

Towards Mitigating Bias in Online Reviews: An Application to Amazon.com

Charlie Walker, Scott Buttinger

Stanford University



Abstract

Our project uses a Bayesian model and an Amazon graph to:

- Estimate bias in Amazon reviews
- Estimate true product quality from observed ratings
- Expand upon previous work by incorporating variance scaling

Introduction

Why Estimate Bias in Online Reviews?

- Reviews are a **signal of product quality**
- Commerce is heavily dependent on **quality of information**
- Current reviews are a **noisy signal** of quality: can we make them better?

Dataset

Amazon.com

- Reviewer ID
- Reviewed product ASIN (Amazon Standard Identification Number)
- Number of “helpful” votes
- Product category
- Overall (i.e. average) product rating

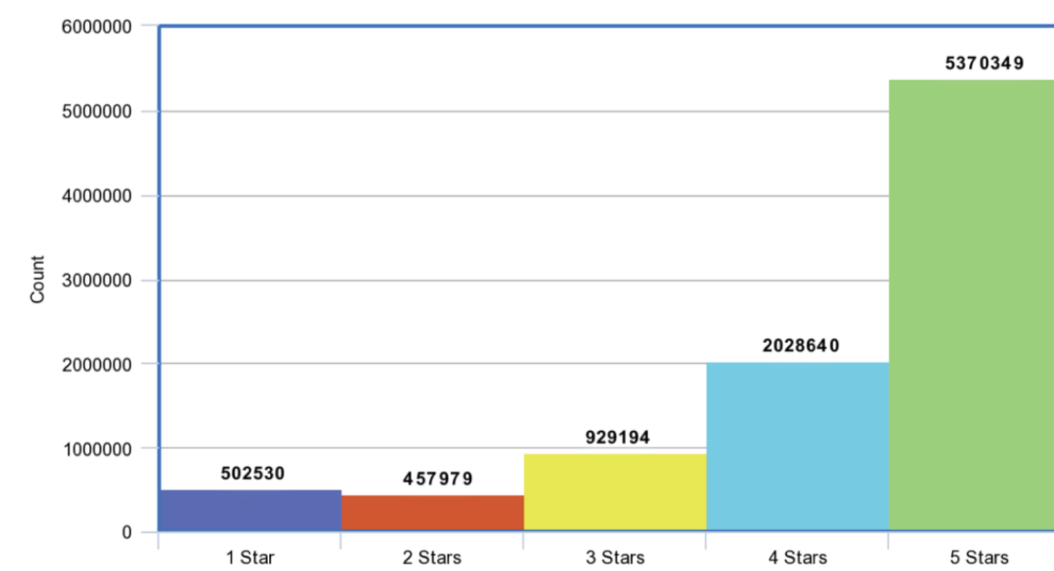


Figure: Histogram of Star Ratings Given

Amazon product ratings have a “J-shaped” distribution, as most reviews give a 5-star rating. This obfuscates the true value of a product.

Model

We observe review r_{pu} for product p and user u . Our goal is to estimate a user bias b_u and product “goodness” g_p from our observed data. Our model can be summarized with:

$$\begin{aligned} g_p &\sim \mathcal{N}(3, 1) & (1) \\ b_u &\sim \mathcal{N}(0, c) & (2) \\ r_{pu} &\sim \mathcal{N}(g_p + b_u, 1/A_{pu}) & (3) \\ A_{pu} &= \begin{cases} 1 & u \text{ reviewed } p \\ 0 & \text{otherwise} \end{cases} & (4) \end{aligned}$$

The above, our baseline model, is extended by replacing A_{pu} with $d \sim \mathcal{N}(a_d, b_d)$ for Model 1, and $H_{pu} \in [1, 2]$, the helpfulness of rating r_{pu} .

We use STAN, a high-performance engine for Bayesian inference based on the No-U-Turn Sampler (NUTS) to estimate reviewer bias and true product quality.

