P.O.R.T. W.E.S.

The Potentially Overzealous Researcher of Transportation and his Waterborne Expense Statistics

Introduction:

My title could very well have been data monkey. However, few chimps can use quantitative analysis software and even fewer lemurs can comprehend the size of the World War 2 Civilian Production Administration's data set being used by economic historian, Gillian Brunet. The data set that I spent my summer managing and parsing through for contracts related to sea vessels has over 190,000 observations, each with a minimum 30 variables (dummy variables were often added and removed for management purposes), and a total value, in nominal 1946 U.S. dollars, of $183 billion. Of these contracts, we found 3,984 related to sea vessel manufacturing and repairing (not including individual parts), which contribute at least $22.6 billion of that sum. I have also spent two weeks going through the data managing every contract related to engines, categorizing almost another 1,400 contracts. With all this data, properly categorized based on standard industrial code classifications, Professor Brunet is able to analyze the economic impacts of U.S. war contracts during WW2 on the county level.

SIC Codes:

To first understand my work with World War 2 sea vessels and engines, it is important to explain how the Standard Industrial Classification codes work. The basis of the codes is very comparable to the Dewey decimal system, where each number between 1-9 is a category of industry type, followed by anther digit that indicates subdivisions of that industry and so on. Most of my work pertained to manufacturing, which the Central Statistical Board created the basis of as early as 1938. However the SIC codes have been evolving since then, so the categorization applicable is significantly different from the modern 2020 classifications. Where the extent of the coding in the 1947 Census of Manufacturers was a four digit SIC code, there are now extended codes up to 8 digits. To keep the 1947 Census of Manufacturers as a viable checklist (it reports the data by SIC code), we used the same SIC classification definitions that were applicable in 1947.

Off to the left is a 1947 Census of Manufacturers chart that lists the number of establishments and production workers for all the subcategories of transportation manufacturing being compared between 1939 and 1947. There, one can see that the number of establishments that manufactured ships and boats went from 608 to 1,010 over that eight year span. The census also includes an appendix that explains the definition of each SIC code, of which there is an example bellow:

1947 Code	1939 Code	Title	1947 Number of establishments	1947 Production workers	1939 Number of establishments	1939 Production workers
3717		Motor vehicles and parts	963	558.7		
	1810*	Motor vehicles, bodies and parts (see 3713, 3715) [4]			942	388.1
3713		Truck and bus bodies	603	22.9		
	1810*	Motor vehicles, bodies and parts (see 3715, 3717)				
3715		Truck trailers	119	9.4	99	4.0
	1810*	Motor vehicles, bodies and parts (see 3713, 3717) [4]			99	4.0
3716	1830	Automobile trailers [4]	219	8.3	87	1.8
3721		Aircraft	58	110.3	44	34.2
	1921*	Aircraft and parts (see 3722, 3723, 3729) [4]			44	34.2
3722		Aircraft engines	57	35.9	20	9.4
	1921*	Aircraft and parts (see 3721, 3723, 3729) [4]			20	9.4
3723		Aircraft propellers	13	4.8		
	1921*	Aircraft and parts (see 3721, 3722, 3729)			67	5.2
3729		Aircraft equipment, n.e.c.	204	11.6		
	1921*	Aircraft and parts (see 3721, 3723) [4]				
3731	1931	Ship building and repairing [12]	301	114.7	608	69.2
3732	1932	Boat building and repairing [12]	809	17.3		
3741	1911	Locomotives and parts [4]	36	25.4	20	7.0
3742		Railroad and street cars	86	50.3	48	17.7
	1912*	Railroad and street cars (see 3591, n.s.) [4]			48	17.7
3751	1941	Motorcycles and bicycles	76	13.7	35	7.0
3799	1999	Transportation equipment, n.e.c.	167	3.9	41	1.3

373 **SHIPS AND BOATS**

3731 **Ship building and repairing**
Establishments primarily engaged in building and repairing all types of ships, barges, canal boats and lighters of more than 65 feet in length, whether propelled by sail or motor power or towed by other craft. Establishments primarily engaged in fabricating structural assemblies or components for ships, or subcontractors engaged in ship painting, joinery, carpentry work, electrical wiring installation, etc., are not classified in this industry. Shipyards engaged in repair work, either exclusively or as their major activity, are included in this industry.

3732 **Boat building and repairing**
Establishments primarily engaged in building and repairing all types of boats of 65 feet or less in length, such as motorboats, sailboats, rowboats, lifeboats, and canoes. Small establishments (less than 10 employees) repairing boats, but not engaged in other manufacturing activity, are excluded from the scope of this census.

