The Causal Relationship between Birth Weight Disparities and Smoking Levels

Hui Qi

Department of Economics

University of California, Davis

June 7, 2021

Abstract

This study examined how the smoking conditions of a pregnant mother are associated with her baby's birth weight. The dataset is called "BWGHT50: N=694, cross-sectional individual data on birth weights (50% sample)" from the Wooldridge data sets in the book *Econometric Analysis of Cross Section and Panel Data*, written by Jeffrey Wooldridge. Moreover, it contains data on birth weights with 694 sample individuals and 14 variables. Although many researchers have already done many experiments and surveys on finding the possible causes of low birth weight, this paper wants to consider mainly three aspects: family condition, government regulation on cigarettes, the infant's background, like his or her gender, race, and birth order. In conclusion, this dataset and its analysis aim to find the causal relationship between birth weight disparities and smoking levels to provide suggestions on how to create an environment such that a mother could give birth to a healthy infant.

Introduction

The Centers for Disease Control and Prevention (CDC) website introduces that if a pregnant mother smokes, she puts her health and her baby's health at risk. What's more, even if she smoked and had a healthy pregnancy in the past, her next pregnancy may not guarantee being healthy again. ("Smoking, pregnancy, and babies", Cdc.gov)

People always say that children are the hope of the future. Parents hope their kids grow up with a healthy and strong body. So, giving birth to an infant that has a healthy birth weight is significant for human beings. A considerable large number of governments in the world make great efforts to support the next generation, like lowing tuition fees for local citizens, providing children's health care, etc. The US government puts effort into reducing the pressure from raising kids in poor and middle-class families during the COVID-19. Recently, the American Rescue Plan Act of 2021 offered a child tax credit to American taxpayers who had qualifying dependent children. It returned some parts of the tax to the taxpayers per month based on the numbers and ages of the kids and parents' income. This section of the act could support families who suffered during the pandemic and needed to spend more time taking care of their children instead of working because of the closure of the child-care center or schools. (Michelle P. Scott, "Child Tax Credit")

Therefore, this paper tries to find the causal relationship between birth weight disparities and smoking levels. It wants to provide some suggestions to governments to reduce the probability of giving birth to a baby having a congenital disease or a weaker immune system caused by the smoking action of pregnant mothers. According to the University of Rochester Medical Center, the low birth weight indicates that babies are born weighing less than 5 pounds, 8 ounces (2500 grams) compared to 8 pounds for average newborns. A low-birth-weight baby may be healthy although they are small. However, low birth weight is most often caused by being born too early (premature birth). The baby has less time in the mother's womb to grow and gain weight. "Low-birth-weight babies may have a harder time eating, gaining weight, and fighting infection." Even worse, some of them may be at risk for diseases and physical defects, like SIDA (sudden infant death syndrome), yellow color to the skin or eyes (jaundice), etc. As a result, it is significant to find the reasons for low birth weight and take action to reduce the possibility. (Urmc.rochester.edu)

There is much past research introducing different elements related to low birth weight. "Community income, smoking, and birth weight disparities in Wisconsin", posted by J Natl

Black Nurses Assoc in the US National Library of Medicine National Institutes of Health, examines the relationship among race, community income, smoking status, and birth weight. “The dataset came from the 1990 U.S. census zip-code level data for individual birth records in the 1998 and 1999 Vital Records Birth file for the state of Wisconsin.” This research suggests that smokers living in lower-middle-income communities have elevated risks of having a low birth-weight infant compared to mothers living in richer communities. In the dataset for this paper, there is a variable called “white”. It is a binary variable and is equal to 1 if the infant is white. Although the dataset does not include variables with more races, this paper is going to include the white variable in the regression to find out whether kids born in the white family will averagely have a healthier birth weight.

While in the “Association between Maternal Smoking during Pregnancy and Low Birthweight: Effects by Maternal Age”, researchers wanted to examine “the association between maternal smoking and low birth weight by maternal age”. The research was a questionnaire survey, and 92641 mothers completed the information when their infants were born between 2004 and 2010 in Okinawa, Japan. Those results indicated that older mothers with smoking habits were more likely to give birth to low-birth-weight infants. However, it was unclear whether the influence was because of age or other age-related factors. Although the dataset for this paper does not provide the ages of mothers when giving birth, it includes a variable called “parity”, which is the birth order of a child. This paper will add this variable into one of the regressions and investigate its effect on birth weight.

According to the article “Effects of Maternal Cigarette Smoking on Birth Weight and Preterm Birth -- Ohio, 1989” posted as a morbidity and mortality weekly report on the CDC website, most states began providing more instructions on maternal behaviors during pregnancy and information about complications of pregnancy in 1989. The Ohio Department of Health (ODH) conducted similar research in 1898 to the dataset used in this paper. It examined the proportion of low birth weight (LBW), very low birth weight (VLBW), and preterm births that were attributable to maternal cigarette smoking. The ODH study included live infants born in Ohio hospitals from January 1 to June 30, 1989. It was restricted to singleton infants of white (n=62732) and black (n=111407) mothers. The study pointed out that “infants born to smokers were more than twice as likely to have LBW as were infants born to nonsmokers". Moreover, smoking by the pregnant mother also increased the risk for VLBW and preterm birth.

