metric as the "fitted value R^2" below). The manual nature of the authors model selection criteria is a weakness in their original analysis; not only does it rely on informal judgement as to which subsets of variables to test, but it becomes computationally intractable to perform with a larger set of dependent variables to choose from.

Given the wealth of data that I use in this study (discussed below in section 3.3), a method is needed to efficiently choose which variables to select for a final model, as well as to similarly sort through possible functional forms in how these variables might interact with one another. Modern advances in machine learning offer sophisticated solutions to these problems. Not only can they perform variable selection optimise the bias variance trade-off but they can also choose this to maximise out-of-sample predictive success. This is important given my primary goal of estimating FTA effects ex-ante. Furthermore, as machine learning methods get increasingly complex, they can also go beyond the linear parametric functional forms that are standard in econometric analysis to take into account non-linear impacts of variables and interactions that might occur between variables. I exploit this capability to explore the importance of variable choice for out-of-sample predictive success. I also examine the importance of functional form, in particular whether there is added benefit to non-parametric methods or whether the standard parametric assumption of linearity is sufficient for success.

Models are run at each different level of aggregation and on data samples ranging from the full sample of first stage ex-post estimates, a subset of those which are statistically significant, and a subset of those estimates that are not outliers.

3.2.1 Linear Parametric Models

The first class of models that I explore maintain the linear parametric form used in the original paper's model.

One of the most common linear models used in the machine learning literature is the "Least Absolute Shrinkage and Selection Operator", or LASSO, model. Adopting the above notation, this procedure chooses an estimator, $\hat{\kappa}$, according to the following objective function:

$$\hat{\kappa} = argmin_{k \in \mathbb{R}^p}(\frac{1}{n}\sum_{a=0}^{n}(\hat{\beta}_a - x'_{ajt}k^2) + \lambda||k||_1) \tag{2}$$

.

This represents standard minimisation of the sum of squared errors with the addition of the penalty term $\lambda||k||_1$, where $|||k||_1$ is the L_1 norm (i.e. the sum of absolute values for the parameters) and λ is a hyperparameter to be chosen.

Two approaches are used for tuning the hyperparameter, λ, cross-validation and the approach put forward by [Belloni et al., 2012]. For the former, I follow the approach of the original paper and use leave-one-out cross-validation. Not only does this ensure comparability with the baseline results of the original paper but it makes optimal use of the relatively small sample size of FTAs that will be estimated. The optimal lambda for this approach, which I term λ_{min}, is then chosen as that which maximises the RMSE across the cross-validated samples. The benefit of the cross-validation approach to hyperparameter tuning is that it is focused on the primary goal here of out-of-sample prediction, whereas by contrast the latter approach (which I refer to here as λ_{BCCH}) represents a theory-driven approach to the optimal choice of λ which, by imposing a lighter penalty, is less sensitive to dataset characteristics.

Given that each observation in the second stage estimation represents an FTA impact, there is likely to be some collinearity between variables; given that FTAs represent bundles of provisions and that there are many more provisions than FTAs, it is likely that provisions will often appear together. This can lead to important variables being deselected and the predictive success diminished. Furthermore the existence of many more predictor variables than observations makes this problem all the more acute.

A generalisation of LASSO, known as elastic net, can help overcome these issues if they are present. Elastic net estimation generalises the penalty term as can be seen in its objective function:

$$\hat{\kappa} = argmin_{k \in \mathbb{R}^p}\left(\frac{1}{n}\sum_{a=0}^{n}(\hat{\beta}_a - x'_{ajt}k^2) + \lambda(\omega||k||_1 + (1-\omega)||k||_2)\right) \tag{3}$$

.

The penalty now includes the L_2 norm of the coefficients as well as a blending parameter, ω, which weights the importance of the two norms. The L_2 norm adds a stricter penalty for large parameters meaning that correlated variables can remain if they add information[4]. For my estimation, I choose ω via cross-validation and combine it with a value of λ also chosen through cross-validation[5].

3.2.2 Random Forest Models

Moving beyond the linear functional forms used above, I explore more general models of the form

$$\hat{\beta}_a = m(x_{ija}) + \epsilon_{ijt}.$$

Non-parametric models such as these have the benefits of greater flexibility in that they allow for non-linear impacts of variables as well as more complex

[4]Here LASSO represents the special case where $\omega = 1$ whilst $\omega = 0$ represents another model, known as RIDGE, which is also common in the machine learning literature.

[5]Given that the hyperparameter tuning for ω is necessarily data-driven, it adds little to examine the success of elastic nets using $\lambda = \lambda_{BCCH}$.

patterns of interaction between the variables. This could prove particularly important given the inclusion of detailed FTA provision data in the study; given that many clauses in FTA relate to each other, their impact on trade is likely to show up as an interaction effect.

Regression tree models find the optimal groupings of the independent variables, based on their in-group dependent variable homogeneity. This identifies patterns within the data and ensures that the variability of estimates is minimised. Fitting such detailed models based on the characteristics of the data risks over-fitting to the data and so a further extension is given by "Random Forest" models. This group of algorithms draw a series of bootstrapped samples from the data as well as a random subset of independent variables to use. A regression tree is then trained on each data sample to give a "forest" of trees. Whilst each tree is tailored to the anachronisms of its data sample, taking the average across the trees' predictions has shown to markedly improve out-of-sample predictive success.

As with all machine learning algorithms, its success is governed by hyperparameters which need to be chosen. The key parameters for random forests are the number of trees ($ntrees$) that are grown (i.e. the number of bootstrapped samples averaged across) and the number of independent variables ($nvar$) that are selected for use in growing each tree. For the former, I choose $ntrees = 500$ and I conduct tests to show that this is sufficient. For the latter, I choose $nvar = \frac{dim(x_{ija})}{3} = 380$, which is a standard choice in the literature.

3.2.3 Model Evaluation

As discussed above, posing the problem of ex-ante FTA impact estimation as an out-of-sample prediction problem allows for the model selection criteria to naturally be defined as the RMSE of each model. For each model, I calculate the training RMSE, estimated on the full sample, as well as the testing RMSE, estimated as in [Baier et al., 2019] using leave-one-out sample splitting. These are the primary metrics that I use to evaluate the importance of different model characteristics for ex-ante estimation.