As you might've noticed, the definitions are more designed to include specific manufacturers than actual products, which would become its own source of difficulty as conditions based on "major activity" aren't applicable. There is also the specific difficulty regarding sea vessel manufacturing (which plagued my summer) that individual ship parts aren't classified under "Ships and Boats." Therefore each of those, no matter how essential to a ship or boat's functionality, are completely different SIC codes, very distant from the sub category of transportation, SIC 37. This problem led to the creation of a new dummy variable, ShipPart, that was applied to contracts across all types of SIC codes.

Data Mining:

The dataset variables first presented to me basically covered the manufacturer name, location, and product for the contract, as well as what government agency (and usually its sub divisional branch) was responsible for the purchase. Then it was my job to fill in the SIC code variables, divided by each digit subcategory, and a couple other descriptive variables such as whether a contract was for a rationed, intermediary, or militantly purposed good. Later on, ShipPart, whether a product was part of a ship or boat, was added to this list. Though this is plenty of information to go off of, there is little uniformity to how products are listed.

More Data Mining:

For example, there were 368 contracts that included the string "BARGE" (everything in the data set is capitalized) in its Product Variable, with a total of 62 different Products labels related to "BARGE." Finding, and then determining if there was a significant difference between, "BARGES FIRE," "BARGES DUMP," "FLUATING CRANE BARGES," and "DERRICK BARGES S1BG1" proved much more difficult than anticipated, particularly because the criteria of 65 feet varied between barge type. What I had to do was use the string position command in Stata to create a dummy variable that identified whether or not a contract included the string "BARGE" in its product input, and cross examine every contract for a barge to a dataset I made using online resources that would indicate size , building material, and such. I did this for various other types of sea vessel based string searches, at one point making a dataset on google sheets with almost 5,000 observations.

