**Does Job satisfaction rise and fall with economy?**

**Course:** SOC-570A

**Semester:** Fall 2024

**Version:** Version-4 (Final)

**Author:** Aditya Bandimatt

**Date:** 2024-12-15

1. **Please write the full citation for the article you replicated and attach a PDF of it.**

This is a replication of the paper published by Emily C. Binachi and others in ‘Academy of Management Journal’ in 2023. The paper explores the hypothesis that during bad economic times, people are likely to see their own jobs favorably. Findings from this paper suggests that job satisfaction is a function of not only intra organizational features, but is also influenced by outside conditions.

Emily C. Bianchi, Chris C. Martin and Ren Li . Published Online:19 Apr 2023. Academy of Management JournalVol. 66, No. 2. <https://doi.org/10.5465/amj.2019.0405>

1. **Attach your code and all the data.**

Following documents are submitted:

1. AB-SOC570\_Final\_Report.docx (This file)
2. AB\_SOC570-JSat.rmd
3. AB-SOC570-JSat.html
4. gss\_jsat\_7416.csv
5. inv\_uempdata.csv
6. dls\_uempdata.csv
7. article\_AMJ.pdf (original paper)
8. **Report your version of the descriptive statistics table next to the one in the original article.**

**Descriptive Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Original Article** | | **Replication** | |
| **Variables** | **Mean** | **SD** | **Mean** | **SD** |
| National unemployment rate | 6.28 | 1.51 | 6.31 | 1.51 |
| State unemployment rate | 6.20 | 2.00 | - | - |
| Job satisfaction | 3.34 | 0.77 | 3.34 | 0.77 |
| Year | 1994.65 | 11.82 | 1994.33 | 11.96 |
| Age | 40.50 | 12.05 | 40.49 | 12.02 |
| Male | 0.55 | 0.50 | 0.56 | 0.50 |
| Income | $26,067 | $29,912 | $20,852 | $6468 |
| Race White | - | - | 0.81 | 0.39 |
| Race Black | - | - | 0.13 | 0.34 |
| Education Level | - | - | 2.64 | 1.21 |
| Total observations (n) | 23,335 | | 23,707 | |

* The original article uses ‘State unemployment rate’, ‘Industry segments’, ‘unemployment rates from UK’ for study-2 and study-3, these were out of scope for this replication.
* As depicted in the table, ‘Total Observations’, as well as the ‘Mean’ and ‘Standard Deviation’ for the variables match with replication except for ‘Income’.
* I had challenges finalizing the ‘Income’ variable from the GSS dataset. This variable is split across multiple questions that have changed over the period of study (1974 to 2016). There are also separate data collection on ‘Family Income’ and ‘Individual Income’. I have taken the variable that most closely matched the Mean and Standard Deviation from the original paper. It was not clear how authors aggregated the ‘Income’ variable. Their dataset has a lot wider dispersion of ‘Income’ values as can be seen form large standard deviation.
* Another challenge I had was with the variable ‘National unemployment rate’. I sourced it from two separate sources and compared them before finalizing the one from ‘US Bureau of Labor Statistics’ as it was closest to the original article.
* I also ran analysis for additional variables of interest that were not used by authors. These variables were used to explore their influence on the question at hand about job satisfaction and state of the economy. The two additional variables explored were: ‘Race’ and ‘Education Level’

1. **Report your version of the regression tables(s) you replicated.**

**Correlations**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Original** | **Unemp** | **Jobsat** | **Year** | **Age** | **Male** | **Income** |
| **Unemp** |  |  |  |  |  |  |
| **Jobsat** | 0.01\* |  |  |  |  |  |
| **Year** | -0.34\*\*\* | 0.02\*\* |  |  |  |  |
| **Age** | -0.02\*\* | 0.12\*\*\* | 0.13\*\*\* |  |  |  |
| **Male** | 0.01 | 0.00 | -0.05\*\*\* | 0.02\* |  |  |
| **Income** | -0.03\*\*\* | 0.13\*\*\* | 0.03\*\*\* | 0.24\*\*\* | 0.26\*\*\* |  |

