The Impact of International Shipping on Product Ratings: A Bayesian Analysis

— Project Report —
Advanced Bayesian Data Analysis

Dia Tasneem Naser

Sourav Poddar

March 14, 2025

TU Dortmund University

Contents

1 Introduction

In the globally expanded landscape of e-commerce and online retail markets, product ratings are a crucial measure of customer satisfaction and significantly influence consumer purchasing behavior. While multiple factors, from pricing and product quality to consumer demographics (gender, age), are well documented, one area that warrants deeper exploration is the possible impact of overseas shipping on customer ratings. "Is international shipping a negligible element, or does it cause a decline in customer satisfaction?" Therefore, for businesses aiming to improve customer happiness and maximize sales techniques, it is pivotal to comprehend this factor on product ratings.

This study's motivation stems from online retail's growing reliance on data-driven decision-making. When businesses expand globally, customer perceptions might be influenced by increased shipping prices, longer delivery times, and customs delays. While some buyers accept these inconveniences, others may express dissatisfaction through lower ratings, negatively affecting a business. By analyzing this association, we aim to provide insights that will benefit merchants in improving their shipping regulations and market experiences.

To explore this relationship, we utilized Bayesian methods for their robust statistical approach towards handling uncertainty and integrating prior knowledge into the analysis. Given that the ratings are inherently ordered, implementing Bayesian ordinal regression models quantifies the uncertainty associated with the effect of international shipping and provides more reliable probabilistic estimates than traditional frequentist methods. Using hierarchical models, we explored how buyer demographics and product categories respond to shipping-related influences. Finally, we validated the performance of the models with Leave-One-Out (LOO) cross-validation.

In Section 2, insights about data filtration and data sources are provided. Section 3 describes the concept of Bayesian ordinal regression, appropriate prior selection, and the description of fitted models. In Section 4, diagnoses of model convergence were discussed. Sections 5 and 6 represent model comparison and prior sensitivity analysis. Limitations and potential improvements were discussed in Section 7. Section 8 contains a comprehensive summary of the whole project.

2 Data

2.1 Dataset Description

The dataset is extracted from Kaggle's Global Retail Sales Data: Orders, Reviews & Trends, publicly accessible at `https://www.kaggle.com/datasets/adarsh0806/influencer-merchandise-sales`.It consists of $n = 7,394$ observations and 15 features, encompassing meticulous records of sales pricing, shipping attributes, buyers' demographics, e-commerce transactions, and customers' feedback through reviews and ratings. Even though this dataset has been previously utilized for analyzing retail sales trends and customer judgment, it hasn't been used to study the impact of international shipping on customers' product ratings. This retail dataset has no missing values, preserving high data integrity.

The features *Product.Category* [Clothing, Ornaments, Other],*Product.ID* (unique identifier), *Buyer.Gender* [Female, Male], *International.Shipping* [Yes, No] (binary), *Order.ID* (unique identifier), and *Order.Location* are categorical.

Other variables, such as *Buyer.Age*, *Sales.Price*, *Shipping.Charges*, *Sales.Per.Unit*, and *Total.Sales* are numeric continuous, *Quantity* (discrete), *Order.Date* (temporal), and *Review* (textual and unstructured). Target variable *Rating* is an ordinal variable, scaling from 1 to 5, reflecting customer satisfaction.

2.2 Data Pre-Processing

In data pre-processing, we began by selecting features based on their exploratory analysis to align the dataset with research objectives, ensuring relevance to product Rating. Choosing features like *International.Shipping*, *Shipping.Charges*, *Product.ID*, *Product.Category*, *Buyer.Age*, *Buyer.Gender*, *Sales.Per.Unit*, *Sales.Price*, *Quantity*, and *Total.Price* was admissible as they may contribute to variations in *Rating*. Our main predictor, *International.Shipping*, differentiates between international and domestic orders. *Buyer.Gender* and *Buyer.Age* offer insights into the demographic importance of rating behavior, whereas *Product.Category* elucidates differences across merchandise types. Economic variables that may influence consumer views include *Sales.Price*, *Shipping.Charges*, *Quantity*, and *Total.Sales*.

The target variable *Rating* is ordinal; we encoded it as an ordered factor $1 < 2 < 3 < 4 < 5$ to preserve its hierarchical structure within the Bayesian Ordinal Regression framework. *International.Shipping*, *Buyer.Gender*, *Product.Category*, and *Product.ID* were factorized to ensure proper handling of the categorical variables. In addition, to maintain model identifiability, we dummy encoded *International.Shipping* ("yes" = international), *Buyer.Gender* (" Female "), and *Product.Category* ("1" = clothing). This dummy encoding is called the reference category. Hence, these scrupulous pre-processing steps ensure that the selected relevant features are appropriately structured for a robust analysis of the impact of international shipping on product ratings.

3 Models

This section demonstrates the Bayesian Ordinal Regression Model as the framework for analyzing the association between international shipping and product ratings. Each model is explained in detail, encompassing its formulations, prior selection, and interpretations with posterior distributions. By gradually enhancing model complexity, we integrated buyer demographics, product attributes, and hierarchical structures for deeper knowledge about the factors influencing customer satisfaction. This section serves as a comprehensive guide for the probabilistic modeling of product ratings, offering valuable insights for e-commerce and global trade applications.

3.1 Bayesian Ordinal Regression

Bayesian ordinal regression is a statistical modeling approach designed for ordinal response variables suitable for analyzing customer ratings, Likert-scale responses, and survey results. Given that the target variable, *Rating*, is an ordinal variable, scaling from 1 to 5, it requires an approach that rationalizes the inherent ranking structure of the data. The cumulative logit model (McCullagh 1980, p. 109) appropriately models the probability of an observation falling within or below a given category, thus preserving the ordinal nature of the ratings while enabling probabilistic inference regarding the effects of international shipping and other covariates.

Ordinal regression assumes that the observed ordinal response y_i arises from an unobserved continuous latent variable y_i^*. This latent variable represents an individual's underlying level of satisfaction, which is mapped onto discrete categories based on a set of threshold parameters, α_k.

The formulation of the latent variable is

$$y_i^* = X_i\beta + \epsilon_i, \quad \epsilon_i \sim Logistic(0,1)$$

Here, the latent continuous variable is denoted by y_i^*, determining the observed ordinal category. X_i is the vector of predictor variables such as international shipping, product category, buyer age, buyer gender, and sales. β are the regression coefficients that show the effects of predictors on the response variable. Cumulative logit transformation of probabilities is ensured by the standard logistic distribution ϵ_i.

