HW2

Alan Wu
MKTG 7760: Probability Models in Marketing

February 5, 2026

1 Question 1

Derive the probability mass function and the forward recursion formula for an NBD model for a period of length t.

To derive the pmf and forward recursion for an observation of length t, where there is non-unit observation, we can take advantage of the poisson distribution with rate λt:

$$P(X(t) = x|\lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

We will use the same mixing distribution: the gamma distribution which has the following pdf:

$$g(\lambda) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}$$

Our ultimate solution is to derive $P(X(t) = x|r, \alpha)$, which can be derived from below:

$$\begin{aligned} P(X(t) = x|r,\alpha) &= \int_{\lambda=0}^{\infty} P(X(t) = x|\lambda) g(\lambda|r,\alpha) d\lambda \\ &= \int_{\lambda=0}^{\infty} \frac{(\lambda t)^x e^{-\lambda t}}{x!} \cdot \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)} d\lambda \\ &= \frac{t^x \cdot \alpha^r}{x! \cdot \Gamma(r)} \cdot \int_{\lambda=0}^{\infty} \lambda^{x+r-1} \cdot e^{-\lambda(\alpha+t)} d\lambda \end{aligned}$$

From here, we notice that we can transform the integral as the cdf of a gamma distribution with parameters r = r+x and $\alpha = \alpha + t$ by multiplying by constants:

$$\frac{t^x \cdot \alpha^r}{x! \cdot \Gamma(r)} \cdot \int_{\lambda=0}^{\infty} \lambda^{x+r-1} \cdot e^{-\lambda(\alpha+t)} d\lambda$$

$$= \frac{t^x \cdot \alpha^r \cdot \Gamma(r+x)}{x! \cdot \Gamma(r) \cdot (\alpha+t)^{r+x}} \cdot \int_{\lambda=0}^{\infty} \frac{\lambda^{x+r-1} \cdot e^{-\lambda(\alpha+t)} \cdot (\alpha+t)^{r+x}}{\Gamma(r+x)} d\lambda$$

$$= \frac{t^x \cdot \alpha^r \cdot \Gamma(r+x)}{x! \cdot \Gamma(r) \cdot (\alpha+t)^{r+x}}$$

$$= \frac{\Gamma(r+x)}{\Gamma(r)x!} \cdot (\frac{\alpha}{\alpha+t})^r \cdot (\frac{t}{a+t})^x$$

Notice that for this formula, when t = 1, it absorbs back to the unit-length observation window formula for the nbd model.

To derive the forward-recursion formula for the t length observation, we can use previous results and compute $\frac{P(X(t)=x|r,\alpha)}{P(X(t)=x-1|r,\alpha)}$ for when $x > 0$. For when x = 0, we have the following:

$$P(X(t) = 0|r, \alpha) = (\frac{\alpha}{\alpha+t})^r$$

And for x ¿ 0:

$$\frac{P(X(t) = x|r, \alpha)}{P(X(t) = x-1|r, \alpha)} = \frac{\frac{\Gamma(r+x)}{\Gamma(r)x!} \cdot (\frac{\alpha}{\alpha+t})^r \cdot (\frac{t}{a+t})^x}{\frac{\Gamma(r+x-1)}{\Gamma(r)(x-1)!} \cdot (\frac{\alpha}{\alpha+t})^r \cdot (\frac{t}{a+t})^{x-1}}$$

$$= \frac{\Gamma(r+x)}{\Gamma(r+x-1)} \cdot \frac{(x-1)!}{x!} \cdot \frac{t^x \cdot (\alpha+t)^{x-1}}{t^{x-1} \cdot (\alpha+t)^x}$$

$$= \frac{r+x-1}{x} \cdot \frac{t}{\alpha+t}$$

And the final derivation becomes: $P(X(t) = x|r, \alpha) = \frac{t(r+x-1)}{x(\alpha+t)} \cdot P(X(t) = x-1|r, \alpha)$.

Putting both expressions together, we get:

$$P(X(t) = x|r, \alpha) = \begin{cases} (\frac{\alpha}{\alpha+t})^r & \text{for } x = 0 \\ \frac{t(r+x-1)}{x(\alpha+t)} \cdot P(X(t) = x-1|r, \alpha) & \text{for } x > 0 \end{cases}$$

2 Question 2

On Canvas, you'll see the "HW prescription data" spreadsheet, containing data on the number of prescriptions for a particular drug written by a sample of 1923 doctors in a given month. Fit an NBD model and show the expected distribution for the number of prescriptions over a 12-month period (assuming stationarity).

