The Effects of Individual Welfare Income on Self-Reported Health Status

Authors: Bella Davies, Mariah Hastings, Isabel Yin
Boston University
May 2022

Abstract: This research paper seeks to examine whether a relationship exists between individuals' welfare income and their self-reported health status during 2010 and 2014. This relationship is important to study since it can provide insight into how welfare recipients' health is affected by the amount of welfare income received, and how it can be beneficial for the recipients of welfare income to have consistent and reliable access to nutritious food or healthcare. If a positive relationship exists between welfare income and better health status, this may point to the necessity of revising or implementing new policies regarding welfare distribution and accessibility. A lack of reliable access to food (a significant issue faced by many who require welfare assistance) is associated with many health issues, including but not limited to birth defects, cognitive and behavior issues, anemia, and more. Insufficient amounts of welfare can also impede individuals' access to medical care, worsening their health status. Past studies have found a possible positive and/or weak relationship between welfare income and health status, however these studies primarily examined welfare income from the Supplemental Nutrition Assistance Program (SNAP), and not total welfare income (which includes other government assistance, some temporary) received by individuals. We chose to look at the years 2010 and 2014, the years that follow the start and end of the American Recovery and Reinvestment Act (ARRA), which began in 2009 and had many of its temporary welfare increases, including for food stamps, expire in 2013. To study this effect, we used data from the Current Population Survey (CPS) to look at individual data from these two years for multiple x variables, including *inc_welfr*, *educ_rev*, *married*, *west*, *south*, *midwest*, *mthwelfr*, *famsize*, *sex*, *age,* and *health* for the y variable. We used this data to run simple and multiple regression models to determine whether these factors result in a difference in individuals' self-reported health. Using the linear-log functional form of the multiple regression model, we found that there is a weak and positive but statistically insignificant association between welfare income and health.

Section I:

This paper examines the annual amount of welfare, or pre-tax income received by individuals from various public assistance programs, and their self-reported health status during the calendar years 2010 and 2014, in order to determine how the amount of welfare a person receives can affect their health status. We chose to use the years 2010 and 2014 because the provisions and welfare increases stipulated by the American Recovery and Reinvestment Act of 2009 should have been in full effect in 2010, and 2014 is the year right after the food stamps expired in 2013. Most other temporary welfare payments or increases created by the act would have expired by 2014 as well. Using the years 2010 and 2014 allows us to compare the effects of increased welfare income in 2010 versus decreased welfare income in 2014.

Previous studies have found a possible correlation between welfare income and health status, showing a positive association where increased amounts of welfare income are associated with a better or higher self-reported health status while decreased amounts of welfare income are associated with a worse or lower self-reported health status (Narain et al., 2017; Gregory & Deb, 2014). The data from the years 2010 and 2014 should show a similar trend, since during February of 2009, the American Recovery and Reinvestment Act was signed and put into effect in order to alleviate the effects of the Great Recession (Katare & Kim, 2017). The act included 3.2 billion in temporary welfare payments, 14.2 billion for $250 one-time payments to certain welfare recipients, 19.9 billion for the food stamp program, and the Supplemental Nutrition Assistance Program or SNAP benefits for a household of four increased by $80 per month (Katare & Kim, 2017). Then on November 1st, 2013 when food stamp benefits expired, there was a significant cut in food stamps across the country, which caused a family of four to lose $36 per month (Katare & Kim, 2017). This led to a rise in food insecurity, or an increased amount of individuals without reliable or steady access to affordable, nutritious food of a sufficient quantity in the years from 2010 to 2014 (Katare & Kim, 2017).

Food insecurity is linked to a large variety of health issues, including birth defects, cognitive and behavior problems, lower nutrient intakes, anemia, and much more (Zhang et al., 2021). These health issues may have been exacerbated by some individuals' lack of accessibility to good healthcare as a result of decreased welfare. Due to the increased welfare benefits from the American Recovery and Reinvestment Act of 2009 and the food stamp cut and expiration of other temporary welfare increases by 2013, the expected association is a positive trend between total individual welfare income and self-reported health status. This is significant as it would demonstrate how changes in government policies on welfare are correlated with the health status of individuals receiving welfare income.

The model used in this paper is a linear probability model because the x-variable, welfare income, is continuous while the y-variable, health, is converted into a dummy variable. As a result, all coefficients in the model are interpreted as probabilities. We found the best fitting regression model to be the linear-log multiple regression model, as its sum of squared errors was the least compared to all other functional forms. Our findings aligned with our expectations, as the multiple linear-log model showed a weak and insignificant positive relationship between welfare income and the probability of self-reporting good health status, which aligns with the study conducted by Miller and Morrisey (2017), discussed below in Section II.

The outline of this paper is as follows. **Section II** is the literature review, which analyzes three different sources examining similar studies that compare various forms of welfare income with self-reported health status using a regression model. **Section III** presents the simple and multiple econometric regression models in the linear, quadratic, cubic, and linear-log functional forms. **Section IV** discusses data set information and descriptive statistics. **Section V** explores the results of the simple and multiple regression models and their comparable functional forms. **Section VI** discusses the conclusions and their implications. **Section VII** includes all references used in this paper, and **Section VIII** is the appendix for the Stata do-file and all tables and graphs used.

Section II: Literature Review

Our study examines the specific relationship between total individual welfare income and self-reported health status. Previous studies with similar topics have been conducted on the effects of welfare reform or public assistance income (such as SNAP benefits) on health, which serve as the past literature on which we are basing some of our expectations for our own study. One such past study examined the impact of The Personal Responsibility Work Opportunity and Reconciliation Act (PRWORA) of 1996, which implemented time limits for receiving welfare cash benefits and imposed sanctions on cash benefits for failures to meet work requirements, on, among other variables, the self-reported health status of low education single mothers (Narain et al., 2017). While previous studies found no effect of PRWORA on health outcomes, Narain et al. (2017) notes that the time periods of data used in those studies occurred before all states had imposed time limits and before people were actually timing out of their welfare benefits, suggesting a possible limitation of those studies (Narain et al., 2017). Using data from the Survey of Income and Program Participation from 1991-2009 (after all states had implemented these welfare time limits) and OLS regression models, Narain et al. (2017) concluded that PRWORA led to a 7% increase in self-reported poor health for low education white single mothers. While not directly related to our study, the implementation of the time limits in PRWORA indicated a decrease in welfare cash income for welfare recipients who timed out, suggesting that decreased welfare income could lead to worse self-reported health among this sample being studied.

Also measuring the effects of welfare on health, Gregory and Deb (2014) examined the specific impact of SNAP welfare income on several measures of health including self-reported health outcomes, finding a significant increase in ‘excellent’ or ‘very good’ responses in self-reported health (Gregory & Deb, 2014). This study attempted to explore this relationship outside of diet-related outcomes and with a nationally representative sample, leading to different conclusions than similar studies that found worse self-reported health outcomes for SNAP recipients than non-SNAP recipients (Gregory & Deb, 2014). Using data from the Medical Expenditure Panel Survey (MEPS) from 1999–2008 (a probability sample of the US non-institutionalized population) and multivariate regression models, Gregory and Deb (2014) found that SNAP participants had better health as measured by self-reported health data, doctor visits, and sick days taken (Gregory & Deb, 2014). Since food stamps are often included in welfare income, this study indicates a possible positive relationship between welfare income and better self-reported health status.

Similarly, Miller and Morrissey (2017) explored the effect of SNAP benefits on health outcomes for adults and children. Using data from the National Health Interview Study (NHIS) from 2008 to 2014, least square IV models, and regression models, Miller and Morrissey found improved health outcomes from SNAP receipt (due to increased affordability for medical care resulting from receiving welfare) after accounting for SNAP variation across states (Miller & Morrissey, 2017). However, they also noted that the effects of the temporary SNAP increase in the American Recovery and Reinvestment Act (ARRA), which they studied in their regression model, did not have significant changes on health (Miller & Morrissey, 2017). Since we are looking at years before and after the implementation of the SNAP benefits expansion (2010 and 2014) in our study, this study contradicts our initial expectations by suggesting a potentially weak or insignificant relationship between welfare income and improvements in health status. Differences between the results of our model and the findings of this study could potentially be because we are accounting for not only food stamp amount variations, but total individual welfare income changes (which includes other government assistance or temporary welfare payments) from 2010 to 2014 that may have also impacted self-reported health status. Overall, these three past literatures' findings point to the likelihood of a weak and/or positive relationship between the variables of total welfare income and self-reported health.

