5. Count data models

5.1 Poisson regression model

Count data variables as dependent variables in a microeconometric analysis:

These quantitative variables are discrete, restricted to non-negative integers, and refer to events within a fixed time interval. Since they are not qualitative, one might think to use linear regression models. However, due to the dominance of zeros and small values in many studies and the discrete nature of these variables, the analysis can be improved since, for example, OLS estimations in linear regression models are generally inconsistent if the data generating process follows a Poisson regression model as discussed below.

Examples for microeconometric analyses with count data models:

- Analysis of the number of visits of a person to a hospital within one year
- Analysis of the number of journeys of a household within five years
- Analysis of the number of incidents of an airline per 1000 scheduled departures within one year
- Analysis of the number of patents of a firm within three years
- Analysis of the number of strikes in an industry within one year
- Analysis of the number of adaptation measures of ski lift operators in response to climate change

Similar to binary probit and logit models in the case of binary response models, the multinomial logit model in the case of multinomial response models, and ordered probit and logit models in the case of ordered response models, the Poisson regression model is the most prominent and important approach in the class of count data models. It is based on the assumption that the dependent variable y_i can take the values j = 0, 1, 2,… and is Poisson distributed with parameter λ_i which is related to the vector $x_i = (x_{i1},\ldots, x_{ik})'$ of k explanatory variables (including a constant) and the corresponding k-dimensional parameter vector $\beta = (\beta_1,\ldots, \beta_k)'$. The probability function of y_i is as follows (i = 1,…, n):

$$f_i(y_i; x_i, \beta) = P(y_i = j|x_i, \beta) = \frac{e^{-\lambda_i}\lambda_i^{\,j}}{j!} \quad \text{for } j = 0, 1, 2,...$$

The most common formulation of λ_i is:

$$\log\lambda_i = \beta' x_i \quad \text{or} \quad \lambda_i = e^{\beta' x_i}$$

It follows for the probability function of y_i (i = 1,…, n):

$$f_i(y_i; x_i, \beta) = P(y_i = j|x_i, \beta) = \frac{e^{-e^{\beta' x_i}} e^{j\beta' x_i}}{j!} \quad \text{for } j = 0, 1, 2,...$$

The assumption of a Poisson distributed random variable necessarily implies the equality of the expected value and the variance. In the case of the Poisson regression model it follows:

$$E(y_i|x_i, \beta) = Var(y_i|x_i, \beta) = \lambda_i = e^{\beta' x_i}$$

This equality implies the equidispersion property of Poisson regression models which is in many cases of overdispersion or underdispersion in count data models (see later) problematic. The partial derivative of this expected value with respect to an explanatory variable x_{ih} is (see also later):

$$\frac{\partial E(y_i|x_i, \beta)}{\partial x_{ih}} = \frac{\partial e^{\beta' x_i}}{\partial x_{ih}} = e^{\beta' x_i}\beta_h = E(y_i|x_i, \beta)\beta_h$$

On the basis of the probabilities as discussed above, the k parameters $\beta_1,\ldots, \beta_k$ in the Poisson regression model can be estimated by ML. Based on a random sample (x_i, y_i) for i = 1,…, n observations, the log-likelihood function therefore is:

$$\log L(\beta) = \sum_{i=1}^{n}(-\lambda_i + \log\lambda_i^{y_i} - \log y_i!) = \sum_{i=1}^{n}(-e^{\beta' x_i} + y_i\beta' x_i - \log y_i!)$$

The score has the following form:

$$\frac{\partial \log L(\beta)}{\partial \beta} = s(\beta) = \sum_{i=1}^{n} s_i(\beta) = \sum_{i=1}^{n}(y_i - e^{\beta' x_i})x_i$$

The ML estimator again solves the first-order conditions for maximizing the log-likelihood function. Thus by equalizing the score with zero it follows:

$$\hat{\beta} = \arg\underset{\beta}{\text{solves}}\left[\sum_{i=1}^{n} s_i(\beta) = \sum_{i=1}^{n}(y_i - e^{\beta' x_i})x_i = 0\right]$$

It can be shown that the log-likelihood function in Poisson regression models is globally concave so that the iterative optimization process leads to a unique (global) maximum (and not a minimum).

