Matt Draugelis & Zach Dowhower

ECON 550

Winter 2025

Can Money Buy Happiness: An Econometric Analysis

I. Introduction

Are you happy right now? What if we asked you what would make you happier? Is it a higher salary? A brand-new car? Is it quality time with a friend or a significant other? We could go on, but, if there was one single unifying answer, we'd have cracked the code for a perfectly happy society. However, that's not the case, at least in the realm of reputable, longitudinal data in the United States. While the underlying answer hasn't been cracked by ChatGPT, as of the date this analysis was conducted, we are setting out in this paper to identify if money buys happiness, or in this study, life satisfaction, measured on a scale from one, not satisfied, to seven, completely satisfied. Specifically, with the addition of various independent control variables in this study, such as the poverty threshold, educational attainment, employment status, and family size, paired across the same pool of individuals with their own unique characteristics, we set out to establish if life satisfaction, either high or low scores, can be predicted through total family income.

This paper uses data from the U.S. Bureau of Labor Statistics' National Longitudinal Survey (NLS), which comprises just over 9,900 men and women born between 1957 and 1964. Our primary independent variable is total family income, used to identify the causal relationship. Our initial hypothesis going into this empirical study is that total family income is statistically significant in predicting one's life satisfaction; said differently, we believe higher incomes lead to higher life satisfaction scores in this sample. As mentioned, we use multiple control variables to isolate the effect on life satisfaction. These variables were selected after reviewing the relevant literature and evaluating the data availability and granularity in the NLS dataset. This empirical study and its accompanying model specifications are specifically created to measure the effect of income on life satisfaction, and as such, we believe this study will be able to identify causation and accurately predict life satisfaction as these controls are applied to the model throughout the years in which data is available.

II. Literature Review

To begin our analysis, we start with an examination into the existing body of research on finding causal effects of life satisfaction throughout various predictive variables. Beginning with earlier literature using data from a previous NLS dataset, NLSY72, Schieve (1992) found, through various regression techniques, evidence that, "education attainment has a statistically significant but trivial impact on satisfaction and is a weak predictor of life satisfaction" (Schieve, 1992). Further, after controlling other variables in his study, including demographics and work-related factors, life satisfaction is instead influenced by variables he did not control for in his study. Using this paper as a baseline provides a starting place for our subset of data from the NLSY79 dataset, which will only cover years 2014 to 2022, allowing us to build upon and identify a causal effect and predictor of life satisfaction. Throughout time, life satisfaction has continually been studied in economics, psychology, and politics to quantify the relationship. Moŵisch et al. (2020) note, in 2020, past literature shows a positive association between education and well-being (Moŵisch et al., 2020). However, the magnitude of the correlation to predict the effect of education on life satisfaction varies throughout studies and time (Moŵisch et al., 2020).

To attempt to illustrate the varying literature, we provide a non-extensive look at exploring the causal effect of education on predicting life satisfaction. Leonard (1981) found education to be correlated positively with life satisfaction, while Lee and Yang (2023) found the effect of education on happiness to be gender dependent – as educational attainment increases, women have a statistically significant higher level of happiness, but for men there was no statistically significant association found between the two variables (Lee & Yang, 2023; Leonard, 1981). To finish the discussion of education's effect on life satisfaction, we think it's important to mention Hartog and Oosterbeek's 1998 paper, "Health, wealth, and happiness: why pursue a higher education," the first study to use the University of Amsterdam's longitudinal survey, similar to our NLS dataset; Hartog and Oosterbeek found the highest levels of education are not the happiest, instead the effect of schooling on happiness is parabolic (Hartog and Oosterbeek, 1998). This finding, a parabolic effect of education on happiness, is a key assumption we expect to find in our own study; we believe at some point, having more education does not statistically associate with higher levels of happiness.

Turning now to the literature on employment and job industry, Bertermann et al. (2023) look at how employment status correlates with life satisfaction, to which they find a positive effect for being employed and high educational attainment (Bertermann et al., 2023). In contrast, in the same study, those with high educational attainment and unemployment report statistically significant lower life satisfaction scores (Bertermann et al., 2023). The relevance for our paper is seeing if these results hold for retired individuals, which do make up a subsection of the individuals in our sample data. Further research shows the result regarding unemployed individuals holds in other studies. L Winkelmann and R Winkelmann (1998) suggest men's unemployment imposes larger burden on individuals in a "non-pecuniary" way, or not able to be measured monetarily, finding a significant effect on social relationships and status (Winkelmann & Winkelmann, 1998). In contrast, our sample will be both men and women and allow us to see if we see similar relationships between employed and unemployed individuals and their life satisfaction scores for both genders.

