LEARNING PRICE PROMOTION EFFECTS ON RECURRING SELL-IN PURCHASES FROM SIMULATED STORE LEVEL SALES DATA

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

INDUSTRIAL ENGINEERING

By

Pelin Keşrit

June 2021

LEARNING PRICE PROMOTION EFFECTS ON RECURRING SELL-IN PURCHASES FROM SIMULATED STORE LEVEL SALES DATA

By Pelin Keşrit

June 2021

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Savaş Dayanık(Advisor)

Semih Onur Sezer

Alper Şen

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

LEARNING PRICE PROMOTION EFFECTS ON RECURRING SELL-IN PURCHASES FROM SIMULATED STORE LEVEL SALES DATA

Pelin Keşrit
M.S. in Industrial Engineering
Advisor: Savaş Dayanık
June 2021

When a product is put on promotion to increase its sales, this causes a decrease in the sales of another product in the same product group. This phenomenon happens usually when the promoted product is a substitute for the other product. In this study, we focus on the wholesaler's revenue maximization problem over the given planning horizon. For this purpose, we constructed a Bayesian hierarchical model for the order quantities observed in the store level data for substitutable products. Order quantities are assumed to have Poisson distributions whose means depend on season, prices and previously ordered quantities for all products in the same group. The customers are assumed to have different price sensitivities, and consumption rates implicit in their historical order quantities. Using a hybrid of different Markov Chain Monte Carlo methods, we update model parameter posterior distributions and predict each retailer's order quantities in the future. We verified on simulated sales data that the MCMC methods work.

Keywords: Bayesian hierarchical models, Markov Chain Monte Carlo, Substitution, Promotion, Marketing.

ÖZET

FİYAT PROMOSYONLARININ TOPTAN SATIN ALMALAR ÜZERİNDEKİ ETKİLERİNİN SİMULE EDİLMİŞ MAĞAZA SATIŞ VERİLERİNDEN ÖĞRENİLMESİ

Pelin Keşrit
Endüstri Mühendisliği, Yüksek Lisans
Tez Danışmanı: Savaş Dayanık
Haziran 2021

Bir ürünün satışlarını artırmak için bir müşteriye promosyonunun yapılması, aynı ürün grubundaki başka bir ürünün satışlarının düşmesine neden olur. Bu olgu genellikle promosyonu yapılan ürün, diğer ürünün ikamesi olduğu zaman ortaya çıkar. Bu çalışmada, toptancının belirli bir planlama dönemi içerisindeki kâr eniyileme problemine odaklanılmıştır. Bu amaçla, eşdeğer ürünlerin mağaza düzeyindeki verilerde gözlenen sipariş miktarları için Bayesci hiyerarşik bir model kurulmuştur. Sipariş miktarlarının, ortalaması dönemsellik, fiyat ve önceki sipariş miktarına bağlı olan Poisson dağılımına sahip olduğu varsayılmıştır. Müşterilerin farklı fiyat duyarlılıklarına ve izini geçmiş sipariş miktarlarında görebileceğimiz farklı tüketim oranlarına sahip oldukları varsayıldı. Hibrit bir Markov Zinciri Monte Carlo metodu kullanarak, model parametrelerinin sonsal dağılımlarının nasıl güncellenebileceğini ve her perakendecinin gelecek sipariş miktarını nasıl tahmin edebileceğini gösterilmiştir. MZMC yöntemlerinin doğrulaması bir benzetim modelinden üretilen satış verileri üzerinde sağlanmıştır.

Anahtar sözcükler: Bayesci hiyerarşik modeller, Markov Zinciri Monte Carlo, İkame, Promosyon, Pazarlama.

Acknowledgement

First and foremost, I would like to express my gratitude for my advisor Prof. Savaş Dayanık. His mentorship, patience and wisdom guided me throughout this study. Without his guidance and support, this thesis would not be possible.

I would like to thank Assoc. Prof. Semih Onur Sezer and Assoc. Prof. Alper Şen for accepting to be in my thesis committee.

I would like to express my deepest gratitude for my parents Berrin Keşrit and Şükrü Keşrit and my sister Yasmin Keşrit. They have always been an inspiration to me. Their endless love and support have given me the strength to achieve my goals and helped me become the person I am today.

I also would like to express my appreciation for the supportive and nurturing environment that Bilkent University provided throughout my academic journey. I would like to thank all my professors, colleagues and friends for creating a community that I am so proud to be a part of.

Contents

List of Figures

Chapter 1

Introduction

From the perspective of a retailer, offering promotions to customers may seem like an effective tool to boost the sales in the short term. However, implementing such policies could result in the opposite of what was intended for such as post-promotion dips or the cannibalization effects. The factors that determine the success of a promotional activity may vary from the type of the promotion (feature, display, price reduction, etc.) to the level of discount offered or the timing and frequency of the promotions to the selection of the products that will be promoted. These factors may have different characteristics depending on the store (size, capacity, etc.), customer (behavioral, demographic, price sensitivity, etc.) and product (perishability, size, quality etc.) features. The impact of those factors on the retailer's profit may also differ with respect to the length of the planning horizon. There are many studies in the literature that focus on promotions and profit maximization and discuss the aforementioned effects.

Blattberg and Neslin [1] study the effects of promotional activities under three main categories; immediate, intermediate, and long term effects. Immediate effects span the particular week or weeks the promotion is made. Multicollinearity, asymmetric cross elasticities, brand switching, comparison of regular and promotional price cut elasticities are thought as immediate effects. Blattberg and Neslin use the term *multicollinearity* to explain the interaction between different promotional tools, such as price cuts, feature and display. They stress that those means are often used together, which makes it harder to separately estimate their effects. They also mention that the promotional price-cuts cause brand switching, which is the main source of the increase in the sales volume of the promoted product. If the price cut is temporary, it motivates the customer to stockpile the brand. This occurs when the customer perceives that he or she is getting a bargain that drives their transactional utility. A regular price reduction does not affect and alter the behavior of the customer in the same way.

On the other hand, *the intermediate effects* consider the weeks or months surrounding the promotion. Blattberg and Neslin note that purchase reinforcement, promotion usage effect, and purchase acceleration are the most commonly observed phenomena in the intermediate term. They argue that there are conflicting results in the literature on whether frequent promotions have a negative effect on the customers because they may form a habit to purchase the brand only when promoted. Promotions can also alter the interpurchase time, and these effects are captured better with household-level data compared to market-level data. Finally, *long term effects* correspond to the years after promotions were implemented. Blattberg and Neslin point out the lack of research that studies those effects and emphasize the importance of developing a consumer franchise so that a brand can have in the long loyal buyers run who will not substitute that brand with some other in its absence.

There is a number of studies on the long-term effects of the price on the demand and the pattern of competition. Blattberg and Wisniewski [2] adopt a price-tier model which considers the utility associated with price tiers and correspond to distinct retail price groups that consist of the premium, moderate and generic brands. They stress that these price tiers reflect the general opinion about the perceived quality and may differ among customers, thus the utility function includes a parameter that takes into account the consumer's willingness for quality. It is also stated that promoted brand can attract customers from the brands in its own tier and the tiers below, but the reverse is rare; hence, the cross elasticities are asymmetric. Building on this concept, they illustrate relative price distributions under uniform, normal and U-shaped assumptions. Along with variables that indicate whether a brand was advertised or promoted at a certain time period and the prices, Blattberg and Wisniewski also include seasonality parameter for the holiday periods and a delay decay variable for multi-week deals to estimate the total sales that are aggregated over several stores in the price zone.

Kumar and Leone [3] present their study as one of the first to investigate the impact of retail promotion on store substitution whilst using scanner data. They argue that there has been a shift in power from manufacturers to retailers as one's access to daily information enhanced with the installation of scanners into the stores. The manufacturer and the retailer's objectives for increasing profit seem to intersect at only one point: the brand substitution within a store. Kumar and Leone used their observations about brand substitution on the hierarchical model where they examined store substitution. This analysis provided information on consumers' store choice process and further insight on identifying which stores within a geographical cluster actually compete on the basis of retail promotions and whether the effect of a retail store's promotion can extend beyond geographical market boundaries. They pooled the data across stores, aiming to reduce the level of collinearity and provide more degrees of freedom based on the pooling tests conducted by Bass and Wittink (1975) who found out that the coefficients for each brand's model was equal across all stores in the scope. Kumar and Leone implemented the Fuller-Battese (1974) procedure and found that the effect of promotional activities on store substitution was smaller compared to within store brand substitution.

Musalem et al. [4] reviews the findings in the literature in terms of modeling and empirically validating consumer choices in response to operational decisions. They gather their study under five main categories: inventory and assortment decisions, capacity and service level decisions in service industries, dynamic pricing decisions and revenue management, supply chain coordination, balancing flexibility and tractability.

Kök and Fisher [5] focus on the product assortment planning problem under the substitution. They use data from a supermarket chain in which the products are broken down into subcategories that assume minimal difference and a high substitution rate within each. They propose an iterative optimization heuristic to maximize the expected gross profits with respect to shelf space constraints. Kök and Fisher also manage to estimate the demand under different substitution scenarios using the sales data.

Promotional activities occurring between the wholesaler and the retailer have different characteristics than those between the retailer and the end consumer. Therefore, it is important to identify the factors that affect the decision making processes of the retailers, in order to maximize the wholesaler's profits without running into issues with cannibalization effect. Our proposed algorithm contributes to the literature by capturing the effect that seasonality, price and inventory have on retailers' demand for the product itself as well as for its substitutes.

We begin by simulating store level sales data using the characteristics of the real store level data in order to capture the promotional effects. Then, by using Bayesian hierarchical modeling techniques we identify the underlying conditional probability distributions of the purchases. We learn the price promotion effects from the simulated sales data by implementing a hybrid of Monte Carlo Markov Chain methods.

We begin by introducing the problem in Section 2, then we describe the characteristics of the data and the calculations of the posterior distributions of the parameters. In Section 3, we present the data simulation method. We then proceed to the results in Section 4 and conclude by discussing the implications of the study and opportunities for further research in Section 5.

Chapter 2

Problem Definition

In this study, we focus on the wholesaler's revenue maximization problem over a given planning horizon, by determining the price of the products to be offered to the retailers. We are told the promotional activities that occurred over T periods to m retailers (later referred to as customers) regarding n products that are categorized into different hierarchical levels by similarities of their features. There are three levels of hierarchies in the raw data. These are high, medium, and low levels, which are ordered from the least to the most common features the products in those levels share, respectively. Among those hierarchical levels, the products are further classified into smaller groups. The study focuses on every group on low level separately because the between-group product cannibalization effect is assumed to be negligible compared to in-group interactions on the same level of hierarchy.

The cannibalization effect occurs when a product is promoted to a customer to increase its sales, but causes a decrease in the sales of another product in the same product group. This phenomenon happens usually when the promoted product is a substitute of the other product. The substitution rate is closely related to the number of common features that products share. The magnitude of the cannibalization effect may differ within product pairs. If the promoted product is a generic product, then the amount of decrease in the sales of a premium product can be different than that in the opposite case. Brand loyalty, perception of quality and budget may be some of the factors that determine an individual's decision making process when presented with different promotional scenarios by the retailer. However, in the wholesaler's revenue maximization problem, the products are promoted to the retailers who aim to maintain some level of product assortment in order to be able to satisfy the demands of its end-customers with different preferences. From this point of view, in this study, it is not possible to entirely rely on the factors that mainly focus on personal preferences which are commonly used in retailer's revenue optimization problem. Instead, we assume that these are implicitly included in the hidden factors that characterize the retail stores. It is also assumed that retail stores pass the promotional discounts they receive from the wholesaler to their end-customers although the retailers' timing of the promotional activity and the rate of discount it offers may not directly reflect the attributes of the wholesaler's promotional strategy.

2.1 Data

The data belongs to a wholesaler who has historical order data (time and quantity pairs for every product) of its customers. The data also consists of the product group information, customer addresses, descriptions of different types of promotional activities and their respective discount rates. The wholesaler has 647 distinct products serving to a total of 89150 customers. Each product is categorized in a hierarchical manner according to its similarities with other products. There exist 63 distinct groups in the lowest level of hierarchy.

The original data consisted of daily promotions. We aggregated them into weekly periods because most of the products were not procured by the customers on a daily basis. The original data only included the percentage changes in the original sales price during a promotional activity, but did not disclose the actual prices. Therefore, we normalized the base prices for products to one. In other words, not the price itself but the change in price affects the next purchase in our models. Taking the raw data as a basis, we simulate store level sales data that captures the desired promotional effects with similar characteristics.

The model considers the effect of price and the previous purchase quantity as the main explanatory variables to predict the next purchase quantity. In other words, the customer is assumed to react to a possible promotional activity given his/her past behavior and price changes. The effect of price is examined through self- and cross-price elasticities of the products that belong to the same product group. Effect of prices on expected purchase quantities depend on a variety of common observed factors, whose effects on purchase quantities are unobserved random variables. These factors are customer specific and include store size and location. Recall that the customers of the wholesaler are the retail stores. The previous purchase quantity is assumed to implicitly reflect the customer inventory. The effect of price and previous purchase quantity are modeled as *log Normally distributed* random variables whose mean vector and covariance matrices are also random variables. By employing a Bayesian hierarchical structure, the model aims to capture the widely different or heterogeneous purchase behavior of customers with relative small number of parameters. We will describe the model in more detail in the next subsection.

2.2 Hierarchical Bayesian Model

The hierarchies and relations between the parameters of the wholesaler's revenue maximization problem are displayed in Figure 2.1.

Observed variables have dark nodes, hidden variables have white nodes. The variables are described in Table 2.2.

As seen in Figure 2.1, at a given time t, the price p^{jt} and the quantities sold q^{jt} are observed for the periods $1, \ldots, t$. For the upcoming time periods, the prices becomes decision variables in order to influence the quantities that will be purchased by the customers.

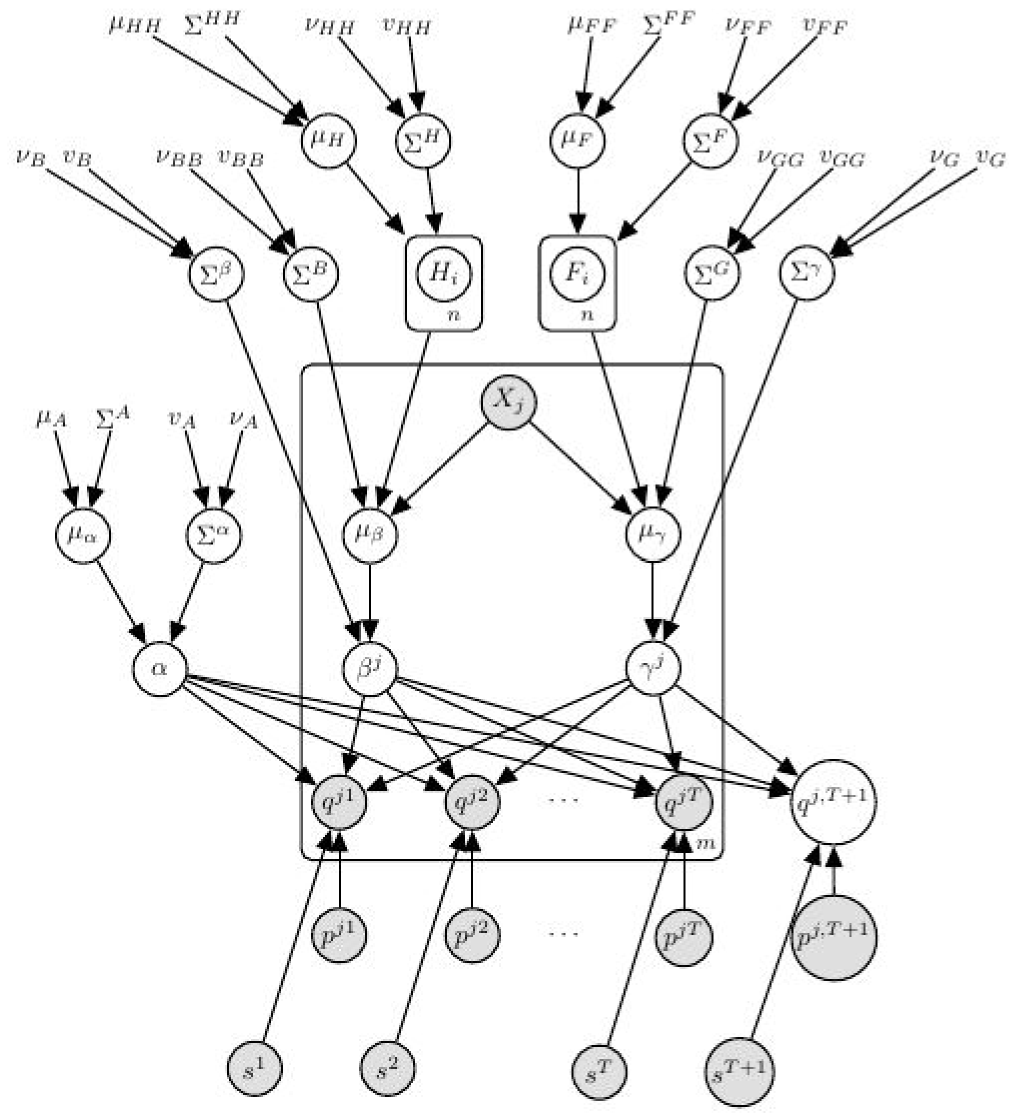

Figure 2.1: Graphical model of Bayesian hiearachical model

Notation	Description
$i = 1, \dots, n$	denote products
$j = 1, \dots, m$	denote customers (retail stores)
$t = 1, \dots, T+1$	denote periods (T as the current time period)
q^{jt}	nx1 vector of quantities ordered from n products by customer j in period t
p^{jt}	nx1 vector of prices of n products offered to customer j in period t
s^t	4x1 vector denoting the season of period t
α	4x1 vector denoting the effect of seasonality
β^{jt}	nxn matrix denoting the effect of price for customer j in period t
γ^{jt}	nxn matrix denoting the effect of previous purchase quantity for customer j in period t
X_j	fx1 vector of hidden factors for customer j
H_i	nxf matrix of unobserved random influence of the factors on the effect of price on the buying tendency for product i
F_i	nxf matrix of unobserved random influence of the factors on the forward buying tendency for product i
μ_α	4x1 vector denoting the mean parameter of α
Σ_α	4x4 matrix denoting the variance parameter of α
μ_β	nxn matrix denoting the mean parameter of β
Σ_β	n^2xn^2 matrix denoting the variance parameter of β
μ_γ	nxn matrix denoting the mean parameter of γ
Σ_γ	n^2xn^2 matrix denoting the variance parameter of γ
μ_H	nxf matrix denoting the mean parameter of i.i.d H_i's
Σ_H	nfxnf matrix denoting the variance parameter of i.i.d H_i's
μ_F	nxf matrix denoting the mean parameter of i.i.d F_i's
Σ_F	nfxnf matrix denoting the variance parameter of i.i.d F_i's
Σ_B	n^2xn^2 matrix denoting the variance parameter of μ_β
Σ_G	n^2xn^2 matrix denoting the variance parameter of μ_γ

Figure 2.2: Descriptions of the notation

We assume that quantities q_i^{jt} of product i for customer j at time t has a *Poisson distribution* with rate λ_i^{jt}, because the products are packaged goods and thus the quantity can only take integer values. We model the *Poisson* purchase rates by

$$\log(\lambda_i^{jt}) = (p^{jt})^T \beta_i^j + (q^{j,t-1})^T \gamma_i^j + (s^t)^T \alpha, \tag{2.1}$$

where α denotes the effect of seasonality (seasons are denoted by s), $\beta^{(j)}$ denotes the effect of price, and $\gamma^{(j)}$ denotes the effect of previous purchase quantity for customer j. They are all assumed to be *Normally distributed* with parameters μ_α, Σ^α ; vec $\mu_\beta^{(j)}$, Σ^β and vec $\mu_\gamma^{(j)}$, Σ^γ respectively. The parameters μ_α, μ_β, and μ_γ are also assumed to be *Normally distributed* random variables which their parameters are defined as

$$\begin{aligned} \mu_\alpha|\mu_A, \Sigma^A &\sim \text{ Normal } (\ \mu_A, \Sigma^A), \\ \text{vec } \mu_\beta^{(j)}|H, \Sigma^B &\sim \text{ Normal } (\ H_{1:\text{n}} X_j, \Sigma^B), \\ \text{vec } \mu_\gamma^{(j)}|F, \Sigma^\mu &\sim \text{ Normal } (\ F_{1:\text{n}} X_j, \Sigma^\mu), \end{aligned} \tag{2.2}$$

where x_j denotes the common hidden factors for customer j, the $n \times f$ matrix H_i denotes the unobserved random influence of the factors on the effect of price on the buying tendency for product i and the $n \times f$ matrix F_i denotes the unobserved random influence of the factors on the forward buying tendency for product i. $H_1, \ldots, H_n$ and $F_1, \ldots, F_n$ are assumed to be independent and identically distributed (i.i.d) Normal random variables with parameters vec μ_H, Σ^H and *vec* μ_F, Σ^F, respectively.