Section III: Econometric Model

Simple Linear Regression Models:

(1) goodhealth = $\beta_1 + \beta_2(adj\_welfr) + e$

(2) goodhealth = $\beta_1 + \beta_2(adj\_welfr\_sq) + e$

(3) goodhealth = $\beta_1 + \beta_2(adj\_welfr\_cu) + e$

(4) goodhealth = $\beta_1 + \beta_2(adj\_welfr\_log) + e$

The fundamental goal of this paper is to determine whether higher welfare income is related to better health status. Our simple regression models are our foundation for future multiple regression models investigating this relationship—we did not control for the variables of age, region, number of family members in a household, marital status, sex, and education in these simple regression models. Our parameter of interest for our simple regression is β_2, which is the population coefficient of our main independent variable adj_welfr. Our nonexperimental data verifies the OLS assumptions, so we are able to regress goodhealth on adj_welfr using simple regression models. For our model, these OLS assumptions include goodhealth being normally distributed (following the Central Limit Theorem since our sample size is sufficiently large), with a mean of $E(goodhealth|adj\_welfr) = \beta_1 + \beta_2(adj\_welfr)$, variance $Var(goodhealth|adj\_welfr) = \sigma^2$, and the fact that there is no relationship/correlation between different individual's health measures that have been collected.

Our first simple linear population regression, Model (1), shows a unit increase in the independent variable adj_welfr leads to an increase in the dependent variable goodhealth, or the probability that a person will have a good measure of self-reported health. The first simple linear regression model aligns with our initial predictions based on past literature: a weak but positive relationship between adj_welfr and goodhealth—the probability of having good health rises with increases in adj_welfr.

Models (2), (3), and (4) include different functional forms we tested in order to account for potential functional form errors produced by Model (1). Model (2) is a quadratic regression model using adjusted welfare squared (adj_welf_sq) to observe its effect on the probability a person has a good measure of health (goodhealth). Model (3) is a cubic regression model using adjusted welfare cubed (adj_welf_cu) to observe its effect on the probability a person has a good measure of health (goodhealth). Model (4) is a linear-log regression model using the log of the adjusted welfare to observe its effects on the probability a person has a good measure of health (goodhealth).

The best simple regression model is Model (2), or the quadratic simple regression model. This is because the SSE, or the sum of squared errors, is the lowest in this model compared to the linear, cubic, and linear-log functional forms, or Models (1), (3), and (4), respectively. The SSE of Model (1) is equal to 564.414004, the SSE of Model (3) is equal to 564.020909, the SSE of Model (4) is equal to 565.367483 while the SSE of Model (2) is equal to 563.640473. The lower SSE of the quadratic model shows that this model has the best fit to the actual data and is the best functional form to show the relationship between adjusted welfare income and the probability a person has a good measure of health.

Multiple Regression Models:

$$(5)\ \text{goodhealth} = \beta_1 + \beta_2(adj\_welfr) + \beta_3(educ\_rev) + \beta_4(married) + \beta_5(west) + \beta_6(south) + \beta_7(midwest) + \beta_8(mthwelfr) + \beta_9(famsize) + \beta_{10}(sex\_rev) + \beta_{11}(age) + \beta_{12}(adj\_welfr\_famsize) + \beta_{13}(adj\_welfr\_educ\_rev) + e$$

$$(6)\ \text{goodhealth} = \beta_1 + \beta_2(adj\_welfr\_sq) + \beta_3(educ\_rev\_sq) + \beta_4(married) + \beta_5(west) + \beta_6(south) + \beta_7(midwest) + \beta_8(mthwelfr\_sq) + \beta_9(famsize\_sq) + \beta_{10}(sex\_rev) + \beta_{11}(age\_sq) + \beta_{12}(adj\_welfr\_famsize\_sq) + \beta_{13}(adj\_welfr\_educ\_rev\_sq) + e$$

$$(7)\ \text{goodhealth} = \beta_1 + \beta_2(adj\_welfr\_cu) + \beta_3(educ\_rev\_cu) + \beta_4(married) + \beta_5(west) + \beta_6(south) + \beta_7(midwest) + \beta_8(mthwelfr\_cu) + \beta_9(famsize\_cu) + \beta_{10}(sex\_rev) + \beta_{11}(age\_cu) + \beta_{12}(adj\_welfr\_famsize\_cu) + \beta_{13}(adj\_welfr\_educ\_rev\_cu) + e$$

$$(8)\ \text{goodhealth} = \beta_1 + \beta_2(adj\_welfr\_log) + \beta_3(educ\_rev\_log) + \beta_4(married) + \beta_5(west) + \beta_6(south) + \beta_7(midwest) + \beta_8(mthwelfr\_log) + \beta_9(famsize\_log) + \beta_{10}(sex\_rev) + \beta_{11}(age\_log) + \beta_{12}(adj\_welfr\_famsize\_log) + \beta_{13}(adj\_welfr\_educ\_rev\_log) + e$$

Model (5) is a linear multiple regression model using our main x variable, adjusted welfare, along with the additional x variables education level, marital status, living in the west, south, or midwest regions, months received welfare, number of family members in household, sex, age, and two interaction terms (discussed below) to observe the effects of these variables on the probability of a good measure of health. We chose to include these additional x variables in Models (5), (6), (7), and (8) because they control for what we feel is the most endogeneity in our simple regression models by taking into account the effects of region, education, marital status, etc. on the probability of good health.

Models (6), (7), and (8) include different functional forms we tested in order to account for potential functional form errors produced by our linear multiple regression Model (5). Model

(6) is a quadratic multiple regression model using the square of all our continuous variables (adj_welf_sq, educ_rev_sq, mthwelfr_sq, famsize_sq, age_sq, adj_welfr_famsize_sq, and adj_welfr_educ_rev_sq) and same dummy variables (sex, married, south, west, midwest) to observe the effect of squaring the continuous variables on goodhealth. Model (7) is a cubic regression model using the cube of all our continuous variables (adj_welf_cu, duc_rev_cu, mthwelfr_cu, famsize_cu, age_cu, adj_welfr_famsize_cu, and adj_welfr_educ_rev_cu) and same dummy variables (sex, married, south, west, midwest) to observe the effect of taking the cube of the continuous variables on goodhealth. Model (8) is a lin-log regression model using the log of all our continuous variables (adj_welf_log, educ_rev_log, mthwelfr_log, famsize_log, age_log, adj_welfr_famsize_log, and adj_welfr_educ_rev_log) and same dummy variables (sex, married, south, west, midwest) to observe the effect of taking the log of the continuous variables on goodhealth.

The best multiple regression model is Model (8), or the linear-log multiple regression model. This is because the SSE, or the sum of squared errors, is the lowest in this model compared to our other multiple regression models, or Models (5), (6), and (7). The SSE of Model (5) is equal to 483.438778, the SSE of Model (6) is equal to 493.70597, and the SSE of Model (7) is equal to 507.698515, while the SSE of Model (8) is equal to 476.170119. Therefore, Model (8) is the best multiple regression model to explain the relationship between adjusted welfare income and a good measure of self-reported health. This is due to the fact that the logarithmic function corrected the right skewness of the adjusted welfare income variable, adj_welfr, and the addition of variables controlled for more endogeneity than in the simple regression models. As seen in Graph 1, the density function of adj_welfr shows a very right skewed distribution. After taking the natural logarithm of adj_welfr as represented by the variable adj_welfr_log, the density function shows a more normal distribution, albeit a slight left skewness. This can be seen in Graph 2.

Interaction Terms:

In our multiple regression models, we included two interaction terms: adj_welfr_famsize and adj_welfr_educ_rev. We included adj_welfr_famsize in order to check whether the effect of adjusted welfare income varies with the number of people living in a household. For Model (8), the effect of an extra person living in the household on the probability a person has a good measure of health is 0.055583 for a person receiving the minimum amount of adjusted welfare income[1] and 0.798982 for a person receiving the median amount of adjusted welfare income[2]. This suggests a higher probability of good health for an extra person living in a household that receives a higher amount of welfare.

We also included adj_welfr_educ_rev in order to check whether the effect of adjusted welfare income varies with the number of years of education. For Model (8), the effect of an extra year of education on the probability a person has a good measure of health is 0.2283595 for a person receiving no adjusted welfare income[3] and 0.219158 for a person receiving the median

[1] $\partial \widehat{goodhealth}/\partial famsize \mid_{adj\_welfr = 0.715} = 0.055369 + 0.000299(0.715)$

[2] $\partial \widehat{goodhealth}/\partial famsize \mid_{adj\_welfr = 2487} = 0.055369 + 0.000299(2487)$

[3] $\partial \widehat{goodhealth}/\partial educ\_rev \mid_{adj\_welfr = 0} = 0.2283595 - 0.0128688(0)$

amount of adjusted welfare income[4]. This suggests a slightly higher probability of good health for a person with an extra year of education who is receiving less welfare, which may be explained by the idea that those with higher education who receive less welfare may also have higher paying jobs that enable them to have better health status.