For the benefit of comparison, I also report the "fitted R^2" value used by the original paper and discussed above. Defining it more formally, I regress the predicted values from the second stage stage, denoted $\tilde{\beta}$, on the estimated coefficients from the first stage, $\hat{\beta}$:

$$\tilde{\beta} = \gamma_0 + \gamma_1 \hat{\beta} \tag{4}$$

The R^2 of this regression is then taken as the "fitted value R^2". As well as comparability with the estimates of [Baier et al., 2019], this regression also allows for effective visualisation of the predictive success, and through γ_0 provides a measure of bias that might occur in the model.

3.3 Data Overview

For the first-stage estimates, I follow the original paper and use total bilateral goods trade flows for the years 1988-2006 in a dataset that included domestic absorption. Given goal of including information on the contents of FTAs, data for the FTA treatment variables is taken from the World Bank's latest release of their Content of Deep Trade Agreements database [Mattoo et al., 2020]. This provides information on which countries have signed bilateral FTAs which can then be used to construct the lower-level treatment variables.

For the second stage, I separate variables into 3 categories: FTA-specific variables, country-specific variables, and pair specific variables. For FTA-specific variables, the World Bank dataset discussed above provides data on over 1000 different provisions. Whilst some previous datasets provided systematic detail on the contents of FTAs at a higher level, the original paper made no use of these, despite having found that 35.5% of variation across directional-pair level estimates is attributable to FTA specific factors. Therefore adding these variables has the potential to greatly improve predictive performance.

For country-specific variables, the original paper restricted the set of variables to the log of GDP. I expand the set of variables here to a much larger set of national accounting variables using the Penn World Tables [Feenstra et al., 2015]. Another addition of potential importance is adding in data on the size of existing non-tariff barriers (NTBs) that exist in the country absent an FTA. This makes intuitive sense that it may affect the realised trade impacts of an FTA; if there are low barriers to begin with, there may be limits to what can be achieved by an agreement whereas high barriers allow more to be removed. This again was not considered in the original analysis as there was no widely agreed methodology for estimating these barriers. Given the uncertainty around any estimates of NTBs, the data I choose for this task was made on the grounds of country coverage. For goods NTBs, I use World Bank data based which has a country coverage limited to just under 70 countries [Kee and Nicita, 2022]. Whilst the trade data used only covers manufacturing, some research exists establishing a relationship between service barriers and the ability of countries to trade in goods (consider the importance of transport or finance in facilitating goods trade) [Hoekman and Shepherd, 2017]. I therefore choose to include measures of services barriers estimated in previous work [Jafari and Tarr, 2017]. Both NTB datasets are flawed by not having time varying estimates of barriers, however if it is assumed that the ranking of restrictiveness hasn't changed substantially over the past decades, the estimates still provide important information on MFN restrictiveness.

For pair-specific variables, the original authors considered the full range of standard gravity variables in their causal analysis (distance, common language, common legal system etc.) but restrict themselves to using distance when it comes to the ex-ante portion of their analysis. Being guided by the statistical significance in a causal analysis may be a misguided approach to choosing a prediction model however, as a variable which doesn't meet the conventional cut-offs for statistical significance may still be important when considering out-of-sample performance metrics. Given the ability of the machine learning methods to sort through variables efficiently in terms of their contribution to predictive success, I therefore return to using the full set of gravity variables provided in the dataset by USITC.

4 Results

4.1 First Stage Ex-Post Estimates

Estimating the above first-stage regression specification at each level of aggregation gives a large set of estimated impacts. Table 1 shows the number of treatments that were in the original dataset as well as how many were able to be identified[6].

Level	No. Present in Data	No. Identified
FTA Level	78	78
Symmetrical Pair Level	1052	936
Directional Pair Level	2104	1881

Table 1: The no. FTAs in the data & how many were identified in the first stage estimation

This set of estimates is larger than that estimated by Baier et al (2019), as a result of the change in data source for FTA treatment variables. In the dataset used by those authors, 65 individual FTAs are present, leading to 455 symmetric-pair estimates, and 910 directional-pair estimates (of which 908 are identified). Discussion of the difference.

Figure 1 displays the distribution of the first-stage estimates as well as the 95% confidence interval around each estimate[7]. This replicates figures from the original paper and shows a broadly similar of results. Note that these y-axes here is the coefficient of the gravity model which doesn't have a direct interpretation, these can be converted into percentage trade impact using the formula $\Delta y_{ijt} = (\exp(\beta) - 1) \times 100$. As can be seen, each level of disaggregation exhibits estimates that are positive and estimates that are negative. Given the conversion formula just given, this translated into there being positive and negative trade impacts as well. Each set of estimates have a median estimate that is positive, with mean results for the first two levels positive and for the directional pair level being negative. This means that trade increases in the majority of cases across at each level of aggregation, although extreme negative values in the directional-pair level results has brought a negative mean. As with

[6] Some FTAs are unable to be identified as they don't change across the time period. In this case, they are indistinguishable from the country-pair fixed effects and so drop out of the model.

[7] Values for the y-axis on the middle and bottom panel are adjusted set to allow great detail of the distribution for the majority of results. This means that estimates on the extreme of the distributions are omitted in the visualisation.

