Effect of Wage on Household Size

Georgia Institute of Technology

ECON 3161 Econometrics

Fall 2021

Prof. Dr. Shatakshee Dhongde

Miah Schultz

Abstract

In this analysis, the relationship between household size and wage is investigated. The measure of household size that is used is the number of dependents, because of the economic relationship associated with wage and number of dependents. In the United States, there is an increasing opportunity cost associated with an increasing number of dependents. Time must be allocated between household production, which increases with increasing dependents, and hours worked in the labor force. This analysis attempts to derive the relationship between wage and number of dependents, investigating the proposed negative relationship between income level and household size.

I. Introduction

Understanding the relationship between household size and wage is important in understanding how individuals in households allocate time between working in the labor force and working in the household. Time spent working in the household is often not included in many indicators of economic size and performance. For example, gross domestic product (GDP) measures aggregate output in a country in a given time period. GDP can be measured by adding up all of the final goods and services produced, adding up all of the incomes in the economy, or adding the value added in the economy in a given period. These three different measures of GDP are all equivalent and all exclude some indication of household production. Because household production is not included in GDP, it is important to have some other indication of household production in an economy. If a relationship can be determined between wage and household size, then this could be a possible indication of time spent toward household production.

This analysis tests the relationship between the number of dependents and wage. There is an increasing opportunity cost associated with an increasing number of dependents in a household. Paying for basic needs for a dependent, such as food and housing, and more time required to be spent in household production associated with an increasing number of dependents are such contributors to the increasing opportunity cost. Time spent in household production could instead be time spent earning income. For this reason, wage could be an indicator of household size. People with higher wages do not want to give up the time associated with a higher number of dependents, therefore resulting in a lower number of dependents. This analysis attempts to derive a negative relationship between the number of dependents and wage, with increasing wage resulting in a decreasing number of dependents.

II. Literature Review

One approach to determining the relationship between wage and household size is to observe the economic reasoning behind individuals choosing to have children. In Cremer, et al.'s (2003) analysis, they derived a relationship between tax breaks and number of dependents. Cremer, et al. (2003) observed that families with low incomes can stay above the poverty line due to tax breaks received from having an increasing number of children. In this case, tax breaks serve as an incentive for low income families to have a larger number of dependents in order to stay out of poverty. This incentive is not as strong for high income families because they do not need the tax breaks from having children to stay out of poverty like the low income families. This analysis adds another angle to observe why there is potentially a negative relationship between wage and number of dependents. Cremer, et al. (2003) derived a negative relationship between income and number of dependents due to the incentive of the tax break for low income families, which provides support for the economic reasoning behind the relationship that will be investigated in this analysis.

Another approach to deriving the relationship between household size and wage is to observe household production. Goussé, et al.'s (2017) analysis provides foundational support for the idea that time spent in household production has an opportunity cost of the forgone wage that could have been earned if the time was spent working. Goussé, et al. (2017) found that married couples that spend more combined time in household production earned lower combined wages. This negative relationship between wages and time spent in household production represents the trade-off associated with this relationship. Because an increasing number of dependents requires more time spent in household production, there would also be a predicted negative relationship between number of dependents and wage.

Rahman's (2013) approach to the relationship between wage and household size is more direct. In the multiple regression analysis, Rahman (2013) found a strong positive relationship between poverty and household size. Poverty is defined by a household's lack of ability to earn a given income level. Because wages determine a worker's income level, poverty can also be defined in terms of the lack of ability to earn a given wage. In other words, a lower wage means a greater risk of poverty. This would produce a negative relationship between wage and poverty, which is also derived in Rahman's (2013) analysis. Because of the positive relationship between household size and poverty, there would be a negative relationship between wage and household size.

This analysis attempts to directly derive the relationship between income level and household size, by regressing the number of dependents as a function of wage. This analysis is unique because it looks at the direct relationship between household size and income level, using number of dependents as the measure of household size and wage as the measure of income level. In previous studies, the relationship between household size and wage has been either not directly tested or different measures have been used. For example, Rahman (2013) used poverty as a measure of income level instead of wage. Groesbeck and Israelsen (1994) used birth rates as a measure of household size instead of number of dependents. Cremer, et al. (2003) used number of children as a measure of number of dependents, which might vary slightly due to dependents in a household being present that are not children. The studies by Cremer, et al. (2003) and Goussé, et al. (2017) provide a significant foundation for this paper's economic reasoning for the relationship between household size and wage through the relationships derived in their studies, but they do not directly test the relationship between wage and household size.

