**New Estimates of Offshoring/Intermediate Inputs, v1.0**

**Data Construction**

* Intermediate import data were downloaded from the BEA: (including from here: <http://www.bea.gov/industry/iedguide.htm>) for the years 1997, 2002, and 2007. (Note: To my knowledge, these were the only years in which intermediate imports were computed directly by the BEA; although 2012 may be available/will be available soon.)
* These data used NAICS codes which are specific to the IO database (we call it IONAICS), which differ slightly each year. Thus, we created a crosswalk between the IONAICS codes in each year and NAICS codes from the Annual Survey of Manufactures (in order to match this data to ASM data and use it as a panel).
* We then tried extrapolating to the intervening years using WITS data from the period 1989 to 2010. If there was only Intermediate Import data for one benchmark year, then we just did a simple extrapolation using the trend of overall imports in that sector. However, we noticed that the share of intermediate imports in total imports fell in each subsequent benchmark year. This was to be expected, given that imports grew much faster than domestic production, implying that it is only natural that the share of imports which are intermediate would decline over time. Thus, a simple extrapolation will tend to lead to upward bias over time. (This can easily be confirmed with the data since we have three benchmark years.)
* Thus, a preferable solution is to use the benchmark estimates for the concentration of intermediates in imports, and then estimate the evolution of (intermediate concentration = imported intermediates/total imports) by using the following regression:
  + ,
  + Where t = 1997, 2002, 2007, and i is all the combinations of different importing and using industries – (we end up with 302 of each using the ASM NAICS classification). = Intermediate Input Concentration of imports, = Imported Intermediates/Total Imports.
  + We found, as we expected, that increases in imports were associated with a decrease in the concentration ratio, and that an increase in shipments was correlated with an increase in the ratio. This is what we expected for imports, as the more imports increase, other things equal, we would expect the share of imports used in production should fall. Production is less intuitive, as there could be effects in either direction, but as production increases, perhaps demand for intermediate inputs in this sector increases almost mechanically, as we find a strong relationship here. We also tried controlling for a weighted average of demand changes of other input-using sectors, computed for each intermediate input sector. However, we found a weaker relationship here which flipped sign on different samples, so we elected not to use. Part of the problem is that this variable and the change in shipments of the importing sector are highly correlated.
* As there were missing observations, even in the benchmark years, we then filled in the missing using:
* However, outside the benchmark years, notice that we can’t use this formula exactly, since we wouldn’t have 5 year lags of the input concentration variable. For the rest of the years, we can only use changes in imports and shipments. Thus, we wondered if this is really an improvement over, for example, simply guessing about the import concentration after 2007 using the 2007 value for import concentration, or adjusting based on the growth in overall imports. To test this, we compared the mean absolute error of estimates for 2007 using several different methods: (1) just use the input concentration in 2002, (2) update the 2002 import concentration numbers just using changes in the ratio of imports to shipments, (3) predict changes from import concentration in 2002 using the formula . (4) We also checked the mean squared error of , although note that we can’t use this formula exactly for the years after 2007.
* Results: Not surprisingly, Method (4) worked the best, followed by (3), then (1), and last was (2). Thus, we used Method (4) for the missing data in the benchmark years, and method (3) for the other years.

|  |  |  |
| --- | --- | --- |
|  | **Description** | **Mean Absolute Error** |
| 1 | Import Conc. In 2002 | 0.982 |
| 2 | Updated using ratio of imports/shipments | 0.985 |
| 3 | Extrapolation using Reg Results (excluding lags) | 0.971 |
| 4 | Extrapolation using Reg Results (using lags) | 0.870 |

* Lastly, we filled in zeros in cases where the only benchmark year values available were zero, when there were consecutive benchmark year zero values, or before the first benchmark value when the first benchmark was a zero, or after the last benchmark value when it, too, was a zero. (Thus, if the BEA reported 0 intermediate imports in 2007, we inputed 0 for 2008-2010 as well. If 1997 was missing, and 2002 was 0, then we imputed 0 values before 2002 as well.)

**Variable Description**

**First File – Bilateral Sector X Sector File of Intermediate inputs: InterImportsfull.dta**

**naics\_mp** – NAICS code of importing sector (matchable with ASM NAICS)

**naics** – NAICS code of using sector (matchable with ASM NAICS)

**year** – the year, spanning 1989-2010

**imports** – Total Nominal Imports for sector naics\_mp (from [World Bank/WITS](http://wits.worldbank.org/))

**mpinports** –Imported Intermediate Imports (imports are classified by naics\_mp, imported and used by the sector listed in the variable naics. We used BEA data for the benchmark years, and our own estimates for the other years.)

**Second File – Sectoral Intermediate Inputs (total for each using sector): IntermediateImportsweb.dta**

**naics** – NAICS code of using sector (matchable with ASM NAICS)

**year** – the year, spanning 1989-2010

**tmpinports** –Total Imported Intermediate Imports (Using BEA data for the benchmark years, and our estimates for the other years)