\* p < .05

\*\* p < .01

\*\*\* p < .001, two-tailed

**Correlations**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Replication** | **Unemp** | **Jobsat** | **Year** | **Age** | **Male** | **Income** |
| **Unemp** |  |  |  |  |  |  |
| **Jobsat** | 0.02\* |  |  |  |  |  |
| **Year** | -0.33\*\*\* | 0.02\*\* |  |  |  |  |
| **Age** | -0.02\*\* | 0.12\*\*\* | 0.13\*\*\* |  |  |  |
| **Male** | 0.02\* | 0.00 | -0.05\*\*\* | 0.02\* |  |  |
| **Income** | -0.20\*\*\* | 0.09\*\*\* | 0.42\*\*\* | 0.16\*\*\* | 0.08\*\*\* |  |

\* p < .05

\*\* p < .01

\*\*\* p < .001, two-tailed

‘Pearson correlation coefficient’, ‘direction of the correlation’ (positive or negative) as well as ‘significance level’ match perfectly between the original article and replicated values. The average absolute magnitude of the coefficients vary by approx. 38% for ‘income’ variable. This can be explained as Mean and Std. Deviation for ‘income’ variable was smaller compared to the original.

**OLS Regressions Predicting Job Satisfaction**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Model-1** | | **Model-2** | | **Notes** |
| **Variables** | Original | Replicated | Original | Replicated |
| Unemp | 0.011\*  (0.005) | 0.013\*  (0.005) | 0.011\*  (0.004) | 0.012\*  (0.005) |  |
| Year | 0.002\*  (0.001) | 0.002\*  (0.001) | 0.001  (0.001) | -0.000  (0.001) | *Change in direction* |
| Age |  |  | -0.008\*\*\*  (0.002) | 0.002  (0.002) | *Change in direction, coefficient & significance* |
| Age^2 |  |  | 0.000\*\*\*  (0.000) | 0.000\*  (0.000) | *Reduced significance* |
| Male |  |  | -0.048\*  (0.015) | -0.006  (0.013) | *Reduced coefficient value* |
| Pay (log) |  |  | 0.106\*\*\*  (0.007) | 0.065\*\*\*  (0.010) |  |
| Observations | 23,335 | 23,707 | 23,335 | 23,707 |  |
| R^2 | 0.015 | 0.0009 | 0.027 | 0.017 | *Reduced explainability for model-1* |

\* p < .05

\*\* p < .01

\*\*\* p < .001, two-tailed

Values in brackets indicate *‘clustered standard errors’. Clustering is done on within the year.*

All reported standard errors are clustered standard errors (clustered on variable ‘year’) allowing for correlation between observations within the same year. As can be seen from the detailed summary outputs (AB-SOC570-JSat.html), these are on average higher than the un-clustered standard errors. But they provide a more robust estimation of uncertainty. The *t-statistics*, *p-value* and *significance* reported are also calculated based on the *clustered standard error*. I have followed this methodology throughout this analysis to align with the approach authors have taken in the original paper.

**Model-1**

The estimated effect for **Unemployment Rate** is positive and significant on **Job Satisfaction** (b = 0.011; SE = 0.005; p < 0.05). A one-unit increase in Unemployment Rate is associated with an average 0.011 unit (on a 4 point scale) increase in self-reported job satisfaction, assuming all other variables remain constant. The replicated values match this well except for the coefficient value (replicated value is 0.002 units higher).

The estimated effect for **Year** is positive and significant on **Job Satisfaction** (b = 0.002; SE = 0.001; p < 0.05). Each passing year is associated with an average 0.002 unit (on a 4 point scale) increase in self-reported job satisfaction, assuming all other variables remain constant. The replicated values match this perfectly in all respects – coefficient, clustered standard error, direction, and significance.

The **R-squared** value is lower in the replicated model-1 (0.015 vs 0.009) which means that the *variability of dependent variable (Job Satisfaction) that can be explained by the independent variables – Unemployment Rate and Year* is lower in the replicated model.

**Model-2**

The estimated effect for **Unemployment Rate** is positive and significant on **Job Satisfaction** (b = 0.011; SE = 0.004; p < 0.05). A one-unit increase in Unemployment Rate is associated with a 0.011-unit (on a 4-point scale) increase in self-reported job satisfaction on average, assuming all other variables remain constant. The replicated values match this well (replicated value is 0.001 units higher for the coefficient and clustered standard error).