The ML estimators $\hat{\beta}_h$ can again not be interpreted as in the linear regression model, i.e. not as estimators of the effect of the respective explanatory variable x_{ih}. However, one possible interpretation of the parameter estimators in Poisson regression models refers to the estimator of (partial) marginal or discrete mean effects. The estimator of a (partial) marginal mean effect refers to the estimated (partial) effect of a (continuous) explanatory variable x_{ih} on the expected value of the variable y_i (in line with the respective estimator of the marginal effect in the linear regression model, which is $\hat{\beta}_h$) (i = 1,…, n):

$$\frac{\partial E(y_i|x_i, \hat{\beta})}{\partial x_{ih}} = e^{\hat{\beta}' x_i}\hat{\beta}_h$$

Therefore, $\hat{\beta}_h$ indeed indicates the direction of this estimated effect of x_{ih}, but the extent depends on x_i. If this value is divided by $\exp(\hat{\beta}' x_i)$ it follows:

$$\hat{\beta}_h = \frac{\partial E(y_i|x_i, \hat{\beta})}{\partial x_{ih}} \frac{1}{e^{\hat{\beta}' x_i}} = \frac{\partial E(y_i|x_i, \hat{\beta})}{\partial x_{ih}} \frac{1}{E(y_i|x_i, \hat{\beta})} = \frac{\partial \log E(y_i|x_i, \hat{\beta})}{\partial x_{ih}}$$

Thus, $\hat{\beta}_h$ can be interpreted as an estimated semi-elasticity, i.e. if x_{ih} increases by one unit, $100\hat{\beta}_h$ indicates the approximately estimated percentage change of $E(y_i|x_i, \beta)$. In addition, if x_{ih} is in logarithmic form, $\hat{\beta}_h$ is an estimated elasticity.

The estimator of a discrete mean effect, i.e. the estimator of a discrete change of $E(y_i|x_i, \beta)$ due to a discrete change Δx_{ih} of an explanatory variable x_{ih} in Poisson regression models is:

$$\Delta E(y_i|x_i, \hat{\beta}) = e^{\hat{\beta}'x_i + \hat{\beta}_h \Delta x_{ih}} - e^{\hat{\beta}'x_i}$$

However, the common interpretation of the $\hat{\beta}_h$ also in Poisson regression models refers to estimated marginal and discrete probability effects. While the estimation of discrete probability effects is again based on differences in estimated probabilities, it follows for the estimator of (partial) marginal probability effects:

$$\frac{\partial P(y_i = j|x_i, \hat{\beta})}{\partial x_{ih}} = P(y_i = j|x_i, \hat{\beta})\left[j - e^{\hat{\beta}'x_i} \right]\hat{\beta}_h$$

Interpretation:

- For $j < \exp(\hat{\beta}'x_i)$: A positive (negative) $\hat{\beta}_h$ implies a decrease (increase) of $P(y_i = j|x_i, \hat{\beta})$ if x_{ih} increases. For $j > \exp(\hat{\beta}'x_i)$: A positive (negative) $\hat{\beta}_h$ implies an increase (decrease) of $P(y_i = j|x_i, \hat{\beta})$ if x_{ih} increases.
- The sign of the estimated probability effects is thus either first negative and changes to a positive value or first positive and changes to a negative value when y_i moves from small to high values and crosses $\exp(\hat{\beta}'x_i)$, which is similar to the corresponding property in ordered probit and logit models

→ On this basis, it is again possible to estimate average marginal and discrete probability effects of an explanatory variable x_{ih} across all i as well as marginal and discrete effects of x_{ih} at the mean of the explanatory variables

Example: Determinants of the fertility of women (I)

By using a Poisson regression model, the effect of the following explanatory variables on the number of children ever born (kids) for 5150 women aged 40 years or older in the US is examined on the basis of pooled cross-sectional data from the US General Social Survey during the period 1974-2002 (with four-year intervals):

- Years of education (educ) as mainly interesting explanatory variable
- Dummy variable for the race (white) that takes the value one if the woman is white
- Dummy variable for the immigrant status (immigrant) that takes the value one if the woman or both parents of the woman were born abroad
- Dummy variable for low income in the youth (lowincome16) that takes the value one if the income of the woman was below average income at the age of 16
- Dummy variable for living in the city in the youth (city16) that takes the value one if the woman lived in a city at the age of 16
- A linear time trend (time) in order to control for fertility changes over time