2.1 Original Distribution vs. 12-month Distribution

After the model was fitted, we obtain the two distribution graphs for 1-month observation period and the forecasted distribution after 12 months. Notably, the results show that the optimal model outputs parameter values of $(r, \alpha) = (0.108, 0.320)$. Full results are in the appendix section.

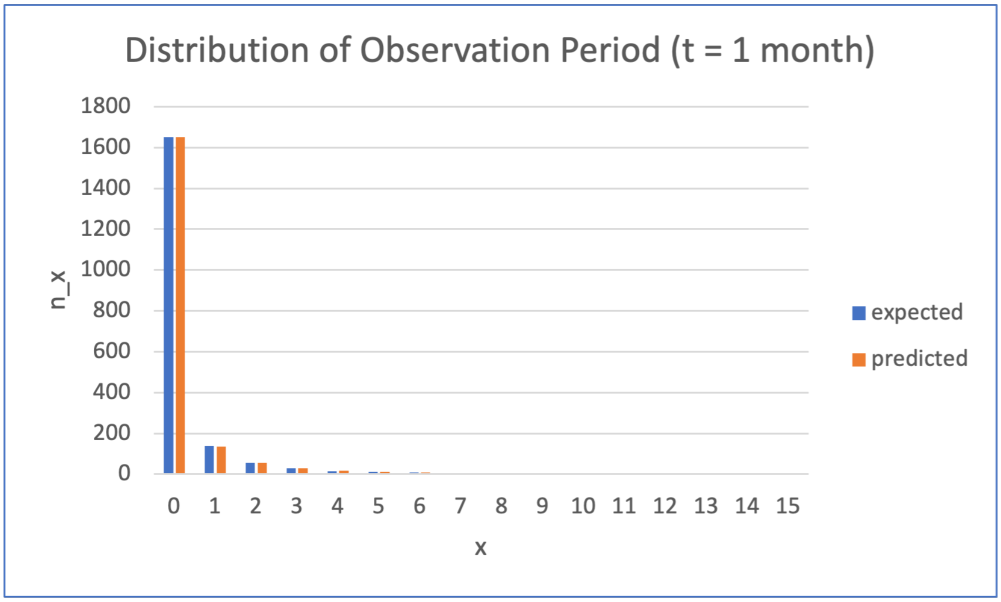


Figure 1: nbd Model on prescription data during observation period with expected and predicted counts