Although, the parameters $\beta^{(j)}$,$\gamma^{(j)}$ along with their means $\mu_\beta^{(j)}$ and $\mu_\gamma^{(j)}$ are customer specific, all of the variances present in the model are designed to be the same for every customer. Variance matrices $\Sigma^\alpha, \Sigma^\beta, \Sigma^B, \Sigma^\mu, \Sigma^\gamma, \Sigma^H$ and Σ^F are assumed to have *Inverted Wishart prior distribution* with parameters ν_A, υ_A ; ν_B, υ_B ; ν_{BB}, υ_{BB} ; ν_μ, υ_μ ; $\nu_\gamma, \upsilon_\gamma$; ν_{HH}, υ_{HH} ; ν_{FF}, υ_{FF}, respectively.

Hyper-parameters of the hierarchical model are chosen without taking into account any information about the data in order to ensure that these parameters do not force the algorithm to start from a particular local optimum. In other words, this approach helps the algorithm's performance and convergence to Bayes optimal parameters, remain unaffected by the choice of hyper parameters.

2.3 Calculation of the posterior distributions of parameters

The parameters of the Bayesian hierarchical model are learned by employing Markov Chain Monte Carlo (MCMC) methods that provide simulation based estimates. Once the posterior distributions of the random variables in the model are formulated, MCMC methods allow us to form a Markov chain that converges to the posterior distribution of (hyper-) parameters as the number of iterations increases, starting from a random point in the parameter space. Transition from one state to another relies on the conditional distribution of a parameter given other parameters in the model that either directly affects or are directly affected by the parameter in consideration. The conditional probability distribution of a parameter is proportional to the joint distribution, hence whilst constructing full conditional probability density function (pdf) of a parameter, the joint pdf was utilized.

The forms of full conditional pdfs of the parameters of the hierarchical model varied depending on their prior distributions and their relations with other parameters. This caused the need for combining a variety of MCMC methods in order to generate the algorithm that perfectly fits the necessities of the model without making it excessively complex. Therefore, for the parameters which the target posterior distribution has a conjugate prior, the Gibbs sampling was used, which is a special case of Metropolis-Hastings algorithm. The Gibbs algorithm starts from an initial set of parameters and at each iteration, it randomly samples a new set of parameters from a predetermined probability distribution of a known form. It updates the old set of parameters that are used to sample the new ones at the end of each iteration. The full conditional pdfs of the parameters that correspond to the mean and variances of α, β, γ, H_i, F_i together with H_i and F_i themselves are found to be of a known form: more specifically either *Normally distributed* or have *Inverted Wishart distribution.* The derivations of full conditional pdfs can be found in Appendix A. The basic structure of the Gibbs sampling that was utilized in this study is as follows:

$$
\begin{aligned}
&\text{Set } \theta_0. \\
\text{Sample from } &\theta_{11} \sim f_1(\theta_1|\theta_{02}, \cdots, \theta_{0k}), \\
&\theta_{12} \sim f_2(\theta_2|\theta_{11}, \cdots, \theta_{0k}), \\
&\vdots \\
&\theta_{1k} \sim f_k(\theta_1|\theta_{11}, \cdots, \theta_{1,k-1}). \\
&\text{Repeat.}
\end{aligned}
\tag{2.3}
$$

On the other hand, the full conditional pdfs of the parameters α, β and γ are not of known form. Since direct sampling was not possible for those cases, Gibbs sampling was not useful. Therefore, Random-Walk Metropolis algorithm was employed instead. This algorithm also generates a reversible Markov chain of random samples from a given probability distribution. Random-Walk Metropolis algorithm considers a set of parameters as a starting point and at each iteration, it generates a new set of parameters by adding randomly generated normal numbers to the original set of parameters. Then, the Metropolis algorithm computes the probability of accepting the new set of parameters by comparing the ratio of the likelihood times the density functions of the old and the new set of parameters. The new set of parameters are either accepted and used to replace the previous ones or they are rejected an the algorithm proceeds with the current parameters depending on the acceptance probability. The basic structure of the Metropolis algorithm that was utilized in this study is as follows:

$$
\begin{aligned}
&\text{Start with } \theta_{\text{now}}. \\
&\text{Draw } \theta_{\text{next}} = \theta_{\text{now}} + \epsilon, \epsilon \sim \text{Normal}(0, \Sigma). \\
&\text{Compute } \alpha = \min\{1, (\pi_{\text{next}}\, q_{\text{next,now}})/(\pi_{\text{now}}\, q_{\text{now,next}})\}. \\
&\text{With probability } \alpha, \text{set } \theta_{\text{next}} = \theta_{\text{next}},\ \text{or else } \theta_{\text{next}} = \theta_{\text{now}}. \\
&\text{Repeat.}
\end{aligned}
\tag{2.4}
$$

In order to be able to draw samples of a random variable, one needs to approximate its density function to a known form. Constructing second order Taylor approximation of the logarithm of the full conditional pdf at the maximum a posteriori estimation (map) we obtain

$$l(x) \approx l(x_{\text{map}}) + \frac{1}{2}(x - x_{\text{map}})^T H_l(x_{\text{map}})(x - x_{\text{map}}), \tag{2.5}$$

where H_l denotes the Hessian matrix of the logarithm of the full conditional pdf at x_{map}. The expression (2.5) is the same as the logarithm of *Normal* density function with mean vector x_{map} and covariance matrix $H_l(x_{\text{map}})^{-1}$.

Combining Gibbs sampling and Metropolis-Hastings algorithm in the process of learning the parameters enables us to tailor the random variable generation process to the specific requirements of a relatively complex Bayesian hierarchical model. One significant advantage this combination of methods provides is that the outcome of the Gibbs sampling always differs at each iteration and therefore delivers a new set of parameters as an input for the Metropolis-Hastings algorithm, regardless of whether the latter accepted its own previously proposed value and supplied a new input for the other or not.

After the parameters were learned through various MCMC methods, the algorithm was tested on simulated sales data in terms of its ability to predict the purchase quantity of a product by a customer at a certain time period given the price and previous purchase quantity. Simulation of sales data will be described in the next chapter.

Chapter 3

Simulation of Store Level Sale Data

The proposed solution method that is a combination of MCMC models and optimization was implemented on simulated data. Firstly, the problem was simplified both in terms of dimension and the number of the parameters that contribute to the prediction of the quantity sold. Once a basic model was constructed, both observed and unobserved variables, were generated according to the predetermined prior distributions with randomly picked hyper-parameters. Later, the aforementioned algorithm was implemented and tested on its ability to learn the actual unobserved parameters. The unobserved parameters were utilized to generate the observed data. It was also evaluated based on its performance on accurately predicting the quantities that would be ordered by the customers at the upcoming time periods, given the price. This procedure was repeated for low, medium and high variation of simulated data. By doing so, we were able to observe and report how the algorithm performed under different scenarios.

3.1 The Effect of Price

The first sub-model considers the effect of price as the sole factor that determines the quantity q_i^{jt} of the product i at time t. The model considers a simple scenario of single customer setting. There are two products with similar features and are substitutes of one other. Two products were given different prices in order to be differentiated, implicitly marking one as the generic and the other as the premium quality product. By doing so, we were also able to observe the difference in cross price elasticities and the cannibalization effect. The influence that promoting the generic product has on the sales of the premium product was expected to be different than the opposite. The graphical model of simpler sub-problem is given in Figure 3.1.

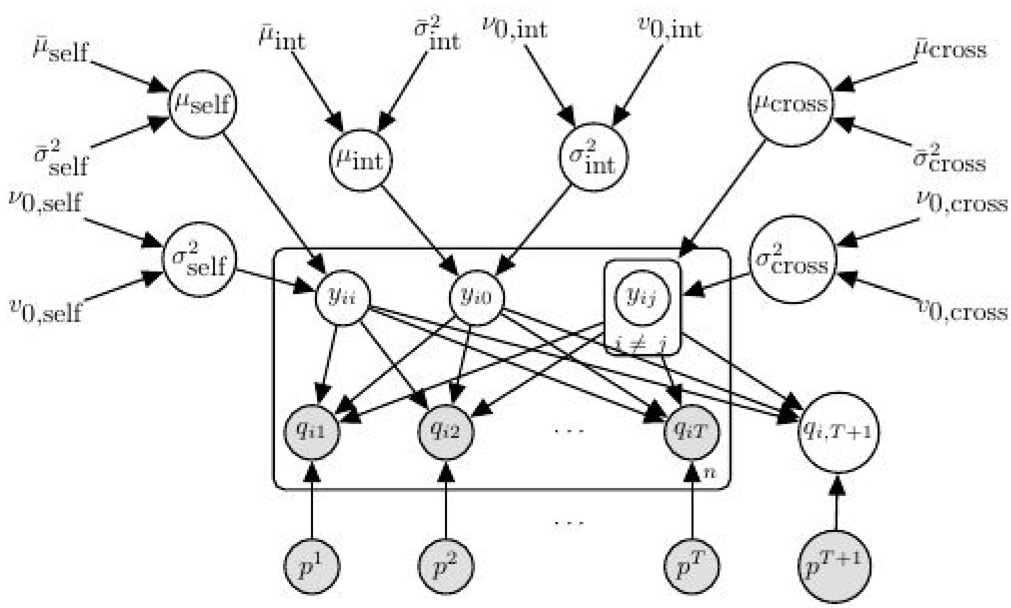

Figure 3.1: Graphical model of Bayesian hiearachical sub-model 1

Observed variables have dark nodes, hidden variables have white nodes. The variables are described in Table 3.2.

The effect of price, previously denoted by β is now referred to as y and is divided into two sub-parts: the self and cross price effects. y_{self} denotes the effect

Notation	Description
$i = 1, \ldots, n$	denote products
$t = 1, \ldots, T+1$	denote periods (T as the current time period)
q^t	nx1 vector of quantities ordered from n products in period t
p^t	nx1 vector of prices of n products offered in period t
y_{self}	nx1 vector denoting the effect of a product's own price (consists of y_{ii}'s)
y^j_{cross}	$(n-1)$x1 vector denoting the effects of the prices of the other products have on product j (consists of y_{ij}'s)
y_0	nx1 vector denoting the intercept (consists of y_{i0}'s)
μ_{self}	1x1 parameter denoting the mean of y_{self}
μ_{cross}	1x1 parameter denoting the mean of y_{cross}
μ_{int}	1x1 parameter denoting the mean of y_0
σ^2_{self}	1x1 parameter denoting the variance of y_{self}
σ^2_{cross}	1x1 parameter denoting the variance of y_{cross}
σ^2_{int}	1x1 parameter denoting the variance of y_0

Figure 3.2: Descriptions of the notations used in sub-model 1

of a product's own price has on its quantity purchased by the customer. Similarly, y_{cross} denotes the effects of the prices of the other products have on a product. The intercept y_0 differs across products. Similar to the model of Section 2, the quantity q_{it} of product i at time t has a *Poisson distribution* with rate λ expressed as in

$$\log(\lambda_i) = y^i_{\text{self}}(p_{it} - \bar{p}_i) + \sum_{i \neq j} y^j_{\text{cross}}(p_{jt} - \bar{p}_j) + y^i_0. \tag{3.1}$$

Here, $\bar{p}$ is the average price in the absence of promotional activities. The variables y_{cross} and y_0 are assumed to be *Normally distributed* with the parameters indicated in the graphical model in Figure 3.1. If none of the products are promoted at a particular time period, then the difference $(p - \bar{p})$ becomes zero in (3.1), and y_0 remains as the only parameter that determines λ.

Although y_{cross} is a free variable by definition, its sign might be foreseen. If a pair of products are substitutes of each other, promoted one would cause a decrease the sales of the other; hence, y_{cross} would take a positive value. On the other hand, if a pair of products are usually consumed together by the end customers, in other words if they tend to be bought simultaneously, then an increase in the price of one would result in a decrease in the sales of the other; thus, y_{cross} would take a negative value. Assuming that an increase in the price of a product will have a negative impact on the sales of that product itself, the variable y_{self} was designed to take negative values only and to ensure this, its distribution was defined as in

$$\log(-y_{\text{self,i}}) \sim \text{ Normal } (\mu_{\text{self}}, \sigma^2_{\text{self}}). \tag{3.2}$$

The full conditional probability density functions of the parameters in the model and their derivations for the implementation of Gibbs and Metropolis algorithms are in Appendix B.

The second order Taylor approximation of the logarithm of the full conditional pdf at the maximum a posteriori estimation (map) of the parameters y_{self}, y_{cross} and y_0, give approximate *Normal distributions* used for the implementation of the Metropolis algorithm. However, in the simulation study, instead of generating random *Normal* variables to propose new values to y_{self}, y_{cross} and y_0 with

Metropolis algorithm, we developed an importance function and used it to generate multivariate *Student's t distribution* samples as suggested by Zellner and Rossi [6]. The degrees of freedom, mean and the variance of the multivariate t *distribution* were defined as in

$$t \sim \mathrm{MSt}(d, x_{\mathrm{map}}, s(-H_l(x_{\mathrm{map}}))^{-1}). \tag{3.3}$$

Here, s denotes the scaling factor that helps tuning the Hessian with the degrees of freedom d. Rossi et al. [7] state that one should choose d greater than 5 if one aims to fatten the tails of the t distribution sufficient enough to minimize variation in the weights, otherwise low degree of freedom would result in very peaked distribution with narrow tails that do not make good importance functions. For this reason, in the preliminary simulation, a variety of values for degrees of freedom d and scaling factors s were tested. We calculated the variance of target variables and choose d and s values so as to give the smallest variance. A similar approach was followed in the original model of Figure 2.1.

Aiming to simulate the effects of different prices on the quantities sold, we constructed a probability function that would determine the price of a product at a particular time depending on a variety of elements. The price of one product was accepted to be independent of the pricing of the other products for the sake of simplicity. For each product, the average price $\bar{p}$ in the absence of promotional activities was accepted as the base price. As in the original store-level data, the promotional activities were expressed as percentage discounts on the original sales price. We specified a range of predetermined discount rates that are the same for every product. The probability of a promotion in the next period increases if the number of weeks passed since that product was last promoted increases. It is assumed that sales and promotions take place in a weekly manner throughout this study. In order to limit the number of weeks the wholesaler can sell its products with the base price and therefore increase the frequency of promotions, we assumed that the longest inter-promotional time can be at most W weeks. For the special case of W=5 and the number of possible distinct discount rates equals to 2, the probabilities associated with the promotional activities are defined as

$$
\begin{aligned}
&P(\text{a promotional activity will occur}) = \min\left(\frac{\log(w_{it})}{\exp(0.5)} + \epsilon, 1\right) \\
&P(\text{the discount rate} = r_1 | \text{a promotional activity will occur}) = \max\left(1 - \frac{w_{it}}{W}, 0\right) \\
&P(\text{the discount rate} = r_2 | \text{a promotional activity will occur}) = \max\left(\frac{w_{it}}{W}, 0\right),
\end{aligned}
\tag{3.4}
$$

where w_{it} is the number of weeks passed at time t since the product i was last promoted. The term ϵ denotes the noise, added in order to ensure that making a promotional activity in two consecutive weeks is still possible, yet it is discouraged. The variables r_1 and r_2 denote the predetermined percentage discount rates one of which is multiplied with the base price $\bar{p}$ to form the final price of a product if a promotion occurs at that time period. In this particular case, it is assumed that r_2 is greater than r_1, so that the longer the product is unpromoted, the higher the chance of a greater discount rate at the next promotion.

3.2 The Effect of Previous Purchase Quantity

The second sub-model considers the effect of previous purchase quantity together with the effect of price as the only factors that determine the quantity q of the product i at time t. Except for the newly added previous purchase quantity parameter, the setting of this preliminary simulation is identical to the previous one in Chapter 3.1. There again are two products with distinct prices, one of generic and one of premium quality, there is one customer and the problem is set on the same planning horizon. The simpler sub-problem was modeled as in Figure 3.3.

Observed variables have dark nodes, hidden variables have white nodes. The variables are described in Table 3.4.

The notation is mostly the same with Chapter 3.1, the effect of price is referred to as y and is divided into two sub-parts: the self and the cross price effect which

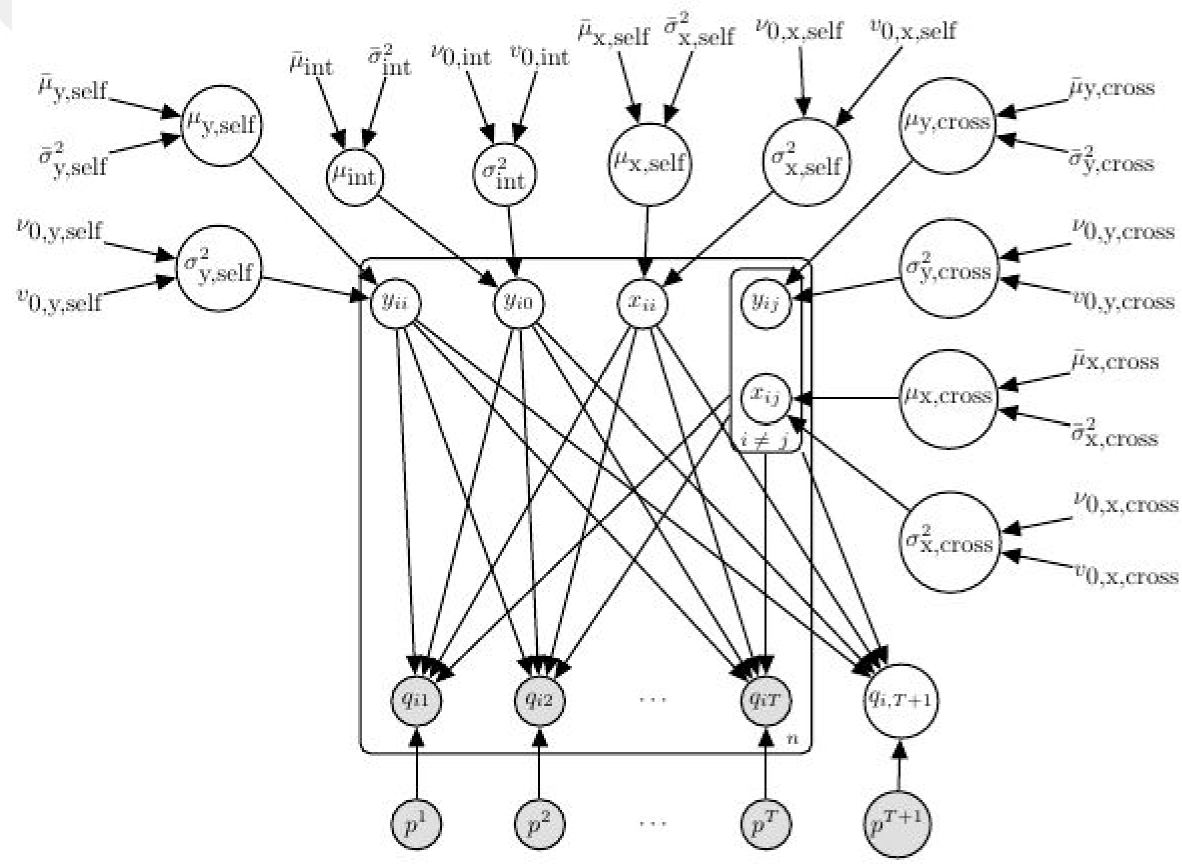

Figure 3.3: Graphical model of Bayesian hiearachical sub-model 2

Notation	Description
$i = 1, \ldots, n$	denote products
$t = 1, \ldots, T+1$	denote periods (T as the current time period)
q^t	nx1 vector of quantities ordered from n products in period t
p^t	nx1 vector of prices of n products offered in period t
y_{self}	nx1 vector denoting the effect of a product's own price (consists of y_{ii}'s)
y^j_{cross}	$(n-1)$x1 vector denoting the effects of the prices of the other products have on product j (consists of y_{ij}'s)
y_0	nx1 vector denoting the intercept (consists of y_{i0}'s)
x_{self}	nx1 vector denoting the effect of a product's own previously purchased quantity (consists of x_{ii}'s)
x^j_{cross}	$(n-1)$x1 vector denoting the effect of the previously purchased quantity of the other products have on product j (x_{ij}'s)
$\mu_{\text{y,self}}$	1x1 parameter denoting the mean of y_{self}
$\mu_{\text{y,cross}}$	1x1 parameter denoting the mean of y_{cross}
μ_{int}	1x1 parameter denoting the mean of y_0
$\sigma^2_{\text{y,self}}$	1x1 parameter denoting the variance of y_{self}
$\sigma^2_{\text{y,cross}}$	1x1 parameter denoting the variance of y_{cross}
σ^2_{int}	1x1 parameter denoting the variance of y_0
$\mu_{\text{x,self}}$	1x1 parameter denoting the mean of x_{self}
$\mu_{\text{x,cross}}$	1x1 parameter denoting the mean of x_{cross}
$\sigma^2_{\text{x,self}}$	1x1 parameter denoting the variance of x_{self}
$\sigma^2_{\text{x,cross}}$	1x1 parameter denoting the variance of x_{cross}

Figure 3.4: Descriptions of the notations used in sub-model 2

correspond to y_{self} and y_{cross} respectively, whereas y_0 is the intercept. Additionally, the effect of previous purchase quantity, previously abbreviated with the symbol γ is now referred to as x and is divided into two sub-parts: the self and the cross effect. x_{self} denotes the effect of a product's own previously purchased quantity has on its quantity to be purchased by the customer, where x_{cross} denotes the effect of the previously purchased quantity of the other products have on a product. Similar to the actual model, the quantity q of product i at time t has a *Poisson distribution* with rate λ described by

$$\log(\lambda_i) = y^i_{\text{self}}(p_{it} - \bar{p}_i) + \sum_{i \neq j} y^j_{\text{cross}}(p_{jt} - \bar{p}_j) + y^i_0 + x^i_{\text{self}}\, q_{i,t-1} + \sum_{i \neq j} x^j_{\text{cross}}\, q_{j,t-1}. \tag{3.5}$$

The variable x_{cross} is assumed to be *Normally distributed* with the parameters that are indicated in the graphical model in Figure 3.3. The variables x_{self} and x_{cross} together implicitly reflect the inventory capacity and shelf space the customer has. For this reason, a high previous purchase quantity is generally expected to have a negative effect on the quantity that the customer will buy in the next period. Although the consumption rate of the end customers and marketing strategy of the retailer may be volatile and unknown by the wholesaler, it is assumed that the retailer has a responsibility to maintain some level of product assortment and passes the promotional discounts that it receives to its end customers. Thus, the effect of previous purchase quantity is esteemed as informative and useful in regard to learning the retailer's order pattern.