The previous study mentioned above discussed the relationship among race, mother’s age, smoking, mother’s education, and birth weight. Besides most variables from the previous research, the dataset used in this paper also includes the tax and price of cigarettes in the home state in 1988. The purpose of this paper is to figure out the association between birth weight and the number of cigarettes smoked by the pregnant mother. It will also investigate whether higher tax and price of cigarettes are related to the number of cigarettes being smoked and thus, affect the birth weight. From the past research, many reasons are contributing to the low birth weight, and many reasons are correlated. For instance, parents’ education level may affect the families’ income, and both of them influence the smoking condition. In this paper, the focus is the causal relationship between the number of cigarettes smoked per day by pregnant mothers to the birth weight. So other variables will be the control variables to help investigate the relationship and how significant the influence is. Many ordinary least-squares regressions with multiple regressors and nonlinear function forms will be manipulated to explore the cause and effect.

Furthermore, this research data has some biases that are hard to eliminate in reality. Even though the data is from the same state, the respondents may smoke different brands of cigarettes, which may have different levels of ingredients producing diverse harm. For example, poor people in an area with high cigarette prices may buy illicit cigarettes that are cheaper but may cost more harm to smokers' health. What's more, even if a person is poor, she may still spend much on buying cigarettes if she is addicted to them. She may buy cheaper cigarettes, which are more harmful to her body and her unborn baby. Therefore, local prices of cigarettes may have less effect on the number of cigarettes smoked per day by pregnant women and the infant's birth weight.

Data

The dataset is called "BWGHT50: N=694, cross-sectional individual data on birth weights (50% sample)" from the Wooldridge data sets in the book *Econometric Analysis of Cross Section and Panel Data,* written by Jeffrey Wooldridge. According to the MIT website, "*Econometric Analysis of Cross Section and Panel Data* was the first graduate econometrics text to focus on microeconomic data structures, allowing assumptions to be separated into population and sampling assumptions." This cross-sectional individual data on birth weights has 694 sample individuals and 14 variables. With these many variables, it is helpful to see which variables are significant in reducing infants' birth weight. The dependent variable is the baby's birth weight in ounces. The independence variables include family's income in $1000s in 1988, numbers of cigarettes smoked per day while the mother is pregnant, the local price and tax of cigarettes in the home state in 1988, the father's and mother's years of education, the birth order of the child, the gender and race of the child, log of birth weight, birth weight in pounds, packs of smoked cigarettes, and log of the family's income. Figure 1 lists all the variables, and there is a summary of the mean, standard deviation, minimum, and maximum of each variable in Figure 2.

Model

In figure 3, the scatter plot between the independent variable cigs and the dependent variable birth weight indicates a negative relationship. If a pregnant mother smoked more cigarettes per day, she is likely to have an averagely lower birth weight infant. In addition, Figure 4 is the scatter plot between cigs and log of birth weight, and it also presents a negative relationship.

$$bwght = \beta_0 + \beta_1 \times cigs + \beta_2 \times fathedu + \beta_3 \times mothedu + \beta_4 \times faminc + \beta_5 \times mothedu^2 + u$$

The first OLS regression model wants to figure out how the family environment together with the number of cigarettes smoked per day could influence the birth weight. For the family environment, the regression considers family income, father's education level, and mother's education level, especially emphasizing the mother's education level. So, it includes the squared value of each mother's education level. As a result, the independent variables are cigs, fathedu, mothedu, faminc, and squared value of mothedu. The dependent variable is bwght. The guess is that if the mother has a higher education level, she is more likely to access the knowledge of some precautions, like the negative effect of smoking, such that she can better protect her unborn baby. Therefore, she is less likely to have a low-birth-weight infant. Furthermore, if a family is not poor, the mother can go to a more expensive hospital and take regular pregnancy examinations. Then, a wealthy family may have an average higher birth weight.

$$lbwght = \beta_0 + \beta_1 \times cigs + \beta_2 \times cigtax + \beta_3 \times cigprice + \beta_4 \times faminc + u$$

$$lbwght = \log(bwght)$$

The second OLS regression focuses on the local policy. It wants to find out whether cigarette taxes, cigarette prices, number of cigarettes, and family income affect percentage decrease in birth weight. So, the independent variables are cigs, cigtax, cigprice, and faminc. The dependent variable is the log of birth weight, lbwght.

