Rohan Arora December 15, 2022

Final Project Econometrics

Introduction:

Food insecurity, or the lack of consistent access to enough food to support a healthy and active life, is a significant problem in the United States. It is estimated that millions of people in the country struggle with food insecurity, particularly those who live in low-income communities. This lack of access to healthy and nutritious food can adversely affect individuals and communities. For example, food insecurity can make it more difficult for children to perform well in school and even lead to developmental problems. At the same time, asthma is a common chronic respiratory condition that affects millions of people in the United States. It can cause symptoms such as shortness of breath, coughing, and chest tightness, ranging from mild to severe. It can lead to decreased productivity, missed school or work days, and a higher risk of hospitalization or emergency room visits. Also, asthma can burden families financially, as the cost of treating the condition can increase over time. Asthma and food insecurity are severe public health concerns in the United States that need to be addressed.

My research question is, "What effect does food insecurity have on asthma?" The data analysis findings did find that there was a statistically significant positive relationship between the two. When food insecurity scores increase and worsen, the dataset's probability of asthma increases. It is essential to study this effect as these conditions are common and can significantly impact a person's quality of life. Studying this relationship can provide important information to policymakers and public health officials, who can use this knowledge to develop programs and policies that address these issues. Overall, studying the relationship between asthma and food insecurity can help improve the health and well-being of individuals and communities.

Model:

This study investigates the relationship between asthma and food insecurity. The dependent variable is a binary variable indicating whether the respondent has asthma. The independent variable representing food insecurity is a variable that indicates the respondent's level of food insecurity. In addition to these two variables, I am also considering the effects of earnings and family size on food insecurity and asthma. In a study in the National Library of Medicine on food insecurity and poverty, researchers found "evidence of a strong and statistically

significant association between poverty and food insecurity" (Wight). Moreover, there is a relationship between income and food insecurity, as when income increases, the more expendable income one has to spend on food. In a study from the National Library of Medicine, researchers found "lower income levels [were] associated with progressively poorer asthma outcomes" (Cardet). As a result, there is a relationship between earnings on food insecurity and asthma.

The other variable, family size, is negatively correlated with food insecurity, as having more family members increases the cost of providing for a family. A study on the effect of family structure and food insecurity found that "that children growing up in complex family households are more vulnerable to food insecurity" (Balistreri). The number of children in the house is associated with asthma levels, as a study found that "exposure of young children to older children at home…protects against the development of asthma" (Ball). So, the more extensive the family size, the more likely the respondent has children in the house, which helps protect from the development of asthma. As a result, I expect to see a relationship between family size, food insecurity, and asthma in our study.

Data:

In this study, I am examining the relationship between asthma and food insecurity. Asthma is the dependent variable represented by a binary variable, with "1" indicating that the respondent has asthma and "0" indicating that they do not. I created this variable by looking at the responses to the asthmaev variable and categorizing the responses into a binary variable. One of the regressors in the model is the foodinsec variable, which was created from the fsstatdet variable in the original dataset. Fsstatdet divides the raw score of the responses to food insecurity into four categories: high food security (a score of 0), marginal food security (a score between 1 and 2), low food security (a score between 3 and 5), and very low food security (a score between 6 and 10). The new foodinsec variable takes the average of those ranges and gives that average to the respondent's food insecurity variable. The other initial regressor is the estearnings variable, which represents the respondent's estimated earnings. The earnings variable provided a range of income for each respondent. This variable was created by taking the income ranges provided and calculating the average of each range. The final regressor is the family size provided in the

original dataset, which states the number of family members living in the respondent household at the time of the survey.

Summary Statistics					
Variable Name	Mean	Std. dev.	Min	Max	Description
asthma	0.1319583	0.3384504	0	1	Binary variable: 0 = No Asthma, 1 = Has Asthma
foodinsec	0.6200838	1.771961	0	8	Food Insecurity rating ranging from 1(not food insecure) to 10 (most food insecure) was categorized into 4 different sections referring to the level of food insecurity. This new variable takes the average of the scores per assigned level of food insecurity.
estearnings	42680.46	24542.76	2500	750000	Estimated earnings based on responses to a range of yearly earnings
logearnings	10.38569	0.8942095	7.824	11.225	A log of the estearnings variable
familysize	1.728505	1.573941	0	12	Number of family members in household not including the respondent
northeast	0.1613517	0.3678605	0	1	Binary variable: 0 = Not in Northeast, 1 = In Northeast
midwest	0.2292329	0.4203456	0	1	Binary variable: 0 = Not in Midwest, 1 = In Midwest
west	0.2407347	0.4275359	0	1	Binary variable: 0 = Not in West, 1 = In West
south	0.3686807	0.482454	0	1	Binary variable: 0 = Not in South, 1 = In South

Results:

The regression constant is .1517853 and was statistically significant, indicating that when all variables are equal to zero, the risk of asthma is 15.17%. The coefficient for food insecurity was .0089893 and was also statistically significant, suggesting that for every increase in food insecurity rating, the chance of having asthma should increase by 0.89893% in this dataset. The next coefficient was estimated earnings, which was -3.51e-07 and was statistically significant. This coefficient indicates that for every unit dollar increase in earnings, the risk of asthma in this sample should decrease by 3.51e-07. The last coefficient was the family size, which was -.0041937 and was also statistically significant. This coefficient suggests that for every additional family member, the chance of having asthma decreases by 0.41937% in the dataset. The initial

research on the variables in the model section supports these coefficient values. An R^2 of 0.0031 on a regression indicates that the model explains 0.0031, or 0.31%, of the variance in the dependent variable. This means that the independent variables included in the model can explain a small amount of the variation in the dependent variable, but there is still a large amount of unexplained variance.

```
. regress asthma foodinsec estearnings familysize, r

Linear regression                               Number of obs     =     13,832
                                                F(3, 13828)       =      12.01
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0031
                                                Root MSE          =     .34137
```

asthma	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
foodinsec	.0089893	.0021423	4.20	0.000	.0047901	.0131886
estearnings	-3.51e-07	1.24e-07	-2.83	0.005	-5.95e-07	-1.08e-07
familysize	-.0041937	.0020418	-2.05	0.040	-.008196	-.0001915
_cons	.1517853	.0071158	21.33	0.000	.1378372	.1657333

Robustness:

To improve our regression model, I considered two alternative specifications. The first was to take the log of the estimated earnings. This made the earnings variable easier to interpret, showing the effect of a $10,000 increase instead of a $1 increase. The second specification was to create a new variable for the observation region, Northeast, West, South, and Midwest, which I expected to be correlated with food insecurity. This is because factors affecting asthma and food insecurity prevalence and severity may vary by region. For example, air quality and access to healthy, affordable food can differ across the US, potentially impacting asthma symptoms and food insecurity. Because these factors may vary by region, I expected the prevalence of food insecurity and asthma to be correlated with the US region.

In the revised model specification, the old variables showed only minor changes. The foodinsec variable changed from .0089893 to .0087877, and the familysize variable decreased from -.0041937 to -.0044695, indicating both variables had only a slightly reduced effect on asthma in the new model. This is likely due to the new regressors explaining some of the variances in the dependent variable that were previously explained by the original regressors. The

constant of the adjusted regression is statistically significant, and it increased from 0.155979 to .2700009. This increase suggests that the predicted probability of asthma increased in this model when all the variables are 0. This constant rise is caused by the new earnings variable, which takes the log of the estimated earnings. In this case, the constant term a represents the expected value of Y when ln(estearnings)=0, which is not the same as the expected value of Y when estearnings=0. The new variable representing the log of the estearnings had a coefficient of -.0138759 and was statistically significant. This indicates that in the dataset, when earnings increase by $10,000, the probability of having asthma decreases by 1.38759%. There were four regions, but one region, the South, was ignored to avoid the dummy variable trap and act as a reference to the other region categories. The coefficients for the dummy variables representing the Northeast and West regions were statistically significant, with values of .023318 and .0257587. This suggests that when living in these regions, the probability of having asthma is around 2 percentage points higher compared to the South. The last included region regressor was the West, which was not statistically significant. If a binary regressor is not statistically significant in a regression model, the estimated coefficient for that variable is not significantly different from zero. The R^2 of the regression was higher than the original specification, which suggests that the independent variables included in the model can explain more of the variation in the dependent variable. However, there is still a large amount of unexplained variance.

```
. regress asthma foodinsec logearnings familysize northeast midwest west, r

Linear regression                               Number of obs     =     13,832
                                                F(6, 13825)       =       9.53
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0048
                                                Root MSE          =     .34112
```

asthma	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
foodinsec	.0087877	.002136	4.11	0.000	.0046009	.0129745
logearnings	-.0138759	.0035439	-3.92	0.000	-.0208225	-.0069293
familysize	-.0044695	.0020428	-2.19	0.029	-.0084737	-.0004654
northeast	.023318	.0088593	2.63	0.008	.0059526	.0406834
midwest	.0058528	.0074575	0.78	0.433	-.008765	.0204705
west	.0257587	.0077573	3.32	0.001	.0105534	.040964
_cons	.2700009	.037619	7.18	0.000	.1962625	.3437392

Conclusion:

This analysis of the 2018 National Health Interview Survey dataset provides insight into the relationship between food insecurity and other factors on asthma. I used regression analysis

to find a statistically significant relationship between food insecurity and asthma. Specifically, we found that for every increase in food insecurity rating, the probability of having asthma increases by 0.87877% in this dataset. In addition, income was found to play a significant role in decreasing the likelihood of asthma, as a $10,000 increase in earnings was associated with a 1.38759% decrease in having asthma. The following variable we looked at was the family size, which was also statistically significant and suggested that for every additional member in the family size, the probability of asthma decreased by 0.44695%. The final variable we looked at was the respondent's region, where the Northeast and West regions had statistically significant differences from the South. These variables showed us that these regions do affect asthma. While the last region, the Midwest, was not statistically significant, which suggests that the Midwest does not significantly predict the outcome of asthma.

