Sri Lankan Journal of Applied Statistics, Vol 26 (Special Issue)
*Selected Papers from International Statistics Conference (ISC) 2024*

# Utilizing Bayesian Regression Analysis and Optimization Approaches to Identify Main Drivers in Inventory Management

**K. A. I. H. K. Arachchi*** **and A. W. L. P. Thilan**

Department of Mathematics, Faculty of Science,
University of Ruhuna, Sri Lanka
*Corresponding author: ishanihimaya@gmail.com

Received: 22nd June, 2025/ Revised: 04th October, 2025/ Published: 30th November, 2025



## ABSTRACT

*Inventory management is a crucial research area focusing on stock optimization, cost reduction, and supply chain efficiency, with particular attention given to optimizing practices and identifying key factors. If inventory management fails to maintain a sustainable level, it directly hampers the company. To address this problem, we propose a new methodology that combines Bayesian regression analysis with optimization techniques. The historical sales data used in this study was obtained from an online database platform called "Kaggle". A Bayesian logistic regression approach was utilized to incorporate prior knowledge from this historical data into the Bayesian design framework. The Bayesian optimal experimental design was then obtained by maximizing the expected utility function, with the Kullback–Leibler divergence selected as the utility criterion to support precise estimation of model parameters. The resulting optimal experiment set represents the selection of an optimal covariate set aimed at improving the prediction of the dependent variable. To determine the optimal design, the Approximate Coordinate Exchange (ACE) optimization algorithm was employed, starting with three randomly generated initial designs. This study demonstrates the efficacy of Bayesian regression analysis and optimization techniques in identifying and prioritizing key drivers impacting inventory management. By incorporating prior knowledge and accounting for uncertainty, businesses can achieve more accurate demand forecasting, improved inventory turnover, and reduced holding costs, ultimately driving superior supply chain efficiency and organizational performance.*

**Keywords: Approximate coordinate exchange, Kullback-Leibler divergence utility, Logistic regression approach, Optimal design**

## 1 Introduction

Inventory is a company's most important asset, often representing up to half of its expenses or even its total capital investment. As such, inventory stands as a vital asset in the company. The types of inventory vary depending on the quantity of goods or the nature of the company. Effective inventory management is crucial as it allows businesses to accurately track stock levels, place precise orders, and ensure timely fulfillment of customer needs. In light of the increasing demands of modern consumers, inventory management concepts have been extensively reviewed in the literature (Munyaka and Yadavalli, 2022; Demeter and Golini, 2014; Baker, 2007). Demand is a critical variable in inventory management, influencing and being influenced by inventory practices. The importance of effective inventory management lies in its ability to meet the needs of both businesses and consumers, ensuring efficiency and satisfaction on both ends.

In order to manage inventory more effectively, the factors affecting it should be studied and the factors that make the most significant changes should be identified. Inventory management is influenced by numerous external and internal factors. These factors, which affected inventory levels, fell into four categories: market factors, internal operations, supply chain characteristics, and business strategy (Demeter and Golini, 2014). Identifying and understanding the impact of these factors is crucial for devising effective inventory management strategies.

By effectively monitoring the factors affecting an inventory and optimizing those identified factors, one should be able to identify risks and trends in predicting future sales for the company. Inventory optimization is one of the best methods for reducing capital investment while maintaining sufficient inventory levels. Identifying main drivers (factors) in inventory is key to maintaining more favorable inventory levels (Teplická and Čulková, 2020). However, despite the extensive literature on optimizing inventory management and identifying primary drivers, a broadly applicable method for assessing the impact of factors on inventory remains elusive. As noted earlier, these factors vary significantly across companies and customer needs. Therefore, it is imperative to identify the critical factors affecting inventory and to optimize them accordingly.

Generally, the regression analysis approach is employed to identify the relationships between relevant factors affecting inventory (Hoppenheit and Günthner, 2015; Vasilev and Milkova, 2022). Bayesian logistic regression is particularly important in this context due to its ability to incorporate prior knowledge, update such prior knowledge with new data, and provide probabilistic predictions. This approach enables us to examine how changes in quantitative variables and different categorical factors affect inventory lev-

els, providing a comprehensive understanding of the interplay between these variables and the associated uncertainties. Using Bayesian methods helps us uncover detailed connections among factors influencing inventory and make strong decisions that consider uncertainties, effectively optimizing stock levels.

Once we identify the relationships between inventory factors using regression analysis, it is important to optimize the process in order to find the most influential factors. In Experimental design, Bayesian optimization systematically explores these relationships to refine inventory strategies for maximum efficiency and resource use. The Bayesian framework improves this process by handling model uncertainty and assessing how informative factors are (Thilan et al., 2021, 2022, 2023). While regression techniques fit models to data, utility evaluation in Bayesian optimization learns from data and applies it to future scenarios. This approach aids in making informed decisions by predicting how the model might perform with unseen data. By evaluating utility, we ensure that strategies are data-driven and adaptable, which helps us prepare for future trends. This comprehensive preparation leads to more robust decision-making and better alignment with evolving conditions.

Constructing Bayesian models for complex nonlinear systems presents computational challenges, particularly in numerically optimizing expected utility functions across high-dimensional design spaces (Overstall and Woods, 2017). In Bayesian experimental design, utility functions typically rely on posterior distributions. A widely adopted criterion for Bayesian design is the Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951), which quantifies the difference between two probability distributions. Laplace approximations have emerged as efficient methods for approximating posterior distributions in Bayesian design, providing fast alternatives (Ryan et al., 2015, 2016). The Bayesian optimization approach is instrumental in identifying factors that optimize inventory management strategies, leveraging these advanced computational techniques to navigate complex decision spaces effectively.

This paper presents a new approach to find and improve key factors in the inventory management using past sales data. Using Bayesian regression analysis, the study uncovers connections between inventory factors and uses Bayesian optimization to identify the most important drivers. Unlike traditional methods, Bayesian optimization includes utility evaluation to capture detailed insights from the data and predict future situations. This method identifies key inventory factors and sets up a framework for continuous enhancements in inventory management. It helps companies adjust inventory levels effectively using real-world data, improving efficiency and strategic decisions in fast-changing business settings.

