Modeling the number of daily orders for a logistics company using Bayesian linear regression

Lowell Monis[*,†]

Abstract. This report presents a Bayesian approach to model the number of daily orders for a Brazilian logistics company using time-series data from a period of 60 days.

Introduction

In today's rapidly evolving business landscape, accurate demand forecasting has become a critical component of operational efficiency and strategic planning. For logistics companies, the ability to predict daily order volumes enables optimal resource allocation, workforce scheduling, and service quality maintenance. Underestimation of demand can lead to service delays, customer dissatisfaction, and lost revenue opportunities, while overestimation results in unnecessary labor costs and inefficient resource utilization. The challenge of demand forecasting is particularly acute in dynamic operational environments where multiple factors–ranging from temporal patterns to order characteristics and sector-specific demands–interact in complex ways to influence daily workload.

The Daily Demand Forecasting Orders dataset represents real operational data from a Brazilian logistics company, collected over a sixty-day period of actual business operations. This dataset captures the multifaceted nature of daily demand through twelve operational and categorical variables that collectively describe the scheduling context, order urgency profiles, order type distributions, and sector-specific order volumes. Unlike synthetic or simulated datasets, this collection reflects the genuine complexities and interdependencies present in real-world logistics operations, where order patterns emerge from the interplay of calendar effects, client behavior, and operational constraints.

Traditional approaches to demand forecasting often rely on time series methods or classical regression techniques that provide point estimates of expected demand. While these methods have proven useful in many contexts, they typically fail to quantify the uncertainty inherent in predictions or to account for the complex dependencies among predictor variables. Furthermore, classical variable selection methods such as stepwise regression or information criterion-based approaches select a single "best" model, ignoring model uncertainty and potentially discarding valuable predictive information contained in alternative model specifications.

Bayesian methods offer a principled alternative framework that addresses these limitations. By treating model parameters as random quantities with probability distributions rather than fixed unknown constants, Bayesian regression provides not only point predictions but complete posterior distributions that characterize uncertainty in both parameter estimates and future predictions. Moreover, Bayesian model selection and model averaging techniques allow us to quantify the importance of each predictor variable through posterior inclusion probabilities, providing interpretable measures of variable relevance while accounting for correlations and redundancies among features.

In this analysis, a Bayesian linear regression approach is employed to model daily total orders as a function of the twelve operational features. The invariant g-prior specification is adopted for regression coefficients, which ensures that inference remains consistent under linear transformations of the predictor variables–a desirable property when working with features measured on different scales. This approach enables both parameter estimation and principled variable selection within a unified probabilistic framework.

Objectives

This study addresses the following question:

> Can an accurate Bayesian linear regression model be developed to predict daily total orders from operational and sector-specific features, and which of the twelve predictor variables are most important for explaining variation in daily demand?

Specifically, this study will:

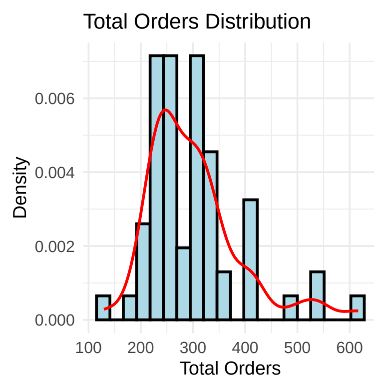
- Build a Bayesian linear regression model using all 12 available features and assess its predictive accuracy
- Quantify the importance of each feature through posterior inclusion probabilities
- Identify the minimal set of predictive features that capture the essential structure of daily demand

The Data

The Daily Demand Forecasting Orders (Ferreira et al. (2016)) dataset was collected from a Brazilian logistics company over a continuous sixty-day operational period. It contains 60 daily observations, each characterized by twelve predictor variables and one target variable representing the total number of orders requiring daily treatment. This dataset was originally collected for academic research purposes at Universidade Nove de Julho and represents an authentic record of operational demand patterns in a real business environment.

```
## 'data.frame':    60 obs. of  13 variables:
##  $ week                       : int  1 1..
##  $ day                        : int  4 5..
##  $ non.urgent                 : num  316..
##  $ urgent                     : num  223..
##  $ typeA                      : num  61...
##  $ typeB                      : num  175..
##  $ typeC                      : num  302..
##  $ fiscal.sector              : num  0 0..
##  $ traffic.controller.sector: int  655..
##  $ banking1                   : int  449..
##  $ banking2                   : int  188..
##  $ banking3                   : int  147..
##  $ target                     : num  540..
```

The response variable, total orders, represents the aggregate number of orders that must be processed on a given day. This is the primary quantity of interest for operational planning and resource allocation decisions.