Latent Variable Distributions

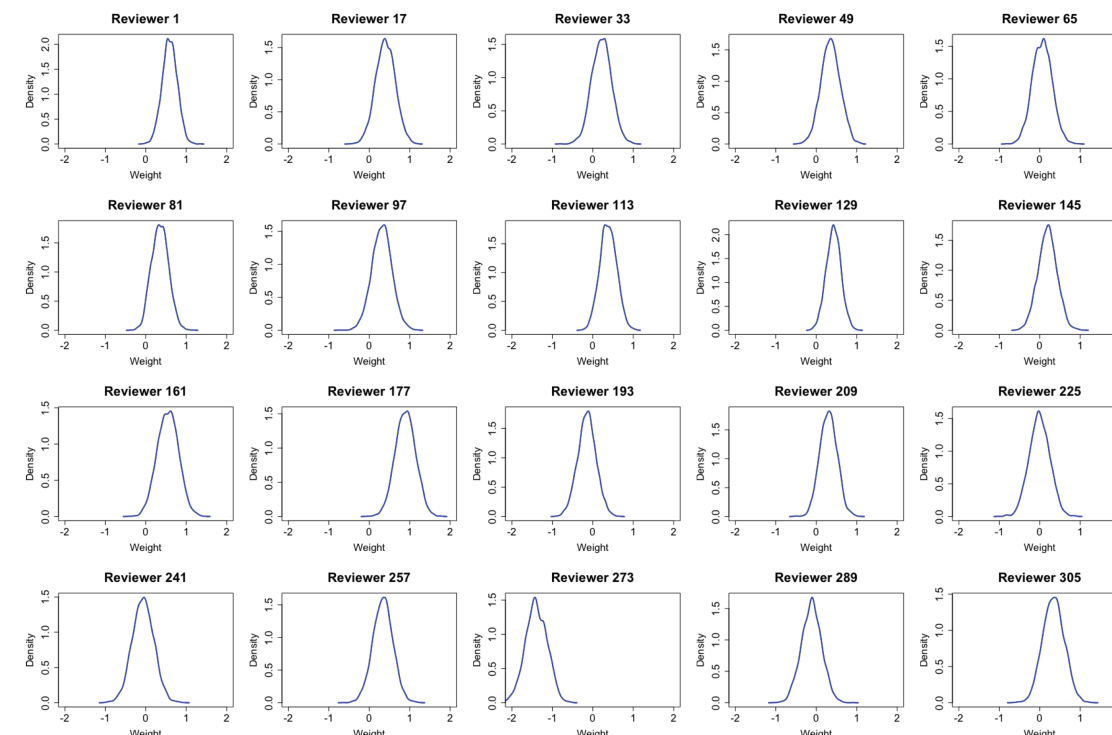


Figure: Posterior Distribution of Reviewer Bias

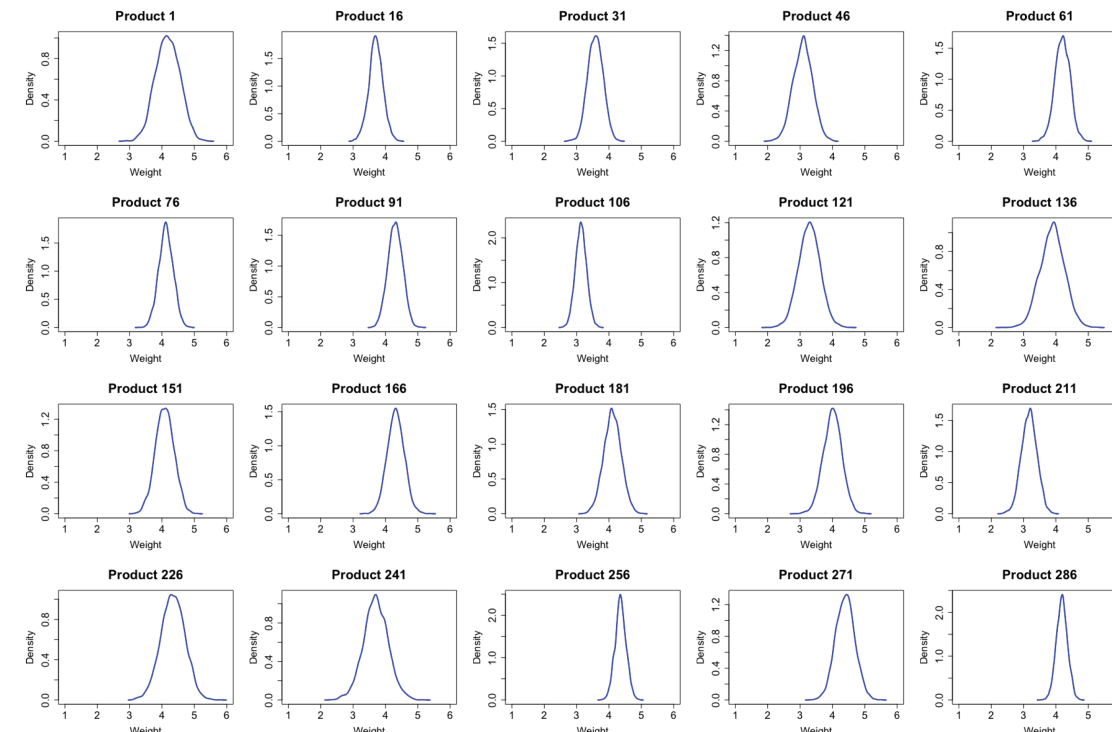


Figure: Posterior Distribution of Adjusted Product Rating

A sample of reviewer bias and true product quality estimations.