III. Data

As previously stated, the independent variable for this study is wage and the dependent variable is the number of dependents (numdep). The control variables used in this analysis are educ, exper, tenure, married, nonwhite, and female. See the table "Variable Descriptions" provided below for information about these control variables. For choosing control variables in this study, the focus was primarily on factors that are correlated with the dependent variable, numdep. To determine the relationship between number of dependents and wage, all other factors that affect the number of dependents need to be held constant to ensure that the measure is not including differences due to other factors. Because the number of dependents in a household can differ between different ethnicities, genders, and marital statuses, variables that represent these different categories were chosen as the control variables in the study (Groesbeck 1994). The variables in the data set that were used as explanatory variables to eliminate bias based off of these three categories are married, nonwhite, and female. Number of dependents also varies with education and experience, so educ, exper, and tenure were included in this analysis as control variables. However, because education and experience also determine wage, they were only included in some of the regression models.

Table 1. Variable Descriptions

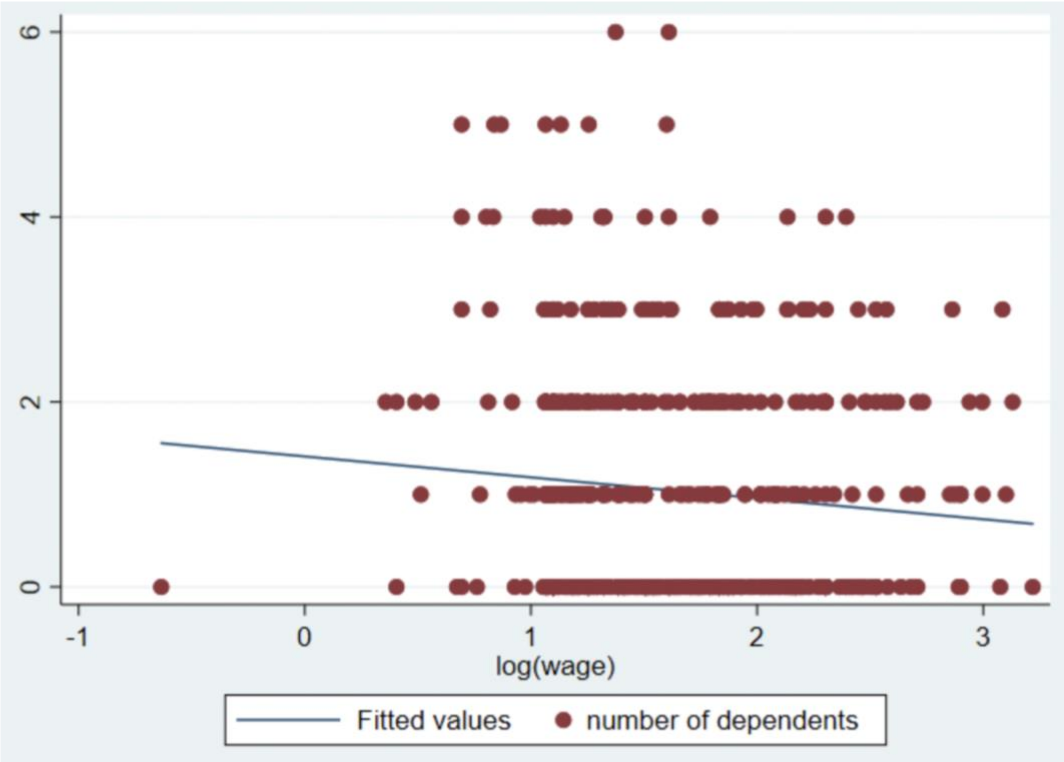
Variable Name	Description	Units	Source	Number of Observations
numdep	Number of dependents	People	1976 Current Population Survey	526
wage	Average hourly earnings	USD	1976 Current Population Survey	526
lwage	Log(wage)	-	1976 Current Population Survey	526
educ	Years of education	Years	1976 Current Population Survey	526
exper	Years of potential experience	Years	1976 Current Population Survey	526
tenure	Years with current employer	Years	1976 Current Population Survey	526
nonwhite	=1 if nonwhite, =0 otherwise	-	1976 Current Population Survey	526
female	=1 if female, =0 otherwise	-	1976 Current Population Survey	526
married	=1 if married, =0 otherwise	-	1976 Current Population Survey	526