The twelve predictor variables fall into four conceptual categories:

Calendar Variables:

- Week of the Month: A categorical variable indicating which week of the month the observation falls into (values 1.0 through 5.0).
- Day of the Week: A categorical variable representing the day of the business week (values 2.0 through 6.0, corresponding to Monday through Friday).

Order Urgency Profile:

- Non-urgent Orders: The count of orders classified as non-urgent for that day.

- Urgent Orders: The count of orders requiring expedited processing or same-day handling.

Order Type Distribution:

- Order Type A: The count of orders classified under category A.
- Order Type B: The count of orders classified under category B.
- Order Type C: The count of orders classified under category C.

Sector-Specific Orders:

- Fiscal Sector Orders: The count of orders originating from fiscal or financial administrative operations.
- Orders from Traffic Controller Sector: The count of orders associated with traffic management or vehicle-related administrative services.
- Banking Orders (1): The count of orders from banking operations, first category.
- Banking Orders (2): The count of orders from banking operations, second category.
- Banking Orders (3): The count of orders from banking operations, third category.

The following is a descriptive summary of the data.

```
##       week            day
##  Min.   :1.000   Min.   :2.000
##  1st Qu.:2.000   1st Qu.:3.000
##  Median :3.000   Median :4.000
##  Mean   :3.017   Mean   :4.033
##  3rd Qu.:4.000   3rd Qu.:5.000
##  Max.   :5.000   Max.   :6.000
##    non.urgent          urgent
##  Min.   : 43.65   Min.   : 77.37
##  1st Qu.:125.35   1st Qu.:100.89
##  Median :151.06   Median :113.11
##  Mean   :172.55   Mean   :118.92
##  3rd Qu.:194.61   3rd Qu.:132.11
##  Max.   :435.30   Max.   :223.27
##      typeA             typeB
##  Min.   : 21.83   Min.   : 25.12
##  1st Qu.: 39.46   1st Qu.: 74.92
##  Median : 47.17   Median : 99.48
##  Mean   : 52.11   Mean   :109.23
##  3rd Qu.: 58.46   3rd Qu.:132.17
##  Max.   :118.18   Max.   :267.34
##      typeC        fiscal.sector
##  Min.   : 74.37   Min.   :  0.000
##  1st Qu.:113.63   1st Qu.:  1.243
##  Median :127.99   Median :  7.832
##  Mean   :139.53   Mean   : 77.396
##  3rd Qu.:160.11   3rd Qu.: 20.361
##  Max.   :302.45   Max.   :865.000
##  traffic.controller.sector
##  Min.   :11992
##  1st Qu.:34994
##  Median :44312
##  Mean   :44504
##  3rd Qu.:52112
##  Max.   :71772
##     banking1         banking2
##  Min.   :  3452   Min.   : 16411
##  1st Qu.: 20130   1st Qu.: 50681
##  Median : 32528   Median : 67181
##  Mean   : 46641   Mean   : 79401
##  3rd Qu.: 45119   3rd Qu.: 94788
##  Max.   :210508   Max.   :188411
##     banking3        target
##  Min.   : 7679   Min.   :129.4
##  1st Qu.:12610   1st Qu.:238.2
##  Median :18012   Median :288.0
##  Mean   :23115   Mean   :300.9
##  3rd Qu.:31048   3rd Qu.:334.2
##  Max.   :73839   Max.   :616.5
```

All predictor variables are recorded as either continuous or integer counts, except the calendar variables, which are discrete. The calendar variables are, however, represented as continuous numerical values despite being inherently categorical. The target variable ranges from a minimum of 129.4 to a maximum of 616.5 orders per day, with a mean of approximately 300.9 orders.

The dataset contains no missing values, and all variables are measured on compatible scales. However, standardization of predictors is performed prior to modeling to ensure that the invariant g-prior treats all regression coefficients on a comparable scale and to facilitate interpretation of coefficient magnitudes.

Preprocessing

First, the response vector and the feature matrix are prepared. Following through the observations made earlier about the difference in the scales of each feature, the feature matrix can be standardized to $\mathcal{N}(0,1)$ to facilitate the invariant g-prior, and make the interpretation of coefficients more meaningful.

$$\mathbf{X}_{\text{scaled}} = \frac{\mathbf{X} - \mu}{\sigma}$$

The design matrix needs to be constructed as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,12} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,12} \end{bmatrix}_{n \times p}$$

where $n = 60$ days, $p = 13$ parameters (1 intercept + 12 features), and $x_{i,j}$ represents the value of feature j for each day i.