Schmitt and Mellon (1980) note the compensatory model, suggesting individuals with "boring" jobs seek out "interesting nonwork activities," implying a negative relationship between employment and life satisfaction for those workers (Schmitt and Mellon, 1980). However, with a spillover model, the authors find a positive relationship between a job and, as a result, their industry, and life satisfaction, finding satisfaction in one domain of an individual's life should spill over into other areas (Schmitt and Mellon, 1980). Further, under the traditional assumption, men being the "breadwinner" will show up in more job industries tied with physical labor, which the research by Schmitt and Mellon (1980) says produces a lower health and ultimately lower life satisfaction overall. Conversely, individuals employed in industries they are passionate about yield higher life satisfaction scores when controlling for the job industry (Schmitt and Mellon, 1980). While our study does not look to aid in this literature, we feel it's an important study to consider. Lastly in employment literature, Grün et al. (2010) find, "job quality only matters to some extent, and often people in bad jobs are still better off than those who remain unemployed" (Grün et al., 2010). Further, the effect is, "statistically significant for most indicators of job quality, except for workers with low job satisfaction and for those whose new job is much worse than their pre-unemployment job" (Grün et al., 2010).

As we transition to the literature on poverty, our analysis will allow us to track the same individual over the period and see by controlling directly for poverty, if a rise out of poverty statistically changes life satisfaction scores. A study by Samman and Santos (2013) in Chile found poor people were more dissatisfied than non-poor people, even if they had economic gains while they remained under the poverty line, while looking at the effect on overall life satisfaction (Samman and Santos, 2013). In contrast, "the people who fell into poverty were also no more satisfied than the people who had already been in poverty" (Samman and Santos, 2013). Additionally, Samman and Santos' (2013) data replicates, almost identically, the same scales as our life satisfaction variable from the NLS dataset, ranging from scores of one to four compared.

Lastly, as we tie literature up, we've yet to discuss the relevant literature on gender, age, and race. Throughout all the previous literature reviewed so far, demographic characteristics were either covered directly (specifically with age and gender), or already present in the studies, like the makeup of race in the individuals studied and their resulting data. As previously mentioned, our sample is primarily a subsample of the same age range, thus we do not control for age as there is no variation in the data. While gender is seen in the literature, both at the explicit forefront and at the underlying mechanism of other studies, we have made the decision to not control explicitly for gender. Within our sample, our data is already split at 50/50 for gender makeup; thus, in combination with our causal area of focus for this study, we do not include it. Lastly, apart from the literature already reviewed, we have not found significant studies individually on race as it affects life satisfaction. As a result, due to both the lack of literature on the topic and fine-tuned data granularity in the NLS dataset, we move forward by not controlling for race in our study.

III. Data

All data used within this project comes from the NLS website, using the NLSY79 cohort of respondents, a dataset tracking individuals throughout their lives. The respondents are all born between 1957 and 1964 and, at the time of the survey, currently reside within the United States, totaling just over 12,000 people. Initially, there was an oversampling of Hispanics, economically disadvantaged non-Hispanics & non-Blacks, as well as military youth; this issue has since been corrected as of 1990, leaving roughly 9,900 respondents. As a result, by the time life satisfaction began being recorded in 2014, this is no longer a concern. Lastly, due to the fact of life satisfaction's late addition to the survey, every respondent in our dataset is over the age of 50, which is an important caveat to using this dataset and the reason in which we do not control explicitly for age in our later study.

Once exporting our data directly from the NLS website into a Stata datafile, we began data processing and cleaning. Throughout the process, missing variables were dropped, and all variables were relabeled and renamed for easy usage. New binary variables were created, such as *SATISFIED*, which represents all respondents who have a *LIFE_SATIS* score over 4, a subjective threshold we defined to create a binary cutoff threshold. Worth pausing, *SATISFIED* is heavily skewed towards being satisfied, as its mean is 0.883, implying a linear probability model of this variable has limitations in predicting a generalizable result as it relates to holding in the real world, addressed later in Results. Additionally, in this stage, we cleaned data by backfilling variables in which data was unclear, e.g., *COVID_STIM*, to have all missing values recoded accordingly. Furthermore, there are other changes to the explanatory variables that we performed. Specifically, for our education variable, in an effort to correct inconsistent data, we explicitly change *COLLEGE_EDUCATED* into a binary variable, denoting which individuals have more than 12 years of schooling. While this approach may sacrifice minimal variation in an explanatory variable, the data itself is inconsistent in its measurement. To be explicit, up until but not including the first year of our sample's interest, this variable decreases from 7,000 responses to just 174. Further, after 2014 and up through 2020, the variable dwindles down to just 45 observations; 2022 data is not available for this variable, which again highlights the inconsistency of this variable. Through these identified challenges with the data, we chose to be explicit in a binary college cutoff at the 12-year mark of education.

After all processing and cleaning, we transformed our dataset into a panel dataset, due solely to the fact our data is spanning over a 10-year range of the same individuals; by initializing a panel dataset, we can track, over time, how our model's data changes, and, in particular, whether these changes statistically change our model's prediction and fit. Within this dataset, while we try to track significant life events, e.g., falling into poverty, becoming unemployed, getting divorced, we ultimately can't control for all major life events in which we think there may be a statistically significant relationship with, like, for e.g., a death in the family.

In the Summary Statistics Table (Table 1) below, we have reproduced all variables used throughout our study, noting the number of observations (N), mean, median, standard deviation (S.D.), and minimum and maximum value. Within this table, beginning at the top, our main dependent variable is listed first, *LIFE_SATIS*, followed by the additional supplemental dependent measure created in the prior step, *SATISFIED*. After these two follows the highlighted independent variable of focus being used to identify the causal relationship in our study, *TOTAL_FAM_INC.* After these three variables, our control variables follow; as mentioned previously, these control variables were selected to explicitly control for characteristics in which we both think statistically impact life satisfaction and have data available in the NLS dataset for.