Assuming that an excess purchase of the quantity of a product will have a negative impact on the sales of that product in the upcoming period, the variable x_{self} was designed to take negative values only and to ensure this, its distribution was defined as

$$\log(-x_{\text{self,i}}) \sim \ \mathit{Normal}\ (\mu_{\text{x,self}}, \sigma^2_{\text{x,self}}). \tag{3.6}$$

The full conditional probability density functions of the parameters in the model and their derivations prepared for the implementation of Gibbs and Metropolis Algorithms can be found in the Appendix C.

Chapter 4

Learning Price Promotion Effects from Simulated Sales Data

We tested the aforementioned methodology in three steps. We first tested it on the simulated sales data and the simplified versions of the original model discussed in Chapter 3. The first simplified model only considered the effect of price as the determining factor for the quantity prediction whereas the second simplified model reflected the customer inventory with the effect of previous purchase quantity by adding this factor to the first simplified model. The simulation data were generated according to the randomly selected initial hyper-parameters separately for each model with predetermined base prices in a one customer and two products setting. The base price was assumed to remain same throughout the planning horizon and unaffected of inflation or other market fluctuation for the sake of simplicity. These steps were repeated for low, medium and high variances of target random variables in order to test the methodology's performance under a variety of conditions. The simulation models were also evaluated according to their convergence to the generating parameters when the starting values of the parameters are taken as their maximum a-posteriori estimators under the full knowledge of other true parameters or randomly chosen.

The third step was to test the methodology on the actual store-level data where there were no information available on the nature of the parameters that are assumed to determine the quantity sold in each time period to each customer. For this reason, the original model was evaluated solely on its ability to precisely predict the quantity of products to be sold and therefore implicitly how a customer will react to the prices offered by a wholesaler. During the implementation of the algorithm, each lower hierarchical group were handled separately.

For the model that only considers the effect of price, the price and the quantity pairs were simulated for 124 weeks, first 100 were used to train the model and learn the parameters while the remaining 24 were the test data and was compared to the predicted values. During the implementation step of the algorithm, in order to randomly generate multivariate *Student's t distribution* values for y_{self}, y_{cross} and y_0, the importance function was formed. Furthermore, aiming to apply this method in an effective and useful way, three different scaling factors s and degrees of freedom d were tested for each parameter. Then the combination that yielded the least numerical error were chosen, and the algorithm proceeded with those values. The best s and d pairs and their respective errors for every parameter for three different variation scenarios can be found in Table 4.1.

For low, medium and high variations the pair (s, d) were chosen as (0.5,10), (0.5,5) and (0.5,10). It was observed that smaller scaling factors and relatively larger degrees of freedom parameters yielded smaller error values. However, there were no major changes observed in the error values for a given set of (s, d) pair as the variation increased. The table comparing the true parameters and the algorithm output after 2000 iterations that uses the best s and d pairs mentioned above were constructed as in Table 4.2.

The parameters that were generated according to the Gibbs algorithm were processed in every iteration. Therefore, they do not have acceptance percentages to be displayed on the Table 4.2. We observed that as the variation increases, the predicted values of the parameters gets closer to the true values.

It was observed that independent of the starting point, the algorithm converged to true parameters in 5000 replications. Figures 4.3, 4.4, and 4.5 show the

Low Variation			
Parameter	s	d	Error
y_{self} generic	0.5	10	0.00201
y_{self} premium	0.5	7	0.00315
y_{cross} generic	0.5	10	0.00093
y_{cross} premium	0.5	5	0.00167
y_0 generic	0.5	10	0.00049
y_0 premium	0.5	7	0.00051
Medium Variation			
y_{self} generic	0.5	5	0.00602
y_{self} premium	0.5	5	0.00271
y_{cross} generic	0.5	10	0.00177
y_{cross} premium	0.5	5	0.00175
y_0 generic	0.5	10	0.00075
y_0 premium	0.5	7	0.00066
High Variation			
y_{self} generic	0.5	10	0.00227
y_{self} premium	0.5	10	0.00268
y_{cross} generic	0.5	7	0.00083
y_{cross} premium	0.5	10	0.00336
y_0 generic	0.5	10	0.00033
y_0 premium	0.5	10	0.00125

Figure 4.1: Best degrees of freedom and scaling factor pairs for model parameters

plots of Metropolis algorithm output and true values of the effect of price y_{self}, y_{cross} and the intercept y_0 under low, medium and high variation scenarios. The true parameters are indicated with a red line in each graph and the data points denote the samples from Metropolis algorithm at each iteration. Under low and medium variation, the algorithm rapidly converges to the true values for y_{self} and y_0 but overestimates under high variation. The algorithm performs well under low, medium and high variations for y_{cross}.

Figure 4.6 show the Gibbs algorithm output and true values of the mean parameters of y_{self}, y_{cross} and y_0, which correspond to μ_{self}, μ_{cross} and μ_0 respectively. Figure 4.7 shows the Gibbs algorithm output and true values of the variance parameters of $y$self, y_{cross} and y_0, which correspond to σ^2_{self}, σ^2_{cross} and σ^2_0 respectively.The true parameters are indicated with a red line in each graph and the data points denote the samples from Gibbs algorithm at each iteration. Gibbs algorithm successfully converges to true values of μ_{self} under low and medium variations, but underestimates the true parameters under high variation. Moreover,

Low Variation					
		Algorithm			
Parameter	True Value	50% Quantile	Mean	Standard Deviation	Acceptance %
y_{self} generic	-0.7167	-0.7718	-0.7715	0.0417	88.54
y_{self} premium	-0.5449	-0.3959	-0.3973	0.0380	88.54
y_{cross} generic	0.1868	0.1909	0.1911	0.0321	92.45
y_{cross} premium	0.3158	0.4034	0.4042	0.0463	92.15
y_0 generic	3.7220	3.7086	3.7087	0.0121	87.14
y_0 premium	3.4763	3.5221	3.5224	0.0128	87.14
μ_{self}	-0.3242	-0.5590	-0.5483	0.3185	-
σ^2_{self}	0.4863	0.1661	0.2742	2.2463	-
μ_{cross}	0.2500	0.2732	0.2698	0.1241	-
σ^2_{cross}	0.8547	0.1826	0.2894	2.2393	-
μ_{int}	3.4021	3.4728	3.4476	0.3284	-
σ^2_{int}	0.1006	0.1700	0.2803	2.2404	-
Medium Variation					
y_{self} generic	-0.3361	-0.4636	-0.4654	0.0778	85.14
y_{self} premium	-0.5389	-0.5764	-0.5781	0.0408	85.14
y_{cross} generic	1.5151	1.3621	1.3647	0.0673	90.70
y_{cross} premium	0.8331	0.7787	0.7781	0.0558	90.05
y_0 generic	2.8460	2.7982	2.7981	0.0203	84.94
y_0 premium	3.2997	3.2949	3.2950	0.0153	84.94
μ_{self}	-0.3243	-0.6072	-0.6039	0.3303	-
σ^2_{self}	0.5411	0.1687	0.2793	2.2432	-
μ_{cross}	0.2500	0.9879	0.9646	0.1434	-
σ^2_{cross}	0.8116	0.1692	0.2770	2.2438	-
μ_{int}	3.4021	2.9624	2.9484	0.3249	-
σ^2_{int}	1.0148	0.1777	0.2860	2.24070	-
High Variation					
y_{self} generic	-0.5063	-0.5055	-0.5059	0.0264	88.49
y_{self} premium	-1.1985	-0.9562	-0.9551	0.0729	88.49
y_{cross} generic	1.3892	1.4933	1.4924	0.0269	91.35
y_{cross} premium	-0.4859	-0.0650	-0.0648	0.1062	94.20
y_0 generic	4.6838	4.7109	4.7108	0.0075	88.09
y_0 premium	1.9377	2.0988	2.0987	0.0297	88.09
μ_{self}	-0.3243	-0.3353	-0.3308	0.3255	-
σ^2_{self}	1.8234	0.1709	0.2793	2.2405	-
μ_{cross}	0.2500	0.6814	0.6696	0.1040	-
σ^2_{cross}	0.8008	0.1152	0.2063	2.2372	-
μ_{int}	3.4021	3.3396	3.3219	0.2605	-
σ^2_{int}	2.0110	0.1025	0.1846	2.2372	-

Figure 4.2: Predicted and true parameters of sub-model

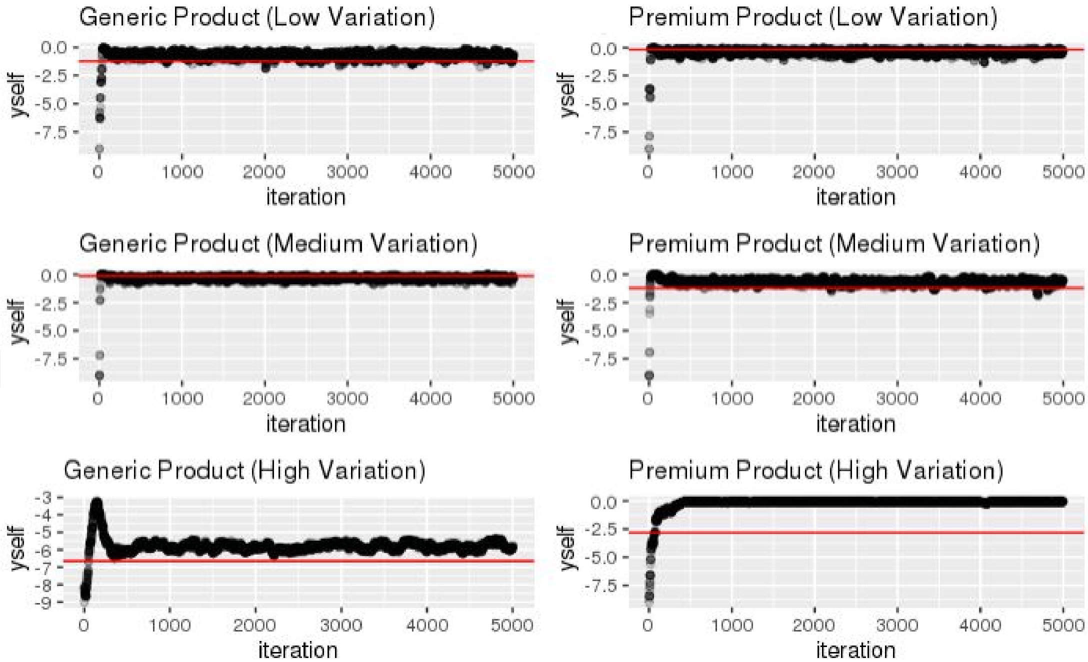

Figure 4.3: Predicted and true values of y_{self}, the effect of product's price on its own demand

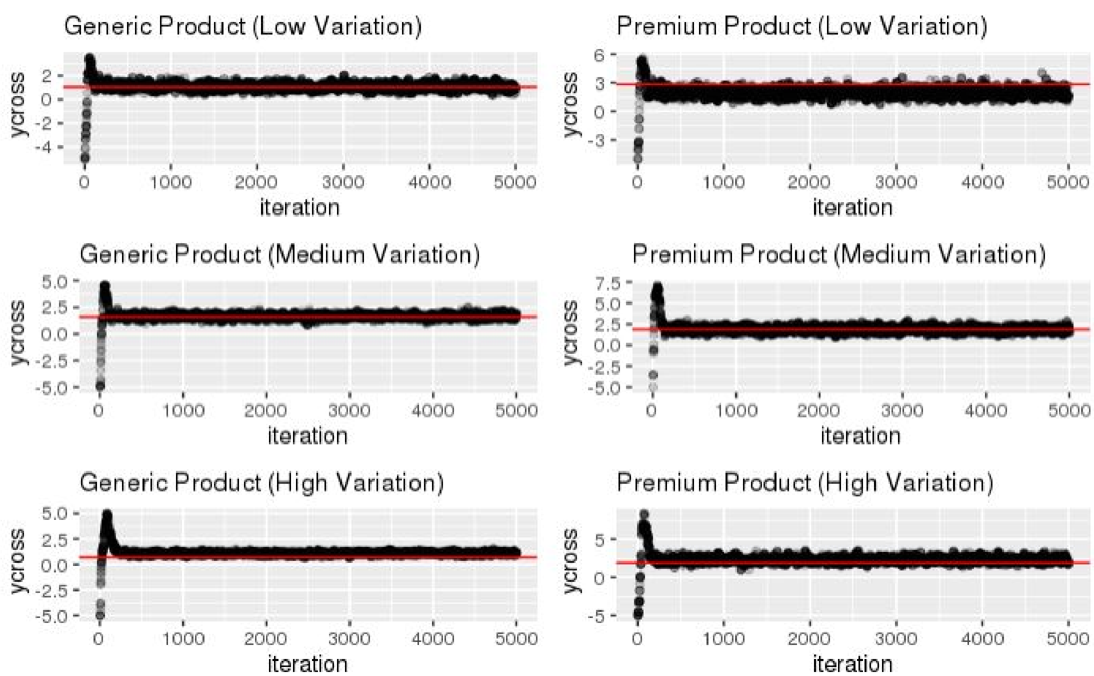

Figure 4.4: Predicted and true values of y_{cross}, effects of the prices of products on the demand of a competing product

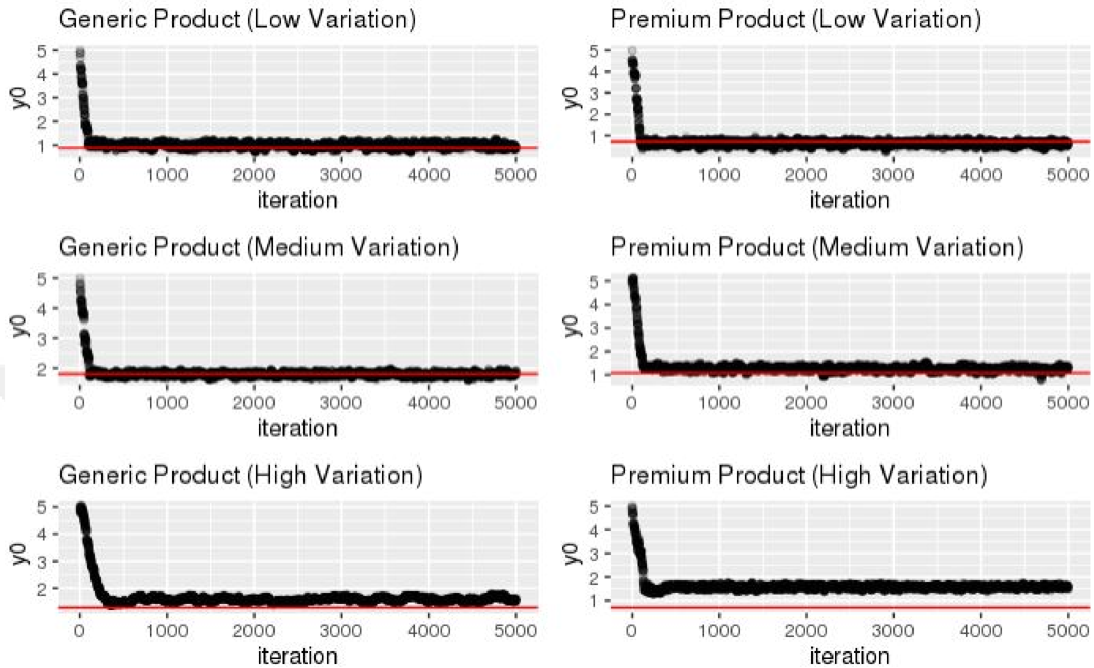

Figure 4.5: Algorithm output and true values of y_0, the intercept

the algorithm successfully estimates μ_{cross} and μ_0, except for slight overestimation issues in high variation. For variance parameters, Gibbs algorithm generally perform well under different scenarios. However, σ^2_{self} is gradually overestimated as the variation increased.

The results show that for the parameters that denote the effect of price, Random Walk Metropolis algorithm performs well and is able to converge to true values under low, medium and high variation scenarios. Similarly, Gibbs algorithm is able to learn true mean and variance parameters of the model under medium and high variation scenarios, but tends to underestimate at low variation. It is observed that the over and under estimation of the price effect parameters reflect similar convergence issues with mean and variance parameters. These results indicate that for certain set of parameters and under particular variation scenarios, both Random Walk Metropolis and Gibbs algorithms do not perform as well as the remaining cases. This enables us to comprehend the limits and the applicability of these methods under different conditions.

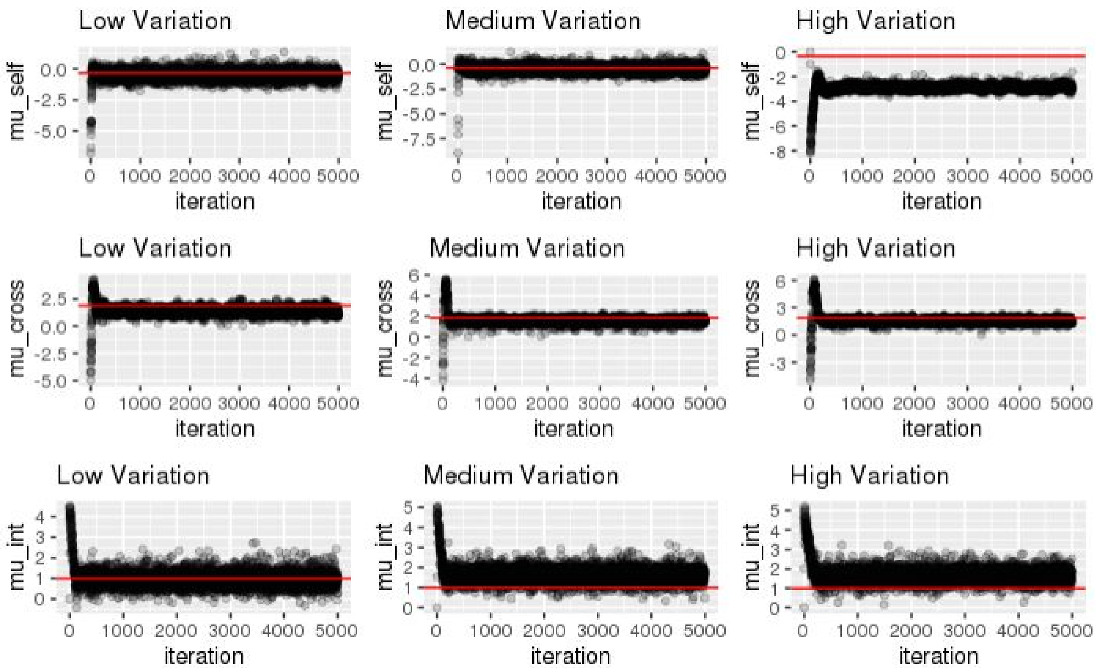

Figure 4.6: Predicted and true values of mean parameters

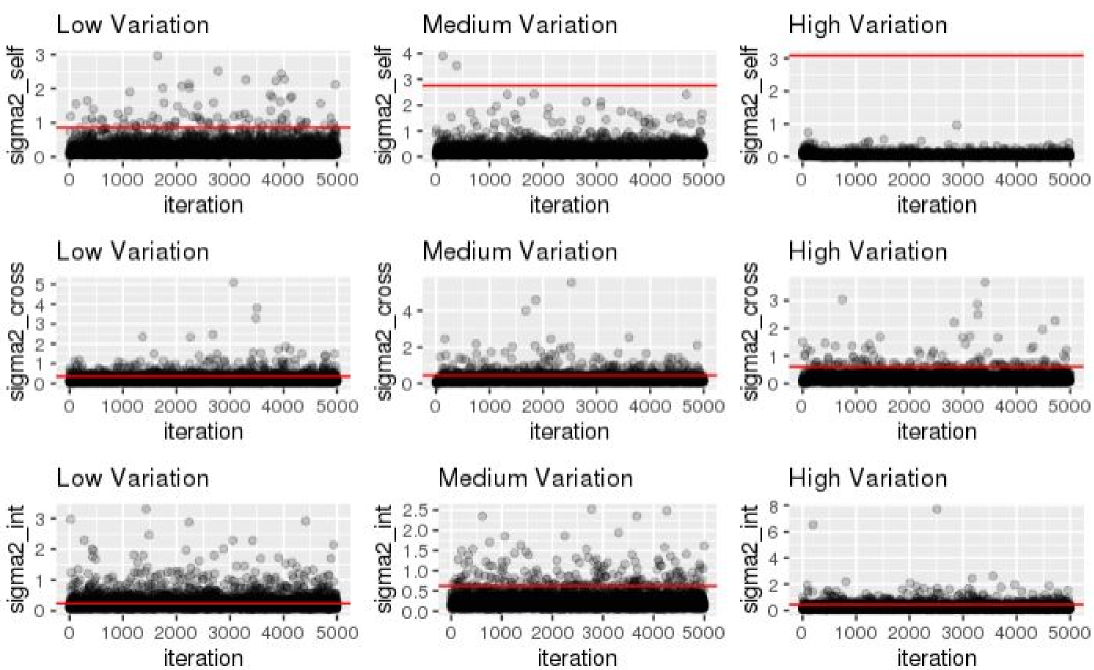

Figure 4.7: Predicted and true values of variance parameters

Next, the algorithm otuput of the parameters of the model were used to predict the purchase quantity of the product for the training data (100 weeks) and the test data (24 weeks). The results regarding the training data under low, medium and high variance scenarios are illustrated in Figures 4.8, 4.10, and 4.12 respectively. The original purchase quantities were indicated with black and the algorithm output was plotted using different colors for generic and premium products.