## 2 Methodology

### 2.1 Data

In this study, we analyzed historical sales data from a company. The data was sourced from "Kaggle," an online platform managed by the data science community under Google LLC (Jamess, 2007). The dataset includes variables like *Sales*, *Marketing type*, *Register price*, *Item count*, *User price*, and *Net price*, chosen based on insights from existing literature. It covers historical sales records from the past six months and current active inventory data, giving a comprehensive view of the company's sales trends and inventory management practices over time.

Before analyzing the data, preprocessingsteps were taken to improve the quality of the data. This involved removing outliers and duplicate entries from the dataset.

Table 1: Types of variables.

| Variable | Type | Description |
|---|---|---|
| *Sales* | Categorical | Item Sold or not<br>Sold – 1<br>Not – 0 |
| *Marketing type* | Categorical | Types of marketing to promote item Direct marketing – D; Search engine marketing - S |
| *Register price* | Numerical | Initial marketing price |
| *Item count* | Numerical | Number of items for sale in the stock |
| *User price* | Numerical | Price at which a product is sold to the end customer from a market |
| *Net price* | Numerical | Price of add taxes and subtract discounts |

Before proceeding with the analysis, categorical data were converted to numerical form using encoding techniques. Specifically, the "*Marketing type*" variable, a categorical factor with two categories, "D" and "S" was encoded by substituting each category with numerical values. As recommended by Baruah (2023), this encoding process was essential for extracting meaningful insights from the categorical variables.

## 2.2 Design framework

The design framework considered in this work consists of three main stages. The first stage involves Quantifying prior information by fitting a statistical model to historical sales data and obtaining the model's posterior distribution. This posterior distribution serves as prior information for the second stage, which leverages this information to Assessing design through simulated potential future data generated based on the formulated prior. The third stage outlines the Optimization and Evaluation, focusing on identifying and leveraging critical factors that influence inventory management (Thilan et al., 2021).

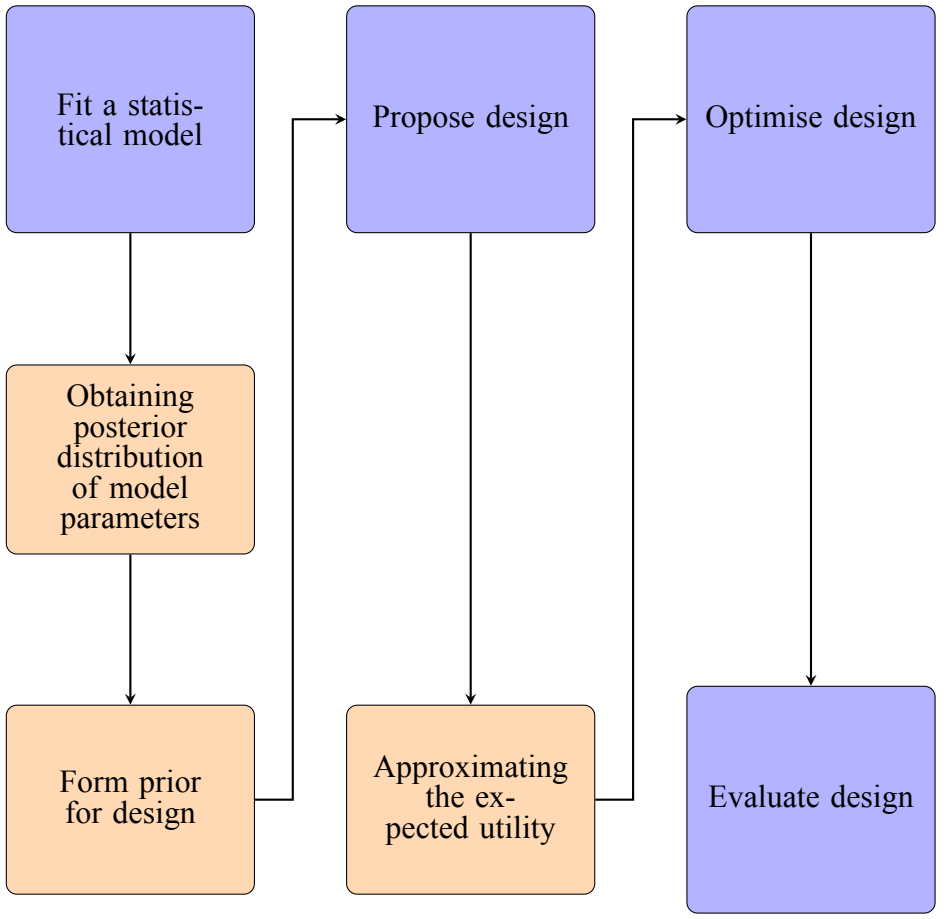


Fig. 1: Diagram of the proposed Bayesian design framework. This consists of three stages: Quantifying prior information (left), Assessing design (middle), and Optimization and Evaluation (right).

### 2.2.1 Quantifying prior information

#### 2.2.1.1 Fit a statistical model

In data modeling, selecting a statistical model involves predicting observed values and defining the relationship between a response variable and explanatory variables. In inventory management, regression analysis reveals relationships within historical data. Specifically, Bayesian logistic regression was applied to historical sales data due to its binary response variable. This method is appropriate for discrete outcomes, calculating probabilities for events where the variable can only be 0 or 1, with probability values ranging between 0 and 1. The resulting sigmoid curve illustrates the relationship between the variables effectively.

The logistic regression model to describe the relationship between the dependent variable (inventory status) and the available covariates can be expressed as follows in general:

$$log\left(\frac{P\left(y=1\right)}{1-P\left(y=1\right)}\right)=\beta_0+\beta_1x_1+\ldots+\beta_nx_n, \tag{1}$$

where $P\left(y=1\right)$ is the probability of the $y$ being equal to 1, $y$ represent sales (Depict whether the product is sold or not), $\beta_0$ is intercept, , $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients of the explanatory variables, $x_1, x_2, \ldots, x_n$ are the explanatory variables.

