Online Robustness Appendix for:

Quantitative Trade Models in *Annual Review of Economics*

Timothy J. Kehoe, Pau S. Pujulàs, Jack Rossbach

February 2017

Robustness Appendix

This appendix accompanies the main paper and explores the robustness of our results.

A.1 Log-Differences versus Percentage Changes

When evaluating the accuracy of the GTAP predictions in Table 1, we defined growth in terms of percentage changes. When evaluating the accuracy of the C-P predictions in Table 2, we defined growth in terms of log differences. There are two reasons we chose to use a different definition for growth between the two tables. First, we chose each definition to be consistent with the definition used by the original papers. GTAP papers use percentage changes, while C-P use log differences. Second, it shows that the overall shortcomings of AGE models does not appear dependent on how exactly we define growth (percentage changes or log differences).

To further explore the robustness of our results, we show that Tables 1 and 2 are overall similar when the definitions of growth are switched between them. Note that we can move between percentage changes and log differences with the following equations.

$$\text{log diff} = 100 * \log\left(\frac{\%\text{ change}}{100} + 1\right), \tag{A.1}$$

$$\%\text{ change} = 100 * \left(\exp\left(\frac{\text{log diff}}{100}\right) - 1\right), \tag{A.2}$$

Below are the tables show the results of this exercise.

Table A.1: Evaluation of GTAP Predictions (with Log Differences)

Exporter	Importer	GTAP Correl.	LTP Correl.
United States	Australia	0.63	0.49
Australia	United States	–0.18	0.45
United States	Chile	0.32	0.56
Chile	United States	0.15	0.44
China	Chile	0.51	0.78
Chile	China	–0.09	–0.08
China	New Zealand	–0.38	0.56
New Zealand	China	–0.17	0.30
Simple Average		0.10	0.44

Overall, Table A.1 shows that the LTP correlation stays almost the same as in Table 1 in the paper. The biggest changes are that GTAP appears to perform better for Chinese exports to Chile and U.S. exports to Australia when expressed in log-differences; whereas Chilean exports to China perform worse. Overall, the average LTP correlation is similar across the two tables (0.44 compared to 0.49 in Table 1). The average GTAP correlation is slightly higher with log-differences (0.10 compared to –0.00), however, it still performs significantly worse than the LTP methodology with both definitions.

Table A.2: Evaluation of NAFTA Predictions (with Percentage Changes)

Exporter	Importer	C-P Correl.	C-P Correl. (Only NAFTA Tariffs)	C-P Correl. (No IO Structure)	LTP Correl.
Canada	Mexico	−0.47	−0.47	−0.42	0.34
Canada	United States	0.63	0.62	0.64	0.05
Mexico	Canada	−0.46	−0.43	−0.49	0.76
Mexico	United States	−0.11	−0.10	−0.16	0.14
United States	Canada	0.38	0.16	0.20	0.20
United States	Mexico	0.96	0.96	0.97	−0.02
Simple Average		0.16	0.12	0.12	0.28

The results in Table A.2 are likewise similar to Table 2, with the LTP methodology outperforming the C-P predictions. By far the biggest change is for U.S. exports to Mexico, which exhibit negative correlation for LTP and near perfect correlation for C-P. This result is driven entirely by growth in exports in the petroleum industry, which grew by over 1500 percent between 1991 and 2006 (the second most growth was in chemicals, which grew by 200 percent). Despite this huge growth, it actually grew less than predicted, as C-P predicted an increase in growth of nearly 3000 percent (the second highest predicted growth was in electrical machinery, at nearly 200 percent). The reason C-P predict this large increase is due to a much higher estimated trade elasticity for the petroleum industry than for any other industry (over 50, no other industry elasticity was over 20). If the petroleum industry is dropped when evaluating the performance of each methodology, then the table appears as below in Table A.3.

Table A.3: Evaluation of NAFTA Predictions
(with Percentage Changes / Petroleum excluded for US Exports to Mexico)

Exporter	Importer	C-P Correl.	C-P Correl. (Only NAFTA Tariffs)	C-P Correl. (No IO Structure)	LTP Correl.
Canada	Mexico	−0.47	−0.47	−0.42	0.34
Canada	United States	0.63	0.62	0.64	0.05
Mexico	Canada	−0.46	−0.43	−0.49	0.76
Mexico	United States	−0.11	−0.10	−0.16	0.14
United States	Canada	0.38	0.16	0.20	0.20
United States	Mexico	−0.37	−0.32	−0.19	0.45
Simple Average		−0.06	−0.09	−0.07	0.35

As we can see, simply excluding the petroleum industry changes the correlation between their base model and actual growth for US exports to Mexico from +0.96 to −0.37. One of the effects of log differences is that it lessens the overall influence of extreme growth rates for individual industries compared to using percentage changes.