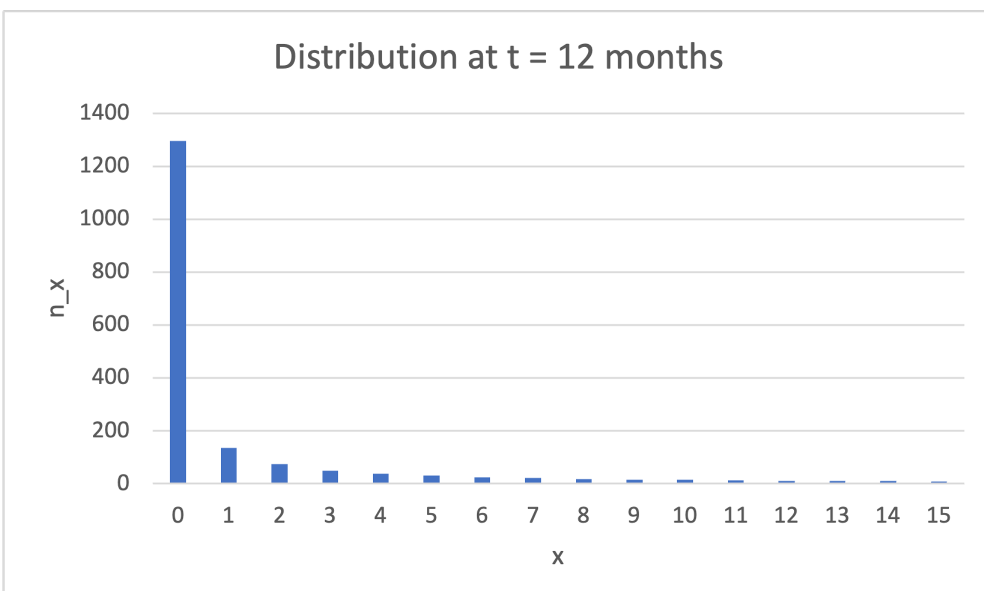


Figure 2: nbd Model on prescription data forecasted to 12 months

2.2 Model Fit

We observe and note the model fit to be very good by the chi-sq goodness of fit test. After rolling up categories to 0 - 5+, due to low expected counts of subsequent categories, we obtain a p-value of 1 with the chi-sq goodness of fit test and the following visual:

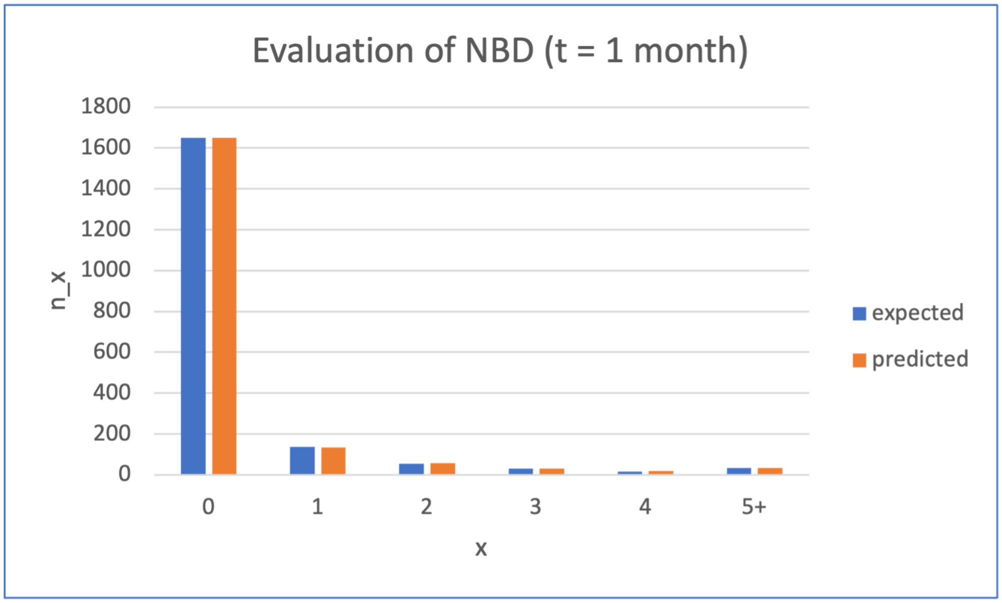


Figure 3: nbd Model on prescription data evaluation rolled up to 5+

Visually, there is practically no difference between the expected and the predicted counts during the observation period. We are unable to evaluate the forecasted model due to there being no ground truth at t = 12 months.

2.3 Conclusions

The transition from a 1-month to a 12-month observation period demonstrates how the NBD models consumer accrual, specifically showing that the "reach" of a product or brand increases as the time horizon expands. As seen in the provided data, the count of non-buyers (n_0) drops from approximately 1,650 to 1,300, a mathematical consequence of the probability of zero transactions, $P(0, t) = (\frac{\alpha}{\alpha+t})^r$, decreasing as t grows.

This shift effectively migrates mass from the zero-class into the positive transaction classes, resulting in a significantly heavier tail where "heavy buyers" become visible and the mean frequency scales linearly by a factor of 12. While the short-term window is dominated by zero-inflation, the long-term distribution provides a more robust view of the underlying purchase intensities and the true penetration of the market.

If we compute the reach, frequency, and GRPs, this shift is even more noticeable:

The transition from a 1-month to a 12-month period illustrates the accrual property of the NBD model, where the brand's reach more than doubles from 14.18% to 32.56%.

Observation Period	Reach (%)	Average Frequency	GRPs
1 Month	14.18%	2.38	34
12 Months	32.56%	12.44	405

Table 1: Comparison of NBD Model Metrics: 1-Month vs. 12-Month Performance

This expansion is driven by the steady conversion of non-buyers into the transaction pool, causing the zero-frequency count (n_0) to drop as the probability $P(0, t)$ decreases over time. Consequently, the average frequency climbs to 12.44, reflecting the accumulation of repeat purchases, while the total GRPs scale from 34 to 405, highlighting the compounding effect of deeper market penetration and sustained customer intensity.