#### 2.2.1.2. Obtaining posterior distribution of model parameters

Bayesian methods were chosen for fitting the model due to their ability to incorporate prior knowledge and uncertainty into the estimation process. This approach allows for more flexible modeling and robust inference, particularly useful in situations where data may be limited or noisy. Within a Bayesian framework, joint posterior distribution $p\left(\boldsymbol{\beta}|\mathbf{y_0}, \mathbf{d_0}\right)$ was estimated by adopting a vague prior such that all inferences are essentially data driven; see Table 2 for the adopted notations. In this study, model parameters $\boldsymbol{\beta}$ represent the regression coefficients in the Equation (1). The $\mathbf{d_0}$ represents initial design and $\mathbf{y_0}$ represents historical sales data.

#### 2.2.1.3. Form prior for design

In the Bayesian design framework, designs are evaluated based on simulated future data scenarios. Future data for next stage was generated by using posterior distribution of historical data. In the subsequent section of the framework, the posterior distribution derived from the fitted statistical model mentioned earlier serves as the prior information or distribution.

Table 2: Definitions of distribution functions.

| Key component | Distribution | Notation |
|---|---|---|
| Quantifying prior information | Prior | $p(\boldsymbol{\beta})$ |
| | Likelihood function | $p(\mathbf{y_0}\|\boldsymbol{\beta}, \mathbf{d_0})$ |
| | Posterior | $p(\boldsymbol{\beta}\|\mathbf{y_0}, \mathbf{d_0})$ |
| Propose design | Prior | $p(\boldsymbol{\beta}\|\mathbf{y_0}, \mathbf{d_0})$ |
| | Likelihood function | $p(\mathbf{y}\|\mathbf{d}, \boldsymbol{\beta}, \mathbf{y_0}, \mathbf{d_0})$ |
| | Marginal likelihood | $p(\mathbf{y}\|\mathbf{d}, \mathbf{d_0}, \mathbf{y_0})$ |
| | Posterior of model parameters | $p(\boldsymbol{\beta}\|\mathbf{d}, \mathbf{y}, \mathbf{y_0}, \mathbf{d_0})$ |

### 2.2.2 Assessing designs

The primary objective of this study is to identify the main drivers of inventory management. An optimal design is then selected to achieve this objective. Here, we define a design as $\mathbf{d} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)^t$, where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are the initial random selections of values for the covariates in the selected model. Random selection ensures fairness by starting the optimization with three initial random designs. This approach prevents biases and reduces the risk of converging to a local maximum, ensuring a more comprehensive evaluation of potential solutions.

Kullback-Leibler divergence (KLD) utility function was selected as utility function for Bayesian design criterion and parameter estimation. KLD utility assesses the divergence between two distributions (Kullback and Leibler, 1951; Friel and Pettitt, 2008). KLD is rooted in information theory, providing a quantifiable measure of how much information is lost when approximating one distribution by another. In what follows, using KLD to identify the optimal design is a powerful approach. In this context, KLD helps select designs that maximize the expected information gain about the model parameters. The prior and posterior distributions can be compared using KLD to quantify the difference in probability distributions (Friel and Pettitt, 2008). It is an information based utility, which is given by:

$$U(\mathbf{d}, \mathbf{y}|\mathbf{y_0}, \mathbf{d_0}) = \int_{\boldsymbol{\beta}} p(\boldsymbol{\beta}|\mathbf{d}, \mathbf{y}, \mathbf{y_0}, \mathbf{d_0}) \log p(\mathbf{y}|\mathbf{d}, \boldsymbol{\beta}, \mathbf{y_0}, \mathbf{d_0})\ d\boldsymbol{\beta} - \log p(\mathbf{y}|\mathbf{d}, \mathbf{d_0}, \mathbf{y_0}). \quad (2)$$

This utility function $U(\mathbf{d}, \mathbf{y}|\mathbf{y_0}, \mathbf{d_0})$ measures the worth of choosing $\mathbf{d}$ that yields data $\mathbf{y}$, where $\mathbf{y_0}$ and $\mathbf{d_0}$ represent historical data and the initial design, respectively.

In the utility function defined in equation 2, $\mathbf{y}$ represents simulated future data. Since the actual observed data is unknown, the utility function is evaluated as an expectation. The expected utility function is defined as follows:

$$E\left(U\left(\mathbf{d}, \mathbf{y}|\mathbf{y_0}, \mathbf{d_0}\right)\right) = \int_{\mathbf{y}} U\left(\mathbf{d}, \mathbf{y}|\mathbf{d}, \mathbf{y}\right) p\left(\mathbf{y}|\mathbf{d}, \mathbf{d_0}, \mathbf{y_0}\right)\, d\mathbf{y}. \tag{3}$$

To find the optimal design, we seek to determine the expected utility. Evaluating the expectation of the utility function directly can be challenging. Therefore, a numerical method, Monte Carlo integration, was employed to estimate the expected utility, defined as follows:

$$E\left(U\left(\mathbf{d}, \mathbf{y}|\mathbf{y_0}, \mathbf{d_0}\right)\right) \approx \frac{1}{B}\sum_{b=1}^{B} U\left(\mathbf{d}, \mathbf{y}^{(b)}|\mathbf{y_0}, \mathbf{d_0}\right), \tag{4}$$

where $B(\geq 100)$ is the controlling parameter for Monte Carlo integration. The notations for the prior distribution, posterior distribution, and likelihood function used in the design framework are detailed in Table 2. During the Quantifying prior information stage, we derive the posterior distribution using a vague prior and the likelihood based on historical sales data. This posterior distribution then serves as the prior distribution in the subsequent design stage. Here, the posterior distribution of model parameters is obtained using the likelihood based on simulated future data and the prior distribution formulated in the previous stage.