A set of threshold parameters α_k $(k = 1, ..., K-1)$ defines the transition points between adjacent categories. Since the response variable Rating is scaled from 1 to 5 and may lack quantifiable distance between the levels, formulating this threshold parameter is important to determine the observed outcome y_i. These thresholds are estimated from the data, ensuring the ordinal structure. The likelihood (Ravula 2023, p. 149) of a response falling into a specific category k or below is formulated by:

$$P(y_i \leq k) = \frac{1}{1 + \exp\left(-(\alpha_k - X_i\beta)\right)}$$

In the cumulative logit model, the log-odds transformation is applied to the cumulative probabilities of an ordinal response y_i, ensuring the cumulative probabilities $P(y_i \leq k)$ increase monotonically across categories while preserving the ordinal structure. Hence, the log-odds of an observation falling in category k or below, relative to being in a higher category, is expressed as:

$$\text{logit}(P(y_i \leq k)) = \log\left(\frac{P(y_i \leq k)}{1 - P(y_i \leq k)}\right) = \alpha_k - X_i\beta$$

3.2 Prior Selection

In the Bayesian approach, the choice of prior distribution is essential for integrating domain knowledge or assumptions about the possible values of model parameters, as it impacts posterior distributions, influencing uncertainty quantification and model interpretability. Owing to that, we opted for normal priors due to their mathematical convenience and ability to impose reasonable constraints on parameters. Normal distributions are symmetric around the mean (μ) when centered at zero, indicating that the predictors don't show any preference for either positive or negative outcomes. The probability density function of a normal distribution is expressed as

$$p(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta - \mu)^2}{2\sigma^2}\right)$$

In our Bayesian ordinal regression models, we used weakly informative normal priors for both regression coefficients denoted by $\beta_j \sim \mathcal{N}(0, 5)$ and intercepts by $\alpha_k \sim \mathcal{N}(0, 5)$, allowing moderate variability while regularizing extreme values in cases of limited data (Gelman et al. 2013) Additionally, for group-level variations in hierarchical ordinal regression, we selected Student's t-distribution priors expressed as $\sigma \sim Student_t(3, 0, 2)$. *Student-t* priors allow heavy-tailed distributions, making the model robust to outliers with a meaningful shrinkage towards the group mean. Besides maintaining flexibility and stability, these priors are also responsible for preventing the overfitting of the models.

3.3 Model Description

3.3.1 Model 1

The first ordinal regression model, specified as `ordinal_model_fit1`, incorporates variables such as *International.Shipping, Buyer.Age, Sales.Price*, and *Quantity* to analyze the relationship with the response variable *Rating.* This is a simple baseline ordinal model, utilizing the cumulative logit link, investigating the effect of the predictors on the log-odds of falling into a specific rating category or below, compared to higher ratings.

$$\text{Rating} \sim \text{International.Shipping} + \text{Buyer.Age} + \text{Sales.Price} + \text{Quantity}$$

Table 1 summarizes the estimations of model 1. The estimated threshold parameters (intercepts) depict the cut-off points that distinguish adjacent rating levels. Intercept[1] estimates -1.65 with 95% CI [-1.88, -1.42], representing the log-odds ratio between rating 1 and 2. Similarly, the subsequent intercepts, Intercept[2] (-0.99, 95% CI: [-1.21, -0.76]), Intercept[3] (-0.29, 95% CI: [-0.52, -0.06]), and Intercept[4] (0.97, 95% CI: [0.75, 1.20]), define the progressive log-odds thresholds between higher rating categories. These estimated intercepts show monotonic increments across rating categories while preserving the ordered structure.

Table 1: Estimates of Model 1

Covariate	**Estimate**	**Est.Error**	**95% CI**	**Rhat**	**Bulk_ESS**	**Tail_ESS**
Intercept[1]	-1.65	0.12	[-1.88, -1.42]	1.00	4915	3406
Intercept[2]	-0.99	0.12	[-1.21, -0.76]	1.00	4870	3063
Intercept[3]	-0.29	0.12	[-0.52, -0.06]	1.00	5061	3398
Intercept[4]	0.97	0.12	[0.75, 1.20]	1.00	5230	3241
International.ShippingNo	0.03	0.05	[-0.06, 0.12]	1.00	3701	2710
Buyer.Age	0.00	0.00	[-0.00, 0.01]	1.00	5309	3503
Sales.Price	-0.00	0.00	[-0.00, 0.00]	1.00	4257	3018
Quantity	0.02	0.02	[-0.02, 0.06]	1.00	4093	2724

The regression coefficients estimate the effect of covariates in the log-odds ratio of receiving a lower rating. The estimation for *International.ShippingNo* is 0.03, suggesting a negligible effect on customers' ratings. 95% CI [-0.06, 0.12] includes 0, which strengthens the interpretation. The variable *Buyer.Age* exhibits an estimate of 0.00 encompassing 0 in 95% CI, indicating no significant relationship with rating variation. Similarly, *Sales.Price* resulting in -0.00 with 95% CI: [-0.00, 0.00] reflects minimal influence on rating assignment. Finally, *Quantity* is associated with an estimated effect of 0.02 (95% CI: [-0.02, 0.06]), though the credible interval overlapping zero suggests substantial uncertainty.

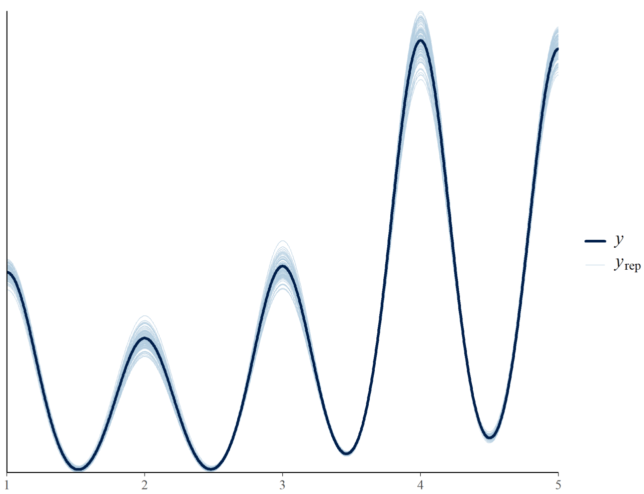


Figure 1: Posterior Predictive Check of Model 1

As shown in **Figure 1**, the posterior predictive distribution of model 1, demonstrating that the replicated data (y_{rep}) closely follows the observed data (y), indicating a good model fit with some uncertainties in predictions. Despite the low standard errors, the credible intervals for all predictors contain 0, exhibiting insignificant effects of included features. This finding emphasizes a more complex model.