A.2 Effects of Outlier Observations

When evaluating the GTAP predictions we have excluded two industries, petroleum exports from the United States to Chile, and cattle meat exports from Australia to the United States. Petroleum grew from 1.9 percent of U.S. exports to Chile to 33.6 percent of U.S. exports to Chile between 2002 and 2015. This growth in exports was due to innovations in hydraulic fracturing techniques, which led to greatly increased oil production and exports in the United States and has little to do with changes in tariffs (there was little change in U.S. petroleum exports to Australia, and so it is not excluded as an outlier).

Australian exports of cattle meat to the United States are also excluded as an outlier industry. Cattle meat was the top export from Australia to the United States in both 2002 and 2015; however, a tariff rate quota has remained in place even after the signing of the trade agreement. Australia has continually used 99 percent of its quota, which has increased by approximately 10 percent between 2002 and 2015. Growth in the trade value of cattle meat exports from Australia to the United States was driven primarily by increases in the price of beef, which was due in part to the spread of bovine spongiform encephalopathy, commonly known as mad cow disease. Mad cow disease was first discovered in the United States in 2003 (it has not been discovered in Australia), and between 2002 and 2015, the average price of cattle meat imports from all countries to the United States increased by 123 percent compared with just a 32 percent increase in the U.S. Consumer Price Index over the same time frame. In fact, cattle meat imports by net weight were actually lower throughout 2003–2014 than they were in 2002, returning to their pre mad cow disease levels only in 2015.

Table A.4 shows that, while excluding these industries does affect the correlations for each of those importer exporter pairs, removing them has little effect on the overall performance of the GTAP predictions and LTP methodology, with the LTP methodology performing substantially better even when these outlier industries are not removed from the analysis.

Table A.4: Evaluation of GTAP Predictions (Outliers Left In)

Exporter	Importer	GTAP Correl.	LTP Correl.
United States	Australia	0.27	0.55
Australia	United States	0.49	0.08
United States	Chile	0.05	–0.05
Chile	United States	0.03	0.48
China	Chile	0.14	0.61
Chile	China	0.04	0.07
China	New Zealand	–0.36	0.61
New Zealand	China	–0.09	0.48
Simple Average		0.07	0.35

Note that including cattle meat exports for Australian exports to the United States makes the GTAP predictions appear to perform much better for that pair, however, it is a classic case of getting things right for the wrong reason. GTAP predicts a large increase in the quantity of beef exports, due to a complete removal of the tariff quota (a mentioned shortcoming of our methodology). In actuality, the trade agreement between Australia and the United States left the quota largely intact, and the increase in trade value was due to a worldwide increase in the price of beef. If we instead ran GTAP without simulating a complete elimination of all trade barriers (a complete removal is a reasonable approximation for the other industries and trade agreements we evaluate), it would be unable to capture to capture the increased value of exports of cattle meat.

A.3 Weighted versus Unweighted Correlation

Another choice we made was to use a weighted correlations when evaluating the models, which has the effect of placing more importance on some industries compared to others. Our view is that it is natural to employ weighted correlations, since policy makers tend to care more about how trade agreements will impact industries that exhibit large amounts of trade compared to industries that exhibit very little trade overall. Perhaps in part because it is not always clear which weighting scheme should be used, it is also common to use standard unweighted correlations when evaluating the performance of models. The tables below show how our evaluation of the GTAP model and C-P model is impacted by not weighting industries.

Table A.5: Evaluation of GTAP Predictions (Unweighted Correlations)

Exporter	Importer	GTAP Correl.	LTP Correl.
United States	Australia	–0.01	0.35
Australia	United States	–0.04	0.65
United States	Chile	–0.03	0.42
Chile	United States	0.00	0.27
China	Chile	0.42	0.37
Chile	China	–0.07	0.13
China	New Zealand	–0.17	0.37
New Zealand	China	–0.14	0.26
Simple Average		–0.00	0.35

Table A.5 shows that there are small changes for most of the GTAP country pairs, with the largest improvement in the performance of the GTAP models coming for China's exports to Chile (increasing from 0.14 to 0.42). Overall, however, the average correlation between the GTAP predictions and actual changes remains approximately zero, while the LTP methodology performs substantially better regardless of whether industry weighting is used or not.