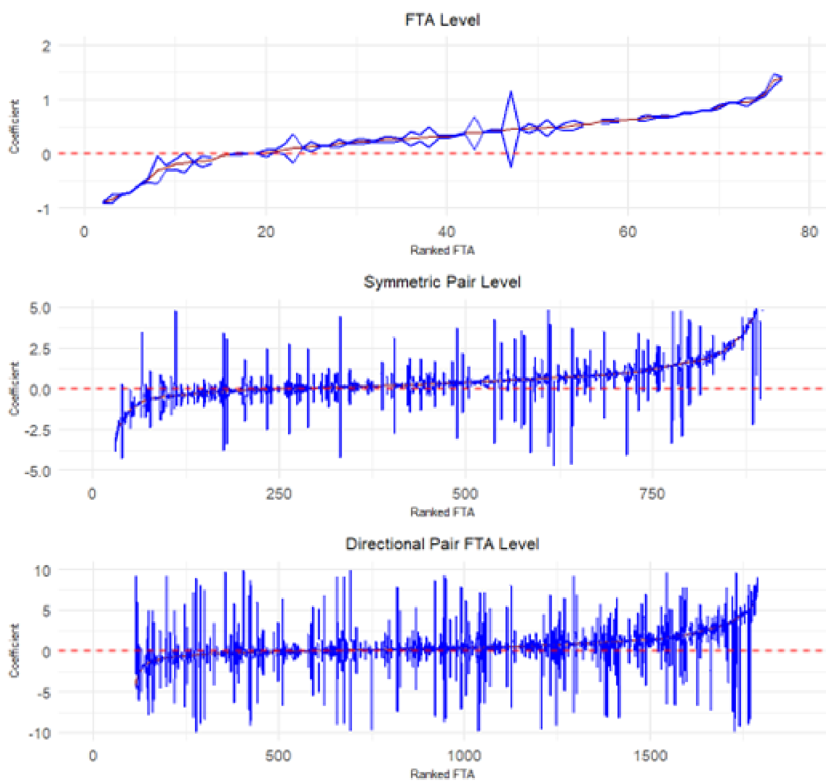


Figure 1: Distribution of estimated FTA coefficient values, with 95% CIs

	FTA Level	Symmetrical Pair Level	Directional Pair Level
Full Data	0.623	47.89	2205.4
Interquartile Range	0.166	0.268	0.333

Table 2: Standard deviations for the estimated coefficients

the original paper, a large amount of variance is present across the estimated coefficients[8]. The extreme values calculated for the lower values make this harder to observe this visually in the above figures. Therefore, in Table 2 I calculate the standard deviation of each set of results, as well as a standard deviation calculated only using the interquartile range.

The extreme values in the more disaggregated regressions can be seen to have a large impact on the variability in the dataset. Disregarding these extreme values, a large amount of variability can still be seen in the results. The standard deviation preserves the units of the coefficients, but these too can be converted into trade impacts. Non-outlying coefficients can be seen to be deviate on average within the IQR by 16.6, 26.8, & 33.3 percentage points which translates to a deviation in the trade impact of 18, 31, and 39 percent respectively.

Figure 1 also reports the 95% confidence interval for each of the individual coefficients. As with the original paper, the more disaggregated the set of results being considered, the greater the variability in the coefficient standard errors. This results from the more disaggregated coefficients having fewer observations to be used in identification and hence a lower degree of statistical precision possible. Across each level of aggregation, a large proportion of the coefficients can be seen to be significant although many can be seen to insignificant, particularly at the lower levels of disaggregation.

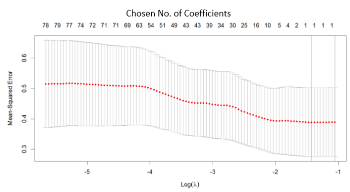
4.2 Second Stage Ex-Post Estimates

4.2.1 Linear Parametric Estimates

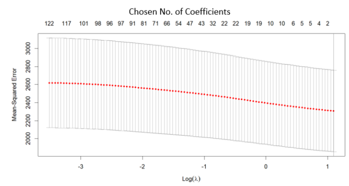
Applying the LASSO approach at each level of aggregation, results for the number of variables selected and chosen lambdas are reported in the Figure 2 below.

In each panel of the figure, the vertical axis shows mean square error and the horizontal axes show the number of coefficients chosen by the algorithm and the value of lambda that corresponds to this respectively. The vertical line show value of the hyperparameter, lambda, that maximising the out-of-sample prediction, as well as the lowest value of lambda that is within one standard deviation (a tougher penalty for more conservative estimates). The

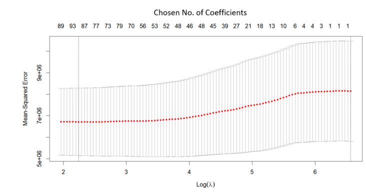
[8]Note that these are the estimated coefficients rather than the estimated effects on trade. To transform them to a trade impact one would merely exponentiate the coefficients and minus 1.



(a) FTA Level



(b) Symmetric Pair Level



(c) Directional Pair Level

Figure 2: Cross-Validation Curves for LASSO Models

initial two levels of aggregation show that very little information is provided by any of the explanatory variables; both choose models with only one variable, the intercept, which suggests that the uncertainty around the estimate is white noise. This may occur due to the relatively small number of observations that are available relative to the explanatory variables that are being used to fit the models. However, the panel for directional pair level estimates show a different picture. It shows that useful information is contained in a fair number of variables (circa 94). Whilst these results are dependent on the cross-validated LASSO specification, I use this as an indicator that the directional pair level results offer the most promise for successful ex-ante estimation of FTA impacts and they therefore become my focus for the remainder of this paper[9].

Digging into the results on the directional-pair level, the results for these are contained in Table 3. Along with the core model comparison metrics discussed in section 3.2.3, I also report the number of variables selected by each model, and the observations used in each estimation.

The results for the two hyperparameter selection approaches are displayed in columns odd and even columns of Table 3 respectively. Comparing first against the results from the original paper, model (1) gives a fitted R^2 of 0.188 whilst the results for the alternate λ_{min} in (2) gives a value of 0.199. This compares favourably with the original paper's result of 0.178 although only marginally. However, the RMSEs calculated for each are large with both exceeding 2500 for the testing RMSE. For both training and testing RMSE, the λ_{BCCH} model displays marginally better out-of-sample predictive success however this improvement is negligible in terms of the size of the error found. These large RMSE results likely reflects the presence of many outliers, as can be seen in the left-

[9]This acts as further confirmation of the results found in [Baier et al., 2019] who, as discussed above, estimated that much of the variation in impacts was within-FTA variation

Data Sample	**Full Data**		**Drop Outliers**		**Drop Insignificants**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Choice of λ	λ_{min}	λ_{BCCH}	λ_{min}	λ_{BCCH}	λ_{min}	λ_{BCCH}
λ	0.014	0.001	0.014	0.002	8.87	0.002
No. of Variables Selected	94	833	47	92	98	832
Training RMSE	2333.645	2315.335	0.4726	0.4279	2759.823	2748.707
Testing RMSE	2591.011	2585.818	0.5475	0.5710	3160.21	3205.325
Fitted R^2	0.1876	0.1985	0.3278	0.3187	0.2076	0.2031
Obs.	667		453		429	

Table 3: LASSO Results

column panels of Figure 3. To explore the importance that outlier datapoints hold for predictive success, as well as the potential that the results are being driven by spurious estimates identified off of a small number of values, columns (3)-(6) repeat the analysis from columns (1) and (2) on a restricted sample of observations; dropping outlier estimates and dropping statistically insignificant estimates.