3.3.2 Model 2

We extended the previous model by introducing an interaction term between *International.Shipping* and *Product.Category*,referred to as `ordinal_model_fit2` allow us to dive even deeper to inspect the effect of shipping on product ratings with varieties of products. The model is specified as:

$$\text{Rating} \sim \text{International.Shipping} * \text{Product.Category}$$

Summarizing the outcomes from **Table 2**, we observe that the intercepts might be dominant in formulating the distribution of ratings. Intercept[1] is estimated at 1.73 with 95% CI [-1.81, -1.64] and presents the log-odds threshold between rating 1 and 2. Intercept[2] estimates the log-odds of -1.07 (95% CI: [-1.15, -0.99]) between rating 2 and 3. Similarly, Intercept[3] (-0.37, 95% CI: [-0.44, -0.29]) and Intercept[4] (0.90, 95% CI: [0.82, 0.97]) define the progressive log-odds cutoffs between higher ratings, with a low standard error of 0.04.

Table 2: Estimates of Model 2

Covariate	**Estimate**	**Est.Error**	**95% CI**	**Rhat**	**Bulk_ESS**	**Tail_ESS**
Intercept[1]	-1.73	0.04	[-1.81, -1.64]	1.00	3301	3217
Intercept[2]	-1.07	0.04	[-1.15, -0.99]	1.00	3932	3561
Intercept[3]	-0.37	0.04	[-0.44, -0.29]	1.00	4073	3378
Intercept[4]	0.90	0.04	[0.82, 0.97]	1.00	3763	3499
International.ShippingNo	0.10	0.06	[-0.02, 0.23]	1.00	2801	3067
Product.Category2	-0.03	0.06	[-0.11, 0.08]	1.00	3425	3053
Product.Category3	0.18	0.07	[0.05, 0.31]	1.00	3551	2940
International.ShippingNo:Product.Category2	-0.08	0.10	[-0.28, 0.12]	1.00	2776	2600
International.ShippingNo:Product.Category3	-0.26	0.12	[-0.50, -0.02]	1.00	3148	3234

Table 2, also shows that the predictors, *International.Shipping* and *Product.Category* don't have a major influence on customer satisfaction levels. The regression coefficient for *International.ShippingNo* is 0.10, indicating a slight tendency towards a lower rating even though it includes 0 in 95% credible interval. From the estimation (-0.03) of *Product.Category2*, it can be concluded that the variable has minimal effect on rating variations. In contrast, *Product.Category3* has an estimated effect of 0.18 (95% CI: [0.05, 0.31]); this positive coefficient means that products in this category are more likely to receive lower ratings compared to other categories. The interaction term exhibits more complex insights into the effect of international shipping across various products on ratings. When product category 2 is domestically shipped it still doesn't imply any effect.

However, the interaction term *International.ShippingNo:Product.Category3* estimates -0.26 with 95% CI: [-0.50, -0.02]. This negative coefficient indicates that when products in category 3 are not shipped internationally, the log-odds of receiving a higher rating increase.

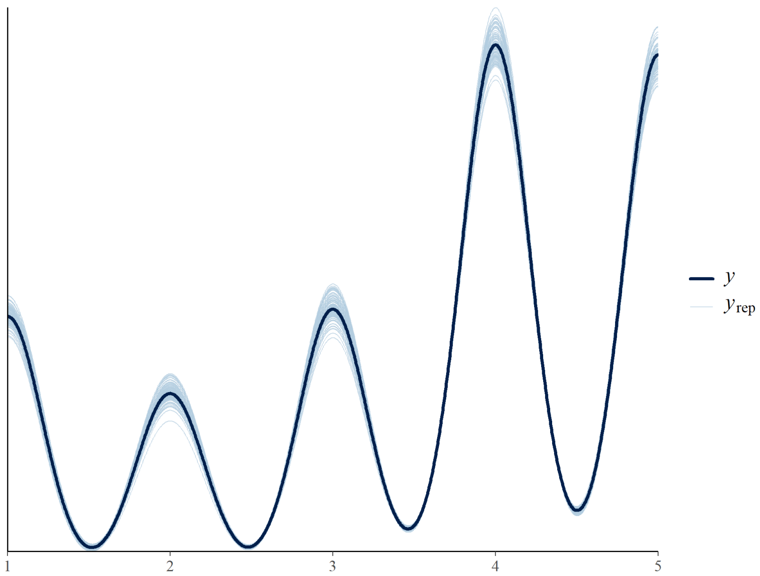


Figure 2: Posterior Predictive Check of Model 2

The posterior predictive check in **Figure 2** indicates that the replicated data (y_{rep}) closely follows the observed data (y), validating the model's fit. The predictive intervals follow empirical data distribution, suggesting the interaction between *International.Shipping* and *Product.Category* is well captured.

In model 2, the inclusion of 0 in the credible interval for several predictors, reinforcing the uncertainty in parameter estimates, establishes the need for further refinement, such as hierarchical structures.

3.3.3 Model 3

The third model, referred to as the *hierarchical_model*, incorporates random effects to account for variability at different grouping levels. Besides key predictors like *International.Shipping* and *Buyer.Age*, this model also considers category-specific and product-specific deviations, as represented by the random intercepts assigned to each *Product.Category* and nested *Product.ID* (Bürkner 2018). The model is expressed as:

$$\text{Rating} \sim \text{International.Shipping} + \text{Buyer.Age} + (1 \mid \text{Product.Category/Product.ID})$$

Illustrating the findings of the Bayesian hierarchical ordinal model from **Table 3**, we discovered that the random effects comprehend variabilities across product categories and individual products. The standard deviation of the intercept across *Product.Category* estimates at 0.22, where the 95% CI ranges in [0.01, 1.04]. This estimation depicts a moderate variation in *Rating* distribution among categories. Moreover, the nested random effect for *Product.ID* within *Product.Category* has a standard deviation of 0.04 (95% CI: [0.00, 0.11]), suggesting minimal product-level deviation from category-level *Rating* tendencies.