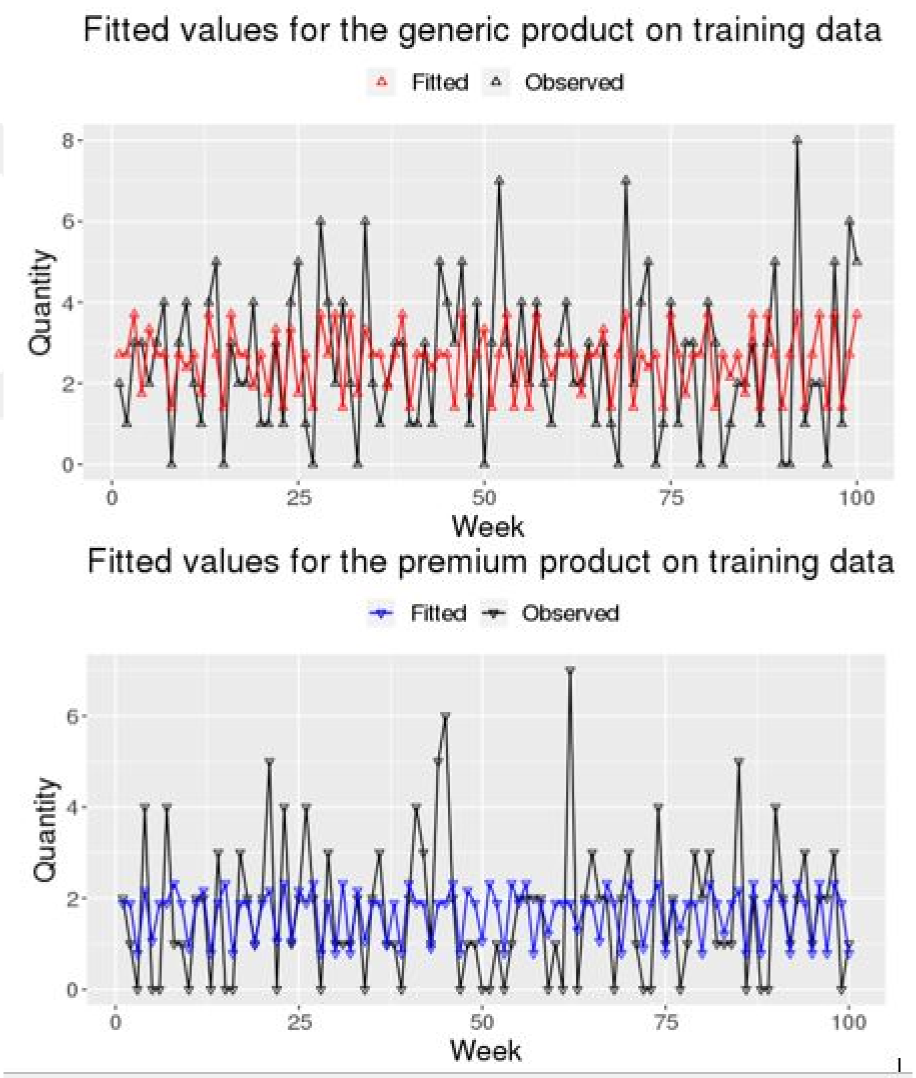

Figure 4.8: Fitted and observed values of purchase quantities under low variation on training set

The results regarding the test data under low, medium and high variance scenarios are illustrated in Figures 4.9, 4.11 and 4.13 respectively. The original

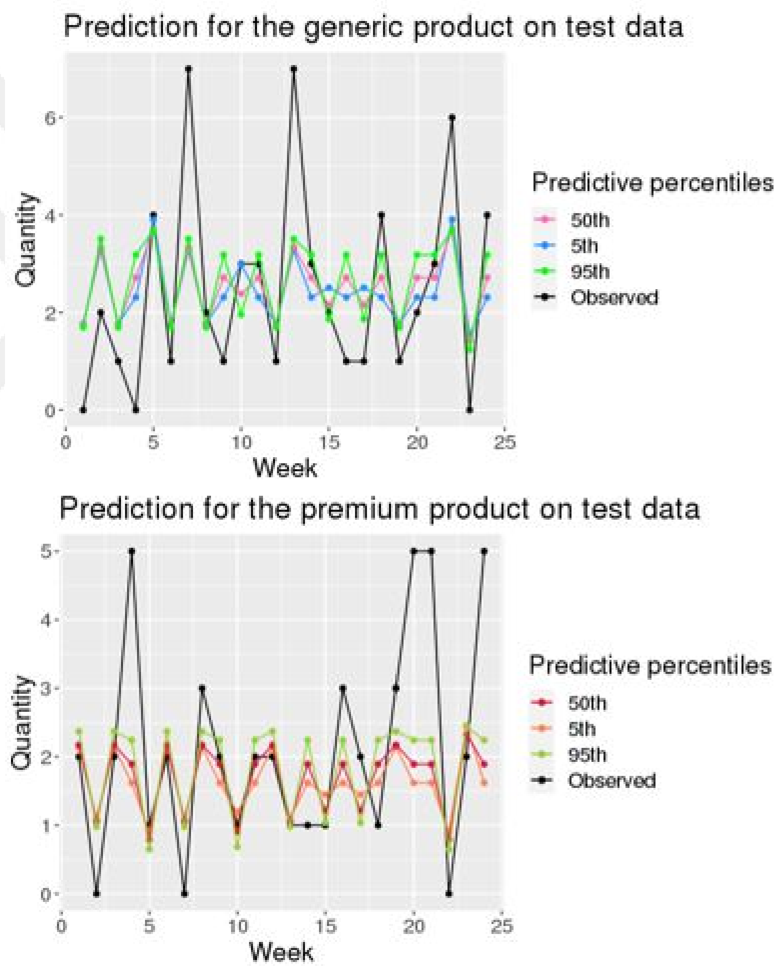

Figure 4.9: Predicted and true values of purchase quantities under low variation on test set

purchase quantities were indicated with black and the algorithm output was plotted using different colors for generic and premium products.The algorithm values were plotted using multiple indicators, the plots include median and 5%, 50%, 90% quantile values, in order to be able to test if the original values lay within these intervals.

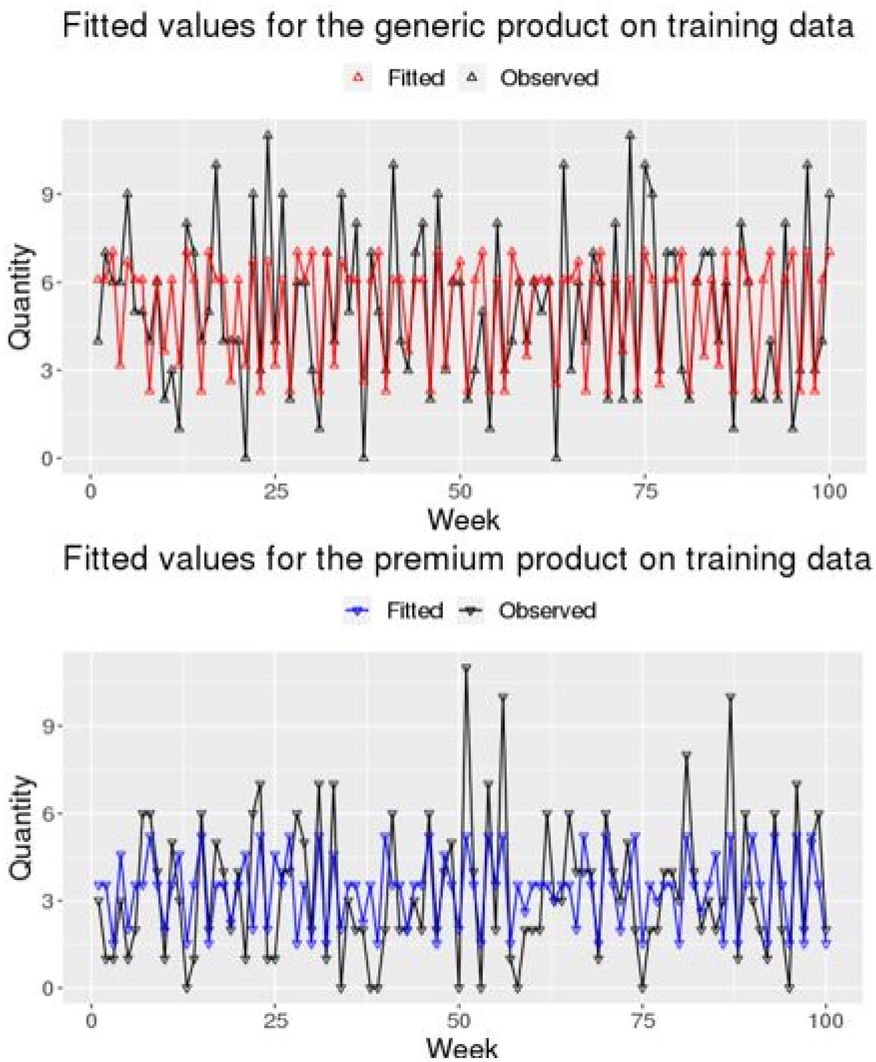

Figure 4.10: Fitted and observed values of purchase quantities under medium variation on training set

The plots show that the algorithm is able to predict the purchase quantity of both of the products in the training and test data sets more accurately as the variation increases. The algorithm also performs better in terms of predicting the purchase quantity of the generic product, compared to premium product. These

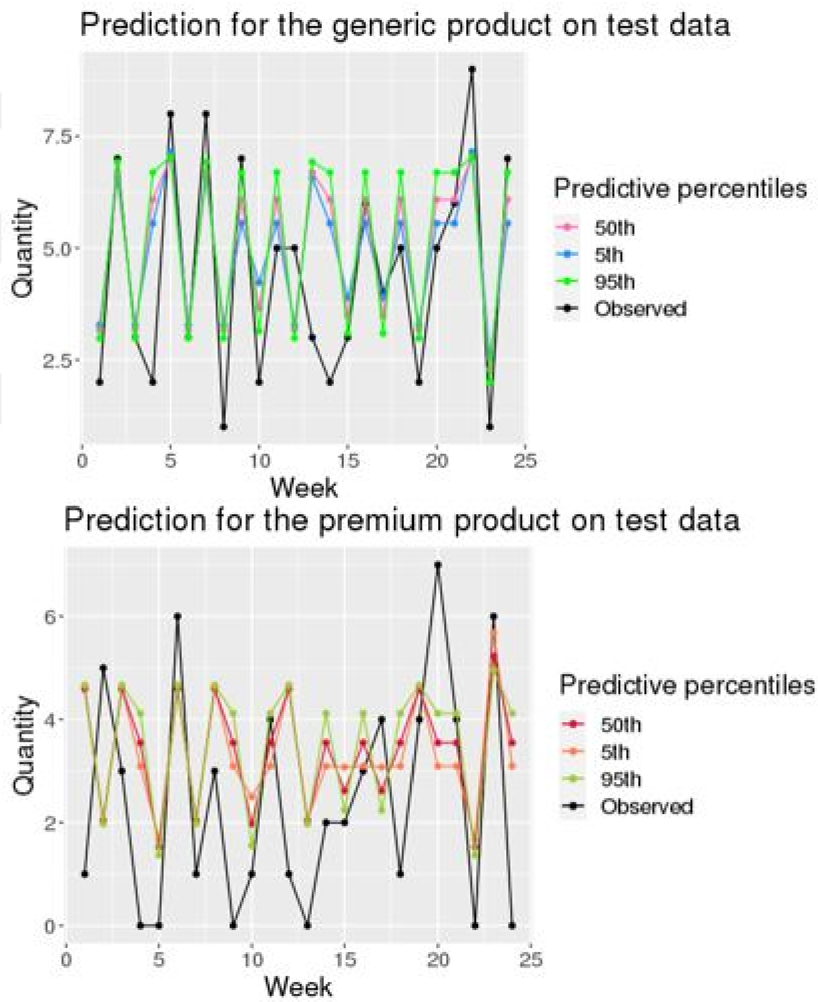

Figure 4.11: Predicted and true values of purchase quantities under medium variation on test set

aspects allow for the algorithm to be applicable to larger data sets with higher variation among the purchase quantities, which is often encountered in real life scenarios.

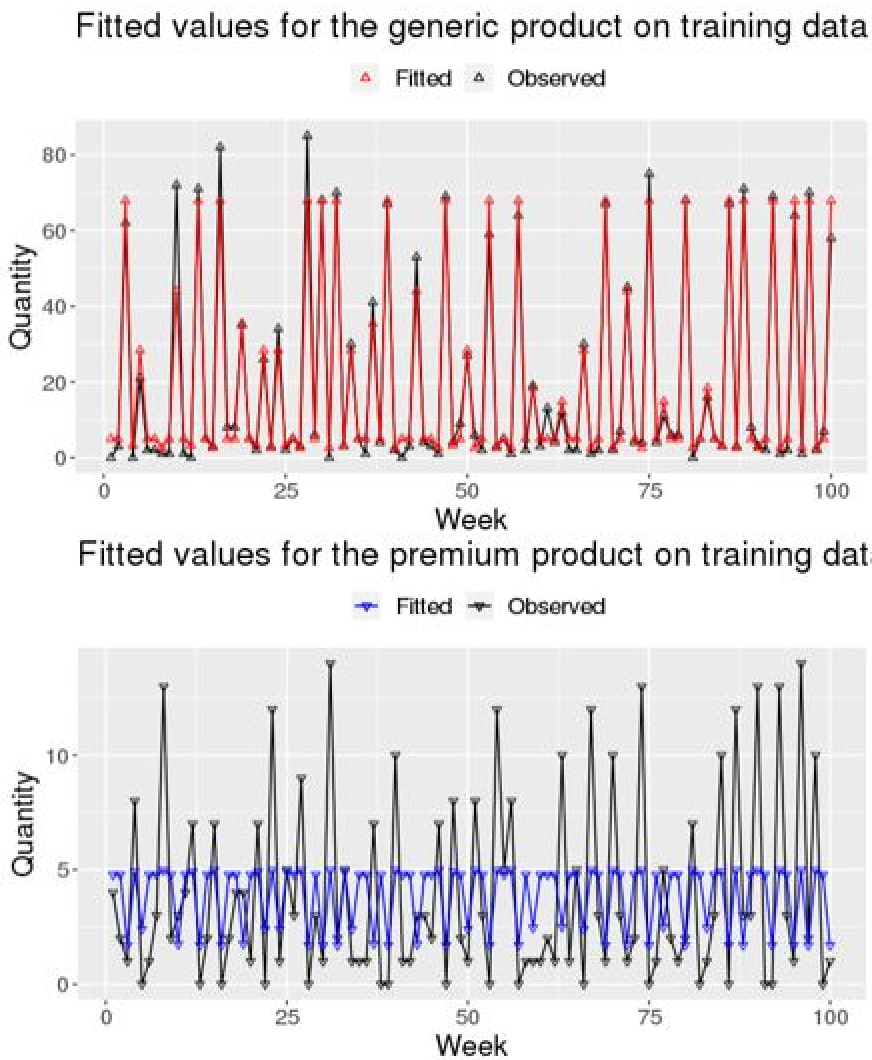

Figure 4.12: Fitted and observed values of purchase quantities under high variation on training set

The auto-correlation plots of the parameters related to the effect of price and the intercept under different variation scenarios are shown in Figures 4.14, 4.15 and 4.16, respectively. The plots show that for higher levels of variation, a higher level of auto-correlation was observed for both of the products. The high auto-correlation observed in these plots reflect the high acceptance rates observed in Table 4.2. On the other hand, it was observed that auto-correlation was higher

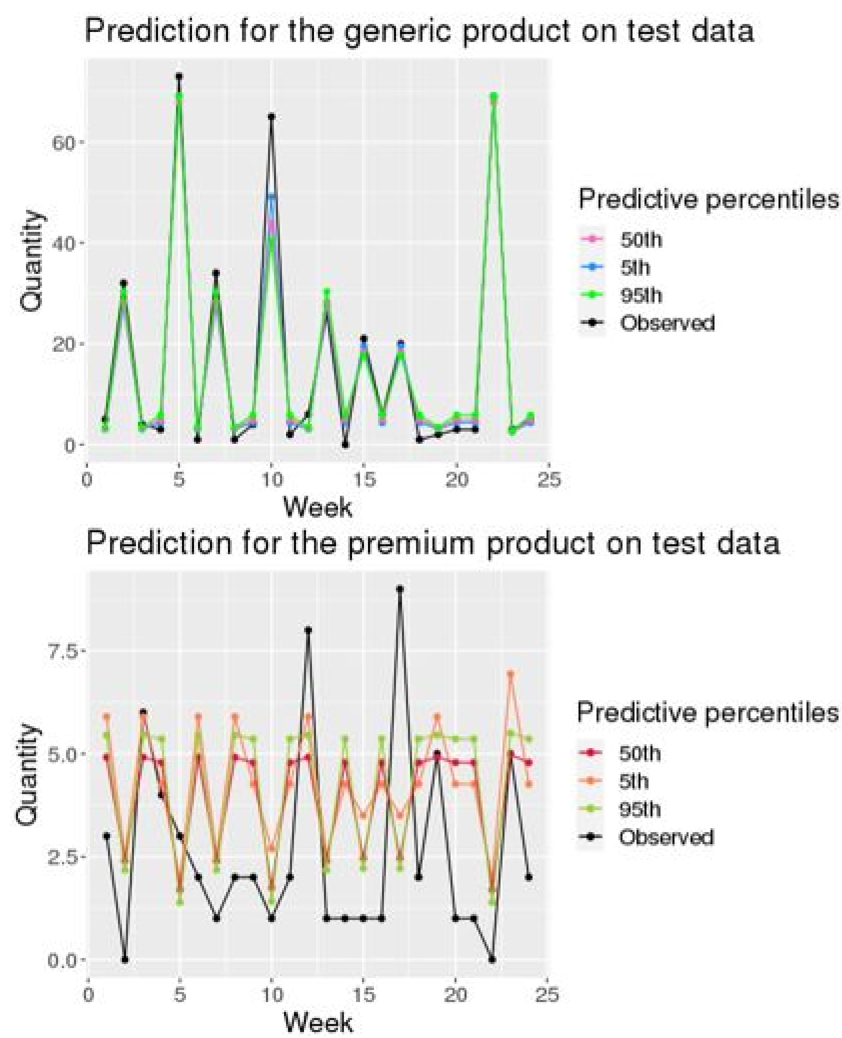

Figure 4.13: Predicted and true values of purchase quantities under high variation on test set

for premium product, compared to generic product.

