

Zero Inflated Negative Binomial Regression

Zero Inflated Negative Binomial Regression

=====
Dataset: D:\DATA ANALYSIS\H Regression Tests and Models\Zero Inflated Negative Binomial Regression\dataset.csv
Rows in original dataset: 649
Rows used in model: 649
Outcome: absences
Count predictors: G1, G2, G3, age, studytime, failures, school, sex
Zero-process predictors: failures, studytime, G1, G2, G3
Estimation mode: direct_zinb
Method note: Direct statsmodels ZeroInflatedNegativeBinomialP fitted with logit inflation and NB2 variance.

Model fit summary

AIC: 2935.6222
BIC: 3007.2291
Observed zero count: 244
Observed zero rate: 0.3760
Model-implied expected zero count: 241.1389
Mean predicted probability of zero: 0.3716
MAE: 3.2096
RMSE: 4.3729
Observed-fitted correlation: 0.3335

Interpretation guide

Count-component coefficients use a log link; exp(coefficient) is a count rate ratio.
Inflation-component coefficients use a logit link; exp(coefficient) is an odds ratio for excess-zero membership.
A p-value below 0.05 and a 95% exponentiated confidence interval not crossing 1 provide stronger evidence of a predictor effect.

Model Fit Statistics

model	n	log_likelihood	aic	bic	df_model	df_residual	observed_zero_count	observed_zero_mean	mean_observed_count	mean_fitted_count	method_note
Negative Binomial	649.0000	-1451.8111	2935.6222	3007.2291	8.0000	640.0000	244.0000	0.3760	Direct stats	3.6595	Zero Inflation
Poisson GLM comparison	649.0000	-2221.7328	4461.4655	-1073.9446	8.0000	640.0000	244.0000	0.3760	3.6595		Poisson log-link comparison
Negative Binomial GLM comparison	649.0000	-1528.5943	3075.1887	-3218.6268	8.0000	640.0000	244.0000	0.3760	3.6595		Negative-binomial log-link comparison

Zero-Count Diagnostics

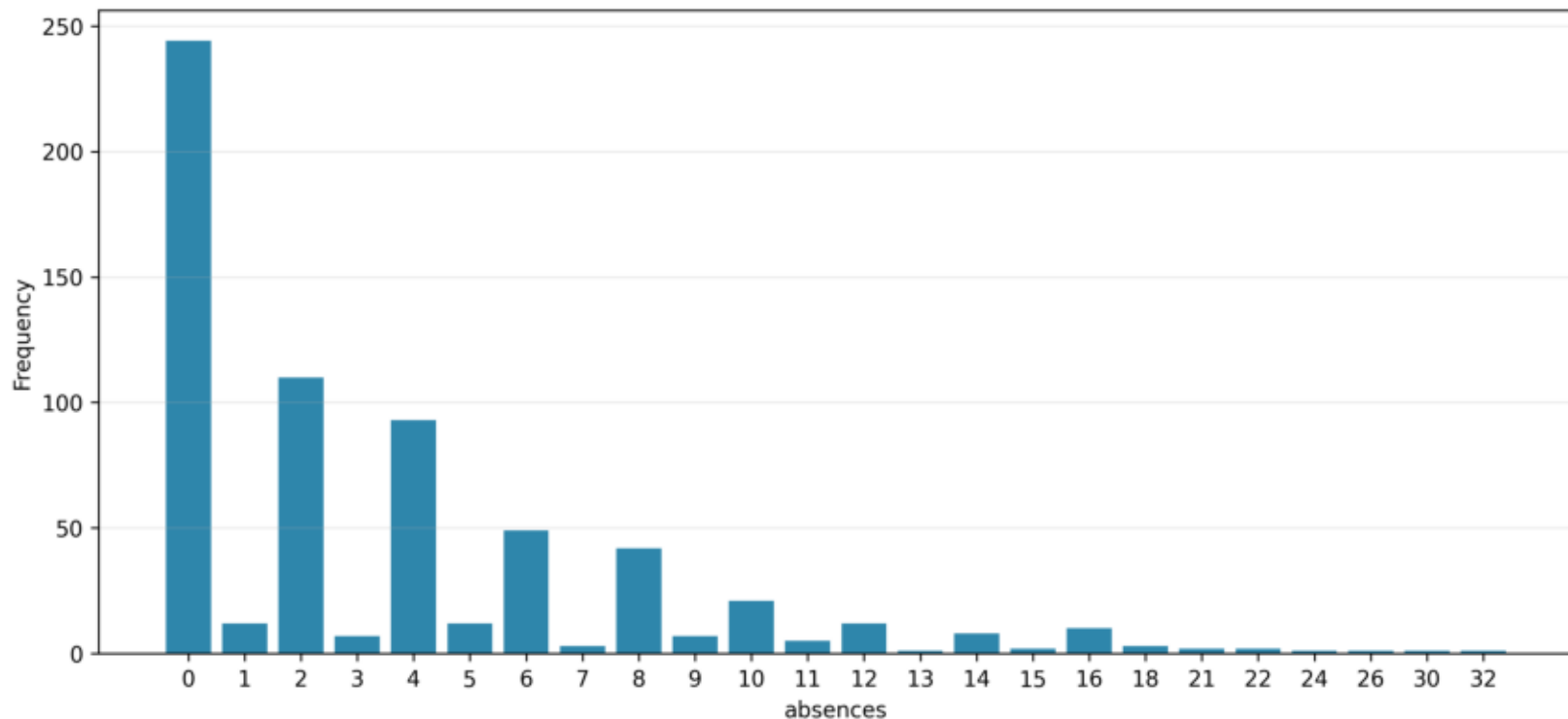
n	observed_zero_count	observed_nonzero_count	observed_zero_rate	expected_zero_count_from_mean	mean_predicted_probability	rounded_fitted_zero_rate	note
649.0000	244.0000	405.0000	0.3760	241.1389	0.5746	0.0000	Observed and model-implied zero frequencies should be reasonably close