3 Question 3

Consider an NBD model for which the variance is twice the mean, and the mean is twice the value of P(0). What are the values of r and α? Show the key steps of your analysis.

We can solve this algebraically with system of equations. We note that we know the following:

- The mean of a gamma distribution with parameters r, α is $\frac{r}{\alpha}$. That is, $E(X) = \frac{r}{\alpha}$, given $X \sim \Gamma(r, \alpha)$
- The variance of a gamma distribution with parameters r, α is $\frac{r}{\alpha} + \frac{r}{\alpha^2}$. That is, $V(X) = \frac{r}{\alpha} + \frac{r}{\alpha^2}$, given $X \sim \Gamma(r, \alpha)$.
- $P(X = 0|r, \alpha) = (\frac{\alpha}{\alpha+1})^r$, for unit observation period.

Therefore, we formulate the following systems of equations:

- The statement $V(X) = 2 \cdot E(X)$ implies that $\frac{r}{\alpha} + \frac{r}{\alpha^2} = 2 \cdot \frac{r}{\alpha}$
- The statement $E(X) = 2 \cdot P(0)$ implies that $\frac{r}{\alpha} = 2 \cdot (\frac{\alpha}{\alpha+1})^r$

We solve the first equation:

$$\begin{aligned} \frac{r}{\alpha} + \frac{r}{\alpha^2} &= 2 \cdot \frac{r}{\alpha} \\ r + \frac{r}{\alpha} &= 2r \\ \frac{r}{\alpha} &= r \\ \alpha &= 1 \end{aligned}$$

Then we plug in $\alpha = 1$ into the second equation

$$
\begin{aligned}
\frac{r}{\alpha} &= 2 \cdot (\frac{\alpha}{\alpha+1})^r \\
r &= 2 \cdot (\frac{1}{2})^r \\
r &= 1
\end{aligned}
$$

Therefore, we obtain parameters $(r, \alpha) = (1, 1)$.

4 Question 4

The table below shows the number of surveys filled out by a sample of 1865 Americans in 1995: Fit an NBD model to these data. Think carefully about how you'll handle that "3-5" cell...

We attempt to handle the extra rolled up feature in the data by computing individual probabilities of the bucket and then summing those probabilities up when computing the log likelihood, to give attribution to each potential category within the rolled up bucket.

Further, for the last 6+ bucket, since probabilities sum to 1, we take the remaining probabilities from the sum from previous buckets such that now the log likelihood is defined as, given the counts n_x:

- $LL_x = n_x \cdot ln(P(X = x|r, \alpha))$, for $x = 0, 1, 2$
- $LL_{3-5} = n_{3-5} \cdot ln[P(X = 3|r, \alpha) + P(X = 4|r, \alpha) + P(X = 5|r, \alpha)]$
- $LL_{6+} = n_6 \cdot ln((1 - \sum_{x=1}^{5} P(X = x|r, \alpha))$

By using this methodology, we account for the rolled up categories in the 3-5 bucket and also any remaining probabilities left unaccounted for by less than 6.

4.1 Results

The fit of the nbd results in $(r, \alpha) = (0.391, 0.248)$ with chisq p-value of $p << 0.001$.

