Int. J. Anal. Appl. (2026), 24:213

International Journal of Analysis and Applications

# Diagnostic Tools for Distributional Negative Binomial Regression: Residuals, Component Plots, and Model Comparison

**Mohieddine Rahmouni**[*]

*Applied College, King Faisal University, Al-Ahsa, Saudi Arabia*

[*]*Corresponding author:* mrahmouni@kfu.edu.sa

**Abstract.** Distributional negative binomial (NB) regression lets both the conditional mean and the overdispersion parameter depend on covariates, but applied work still lacks a compact, model-based diagnostic workflow for this additional flexibility. We develop three complementary post-fit tools. First, randomized quantile residuals (RQRs) provide a single normal-scale check of the fitted count distribution and mean equation: with known parameters they are exactly standard normal under correct specification, while with estimated parameters they are best interpreted as approximate graphical diagnostics. In a targeted simulation, the RQR QQ plot detects omitted mean structure (KS $D = 0.042$, $p = 0.002$ at $n = 2000$) and returns to the reference line when the mean is correctly specified (KS $D = 0.010$, $p = 0.988$). Second, we introduce score-based component-plus-residual plots for the dispersion equation, allowing analysts to identify slope misfit, nonlinear effects, and stratum-level departures in the log-dispersion predictor. Third, we compare AIC, BIC, and likelihood-ratio checks for choosing among candidate dispersion specifications. In the simulation design considered here, BIC and likelihood-ratio checks perform best once $n \geq 300$; at $n = 1000$ both select the true dispersion specification in 99% of replications, whereas AIC selects an overfitted specification in 15%. An application to `NMES1988` physician-visit counts illustrates the full workflow and shows why global RQR diagnostics should be combined with targeted checks of the dispersion equation.

## 1. Introduction

The negative binomial (NB) regression model is a cornerstone of applied count-data analysis. In its standard form it accommodates extra-Poisson variation through a single constant dispersion parameter $\alpha$, yielding conditional variance $\mu_i + \mu_i^2/\alpha$ [1, 2]. A natural extension allows $\alpha_i$ to depend on covariates through its own log-linear regression equation, producing the distributional NB model studied by Smyth (1989) [3] and accommodated by the GAMLSS framework [4,5]. The model has seen growing uptake in settings where the *regularity* of counts—not just their expected

---

Received: Apr. 27, 2026.

2020 *Mathematics Subject Classification.* 62J12; 62F03; 62F10; 62P10.

*Key words and phrases.* count regression; negative binomial; distributional regression; randomized quantile residuals; component-plus-residual plots; model selection; overdispersion.

https://doi.org/10.28924/2291-8639-24-2026-213

ISSN: 2291-8639

level—is substantively important. For example, two individuals or sites may have similar expected counts but very different conditional variability, and this difference may itself be of scientific or policy interest.

Despite this uptake, the diagnostic toolkit for distributional NB regression lags well behind its adoption. For the constant-dispersion NB, practitioners have Pearson and deviance residuals, half-normal plots, and goodness-of-fit tests [1,2]. For the varying-dispersion extension the situation is weaker. Standard residuals do not account for the observation-specific $\hat{\alpha}_i$ in their standardization, so their sampling distributions are non-standard. There are no established plots for assessing the dispersion equation's functional form. And the literature offers no systematic comparison of procedures for choosing among candidate dispersion specifications.

The contribution of this paper is therefore practical as well as methodological. We do not propose a new estimator; instead, we provide diagnostics that can be run after a standard distributional NB fit and interpreted in a sequence that mirrors applied model building. This matters because misspecification can enter through different channels: mean misspecification changes fitted conditional probabilities, dispersion misspecification changes the variance structure, and overfitted dispersion formulas can create spurious heterogeneity.

The proposed workflow has three steps.

1. *Randomized quantile residuals* (Section 3): check whether the NB family and the mean equation are adequate.
2. *Partial residual plots for the dispersion equation* (Section 4): assess whether the log-linear dispersion specification is adequate and identify nonlinear effects.
3. *Model comparison* (Section 5): use information criteria as the primary ranking tools and likelihood-ratio tests (LRTs) as confirmatory checks against constant dispersion.

The application (Section 6) demonstrates the full three-step workflow on the `NMES1988` physician-visit data of Deb and Trivedi (1997) [6]. Each diagnostic can be used independently, but the recommended order is sequential: first check the fitted distribution globally, then inspect the dispersion linear predictor locally, and only then compare alternative dispersion specifications.