Results

Rank Changes

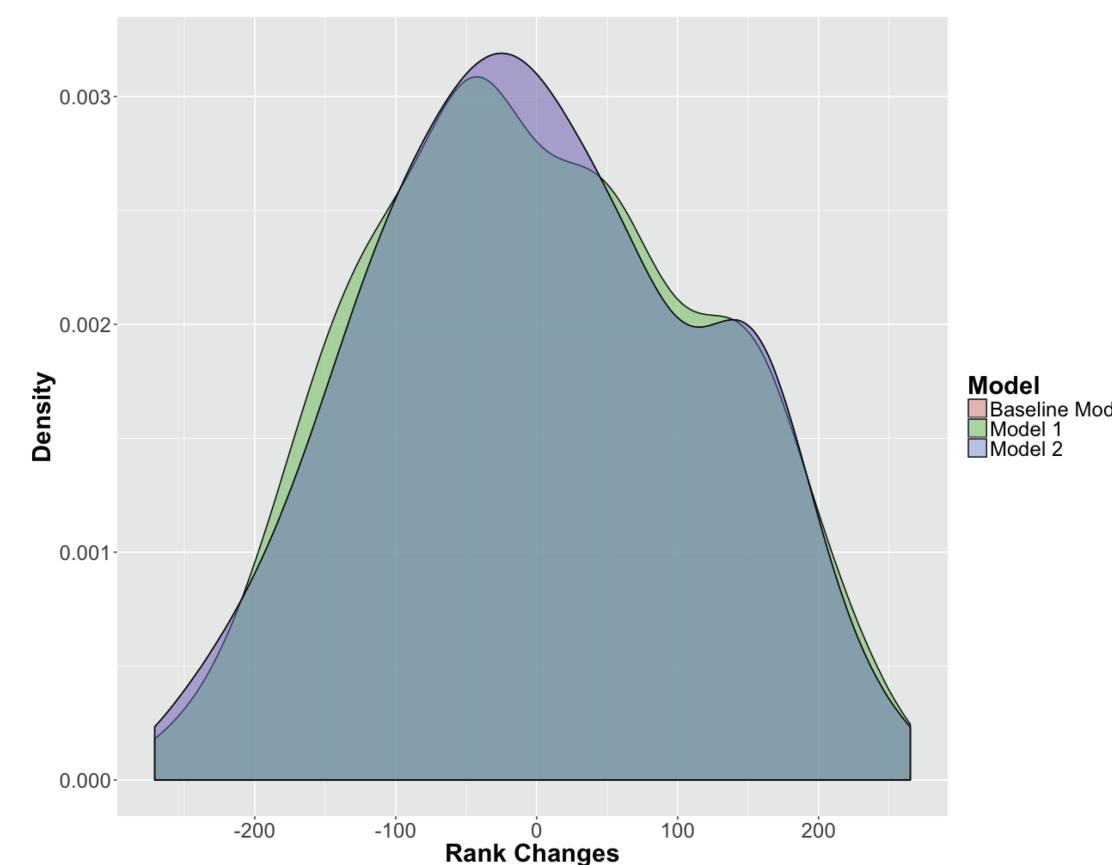


Figure: Change in Product Rank

Recalculating product ratings based on estimated reviewer bias and true product quality dramatically alters the ranking of products on Amazon.