Table A.6: Evaluation of NAFTA Predictions (Unweighted Correlations)

Exporter	Importer	C-P Correl.	C-P Correl. (Only NAFTA Tariffs)	C-P Correl. (No IO Structure)	LTP Correl.
Canada	Mexico	–0.23	–0.22	–0.23	0.21
Canada	United States	0.62	0.57	0.57	0.23
Mexico	Canada	–0.52	–0.52	–0.50	0.27
Mexico	United States	0.04	0.06	0.05	0.27
United States	Canada	0.47	0.19	0.24	0.12
United States	Mexico	0.70	0.69	0.76	–0.05
Simple Average		0.18	0.13	0.15	0.18

We then evaluate the C-P predictions using unweighted correlations. The results in Table A.6 appear in line with Table A.2 (even though Table A.2 uses log differences) and slightly better than Table 2. The LTP predictions, however, appear to perform significantly worse than before and only on par with the C-P correlations, which is surprising given the otherwise robustness of their performance. Closer inspection reveals this shortcoming is primarily due to poor performance for the LTP methodology on industries with very little trade. If we compute the unweighted correlations excluding industries that do not account for on average 1 percent of trade in 1991 and 2006, we find the LTP methodology performs significantly better while the performance of the C-P predictions are largely unchanged. The results of this exercise are reported below in Table A.7.

Table A.7: Evaluation of NAFTA Predictions
(Unweighted Correlations; Industries Averaging >1% of Trade in 1991 and 2006)

Exporter	Importer	C-P Correl.	C-P Correl. (Only NAFTA Tariffs)	C-P Correl. (No IO Structure)	LTP Correl.
Canada	Mexico	–0.27	–0.26	–0.24	0.53
Canada	United States	0.62	0.59	0.58	0.30
Mexico	Canada	–0.56	–0.54	–0.53	0.69
Mexico	United States	0.07	0.09	0.06	0.27
United States	Canada	0.47	0.19	0.24	0.12
United States	Mexico	0.73	0.72	0.80	0.01
Simple Average		0.18	0.13	0.15	0.32

Why exactly the LTP methodology performs poorly for capturing changes in industries with small amounts of trade is not clear and merits further exploration in future research.

One final topic of exploration is why the C-P correlations perform better with unweighted correlations than with weighted correlations. We find that these improvements are driven primarily by a few industries with extreme changes (similar to the results in Tables A.2 and A.3). Compared to the C-P predictions, the LTP methodology appears less driven by extreme movements. In the below table we exclude industries that we define as exhibiting extreme changes in growth (log differences less than -100 or greater than 200).

Table A.8: Evaluation of NAFTA Predictions
(Unweighted Correlations; Industries Averaging >1% of Trade in 1991 and 2006; Industries with Extreme Growth Rates Excluded)

Exporter	Importer	C-P Correl.	C-P Correl. (Only NAFTA Tariffs)	C-P Correl. (No IO Structure)	LTP Correl.
Canada	Mexico	–0.23	–0.22	–0.27	0.34
Canada	United States	0.64	0.63	0.64	0.18
Mexico	Canada	–0.50	–0.53	–0.45	0.48
Mexico	United States	0.06	0.11	0.00	0.18
United States	Canada	0.47	0.19	0.24	0.12
United States	Mexico	–0.37	–0.32	–0.15	0.39
Simple Average		0.01	–0.02	0.00	0.28

Table A.8 shows that in this case, the LTP methodology performs nearly identically to before, however the C-P methodology performs much worse. The impact of industries with extreme changes on the results are less evident when we use weighted correlations partly since industries with large trade flow values appear less likely to exhibit extreme movements — the petroleum example from Tables A.2 and A.3 withstanding.

A.4 Different Base Years

A final concern might be that our choice of base year might drive our result. In the paper we chose base years several years before the trade agreements were actually implemented. This is due to the observation that trade flows often begin reacting to a trade agreement before the trade agreement is actually implemented. Below we show our results are robust to alternative base years. In the main paper we evaluated the GTAP predictions using 2002 as our base period and 2015 as our end period. Below we change the base period to 2004, which immediate precedes many of the trade agreements. When determining the set of least traded products, we still sort over three years (2002–2004) to avoid classifying products with lumpy trade as least traded products, although this has little effect on our overall results compared if we sorted only over trade in 2004. Table A.9 below reports the results of the GTAP evaluation when we use 2004 as our base year.

