Negative Binomial

$i =$ rows

$j =$ columns

D_{ij}= cumulative claims

C_{ij}= cumulative claims

In R this is named *loss:*

	1	2	3	4	5	6	7	8	9	10
1	5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
2	106	4285	5396	10666	13782	15599	15496	16169	16704	
3	3410	8992	13873	16141	18735	22214	22863	23466		
4	5655	11555	15766	21266	23425	26083	27067			
5	1092	9565	15836	22169	25955	26180				
6	1513	6445	11702	12935	15852					
7	557	4020	10946	12314						
8	1351	6947	13112							
9	3133	5395								
10	2063									

So $D_{11} = 5012$, $D_{12} = 8269$ …

In R this is called *All:*

	1	2	3	4	5	6	7	8	9	10
1	0	8.51959	9.020269	9.29716	9.376278	9.51333	9.691593	9.798627	9.831347	6.481577
2	0	4.663439	8.362876	8.593413	9.274816	9.531119	9.654962	9.648337	9.690851	
3	0	8.134468	9.104091	9.5377	9.689118	9.838149	10.00848	10.03728		
4	0	8.640295	9.354874	9.665611	9.964865	10.06156	10.16904			
5	0	6.995766	9.165866	9.670041	10.00645	10.16412				
6	0	7.32185	8.77106	9.367515	9.467692					
7	0	6.322565	8.299037	9.300729						
8	0	7.2086	8.846065							
9	0	8.049746								
10	0									

So $8.51959 = ln(5012)$ represents $\log(D_{i,j-1})$ where $i = 1,\ j = 2$

(using C_{ij} or D_{ij} makes no difference)

Now,

$$E[C_{ij}] = m_{ij} = (\lambda_j - 1)D_{i,j-1}$$

$$\eta_{ij} = \log(m_{ij}) = \log(\lambda_j - 1) + \log(D_{i,j-1})$$

$\log(\lambda_j - 1) = c + \alpha_{j-1}$ with: $\alpha_1 = 0, j \geq 2$

$$\log(m_{ij}) = c + \alpha_{j-1} + \log(D_{i,j-1})$$

For $1 \leq i \leq 10\ and\ j = 1$

$$\log(m_{11}) = c$$

$$\log(m_{21}) = c$$

$$\vdots$$

$$\log(m_{10,1}) = c$$

Since $\alpha_{j-1} = \alpha_0$does not exists.

Then for $1 \leq i \leq 9\ and\ j = 2$

$$\log(m_{12}) = c + \alpha_1 + \log(D_{1,1})$$

$$\log(m_{22}) = c + \alpha_1 + \log(D_{2,1})$$

$$\vdots$$

$$\log(m_{92}) = c + \alpha_1 + \log(D_{9,1})$$

But $\alpha_1 = 0$

Then for $1 \leq i \leq 8\ and\ j = 3$

$$\log(m_{13}) = c + \alpha_2 + \log(D_{1,2})$$

$$\vdots$$

$$\log(m_{83}) = c + \alpha_2 + \log(D_{8,2})$$

And so on until $i = 1,\ j = 10$

$$\log(m_{1,10}) = c + \alpha_9 + \log(D_{1,9})$$

In R:

```
> loss  #the first table above
> claims <- as.vector(loss)
```

```
> n.origin <- nrow(loss)

> n.dev <- ncol(loss)-1   #since α1 … α9

> dev <- factor(col <- rep(0:n.dev, each=n.origin))

> All    #the second table above

> cum <- as.vector(All)

> mack <- data.frame(claims=claims,dev=dev,cum=cum)

> library(MASS)

> model.nb1 <- glm.nb(claims ~ dev + offset(cum), data = mack)

> summary(model.nb1)
```

The parameters:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   7.7787     0.1865   41.72   <2e-16 ***
dev1         -5.6752     0.2709  -20.95   <2e-16 ***
dev2         -7.2505     0.2796  -25.93   <2e-16 ***
dev3         -7.5053     0.2905  -25.84   <2e-16 ***
dev4         -7.6107     0.3044  -25.00   <2e-16 ***
dev5         -7.6592     0.3228  -23.72   <2e-16 ***
dev6         -7.7363     0.3487  -22.19   <2e-16 ***
dev7         -7.7449     0.3880  -19.96   <2e-16 ***
dev8        -10.6810     0.4571  -23.37   <2e-16 ***
dev9         -8.8398     0.6217  -14.22   <2e-16 ***
```

But they are not correct.

dev1 which represent α_1 should be zero and the result should be

	Parameter estimates
Constant	0.6928
α_2	-1.1652
α_3	-1.9989
α_4	-2.4550
α_5	-2.8698
α_6	-3.8645
α_7	-4.0961
α_8	-4.7711
α_9	-5.3796

I think that the problem is when I'm specifying the data.frame.

```
> mack
    claims dev       cum
1     5012   0  0.000000
2      106   0  0.000000
3     3410   0  0.000000
4     5655   0  0.000000
5     1092   0  0.000000
6     1513   0  0.000000
7      557   0  0.000000
8     1351   0  0.000000
9     3133   0  0.000000
10    2063   0  0.000000
11    8269   1  8.519590
12    4285   1  4.663439
13    8992   1  8.134468
14   11555   1  8.640295
15    9565   1  6.995766
16    6445   1  7.321850
17    4020   1  6.322565
18    6947   1  7.208600
19    5395   1  8.049746
20      NA   1        NA
21   10907   2  9.020269
22    5396   2  8.362876
23   13873   2  9.104091
24   15766   2  9.354874
25   15836   2  9.165866
26   11702   2  8.771060
27   10946   2  8.299037
28   13112   2  8.846065
```

...and so on.

I need to tell R that under the column named *dev*, the first ten zeros, should not take them as a label of a parameter. Since as I specified before, the first ten claims, does not depend on a parameter but only on the intercept.

How can I do it please?