Hypothesis Testing:

As seen in Table 6, our hypotheses for testing Model (2) and Model (8) are:

$$H_0: \beta_2 \leq 0$$
$$H_1: \beta_2 > 0$$

We created the null hypothesis using our past literature as a guide. Our referenced literature supports the idea that there is either a weak, insignificant relationship or positive relationship between welfare income and good measures of health (Miller & Morrisey, 2017, Gregory & Deb, 2014). Through our regression, we are hoping to test how weak or positive this relationship is. By testing our β_2 with these hypotheses, we will be able to conclude if our parameter of interest is statistically significant for this relationship or if the correlation is weak and statistically insignificant, as one of our referenced studies proposes. We chose to test these hypotheses on Models (2) and (8), as these models are the best representations from our simple and multiple regressions of the relationship between adjusted welfare income and the probability of a person having a good measure of health due to their relatively low SSE values.

Model (2) had a calculated t-statistic of 3.47 under the right-tailed test stated above[5]. This would equal a p-value of 0.0002. Under a t-distribution with 2,712 degrees of freedom and an alpha of 0.005, the critical t-value is equal to 2.576. Because the t-statistic is greater than the critical value and the p-value is less than alpha, we reject the null hypothesis, and conclude that the β_2 value is more than zero in the population at the 90%, 95%, 97.5% and 99.5% confidence levels. This would support past literature suggesting a positive relationship between welfare income and a good measure of self-reported health.

Model (8) had a calculated t-statistic of 0.65 under the right-tailed test stated above[6]. This would equal a p-value of 0.2877. Under a t-distribution with 2,712 degrees of freedom and an alpha of 0.005, the critical t-value is equal to 2.576. Because the t-statistic is less than the critical value and the p-value is more than alpha, we reject the null hypothesis, and conclude that the β_2 value is less than or equal to zero in the population at the 90%, 95%, 97.5% and 99.5% confidence levels. This would support the Miller & Morrisey study in 2017 suggesting a weak and insignificant relationship between welfare income and a good measure of self-reported health.

While the hypothesis test of Model (2) yielded statistically significant results, Model (8) is the best fitting model out of our simple and multiple regressions, as it controls for more endogeneity than the simple regression model and has the lowest SSE value out of our other

[4] $\partial \widehat{goodhealth}/\partial educ\_rev \mid_{adj\_welfr = 0.715} = 0.2283595 - 0.0128688(0.715)$

[5] t-stat = $(1.03 \times 10^{-9} - 0)/2.96 \times 10^{-10} = 3.74$

[6] t-stat = $(1.03 \times 10^{-9} - 0)/2.96 \times 10^{-10} = 3.74$

multiple regression models. Therefore, we accept the conclusions from the hypothesis test of Model (8), and conclude that the β_2 value is less than or equal to zero in the population at the 90%, 95%, 97.5% and 99.5% confidence levels, thus supporting the Miller & Morrisey study findings from 2017.

Section IV: Data and Descriptive Statistics

The data set includes eleven main variables, all pulled from IPUMS CPS data in yearly (ASEC) format. Data from the years 2010 and 2014 were chosen to fully showcase the expansion and cut, respectively, of welfare income in the years 2009 and 2013 as described in past literature. Our main dependent variable is named goodhealth, using cleaned data from the original IPUMS variable health. The original variable health included 5 codes, with 1 equalling poor health, 2 equalling fair health, 3 equalling good health, 4 equalling very good health, and 5 equalling excellent health. The variable health was then converted into a dummy variable, goodhealth, with 1 taking the value of good, very good, and excellent health, and 0 taking the value of fair and poor health. Because we use the dummy variable goodhealth as our dependent (y) variable, our model is considered a linear probability model. Our main independent variable is named adj_welfr, which uses cleaned data from the IPUMS variable incwelfr. The original variable incwelfr housed the amount of welfare income a person received during the specified year. Because the variable includes yearly income data across the years 2010 and 2014, it must be multiplied by the variable cpi99, which is the CPI value of 1999 to account for inflation. This variable was cleaned by dropping observations coded 999999, as they were recorded as NIU (not in universe). Because of our desire to show the relationship between receiving welfare income and measure of health, our data set was further cleaned to only include welfare income values greater than zero. This drop in observations also makes it more feasible for our models to be comparable across all functional forms. Our variable adj_welfr shows the adjusted welfare income data after these changes have been made. The variables adj_welfr_sq, adj_welfr_cu, and ad_welr_log are the squared, cubed and natural log values of the adjusted welfare variable, respectively.

The remaining nine main variables include age, sex_rev, married, famsize, mthwelfr, midwest, south, west, educ_rev as described in Table 2 of the Appendix. These variables were also cleaned and re-coded to the appropriate continuous and dummy variables. Our multiple regression models include these variables to combat endogeneity of adjusted welfare income. With these variables, age, sex, number of family members in the household, months received welfare income, region, and education are all controlled for. Our multiple regression model also includes the previously mentioned interaction terms between adjusted welfare and number of family members in the household and between adjusted welfare and number of years of education. While our regression models will not conclude causal relationships between welfare income and the probability of a good measure of health, these added variables will aid in our conclusions of a non-causal relationship.

As seen in Table 2, the number of observations for all variables was within the range 2,689 to 2,713. Noticeable means include the average adjusted welfare income of $2,487, average probability of obtaining a good measure of health of 0.703, average number of people living in a household of 3.633, and the average education level being around 12th grade. These are important measures to note, as they serve as baseline indicators for our data set and sample parameters.

As seen in Table 3, there is a positive difference between the health statuses of individuals receiving the top 50% of adjusted welfare income versus the individuals receiving the lower 50% of adjusted welfare income. People receiving the top 50% of welfare income have a mean probability of having a good measure of health that is 0.015 units higher than people who receive the lower 50% of adjusted welfare income. This would support some of the past findings that report a weak but positive relationship between adjusted welfare income and the probability of a good measure of health.

Section V: Estimation Results

Simple Regression (Table 4: Models 1, 2, 3, 4)

Linear regression (Model 1):
In Model 1, the b_1 value is equal to 0.8678851, which is the probability of a person having a good measure of health when the amount of welfare income equals 0 dollars. The b_2 value is 0.0000108, which means the probability of having a good measure of health increases by 0.0108 as the amount of welfare income increases by $1,000. The slope of this model and its interpretation is the same as the b_2 value[7]. The elasticity of this model is equal to 0.038207, which means as the adjusted welfare income increases by 1%, the probability of a good measure of health also increases by 0.038207%, evaluated at the mean[8]. The semi-elasticity of this model is equal to 0.001536, which means as the adjusted welfare income increases by $1, the probability of a good measure of health increases by 0.001536%, evaluated at the mean[9].

Quadratic regression (Model 2):
In Model 2, the b_1 value is equal to 0.8673395, which is the probability of a person having a good measure of health when the amount of welfare income equals 0 dollars. The b_2 value is 1.00×10^{-9}, which is how much the probability of having a good measure of health decreases as the amount of welfare income squared increases by $1. The slope of this model is 4.974×10^{-6} which means as adjusted welfare increases by $1, the probability of a person having a good measure of health increases by 4.974×10^{-6}, evaluated at the mean[10]. The elasticity of this model is 0.017596, which means as adjusted welfare income increases by 1%, the probability of a person having a good measure of health increases by 0.017596%, evaluated at the means[11]. The semi-elasticity of this model is 0.00070956, which means that as adjusted welfare income increases by $1, the probability of a person having a good measure of health increases by 0.0007056%, evaluated at the mean[12].

Cubic regression (Model 3):

---

[7]Slope = b_2 in linear regression model
[8]Elasticity = 0.0000108 • (2487/0.701)
[9]Semi-elasticity = 0.0000108 • (1/0.701) • 100
[10]Slope = 2 • (1.00×10^{-9}) • 2487
[11]Elasticity = 1.00×10^{-9} • 2487/0.701
[12]Semi-elasticity = 1.00×10^{-9} • (1/0.701) • 100

In Model 3, the b_1 value is equal to 0.8672415, which is the probability of a person having a good measure of health when the amount of welfare income equals 0 dollars. The b_2 value is 4.16 x 10^{-15}, which is how much the probability of having a good measure of health decreases as the amount of welfare income cubed increases by \$1. The slope of this model is 7.7 x 10^{-8}, which means that as adjusted welfare income increases by \$1, the probability of a person having a good measure of health increases by 7.7 x 10^{-8}, evaluated at the mean[13]. The elasticity of this model is 0.0002739, which means that as adjusted welfare income increases by 1%, the probability of a person having a good measure of health increases by 7.7 x 10^{-8}%, evaluated at the means[14]. The semi-elasticity of this model is 0.00001098, which means that as adjusted welfare income increases by \$1, the probability of a person having a good measure of health increases by 0.00001098%, evaluated at the mean[15].