To do this, the features are bound with an all-ones vector column-wise:

$$\mathbf{X} = [\mathbf{1}_n \mid \mathbf{X}_{\text{features}}] \in \mathbb{R}^{n \times p}$$

Model Setup

Since the response variable for this analysis is continuous, representing the total number of daily orders, a linear regression model is the appropriate choice for prediction. The goal is to estimate the expected number of orders given the observed operational feature measurements. A Bayesian approach will be employed to quantify uncertainty in both predictions and parameter estimates.

For each observation $i = 1, \ldots, n$ where $n = 60$ daily observations, the response variable Y_i follows a normal distribution:

$$Y_i \mid \mathbf{x}_i, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$$

Here:

- $Y_i \in \mathbb{R}$ represents the total number of orders for day i
- $\mathbf{x}_i \in \mathbb{R}^p$ is the vector of 12 standardized feature measurements for day i plus an intercept term, so $p = 13$- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{12})^T \in \mathbb{R}^{13}$ is the vector of regression coefficients to be estimated
- β_0 is the intercept representing the baseline expected number of orders
- β_j for $j = 1, \ldots, 12$ quantifies the change in expected total orders associated with a one-standard-deviation increase in the j-th feature, holding all other features constant
- $\sigma^2 > 0$ is the error variance, representing the variability in daily orders not explained by the predictor variables

The expected number of orders for day i is modeled as a linear combination of the features:

$$\mathbb{E}[Y_i \mid \mathbf{x}_i, \boldsymbol{\beta}] = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \sum_{j=1}^{12} \beta_j x_{i,j}$$

This specification assumes that the relationship between the predictors and the response is linear and additive, and that the errors $\epsilon_i = Y_i - \mathbf{x}_i^T \boldsymbol{\beta}$ are independent and identically distributed as $\mathcal{N}(0, \sigma^2)$.

Under the assumption that observations are independent, the joint likelihood function for the complete data $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ given the design matrix $\mathbf{X}$, coefficient vector $\boldsymbol{\beta}$, and error variance σ^2 is:

$$\begin{aligned} p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2} \\ &= (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})} \end{aligned}$$

The last line uses vector notation where $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2$ is the residual sum of squares.

For computational purposes, one works with log-likelihood:

$$\log p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\log p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$$

This form will be useful for deriving posterior distributions and for computational implementation.

Prior Distribution

In Bayesian inference, prior distributions encode beliefs about parameters before observing the data. For this analysis, one adopts a semi-conjugate prior structure that separates the prior on the regression coefficients from the prior on the error variance.

The error variance σ^2 is assigned an inverse-gamma prior:

$$\sigma^2 \sim \text{Inverse-Gamma}(a, b)$$

with probability density function:

$$p(\sigma^2) = \frac{b^a}{\Gamma(a)}(\sigma^2)^{-(a+1)}\exp\left\{-\frac{b}{\sigma^2}\right\}, \quad \sigma^2 > 0$$

where $a > 0$ is the shape parameter and $b > 0$ is the scale parameter, while)$\Gamma(\cdot)$ is the gamma function.

For this analysis, one chooses: $a = 0.5$, a weakly informative choice that places minimal prior constraint, and $b = 0.5 \times \text{MSE}_{\text{full}}$, where MSE_{full} is the mean squared error from the ordinary least squares fit of the full model. This specification centers the prior for σ^2 near the empirical residual variance while maintaining substantial prior uncertainty.

For the regression coefficient vector $\boldsymbol{\beta}$, one adopts Zellner's invariant g-prior:

$$\boldsymbol{\beta} \mid \sigma^2 \sim \mathcal{N}\left(\boldsymbol{\mu}_0, g\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

where $\boldsymbol{\mu}_0 = \mathbf{0}_p$ is a p-dimensional zero vector, representing no prior preference for positive or negative effects, $g > 0$ is a scalar hyperparameter controlling the prior variance, and $(\mathbf{X}^T\mathbf{X})^{-1}$ is the inverse of the information matrix from the design.