Table 3: Estimates of Model 3

Covariate	Estimate	Est.Error	95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Multilevel Hyperparameters:							
Product.Category (Levels: 3)							
sd(Intercept)	0.22	0.29	0.01	1.04	1.01	874	1313
Product.Category:Product.ID (Levels: 13)							
sd(Intercept)	0.04	0.03	0.00	0.11	1.00	1887	2186
Regression Coefficients:							
Intercept[1]	-1.66	0.19	-2.01	-1.24	1.00	1648	748
Intercept[2]	-1.00	0.19	-1.35	-0.58	1.00	1626	707
Intercept[3]	-0.30	0.19	-0.65	0.11	1.00	1607	747
Intercept[4]	0.96	0.19	0.61	1.37	1.00	1600	790
International.ShippingNo	0.03	0.04	-0.05	0.12	1.00	5773	2711
Buyer.Age	0.00	0.00	-0.00	0.01	1.00	8118	2812

Consecutively, Intercept[1] estimates -1.66 (95% CI: [-2.01, -1.24]), Intercept[2] (-1.00, 95% CI: [-1.35, -0.58]), Intercept[3] (-0.30, 95% CI: [-0.65, -0.11]), and Intercept[4] (0.96, 95% CI: [0.61, 1.37]), define the log-odds threshold differentiating higher rating levels. In addition to that, the fixed effects are *International.Shipping* and *Buyer.Age* have negligible effect on *Rating* as the credible intervals for all fixed predictors include zero.

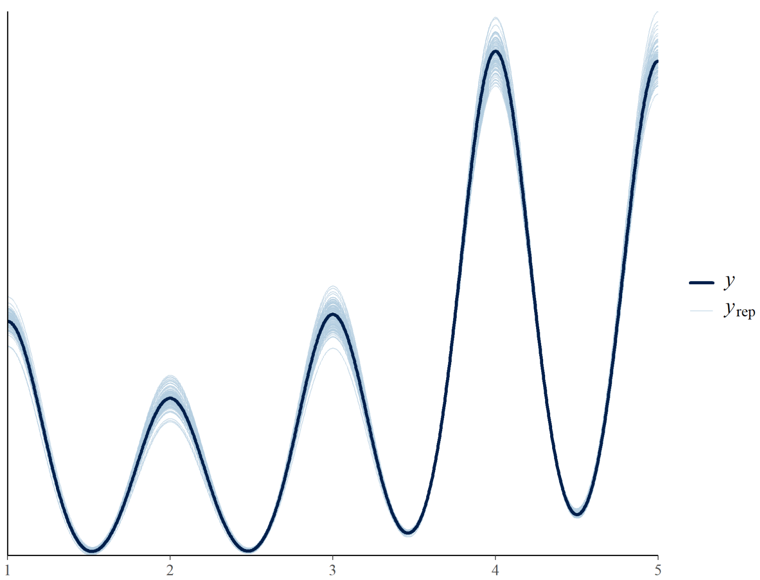
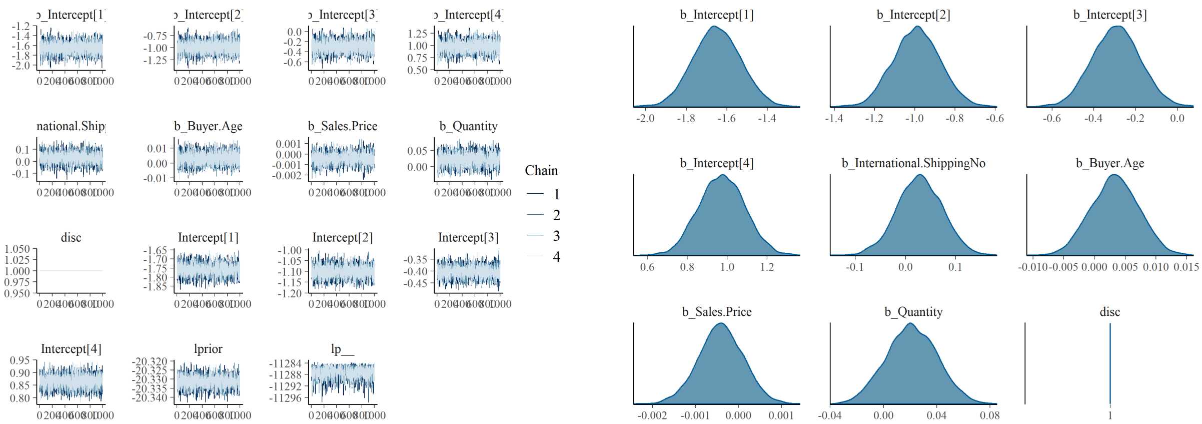


Figure 3: Posterior Predictive Check of Model 3

The posterior predictive check in **Figure 3** exhibits strong alignment between observed data (y) and the predicted data(y_{rep}), supporting the model's ability to capture the underlying structure of the data. Incorporating hierarchical elements improved the capacity of the model to generalize group-level variations.

4 Convergence Diagnostics

The convergence diagnostics for Model 1 (`ordinal_model_fit1`) were evaluated by Monte Carlo Markov Chain (**MCMC**) trace plots, density plots, and key metrics such as **Rhat** and Effective Sample Size (ESS). **Figure 4(a)**, containing trace plots, illustrates that all four chains are well-mixed and maintain a stationary behavior. The density plots from **Figure 4(b)** confirm efficient posterior exploration, displaying smooth, unimodal distributions without irregularities.**Table 1** shows that **Rhat** is approximately 1.00 for all parameters, citing equilibrium among the chains. Moreover, **ESS** (Bulk_ESS and Tail_ESS) are sufficiently large, ensuring satisfactory convergence and reliable inference.

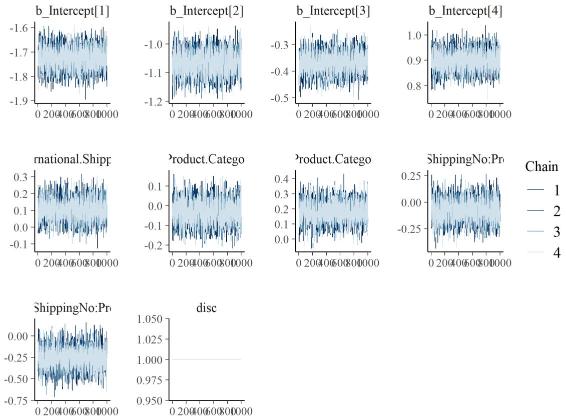


(a) Trace Plots of Model 1 (b) Density Plots of Model 1

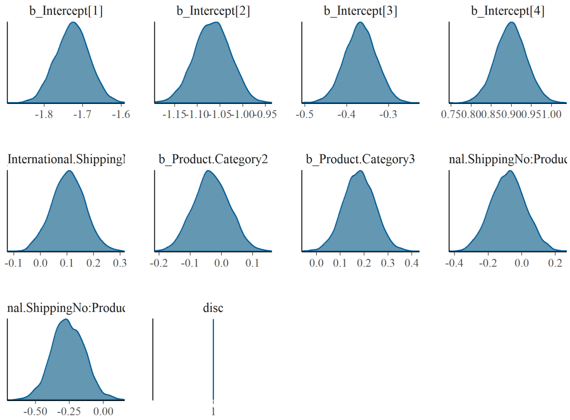
Figure 4: Convergence Diagnostics: Trace and Density Plots for Model 1

Visualizing the trace plots in **Figure 5(a)**, Model 2 has stable and well-mixed MCMC chains across all parameters. Strong overlaps between chains without any evident drift suggest a competent exploration of the posterior distribution. This finding is reinforced by the density plots in **Figure 5(b)**, which show unimodal and symmetric distributions. Additionally, the key metrics **Rhat** values are approximately 1.00 for all parameters, confirming convergence across all chains, and the large values of Bulk_ESS and Tail_ESS indicate low autocorrelation, which is desirable for effective mixing and convergence of the model.