The algorithm provides an approximate method for evaluating expected utility. It begins by initializing the parameters and a design (line 1). The process then enters a loop where parameters are simulated from the prior distribution (line 3). Binomial data representing sales are simulated using a specified model equation (line 4). For future covariate data, continuous covariates are simulated by fitting a Log-normal distribution to historical data, while binary covariates are generated using a binomial distribution that reflects the proportions from historical data by following the procedure discussed in Thilan et al. (2021). The posterior distribution is estimated via Laplace approximation based on the simulated data (line 5), a computationally intensive step that requires multiple executions. Kullback-Leibler divergence (KLD) utility values are then evaluated using the simulated values and the estimated posterior distribution (line 6), and these KLD values are stored (line 7). Finally, the

**Algorithm 1** Approximate the expected utility

1. Initialise $\mathbf{d} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_5)^t$, $\mathbf{d_0}, \mathbf{y_0}, \boldsymbol{\beta}$
2. **for** b=1 to $B$ **do**
3. Simulate $\boldsymbol{\beta}^{(b)} \sim p\left(\boldsymbol{\beta}|\mathbf{y_0}, \mathbf{d_0}\right)$
4. Simulate $\mathbf{y}^{(b)} \sim p\left(\mathbf{y}|\boldsymbol{\beta}^{(b)}, \mathbf{y_0}, \mathbf{d_0}, \mathbf{d}\right)$
5. Estimate $p\left(\boldsymbol{\beta}|\mathbf{y}, \mathbf{d}, \mathbf{y_0}, \mathbf{d_0}\right)$ via Laplace approximation
6. Evaluate KLD utility $U\left(\mathbf{d}, \mathbf{y}^{(b)}|\mathbf{y_0}, \mathbf{d_0}\right)$
7. store $U\left(\mathbf{d}, \mathbf{y}^{(b)}|\mathbf{y_0}, \mathbf{d_0}\right)$
8. **end for**
9. output $E\left(U\left(\mathbf{d}, \mathbf{y}|\mathbf{y_0}, \mathbf{d_0}\right)\right) \approx \frac{1}{B}\sum_{b=1}^{B} U\left(\mathbf{d}, \mathbf{y}^{(b)}|\mathbf{y_0}, \mathbf{d_0}\right)$

average KLD utility values are used as an approximate measure of expected utility (line 9).

### 2.2.3 Optimization and Evaluation

This section explains the method used to select the optimal design and compare it with other optimized designs. In this process, each variable was omitted one at a time to evaluate its impact on the overall design.

**2.2.3.1. Optimize design**
An integral aspect of this study is optimizing design selection. The optimal design (denoted as $\mathbf{d}^*$) is identified as the one that maximizes the expected utility computed in the previous section, as illustrated below:

$$\mathbf{d}^* = \arg\max_{\mathbf{d}} \quad E\left(U\left(\mathbf{d}, \mathbf{y}|\mathbf{y_0}, \mathbf{d_0}\right)\right). \tag{5}$$

When the expected utility reaches its peak, the optimal design is chosen, ensuring the most efficient and effective outcomes. Selection of the optimal design was based on boxplots depicting expected utility evaluations across various designs; the boxplot exhibiting the highest median was selected as the optimal design.

**2.2.3.2. Evaluate design**
Variables were omitted one at a time to identify optimal designs and compare them with the design obtained from the full model. This process was used to assess the contribution of each variable to the overall design. For each omitted variable, the procedure was repeated to generate a corresponding optimal de-

sign. These designs were then compared with the optimal design based on the full model to evaluate the impact of each omission. Through this comparison, the most informative variables were identified.

## 3 Results

### 3.1 Quantifying prior information

The full model was constructed by incorporating all the available covariates, as detailed below.

$$\begin{aligned} log\left(\frac{P\left(y=1\right)}{1-P\left(y=1\right)}\right) =& 0.8973 - 3.2518 \times \textit{Marketing type} \\ & - 0.0012 \times \textit{Register price} + 0.0218 \times \textit{Item count} \\ & + 0.0059 \times \textit{User price} + 0.0005 \times \textit{Net price}. \end{aligned}$$

The adequacy of the fitted model was assessed using McFadden's pseudo $R^2$ (Smith and McKenna, 2013), yielding a value of 0.28. McFadden's pseudo $R^2$ values between 0.2 and 0.4 are generally considered indicative of a reasonable good model fit (Allison, 2013).

Fitted logistic regression model provides valuable insights into the factors influencing the likelihood of a product being sold. In this analysis, the dependent variable represents whether a product was sold or not, with coefficients indicating the strength and direction of these influences. *Marketing type* emerges as a significant predictor, evidenced by its substantial negative coefficient of -3.2518. This suggests that different marketing strategies have a pronounced effect on the probability of sales. Conversely, *Register price* and *Net price* have coefficients close to zero, showing they have little direct impact on the odds of a sale. On the other hand, *Item count* and *User price* have positive coefficients (0.0218 and 0.0059, respectively), indicating that more items and higher user prices are linked to increased odds of sales. These findings highlight the crucial role of marketing strategy in driving sales, while also showing the subtle effects of pricing and product availability on sales probability.

### 3.2 Optimal design selection

In the Bayesian optimal experimental design, utility evaluation translates relationships between predictors and outcomes into practical insights. While coefficients show the direction and strength of these relationships, utility evaluation assesses how changes in predictors affect the actual probability of outcomes, like product sales, giving a clearer picture of their real-world impact.

In this study, we aimed to identify the optimal design for a specific scenario by evaluating designs, which started with three initial options. Figure

2 illustrates the outcomes of the design selection process for the full model, while Table 3 summarizes the utility evaluation results. According to Table 3, Design 2 is the optimal design choice, with the highest mean value of 31.3553 and the standard deviation of 0.0052. The individual values for these designs are presented in Appendix B, along with a comprehensive description of their similarities, differences, and the implications of overlap in the box plot shown in Figure 2.

Table 3: Summary of utility values for three initial random designs

| Designs | Mean | Standard deviation |
|---|---|---|
| 1 | 31.3511 | 0.0042 |
| 2 | 31.3553 | 0.0052 |
| 3 | 31.3516 | 0.0039 |

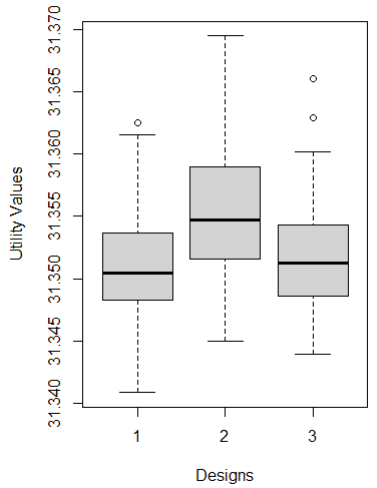


Fig. 2: Utility evaluation results for three initial random designs of the full model.

