**Explanation of Files for Table 1 and Instructions for Mapping U.S. Industrial Skill Intensity Data to U.S. CEX Consumption Data through the U.S. Input-Output Structure**

1. **Files:**

MakeTable1.do – this creates the regression results in Table 1 and outputs them to “Table 1 results.log”. These results are quarterly, but the monthly analogs are outputted to a “monthlyresults.log”. It uses “VAdataforCEXmerge.dta" and also the monthly interview data from the 2012 CEX, and creates an intermediate file called “CEXdemographicdata.dta”

MakeTable1Analog\_DiaryData.do – this creates the analog of the regression results in Table 1 but for the CEX diary sample. It outputs the results to “MonthlyResultsDiary.log" and “QuarterlyResultsDiary.log”. It uses “VAdataforCEXmerge.dta" and also the monthly interview data from the 2012 CEX, and creates an intermediate file called “CEXdemographicdataDiary.dta”

CEXtoVAdocumentation.xlsx – this Excel workbook contains all needed data to reassemble “VAdataforCEXmerge.dta”, including the relevant input-output matrixes used in MatrixManipulations.m, the relevant high-skill intensive and low-skill intensive industry classification, and the mapping steps for mapping PCE categories to CEX categories (as explained below).

VAdataforCEXmerge.dta – this is a data file that contains the final output from the input-output matrixes and concordance mappings as explained below. It is used in combination with the CEX data in both “MakeTable1.do” and “MakeTable1Analog\_DiaryData.do”

MatrixManipulations.m – This is a Matlab code that does the required matrix algebra from the Input-Output tables to yield the industry value added contained in various contained in various final expenditure categories. The output of this is reproduced in ‘VA Bundle Matrix’ worksheet in “CEXtoVAdocumentation.xlsx”.

1. **Instructions for Mapping U.S. Industrial Skill Intensity Data to U.S. CEX Consumption Data through the U.S. Input-Output Structure**

There are several important steps to mapping the data together.

We start by defining a binary variable for high-skill intensive industries as well as a continuous measure of the fraction of labor that is high-skill intensive. The definitions for high-skill are done at a level that allows us to match up with the industry classifications in the EU KLEMS “Basic Tables” data. We define “Financial Intermediation”, “Business Services", “Education”, and “Health and Social Work”. We map the disaggregate components of these SIC industries to the 69 industries defined by the U.S. input-output classification to create a dummy variable for which of the 69 NIPA industries are high-skill intensive. These final decisions are all given in the file in the worksheet ‘HS\_Cutoff for CEX’ in “CEXtoVAdocumentation.xlsx”. (Red indicates low-skill intensive industries, while green indicates high-skill intensive industries.)

The second thing that needs to be done is to translate 76 personal consumption expenditure categories (NIPA lines) into the 71 Input-Output commodity categories for the IO system (69 industries plus “used” and “other”). This involves two steps (1) attributing the expenditure to a final goods/services producing industry, and (2) attributing the appropriate transportation and distribution margins for the final goods/services into the appropriate industries. We start this by using the data in worksheet ‘PCEBridge 2012’ in “CEXtoVAdocumentation.xlsx”. It maps NIPA lines to commodity codes and it also gives the totals of producer value (produced in the appropriate commodity code), as well as transportation, retail, and wholesale margins for the year 2012. Using these data, we can construct the fraction of a dollar in final expenditures (purchaser value) that goes to each industry. However, the table does not tell the breakdown of transportation costs across various transportation industries, nor does it appropriate retail and wholesale costs. We impute transportation costs at a finer classification by assuming the same split we observe across five distribution industries in the 2002 benchmark data, which are in worksheet ‘2002 PCE\_Bridge\_Summary’ in “CEXtoVAdocumentation.xlsx”. We also choose a mapping for retail costs into subindustries of retail (automotive, food and drink, general merchandise and other retail).[[1]](#footnote-1)

We thus form a matrix, in which each column represents a NIPA line from the personal consumption expenditures and each row is an industry. Element ij represents the total (current value) final output demanded for an industry (or commodity) i from expenditures in NIPA line j. After normalizing columns to some to one, we get the value of demand for final output for each industry (or commodity) that stems from one dollar of final expenditure in a given PCE category.