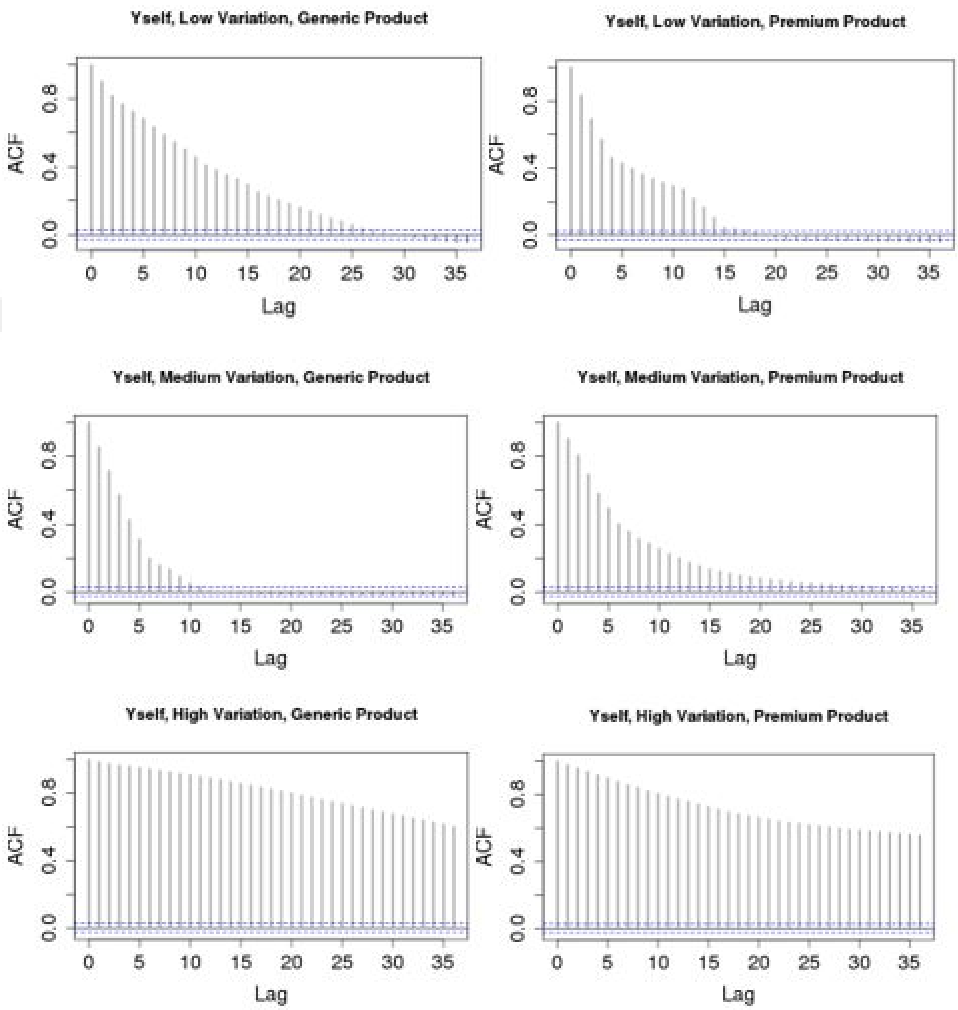

Figure 4.14: Auto-correlation plots of y_{self}

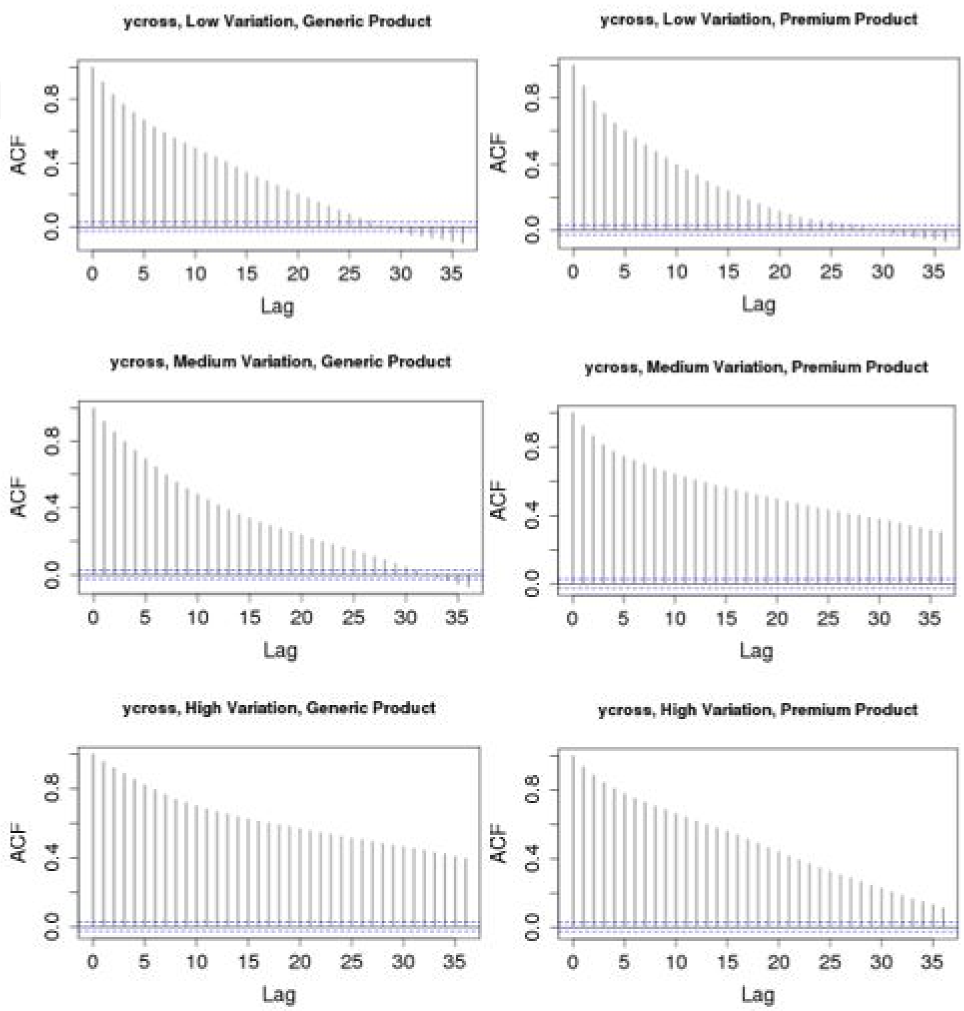

Figure 4.15: Auto-correlation plots of y_{cross}

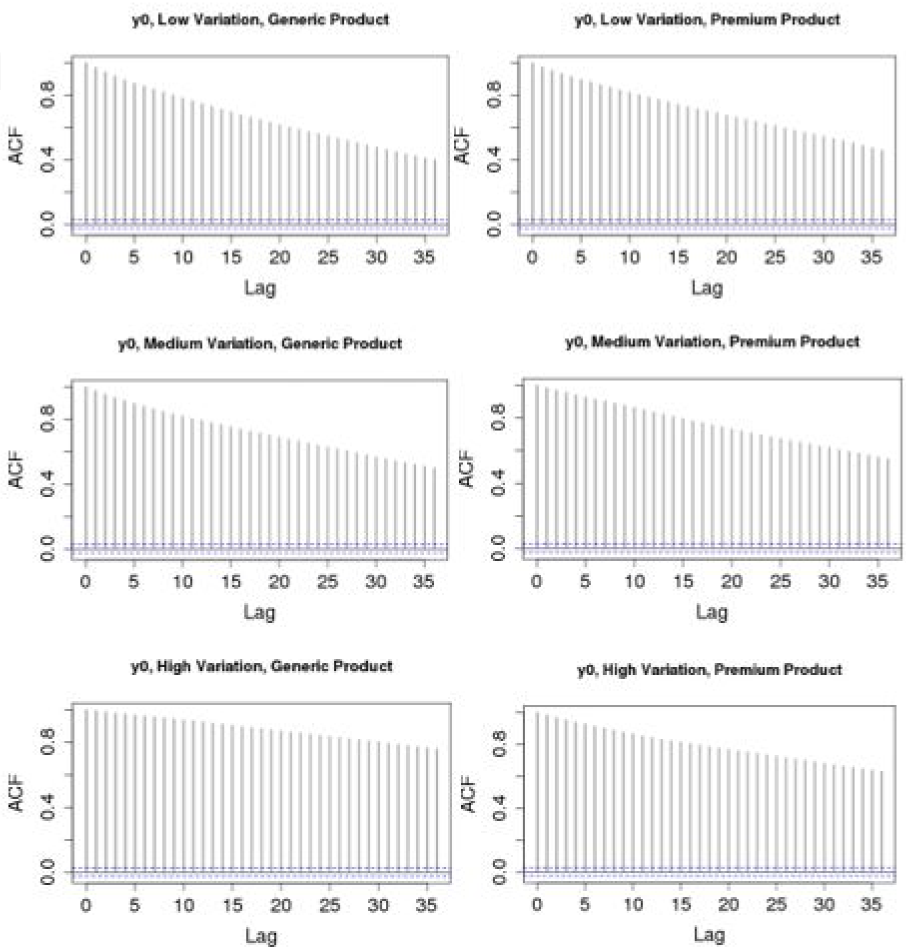

Figure 4.16: Auto-correlation plots of y_0

Chapter 5

Conclusion

From the perspective of the wholesaler, offering price promotions to the retailers seem like a beneficial tool that could increase the profits on the short run. However, depending on how, how much and when price promotions are offered, this expected increase on the sales of the promoted product can also alter the sales of its substitutes and thus, may not be profitable in the long run. This study showed that the success of price promotions depend on several factors including the regular price of the product as well as its previous purchase quantity, which implicitly informs the decision maker of the retailer's inventory and the end customer demand. The results also demonstrated that the purchases of products in a wholesaler's assortment can be modeled using Bayesian hierarchical methods and the parameters of the probability distributions can be learned using Monte Carlo Markov Chain methods. By doing so, this study provides a basis for the utilization of Bayesian hierarchical modeling on store level sales data and prediction of purchase quantities of the retailers as an input for the wholesaler's revenue maximization problem.

The results of the study showed that the proposed algorithm was able to learn the underlying parameters when implemented on simulated data. As discussed in Section 4, the hybrid MCMC algorithm is able to predict the model parameters with a small error rate when the algorithm starts from a random point. Table

4.1 shows that a small scaling factor and a larger degrees of freedom pair yields a lower error of the estimates. It was observed that the algorithm performs better as the variance increases. These findings indicate that the algorithm converges to Bayes optimal parameters and is able to predict the order quantities of the customers.

The results of this study suggest that the proposed hybrid MCMC algorithm can predict the purchase quantity of simulated sales data. As a future research direction, it is possible to implement the same methodology on real store level sales data. By carrying out several data processing tasks like grouping products in the raw data into hierarchical subgroups and identifying the substitution relationships among them, real store level data can be used as an input for the aforementioned algorithm and promotional effects can be observed. Learning the parameters of the predetermined probability distributions allows the decision maker to be able to make predictions about the purchase quantities of the retailers, for a given set of prices.

Using the parameter input provided by this study, one can also concentrate on the profit maximization problem of the wholesaler. In the absence of information about the costs associated with implementing different pricing policies, it transforms into a revenue maximization problem. The objective is to maximize the revenue defined as the multiplication of the price and the order quantity, summed over all products and customers over the planning horizon, subject to the constraint that the order quantity equals to the expected mean of the *Poisson* purchase rate. This value can be denoted as a function of the observed variables of the seasonality, price, previous purchase quantity and their respective effects on order quantity, calculated using the hybrid MCMC algorithm.

Appendix A

Calculations and Derivations

A.1 Full conditional PDF of α

$$
\begin{aligned}
&\text{f}(\alpha |\text{ everything else (including q) }) \propto \text{f}(\alpha, \text{everything else}) \\
&\propto \text{f}(\ \alpha \,|\, \mu_\alpha\ ,\ \Sigma^\alpha\ ) \prod_{j=1}^{m} \prod_{t=1}^{T} \prod_{i=1}^{n} \text{f}(\ q_{it}^{(j)} \,|\, \alpha, \beta^{(j)}, \gamma^{(j)}, p_{it}^{(j)}) \\
&= \frac{1}{(2\pi)^{\frac{4}{2}} |\Sigma^\alpha|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\alpha - \mu_\alpha)^T (\Sigma^\alpha)^{-1} (\alpha - \mu_\alpha)\} \prod_{j=1}^{m} \prod_{t=1}^{T} \prod_{i=1}^{n} \frac{[\lambda_{ijt}]^{q_{it}^{(j)}} \exp\{-\lambda_{ijt}\}}{q_{it}^{(j)}} \\
&= \frac{1}{(2\pi)^{\frac{4}{2}} |\Sigma^\alpha|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\alpha - \mu_\alpha)^T (\Sigma^\alpha)^{-1} (\alpha - \mu_\alpha)\} \times \\
&\prod_{j=1}^{m} \prod_{t=1}^{T} \prod_{i=1}^{n} \frac{1}{q_{it}^{(j)}!} \left[\exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\} \right]^{q_{it}^{(j)}} \times \\
&\exp\left\{ - \exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\} \right\} \\
&\propto \exp\{-\frac{1}{2}[\alpha^T (\Sigma^\alpha)^{-1} \alpha - 2\alpha^T (\Sigma^\alpha)^{-1} \mu_\alpha]\} \times \\
&\prod_{j=1}^{m} \prod_{t=1}^{T} \prod_{i=1}^{n} \left[\exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\} \right]^{q_{it}^{(j)}} \times \\
&\exp\left\{ - \exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\} \right\}
\end{aligned}
\tag{A.1}
$$

A.2 Full conditional PDF of μ_α

$$
\begin{aligned}
&\text{f}(\mu_\alpha|\text{ everything else (including q) }) \propto \text{f}(\mu_\alpha, \text{everything else})\\
&\propto \text{f}(\ \mu_\alpha \,|\, \mu_A\ ,\ \Sigma^A\ )\text{f}(\ \alpha \,|\, \mu_\alpha\ ,\ \Sigma^\alpha\ )\\
&= \frac{1}{(2\pi)^{\frac{4}{2}}|\Sigma^A|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(\mu_\alpha-\mu_A)^T(\Sigma^A)^{-1}(\mu_\alpha-\mu_A)\}\times\\
&\frac{1}{(2\pi)^{\frac{4}{2}}|\Sigma^\alpha|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(\alpha-\mu_\alpha)^T(\Sigma^\alpha)^{-1}(\alpha-\mu_\alpha)\}\\
&\propto \exp\{-\frac{1}{2}[(\mu_\alpha-\mu_A)^T(\Sigma^A)^{-1}(\mu_\alpha-\mu_A)+(\alpha-\mu_\alpha)^T(\Sigma^\alpha)^{-1}(\alpha-\mu_\alpha)]\}\\
&\propto \exp\{-\frac{1}{2}[\mu_\alpha^T(\Sigma^A)^{-1}\mu_\alpha-2\mu_A^T(\Sigma^A)^{-1}\mu_\alpha+\mu_\alpha^T(\Sigma^\alpha)^{-1}\mu_\alpha-2\alpha^T(\Sigma^\alpha)^{-1}\mu_\alpha]\}\\
&= \exp\{-\frac{1}{2}[\mu_\alpha^T((\Sigma^A)^{-1}+(\Sigma^\alpha)^{-1})\mu_\alpha-2(\mu_A^T(\Sigma^A)^{-1}+\alpha^T(\Sigma^\alpha)^{-1})\mu_\alpha]\}
\end{aligned}
\tag{A.2}
$$

which is the same as the pdf of Normal (μ, Σ) where

$$
\begin{aligned}
&\Lambda := \Sigma^{-1} = (\Sigma^A)^{-1} + (\Sigma^\alpha)^{-1}\\
&\mu = \Sigma[((\Sigma^A)^{-1})^T\mu_A + ((\Sigma^\alpha)^{-1})^T\alpha]\\
&\ = ((\Sigma^A)^{-1} + (\Sigma^\alpha)^{-1})^{-1}[((\Sigma^A)^{-1})^T\mu_A + ((\Sigma^\alpha)^{-1})^T\alpha]
\end{aligned}
\tag{A.3}
$$

A.3 Full conditional PDF of Σ^{α}

$$
\begin{aligned}
&\text{f}(\ \Sigma^{\alpha}|\text{ everything else}) \propto \text{f}(\ \Sigma^{\alpha}, \text{ everything else})\\
&\propto \text{f}(\ \Sigma^{\alpha}|\nu_A, V_A\ ) * \text{f}(\ \alpha|\mu_{\alpha}, \Sigma^{\alpha})\\
&= 2^{\frac{\nu_A * 4}{2}} \pi^{\frac{4(4-1)}{4}} \prod_{\text{i}=1}^{4} \Gamma\left(\frac{\nu_A - 1 - i}{2}\right)^{-1} \times |V_A|^{\frac{\nu_A}{2}} |\Sigma^{\alpha}|^{\frac{-(\nu_A+4+1)}{2}} \exp\left\{\text{tr}\left(\frac{-1}{2} V_A (\Sigma^{\alpha})^{-1}\right)\right\}\\
&\frac{1}{(2\pi)^{\frac{4}{2}} |\Sigma^{\alpha}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left[(\alpha-\mu_{\alpha})^T (\Sigma^{\alpha})^{-1} (\alpha - \mu_{\alpha})\right]\right\}\\
&\propto |\Sigma^{\alpha}|^{\frac{-(\nu_A+4+1)}{2}} \times \exp\left\{-\text{tr}\left(\frac{V_A(\Sigma^{\alpha})^{-1}}{2}\right)\right\} \times\\
&|\Sigma^{\alpha}|^{-\frac{1}{2}} \exp\left\{-\text{tr}\left(\frac{[(\alpha-\mu_{\alpha})(\alpha-\mu_{\alpha})^T(\Sigma^{\alpha})^{-1}]}{2}\right)\right\}\\
&\propto |\Sigma^{\alpha}|^{\frac{-(\nu_A+4+2)}{2}} \exp\left\{-\text{tr}\left(\frac{(V_A + (\alpha-\mu_{\alpha})(\alpha-\mu_{\alpha})^T(\Sigma^{\alpha})^{-1}}{2}\right)\right\}\\
&\text{which is the pdf of Inverted Wishart } (\nu_A + 1, V_A + (\alpha-\mu_{\alpha})(\alpha-\mu_{\alpha})^T)
\end{aligned}
\tag{A.4}
$$

A.4 Full conditional PDF of μ_{β}^{j}

$$
\begin{aligned}
&\text{f}(\text{ vec } \mu_{\beta}^{(j)}|\text{ everything else (including q) }) \propto\ \text{f}(\text{ vec } \mu_{\beta}^{(j)}\ , \text{ everything else})\\
&\propto\ \text{f}(\text{ vec } \mu_{\beta}^{(j)}|X_j\ , H_{\text{ 1: n}}\ , \Sigma^{B}\ )\ \text{f}\ (\text{ vec } \beta^{(j)}|\text{ vec } \mu_{\beta}^{(j)}, \Sigma^{\beta}\ )\\
&= \frac{1}{(2\pi)^{\frac{n^2}{2}} |\Sigma^{B}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \mu_{\beta}^{(j)} - H_{\text{ 1: n}} X_j)^T (\Sigma^{B})^{-1} (\text{vec } \mu_{\beta}^{(j)} - H_{\text{ 1: n}} X_j)\} \times\\
&\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma^{\beta}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \beta^{(j)} - \text{vec } \mu_{\beta}^{(j)})^T (\Sigma^{\beta})^{-1} (\text{vec } \beta^{(j)} - \text{vec } \mu_{\beta}^{(j)})\}\\
&\propto \exp\{-\frac{1}{2}[\ (\text{ vec}\mu_{\beta}^{(j)})^T (\Sigma^{B})^{-1}\ (\text{vec } \mu_{\beta}^{(j)}) - 2(H_{\text{ 1: n}} X_j)^T (\Sigma^{B})^{-1} (\text{vec } \mu_{\beta}^{(j)})\\
&+ (\text{ vec}\mu_{\beta}^{(j)})^T (\Sigma^{\beta})^{-1} (\text{ vec } \mu_{\beta}^{(j)}) - 2(\text{ vec } \beta^{(j)})^T (\Sigma^{\beta})^{-1} (\text{ vec}\mu_{\beta}^{(j)})]\}\\
&= \exp\{-\frac{1}{2}[\ (\text{vec } \mu_{\beta}^{(j)})^T ((\Sigma^{B})^{-1} + (\Sigma^{\beta})^{-1})(\text{vec } \mu_{\beta}^{(j)})\\
&- 2[(H_{\text{ 1: n}} * X_j)^T (\Sigma^{B})^{-1} + (\text{vec } \beta^{(j)})^T (\Sigma^{\beta})^{-1}](\text{vec } \mu_{\beta}^{(j)})]\}
\end{aligned}
\tag{A.5}
$$

which is the same as the pdf of Normal (μ, Σ) where

$$\Lambda := \Sigma^{-1} = (\Sigma^B)^{-1} + (\Sigma^\beta)^{-1}$$

$$\begin{aligned}
\mu &= \Sigma[(\Sigma^B)^{-1} H \text{ 1: n } * X_j + (\Sigma^\beta)^{-1}\text{vec } \beta^j] \\
&= ((\Sigma^B)^{-1} + (\Sigma^\beta)^{-1})^{-1}[(\Sigma^B)^{-1} H \text{ 1: n} * X_j + (\Sigma^\beta)^{-1}\text{vec } \beta^j] \\
&= \Lambda^{-1}[\Lambda^B H \text{ 1: n} * X_j + \Lambda^\beta \text{vec } \beta^j] \\
&= \frac{\Lambda^B}{\Lambda^B + \Lambda^\beta} H_{\text{ 1: n}} * X_j + \frac{\Lambda^\beta}{\Lambda^B + \Lambda^\beta} \text{ vec } \beta^{(j)}
\end{aligned} \tag{A.6}$$