Linear-log regression (Model 4):
In Model 4, the b_1 value is equal to 0.6034597, which is the probability of a person having a good measure of health when the amount of welfare income equals 1 dollars. The b_2 value is 0.0136383, which means that the probability of having a good measure of health increases by 0.000136383 as the amount of welfare income increases by 1%. The slope of this model is 5.484 x 10^{-6}, which means that as adjusted welfare income increases by \$1, the probability of a person having a good measure of health increases by 5.484 x 10^{-6}, evaluated at the mean[16]. The elasticity of this model is 0.019456, which means that as the adjusted welfare income increases by 1%, the probability of a person having a good measure of health increases by 0.019456%, evaluated at the means[17]. The semi-elasticity of this model is 0.0007823, which means that as adjusted welfare income increases by \$1, the probability of a person having a good measure of health increases by 0.0007823%, evaluated at the mean[18].

While the findings of our simple regression models show an increase in the probability of good health status with an increase in welfare income, we cannot conclude that this potential positive relationship points to a direct effect of increased welfare income on health status because endogeneity exists in this relationship. Our multiple regression models, as detailed below, attempt to account for some of this endogeneity by incorporating other x variables in our regression. For instance, one such x variable which makes our main x variable adjusted welfare income endogenous is education level (educ_rev), which is an omitted variable in our simple regression model that can be correlated with both adjusted welfare income and health status (when controlling for adjusted welfare income). Endogeneity in this case can be described by corr(educ_rev, adj_welf) > 0 (as education level increases, people may have higher paying jobs and qualify for less welfare) and corr(educ_rev, goodhealth) > 0, controlling for adjusted welfare income (as education level increases, people may be more educated on their health/healthy habits and thus maintain better health status).

Multiple Regression (Table 5: Models 5, 6, 7, and 8)

---

[13]Slope = 3 • (4.16 x 10^{-15}) • 2487^2 where $x = \bar{x}$
[14]Elasticity = 4.16 x 10^{-15} • 2487/0.701
[15]Semi-elasticity = 4.16 x 10^{-15} • (1/2487) • 100
[16]Slope = 0.0136383/2487
[17]Elasticity = 0.0136383 • 2487/0.701
[18]Semi-elasticity = 0.0136383 • (1/0.701) • 100

In Models 5, 6, 7 and 8, we attempt to evaluate the relationship between adjusted welfare income and the probability of a good measure of health while controlling for other variables that might impact the direct relationship. These variables include age, education level, adjusted welfare income, months received welfare income, and number of family members living in the household. The model also includes two interaction terms meant to evaluate the relationship between adjusted welfare income and number of people living in the household, and the relationship between adjusted welfare income and education level.

Linear Multiple Regression (Model 5):
In Model 5, the b_1 value is 0.8876198, which is the probability of an unmarried male living in the northeast having good health when all other variables (age, education level, adjusted welfare income, months received welfare income, number of family members living in household, the interaction term between adjusted welfare income and number of family members living in the household, and the interaction term between adjusted welfare income and education level) is equal to 0 units. For our model, this b_1 value isn't really applicable since the minimum age in our dataset is 15 years. The b_2 value is 0.000023, which means that the probability of a person having good health increases by 0.023 as adjusted welfare income increases by $1,000, ceteris paribus. The b_3 value is 0.013046, which means that the probability of a person having good health increases by 0.013046 as education level increases by 1 year, ceteris paribus. The b_4 value is 0.0505203, which means that the probability of a person having good health is 0.0505203 higher when a person is married versus unmarried, ceteris paribus. The b_5 value is 0.0405942, which means that the probability of a person having good health is 0.0405942 higher when a person lives in the west versus the non-west, ceteris paribus. The b_6 value is 0.0467196, which means that the probability of a person having good health is 0.0467196 higher when a person lives in the south versus the non-south, ceteris paribus. The b_7 value is -0.0068615, which means that the probability of a person having good health is 0.0068615 lower when a person lives in the midwest versus the non-midwest, ceteris paribus. The b_8 value is -0.0083918, which means that the probability of a person having good health decreases by 0.0083918 as months received welfare income increases by 1 month, ceteris paribus. The b_9 value is 0.0124796, which means that the probability of a person having good health increases by 0.0124796 as the number of family members in the household increases by 1 family member, ceteris paribus. The b_{10} value is 0.0250831, which means that the probability of a person having good health is 0.0250831 higher when the person is female versus male, ceteris paribus. The b_{11} value is -0.0108426, which means that the probability of a person having good health decreases by 0.0108426 as age increases 1 year, ceteris paribus. The b_{12} value is 8.55×10^{-7}, which means that the probability of a person having good health increases by 8.55×10^{-7} as the interaction term between adjusted welfare income and number of family members in the household increases by 1 unit, ceteris paribus. The b_{13} value is -1.06×10^{-6}, which means that the probability of a person having good health decreases by 1.06×10^{-6} as the interaction term between adjusted welfare income and education level increases by 1 unit, ceteris paribus.

Quadratic Multiple Regression (Model 6):
In Model 6, the b_1 value is 0.7654234, which is the probability of an unmarried male living in the northeast having good health when the squared value of all other variables (age, education level, adjusted welfare income, months received welfare income, number of family members living in

household, the interaction term between adjusted welfare income and number of family members living in the household, and the interaction term between adjusted welfare income and education level) is equal to 0 units. For our model, this b_1 value isn't really applicable since the minimum age in our dataset is 15 years. The b_2 value is 1.21 x 10^{-9}, which means that the probability of a person having good health increases by 1.21 x 10^{-9} as the square of adjusted welfare income increases by 1 unit, ceteris paribus. The b_3 value is 0.0004857, which means that the probability of a person having good health increases by 0.0004857 as the square of education level increases by 1 unit, ceteris paribus. The b_4 value is 0.0471518, which means that the probability of a person having good health is 0.0471518 higher when a person is married versus unmarried, ceteris paribus. The b_5 value is 0.0471518, which means that the probability of a person having good health increases by 0.0471518 when a person lives in the west versus the non-west, ceteris paribus. The b_6 value is 0.047741, which means that the probability of a person having good health increases by 0.047741 when a person lives in the south versus the non-south, ceteris paribus. The b_7 value is -0.0034416, which means that the probability of a person having good health decreases by 0.00034416 when a person lives in the midwest versus the non-midwest, ceteris paribus. The b_8 value is -0.000464, which means that the probability of a person having good health decreases by 0.000464 as the square of months received welfare income increases by 1 unit, ceteris paribus. The b_9 value is 0.0013032, which means that the probability of a person having good health increases by 0.0013032 as the square of the number of family members in the household increases by 1 unit, ceteris paribus. The b_{10} value is 0.029457, which means that the probability of a person having good health increases by 0.029457 when the person is female versus male, ceteris paribus. The b_{11} value is -0.0001175, which means that the probability of a person having good health decreases by 0.0001175 as the square of age increases 1 unit, ceteris paribus. The b_{12} value is -1.35 x 10^{-12}, which means that the probability of a person having good health decreases by 1.35 x 10^{-12} as the squared interaction term between adjusted welfare income and number of family members in the household increases by 1 unit, ceteris paribus. The b_{13} value is -1.06 x 10^{-12}, which means that the probability of a person having good health decreases by 1.06 x 10^{-12} as the squared interaction term between adjusted welfare income and education level increases by 1 unit, ceteris paribus.

Cubic Multiple Regression (Model 7):

In Model 7, the b_1 value is 0.7190435, which is the probability of a non-married male from the northeast having a good measure of health when the cubed values of all other variables (age, education level, adjusted welfare income, months received welfare income, number of family members living in household, the interaction term between adjusted welfare income and number of family members living in the household, and the interaction term between adjusted welfare income and education level) is equal to 0 units. For our model, this b_1 value isn't really applicable since the minimum age in our model is 15 years. The b_2 value is 4.57×10^{-14}, which means the probability of having a good measure of health increases by 4.57×10^{-14} as the cube of adjusted welfare increases by 1 unit, ceteris paribus. The b_3 value is 0.0000229, which means the probability of having a good measure of health increases by 0.0000229 as the cube of education level increases by 1 unit, ceteris paribus. The b_4 value is 0.0349016, which means that the probability of having a good measure of health is 0.0349016 higher for a person who is married relative to non-married, ceteris paribus. The b_5 value is 0.0552447, which means

the probability of having a good measure of health is 0.0552447 higher for a person from the west relative to someone not from the west, ceteris paribus. The b_6 = 0.0508043 means that the probability of having a good measure of health is 0.0508043 higher for a person from the south relative to someone not from the south, ceteris paribus. The b_7 value is -0.0005474, which means that the probability of having a good measure of health is 0.0005474 lower for a person from the midwest relative to someone not from the midwest, ceteris paribus. The b_8 value is -0.0000355, which means that the probability of having a good measure of health decreases by 0.0000355 as the cube of months received welfare income increases by 1 unit, ceteris paribus.
The b_9 value is 0.0001051, which means the probability of having a good measure of health increases by 0.0001051 when the cube of the number of family members increases by 1 unit, ceteris paribus. The b_{10} value is 0.0392684, which means that the probability of having a good measure of health is 0.0392684 higher for females relative to males, ceteris paribus.The b_{11} value is -1.40×10^{-6}, which means the probability of having a good measure of health decreases by 1.40×10^{-6} when the cube of age increases by 1 unit, ceteris paribus. The b_{12} value is -2.44×10^{-17}, which means that the probability of having a good measure of health decreases by -2.44×10^{-17} as the cubed interaction term between adjusted welfare income and number of family members in the household increases by 1 unit, ceteris paribus. The b_{13} value is 5.42×10^{-18}, which means the probability of having a good measure of health increases by 5.42×10^{-18} as the cubed interaction term between adjusted welfare income and education level increases by 1 unit, ceteris paribus.