Variable	**N**	**Mean**	**Median**	**S.D.**	**Min**	**Max**
LIFE_SATIS	33702	5.811139	6	1.367415	1	7
SATISFIED*	33702	0.8828259	1	0.3216323	0	1

TOTAL_FAM_INC**	28549	90.52364	60	118.6415	0	922.631
POV_STATUS*	28549	0.1646643	0	0.3708838	0	1
EMPLOYED*	33644	0.6368149	1	0.4809247	0	1
COVID_STIM*	63430	0.1155132	0	0.3196427	0	1
MARIT_STATUS	33800	1.576243	1	1.364744	0	6
COLLEGE_EDUCATED*	37765	0.4583609	0	0.4982968	0	1
FAM_SIZE	33808	2.284075	2	1.265619	1	15

Table 1: Summary Statistics of Key Variables
Note (*): Binary variables are marked with a single asterisk
Note (**): TOTAL_FAM_INC is scaled in thousands to provide a cleaner interpretation of our results

Before wrapping the Data section, we have important considerations about our data to state. First, *COVID_STIM* is a variable that have implicit considerations and assumptions made with respect to how it's being measured. Initially, the extracted variable itself is a binary variable denoting if someone had received a COVID-19 stimulus check in the two years since the previous recorded survey; the measuring of this variable allows for no distinction between either the amount or frequency of the payment. Within this environment, there is also no clear distinction between the amount of stimulus an individual and/or household received at the time the survey was conducted. In addition, since *TOTAL_FAM_INC* is self-reported, it is unclear how individuals reported this supplemental stimulus, how much they received, and if they even reported it at all in this response.

IV. Results & Discussion

Approach I: OLS Estimation Model

To begin our analysis, we start with two ordinary least squares (OLS) regression models: 1) *SATISIFIED* on *TOTAL_FAM_INCOME*, where *SATISFIED* is a dummy variable representing life satisfaction (0 = Unsatisfied, 1 = Satisfied), and 2) *LIFE_SATIS* on *TOTAL_FAM_INCOME*, where *LIFE_SATIS* represents life satisfaction, as previously highlighted, scaled from one to seven. This acts as a baseline so we can see how independently the way in which scores are predicted by how much money is flowing into a given household. By establishing these first relationships, we believe we can move forward and utilize more complex econometric methods along with additional controls. We acknowledge that while using a binary dependent variable for an OLS estimation there are limitations, which complicate any generalizations. Specifically, when a binary variable is used in a model as the dependent variable, the coefficients produced are measured in percentages rather than a direct unit-by-unit change. Still, we believe these two models provides a starting point for exploring this relationship.

Using robust standard errors to address potential heteroskedasticity, we will estimate the following models below:

$$Y_{SATISFIED} = \beta_0 + \beta_1 TOTAL\_FAM\_INC + \epsilon$$

$$Y_{LIFE\_SATIS} = \beta_0 + \beta_1 TOTAL\_FAM\_INC + \epsilon$$

where β_0 represents the baseline score for each measure of satisfaction and $\beta_1 TOTAL\_FAM\_INC$ represents the change for each additional \$1,000 increase in total family income. Below are the results for each respective model:

Linear regression

Number of obs	=	28,506
F(1, 28504)	=	808.27
Prob > F	=	0.0000
R-squared	=	0.0167
Root MSE	=	.31398

SATISFIED	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
TOTAL_FAM_INC	.0003451	.0000121	28.43	0.000	.0003213	.0003689
_cons	.8557032	.0025056	341.51	0.000	.850792	.8606144

Figure 1: Regress SATISFIED (dummy) on TOTAL_FAM_INCOME (thousands of dollars)

Linear regression

Number of obs	=	28,506
F(1, 28504)	=	634.78
Prob > F	=	0.0000
R-squared	=	0.0157
Root MSE	=	1.3304

LIFE_SATIS	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
TOTAL_FAM_INC	.0014162	.0000562	25.19	0.000	.001306	.0015264
_cons	5.690063	.0106842	532.57	0.000	5.669122	5.711005

Figure 2: Regress LIFE_SATIS on TOTAL_FAM_INCOME (thousands of dollars)

Figure (1), a regression of *SATISFIED* on *TOTAL_FAM_INC*, produces a *_cons* coefficient, interpreted as 85.6% of people are satisfied in this sample when total family income is \$0. When interpreting the *TOTAL_FAM_INCOME* ‘s coefficient, we must consider it in terms of a percentage, so the model’s displayed result must be multiplied by 100 to get the true interpreted effect. After adjusting to percentage, this model predicts a 0.03% increase in life satisfaction scores for each additional thousands of dollars in total family income gained; to make this result more intuitive, we can also say this model predicts a 0.3% increase in life satisfaction scores for every \$10,000 increase in total family income.