For models that drop the outlying values, the fitted value R^2 increases markedly with the λ_{min} results having a value above 0.3278 and the λ_{min} results having a value of 0.3187. Even more notably, the RMSE falls dramatically to fall within the region that many standard trade estimates from the previous literature fall within[10]. Here, the estimates show the cross-validation approach to hyperparameter performs better out-of-sample as one might expect, although again this is shown not to matter greatly in larger context of methodological choices when compared to the improvement due to a change in sample. By contrast the results focusing on the removal of insignificant results also have improved fitted value R^2s but it has a worse RMSE compared with the core models. This demonstrates that the issues stemming from outliers is not a result of these estimates being insignificant and poorly identified. Instead, it suggests that there are certain FTAs which, under LASSO, are unable to be explained by the same model that can predict non-outlying estimates.

The above results underline the importance of sample choice in the estimation. As the results for the cross-validation model removing outlying data are the most successful, I discuss its results regarding which variables were selected by the LASSO in greater depth[11]. Whilst causal claims can't be made about these estimates, what is notable about the selection is that it confirms the importance of variables found to be important by the original paper, but also demonstrates the importance of the extended variable set used in this study. For trade-partner specific variables (i.e. gravity variables), distance is again found to be important, although it is now found to have a positive impact

[10]To put these estimates into context, they represent a trade impact of approximately +/- 73% and +/- 77% respectively.

[11]A detailed list of selected variables is given in Appendix A

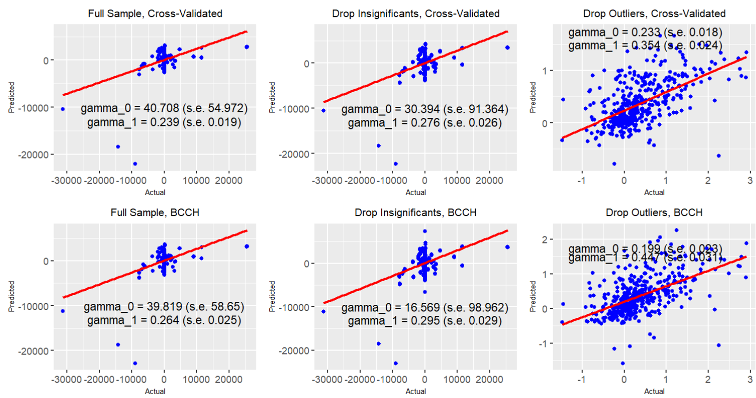


Figure 3: Comparison of first stage ex-post estimates with LASSO ex-ante estimates

where [Baier et al., 2019] found it to have a negative impact. In addition to this exporter longitude is found to be important. In contrast to the original paper, importer and exporter GDPs are not selected, although interestingly the other national accounting variables are found to add explanatory power. Many of these correspond to the individual components of GDP, such as the level importer (domestic) consumption or the level of exporter's general exports being important, as well as shares of GDP components in GDP. Also interestingly, several measures of the price level in both importer and exporter partners are selected. Not only does this match expectations from a-priori trade theory that terms-of-trade matter, it matches the findings of the original paper that they are a determinant of FTA impacts. As in the original paper, the pre-FTA trade barrier levels are found to be important again. This is the case for both importer and exporter barriers. What is notable is that all the pre-FTA NTBs selected are services barriers and almost all have a negative impact on FTA impact. Accounting, banking, and transport services feature most prominently, suggesting that a country's regulatory framework is almost certainly relevant to how FTAs will impact goods trade. Most importantly, a large number of FTA provision level variables are selected as relevant, underlining the importance of considering these where the original paper omitted them. These provisions are chosen from across the chapters of an FTA treaty.

Turning to the more general elastic net specification, results for these are outlined in Table 4. As discussed above, both hyperparameters here are tuned using cross-validation and they are reported, again with the number of variables selected.

As can be seen, the elastic net tends to select a larger number of explanatory variables than counterpart cross-validated estimates from the LASSO results in Table 3. This suggests that the collinearity issues discussed in section 3.2.1 above are indeed relevant; by imposing an increasing penalty for large

Data Sample	Full Sample	Drop Outliers	Drop Insignificants
	(1)	**(2)**	**(3)**
ω	0.1	0.55	0.1
λ	71.115	0.024	92.928
No. of Variables Selected	149	56	150
Training RMSE	2338.018	0.474	2771.711
Testing RMSE	2586.328	0.546	3156.54
Fitted R^2	0.186	0.33	0.204
Obs.	667	453	429

Table 4: Elastic Net Results

coefficients, variables that add less explanatory power but remain relevant are preserved. Again a divide can be seen between the results for full sample and sub-sample excluding insignificants (columns (1) and (3) respectively) on the one hand, and results for the sub-sample excluding outliers on the other. For the former, the cross-validation procedure has selected a value for the blending parameter, ω, that is close to a RIDGE specification (i.e. less weight on the L_1 norm that defines LASSO). By contrast, the results in column (2) show that both parts of the penalty term are given similar weighting, indicating that this sample contains consistent evidence that some of the less important and highly correlated variables still add explanatory power. This translates into higher fitted R^2 values, as well as improved RMSEs for the no outlier sample. By contrast, the results in columns (1) and (3) remain strikingly similar to their LASSO counterparts, suggesting again that the outlying FTA estimates don't share the information from these variables in the same way.

Having found that the model estimated without outliers is again the most successful for out-of-sample prediction, I discuss its selected variables in greater depth as well[12]. As would be expected, the variables selected by the equivalent LASSO are a subset of the ones selected by the elastic net. Of the nine additional variables selected, 8 are additional FTA provision variables, and one is the total level of exporter Services AVEs. As was discussed above, the elastic net has the benefit that it helps distinguish between highly correlated variables. That FTA provisions were added lends support that this is indeed an important aspect to ex-ante estimation, given that the across-FTA variation is limited by the sample of FTAs (78, as reported in Table 1). The addition of total exporter barriers to individual barriers already present also serves to highlight that there are important subtleties in how AVEs affect FTA impact. Out-of-sample improvement therefore comes from picking up the more subtle information in these variables, as well as from relying less on the larger LASSO coefficients that are now penalised more heavily.

