Using Taylor's Law to Estimate Variance in Annual Unemployment by State

David A. Swanson[1] and Jeff Tayman[2]

[1]*Department of Sociology, University of California Riverside, 900 University Ave, Riverside, CA 92521, U.S.A.*

[2]*Department of Economics, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, U.S.A.*

Abstract: Taylor's law (TL) is a widely observed empirical pattern that relates variance to the the mean of a set of non-negative measurements via an approximate power function: $\text{variance}_g \approx a \times (\text{mean}_g)^b$, where g indexes the group of measurements. While widely observed, we have not found an application of TL to annualized state unemployment data. Thus, in this paper, we construct a model using TL to estimate of the variance in the 2018 annualized number of employed by state using the mean number. Our "in-sample" set consists of 38 states with 25 or more counties. We then test the model by estimating the variance in the 2018 annualized number of employed by state using the mean number. Our "out-of-sample" test set consists of the 12 states with fewer than 25 counties. Variance in the numbers of annualized unemployed by county within each state is important because it is a summary measure of how disproportionate unemployment is spread across counties. This suggests that policy-based efforts to reduce unemployment inequality among counties in a given state might better serve a state with a high mean level of county unemployment than a state with a low mean level of county unemployment because Taylor's Law shows that there is a higher level of unemployment inequality in the former than in the latter.

JEL Classification: B41, C13, C18.

1. BACKGROUND

Variance in the numbers of annualized unemployed by county within each state is important because it is a summary measure of how disproportionately unemployment is spread across counties. The higher the variance, the more disproportionate the distribution of unemployment, which is an overlooked indicator of an important policy issue, namely that areas of high unemployment, like areas of low unemployment, tend to be spatially clustered, suggesting that unemployment is persistent across space and time regimes (Cracolici, Cuffaro, and Nijkamp, 2007). In a related vein, Swanson, Tayman, and Byran (2018) found, for example, that 2017 county unemployment numbers were not only far more right-skewed in Arizona than in New Mexico, but that both the mean and variance were higher in Arizona (6.29, and 3.84, respectively) than New Mexico (6.24 and 1.35, respectively), which suggests that unemployment is more disproportionately spread across counties in Arizona than in New Mexico.

One possible indicator of the disproportionate distribution of unemployment is Taylor's law (TL), which is a widely observed empirical pattern (Cohen 2916, Cohen 2017, Cohen and Courgeau 2017, Cohen Bohk-Ewald and Rau 2018, Demers 2018, Swanson and Tedrow, 2022, Taylor 1961, Tokeshi 1995) that relates the variances to the means of sets of non-negative measurements via an approximate power function: $\text{variance}_g \approx a \times (\text{mean}_g)^b$, where g indexes the group of measurements (Reuman et al. 2017: 6788). While widely, observed, we have not found an application of TL to annualized state unemployment data.

In testing to see if TL can be used to estimate variance in unemployment, we use the state average of county unemployment data to construct a logarithmic regression model that estimates the variance in unemployment from mean unemployment by state in accordance with TL.

2. DATA AND METHOD

In this study, we use a (non-random) "in-sample" of 38 states, where each state has 25 or more counties. The means represent the arithmetic average number of annualized numbers of unemployed by county within each state for the year 2018 (Bureau of Labor Statistics, 2020a). Table **1** provides these 38 measurements, showing (1) μ, the mean n of annualized unemployed by state) and (2) σ^2, variance in the number of annualized unemployed persons by state.

Natural logarithms were calculated for each measurement in Table **1** and placed in a bivariate regression framework:

$\ln(\sigma^2) = \ln(a) + b(*\ln(\mu)) + \varepsilon[1]$

where

$\sigma 2$ = Variance in state unemployment as calculated across counties within each state

μ= Mean state unemployment as calculated across counties within each state

*Address correspondence to this author at the Center for Studies in Demography & Ecology, University of Washington, Seattle, WA 98195-3412 U.S.A; E-mail: dswanson@ucr.edu

Table 1. In-Sample Data used in Model Construction.