The source of the data set is the course textbook, *Introductory Econometrics: A Modern Approach*, by J. Wooldridge. The data set is from the 1976 Current Population Survey, collected by Henry Farber and J. Wooldridge in 1988 when they were colleagues at MIT. There are 526 observations in the data set. All of the variables used in this analysis contain all 526 observations, so no modifications were necessary when determining the sample size. The variables wage, tenure, educ, exper, numdep, nonwhite, married, and female have values for each of the 526 observations contained in the data set; therefore, the sample size for this analysis is 526. No adjustments were made to the data set for this analysis, because all the variables of interest were contained in the original data set, and the course textbook offered the data set in a format useable by STATA. lwage, which represents the log of the wage, will be used as the measure of wage in the regression models, because the relationship between a one percent increase in wage and a decrease in number of dependents is stronger than the relationship between a one dollar increase in wage and a decrease in number of dependents. Because the magnitudes of the coefficients when lwage is used are larger, the relationships between the variables used in the regression models are more clear. The variable lwage was contained in the original data set, so no modifications were necessary to genereate this variable.

Table 2. Summary Statistics				
Variable	Mean	Standard Deviation	Minimum	Maximum
numdep	1.04	1.26	0	6
wage	5.90	3.69	0.53	24.98
lwage	1.62	0.53	-0.63	3.22
educ	12.56	2.77	0	18
exper	17.02	13.57	1	51
tenure	5.10	7.22	0	44
nonwhite	0.10	0.30	0	1
female	0.48	0.50	0	1
married	0.61	0.49	0	1
Number of Observations = 526				

As seen in the “Summary Statistics” table above, all variables included in the model have 526 observations, so no modifications were needed to obtain a large sample size. The sample size includes all 526 observations from the data set in this analysis. Married, nonwhite, and female all have values of either 0 or 1 because they are categorical variables. Therefore, the mean for these variables indicates what portion of the sample falls into the categories. For example, for the variable married, observations were given a value of 1 if married and 0 otherwise. Because the mean is 0.61, this means that over half of the observations in the data set fall under a married status.

Graph 1. Scatterplot of numdep (y) vs. lwage (x)



Graph 1 depicts a scatter plot of the two main variables of interest in this analysis, lwage and numdep. lwage is represented on the x-axis and is the primary independent variable for this analysis, and the dependent variable, numdep, is represented on the y-axis. As seen in Graph 1, the line of best fit between lwage and numdep is downward sloping, indicating a negative relationship between numdep and lwage, which is consistent with the hypothesis for this analysis. The significance of the negative relationship between the two primary variables of interest will be determined in this analysis in the "Results" section after the composition of a simple regression model and several multiple regression models using numdep as the dependent variable and lwage as the primary independent variable.

There are a set of Classical Linear Model (CLM) assumptions that must be evaluated as to whether this analysis meets the assumptions:

1. Assumption MLR.1 (Linear in Parameters): In this analysis, the variables of interest are numdep and wage. The model for the population can be written as numdep = $\beta_0 + \beta_1$(lwage) + u , where β_0 and β_1 are unknown parameters (constants) of interest and u is an unobserved random error

term. Because the analysis between numdep and lwage can be written as a linear model, assumption MLR.1 is met in this analysis.

2. Assumption MLR.2 (Random Sampling): The source of this data was from a Current Population Survey in 1976. It can be assumed that this produced a random sample of 576 observations used in this analysis. This assumption means that the data can be used to estimate the ceteris paribus effects of lwage on numdep and that the data are representative of the population described in MLR.1.
3. Assumption MLR.3 (No Perfect Collinearity): If there is sample variation in each independent variable and no exact linear relationship between independent variables, the ceteris paribus effect of lwage on numdep can be computed. This analysis meets this assumption, because there are no exact linear relationships between lwage, educ, exper, tenure, married, nonwhite, and female. As seen in Table 3 below, the highest correlation coefficient is 0.50 between tenure and exper, which is not close to a perfect linear relationship of 1.00; therefore, MLR.3 is satisfied.

Table 3. Correlation Coefficients For All Variables Used

	numdep	lwage	educ	exper	tenure	nonwhite	female	married
numdep	1.00							
lwage	-0.10	1.00						
educ	-0.22	0.43	1.00					
exper	-0.06	0.11	-0.30	1.00				
tenure	-0.03	0.33	-0.06	0.50	1.00			
nonwhite	0.08	-0.04	-0.08	0.01	0.01	1.00		
female	0.03	-0.37	-0.09	-0.04	-0.20	-0.01	1.00	
married	0.15	0.27	0.07	0.32	0.24	-0.06	-0.17	1.00

