

Online Appendix to:
Using the New Products Margin to Predict the Industry-Level Impact of Trade Reform
Kehoe, Rossbach, and Ruhl
December 2014

Section A: Data

1. Product Level Trade Data

1.1 KORUS Trade Data

1.1.1A Bilateral KORUS Trade Data

The data covering bilateral trade, by 5-digit SITC Rev. 2, between the United States and Korea are from COMTRADE. The data are for Korea-United States and United States-Korea (exporter-importer). They are reported by the importer, and cover 2005-2007.

Original File:

The file is “korus_bilateral0507.csv”

1.1.1B KORUS Trade with World

The data covering bilateral and world trade for Korea and the United States, by 5-digit SITC Rev. 2, are from COMTRADE. The data for Korean trade covers both imports and exports between Korea and the United States and Korea and the World, with all data reported by Korea, covering the period 2005-2007. The data for U.S. trade covers both imports and exports between the United States and Korea and the United States and the World, with all data reported by the United States, covering the period 2005-2007.

Original Files:

The file for Korean trade is “KOREA_0507.csv”

The file for U.S. trade is “USA_0507.csv”

1.1.2 Processing

The observations are listed by importer and exporter by using the reporter, partner, and trade flow from the original data. Additionally, the four digit leaf codes are transformed into 5 digit codes by adding a zero onto the end.

The bilateral KORUS trade data is formatted using the stata program “FormatData_KORUS.do” and creates the file “KORUS.dta”

The trade data for Korea’s trade with the United States and the World is formatted using the Stata program “FormatData_KORUS_KOREA_WORLD.do” and creates the file “KOREA_WORLD.dta”

The trade data for the United States’ trade with Korean and the World is formatted using the Stata program “FormatData_KORUS_USA_WORLD.do” and creates the file “USA_WORLD.dta”

1.1.3 Processed File Structure

Each observation in the processed datasets (“KORUS.dta”, “KOREA_WORLD.dta”, “USA_WORLD.dta”) is the reported flow of goods in a 5-digit commodity code during a certain year between the importer and the exporter. The columns of the data are, in order: the importer, the exporter,

the SITC commodity code (with leading zeros suppressed), the year in which the trade occurred, and the value of the trade flow (in thousands of USD).

The COMTRADE country codes are used for exporter and importer. 410 is Korea, 842 is the United States, and 0 is the World. A complete list of COMTRADE country codes is available here: <http://unstats.un.org/unsd/tradekb/Knowledgebase/Comtrade-Country-Code-and-Name>

1.2 NAFTA Trade Data

1.2.1A Bilateral NAFTA Trade Data

The data covering bilateral trade, by 5-digit SITC Rev. 2, between the United States, Canada, and Mexico are from COMTRADE.

The data are for Canada-Mexico, Canada-United States, United States-Canada, Mexico-Canada, Mexico-United States, and United States-Mexico. They are reported by the importer, and cover 1988-2009 (although we use 1989 as our base period and 2009 as our end period).

Original Files:

The naming structure is NNNMMMYYZZ.csv with the exporting country (NNN) the importing country (MMM) the first year of the data (YY) and the ending year of the data (ZZ).

canmex8809.csv	mexcan8809.csv	mexusa8809.csv
canusa8809.csv	usacan8809.csv	usamex8809.csv

1.2.1B Canadian Trade Data with the World

The data covers Canada's total exports and imports with the United States, Mexico, and the World as a partner, by 5-digit SITC Rev. 2. The data covers the period 1988-2009, and all trade data is reported by Canada.

Original File:

The file is "Canada_CoxHarris.csv"

1.2.2 Processing

The observations are listed by importer and exporter by using the reporter, partner, and trade flow from the original data. Additionally, the four digit leaf codes are transformed into 5 digit codes by adding a zero onto the end.

The data from the individual bilateral trade files are combined into a single stata file "NAFTA.dta" using the Stata program "FormatData_NAFTA.do"

The data for Canadian trade with Mexico, the United States, and the World is formatted using the Stata program "FormatData_NAFTA_CoxHarris.do" and produces the file "NAFTA_CoxHarris.dta"

1.2.3 Processed File Structure

Each observation in the processed datasets ("NAFTA.dta", "NAFTA_CoxHarris.dta") is the reported flow of goods in a 5-digit commodity code during a certain year between the importer and the exporter. The columns of the data are, in order: the importer, the exporter, the SITC commodity code (with leading zeros suppressed), the year in which the trade occurred, and the value of the trade flow (in thousands of USD), the quantity of trade (volume), the quantity unit code (which tells us what units quantity is in), and

the net weight in kg – which is equal to quantity if quantityunitcode is weight in kilograms (code 8) and zero otherwise.