[12] A detailed list of selected variables is given in Appendix A

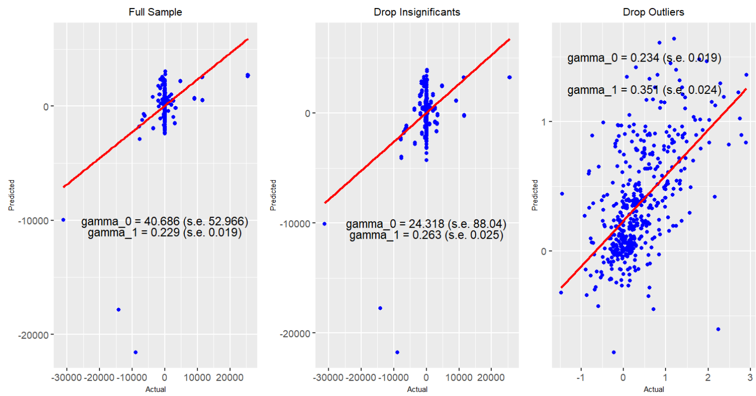


Figure 4: Comparison of first stage ex-post estimates with Elastic Net ex-ante estimates

4.2.2 Random Forest Estimates

The parametric models examined in the previous section found that using data without excluding outlier estimates led to poor out-of-sample predictive success. As was suggested above, this could suggest that the outlying observations are drawn from a separate data generating process which therefore increases the variance of the models compared to those that exclude them. An alternative explanation is that the underlying data generating process is more complex than that imposed by linear parametric assumptions.

Before discussing results from the random forest models, I briefly review the choice of $ntrees = 500$. As outlined in [Probst and Boulesteix, 2017], the choice of $ntrees$ presents a trade-off between accuracy and computational feasibility. As long as $ntrees$ is large enough for the out-of-bag error rate to converge, researchers should choose $ntrees$ to a computationally feasible level[13]. Figure 5 shows the out-of-bag error for models estimated on each data sample, as a function of $ntrees$. This shows clearly that $ntrees = 500$ is sufficient for error convergence on each sample and so the choice presents no issue of overfitting.

Results for the random forest models are displayed in Table 5. As can be seen, both the training RMSE and testing RMSE for the full sample and sub-sample dropping insignificants fall substantially. Whilst, the error size remains substantial, this lends immediate support to the possibility that the outliers are in reality outcomes from a more complex non-linear model. That the ability of the models to capture the outliers successfully is so greatly improved is emphasised by the fitted R^2 values that reach almost 80% of variation explained. This can further be seen by contrasting Figure 6 against Figures 3 and 4; the outliers that were so prominent are now greatly diminished. Whilst beyond the scope of

[13]Out-of-bag error refers to the ability for a model trained on a bootstrapped sample to predict observations outside of the sample

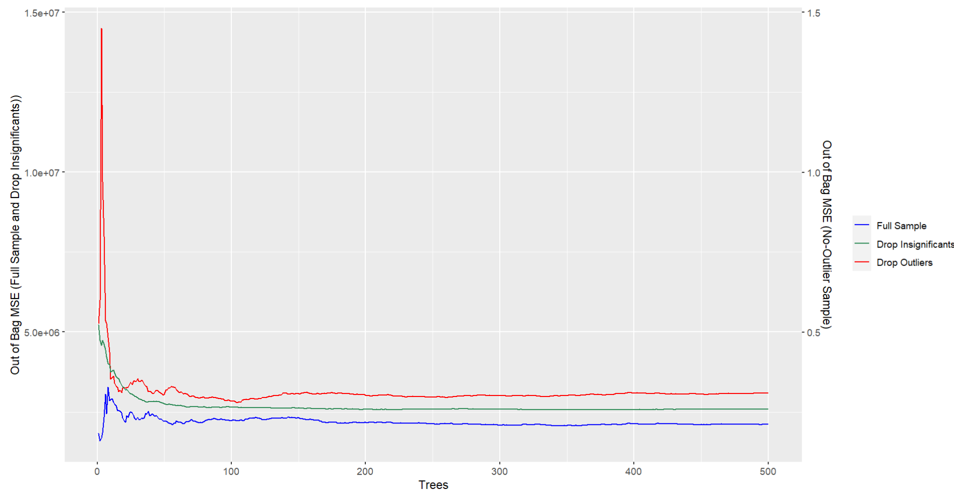


Figure 5: Random forest out-of-bag error for each sample

this paper, this suggests an interesting avenue for future research to investigate whether it is possible to further integrate these outliers into the same model as non-outlying estimates.

Nonetheless, the RMSEs for the sub-sample excluding outliers still improve substantially. As the results in column (2) show, this combination reaches the smallest RMSE, both in terms of the training data and testing data, of all models examined. The testing RMSE is only marginally better than the comparable elastic net result in column (2) of Table 4. This suggests that, whilst there there may be some benefit to removing the parametric assumption, the benefit is not large.

Data Sample	**Full Sample**	**Drop Outliers**	**Drop Insignificants**
	(1)	**(2)**	**(3)**
Training RMSE	695.61	0.223	902.629
Testing RMSE	1405.046	0.505	1722.395
Fitted R^2	0.799	0.427	0.777
Obs.	667	453	429

Table 5: Random Forest Results

It may be observed that, whilst the RMSEs for the sample excluding outliers are vastly superior than those for the alternative samples, the fitted R^2 is lower (although it should be noted that it still outperforms all parametric models). This occurs because of a bias that is found for the ex-ante estimates derived from the outlier free sample. This can be seen through the results for estimating equation (4), displayed as an annotation in Figure 6. The results for the intercept are significant for the sample excluding outliers, as they were for elastic net and LASSO models too. This suggests that the outliers are skewed compared to the wider estimates and that removing them discards important information. Related to this, the presence of heteroscedasticity can be seen visually in the right-hand panel which may be creating issues for the estimation. As above, this presents avenues for future research to investigate how this can be taken into account for improved prediction.

