Results for predictors of spelling mistakes using the Bayesian Poisson model with horseshoe priors

Adrian Barnett

26 March, 2025

We modelled the counts of spelling errors per abstract. The sample size was 157,731 abstracts.

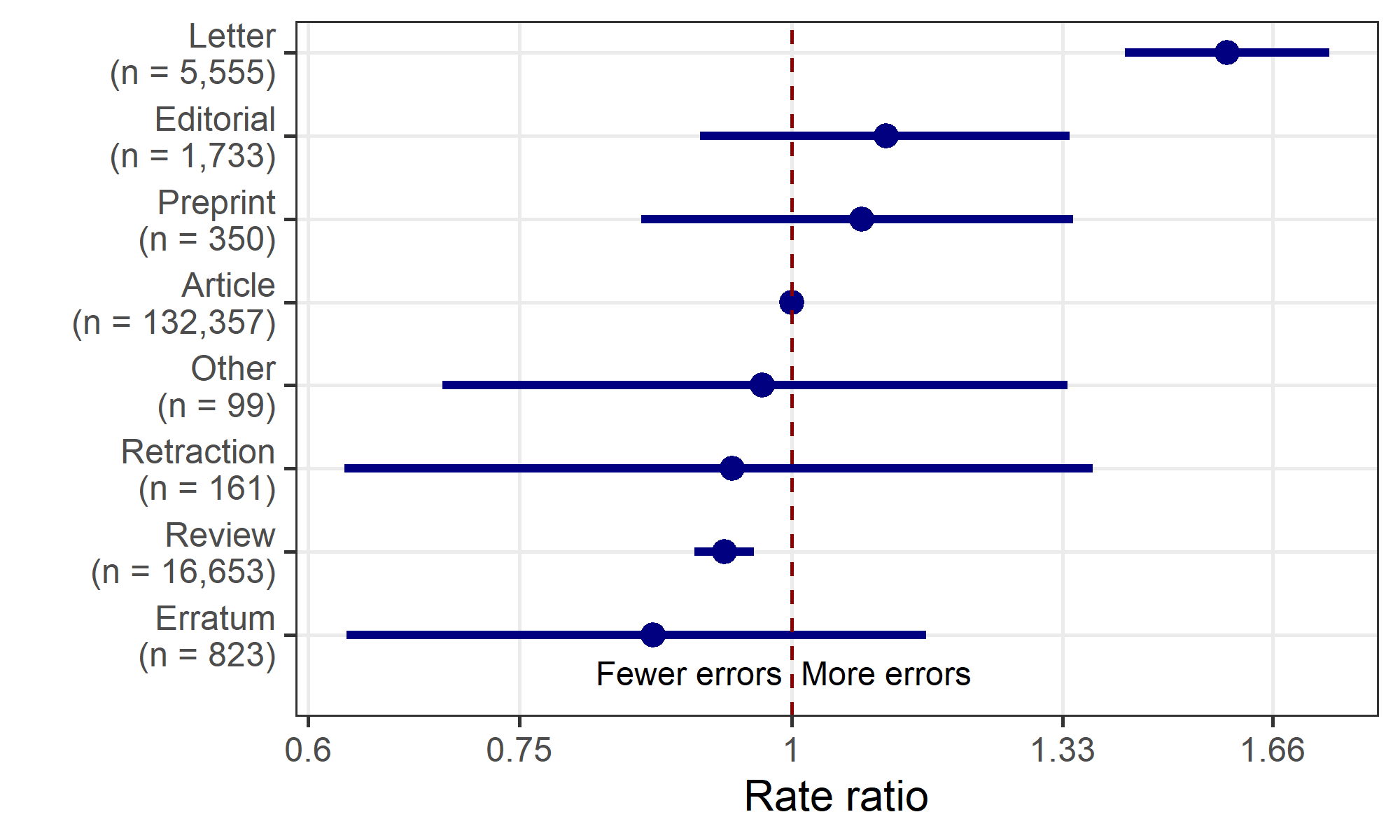
## Article types

#### Table

| **var** | **mean** | **lower** | **upper** | **posterior probability** |
| --- | --- | --- | --- | --- |
| Article (reference) | 1.00 |  |  |  |
| Editorial | 1.10 | 0.91 | 1.34 | 0.3066 |
| Erratum | 0.86 | 0.62 | 1.15 | 0.3280 |
| Letter | 1.58 | 1.42 | 1.76 | 0.0001 |
| Other | 0.97 | 0.69 | 1.34 | 0.8658 |
| Preprint | 1.08 | 0.85 | 1.35 | 0.5244 |
| Retraction | 0.94 | 0.62 | 1.37 | 0.7628 |
| Review | 0.93 | 0.90 | 0.96 | 0.0001 |