Coefficient Table

	model	component	term	coefficient	standard_error	z_value	p_value	ci95_lower	ci95_upper	exp_coefficient	exp_ci95_lower	exp_ci95_upper	decision_alpha_0.05
cd	Negative Binomial	multinomial logit coefficients	const	-0.7673	0.1078	-7.1208	0.0000	-0.9785	-0.5561	0.4643	0.3759	0.5734	Statistically significant
cd	Negative Binomial	multinomial logit coefficients	failures	-0.1033	0.1184	-0.8718	0.3833	-0.3354	0.1289	0.9019	0.7151	1.1371	Not statistically significant
cd	Negative Binomial	multinomial logit coefficients	studytime	0.1008	0.0978	1.0299	0.3031	-0.0910	0.2925	1.1060	0.9130	1.3396	Not statistically significant
cd	Negative Binomial	multinomial logit coefficients	G1	0.7229	0.2112	3.4235	0.0006	0.3090	1.1368	2.0604	1.3621	3.1167	Statistically significant
cd	Negative Binomial	multinomial logit coefficients	G2	0.5874	0.3082	1.9056	0.0567	-0.0168	1.1915	1.7993	0.9834	3.2921	Not statistically significant
cd	Negative Binomial	multinomial logit coefficients	G3	-1.1996	0.3051	-3.9324	0.0001	-1.7975	-0.6017	0.3013	0.1657	0.5479	Statistically significant
cd	Negative Binomial	mean log-link coefficients	const	1.8142	0.0569	31.8640	0.0000	1.7026	1.9258	6.1361	5.4881	6.8604	Statistically significant
cd	Negative Binomial	mean log-link coefficients	G1	0.0486	0.0853	0.5695	0.5690	-0.1186	0.2158	1.0498	0.8881	1.2406	Not statistically significant
cd	Negative Binomial	mean log-link coefficients	G2	0.0758	0.1444	0.5252	0.5995	-0.2072	0.3588	1.0788	0.8129	1.4316	Not statistically significant
cd	Negative Binomial	mean log-link coefficients	G3	-0.3349	0.1563	-2.1424	0.0322	-0.6413	-0.0285	0.7154	0.5266	0.9719	Statistically significant
cd	Negative Binomial	mean log-link coefficients	age	0.1490	0.0412	3.6126	0.0003	0.0681	0.2298	1.1606	1.0705	1.2583	Statistically significant
cd	Negative Binomial	mean log-link coefficients	studytime	-0.0907	0.0409	-2.2160	0.0267	-0.1709	-0.0105	0.9133	0.8429	0.9896	Statistically significant
cd	Negative Binomial	mean log-link coefficients	failures	-0.0009	0.0417	-0.0217	0.9827	-0.0827	0.0809	0.9991	0.9206	1.0841	Not statistically significant
cd	Negative Binomial	mean log-link coefficients	school_MS	-0.5268	0.0910	-5.7859	0.0000	-0.7052	-0.3483	0.5905	0.4940	0.7059	Statistically significant
cd	Negative Binomial	mean log-link coefficients	sex_M	0.0050	0.0810	0.0616	0.9509	-0.1537	0.1636	1.0050	0.8576	1.1776	Not statistically significant
cd	Negative Binomial	binomial overdispersion	alpha	0.3831	0.0515	7.4333	0.0000	0.2821	0.4841	1.4668	1.3259	1.6226	Statistically significant