The descriptions for each quantityunitcode are included in the file, and are also available here:

<http://unstats.un.org/unsd/tradekb/Knowledgebase/UN-Comtrade-Reference-Tables>

The COMTRADE country codes are used for exporter and importer. 124 is Canada, 484 is Mexico, 842 is the United States, and 0 is the World. A complete list of COMTRADE country codes is available here: <http://unstats.un.org/unsd/tradekb/Knowledgebase/Comtrade-Country-Code-and-Name>

2. SITC-to-ISIC Concordance

2.1 Original files

We adapt the 4-digit SITC (Rev.2) to 3-digit ISIC (Rev.2) created by MA Muendler. The file is “sitc2isic.dta” and is available with documentation here:

<http://dss.ucsd.edu/~muendler/html/resource.html#sitc2isic>

The file “SITC_Leaf_Codes.txt” contains a list of all SITC (Rev.2) leaf commodity codes, which is all 5 digit codes as well as all 4 digit codes without any 5 digit sub-codes. Those 4 digit codes have been converted to 5 digit codes by adding a zero onto the end.

There are 27 5-digit SITC (Rev.2) codes that are not listed in the concordance or are listed as “Unclassified.” These goods we classify by hand according to their corresponding ISIC2 industry, with the exception of SITC code 91100: “Postal packages not classified according to kind”, which we leave as unclassified and exclude from our analysis. The list of the 27 SITC codes and the ISIC2 codes we assigned to them is listed in “sitc2isic_missing.dta”.

2.2 Processing

We create 5-digit SITC (Rev.2) to 3-digit ISIC (Rev.2) concordance by replacing non-leaf 4-digit codes with all of their 5-digit subcodes. This is done using the stata program “FormatConcordance.do”, which creates the file “SITC5_to_ISIC3.dta”

2.3 Processed File Structure

The first column has SITC leaf codes (leading zeros suppressed, 4-digit codes converted to 5-digits by adding a zero onto the end) and the second column has the ISIC category to which the SITC code contributes.

3. Miscellaneous Data

3.1 Nominal GDP in USD

Nominal GDP in USD for Canada, Mexico, and the United States, over the years 1988-2009 is constructed using Nominal GDP in NCU and converting it to USD using the period average exchange rate which is from IMF IFS. Nominal GDP in USD for the World is downloaded from WDI Online. The original file and calculations are in “IFS GDP.xls” and the processed file is “GDP_input.txt”

3.2 ISIC-to-Industry Concordance

The file “isic2db.csv” contains the concordance from 3-digit ISIC (Rev.2) industry codes to the DBS industries.

Based on information from here: <http://www.fordschool.umich.edu/rsie/model/versions.html#sectors>

The file “isic2other.csv” contains our constructed concordances from 3-digit ISIC (Rev.2) industry codes to the Cox-Harris model industries, the Sobarzo model industries, the Kiyota-Stern model industries, and the Yaylaci-Shikher model industries.

The Cox-Harris and Sobarzo concordances are adapted from the concordances listed in Kehoe (2003): <http://www.econ.umn.edu/~tkehoe/papers/NAFTAevaluation.pdf>

Kiyota and Stern do not provide a concordance, so we develop one. Yaylaci and Shiker provide a concordance between their industries and 2-digit ISIC rev. 3 codes, which we adapt to be in terms of 3-digit ISIC rev. 2 codes.

3.3 KORUS Projections (BDS, Cox-Harris, and Sobarzo)

The file “KORUS_KS_input.txt” contains the projections for the effects of KORUS on Korean and U.S. trade with the World from Kiyota and Stern’s paper “Economic Effects of a Korea-U.S. Free Trade Agreement.”

The file “KORUS_YS_input.txt” contains the projections for the effects of KORUS on bilateral trade between Korea and the United States from Yaylaci and Shikher’s paper “What Would KORUS FTA Bring?”

3.4 NAFTA Projections (BDS, Cox-Harris, and Sobarzo)

The files “bds_input.txt”, “sobarzo_input.txt” and “coxharris_input.txt” contain the projections for the effects of NAFTA from, respectively, the BDS model, the Sobarzo model, and the Cox-Harris Model. These projections are taken from *Modeling North American Economic Integration* (1995) edited by P.J. Kehoe and T.J. Kehoe; and are also available in Kehoe (2003): <http://www.econ.umn.edu/~tkehoe/papers/NAFTAevaluation.pdf>

3.5 Level Tariff Data

The files “NAFTAtariffs_product.csv” contains tariff data from the TRAINS database accessed through WITS (see section 4 below) for imports into Canada and the United States in 1989 and 2009. “NAFTAtariffs_product_mex.csv” contains the same thing for imports into Mexico for 1991 and 2009. The tariffs are the simple average of effectively applied tariff rates at the 5-Digit SITC Rev. 2 level.