Linear-log Multiple Regression (Model 8):
In Model 8, the value of b_1 = 1.396314, which is the probability of an unmarried male from the northeast having a good measure of health when when the values of all other variables (age, education level, adjusted welfare income, months received welfare income, number of family members living in household, the interaction term between adjusted welfare income and number of family members living in the household, and the interaction term between adjusted welfare income and education level) is equal to 1 unit. For our model, this b_1 value isn't really applicable since the minimum age in our model is 15 years. The b_2 value is 0.0530829, which means that when adjusted welfare income increases by 1%, the probability of having a good measure of health increases by $\frac{0.0530829}{100} = 0.000530829$, ceteris paribus. The b_3 value is 0.2283595, which means that when years of education increases by 1%, the probability of having a good measure of health increases by $\frac{0.2283595}{100} = 0.002283595$, ceteris paribus. The b_4 value is 0.057915, which means that the probability of having a good measure of health is 0.057915 higher for a person who is married relative to non-married, ceteris paribus. The b_5 value is 0.0419862, which means that the probability of having a good measure of health is 0.0419862 higher for a person from the west relative to someone not from the west, ceteris paribus. The b_6 value is 0.0413366, which means that the probability of having a good measure of health is 0.0413366 higher for a person from the south relative to someone not from the south, ceteris paribus. The b_7 value is -0.0121155, which means that the probability of having a good measure of health is 0.0121155 lower for a person from the midwest relative to someone not from the midwest, ceteris paribus. The b_8 value is -0.0464698, which means that when number of months receiving welfare increases by 1%, the probability of having a good measure of health decreases by $\frac{0.0464698}{100} =$

0.000464698 , ceteris paribus. The b_9 value is 0.05879, which means that when number of family members in household increases by 1%, the probability of having a good measure of health increases by $\frac{0.05879}{100} =$ 0.0005879, ceteris paribus. The b_{10} value is 0.0315911, which means that the probability of having a good measure of health is 0.0315911 higher for females relative to males, ceteris paribus. The b_{11} value is -0.4086438, which means that when age increases by 1%, the probability of having a good measure of health decreases by $\frac{0.4086438}{100}$ = 0.004086438 , ceteris paribus.The b_{12} value is -0.0001571, which means that when the interaction term between adjusted welfare income and number of family members in the household increases by 1%, the probability of having a good measure of health decreases by $\frac{0.0001571}{100}$ = 0.000001571, ceteris paribus. The b_{13} value is -0.0128688, which means that when the interaction term between adjusted welfare income and education level increases by 1%, the probability of having a good measure of health decreases by $\frac{0.0128688}{100}$ = 0.000128688, ceteris paribus.

With these interpretations, Model 8 supports past literature in describing a weak and insignificant relationship between welfare income and measures of good health. With all other variables included to decrease endogeneity, (age, education level, regions, sex, marrital status, adjusted welfare income, months received welfare income, number of family members living in household, the interaction term between adjusted welfare income and number of family members living in the household, and the interaction term between adjusted welfare income and education level equal) the b_2 value shows an increase of 0.023 in probability of good health as welfare income increases by $1,000. This shows a weak relationship between the amount of welfare income individuals receive and a measure of good health. With the right-tailed hypothesis test results from Table 6, this model supports the finding that the relationship between welfare income and good measures of self-reported health status is weak and statistically insignificant, as proposed in the study conducted by Miller & Morrisey in 2017.

Section VI: Conclusions

After studying the relationship between welfare income and health status in the calendar years 2010 and 2014, we found that the best regression model for the relationship between good self-reported measures of health and adjusted welfare income is the multiple linear-log regression, or Model 8. This is because this model controls for more endogeneity in comparison to the simple regression models (Models 1, 2, 3, and 4), and has the lowest SSE value out of our other multiple regression models (Models 5, 6, and 7). In using Model 8, we were able to find the b_1 and b_2 values that minimize the sum of squared errors. The slightly positive b_2 value of Model 8 shows the most accurate relationship we found between welfare income and health status: there is a weak but positive relationship between the amount of welfare income and self-reported health status in this sample data. However, as seen through the hypothesis testing in Table 6, the relationship is statistically insignificant in the population, a finding supported in Miller & Morrisey's 2017 study.

The use of multiple regression in Model 8 is to control for endogeneity. For instance, our simple quadratic regression model, Model 2, yielded a statistically significant b_2 value, as seen in Table 6. This model has the lowest SSE value of the simple regression models, so it was used in our hypothesis testing. This would conclude that the relationship between adjusted welfare income and good measures of health would be positive in the population at the 90%, 95%, 97.5%, and 99.5% confidence levels. It is not until after adding variables that explain marital status, regional differences, sex, age, number of people living in the household, education level, and months received welfare does a better fitting model with a higher control for endogeneity come to light, as seen in Model 8. With the addition of these variables, the total error in the model is reduced, as some omitted variables that could be influencing the simple regression models are controlled for. Although Model 8 yields a statistically insignificant population value of a positive β_2 in contrary to the hypothesis testing results of Model 2, it is a more accurate representation of the true relationship between welfare income and the probability of a good measure of health, as factors that influence the adjusted welfare income are controlled for and the threat of endogeneity is decreased. It is worth noting that although endogeneity is decreased with the addition of independent variables in Model 8, error due to endogeneity is never zero. This is why the observed relationship between adjusted welfare income and the probability of a good measure of health in this paper is not a causal relationship.

Multicollinearity between the variables age and education level is worth mentioning. There is likely to be a positive correlation between age and education level, as the older a person is the more likely they have a higher level of education. As observed in Table 2, the average age of our data set is 36.40 years old, and the average education level is around 12th grade. With a younger data set, we could expect to see a lower level of education. A multicollinearity between education level and age could reduce the statistical significance of an independent variable, and is thus a limitation in our study.

The findings in our paper can be extended to welfare and SNAP policy decisions in the future. Although SNAP and welfare benefits increase the amount of disposable income to recipients, our study finds that this increase in income does not have a significant or positive effect on the health status of recipients. While welfare payments are targeted in decreasing food insecurity and are thus an extension of improving health, our study finds they do not truly achieve this. This could be due to the fact that the welfare income is not substantial enough to cover the costs of nutritious foods, or the recipients prioritize other expenses over health related ones. Alternatively, other studies with more variables, data, or econometric analysis could disprove our findings. Either way, more research into why our findings show little effect on health status could help policymakers better formulate SNAP and other welfare programs, allowing researchers to explore more effective methods for improving health status for individuals.

Section VII: References

Flood, Sarah, et al. Integrated Public Use Microdata Series, Current Population Survey: Version 8.0. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D030.V8.0

Gregory, Christian A., and Partha Deb. "Does Snap Improve Your Health?" *Food Policy*, vol. 50, 11 Nov. 2015, pp. 11–19., https://doi.org/10.1016/j.foodpol.2014.09.010.

Katare, Bhagyashree, and Jiyoon Kim. "Applied Economic Perspectives and Policy." *Effects of the 2013 SNAP Benefit Cut on Food Security*, vol. 39, no. 4, 25 Apr. 2017, pp. 662–681., https://doi.org/10.1093/aepp/ppx025.

Miller, Daniel P, and Tayrn Morrissey. "Using Natural Experiments to Identify the Effects of SNAP on Child and Adult Health." *University of Kentucky Center for Poverty Research*, 31 Jan. 2017.

Narain, Kimberly, et al. "The Impact of Welfare Reform on the Health Insurance Coverage, Utilization and Health of Low Education Single Mothers." *Social Science & Medicine*, vol. 180, 16 Mar. 2017, pp. 28–35., https://doi.org/10.1016/j.socscimed.2017.03.021.

Zhang, Jun, et al. "International Journal of Environmental Research and Public Health." *Does Supplemental Nutrition Assistance Program Reduce Food Insecurity among Households with Children?*, vol. 18, no. 6, 19 Mar. 2021, p. 3178., https://doi.org/10.3390/ijerph.

