Eco 6352
Applied Econometrics

Professor Tom Fomby
Spring 2016

MIDTERM EXAM I

Name Mr. Key

Instructions: Put your name in the space above. This exam is scheduled for 1 hour and 20 minutes. It is worth 50 points with the questions carrying the following weights:

Q1: 3 pts.
Q2: 3
Q3: 3
Q4: 3
Q5: 3
Q6: 5
Q7: 5
Q8: 5
Q9 a) 3 b) 2 c) 3 = 8
Q10 a) 3 b) 3 c) 3 d) 3 = 12

1. Consider the following probability distribution: $P(y = j) = \frac{\exp((-\lambda)\lambda^j}{j!}$, j = 0,1,2,⋯ . This probability distribution serves as a basis for which conditional probability model that we have studied this semester? How is this pdf hyper-parameterized to make it a workable model?

3

This is the pdf that the Poisson Count Model is based on with $\lambda = \exp(X'\beta)$.

2. Consider the following probability distribution: $f(y;\pi) = \pi^y(1-\pi)^{1-y}, y = 0,1$. This probability distribution serves as a basis for which conditional probability model that we have studied this semester? How is this pdf hyper-parameterized to make it a workable model?

3

This is the pdf upon which the Logit and Probit binary choice models are built. For the Logit model $\Pi = \frac{\exp(X'\beta)}{1+\exp(X'\beta)}$. For the Probit model $\Pi = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx$

3. Consider the following probability distribution: $f(t;\lambda) = \lambda\exp(-\lambda t), t > 0$. This probability distribution serves as a basis for which conditional probability model that we have studied this semester? How is this pdf hyper-parameterized to make it a workable model?

3

This is the exponential pdf and serves as the basis for the exponential hazard (survival) model. The hyper-parametrization is $\lambda = \exp(X'\beta)$.

4. Name the 3 basic maximum likelihood based test procedures?

3

i) The Likelihood Ratio test
ii) The Wald Test
iii) The Score (Lagrangian) Test

5. Consider the following notations:

3

a) $\hat{\theta}_l \approx Normal(\theta_l, \hat{\sigma}_{ll})$ represents the Wald test

b) $2\left(logL(\hat{\theta}_u) - logL(\hat{\theta}_r)\right) \approx \chi^2_q$ represents the Likelihood Ratio test

c) $\frac{1}{\sqrt{n}} \cdot \sum_{i=1}^{n} s(\hat{\theta}_r; y_i) \approx \chi^2_q$ represents the Score test

Fill in the above blanks with the appropriate test label that you answered in question 4 above.

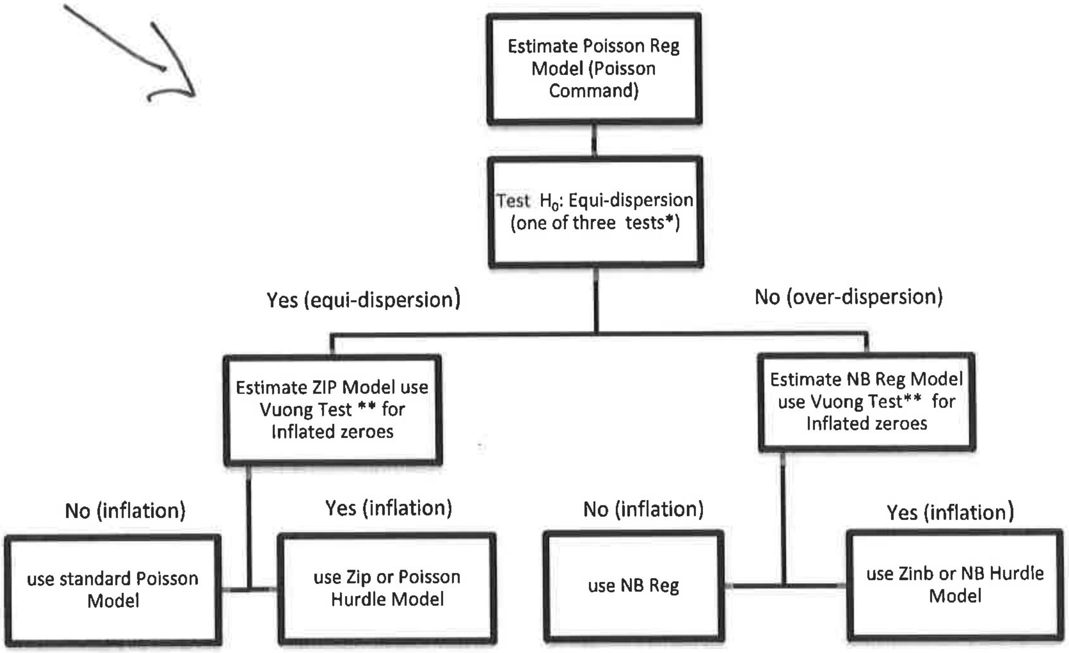
6. A few lectures ago I gave you a Flow (Decision) diagram (see Count Model Flow Chart.pdf) for how one might proceed in modeling count data. Reproduce that diagram as best you can in the space below:

5

See the accompanying diagram
That I previously sent to the class.

something like this is what I was looking for:

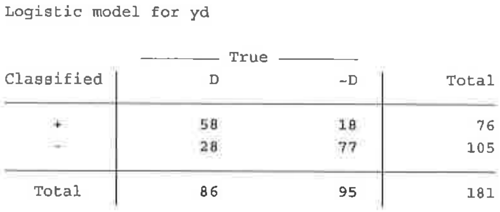
Flow Chart for Analyzing Count Data


Estimate Poisson Reg Model (Poisson Command)
Test H_0: Equi-dispersion (one of three tests*)
Yes (equi-dispersion)
No (over-dispersion)
Estimate ZIP Model use Vuong Test ** for Inflated zeroes
Estimate NB Reg Model use Vuong Test** for Inflated zeroes
No (inflation)
Yes (inflation)
No (inflation)
Yes (inflation)
use standard Poisson Model
use Zip or Poisson Hurdle Model
use NB Reg
use Zinb or NB Hurdle Model

* Cameron & Trevidi test (1990) and Wooldridge test (1996) involve the residuals of the standard Poisson model. On the other hand, the nbreg procedure is used to test the hypotheses H_0: $\alpha = 0$ (equi-dispersion) versus H_1: $\alpha \neq 0$. If $\alpha > 0$ then the Negative Binomial model is suggested. (Alternatively, one could produce robust standard errors for the Poisson regression coefficients using the sandwich variance-covariance matrix of the estimated coefficients (i.e. Quasi-Maximum Likelihood estimation)).

** Significant positive value indicates Zip or Zinb model is preferred. A significant negative value implies that the non-excess zeroes (standard) count models are preferred.

7. Fill in the following blanks using the following classification table based on a logit model for distressed firms (Exercise 2):



```
Logistic model for yd

              ──────── True ────────
Classified │        D            ~D  │      Total
───────────┼──────────────────────────┼───────────
     +     │       58            18  │         76
     -     │       28            77  │        105
───────────┼──────────────────────────┼───────────
   Total   │       86            95  │        181
```