The files “NAFTAtariffs_industry.csv” and “NAFTAtariffs_industry_mex.csv” are the exact same as the above files, except the tariffs are the simple average of effectively applied tariff rates at the 3 digit ISIC rev. 2 level.

The file “NAFTA_KORUS_TARIFFS.dta” contains the simple average of effectively applied tariff rates at the TOTAL level, between NAFTA country pairs and KORUS country pairs for 1988-2012.

4. Gravity Regression Data

The files “Gravity1989.dta” and “Gravity2005.dta” contain the variables used run the gravity regressions in the paper. The structure of the files list observations for importer-exporter pairs, where the importer and exporter are listed by their 3-digit iso codes. The variable tariffs gives the simple average for effectively applied tariff rates for exports from the exporter to the importer. The tariff data is from the TRAINS database, accessed through WITS (<http://wits.worldbank.org/>). The variable value gives aggregate exports from the exporter to the importer in the specified year, and the data is from COMTRADE (<http://comtrade.un.org/>). The other data is from the CEPII gravity database (http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8), and includes importer GDP (imp_gdp_nom), importer GDP per capita (imp_gdp_cap), exporter GDP (exp_gdp_nom), exporter GDP

per capita (`exp_gdp_cap`), population weighted distance between the exporter and importer (`distw`), whether the importer and exporter share a border (`contig`), a common language (`comlang_ethno`), or a colonial relationship (`colony`). The sample is all possible importer-exporter pairs that have data on all variables for the given year.

5. Gamma Estimate with 1978-1987 Data

The file “`SITC_least_78_87.dta`” contains data on aggregate trade flows for all products (`total_*`), least traded products (`least_val_*`), as well as the number of all products and least traded products traded (`all_products_*`, `least_products_*`) where `*` is the year, 1978 for the base year and 1987 for the final year. The file also contains growth rates for the value of trade of all products (`growth_total_78_87`) and least traded products (`growth_least_78_87`). This file was constructed using bilateral trade flows at the 5-digit SITC Rev. 2 level for all importer-exporter pairs with data availability in COMTRADE, and formatted into “`SITC_least_78_87.dta`” using code adapted from the Kehoe and Ruhl (2013) data appendix. These files are available by contacting the authors, but are not included here due to their file size.

Section B: Programs and Descriptions

All programs discussed here are in the folder “Programs”. The name of the program being described is bolded and followed by an explanation of what it does and how it works. It’s also broken down into subsections which are underlined and correspond to the subfolder the programs are in.

Note: before doing the analysis we drop the codes for “Coal gas, water gas, and similar gases” (34150), electric current (35100), and “special transactions, commodity not classified according to class” (93100) since they aren’t included in our industry analysis.

Note2: must edit the base file directory in each program before running it. Change “C:\DataAppendix” to your location.

Note3: Tables 3 and 8 are constructed by hand using the estimates of γ , λ_t , and initial tariff levels taken from the file “NAFTA_KORUS_TARIFFS.dta”.

Gravity Regressions:

Gravity_Regression 1989.do

Uses the file “Gravity1989.dta” to run the gravity regressions specified in the paper (equation 5) used to make table 1, both the base specification and with additional controls, and outputs the results to the file “Table 1 Gravity1989.txt”

Gravity_Regression 2005.do

Uses the file “Gravity2005.dta” to run the gravity regressions specified in the paper (equation 5) used to make table 1, both the base specification and with additional controls, and outputs the results to the file “Table 1 Gravity2005.txt”

Gamma Estimate:

Gamma_78_87_estimate.do

Uses the file “SITC_least_78_87.dta” and reproduces the exercise from Kehoe and Ruhl (2013) to estimate gamma (used in the right side of equation 4) using 1978-1987 data. Outputs the estimate of gamma to the file “Gamma Estimate 1978-1987.txt”

Pre-Formatting:

FormatConcordance.do

Creates the 5 digit SITC to 3 digit ISIC concordance using the list of 5 digit SITC codes (“SITC_Leaf_Codes.txt”), the 4 digit SITC to 3 digit ISIC concordance from Muendler (2009) (“sitc2isic.dta”), and our changes to Muendler’s concordance (“sitc2isic_missing.dta”). Outputs the file “SITC5_to_ISIC3.dta”