Next, we will sequentially omit variables one at a time to find the optimal design and compare it with the full model based designs. This process will help identify the most important variables influencing the outcomes.

### 3.2.1 Optimal design after omitting the *Marketing type* factor

First, we excluded the *Marketing type* factor, which has two categories: D (direct marketing) and S (search engine marketing). Figure 3 displays the utility evaluation results, while Table 4 summarizes the utility values for three designs without considering the *Marketing type*. According to Table 4, design 1 emerges as the optimal design, with the highest mean utility value of 22.0101, despite having a slightly higher standard deviation of 0.0060 compared to the other designs. The higher mean suggests that design 1 performs better overall. Additionally, Figure 3 confirms this by showing that design 1 has the highest median utility value. The standard deviations across all designs are quite small, indicating that they all perform similarly, with very little variability in the results.

Table 4: Summary of utility values for three initial random designs after *Marketing type* omitted.

| Designs | Mean | Standard deviation |
|---|---|---|
| 1 | 22.0101 | 0.0060 |
| 2 | 21.9954 | 0.0020 |
| 3 | 22.0073 | 0.0054 |

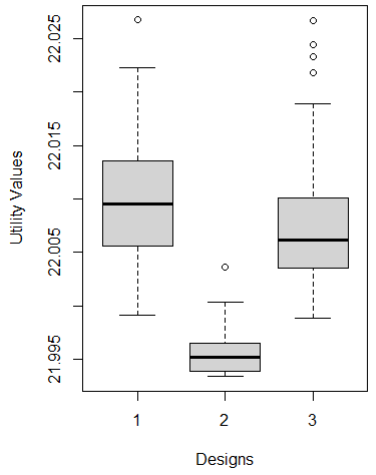


Fig. 3: Utility evaluation results for three initial random designs after omitting the *Marketing type*.

The significant drop in utility value observed when omitting the *Marketing type* variable in logistic regression highlights its crucial role in predicting outcomes, especially in determining whether a product is sold. The initially high utility value likely came from the variable's large coefficient, indicating a strong impact on sales probability based on different marketing strategies. This decrease from 31.3553 in the full model to 22.0101 without the *Marketing type* variable indicates a significant loss in predictive accuracy. These results emphasize the importance of including *Marketing type* in logistic regression models for effective decision-making and strategy formulation. It underscores the necessity of considering all relevant factors to maximize utility and enhance the reliability of logistic regression analyses in optimizing business outcomes.

### 3.2.2 Optimal design after omitting the *Register price* factor

Next, we excluded the *Register price* factor from our model. Table 5 shows that design 3 is the optimal design choice, with the highest mean utility value of 31.3762, supported by Figure 4 showing its highest median utility. The slight rise in expected utility compared to the full model could be attributed to Monte Carlo error, which introduces variability in the results. This does not impact the overall performance without the *Register price* factor and does not result in a loss of predictive accuracy.

Table 5: Summary of utility values for three initial random designs after *Register price* omitted.

| Designs | Mean | Standard deviation |
|---|---|---|
| 1 | 31.3504 | 0.0024 |
| 2 | 31.3538 | 0.0045 |
| 3 | 31.3763 | 0.0067 |

Similarly, we systematically drop one variable at a time and evaluate the utility for finding the optimal design in each scenario (see Appendix C for these comparisons). Finally, we compare the optimal designs obtained under each scenario with the optimal design based on the full model. Figures 5 illustrate the comparison between the full model and models with one factor omitted at a time. It suggests that omitting the *Marketing type* factor leads to a significant loss of information compared to both the full model and other models. This underscores the crucial role of the *Marketing type* variable in optimizing predictive accuracy and enhancing decision-making in inventory management using the logistic regression model in this study.

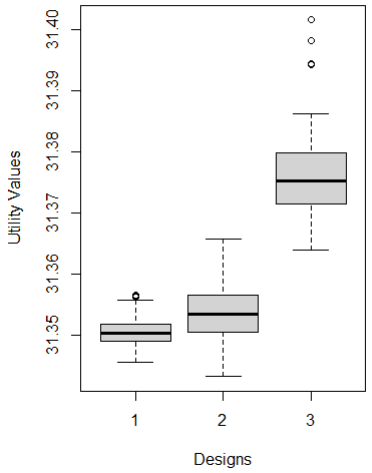


Fig. 4: Utility evaluation results for three initial random designs after omitting the *Register price*.

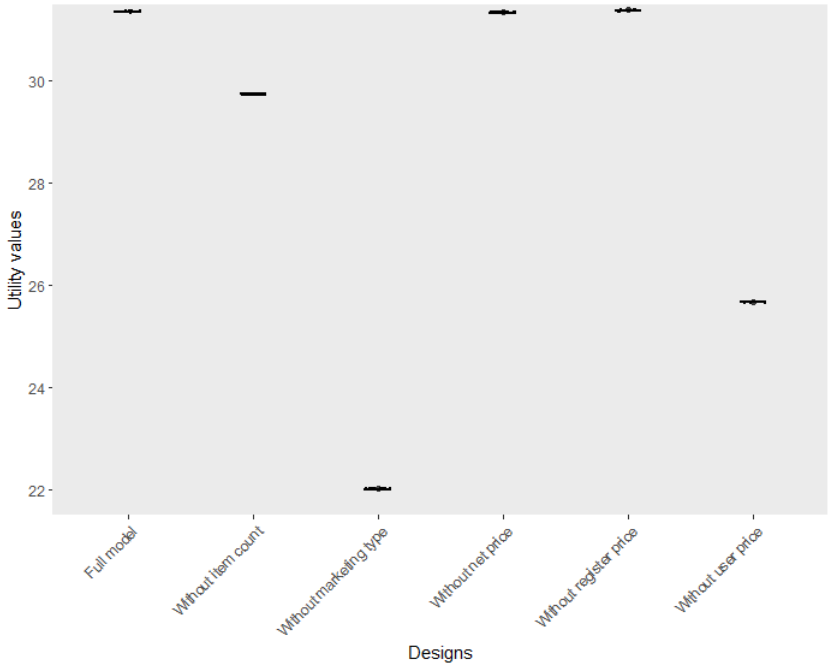


Fig. 5: The utility comparison between the full model and models that omit one factor at a time.