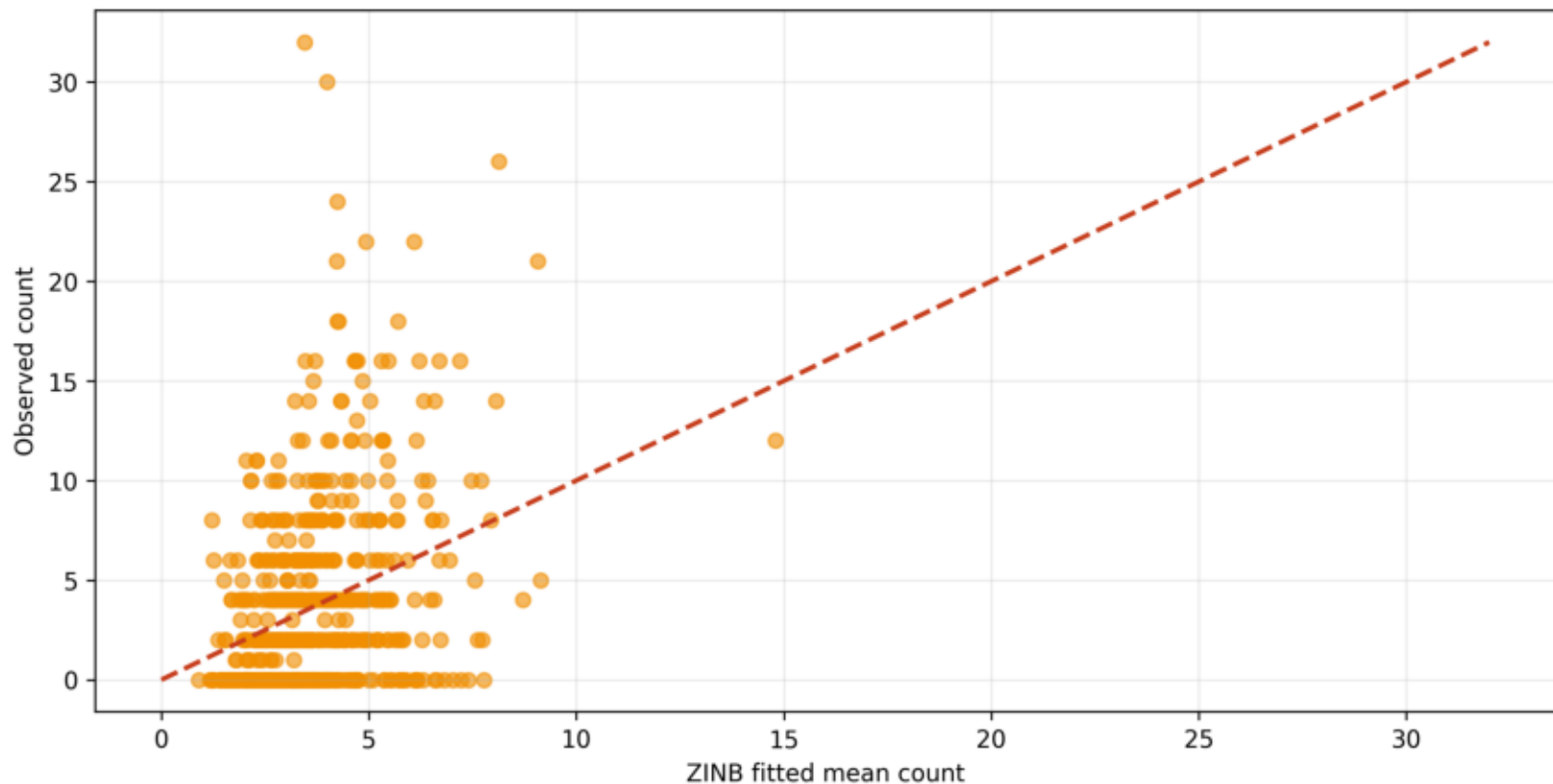
Count Outcome Distribution

A large frequency at zero supports evaluating a zero-inflated count model.