If the regression can prove that higher cigarette taxes and prices have significance in improving percentage birth weight, this research could provide suggestions to other researchers and the government to reset local cigarette tax policy to help protect the health of unborn babies. The influence of family income on birth weight may be complicated in this regression. A family with a higher income may have more money to spend on smoking. They may buy more expensive but less harmful cigarettes. On the contrary, a lower-income family may not have much money to buy cigarettes, but cigarettes may be their way to release pressure. Thus, they may instead smoke many cheaper but harmful and illegal cigarettes.

$$lbwght = \beta_0 + \beta_1 \times cigs + \beta_2 \times male + \beta_3 \times white + \beta_4 \times male \times white + \beta_5 \times parity + u$$

$$lbwght = \log(bwght)$$

The third OLS regression model adds gender, race, and parity into consideration. The independent variables are cigs, male, white, intersection of male and white, and parity. The dependent variable is the log of birth weight, lbwght. The variables male and white are binary. The male variable is equal to 1 if the infant is a boy. It supposes that male infants may have an averagely higher birth weight than female infants. The white variable is equal to 1 if the infant is white. The purpose is to find out whether a kid born in a white family may have a higher average birth weight.

$$lbwght = \beta_0 + \beta_1 \times cigs + \beta_2 \times white + \beta_3 \times parity + u$$

$$lbwght = \log(bwght)$$

From the result of the above regression models, the fourth OLS regression model only takes the numbers of cigarettes smoked per day, gender, and the order of birth into consideration. So, the independent variables are cigs, white, and parity. The dependent variable is the log of birth weight in ounces. This regression will check again whether the three independent variables have a significant influence on the percentage change of birth weight.

From Figure 5, the average birth weight for the dataset is 118.9 ounces compared to the average 128-ounce-birthweight of infants from the University of Rochester Medical Center. It is 7.1% lighter than the average birth weight of a much larger population. Since 118.9-ounce is much smaller than the standard average birth weight, the dataset for this paper is averagely lighter. Therefore, it may be better to use the median value of the birth weight in this dataset as a standard to analyze and produce a regression model. As a result, create a new variable called bwght1. Bwght1 is 1 if the birth weight is larger than the median birth weight, 120 ounces. Otherwise, it is 0.

$$\Pr(bwght1 = 1|cigs, white, parity) = \Phi(\beta_0 + \beta_1 \times cigs + \beta_2 \times white + \beta_3 \times parity)$$

The fifth regression model is called the probit regression model. The independent variables are cigs, white, and parity. The dependent variable is the newly created binary variable bwght1. Each β_j will help measure the change in the probability of bwght1=1 when X_j increases by some number with normal distribution, holding other factors fixed. It may be even clearer to distinguish that by smoking one more cigarette at different smoking conditions, how less likely would the probability of the birth weight be greater or equal to the median value.

Result

From Figure 6 for the first regression, the coefficient of cigs is -0.72. It represents that a one-unit increase in the number of cigarettes smoked per day by a pregnant mother will reduce her infant's birth weight by an average of 0.72 ounces if other independent variables and the error term are fixed. Its t-ratio is -5.78, which is less than -1.96 under the 5%

significance level. Therefore, the null hypothesis that the coefficient of cigs is equal to zero will be rejected. Thus, the number of cigarettes smoked per day has a significant influence on birth weight. Even in the real world, a 0.72-ounce drop is significant, considering that the birth weight for average newborns is 128 ounces. While the t-ratio for fatheduc, motheduc, faminc, and motheduc2 are all between -1.96 and 1.96 under a 5% significant level. Then, they fail to reject the null hypothesis that the coefficients of fatheduc, motheduc, faminc, and motheduc2 are zero. (motheduc2 = motheduc × motheduc) Also, testing the restriction of the coefficients of fatheduc, motheduc, faminc, motheduc2 equal to 0, the f statistic is 0.74 less than 2.3719 with 4 restrictions and the 5% significance level (see fig. 7&8). Therefore, they fail to reject the null hypothesis that the coefficients of independent variables, except cigs and constant, are zero. These variables are not significant in the model. Some problems are occurring in this regression. First, there is no record of some fathers' education level, so the observations' number decreased to 588. Moreover, although both father and mother's education are not significant, it seems that in contrast to the guess, the father's education is more influential than the mother's education because the absolute value of the coefficient of fatheduc is 1.61 that is greater than the 0.14 of the absolute value of the coefficient of motheduc.

From Figure 9 of the second regression model, the coefficient of cigs is -0.0049138 which interprets that if a pregnant mother smokes one more cigarette, her baby in her womb is estimated to lose 0.49% weight with other independent variables fixed. Its t-ratio is -4.85, which is less than -1.96 with the 5% significant level, so cigs is significant in the regression model. However, the p-value of the other independent variables, cigtax, cigprice, faminc, are greater than 0.05 with the 5% significant level. Therefore, they fail to reject the null hypothesis that the coefficients of the other independent variables are equal to 0. Among these three independent variables, family income has more influence on the dependent variable.

From the Figure 10 of the third OLS regression, the coefficient of cigs is -0.005345. It indicates that if a pregnant mother smokes one more cigarette, her baby in her womb is estimated to lose 0.53% weight when other independent variables are unchanged. The coefficient of white is 0.0603268. Hence, if the infant is white, he or she is estimated to have 6% more birth weight with other independent variables fixed. The 6% more birth weight is quite a lot in the real world because the average birth weight is 8 pounds that are 128 ounces. The coefficient of parity is 0.0180721, which represents that if a mother has one more baby, her newborn baby will have on average 1.8% more birth weight than the previous one with other conditions the same. A possible reason is that the mother has more experience in taking care of herself and the baby. In addition, the t-ratio of the other independent variables, male and malewhite, is between negative 1.96 and positive 1.96 with the 5% significant level. Therefore, they fail to reject the null hypothesis that their coefficients are equal to 0. There is no evidence that male and malewhite are significant in affecting birth weight.