$\forall$ customer j

A.5 Full conditional PDF of Σ^β

$$\begin{aligned}
&\text{f}(\ \Sigma^\beta |\text{ everything else}) \propto \text{f}(\ \Sigma^\beta, \text{ everything else}) \\
&\propto \text{f}(\ \Sigma^\beta | \nu_B, V_B\ ) * \text{f}(\text{ vec } \beta^{(j)} | \mu_\beta^{(j)}, \Sigma^\beta) \\
&= 2^{\frac{\nu_B * n^2}{2}} \pi^{\frac{n^2(n^2-1)}{4}} \prod_{\text{i=1}}^{n^2} \Gamma\left(\frac{\nu_B - 1 - i}{2}\right)^{-1} \times |V_B|^{\frac{\nu_B}{2}} |\Sigma^\beta|^{\frac{-(\nu_B + n^2 + 1)}{2}} \exp\left\{\text{tr}\left(\frac{-1}{2} V_B (\Sigma^\beta)^{-1}\right)\right\} \\
&\frac{1}{(2\pi)^{\frac{nm}{2}} |\Sigma^\beta|^{\frac{m}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^m \left[(\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})^T (\Sigma^\beta)^{-1} (\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})\right]\right\} \\
&\propto |\Sigma^\beta|^{\frac{-(\nu_B + n^2 + 1)}{2}} \times \exp\left\{-\text{tr}\left(\frac{V_B(\Sigma^\beta)^{-1}}{2}\right)\right\} \times |\Sigma^\beta|^{-\frac{m}{2}} \exp\left\{-\text{tr}\left(\frac{\sum_{j=1}^m [S_\beta^{(j)} (\Sigma^\beta)^{-1}]}{2}\right)\right\} \\
&\propto |\Sigma^\beta|^{\frac{-(\nu_B + n^2 + 1 + m)}{2}} \exp\left\{-\text{tr}\left(\frac{(V_B + \sum_{j=1}^m S_\beta^{(j)})(\Sigma^\beta)^{-1}}{2}\right)\right\}
\end{aligned}$$

which is the pdf of Inverted Wishart $(\nu_B + m, V_B + \sum_{j=1}^m S_\beta^{(j)})$

$$S_\beta^{(j)} := (\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})(\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})^T \tag{A.7}$$

A.6 Full conditional PDF of Σ^B

$$
\begin{aligned}
&\text{f}(\ \Sigma^B|\text{ everything else}) \propto \text{f}(\ \Sigma^B,\text{ everything else})\\
&\propto \text{f}(\ \Sigma^B|\nu_{BB}, V_{BB}) \prod_{j=1}^{m} \text{f}(\text{ vec } \mu_\beta^{(j)}|X_j\ ,\ H_{\text{ 1: n}}\ ,\ \Sigma^B)\\
&= 2^{\frac{\nu_{BB}*n^2}{2}} \pi^{\frac{n^2(n^2-1)}{4}} \prod_{\text{i}=1}^{n^2} \Gamma\left(\frac{\nu_{BB}-1-i}{2}\right)^{-1} \times |V_{BB}|^{\frac{\nu_{BB}}{2}} |\Sigma^B|^{\frac{-(\nu_{BB}+n^2+1)}{2}} \exp\left\{\text{tr}\left(\frac{-1}{2} V_{BB}(\Sigma^B)^{-1}\right)\right\}\\
&\frac{1}{(2\pi)^{\frac{nm}{2}}|\Sigma^B|^{\frac{m}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\left[(\text{vec } \mu_\beta^{(j)} - H_{\text{ 1: n}} * X_j)^T (\Sigma^B)^{-1} (\text{vec } \mu_\beta^{(j)} - H_{\text{ 1: n}} * X_j)\right]\right\}\\
&\propto |\Sigma^B|^{\frac{-(\nu_{BB}+n^2+1)}{2}} \times \exp\left\{-\text{tr}\left(\frac{V_{BB}(\Sigma^B)^{-1}}{2}\right)\right\} \times\\
&|\Sigma^B|^{-\frac{m}{2}} \exp\left\{-\text{tr}\left(\frac{\sum_{j=1}^{m}[S_B^{(j)}(\Sigma^B)^{-1}]}{2}\right)\right\}\\
&\propto |\Sigma^B|^{\frac{-(\nu_{BB}+n^2+1+m)}{2}} \exp\left\{-\text{tr}\left(\frac{(V_{BB}+\sum_{j=1}^{m} S_B^{(j)})(\Sigma^B)^{-1}}{2}\right)\right\}\\
&\text{which is the pdf of Inverted Wishart } (\nu_{BB}+m, V_{BB}+\sum_{j=1}^{m} S_B^{(j)})\\
&S_B^{(j)} := (\text{vec } \mu_\beta^{(j)} - H_{\text{ 1: n}} * X_j)(\text{vec } \mu_\beta^{(j)} - H_{\text{ 1: n}} * X_j)^T
\end{aligned}
\tag{A.8}
$$

A.7 Full conditional PDF of μ_H

$$
\begin{aligned}
&\text{f} (\ \mu_H |\ \text{everything else}) \propto \text{f} (\ \mu_H, \text{everything else}) \\
&\propto \text{f} (\ \text{vec}\ \mu_H | \mu_{HH}, \Sigma^{HH}) \times \prod_{i=1}^{n} \text{f} (\ \text{vec}\ H_i | \mu_H, \Sigma^H) \\
&= \frac{1}{(2\pi)^{\frac{nf}{2}} |\Sigma^{HH}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec}\ \mu_H - \mu_{HH})^T (\Sigma^{HH})^{-1} (\text{vec}\ \mu_H - \mu_{HH})\} \\
&\frac{1}{(2\pi)^{\frac{n^2 f}{2}} |\Sigma^H|^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \left[(\text{vec}\ H_i -\ \text{vec}\ \mu_H)^T (\Sigma^H)^{-1} (\text{vec}\ H_i -\ \text{vec}\ \mu_H) \right] \right\} \\
&\propto \exp\{-\frac{1}{2}[\ (\text{vec}\ \mu_H)^T (\Sigma^{HH})^{-1} (\ \text{vec}\ \mu_H) - 2(\mu_{HH})^T (\Sigma^{HH})^{-1} (\ \text{vec}\ \mu_H) \\
&+ \text{n}[(\ \text{vec}\ \mu_H)^T (\Sigma^H)^{-1} (\ \text{vec}\ \mu_H)] - 2 \left[\sum_{i=1}^{n} (\ \text{vec}\ H_i)^T (\Sigma^H)^{-1} (\ \text{vec}\ \mu_H) \right] \} \\
&= \exp\{-\frac{1}{2}[(\ \text{vec}\ \mu_H)^T ((\Sigma^{HH})^{-1} + \text{n}(\Sigma^H)^{-1})(\ \text{vec}\ \mu_H) \\
&- 2 \left((\mu_{HH})^T (\Sigma^{HH})^{-1} + \sum_{i=1}^{n} (\ \text{vec}\ H_i)^T (\Sigma^H)^{-1} \right) (\ \text{vec}\ \mu_H)]\}
\end{aligned}
$$

which is the same as the pdf of Normal (μ_{HHH}, Σ^{HHH}) where

$$
\begin{aligned}
&\Lambda^{HHH} := (\Sigma^{HHH})^{-1} = (\Sigma^{HH})^{-1} + \text{n}(\Sigma^H)^{-1} \\
&\mu_{HHH} = \Sigma^{HHH}[(\Sigma^{HH})^{-1}\mu_{HH} + \sum_{i=1}^{n}[(\Sigma^H)^{-1}\text{vec}\ H_i]] \\
&= ((\Sigma^{HH})^{-1} + \text{n}(\Sigma^H)^{-1})^{-1}[(\Sigma^{HH})^{-1}\mu_{HH} + \sum_{i=1}^{n}[(\Sigma^H)^{-1}\text{vec}\ H_i]] \\
&= (\Lambda^{HHH})^{-1}[(\Lambda^{HH})\mu_{HH} + \sum_{i=1}^{n}(\Lambda^H \text{vec}\ H_i)]
\end{aligned}
\tag{A.9}
$$

A.8 Full conditional PDF of Σ^H

$$
\begin{aligned}
&\text{f}(\ \Sigma^H|\text{ everything else}) \propto \text{f}(\ \Sigma^H, \text{ everything else})\\
&\propto \text{f}(\ \Sigma^H|\nu_{HH}, V_{HH}\ ) \times \text{f}(\text{ vec } H_i|\mu_H, \Sigma^H)\\
&= 2^{\frac{\nu_{HH}*nf}{2}} \pi^{\frac{nf(nf-1)}{4}} \prod_{\mathrm{i=1}}^{nf} \Gamma\left(\frac{\nu_{HH}-1-i}{2}\right)^{-1} \times |V_{HH}|^{\frac{\nu_{HH}}{2}} \times\\
&|\Sigma^H|^{\frac{-(\nu_{HH}+nf+1)}{2}} \exp\left\{\text{tr}\left(\frac{-1}{2}V_{HH}(\Sigma^H)^{-1}\right)\right\}\\
&\frac{1}{(2\pi)^{\frac{n^2f}{2}}|\Sigma^H|^{\frac{n}{2}}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}\left[(\text{vec } H_i - \text{ vec } \mu_H)^T(\Sigma^H)^{-1}(\text{vec } H_i - \text{ vec } \mu_H)\right]\}\\
&\propto |\Sigma^H|^{\frac{-(\nu_{HH}+nf+1)}{2}} \times\\
&\exp\left\{-\text{tr}\left(\frac{V_{HH}(\Sigma^H)^{-1}}{2}\right)\right\} \times |\Sigma^H|^{-\frac{n}{2}} \exp\left\{-\text{tr}\left(\frac{\sum_{i=1}^{n}[S_H^{(i)}(\Sigma^H)^{-1}]}{2}\right)\right\}\\
&\propto |\Sigma^H|^{\frac{-(\nu_{HH}+nf+1+n)}{2}} \exp\left\{-\text{tr}\left(\frac{(V_{HH}+\sum_{i=1}^{n} S_H^{(i)})(\Sigma^H)^{-1}}{2}\right)\right\}\\
&\text{which is the pdf of Inverted Wishart } (\nu_{HH}+n, V_{HH}+\sum_{i=1}^{n} S_H^{(i)})\\
&S_H^{(i)} := (\text{vec } H_i - \text{vec } \mu_H)(\text{vec } H_i - \text{vec } \mu_H)^T
\end{aligned}
\tag{A.10}
$$

A.9 Full conditional PDF of $\mu_\gamma^{(j)}$

$$
\begin{aligned}
&\text{f(vec } \mu_\gamma^{(j)} | \text{ everything else (including q)) } \propto \text{ f(vec } \mu_\gamma^{(j)} \text{ , everything else)} \\
&\propto \text{ f(vec } \mu_\gamma^{(j)} | X_j \text{ , } F_{\text{ 1: n}} \text{ , } \Sigma^G \text{) } \times \text{ f (vec } \gamma^{(j)} | \text{ vec } \mu_\gamma^{(j)}, \Sigma^\gamma \text{)} \\
&= \frac{1}{(2\pi)^{\frac{n^2}{2}} |\Sigma^G|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \mu_\gamma^{(j)} - F_{\text{ 1: n}} X_j)^T (\Sigma^G)^{-1} (\text{vec } \mu_\gamma^{(j)} - F_{\text{ 1: n}} X_j)\} \\
&\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma^\gamma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \gamma^{(j)} - \text{vec } \mu_\gamma^{(j)})^T (\Sigma^\gamma)^{-1} (\text{vec } \gamma^{(j)} - \text{vec } \mu_\gamma^{(j)})\} \\
&\propto \exp\{-\frac{1}{2}[\text{ (vec} \mu_\gamma^{(j)})^T (\Sigma^G)^{-1} \text{ (vec } \mu_\gamma^{(j)}) - 2(F_{\text{ 1: n}} X_j)^T (\Sigma^G)^{-1} (\text{vec } \mu_\gamma^{(j)}) \\
&+ \text{ (vec} \mu_\gamma^{(j)})^T (\Sigma^\gamma)^{-1} (\text{ vec } \mu_\gamma^{(j)}) - 2(\text{ vec } \gamma^{(j)})^T (\Sigma^\gamma)^{-1} (\text{ vec} \mu_\gamma^{(j)})]\} \\
&= \exp\{-\frac{1}{2}[\text{ (vec } \mu_\gamma^{(j)})^T ((\Sigma^G)^{-1} + (\Sigma^\gamma)^{-1}) (\text{vec } \mu_\gamma^{(j)}) \\
&- 2[(F_{\text{ 1: n}} X_j)^T (\Sigma^G)^{-1} + (\text{vec } \gamma^{(j)})^T (\Sigma^\gamma)^{-1}] (\text{vec } \mu_\gamma^{(j)})]\} \\
&\text{which is the same as the pdf of Normal (} \mu\mu, \Sigma\Sigma) \text{ where} \\
&\Lambda\Lambda := \Sigma\Sigma^{-1} = (\Sigma^G)^{-1} + (\Sigma^\gamma)^{-1} \\
&\mu\mu = \Sigma\Sigma[(\Sigma^G)^{-1} F_{\text{ 1: n}} \ X_j + (\Sigma^\gamma)^{-1} \text{vec } \gamma^j] \\
&= ((\Sigma^G)^{-1} + (\Sigma^\gamma)^{-1})^{-1} [(\Sigma^G)^{-1} F_{\text{ 1: n}} X_j + (\Sigma^\gamma)^{-1} \text{vec } \gamma^j] \\
&= \Lambda\Lambda^{-1} [\Lambda^\mu F_{\text{ 1: n}} X_j + \Lambda^\gamma \text{vec } \gamma^j] \\
&= \frac{\Lambda^\mu}{\Lambda^\mu + \Lambda^\gamma} F_{\text{ 1: n}} X_j + \frac{\Lambda^\gamma}{\Lambda^\mu + \Lambda^\gamma} \text{ vec } \gamma^{(j)} \\
&\forall \text{ customer j}
\end{aligned}
\tag{A.11}
$$

A.10 Full conditional PDF of Σ^G

$$
\begin{aligned}
&\text{f}(\ \Sigma^G|\text{ everything else}) \propto \text{f}(\ \Sigma^G,\text{ everything else})\\
&\propto \text{f}(\ \Sigma^G|\nu_{GG}, V_{GG}) \prod_{j=1}^{m} \text{f}(\ \text{vec}\ \mu_\gamma^{(j)}|X_j\ ,\ F_{\text{1: n}}\ ,\ \Sigma^G)\\
&= 2^{\frac{\nu_{GG}*n^2}{2}} \pi^{\frac{n^2(n^2-1)}{4}} \prod_{\text{i}=1}^{n^2} \Gamma\left(\frac{\nu_{GG}-1-i}{2}\right)^{-1} \times |V_{GG}|^{\frac{\nu_{GG}}{2}} \times\\
&|\Sigma^G|^{\frac{-(\nu_{GG}+n^2+1)}{2}} \exp\left\{\text{tr}\left(\frac{-1}{2} V_{GG}(\Sigma^G)^{-1}\right)\right\}\\
&\frac{1}{(2\pi)^{\frac{nm}{2}} |\Sigma^G|^{\frac{m}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\left[(\text{vec}\ \mu_\gamma^{(j)} - F_{\text{1: n}} X_j)^T (\Sigma^G)^{-1} (\text{vec}\ \mu_\gamma^{(j)} - F_{\text{1: n}} X_j)\right]\right\}\\
&\propto |\Sigma^G|^{\frac{-(\nu_{GG}+n^2+1)}{2}} \times \exp\left\{-\text{tr}\left(\frac{V_{GG}(\Sigma^G)^{-1}}{2}\right)\right\} \times\\
&|\Sigma^G|^{-\frac{m}{2}} \exp\left\{-\text{tr}\left(\frac{\sum_{j=1}^{m}[S_\mu^{(j)}(\Sigma^G)^{-1}]}{2}\right)\right\}\\
&\propto |\Sigma^G|^{\frac{-(\nu_{GG}+n^2+1+m)}{2}} \exp\left\{-\text{tr}\left(\frac{(V_{GG}+\sum_{j=1}^{m} S_\mu^{(j)})(\Sigma^G)^{-1}}{2}\right)\right\}\\
&\text{which is the pdf of Inverted Wishart } (\nu_{GG}+m, V_{GG}+\sum_{j=1}^{m} S_\mu^{(j)})\\
&S_\mu^{(j)} := (\text{vec}\ \mu_\gamma^{(j)} - F_{\text{1: n}} X_j)(\text{vec}\ \mu_\gamma^{(j)} - F_{\text{1: n}} X_j)^T
\end{aligned}
\tag{A.12}
$$

A.11 Full conditional PDF of Σ^γ

$$
\begin{aligned}
&\text{f}(\ \Sigma^\gamma |\text{ everything else}) \propto \text{f}(\ \Sigma^\gamma, \text{ everything else})\\
&\propto \text{f}(\ \Sigma^\gamma|\nu_G, V_G\ ) \times \text{f}(\ \text{vec }\gamma^{(j)}|\mu_\gamma^{(j)}, \Sigma^\gamma)\\
&= 2^{\frac{\nu_G * n^2}{2}} \pi^{\frac{n^2(n^2-1)}{4}} \prod_{\text{i}=1}^{n^2} \Gamma\left(\frac{\nu_G - 1 - i}{2}\right)^{-1} \times |V_G|^{\frac{\nu_G}{2}} \times\\
&|\Sigma^\gamma|^{\frac{-(\nu_G+n^2+1)}{2}} \exp\left\{\text{tr}\left(\frac{-1}{2} V_G (\Sigma^\gamma)^{-1}\right)\right\}\\
&\frac{1}{(2\pi)^{\frac{nm}{2}}|\Sigma^\gamma|^{\frac{m}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\left[(\text{vec }\gamma^{(j)} - \text{vec }\mu_\gamma^{(j)})^T(\Sigma^\gamma)^{-1}(\text{vec }\gamma^{(j)} - \text{vec }\mu_\gamma^{(j)})\right]\right\}\\
&\propto |\Sigma^\gamma|^{\frac{-(\nu_G+n^2+1)}{2}} \times \exp\left\{-\text{tr}\left(\frac{V_G(\Sigma^\gamma)^{-1}}{2}\right)\right\} \times\\
&|\Sigma^\gamma|^{-\frac{m}{2}} \exp\left\{-\text{tr}\left(\frac{\sum_{j=1}^m [S_\gamma^{(j)}(\Sigma^\gamma)^{-1}]}{2}\right)\right\}\\
&\propto |\Sigma^\gamma|^{\frac{-(\nu_G+n^2+1+m)}{2}} \exp\left\{-\text{tr}\left(\frac{(V_G + \sum_{j=1}^m S_\gamma^{(j)})(\Sigma^\gamma)^{-1}}{2}\right)\right\}\\
&\text{which is the pdf of Inverted Wishart } (\nu_G + m, V_G + \sum_{j=1}^{m} S_\gamma^{(j)})\\
&S_\gamma^{(j)} := (\text{vec }\gamma^{(j)} - \text{vec }\mu_\gamma^{(j)})(\text{vec }\gamma^{(j)} - \text{vec }\mu_\gamma^{(j)})^T
\end{aligned}
\tag{A.13}
$$

A.12 Full conditional PDF of μ_F

$$
\begin{aligned}
&\text{f (} \mu_F| \text{ everything else}) \propto \text{f (} \mu_F, \text{ everything else})\\
&\propto \text{f (vec } \mu_F|\mu_{FF}, \Sigma^{FF}) \times \prod_{i=1}^{n} \text{f (vec } F_i|\mu_F, \Sigma^F)\\
&= \frac{1}{(2\pi)^{\frac{nf}{2}}|\Sigma^{FF}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \mu_F - \mu_{FF})^T(\Sigma^{FF})^{-1}(\text{vec } \mu_F - \mu_{FF})\}\\
&\frac{1}{(2\pi)^{\frac{n^2 f}{2}}|\Sigma^F|^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[(\text{vec } F_i - \text{ vec } \mu_F)^T(\Sigma^F)^{-1}(\text{vec } F_i - \text{ vec } \mu_F)\right]\right\}\\
&\propto \exp\{-\frac{1}{2}[\text{ (vec } \mu_F)^T(\Sigma^{FF})^{-1}(\text{ vec } \mu_F) - 2(\mu_{FF})^T(\Sigma^{FF})^{-1}(\text{ vec } \mu_F)\\
&+ \text{n}[(\text{ vec } \mu_F)^T(\Sigma^F)^{-1}(\text{ vec } \mu_F)] - 2\left[\sum_{i=1}^{n}(\text{ vec } F_i)^T(\Sigma^F)^{-1}(\text{ vec } \mu_F)\right]\}\\
&= \exp\{-\frac{1}{2}[(\text{ vec } \mu_F)^T((\Sigma^{FF})^{-1} + \text{n}(\Sigma^F)^{-1})(\text{ vec } \mu_F)\\
&- 2\left((\mu_{FF})^T(\Sigma^{FF})^{-1} + \sum_{i=1}^{n}(\text{ vec } F_i)^T(\Sigma^F)^{-1}\right)(\text{ vec } \mu_F)]\}
\end{aligned}
$$

which is the same as the pdf of Normal (μ_{FFF}, Σ^{FFF}) where

$$
\begin{aligned}
&\Lambda^{FFF} := (\Sigma^{FFF})^{-1} = (\Sigma^{FF})^{-1} + \text{n}(\Sigma^F)^{-1}\\
&\mu_{FFF} = \Sigma^{FFF}[(\Sigma^{FF})^{-1}\mu_{FF} + \sum_{i=1}^{n}[(\Sigma^F)^{-1}\text{vec } F_i]]\\
&= ((\Sigma^{FF})^{-1} + \text{n}(\Sigma^F)^{-1})^{-1}[(\Sigma^{FF})^{-1}\mu_{FF} + \sum_{i=1}^{n}[(\Sigma^F)^{-1}\text{vec } F_i]]\\
&= (\Lambda^{FFF})^{-1}[(\Lambda^{FF})\mu_{FF} + \sum_{i=1}^{n}(\Lambda^F\text{vec } F_i)]
\end{aligned}
\tag{A.14}
$$