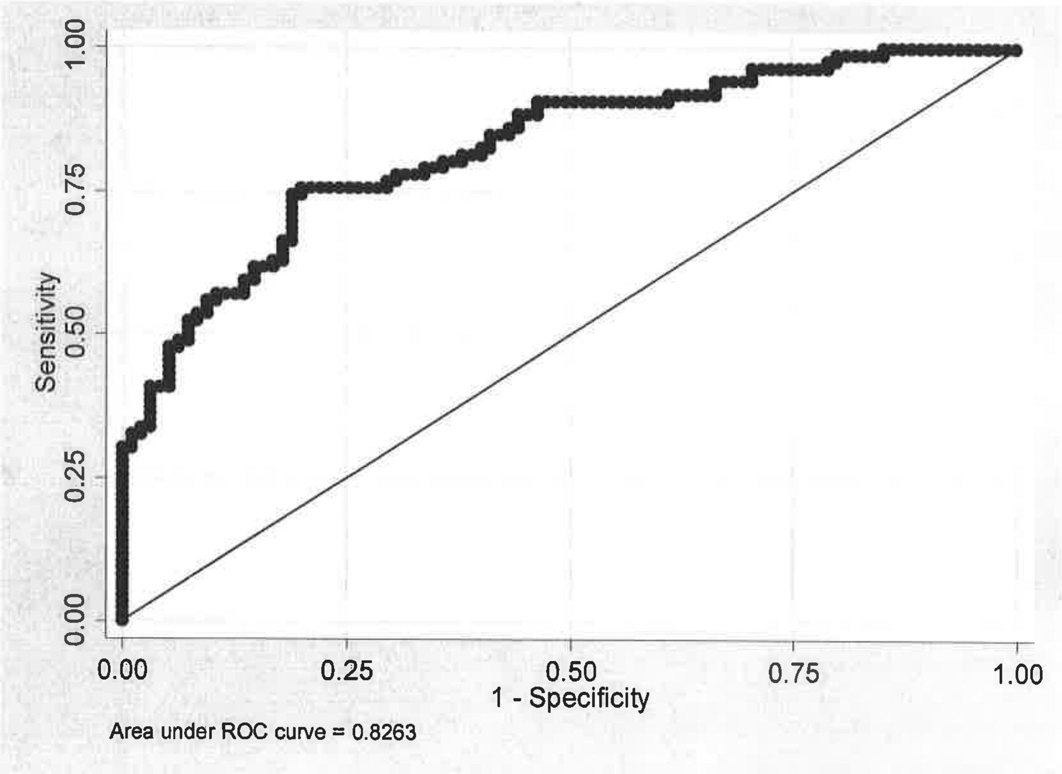
(1) Accuracy Rate = .746. (Show your work below.) $Acc = \frac{58+77}{181} = .746$

(1) Sensitivity Rate = .674. (Show your work below.) $\text{sensitivity} = \frac{58}{58+28} = .674$

(1) Specificity Rate = .810. (Show your work below.) $\text{specificity} = \frac{77}{18+77} = .810$

(2) If the cutoff probability is increased we know that Sensitivity will (increase/[✓ decrease]) and Specificity with ([✓ increase]/decrease). Choose the correct alternatives.

8. Consider the following curve that was produced in Exercise 2 (the distressed firm logit model).



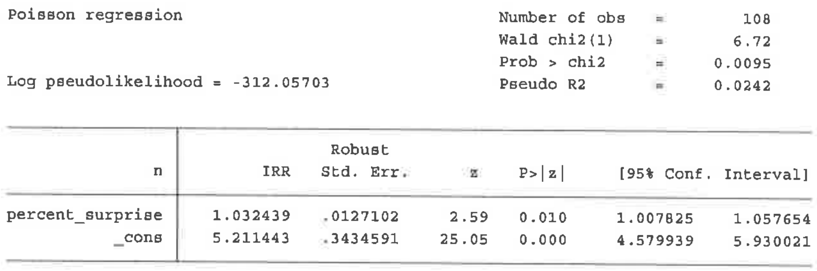
45° line = RoC of Naive Predicter

Two purposes of RoC curves: 1) To determine which classifier is best among a group of competing classiers (the one with the largest area under the RoC is best and 2) to determine an optimal cut-off probability.

What does the dark curve represent? What does the 45-degree line represent? Name at least two usages of this curve and briefly tell me how the curve is generated.

5 The dark curve represents the RoC curve. It is the points (1-spec., sens.) obtained by recording the 1-spec. and sens. nos. for a succession of classification tables generated by using a series of cut-off probabilities from 0.01 → 0.99.