Figure (2), regressing *LIFE_SATIS* on *TOTAL_FAM_INC*, has a more straightforward interpretation. Our *_cons* coefficient represents a predicted life satisfaction score of 5.69 when total family income is $0. Lastly, our coefficient for total family income represents how, for each additional thousand dollars of income reported, life satisfaction scores change. Explicitly, in this model, a $1,000 increase in total family income results in a .001 increase in someone's life satisfaction score; again, intuitively we can think about the results of this model as predicting an increase, on average, of 0.01in life satisfaction scores for every $10,000 increase in total family income.

Throughout both these models, we see high predicted life satisfaction scores, on average, when total family income is $0. Simultaneously within both models, we see *TOTAL_FAM_INC* remains statistically significant, highlighting that the relationship between total family income and life satisfaction scores are statistically significant. In addition, within both models, we notice a low R^2 score, indicating that a very small amount of life satisfaction scores can be explained by our model, in this case, total family income; in other words, our model's fit is weak, which prompts future expansions below.

Overall, these two OLS regressions showcase a starting point to examine the causal relationship, but they are not without limitations in what can be done with them. Further, no amount of additional control variables in these models will improve our model's fit, even if it does mitigate the clear omitted variable bias, solely since our data has been transformed into panel data. Importantly, OLS does not take into consideration our data is composed of the same individual changing year over year. While OLS allows for the control of year-specific effects in its model, it cannot handle time-constant unobserved individual factors as OLS assumes they are uncorrelated with our independent variable, total family income. With this limitation noted, we are going to move forward and introduce Fixed Effects (FE) models in Approach II, which we will use to account for panel structure, within-individual effects, and year fixed effects simultaneously alike.

Approach II: Fixed Effects Model

Moving to a Fixed Effects (FE) model, we begin by declaring our data as panel data in STATA, using *ID* to distinguish unique individuals and introduce *YEAR* to follow the same person over different years. By using FE, our estimator will eliminate the bias caused by time-constant unobserved individual-specific factors that are correlated with our independent variables; that is, FE will difference away the unobserved characteristics in our sample. Next, in our dependent variable, we swap out *SATISFIED* for *LIFE_SATIS*, the true reported data on life satisfaction, ranging between 1 and 7 (1 = not at all satisfied, 7 = completely satisfied), with each unit increase denoting an increase in life satisfaction. Lastly, we cluster *ID*, letting STATA know to cluster the standard errors at the individual level, that is, each *ID*. This last piece of setting up our model is crucial because observations within the same individual are likely to be correlated over time, that is, one respondent may be naturally more optimistic about their life compared to the rest of the data. By clustering, we increase the accuracy of our t-tests and their associated p-values; without clustering, our standard errors would be noticeably smaller and lead us to potentially find statistically significant results that aren't real, that is, a Type I error or a false positive. For our first FE model, we will estimate the following model:

$$Y_{LIFE\_SATIS} = \alpha_i + \gamma_t + \beta_1 TOTAL\_FAM\_INC + \varepsilon_{it}$$