This ability to search for strings inputs by product title was a crucial development in my process to create categorical variables to subdivide the dataset. As the subcategories grew based off vastness of the specific product or criteria being searched, I began subdividing from previous dummy variables and so on. A great example of this can be found in the work I did for the machinery SIC codes, over on the right. I created a dataset pertaining to everything "ENGINE" and then split it two ways, amongst different SIC codes for machinery, and then again amongst SIC codes for transportation, electrical equipment, and instruments, as Engines that are part of certain transportation modes are to be classified as transportation parts and certain engine applicators or accessories were separate from machinery, SIC .35. This required the creation of variables for the sole purpose of identifying whether a contract was being altered after each subdivision for the purpose of recollecting the data that remained unclassified post subset transformations. Overall there were 11 different categorical variables created for the purpose of management of all "ENGINE" related products.

```
gen eng =1 if strpos(Product, "ENGINE")
replace eng=1 if  strpos(Product, "ENGN")
replace eng=0 if eng==1 & sic2==35
replace eng=0 if eng==1 & strpos(Product, "ENGINEER")
replace eng=0 if eng==1 & strpos(Product, "PART")
replace eng=0 if eng==1 & strpos(Product, "FITS")
replace eng=0 if eng==1 & strpos(Product, "SERVICE")
replace eng=0 if eng==1 & strpos(Product, "SERV")
replace eng=0 if eng==1 & strpos(Product, "ROOM")
replace eng=0 if eng==1 & strpos(Product, "MATERIAL")
replace eng=0 if eng==1 & strpos(Product, "WRECHES")
replace eng=0 if eng==1 & strpos(Product, "KIT")
replace eng=0 if eng==1 & strpos(Product, "TOOL")
replace eng=0 if eng==1 & strpos(Product, "FUEL")
replace eng=0 if eng==1 & strpos(Product, "WASHING")
replace eng=0 if eng==1 & strpos(Product, "CONL")
replace eng=0 if eng==1 & strpos(Product, "SEAL")
replace eng=0 if eng==1 & strpos(Product, "COVER")
// envelopes used for transporting engines
replace eng=0 if eng==1 & strpos(Product,"ENVELOPE")
replace eng=0 if eng==1 & strpos(Product, "DENTAL")
// Mounts and such
replace eng=0 if eng==1 & strpos(Product, "STAND")
replace eng=0 if eng==1 & strpos(Product, "CRADLE")
replace eng=0 if eng==1 & strpos(Product, "HOIST")
replace eng=0 if eng==1 & strpos(Product, "HOIST")
replace eng=0 if eng==1 & Product== "ENGINE MOUNTING BOOMS"
replace eng=0 if eng==1 & Product=="ENGINE DRIVEN DRILLS"
// FAUX CORP worked specifically with engine gears, while

// I have commented out code that made zero changes

// eng =2 is for repairs, conversions, and modifications
replace eng=2 if eng==1 & strpos(Product, "REPAIR")
replace eng=2 if eng==1 & strpos(Product, "MVM")
replace eng=2 if eng==1 & strpos(Product, "CONV")
replace eng=2 if eng==1 & strpos(Product, "MODIF")
replace eng=2 if eng==1 & strpos(Product, "REMORK")
replace eng=2 if eng==1 & strpos(Product, "REBUILT")
replace eng=2 if eng==1 & strpos(Product, "REBUILT")
replace eng=2 if eng==1 & strpos(Product, "REMANUFACTUR")
replace eng=2 if eng==1 & strpos(Product, "REMFG")

// eng=3 is 3511
replace eng=3 if eng==1 & strpos(Product, "TURBINE")
//replace eng=2 if strpos(Product, "WATER WHEEL")
//replace eng=2 if strpos(Product, "WATERWHEEL")
//replace eng=2 if eng==1 &  strpos(Product, "TURBOGENERATOR")
//replace eng=2 if eng==1 &  strpos(Product, "TURBO GENERATOR")
//replace eng=2 if eng==1 &  strpos(Product, "STEAM")

// eng=4 is 3519
replace eng=4 if eng==1 & strpos(Product, "DIESEL")
//replace eng=4 if eng==1 & strpos(Product, "COMBUSTION")
replace eng=4 if eng==1 & strpos(Product, "GAS")
replace eng=4 if eng==1 & strpos(Product, "OIL")
replace eng=4 if eng==1 & strpos(Product, "KEROSENE")

// eng=5 is 3561
replace eng=5 if eng==1 & strpos(Product, "PUMP")
replace eng=5 if eng==1 & strpos(Product, "COMP")

// eng = 6 is for 3614
replace eng=6 if eng==1 & strpos(Product, "MOTOR")
replace eng=6 if eng==1 & strpos(Product, "GENERATOR")

// eng=7 is 3566
replace eng=7 if eng==1 & strpos(Product, "TRANSMISSION")
replace eng=7 if eng==1 & strpos(Product, "GEAR")
replace eng=7 if eng==1 & strpos(Product, "DRIVE")
// replace eng=7 if strpos(Product, "SPEEDRIVE")
//replace eng=7 if eng==1 & strpos(Product, "SHAFT")
replace eng=7 if eng==1 & strpos(Product, "CLUTCH")
replace eng=7 if eng==1 & strpos(Product, "BEARING")

// eng= 8 is for 3599 or 3641 but is also the catch all so... ya
//replace eng=8 if eng==8 & strpos(Product, "INVET")
replace eng=8 if eng==1 & strpos(Product, "CONTR")
replace eng=8 if eng==1 & strpos(Product, "START")
replace eng=8 if eng==1 & strpos(Product, "GAGE")
replace eng=8 if eng==1 & strpos(Product, "GAUGE")
replace eng=8 if eng==1 & strpos(Product, "THERMOMETER")
replace eng=8 if eng==1 & strpos(Product, "TEST")
replace eng=8 if eng==1 & strpos(Product, "EQUIPMENT")
replace eng=8 if eng==1 & strpos(Product, "PLUG")
// MAYBE:
replace eng=8 if eng==1 & strpos(Product, "ACCESS")

// eng= 9 is for 3591 valves and fittings
replace eng=9 if eng==1 & strpos(Product, "VALVE")
replace eng=9 if eng==1 & strpos(Product, "FITTING")

// zeng is for allowing the transportation industries to see all the machinery
// including engine. Called Zeng cause I'm getting all zengy with it (whatever that means!)
gen zeng =1 if eng!=. & eng!=0

// Miscellaneous needing checking
replace eng=0 if Product=="CHEMICAL FOAM ENGINES"
replace eng=0 if Product=="ENGINE CYLINDER REP-IR"
replace eng=0 if strpos(Product, "TANK")
replace eng=0 if strpos(Product, "LATHE")
replace eng=0 if Product=="REFRIGERATION ENGINES"

//aeng is for 3722
gen aeng=0 if strpos(Product, "AIRPLANE") | strpos(Product, "AIRCRAFT") | strpos(Product, "JET") | strpos(Product, "PLANE")
replace aeng=1 if aeng==0 & zeng==1
replace aeng=0 if aeng==1

// teng is for trucks and tractors
gen teng=0 if strpos(Product, "TRUCK") |strpos(Product, "TRACTOR") | strpos(Product, "ZENE")
replace teng=1 if zeng==1 & teng==0
replace eng=0 if teng==1
// double check tire engine definition

// ceng is for general cars and automobiles
gen ceng=0 if strpos(Product, "AUTO") | strpos(Product, "CAR") | strpos(Product, "VEHICLE")
replace ceng=1 if zeng==1 & ceng==0
replace eng=0 if ceng==1

// reng= is for trains and locomotives and railroads
gen reng=0 if  strpos(Product, "TRAIN") | strpos(Product, "LOCOM") |strpos(Product, "RAILROAD") |strpos(Product, "RAIL ROAD")
replace reng=11 if zeng==1 & reng==0
replace eng=0 if reng==1
```