## 4 Discussion

The findings of this study show the advantages of using Bayesian regression analysis and optimization techniques to identify the main factors affecting inventory management. By employing a Bayesian framework, we could estimate the parameters of our regression model while incorporating prior knowledge and uncertainty into the analysis. Thus, this type of framework will provide more robust and reliable insights especially when limited data or complex relationships between variables.

One key advantage we observed was the ability to identify and prioritize the main factors affecting inventory management. Using Bayesian regression analysis, we could measure how strongly different factors influenced inventory performance. This allowed us to identifying the most important drivers, such as *Marketing type*, giving valuable insights for decision-makers to optimize inventory policies and operations. Moreover, applying these methods in inventory management will assist organizations in adjusting to market changes and achieving optimal outcomes in terms of cost-efficiency and customer satisfaction.

Using Bayesian regression analysis and optimization techniques provides several practical benefits for inventory management. Firstly, it allows for more accurate demand forecasting and inventory planning by considering uncertainty and variability in demand patterns. This helps reduce the risk of stockouts or overstocking, leading to better inventory turnover and lower holding costs. Additionally, it supports better decision-making by offering actionable insights into the key factors affecting inventory performance and the best strategies for improvement.

Future studies should continue exploring how Bayesian regression analysis and optimization techniques can be applied across various industries and scenarios to enhance inventory management practices. By refining these methods and discovering new applications for them, researchers can enhance inventory management and strengthen supply chains. Our study demonstrates that Bayesian methods could transform inventory management practices and greatly improve organizational performance.

## 5 Conclusions and Recommendations

In conclusion, this study shows the effectiveness of using Bayesian regression analysis and optimization techniques to identify key factors in inventory management. By using a Bayesian framework, we could incorporate prior knowledge and uncertainty, leading to more robust and reliable insights. This

approach helped us understand the strength and direction of relationships between factors like marketing type and inventory performance, providing valuable information for optimizing inventory policies.

The use of Bayesian optimization algorithms allowed us to enhance inventory management strategies, reducing costs and maximizing service levels. These methods offer practical benefits, such as more accurate demand forecasting, improved inventory turnover, and lower holding costs, which traditional methods may not capture. Our findings highlight the importance of adopting Bayesian approaches to achieve better results and improve supply chain efficiency and organizational performance.

Businesses can apply this methodology by combining utility evaluation with Bayesian regression analysis to gain practical insights into how predictors impact outcomes, beyond traditional coefficient interpretation. This helps identify and prioritize key drivers, like marketing strategies, to optimize operations. Using Bayesian optimization algorithms, businesses can refine strategies to minimize costs and maximize metrics like service levels and customer satisfaction. Additionally, utility evaluation for demand forecasting and inventory planning reduces the risks of stock-outs and overstocking. Continuous monitoring and adjustments based on utility evaluation results, along with investing in staff training and analytical tools, will enhance data-driven decision-making and improve overall business performance.

The R code used for the analyses in this study can be accessed via the following GitHub repository:
`https://github.com/HimayaKuruppu/Inventory_Management.git`

**References**

Agresti A. (1992) Analysis of Ordinal Paired Comparison Data, *Journal of the Royal Statistical Society*, 41(2):287-297.

Allison, P., 2013. What's the best R-squared for logistic regression. *Statistical Horizons*, 13: 126-129.

Baker, P., 2007 An exploratory framework of the role of inventory and warehousing in international supply chains. *The International Journal of Logistics Management*, 18(1): 64-80.

Baruah,I. D,2023 All you need to know about encoding techniques `https://medium.com/anolytics/all-you-need-to-know-about-encoding-techniques-b3a0af68338b` (Accessed: 2023)

Culkova, K., 2020 Using of optimizing methods in inventory management of the company. *Acta logistica*, 7(1): 9-16.

Demeter, K. and Golini, R., 2014 Inventory configurations and drivers: An international study of assembling industries. *International Journal of Production Economics*, 157: 62-73.

Fahrmeir L. and Tutz G. (1994) *Multivariate Statistical Modeling Based on Generalized Linear Models*, New York: Springer-Verlag. DOI: 10.1007/978-1-4899-0010-4

Friel, N. and Pettitt, A. N., 2008 Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(3): 589-607.

Hoppenheit, S. and Günthner, W. A., 2015 Identifying Main Drivers on Inventory using Regression Analysis. In Operational Excellence in Logistics and Supply Chains: Optimization Methods, Data-driven Approaches and Security Insights. Proceedings of the Hamburg International Conference of Logistics (HICL), Vol. 22 (pp. 141-169). Berlin: epubli GmbH.

Jamess 2007 Historical sales and active inventory, Available at: `https://www.kaggle.com/datasets/flenderson/sales-analysis`(Accessed: 2007).

Kannisto, V., Lautisten, J., Thatcher, A. R., Vaupel, J. W. (1994) 'Reductions in mortality at advanced ages: Several decades of evidence from 27 countries', *Population and Development Review*, 20:793–809.

Kullback, S. and Leibler, R. A., 1951 On information and sufficiency. *The annals of mathematical statistics*, 22(1): 79-86.

Munyaka J. and Yadavalli V. S. S. (2022) Inventory management concepts and implementations: a systematic review, *South African Journal of Industrial Engineering*, 33(2): 15-36.

Overstall, A.M. and Woods, D. C., 2017 Bayesian design of experiments using approximate coordinate exchange. *Technometrics*, 59(4): 458-470.

Ryan, E. G., Drovandi, C. C. and Pettitt, A. N., 2015 Fully Bayesian experimental design for pharmacokinetic studies. *Entropy*, 17(3): 1063-1089.

Ryan, E. G., Drovandi, C. C., McGree, J. M. and Pettitt, A. N., 2016 Fully Bayesian optimal experimental design: A review. *International Statistical Review*, 84: 128-154.

Smith, T. J. and McKenna, C. M., 2013 A comparison of logistic regression pseudo R2 indices. *Multiple Linear Regression Viewpoints*, 39(2): 17-26.

