# Advanced Econometrics 1
# Computer Lab Exercise Week 9:
# Actual Size of LR, LM, and Wald tests

In this computer exercise we investigate the actual size properties of the likelihood ratio, Lagrange-multiplier and Wald tests for a restriction in a Poisson regression. In particular, we look at the effect of a reformulation of the null hypothesis on the Wald test.

The exercise is based on Section 7.4 in Cameron & Trivedi, where they consider a dataset of 200 (simulated) observations on three regressors $x_2$, $x_3$ and $x_4$, and a dependent count (non-negative integer-valued) variable $y$.

The data are collected in the text file `Chapter7data.txt`. As explained in Section 7.4, the regressors $\mathbf{x}$ are independent standard normal, and $y|\mathbf{x}$ follows a Poisson distribution with parameter $\lambda(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}'\boldsymbol{\beta})$, where $\mathbf{x} = (1, x_2, x_3, x_4)'$ and the true value $\boldsymbol{\beta}_0$ is $\frac{1}{10}(0, 1, 1, 1)'$.

The density for observation $i$ is:

$$pdf\ (y_i \mid x_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$$

with $\lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$. This implies that the (log-) likelihood function for $N$ independent observations equals:

$$\ell(\beta \mid y; X) = \sum_{i=1}^{N} -\exp\left\{x_i'\beta\right\} + y_i x_i'\beta - \ln(y_i!),$$

but the last term could be dropped as it does not depend on parameters.

We will be interested in testing the hypothesis $H_0 : \beta_3 = \beta_4$.

As a strarting point, you can use the Matlab *script file* `CompLabEx5.m` that uses two Matlab *function files* `loglikpoisson.m` and `rbr.m`. All these files should be in the **same** folder on your network drive.

- `loglikpoisson.m` : calculates minus the log-likelihood and score vector and the Hessian for the Poisson regression model.
- `CompLabEx5.m` :
  first part reproduces some of the results in Table 7.1.
  second part is a Monte Carlo experiment, which initially has been commented out.
  Both parts of the program will need to be completed.
- `Chapter7data.txt:` contains relevant data
- `rbr.m` : does row-by-row multiplication

0. Start the Matlab program via the Start menu in Windows, open the file `CompLabEx5.m` via the menu bar in Matlab, and run it. The program first loads the data from the text file, and creates an $N \times 1$ dependent vector $y$ and an $N \times 4$ regressor matrix $X$.
   Next, it estimates the parameters via maximum likelihood, first unconstrained, and then under the restriction
   $$H_{30} : \beta_3 - \beta_4 = 0,$$
   Show that this restriction can be been incorporated in the model by replacing
   $$\mathbf{x}'\boldsymbol{\beta} = \beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4$$

by

$$\mathbf{x}_r'\boldsymbol{\beta}_r = \beta_1 + x_2\beta_2 + (x_3 + x_4)\beta_3.$$

Therefore the matrix $Xr$ has three columns, consisting of observations on $x_1 = 1$, $x_2$ and $(x_3 + x_4)$. Using these 3 variables only, we can estimate the restricted model (under $H_0$).
The program calculates the unrestricted likelihood, MLE's, together with standard errors and $t$-statistics.
It then calculates the restricted likelihood and likelihood ratio (LR) test statistic. It prints these to the Command Window, together with a $p$-value from the $\chi^2(1)$ distribution. (what is the justification?).
[The `num2str` function in Matlab converts a number to a string, which is useful in `disp` commands, printing results to the Command Window; see p. 6 of "Getting Started with Matlab".]

1. Extend the program to calculate the Wald test statistic.
   The unrestricted MLE `b` and its estimated variance matrix `Vb` are already computed by the program. Therefore, all you need to do is to apply the formula

   $$\mathrm{W} = h(\widehat{\boldsymbol{\beta}}_u)' \left(\widehat{\mathbf{r}}\ \widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_u]\ \widehat{\mathbf{r}}'\right)^{-1} h(\widehat{\boldsymbol{\beta}}_u).$$

   Show that for

   $$H_{30} : \beta_3 - \beta_4 = 0$$

   this becomes

   $$\mathrm{W}_{30} = \frac{(\widehat{\beta}_3 - \widehat{\beta}_4)^2}{\widehat{\mathbf{r}}\ \widehat{\mathsf{V}}[\widehat{\boldsymbol{\beta}}_u]\widehat{\mathbf{r}}'}$$

   where you still need to determine the $1 \times 4$ vector $\mathbf{r} = \partial h(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$, with $h(\boldsymbol{\beta}) = \beta_3 - \beta_4$.
   Print the result to the Command Window, together with its $p$-value, analogous to the LR test.