A.13 Full conditional PDF of Σ^F

$$
\begin{aligned}
&\text{f}(\ \Sigma^F|\text{ everything else}) \propto \text{f}(\ \Sigma^F,\text{ everything else})\\
&\propto \text{f}(\ \Sigma^F|\nu_{FF}, V_{FF}\ ) \times \text{f}(\text{ vec } F_i|\mu_F, \Sigma^F)\\
&= 2^{\frac{\nu_{FF}*nf}{2}}\pi^{\frac{nf(nf-1)}{4}}\prod_{\text{i=1}}^{nf}\Gamma\left(\frac{\nu_{FF}-1-i}{2}\right)^{-1}\times|V_{FF}|^{\frac{\nu_{FF}}{2}}\times\\
&|\Sigma^F|^{\frac{-(\nu_{FF}+nf+1)}{2}}\exp\left\{\text{tr}\left(\frac{-1}{2}V_{FF}(\Sigma^F)^{-1}\right)\right\}\\
&\frac{1}{(2\pi)^{\frac{n^2 f}{2}}|\Sigma^F|^{\frac{n}{2}}}\exp\{-\frac{1}{2}\sum_{i=1}^{n}\left[(\text{vec } F_i - \text{ vec } \mu_F)^T(\Sigma^F)^{-1}(\text{vec } F_i - \text{ vec } \mu_F)\right]\}\\
&\propto |\Sigma^F|^{\frac{-(\nu_{FF}+nf+1)}{2}}\times\exp\left\{-\text{tr}\left(\frac{V_{FF}(\Sigma^F)^{-1}}{2}\right)\right\}\times\\
&|\Sigma^F|^{-\frac{n}{2}}\exp\left\{-\text{tr}\left(\frac{\sum_{i=1}^{n}[S_F^{(i)}(\Sigma^F)^{-1}]}{2}\right)\right\}\\
&\propto |\Sigma^F|^{\frac{-(\nu_{FF}+nf+1+n)}{2}}\exp\left\{-\text{tr}\left(\frac{(V_{FF}+\sum_{i=1}^{n}S_F^{(i)})(\Sigma^F)^{-1}}{2}\right)\right\}\\
&\text{which is the pdf of Inverted Wishart } (\nu_{FF}+n, V_{FF}+\sum_{i=1}^{n}S_F^{(i)})\\
&S_F^{(i)} := (\text{vec } F_i - \text{vec } \mu_F)(\text{vec } F_i - \text{vec } \mu_F)^T
\end{aligned}
\tag{A.15}
$$

A.14 Full conditional PDF of $\beta^{(j)}$

$$
\begin{aligned}
&\text{f (} \beta^{(j)} \text{ — everything else)} \propto \text{f(} \beta^{(j)}\text{, everything else)}\\
&\propto \text{f(vec}\beta^{(j)} |\ \text{vec } \mu_\beta^{(j)}, \Sigma^\beta) \prod_{t=1}^{T}\prod_{i=1}^{n} \text{f(} q_{it}^{(j)} \,|\, \alpha, \beta^{(j)}, \gamma^{(j)}, p_{it}^{(j)})\\
&= \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma^\beta|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})^T(\Sigma^\beta)^{-1}(\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})\}\times\\
&\prod_{t=1}^{T}\prod_{i=1}^{n} \frac{[\lambda_{ijt}]^{q_{it}^{(j)}} \exp\{-\lambda_{ijt}\}}{q_{it}^{(j)}}\\
&= \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma^\beta|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})^T(\Sigma^\beta)^{-1}(\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})\}\\
&\prod_{t=1}^{T}\prod_{i=1}^{n} \frac{\left[\exp\{(p_{\cdot\text{t}}^{(j)})^T\beta_i^{(j)} + (q_{\cdot\text{t-1}}^{(j)})^T\gamma_i^{(j)} + (s^t)^T\alpha\}\right]^{q_{it}^{(j)}}}{q_{it}^{(j)}!}\times\\
&\exp\left\{-\exp\{(p_{\cdot\text{t}}^{(j)})^T\beta_i^{(j)} + (q_{\cdot\text{t-1}}^{(j)})^T\gamma_i^{(j)} + (s^t)^T\alpha\}\right\}\\
&\propto \exp\{-\frac{1}{2}(\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})^T(\Sigma^\beta)^{-1}(\text{vec } \beta^{(j)} - \text{vec } \mu_\beta^{(j)})\}\\
&\prod_{t=1}^{T}\prod_{i=1}^{n} \left[\exp\{(p_{\cdot\text{t}}^{(j)})^T\beta_i^{(j)} + (q_{\cdot\text{t-1}}^{(j)})^T\gamma_i^{(j)} + (s^t)^T\alpha\}\right]^{q_{it}^{(j)}}\times\\
&\exp\left\{-\exp\{(p_{\cdot\text{t}}^{(j)})^T\beta_i^{(j)} + (q_{\cdot\text{t-1}}^{(j)})^T\gamma_i^{(j)} + (s^t)^T\alpha\}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[(\text{vec } \beta^{(j)})^T(\Sigma^\beta)^{-1}(\text{ vec } \beta^{(j)}) - 2(\text{vec } \beta^{(j)})^T(\Sigma^\beta)^{-1}(\text{ vec } \mu_\beta^{(j)})\right]\right\}\\
&\prod_{t=1}^{T}\prod_{i=1}^{n} \left[\exp\{(p_{\cdot\text{t}}^{(j)})^T\beta_i^{(j)} + (q_{\cdot\text{t-1}}^{(j)})^T\gamma_i^{(j)} + (s^t)^T\alpha\}\right]^{q_{it}^{(j)}}\times\\
&\exp\left\{-\exp\{(p_{\cdot\text{t}}^{(j)})^T\beta_i^{(j)} + (q_{\cdot\text{t-1}}^{(j)})^T\gamma_i^{(j)} + (s^t)^T\alpha\}\right\}\\
&\forall \text{ customer j}
\end{aligned}
\tag{A.16}
$$

A.15 Full conditional PDF of $\gamma^{(j)}$

$$
\begin{aligned}
&\text{f (} \gamma^{(j)} \text{ — everything else)} \propto \text{f(} \gamma^{(j)}\text{, everything else)}\\
&\propto \text{f(vec } \gamma^{(j)} | \text{ vec } \mu_\gamma^{(j)}, \Sigma^\gamma) \prod_{t=1}^{T} \prod_{i=1}^{n} \text{f(} q_{it}^{(j)} \,|\, \alpha, \beta^{(j)}, \gamma^{(j)}, p_{it}^{(j)})\\
&= \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma^\gamma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \gamma^{(j)} - \text{vec } \mu_\gamma^{(j)})^T (\Sigma^\gamma)^{-1} (\text{vec } \gamma^{(j)} - \text{vec } \mu_\gamma^{(j)})\} \times\\
&\prod_{t=1}^{T} \prod_{i=1}^{n} \frac{[\lambda_{ijt}]^{q_{it}^{(j)}} \exp\{-\lambda_{ijt}\}}{q_{it}^{(j)}}\\
&= \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma^\gamma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\text{vec } \gamma^{(j)} - \text{vec } \mu_\gamma^{(j)})^T (\Sigma^\gamma)^{-1} (\text{vec } \gamma^{(j)} - \text{vec } \mu_\gamma^{(j)})\}\\
&\prod_{t=1}^{T} \prod_{i=1}^{n} \frac{\left[\exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\}\right]^{q_{it}^{(j)}}}{q_{it}^{(j)}!} \times\\
&\exp\left\{- \exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\}\right\}\\
&\propto \exp\{-\frac{1}{2}(\text{vec } \gamma^{(j)} - \text{vec } \mu_\gamma^{(j)})^T (\Sigma^\gamma)^{-1} (\text{vec } \gamma^{(j)} - \text{vec } \mu_\gamma^{(j)})\}\\
&\prod_{t=1}^{T} \prod_{i=1}^{n} \left[\exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\}\right]^{q_{it}^{(j)}} \times\\
&\exp\left\{- \exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\}\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[(\text{vec } \gamma^{(j)})^T (\Sigma^\gamma)^{-1} (\text{ vec } \gamma^{(j)}) - 2(\text{vec } \gamma^{(j)})^T (\Sigma^\gamma)^{-1} (\text{ vec } \mu_\gamma^{(j)})\right]\right\}\\
&\prod_{t=1}^{T} \prod_{i=1}^{n} \left[\exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\}\right]^{q_{it}^{(j)}}\\
&\times \exp\left\{- \exp\{(p_{\cdot \text{t}}^{(j)})^T \beta_i^{(j)} + (q_{\cdot \text{t-1}}^{(j)})^T \gamma_i^{(j)} + (s^t)^T \alpha\}\right\}\\
&\forall \text{ customer j}
\end{aligned}
\tag{A.17}
$$

A.16 Full conditional joint PDF of $H_1 \cdots H_n$

$$
\begin{aligned}
&\text{f (} H_i \text{ — everything else)} \propto \text{f (} H_i \text{ , everything else)}\\
&\propto \prod_{j=1}^{m} \text{f (vec } \mu_\beta^{(j)} | H_{\text{ 1:n}}, X_j, \Sigma^B) \times \prod_{i=1}^{n} \text{f (vec } H_i | \mu_H, \Sigma^H)\\
&= \frac{1}{(2\pi)^{\frac{mn^2}{2}} |\Sigma^B|^{\frac{m}{2}}} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{m} (\text{ vec } \mu_\beta^{(j)} - H_{\text{ 1:n}} X_j)^T (\Sigma^B)^{-1} (\text{ vec } \mu_\beta^{(j)} - H_{\text{ 1:n}} X_j) \right\}\\
&\frac{1}{(2\pi)^{\frac{n^2 f}{2}} |\Sigma^H|^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \left[(\text{vec } H_i - \text{ vec } \mu_H)^T (\Sigma^H)^{-1} (\text{vec } H_i - \text{ vec } \mu_H) \right] \right\}\\
&\propto \exp\{-\frac{1}{2}\{\sum_{j=1}^{m} \begin{bmatrix} \mu_\beta^{(j),1} - H_1 X_j \\ \mu_\beta^{(j),2} - H_2 X_j \\ \vdots \\ \mu_\beta^{(j),n} - H_n X_j \end{bmatrix}^T \begin{bmatrix} \Lambda_{11}^B & \cdots & \Lambda_{1n}^B \\ \vdots & \Lambda_{ik}^B & \vdots \\ \Lambda_{n1}^B & \cdots & \Lambda_{nn}^B \end{bmatrix} \begin{bmatrix} \mu_\beta^{(j),1} - H_1 X_j \\ \mu_\beta^{(j),2} - H_2 X_j \\ \vdots \\ \mu_\beta^{(j),n} - H_n X_j \end{bmatrix}\\
&+ \sum_{i=1}^{n} \left(\begin{bmatrix} H_i^1 \\ H_i^2 \\ \vdots \\ H_i^f \end{bmatrix} - \begin{bmatrix} \mu_H^1 \\ \mu_H^2 \\ \vdots \\ \mu_H^f \end{bmatrix} \right)^T \begin{bmatrix} \Lambda_{11}^H & \cdots & \Lambda_{1f}^H \\ \vdots & \Lambda_{lr}^H & \vdots \\ \Lambda_{f1}^H & \cdots & \Lambda_{ff}^H \end{bmatrix} \left(\begin{bmatrix} H_i^1 \\ H_i^2 \\ \vdots \\ H_i^f \end{bmatrix} - \begin{bmatrix} \mu_H^1 \\ \mu_H^2 \\ \vdots \\ \mu_H^f \end{bmatrix} \right) \}\}\\
&= \exp\{-\frac{1}{2}[\sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{n} (\mu_\beta^{(j),i} - H_i X_j)^T \Lambda_{ik}^B (\mu_\beta^{(j),k} - H_k X_j)\\
&+ \sum_{i=1}^{n} [(\text{vec } H_i - \text{ vec } \mu_H)^T (\Sigma^H)^{-1} (\text{vec } H_i - \text{ vec } \mu_H)]]\}
\end{aligned}
\tag{A.18}
$$

$$
\begin{aligned}
&= \exp\{-\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}[(\mu_{\beta}^{(j),i} - \sum_{l=1}^{f} x_j^l\, H_i^l)^T \Lambda_{ik}^B\, (\mu_{\beta}^{(j),k} - \sum_{r=1}^{f} x_j^r H_k^r)] \\
&+ \sum_{i=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}[(H_i^l - \mu_H^l)^T \Lambda_{lr}^H (H_i^r - \mu_H^r)]\} \\
&\propto \exp\{-\frac{1}{2}\{\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f} x_j^l x_j^r (H_i^l)^T \Lambda_{ik}^B H_k^r - 2\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{r=1}^{f} x_j^r (\mu_{\beta}^{(j),i})^T \Lambda_{ik}^B H_k^r \\
&+ \sum_{i=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(H_i^l)^T \Lambda_{lr}^H H_i^r - 2\sum_{i=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(\mu_H^l)^T \Lambda_{lr}^H H_i^r\}\} \\
&= \exp\{-\frac{1}{2}\{\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(H_i^l)^T [\sum_{j=1}^{m} x_j^l x_j^r] \Lambda_{ik}^B H_k^r - 2\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{r=1}^{f} x_j^r (\mu_{\beta}^{(j),i})^T \Lambda_{ik}^B H_k^r \\
&+ \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(H_i^l)^T \Lambda_{lr}^H \delta_{ik}\, H_i^r - 2\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(\mu_H^l)^T \Lambda_{lr}^H H_k^r\}\} \\
&= \exp\{-\frac{1}{2}\{\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(H_i^l)^T m\, C_{lr} \Lambda_{ik}^B H_k^r - 2\sum_{k=1}^{n}\sum_{r=1}^{f}[\sum_{j=1}^{m}\sum_{i=1}^{n} x_j^r (\mu_{\beta}^{(j),i})^T \Lambda_{ik}^B] H_k^r \\
&+ \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(H_i^l)^T \Lambda_{lr}^H \delta_{ik}\, H_i^r - 2\sum_{k=1}^{n}\sum_{r=1}^{f}[\sum_{l=1}^{f}(\mu_H^l)^T \Lambda_{lr}^H] H_k^r\}\} \\
&= \exp\{-\frac{1}{2}\{\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(H_i^l)^T \left[m\, C_{lr} \Lambda_{ik}^B + \delta_{ik} \Lambda_{lr}^H\right] H_k^r \\
&- 2\sum_{k=1}^{n}\sum_{r=1}^{f}\left[\sum_{j=1}^{m}\sum_{i=1}^{n} x_j^r (\mu_{\beta}^{(j),i})^T \Lambda_{ik}^B + \sum_{l=1}^{f}(\mu_H^l)^T \Lambda_{lr}^H\right] H_k^r\}\}
\end{aligned}
\tag{A.19}
$$

which is the same as the pdf of Normal (μ_H^H, Σ_H^H)

$\Lambda^H_{il,kr} := (\Sigma^H_{il,kr})^{-1} = m\, C_{lr}\Lambda^B_{ik} + \delta_{ik}\Lambda^H_{lr}$

$$\mu^H_{kr} = \Sigma^{Hi}[\sum_{j=1}^{m}\sum_{i=1}^{n} x_j^r \Lambda^B_{ik}\, \mu_\beta^{(j),i} + \sum_{l=1}^{f}\Lambda^H_{lr}\mu^l_H]$$

$$= (m\, C_{lr}\Lambda^B_{ik} + \delta_{ik}\Lambda^H_{lr})^{-1}[\sum_{j=1}^{m}\sum_{i=1}^{n} x_j^r \Lambda^B_{ik}\, \mu_\beta^{(j),i} + \sum_{l=1}^{f}\Lambda^H_{lr}\mu^l_H]$$

$$\begin{bmatrix} H_1^1 \\ H_1^2 \\ \vdots \\ H_1^f \\ \\ \vdots \\ \\ H_n^1 \\ H_n^2 \\ \vdots \\ H_n^f \end{bmatrix} \sim \text{ Normal } \left(\mu_H^H = \begin{bmatrix} \mu^H_{11} \\ \mu^H_{12} \\ \vdots \\ \mu^H_{1r} \\ \\ \vdots \\ \\ \mu^H_{k1} \\ \mu^H_{k2} \\ \vdots \\ \mu^H_{kr} \end{bmatrix}, \Lambda^H_H = \begin{bmatrix} \Lambda^H_{11,11}\Lambda^H_{11,12}\cdots\Lambda^H_{11,1f}\Lambda^H_{11,21}\cdots\Lambda^H_{11,nf} & & \\ & & \\ \vdots & \Lambda^H_{il,kr} & \vdots \\ & & \\ \Lambda^H_{nf,11} & \cdots & \Lambda^H_{nf,nf} \end{bmatrix} \right) \tag{A.20}$$

A.17 Full conditional joint PDF of $F_1, \cdots F_n$

f (F_i — everything else) $\propto$ f (F_i , everything else)

$$\propto \prod_{j=1}^{m} \text{f (vec } \mu_\gamma^{(j)} | F_{\text{1:n}}, X_j, \Sigma^G) \times \prod_{i=1}^{n} \text{f (vec } F_i | \mu_F, \Sigma^F)$$

$$= \frac{1}{(2\pi)^{\frac{mn^2}{2}} |\Sigma^G|^{\frac{m}{2}}} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{m} (\text{ vec } \mu_\gamma^{(j)} - F_{\text{1:n}} * X_j)^T (\Sigma^G)^{-1} (\text{ vec } \mu_\gamma^{(j)} - F_{\text{1:n}} * X_j) \right\}$$

$$\frac{1}{(2\pi)^{\frac{n^2 f}{2}} |\Sigma^F|^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \left[(\text{vec } F_i - \text{ vec } \mu_F)^T (\Sigma^F)^{-1} (\text{vec } F_i - \text{ vec } \mu_F) \right] \right\}$$

$$\propto \exp\{-\frac{1}{2}\{\sum_{j=1}^{m} \begin{bmatrix} \mu_\gamma^{(j),1} - F_1 X_j \\ \mu_\gamma^{(j),2} - F_2 X_j \\ \vdots \\ \mu_\gamma^{(j),n} - F_n X_j \end{bmatrix}^T \begin{bmatrix} \Lambda_{11}^\mu & \cdots & \Lambda_{1n}^\mu \\ \vdots & \Lambda_{ik}^\mu & \vdots \\ \Lambda_{n1}^\mu & \cdots & \Lambda_{nn}^\mu \end{bmatrix} \begin{bmatrix} \mu_\gamma^{(j),1} - F_1 X_j \\ \mu_\gamma^{(j),2} - F_2 X_j \\ \vdots \\ \mu_\gamma^{(j),n} - F_n X_j \end{bmatrix}$$

$$+ \sum_{i=1}^{n} \left(\begin{bmatrix} F_i^1 \\ F_i^2 \\ \vdots \\ F_i^f \end{bmatrix} - \begin{bmatrix} \mu_F^1 \\ \mu_F^2 \\ \vdots \\ \mu_F^f \end{bmatrix} \right)^T \begin{bmatrix} \Lambda_{11}^F & \cdots & \Lambda_{1f}^F \\ \vdots & \Lambda_{lr}^F & \vdots \\ \Lambda_{f1}^F & \cdots & \Lambda_{ff}^F \end{bmatrix} \left(\begin{bmatrix} F_i^1 \\ F_i^2 \\ \vdots \\ F_i^f \end{bmatrix} - \begin{bmatrix} \mu_F^1 \\ \mu_F^2 \\ \vdots \\ \mu_F^f \end{bmatrix} \right) \}\}$$

$$= \exp\{-\frac{1}{2}[\sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{n} (\mu_\gamma^{(j),i} - F_i X_j)^T \Lambda_{ik}^\mu (\mu_\gamma^{(j),k} - F_k X_j)$$

$$+ \sum_{i=1}^{n} [(\text{vec } F_i - \text{ vec } \mu_F)^T (\Sigma^F)^{-1} (\text{vec } F_i - \text{ vec } \mu_F)]]\} \tag{A.21}$$