Section VIII: Appendix

Table 1: Description of Variables

Variable Name	**Description**
incwelfr	Welfare (public assistance) income
health	Health status
goodhealth	Dummy variable which = 1 if good health status (excellent, very good, good) and = 0 if not good health status (fair, poor)
adj_welfr	Adjusted welfare income in 1999 dollars
adj_welfr_sq	Adjusted welfare income in 1999 dollars squared
adj_welfr_cu	Adjusted welfare income in 1999 dollars cubed

adj_welfr_log	Natural log of adjusted welfare income in 1999 dollars
educ_rev	Education level
educ_rev_sq	Education level squared
educ_rev_cu	Education level cubed
educ_rev_log	Natural log of education level
married	Dummy variable where = 1 if person is married and = 0 if person is not married
west	Dummy variable where = 1 if living in west region and = 0 if living in non-west region
south	Dummy variable where = 1 if living in south region and = 0 if living in non-south region
midwest	Dummy variable where = 1 if living in midwest region and = 0 if living in non-midwest region
mthwelfr	Number of months received welfare income
mthwelfr_sq	Number of months received welfare income squared
mthwelfr_cu	Number of months received welfare income cubed
mthwelfr_log	Natural log of number of months received welfare income
famsize	Number of family members in household
famsize_sq	Number of family members in household squared
famsize_cu	Number of family members in household cubed
famsize_log	Natural log of number of family members in household
sex	Dummy variable where = 1 if female and = 0 for male

age	Age of person
age_sq	Age of person squared
age_cu	Age of person cubed
age_log	Natural log of age of person
adj_welfr_famsize	Interaction term between adjusted welfare income and number of family members in home
adj_welfr_famsize_sq	Interaction term between adjusted welfare income and number of family members in home squared
adj_welfr_famsize_cu	Interaction term between adjusted welfare income and number of family members in home cubed
adj_welfr_famsize_log	Natural log of interaction term between adjusted welfare income and number of family members in home
adj_welfr_educ_rev	Interaction term between adjusted welfare income and education level
adj_welfr_educ_rev_sq	Interaction term between adjusted welfare income and education level squared
adj_welfr_educ_rev_cu	Interaction term between adjusted welfare income and education level cubed
adj_welfr_educ_rev_log	Natural log of interaction term between adjusted welfare and education level
fiftypercent_adj_welfr	Dummy variable where 1 = top 50% of adjusted welfare income received and 0 = bottom 50% of adjusted welfare income received

Table 2: Descriptive Statistics of the Variables

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max

age	2,713	36.40	13.80	15	85
famsize	2,713	3.633	1.813	1	14
incwelfr	2,713	3,327	3,116	1	25,000
mthwelfr	2,713	9.264	3.859	1	12
health	2,713	2.874	1.186	1	5
adj_welfr	2,713	2,487	2,339	0.715	19,425
goodhealth	2,713	0.703	0.457	0	1
midwest	2,713	0.203	0.402	0	1
south	2,713	0.229	0.420	0	1
west	2,713	0.332	0.471	0	1
sex_rev	2,713	0.818	0.386	0	1
married	2,713	0.244	0.430	0	1
educ_rev	2,713	11.65	2.421	0	21
adj_welfr_famsize	2,713	9,590	11,758	0.715	155,400
adj_welfr_educ_rev	2,713	28,965	29,159	0	252,525
adj_welfr_sq	2,713	1.165e+07	2.957e+07	0.511	3.773e+08
adj_welfr_cu	2,713	8.893e+10	4.855e+11	0.366	7.330e+12
adj_welfr_log	2,713	7.319	1.238	-0.335	9.874
age_sq	2,713	1,515	1,194	225	7,225
famsize_sq	2,713	16.48	17.30	1	196
educ_rev_sq	2,713	141.6	50.27	0	441
mthwelfr_sq	2,713	100.7	58.18	1	144
adj_welfr_famsize_sq	2,713	2.302e+08	8.789e+08	0.511	2.415e+10

adj_welfr_educ_rev_sq	2,713	1.689e+09	4.731e+09	0	6.377e+10
age_cu	2,713	71,247	88,575	3,375	614,125
famsize_cu	2,713	89.51	166.9	1	2,744
educ_rev_cu	2,713	1,767	914.7	0	9,261
mthwelfr_cu	2,713	1,155	753.9	1	1,728
adj_welfr_famsize_cu	2,713	1.078e+13	1.015e+14	0.366	3.753e+15
adj_welfr_educ_rev_cu	2,713	1.684e+14	9.905e+14	0	1.610e+16
age_log	2,713	3.526	0.371	2.708	4.443
famsize_log	2,713	1.156	0.545	0	2.639
educ_rev_log	2,689	2.443	0.218	1.386	3.045
mthwelfr_log	2,713	2.066	0.672	0	2.485
adj_welfr_famsize_log	2,713	8.557	4.387	-0.653	21.64
adj_welfr_educ_rev_log	2,689	17.87	3.431	-0.930	26.58

Table 3: Descriptive statistics according to median adjusted welfare income

	(1)	(2)	(3)	(4)	(5)
	fiftypercent_adj_welfr: 0		fiftypercent_adj_welfr: 1		
VARIABLES	N	mean	N	mean	difference in means
age	1,618	36.72	1,095	35.92	-0.8
famsize	1,618	3.466	1,095	3.879	0.413

incwelfr	1,618	1,467	1,095	6,076	4609
mthwelfr	1,618	7.868	1,095	11.33	3.462
health	1,618	2.880	1,095	2.864	-0.016
adj_welfr	1,618	1,097	1,095	4,540	3443
goodhealth	1,618	0.697	1,095	0.712	0.015
midwest	1,618	0.231	1,095	0.161	-0.07
south	1,618	0.282	1,095	0.151	-0.131
west	1,618	0.258	1,095	0.442	0.184
sex_rev	1,618	0.792	1,095	0.857	0.065
married	1,618	0.229	1,095	0.266	0.037
educ_rev	1,618	11.68	1,095	11.61	-0.07
adj_welfr_famsize	1,618	3,894	1,095	18,008	14114
adj_welfr_educ_rev	1,618	12,721	1,095	52,967	40246
adj_welfr_sq	1,618	1.689e+06	1,095	2.637e+07	24681000
adj_welfr_cu	1,618	2.998e+09	1,095	2.159e+11	2.12902E+11
adj_welfr_log	1,618	6.634	1,095	8.332	1.698
age_sq	1,618	1,555	1,095	1,455	-100
famsize_sq	1,618	15.30	1,095	18.22	2.92
educ_rev_sq	1,618	141.9	1,095	141.2	-0.7
mthwelfr_sq	1,618	79.70	1,095	131.7	52

adj_welfr_famsize_sq	1,618	2.737e+07	1,095	5.298e+08	502430000
adj_welfr_educ_rev_sq	1,618	2.365e+08	1,095	3.835e+09	3598500000
age_cu	1,618	74,820	1,095	65,968	-8852
famsize_cu	1,618	82.29	1,095	100.2	17.91
educ_rev_cu	1,618	1,771	1,095	1,761	-10
mthwelfr_cu	1,618	886.6	1,095	1,551	664.4
adj_welfr_famsize_cu	1,618	2.720e+11	1,095	2.630e+13	2.6028E+13
adj_welfr_educ_rev_cu	1,618	5.218e+12	1,095	4.096e+14	4.04382E+14
age_log	1,618	3.529	1,095	3.520	-0.009
famsize_log	1,618	1.098	1,095	1.240	0.142
educ_rev_log	1,609	2.442	1,080	2.446	0.004
mthwelfr_log	1,618	1.837	1,095	2.403	0.566
adj_welfr_famsize_log	1,618	7.344	1,095	10.35	3.006
adj_welfr_educ_rev_log	1,609	16.19	1,080	20.38	4.19

Table 4: Simple Regression Models

	(1)	(2)	(3)	(4)
VARIABLES	Model 1	Model 2	Model 3	Model 4

adj_welfr	1.08e-05***			
	(3.75e-06)			
adj_welfr_sq		1.03e-09***		
		(2.96e-10)		
adj_welfr_cu			5.76e-14***	
			(1.80e-14)	
adj_welfr_log				0.0136*
				(0.00708)
Constant	0.676***	0.691***	0.698***	0.603***
	(0.0128)	(0.00941)	(0.00890)	(0.0526)
Observations	2,713	2,713	2,713	2,713
R-squared	0.003	0.004	0.004	0.001