4. Assumption MLR.4 (Zero Conditional Mean): MLR.4 assumes that the unobserved factors are, on average, unrelated to the explanatory variables so that the unbiasedness of each OLS estimator for the corresponding population parameter holds true. For this analysis, MLR.4 is not likely to be perfectly satisfied due to omitted variable bias. One potential source of omitted variable bias for this analysis would be the physical size of the house measured in square footage. Wage and square footage of the house are likely to be positively correlated, because as people earn more, they can afford to purchase a larger house. If δ_1 represents the slope of the simple regression of square footage of a house on wage, then δ_1 is likely to be greater than 0 because if the positive relationship between wage and square footage. If β_8 represents the OLS estimator for house size if it were to be included in the model of regressing numdep on all of the independent variables in

the analysis, then β_8 is likely to be positive. As physical house size increases, there is more space for more dependents, giving β_8 a predicted positive value. Because it is predicted that $\delta_1 > 0$ and $\beta_8 > 0$, there would be a positive omitted variable bias, meaning the OLS estimators in this analysis are overestimated, due to the positive impact of house size. Including a variable representing physical house size in the model could eliminate this bias; however, data on square footage of the house was not included in the data set. Therefore, MLR.4 is not likely to be satisfied for this analysis, indicating that the results produced should be taken with caution.

5. Assumption MLR.5 (Homoskedasticity): MLR.5 states that the error u has the same variance given any values of the explanatory variables. To check whether the variance in the residuals is homoskedastic, a scatter plot can be generated plotting the residuals vs. the fitted values from regressing numdep on lwage. No clear pattern should be observed in the scatter plot for MLR.5 to be satisfied. As seen in A6 in the Appendix, there is some sort of negative relationship pattern indicating variance is not constant. MLR.5 is not satisfied for this model, so it cannot be assumed that u has the same variance for any values of the explanatory variables, which also indicates that numdep does not have the same variance for any values of the explanatory variables. Results of this analysis should be taken with caution.

   Graph 2. Residuals vs. Fitted Values
6. Assumption MLR.6 (Normality): MLR.6 states that the population error u is independent of the explanatory variables and is normally distributed with zero mean and variance σ^2. If MLR.6 is satisfied, it would indicate that the predicted values for numdep are also normally distributed. A histogram can be generated using the fitted values from the regression of numdep on lwage to check if the predicted values are normally distributed. As seen in A7 in the Appendix, the histogram is somewhat normally distributed with a slight left skew. However, MLR.6 can be dropped if the sample size is reasonably large. Because the sample size for this analysis is 526, MLR.6 can be dropped for this analysis.

IV. Results

Simple Regression Model:

numdep = β_0 + β_1(lwage) + u

After simple regression analysis using STATA: numdep = 1.41 - 0.23(wage)

Table 4. Simple Regression Results – Dependent Variable: numdep		
Independent Variable	Coefficient	Standard Error
lwage	-0.23	0.10

Constant	1.41	0.18
Observations = 526; R-squared = 0.01		

Table 4 shows the results of a simple regression model with numdep as the dependent variable and lwage as the independent variable. In the model produced by this regression, a one percent increase in wage decreases the number of dependents by 0.23. This shows a negative relationship between wage and number of dependents, which is consistent with the economic idea that there is an increasing opportunity cost associated with an increasing number of dependents.

Multiple Regression Model 1:

numdep = β_0 + β_1(lwage)+ β_2(educ)+ β_3(exper)+ β_4(tenure)+ β_5(married)+ β_6(nonwhite)+ β_7(female)+ u

After multiple regression analysis using STATA: numdep = 2.58 + 0.00(lwage) - 0.13(educ) – 0.02(exper) + 0.00(tenure) + 0.66(married) + 0.30(nonwhite) + 0.12(female)

Table 5. Multiple Regression Results – Dependent Variable: numdep		
Independent Variables	Coefficient	Standard Error
lwage	0.00	0.13
educ	-0.13	0.02
exper	-0.02	0.00
tenure	0.00	0.01
married	0.66	0.12
nonwhite	0.30	0.17
female	0.12	0.11
Constant	2.58	0.30
Observations = 526; R-squared = 0.12		

Table 5 shows the results of a multiple regression model using all of the variables in this analysis. Numdep is the dependent variable, and lwage, educ, exper, educ, tenure, married, nonwhite, and female are the independent variables for this multiple regression model. All of the variables were used in this model to determine an initial relationship between numdep and the control variables to determine which control variables have a significant impact on numdep. As seen in Table 8, which shows the combined results for all of the models, married, educ, and exper are the most significant control variables, while lwage is not significant in this model, even at the 10% level. Education and experience both had a significant negative relationship with number of dependents, where a one year decrease in education increases number of dependents by 0.13 holding all other factors constant and a one year decrease in experience increases number of dependents by 0.02 holding all other factors constant.