## 2. Distributional NB regression

Let $Y_i \in \{0, 1, 2, \ldots\}$ be a count response for observation $i = 1, \ldots, n$, with mean covariates $x_i \in \mathbb{R}^p$ and dispersion covariates $z_i \in \mathbb{R}^q$. The distributional NB model specifies

$$Y_i \mid x_i, z_i \sim \mathrm{NB}(\mu_i, \alpha_i), \qquad \log \mu_i = \beta_0 + x_i^\top \beta, \qquad \log \alpha_i = \eta_0 + z_i^\top \eta, \tag{2.1}$$

with probability mass function

$$\Pr(Y_i = y \mid x_i, z_i) = \frac{\Gamma(y + \alpha_i)}{\Gamma(\alpha_i)\, y!} \left( \frac{\alpha_i}{\alpha_i + \mu_i} \right)^{\alpha_i} \left( \frac{\mu_i}{\alpha_i + \mu_i} \right)^{y}, \quad y = 0, 1, 2, \ldots, \tag{2.2}$$

so $\mathbb{E}(Y_i) = \mu_i$ and $\mathrm{Var}(Y_i) = \mu_i + \mu_i^2/\alpha_i$. Larger $\alpha_i$ implies less overdispersion. Setting $\eta = 0$ recovers the constant-dispersion NB. The log-likelihood

$$\ell(\theta) = \sum_{i=1}^{n}\Big[\log\Gamma(Y_i+\alpha_i) - \log\Gamma(\alpha_i) - \log(Y_i!) + \alpha_i \log\tfrac{\alpha_i}{\alpha_i+\mu_i} + Y_i \log\tfrac{\mu_i}{\alpha_i+\mu_i}\Big] \tag{2.3}$$

is smooth under this parameterisation and can be maximised by standard numerical methods. Under the regularity and identifiability conditions for NB and mixed-Poisson regression, the MLE $\hat{\theta}$ is consistent and asymptotically normal [7]. Throughout the paper we use the mean–shape parameterisation in (2.2); software that reports reciprocal dispersion, such as some GAMLSS families, must therefore be translated before computing residuals or partial residuals. This convention is used consistently in the derivations, simulations, and code.

## 3. RANDOMIZED QUANTILE RESIDUALS

3.1. **Definition and exact normality.** For a continuous response the probability integral transform maps the fitted CDF to a $\mathrm{Uniform}(0,1)$ variate, and $\Phi^{-1}$ then gives a standard normal. For a discrete response the CDF has jumps, and the transform is not unique. Dunn and Smyth (1996) [8] resolve this by randomizing uniformly within each jump: for $F_i = F(\cdot;\mu_i,\alpha_i)$ the CDF of $\mathrm{NB}(\mu_i,\alpha_i)$, define

$$a_i = F_i(Y_i - 1)\ (0 \text{ if } Y_i = 0), \quad b_i = F_i(Y_i), \quad U_i \sim \mathrm{Uniform}(a_i,b_i), \quad r_i^{\mathrm{Q}} = \Phi^{-1}(U_i). \tag{3.1}$$

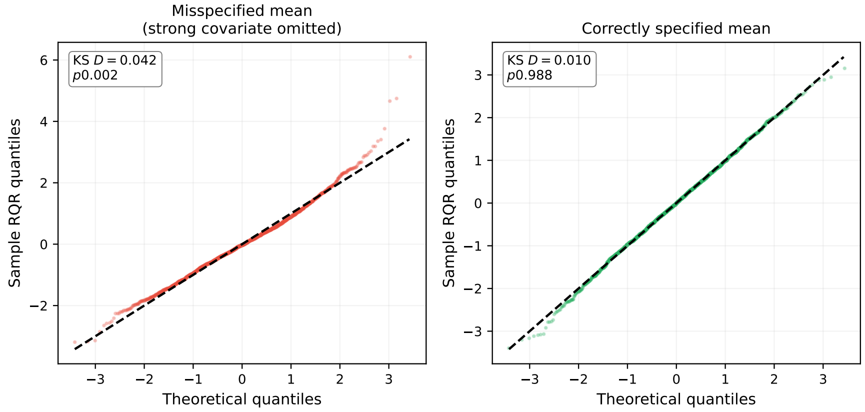
**Proposition 3.1.** *Under the correctly specified model* (2.1)–(2.2) *with parameters* $\theta_0$*, the RQR* $r_i^{\mathrm{Q}}$ *defined in* (3.1) *satisfies* $r_i^{\mathrm{Q}} \sim \mathcal{N}(0,1)$ *exactly, and* $(r_1^{\mathrm{Q}},\ldots,r_n^{\mathrm{Q}})$ *are mutually independent conditional on* $(x_i,z_i)$*.*