The ML estimation of the Poisson regression model with STATA leads to the following results:

Example: Determinants of the fertility of women (II)

```
poisson kids educ white immigrant lowincome16 city16 time

Poisson regression                              Number of obs     =      5,150
                                                LR chi2(6)        =     484.28
                                                Prob > chi2       =     0.0000
Log likelihood = -10116.635                     Pseudo R2         =     0.0234

------------------------------------------------------------------------------
        kids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0441974   .0028098   -15.73   0.000    -.0497046   -.0386902
       white |  -.1365916   .0229592    -5.95   0.000    -.1815909   -.0915924
   immigrant |  -.0808302   .0276367    -2.92   0.003    -.1349972   -.0266632
 lowincome16 |   .0114575   .0211581     0.54   0.588    -.0300117    .0529266
      city16 |   -.055456   .0190296    -2.91   0.004    -.0927534   -.0181586
        time |  -.0218599   .0040518    -5.40   0.000    -.0298012   -.0139186
       _cons |   1.708038   .0387383    44.09   0.000     1.632112    1.783963
------------------------------------------------------------------------------
```

Example: Determinants of the fertility of women (III)

Interpretation:

- The value of 484.28 of the likelihood ratio test statistic means that the null hypothesis that all six parameters of the explanatory variables are zero (which would imply that no explanatory variable has an effect on the number of children ever born) can be rejected at any common significance level
- The parameter estimate for educ is negative and highly significantly different from zero due to the z statistics of -15.73 and thus implies that education has a strong significantly negative effect on the fertility of women
- The parameter estimate of -0.0442 implies that an increase of the years of education by one (unit) leads to an approximately estimated decrease of the expected (or mean) number of children ever born by $100 \cdot 0.0442 = 4.42\%$
- Similarly, white women, women with a migration background, and women who lived in a city at the age of 16 have a significantly lower fertility
- Furthermore, the time trend is highly significantly negative which implies strong decreasing numbers of children ever born over time (due to the inclusion of the time trend, spurious effects can be avoided and the effects of the other explanatory variables can be considered detrended)

Example: Determinants of the fertility of women (IV)

Wald and likelihood ratio tests:

As an example, the null hypothesis that neither white nor immigrant nor lowincome16 has any effect on the number of children ever born, i.e. that the three corresponding parameters are zero, is tested. The command for the Wald test in STATA is:

```
test white immigrant lowincome16

 ( 1)  [kids]white = 0
 ( 2)  [kids]immigrant = 0
 ( 3)  [kids]lowincome16 = 0

           chi2(  3) =   46.68
         Prob > chi2 =    0.0000
```

The corresponding commands for the likelihood ratio test in STATA are then:

```
estimates store unrestricted

poisson kids educ city16 time

estimates store restricted

lrtest unrestricted restricted

Likelihood-ratio test                                 LR chi2(3)  =     45.84
(Assumption: restricted nested in unrestricted)       Prob > chi2 =    0.0000
```

Example: Determinants of the fertility of women (V)

The estimation of the average marginal probability effects of educ across all 5150 observations for one, two, and three children ever born leads to the following (shortened) STATA results:

```
margins, dydx(educ) predict(pr(1))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0128414   .0007696    16.69   0.000      .011333    .0143499
------------------------------------------------------------------------------

margins, dydx(educ) predict(pr(2))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0057267   .0003731    15.35   0.000     .0049955     .006458
------------------------------------------------------------------------------

margins, dydx(educ) predict(pr(3))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0035619   .0002723   -13.08   0.000    -.0040957   -.0030281
------------------------------------------------------------------------------
```

Example: Determinants of the fertility of women (VI)

The estimation of the marginal probability effects of educ at the means of the explanatory variables across all 5150 observations for one, two, and three children ever born leads to the following (shortened) STATA results:

```
margins, dydx(educ) atmeans predict(pr(1))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0136351   .0008653    15.76   0.000     .0119392     .015331
------------------------------------------------------------------------------

margins, dydx(educ) atmeans predict(pr(2))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0060802   .0004203    14.47   0.000     .0052565    .0069039
------------------------------------------------------------------------------

margins, dydx(educ) atmeans predict(pr(3))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0043719   .0003601   -12.14   0.000    -.0050776   -.0036662
------------------------------------------------------------------------------
```