The next thing to do is translate these demands for *final* output into demands for *total* output by industry using the total requirements table for 2012, which is the worksheet ‘IxC\_TR\_2012’, which is the ‘2012’ sheet in the summary BEA table “IxC\_TR\_1997-2012\_Summary.xlsx”. Each element of this table shows the total amount of industry i output required in the production of one dollar of industry j output. This industry-by-commodity table translates final output of commodity i into total output required from each industry. Finally, we get the value-added demand from output in that industry using the direct requirements table, the worksheet ‘CxI\_DR\_2012’, which is ‘2012’ sheet in the summary BEA table “CxI\_DR\_1997-2012\_Summary.xlsx”. Rows correspond to factors and intermediate inputs, so that element ij is the (current value) value of commodity/factor i used to produce one dollar of output industry j output. We the sum of V001 (compensation by employees), V002 (taxes on production and imports, less subsidies), and V003 (gross operating surplus) to get value added as a fraction of total output in each industry, which we then use to get the value added from each industry in a dollar of each PCE NIPA line. This is a 69x76 (industry x PCEline) matrix saved in the worksheet ‘VA Bundle Matrix’ in the file “CEXtoVAdocumentation.xlsx”.

Lastly, we multiply this matrix of value added in each industry per dollar of PCE commodity by the vector of high skill-intensive dummies for each industry (i.e., “hsect” in ‘HS Cutoff for CEX’) to get value-added in high-skill intensive sector. This results in a vector that tells us the value added in the high skill-intensive sector that is embodied in one dollar of final output for each of NIPA PCE lines. This resulting vector is saved in the sheet ‘Data for CEX Merge’ in the file “CEXtoVAdocumentation.xlsx”. This is the output of the Matlab code “MatrixManipulations.m”.

The final step is to creatae a corresponde to allow us to merge this vector with the CEX data. We start with the household level Consumer Expenditure Survey (CEX) data for the United States from 2012. We restrict ourselves to months in the primary interview sample (spread out over contained in the files mtbi121x.dta, mtbi122.dta, mtbi123.dta, mtbi124.dta, and mtbi131.dta) and each observation is a household-month observation. We combine these expenditure data with demographic data from the ‘fmli’ files (fmli121x.dta, fmli122.dta, fmli123.dta, fmli124.dta, and fmli124.dta) to get household composition, education levels, and after tax income.

Merging the data requires mapping the “UCC” codes on expenditures in the CEX to the “PCE codes”. We do this using the BLS-provided Concordance Mapping of CE UCC codes to PCE Series from Table 2.4.5U, which is worksheet titled ‘UCC-PCE Item Mapping’ in “CEXtoVAdocumentation.xlsx”. There are three caveats, however.

First, the BLS Concordance mapping attributes different weights to CEX expenditures in some cases. These are recorded as a variable “weight”, which is used in aggregating expenditures.

Second, several UCC categories are not mapped because they cannot be fully mapped into disaggregate PCE series. For many of these, the expenditures would all fall under the same category at the level of aggregation (76 PCE lines) we use in our input-output analysis, and so we assign these by hand. Our original hand mapping was less complete than the one later constructed by Comin, Lashkari, and Mestieri (2019), which included our mappings plus additional hand mappings as well as those deemed as “expert judgement” from a BEA’s updated correspondence. We have therefore adopted their mapping.

Third, this table only maps UCC codes into PCE Series *names* and using designations in Table 2.4.5U*.* The sheets ‘Table 2.4.5U’ is needed to map it into PCE lines, and the ‘Table 2.4.5U to Table 2.4.5 Map’ (also BLS-provided) is needed to map these into the final NIPA lines of BEA NIPA Table 2.4.5. The final mapping, including all stages, is available in the worksheet ‘VAdataforCEXmerge’ in the file in the “CEXtoVAdocumentation.xlsx” and the Stata file ‘VAdataforCEXmerge.dta’ . The column “source” indicates the source of the mapping.

1. NIPA lines 5-7 (new autos, used autos, and automotive parts) are mapped to auto retailers. NIPA lines 27-28 (food and non-alcoholic drink for off premises consumption and alcoholic drinks for off premise consumption) are mapped into food and drink retailers. Everything else is mapped into miscellaneous retailers, with nothing mapped into general merchandise stores. [↑](#footnote-ref-1)