The proof is given in Appendix A. Proposition 3.1 is a known-parameter result: it holds exactly for any finite $n$, not merely asymptotically, despite the integer-valued response. In applications the parameters are replaced by estimates, so fitted RQRs are best regarded as approximately standard normal residuals. QQ plots should therefore be read primarily as graphical diagnostics, and formal Kolmogorov–Smirnov or Shapiro–Wilk tests should be treated as approximate checks unless calibrated by parametric simulation. One practical calibration is a parametric bootstrap: simulate responses from the fitted model, refit the model, recompute RQRs, and overlay pointwise QQ envelopes. For reproducibility, the random seed used for RQR randomization should be fixed. In the simulations and application below, the mid-CDF approximation $\tilde{u}_i = (a_i + b_i)/2$ gave visually similar QQ plots without randomization; for highly discrete or zero-heavy data, randomized RQRs or simulation envelopes are preferable.

3.2. **What RQR detects—and what it does not.** The RQR is a function of the *marginal* CDF of $Y_i$, which integrates over the full conditional distribution. This has two consequences.

Mean misspecification is detected reliably. If the mean equation omits a covariate or uses the wrong functional form, $\hat{\mu}_i$ is biased for some observations. The RQR will then depart from $\mathcal{N}(0,1)$: the QQ plot shows systematic curvature or heavy tails, and approximate formal checks may reject. Figure 1 confirms this: at $n = 2000$, omitting a predictor with true slope $\beta = 0.8$ produces a QQ plot with a pronounced S-curve and extreme upper-tail outliers (KS $D = 0.042$, $p = 0.002$), while fitting the correct mean model restores the plot to the identity (KS $D = 0.010$, $p = 0.988$).

Moderate dispersion misspecification may not be detected. If dispersion genuinely varies with covariates but a constant-dispersion model is fitted, the departure in the RQR is attenuated when the true $\alpha_i$ varies modestly: the marginal distribution of $Y_i$ is then close to a constant-$\alpha$ NB at the geometric mean of the true $\alpha_i$, and the constant-dispersion RQR QQ plot may look approximately correct. The practical implication is that the RQR is most useful as a check on the distributional *family* and the *mean* equation; for the dispersion equation specifically, the partial residual plots of Section 4 are the appropriate tool.



Misspecified mean
(strong covariate omitted)

Correctly specified mean

Figure 1. RQR QQ plots demonstrating detection of mean misspecification ($n = 2000$, simulated). Left: a strong mean covariate ($\beta = 0.8$) is omitted; the plot shows a pronounced S-curve with extreme upper-tail outliers (KS $D = 0.042$, $p = 0.002$). Right: the correct mean is fitted; the plot is indistinguishable from the identity (KS $D = 0.010$, $p = 0.988$). Dashed line: 45° reference.

## 4. Partial residual plots for the dispersion equation

4.1. **Motivation, definition, and interpretation.** In standard regression, component-plus-residual (CPR) plots reveal the partial effect of a single predictor and detect nonlinearity [9]. For the mean equation of (2.1), standard CPR plots apply directly, using the working residuals from the last IRLS step. For the *dispersion equation* an analogue requires a suitable dispersion working residual—one that plays the same structural role in the score equation for $\eta$ as the IRLS residual plays in the score equation for $\beta$.

The diagnostic below is score-based rather than an exact IRLS component plot. It uses the dispersion score residual as a local measure of whether the observed counts are more or less dispersed than the fitted model predicts after the mean has been accounted for. The score contribution for $\eta$ at the fitted model $\hat{\theta}$ is

$$\frac{\partial \ell_i}{\partial \eta} = z_i \hat{\alpha}_i \underbrace{\Big[\psi(Y_i + \hat{\alpha}_i) - \psi(\hat{\alpha}_i) + \log \tfrac{\hat{\alpha}_i}{\hat{\alpha}_i + \hat{\mu}_i} + 1 - \tfrac{Y_i + \hat{\alpha}_i}{\hat{\alpha}_i + \hat{\mu}_i}\Big]}_{=: r_i^{\phi}},$$

where $\psi(\cdot)$ is the digamma function. We call $r_i^{\phi}$ the *dispersion working residual*. The MLE score conditions enforce $\sum_i \hat{\alpha}_i r_i^{\phi} = 0$ and $\sum_i z_i \hat{\alpha}_i r_i^{\phi} = \mathbf{0}$, so $r_i^{\phi}$ has weighted mean zero—mirroring the zero-mean property of IRLS residuals—and is the natural analogue of the mean working residual for the dispersion equation.