There are several potential drawbacks to using this sample to draw conclusions. One potential drawback is that the study did not provide quantitative data for multiple variables but categorical data. The food insecurity variable in the dataset was a categorical variable that split food insecurity scores into four categories. Each category represents some range of food insecurity scores; for example, people with very low food insecurity fell between 6 and 10. The data did not provide the individual's actual score; instead, it provided the food-insecure score range the respondent fell into. This was also the case for income, where a respondent was given a range of income instead of their actual income. As a result, the regression does not fully eliminate omitted variable bias when using these variables. In addition, these variables give an incomplete estimate of their effect, and it does not give us an entirely accurate representation of the sample and the regression. OVB is a threat to the internal validity of the regression, so we need to address this in future data analysis. Additionally, if there is OVB, the estimator is biased and inconsistent as it violates the least squares assumption. To address this problem with OVB, I could look for the data omitted from the regression analysis; for example, the codebook mentioned a raw score for food insecurity rating. If I can’t find the data to address the omitted variable, I can try including one or more control variables in the model instead. If these control variables achieve conditional mean independence, their inclusion in the model can help eliminate the potential bias in the coefficient for the variable of interest. However, suppose a variable is included in the regression model when it should not be. In that case, the precision of the

estimators for the other regression coefficients may be reduced, so I have to be careful in selecting these variables.

There is also the threat of measurement error for the variables earnings and food insecurity, which threatens internal validity. There are many possible sources of measurement error. The data used in this report is collected through a survey so that the respondent might have given the wrong answer—for example, one question in the survey involved last year's earnings. A respondent might not know his or her exact earnings or misstate the amount for another reason. When measurement errors occur in a regression, it can cause biased coefficient estimates, lower R^2, and higher standard errors. The most effective way to address measurement errors is to obtain accurate measurements of the relevant variables. If this is not possible, econometric techniques such as instrumental variables regression can be used to reduce the bias caused by measurement errors. This method uses an additional variable, the instrumental variable, that is correlated with the true value of X, the earnings or food insecurity variable, but is not correlated with the measurement error. Because the study threatens internal validity, the statistical inferences about causal effects are invalid for the studied population.

The dataset threatens external validity as there is a population difference. 13.2% of the dataset had asthma, which is much higher than the national average. According to the Asthma and Allergy Foundation of America (AAFA), approximately 1 in 13 people in the United States have asthma or 7.7% of the US. This difference could have resulted from people with asthma being more likely to answer the asthma question as it applied to them. At the same time, someone who did not have asthma might have glossed over the question in the survey. Therefore, the sample data collected differed from the population of interest, threatening external validity. To minimize these threats, it is best to address them at the beginning of the study before the data are collected. The future dataset should be more representative of the US population to address the threat to external validity. Until these external and internal threats are addressed, the regression will not provide a sufficiently accurate estimate of the relationship between asthma and food insecurity.

Bibliography:

Cardet, J. C., Louisias, M., King, T. S., Castro, M., Codispoti, C. D., Dunn, R., Engle, L., Giles, B. L., Holguin, F., Lima, J. J., Long, D., Lugogo, N., Nyenhuis, S., Ortega, V. E.,

Ramratnam, S., Wechsler, M. E., Israel, E., Phipatanakul, W., & Vitamin D Add-On Therapy Enhances Corticosteroid Disparities Working Group members on behalf of the AsthmaNet investigators (2018). Income is an independent risk factor for worse asthma outcomes. The Journal of allergy and clinical immunology, 141(2), 754–760.e3. https://doi.org/10.1016/j.jaci.2017.04.036

Balistreri K. S. (2018). Family Structure and Child Food Insecurity: Evidence from the Current Population Survey. Social indicators research, 138(3), 1171–1185. https://doi.org/10.1007/s11205-017-1700-7

Ball, T. M., Castro-Rodriguez, J. A., Griffith, K. A., Holberg, C. J., Martinez, F. D., & Wright, A. L. (2000). Siblings, day-care attendance, and the risk of asthma and wheezing during childhood. The New England journal of medicine, 343(8), 538–543. https://doi.org/10.1056/NEJM200008243430803

Wight, V., Kaushal, N., Waldfogel, J., & Garfinkel, I. (2014). Understanding the Link between Poverty and Food Insecurity among Children: Does the Definition of Poverty Matter?. Journal of children & poverty, 20(1), 1–20. https://doi.org/10.1080/10796126.2014.891973

Asthma and Allergy Foundation of America (AAFA). (2022). Asthma facts. Retrieved from https://aafa.org/asthma/asthma-facts/