In the convergence diagnosis for Model 3, we observed sampling inefficiencies as initially there were 32 divergent samples. So, in re-iteration, we adjusted the control with `adapt_delta` to 0.999 and `max_treedepth` to 20, which resulted in 3 divergent transitions,



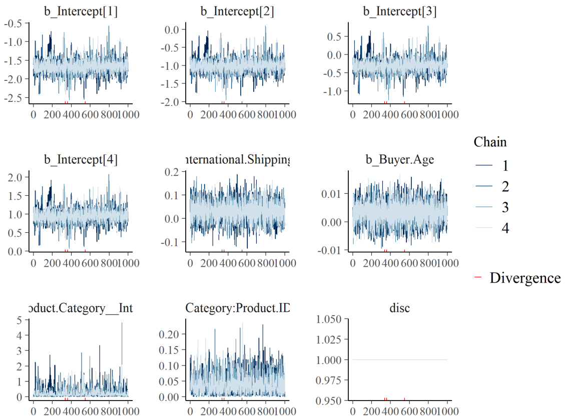
(a) Trace Plots of Model 2



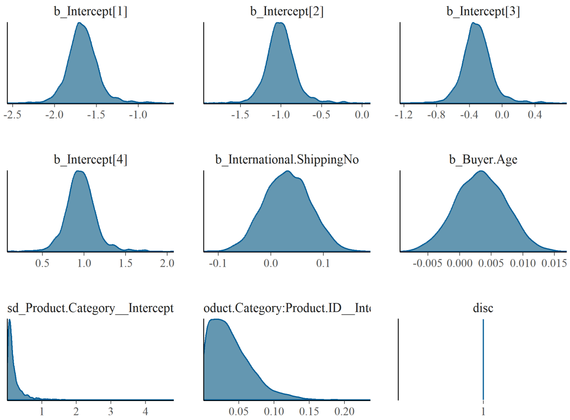
(b) Density Plots of Model 2

Figure 5: Convergence Diagnostics: Trace and Density Plots for Model 2

improving the model stability.In **Figure 6(a)**, the MCMC trace plots visualize drifts indicating slow mixing and potential non-stationarity. Therefore, some chains struggled to fully explore efficient posterior distribution. The density plots in **Figure 6(b)** show unimodal distributions for most of the parameters except for the group-level standard deviations with some skewness. However, the Rhat values of 1.00 and ESS values from **Table 3** confirm convergence. To conclude, reparameterization (increase iterations to 4000) or alternative priors might be needed for improving stability.



(a) Trace Plots of Model 3



(b) Density Plots of Model 3

Figure 6: Convergence Diagnostics: Trace and Density Plots for Model 3

5 Model Comparison

Leave-One-Out Cross Validation (LOO-CV) was utilized to evaluate the predictive performances of our fitted models. The Expected Log Predictive Density (ELPD) is used as the primary evaluation metric, calculated as:

$$ELPD = \sum_{i=1}^{n} \log p(y_i|y_{-i}), \quad LOOIC = -2ELPD$$

We have compared our three fitted models based on their ELPD. From **Table 4**, it is evident that Model 2 (`ordinal_model_fit2`) is established as the reference model, having an ELPD difference (`elpd_diff`) of 0.0 with a standard error (`se_diff`) of 0.0. Compared to Model 2, the predictive performance of Model 3 (`hierarchical_model`) moderately decreased, with `elpd_diff` of -1.9 and `se_diff` of 2.4, indicating slight overfit regardless of incorporating group-level variations. Model 1 (`ordinal_model_fit1`) with all features resulted in the worst performance, with `elpd_diff` of -3.4 and `se_diff` of 3.6, showing an inability to capture meaningful patterns in the data. Hence, Model 2 (`ordinal_model_fit2`) is the best-performing model.

Table 4: Model Comparison with LOO-CV

Model	**elpd_diff**	**se_diff**
ordinal_model_fit2	0.0	0.0
hierarchical_model	−1.9	2.4
ordinal_model_fit1	−3.4	3.6

6 Prior Sensitivity Analysis

Utilizing LOO-CV, Model 2 (*ordinal_model_fit2*) has the best predictive accuracy. To validate the robustness of our best model, we did a prior sensitivity analysis by **re-iterating** the model with two different priors. Now, we applied $\mathcal{N}(0, 1)$ a tighter weakly informative prior and $\mathcal{N}(0, 10)$ a diffuse weakly informative prior on both the regression coefficients and intercepts. The fundamental objective of this prior sensitivity test was to explore whether the choice of priors notably influenced the posterior distributions, model convergence, or predictive accuracy.

The posterior density plots in **Figure 7** compare estimates across different priors and illustrate high overlaps in parameter distributions. This suggests that the posterior is highly data-driven rather than the prior assumptions, indicating the robustness of the model. Moreover, the traceplots from **Figure 8**a and **Figure 8**b show the MCMC chain convergence of models using prior $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, 10)$ consecutively. These plots demonstrate well-mixed chains with no divergence issue across iterations and effective convergence behavior irrespective of prior specification.