The estimated effect of **Year** is positive but in but not significant on **Job Satisfaction**. The magnitude of the coefficient has also decreased from model-1 (b = 0.001, SE = 0.001, p > 0.05). This means that with every passing year, self-reported the job satisfaction on a 4-point scale will only increase by 0.001 units on average, assuming all other variables remain constant. The replicated values match in terms of significance and (clustered) standard error. But the magnitude of the coefficient has reduced to almost 0 and the direction is also indicating negative trend (which means in the replicate model, job satisfaction will decrease by a much reduced fraction with passing year, when all other variables are constant).

The estimated effect of **Age** is negative and highly significant on **Job Satisfaction** (b = -0.008, SE = 0.002, p < 0.001). This means that with every one-unit increase in age, the self-reported job satisfaction (on a 4-point scale) will decrease by 0.008 on average, assuming all other variables remain constant including the unemployment rate. The replicated value however, has changes in terms of direction, magnitude of the coefficient and the significance level. In the replication, age is not coming up as significant.

The estimated effect of **Age-squared** is positive and highly significant even though the magnitude of the coefficient is very small (b = 0.000, SE = 0.000, p < 0.001). Understanding this along with the age variable, we can conclude that age has non-linear relation with job satisfaction. As people get older, the trend reverses and self-reported job satisfaction starts increasing with every unit increase in age, on average (assuming all other variables remain constant). This is statistically significant and meaningful feature of the data. Replicated model is not able to capture the reversal of this trend between Age and Age-squared or its significance.

The estimated effect of **Gender(Male)** is negative and statistically significant on **Job Satisfaction** (b = -0.048, SE = 0.015, p < 0.05). This means that controlling for all other variables including unemployment rate, compared to other genders (Female and Others), Males self-report job satisfaction at 0.048 units lower (on a 4-point job satisfaction scale) on average. The replicated model is capturing this trend but the coefficient is smaller ( -0.006 compared to -0.048).

The estimated effect of **Pay(log)** is positive and statistically highly significant on **Job Satisfaction** (b = 0.106, SE = 0.007, p < 0.001). This means that with every proportional unit increase in the pay(log), self-reported job satisfaction raises by 0.106 proportional units on average, assuming all other variables remain constant. The replicated model is capturing this relationship in terms of the magnitude, direction as well as the significance level.

**The R-square** values are comparable between original and replicated models (0.03 v/s 0.02). This means that replicated model explain 2% of the variability in job satisfaction with the selected independent variables (compared to 3% in the original model)

In summary, the replicated model is capturing the magnitude of coefficients, overall direction of the relationships and significance of the original model ( except for the non-linear relationship with the Age variable which could be due to differences in the standard deviation, size of observations and possible multicollinearity in the replicated data ).

1. **Run a full set of diagnostics for your version of the regression you replicated.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Model-1** | **Model-2** | **Model-3** |
| **Heteroskedasticity** | BP = 14.91  DF = 2  p-value < 0.001 | BP = 107.1  DF = 6  p-value < 0.001 | BP = 124.64  DF = 9  p-value < 0.001 |
| **Multicollinearity** | None | Age and Age^2 | Age and Age^2 |
| **Normality** | A = 2014.2  p-value < 0.001 | A = 1271.9  p-value < 0.001 | A = 1108.6  p-value < 0.001 |
| **RESET** | RESET = 6.211  p-value < 0.05 | RESET = 19.752  p-value < 0.001 | RESET = 4.24  p-value < 0.05 |

**Multicollinearity:**

Model-1 does not have any multicollinearity. Model-2 and Model-3 have multicollinearity between Age and Age^2. This is obvious as the both variables are the same with a multiplication factor of 2.

**Heteroskedasticity:**

* Null Hypothesis (H₀): The residuals have constant variance (homoskedasticity).
* Alternative Hypothesis (H₁): The residuals have non-constant variance (heteroskedasticity).

The p-values are small for all the models and also evidence for variance in the residuals not being constant across all levels of independent variables is high. We have to reject the null hypothesis in all three cases. Heteroskedasticity is noticed in all the models.

