Thomas Marron

Quantitatively Testing the Effects of Baumol's Cost Disease in the Service Economy

Econ 6060 – Spring 2024

Table of Contents

(1) Abstract

Baumol's Cost Disease dictates that unproductive industries must be compensated enough to remain competitive with other occupations in the labor market, causing a disparity between productivity and wages in the unproductive industry. I test this theory in the service economy by using panel data for ten subsectors from the Bureau of Labor Statistics to determine how statistically significant productivity is in a regression analysis on hourly compensation. The findings go against the existence of the cost disease, but I add caveats and determine a strategy to further test the hypothesis.

(2) Introduction

In 1965, William Baumol and William Bowen, both economists at Princeton, wrote an article about the performing arts (namely theater) and their economic problems. In the article, they identify a productivity problem with the industry: in the performing arts, the work *is* the service (Baumol, Bowen, 1965). This will be extrapolated on more in my review of related literature, but, in short, this means that technological progress over time does very little to enhance the productivity of the performing arts. Thus, if people are paid in direct proportion to their output, then we should expect that the unproductive industries should be left behind in terms of wages. Baumol and Bowen take note of that seemingly logical conclusion, but point out that, wages must increase relatively in line with the rest of the economy, otherwise the actor or actress would be incentivized to switch jobs to industries with higher pay. This phenomenon, colloquially referred to as the Baumol Effect or Baumol's Cost Disease, can be transferred to other industries where technological progress does not drive productivity. In figure 1, we note that educational costs and healthcare costs have greatly outpaced inflation, which can, in part, be explained by the Baumol effect. I will elaborate on these examples in the review of related literature.

In this paper, I test empirically to what extent Baumol's Cost Disease exists, in the modern day, in the service industry using data from the Bureau of Labor Statistics (BLS). In line with the theory, I perform regressions using the statistical software *E-Views* to determine the relationship between productivity and hourly compensation in the service industry. My techniques are expanded upon in my econometric methodology and testing.

(3) Review of Related Literature

I became interested in the Baumol Effect after reading a summary of his works, focusing on both the Baumol Effect and on his theories of entrepreneurship (Tom Butler-Bowdon, 2017). After finding both time series and cross-sectional data from the BLS, I researched deeper. As noted above, the Baumol Effect specifically started as a defined phenomenon after the published paper on the performing arts and the necessity of increasing costs as a result of increasing input costs (Baumol and Bowen, 1965). From there, in 1967 Baumol refined his argument, putting the cost disease into a model in "Macroeconomics of Unbalanced Growth" and extending the effects of the cost disease to include restaurants, leisure activities, and public works projects (Baumol and Bowen, 1967).

Baumol gained popularity as a result of his research and this theory eventually gained both admirers and critics. In 2000, Alan Kruger interviewed Baumol about the cost disease and the role of welfare economics and government intervention in the markets, where Baumol cited the great expansion of spending on healthcare and education, the most common examples to point to of the cost disease (Kruger 2001). The theory has also received a fair share of critics, most notably Tyler Cowen of George Mason University, who argues the goods themselves change over time (as an example, televisions in the 1980s are very different from televisions today), and thus any comparison between the two does not properly isolate the cost disease, making it difficult to prove the existence of the effect (Cowen 1996).

Specifically in healthcare and education, Baumol, backed by quantitative references, argued for the existence the effect in the United States health industry (Baumol 1993). Again in 2012, outside researchers empirically tested the effects by determining the effects of health care cost growth when wage-increases exceeded productivity gains (Bates and Santerre, 2012). They concluded statistically significant effects in both time and state-fixed tests, suggesting the existence of the Baumol Effects, and a recent test confirms the hypothesis empirically (Hartwig and Krämer, 2023).

(4) Econometric Methodology and Data Acquisition

The main question I wish to answer is “to what extent is productivity statistically significant in determining hourly compensation in specific industries?” In my data collection, I used times series data from 1978-2022 from ten different industries, creating a panel data set. In some cases, data for the first few years is nonexistent, but in all cases, we have at least twenty years’ worth of time series data to analyze.

The data from the BLS contained 79 different sectors and industries, all containing data averages on labor productivity, output per worker, real sectoral output, hours worked, employment, hourly compensation, and unit labor costs, organized to an index such that the value for the year 2012 is always 100 (see Figure 2 for an example of what the exported data looks like in Excel).

