Manufacturing Productivity in 2012

Mike Silva

April 2015

Introduction

This study is an attempt to update on the previous work of Brunot and Kurrea [1]. This paper uses data from the 2012 Economic Census to measure manufacturing sector labor productivity in American metropolitan areas. It first presents 2012 manufacturing productivity data for a broad range of metro areas and then explores possible determinants of productivity.

[1] http://eriedata.bd.psu.edu/AUBER%202012%20Paper–Honolulu–Productivity%20across%20MSAs%20%20FINAL.pdf

Study Data

2012 Economic Census

Consistent with Brunot and Kurrea's works this study uses data from the Econmic Census. The metro areas published in the 2012 Economic Census are those delineated by the Office of Management and Budget (OMB) in 2013 based in part on the results of the 2010 Census of Population and Housing. This data was downloaded from American FactFinder.[2]

[2] Online at http://factfinder.census.gov/

Local Area Personal Income and Employment

Earnings in each subsector as a share of total earnings in manufacturing in each MSA in 2007 was peviously used as a measure of industry mix. Data for this variable are from the Bureau of Economic Analysis's Regional Economic Information System (REIS). [3]

The BEA statistical areas are defined by OMB in bulletin no. 13-01 issued February 28, 2013, and the definitions are updated as new information warrants.[4] These definitions are consistent with those of the 2012 Economic Census.

[3] Table CA5N download from https://www.bea.gov/regional/downloadzip.cfm

[4] See https://www.bea.gov/regional/docs/msalist.cfm

American Community Survey

The American Community Survey (ACS) was used in estimating the educational attainment of the population. Data was downloaded from American FactFinder[5].

[5] Table S1501, Educational Attainment. Downloaded from http://factfinder.census.gov/

Population Estimates

The Census Bureau's Population Estimates Program estimates of the population for the metro areas was used in the measure of external economies of scale.

Annual Survey of State Government Tax Collections

The Census Bureau's Annual Survey of State Government Tax Collections for 2012 was used to estimate the business tax impacts. [6]

[6] Online at http://www.census.gov/govs/statetax/historical_data_2012.html

County Business Patterns

County Business Patterns data was used to scale the state state level business tax metrics referenced in the preceding section. [7]

[7] Downloaded from http://www.census.gov/econ/cbp/download/12_data/

Data Processing

First we read in the Economic Census data. The first line contains meta data so we skip it. We initially defined all variables as characters so as to not loose any data.

```
econ.census <- read.csv("data/ECN_2012_US_31A1_with_ann.csv",
    skip = 1, colClass = c(rep("character", 18)))
```

There are 381 records initially in the data set. First we will rename some of the variable names to make them clearer.

```
library(dplyr)
econ.census <- econ.census %>% rename(GeoFIPS = Id2) %>%
    rename(Metro = Geographic.area.name) %>% rename(Value.added = Value.added...1.000.) %>%
    rename(Production.workers.annual.hours = Production.workers.annual.hours..1.000.) %>%
    rename(Manufacturing.employment = Number.of.employees) %>%
    rename(Total.capital.expenditures = Total.capital.expenditures...1.000.)
```

Next we will clean up the metro area names removing the redundant "Metro Area" form each of their names.

```
econ.census <- econ.census %>% mutate(Metro = gsub(" Metro Area",
    "", Metro))
```

Next we will change the variable types to numeric. Nondisclosed data will create NA's in this process.

```
econ.census <- econ.census %>% mutate(Value.added = as.numeric(Value.added)) %>%
    mutate(Production.workers.annual.hours = as.numeric(Production.workers.annual.hours)) %>%
    mutate(Number.of.establishments = as.numeric(Number.of.establishments)) %>%
    mutate(Establishments.with.20.employees.or.more = as.numeric(Establishments.with.20.employees.or.more)
    mutate(Manufacturing.employment = as.numeric(Manufacturing.employment)) %>%
    mutate(Total.capital.expenditures = as.numeric(Total.capital.expenditures))
```

Productivity

Following Brunot and Kurrea's definition, productivity is defined as the value added per hour of production worker labor.

```
econ.census <- econ.census %>% mutate(Productivity = Value.added/Production.workers.annual.hours)
```

There are 8 out of 381 MSAs that do not have productivity data. We will remove them from the study data set.

```
econ.census <- econ.census[!is.na(econ.census$Productivity),
    ]
```

Capital

We also created the change in capital stock per production worker hour variable using the same definition used by Brunot and Kurrea.

```
econ.census <- econ.census %>% mutate(Capital.stock.per.production.worker.hour = Total.capital.expenditure
```

Of the 373 with productivity data, 9 did not have capital stock per production worker hour data.