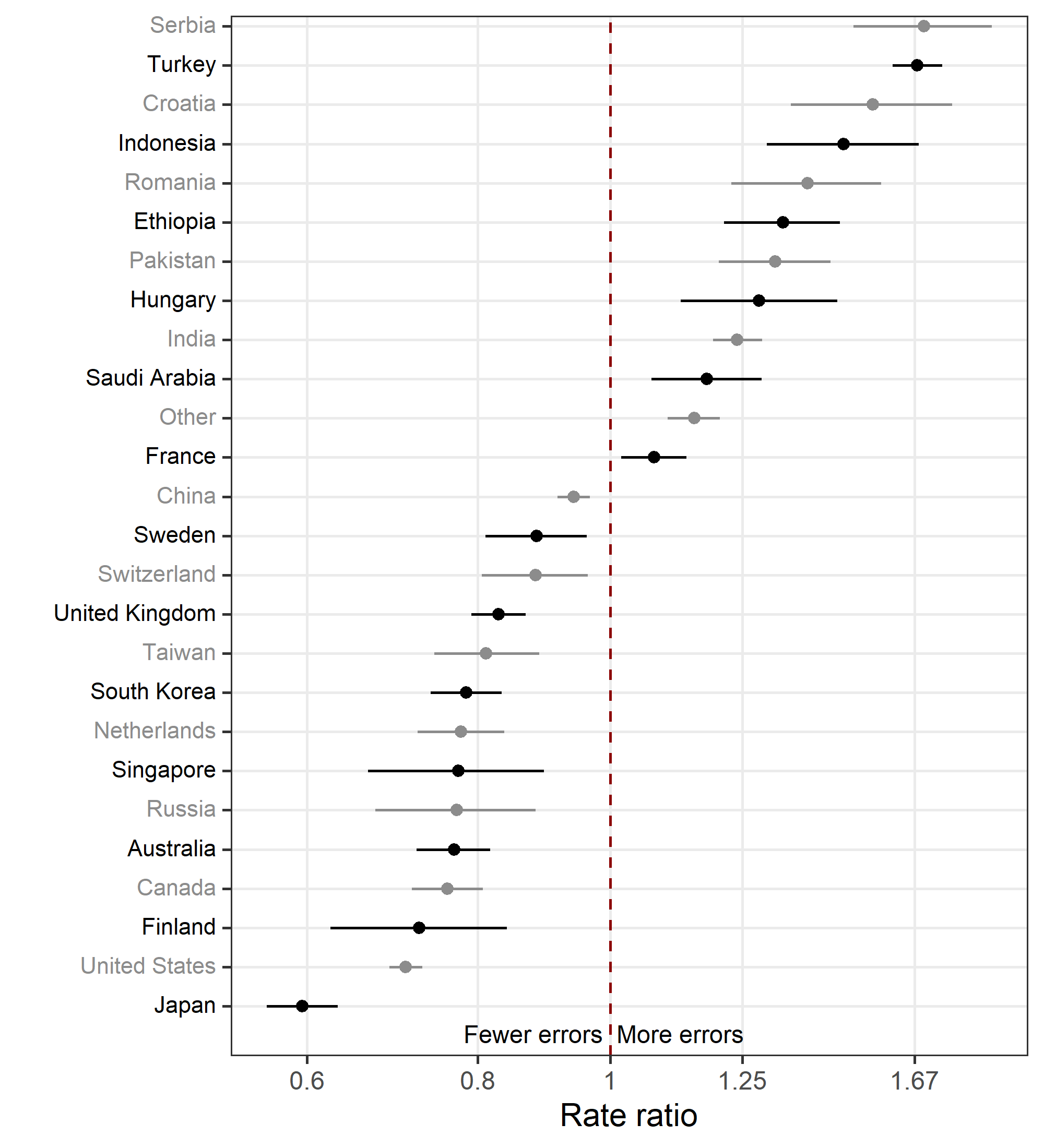
The table shows the mean, 95% credible interval, and posterior probability that the estimate is zero.