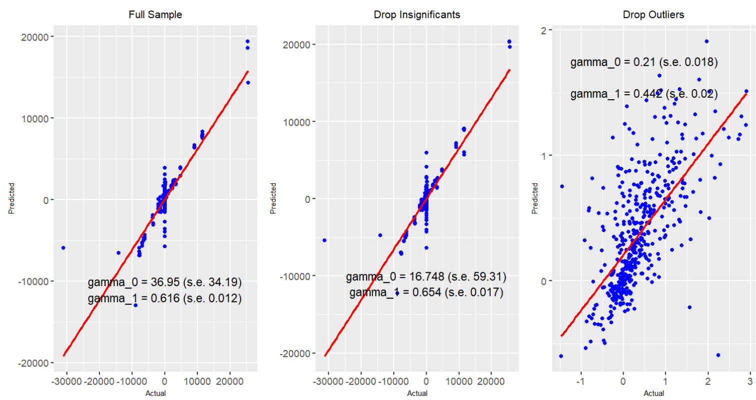


Figure 6: Comparison of first stage ex-post estimates with Random Forest ex-ante estimates

Again, given the superior performance of the model estimated with the outlying data, I explore the data further. Where the parametric models have coefficients that can be inspected for this purpose, by their nature this is not possible for non-parametric models. Instead, variable performance can be assessed by measuring how group purity increases when a variable is included in a model[14]. Variable importance results for the model are shown in figure 7[15]. The units for increase in node purity is mean square error within-group (i.e. trees which include distance as a variable have a mean square error for each of its groups which is on average roughly 11.3 lower than trees which don't include it). Out of the 34 variables displayed, 16 were selected as important by the previous elastic net model whilst 18 are newly selected. Interestingly, where FTA provision variables made up half the count of elastic net variables selected, only 5 are in the top 34 most important variables in the random forest. This includes 2 variables that are newly selected. Of the remaining newly selected variables, national account make up half the count, whilst additional forms of AVE or gravity variable not selected in the elastic net are also selected. This represents a shift in which variables are important, away from FTA provision variables and towards country-specific variables. However, on the whole the

[14]As discussed in section 3.2.2, non-parametric models aim to form groups of observations to optimise in-group homogeneity. Therefore, if a variable's inclusion allows greater purity within-group, it adds important information

[15]Whilst this measure is defined for all 1142 variables, for visualisation purposes, I restrict consideration to those variables which increase node purity by an average of at least 1.5.

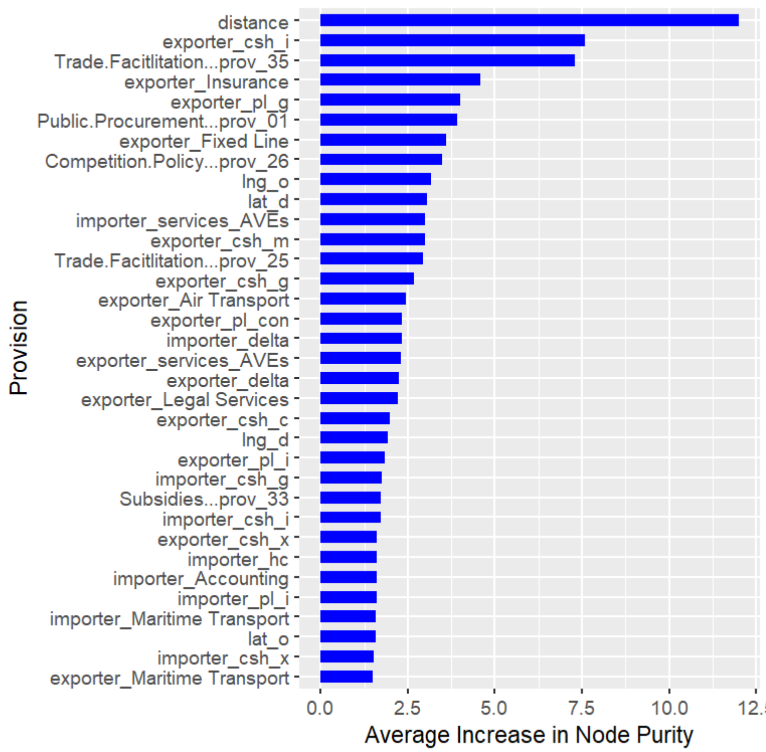


Figure 7: Random Forest Variable Importance

results continue to affirm the importance of the wider variable set used in this analysis. There is strong evidence from across the model specifications that wider national accounting variable than importer and exporter GDP matter for trade as well as for regulatory barriers that are in place in both exporter and importer countries. Furthermore, even if the random forest doesn't select FTA provisions within its most important explanatory variables, when considered together across provisions, the contribution is substantial.

As a final point, I highlight that these importance variables, as with the estimated coefficients for LASSO and elastic nets, can't be considered causal. Just because a variable contains information that helps to predict the ex-post effects from the first stage analysis, this can be for a variety of non-causal reasons. For example, Figure 6 shows that longitude and latitude for both exporter and importer country have explanatory power. This doesn't suggest that country's geographic position determines their trade, so much as it may reflect regional preferences for international trade evidenced in historic FTAs. Similarly, the importance of the national accounting variables may not be picking up an impact that having a particularly large portion of GDP going towards consumption causes trade, so much as certain types of economy may be able to afford a lot of consumption and that type may also trade more. This is not important for the concern of this paper, to examine the important factors to consider when making ex-ante estimates, but it does highlight that out-of-sample prediction may break down as the structure of the global economy shifts.

4.3 Illustrative FTA Examples: NAFTA

To give a concrete illustration of how the choice of data sample and model impact the success of the ex-ante estimates, I provide an example of how the ex-ante estimates made above compare with the original ex-post estimates. For this I use the particular example of NAFTA.

Table 6 shows the LASSO results comparison for the full sample and the sub-sample dropping outliers respectively. Each table shows the ex-post result ($\hat{\beta}$) and the ex-ante result ($\tilde{\beta}$) as well as a conversion of the coefficient into trade impacts in percentage terms. As can be seen, the results for the full sample demonstrate ex-ante estimates that take relatively extreme values when compared to the respective ex-post estimates. The exponential form of the gravity model then translates these extreme values into even higher trade impacts with all results taking either the extreme negative (100% removal of trade) or increasing trade by an incalculably large quantity. As expected, this picture is reversed for the results estimated on the sample with outliers removed. Whilst estimates and trade impacts are still large they become comparable with the respective ex-post estimates and take values which are not out of step with the existing gravity estimates of FTA impacts in the trade literature.

Table 7 provides the counterpart results for the elastic net and random forest models estimated without outliers. The were the two most successful models in terms of out-of-sample predictive success. As can be seen, there is still a non-negligible degree of error found in these estimates, but when compared to the equivalent sample results in Table 6, they represent a vast improvement.

