

Negative Binomial Regression

Second Edition

JOSEPH M. HILBE

*Jet Propulsion Laboratory,
California Institute of Technology and
Arizona State University*



CAMBRIDGE
UNIVERSITY PRESS

Contents

<i>Preface to the second edition</i>	<i>page xi</i>
1 Introduction	1
1.1 What is a negative binomial model?	1
1.2 A brief history of the negative binomial	5
1.3 Overview of the book	11
2 The concept of risk	15
2.1 Risk and 2×2 tables	15
2.2 Risk and $2 \times k$ tables	18
2.3 Risk ratio confidence intervals	20
2.4 Risk difference	24
2.5 The relationship of risk to odds ratios	25
2.6 Marginal probabilities: joint and conditional	27
3 Overview of count response models	30
3.1 Varieties of count response model	30
3.2 Estimation	38
3.3 Fit considerations	41
4 Methods of estimation	43
4.1 Derivation of the IRLS algorithm	43
4.1.1 Solving for $\partial \mathcal{L}$ or U – the gradient	48
4.1.2 Solving for $\partial^2 \mathcal{L}$	49
4.1.3 The IRLS fitting algorithm	51
4.2 Newton–Raphson algorithms	53
4.2.1 Derivation of the Newton–Raphson	54
4.2.2 GLM with OIM	57

4.2.3	Parameterizing from μ to $x'\beta$	57
4.2.4	Maximum likelihood estimators	59
5	Assessment of count models	61
5.1	Residuals for count response models	61
5.2	Model fit tests	64
5.2.1	Traditional fit tests	64
5.2.2	Information criteria fit tests	68
5.3	Validation models	75
6	Poisson regression	77
6.1	Derivation of the Poisson model	77
6.1.1	Derivation of the Poisson from the binomial distribution	78
6.1.2	Derivation of the Poisson model	79
6.2	Synthetic Poisson models	85
6.2.1	Construction of synthetic models	85
6.2.2	Changing response and predictor values	94
6.2.3	Changing multivariable predictor values	97
6.3	Example: Poisson model	100
6.3.1	Coefficient parameterization	100
6.3.2	Incidence rate ratio parameterization	109
6.4	Predicted counts	116
6.5	Effects plots	122
6.6	Marginal effects, elasticities, and discrete change	125
6.6.1	Marginal effects for Poisson and negative binomial effects models	125
6.6.2	Discrete change for Poisson and negative binomial models	131
6.7	Parameterization as a rate model	134
6.7.1	Exposure in time and area	134
6.7.2	Synthetic Poisson with offset	136
6.7.3	Example	138
7	Overdispersion	141
7.1	What is overdispersion?	141
7.2	Handling apparent overdispersion	142
7.2.1	Creation of a simulated base Poisson model	142
7.2.2	Delete a predictor	145
7.2.3	Outliers in data	145
7.2.4	Creation of interaction	149

7.2.5	Testing the predictor scale	150
7.2.6	Testing the link	152
7.3	Methods of handling real overdispersion	157
7.3.1	Scaling of standard errors / quasi-Poisson	158
7.3.2	Quasi-likelihood variance multipliers	163
7.3.3	Robust variance estimators	168
7.3.4	Bootstrapped and jackknifed standard errors	171
7.4	Tests of overdispersion	174
7.4.1	Score and Lagrange multiplier tests	175
7.4.2	Boundary likelihood ratio test	177
7.4.3	R_p^2 and R_{pd}^2 tests for Poisson and negative binomial models	179
7.5	Negative binomial overdispersion	180
8	Negative binomial regression	185
8.1	Varieties of negative binomial	185
8.2	Derivation of the negative binomial	187
8.2.1	Poisson–gamma mixture model	188
8.2.2	Derivation of the GLM negative binomial	193
8.3	Negative binomial distributions	199
8.4	Negative binomial algorithms	207
8.4.1	NB-C: canonical negative binomial	208
8.4.2	NB2: expected information matrix	210
8.4.3	NB2: observed information matrix	215
8.4.4	NB2: R maximum likelihood function	218
9	Negative binomial regression: modeling	221
9.1	Poisson versus negative binomial	221
9.2	Synthetic negative binomial	225
9.3	Marginal effects and discrete change	236
9.4	Binomial versus count models	239
9.5	Examples: negative binomial regression	248
	Example 1: Modeling number of marital affairs	248
	Example 2: Heart procedures	259
	Example 3: Titanic survival data	263
	Example 4: Health reform data	269
10	Alternative variance parameterizations	284
10.1	Geometric regression: NB $\alpha = 1$	285
10.1.1	Derivation of the geometric	285
10.1.2	Synthetic geometric models	286