The *partial residual* for the $j$-th dispersion covariate is

$$\hat{p}_{ij} = \hat{\eta}_j z_{ij} + r_i^{\phi}. \tag{4.1}$$

Plotting $\hat{p}_{ij}$ against $z_{ij}$ with a smoother and the reference line $\hat{\eta}_j z_{ij}$ gives a visual check of the fitted contribution of $z_{ij}$ to $\log \alpha_i$. Three reading rules guide interpretation: (i) if the smoother tracks the reference line, the log-linear specification is adequate at the resolution of the plot; (ii) if the smoother is systematically steeper or flatter, the linear effect may be misestimated or confounded with other dispersion covariates; (iii) if the smoother curves, a nonlinear term—such as a spline or a knot—is warranted.

An important advantage over a simple scatter of $r_i^{\phi}$ against $z_{ij}$ is that the partial residual incorporates the fitted slope into the vertical axis: the analyst can directly compare the smoother with the reference line and assess *magnitude* as well as direction. For binary $z_{ij}$, the plot reduces to a two-stratum display. Under a well-specified log-linear dispersion equation, the stratum-level smoother should be broadly consistent with the fitted component $\hat{\eta}_j z_{ij}$, but exact equality of unweighted group means is not guaranteed after adjustment for other covariates. The plot should therefore be read as a model-checking device rather than as a formal test; large apparent departures can be followed by refitting with the suggested nonlinear or interaction terms and comparing the resulting likelihood, BIC, and RQR plots.

4.2. **Relationship to the null-model residual plot.** If $r_i^{\phi}$ is computed at the *null* (constant-dispersion) MLE rather than the varying-dispersion MLE, the plot (4.1) reduces to a scatter of null dispersion residuals against $z_j$ with no reference line. Such a plot is useful as a *pre-fit* diagnostic, motivating the fit of a varying-dispersion model. Computing $r_i^{\phi}$ at the varying-dispersion MLE and overlaying the reference line provides the corresponding *post-fit* diagnostic, validating the fitted specification. The two plots therefore bracket the modelling step: one motivates fitting the varying-dispersion model, the other validates it.

## 5. Model comparison: simulation study

5.1. **Design.** The true model uses $p = 3$ mean covariates (two continuous, one binary, with slopes $\beta = (0.3, -0.2, 0.4)^{\top}$ and intercept $\beta_0 = 0.8$) and a single active dispersion covariate $z_1$ distinct from

$x$ (slope $\eta_1 = 0.55$, baseline $\exp(\eta_0) = 1.5$). All covariates are drawn independently. We compare five candidate dispersion specifications: $M_0$ (constant, $\eta = 0$); $M_1$ ($z = z_1$, **true**); $M_2$ ($z = (z_1, z_2)$, overfitted); $M_3$ ($z = z_2$, wrong covariate); $M_x$ ($z = x$, shared). The three procedures are: **AIC** and **BIC** (which rank the fitted candidate models by the corresponding criterion); and the **LRT**, used as a confirmatory test of improvement over the constant-dispersion null. Specifically, for each varying-dispersion candidate we test $H_0 : \eta = 0$ at $\alpha = 0.05$ using $2[\ell(\hat{\theta}_{\text{alt}}) - \ell(\hat{\theta}_0)] \sim \chi^2_q$ and record the lowest rejecting $p$-value, with $M_0$ retained if no candidate rejects. This LRT rule is included as a pragmatic benchmark rather than as a general selector among non-nested alternatives. Sample sizes are $n \in \{100, 300, 1000\}$ with $R \in \{200, 200, 150\}$ replications.

5.2. **Results.** Table 1 reports the selection rates.

Correct identification. At $n = 300$, AIC selects $M_1$ in 80%, BIC in 91%, and the LRT in 90%. At $n = 1000$ the rates are 85%, 99%, and 99%.