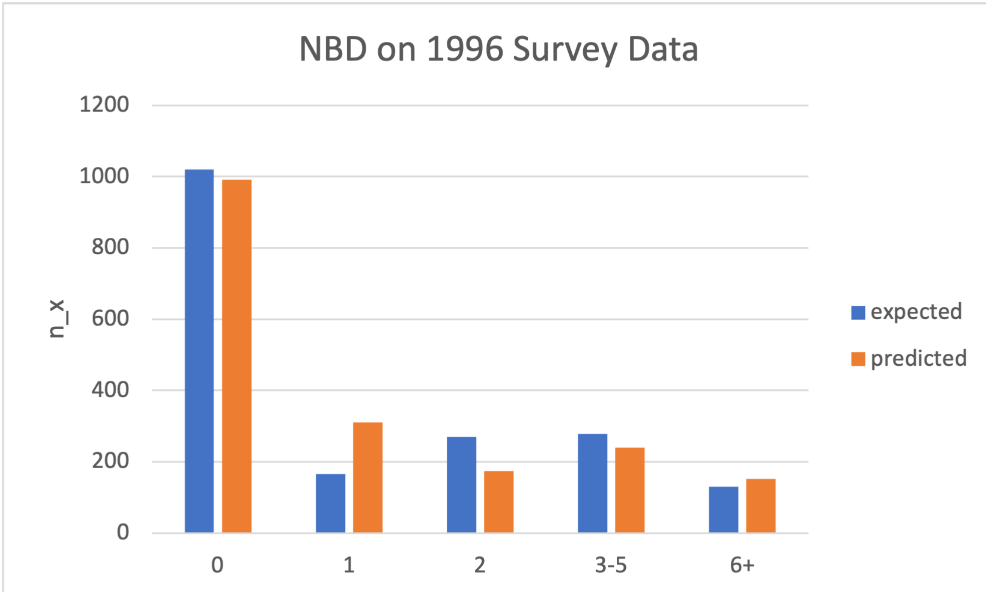


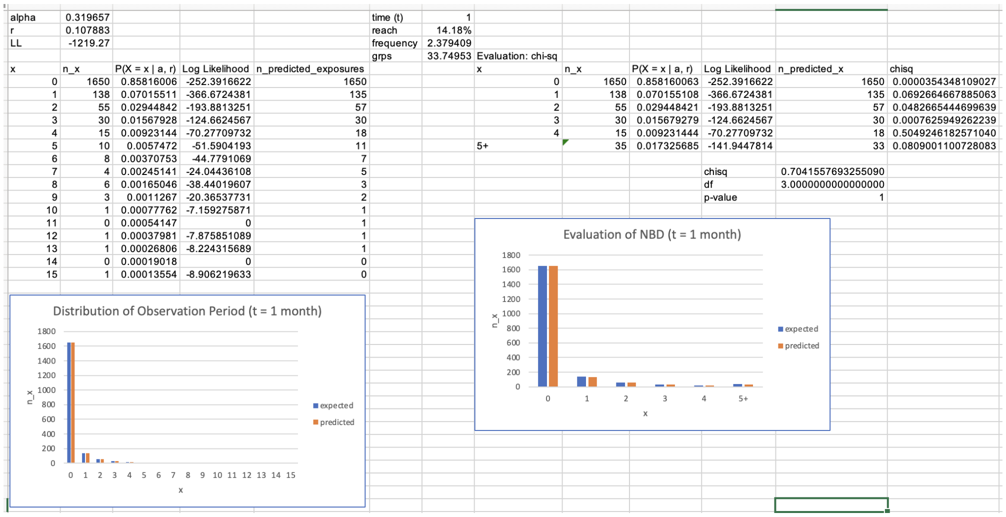
Figure 4: nbd Model on 1995 survey data during observation period with expected and predicted counts

4.2 Conclusions

Based on the goodness-of-fit test for the model, the results indicate a poor fit to the observed survey data. Specifically, the model significantly underestimates the frequency of the ”0” category (predicting 990 vs. 1020 observed) and fails to capture the variance in the higher-frequency bins, such as category ”1” (predicting 311 vs. 166 observed) and category ”2” (predicting 173 vs. 270 observed). Both visual observation and goodness-of-fit test confirm these conclusions.

5 Appendix

5.1 Total Excel Tables



alpha	0.319657
r	0.107883
LL	-1219.27

time (t)	1
reach	14.18%
frequency	2.379409
grps	33.74953

x	n_x	P(X = x | a, r)	Log Likelihood	n_predicted_exposures
0	1650	0.85816006	-252.3916622	1650
1	138	0.07015511	-366.6724381	135
2	55	0.02944842	-193.8813251	57
3	30	0.01567928	-124.6624567	30
4	15	0.00923144	-70.27709732	18
5	10	0.0057472	-51.5904193	11
6	8	0.00370753	-44.7791069	7
7	4	0.00245141	-24.04436108	5
8	6	0.00165046	-38.44019607	3
9	3	0.0011267	-20.36537731	2
10	1	0.00077762	-7.159275871	1
11	0	0.00054147	0	1
12	1	0.00037981	-7.875851089	1
13	1	0.00026806	-8.224315689	1
14	0	0.00019018	0	0
15	1	0.00013554	-8.906219633	0