Important properties of the invariant g-prior include mean-centering at zero, indicating no prior belief that any feature systematically increases or decreases total orders, a variance structure proportional to data with prior covariance being proportional to the sampling covariance of the MLE (thus, features with less information in the data receive larger prior variances, and the prior accounts for correlations among predictors through the full inverse matrix), and the invariance property, due to which the prior is invariant to linear transformations of the predictors.

The hyperparameter g controls the relative importance of the prior versus the data. Larger values of g indicate more prior variance (less informative prior), while smaller values represent stronger prior beliefs. For this analysis $g = n = 60$ is adopted, which provides a weakly informative prior that allows the data to largely determine the posterior while maintaining some regularization to prevent overfitting.

Joint Prior Distribution

The complete prior specification is:

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta} \mid \sigma^2) \cdot p(\sigma^2)$$

Explicitly:

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma^2) &= \mathcal{N}\left(\boldsymbol{\beta} \mid \mathbf{0}, g\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right) \cdot \text{I-G}(\sigma^2 \mid a, b) \\ &= \frac{1}{(2\pi)^{\frac{p}{2}}(g\sigma^2)^{\frac{p}{2}}|\mathbf{X}^T\mathbf{X}|^{-\frac{1}{2}}} e^{-\frac{1}{2g\sigma^2}\boldsymbol{\beta}^T(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta}} \\ &\times \frac{b^a}{\Gamma(a)}(\sigma^2)^{-(a+1)}e^{-\frac{b}{\sigma^2}} \end{aligned}$$

This semi-conjugate structure does not lead to closed-form full conditional distributions for both parameters simultaneously, but it does yield tractable marginal and conditional posteriors that facilitate efficient computation.

Posterior Distribibtion

By Bayes' theorem, the posterior distribution combines the likelihood and prior:

$$p(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}, \mathbf{X}) = \frac{p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \cdot p(\boldsymbol{\beta}, \sigma^2)}{p(\mathbf{y} \mid \mathbf{X})}$$

The denominator $p(\mathbf{y} \mid \mathbf{X}) = \int\int p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}, \sigma^2)\, d\boldsymbol{\beta}\, d\sigma^2$ is the marginal likelihood, which serves as a normalizing constant.

Working with the unnormalized posterior (which is proportional to the numerator):

$$p(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}, \mathbf{X}) \propto p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \cdot p(\boldsymbol{\beta} \mid \sigma^2) \cdot p(\sigma^2)$$

Under the g-prior, the full conditional distribution of $\boldsymbol{\beta}$ given σ^2 and the data has a closed form. To derive this, we combine the normal likelihood with the normal prior:

$$p(\boldsymbol{\beta} \mid \sigma^2, \mathbf{y}, \mathbf{X}) \propto e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} e^{-\frac{1}{2g\sigma^2}\boldsymbol{\beta}^T(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta}}$$

$$p(\boldsymbol{\beta} \mid \sigma^2, \mathbf{y}, \mathbf{X}) \propto e^{-\frac{1}{2\sigma^2}\left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \frac{1}{g}\boldsymbol{\beta}^T(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta}\right]}$$

Expanding the quadratic forms and completing the square, one obtains:

$$\boldsymbol{\beta} \mid \sigma^2, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}_n, \boldsymbol{\Sigma}_n)$$

where:

$$\boldsymbol{\Sigma}_n = \frac{g}{g+1}\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

$$\boldsymbol{\beta}_n = \frac{g}{g+1}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

The posterior mean $\boldsymbol{\beta}_n$ can be related to the ordinary least squares (OLS) estimate:

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Thus:

$$\boldsymbol{\beta}_n = \frac{g}{g+1}\hat{\boldsymbol{\beta}}_{\text{OLS}}$$

This shows that the posterior mean is a shrinkage estimator, pulling the OLS estimate toward zero by the factor $\frac{g}{g+1}$. With $g = n = 60$, this shrinkage factor is $60/61 \approx 0.984$, indicating very mild shrinkage–the posterior mean is nearly identical to the OLS estimate, but with a small amount of regularization.

A remarkable feature of the g-prior is that the marginal posterior distribution of σ^2 (integrating out $\boldsymbol{\beta}$) has a closed form. Specifically:

$$\sigma^2 \mid \mathbf{y}, \mathbf{X} \sim \text{Inverse-Gamma}(\tilde{a}, \tilde{b})$$

where $\tilde{a} = a + \frac{n}{2}$, $\tilde{b} = b + \frac{1}{2}\left[\mathbf{y}^T\mathbf{y} - \frac{g}{g+1}\mathbf{y}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}\right]$.