2. Now do the same for the Wald test statistic for

   $$H_{40} : \beta_3/\beta_4 - 1 = 0.$$

   The two hypotheses $H_{30}$ and $H_{40}$ are equivalent (except that $H_{40}$ excludes parameter values with $\beta_3 = \beta_4 = 0$, but that can be neglected here) so the likelihood ratio test does not need to be recalculated. The Wald test, however, is not invariant to such equivalent reformulations, so the result will be different. Again, to calculate $\mathrm{W}_{40}$, say, you will have to determine $\mathbf{r} = \partial h(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$, where now $h(\boldsymbol{\beta}) = \beta_3/\beta_4 - 1$ and in this case $\mathbf{r}$ does depend on the parameters, so it will have to be evaluated at $\widehat{\boldsymbol{\beta}}_u$. Again, print the result to the Command Window, together with the (asymptotic) $p$-value.

3. Next, calculate the LM test statistics for $H_{30}$ and hence $H_{40}$ (they yield the same LM statistic). Use the score version of the test:

   $$\mathrm{LM} = \mathbf{s}_N(\widetilde{\boldsymbol{\beta}}_r)'(\widetilde{\mathcal{I}}_N)^{-1}\mathbf{s}_N(\widetilde{\boldsymbol{\beta}}_r).$$

   The restricted MLE $\widetilde{\boldsymbol{\beta}}_r$ already has been computed and the score vector:

   $$\mathbf{s}_N(\widetilde{\boldsymbol{\beta}}_r) = \mathbf{X}'(\mathbf{y} - \widetilde{\boldsymbol{\lambda}}),$$

   is easily calculated with $\widetilde{\boldsymbol{\lambda}}$ the $N \times 1$ vector with observations on $\widetilde{\lambda}_i = \exp(\mathbf{x}_i'\widetilde{\boldsymbol{\beta}}_r)$; in Matlab this is easily computed as `exp(Xr*br)`.
   The information matrix, evaluated at $\widetilde{\boldsymbol{\beta}}_r$ is given by

   $$\widetilde{\mathcal{I}}_N = \sum_{i=1}^{N} \widetilde{\lambda}_i \mathbf{x}_i \mathbf{x}_i' = \mathbf{X}' \operatorname{diag}(\widetilde{\boldsymbol{\lambda}})\mathbf{X}.$$

This can be done in Matlab using the `rbr` function, see also Computer Lab Exercise 3. With these ingredients, you can compute the test. Again, print the result to the Command Window, together with the approximate $p$-value.

4. The p-values have been calculated based on the asymptotic approximation to the distribution of the teststatistic. We only have a finite sample, 200 observations in this case, and the approximation might not be very good. We therefore wish to explore the behaviour of these four tests (LR-, two versions of the Wald- and the LM-test) in finite samples by a Monte Carlo simulation experiment.

   Uncomment the second part of the program (i.e., remove `%{` and `%}`). In this part, for each of the `R` ($=1000$) replications, we generate $y$ from the Poisson distribution with intensity vector `lambda0 = exp(X*beta0)`, and then recalculate the LR statistic for this realization of $y$, which is then saved as the $i^{th}$ component of the vector `LR`. After the replications, we calculate the percentage of rejections (which is the *actual* significance level, because the true value satisfies $H_{30}\&H_{40}$) by `mean(LR>c)`, where `c` is the 5% critical value from the $\chi^2(1)$ distribution, and where `(LR>c)` is a vector of zeros and ones (depending on whether `LR(i)<c(i)` or `LR(i)>c(i)`), so that its average is the percentage of rejections.

   Create $R \times 1$ vectors of zeros `W30`, `W40` and `LM` before the loop (for the two versions of the Wald test and the LM test), and recalculate each of these statistics inside the loop, copying the relevant parts of the code that you have created above. Finally, compute the Monte Carlo significance levels for each the tests. What you should find (after running the resulting program) is that LR, W30 and LM have a significance level close to the nominal 5%, but the W40 test (the Wald test for $H_{40}$) has a rejection frequency close to 10% instead of the nominal level.

5. How would you determine an appropriate critical value based on the MC simulations?

6. How would you investigate the power of the tests?