The fourth OLS regression model in Figure 11 only includes cigs, white, parity as the independent variables. The coefficient of cigs is -0.0054286. Hence, if a pregnant mother smokes one more cigarette, her baby's birth weight is estimated to drop 0.54% when other independent variables remain unchanged. The coefficient of white is 0.0615256, which interprets as if the infant is white, he/she is estimated to have 6.15% more birth weight with other independent variables fixed. The coefficient of parity is 0.017888. It indicates that if a mother gives birth one more time, her baby will have on average 1.8% more birth weight than the previous one with other conditions unchanged. All the coefficients have little change compared to the third OLS regression model. What's more, the p-value of all the variables in the figure is less than 0.05 under the 5% significant level. Therefore, there is evidence that the coefficients of cigs, white, and parity are not zero. As a result, cigs, white, and parity significantly influence birth weight.

Furthermore, since birth order is a new finding that the other related research mentions in the introduction do not include, this paper does a single restriction hypothesis test on parity (see fig. 11). The f statistic is 5.65, which is the squared result of the t statistic 2.38. Since both test results are larger than each respective critical value under the 5% confidence level, the null hypothesis that the coefficient of parity on the log of birthweight is 0 is rejected. Therefore, parity indeed has a significant influence on the log of birth weight. However, there may be a potential endogeneity issue in the fourth OLS regression analysis. Although from the previous regression model, it has been proven that there is no evidence that family's income, parents' education level, or price of cigarettes are significant in reducing birth weight. But these factors are in the error term in the fourth OLS regression. More importantly, they are actually correlated with the parameter of interest, which is the number of cigarettes smoked per day by pregnant mothers. Family income, parents' education degree, and price of cigarettes will affect the number of cigarettes to some degree. To solve this potential endogeneity problem, the gender of the infant and local tax of the cigarettes can be utilized as the instrumental variables in the regression model. Human beings cannot control the gender of the newborn infant. Even with modern medicine, parents may know the infant's gender after it has been in the womb for 14 weeks. Therefore, the infant's gender is random and irrelevant with the error term. Moreover, in the United States, tobacco is taxed by federal, state, and local governments. Thus, the error term is independent of the cigarettes' tax. However, the gender of the infant and local tax of cigarettes may have some influence on the number of cigarettes smoked per day. For instance, in some areas, families value boys more than girls. Hence, if the mother is tested to know that her infant is a boy, she may pay more attention to her infant. She may stop smoking and she will be far away from people who smoke. Furthermore, a high local tax on cigarettes may increase the price of cigarettes and reduce its amount of sales.

Figure 12 presents the regression result for the fifth regression model. Then, from Figure 13, based on the estimated coefficients, if the cigs number changes from 0 to 1, then the probability of $bwght1 = 1$ will decrease by 1.50%. If the cigs number increases from 10 to 11, then the probability of $bwght1 = 1$ will decrease by 1.42%. Figure 14 provides the multiple restrictions hypothesis that the coefficients of cigs, white, and parity are 0. The chi-square result is 27.32, large enough to reject the null hypothesis with the 5% significance level. Therefore, cigs, white, parity are significant in this probit regression model. Moreover, figure 15 demonstrates a single restriction hypothesis on cigs. The ci-square result is 18.28, which is large enough to reject the null hypothesis that the coefficient of cigs is equal to 0. As a result, it proves that the variable cigs is significant in regression the probability of $bwght1 = 1$.

Conclusion

The purpose of this paper is to find the causal relationship between the number of cigarettes smoked per day by the pregnant mother and her infant's birth weight. From the regression models and analyses, smoking cigarettes will cause a decrease in birth weight. It is shown from the result of regression models that if a pregnant mother smokes one more cigarette, her baby's birth weight is estimated to drop 0.54% when other independent variables remain unchanged. Moreover, if the infant is white, he/she is estimated to have 6.15% more birth weight with the other independent variables fixed. The birth order also affects the birth weight. The second birth infant will have an average of 1.8% more birth weight than the first birth infant if other factors the same. As mentioned in the introduction, "Association between Maternal Smoking during Pregnancy and Low Birthweight: Effects by

Maternal Age" points out that older mothers with smoking habits were more likely to give birth to low-birth-weight infants. However, this paper indicates that with more birth order, the new-birth infant will have an averagely higher birth weight than the previous one. So, how does the age of the pregnant mother and her birth order together affect the birth weight may be a further research topic.

Although this paper indicates one of the main reasons that cause the low birth weight, the factors that lead to low birth weight are much more complicated. If pregnant mothers do not follow a healthy diet during pregnancy and drink alcohol, smoke, or use drugs, the infants will be very likely to have lower birth weight. One regression model in this paper takes the price and tax of the cigarettes into account, but there is no evidence that they are significant. Therefore, besides regulating the tax on cigarettes, which many governments have done to reduce smoking, it may be helpful to provide subsidies for mothers to go to a legal hospital and take regular pregnancy examinations. It is better if every mother could learn the importance of giving birth to a healthy infant and how to take good care of the unborn infant from doctors and nurses.