Table A.9: Evaluation of GTAP Predictions (2004–2015)

Exporter	Importer	GTAP Correl.	LTP Correl.
United States	Australia	0.11	0.55
Australia	United States	–0.16	0.40
United States	Chile	0.25	0.48
Chile	United States	0.01	0.45
China	Chile	0.11	0.36
Chile	China	0.05	0.14
China	New Zealand	–0.09	0.74
New Zealand	China	–0.11	0.63
Simple Average		0.02	0.47

Table A.9 shows there are few major changes when we change the base year. Overall, the GTAP predictions and LTP predictions perform nearly identically with a base year of 2004 as they do with the base year of 2002 used in Table 1 of the paper.

Table A.10: Evaluation of NAFTA Predictions (1993–2006)

Exporter	Importer	C-P Correl.	C-P Correl. (Only NAFTA Tariffs)	C-P Correl. (No IO Structure)	LTP Correl.
Canada	Mexico	0.09	0.07	0.05	0.33
Canada	United States	0.32	0.28	0.34	0.14
Mexico	Canada	–0.85	–0.83	–0.89	0.74
Mexico	United States	0.27	0.29	0.29	–0.17
United States	Canada	0.39	0.10	0.14	0.26
United States	Mexico	0.50	0.48	0.60	0.18
Simple Average		0.12	0.07	0.09	0.25

We can also repeat the evaluation of the C-P predictions using a different base year. In the paper we used a base year of 1991, well before the implementation of NAFTA, and Table A.10 shows the results for using 1993 as a base year (again trade is sorted over 1991–1993 when defining the set of least traded goods). The results for the various importer export pairs are indeed affected by the change in base year, which highlights the changes in trade flows that occurred in the lead up to NAFTA's implementation (although NAFTA was enacted in 1994, negotiations were largely complete well before this, a fact that is highlighted by the ceremonial signing of NAFTA in 1992). For example, for exports from Mexico to the United States, the unweighted correlation between industry trade flows in 1991 and 1993 is only 0.87. This correlation is even lower than the correlation between trade flows in 1993 and 1995, the years immediately preceding and succeeding the implementation of NAFTA. Regardless of the base year, our exercise shows that the LTP methodology still outperforms the C-P predictions on average.

While this appendix shows that our overall results are largely unchanged by outliers, base years, weighting schemes, and how we define growth; it also shows that for some individual country pairs these choices substantially alter the apparent accuracy of various predictions and methodologies. Thus, it is important for researchers and policy makers to think carefully about how the performance of their models should be evaluated. Beyond that, it is important that we carry out such evaluations, so that shortcomings can be identified and we can improve the models.

A.5 Level Predictions for the LTP Methodology and an Alternative Measure of Accuracy

While we have focused on using weighted correlations with data to evaluate model predictions for the industry level impact of trade reform, there are many other metrics that might be used to evaluate the predictions of models. Correlations are simple to understand, capture our point well that current models fail to predict which industries will be impacted most and least by trade reform, and are robust to a wide range of changes, as illustrated by this appendix. Correlations do not, however, tell us much about how the predicted level of growth for an industry compared to the actual level of growth for that industry. To address this drawback, we evaluate the accuracy of the

AGE models by computing the absolute value of the percentage difference between the growth of each industry in the data and the predicted growth. We take the absolute value, so over predicting growth in one industry does not offset under predicting growth in another industry.

To compute our new measure of accuracy, we first generate predictions for the level of growth for each industry based on the LTP methodology. We do this following Kehoe et al. (2015). In particular, we predict that the growth rate in each industry will be a linear function of the share of least traded products in the industry

$$\hat{z}_{ijk}^{t,t'} = \alpha_{ij}^{t,t'} + \beta_{ij}^{t,t'} s_{ijk}^{t}, \tag{A.3}$$

where $\alpha_{ij}^{t,t'}$ is our predicted growth rate for nonLTP and $\alpha_{ij}^{t,t'} + \beta_{ij}^{t,t'}$ is our predicted growth rate for nonLTP, and s_{ijk}^{t} is the export share of LTP for industry k in the base period. Note that any parameterization of (A.3) that has $\beta_{ij}^{t,t'} > 0$ will generate identical weighted correlations with the data, which is why we did not need to generate such predictions to compute our earlier tables.

