

Models compared, starting from the left:

1. Zero-inflated Poisson GLM
2. Zero-inflated Poisson GAM
3. Zero-inflated Poisson LRMoE, with 4 latent classes
4. Zero-inflated Gamma Count LRMoE, with 4 and 5 latent classes
5. Hurdle Poisson LRMoE, with 3, 4, 5, 6 latent classes for both Bernoulli and (truncated) Poisson parts

The following provides chi-square statistics for the models considered:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Claim#	observed		ZIPoisson GLM	ZIPoisson GAM		4 zi-Poisson LRMoE	4 zi-Gamma Count LRMoE	5 zi-Gamma Count LRMoE		3 Hurdle Poisson LRMoE	4 Hurdle Poisson LRMoE	5 Hurdle Poisson LRMoE	6 Hurdle Poisson LRMoE
0	397779		397697.76	397778.97		397859.24	397789.42	397786.6		397787.15	397779.03	397796.8	397768.12
1	14633		14785.76	14630.58		14467.26	14622.39	14624.71		14621.35	14625.43	14615.56	14638.57
2	726		660.18	728.44		809.9	721.08	722.19		717.3	723.78	712.4	715.06
3	28		24.44	29.84		31.61	35.86	35.15		40.83	38.7	41.7	44.48
4+	3		0.86	1.17		0.99	0.25	0.35		2.37	2.06	2.54	2.77
total	413169		413169	413169		413169	413169	413169		413169	413169	413169	413169
		chi square	14.00078021	2.984339507		15.09957794	32.01434098	21.54363988		4.314006	3.398057394	4.865502488	6.294789349
		left out last	8.675663933	0.122031814		11.01866885	1.764340983	1.479354165		4.146537645	2.969125355	4.782195401	6.275691876
		binned last	9.441293918	0.008576594		10.6849187	0.764667764	0.595367028		3.560340541	2.347763732	4.243659315	5.758416934

The row “left out last” refers to chi-square not including claim number = 4+, since all models generally have a poor fit for this category.

The row “binned last” refers to adjusted chi-square with claim number = 3 and 4+ combined, following the convention that requires expected counts to be at least 5. (briefly mentioned here: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2703493/> still looking for formal justification)

It is obvious that the zero-inflated Poisson GAM performed the best, followed by the zero-inflated Gamma Count LRMoe models. However even these models have missed out the last category claim number = 4+. The improvement obtained moving from zero-inflated Poisson to hurdle Poisson is quite significant: chi-square statistics improved for both adjusted and unadjusted, and a better fit is obtained for the last category.