References

F Distribution Tables. http://www.socr.ucla.edu/Applets.dir/F_Table.html. Accessed 5 June 2021.

"Low Birth Weight." University of Rochester Medical Center, www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=90&contentid=p02382. Accessed 16 May 2021.

Morbidity and mortality weekly report. "Effects of Maternal Cigarette Smoking on Birth Weight and Preterm Birth -- Ohio, 1989." Centers for Disease Control and Prevention, 28 Sep 1990, www.cdc.gov/MMWR/preview/mmwrhtml/00001782.htm. Accessed 16 May 2021.

Scott, Michelle P. "What Is the Child Tax Credit?" Investopedia, 7 May 2021, www.investopedia.com/terms/c/childtaxcredit.asp. Accessed 16 May 2021.

Sims, Mario et al. "Community income, smoking, and birth weight disparities in Wisconsin." Journal of National Black Nurses' Association: JNBNA vol. 18,2 (2007): 16-23.

"Smoking, Pregnancy, and Babies." Centers for Disease Control and Prevention, 15 Feb. 2021, www.cdc.gov/tobacco/campaign/tips/diseases/pregnancy.html. Accessed 16 May 2021.

"Wooldridge data sets" http://fmwww.bc.edu/ec-p/data/wooldridge/datasets.list.html. Accessed 5 June 2021.

Zheng, Wei et al. "Association between Maternal Smoking during Pregnancy and Low Birthweight: Effects by Maternal Age." PloS one vol. 11,1 e0146241. 21 Jan. 2016, doi:10.1371/journal.pone.0146241

Appendix

A.1 Tables and Figures

Figure 1

1.	faminc	1988 family income, $1000s
2.	cigtax	cig. tax in home state, 1988
3.	cigprice	cig. price in home state, 1988
4.	bwght	birth weight, ounces
5.	fatheduc	father's yrs of educ
6.	motheduc	mother's yrs of educ
7.	parity	birth order of child
8.	male	=1 if male child
9.	white	=1 if white
10.	cigs	cigs smked per day while preg
11.	lbwght	log of bwght
12.	bwghtlbs	birth weight, pounds
13.	packs	packs smked per day while preg
14.	lfaminc	log(faminc)

Figure 2

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
faminc	694	29.03602	18.5336	.5	65
cigtax	694	19.76657	7.470933	2	38
cigprice	694	131.0408	10.05966	103.8	152.5
bwght	694	118.9323	20.60891	38	271
fatheduc	589	13.20204	2.650554	3	18
motheduc	693	12.886	2.354431	3	18
parity	694	1.636888	.9041317	1	6
male	694	.5043228	.5003419	0	1
white	694	.7824207	.4128973	0	1
cigs	694	2.412104	6.435622	0	50
lbwght	694	4.762425	.1852552	3.637586	5.602119
bwghtlbs	694	7.433267	1.288057	2.375	16.9375
packs	694	.1206052	.3217811	0	2.5
lfaminc	694	3.084411	.8881126	-.6931472	4.174387

.