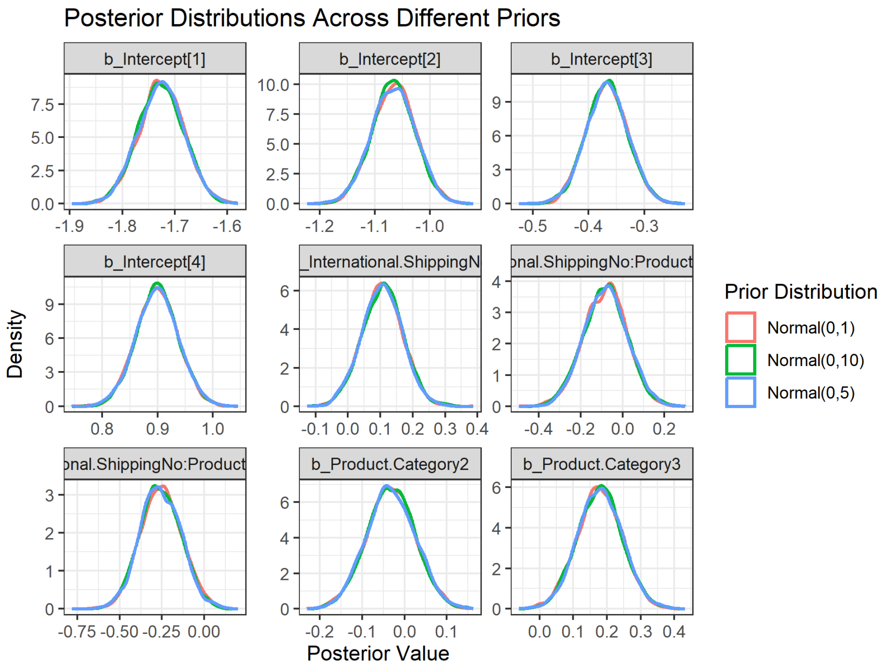
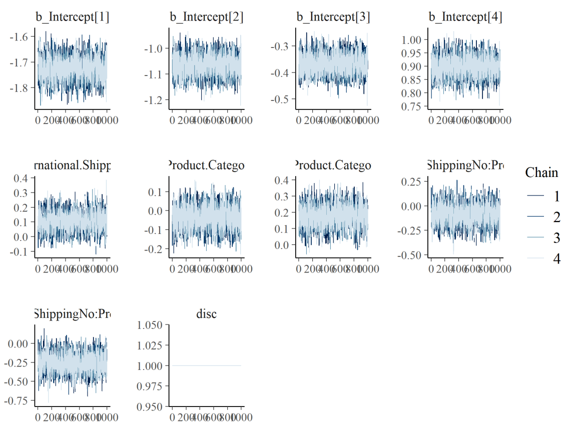
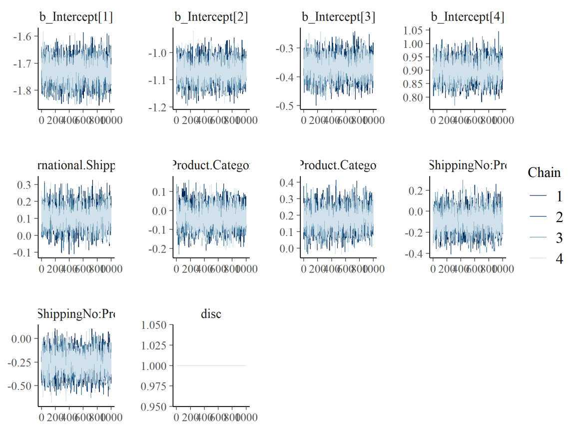


Figure 7: Posterior Distribution Across Different Priors



(a) Trace plot of ordinal model fit2 normal 1



(b) Trace plot of ordinal model fit2 normal 10

Figure 8: Trace Plots of Model 2 with Different Priors

We also performed prior predictive checks vs. posterior predictive checks using these two priors. From **Figure 13**, **Figure 14**, and **Figure 15**, we observed that compared to our best predictive model, *ordinal_model_fit2*, fitted with $\mathcal{N}(0,5)$, the other two priors constrained the predictive performance, meaning they regularized parameter estimates more effectively, which aligns with their stronger or weaker informativeness.

In conclusion, it can be stated from the prior sensitivity analysis that the posterior inference remains stable across different prior choices, strengthening the reliability of the model's robustness.

7 Limitations and Potential improvements

During the Bayesian ordinal regression model deployment phase, we faced several challenges, including minimal variations in the dataset and low computational resources, which greatly impacted the overall interpretability and efficacy of the analysis. The first constraint was the limited set of predictors, leaving out the crucial influential variables such as marketing influence and product-specific quality measures. Furthermore, temporal variations such as seasonality, delivery duration, promotional periods, and trends over time were missing, which could have significantly molded the customers' satisfaction. Secondly, the exceptionally high computational complexity for Bayesian inference, specifically for the hierarchical model, also posed challenges requiring substantial processing time and resources. Therefore, future improvements need expansion of the dataset with time-sensitive and customer-segmentation variables and optimizing computational efficiency through parallel processing to enhance scalability and interpretability.

8 Conclusion

In summary, we implemented the Bayesian ordinal regression approach to analyze the impact of an unexplored factor shipping on product ratings, including buyer demographics and sales attributes. We handled the dataset meticulously by factoring the categorical variables for proper modeling of our ordered target variable. Employing the cumulative logit model, we fitted the baseline ordinal model for the initial analysis of the impact of international shipping on product rating. We gradually increased the model complexities by introducing interaction terms and a hierarchical structure. Overall findings from the analysis suggest that international shipping had a negligible effect on rating distribution, whereas categorical-level and product-level variations had a substantial influence on ratings. The best-performing model included interaction terms through LOO validation.

9 Reflection on own Learnings

This project strengthened our knowledge of the Bayesian Ordinal Regression approach, particularly handling the ordinal data and the proportional odds model. Besides familiarizing ourselves with the computational difficulties of hierarchical structure, we also focused on how important structured variability, model validation, and convergence indicators are.

References

Bürkner, P. C. (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms". In: *The R Journal* 10.1, pp. 395–411. DOI: 10.32614/RJ-2018-017.

Gelman, Andrew et al. (2013). *Bayesian Data Analysis*. 3rd ed. Chapman and Hall/CRC. ISBN: 978-1439840955. DOI: 10.1201/b16018.

McCullagh, P. (1980). "Regression Models for Ordinal Data". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 42.2, pp. 109–142. DOI: 10.1111/j.2517-6161.1980.tb01109.x.

Ravula, P. (2023). "Impact of delivery performance on online review ratings: the role of temporal distance of ratings". In: *J Market Anal.* 11.2, pp. 149–159. DOI: 10.1057/s41270-022-00168-5.