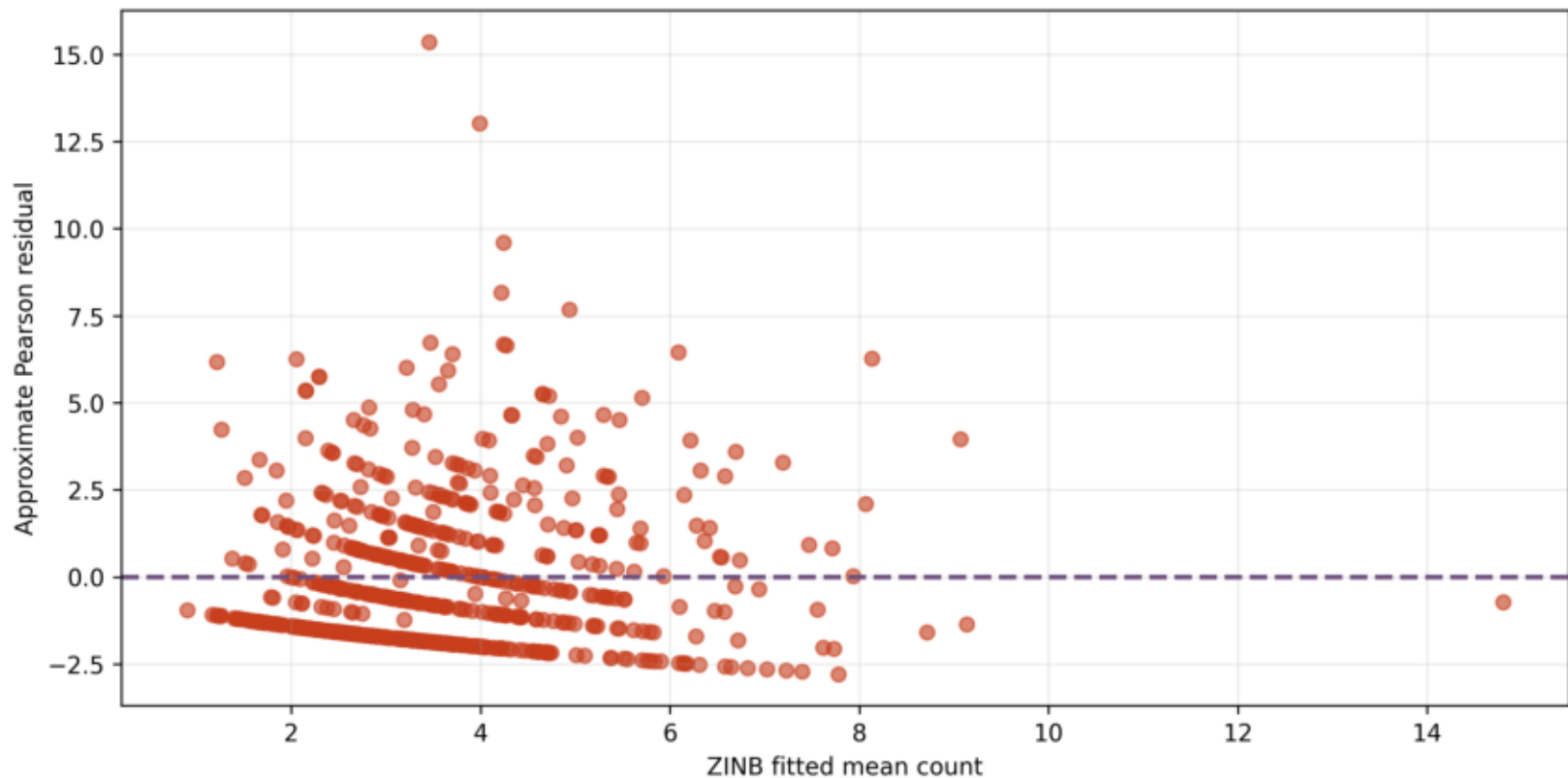
Observed vs Fitted Counts

Points near the diagonal indicate closer agreement between observed and fitted counts.