The term in brackets can be simplified using matrix identities. Define the sum of squared residuals from the g-prior fit:

$$\text{SSR}_g = \mathbf{y}^T\mathbf{y} - \frac{g}{g+1}\mathbf{y}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

This can be interpreted as:

$$\begin{aligned}\text{SSR}_g &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\beta}_n \\ &= \text{Total SS} - \text{Explained SS under g-prior}\end{aligned}$$

Thus, $\tilde{b} = b + \frac{1}{2}\text{SSR}_g$.

Constructing the Full Model

The closed-form conditional and marginal posteriors enable efficient posterior sampling via composition sampling (also called direct sampling), which is exact and requires no Markov chain convergence:

For $s = 1, \ldots, S$ iterations:

- Sample σ^2: Draw $(\sigma^2)^{(s)}$ from the marginal posterior $(\sigma^2)^{(s)} \sim \text{Inverse-Gamma}(\tilde{a}, \tilde{b})$ This is implemented as: $(\sigma^2)^{(s)} = \frac{1}{\text{Gamma}(\tilde{a},\tilde{b})}$
- Sample $\boldsymbol{\beta}$ given σ^2: Draw $\boldsymbol{\beta}^{(s)}$ from the conditional posterior $\boldsymbol{\beta}^{(s)} \mid (\sigma^2)^{(s)} \sim \mathcal{N}\left(\boldsymbol{\beta}_n, \boldsymbol{\Sigma}_n^{(s)}\right)$ where $\boldsymbol{\Sigma}_n^{(s)} = \frac{g}{g+1}(\sigma^2)^{(s)}(\mathbf{X}^T\mathbf{X})^{-1}$

The collection $\{(\boldsymbol{\beta}^{(1)}, (\sigma^2)^{(1)}), \ldots, (\boldsymbol{\beta}^{(S)}, (\sigma^2)^{(S)})\}$ consists of S independent draws from the exact joint posterior $p(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}, \mathbf{X})$.

Now, one can proceed with posterior inference. With S posterior samples in hand, one can approximate any posterior quantity of interest.

The posterior mean,

$$\mathbb{E}[\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}] \approx \hat{\boldsymbol{\beta}} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\beta}^{(s)}$$

The 95% credible intervals can be determined via quantiles of the samples.

The results of the full model are as follows:

```
##                   Predictor Post_Mean
## 1                            295.9491
## 2                      week   -0.0654
## 3                       day   -0.0333
## 4                non.urgent   -0.5450
## 5                    urgent   -0.2885
## 6                     typeA   18.7177
## 7                     typeB   50.0077
## 8                     typeC   41.0553
## 9             fiscal.sector   -0.1095
## 10 traffic.controller.sector    0.1088
## 11                 banking1    0.1658
## 12                 banking2    0.1299
## 13                 banking3    0.0098
##    CI_Lower CI_Upper Significant
## 1  285.8942 305.9242         Yes
## 2  -13.6963  13.4147          No
## 3  -14.6884  14.8050          No
## 4  -78.5754  74.1302          No
## 5  -28.9989  28.9436          No
## 6   -3.2515  40.7543          No
## 7   27.4073  73.1069         Yes
## 8    2.7302  78.8966         Yes
## 9  -13.0382  12.4176          No
## 10 -16.7325  17.6984          No
## 11 -36.6977  38.2690          No
## 12 -36.8792  38.3887          No
## 13 -12.6985  12.9526          No
```

From the full model, the intercept, and the orders of type B and type C, are the only features that are significant, in the sense that their 95% credible intervals do not include zero, which means that there is no uncertainty about the direction in which the coefficient plays a role on the target variable. For example, for every 1-SD increase in `typeB`, there is a positive increase in daily orders. Now, if this included zero, one could not be certain if the value would increase or decrease.

Model Selection

Variable Selection via Gibbs Samopling

Not all twelve predictors may contribute meaningfully to predicting daily orders. Bayesian model selection assigns posterior inclusion probabilities to each feature, quantifying which variables are truly important while accounting for collinearity and model uncertainty.

One uses a Gibbs sampler to explore the space of $2^{12} = 4096$ possible models. Binary indicators $z_j \in \{0, 1\}$ denote whether predictor j is included. The intercept is always retained. At each iteration, the inclusion indicators are updated by comparing marginal likelihoods across models. Regression coefficients conditional on the selected variables are then sampled. Functions from external R scripts are used.