A Appendix

A.1 Additional Figures

A.1.1 Visualization of Conditional Effect

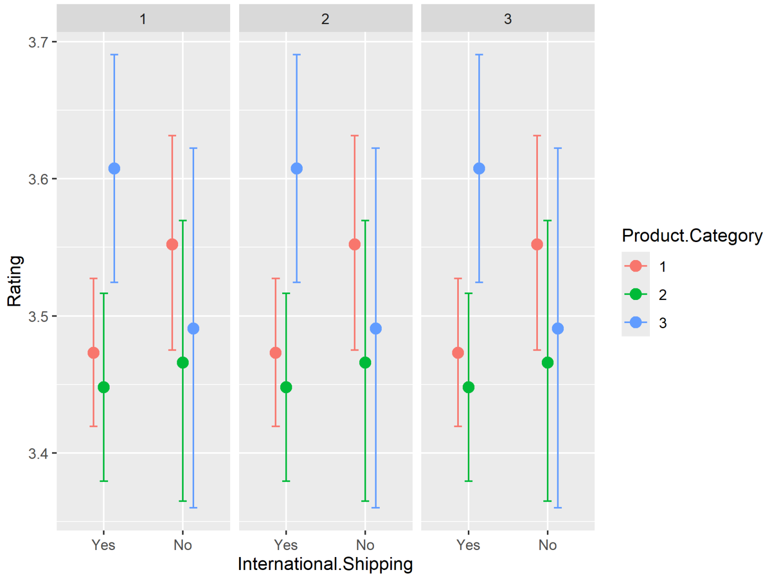


Figure 9: Interaction Effect on Rating of Model 2

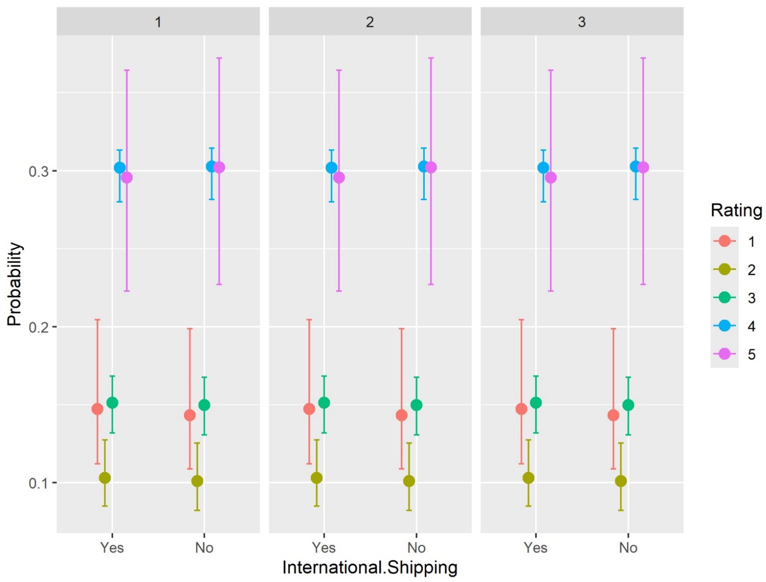


Figure 10: Effect of International Shipping on Ratings by Product Category (Model 3)

A.1.2 Visualization of Random Effects

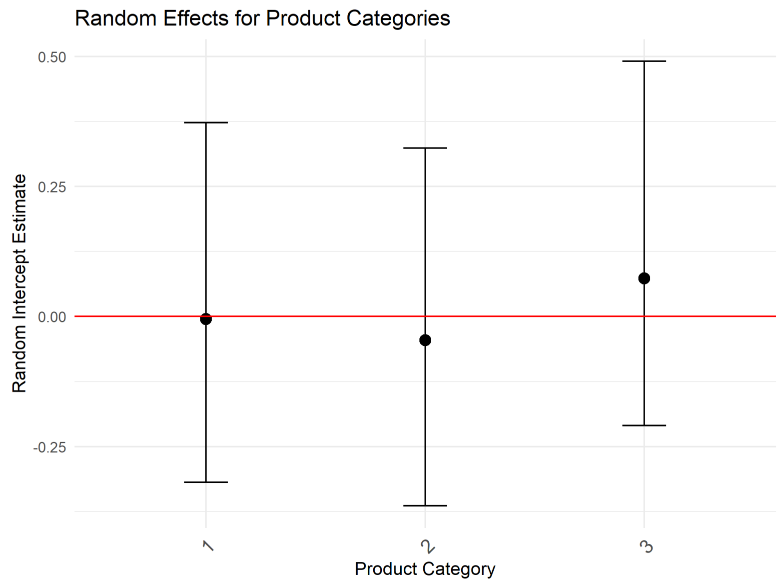


Figure 11: Random Effects For Product Category in Model 3 (Hierarchical Model)

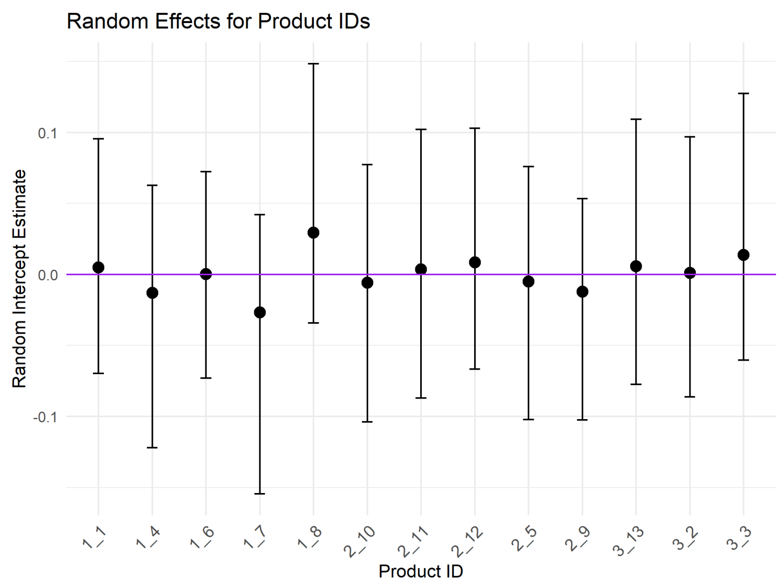


Figure 12: Random Effects For Product IDs in Model 3 (Hierarchical Model)