Multiple Regression Model 2:

numdep = $\beta_0 + \beta_1$(lwage) + β_2(educ) + β_3(exper) + β_4(married) + u

After multiple regression analysis using STATA: numdep = 2.73 – 0.03(lwage) – 0.13(educ) – 0.02(exper) + 0.64(married)

Table 6. Multiple Regression Results – Dependent Variable: numdep		
Independent Variables	Coefficient	Standard Error
lwage	-0.03	0.11
educ	-0.13	0.02
exper	-0.02	0.00
married	0.64	0.12
Constant	2.73	0.28
Observations = 526; R-squared = 0.12		

Table 6 shows the results of a multiple regression analysis regressing numdep as a function of lwage, educ, exper, and married. The independent variables used in this model were the variables that were significant at the 1% level from Multiple Regression Model 1 in addition to lwage, because lwage is the main independent variable of interest for this analysis overall. There is a negative coefficient for lwage, which is consistent with the hypothesis for this analysis; however, lwage is not a significant variable in this regression model, even at the 10% level. educ, exper, and married all remained significant at the 1% level for this model. Because education level and years of experience have an impact on wage, lwage may not be significant in this model because educ and exper are very significant. An F-test can be performed to determine whether lwage, educ, and exper are jointly significant in the model. The F-test with these 3 restrictions is conducted in the Extensions section.

Multiple Regression Model 3:

numdep = $\beta_0 + \beta_1$(lwage) + β_2(married) + β_3(nonwhite) + β_4(female) + u

After multiple regression analysis using STATA: numdep = 1.22 – 0.33(lwage) + 0.52(married) + 0.35(nonwhite) + 0.04(female)

Table 7. Multiple Regression Results – Dependent Variable: numdep		
Independent Variables	Coefficient	Standard Error
lwage	-0.33	0.11
married	0.52	0.11
nonwhite	0.35	0.18

female	0.04	0.12
Constant	1.22	0.22
Observations = 526; R-squared = 0.05		

Table 7 shows the results of a multiple regression model with numdep as the dependent variable and lwage, married, nonwhite, and female as independent variables. This regression model is able to isolate the effect of wage on household size holding married, nonwhite, and female constant. In the multiple regression model, a one percent increase in wage decreases the number of dependents by 0.33 dependents holding all other factors constant. This is consistent with the negative relationship between wage and number of dependents. lwage becomes significant in this model at the 1% level after dropping educ and exper because of the results of the F-test conducted in the Extensions section.

Table 8. Combined Results – Dependent Variable: numdep				
Independent Variables	Simple Regression	Multiple Regression 1	Multiple Regression 2	Multiple Regression 3
lwage	-0.23** (1.41)	0.00 (0.13)	-0.03 (0.11)	-0.33*** (0.11)
educ	-	-0.13*** (0.02)	-0.13*** (0.02)	-
exper	-	-0.02*** (0.00)	-0.02*** (0.00)	-
tenure	-	0.00 (0.01)	-	-
married	-	0.66*** (0.12)	0.64*** (0.12)	0.52*** (0.11)
nonwhite	-	0.30* (0.17)	-	0.35** (0.18)
female	-	0.12 (0.11)	-	0.04 (0.12)
Constant	1.41*** (0.18)	2.58*** (0.30)	2.73*** (0.28)	1.22*** (0.22)

Number of Observations	526	526	526	526
R-squared	0.01	0.12	0.12	0.05
Adjusted R-squared	0.01	0.11	0.11	0.04

Significance levels: 10%*, 5%**, 1%***

Table 9. Statistical Inference Using Multiple Regression Model 1					
Variable	Coefficient	Standard Error	t-value	p-value	95% Confidence Interval
lwage	0.00	0.13	0.02	0.98	(-0.25, 0.25)
educ	-0.13	0.02	-5.84	0.00***	(-0.18, -0.09)
exper	-0.02	0.00	-4.47	0.00***	(-0.03, -0.01)
tenure	0.00	0.01	0.40	0.69	(-0.01, 0.02)
married	0.66	0.12	5.67	0.00***	(0.43, 0.89)
nonwhite	0.30	0.17	1.75	0.08*	(-0.04, 0.63)
female	0.12	0.11	1.03	0.30	(-0.11, 0.34)
Constant	2.58	0.30	8.56	0.00***	(1.99, 3.17)