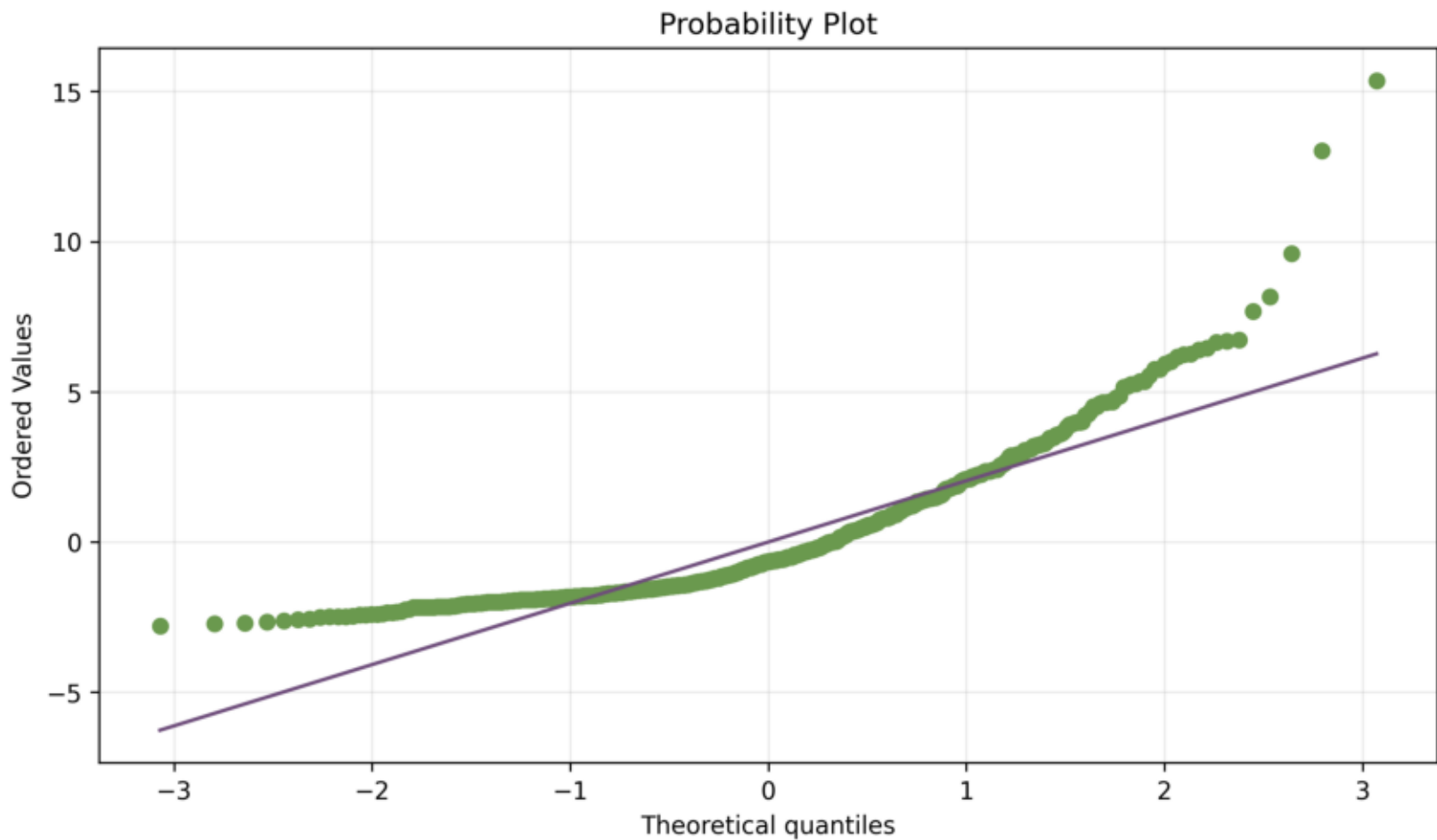
Pearson Residuals vs Fitted Values

Residuals should be distributed around zero without a strong systematic pattern.