Figure 3

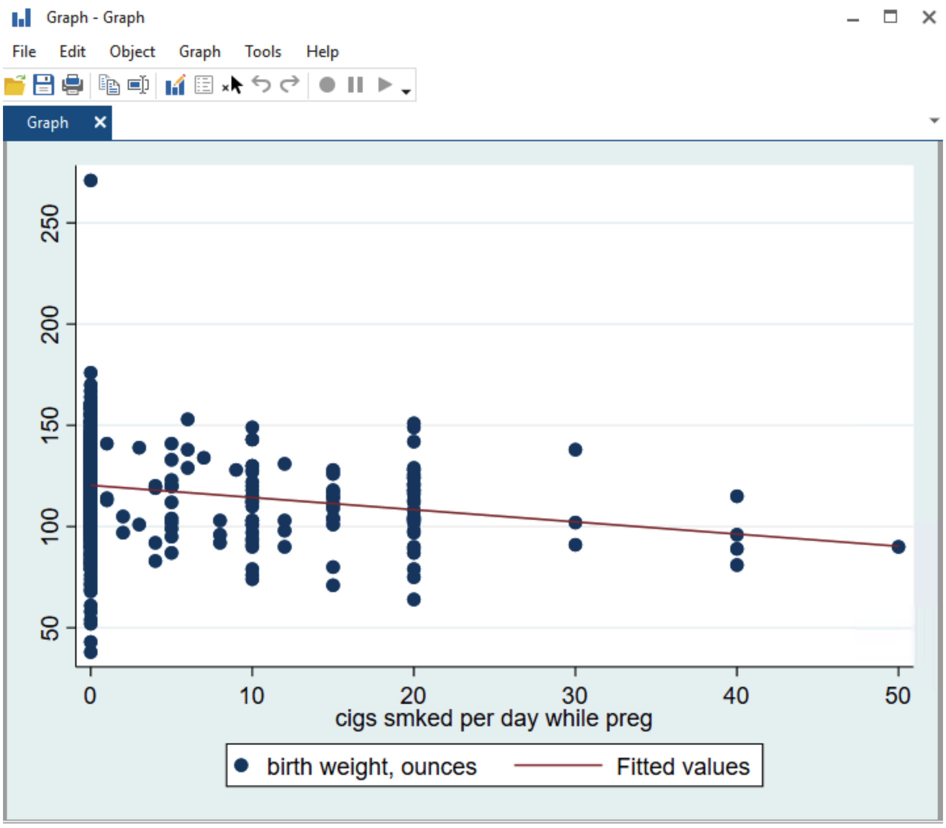


Figure 4

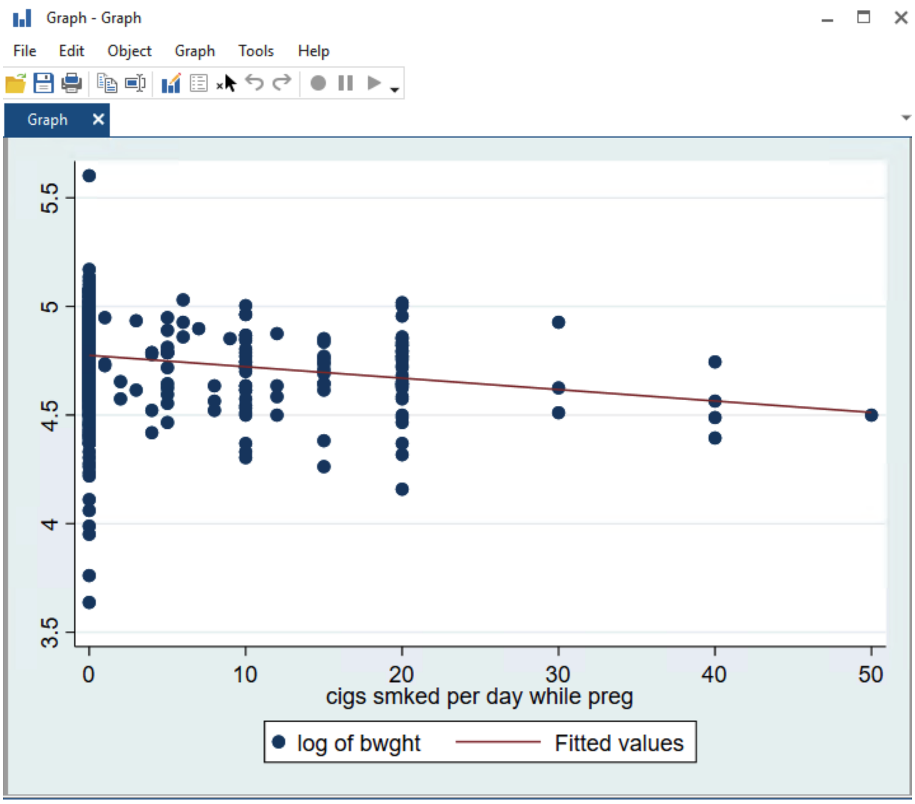


Figure 5

```
. summarize bwght,detail

                    birth weight, ounces
-------------------------------------------------------------
      Percentiles      Smallest
 1%           64             38
 5%           84             43
10%           94             52       Obs                 694
25%          106             54       Sum of Wgt.         694

50%          120                      Mean           118.9323
                        Largest       Std. Dev.      20.60891
75%          132            167
90%          143            170       Variance        424.727
95%          150            176       Skewness       .2012538
99%          161            271       Kurtosis       7.387228

.
end of do-file
```

Figure 6

```
. do "C:\Users\huiqi\AppData\Local\Temp\STD2910_000000.tmp"

. regress bwght cigs fatheduc motheduc faminc motheduc2, vce(robust)

Linear regression                               Number of obs     =        588
                                                F(5, 582)         =       7.78
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0506
                                                Root MSE          =     19.677
```

bwght	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
cigs	-.7211953	.1246803	-5.78	0.000	-.9660734	-.4763171
fatheduc	.6315896	.393327	1.61	0.109	-.1409237	1.404103
motheduc	-.2736997	1.925926	-0.14	0.887	-4.056312	3.508912
faminc	.0005446	.0509841	0.01	0.991	-.0995907	.1006799
motheduc2	-.0103603	.076113	-0.14	0.892	-.1598499	.1391293
_cons	118.5375	12.41941	9.54	0.000	94.14517	142.9298

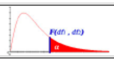
.