where α_i captures the unobserved, time-invariant individual effects, γ_t captures the year fixed effects for each two-year data point. The results from this model are:

```
Fixed-effects (within) regression               Number of obs     =     28,506
Group variable: ID                              Number of groups  =      7,446

R-squared:                                      Obs per group:
     Within  = 0.0075                                         min =          1
     Between = 0.0291                                         avg =        3.8
     Overall = 0.0141                                         max =          5

                                                F(5, 7445)        =      38.30
corr(u_i, Xb) = 0.0712                          Prob > F          =     0.0000

                                 (Std. err. adjusted for 7,446 clusters in ID)
------------------------------------------------------------------------------
             |               Robust
  LIFE_SATIS | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
TOTAL_FAM_INC|   .0003729   .0000706     5.28   0.000     .0002344    .0005114
             |
        YEAR |
        2016 |  -.0806049   .0186168    -4.33   0.000    -.1170992   -.0441107
        2018 |  -.0076195   .0187509    -0.41   0.684    -.0443765    .0291375
        2020 |  -.0657188   .0196337    -3.35   0.001    -.1042063   -.0272312
        2022 |   .1264354   .0190667     6.63   0.000     .0890593    .1638114
             |
       _cons |   5.792044   .0127284   455.05   0.000     5.767092    5.816995
-------------+----------------------------------------------------------------
     sigma_u |   1.121111
     sigma_e |  .96034813
         rho |  .57677768   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 3: Simple Fixed Effects Model

Similar to our previous model, *TOTAL_FAM_INC* has a coefficient of 0.0003729, representing the change in *LIFE_SATIS* for every $1,000 increase, or about an increase of 0.003729 in satisfaction for every extra family income of $10,000, and is statistically significant. However, the interpretation of this coefficient is different as we're now using *LIFE_SATIS*, which is the actual range of satisfaction scores instead of a binary variable. Worth highlighting here is these results are merely coincidental since *SATIS* and *LIFE_SATIS* are measured two entirely different ways. It's important to note, this coefficient does not tell us about the difference in life satisfaction between people with different income levels, instead it only tells us how changes in income affect changes in life satisfaction within the same person. Our four dummies denote potential shocks to life satisfaction across everyone, on average, in comparison to the base year 2014. Three dummies, 2016, 2018, and 2020 are negative, with 2020 having a noticeable negative shock in comparison to the base 2014 year; in contrast, 2022, the last year in our data's sample, there was, on average, a positive shock of values in contrast to our base 2014 year.

Additionally, there a few more things worth noting here beginning with the p-values in our model. 2018 is the only year in our model that is not statistically significant with a p-value of 0.684, indicating there was no significant difference in life satisfaction compared to the base year, controlling for income and individual fixed effects. Next, in 2020, as noted previously, we saw a large drop in life satisfaction scores on average in comparison to the base year 2014, which likely reflects the impact of the COVID-19 global pandemic. Lastly, 2022 was significantly higher at a coefficient of 0.126 units after controlling for income and individual fixed effects in comparison to the base year 2014, possibly representing the recovery from the pandemic or a positive change of lifestyle post-pandemic. As with most FE models, the *_cons* coefficient may not be useful since it represents when family income is $0, all year dummies are zero, and for an individual with a fixed effect of zero. That is, to interpret *_cons* here, we would be predicting the *LIFE_SATIS* score of an individual in 2014 with a $0 *TOTAL_FAM_INC,* producing a *LIFE_SATIS* score of 5.79.

Lastly to note with this model, our sigma values here represent the standard deviation of fixed effects, both of individual fixed effects (*sigma_u*) and the error term (*sigma_e*). Our *rho* value of 0.577 is the fraction of total variance in life satisfaction due to individual fixed effects, meaning roughly 58% of the variance in life satisfaction is due to the differences between individuals and 42% is due to the variation within individuals over time. All of this, however, only puts our overall R^2 at 0.0141, meaning our model only explains around 1.41% of the variation in life satisfaction, so we move forward by expanding this model to include additional controls.

To build onto our previous FE model, we now control directly for other variables we believe affect life satisfaction with a vector, **X**, of controls, controlled at the individual and time fixed level: marital status, employment status, poverty status, college status, family size, and the effect of the government implemented COVID stimulus checks, which, as we discussed above in the Data section (and later in Limitations), boosted income when received. Specifically,

$$Y_{LIFE\_SATIS} = \alpha_i + \gamma_t + \beta_1 TOTAL\_FAM\_INC + \mathbf{X}_{it}\beta + \varepsilon_{it}$$

Replicating the previous technique, clustering at the *ID* level again, our results produce:

Fixed-effects (within) regression — Number of obs = **27,014**
Group variable: **ID** — Number of groups = **6,894**

R-squared: — Obs per group:
Within = **0.0129** — min = **1**
Between = **0.0692** — avg = **3.9**
Overall = **0.0432** — max = **5**

F(**13, 6893**) = **19.35**
corr(u_i, Xb) = **0.0959** — Prob > F = **0.0000**

(Std. err. adjusted for **6,894** clusters in **ID**)

LIFE_SATIS	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
TOTAL_FAM_INC	.0002089	.000071	2.94	0.003	.0000697	.0003481
MARIT_STATUS						
1: 1 MARRIED	.1759009	.1735025	1.01	0.311	-.1642175	.5160193
2: 2 SEPARATED	-.1435157	.1937769	-0.74	0.459	-.5233781	.2363468
3: 3 DIVORCED	.0707286	.1772221	0.40	0.690	-.2766814	.4181386
6: 6 WIDOWED	-.1018785	.1938414	-0.53	0.599	-.4818673	.2781104
EMPLOYED	.1417242	.0248098	5.71	0.000	.0930893	.1903591
POV_STATUS	-.1403253	.0380892	-3.68	0.000	-.2149918	-.0656587
COLLEGE_EDUCATED	0	(omitted)				