#### Plot



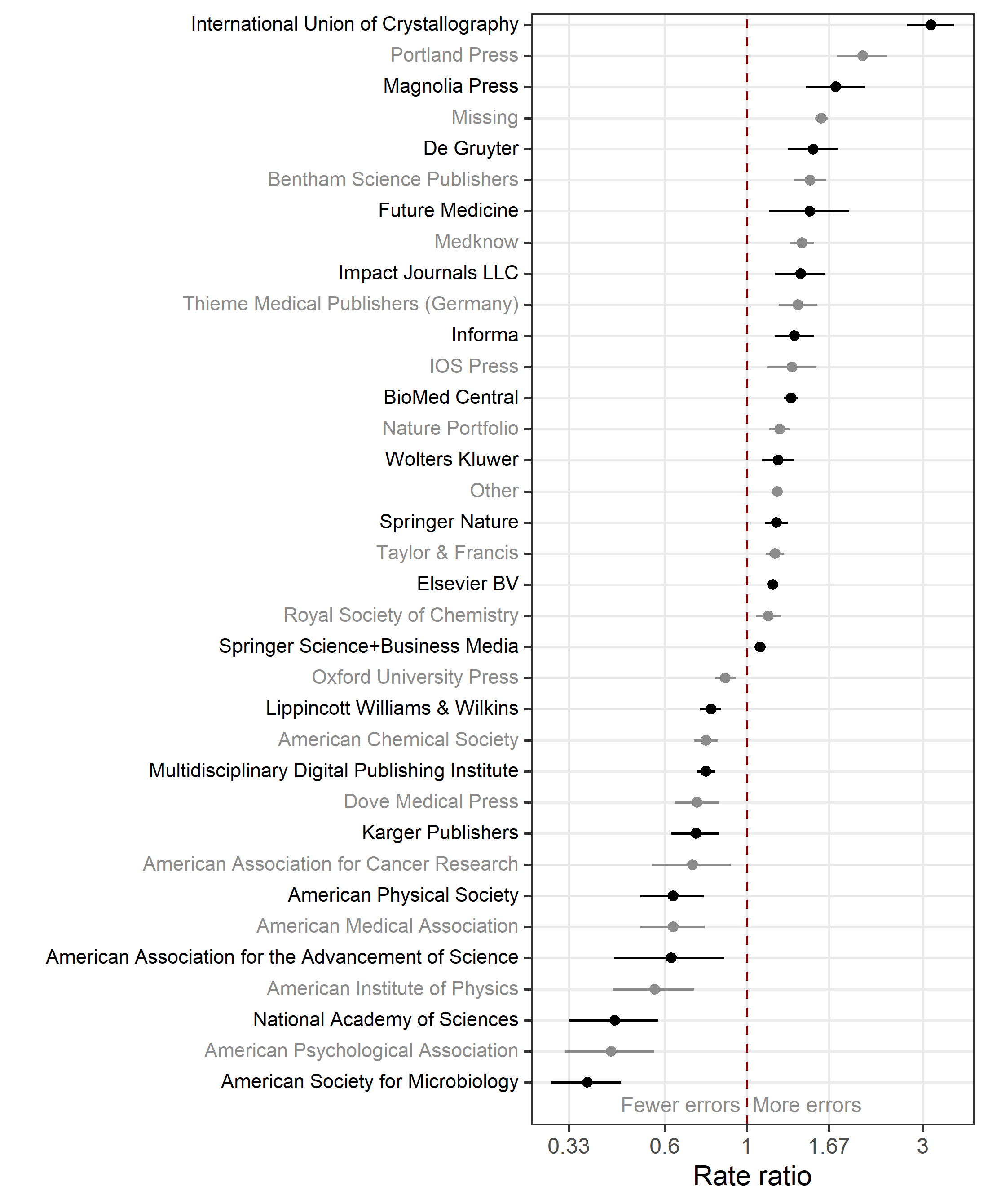
A type of “Article” was the reference category.