Figure 7

```
. do "C:\Users\huiqi\AppData\Local\Temp\STD2910_000000.tmp"

. test (fatheduc=0) (motheduc=0)(faminc=0)(motheduc2=0)

 ( 1)  fatheduc = 0
 ( 2)  motheduc = 0
 ( 3)  faminc = 0
 ( 4)  motheduc2 = 0

       F(  4,   582) =    0.74
            Prob > F =    0.5663

.
end of do-file
```

Figure 8

F Table for α = 0.05



/	df₁=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df₂=1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

Figure 9

```
. do "C:\Users\huiqi\AppData\Local\Temp\STD2910_000000.tmp"

. regress lbwght cigs cigtax cigprice faminc, vce(robust)

Linear regression                               Number of obs     =        694
                                                F(4, 689)         =       7.99
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0377
                                                Root MSE          =     .18226

------------------------------------------------------------------------------
             |               Robust
      lbwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.0049138   .0010123    -4.85   0.000    -.0069014   -.0029262
      cigtax |  -.0001648    .001747    -0.09   0.925    -.0035949    .0032652
    cigprice |   .0003315    .001302     0.25   0.799    -.0022248    .0028878
      faminc |   .0006359   .0003732     1.70   0.089    -.0000968    .0013686
       _cons |   4.715632   .1415975    33.30   0.000     4.437618    4.993646
------------------------------------------------------------------------------

.
end of do-file
```

Figure 10

```
. do "C:\Users\huiqi\AppData\Local\Temp\STD2910_000000.tmp"

. regress lbwght cigs male white malewhite parity ,vce(robust)

Linear regression                               Number of obs     =        694
                                                F(5, 688)         =       9.70
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0622
                                                Root MSE          =     .18005

------------------------------------------------------------------------------
             |               Robust
      lbwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |   -.005345   .0010411    -5.13   0.000    -.0073891   -.0033008
        male |   .0219866   .0322915     0.68   0.496    -.0414151    .0853883
       white |   .0603268   .0238138     2.53   0.012     .0135704    .1070832
   malewhite |   .0040163   .0356381     0.11   0.910    -.0659563    .0739888
      parity |   .0180721   .0074956     2.41   0.016     .0033552     .032789
       _cons |   4.685884   .0249209   188.03   0.000     4.636954    4.734814
------------------------------------------------------------------------------

.
end of do-file
```

Figure 11

```
. regress lbwght cigs white parity  ,vce(robust)

Linear regression                               Number of obs     =        694
                                                F(3, 690)         =      14.50
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0576
                                                Root MSE          =     .18024

------------------------------------------------------------------------------
             |               Robust
      lbwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.0054286   .0010346    -5.25   0.000    -.0074599   -.0033973
       white |   .0615256   .0180394     3.41   0.001      .026107    .0969443
      parity |    .017888   .0075249     2.38   0.018     .0031136    .0326624
       _cons |     4.6981   .0213582   219.97   0.000     4.656165    4.740034
------------------------------------------------------------------------------

.
end of do-file

. do "C:\Users\huiqi\AppData\Local\Temp\STD2910_000000.tmp"

. test (parity = 0)

 ( 1)  parity = 0

       F(  1,   690) =    5.65
            Prob > F =    0.0177

.
end of do-file

.
```

Figure 12

```
. probit bwght1 cigs white parity, vce(robust)

Iteration 0:   log pseudolikelihood = -480.69538
Iteration 1:   log pseudolikelihood = -464.86497
Iteration 2:   log pseudolikelihood = -464.81642
Iteration 3:   log pseudolikelihood = -464.81641

Probit regression                               Number of obs     =        694
                                                Wald chi2(3)      =      27.32
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -464.81641               Pseudo R2         =     0.0330

------------------------------------------------------------------------------
             |               Robust
      bwght1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |   -.038017   .0088908    -4.28   0.000    -.0554427   -.0205913
       white |   .3403927   .1192315     2.85   0.004     .1067034    .5740821
      parity |   .1118542   .0550899     2.03   0.042     .0038801    .2198284
       _cons |  -.4043707   .1423681    -2.84   0.005     -.683407   -.1253344
------------------------------------------------------------------------------
```

Figure 13

```
. margins, dydx(cigs) at (cigs = (0 1 5 10 15 20))

Average marginal effects                        Number of obs     =        694
Model VCE    : Robust

Expression   : Pr(bwght1), predict()
dy/dx w.r.t. : cigs

1._at        : cigs            =           0

2._at        : cigs            =           1

3._at        : cigs            =           5

4._at        : cigs            =          10

5._at        : cigs            =          15

6._at        : cigs            =          20

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
cigs         |
         _at |
          1  |  -.0149487   .0034753    -4.30   0.000    -.0217602   -.0081371
          2  |  -.0149637   .0034892    -4.29   0.000    -.0218024    -.008125
          3  |  -.0148139   .0034001    -4.36   0.000    -.0214781   -.0081498
          4  |  -.0141729    .002984    -4.75   0.000    -.0200213   -.0083244
          5  |  -.0130905   .0023012    -5.69   0.000    -.0176009   -.0085801
          6  |  -.0116724    .001474    -7.92   0.000    -.0145614   -.0087834
------------------------------------------------------------------------------
```

Figure 14

```
. test(cigs = 0)(white = 0)(parity = 0)

 ( 1)  [bwght1]cigs = 0
 ( 2)  [bwght1]white = 0
 ( 3)  [bwght1]parity = 0

           chi2(  3) =   27.32
         Prob > chi2 =    0.0000

.
```