Thilan, A. W. L. P., Fisher, R., Thompson, H., Menendez, P., Gilmour, J. and McGree, J.M., 2022 Adaptive monitoring of coral health at Scott Reef where data exhibit nonlinear and disturbed trends over time. *Ecology and Evolution*, 12(9):e9233.

Thilan, A. W. L. P, Menéndez, P. and McGree, J. M., 2023 Assessing the ability of adaptive designs to capture trends in hard coral cover. *Environmetrics*, 34(6): e2802.

Thilan, A. W. L. P., Peterson, E., Menéndez, P., Caley, J., Drovandi, C., Mellin, C. and McGree, J., 2024 Bayesian design methods for improving the effectiveness of ecosystem monitoring. *Environmental and Ecological Statistics*, 31(4): 893-919.

Vasilev, J. and Milkova, T., 2022 Optimisation Models for Inventory Management with Limited Number of Stock Items. *Logistics*, 6(3): 54.

## Appendix A

The basic statistics of the selected factors are shown in Table A1.

Table 6: Table A1: Descriptive statistics of numerical variable.

| Variable | Minimum | Median | Mean | Skewness | Maximum | Sd |
|---|---|---|---|---|---|---|
| Register price | 2.99 | 89.96 | 106.04 | 1.291 | 471.23 | 70.073 |
| Item count | 1.00 | 30.00 | 37.33 | 1.651 | 152.00 | 24.198 |
| User price | 0.00 | 53.94 | 64.38 | 1.234 | 402.13 | 47.226 |
| Net price | 3.96 | 39.83 | 51.19 | 2.851 | 602.98 | 46.449 |

Table 7: Table A2: Descriptive statistics of categorical variable.

| Marketing type | Frequency | Proportions |
|---|---|---|
| Sold | 4190 | 0.4190 |
| Not | 5809 | 0.5810 |

## Appendix B: Optimal designs

### B.1 Optimal designs for the full model

In this appendix, we explore the optimal/sub-optimal designs for each random design choice considered under each scenario adopted in this study. The optimal/sub-optimal designs (hereafter referred to as optimal designs) for the full model are given below.

>Optimal design 1

| | [,1] | [,2] | [,3] | [,4] | [,5] |
|---|---|---|---|---|---|
| [1,] | 1 | 81.22850 | 107 | 112.13186 | 31.88998 |
| [2,] | 2 | 39.80136 | 119 | 45.86405 | 26.88234 |
| [3,] | 2 | 39.80136 | 119 | 45.86405 | 26.88234 |
| [4,] | 2 | 39.80136 | 119 | 45.86405 | 26.88234 |
| [5,] | 1 | 81.22850 | 107 | 112.13186 | 31.88998 |
| [6,] | 2 | 39.80136 | 119 | 45.86405 | 26.88234 |

>Optimal design 2

| | [,1] | [,2] | [,3] | [,4] | [,5] |
|---|---|---|---|---|---|
| [1,] | 1 | 132.02817 | 40 | 78.14091 | 86.09848 |
| [2,] | 2 | 57.33859 | 137 | 108.73217 | 23.89565 |
| [3,] | 2 | 46.82812 | 130 | 17.63174 | 243.43249 |
| [4,] | 2 | 46.82812 | 130 | 17.63174 | 243.43249 |
| [5,] | 2 | 46.82812 | 130 | 17.63174 | 243.43249 |
| [6,] | 2 | 57.33859 | 137 | 108.73217 | 23.89565 |

>Optimal design 3

| | [,1] | [,2] | [,3] | [,4] | [,5] |
|---|---|---|---|---|---|
| [1,] | 1 | 89.74444 | 60 | 14.17698 | 56.56094 |
| [2,] | 1 | 85.02470 | 101 | 12.96414 | 247.40324 |
| [3,] | 1 | 110.67050 | 81 | 112.32432 | 260.12944 |
| [4,] | 2 | 88.46071 | 119 | 67.66744 | 29.75634 |
| [5,] | 2 | 88.46071 | 119 | 67.66744 | 29.75634 |
| [6,] | 2 | 88.46071 | 119 | 67.66744 | 29.75634 |

Optimal design 1 displays repeated configurations, notably in rows 1 and 5, where similar covariate settings are consistently selected. This repetition suggests that the overlap in the boxplots may stem from these configurations yielding comparable utility values. In contrast, optimal design 2 introduces distinct covariates, which differ significantly from that in design 1. Optimal design 3 offers another unique set of covariates, providing further variation. While some covariates may overlap with previous designs, the distinct values of each optimal design suggest differing utilities that merit further exploration. In conclusion, since the designs appear to be distinct, the overlapping in boxplots likely results from Monte Carlo errors associated with parameters such as $B$, $N1$, and $Q$.

### B.2 Optimal designs for the *Marketing type* factor

In the scenario without the *Marketing type* factor, optimal design 1 exhibits repeated configurations, indicating that different design setups can yield similar utility values, which may explain the observed overlap in the boxplots. In contrast, optimal design 2 incorporates distinct covariates, setting it significantly apart from design 1. Meanwhile, optimal design 3 presents another unique set of covariates, further contributing to the variability among the designs.

A similar reasoning applies to the remaining four models. Each design features its own distinct set of covariates, though some overlap in covariate settings persists across the designs, potentially contributing to the boxplot overlap. The variations among these designs, as seen in optimal designs 1, 2, and 3, highlight the influence of specific covariates on utility values. The patterns observed in the boxplots for the remaining designs likely arise from the overlapping covariate settings and their effects on utility values, similar to the factors discussed in both the full model and the scenario without the marketing type factor.