**Normality:**

* **Null Hypothesis (H₀)**: The data (residuals) follow a normal distribution.
* **Alternative Hypothesis (H₁)**: The data do not follow a normal distribution.

High test statistics and low p-value indicates that we need to reject null-hypothesis for all three models. This means that the model’s residuals are not normally distributed.

**RESET (Ramsey test):**

* Null Hypothesis (H₀): The model is correctly specified.
* Alternative Hypothesis (H₁): The model is not correctly specified.

High values of RESET and low p-values indicate that we will have to reject null hypothesis for all three models. This indicates that al the models are not correctly specified and potentially omitting non-linear relationships and have incorrect functional forms.

**5a. If you see any issues, try to fix them. Report the new “fixed” regression.**

Following two treatments were done to the dataset to address the issues:

1. Remove **Age** from the model.
2. Remove outliers – Remove IQR+1.5 and IQR-1.5 range data from the dataset on Age column.

**OLS Regressions Predicting Job Satisfaction**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Model-2** | | **Model-2A** | | **Notes** |
| **Variables** | Original | Replicated | Original | Replicated |
| Unemp | 0.011\*  (0.004) | 0.012\*  (0.005) | - | 0.012\*  (0.005) |  |
| Year | 0.001  (0.001) | -0.000  (0.001) | - | -0.000  (0.001) |  |
| Age | -0.008\*\*\*  (0.002) | 0.002  (0.002) | - | - |  |
| Age^2 | 0.000\*\*\*  (0.000) | 0.000\*  (0.000) | - | **0.000\*\*\***  **(0.000)** | *Significance has improved and now aligned with original model* |
| Male | -0.048\*  (0.015) | -0.006  (0.013) | - | -0.006  (0.013) |  |
| Pay (log) | 0.106\*\*\*  (0.007) | 0.065\*\*\*  (0.010) | - | 0.066\*\*\*  (0.009) |  |
| Observations | 23,335 | 23,707 | - | 23,664 |  |
| R^2 | 0.027 | 0.017 | - | 0.017 |  |

\* p < .05

\*\* p < .01

\*\*\* p < .001, two-tailed

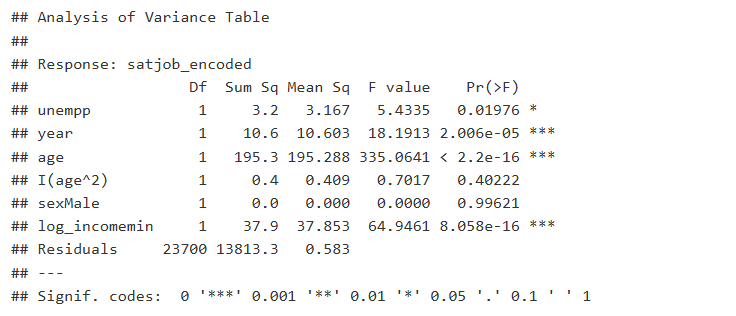
Values in brackets indicate *‘clustered standard errors’. Clustering is done on within the year.*

After the removal of outliers and multicollinearity for **Age** variable, the significance of **Age^2** variable has improved and now aligns with the original model. Rest of the parameters of the model including R^2 have remained the same.

***Multicollinearity* is removed in the Model-2A**. However, issues with *Heteroskedasticity* has also decreased marginally (BP reduced to 101 from 107). Issues with *Normality* and *Functional* *Specification* persists.

1. **Using the sum of squares of your version of the original model.**

Variance table for replicated model-2:



The two variables that are explaining most of the variability in job satisfaction are **Age** (sum of squares 195.3) and **Log of Income (or Pay)** withsum of squares 37.9. Both of them are also highly significant. Two variables that explain the variability least are **Gender (Male)** and **Age^2**.

Age came out as not significant in the replicated model. However, it is playing a very significant role in explaining the variability of ‘Job Satisfaction’. This needs to be explored more. This could be happening due to two factors:

1. Clustering: We clustered the model within the year variable. The variance contributed by Age may not be consistent across years, contributing to this effect.
2. Heteroskedasticity: could also be contributing to this effect. Since Anova calculates the explainability across the dataset, Age is coming out on top but not in the clustered summary.

Further fine-tuning and exploring the mode with additional variable ( Age \* Year) might help explain this better.