Every sector or industry also contains a number according to the North American Industry Classification System (NAICS). Every sector contains two digits; subsector, three digits; industry group, four digits; and industry, five digits. Of the 79 different sectors and industries, my data set contained multiple industries within a certain subsector, and in some cases, only one sector. For example, inside newspapers (NAICS 5111), we are given newspaper publishers (NAICS 51111) and periodical publishers (NAICS 51112). I selectively pruned data in this case to both make the data simpler to handle and the outcomes easier to see. For example, I chose to include the publishing industry (NAICS 511) in my regression analysis, which encompasses all the above-mentioned industry groups and industries. Below are the ten industries I selected to test:

- 22 Utilities
- 481 Air transportation
- 484 Truck transportation
- 491 Postal service
- 493 Warehousing and storage
- 511 Publishing Industries
- 515 Broadcasting
- 6215 Medical Laboratories
- 722 Food Service and drinking places
- 8111 Automotive repair and maintenance

In addition to taking data from the BLS, I also added other relevant variables to prevent omitted variable bias: minimum wage, core consumer price index, unemployment rate, federal funds rate, and GDP; all taken from the Federal Reserve Economic Database of Saint Louis (FRED).

(5) Empirical Testing

I plan on testing Baumol's hypothesis by examining the relationship between productivity and hourly compensation. If the Baumol Effect is true, then in a productive industry, productivity would be closely linked to hourly compensation, and thus statistically significant in a regression. On the other hand, and more interestingly, in an unproductive industry we should expect productivity to be statistically insignificant in a regression determining hourly compensation. This insignificance is due to wages being infected with the cost disease. Otherwise, the theory states, a worker will be incentivized to leave their job for a better paying one.

To begin the empirical testing of the theory, I first ran a pooled least squares regression to test the relationship between hourly compensation and the other variables included in the data set. Below is the regression output transferred from EViews after removing data for 'output per worker' and 'unit labor costs' due to multicollinearity concerns. In the final regressions, I removed all insignificant variables one at a time until I was left with all statistically significant variables. A full list of variables in my first regressions are in the Appendix. After cleaning the model up, we can easily see that productivity, with a p-value of 0.000 compared to $\alpha = 5\%$, is statistically significant in our model in determining hourly compensation.

Dependent Variable: HOURLY_COMPENSATION_?
Method: Pooled Least Squares
Date: 04/10/24 Time: 16:27
Sample (adjusted): 1988 2022
Included observations: 35 after adjustments
Cross-sections included: 10
Total pool (unbalanced) observations: 333
Convergence achieved after 8 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-35.75932	23.00160	-1.554645	0.1210
PRODUCTIVITY_?	0.237007	0.029591	8.009369	0.0000
EMPLOYMENT_?	0.511552	0.078475	6.518673	0.0000
HOURS_?	-0.388753	0.068297	-5.692068	0.0000
AR(1)	1.026447	0.006975	147.1537	0.0000

Fixed Effects (Cross)	
22--C	-13.60694
481--C	3.016359
484--C	4.640178
491--C	-0.743189
493--C	62.19052
511--C	-49.39531
515--C	10.36545
6215--C	22.64036
722--C	-16.53253
8111--C	-8.499572

Effects Specification

Cross-section fixed (dummy variables)

Root MSE	2.828446	R-squared	0.990182
Mean dependent var	84.22446	Adjusted R-squared	0.989782
S.D. dependent var	28.58830	S.E. of regression	2.889846
Akaike info criterion	5.001416	Sum squared resid	2664.035
Schwarz criterion	5.161518	Log likelihood	-818.7358
Hannan-Quinn criter.	5.065258	F-statistic	2474.784
Durbin-Watson stat	1.660568	Prob(F-statistic)	0.000000

Now, I will separate the 10 industries into three groups, high productivity, low productivity, and medium productivity. See figure 2 to look at the data. I will calculate the average growth rate and separate them into buckets.