>Optimal design 1

| | [,1] | [,2] | [,3] | [,4] |
|---|---|---|---|---|
| [1,] | 32.10774 | 33 | 12.09312 | 339.89036 |
| [2,] | 55.68503 | 7 | 16.10191 | 52.77231 |
| [3,] | 32.10774 | 33 | 12.09312 | 339.89036 |
| [4,] | 32.10774 | 33 | 12.09312 | 339.89036 |
| [5,] | 25.02907 | 142 | 66.28845 | 258.98996 |
| [6,] | 32.10774 | 33 | 12.09312 | 339.89036 |

>Optimal design 2

| | [,1] | [,2] | [,3] | [,4] |
|---|---|---|---|---|
| [1,] | 84.34505 | 112 | 31.97465 | 31.49082 |
| [2,] | 117.19832 | 79 | 84.98876 | 306.75668 |
| [3,] | 117.19832 | 79 | 84.98876 | 306.75668 |
| [4,] | 84.34505 | 112 | 31.97465 | 31.49082 |
| [5,] | 49.62425 | 113 | 100.46166 | 362.71249 |
| [6,] | 49.62425 | 113 | 100.46166 | 362.71249 |

>Optimal design 3

| | [,1] | [,2] | [,3] | [,4] |
|---|---|---|---|---|
| [1,] | 66.94145 | 18 | 13.42770 | 300.1757 |
| [2,] | 130.26993 | 99 | 4.98874 | 298.5396 |
| [3,] | 130.26993 | 99 | 4.98874 | 298.5396 |
| [4,] | 130.26993 | 99 | 4.98874 | 298.5396 |
| [5,] | 66.94145 | 18 | 13.42770 | 300.1757 |
| [6,] | 66.94145 | 18 | 13.42770 | 300.1757 |

### B.3 Optimal designs for the *Register price* factor

>Optimal design 1

| | [,1] | [,2] | [,3] | [,4] |
|---|---|---|---|---|
| [1,] | 1 | 49 | 138.14055 | 3.994955 |
| [2,] | 1 | 34 | 39.43418 | 384.192751 |
| [3,] | 2 | 93 | 12.17050 | 49.061301 |
| [4,] | 1 | 34 | 39.43418 | 384.192751 |
| [5,] | 1 | 34 | 39.43418 | 384.192751 |
| [6,] | 2 | 25 | 42.44417 | 160.01861 |

>Optimal design 2

| | [,1] | [,2] | [,3] | [,4] |
|---|---|---|---|---|
| [1,] | 1 | 76 | 52.06419 | 107.15437 |
| [2,] | 2 | 126 | 92.72598 | 76.80402 |
| [3,] | 2 | 126 | 92.72598 | 76.80402 |
| [4,] | 2 | 117 | 50.38966 | 117.07664 |
| [5,] | 2 | 120 | 65.44532 | 36.38839 |
| [6,] | 2 | 120 | 65.44532 | 36.38839 |

>Optimal design 3

| | [,1] | [,2] | [,3] | [,4] |
|---|---|---|---|---|
| [1,] | 1 | 135 | 23.85439 | 354.3417 |
| [2,] | 1 | 135 | 23.85439 | 354.3417 |
| [3,] | 1 | 135 | 23.85439 | 354.3417 |
| [4,] | 1 | 135 | 23.85439 | 354.3417 |
| [5,] | 1 | 135 | 23.85439 | 354.3417 |
| [6,] | 2 | 8 | 10.69078 | 218.5830 |

## Appendix C: Optimal design after omitting the factor one at a time

### C.1 Optimal design after omitting the *Item count* factor

Table C1: Summary of utility values for three initial random designs after *Item count* omitted.

| Designs | Mean | Standard deviation |
|---|---|---|
| 1 | 29.7353 | 0.0027 |
| 2 | 29.7410 | 0.0034 |
| 3 | 29.7351 | 0.0025 |

After omitting the *Item count* factor from our full model, design 2 emerges as the optimal choice, as shown in Table C1. It has the highest mean utility value of 29.7410, with a standard deviation of 0.0034, indicating both superior performance and notable consistency. The small standard deviation suggests that the utility values are closely clustered around the mean, reinforcing the reliability of this design. Figure C1 further supports this, showing that design 2 also has the highest median utility values.

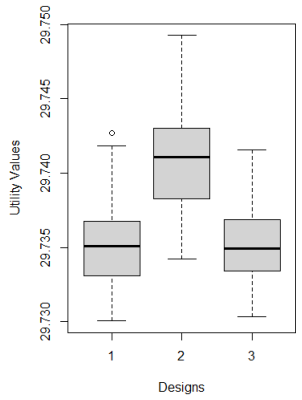


Figure C1: Utility evaluation results for three initial random designs after omitting the *Item count*.

## C.2 Optimal design after omitting the *User price* factor

After omitting the *User price* factor from our model, design 1 stands out as the optimal choice, as shown in Table C2. It has the highest mean utility value of 25.6686 and the lowest standard deviation of 0.0023, indicating strong performance and minimal variability. Figure B2 also supports this, with design 3 showing the highest median utility value. Thus, design 3 is the optimal design after removing the *User price* factor.

Table C2: Summary of utility values for three initial random designs after *user price* omitted.

| Designs | Mean | Standard deviation |
|---|---|---|
| 1 | 25.6675 | 0.0038 |
| 2 | 25.6637 | 0.0024 |
| 3 | 25.6686 | 0.0023 |

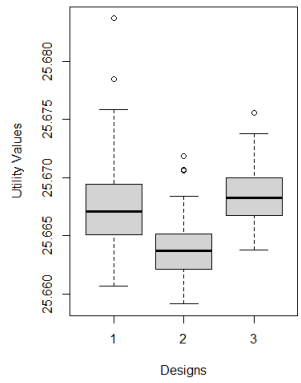


Figure C2: Utility evaluation results for three initial random designs after omitting the *User price*.

### C.3 Optimal design after omitting the *Net price* factor

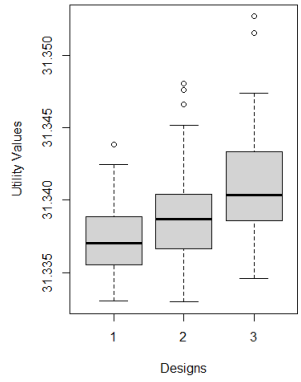


Figure C3: Utility evaluation results for three initial random designs after omitting the user price.

Table C3: Summary of utility values for three initial random designs after *Net price* omitted.

| Designs | Mean | Standard deviation |
|---|---|---|
| 1 | 31.3373 | 0.0022 |
| 2 | 31.3388 | 0.0031 |
| 3 | 31.3410 | 0.0035 |