Overfitting. AIC selects the spurious $M_2$ in 15% of replications at $n = 1000$, reflecting its fixed $2k$ penalty that does not grow with $n$. BIC and the LRT each reduce this to 1%.

Underfitting at small $n$. At $n = 100$, BIC selects $M_0$ in 44% of replications (vs. 12% for AIC and 34% for the LRT), reflecting its well-known conservatism.

Practical recommendation. The results support a practical hierarchy, conditional on this simulation design. **BIC** and the **LRT** are most reliable once $n \geq 300$, with high correct-identification rates and little overfitting. The LRT remains best interpreted as a confirmatory check for improvement over constant dispersion rather than as a general ranking method for non-nested alternatives. **AIC** can be useful as an exploratory criterion, especially at small $n$, but it should be used with caution in larger samples because its fixed penalty is more prone to retaining spurious dispersion covariates. All requested simulation replications completed with no fitting failures.

Table 1. Model selection rates (%) over $R$ replications. $M_0$: null; $M_1$: true; $M_2$: overfitted; $M_3$: wrong covariate; $M_x$: shared ($z = x$). Bold: true model. For the LRT columns, the reported model is the rejecting alternative with the smallest $p$-value against $M_0$, or $M_0$ if no alternative rejects.

| | AIC (%) | | | | | BIC (%) | | | | | LRT (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ ($R$) | $M_0$ | $\mathbf{M_1}$ | $M_2$ | $M_3$ | $M_x$ | $M_0$ | $\mathbf{M_1}$ | $M_2$ | $M_3$ | $M_x$ | $M_0$ | $\mathbf{M_1}$ | $M_2$ | $M_3$ | $M_x$ |
| 100 (200) | 12 | **59** | 18 | 4 | 7 | 44 | **50** | 4 | 2 | 0 | 34 | **46** | 11 | 2 | 6 |
| 300 (200) | 1 | **80** | 18 | 0 | 1 | 7 | **91** | 2 | 0 | 0 | 2 | **90** | 6 | 0 | 1 |
| 1000 (150) | 0 | **85** | 15 | 0 | 0 | 0 | **99** | 1 | 0 | 0 | 0 | **99** | 1 | 0 | 0 |

## 6. Application: `NMES1988` physician-visit counts

The `NMES1988` data [6] comprises $n = 4{,}406$ non-institutionalised Medicare beneficiaries aged 66 and older from the 1987–88 U.S. National Medical Expenditure Survey, distributed with R's

`AER` package. The outcome `visits` is the number of physician office visits in the prior year (mean 5.77, variance 45.68, variance-to-mean ratio 7.91). We use clinical predictors (`chronic`, `hlth_poor`, `hlth_excellent`, `adl_limited`, `age_c`) in the mean equation. For dispersion we consider two scientifically plausible channels: demographic/insurance covariates (`female`, `private`, `medicaid`, `age_c`) and a shared clinical specification using the mean covariates. The goal is not to claim a final health-economic model, but to show how the diagnostics separate global distributional adequacy, local dispersion-equation checks, and specification choice.

6.1. **Step 1: Assess overall model adequacy via RQR.** We fit both the constant-dispersion NB ($\hat{\alpha} = 1.127$) and the varying-dispersion NB with demographic dispersion covariates. Figure 2 shows normal QQ plots of the RQR for each fit. Both are close to the diagonal, providing no strong evidence against the NB family for these data. The constant-dispersion model shows mild right-tail excess beyond theoretical quantile 3, consistent with a handful of high-visit observations; this excess is slightly reduced under the varying-dispersion model. As a supplement to the QQ plots, a Shapiro–Wilk test on a fixed random subsample of 500 observations returns $W = 0.994$ ($p = 0.065$) for the constant-dispersion model and $W = 0.994$ ($p = 0.064$) for the varying-dispersion model. We use the subsample test to avoid making a large-sample normality test the primary diagnostic; the fitted-parameter caveat in Section 3 also applies. The two QQ plots are nearly indistinguishable, as expected from Section 3.2: the fitted $\hat{\alpha}_i$ range (0.7–2.0) is moderate, and moderately varying dispersion does not produce a detectable departure in the marginal RQR. We conclude that the NB family is broadly adequate, with mild upper-tail excess; the partial residual plots below assess the dispersion equation.