Industry	NAICS Classification Number	Productivity growth rate per year
Utilities	22	2.80%
Air Transportation	481	3.10%
Truck Transportation	484	0.91%
Postal Service	491	-0.24%
Warehousing and Storage	493	0.34%
Publishing Industries	511	8.98%
Broadcasting	515	5.83%
Medical Laboratories	6215	2.79%
Food Service and Drinking Places	722	0.85%
Automotive Repair and Maintenance	8111	1.06%

High growth rate industries: 3%+ yearly annual growth

- Air Transportation
- Publishing Industries
- Broadcasting

Medium growth rate industries: 1-3% yearly annual growth

- Utilities
- Medical Laboratories
- Automotive Repair and Maintenance

Low growth rate industries: 1%- yearly annual growth

- Truck Transportation
- Postal Service
- Warehousing and Storage
- Food Service and Drinking Places

Now we should expect that productivity in the low growth rate industries will be insignificant and significant in high growth rate industries. Once again using pooled least squares, I will test the effectiveness of productivity on hourly compensation on the most and least productive industries.

Test for "High Growth Rate" Industries:

Dependent Variable: HOURLY_COMPENSATION_?
Method: Pooled Least Squares
Date: 04/17/24 Time: 09:02
Sample (adjusted): 1988 2022
Included observations: 35 after adjustments
Cross-sections included: 3
Total pool (balanced) observations: 105
Convergence achieved after 8 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-15.59264	17.32148	-0.900191	0.3702
PRODUCTIVITY_?	0.240492	0.039841	6.036341	0.0000
EMPLOYMENT_?	0.614932	0.132664	4.635269	0.0000
HOURS_?	-0.332139	0.091562	-3.627486	0.0005
STICKY_CPI	-0.057061	0.011933	-4.781704	0.0000
AR(1)	1.049165	0.013092	80.14027	0.0000
Fixed Effects (Cross)				
481--C	11.28454			
511--C	-21.67754			
515--C	10.39300			

Effects Specification

Cross-section fixed (dummy variables)

Root MSE	3.225110	R-squared	0.992123
Mean dependent var	81.67158	Adjusted R-squared	0.991554
S.D. dependent var	36.51194	S.E. of regression	3.355469
Akaike info criterion	5.332192	Sum squared resid	1092.140
Schwarz criterion	5.534398	Log likelihood	-271.9401
Hannan-Quinn criter.	5.414130	F-statistic	1745.275
Durbin-Watson stat	1.795925	Prob(F-statistic)	0.000000

The p-value of the t-statistic is 0.000, after removing insignificant variables and adjusting for serial correlation, thus confirming our hypothesis.

At the same time, taking the "low growth rate" industries, we should expect productivity to be statistically insignificant.

Dependent Variable: HOURLY_COMPENSATION_?
Method: Pooled Least Squares
Date: 04/17/24 Time: 09:01
Sample (adjusted): 1989 2022
Included observations: 34 after adjustments
Cross-sections included: 4
Total pool (unbalanced) observations: 126
Convergence achieved after 11 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-71.22153	99.66117	-0.714637	0.4763
PRODUCTIVITY_?	0.373510	0.086508	4.317630	0.0000
EMPLOYMENT_?	0.737591	0.142695	5.168997	0.0000
HOURS_?	-0.609007	0.130161	-4.678859	0.0000
AR(1)	1.284019	0.102624	12.51183	0.0000
AR(2)	-0.271037	0.104722	-2.588153	0.0109
Fixed Effects (Cross)				
484--C	-21.39951			
491--C	-12.07743			
493--C	84.22370			
722--C	-41.50791			

Effects Specification

Cross-section fixed (dummy variables)

Root MSE	2.329370	R-squared	0.988775
Mean dependent var	86.18721	Adjusted R-squared	0.988008
S.D. dependent var	22.07387	S.E. of regression	2.417302
Akaike info criterion	4.671930	Sum squared resid	683.6716
Schwarz criterion	4.874522	Log likelihood	-285.3316
Hannan-Quinn criter.	4.754237	F-statistic	1288.288
Durbin-Watson stat	1.844893	Prob(F-statistic)	0.000000

As shown in the regression run above, however, productivity is statistically significant in this case. Although the t-statistic is not as high as in the “productive” pool, using a pooled least squares, the p-value is still 0.0000.

(6) Conclusion

Using panel data, I found statistical significance of productivity in both the more productive and less productive industries. This suggests that Baumol's Cost Disease is not true in for the service economy, as tested by the above ten industries. The findings confuse me, as other empirical tests tend to confirm the effects, at least the studies that I read in my review of literature. That being said, the majority of the other studies focused specifically on the health care industry and looked at the effect as determined by the end result: the price consumers face. In my tests I looked more holistically at the service economy, used a smaller sample size, and also tested the relationship between productivity and wages, instead of productivity and consumer costs. While we expect consumer costs and wages to be highly positively correlated, some differences may exist, contributing to the disparity.