5 Conclusion

This paper aimed to develop a systematic approach to estimating the impact of FTAs ex-ante. By extending a framework put forth by [Baier et al., 2019], three types of feature were identified as important to successful out-of-sample FTA impact estimation; data sample used, explanatory variables used, and analytical model of choice. It was found that this can be the difference between obtaining extreme ex-ante estimates, that are factors larger than existing FTA impact estimates in the literature, and reasonably credible results.

Of particular importance, the choice of FTA estimates used to train a model proved to be the largest factor in minimising RMSEs for results. Further improvements are gained through model choice; whilst random forest models showed the greatest predictive success amongst those considered here, elastic net models were not far behind. Furthermore, in all the results presented here,

			Full Sample				No Outliers			
	Ex-Post Estimate		min		BCCH		min		BCCH	
	$\%\Delta Trade$	$\hat{\beta}$	$\%\Delta Trade$	$\tilde{\beta}$	$\%\Delta Trade$	$\tilde{\beta}$	$\%\Delta Trade$	$\tilde{\beta}$	$\%\Delta Trade$	$\tilde{\beta}$
CAN $\Rightarrow$ MEX	230.037	1.194	-100	-349.195	-100	-586.045	125.828	0.815	151.923	0.924
CAN $\Rightarrow$ USA	84.856	0.612	∞	760.997	∞	688.962	151.314	0.922	79.47	0.585
MEX $\Rightarrow$ CAN	324.939	1.447	∞	1051.406	∞	582.618	226.576	1.183	350.05	1.504
MEX $\Rightarrow$ USA	330.953	1.461	-100	-686.066	-100	603.725	152.372	0.926	115.118	0.766
USA $\Rightarrow$ CAN	65.984	0.507	-100	-1103.67	-100	-282.751	108.571	0.735	173.429	1.006
USA $\Rightarrow$ MEX	71.116	0.537	-100	-576.265	-100	-1043.47	84.869	0.614	113.812	0.76

Table 6: LASSO NAFTA Results Comparison

			Elastic Net		Random Forest	
	Ex-Post Estimate		No Outliers		No Outliers	
	$\%\Delta Trade$	$\hat{\beta}$	$\%\Delta Trade$	$\tilde{\beta}$	$\%\Delta Trade$	$\tilde{\beta}$
CAN $\Rightarrow$ MEX	230.037	1.194	132.1004	0.842	75.76892	0.564
CAN $\Rightarrow$ USA	84.856	0.612	147.9359	0.908	104.0102	0.713
MEX $\Rightarrow$ CAN	324.939	1.447	226.7418	1.184	164.3226	0.972
MEX $\Rightarrow$ USA	330.953	1.461	147.1932	0.905	117.0592	0.775
USA $\Rightarrow$ CAN	65.984	0.507	111.277	0.748	93.47923	0.66
USA $\Rightarrow$ MEX	71.116	0.537	88.70221	0.635	105.0328	0.718

Table 7: Elastic Net and Random Forest NAFTA Results Comparison

evidence was found that the expanded set of variables used in estimation provided key information that was useful for improving results.

Furthermore, in light of the importance that excluding outliers was shown to hold, the random forest gave interesting results that it could better explain these outliers. Given that all of the ex-ante estimates trained without outliers exhibited some degree of bias, this suggests future research may make progress by investigating how these outliers may be integrated for improved prediction.

Whilst estimates obtained here are still exhibit an inability to capture a large part of the variation, this represents a marked improvement on the original paper. Furthermore, whilst many of the alternative approaches used in practice don't evaluate their success in terms of RMSE, this approach offers the promise to be able to compare and contrast ex-ante methods. Through further research, this may offer substantial improvements, as well as give a sense of confidence that is important to informed policy making.

References

[Abbyad and Herman, 2017] Abbyad, N. and Herman, P. (2017). How Much Do Non-Tariff Measures Cost? A Survey of Quantification Methods. Technical Report 2017–05–C, United States International Trade Commission, Washington, DC.

[Baier et al., 2019] Baier, S. L., Yotov, Y. V., and Zylkin, T. (2019). On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. *Journal of International Economics*, 116:206–226.

[Belloni et al., 2012] Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica*, 80(6):2369–2429. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9626.

[Breinlich et al., 2021] Breinlich, H., Corradi, V., Rocha, N., Ruta, M., Santos Silva, J. M. C., and Zylkin, T. (2021). Machine Learning in International Trade Research: Evaluating the Impact of Trade Agreements. Working Paper, World Bank, Washington, DC. Accepted: 2021-04-19T19:38:00Z.

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

[European Commission, 2019] European Commission (2019). Evaluation of the Implementation of the Free Trade Agreement between the EU and its Member States and the Republic of Korea. Technical report, European Commission, Brussels, Belgium.

[Feenstra et al., 2015] Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review*, 105(10):3150–3182.

[Fontagné et al., 2016] Fontagné, L., Mitaritonna, C., and Signoret, J. (2016). Estimated Tariff Equivalents of Services NTMs. Technical Report 2016-20, CEPII, Paris, France.

[Hoekman and Shepherd, 2017] Hoekman, B. and Shepherd, B. (2017). Services Productivity, Trade Policy and Manufacturing Exports. *The World Economy*, 40(3):499–516. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/twec.12333.

[Hofmann et al., 2017] Hofmann, C., Osnago, A., and Ruta, M. (2017). Horizontal Depth. Working Paper, World Bank, Washington, DC. Accepted: 2017-02-22T23:13:01Z.

[Jafari and Tarr, 2017] Jafari, Y. and Tarr, D. G. (2017). Estimates of Ad Valorem Equivalents of Barriers Against Foreign Suppliers of Services in Eleven Services Sectors and 103 Countries. *The World Economy*, 40(3):544–573. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/twec.12329.

[Kee and Nicita, 2022] Kee, H. L. and Nicita, A. (2022). Trade fraud and non-tariff measures. *Journal of International Economics*, 139:103682.

[Kee et al., 2009] Kee, H. L., Nicita, A., and Olarreaga, M. (2009). Estimating Trade Restrictiveness Indices. *The Economic Journal*, 119(534):172–199. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0297.2008.02209.x.

[Larch et al., 2023] Larch, M., Tan, S. W., and Yotov, Y. V. (2023). A simple method to ex-ante quantify the unobservable effects of trade liberalization and trade protection. *Journal of Comparative Economics*.