Evaluation: chi-sq

x	n_x	P(X = x | a, r)	Log Likelihood	n_predicted_x	chisq
0	1650	0.858160063	-252.3916622	1650	0.0000354348109027
1	138	0.070155108	-366.6724381	135	0.0692664667885063
2	55	0.029448421	-193.8813251	57	0.0482665444699639
3	30	0.015679279	-124.6624567	30	0.0007625949262239
4	15	0.009231444	-70.27709732	18	0.5049246182571040
5+	35	0.017325685	-141.9447814	33	0.0809001100728083

chisq	0.7041557693255090
df	3.0000000000000000
p-value	1

Figure 5: nbd Model on prescription data with 1 month observation

alpha	0.3196571			reach	0.3256161
r	0.1078828			frequency	12.437785
t	12			grps	404.99435
x	n_x	P(X = x | a, r	predicted_n_x		
0	1650	0.6743839	1297		
1	138	0.0708667	136		
2	55	0.0382374	74		
3	30	0.0261695	50		
4	15	0.0198054	38		
5	10	0.0158494	30		
6	8	0.0131428	25		
7	4	0.0111702	21		
8	6	0.0096671	19		
9	3	0.0084829	16		
10	1	0.0075256	14		
11	0	0.0067358	13		
12	1	0.0060733	12		
13	1	0.0055098	11		
14	0	0.0050248	10		
15	1	0.0046033	9		

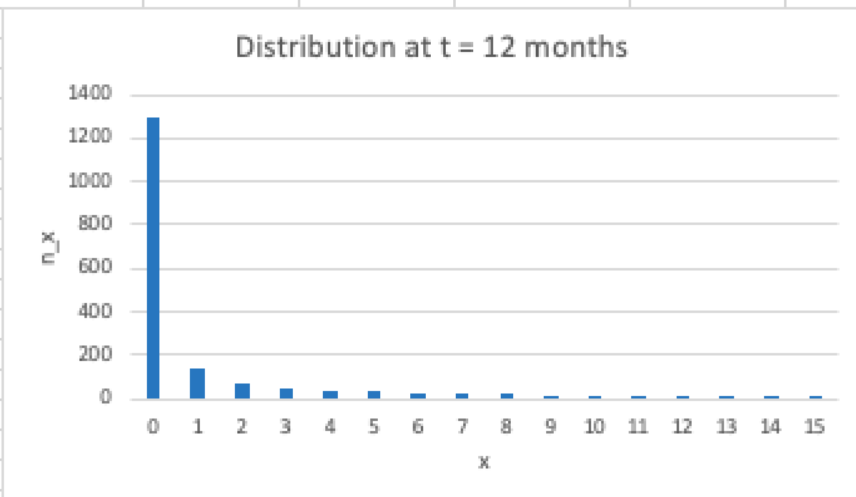


Figure 6: nbd Model on prescription data forecasted 12 month observation

alpha	0.24763044
r	0.3913563
LL	-2484.153
t	1

x	n_x	% of respondent	P(X = x | a,r)	Log Likelihood	Predicted n_x
0	1020	54.7%	0.53107706	-645.505107	990
1	166	8.9%	0.16658807	-297.5103685	311
2	270	14.5%	0.09288943	-641.6132569	173
3			0.05934763		
4			0.04033024		
5			0.02839053		
3-5	279	15.0%	0.12806841	-573.3982138	239
6+	130	7.0%	0.08137703	-326.1260963	152

categories	n_x	P(X = x | a,r)	LL	predicted_n_x	chisq
0	1020	0.531077061	-645.50511	990	0.85557579
1	166	0.166588074	-297.51037	311	126.109988
2	270	0.092889432	-641.61326	173	34.6767837
3-5	279	0.128068407	-573.39821	239	5.77855539
6+	130	0.081377026	-326.1261	152	3.64501939
	chisq	171.0659224			
	df	2			
	pvalue	7.13686E-38			

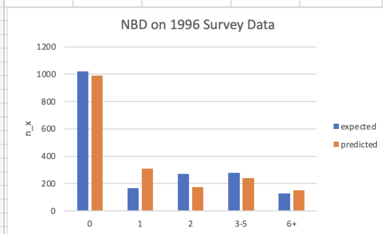


Figure 7: nbd Model on survey data