ECO 6375 Prof. T. Fomby
Econometrics III Fall 2008

EXERCISE 5

POISSON REGRESSION MODEL

Purpose: To learn how to use the **Poisson Regression model** to analyze dependent variables that represent "counts."

Go to my website and download the file install.dat into an EVIEWS file. I analyzed this data as a private consultant so "some of the names have been changed to protect the innocent" and so that I will not compromise the confidentiality of the data. This file contains 72 observations on two variables: time and installmonth. The variable "time" (the first column) is just a time index for your convenience in creating dummy variables. Let us assume that the variable installmonth (the second column) is the number of software systems that computer service firm XYZ (a fictitious name) installs per month. The data spans the period March 1994 through December 1999. There are two major events to consider. First during the period May 1996 through December 1999, company XYZ claims that computer services company RST (another fictitious name) engaged in predatory price cutting and cost it substantial loss of business in doing so. In addition, company XYZ claims that the period October 1998 through March 1999 was unusual with respect to the number of installations they made per month because of the Y2K scare. Using this information, complete the following tasks:

(a) Create the following two "event" dummy variables. Let doct98_mar99 be a dummy variable that is one for the period October 1998 through March 1999 and zero otherwise. This dummy variables controls for the Y2K period. Let dmay96f be a dummy variable that is one for May 1996 forward and zero before May 1996. This dummy variable represents the "predation" period. If there was a Y2K scare, what is the appropriate sign on the dummy variable doct98_mar99? If there was predation, what is the appropriate sign on the dummy variable dmay96f? Explain your reasoning.

(b) Using installmonth as the dependent variable, a constant term, and the two dummy variables doct98_mar99 and dmay96f, estimate a Poisson Regression model using the maximum likelihood technique. Report your estimated equation with coefficient estimates and standard errors. Are the coefficients of the Poisson Regression model statistically significant? Do the coefficients have the expected sign?

One of the restrictions of the Maximum Likelihood Estimation of the Poisson Regression model is that the **mean of the counts must be equal to the variance of the counts**. ("Counts" here mean the number of installments per month.) But in actuality this assumption might not hold. Therefore we need to conduct some tests to see if there is **over- or under-dispersion** of the count data. The data exhibits **overdispersion** when the variance of the counts is greater than the mean of the counts. The data exhibits **underdispersion** when the variance of the counts is less than the mean of the counts.

(c) Conduct the **Cameron and Trivedi (1990) Diagnostic Test for over- or under-dispersion** in the model you estimated in part (b) above. What is the null hypothesis

of this test? What is the alternative hypothesis of this test? Draw the sampling distribution of the test statistic and the acceptance and rejection regions.

(d) Conduct the **Wooldridge (1996) Diagnostic Test for over- or under-dispersion** in the model you estimated in part (b) above. What is the null hypothesis of this test? What is the alternative hypothesis of this test? Draw the sampling distribution of the test statistic and the acceptance and rejection regions.

(e) Regardless of the outcomes of the test results of parts (c) and (d), estimate the Poisson Regression Model of part (b) above using the **GLM** (Generalized Linear Model) option. Do your results change very much from the Maximum Likelihood results you obtained in part (b)? (The GLM method is supposed to adjust the standard errors of the Maximum Likelihood estimates to accommodate over- or under-dispersion in the case that it exists.)

(f) Regardless of the outcomes of the test results of parts (c) and (d), estimate the Poisson Regression Model of part (b) above using the **Huber/White** option. Do your results change very much from the Maximum Likelihood results you obtained in part (b)? (The QML method is supposed to adjust the standard errors of the Maximum Likelihood estimates to accommodate over- or under-dispersion in the case that it exists.)

(g) Given the test results of parts (c) and (d) above, does it appear that the installments data exhibits over- or under-dispersion or is the Poisson assumption of equal mean and variance of the counts supported? To analyze the current data, which estimation technique would you use to test the significance of the Y2K and predation events that are claimed?

(h) Using your chosen Poisson equation, I want you to compare the number of actual installments with the expected number of installments that would have occurred **in the absence of the May 1996 predation event**. How many installations did the plaintiff firm (firm XYZ) lose during the period for which we have data? Carefully explain how you arrived at your answer?

(i) Suppose that the plaintiff firm (firm XYZ) loses $500,000 in profit for each installment lost due to the predatory activity of the defendant firm (firm RST). According to your calculations in part (h) above, what is the **expected** profit lost by the plaintiff firm? If the courts rule that treble damages is appropriate because the case qualifies as an antitrust case, what would the estimated damage be? Do you think is would be worth it for the plantiff firm to hire an expert econometrician at $500 per hour to do these calculations?