Example: Determinants of the fertility of women (VII)

Interpretation:

- The estimated average marginal probability effects of educ and the estimated marginal probability effects at the means of the explanatory variables are very similar
- The estimated average marginal probability effects of 0.0128, 0.0057, and -0.0036 (which are strongly significantly different from zero) imply that an increase of the years of education by one (unit) leads to an approximately estimated increase of the probability of one child or two children ever born by 1.28 or 0.57 percentage points as well as an approximately estimated decrease of the probability of three children ever born by 0.36 percentage points
- Since the (not reported) estimated average marginal probability effect and estimated marginal probability effect at the means of the explanatory variables for zero children are positive and the corresponding estimates for more than three children are negative, the results imply that the sign of the estimated probability effects is first positive for $y_i \leq 2$ and then becomes negative for $y_i > 2$
- All these estimation results are in line with the significantly negative effect of educ on kids

---

Example: Determinants of the fertility of women (VIII)

The estimation of the average probability that kids = 1 for white = 1 and for white = 0 across all 5150 observations (as basis for the estimation of the average discrete probability effect of race) leads to the following STATA results:

```
margins, at(white=1) predict(pr(1))

Predictive margins                                Number of obs   =       5150
Model VCE    : OIM

Expression   : Pr(kids=1), predict(pr(1))
at           : white           =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .2062629   .0028203    73.14   0.000     .2007352    .2117905
------------------------------------------------------------------------------

margins, at(white=0) predict(pr(1))

Predictive margins                                Number of obs   =       5150
Model VCE    : OIM

Expression   : Pr(kids=1), predict(pr(1))
at           : white           =           0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .166371   .0060663    27.43   0.000     .1544812    .1782608
------------------------------------------------------------------------------
```

Example: Determinants of the fertility of women (IX)

In contrast, the estimation of the probability that kids = 5 for white = 1 and for white = 0 at the means of the other explanatory variables across all 5150 observations (as basis for the estimation of the discrete probability effect of race at the mean of the explanatory variables) leads to the following STATA results:

```
margins, at((means)_all white=1) predict(pr(5))

Adjusted predictions                                Number of obs   =       5150
Model VCE    : OIM

Expression   : Pr(kids=5), predict(pr(5))
at           : educ            =    12.24485 (mean)
               white           =           1
               immigrant       =    .1221359 (mean)
               lowincome16     =    .2267961 (mean)
               city16          =    .3733981 (mean)
               time            =    16.74796 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .065879   .0016054    41.04   0.000     .0627325    .0690256
------------------------------------------------------------------------------
```

Example: Determinants of the fertility of women (X)

```
margins, at((means)_all white=0) predict(pr(5))

Adjusted predictions                              Number of obs   =       5150
Model VCE    : OIM

Expression   : Pr(kids=5), predict(pr(5))
at           : educ            =    12.24485 (mean)
               white           =           0
               immigrant       =    .1221359 (mean)
               lowincome16     =    .2267961 (mean)
               city16          =    .3733981 (mean)
               time            =    16.74796 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0906405   .0040638    22.30   0.000     .0826757    .0986054
------------------------------------------------------------------------------
```

These results are in line with the significantly negative effect of white on kids:

- The estimated average probability that a white woman has ever born one child is 100·(0.2063-0.1664) = 3.99 percentage points higher than the corresponding estimate for a non-white woman
- The estimated probability at the means of the other explanatory variables that a white woman has ever born five children is 100·(0.0906-0.0659) = 2.47 percentage points lower than the estimate for a non-white woman

5.2 Negbin regression models

Problems of Poisson regression models:

- The fundamental problem is that the Poisson distribution is parameterized only by the single parameter λ_i so that all moments of the dependent variable y_i are a function of λ_i
- One frequent consequence is that the Poisson regression model often estimates the probability that the dependent variable y_i takes the value zero too small compared with the actual values in the sample (excess zero problem) (in this case more general count data models such as the hurdle count data model or zero-inflated count data model can be applied)
- Another frequent consequence of this problem is that the variance of the dependent variable y_i is higher than its expected value (overdispersion), whereas the Poisson regression model implies equidispersion, i.e. the equality of expected value and variance
- In the case of overdispersion, the ML estimators in the Poisson regression are inefficient and the standard deviations of the estimated parameters are inconsistently estimated (see later)
- Furthermore, overdispersion can be a signal for even stronger misspecifications in the Poisson regression model
- The estimation of the probabilities requires additional parameters in the case of overdispersion

