<div align="center">

Data Appendix
January 2014

</div>

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"

1.2.1C Bilateral NAFTA Trade Data for Using 1988 as a Base Year
The data are exactly the same as 1.2.A, except the files "canmex8809.csv" and "canusa8809.csv" are replaced by "canmex8809_CAN.csv" and "canusa8809_CAN.csv" which are the same flows and trade, respectively Canadian exports to Mexico and Canadian exports to the United States, both over 1988-2009, only reported by Canada instead of by the importer.

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"

The data for using 1988 as a base year is formatted using the Stata program "FormatData_NAFTA_88_07.do" and produces the file "NAFTA_88_07.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

4.1 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"

4.2 ISIC-to-Industry Concordance
The file "isic2dbs.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.

4.3 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?"

4.4 NAFTA Projections (BDS, Cox-Harris, and Sobarzo)
The files "bds_input.txt", "sobarzo_input.tzt" 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

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.

Note 2: must edit the base file directory in each program before running it. Change "C:\DataAppendix" to your location.

<u>**Pre-Formatting:**</u>

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 2 KORUS_ISIC Kor Exports.txt", and "Table 2 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 3 KORUS_KS KOR Exports.txt", and "Table 3 KORUS_KS KOR Imports.txt". The correlations for Table 3 are only displayed on the Stata terminal and not saved to file.

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 4 KORUS_KS USA Exports.txt", and "Table 4 KORUS_KS USA Imports.txt". The correlations for Table 4 are only displayed on the Stata terminal and not saved to file.

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 5 KORUS_YS Kor Exports.txt", and "Table 5 KORUS_YS Kor Imports.txt". The correlations for Table 5 are displayed on the Stata terminal and not saved to file.

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 files "NAFTA_ISIC_RESULTS.dta" and "NAFTA Country Weights.txt", the latter of which is used to manually construct the weighted correlations and weighted regression coefficients. It also outputs the tables "Table 6 ISIC Can to US.txt", "Table 6 ISIC US to Can.txt", "Table 7 ISIC Can to Mex.txt", "Table 7 ISIC Mex to Can.txt", "Table 8 ISIC Mex to US.txt", and "Table 8 ISIC US to Mex.txt". The correlations and regression coefficients for tables 6-8 are displayed on the Stata terminal and not saved to file.

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 10 BDS Can to US.txt", "Table 10 BDS US to Can.txt", "Table 11 BDS Can to Mex.txt", "Table 11 BDS Mex to Can.txt", "Table 12 BDS Mex to US.txt", "Table 12 BDS US to Mex.txt". The correlations and regression coefficients for tables 10-12 are displayed on the Stata terminal and not saved to file.

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.

It outputs the file "CoxHarris_RESULTS.dta" and the tables "Table A1 Cox-Harris Exports.txt" and "Table A1 Cox-Harris Imports.txt". The correlations and regression coefficients for table A1 are displayed on the Stata terminal and not saved to file.

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:

FormatData_NAFTA_88_07.do
Combines the raw bilateral NAFTA trade data in the folder "raw", replacing the files "canmex8809.csv" and "canusa8809.csv" with "canmex8809_CAN.csv" and "canusa8809_CAN.csv", and processing it to create the file "NAFTA_88_07.dta"

NAFTA_LTP_1988_2007.do
Using the files "NAFTA_88_07.dta", "isic2bds.csv", "SITC5_to_ISIC3.dta", "GDP_input.txt", and "bds_input.txt" as input it does both the BDS exercise and the least traded products exercise as described in the paper using the period 1988-2007 for trade involving Canada and 1989-2007 otherwise.

It outputs the files "NAFTA_BDS_88_07_RESULTS" and "NAFTA 88_07 Country Weights.txt", the latter of which is used to create the weighted correlations and weighted regression coefficients for table 15. The correlations and regression coefficients for table 15 are displayed on the Stata terminal and not saved to file.

alphabeta_product_estimation.do.do
Uses the file "NAFTA.dta" as input to do the alternative estimation for $\tilde{\alpha}$ and $\tilde{\beta}$, and outputs the file "alphabetaest.dta" and the table "Table 9 Product Level.txt".

pricequantity analysis.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_{i,T_0,T_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 16 are displayed on the Stata terminal and not saved to file.