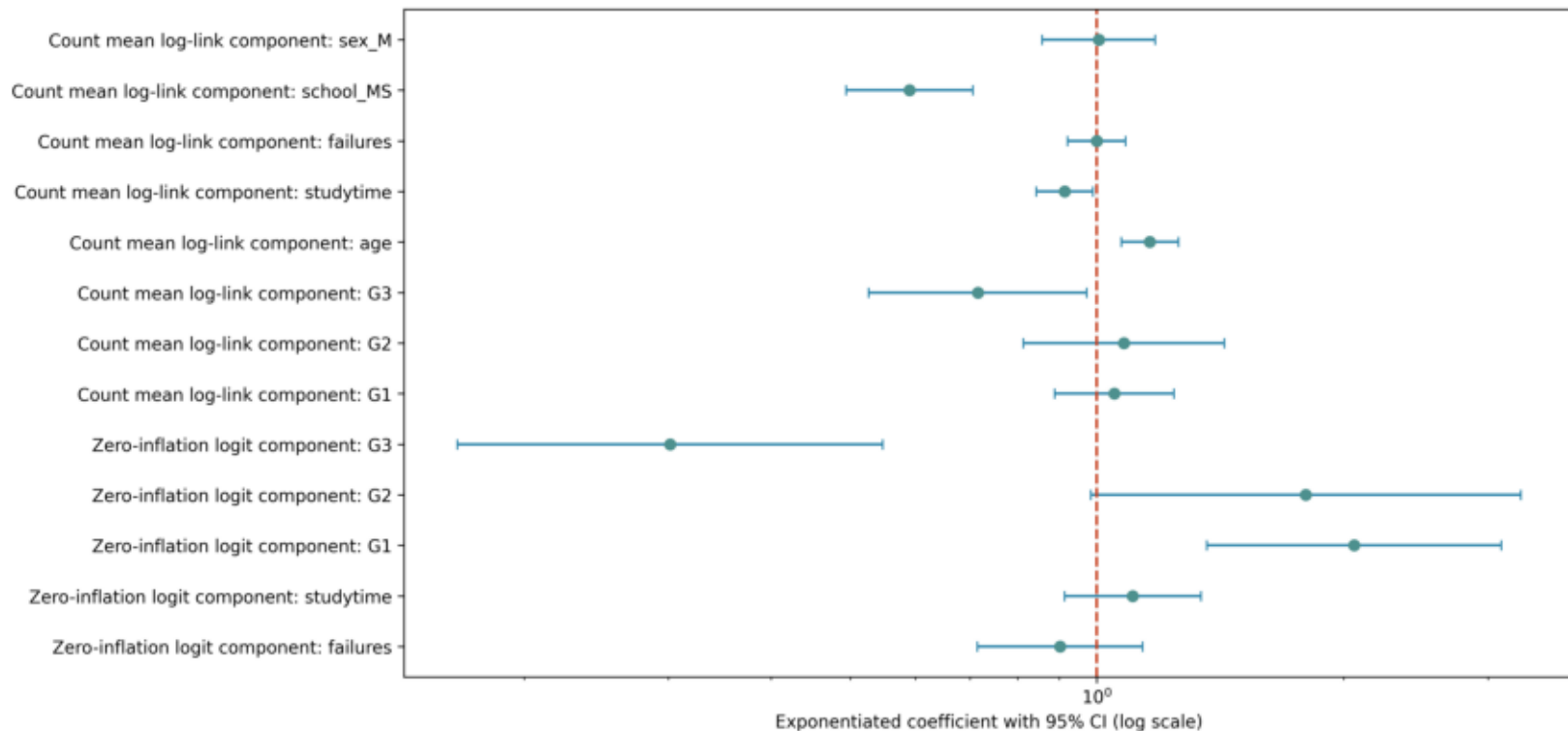
Residual Q-Q Plot

Large tail departures can identify observations that are not well represented by the fitted count model.