Table 9 shows the results for each variable of the calculated t-value, p-value, and 95% confidence interval using Multiple Regression Model 1 that uses all of the variables in the analysis. The statistical significance of each variable was determined using the p-values. For example, the p-value for 0.08 which corresponds to a 8% level of significance. There is one asterisk (*) next to the p-value for nonwhite and one asterisk next to the coefficient of 0.30 for nonwhite in Table 8 for Multiple Regression Model 1 to indicate that nonwhite is significant at the 10% level for this model. Because the p-value = 0.08 = 8% which is less than 10%, the p-value indicates a significance at the 10% level. However, nonwhite is not statistically significant at the 1% or 5% level for this model because the p-value is greater than both 1% and 5%. This analysis was performed for all of the variables in each of the models, and their significance levels are denoted in Table 8 with one asterisk for 10%, two for 5%, and three for 1%. All test statistics were calculated testing the null hypothesis H_0: $\beta_{i=1,2,\ldots} = 0$ against the two tailed alternative H_a: $\beta_i \neq 0$. It is important to note that a two-tailed hypothesis was used, indicating that the tests were performed to determine whether the OLS estimators for the variables present in the models were statistically different than 0 in either direction instead of only one direction, for example if the test had been performed to determine whether the OLS estimators where statistically greater than 0 instead.

V. Extensions

F-test:

Multiple Regression Model 1 will be used as the unrestricted model for the F-test. The restrictions are lwage, educ, and exper; therefore, the restricted model will be numdep = β_0 + β_1(tenure)+ β_2(married)+ β_3(nonwhite)+ β_4(female)+ u. These variables were chosen as the restrictions because education level and years of experience both affect wage. In Multiple Regression Models 1 and 2, lwage is not significant; however, educ and exper are both very significant. If lwage, educ, and exper are jointly significant as a result of the conclusions from the F-test, then educ and exper will be dropped to determine a more direct effect of lwage on numdep, without educ and exper affecting the results because of their significance. The F-statistic will be calculated using $F = [(SSR_r - SSR_{ur})/q] / [SSR_{ur}/(n-k-1)]$, where SSR_r represents the sum of squared residuals of the restricted model, SSR_{ur} represents the sum of squared residuals of the unrestricted model, q represents the number of restrictions of 3, n represents the sample size of 526, and k represents the number of variables in the unrestricted model, which is 7. The critical value used for the F-test will be $c_{0.01} = 3.78$. The null hypothesis H_0: $\beta_{lwage} = 0$, $\beta_{educ} = 0$, $\beta_{exper} = 0$ will be tested against the alternative hypothesis H_a: H_0 is not true.

$$F= [(803.55 - 733.82) / 3] / [733.82 / (526 - 7 - 1)] = 16.41$$

Because F = 16.41 > $c_{0.01}$ = 3.78, the null hypothesis is rejected. It can be concluded at the 1% significance level that lwage, educ, and exper are jointly significant. Multiple Regression Model 3 drops educ and exper from the multiple regression analysis to determine the impact of lwage on numdep without educ and exper affecting the results due to their impact on wage. In Multiple Regression Model 3, lwage was significant at the 1% level after educ and exper were dropped from the model. A one percent increase in wage decreases number of dependents by 0.33 people holding all other factors constant in Multiple Regression Model 3, which can be seen in the Results section.

VI. Conclusions

The results of this analysis support the economic hypothesis that an increase in wage leads to a decrease in number of dependents to a certain degree. The strongest evidence for this relationship resulted from Multiple Regression Model 3, where a one percent increase in wage decreased the number of dependents by 0.33 dependents holding all other factors constant, with -0.33 holding significance at the 1% level. However, the coefficient for lwage was not significant even at the 10% level for either Multiple Regression Model 1 or Multiple Regression Model 2. This could be due to multi-collinearity between lwage, educ, and exper because experience and education heavily impact wage. The results from the F-test indicated that lwage was jointly significant with educ and exper, which is why they were dropped in

Multiple Regression Model 3 to try to observe the significance of lwage on number of dependents without the strong influence of the significance from educ and exper. In addition, MLR.4 – MLR.6 were not strictly satisfied for this analysis, so the significance of the results produced is not very high. The relationship derived between wage and number of dependents in Multiple Regression Model 3 could provide support for future analysis that works to determine whether this relationship is significant economically with a data set or different analysis that satisfies all of the CLM assumptions.