Overdispersion in count data models is often caused by unobserved heterogeneity across observations which is not included in the Poisson regression model since the parameter λ_i only comprises the explanatory variables in x_i and the parameter vector β. However, a corresponding unobserved heterogeneity parameter ε_i (with $u_i = e^{\varepsilon_i}$) can be included in λ_i as follows:

$$\log\lambda_i = \beta' x_i + \varepsilon_i \quad \text{or} \quad \lambda_i = e^{\beta' x_i + \varepsilon_i} = e^{\beta' x_i} e^{\varepsilon_i} = e^{\beta' x_i} u_i$$

If it is again assumed that the dependent variable is Poisson distributed, however, conditional not only on x_i and β, but also on the unobservable error term u_i, it follows for the conditional probability function of y_i (i = 1,…, n):

$$f_i(y_i; x_i, \beta, u_i) = P(y_i = j | x_i, \beta, u_i) = \frac{e^{-e^{\beta' x_i} u_i} (e^{\beta' x_i} u_i)^j}{j!}$$

The conditional expected value and variance of y_i is now $\lambda_i = \exp(\beta' x_i) u_i$ instead of $\lambda_i = \exp(\beta' x_i)$ in the Poisson regression model. For the unconditional probability function, the unobservable error term u_i has to be integrated out, so that it follows for specific density functions $g(u_i)$:

$$f_i(y_i; x_i, \beta) = \int_0^{\infty} \frac{e^{-e^{\beta' x_i} u_i} (e^{\beta' x_i} u_i)^j}{j!} g(u_i) du_i$$

Usually it is assumed that u_i is gamma distributed with the two inherent parameters θ > 0 and γ > 0. With the restriction θ = γ in order to reach the normalization $E(u_i) = 1$, it follows for the probability function:

$$f_i(y_i;\, x_i,\, \beta) = \int_0^{\infty} \frac{e^{-e^{\beta' x_i} u_i} (e^{\beta' x_i} u_i)^{y_i}}{y_i!} \frac{\theta^{\theta} u_i^{\theta-1} e^{-\theta u_i}}{\Gamma(\theta)} du_i$$

$\Gamma(\theta)$ denotes the gamma function. After several transformations it follows:

$$f_i(y_i;\, x_i,\, \beta) = P(y_i = j|x_i,\, \beta) = \frac{\Gamma(y_i + \theta)}{\Gamma(y_i + 1)\Gamma(\theta)} \left(\frac{e^{\beta' x_i}}{e^{\beta' x_i} + \theta} \right)^{y_i} \left(\frac{\theta}{e^{\beta' x_i} + \theta} \right)^{\theta}$$

This probability function stems from a specific form of the negative binomial distribution with the following expected value and variance:

$$E(y_i|x_i,\, \beta) = e^{\beta' x_i}$$

$$Var(y_i|x_i,\, \beta) = e^{\beta' x_i} + \frac{(e^{\beta' x_i})^2}{\theta}$$

Different Negbin regression models with different variances can be derived with different values of θ. The most important specification is the Negbin II model with $1/\theta = \sigma^2$. It follows for the variance of y_i:

$$Var(y_i|x_i,\, \beta) = e^{\beta' x_i} + \sigma^2 \left(e^{\beta' x_i} \right)^2 = e^{\beta' x_i} \left(1 + \sigma^2 e^{\beta' x_i} \right)$$

The assumption of another θ with $1/\theta = \sigma^2/\exp(\beta' x_i)$ leads to the Negbin I model with the following variance of y_i:

$$Var(y_i|x_i,\, \beta) = e^{\beta' x_i} + \frac{\sigma^2 \left(e^{\beta' x_i} \right)^2}{e^{\beta' x_i}} = e^{\beta' x_i} + \sigma^2 e^{\beta' x_i} = e^{\beta' x_i} \left(1 + \sigma^2 \right)$$

Comparison between Negbin and Poisson regression models:

- For $\sigma^2 \rightarrow 0$, both Negbin regression models with overdispersion converge to the Poisson regression model (this property can be used as basis for Wald or likelihood ratio tests to test the validity of the Poisson regression model)
- If the assumptions of the Poisson regression model are violated due to unobserved heterogeneity and this unobserved heterogeneity term is gamma distributed, then the application of Negbin regression models is useful since this leads to efficient estimators. In particular the Negbin II model has been found to be very useful in empirical applications since the squared variance specification is often a good approximation.
- However, the Negbin regression models lead to inconsistent ML estimators if the underlying model assumptions are violated (e.g. if the unobserved heterogeneity term is not gamma distributed) so that they are less robust to misspecifications than the Poisson regression model
- In the case of overdispersion, the ML estimator in the Poisson regression model is (similar to heteroskedasticity in linear regression models) inefficient and the estimators of the standard deviations of the estimated parameters are inconsistent, i.e. they are underestimated so that the z statistics become too high. However, the ML estimators remain consistent under overdispersion and other model specifications so that a quasi ML estimation strategy is often useful which includes robust estimated standard deviations on the basis of the consistent ML estimators under misspecification.

Robust estimations of variances and covariances of estimated parameters:

- According to chapter 1.2, three different estimators for the variance covariance matrix of the estimated parameter vector can be examined
- The default option in ML estimations of econometric models with STATA refers to the exclusive incorporation of the Hessian matrix at the ML estimator
- This default option can also be generated by including the STATA command "vce(oim)" at the end of the command line for an ML estimation
- The second approach, i.e. the outer product of the gradient, that exclusively incorporates the score at the ML estimator can be generated by including the STATA command "vce(opg)"
- The corresponding STATA command for the third robust approach that incorporates both the Hessian matrix and the score at the ML estimator is "robust" or "vce(robust)"
- The use of this estimator is robust to several types of misspecifications on the basis of ML estimations in different econometric models (i.e. not only with respect to an incorrectly assumed equidispersion in Poisson regression models)
- Therefore, the robust estimation of the variances of estimated parameters is the general rule rather than the exception in empirical applications, not only in count data models, but also in binary, multinomial, and ordered response models as discussed in the previous chapters

---

Example: Determinants of the fertility of women (I)

As in the previous example, the effect of the years of education, race, the immigrant status, low income in the youth, living in the city in the youth, and of time on the number of children ever born for 5150 women aged 40 years or older in the US is examined. However, instead of a Poisson regression model, Negbin regression models are used now.

Concerning the ML estimation of the Negbin I model and Negbin II model with STATA, the following peculiarities have to be considered:

- The default of the estimation of Negbin regression models without an additional option is the (more robust) Negbin II model (with different dispersions across the observations)
- For the estimation of the Negbin I model the specific form of the variance has to be made clear (i.e. constant dispersion across the observations)

The ML estimation of these two Negbin regression models with STATA leads to the following results:

Example: Determinants of the fertility of women (II)

```
nbreg kids educ white immigrant lowincome16 city16 time

Negative binomial regression                    Number of obs     =      5,150
                                                LR chi2(6)        =     355.04
Dispersion     = mean                           Prob > chi2       =     0.0000
Log likelihood = -10014.049                     Pseudo R2         =     0.0174

------------------------------------------------------------------------------
        kids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0452325   .0032889   -13.75   0.000    -.0516787   -.0387864
       white |  -.1376176   .0268243    -5.13   0.000    -.1901923   -.0850429
   immigrant |  -.0791429   .0316198    -2.50   0.012    -.1411166   -.0171692
 lowincome16 |   .0116351   .0245593     0.47   0.636    -.0365002    .0597704
      city16 |  -.0549255   .0218053    -2.52   0.012    -.0976631   -.0121878
        time |  -.0218662   .0046713    -4.68   0.000    -.0310217   -.0127107
       _cons |    1.72085   .0459114    37.48   0.000     1.630865    1.810835
-------------+----------------------------------------------------------------
    /lnalpha |   -2.08434   .0862541                     -2.253395   -1.915285
-------------+----------------------------------------------------------------
       alpha |   .1243892   .0107291                       .105042    .1472999
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 205.17               Prob >= chibar2 = 0.000
```