STATE	UNEMPLOYED MEAN	UNEMPLOYED VARIANCE
AL	1,280.3	3,635,067.9
AK	784.9	2,527,819.5
AR	657.5	914,117.1
CA	14,139.6	1,034,757,139.6
CO	1,518.2	9,499,098.9
FL	5,446.0	76,348,079.6
GA	1,260.7	8,138,802.2
ID	566.4	1,132,266.3
IL	2,736.8	119,365,974.5
IN	1,281.1	5,275,890.2
IA	447.0	707,266.5
KS	463.1	1,812,816.3
KY	743.2	2,511,247.6
LA	1,601.5	4,598,748.6
MD	5,293.1	39,191,049.3
MI	2,450.1	30,071,856.9
MN	1,037.8	4,855,656.7
MS	749.5	744,581.3
MO	838.6	4,208,521.8
MT	344.1	332,046.3
NE	321.9	1,222,355.9
NM	1,391.3	6,819,093.3
NY	6,357.6	117,014,189.2
NC	1,984.2	10,900,677.4
ND	195.9	144,082.4
OH	2,943.9	23,011,786.7
OK	818.8	3,605,498.1
OR	2,360.3	12,172,938.7
PA	4,073.6	37,042,046.1
SC	1,758.3	3,572,965.0
SD	213.5	178,014.2
TN	1,204.0	5,427,753.5
TX	2,092.8	65,080,000.7
UT	1,644.1	13,033,666.4
VA	988.8	3,268,765.0
WA	4,344.5	56,045,370.4
WV	743.1	509,160.9
WI	1,259.9	4,850,616.2

and ε = error.

3. RESULTS

The set-up just described yielded the following model:

$$\ln(\sigma^2) = \ln(1.7907) + (1.9251*\ln(\mu)) \quad [2]$$

$r^2 = 0.897$

As can be seen from the coefficient of determination ($r^2 = 0.897$) the regression model fits the data well with the estimated parameters of a = ln(1.7907) and b = 1.9251.

Fig. (**1**) provides a graphic view of this relationship defined by equation [2]

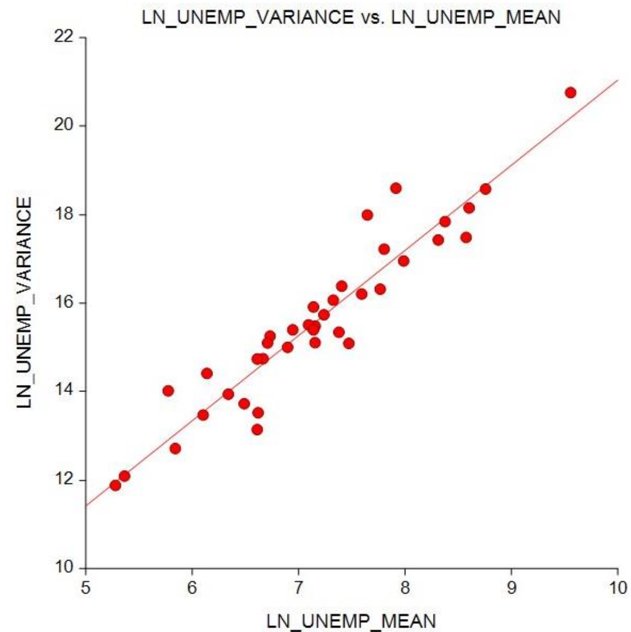

Fig. (1). The relationship between Ln (unemployment) and Ln (Variance).

4. TEST

We test the model by estimating the variance in the 2018 annualized number of employed by state using the mean number (BLS, 2020a. Our "out-of-sample" test set consists of the 12 states with fewer than 25 counties. As shown in Table **2**, the mean algebraic percent error (MALPE) is and the Mean Absolute Percent Error (MAPE) found by using the model are both very low at 2.00% and 4.96%, respectively.

Table 2. Results of the Out-of-Sample Test of Model Accuracy.

UNEMPLOYMENT STATE	MEAN	LN(MEAN)	VARIANCE	LN(VARIANCE)	ESTIMATED LN(VARIANCE) USING TAYLOR'S MODEL	DIFFERENCE EST - ACTUAL	PCT DIFFERENCE	ABS PCT DIFFERENCE
AZ	10,804	9.29	483,371,391	20.00	18.67	-1.33	-6.63%	6.63%
CT	9,847	9.19	62,412,016	17.95	18.49	0.54	3.02%	3.02%
DE	6,077	8.71	12,074,304	16.31	17.56	1.26	7.70%	7.70%
HI	4232.3	8.35	15956755.7	16.59	16.87	0.28	1.69%	1.69%
ME	1,408	7.25	1,223,380	14.02	14.75	0.73	5.21%	5.21%
MA	9,086	9.11	49,376,821	17.71	18.34	0.62	3.51%	3.51%
NV	3,929	8.28	144,915,198	18.79	16.72	-2.07	-11.01%	11.01%
NH	1,957	7.58	3,887,639	15.17	15.38	0.21	1.37%	1.37%
NJ	8,346	9.03	25,020,227	17.04	18.17	1.14	6.68%	6.68%
RI	4,415	8.39	23,495,021	16.97	16.95	-0.02	-0.14%	0.14%
VT	629	6.44	174,511	12.07	13.20	1.13	9.33%	9.33%
WY	494	6.20	226,451	12.33	12.73	0.40	3.24%	3.24%
						MALPE	**2.00%**	
						MAPE		**4.96%**