FAM_SIZE	-.0083954	.0120589	-0.70	0.486	-.0320345	.0152438
COVID_STIM	.0539819	.0244874	2.20	0.028	.005979	.1019848
YEAR						
2016	-.0826365	.0189581	-4.36	0.000	-.1198003	-.0454727
2018	.0025063	.0194359	0.13	0.897	-.035594	.0406066
2020	-.0849436	.0279113	-3.04	0.002	-.1396582	-.0302289
2022	.1271122	.0239356	5.31	0.000	.0801909	.1740334
_cons	5.651528	.151287	37.36	0.000	5.354959	5.948097
sigma_u	1.0837307					
sigma_e	.95597263					
rho	.56239085	(fraction of variance due to u_i)				

Figure 4: Expanded Fixed Effects Model

After expanding our FE model, our *TOTAL_FAM_INC* coefficient has decreased slightly from the previous model to 0.0002089, meaning that, for every $1,000 increase in total family income for a given individual, their life satisfaction score is expected to increase by 0.0002089 on average, holding everything else constant. While this coefficient is small, again it is statistically significant with a p-value of 0.003. If an individual becomes employed, changing the *EMPLOYED* variable from 0 to 1, their life satisfaction score is expected to increase by 0.142 units, on average, while holding all other controls constant and is statistically significant. For *POV_STATUS*, a change from a 0 to 1—indicating falling below the poverty line—is associated with a 0.140 statistically significant unit decrease in life satisfaction. As prefaced, *COVID_STIM* appears to have a positive effect on life satisfaction scores, increasing it by, on average, about 0.054 units while holding all other controls constant; *COVID_STIM* is statistically significant as well. One possible interpretation for this result is that individuals received additional income they were not expecting with these stimulus checks, possibly boosting their current living situation and, in return, their life satisfaction score.

While our year fixed effects dummies have the same interpretation as before, there's some changes worth noting. While 2018's year fixed coefficient was statistically insignificant before, the p-value has increased significantly more in this new model, 0.685 previously to 0.885 now, and now has a small positive effect on life satisfaction scores, instead of a negative effect as seen in the previous model in comparison to the base year 2014. While the other years remain almost identical, it's worth noting, despite the previous finding of the effect of a COVID stimulus, 2020's effect is now more negative than it was in the previous model, despite COVID checks seeming to play a positive effect on life satisfactions on their own when holding all other constants constant. Again, 2016, 2020, and 2022 remain statistically significant at all levels compared to their 2014 baseline value.

In addition to the previous results, when controlling for marital status in reference to the base category, "never married", it turns out all options, married, separated, divorced, and widowed, are all statistically insignificant, having a p-value above 0.05. While this doesn't mean marital status doesn't affect life satisfaction, it does mean that when holding everything else constant, changes in marital status are not significantly associated with changes in life satisfaction within individuals for the people in our sample. With this control, it's also possible there was not a lot of overall variation in this control, that is, people weren't going from single to married, married to widowed, etc., in the selected sample. A change in family size, *FAM_SIZE*, is also insignificant, meaning changes in family size, like adding an additional child to a household, are not significantly associated with changes in life satisfaction. Interestingly, *COLLEGE_EDUCATED* was omitted due to collinearity, meaning its perfectly collinear with one or more of the variables in this model. Because we are using a fixed effects model, if a variable does not change over time for any individual, it's perfectly collinear with the individual fixed effects because each person's fixed effects already capture the average value of all time-constant characteristics. Since this variable doesn't change, the fixed effect absorbs all its effect, leaving no variation to explain any of the variation in life satisfaction scores.

Lastly, our *rho* value, or total variance in life satisfaction due to individual fixed effects, remains roughly 57%. Our R^2 remains low at 0.0432, representing our model only explains around 4% of the variation in life satisfaction. The final statistic we'll mention in our model here is the reported F-statistic, which has a value of 38.30, denoting the model as whole is statistically significant (p-value < 0.0000), meaning all our control variables taken together have a statistically significant effect on life satisfaction scores. With these two FE iterations, we believe we still have not identified a meaningful causal effect of *TOTAL_FAM_INC* on *LIFE_SATIS*. In another attempt to try to capture the causal effect, we will make use of our ordinal dependent variable and move to a Proportional Odds Model in the next section.

Approach III: Proportional Odds Model

In a final attempt to correctly identify if a causal relationship exists between life satisfaction scores and total family income, we introduce a third econometric approach, a proportional odds model (POM), sometimes referred to an ordinal logistic regression. Specifically,

$$\log \frac{\mathrm{P}(Y_{it} \leq k)}{\mathrm{P}(Y_{it} > k)} = \theta_k - (\beta_1 TOTAL\_FAM\_INC + \boldsymbol{X}_{it}\beta + \alpha_i + \gamma_t)$$

where Y_{it} is the ordinal outcome, represented as a cumulative probability of each life satisfaction score category, k indexes the cut off for each category, $\boldsymbol{X}_{it}\beta$ remain our vectors of control variables, and α_i and γ_t represent the individual and time fixed effects respectively.

Since our dependent variable *LIFE_SATIS* is ordinal, that is, it takes on a value of 1 through 7 (1 = not at all satisfied, 7 = completely satisfied), a proportional odds model will fit this ordinal dependent variable under three assumptions: 1) each "step" in the variable are equally sized apart, 2) the odds of being in a lower or higher category life satisfaction score are consistent across all individuals, and 3) higher scores equate to better outcomes. Worth noting, the second assumption is the only assumption being made of the three. Since all three assumptions needed for an ordinal logistic regression are met, the model produces the following results:

```
Iteration 0:  Log pseudolikelihood = -38940.589
Iteration 1:  Log pseudolikelihood = -38412.816
Iteration 2:  Log pseudolikelihood = -38411.804
Iteration 3:  Log pseudolikelihood = -38411.804

Ordered logistic regression                     Number of obs = 27,014
                                                Wald chi2(14) = 599.71
                                                Prob > chi2   = 0.0000
Log pseudolikelihood = -38411.804               Pseudo R2     = 0.0136

                              (Std. err. adjusted for 6,894 clusters in ID)
```

LIFE_SATIS	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
TOTAL_FAM_INC	.0007282	.0001376	5.29	0.000	.0004584	.0009979
MARIT_STATUS						
1: 1 MARRIED	.397807	.0609378	6.53	0.000	.2783711	.517243
2: 2 SEPARATED	.1135675	.1106264	1.03	0.305	-.1032562	.3303912
3: 3 DIVORCED	-.131152	.066184	-1.98	0.048	-.2608703	-.0014337
6: 6 WIDOWED	.0212088	.1009021	0.21	0.834	-.1765557	.2189732
EMPLOYED	.3258348	.0366424	8.89	0.000	.254017	.3976526
POV_STATUS	.0025867	.0555466	0.05	0.963	-.1062825	.111456
COLLEGE_EDUCATED	-.3103991	.0378626	-8.20	0.000	-.3846085	-.2361898
FAM_SIZE	.0116908	.014941	0.78	0.434	-.0175931	.0409747
COVID_STIM	-.1203784	.0409332	-2.94	0.003	-.200606	-.0401509
YEAR						
2016	-.1082474	.0258477	-4.19	0.000	-.158908	-.0575869
2018	.0615699	.0277351	2.22	0.026	.0072102	.1159297
2020	.0738165	.0415396	1.78	0.076	-.0075995	.1552326
2022	.3777606	.0351334	10.75	0.000	.3089005	.4466207
/cut1	-3.4186	.0843413			-3.583906	-3.253294
/cut2	-3.029283	.0804848			-3.18703	-2.871535
/cut3	-2.399964	.0757105			-2.548354	-2.251574
/cut4	-1.721261	.0727395			-1.863827	-1.578694
/cut5	-.2776791	.0700424			-.4149596	-.1403986
/cut6	.8701658	.070068			.7328351	1.007497

Figure 5: Ordinal Logistic Regression

To begin our interpretation, let’s start by the statistics at the top of the output. Our model’s p-value, denoted at “Prob > chi2” is < 0.000, indicating that the model is statistically significant; the model’s statistical significance indicates the independent variables, taken together in our model, are significantly related to life satisfaction. However, our pseudo R^2 is 0.0136, again signaling our model explains only about 1.4% of the variance in the outcome. As opposed to our previous FE model’s clustering, this ordinal logistic regression clusters our robust standard errors around only 6,894 IDs.

Beginning with the output seen in the results figure above, our dependent variable *TOTAL_FAM_INC*’s coefficient, 0.0007282, says that for every $1,000 increase in total family income, the logarithmic odds of having a higher life satisfaction score increases by 0.0007282. However, interpreting an ordinal logistic regression is more nuanced, requiring a calculation of marginal effects for each variable in the model to capture their effect, so we will focus solely on the marginal effects of our independent variable of interest, *TOTAL_FAM_INC*, for both brevity and limitations of familiarity with this model's higher logic. As mentioned, to interpret these results, it's necessary to get the marginal effects, producing:

	dy/dx	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
TOTAL_FAM_INC						
_predict						
1	**-.0000154**	**3.00e-06**	**-5.15**	**0.000**	**-.0000213**	**-9.54e-06**
2	**-6.88e-06**	**1.38e-06**	**-4.99**	**0.000**	**-9.58e-06**	**-4.17e-06**
3	**-.0000173**	**3.37e-06**	**-5.13**	**0.000**	**-.0000239**	**-.0000107**
4	**-.0000304**	**5.83e-06**	**-5.21**	**0.000**	**-.0000418**	**-.000019**
5	**-.000093**	**.0000176**	**-5.28**	**0.000**	**-.0001276**	**-.0000585**
6	**-9.04e-06**	**2.40e-06**	**-3.77**	**0.000**	**-.0000137**	**-4.34e-06**
7	**.0001721**	**.0000325**	**5.29**	**0.000**	**.0001084**	**.0002358**

Figure 6: Total Family Income's Marginal Effects

With the marginal effects above, we can see how, while holding all over variables at their means in the model, increases in total family income affect the probability of life satisfaction scores. For a simple analysis, we see that a $1,000 increase in total family income only positively increases the probability of higher satisfaction score for category 7, very satisfied. Intuitively, these results suggest that increased income is associated with a small shift away from the lower life satisfaction scores and towards the highest life satisfaction score. Lastly to note, all p-values are significant here, noting that there is a statistically significant relationship for every life satisfaction category as it relates to total family income. Importantly, all categories shift away from a lower score towards the highest life satisfaction with increased income, meaning this relationship is consistent across the sample. In other words, if some categories shifted positively and some negatively, there would be a potential mismatch of the overall relationship seen.

However, since all life satisfaction categories shift towards category 7, Very Satisfied, we feel we have established a meaningfully statistically significant relationship in this model.

After examining three different approaches for this study, we move on to the next section, Limitations, in which we detail the limits to these results presented overall.

V. Limitations

Despite confidence in our previous results, we believe there are many limitations that prevent us from claiming these results to be a true causal effect. Specifically, our limitations include external validation, selection bias, data availability and inconsistencies, and model fit, which will be detailed briefly now.