eng variable made to capture "ENG" in data set

Removing inapplicable engine accessories and services

Seperation for Repair Variable

Seperation for SIC 3511

Seperation for SIC 3519

Seperation for SIC 3561

Seperation for SIC 3614

Seperation for SIC 3566

Seperation for SIC 3599 and 3541, but also serves as miscellaneous

Seperation for SIC3591

Compilation of all eng into a seperate varialbe, zeng

Separation of zeng for Airforce

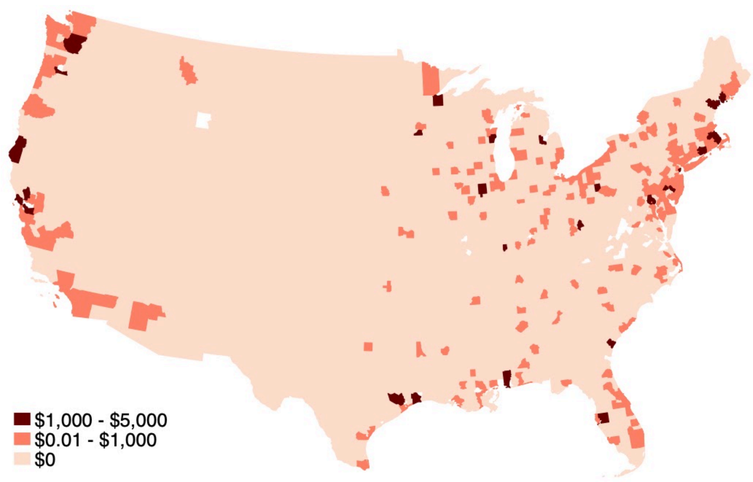
Separation of zeng for trucks

Separation of zeng for cars

Separation of zeng for trains

Findings:

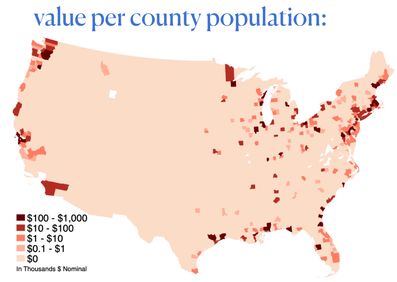
The main purpose of this data will be mapping the economic impacts of World War 2 to the county level. To best exemplify the impact of my coding on Professor Brunet's work, here is an actual map depicting how much money was spent by the U.S. government on ships and boats in terms of total dollars spent by county per capita:



While it is important for county level analysis of fiscal and aggregate multipliers to think in terms of population, that mindset should not be allowed to manipulate perception of what counties produced the largest total values, of sea vessels, during World War 2. Another surprising aspect of the ship and boat related war contracts was how. the Army supposedly had 1,307 Ship and Boat contracts, which felt like a large proportion of a total less than 4,000. However, upon further inspection, 42% of these contracts were for boats versus ships and the Army barely spent around a billion nominal 1946 U.S. dollars, while the Maritime Commission and Navy spent over $9.32 billion and $12 billion, respectively.

While my work does not directly lend itself to analytical regressions, as the ship and boat data has few factors that can be seen as causation for dollars spent, however I have over a six week period accounted for approximately 12% of the WW2 manufacturing costs accounted for by this dataset. The greatest take aways from this apprenticeship have been about how to use Stata and general data mining techniques.

As you can see, most of the spending occurred on the coasts, the Gulf of Mexico, and the Great Lakes, particularly in urban centers like Boston, New Orleans, and San Francisco. Initially I was surprised to see the presence of spending in the central south east of the country, however after more detailed analysis, it becomes clear that the per capita adjustment provides a bias for the rural areas. For instance, Morgan County Alabama inputted less than $20 million total value while Los Angeles County added around $1.3 billion, but due to division by population, they both have a per capita spending on boats and ships of less than $500. Below is a graph showing total county value added rather than value per county population:



References:

U.S. Shipbuilding History, Shipbuilding Records, Tim Colton. shipbuildinghistory.com/.

"History of SIC Codes." *SIC & NAICS Codes, Company Search, Business Lists - SICCODE.com*, 21 Jan. 2020, siccode.com/page/history-of-sic-codes.