## Countries



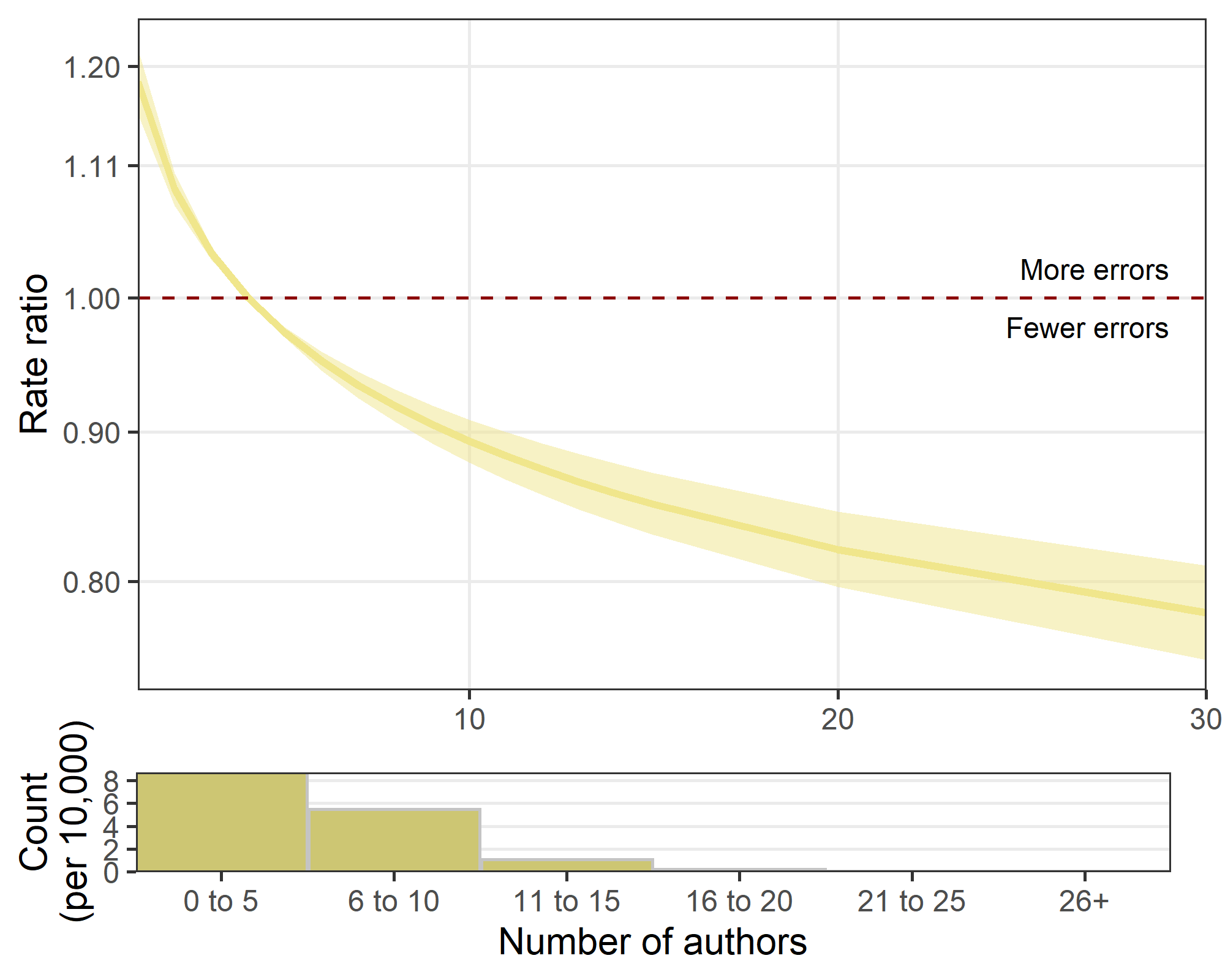
The plot only shows those countries where there is an over 0.99 probability that the mean is not zero.