A.1.3 Prior VS Posterior Predictive Check

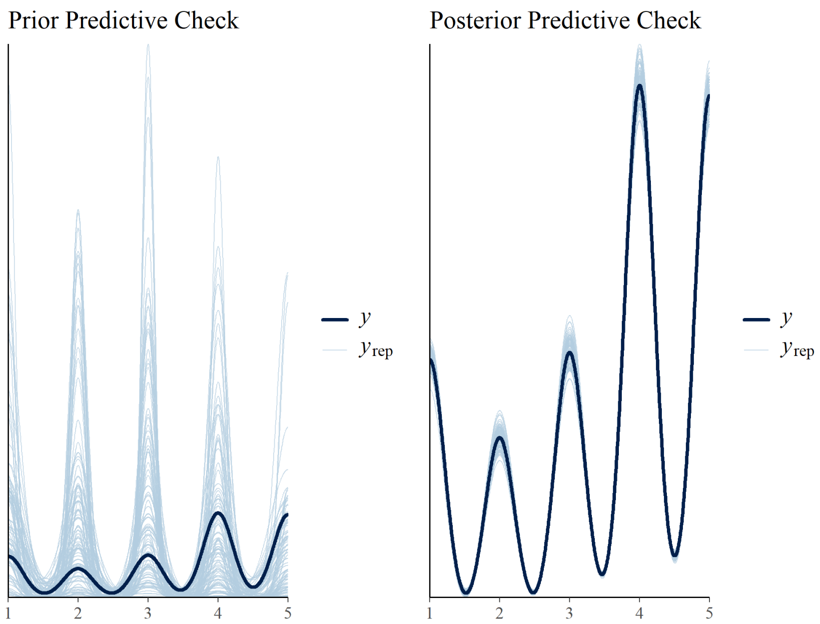


Figure 13: Prior Vs. Posterior with N(0,5) of Model 2

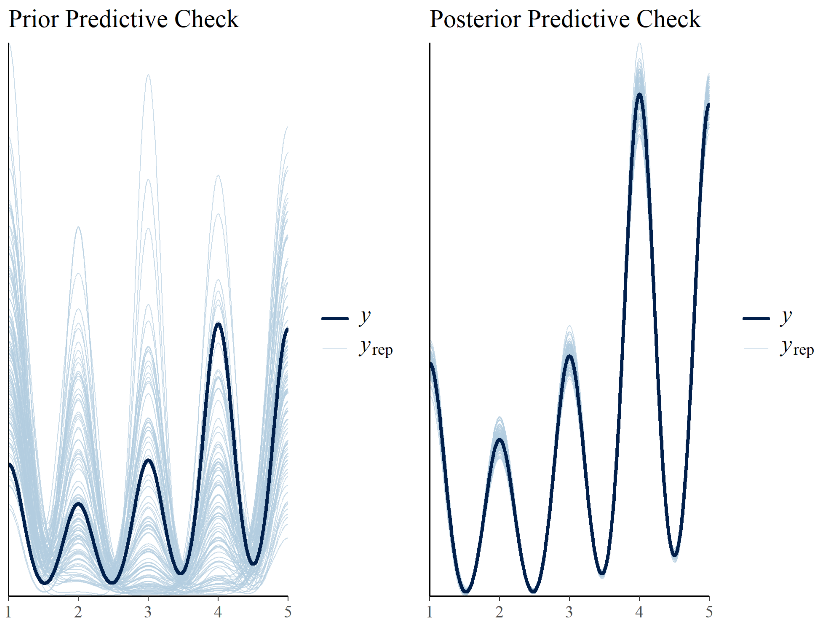


Figure 14: Prior Vs. Posterior with N(0,1) of Model 2

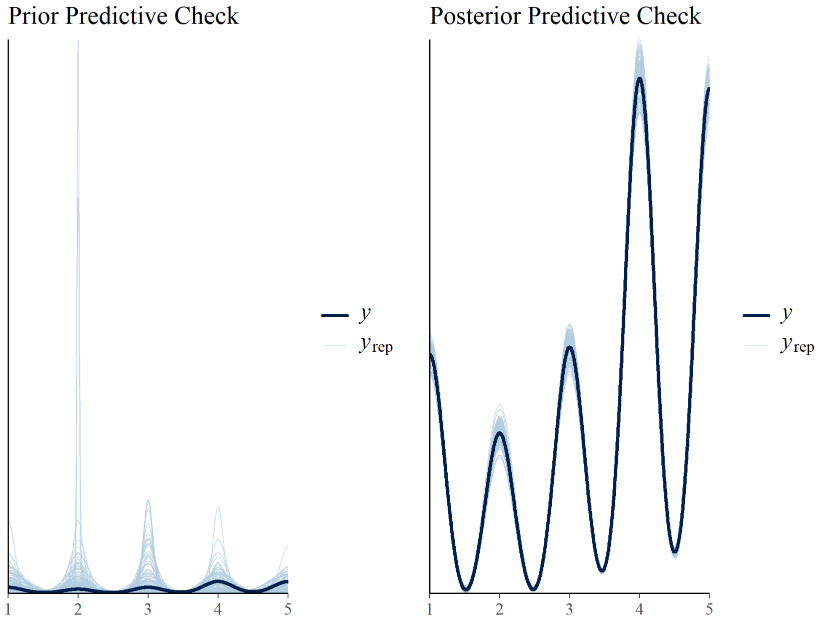


Figure 15: Prior Vs. Posterior with N(0,10) of Model 2

A.1.4 Correlation Matrix

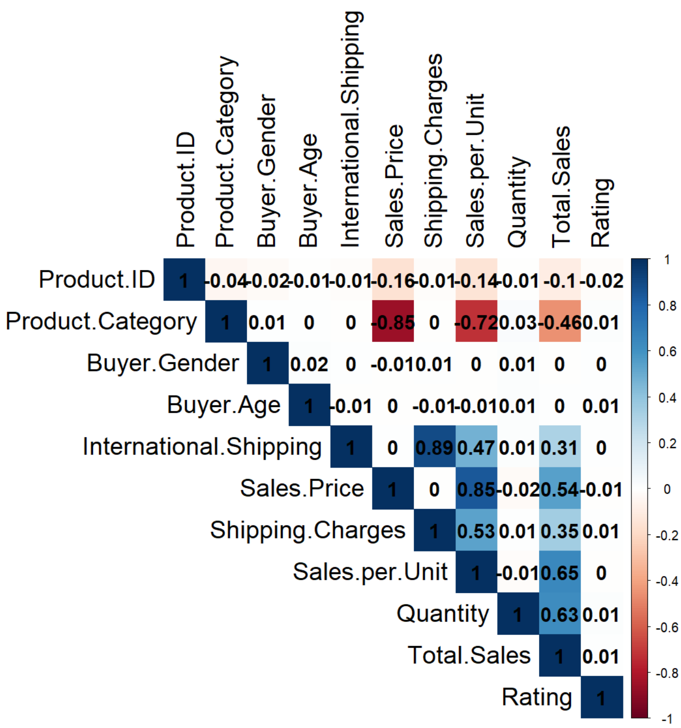


Figure 16: Correlation Matrix