A Statistical Analysis of Regional Economic Impacts of the Manufacturing USA Program

...

SPS 2020 Summer Internship Symposium

AIP | American Institute of Physics

Background

Who: Max Dornfest, Mather Policy - NIST

Where: Fremont, California

What: Senior, UC Berkeley, Research Affiliate, LBNL
Majors: Physics, Political Science
Minor: Public Policy

Why: (should you care) Using econometrics and astrophysics tools to analyze ROI for Manufacturing USA program

Where I worked this summer

National Institute of Standards and Technology's
Office of Advanced Manufacturing

- Major program is Manufacturing USA, a network of 15 public-private manufacturing innovation institutes funded by nine government agencies.

- Manufacturing USA institutes convene business, academia, and stakeholders to work on those hard problems that can not be solved alone.

Research Project

Zigs and Zags

- Started with rare earth metals and electronics components.
- SARS-Covid-19 helped shape the direction of my work.
- Along the way - reviewed a grant, made progress to creating industry “one-pagers” starting with healthcare.
- All of this helped me focus my data analysis efforts on NIIMBL - The National Institute for Innovation in Manufacturing Biopharmaceuticals, one of the 15 Manufacturing USA institutes

Developing paper

Data

- Collaborated with Nico Thomas and Stephen Campbell.
- Used EMSI data a labor market analytics firm.

Language

- Python 3.0 in JupyterLab GUI housed in an Anaconda framework.
- Libraries: StatsModels, Pandas, Stargazer, etc.

Developing paper

Analysis

Standard Regressions (OLS)

- easy to implement.

Regression Discontinuity

- identifies chronological effect of treatment.

Synthetic Control

- allows comparison amongst many data sets.

Expectation

Standard Regressions (OLS)

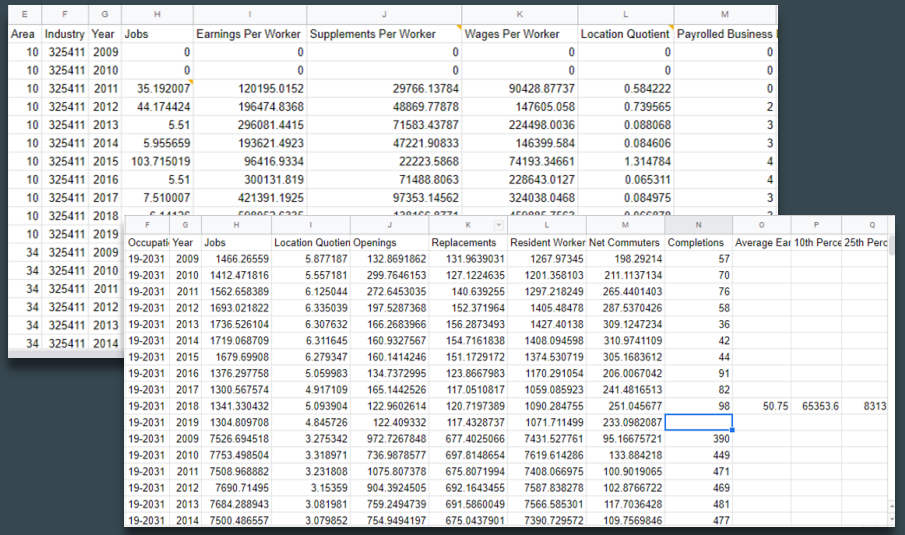
- Highly dependant on variables present in data
- Would like to tease out geographic and chronological relationship
- Each sub-institute was established at a different time

Data (Varied Variables)

Two sets of Data

Industries

North American Industry Classification System (NAICS) is the standard used by Federal statistical agencies



Area	Industry	Year	Jobs	Earnings Per Worker	Supplements Per Worker	Wages Per Worker	Location Quotient	Payrolled Business
10	325411	2009	0	0	0	0	0	0
10	325411	2010	0	0	0	0	0	0
10	325411	2011	35.192007	120195.0152	29766.13784	90428.87737	0.584222	0
10	325411	2012	44.174424	196474.8368	48869.77878	147605.058	0.739565	2
10	325411	2013	5.51	296081.4415	71583.43787	224498.0036	0.088068	3
10	325411	2014	5.955659	193621.4923	47221.90833	146399.584	0.084606	3
10	325411	2015	103.715019	96416.9334	22223.5868	74193.34661	1.314784	4
10	325411	2016	5.51	300131.819	71488.8063	228643.0127	0.065311	4
10	325411	2017	7.510007	421391.1925	97353.14562	324038.0468	0.084975	3
10	325411	2018						
10	325411	2019						
34	325411	2009						
34	325411	2010						
34	325411	2011						
34	325411	2012						
34	325411	2013						
34	325411	2014						