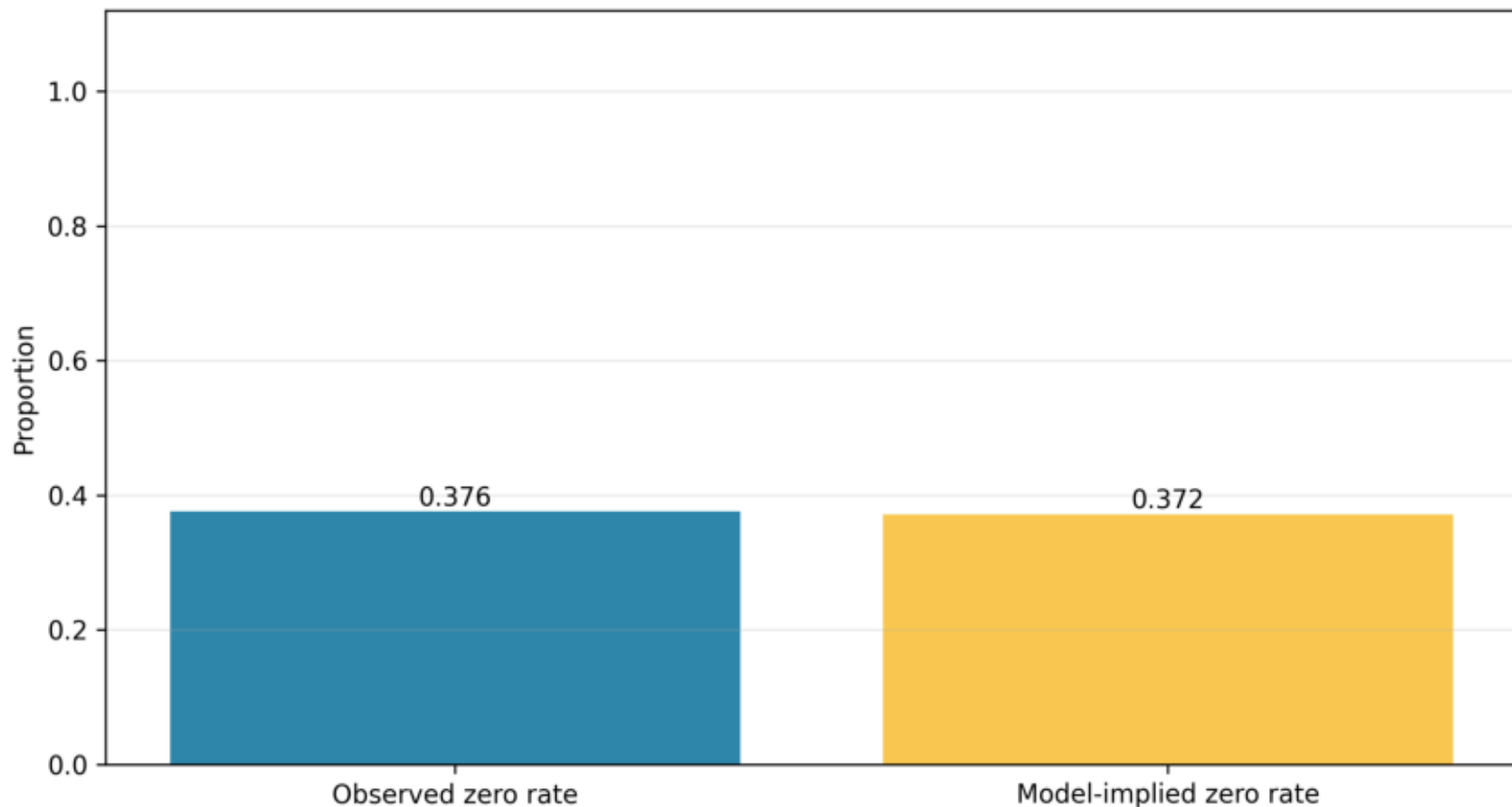
ZINB Exponentiated Coefficients

Count-part values are rate ratios; inflation-part values are odds ratios for excess-zero membership.