Internal Economies of Scale

Employment per establishment is used however we calculated all the same internal economies of scale measures.

```
econ.census <- econ.census %>% mutate(Average.empt.per.establishment = Manufacturing.employment/Number.of.
    mutate(Percent.est.with.20.plus.employees = (Establishments.with.20.employees.or.more/Number.of.establ
        100) %>% mutate(Value.added.per.establishment = Value.added/Number.of.establishments)
```

There are 1 metro areas with no average employment per establishment.

External Economies of Scale

Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2014

All metropolitan statistical area delineations for the 2014 vintage population estimates series follow the Office of Management and Budget's statistical area issued in February 2013.

First we will pull in the data, rename the variables and change the type.

```
pop.est <- read.csv("data/PEP_2014_GCTPEPANNR.US24PR_with_ann.csv",
    colClass = c(rep("character", 14))) %>% rename(GeoFIPS = GC.target.geo.id2) %>%
    rename(Population = respop72012) %>% mutate(Population = as.numeric(Population)) %>%
    select(GeoFIPS, Population)
```

There are 1 record without data so we will drop it.

```
pop.est <- pop.est[!is.na(pop.est$Population),
    ]
```

The first 3 lines can be dropped as they are metadata and national figures (United States, and United States In metropolitan statistical area).

```
pop.est <- pop.est[4:nrow(pop.est), ]
```

There are 389 records in this data set.

Educational Attainment

This data comes from the 2011-2013 American Community Survey. Like any sample survey, the ACS is a household sample survey and is subject to response and coding error, as well as sampling error.

First we will pull in the data and rename the variables.

```
S1501 <- read.csv("data/ACS_13_3YR_S1501_with_ann.csv",
    colClass = c(rep("character", 231))) %>% rename(GeoFIPS = GEO.id2) %>%
    rename(High.school.graduates = HC01_EST_VC10) %>%
    rename(Associate.degree = HC01_EST_VC12) %>%
    rename(Bachelors.degree = HC01_EST_VC13) %>%
    rename(Graduate.degree = HC01_EST_VC14) %>%
    rename(Bachelors.or.higher = HC01_EST_VC17)
```

Next we will drop all unneeded variables and change the data from a character to a number.

```
S1501 <- S1501 %>% select(GeoFIPS, High.school.graduates,
    Associate.degree, Bachelors.degree, Graduate.degree,
    Bachelors.or.higher) %>% mutate(High.school.graduates = as.numeric(High.school.graduates)) %>%
    mutate(Associate.degree = as.numeric(Associate.degree)) %>%
    mutate(Bachelors.degree = as.numeric(Bachelors.degree)) %>%
    mutate(Graduate.degree = as.numeric(Graduate.degree)) %>%
    mutate(Bachelors.or.higher = as.numeric(Bachelors.or.higher))
```

There are initially 382 records in this data set. However the first record contains meta data so we dropped it.

```
S1501 <- S1501[2:nrow(S1501), ]
```

There are now 381 records in this data set.

Innovation

The patent data are from the U.S. Department of Commerce, United States Patent and Trademark Office. We used data for "utility" patents, the most common kind of patent. The Patent Office issues reports on the residence of the first named patent holder, which adds the spatial dimension needed for this study. They note that this is probably an imperfect indicator of the location where the patent work was actually done, since in some cases the first-named patent holder might live in a different place than the location of his/her employer where the work was actually done.