## Publisher



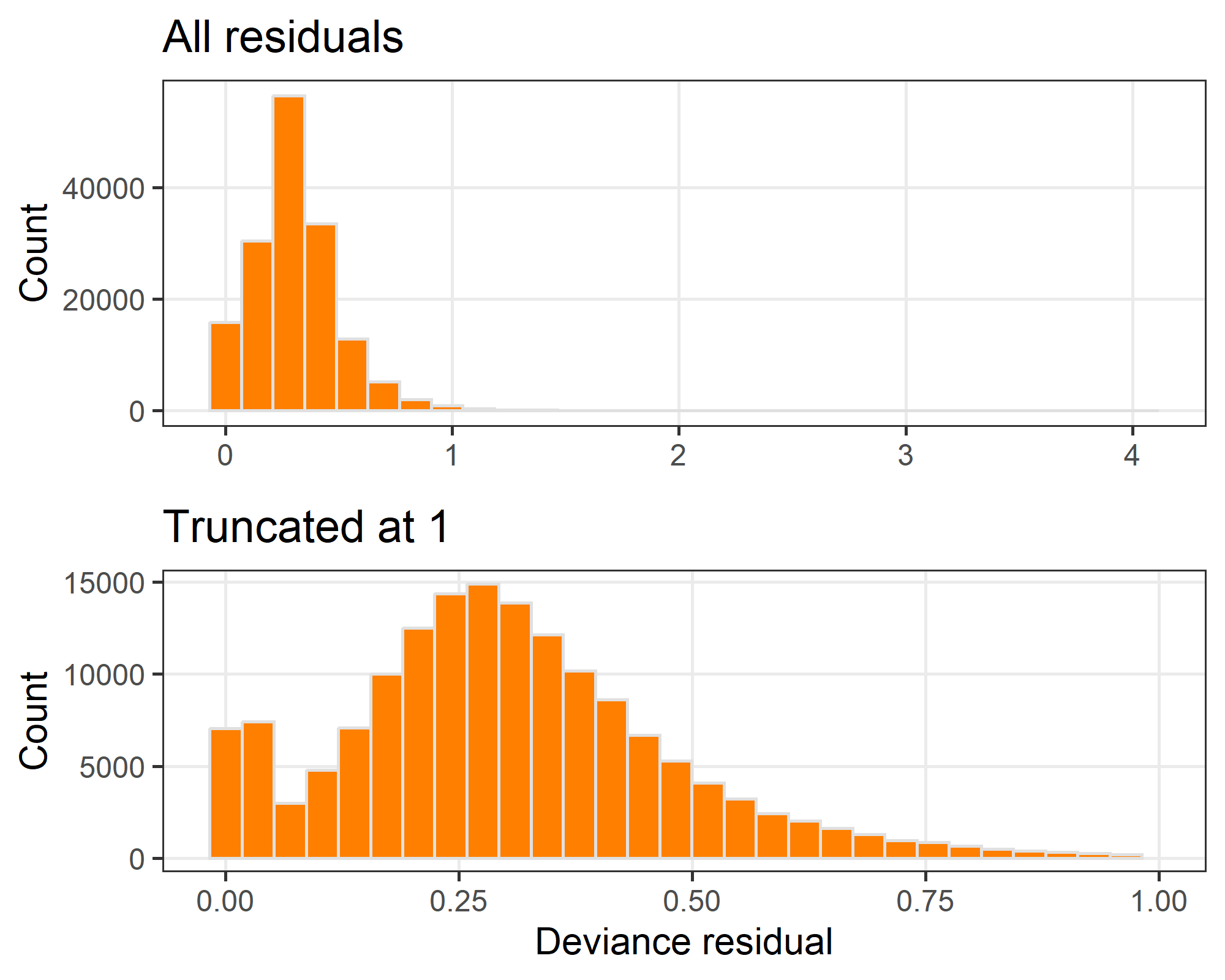
The plot only shows those publishers where there is an over 0.99 probability that the mean is not zero.