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Multiple Regression Models

	(1)	(2)	(3)	(4)
VARIABLES	Model 5	Model 6	Model 7	Model 8
adj_welfr	2.30e-05			
	(2.06e-05)			
educ_rev	0.0130**			
	(0.00527)			

married	0.0505**	0.0424**	0.0349*	0.0579***
	(0.0203)	(0.0201)	(0.0200)	(0.0207)
west	0.0406*	0.0472**	0.0552**	0.0420*
	(0.0222)	(0.0224)	(0.0227)	(0.0223)
south	0.0467*	0.0477**	0.0508**	0.0413*
	(0.0241)	(0.0242)	(0.0245)	(0.0240)
midwest	-0.00686	-0.00344	-0.000547	-0.0121
	(0.0248)	(0.0250)	(0.0253)	(0.0248)
mthwelfr	-0.00839***			
	(0.00238)			
famsize	0.0125*			
	(0.00687)			
sex_rev	0.0251	0.0295	0.0393*	0.0316
	(0.0216)	(0.0217)	(0.0220)	(0.0219)
age	-0.0108***			
	(0.000634)			
adj_welfr_famsize	8.55e-07			
	(1.89e-06)			
adj_welfr_educ_rev	-1.05e-06			
	(1.53e-06)			
adj_welfr_sq		1.21e-09		
		(1.19e-09)		
educ_rev_sq		0.000486***		
		(0.000183)		

mthwelfr_sq	-0.000464***	
	(0.000148)	
famsize_sq	0.00130**	
	(0.000538)	
age_sq	-0.000117***	
	(7.27e-06)	
adj_welfr_famsize_sq	-0	
	(0)	
adj_welfr_educ_rev_sq	-0	
	(0)	
adj_welfr_cu		0
		(0)
educ_rev_cu		2.29e-05**
		(9.42e-06)
mthwelfr_cu		-3.55e-05***
		(1.14e-05)
famsize_cu		0.000105**
		(5.21e-05)
age_cu		-1.40e-06***
		(9.80e-08)
adj_welfr_famsize_cu		-0
		(0)
adj_welfr_educ_rev_cu		0
		(0)

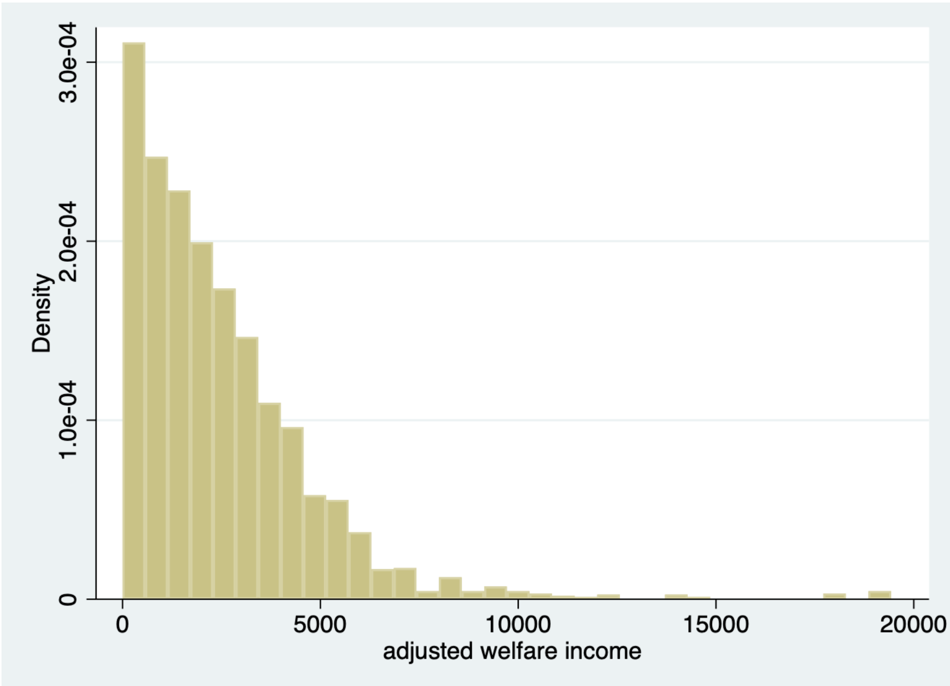
adj_welfr_log				0.0531
				(0.0815)
educ_rev_log				0.228
				(0.241)
mthwelfr_log				-0.0465***
				(0.0140)
famsize_log				0.0588
				(0.0798)
age_log				-0.409***
				(0.0238)
adj_welfr_famsize_log				-0.000157
				(0.0109)
adj_welfr_educ_rev_log				-0.0129
				(0.0324)
Constant	0.888***	0.765***	0.719***	1.396**
	(0.0796)	(0.0418)	(0.0342)	(0.611)
Observations	2,713	2,713	2,713	2,689
R-squared	0.146	0.128	0.103	0.150

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

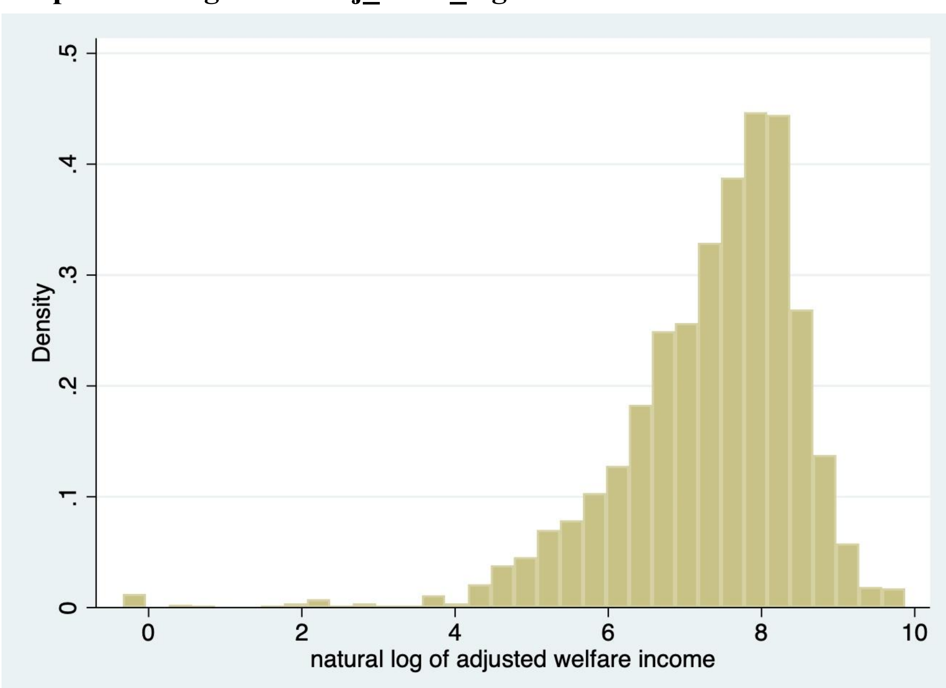
Table 6: Hypothesis Testing

Hypotheses	**Model 2**	**Model 8**
$H_0: \beta_2 \leq 0$ $H_1: \beta_2 > 0$	0.002***	0.2877