9. Consider the below output concerning the strike count in the U.S. economy as it relates to the strength of the business cycle.



```
Poisson regression                              Number of obs     =        108
                                                Wald chi2(1)      =       6.72
                                                Prob > chi2       =     0.0095
Log pseudolikelihood = -312.05703               Pseudo R2         =     0.0242

------------------------------------------------------------------------------
                 |               Robust
               n |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
percent_surprise |   1.032439   .0127102     2.59   0.010     1.007825    1.057654
           _cons |   5.211443   .3434591    25.05   0.000     4.579939    5.930021
------------------------------------------------------------------------------
```

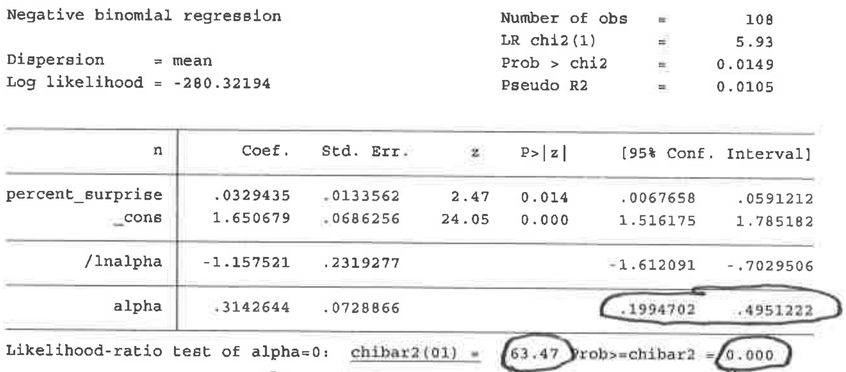
a) Explain to me the meaning of the IRR number associated with the explanatory variable percent_surprise. 1.032439 ⇒ For every one percent increase

3 in percent-surprise, the expected number of strikes increases by 3.2 percent.

b) Why do we choose to report the robust standard errors of this model rather than the usual Maximum Likelihood standard errors?

2 More than likely the count data doesn't have equi-dispersion and without robustifying the ~~the~~ standard errors of the estimates we are likely to have inconsistent statistical inference.

c) Given the following output, which model would you prefer for modeling the strike counts? The Poisson model or the Negative Binomial model. Thoroughly explain your answer.



```
Negative binomial regression                    Number of obs     =        108
                                                LR chi2(1)        =       5.93
Dispersion     = mean                           Prob > chi2       =     0.0149
Log likelihood = -280.32194                     Pseudo R2         =     0.0105

------------------------------------------------------------------------------
               n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
percent_surprise |   .0329435   .0133562     2.47   0.014     .0067658    .0591212
           _cons |   1.650679   .0686256    24.05   0.000     1.516175    1.785182
-----------------+------------------------------------------------------------
        /lnalpha |  -1.157521   .2319277                     -1.612091   -.7029506
-----------------+------------------------------------------------------------
           alpha |   .3142644   .0728866                      .1994702    .4951222
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =  63.47 Prob>=chibar2 = 0.000
```

This is the "Alpha" test for overdispersion. H_0: The counts exhibit equi-dispersion (use Poisson)

H_1: The counts exhibit over-dispersion (use NB)