I also believe that this model hardly scratches the surface in the complexity of Baumol's cost disease. Therefore, this summer I plan on updating and expanding upon my research with three key changes. First, I will lengthen my review or research, including more on Baumol's previous writings, general reactions to his writings, other empirical testing of the hypothesis, and labor economics more broadly. The goal is to become an expert on the theoretical standpoint of Baumol's Cost Disease and to be able to include every variable that could be relevant.

Second, I will increase my sample size. In this study, I merely looked at ten sectors or industries, a long way away from encompassing the entire service economy. I would have liked to find data on manufacturing, finance, or education, but I could not find updated data on the BLS website. In order to definitively test the entire economy, a broader search with more data is needed.

Note: I found out over the summer that productivity is measured as output/wage according to the BLS. I tried to find a workaround without using something expressed in wage (because that would run into problems in a regression where wage is the dependent variable), but will probably have to move to employment as a means of testing.

Third, I want to improve my methodological techniques, including getting better with panel data. I never worked with panel data before starting this study. I actually had to teach myself the theory and practice of pooled least squares during the semester. While I made a great deal of progress in being comfortable working with such a data set, I still have more to learn, and I believe I could run more accurate regressions with the knowledge I have now.

(7) References

Bates, Laurine, and Rexford Santerre. "Does the U.S. Health Care Sector Suffer from Baumol's Cost Disease? Evidence from the 50 States." *Journal of Health Economics*, North-Holland, 31 Dec. 2012, www.sciencedirect.com/science/article/pii/S0167629612001877#:~:text=According%20to%20Hartwig%20%282008%29%2C%20Baumol%27s%20theory%20can%20be,%5B%20%CE%94%20log%20%28W%29%20%E2%88%92%20%CE%94%20log%20%28Y%29%5D.

Baumol, William. *Health Care, Education and the Cost Disease: A Looming Crisis for Public Choice*, 1993, www.jstor.org/stable/pdf/30027203.pdf?refreqid=fastly-default%3A3731006c719fa8ba59dbc310be0db0fc&ab_segments=0%2Fbasic_phrase_search%2Fcontrol&origin=&initiator=&acceptTC=1.

Baumol, William. *Macroeconomics of Unbalanced Growth*, 1967, piketty.pse.ens.fr/files/Baumol1967.pdf.

Baumol, William. *On the Performing Arts: The Anatomy of Their Economic ...*, Mar. 1965, pages.stern.nyu.edu/~wbaumol/OnThePerformingArtsTheAnatomyOfTheirEcoProbs.pdf.

Butler-Bowdon, Tom. *50 Classics Series: 50 Economics Classics*. Nicholas Brealey Publishing, 2017.

Cowen, Tyler. "Why I Do Not Believe in the Cost-Disease - Journal of Cultural Economics." *SpringerLink*, Springer-Verlag, 1996, link.springer.com/article/10.1007/s10824-005-7214-1.

"Federal Reserve Economic Data." *FRED*, Federal Reserve Bank of St. Louis, fred.stlouisfed.org/. Accessed 23 Apr. 2024.

Hartwig, Jochen, and Hagen Krämer. "Revisiting Baumol's Disease: Structural Change, Productivity Slowdown and Income Inequality." *Intereconomics*, 1 Jan. 1970, www.intereconomics.eu/contents/year/2023/number/6/article/revisiting-baumol-s-disease-structural-change-productivity-slowdown-and-income-inequality.html.

"Industries at a Glance: Utilities: NAICS 22." *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, Apr. 2024, www.bls.gov/iag/tgs/iag22.htm.

Kruger, Alan. *An Interview with William J. Baumol.*, 2001, graphics8.nytimes.com/packages/pdf/health/18baumol-doc.pdf.