Example: Determinants of the fertility of women (III)

```
nbreg kids educ white immigrant lowincome16 city16 time, dispersion(constant)

Negative binomial regression                    Number of obs     =      5,150
                                                LR chi2(6)        =     318.49
Dispersion     = constant                       Prob > chi2       =     0.0000
Log likelihood = -10032.325                     Pseudo R2         =     0.0156

------------------------------------------------------------------------------
        kids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0422836   .0031905   -13.25   0.000    -.0485369   -.0360304
       white |  -.1267473   .0262465    -4.83   0.000    -.1781895   -.0753051
   immigrant |  -.0814915   .0313829    -2.60   0.009    -.1430008   -.0199823
 lowincome16 |   .0096098   .0240734     0.40   0.690    -.0375732    .0567929
      city16 |  -.0527135   .0215582    -2.45   0.014    -.0949668   -.0104602
        time |  -.0205964   .0046097    -4.47   0.000    -.0296313   -.0115615
       _cons |     1.6722   .0441904    37.84   0.000     1.585589    1.758812
-------------+----------------------------------------------------------------
    /lndelta |  -1.197586   .0936571                      -1.38115   -1.014021
-------------+----------------------------------------------------------------
       delta |   .3019222   .0282772                      .2512893    .3627573
------------------------------------------------------------------------------
LR test of delta=0: chibar2(01) = 168.62               Prob >= chibar2 = 0.000
```

Example: Determinants of the fertility of women (IV)

Interpretation:

- The estimation results in both Negbin regression models can be identically interpreted as in the Poisson regression model (probabilities and probability effects can also be identically estimated with STATA, see later)
- The parameter estimates are qualitatively and quantitatively extremely similar across the three count data models so that particularly the strong significantly negative effect of education on the fertility of women from the Poisson regression model is confirmed. However, the z statistics are slightly smaller in the Negbin regression models, although this has only small impacts on the significance of the effect of the variables (e.g. for city16).
- These different z statistics can be caused by the existence of overdispersion which is strongly confirmed due to the corresponding Wald and likelihood ratio test statistics in both the Negbin II model and the Negbin I model
- Overall, however, the estimation results from the Poisson regression model seem to be robust, although further robustness checks with alternative count data models could be conducted

An example for the presentation of all these estimation results (in Poisson and Negbin regression models) in empirical studies is as follows:

Example: Determinants of the fertility of women (V)

ML estimates (z statistics), dependent variable: number of children ever born (kids)			
Explanatory variables	Poisson regression model	Negbin II model	Negbin I model
educ	-0.0442*** (-15.73)	-0.0452*** (-13.75)	-0.0423*** (-13.25)
white	-0.1366*** (-5.95)	-0.1376*** (-5.13)	-0.1267*** (-4.83)
immigrant	-0.0808*** (-2.92)	-0.0791** (-2.50)	-0.0815*** (-2.60)
lowincome16	0.0115 (0.54)	0.0116 (0.47)	0.0096 (0.40)
city16	-0.0555*** (-2.91)	-0.0549** (-2.52)	-0.0527** (-2.45)
time	-0.0219*** (-5.40)	-0.0219*** (-4.68)	-0.0206*** (-4.47)
constant	1.7080 (44.09)	1.7209 (37.48)	1.6722 (37.84)
σ	-	0.1244	0.3019

Note: *** (**, *) means that the appropriate explanatory variable has an effect at the 1% (5%, 10%) significance level, n = 5150

Example: Determinants of the fertility of women (VI)

The ML estimation of the corresponding Poisson regression model and the Negbin II model with STATA that includes a robust estimation of the variances and covariances of the estimated parameters, respectively, leads to the following results:

```
poisson kids educ white immigrant lowincome16 city16 time, robust

Poisson regression                                Number of obs   =       5150
                                                  Wald chi2(6)    =     316.80
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -10116.635                 Pseudo R2       =     0.0234

------------------------------------------------------------------------------
             |               Robust
        kids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0441974   .0034358   -12.86   0.000    -.0509314   -.0374633
       white |  -.1365916   .0281747    -4.85   0.000    -.1918131   -.0813702
   immigrant |  -.0808302   .0318329    -2.54   0.011    -.1432215   -.0184389
 lowincome16 |   .0114575    .024799     0.46   0.644    -.0371477    .0600626
      city16 |   -.055456   .0212928    -2.60   0.009    -.0971891   -.0137229
        time |  -.0218599   .0046366    -4.71   0.000    -.0309475   -.0127723
       _cons |   1.708038   .0494595    34.53   0.000     1.611099    1.804977
------------------------------------------------------------------------------
```