We will scrape the web and pull the county level data for the number of utility patents.

```
library(rvest)
pto <- html("http://www.uspto.gov/web/offices/ac/ido/oeip/taf/countyall/usa_county_gd.htm") %>%
    html_nodes("table") %>% html_table() %>% as.data.frame(.) %>%
    select(-Total)
```

Next we need to aggregate the data to the MSA level. Using the MSA definitions used by the BEA [8]

[8] Definitions downloaded from http://www.bea.gov/regional/docs/msalist.cfm

```
library(tidyr)
pto <- read.csv("data/metrolist.csv") %>% select(county.fips,
    msa.fips) %>% rename(FIPS.Code = county.fips) %>%
    merge(., pto) %>% select(msa.fips, X2012) %>%
    rename(GeoFIPS = msa.fips) %>% group_by(GeoFIPS) %>%
    summarise(Patents = sum(X2012))
```

There are 316 metros with patent data. We want to express the figures on a per 100,000 resident basis. We will merge in the population estimates and computed the scaled patent rate.

```
pto <- merge(pto, pop.est) %>% mutate(Patents.per.100000 = Patents/(Population/1e+05)) %>%
    select(-Population)
```

There are 315 metros with patent per 100,000 rates.

Industry Mix

We need the Manufacturing (500 line code) and Nondurable goods manufacturing (530 Line Code) for 2012.

```
CA5N <- read.csv("data/CA5N_2001_2013_MSA.csv",
    colClass = c(rep("character", 20))) %>% select(GeoFIPS,
    LineCode, X2012) %>% rename(Value = X2012) %>%
    filter(ifelse(LineCode %in% c("500", "530"),
        1, 0) == 1)
```

Some of the data has 'E' to represent an estimate. We have removed them from the data and then converted the data to a number.

```
CA5N <- CA5N %>% mutate(Value = as.numeric(gsub("E",
    "", Value))) %>% spread(LineCode, Value)

names(CA5N) <- c("GeoFIPS", "total", "nondurable")
CA5N$Percent.nondurable <- (CA5N$nondurable/CA5N$total) *
    100
```

There are 52 NA's in this data which were removed.

```
CA5N <- CA5N[!is.na(CA5N$Percent.nondurable),
    c("GeoFIPS", "Percent.nondurable")]
```

We then merge the 381 records with the economic census data.

Demographics

those in the 25-34 age group, not long out of college, and those in the 55-64 age group, those contemplating retirement.

```
S0101 <- read.csv("data/ACS_13_3YR_S0101_with_ann.csv",
    skip = 1, colClass = c(rep("character", 219))) %>%
    rename(GeoFIPS = Id2) %>% rename(Total.population = Total..Estimate..Total.population) %>%
    rename(People.25.to.29 = Total..Estimate..AGE...25.to.29.years) %>%
    rename(People.30.to.34 = Total..Estimate..AGE...30.to.34.years) %>%
    rename(People.55.to.59 = Total..Estimate..AGE...55.to.59.years) %>%
    rename(People.60.to.64 = Total..Estimate..AGE...60.to.64.years) %>%
    select(GeoFIPS, Total.population, People.25.to.29,
        People.30.to.34, People.55.to.59, People.60.to.64)
```

Initially there are 381 records. We will change the type to numeric:

```
S0101 <- S0101 %>% mutate(Total.population = as.numeric(Total.population)) %>%
    mutate(People.25.to.29 = as.numeric(People.25.to.29)) %>%
    mutate(People.30.to.34 = as.numeric(People.30.to.34)) %>%
    mutate(People.55.to.59 = as.numeric(People.55.to.59)) %>%
    mutate(People.60.to.64 = as.numeric(People.60.to.64))
```

Now we will create the group aggregates and compute the share of total population:

```
S0101 <- S0101 %>% mutate(People.25.to.34 = People.25.to.29 +
    People.30.to.34) %>% mutate(People.55.to.64 = People.55.to.59 +
    People.60.to.64) %>% select(GeoFIPS, Total.population,
    People.25.to.34, People.55.to.64)
```

There are 0 metros missing a share of population 25 to 34 and 0 metros missing a share of population 55 to 64 estimate.

Business Taxes

We use business taxes paid per employee for 2012. We consider both corporate net income taxes and "occupation and business taxes not elsewhere classified." Data was download using American FactFinder.