Graph 1: Histogram of adj_welfr



Graph 2: Histogram of adj_welfr_log



Do-File:

```
clear all

use "/Users/mariahhastings/Documents/Stata/Empirical Paper/cps_00013.dta"
*data extract is from IPUMS CPS with samples from 2010 and 2014 and variables incwelfr
(income from welfare assistance), cpi99 (cpi value that incwelfr must be multiplied by to control
for inflation), health (health status), region, age, sex, marst (marital status), famsize (number of
own family members living in household), educ (education level), and mthwelfr (number of
months received welfare)*

log using "/Users/mariahhastings/Documents/Stata/Empirical Paper/Multiple Regression
Log.smcl", replace

*to get rid of other variables given automatically in data set*
drop year serial month cpsid asecflag asecwth pernum cpsidp asecwt hflag
br incwelfr health region sex age marst famsize educ mthwelfr

*cleaning data*
replace incwelfr = . if incwelfr == 999999
drop if incwelfr == .
drop if incwelfr == 0

cap gen adj_welfr = cpi99*incwelfr
label variable adj_welfr "adjusted welfare income"

tab health
tab health, nolabel
cap gen goodhealth = 0
replace goodhealth = 1 if health == 1 | health == 2 | health == 3
tab goodhealth
label variable goodhealth "good health status"

tab region, nolabel
tab region
cap gen midwest = 0
replace midwest = 1 if region >= 21 & region <= 23
cap gen south = 0
replace south = 1 if region >= 31 & region <= 34
cap gen west = 0
```

```
replace west = 1 if region >= 41 & region <= 43
tab midwest
tab south
tab west
label variable midwest "living in midwest"
label variable south "living in south"
label variable west "living in west"

tab sex
tab sex, nolabel
cap gen sex_rev = 0
replace sex_rev = 1 if sex == 2
label variable sex_rev "sex: 1 = female, 0 = male"

tab age

tab marst
tab marst, nolabel
cap gen married = 0
replace married = 1 if marst >= 1 & marst <= 2
replace married = . if marst == .
tab married
label variable married "marital status: married or unmarried"

tab famsize

tab educ
tab educ, nolabel
replace educ = . if educ == 1
drop if educ == .
cap gen educ_rev = educ
replace educ_rev = 0 if educ == 2
replace educ_rev = 4 if educ == 10
replace educ_rev = 6 if educ == 20
replace educ_rev = 8 if educ == 30
replace educ_rev = 9 if educ == 40
replace educ_rev = 10 if educ == 50
replace educ_rev = 11 if educ == 60 | educ == 71
replace educ_rev = 12 if educ == 73
replace educ_rev = 13 if educ == 81
```

```
replace educ_rev = 14 if educ == 91 | educ == 92
replace educ_rev = 16 if educ == 111
replace educ_rev = 18 if educ == 123
replace educ_rev = 21 if educ == 124 | educ == 125
label variable educ_rev "revised education level"

tab mthwelfr

*interaction terms*
cap gen adj_welfr_famsize = adj_welfr*famsize
label variable adj_welfr_famsize "interaction term of adjusted welfare income and number of family members in home"

cap gen adj_welfr_educ_rev = adj_welfr*educ_rev
label variable adj_welfr_educ_rev "interaction term of adjusted welfare income and education level"

*linear simple regression*
reg goodhealth adj_welfr
outreg2 using myreg.doc, replace ctitle(Model 1)

*quadratic simple regression*
cap gen adj_welfr_sq = adj_welfr*adj_welfr
label variable adj_welfr_sq "adjusted welfare income squared"

reg goodhealth adj_welfr_sq
outreg2 using myreg.doc, append ctitle(Model 2)

*cubic simple regression*
cap gen adj_welfr_cu = adj_welfr*adj_welfr*adj_welfr
label variable adj_welfr_cu "adjusted welfare income cubed"

reg goodhealth adj_welfr_cu
outreg2 using myreg.doc, append ctitle(Model 3)

*lin-log simple regression*
cap gen adj_welfr_log = ln(adj_welfr)
label variable adj_welfr_log "natural log of adjusted welfare income"
hist adj_welfr
hist adj_welfr_log
```

```
reg goodhealth adj_welfr_log
outreg2 using myreg.doc, append ctitle(Model 4)

*linear multiple regression*
reg goodhealth adj_welfr educ_rev married west south midwest mthwelfr famsize sex_rev age
adj_welfr_famsize adj_welfr_educ_rev
outreg2 using myreg2.doc, replace ctitle(Model 5)

*quadratic multiple regression*
cap gen age_sq = age*age
label variable age_sq "age squared"
cap gen famsize_sq = famsize*famsize
label variable famsize_sq "number of family members in household squared"
cap gen educ_rev_sq = educ_rev*educ_rev
label variable educ_rev_sq "education level squared"
cap gen mthwelfr_sq = mthwelfr*mthwelfr
label variable mthwelfr_sq "number of months received welfare income squared"
cap gen adj_welfr_famsize_sq = adj_welfr_famsize*adj_welfr_famsize
label variable adj_welfr_famsize_sq "interaction term of adjusted welfare income and number of
family members in home squared"
cap gen adj_welfr_educ_rev_sq = adj_welfr_educ_rev*adj_welfr_educ_rev
label variable adj_welfr_educ_rev_sq "interaction term of adjusted welfare income and
education level squared"

reg goodhealth adj_welfr_sq educ_rev_sq married west south midwest mthwelfr_sq famsize_sq
sex_rev age_sq adj_welfr_famsize_sq adj_welfr_educ_rev_sq
outreg2 using myreg2.doc, append ctitle(Model 6)

*cubic multiple regression*
cap gen age_cu = age*age*age
label variable age_cu "age cubed"
cap gen famsize_cu = famsize*famsize*famsize
label variable famsize_cu "number of family members in household cubed"
cap gen educ_rev_cu = educ_rev*educ_rev*educ_rev
label variable educ_rev_cu "education level cubed"
cap gen mthwelfr_cu = mthwelfr*mthwelfr*mthwelfr
label variable mthwelfr_cu "number of months received welfare income cubed"
cap gen adj_welfr_famsize_cu = adj_welfr_famsize*adj_welfr_famsize*adj_welfr_famsize
```

label variable adj_welfr_famsize_cu "interaction term of adjusted welfare income and number of family members in home cubed"
cap gen adj_welfr_educ_rev_cu = adj_welfr_educ_rev*adj_welfr_educ_rev*adj_welfr_educ_rev
label variable adj_welfr_educ_rev_cu "interaction term of adjusted welfare income and education level cubed"

reg goodhealth adj_welfr_cu educ_rev_cu married west south midwest mthwelfr_cu famsize_cu sex_rev age_cu adj_welfr_famsize_cu adj_welfr_educ_rev_cu
outreg2 using myreg2.doc, append ctitle(Model 7)

lin-log multiple regression
gen age_log = ln(age)
label variable age_log "natural log of age"
gen famsize_log = ln(famsize)
label variable famsize_log "natural log of number of family members in household"
gen educ_rev_log = ln(educ_rev)
label variable educ_rev_log "natural log of education level"
gen mthwelfr_log = ln(mthwelfr)
label variable mthwelfr_log "natural log of months received welfare income"
cap gen adj_welfr_famsize_log = ln(adj_welfr)*ln(famsize)
label variable adj_welfr_famsize_log "natural log of interaction term of adjusted welfare income and number of family members in home"
cap gen adj_welfr_educ_rev_log = ln(adj_welfr)*ln(educ_rev)
label variable adj_welfr_educ_rev_log "natural log of interaction term of adjusted welfare income and education level"

reg goodhealth adj_welfr_log educ_rev_log married west south midwest mthwelfr_log famsize_log sex_rev age_log adj_welfr_famsize_log adj_welfr_educ_rev_log
outreg2 using myreg2.doc, append ctitle(Model 8)

descriptive statistics
sum incwelfr health goodhealth adj_welfr adj_welfr_sq adj_welfr_cu adj_welfr_log educ_rev educ_rev_sq educ_rev_cu educ_rev_log married west south midwest mthwelfr mthwelfr_sq mthwelfr_cu mthwelfr_log famsize famsize_sq famsize_cu famsize_log sex_rev age age_sq age_cu age_log adj_welfr_famsize adj_welfr_famsize_sq adj_welfr_famsize_cu adj_welfr_famsize_log adj_welfr_educ_rev adj_welfr_educ_rev_sq adj_welfr_educ_rev_cu adj_welfr_educ_rev_log

outreg2 using summary.doc, replace sum(log) keep(incwelfr health goodhealth adj_welfr adj_welfr_sq adj_welfr_cu adj_welfr_log educ_rev educ_rev_sq educ_rev_cu educ_rev_log

```
married west south midwest mthwelfr mthwelfr_sq mthwelfr_cu mthwelfr_log famsize
famsize_sq famsize_cu famsize_log sex_rev age age_sq age_cu age_log adj_welfr_famsize
adj_welfr_famsize_sq adj_welfr_famsize_cu adj_welfr_famsize_log adj_welfr_educ_rev
adj_welfr_educ_rev_sq adj_welfr_educ_rev_cu adj_welfr_educ_rev_log)

*table 3*
cap gen fiftypercent_adj_welfr = 0
replace fiftypercent_adj_welfr = 1 if adj_welfr >= 2486.58
label variable fiftypercent_adj_welfr "top fifty percent of adjusted welfare income"

bys fiftypercent_adj_welfr: sum incwelfr health goodhealth adj_welfr adj_welfr_sq adj_welfr_cu
adj_welfr_log educ_rev educ_rev_sq educ_rev_cu educ_rev_log married west south midwest
mthwelfr mthwelfr_sq mthwelfr_cu mthwelfr_log famsize famsize_sq famsize_cu famsize_log
sex_rev age age_sq age_cu age_log adj_welfr_famsize adj_welfr_famsize_sq
adj_welfr_famsize_cu adj_welfr_famsize_log adj_welfr_educ_rev adj_welfr_educ_rev_sq
adj_welfr_educ_rev_cu adj_welfr_educ_rev_log

bys fiftypercent_adj_welfr: outreg2 using descriptivestats.doc, replace sum(log) eqkeep(N mean)
keep(incwelfr health goodhealth adj_welfr adj_welfr_sq adj_welfr_cu adj_welfr_log educ_rev
educ_rev_sq educ_rev_cu educ_rev_log married west south midwest mthwelfr mthwelfr_sq
mthwelfr_cu mthwelfr_log famsize famsize_sq famsize_cu famsize_log sex_rev age age_sq
age_cu age_log adj_welfr_famsize adj_welfr_famsize_sq adj_welfr_famsize_cu
adj_welfr_famsize_log adj_welfr_educ_rev adj_welfr_educ_rev_sq adj_welfr_educ_rev_cu
adj_welfr_educ_rev_log)

save "/Users/mariahhastings/Documents/Stata/Empirical Paper/Multiple Regression Cleaned
Data.dta", replace
log close
```