In order to improve and extend the results of this analysis for future use, more recent data could be used to determine whether the negative relationship between wage and number of dependents holds true in a more modern society. Some other factors that could be included in the future are physical household size, as discussed in MLR.4, and a dummy variable indicating whether or not the family chooses to use child care. If data about child care were available, the results of this analysis could be improved, because if a family chooses to use child care, they would no longer have to spend additional time in household production that would prevent them from spending more hours working in the labor force. For this reason, it would be interesting to see whether child care availability and the cost of child care have impacts on whether or not individuals choose to have more or less children depending on their income level. Using the results from this analysis as a structure, more significant results could potentially be produced if child care statistics were used and if a more recent data set that satisfies more of the CLM assumptions was used. These potential future extensions would help to determine if there is a true significance between wage and number of dependents.

References

Cremer, H., Dellis, A., & Pestieau, P. (2003). Family Size and Optimal Income Taxation. *Journal of Population Economics*, 16(1), 37–54. http://www.jstor.org/stable/20007851

Goussé, M., Jacquemet, N., & Robin, J.-M. (2017). MARRIAGE, LABOR SUPPLY, AND HOME PRODUCTION. *Econometrica*, 85(6), 1873–1919. http://www.jstor.org/stable/44955185

Groesbeck, John D. and Israelsen, L. Dwight. (1994). The Relationship Between Household Size, Real Wages, and Labor Force Participation Rates of Men and Women. *Economic Research Institute Study Papers*, Paper 43. https://digitalcommons.usu.edu/eri/43

Rahman, M. A. (2013). HOUSEHOLD CHARACTERISTICS AND POVERTY: A LOGISTIC REGRESSION ANALYSIS. *The Journal of Developing Areas, 47(1)*, 303–317. http://www.jstor.org/stable/23612271

Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach* (7th ed.). South-Western, Cengage Learning.

Appendix

A1: Simple Regression Model

```
. regress numdep lwage

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(1, 524)       =      4.80
       Model |  7.59204039         1  7.59204039   Prob > F        =    0.0289
    Residual |  828.402256       524  1.58092034   R-squared       =    0.0091
-------------+----------------------------------   Adj R-squared   =    0.0072
       Total |  835.994297       525  1.59237009   Root MSE        =    1.2573

------------------------------------------------------------------------------
      numdep |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |  -.2262378   .1032383    -2.19   0.029    -.4290496    -.023426
       _cons |   1.410971   .1763229     8.00   0.000     1.064584    1.757358
------------------------------------------------------------------------------
```

A2: Multiple Regression Model 1

```
. regress numdep lwage educ exper tenure nonwhite female married

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(7, 518)       =     10.30
       Model |  102.172732         7  14.5961045   Prob > F        =    0.0000
    Residual |  733.821565       518  1.41664395   R-squared       =    0.1222
-------------+----------------------------------   Adj R-squared   =    0.1104
       Total |  835.994297       525  1.59237009   Root MSE        =    1.1902

------------------------------------------------------------------------------
      numdep |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |   .0026411   .1265471     0.02   0.983    -.2459675    .2512498
        educ |  -.1332456   .0228011    -5.84   0.000    -.1780397   -.0884515
       exper |  -.0217849   .0048701    -4.47   0.000    -.0313526   -.0122173
      tenure |   .0035045   .0087925     0.40   0.690    -.0137688    .0207779
    nonwhite |   .3014127   .1719686     1.75   0.080    -.0364289    .6392543
      female |   .1166905   .1134744     1.03   0.304    -.1062361     .339617
     married |   .6611716   .1166867     5.67   0.000     .4319342     .890409
       _cons |   2.577114   .3011957     8.56   0.000     1.985398    3.168829
------------------------------------------------------------------------------
```

A3: Multiple Regression Model 2

```
. regress numdep lwage educ exper married
```

Source	SS	df	MS			
Model	96.283998	4	24.0709995	Number of obs	=	526
Residual	739.710299	521	1.41978944	F(4, 521)	=	16.95
				Prob > F	=	0.0000
Total	835.994297	525	1.59237009	R-squared	=	0.1152
				Adj R-squared	=	0.1084
				Root MSE	=	1.1915

numdep	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lwage	-.0260093	.1149675	-0.23	0.821	-.2518662	.1998475
educ	-.1340249	.022633	-5.92	0.000	-.1784881	-.0895617
exper	-.0206236	.0043882	-4.70	0.000	-.0292444	-.0120028
married	.6406089	.1161786	5.51	0.000	.4123729	.8688449
_cons	2.730897	.2815867	9.70	0.000	2.177712	3.284081

A4: Multiple Regression Model 3

```
. regress numdep lwage nonwhite female married
```