FormatData_KORUS.do

Formats the raw KORUS bilateral trade data in the folder “raw”, downloaded from Comtrade, together and processes it to create the file “KORUS.dta” (see section 1.1.1A)

FormatData_KORUS_KOREA_WORLD.do

Formats the raw KORUS trade data for Korea with the United States and the World in the folder “raw”, downloaded from Comtrade, and processes it to create the file “KOREA_WORLD.dta” (see section 1.1.1B)

FormatData_KORUS_USA_WORLD.do

Formats the raw KORUS trade data for United States with Korea and the World in the folder “raw”, downloaded from Comtrade, and processes it to create the file “USA_WORLD.dta” (see section 1.1.1B)

FormatData_NAFTA.do

Combines the raw bilateral NAFTA trade data in the folder “raw”, downloaded from Comtrade, together and processes it to create the file “NAFTA.dta”

FormatData_NAFTA_CoxHarris.do

Combines the raw bilateral trade data in the folder “raw”, replacing canmex8809.csv” and “canusa8809.csv” with “canmex8809_CAN.csv” and “canusa8809_CAN.csv,” and processes it to create the file “NAFTA_CoxHarris.dta”

KORUS Experiments:**KORUS_ISIC.do**

Uses the files “KORUS.dta” and “SITC5_to_ISIC3.dta” as inputs, and does the least traded products exercise described in the paper to generate predictions for bilateral trade between Korea and the United States using ISIC codes for industries.

Outputs the files “korus_ISIC_RESULTS.dta” and the tables “Table B2 KORUS_ISIC Kor Exports.txt”, and “Table B2 KORUS_ISIC Kor Imports.txt”

KOREA_WORLD_KS.do

Uses the files “KOREA_WORLD.dta”, “SITC5_to_ISIC3.dta”, “isic2other.csv”, and “KORUS_KS_input.txt” as inputs, and does the least traded products exercise described in the paper to generate predictions for Korean trade with the United States and the World using Kiyota and Stern’s industry definitions and comparing our projections to their own.

Outputs the files “korus_ks_KOR_RESULTS.dta” and the tables “Table B4 KORUS_KS KOR Exports.txt”, and “Table B4 KORUS_KS KOR Imports.txt”. The correlations for Table B4 are outputted to the file “Table B4 KS-LTP corr.txt”.

USA_WORLD_KS.do

Uses the files “USA_WORLD.dta”, “SITC5_to_ISIC3.dta”, “isic2other.csv”, and “KORUS_KS_input.txt” as inputs, and does the least traded products exercise described in the paper to generate predictions for U.S. trade with Korea and the World using Kiyota and Stern’s industry definitions and comparing our projections to their own.

Outputs the files “korus_ks_USA_RESULTS.dta” and the tables “Table B5 KORUS_KS USA Exports.txt”, and “Table B5 KORUS_KS USA Imports.txt”. The correlations for Table B5 are outputted to the file “Table B5 KS-LTP corr.txt”.

KORUS_YS_BILATERAL.do

Uses the files “KORUS.dta”, “SITC5_to_ISIC3.dta”, “isic2other.csv”, and “KORUS_YS_input.txt” as inputs, and does the least traded products exercise described in the paper to generate predictions for bilateral trade between Korea and the United States using Yaylaci and Shikher’s industry definitions and comparing our projections to their own.

Outputs the files “korus_YS_RESULTS.dta” and the tables “Table B6 KORUS_YS Kor Exports.txt”, and “Table B6 KORUS_YS Kor Imports.txt”. The correlations for Table B6 are outputted to the file “Table 7 YS-LTP corr.txt”.

NAFTA Experiments:

NAFTA_ISIC.do

Using the files “NAFTA.dta”, “SITC5_to_ISIC3.dta”, and “GDP_input.txt” as input to evaluate how effective the least traded products exercise might have been in generating relative industry level predictions, using ISIC codes as industries, for bilateral trade between Canada, Mexico, and the United States following NAFTA.

It outputs the file “NAFTA_ISIC_RESULTS.dta” which is used by the program “NAFTATariff_industry.do”. It also produces the file “NAFTA Country Weights.txt” which gives the weights used to construct weighted averages.

NAFTA_BDS.do

Using the files “NAFTA.dta”, “isic2bds.csv”, “SITC5_to_ISIC3.dta”, “GDP_input.txt”, and “bds_input.txt” as input it evaluates both the Brown, Deardorff, and Stern exercise and the least traded products exercise to see how effective they were in generating relative industry level predictions for bilateral trade between Canada, Mexico, and the United States following NAFTA.