10.1.3	Using the geometric model	290
10.1.4	The canonical geometric model	294
10.2	NB1: The linear negative binomial model	298
10.2.1	NB1 as QL-Poisson	298
10.2.2	Derivation of NB1	301
10.2.3	Modeling with NB1	304
10.2.4	NB1: R maximum likelihood function	306
10.3	NB-C: Canonical negative binomial regression	308
10.3.1	NB-C overview and formulae	308
10.3.2	Synthetic NB-C models	311
10.3.3	NB-C models	315
10.4	NB-H: Heterogeneous negative binomial regression	319
10.5	The NB-P model: generalized negative binomial	323
10.6	Generalized Waring regression	328
10.7	Bivariate negative binomial	333
10.8	Generalized Poisson regression	337
10.9	Poisson inverse Gaussian regression (PIG)	341
10.10	Other count models	343
11	Problems with zero counts	346
11.1	Zero-truncated count models	346
11.2	Hurdle models	354
11.2.1	Theory and formulae for hurdle models	356
11.2.2	Synthetic hurdle models	357
11.2.3	Applications	359
11.2.4	Marginal effects	369
11.3	Zero-inflated negative binomial models	370
11.3.1	Overview of ZIP/ZINB models	370
11.3.2	ZINB algorithms	371
11.3.3	Applications	374
11.3.4	Zero-altered negative binomial	376
11.3.5	Tests of comparative fit	377
11.3.6	ZINB marginal effects	379
11.4	Comparison of models	382
12	Censored and truncated count models	387
12.1	Censored and truncated models – econometric parameterization	387
12.1.1	Truncation	388
12.1.2	Censored models	395

12.2	Censored Poisson and NB2 models – survival parameterization	399
13	Handling endogeneity and latent class models	407
13.1	Finite mixture models	408
13.1.1	Basics of finite mixture modeling	408
13.1.2	Synthetic finite mixture models	412
13.2	Dealing with endogeneity and latent class models	416
13.2.1	Problems related to endogeneity	416
13.2.2	Two-stage instrumental variables approach	417
13.2.3	Generalized method of moments (GMM)	421
13.2.4	NB2 with an endogenous multinomial treatment variable	422
13.2.5	Endogeneity resulting from measurement error	425
13.3	Sample selection and stratification	428
13.3.1	Negative binomial with endogenous stratification	429
13.3.2	Sample selection models	433
13.3.3	Endogenous switching models	438
13.4	Quantile count models	441
14	Count panel models	447
14.1	Overview of count panel models	447
14.2	Generalized estimating equations: negative binomial	450
14.2.1	The GEE algorithm	450
14.2.2	GEE correlation structures	452
14.2.3	Negative binomial GEE models	455
14.2.4	GEE goodness-of-fit	464
14.2.5	GEE marginal effects	466
14.3	Unconditional fixed-effects negative binomial model	468
14.4	Conditional fixed-effects negative binomial model	474
14.5	Random-effects negative binomial	478
14.6	Mixed-effects negative binomial models	488
14.6.1	Random-intercept negative binomial models	488
14.6.2	Non-parametric random-intercept negative binomial	494
14.6.3	Random-coefficient negative binomial models	496
14.7	Multilevel models	500
15	Bayesian negative binomial models	502
15.1	Bayesian versus frequentist methodology	502
15.2	The logic of Bayesian regression estimation	506
15.3	Applications	510

<i>Appendix A: Constructing and interpreting interaction terms</i>	520
<i>Appendix B: Data sets, commands, functions</i>	530
<i>References and further reading</i>	532
<i>Index</i>	541