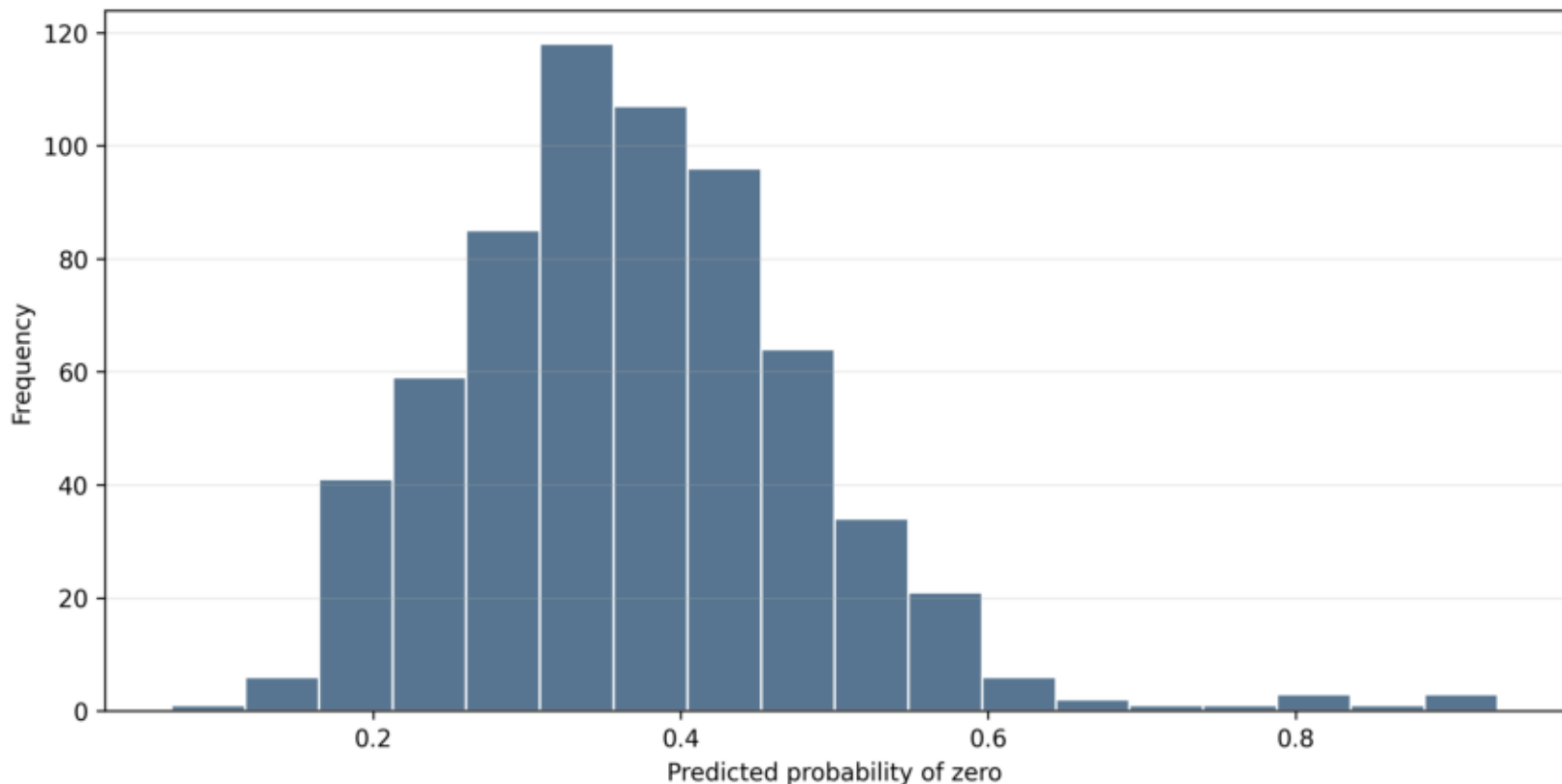
Observed and Model-Implied Zero Rates

Closer bars indicate that the model reproduces the observed frequency of zeros more closely.



Predicted Zero-Probability Distribution

The inflation component estimates which observations are more likely to belong to the excess-zero process.



Count Model AIC Comparison

Use AIC only when models are fitted to comparable observations and likelihoods are available.