Occupati	Year	Jobs	Location Quotien	Openings	Replacements	Resident Worker	Net Commuters	Completions	Average Ear	10th Perce	25th Perc
19-2031	2009	1466.26559	5.877187	132.8691862	131.9639031	1267.97345	198.29214	57			
19-2031	2010	1412.471816	5.557181	299.7646153	127.1224635	1201.358103	211.1137134	70			
19-2031	2011	1562.658389	6.125044	272.6453035	140.639255	1297.218249	265.4401403	76			
19-2031	2012	1693.021822	6.335039	197.5287368	152.371964	1405.48478	287.5370426	58			
19-2031	2013	1736.526104	6.307632	166.2683966	156.2873493	1427.40138	309.1247234	36			
19-2031	2014	1719.068709	6.311645	160.9327567	154.7161838	1408.094598	310.9741109	42			
19-2031	2015	1679.69908	6.279347	160.1414246	151.1729172	1374.530719	305.1683612	44			
19-2031	2016	1376.297758	5.059983	134.7372995	123.8667983	1170.291054	206.0067042	91			
19-2031	2017	1300.567574	4.917109	165.1442526	117.0510817	1059.085923	241.4816513	82			
19-2031	2018	1341.330432	5.093904	122.9602614	120.7197389	1090.284755	251.045677	98	50.75	65353.6	8313
19-2031	2019	1304.809708	4.845726	122.409332	117.4328737	1071.711499	233.0982087				
19-2031	2009	7526.694518	3.275342	972.7267848	677.4025066	7431.527761	95.16675721	390			
19-2031	2010	7753.498504	3.318971	736.9878577	697.8148654	7619.614286	133.884218	449			
19-2031	2011	7508.968882	3.231808	1075.807378	675.8071994	7408.066975	100.9019065	471			
19-2031	2012	7690.71495	3.15359	904.3924505	692.1643455	7587.838278	102.8766722	469			
19-2031	2013	7684.288943	3.081981	759.2494739	691.5860049	7566.585301	117.7036428	481			
19-2031	2014	7500.486557	3.079852	754.9494197	675.0437901	7390.729572	109.7569846	477			

Occupations

Standard Occupational Classification (SOC) system is a federal standard used to classify workers into occupational categories

Industries

E	F	G	H	I	J	K	L	M
Area	Industry	Year	Jobs	Earnings Per Worker	Supplements Per Worker	Wages Per Worker	Location Quotient	Payrolled Business
10	325411	2009	0	0	0	0	0	0
10	325411	2010	0	0	0	0	0	0
10	325411	2011	35.192007	120195.0152	29766.13784	90428.87737	0.584222	0
10	325411	2012	44.174424	196474.8368	48869.77878	147605.058	0.739565	2
10	325411	2013	5.51	296081.4415	71583.43787	224498.0036	0.088068	3
10	325411	2014	5.955659	193621.4923				
10	325411	2015	103.715019	96416.9334				
10	325411	2016	5.51	300131.819				
10	325411	2017	7.510007	421391.1925				
10	325411	2018	6.14126	598052.6335				
10	325411	2019	4.6724605	455769.5163				
34	325411	2009	5067.489538	210498.0898				
34	325411	2010	4832.083198	227156.6118				
34	325411	2011	4685.477477	242851.923				
34	325411	2012	4514.510024	266146.8576				
34	325411	2013	4259.510102	267929.0663				
34	325411	2014	4299.034185	292367.0885				

Important Takeaways

- Area and Industry/Occupation are “nice” *categorical variables*.
- Jobs are not integers.
- With industry dataset I have access to pay, broken down into *three* variables!

Occupations

F	G	H	I	J	K	L	M	N	O	P	Q
Occupati	Year	Jobs	Location Quotien	Openings	Replacements	Resident Worker	Net Commuters	Completions	Average Ear	10th Perce	25th Perc
19-2031	2009	1466.26559	5.877187	132.8691862	131.9639031	1267.97345	198.29214	57			
19-2031	2010	1412.471816	5.557181	299.7646153	127.1224635	1201.358103	211.1137134	70			
19-2031	2011	1562.658389	6.125044	272.6453035	140.639255	1297.218249	265.4401403	76			
19-2031	2012	1693.021822	6.335039	197.5287368	152.371964	1405.48478	287.5370426	58			
19-2031	2013	1736.526104	6.307632	166.2683966	156.2873493	1427.40138	309.1247234	36			
19-2031	2014	1719.068709	6.311645	160.9327567	154.7161838	1408.094598	310.9741109	42			
19-2031	2015	1679.69908	6.279347	160.1414246	151.1729172	1374.530719	305.1683612	44			
19-2031	2016	1376.297758	5.059983	134.7372995	123.8667983	1170.291054	206.0067042	91			
19-2031	2018	1341.330432	5.093904	122.9602614	120.7197389	1090.284755	251.045677	98	50.75	65353.6	831
19-2031	2019	1304.809708	4.845726	122.409332	117.4328737	1071.711499	233.0982087				
19-2031	2010	7753.498504	3.318971	736.9878577	697.8148654	7619.614286	133.884218	449			
19-2031	2011	7508.968882	3.231808	1075.807378	675.8071994	7408.066975	100.9019065	471			
19-2031	2012	7690.71495	3.15359	904.3924505	692.1643455	7587.838278	102.8766722	469			
19-2031	2013	7684.288943	3.081981	759.2494739	691.5860049	7566.585301	117.7036428	481			
19-2031	2014	7500.486557	3.079852	754.9494197	675.0437901	7390.729572	109.7569846	477			