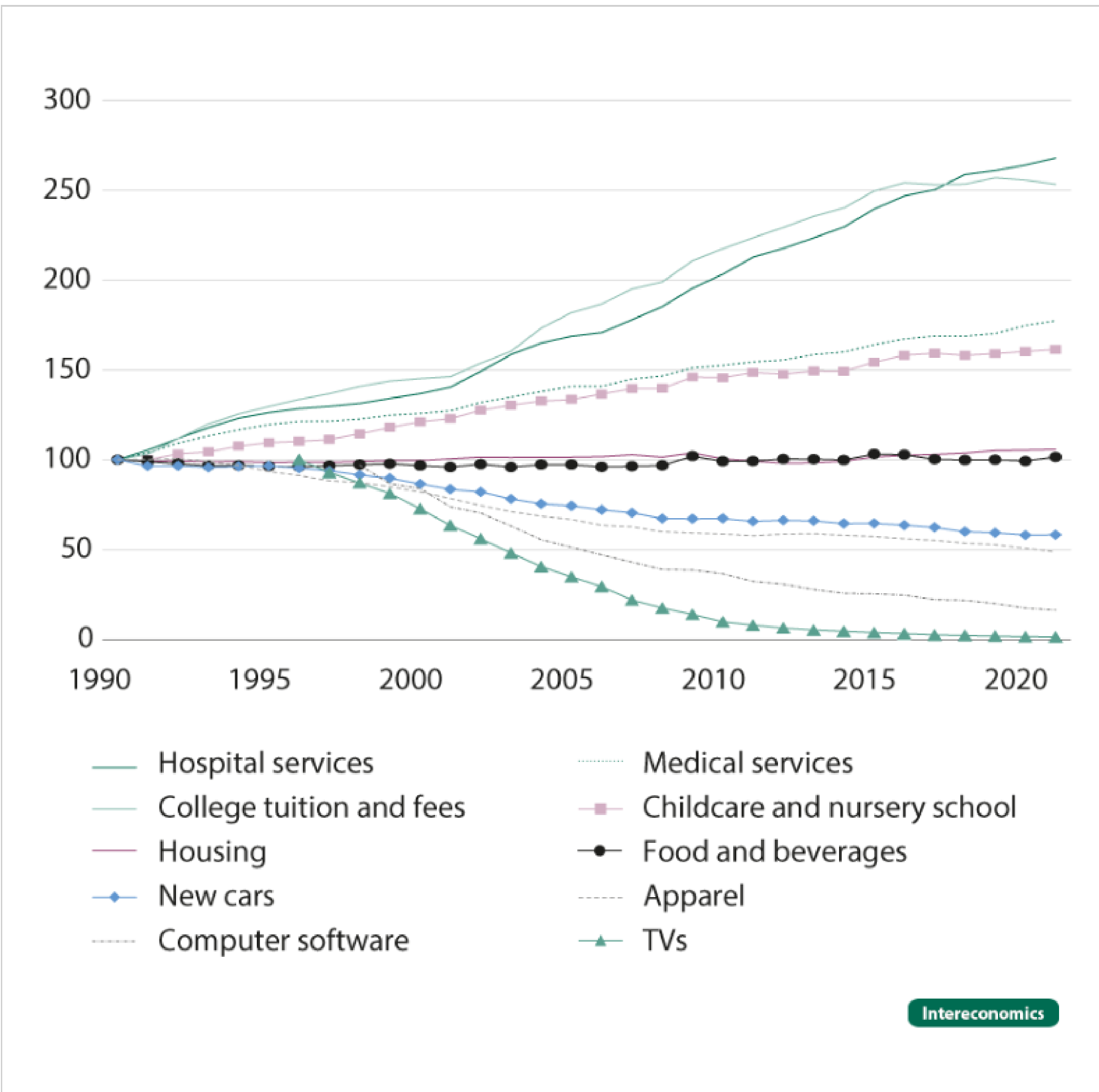
"NAICS & Sic Identification Tools." *NAICS Association*, 2023, www.naics.com/search/.