Figure 15

```
. test(cigs = 0)

 ( 1)  [bwght1]cigs = 0

           chi2(  1) =   18.28
         Prob > chi2 =    0.0000
```

Figure 16

```
. logit bwght1 cigs white parity, vce(robust)

Iteration 0:   log pseudolikelihood = -480.69538
Iteration 1:   log pseudolikelihood = -464.80506
Iteration 2:   log pseudolikelihood = -464.72353
Iteration 3:   log pseudolikelihood =  -464.7235

Logistic regression                             Number of obs     =        694
                                                Wald chi2(3)      =      25.00
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =  -464.7235               Pseudo R2         =     0.0332

------------------------------------------------------------------------------
             |               Robust
      bwght1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.0632542   .0156419    -4.04   0.000    -.0939119   -.0325966
       white |    .552323   .1934071     2.86   0.004     .1732522    .9313939
      parity |   .1798206   .0897361     2.00   0.045      .003941    .3557001
       _cons |  -.6532933   .2303764    -2.84   0.005    -1.104823   -.2017639
------------------------------------------------------------------------------
```

Figure 17

```
. margins, dydx(cigs) at (cigs = (0 1 5 10 15 20))

Average marginal effects                        Number of obs     =        694
Model VCE     : Robust

Expression    : Pr(bwght1), predict()
dy/dx w.r.t. : cigs

1._at         : cigs            =           0

2._at         : cigs            =           1

3._at         : cigs            =           5

4._at         : cigs            =          10

5._at         : cigs            =          15

6._at         : cigs            =          20

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
cigs         |
         _at |
          1  |  -.0155204   .0038083    -4.08   0.000    -.0229845   -.0080563
          2  |  -.0155413    .003827    -4.06   0.000    -.0230422   -.0080404
          3  |   -.015331   .0036894    -4.16   0.000    -.0225621     -.0081
          4  |  -.0144506   .0030898    -4.68   0.000    -.0205065   -.0083947
          5  |   -.013028   .0021831    -5.97   0.000    -.0173067   -.0087492
          6  |  -.0112789   .0012118    -9.31   0.000     -.013654   -.0089038
------------------------------------------------------------------------------
```

Figure 18

```
use "bwght50.dta",clear

label var faminc " 1988 family income, $1000s"
label var cigtax " cig. tax in home state, 1988"
label var cigprice "cig. price in home state, 1988"
label var bwght " birth weight, ounces"
label var fatheduc "father's yrs of educ"
label var motheduc "mother's yrs of educ"
label var parity "birth order of child"
label var male "=1 if male child"
labe  var white "=1 if white"
label var cigs "cigs smked per day while preg"
label var lbwght "log of bwght"
label var bwghtlbs "birth weight, pounds"
label var packs "packs smked per day while preg"
label var lfaminc "log(faminc)"
save bwght50,replace

use bwght50,clear
summarize

//generate logcigs = log(cigs)
generate motheduc2 = motheduc*motheduc
twoway(scatter bwght cigs)(lfit bwght cigs)
twoway(scatter lbwght cigs)(lfit lbwght cigs)
//twoway(scatter bwght logcigs)(lfit bwght logcigs)
//twoway(scatter lbwght logcigs)(lfit lbwght logcigs)

//first OLS regression on family condition
regress bwght cigs fatheduc motheduc faminc c.motheduc#c.motheduc, vce(robust)
test (fatheduc=0) (motheduc=0)(faminc=0)(motheduc2=0)

//second OLS regression on cig tax and price
regress lbwght cigs cigtax cigprice faminc, vce(robust)

//third OLS regression on
generate malewhite = male*white
regress lbwght cigs male white malewhite parity ,vce(robust)

//fourth OLS regression
generate fatheduc2 = fatheduc*fatheduc
regress lbwght cigs white parity  fatheduc fatheduc2,vce(robust)

//fifthe OLS regression
regress lbwght cigs white parity  ,vce(robust)
test (parity = 0)
```

Figure 19

```
generate logparity = log(parity)
generate lnparity = ln(parity)
twoway(scatter bwght logparity)(lfit bwght logparity)
twoway(scatter bwght lnparity)(lfit bwght lnparity)

regress bwght parity c.parity#c.parity,vce(robust)
twoway(scatter bwght parity)(function y = _b[_cons] + _b[parity]*x + _b[c.parity#c.parity]*x^
2, range(0 10))

//linear probability model
//median of birth weight bwght is 120 ounces
summarize bwght,detail
generate bwght1 = 1 if bwght>120
replace bwght1 =0 if bwght <=120

probit bwght1 cigs white parity, vce(robust)
margins, dydx(cigs) at (cigs = (0 1 5 10 15 20))
test(cigs = 0)(white = 0)(parity = 0)
test(cigs = 0)

// logit
logit bwght1 cigs white parity, vce(robust)
margins, dydx(cigs) at (cigs = (0 1 5 10 15 20))
```

A.2 Some Details on Dataset

Figures 16 and 17 provide the logit regression model of bwght1 =1 on cigs, white, and parity. The coefficient results are similar to the results from the probit regression model.