First we will read in the data and rename variables for clarity.

```
STC001 <- read.csv("data/STC_2012_STC001_with_ann.csv",
    skip = 1, colClass = c(rep("character", 34))) %>%
    rename(fipstate = Id2) %>% rename(Corporation.net.income.taxes = Income.Taxes...Corporation.Net.Income
    rename(Occupation.and.business.taxes.nec = License.Taxes...Occupation.and.Business.License..NEC) %>%
    select(fipstate, Geography, Corporation.net.income.taxes,
        Occupation.and.business.taxes.nec)
```

Then we need to change the variable types:

```
STC001 <- STC001 %>% mutate(Corporation.net.income.taxes = as.numeric(Corporation.net.income.taxes)) %>%
    mutate(Occupation.and.business.taxes.nec = as.numeric(Occupation.and.business.taxes.nec))
```

There are 4 states without corporate net income taxes and 0 states without occupational and business taxes.

These data were filled in with zeros (since the X in the data represents "Not applicable" and not missing values), and the United States line was dropped from the table.

```
STC001 <- STC001[STC001$Geography != "United States",
    ] %>% mutate(Corporation.net.income.taxes = ifelse(is.na(Corporation.net.income.taxes),
    0, Corporation.net.income.taxes)) %>% mutate(Occupation.and.business.taxes.nec = ifelse(is.na(Occupati
    0, Occupation.and.business.taxes.nec))
```

First we will merge in in the county business patterns employment totals(all NAICS, all Legal Form of Organization).

```
STC001 <- read.csv("data/cbp12st.txt", colClass = c(rep("character",
    84))) %>% filter(naics == "------") %>% filter(lfo ==
    "-") %>% mutate(emp = as.numeric(emp)) %>%
    select(fipstate, emp) %>% merge(STC001, .)
```

Now we can compute the business tax per employee rates

```
STC001 <- STC001 %>% mutate(Corporate.net.income.tax.per.worker = (Corporation.net.income.taxes *
    1000)/emp) %>% mutate(Other.business.taxes.per.worker = (Occupation.and.business.taxes.nec *
    1000)/emp) %>% mutate(Business.taxes.per.worker = Corporate.net.income.tax.per.worker +
    Other.business.taxes.per.worker) %>% select(-Corporation.net.income.taxes,
    -Occupation.and.business.taxes.nec, -emp)
```

Exploratory Analysis

Productivity

How much does metro manufacturing productivity vary? It ranges from $23 per hour of labor in California-Lexington Park, MD to $1027 in Duluth, MN-WI. That is a 46-fold difference. The following table summarizes the manufacturing productivity values:

	Productivity
1	Min. : 22.58
2	1st Qu.: 101.60
3	Median : 130.43
4	Mean : 156.45
5	3rd Qu.: 171.44
6	Max. :1027.35

Table 1: 2012 Manufacturing Productivity Summary Statistics

Now to examine the top and the bottom of the productivity spectrum

	Metro	Productivity
1	Duluth, MN-WI	1027.35
2	Lake Charles, LA	973.02
3	Baton Rouge, LA	795.15
4	Albuquerque, NM	766.57
5	Beaumont-Port Arthur, TX	693.80
6	Lima, OH	497.15

Table 2: Metros with Highest Manufacturing Productivity

	Metro	Productivity
368	Hilton Head Island-Bluffton-Beaufort, SC	65.79
369	Elkhart-Goshen, IN	62.25
370	Gadsden, AL	60.99
371	Lawton, OK	51.88
372	San Angelo, TX	50.79
373	California-Lexington Park, MD	22.58

Table 3: Metros with Lowest Manufacturing Productivity

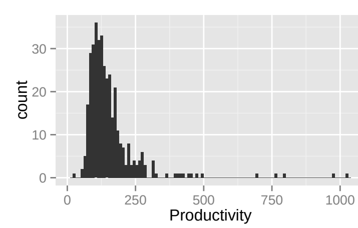

Figure 1: Manufacturing Productivity

Study Data

```
study.data <- econ.census %>% select(GeoFIPS,
    Metro, Productivity, Manufacturing.employment,
    Average.empt.per.establishment, Percent.est.with.20.plus.employees,
    Value.added.per.establishment, Capital.stock.per.production.worker.hour) %>%
    merge(., pop.est, all.x = TRUE) %>% merge(.,
```

```
    S0101, all.x = TRUE) %>% merge(., S1501, all.x = TRUE) %>%
    merge(., pto, all.x = TRUE) %>% merge(., CA5N,
    all.x = TRUE) %>% select(-Patents, -Total.population)
```