Beginning with external validation, or the applicability of these empirical results to the real world, our econometric study faces an important limitation. While we feel we have established a positive relationship between life satisfaction scores and total family income, we cannot say that these results are generalizable to other populations, environments, or even the constraints of the same dataset as more data is made available given the data for our dependent variable has only begun being reported in 2014. Further, because of this later start date, all respondents from the NLSY79 sample are over the age of 50, leading our empirical study to suffer from selection bias. Even within the data used in this study, the environment in which the data has been collected has also changed. Since 2014, technology has become an interwoven fabric to the everyday modern life, including, and not limited to, the usage of social media, which other prominent literature shows varying negative effects on individuals and their health. While this is a limitation in this study, we pose it as an opportunity for future research to compare a different population subset, e.g., born 1990-2000 or 2010-2020, where this changed environment would, we believe, highlight the depth of this limitation greatly.

Within the data for *LIFE_SATIS* itself, there were also important limitations. Specifically, the data itself doesn't have much variance on a year-to-year basis, despite shocks occurring in the environment of this study, such as 2016 and 2020 for various political reasons and/or the global pandemic that occurred. By solely using this data alone, we may have also limited the generalizability that may have been found when compared with other data sources and/or measures of life satisfaction data. In addition, this limited data, along with limited controls, made it difficult to find a good model fit throughout the different econometric approaches we took to correct this shortcoming. With consistently low model fits, it further raises questions of clear causal relationship identification and generalizability for future empirical studies. The combination of these limitations, external validity, selection bias, and lack of better data, paired with natural time constraints, raise caution to the results of this study. Despite all these shortcomings, we feel confident these results are still meaningful progress in identifying a causal relationship for the variables of interest in this study.

VI. Conclusion

Throughout this econometric analysis, we've established three different econometric approaches to estimate the causal effect of *TOTAL_FAM_INC* on *LIFE_SATIS*: OLS, FE, and POM. With that, we believe the Proportional Odds Model (Approach III) is our best approach to identify the true effect that income has on an individual's perceived life satisfaction. Although no model identifies a generalizable result, it is still clear there is a relationship between the two variables of interest in our study, and positive one at that. While this may be disappointing to not have a better causal effect identified, it is clear our initial hypothesis that total family income is statistically meaningful in predicting life satisfaction scores, as seen by our consistent statistically significant variables and their model. Further, this study is not alone; in fact, this is a result seen in similar studies and literature, such as, Schieve (1992), highlighted in our Literature Review, who faced a similar issue in that life satisfaction was influenced best by variables that he didn't control for.

As highlighted, there is room for future improvements in this field of study. Future expansions would highlight the need for improved data; while the NLS dataset is a reputable longitudinal dataset spanning the same individuals over decades, it doesn't come without its own challenges in the data itself, how it's reported, and who the respondents are. Through various control variable additions and modeling specifically for panel data both to account for fixed effects and time observed invariant effects and the ordinal design of our data, we believe we have demonstrated that money does, indeed, buy life satisfaction.

References

Bertermann, A., Kamhöfer, D., & Schildberg-Hörisch, H. (2023). More Education Does Make You Happier – Unless You Are Unemployed. *IZA Institute of Labor Economics - Discussion Paper Series*. https://doi.org/10.2139/ssrn.4578417

Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1979 cohort, 1979-2020 (rounds 1-29). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 2023.

Grün, C., Hauser, W., & Rhein, T. (2010). Is Any Job Better Than No Job? Life Satisfaction and Re-employment. *Journal of Labor Research*, *31*(3), 285–306. https://doi.org/10.1007/s12122-010-9093-2

Hartog, J., & Oosterbeek, H. (1998). Health, Wealth and Happiness: Why Pursue a Higher Education? *Economics of Education Review*, *17*(3). https://doi.org/10.1016/S0272-7757(97)00064-2

Lee, K. S., & Yang, Y. (2022). Educational Attainment and Emotional Well-being in Adolescence and Adulthood. *SSM - Mental Health*, *2*. https://doi.org/10.1016/j.ssmmh.2022.100138

Leonard, W. M. (1982). Successful Aging: An Elaboration of Social and Psychological Factors. *The International Journal of Aging and Human Development*, *14*(3), 223–232. https://doi.org/10.2190/vu7r-bj90-lc2m-u696

Möwisch, D., Brose, A., & Schmiedek, F. (2020). Do Higher Educated People Feel Better in Everyday Life? Insights From a Day Reconstruction Method Study. *Social Indicators Research*, *153*(1), 227–250. https://doi.org/10.1007/s11205-020-02472-y

Samman, E., & Santos, M. E. (2013). Poor and Dissatisfied? Income Poverty, Poverty Transitions and Life Satisfaction in Chile. *Journal of Poverty and Social Justice*, *21*(1), 19–31. https://doi.org/10.1332/175982713x664038

Schieve, A. D. (1992). The contribution of educational attainment to the prediction of life satisfaction in the NLS-72 data base (Order No. 9234017). Available from ProQuest One Academic. (303988598). http://ezproxy2.library.drexel.edu/login?url=https://www.proquest.com/dissertations-theses/contribution-educational-attainment-prediction/docview/303988598/se-2

Schmitt, N., & Mellon, P. (1980). Life and Job Satisfaction: Is the Job Central? *Journal of Vocational Behavior*, *16*(1). https://doi.org/10.1016/0001-8791(80)90037-8

Winkelmann, L., & Winkelmann, R. (1998). Why Are the Unemployed so Unhappy? Evidence From Panel Data. *Economica*, *65*(257). https://doi.org/10.1111/1468-0335.00111