(8) Figures

Figure 1:

Price changes of selected goods and services in the USA relative to the consumer price index (1990-2020)

Index, 1990 = 100



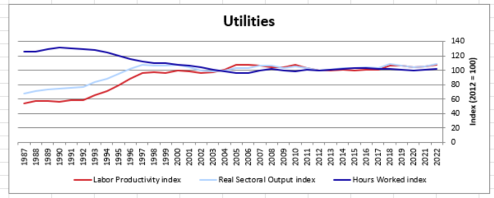
Hartwig, J. and H. M. Krämer (2022)

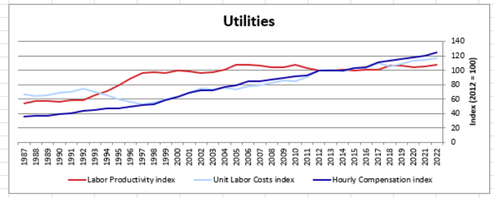
Figure 2:

22 - Utilities

Year	Labor Productivity index	Output per Worker index	Real Sectoral Output index	Hours Worked index	Employment index	Hourly Compensation index	Unit Labor Costs index
2022	107.32	109.152	109.31	101.854	100.145	124.887	116.369
2021	105.282	107.14	105.629	100.329	98.589	120.039	114.017
2020	104.1	105.909	104.012	99.915	98.209	118.355	113.693
2019	106.292	107.679	107.017	100.682	99.385	115.667	108.82
2018	106.692	108.131	108.248	101.459	100.109	112.949	105.865
2017	100.708	102.195	102.675	101.953	100.47	110.507	109.73
2016	100.73	103.041	103.712	102.961	100.651	104.279	103.523
2015	100.239	102.834	103.447	103.201	100.597	103.546	103.299
2014	100.522	103.187	102.945	102.41	99.765	99.166	98.651
2013	99.841	101.194	101.029	101.191	99.837	100.094	100.254
2012	100	100	100	100	100	100	100
2011	102.786	103.152	103.115	100.32	99.964	93.259	90.732
2010	107.263	105.612	105.612	98.461	100	91.62	85.416
2009	103.684	101.642	102.966	99.308	101.302	89.368	86.193
2008	104.568	104.888	106.045	101.412	101.103	86.875	83.08
2007	106.649	106.162	106.278	99.652	100.109	84.475	79.208
2006	107.79	104.278	103.467	95.989	99.222	84.661	78.542
2005	107.393	102.988	103.211	96.106	100.217	78.737	73.317
2004	101.102	97.249	99.184	98.103	101.99	76.943	76.104
2003	97.581	94.085	98.204	100.639	104.378	72.003	73.789
2002	96.646	92.893	100.185	103.662	107.851	71.803	74.295
2001	98.777	96.542	104.68	105.976	108.43	68.563	69.412
2000	99.562	98.867	107.541	108.014	108.774	62.828	63.104
1999	96.62	96.371	106.081	109.792	110.076	58.14	60.173
1998	96.859	96.411	106.98	110.45	110.962	53.164	54.888
1997	95.834	95.592	107.368	112.036	112.319	51.281	53.51
1996	88.14	88.322	102.19	115.94	115.702	49.403	56.05
1995	78.982	79.115	95.345	120.717	120.514	47.417	60.035
1994	71.266	71.224	88.811	124.619	124.692	46.873	65.773
1993	65.046	64.913	83.455	128.301	128.564	45.336	69.699
1992	59.196	58.325	76.599	129.399	131.331	43.96	74.262
1991	58.234	57.179	76.139	130.746	133.158	40.94	70.303
1990	56.446	55.534	74.34	131.701	133.864	38.971	69.041
1989	57.447	55.831	73.95	128.727	132.453	37.397	65.097
1988	56.993	54.889	71.769	125.927	130.753	36.875	64.701
1987	54.18	52.038	68.229	125.931	131.114	36.297	66.994

Return to Table of Contents





(Data taken from the Bureau of Labor Statistics)

(9) Appendix: Variables Tested and Sources

When starting to test my theories and run regressions, I added a multitude of time series variables. Below is a chart of the variables and the source I used in extracting the data.

Name in EViews	Data	Source
Hourly_Compensation_?	An index for the hourly wage of an employee, with the year 2012=100	BLS
Productivity_?	An index for labor productivity index of an employee, with the year 2012=100	BLS
Hours_?	An index for how many hours an employee works a year, adjusted for seasonality, with the year 2012=100	BLS
Federal_funds_rate	The interest rate at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other overnight	FRED
GDP	United States Gross Domestic Product, as defined by the Bureau of Economic Analysis	FRED
Minimum_Wage	The US federal minimum wage	FRED
Sticky_CPI	The increase in consumer price index, not including utilities and food	FRED
Unemployment_Rate	The US unemployment rate	FRED