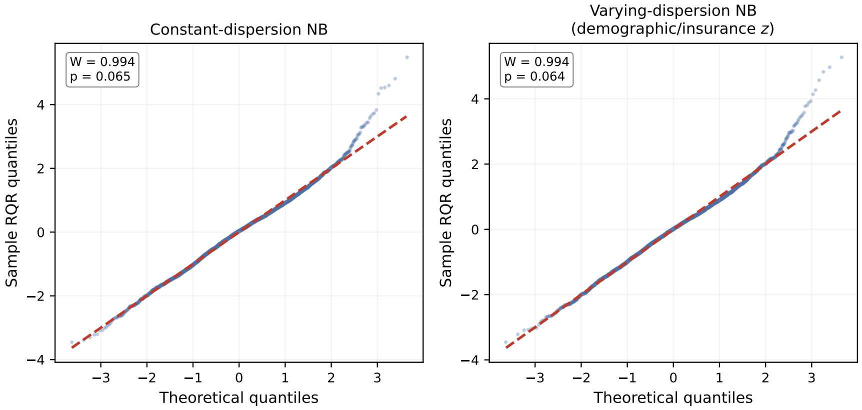


Figure 2. Normal QQ plots of RQR for the constant-dispersion NB (left) and varying-dispersion NB (right), fitted to `NMES1988`. Dashed line: 45° reference. The 500-observation Shapiro–Wilk checks do not reject at the 5% level ($p = 0.065$ and $p = 0.064$), but they are interpreted as supplementary to the graphical diagnostic.

6.2. **Step 2: Assess the dispersion equation via partial residuals.** Figure 3 shows the partial residual plots (4.1) for the four dispersion covariates under the varying-dispersion NB fit.

*Female* ($\hat{\eta} = 0.203$): the smoother is close to the fitted binary contrast, supporting a simple log-linear contribution for this covariate.

*Private insurance* ($\hat{\eta} = 0.615$): the loess smoother tracks the reference line closely across the full covariate range, validating the log-linear specification.

*Medicaid* ($\hat{\eta} = 0.462$): the fitted component is positive, but the smoother at $z = 1$ lies below the fitted reference. This suggests that the log-linear model may modestly overstate the medicaid dispersion effect in the enrolled stratum.

*Age* ($\hat{\eta} = 0.073$): the smoother is near zero for ages below 83 (`age_c`$\leq 3$) but curves downward at older ages, suggesting mild nonlinearity that the log-linear model cannot capture; a quadratic term, spline, or piecewise-linear knot at `age_c`$= 3$ could improve the fit.

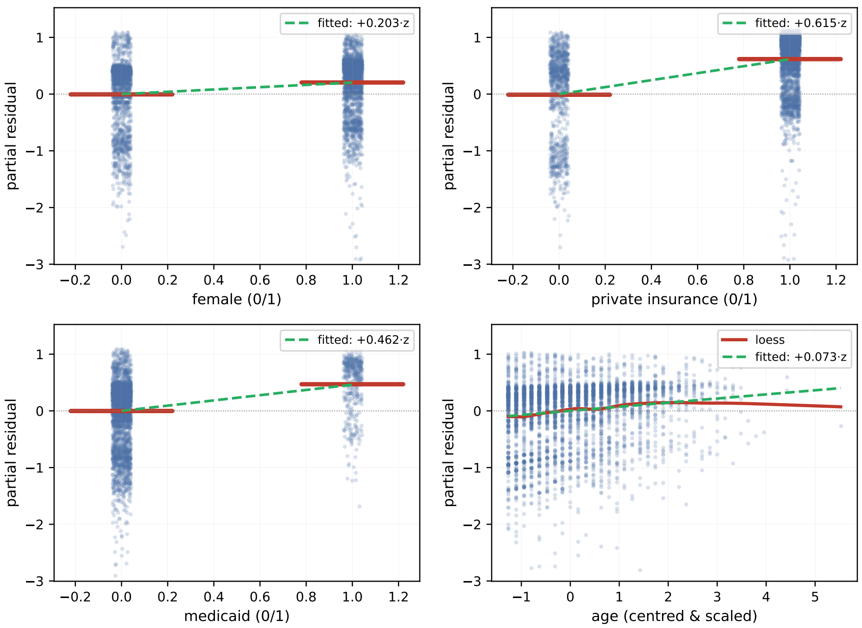


FIGURE 3. Partial residual plots (4.1) for the varying-dispersion NB fit to `NMES1988`. Red: loess smoother. Green dashed: fitted linear component $\hat{\eta}_j z_j$. Closeness of smoother to reference line validates the log-linear dispersion specification; departures indicate potential nonlinearity or stratum-level misfit.