3

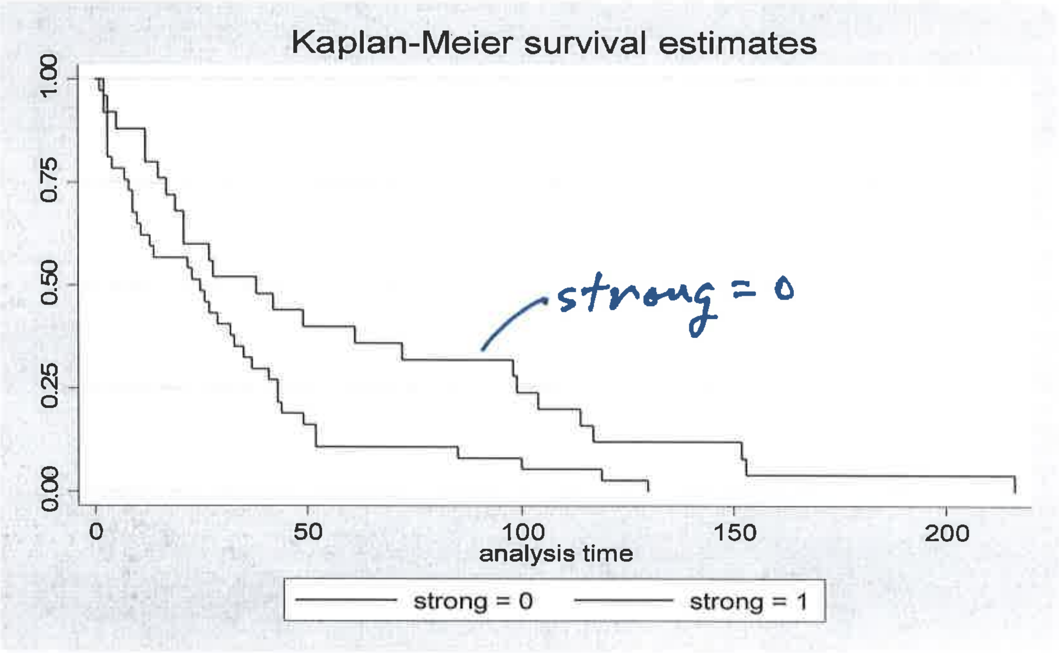
Since the 95% C.I. for Alpha does not encompass zero we conclude that there is over-dispersion in the count data and, therefore, we need to use the NB model in our count analysis. Equivalently, the LR test has p = 0.000.

10. Recall Exercise 4 where we examined the duration of strikes in the U.S. economy as a function of the strength of the economy at the time the strike began.

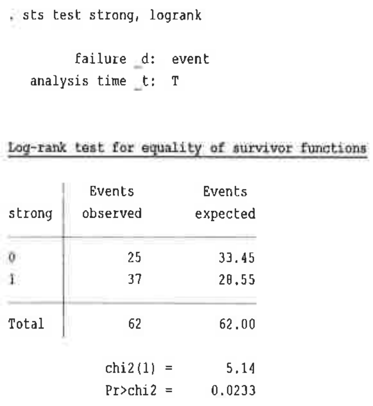
3

a) Consider the following graph. Explain to me the meaning of the graph.

For the two regimes (strong = 0, 1) these show the probability of surviving up to a certain time t. During a weak economy, strikes last longer than during strong economic times.



b) Consider the following STATA output concerning the duration of U.S. strikes:



```
. sts test strong, logrank

         failure _d:  event
   analysis time _t:  T

Log-rank test for equality of survivor functions

       |   Events         Events
strong |  observed       expected
-------+-------------------------
0      |        25          33.45
1      |        37          28.55
-------+-------------------------
Total  |        62          62.00

             chi2(1) =       5.14
             Pr>chi2 =     0.0233
```

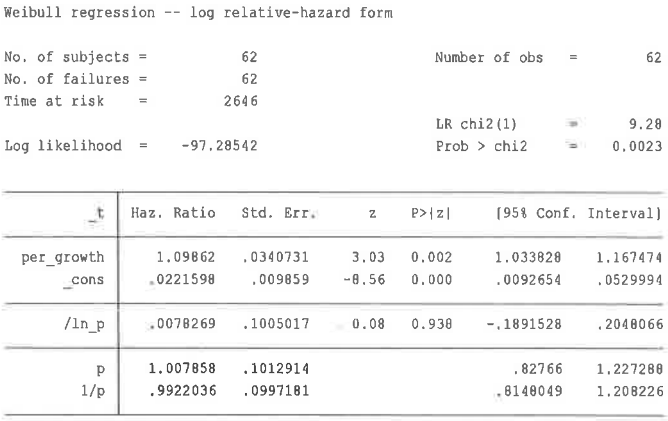
What is implied by this output?