In order to merge in the business taxes data we followed Brunot and Kurre's method of using the state rate for the first state listed in the MSA's name. To do that I will use a function to pull out the state abbreviation:

```
get.state.abbr <- function(name) {
    name <- strsplit(name, ",")
    name <- name[[1]][2]  # Get the state abbreviation
    name <- substr(name, 1, 3)  # Get the first 3 characters from the left
    substr(name, 2, 3)  # Get the last 2 characters from the right
}

study.data <- study.data %>% group_by(Metro) %>%
    mutate(state.abbr = get.state.abbr(Metro))
```

Next I will pull in a crosswalk that will let me get the state fips code.[9]

[9] Found online at: http://www.bls.gov/cew/cewedr10.htm

```
study.data <- read.csv("data/state.fips.csv",
    colClass = c(rep("character", 3))) %>% rename(state.abbr = Postal.Abbr.) %>%
    rename(fipstate = FIPS.Code) %>% select(state.abbr,
    fipstate) %>% merge(study.data, .) %>% merge(.,
    STC001) %>% select(-fipstate, -state.abbr,
    -Geography) %>% arrange(-Productivity)
```

Regresssion Model

```
lm.data <- study.data %>% select(Productivity,
    Population, Average.empt.per.establishment,
    Patents.per.100000, Bachelors.degree, People.55.to.64,
    Corporate.net.income.tax.per.worker, Capital.stock.per.production.worker.hour,
    Percent.nondurable) %>% mutate(Population.Squared = Population *
    Population) %>% rename(Employment.per.Establishment = Average.empt.per.establishment) %>%
    rename(Patents.per.Capita = Patents.per.100000) %>%
    rename(Percent.Bachelor.Degree = Bachelors.degree) %>%
    rename(Percent.55.to.64.Yr.Olds = People.55.to.64) %>%
    rename(Corp.Income.Tax = Corporate.net.income.tax.per.worker) %>%
    rename(Capital.per.Prdn.Worker.Hour = Capital.stock.per.production.worker.hour) %>%
    mutate(Capital.per.Prdn.Worker.Hour.Squared = Capital.per.Prdn.Worker.Hour *
        Capital.per.Prdn.Worker.Hour) %>% rename(Percent.NonDurables = Percent.nondurable) %>%
    select(Productivity, Population, Population.Squared,
```

```
        Employment.per.Establishment, Patents.per.Capita,
        Percent.Bachelor.Degree, Percent.55.to.64.Yr.Olds,
        Corp.Income.Tax, Capital.per.Prdn.Worker.Hour,
        Capital.per.Prdn.Worker.Hour.Squared,
        Percent.NonDurables)

lm.data <- lm.data[complete.cases(lm.data), ]

lm.fit <- lm(Productivity ~ ., lm.data)
```

Model Results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.3398	64.2325	0.11	0.9091
Population	0.0000	0.0000	1.73	0.0846
Population.Squared	-0.0000	0.0000	-1.46	0.1449
Employment.per.Establishment	0.4334	0.3363	1.29	0.1988
Patents.per.Capita	-0.0874	0.1845	-0.47	0.6362
Percent.Bachelor.Degree	2.2326	1.7088	1.31	0.1926
Percent.55.to.64.Yr.Olds	-0.1447	3.6643	-0.04	0.9685
Corp.Income.Tax	-0.0084	0.0308	-0.27	0.7865
Capital.per.Prdn.Worker.Hour	7.6149	0.8924	8.53	0.0000
Capital.per.Prdn.Worker.Hour.Squared	-0.0211	0.0039	-5.40	0.0000
Percent.NonDurables	0.3766	0.3321	1.13	0.2579