Results

R-squared = 0.643

The Durbin-Watson stat is used to test for autocorrelation.

If the Durbin-Watson stat is between 1.5 and 2.5, it is a good sign that there is no autocorrelation

OLS Regression Results

Dep. Variable:	Location Quotient	**R-squared:**	0.643
Model:	OLS	**Adj. R-squared:**	0.643
Method:	Least Squares	**F-statistic:**	1487.
Date:	Mon, 03 Aug 2020	**Prob (F-statistic):**	0.00
Time:	15:42:02	**Log-Likelihood:**	-2246.5
No. Observations:	1650	**AIC:**	4497.
Df Residuals:	1648	**BIC:**	4508.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Jobs	-1.648e-07	5.08e-08	-3.244	0.001	-2.64e-07	-6.52e-08
Year	0.0006	1.21e-05	53.035	0.000	0.001	0.001

Omnibus:	1314.465	**Durbin-Watson:**	0.147
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	27521.677
Skew:	3.668	**Prob(JB):**	0.00
Kurtosis:	21.615	**Cond. No.**	248.

Results

Occupation = Top regression
Industry = Bottom regression

Area [T.10] = Delaware

Delaware is the one state with stable signal across both datasets.

NIIMBL = In Delaware

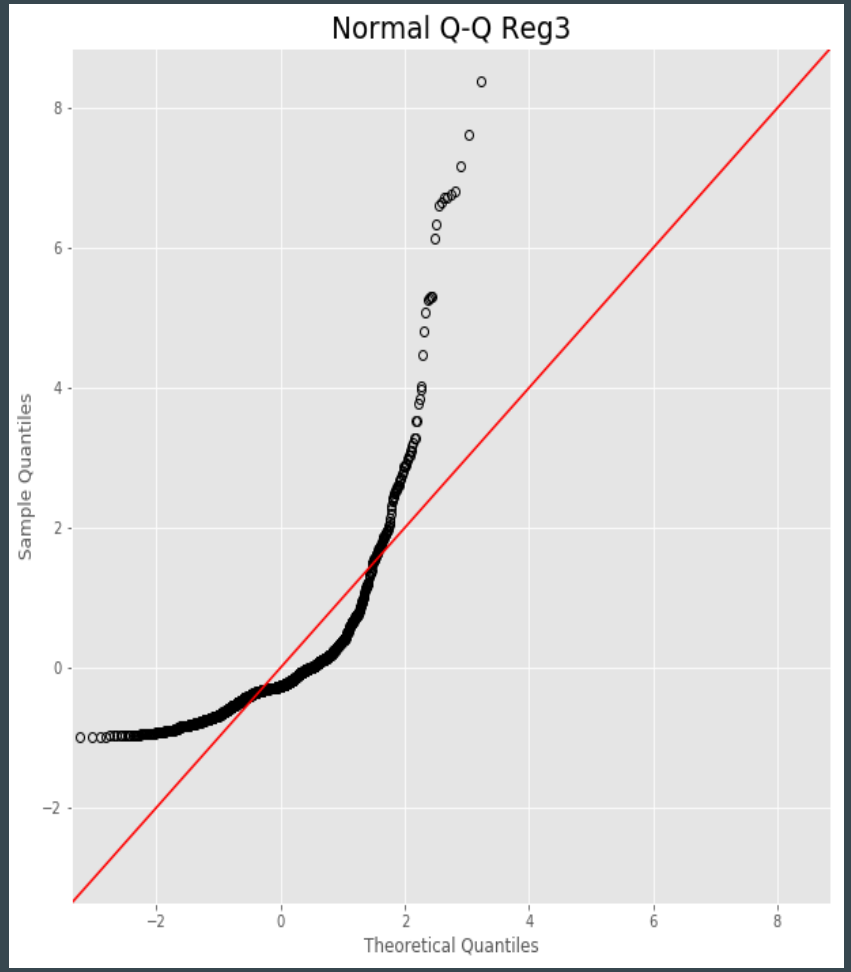
This is a promising start.