6.3. **Step 3: Choose among dispersion specifications.** Table 2 compares three models. AIC and BIC both prefer $M_x$—clinical covariates shared with the mean equation—over $M_{\text{dem}}$, the demographic/insurance specification ($\Delta\text{AIC} = 23.3$ and $\Delta\text{BIC} \approx 17.0$ in favour of $M_x$, using the rounded table values). The LRT rejects constant dispersion against both alternatives decisively ($p < 10^{-18}$), confirming that a varying-dispersion specification improves on the null model. The result is consistent with the interpretation that dispersion in physician-visit counts is more closely tied to health status than to demographics alone. Combining this with Step 2, the demographic specification is not grossly misspecified, but its age component shows enough curvature to warrant a flexible sensitivity analysis. In a final applied analysis, the preferred $M_x$ specification should be subjected to the same partial-residual checks before reporting coefficient-level conclusions.

Table 2. Model comparison for `NMES1988`. $k$: number of parameters. BIC uses $n =$ 4,406 and is computed from the rounded log-likelihoods shown here. LR statistic tests $H_0 : \eta = 0$ for each varying-dispersion specification against the constant-dispersion null.

| Model | Dispersion cov. | $k$ | Log-lik. | AIC | BIC | LR (df) | $p$ |
|---|---|---|---|---|---|---|---|
| $M_0$ | — (constant) | 7 | −12281.9 | 24577.9 | 24622.5 | — | — |
| $M_x$ | clinical ($z = x$) | 12 | −12225.5 | 24475.0 | 24551.7 | 112.9 (5) | $< 10^{-22}$ |
| $M_{\text{dem}}$ | female, private, medicaid, age_c | 11 | −12238.2 | 24498.3 | 24568.7 | 87.5 (4) | $< 10^{-18}$ |

## 7. Discussion

Distributional NB regression is attractive because it separates two scientific questions: which covariates change the expected count, and which covariates change the regularity of that count around its expectation. The cost of this flexibility is a larger misspecification surface. A model can have an adequate mean but a poor dispersion formula, or it can improve the likelihood through dispersion covariates that are not stable under reasonable penalties. The workflow proposed here is designed to make those possibilities visible.

The main practical lesson is that no single diagnostic is sufficient. RQR QQ plots are the right starting point because they check the fitted conditional distribution on a common normal scale and are highly sensitive to mean misspecification. They are not, however, a reliable stand-alone diagnostic for moderate dispersion heterogeneity. The score-based partial residual plots fill that gap by targeting the dispersion linear predictor directly. Finally, AIC, BIC, and LRTs answer a different question: whether a candidate dispersion formula is worth retaining once competing specifications and penalty strength are considered. In the simulations here, BIC is the safest primary ranking criterion once the sample size is moderate, LRTs are useful confirmatory tests against constant dispersion, and AIC is best treated as exploratory.

Several limitations should be kept explicit. First, the exact normality result for RQRs is a known-parameter result; fitted residuals inherit estimation uncertainty, so formal normality tests should be interpreted cautiously or calibrated by parametric bootstrap. Second, the partial residual plot is a qualitative score diagnostic rather than a formal hypothesis test. Its role is to suggest where a log-linear dispersion formula is too rigid, after which the analyst should refit the suggested model and compare it with the original specification. Third, the model-comparison simulation is intentionally simple: it studies one active dispersion covariate, independent covariates, and a correctly specified NB family. Correlated mean and dispersion covariates, stronger tail behavior, zero inflation, and clustered observations may change the operating characteristics of all three tools.

The `NMES1988` application illustrates both the value and the limits of the workflow. The RQR plots provide no strong evidence against an NB family, but the partial residuals reveal that age may not enter the dispersion equation linearly. Model comparison then favours a clinical dispersion specification over a demographic/insurance specification. Taken together, these diagnostics lead to a more defensible analysis than a coefficient table alone: they show what the model fits globally, where it may need local refinement, and which dispersion specification is best supported by penalized likelihood.

