**Problem Set on “Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns,”** developed by Amanda Kowalski, Michael Cunetta, and Tiffany Fan.

This problem set is based on Almond, Doyle, Kowalski, & Williams (2010) *Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns.* It is highly recommended that you read the paper, which can be found in the *Quarterly Journal of Economics,* May 2010, Vol. 125, No.2, pages 591-634.

This problem set is designed to deepen your understanding of the theoretical model and empirical results behind the method of estimating the marginal returns to medical care that the authors use. Included with the problem set questions below is a Stata do-file template that is provided for reference. It was developed with Stata version 11. It is not necessary to use the template in answering the empirical questions, but may be helpful if you have limited experience with Stata.

* The data file for this problem set is available for download at: <http://data.nber.org/lbid/adkw.dta>. By downloading, you agree to the NCHS data rules:

1. Use the data in this dataset for statistical reporting and analysis only.  
2. Make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery.  
3. Not link this dataset with individually identifiable data from other  
NCHS or non- NCHS datasets.

Please complete the questions as clearly and concisely as possible, drawing on what you have learned from the paper.

**1 Comprehension**

Read the paper and answer the following questions:

1.1. Why might a discontinuity exist in health inputs for infants just above vs. just below 1500g?

1.2. What crucial assumption about the health of infants near the Very Low Birth Weight cutoff allows the authors to estimate the marginal returns to medical care?

1.3. What happens to this assumption as one moves away from the cutoff in either direction?

1.4. What are the two ways the study measures health inputs?

1.5. What is the one way the study measures health outputs?

1.6. What is the “nationwide data” that the paper refers to?

1.7. How do the authors link data sets to arrive at their “five-state sample”?

**2 Understanding the Model**

Use the following model for infant *i* weighing *g* grams in year *t*:

Yi = α0 + α1VLBWi + α2VLBWi × (gi − 1500) +α3(1−VLBWi)×(gi −1500)+αt +αs +δXi′ +εi

where Y is an outcome or treatment measure such as one-year mortality or costs, and VLBW is an indicator that the newborn was classified as VLBW (that is, strictly less than 1,500 g). For now, ignore the indicators αt, αs, and δXi’, and **consider Y to be one-year mortality**.

2.1. Define a “dummy variable.”

2.2. What value does VLBW take if an infant weighs more than 1500g? Less than 1500g?

2.3. What is the expected value of Y, in terms of regression coefficients given in the model above, for an infant weighing:

i. 1450 g?

ii. 1550 g?

iii. 1499 g?

iv. 1501 g?

2.4. What is the interpretation of the coefficients *α2* and *α3*?

2.5. Why do *α2* and *α3* differ?

2.6. (2 Points) Graphically, assuming *α2* <0 and *α3*<0, if *α2* > *α3*, how would the slopes differ above vs below the cutoff?

2.7. At the cutoff, there are two intercepts: one for infants just under VLBW classification and one for infants just over it.

i. What are these two intercepts in terms of the coefficients?

ii. What does the difference between the intercepts represent?

iii. Why does this difference exist?

2.8. If you were to graph Y (one year mortality) on the y-axis and g (infant’s weight in grams) on the x-axis, what would the “y-intercept” (g=0) be, in terms of the coefficients?

2.9. In words, what does the y-intercept value represent and why is this value nonsensical?

2.10. Draw and label a graph of the regression model with Y on the y-axis and g on the x-axis, 0≤g≤3000.

2.11. Based on your graph, would you expect the sign of the following coefficients to be positive or negative?

i. α0

ii. α1

iii. α2

iv. α3

**3 Running the Regression**

Save the nationwide mortality data file adkw.dta to an easily accessible location on your computer. If you choose to use the do-file template, open the do-file birthweightpset.do and follow the instructions. Some of the code has been written already; it is your job to fill in the gaps. Once you’re done, turn in your log file, which will contain your list of commands and results. If you get stuck on a particular command, try a Google search. You can also use Stata’s help feature by typing “help [command you don’t understand].” If you choose not to use the do-file template, please record the commands you used in separate do- and log-files. Use your results to answer the following questions:

3.1. Report your estimates of the following coefficients:

i. α0

ii. α1

iii. α2

iv. α3

3.2. Which of these coefficients are statistically significant?

3.3. According to your results, how much does mortality change as birth weight passes from just under to just over 1500g?

3.4. The authors found that the dollar difference in hospital charges at the VLBW cutoff for the five-state sample was $4553. Given this value and the value you just estimated, what is the cost of saving a newborn’s life?

3.5. Cutler and Meara (2000) calculate a quality-adjusted value of a newborn life for newborns born in 1990 near 1,500 g to be approximately $2.7 million. Based off your answer to part 3.4, does the additional expenditure seem “worth it”?

3.6. Print your graph of gestational age vs. birth weight.

i. How does gestational age change as birth weight increases?

ii. What happens to gestational age around the 1500g threshold?

iii. Why is this important for the main result of the study?

**4 Concluding Questions**

4.1. What is novel about the way this study estimated returns to medical care?

4.2. Describe a potential limitation of this study.

4.3. How might this study inform policy?