Source	SS	df	MS			
Model	43.003218	4	10.7508045	Number of obs	=	526
Residual	792.991079	521	1.52205581	F(4, 521)	=	7.06
				Prob > F	=	0.0000
Total	835.994297	525	1.59237009	R-squared	=	0.0514
				Adj R-squared	=	0.0442
				Root MSE	=	1.2337

numdep	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lwage	-.3340116	.112224	-2.98	0.003	-.5544787	-.1135445
nonwhite	.3524867	.1777107	1.98	0.048	.0033692	.7016042
female	.0372955	.1164555	0.32	0.749	-.1914845	.2660756
married	.5173718	.1149622	4.50	0.000	.2915254	.7432181
_cons	1.217111	.2161461	5.63	0.000	.7924861	1.641736

A5: Restricted Model for F-Test

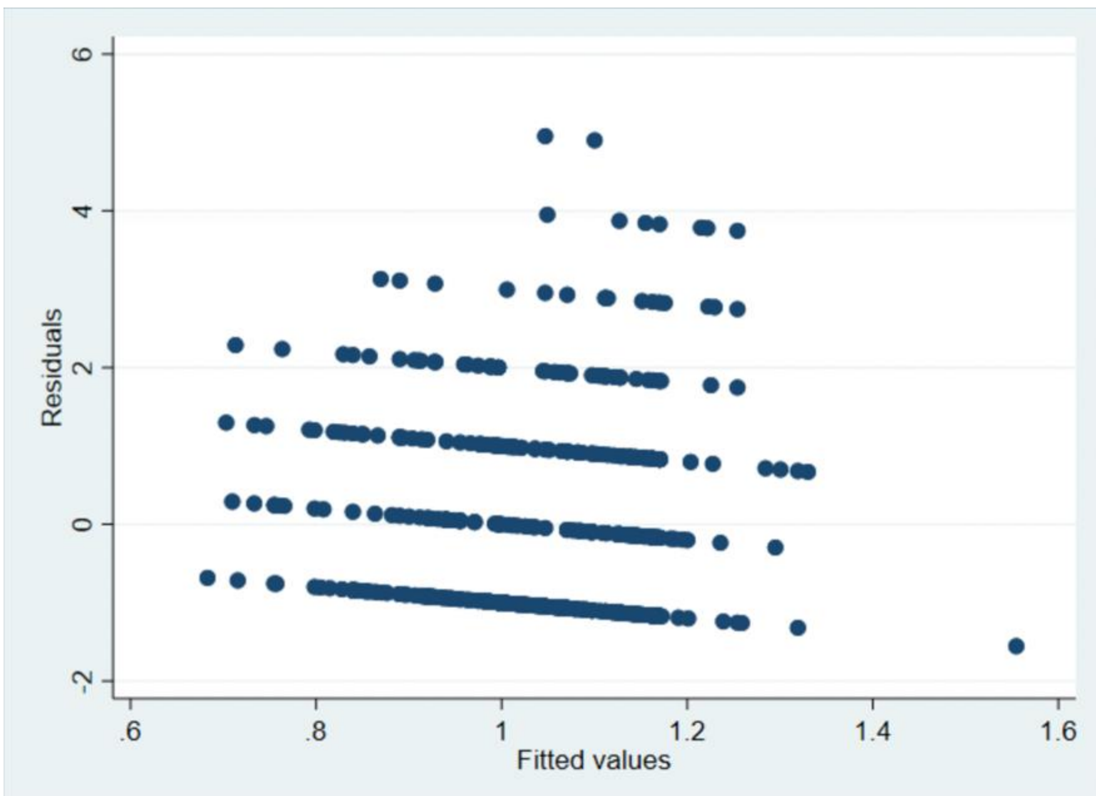
```
. regress numdep tenure nonwhite female married
```

Source	SS	df	MS
Model	32.442582	4	8.11064551
Residual	803.551715	521	1.54232575
Total	835.994297	525	1.59237009

Number of obs	=	526
F(4, 521)	=	5.26
Prob > F	=	0.0004
R-squared	=	0.0388
Adj R-squared	=	0.0314
Root MSE	=	1.2419

numdep	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tenure	-.0107887	.0078379	-1.38	0.169	-.0261866	.0046091
nonwhite	.375583	.1788481	2.10	0.036	.0242309	.7269351
female	.1322885	.1114678	1.19	0.236	-.0866931	.35127
married	.4743308	.1154437	4.11	0.000	.2475385	.7011231
_cons	.7082961	.1151762	6.15	0.000	.4820293	.9345628

A6: Residuals vs. Fitted Values – MLR.5



A7: Histogram of Predicted Values of numdep – MLR.6