[London School of Economics, 2017] London School of Economics (2017). Ex-ante study of the eu-australia and eu-new zealand trade and investment agreements - final report. Technical report, European Commission, Brussels, Belgium.

[Mattoo et al., 2020] Mattoo, A., Rocha, N., and Ruta, M. (2020). *Handbook of Deep Trade Agreements*. World Bank, Washington, DC. Accepted: 2020-07-08T22:34:11Z.

[Petri et al., 2012] Petri, P. A., Plummer, M. G., and Zhai, F., editors (2012). *The Trans-Pacific Partnership and Asia-Pacific integration: a quantitative assessment*. Peterson Institute for International Economics, Washington, DC. OCLC: 810273691.

[Probst and Boulesteix, 2017] Probst, P. and Boulesteix, A.-L. (2017). To tune or not to tune the number of trees in random forest. *The Journal of Machine Learning Research*, 18(1):6673–6690.

[UK Department for International Trade, 2020] UK Department for International Trade (2020). The UK's Approach to Trade Negotiations with the US. Impact Assessment, United Kingdom Department for International Trade, London, United Kingdom.

[UK Department for International Trade, 2021] UK Department for International Trade (2021). UK-Australia FTA: Impact Assessment. Impact Assessment, United Kingdom Department for International Trade, London, United Kingdom.

[United States International Trade Commission, 2019] United States International Trade Commission (2019). U.S.-Mexico-Canada Trade Agreement: Likely Impact on the U.S. Economy and on Specific Industry Sectors. Impact Assessment 4889, United States International Trade Commission, Washington, DC.

[United States International Trade Commission, 2021] United States International Trade Commission (2021). Economic Impact of Trade Agreements Implemented under Trade Authorities Procedures. Methodology Paper 5199, United States International Trade Commission, Washington, DC.

A Appendix A

Variable Name	Estimated Coefficient	Variable Name	Estimated Coefficient
(Intercept)	1.337223	island_o	-0.09875
Trade.Facitlitation...prov_34	-0.00443	importer_cn	-9.2E-10
Trade.Facitlitation...prov_41	-3.1E-16	importer_delta	2.28191
Public.Procurement...prov_01	0.19489	importer_csh_i	-1.72184
Public.Procurement...prov_19	-0.15946	importer_csh_m	-0.22998
Public.Procurement...prov_58	-0.00224	importer_csh_r	0.98709
Public.Procurement...prov_100	-0.22067	importer_pl_n	0.142455
Movement.of.Capitals...prov_77	-0.1536	exporter_hc	0.017377
Subsidies...prov_33	-0.19008	exporter_delta	10.4823
Subsidies...prov_43	0.238616	exporter_xr	1.45E-05
Migration...prov_6	0.162185	exporter_csh_i	-2.70923
Migration...prov_16	0.011763	exporter_pl_i	-0.12949
SPS...prov_39	0.020287	exporter_pl_g	-0.38308
SPS...prov_42	0.026655	exporter_pl_n	-0.17163
Services...mov_prov_cov	0.104861	importer_Accounting	-0.00053
Competition.Policy...prov_07	-0.00703	importer_Road Transport	-0.0044
Competition.Policy...prov_26	-0.12608	importer_Banking	-0.00488
Rules.of.Origin...roo_ver_two	-0.03641	importer_services_AVEs	-0.00264
Rules.of.Origin...roo_vcr_prc	-0.0498	exporter_Accounting	-0.00046
Rules.of.Origin...roo_vcr_fob	0.094304	exporter_Air Transport	0.005561
distance	4.35E-05	exporter_Banking	-0.00307
member_eu_joint	-0.21656	exporter_Mobile Line	-0.02455
lng_o	-0.00056	exporter_Retail	0.005423
exporter_Maritime Transport	-0.00934		

Table 8: Cross-Validation LASSO Selected Variables (No Outlier Sample)

Variable Name	**Estimated Coefficient**	**Variable Name**	**Estimated Coefficient**
(Intercept)	1.377786	distance	4.17E-05
Trade.Facitlitation...prov_34	-0.00012	member_eu_joint	-0.22055
Trade.Facitlitation...prov_41	-1.97E-06	lng_o	-0.00057
Anti.dumping...ad3.e.4	-3.18E-08	island_o	-0.09899
Anti.dumping...ad3.n.3	-0.00653	importer_cn	-1.45E-09
Countervailing.duties...cvd3_b_3	-0.00973	importer_delta	2.594102
TBT...prov_27	-0.01627	importer_csh_i	-1.65804
Public.Procurement...prov_01	0.186689	importer_csh_m	-0.20865
Public.Procurement...prov_19	-0.11042	importer_csh_r	0.962144
Public.Procurement...prov_58	-0.00704	importer_pl_n	0.162407
Public.Procurement...prov_100	-0.18599	exporter_hc	0.020115
IPR...prov_09	-0.02758	exporter_delta	9.949086
Movement.of.Capitals...prov_77	-0.12496	exporter_xr	1.20E-05
Subsidies...prov_33	-0.15929	exporter_csh_i	-2.74353
Subsidies...prov_43	0.125966	exporter_pl_i	-0.16368
Migration...prov_6	0.162401	exporter_pl_g	-0.35067
Migration...prov_16	0.110202	exporter_pl_n	-0.19375
Migration...prov_27	0.012892	importer_Accounting	-0.00067
SPS...prov_20	-0.00255	importer_Road Transport	-0.00425
SPS...prov_21	-0.00303	importer_Banking	-0.00465
SPS...prov_39	0.046001	importer_services_AVEs	-0.00259
SPS...prov_42	0.023211	exporter_Accounting	-0.00072
Services...mov_prov_cov	0.095704	exporter_Air Transport	0.005194
Competition.Policy...prov_07	-0.05094	exporter_Banking	-0.00378
Competition.Policy...prov_26	-0.12073	exporter_Mobile Line	-0.02352
Rules.of.Origin...roo_ver_two	-0.1072	exporter_Retail	0.005418
Rules.of.Origin...roo_vcr_prc	-0.06233	exporter_Maritime Transport	-0.00834
Rules.of.Origin...roo_vcr_fob	0.076252	exporter_services_AVEs	-4.46E-05

Table 9: Elastic Net Selected Variables (No Outlier Sample)