1. **Report the final regression in which you added a variable of interest to you.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Model-3** | | **Notes** |
| **Variables** | Original | Replicated |
| Unemp | - | 0.012\*  (0.005) |  |
| Year | - | -0.000  (0.001) |  |
| Age | - | 0.001  (0.002) |  |
| Age^2 | - | 0.000\*\*  (0.000) | Improved significance level  Closer to original model |
| Male | - | -0.012  (0.013) |  |
| Pay (log) | - | 0.046\*\*\*  (0.010) |  |
| Race Black | - | -0.147\*\*\*  (0.018) | Additional variable. Statistically highly significant |
| Education | - | 0.034\*\*\*  (0.005) | Additional variable. Statistically highly significant |
| Observations | - | 23,707 |  |
| R^2 | - | 0.025 | Improvement of 0.5% compared to model-2 |

\* p < .05

\*\* p < .01

\*\*\* p < .001, two-tailed

Values in brackets indicate *‘clustered standard errors’. Clustering is done on within the year.*

**Additional variables of interest:** For the third model, I chose two variables:

1. Race
2. Education Level

**Race** is acknowledged as one of the important factor that influences the behaviors and interactions in American society. I wanted to analyze how this factor interfaces with job satisfaction and state of the economy (measured in terms of national unemployment rate). In particular, I wanted to understand if given the same economic and other indicators, would race make a difference in how people felt about their jobs?

**Educational level** is one of the key factors in hiring decisions. I wanted to analyze how the education level impacts employee perceptions about job satisfaction when national unemployment rate changes.

As can be seen from the new model, both variables came out as statistically significant.

The estimated effect of **Race(Black)** is negative and statistically highly significant on **Job Satisfaction** (b = -0.147, SE = 0.018, p < 0.001). This means that when all other factors are controlled, including the national unemployment rate, compared to other races (White, Others), Blacks self-report job satisfaction at a 0.147 units (on a 4-point job satisfaction scale) lower on average.

The estimated value of **Education** is positive and statistically highly significant on **Job Satisfaction** (b = 0.034, SE = 0.005, p < 0.001). This means that with every one unit increase in the education level (5 scale education level used in the survey), self-reported job satisfaction increases by 0.034 units (4 scale job satisfaction level) on average, when all other variables remain constant.

Both the new variables and their results make sociological sense. With higher education levels, people will feel more confident of managing the uncertainties of the economy. Therefore, given a state of the economy, more educated people feel more job satisfaction.

In terms of race, Blacks report lower job satisfaction when other variables are controlled. This means that a black males with same income range and same age , same education level, facing same economic condition (measured in national unemployment rate), self-report lower job satisfaction compared to their colleagues of non-black races. This could be due to cultural, social and perceptions of systemic biases. This will need more study to explain the correlation.

The overall direction of all other variables has remained the same (compared to replicated model-2). Significance level has remained same as well expect for Age-squared. For Age-squared variable, significance level has increased and matches more closely to the original model. The overall explainability has also improved from 2% to 2.5%.

1. **Finally, how could your regression analysis described on item 7 above be even better?**

This regression model has relevant and actionable insights:

1. First, it proves that broader national economic indicators (unemployment rate) influence the job satisfaction of employees.
2. It identifies the highly significant role Pay (or compensation) plays in job satisfaction.

These insights will help organizations to plan their compensation components like bonus based on broader economic indicators. This will contribute to higher job satisfaction and hypothetically higher employee retention.

This model can be improved by building a regression with a larger sample data and more variables:

**Dataset:** A larger and statistically more robust dataset

**Variables:**

1. Systemic:
   1. National Unemployment Rate
   2. State Unemployment Rate
   3. Labor laws passed in the year
2. Organizational:
   1. Overall performance and reputation of the organization
   2. Organizational policies – Diversity, Leave and compensation
3. Individual:
   1. Gender
   2. Race
   3. Education Level
   4. Compensation
   5. Supervisor Relationship

This should improve the R^2 / explainability of Job satisfaction of our model and help organizations in improving the key performance indicators of their human resources departments (like employee job satisfaction and retention). It could even contribute in improving organizations’ overall performance in stock market and standing in society.