Before estimating each of the coefficients, we first estimate a predicted growth rate for total exports, $\hat{z}_{ij}^{t,t'}$, by using the tariff rate in the base period, an estimated trade elasticity of -3.511 from Kehoe et al. (2015), and the assumption that tariffs fall to zero following trade liberalization. We then pin down the coefficients with two restrictions. First, we require that the overall growth rate of trade implied by our coefficients equals our prediction for the overall increase in trade

$$\alpha_{ij}^{t,t'} + 0.1 \times \beta_{ij}^{t,t'} = \hat{z}_{ij}^{t,t'}, \tag{A.4}$$

where the 0.1 represents the 10 percent of overall trade accounted for by LTP in the base period. Second we require that the ratio of our predicted growth rate for LTP relative to our predicted growth rate for overall trade is consistent with the average ratio for the growth rate of LTP relative to the overall growth for exports across countries, γ, in the data

$$\frac{\alpha_{ij}^{t,t'} + \beta_{ij}^{t,t'}}{\alpha_{ij}^{t,t'} + 0.1 \times \beta_{ij}^{t,t'}} = \gamma, \tag{A.5}$$

where we take our estimate of $\gamma = 3.591$ from Kehoe & Ruhl (2013), who calculate γ using bilateral trade flows for all countries with available data between 1995 and 2005. These two equations pin down our estimated coefficients, which we use to generate predicted growth for each industry according to (A.3).

After estimating level predictions for each industry, we compute their accuracy using the weighted mean absolute percentage difference, $A_{ij}^{t,t'}$, between the predictions and the data. This is done by first compute the absolute percentage difference between each prediction and the actual growth in the data, then taking the weighted average across industries, where the weights are the same as for the weighted correlations, given by equation (5) in the paper. The formula for $A_{ij}^{t,t'}$ is

$$A_{ij}^{t,t'} = \sum_{k=1}^{K} \omega_{ijk}^{t,t'} \left| \frac{z_{ijk}^{t,t'} - \hat{z}_{ijk}^{t,t'}}{\hat{z}_{ijk}^{t,t'}} \right|. \tag{A.6}$$

Table A.11: Mean absolute percentage difference for GTAP and LTP predictions with data

		Estimated coefficients for LTP predictions		Accuracy of predictions (Mean absolute % difference)	
Exporter	**Importer**	α	β	**GTAP**	**LTP**
United States	Australia	11.82	47.77	1161.4	303.1
Australia	United States	10.13	40.94	5239	515.9
United States	Chile	19.03	76.95	102237.3	468.8
Chile	United States	7.33	29.64	14399.4	615.5
China	Chile	19.09	77.19	556.9	538.6
Chile	China	33.74	136.40	52497.1	892.2
China	New Zealand	10.68	43.17	422.7	302.5
New Zealand	China	33.87	136.94	9241.8	934.5
Average accuracy of predictions				23219.5	571.4

Table A.11 reports the estimated coefficients for the LTP predictions for each of the eight country pairs in our GTAP evaluations, along with the weighted mean value across industries of the absolute percentage difference between the model predictions and the data for both the GTAP model predictions and the LTP predictions. For these results, trade flows in the data are deflated by exporter GDP to partially isolate changes in trade flows due to the trade reform, from increases in trade flows due to other factors, such as population and TFP growth. This does not affect our results for the earlier tables with correlations, but it does matter when computing absolute percentage differences. Our results show that the LTP methodology outperforms the GTAP methodology, often substantially, for each of the free trade agreements evaluated.

Table A.12 reports the estimated coefficients for the LTP methodology and our accuracy measure, the weighted mean absolute percentage difference, for both the CP predictions and the LTP predictions for each of the NAFTA country pairs. We find that the LTP predictions outperform the CP predictions on average and for each of the country pairs except for Mexican exports to the United States. Together with our other results, we take these findings as strong evidence of the importance of incorporating the LTP margin into models when developing predictions for the industry level impact of trade liberalization.

Table A.12: Mean absolute percentage difference for CP and LTP predictions with data

		Estimated coefficients for LTP predictions		Accuracy of predictions (Mean absolute % difference)	
Exporter	**Importer**	α	β	**CP**	**LTP**
Canada	Mexico	41.08	166.06	4438.3	162.2
Canada	United States	4.42	17.87	2004	839
Mexico	Canada	18.68	75.52	336.8	253.6
Mexico	United States	15.42	62.32	232	546.3
United States	Canada	7.84	31.70	800.8	171.8
United States	Mexico	40.56	163.96	151	41.3
Average accuracy of predictions				1327.15	335.7