3

The null hypothesis of the logrank test is that there is no statistical difference in the survival curves. The alternative hypothesis is that the two survival curves are statistically different.

Given that the p-value of the test statistic is less than 0.05 we reject the H_0 and accept H_1 that the survival curves are significantly different from each other.

c) Consider the following STATA output concerning the duration of U.S. strikes:



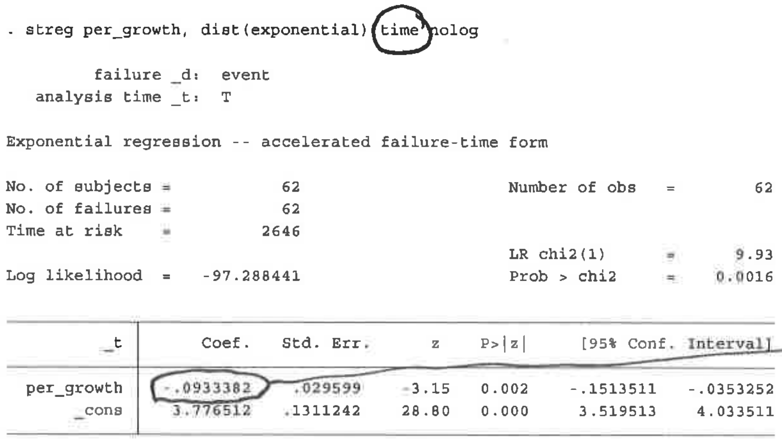
```
Weibull regression -- log relative-hazard form

No. of subjects =          62                  Number of obs   =        62
No. of failures =          62
Time at risk    =        2646
                                               LR chi2(1)      =      9.28
Log likelihood  =   -97.28542                  Prob > chi2     =    0.0023

------------------------------------------------------------------------------
          _t |  Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  per_growth |    1.09862    .0340731     3.03   0.002     1.033828    1.167474
       _cons |   .0221598     .009859    -8.56   0.000     .0092654    .0529994
-------------+----------------------------------------------------------------
        /ln_p |   .0078269    .1005017     0.08   0.938    -.1891528    .2048066
-------------+----------------------------------------------------------------
           p |   1.007858    .1012914                         .82766    1.227288
         1/p |   .9922036    .0997181                       .8148049    1.208226
------------------------------------------------------------------------------
```

Given this output, which duration model do you prefer? The Weibull or Exponential? Thoroughly explain your reasoning. Test $H_0: p = 1$ vs. $H_1: p \neq 1$.

3

If H_0 is supported we accept that the weibull model can be simplified to the Exponential model. Since the 95% C.I. of the p statistic encompasses 1 we accept H_0. The Exponential model is the model of choice.

d) Consider the following STATA output concerning the duration of U.S. strikes:



```
. streg per_growth, dist(exponential) time nolog

         failure _d:  event
   analysis time _t:  T

Exponential regression -- accelerated failure-time form

No. of subjects =           62                   Number of obs   =        62
No. of failures =           62
Time at risk    =         2646
                                                 LR chi2(1)      =      9.93
Log likelihood  =   -97.288441                   Prob > chi2     =    0.0016

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  per_growth |  -.0933382     .029599    -3.15   0.002    -.1513511   -.0353252
       _cons |   3.776512    .1311242    28.80   0.000     3.519513    4.033511
------------------------------------------------------------------------------
```

Remember: This "time ratio" is already in exp(b) - 1 form.

Explain to me the meaning of the coefficient on the variable per_growth. Per_growth represents the growth surprise in the U.S. economy measured in percent terms.

3

What is being modeled here is the expected duration of U.S. strikes as a function of the strength of the economy. The coefficient estimate of -.0933 implies that for every percent increase in growth surprise we expect the expected duration of a strike to fall by 9.3%