Example: Determinants of the fertility of women (VII)

```
nbreg kids educ white immigrant lowincome16 city16 time, vce(robust)

Negative binomial regression                    Number of obs   =       5150
Dispersion           = mean                     Wald chi2(6)    =     324.42
Log pseudolikelihood = -10014.049               Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
        kids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0452325   .0034371   -13.16   0.000    -.0519691    -.038496
       white |  -.1376176   .0277906    -4.95   0.000    -.1920862    -.083149
   immigrant |  -.0791429   .0316932    -2.50   0.013    -.1412605   -.0170253
 lowincome16 |   .0116351   .0245938     0.47   0.636    -.0365679     .059838
      city16 |  -.0549255   .0211495    -2.60   0.009    -.0963777   -.0134732
        time |  -.0218662   .0045796    -4.77   0.000     -.030842   -.0128905
       _cons |    1.72085   .0495233    34.75   0.000     1.623786    1.817914
-------------+----------------------------------------------------------------
    /lnalpha |   -2.08434   .0868808                     -2.254623   -1.914057
-------------+----------------------------------------------------------------
       alpha |   .1243892    .010807                      .1049131    .1474809
------------------------------------------------------------------------------
```

→ The likelihood ratio test statistic with respect to the null hypothesis of equi-dispersion is not reported in this case (i.e. it is only directly reported on the basis of the two other versions of estimated variance covariance matrixes of the estimated parameters)

Example: Determinants of the fertility of women (VIII)

Wald and likelihood ratio tests in the Negbin II model:

The null hypothesis that neither lowincome16 nor city16 has any effect on the number of children ever born, i.e. that the two corresponding parameters are zero, is tested after the ML estimation without a robust estimation of the variance covariance matrix of the estimated parameters (the test results based on a robust estimation are different). The command for the Wald test in STATA is:

```
test lowincome16 city16

 ( 1)  [kids]lowincome16 = 0
 ( 2)  [kids]city16 = 0

           chi2(  2) =    6.71
         Prob > chi2 =    0.0348
```

The corresponding commands for the likelihood ratio test in STATA are then (a likelihood ratio test with a robust estimation of the variance covariance matrix of the estimated parameters is not directly possible):

```
estimates store unrestricted

nbreg kids educ white immigrant time

estimates store restricted

lrtest unrestricted restricted

Likelihood-ratio test                                 LR chi2(2)  =      6.72
(Assumption: restricted nested in unrestricted)       Prob > chi2 =    0.0347
```

Example: Determinants of the fertility of women (IX)

The estimation of the average marginal probability effects of educ across all 5150 observations for one, two, and three children ever born leads to the following (shortened) STATA results in the Negbin II model (no robust estimation of the variance covariance matrix of the estimated parameters):

```
margins, dydx(educ) predict(pr(1))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0106979   .0007563    14.15   0.000     .0092156    .0121802
------------------------------------------------------------------------------

margins, dydx(educ) predict(pr(2))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |     .00402   .0003231    12.44   0.000     .0033868    .0046533
------------------------------------------------------------------------------

margins, dydx(educ) predict(pr(3))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.0025423   .0002361   -10.77   0.000     -.003005   -.0020797
------------------------------------------------------------------------------
```

Example: Determinants of the fertility of women (X)

The estimation of the average probability that kids = 1 for white = 1 and for white = 0 across all 5150 observations leads to the following STATA results in the Negbin II model (no robustly estimated variance covariance parameters):

```
margins, at(white=1) predict(pr(1))

Predictive margins                                Number of obs   =       5150
Model VCE     : OIM

Expression    : Pr(kids=1), predict(pr(1))
at            : white           =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .2167028   .0026777    80.93   0.000     .2114547     .221951
------------------------------------------------------------------------------

margins, at(white=0) predict(pr(1))

Predictive margins                                Number of obs   =       5150
Model VCE     : OIM

Expression    : Pr(kids=1), predict(pr(1))
at            : white           =           0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .1837677   .0059703    30.78   0.000     .1720661    .1954692
------------------------------------------------------------------------------
```