## Author numbers



## Residual checks

#### Histogram of deviance residuals



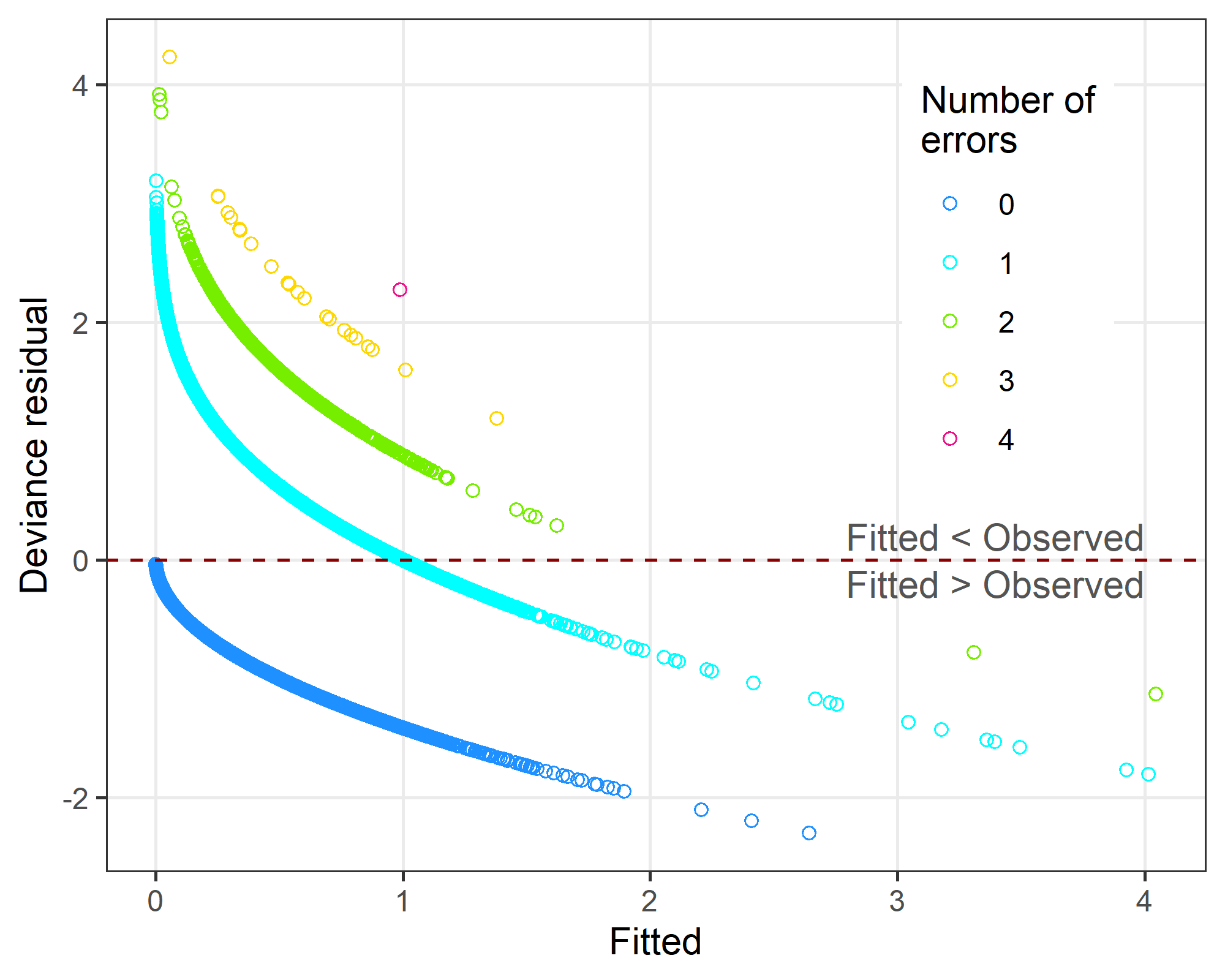
The residual distribution has a positive skew and there are a few large positive residuals of 3 and over. Reducing the scale to exclude these outliers shows a bimodal distribution with a mode at zero and second mode under 0.5.

#### Invesigate large residuals

| **type** | **country** | **word\_count** | **errors** | **fitted** | **deviance** |
| --- | --- | --- | --- | --- | --- |
| Article | BR | 37 | 3 | 0.05703548 | 4.229685 |
| Article | Missing | 6 | 2 | 0.01593881 | 3.919242 |
| Editorial | CA | 12 | 2 | 0.01749565 | 3.871794 |
| Letter | US | 16 | 2 | 0.02149521 | 3.765007 |

These large deviance residuals (over 3.5) are often papers with multiple errors and a small word count. It is relatively unusual for papers with such small word counts to have multiple errors, and the model does not predict this. Other problems are those with six errors.