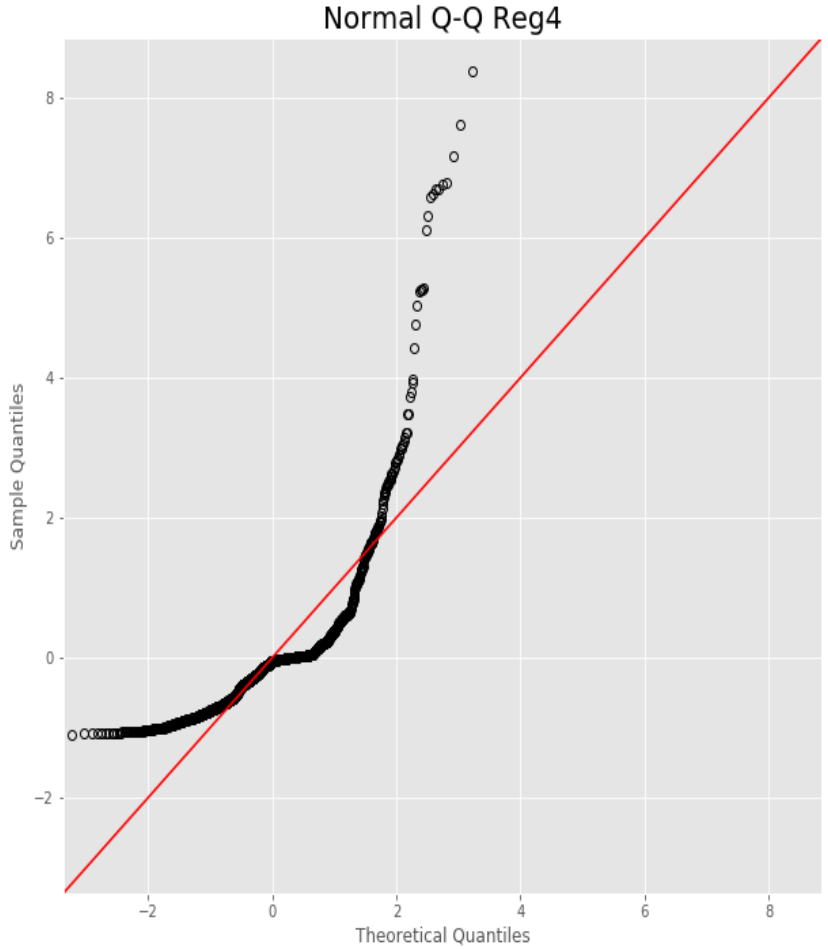
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.2809	0.001	-223.613	0.000	-0.283	-0.278
C(Area)[T.10]	0.4282	0.072	5.954	0.000	0.287	0.569
C(Area)[T.24]	0.4231	0.056	7.580	0.000	0.314	0.533
C(Area)[T.34]	0.4213	0.078	5.390	0.000	0.268	0.575
C(Area)[T.42]	0.1318	0.019	6.925	0.000	0.095	0.169
zscore(Year)	-0.0227	0.024	-0.934	0.350	-0.070	0.025

Omnibus:	1265.574	**Durbin-Watson:**	0.145

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.5374	0.010	-53.249	0.000	-0.557	-0.518
C(Area)[T.10]	0.4716	0.140	3.360	0.001	0.197	0.747
C(Area)[T.24]	0.9261	0.153	6.060	0.000	0.627	1.226
C(Area)[T.34]	1.0316	0.137	7.546	0.000	0.764	1.300
C(Area)[T.42]	0.2577	0.068	3.814	0.000	0.125	0.390
zscore(Year)	-0.0803	0.050	-1.608	0.108	-0.178	0.018

Omnibus:	74.129	**Durbin-Watson:**	0.300
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	127.827
Skew:	1.271	**Prob(JB):**	1.75e-28
Kurtosis:	4.684	**Cond. No.**	5.83


Normal Q-Q Reg3
Sample Quantiles
Theoretical Quantiles
8
6
4
2
0
−2
−2
0
2
4
6
8


Normal Q-Q Reg4
Sample Quantiles
Theoretical Quantiles
8
6
4
2
0
−2
−2
0
2
4
6
8

Concluding Thoughts

Great learning experience

Paper still on going

Recently asked to interview for a quantitative analysis job.

Conclusion: Paper was a great serendipitous choice.

Thank you for the great summer!

Jessica Strickler, Administrative Officer

Robert Rudnitsky, Physicist, Associate Director for Policy

Lisa Fronczek, Electronics Engineer, Manufacturing USA Competition Manager

Steve Campbell, Economist, Manufacturing Extension Partnership

Nico Thomas, Performance Analyst, Manufacturing Extension Partnership

and All of SPS!