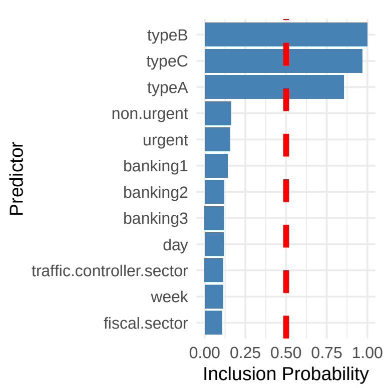
Posterior Inclusion Probabilities

The posterior inclusion probability $P(z_j = 1 \mid \mathbf{y}, \mathbf{X})$ represents the probability that predictor j is truly relevant for explaining daily orders, averaging over all possible models. Values near 1 indicate strong evidence for inclusion; values near 0 suggest the variable is redundant or uninformative.

```
##                       Predictor
## 1                         typeB
## 2                         typeC
## 3                         typeA
## 4                    non.urgent
## 5                        urgent
## 6                      banking1
## 7                      banking2
## 8                      banking3
## 9                           day
## 10 traffic.controller.sector
## 11                         week
## 12                fiscal.sector
##    Inclusion_Probability
## 1                 0.9984
## 2                 0.9672
## 3                 0.8553
## 4                 0.1608
## 5                 0.1567
## 6                 0.1395
## 7                 0.1195
## 8                 0.1174
## 9                 0.1158
## 10                0.1141
## 11                0.1129
## 12                0.1070
```

```
## Features with inclusion probability > 0.5:
```

```
## [1] "typeB" "typeC" "typeA"
```



From the above illustrations, only the orders of type A, B, and C, are in more than half the models that are significant.

Model Averaging

Rather than selecting a single “best” model, one can average predictions across all visited models, weighted by their posterior probabilities. The samples from earlier already incorporate this averaging–coefficients are zero when variables are excluded, producing automatic shrinkage.

```
##                  Predictor Post_Mean
## 1                           295.9889
## 2                     week    0.0115
## 3                      day    0.0365
## 4               non.urgent    2.1756
## 5                   urgent    1.1293
## 6                    typeA   15.6900
## 7                    typeB   49.5664
## 8                    typeC   38.8654
## 9            fiscal.sector    0.0271
## 10 traffic.controller.sector  -0.0060
## 11                banking1    0.4536
## 12                banking2    0.0164
## 13                banking3    0.0461
##    CI_Lower CI_Upper
## 1  285.7962 305.9726
## 2   -4.3292   4.6238
## 3   -4.6035   5.3016
## 4  -10.9217  39.1648
## 5   -6.8433  22.2700
## 6    0.0000  30.2898
## 7   31.2806  65.5657
## 8    0.0000  54.6823
## 9   -3.9383   4.2941
## 10  -4.9503   4.9814
## 11  -6.2968  12.8093
## 12  -7.5526   8.1152
## 13  -4.2553   5.0608
```

Coefficients for variables with low inclusion probabilities have posterior means shrunken toward zero, reflecting uncertainty about their relevance.

Results and Conclusion

Reduced Model with Selected Features

For clearer interpretation and potential improved prediction, one can refit a Bayesian linear regression using only the features with posterior inclusion probability exceeding 0.5.

```
##     Predictor Post_Mean CI_Lower
## 1 (Intercept)    296.00   285.77
## 2       typeB     49.92    36.48
## 3       typeC     40.82    28.68
## 4       typeA     18.60     7.14
##   CI_Upper
## 1   306.51
## 2    63.05
## 3    52.69
## 4    30.46
```

Some metrics for the predictive model is as follows:

```
##
## Prediction Performance Metrics:

##   MSE:         25.61

##   RMSE:        5.06

##   R-squared:   0.9968
```

Discussion

The Bayesian variable selection identified a parsimonious subset of the original 12 predictors that capture the essential predictive information. These happen to be the three order type predictors. These features with high inclusion probabilities represent the operational factors most strongly associated with daily demand variation. The other features were very rarely found in significant models, with posterior probabilities well below 0.5. Model averaging helped create a model that uses all features but only promotes partial contributions from each variable depending on its posterior probability, thus remove the need for creating a selective set of predictors. The selected model achieves more than 99% of the variation in daily orders, demonstrating great predictive capability. The identified features should be prioritized in operational planning and demand forecasting systems.

References

Ferreira, R., Martiniano, A., Ferreira, A., Ferreira, A., and Sassi, R. (2016). "Daily Demand Forecasting Orders." UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5BC8T. 2