Results Continued

Rating Changes

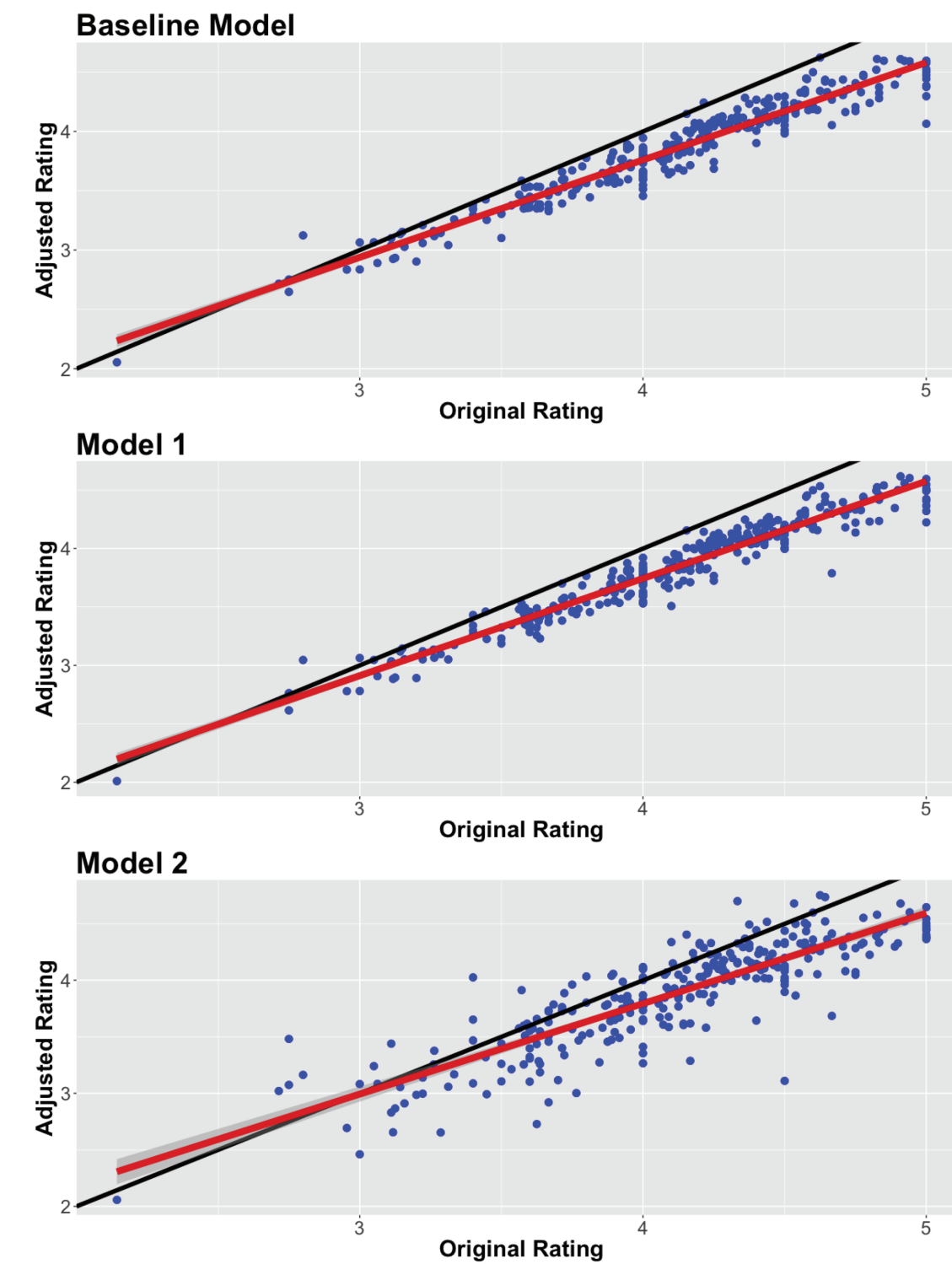


Figure: Change in Product Ratings

Adjusted product rating is generally less than the original rating from the data.

Predictive Accuracy

We evaluated our models using RMSE, calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{|R_{test}|} \sum_{r_{ij} \in R_{test}} (r_{pu} - g_p - b_u)^2} \quad (5)$$

Where R_{test} is the actual product rating given by the reviewer in our test dataset.

Table: RMSE by Model

	Baseline	Model 1	Model 2
RMSE	0.9322	0.8353	0.8403

Conclusion and Next Steps

Our results demonstrate that there are groups of reviewers that systematically rate products higher or lower than their peers, and that this latent bias can be detected with a straightforward Bayesian model.

Next Steps

- More sophisticated variance scales, such as product expertise or seller trustworthiness
- Category specific models
- Explore use cases for calibrated reviews