5. SUMMARY AND CONCLUSION

We find that TL fits the 2018 annualized number of unemployed for the "in-sample" set of states and that the subsequent test quite well. Variance in the numbers of annualized unemployed by county within each state is important because it is a summary measure of how disproportionate unemployment is spread across counties. The higher the variance, the more disproportionate the distribution of unemployment, a policy issue. TL shows that as mean annualized unemployment increases in a given state, so does the variance, which suggests that the burden of unemployment becomes more disproportionately shared across counties in state with a high mean level of unemployment than in a state with a low mean level of unemployment.

This a finding that should be of interest to policy-makers because, as Cracolici, Cuffaro, and Nijkamp (2007) found, it may be the case that a state with a high level of mean county unemployment experiences spatially clustered unemployment that is more persistent across space and time than does a state with a low level of mean county unemployment. It further suggests that if place-based economic development (Partridge and Rickman, 2007) has a potential role for reducing unemployment inequality across counties in a given state then such development efforts might better serve a state with a high mean level of county unemployment than a state with a low mean level of county unemployment because TL shows that there is a higher level of unemployment inequality in the former than in the latter.

REFERENCES

Bureau of Labor Statistic (2020a). Labor force data by county, 2018 annual averages (TXT, XLS), (https://www.bls.gov/lau/#tables)

Bureau of Labor Statistic (2020b). Labor force data by county, 2019 annual averages (TXT, XLS), (https://www.bls.gov/lau/#tables)

Cohen, J. (2016) Statistics of Primes (and Probably Twin Primes) Satisfy Taylor's Law from Ecology, *The American Statistician*, 70:4, 399-404.

Cohen, M. (2017) Non-Asymptotic Mean and Variance Also Approximately Satisfy Taylor's Law, *The American Statistician*, 71:2, 187-187.

Cohen, J., C. Bohk-Ewald, and R. Rau (2018). Gompertz, Makeham, and Siler models explain Taylor's law in human mortality data, *Demographic Research*, 38:29, 773-842

Cohen, J. and D. Courgeau (2017). Modeling distances between humans using Taylor's law and geometric probability, Mathematical Population Studies, 24:4, 197-218.

Cracolici, M., M. Cuffaro, and P. Nijkamp. (2007). Geographical distribution of unemployment: An analysis of provincial differences in Italy, *Growth and Change* 38 (4): 649-670.

Demers, S. (2018). Taylor's Law Holds for Finite OEIS Integer Sequences and Binomial Coefficients, *The American Statistician*, 72:4, 376-378

Partridge, M. and D. Rickman. (2007). Persistent pockets of extreme American poverty and job growth: Is there a place-based policy role? *Journal of Agricultural and Resource Economics*. 32: 201-224.

Reuman, D., L. Zhao, L. Sheppard., P. Reid, and J. Cohen. (2017). Synchrony affects Taylor's Law in Theory and Data, *Proceedings of the National Academy of Sciences*, 114:26, 6788-6793.

Swanson, D., and L. Tedrow. (2022). Taylor's Law and the Relationship between Life Expectancy at Birth and Variance in Age at Death in a Period Life Table, *Population Review*, 61 (1): 31-42

Swanson, D., J. Tayman, and T. Bryan. (2018). A Note on rescaling the arithmetic mean for right-skewed positive distributions, *Review of Economics and Finance* 14 (4):17-24 DOI Article ID: 1923-7529-2018-04-17-08

Taylor, L. R. (1961). Aggregation, Variance and the Mean, *Nature*, 189, 732–735.

Tokeshi, M. (1995), On the Mathematical Basis of the Variance- Mean Power Relationship, *Researches on Population Ecology*, 37, 43–48.

Received: Jun 22, 2022 Revised: Jun 15, 2022 Accepted: Sep 12, 2022