The workflow is deliberately modular. Analysts who need only a global check can stop after the RQR QQ plot; analysts developing a dispersion formula can use the partial residual plots iteratively; and analysts choosing among several plausible formulas can use the model-comparison step to guard against unnecessary complexity. This modularity is important because the dispersion equation is often exploratory in applied count-data analysis.

Natural directions for future work include formal smooth tests for the dispersion equation, bootstrap QQ envelopes tailored to fitted distributional NB models, and extensions to zero-inflated NB, finite-mixture, longitudinal, and spatial count models.

## APPENDIX A. DERIVATIONS

Proof of Proposition 3.1. For any $y \in \{0, 1, 2, \ldots\}$, $U_i \mid (Y_i = y)$ is uniform on $(F_i(y-1), F_i(y))$ by construction (3.1). Since $F_i$ is the true CDF, $\Pr(U_i \leq u) = u$ for all $u \in (0,1)$, so $U_i \sim \text{Uniform}(0,1)$ marginally, giving $r_i^{\mathrm{Q}} = \Phi^{-1}(U_i) \sim \mathcal{N}(0,1)$. Independence follows from the conditional independence of $(Y_i, U_i)$ given the covariates. $\square$

Dispersion working residual. Differentiating $\ell_i$ with respect to $\alpha_i$ gives

$$\frac{\partial \ell_i}{\partial \alpha_i} = \psi(Y_i + \alpha_i) - \psi(\alpha_i) + \log \tfrac{\alpha_i}{\alpha_i + \mu_i} + 1 - \tfrac{Y_i + \alpha_i}{\alpha_i + \mu_i}.$$

The bracketed dispersion working residual used in Section 4 is this quantity, evaluated at the fitted model:

$$r_i^{\phi} = \psi(Y_i + \hat{\alpha}_i) - \psi(\hat{\alpha}_i) + \log \tfrac{\hat{\alpha}_i}{\hat{\alpha}_i + \hat{\mu}_i} + 1 - \tfrac{Y_i + \hat{\alpha}_i}{\hat{\alpha}_i + \hat{\mu}_i}.$$

Because $\partial \alpha_i / \partial(\log \alpha_i) = \alpha_i$, the score contribution for the dispersion linear predictor is $\alpha_i r_i^{\phi}$, and the fitted score conditions enforce $\sum_i \hat{\alpha}_i r_i^{\phi} = 0$ and $\sum_i z_i \hat{\alpha}_i r_i^{\phi} = \mathbf{0}$.

**Funding:** This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU263144].

**Data and availability:** `NMES1988` is publicly available via `data(NMES1988)` in R's `AER` package.

**Conflicts of Interest:** The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

[1] A.C. Cameron, P.K. Trivedi, Regression Analysis of Count Data, Cambridge University Press, 2013. https://doi.org/10.1017/CBO9781139013567.

[2] J.M. Hilbe, Negative Binomial Regression, Cambridge University Press, 2011. https://doi.org/10.1017/CBO9780511973420.

[3] G.K. Smyth, Generalized Linear Models with Varying Dispersion, J. R. Stat. Soc. Ser. B Stat. Methodol. 51 (1989), 47–60. https://doi.org/10.1111/j.2517-6161.1989.tb01747.x.

[4] R.A. Rigby, D.M. Stasinopoulos, Generalized Additive Models for Location, Scale and Shape, J. R. Stat. Soc. Ser. C Appl. Stat. 54 (2005), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x.

[5] M.D. Stasinopoulos, R.A. Rigby, G.Z. Heller, V. Voudouris, F. De Bastiani, Flexible Regression and Smoothing: Using GAMLSS in R, Chapman and Hall/CRC, 2017. https://doi.org/10.1201/b21973.

[6] P. Deb, P.K. Trivedi, Demand for Medical Care by the Elderly: A Finite Mixture Approach, J. Appl. Econom. 12 (1997), 313–336. https://doi.org/10.1002/(sici)1099-1255(199705)12:3<313::aid-jae440>3.0.co;2-g.

[7] J.F. Lawless, Negative Binomial and Mixed Poisson Regression, Can. J. Stat. 15 (1987), 209–225. https://doi.org/10.2307/3314912.

[8] P.K. Dunn, G.K. Smyth, Randomized Quantile Residuals, J. Comput. Graph. Stat. 5 (1996), 236–244. https://doi.org/10.1080/10618600.1996.10474708.

[9] S.N. Wood, Generalized Additive Models: An Introduction with R, CRC Press, 2017. https://doi.org/10.1201/9781315370279.