It outputs the file “NAFTA_BDS_RESULTS.dta” and the tables “Table 4 BDS Can to US.txt”, “Table 4 BDS US to Can.txt”, “Table 5 BDS Can to Mex.txt”, “Table 5 BDS Mex to Can.txt”, “Table 6 BDS Mex to US.txt”, “Table 6 BDS US to Mex.txt”. The correlations and regression coefficients for tables 4-7 are outputted to the file “Table 4-7 corr_reg.txt”. The optimal coefficients for table 2 are outputted to the file “Table 2 opt_coeff.txt”.

The regression coefficients from this file, along with the weights from “NAFTA Country Weights.txt” are used to construct the predicted growth rates for least traded and non-least traded products used in our KORUS predictions.

NAFTA_CoxHarris.do

Using the files “NAFTA_CoxHarris.dta”, “isic2other.csv”, “SITC5_to_ISIC3.dta”, “GDP_input.txt”, and “coxharris_input.txt” as input it evaluates both the Cox-Harris exercise and the least traded products exercise to see how effective they were in generating relative industry level predictions for trade between Canada and the World following NAFTA.

NAFTA_Sobarzo.do

Using the files “NAFTA.dta”, “isic2other.csv”, “SITC5_to_ISIC3.dta”, “GDP_input.txt”, and “sobarzo_input.txt” as input it evaluates both the Sobarzo exercise and the least traded products exercise to see how effective they were in generating relative industry level predictions for trade between Mexico and North America (Canada+United States) following NAFTA.

It outputs the file “Sobarzo_RESULTS.dta” and the tables “Table A2 Sobarzo Exports.txt” and “Table A2 Sobarzo Imports.txt”. The correlations and regression coefficients for table A2 are displayed on the Stata terminal and not saved to file.

Robustness Checks:

NAFTATariff_product.do

Uses the files “NAFTA_ISIC_RESULTS.dta”, “NAFTATariffs_product.csv”, and “NAFTATariffs_product_mex.csv” to compute the optimal coefficients at the product level (5 digit SITC Rev. 2 codes) both with and without controlling for initial tariff levels. The results are used in Table 14 and outputted to the file “Table 14 Product.txt”. This program also outputs the average tariff rates between NAFTA country pairs in 1989 (1991 for Mexican imports) and 2009, which is used for Table 3 and outputted to “Table 3 Tariff Rates.txt”

NAFTATariff_industry.do

Uses the files “NAFTA.dta”, “NAFTATariffs_industry.csv”, and “NAFTATariffs_industry_mex.csv” to compute the optimal coefficients at the industry level (3 digit ISIC Rev 2. Codes) both with and without controlling for initial tariff levels. The results are used in Table 3 and outputted to the file “Table 3 Industry.txt”.

NAFTA_accuracy.do

Uses the file NAFTA_BDS_RESULTS as an input file to construct the weighted average of the absolute values of the percent deviation from predicted growth for the BDS predictions, our LTP predictions, and how well the LTP ‘predictions’ would have done if our estimated coefficients had been the optimal coefficients. This outputs the values for table 8 to “Table 8 Accuracy.txt”.

pricequantity decomposition.do

Uses the files “NAFTA.dta” and “PPI.csv” as input data files to breakdown the change in the value of trade into the share accounted for by changes in quantity and the share accounted for by changes in prices. In particular we compute share of change due to quantity changes $(q_{share})_{iv}^k$, for each product v as:

$$(q_{share})_{iv}^k = \frac{\log(x_{ivT_1}^k / def_{iT_0T_1}^k) - \log(x_{ivT_0}^k)}{\log(q_{ivT_1}^k) - \log(q_{ivT_0}^k)}$$

Where $x_{i,v,t}^k$ and $q_{i,v,t}^k$ are, respectively, the value of exports and the volume (quantity) of exports between from country i to country k at time t , where $T_1 = 2009$ and $T_0 = 1989$, and $def_{iT_0T_1}^k$ is a price deflator computed as:

$$def_{iT_0T_1}^k = \left(\frac{PPI_{i,T_1}}{PPI_{i,T_0}} \right) \left(\frac{e_{i,T_0}}{e_{i,T_1}} \right)$$

Where $PPI_{i,t}$ is the producer price index for country i (the exporter) at time t , and $e_{i,t}$ is the exchange rate of country i with the United States at time t . We then compute the weighted average of $(q_{share})_{iv}^k$ across all products, weighting each product by $x_{ivT_0}^k$, after trimming the 5% of products with the highest and lowest values of $(q_{share})_{iv}^k$ to reduce the influence of outliers.

The q-share values for table A4 are outputted to the file “Table A4 QShare.txt”.