#### Plot of deviance residuals against fitted values



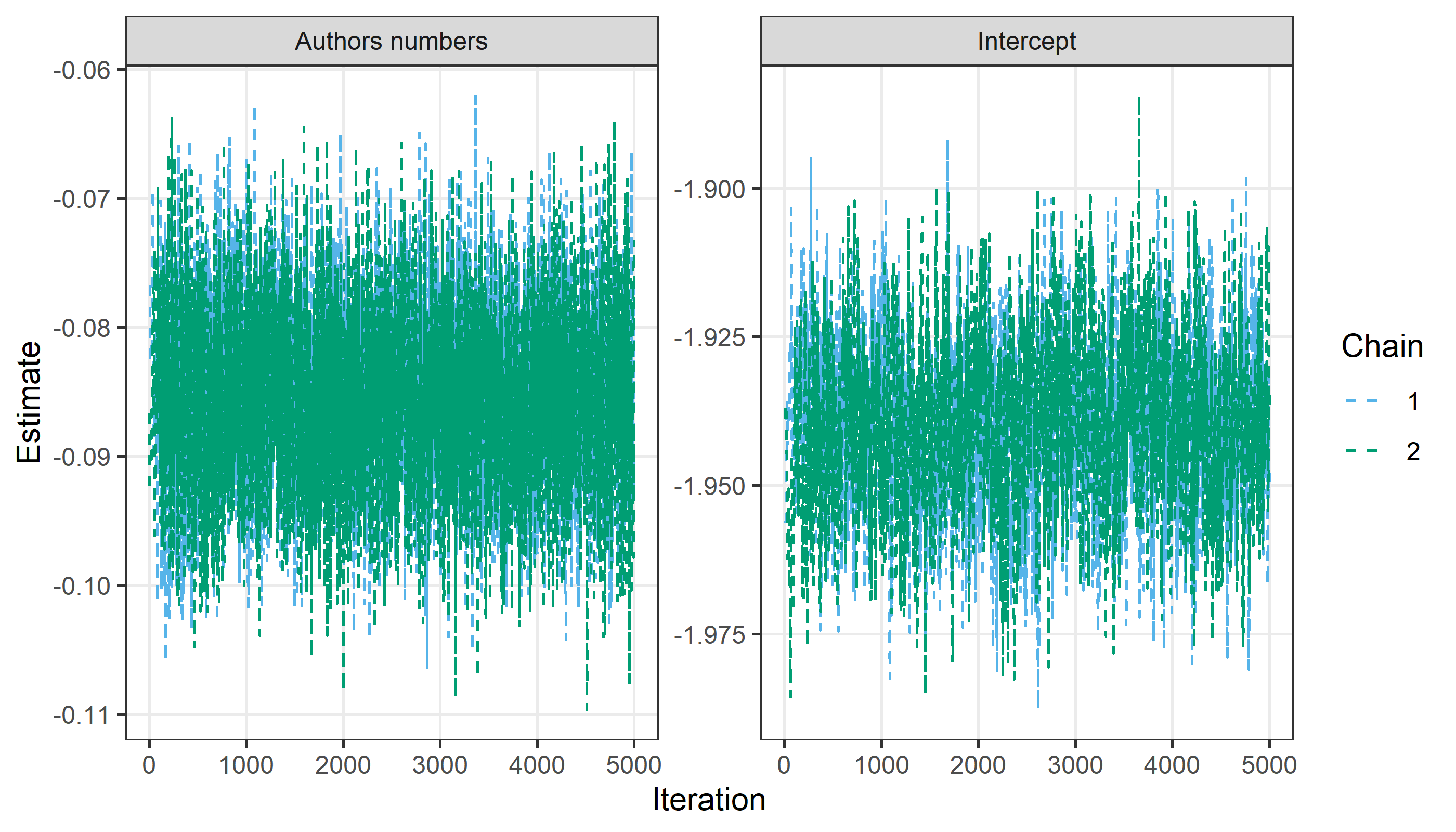
The plot shows a general pattern of positive residuals for small fitted values and negative deviance residuals for large fitted values.

The large residuals for small fitted values are because some abstracts with multiple spelling errors were not captured by the model. This is because there are likely other predictors of spelling errors that were not included in the model, and also that some errors are stochastic and not predictable.

## Check model convergence

We used 5,000 samples thinned by 5 with two chains.

#### Chains for the intercept and author numbers estimates



The chains appear to have converged with good mixing.

#### Densities for the intercept and word count estimates