$$
\begin{aligned}
&= \exp\{-\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}[(\mu_\gamma^{(j),i} - \sum_{l=1}^{f} x_j^l\, F_i^l)^T \Lambda_{ik}^{\mu}\, (\mu_\gamma^{(j),k} - \sum_{r=1}^{f} x_j^r F_k^r)] \\
&+ \sum_{i=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}[(F_i^l - \mu_F^l)^T \Lambda_{lr}^F (F_i^r - \mu_F^r)]\} \\
&\propto \exp\{-\frac{1}{2}\{\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f} x_j^l x_j^r (F_i^l)^T \Lambda_{ik}^{\mu}\, F_k^r - 2\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{r=1}^{f} x_j^r (\mu_\gamma^{(j),i})^T \Lambda_{ik}^{\mu}\, F_k^r \\
&+ \sum_{i=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(F_i^l)^T \Lambda_{lr}^F F_i^r - 2\sum_{i=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(\mu_F^l)^T \Lambda_{lr}^F F_i^r\}\} \\
&= \exp\{-\frac{1}{2}\{\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(F_i^l)^T [\sum_{j=1}^{m} x_j^l x_j^r] \Lambda_{ik}^{\mu} F_k^r - 2\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{r=1}^{f} x_j^r (\mu_\gamma^{(j),i})^T \Lambda_{ik}^{\mu} F_k^r \\
&+ \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(F_i^l)^T \Lambda_{lr}^F \delta_{ik}\, F_i^r - 2\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(\mu_F^l)^T \Lambda_{lr}^F F_k^r\}\} \\
&= \exp\{-\frac{1}{2}\{\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(F_i^l)^T m\, C_{lr} \Lambda_{ik}^{\mu} F_k^r - 2\sum_{k=1}^{n}\sum_{r=1}^{f}[\sum_{j=1}^{m}\sum_{i=1}^{n} x_j^r (\mu_\gamma^{(j),i})^T \Lambda_{ik}^{\mu}] F_k^r \\
&+ \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(F_i^l)^T \Lambda_{lr}^F \delta_{ik}\, F_i^r - 2\sum_{k=1}^{n}\sum_{r=1}^{f}[\sum_{l=1}^{f}(\mu_F^l)^T \Lambda_{lr}^F] F_k^r\}\} \\
&= \exp\{-\frac{1}{2}\{\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{f}\sum_{r=1}^{f}(F_i^l)^T \left[m\, C_{lr}\Lambda_{ik}^{\mu} + \delta_{ik}\Lambda_{lr}^F\right] F_k^r \\
&- 2\sum_{k=1}^{n}\sum_{r=1}^{f}\left[\sum_{j=1}^{m}\sum_{i=1}^{n} x_j^r (\mu_\gamma^{(j),i})^T \Lambda_{ik}^{\mu} + \sum_{l=1}^{f}(\mu_F^l)^T \Lambda_{lr}^F\right] F_k^r\}\}
\end{aligned}
\tag{A.22}
$$

which is the same as the pdf of Normal (μ_F^F, Σ_F^F) where

$$\Lambda_{il,kr}^F := (\Sigma_{il,kr}^F)^{-1} = m\, C_{lr} \Lambda_{ik}^{\mu} + \delta_{ik} \Lambda_{lr}^F$$

$$\mu_{kr}^F = \Sigma^{Fi}[\sum_{j=1}^{m} \sum_{i=1}^{n} x_j^r \Lambda_{ik}^{\mu}\, \mu_{\gamma}^{(j),i} + \sum_{l=1}^{f} \Lambda_{lr}^F \mu_F^l]$$

$$= (m\, C_{lr} \Lambda_{ik}^{\mu} + \delta_{ik} \Lambda_{lr}^F)^{-1}[\sum_{j=1}^{m} \sum_{i=1}^{n} x_j^r \Lambda_{ik}^{\mu}\, \mu_{\gamma}^{(j),i} + \sum_{l=1}^{f} \Lambda_{lr}^F \mu_F^l]$$

$$\begin{bmatrix} F_1^1 \\ F_1^2 \\ \vdots \\ F_1^f \\ \\ \vdots \\ \\ F_n^1 \\ F_n^2 \\ \vdots \\ F_n^f \end{bmatrix} \sim \text{ Normal } \left(\mu_F^F = \begin{bmatrix} \mu_{11}^F \\ \mu_{12}^F \\ \vdots \\ \mu_{1r}^F \\ \\ \vdots \\ \\ \mu_{k1}^F \\ \mu_{k2}^F \\ \vdots \\ \mu_{kr}^F \end{bmatrix}, \Lambda_F^F = \begin{bmatrix} \Lambda_{11,11}^F \Lambda_{11,12}^F \cdots \Lambda_{11,1f}^F \Lambda_{11,21}^F \cdots \Lambda_{11,nf}^F \\ \\ \vdots \qquad \Lambda_{il,kr}^F \qquad \vdots \\ \\ \Lambda_{nf,11}^F \qquad \cdots \qquad \Lambda_{nf,nf}^F \end{bmatrix} \right) \tag{A.23}$$

Appendix B

Calculations and Derivations for Preliminary Model I

B.1 Full conditional PDF of y_{self}

$$
\begin{aligned}
&\text{f}(y_{\text{self}}) \propto \text{f}(y_{\text{self}}|\text{ everything else}) \propto \text{f}(y_{\text{self}}\text{ , everything else})\\
&\propto \text{f}(\log(\text{-y}_{\text{self}})|\mu_{\text{self}} * 1_n\ ,\ \sigma^2_{\text{self}} * I)\left|\frac{\partial \log(-y_{\text{self}})}{\partial y_{\text{self}}}\right| \prod_{\text{i=1}}^{n} \text{f}(\ q_{it}\ |\ y_{\text{self}}, y^i_{cross}, y_0, p)\\
&\propto \frac{1}{|2\pi|^{\frac{n}{2}}|\sigma^2_{\text{self}} I|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\log(-y_{\text{self}}) - \mu_{\text{self}}1_n)^T(\sigma^2_{\text{self}}I)^{-1}(\log(-y_{\text{self}}) - \mu_{\text{self}}1_n)\}\times\\
&|\text{diag}(\ y_{\text{self}})^{-1}| \times (\prod_{i=1}^{n}\prod_{t=1}^{\text{Time}} \frac{\exp\{-\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p_{-i}}^t)^T y^i_{cross} + e_i^T y_0\}\}}{q_{it}!} \times\\
&\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p_{-i}}^t)^T y^i_{cross} + e_i^T y_0\}^{q_{it}})\\
&\propto \exp\{-\frac{1}{2}(\log(-y_{\text{self}}) - \mu_{\text{self}}1_n)^T(\sigma^2_{\text{self}}I)^{-1}(\log(-y_{\text{self}}) - \mu_{\text{self}}1_n)\}(-1)^n \left[\prod_{i=1}^{n} y^i_{\text{self}}\right]^{-1} \times\\
&(\prod_{i=1}^{n}\prod_{t=1}^{\text{Time}} \exp\{-\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p_{-i}}^t)^T y^i_{cross} + e_i^T y_0\}\}\times\\
&\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p_{-i}}^t)^T y^i_{cross} + e_i^T y_0\}^{q_{it}})
\end{aligned}
\tag{B.1}
$$

B.2 Loglikelihood of y_{self}

$$
\begin{aligned}
& l(y_{self}) = \log(f(y_{\text{self}})) \\
& \propto -\frac{1}{2}(\log(-y_{\text{self}}) - \mu_{\text{self}} 1_n)^T (\sigma^2_{\text{self}} I)^{-1} (\log(-y_{\text{self}}) - \mu_{\text{self}} 1_n) - \sum_{\text{i=1}}^{n} \log(-y^i_{\text{self}}) \\
& - \sum_{i=1}^{n} \sum_{\text{t=1}}^{\text{Time}} \exp\{e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}_{-i}^{\,t})^T y^i_{\text{cross}} + e_i^T y_0\} \\
& + \sum_{\text{i=1}}^{n} \sum_{\text{t=1}}^{\text{Time}} q_{it} [e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}_{-i}^{\,t})^T y^i_{\text{cross}} + e_i^T y_0] \\
& = -\frac{1}{2\sigma^2_{\text{self}}}(\log(-y_{\text{self}}) - \mu_{\text{self}} 1_n)^T (\log(-y_{\text{self}}) - \mu_{\text{self}} 1_n) - (\log(-y_{\text{self}}))^T 1_n \\
& - \sum_{i=1}^{n} \sum_{\text{t=1}}^{\text{Time}} \exp\{e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}_{-i}^{\,t})^T y^i_{\text{cross}} + e_i^T y_0\} \\
& + \sum_{\text{i=1}}^{n} \sum_{\text{t=1}}^{\text{Time}} q_{it} [e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}_{-i}^{\,t})^T y^i_{\text{cross}} + e_i^T y_0]
\end{aligned}
\tag{B.2}
$$

B.3 Gradient of y_{self}

$$
\begin{aligned}
d\,l(y_{\text{self}}) &= -\frac{1}{2\sigma^2_{\text{self}}}[d\log(-y_{self})^T(\log(-y_{self}) - \mu_{self}1_n) + (\log(-y_{self}) - \mu_{self}1_n)^T d\log(-y_{self})] \\
&- 1_n^T d\log(-y_{\text{self}}) \\
&- \sum_{i=1}^{n}\sum_{\text{t}=1}^{\text{Time}} d\,(\exp(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0)) \\
&+ \sum_{\text{i}=1}^{n}\sum_{\text{t}=1}^{\text{Time}} q_{it}\; d\,(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0) \\
&= -\frac{1}{2\sigma^2_{\text{self}}}[d\,y^T_{\text{self}}\text{diag}(\frac{1}{y_{\text{self}}})(\log(-y_{\text{self}}) - \mu_{\text{self}}1_n) + (\log(-y_{\text{self}}) - \mu_{\text{self}}1_n)^T\text{diag}(\frac{1}{y_{\text{self}}})dy_{\text{self}}] \\
&- 1_n^T\text{diag}(\frac{1}{y_{\text{self}}}) \\
&- \sum_{i=1}^{n}\sum_{\text{t}=1}^{\text{Time}} d\,(\exp(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0)) \\
&+ \sum_{\text{i}=1}^{n}\sum_{\text{t}=1}^{\text{Time}} q_{it}\; d\,(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0) \\
&= -\frac{1}{\sigma^2_{\text{self}}}(\log(-y_{\text{self}}) - \mu_{\text{self}}1_n + \sigma^2_{\text{self}}1_n)^T\text{diag}((y_{\text{self}})^{-1})dy_{\text{self}} \\
&- \sum_{i=1}^{n}\sum_{\text{t}=1}^{\text{Time}} \exp(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0) e_i^T(p^t - \bar{p}^t)\, e_i^T dy_{\text{self}} \\
&+ \sum_{\text{i}=1}^{n}\sum_{\text{t}=1}^{\text{Time}} q_{it}\; e_i^T(p^t - \bar{p}^t)\, e_i^T dy_{\text{self}} \\
\nabla\, l(y_{\text{self}}) &= -\frac{1}{\sigma^2_{\text{self}}}\text{diag}(y_{\text{self}})^{-1}(\log(-y_{\text{self}}) - \mu_{\text{self}}1_n + \sigma^2_{\text{self}}1_n) \\
&- \sum_{i=1}^{n}\sum_{\text{t}=1}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y^T_{\text{self}}\, e_i + (y^i_{cross})^T(p^t_{-i} - \bar{p}^t_{-i}) + y_0^T e_i\}\, e_i(p^t - \bar{p}^t)^T e_i \\
&+ \sum_{\text{i}=1}^{n}\sum_{\text{t}=1}^{\text{Time}} q_{it}\; e_i(p^t - \bar{p}^t)^T\, e_i
\end{aligned}
\tag{B.3}
$$

B.4 Hessian of y_{self}

$$
\begin{aligned}
Hl(y_{\text{self}}) &= \frac{\partial \nabla l(y_{\text{self}})}{\partial y_{\text{self}}} = \frac{\partial}{\partial y_{\text{self}}} \frac{\partial f(y_{\text{self}})}{\partial y_{\text{self}}^T} = \frac{\partial^2 f(y_{\text{self}})}{\partial y_{\text{self}} \partial y_{\text{self}}^T} \\
&= \frac{\partial}{\partial y_{\text{self}}} \left(\operatorname{diag}(y_{\text{self}})^{-1} (\log(-y_{\text{self}}) - \mu_{\text{self}} 1_n + \sigma_{\text{self}}^2 1_n) + \sum_{\text{i=1}}^{n} \sum_{\text{t=1}}^{\text{Time}} q_{it}\, e_i (p^t - \bar{p}^t)^T\, e_i \right) \\
&- \frac{\partial}{\partial y_{\text{self}}} \left(\sum_{i=1}^{n} \sum_{\text{t=1}}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y_{\text{self}}^T e_i + (y_{\text{cross}}^i)^T (p_{-i}^t - \bar{p}_{-i}^t) + y_0^T e_i\} e_i (p^t - \bar{p}^t)^T e_i \right) \\
&= \frac{\partial}{\partial y_{\text{self}}} \left(\begin{bmatrix} \frac{1}{y_{\text{self}}^1} & 0 & \cdots & 0 \\ \vdots & & \frac{1}{y_{\text{self}}^i} & \vdots \\ 0 & \cdots & 0 & \frac{1}{y_{\text{self}}^n} \end{bmatrix} \begin{bmatrix} \log(-y_{\text{self}}^1) - \mu_{\text{self}} + \sigma_{\text{self}}^2 \\ \log(-y_{\text{self}}^2) - \mu_{\text{self}} + \sigma_{\text{self}}^2 \\ \vdots \\ \log(-y_{\text{self}}^n) - \mu_{\text{self}} + \sigma_{\text{self}}^2 \end{bmatrix} + \sum_{\text{i=1}}^{n} \sum_{\text{t=1}}^{\text{Time}} q_{it}\, e_i (p^t - \bar{p}^t)^T\, e_i \right) \\
&- \frac{\partial}{\partial y_{\text{self}}} \left(\sum_{i=1}^{n} \sum_{\text{t=1}}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y_{\text{self}}^T e_i + (y_{cross}^i)^T (p_{-i}^t - \bar{p}_{-i}^t) + y_0^T e_i\} e_i (p^t - \bar{p}^t)^T e_i \right) \\
&= \frac{\partial}{\partial y_{\text{self}}} \left(\begin{bmatrix} (y_{\text{self}}^1)^{-1} (\log(-y_{\text{self}}^1) - \mu_{\text{self}} + \sigma_{\text{self}}^2) \\ (y_{\text{self}}^2)^{-1} (\log(-y_{\text{self}}^2) - \mu_{\text{self}} + \sigma_{\text{self}}^2) \\ \vdots \\ (y_{\text{self}}^n)^{-1} (\log(-y_{\text{self}}^n) - \mu_{\text{self}} + \sigma_{\text{self}}^2) \end{bmatrix} + \sum_{\text{i=1}}^{n} \sum_{\text{t=1}}^{\text{Time}} q_{it}\, e_i (p^t - \bar{p}^t)^T\, e_i \right) \\
&- \frac{\partial}{\partial y_{\text{self}}} \left(\sum_{i=1}^{n} \sum_{\text{t=1}}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y_{\text{self}}^T e_i + (y_{cross}^i)^T (p_{-i}^t - \bar{p}_{-i}^t) + y_0^T e_i\}\, e_i (p^t - \bar{p}^t)^T e_i \right) \\
&= \begin{bmatrix} -(y_{\text{self}}^1)^{-2} (\log(-y_{\text{self}}^1) - \mu_{\text{self}} + \sigma_{\text{self}}^2) + (y_{\text{self}}^1)^{-2} & 0 & \cdots & 0 \\ \vdots & -(y_{\text{self}}^i)^{-2} (\log(-y_{\text{self}}^i) - \mu_{\text{self}} + \sigma_{\text{self}}^2) + (y_{\text{self}}^i)^{-2} & & \vdots \\ 0 & \cdots & 0 & -(y_{\text{self}}^n)^{-2} (\log(-y_{\text{self}}^n) - \mu_{\text{self}} + \sigma_{\text{self}}^2) + (y_{\text{self}}^n)^{-2} \end{bmatrix} \\
&- \sum_{i=1}^{n} \sum_{\text{t=1}}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y_{\text{self}}^T e_i + (y_{\text{cross}}^i)^T (p_{-i}^t - \bar{p}_{-i}^t) + y_0^T e_i\}\, e_i (p^t - \bar{p}^t)^T e_i\, e_i^T (p^t - \bar{p}^t) e_i^T
\end{aligned}
\tag{B.4}
$$

B.5 Full conditional PDF of $y^i_{\textbf{cross}}\ \forall\, i$

$$
\begin{aligned}
&\mathrm{f}(\ y^i_{\text{cross}}) \propto \mathrm{f}(y^i_{\text{cross}}|\text{everything else}) \propto \mathrm{f}(\ y^i_{\text{cross}}, \text{everything else})\\
&\propto \mathrm{f}(\ y^i_{\text{cross}}|\mu_{\text{cross}}1_{(n-1)}, \sigma^2_{\text{cross}}I)\mathrm{f}(q_{it}|\ y_{\text{self}}, y^i_{\text{cross}}, y_0, p)\\
&\propto \frac{1}{|2\pi|^{\frac{n-1}{2}}|\sigma^2_{\text{cross}}I|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})^T(\sigma^2_{\text{cross}}I)^{-1}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})\}\times\\
&(\prod_{\text{t=1}}^{\text{Time}}\frac{\exp\{-\exp(e_i^T y_{\text{self}} e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0)\}}{q_{it}!}\times\\
&\exp\{e_i^T\ y_{\text{self}}\ e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0\}^{q_{it}})\\
&\propto \exp\{-\frac{1}{2}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})^T(\sigma^2_{\text{cross}}I)^{-1}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})\}\times\\
&(\prod_{\text{t=1}}^{\text{Time}}\exp\{-\exp(e_i^T y_{\text{self}} e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0)\}\times\\
&\exp\{e_i^T\ y_{\text{self}}\ e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0\}^{q_{it}})
\end{aligned}
\tag{B.5}
$$

B.6 Loglikelihood of $y^i_{\textbf{cross}}\ \forall\, i$

$$
\begin{aligned}
l(y^i_{\text{cross}}) &= \log \mathrm{f}(y^i_{\text{cross}})\\
&\propto -\frac{1}{2}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})^T(\sigma^2_{\text{cross}}I)^{-1}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})\\
&- \sum_{\text{t=1}}^{\text{Time}}\exp(e_i^T y_{\text{self}} e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0)\\
&+ \sum_{\text{t=1}}^{\text{Time}} q_{it}[e_i^T\ y_{\text{self}}\ e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0]\\
&= -\frac{1}{2\sigma^2_{\text{cross}}}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})^T(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})\\
&- \sum_{\text{t=1}}^{\text{Time}}\exp(e_i^T y_{\text{self}} e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0)\\
&+ \sum_{\text{t=1}}^{\text{Time}} q_{it}[e_i^T\ y_{\text{self}}\ e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0]
\end{aligned}
\tag{B.6}
$$

B.7 Gradient of $y^i_{\text{cross}}\ \forall\, i$

$$
\begin{aligned}
&d(l(y^i_{\text{cross}})) = -\frac{1}{2\sigma^2_{\text{cross}}}[d(y^i_{\text{cross}})^T(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)}) + (y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})^T d(y^i_{\text{cross}})] \\
&- \sum_{\text{t=1}}^{\text{Time}} d[\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T\, y_0\}] \\
&+ \sum_{\text{t=1}}^{\text{Time}} q_{it}\, d[e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T\, y_0] \\
&= -\frac{1}{\sigma^2_{\text{cross}}}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)})^T d(y^i_{\text{cross}}) \\
&- \sum_{\text{t=1}}^{\text{Time}} \exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T\, y_0\}(p^t_{-i} - \bar{p}^t_{-i})^T d(y^i_{cross}) \\
&+ \sum_{\text{t=1}}^{\text{Time}} q_{it}\,(p^t_{-i} - \bar{p}^t_{-i})^T d(y^i_{cross}) \\
&\nabla\, l(y^i_{\text{cross}}) = -\frac{1}{\sigma^2_{\text{cross}}}(y^i_{\text{cross}} - \mu_{\text{cross}}1_{(n-1)}) + \sum_{\text{t=1}}^{\text{Time}} q_{it}\,(p^t_{-i} - \bar{p}^t_{-i}) \\
&- \sum_{\text{t=1}}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y^T_{\text{self}}\, e_i + (y^i_{cross})^T(p^t_{-i} - \bar{p}^t_{-i}) + y_0^T\, e_i\}(p^t_{-i} - \bar{p}^t_{-i})
\end{aligned}
\tag{B.7}
$$

B.8 Hessian of $y^i_{\text{cross}}\ \forall\, i$

$$
\begin{aligned}
&Hl(y^i_{\text{cross}}) = \frac{\partial\, \nabla\, l(y^i_{\text{cross}})}{\partial\, y^i_{\text{cross}}} = -\frac{1}{\sigma^2_{\text{cross}}} I_{(n-1)} \\
&- \sum_{\text{t=1}}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y^T_{\text{self}}\, e_i + (y^i_{cross})^T(p^t_{-i} - \bar{p}^t_{-i}) + y_0^T\, e_i\}(p^t_{-i} - \bar{p}^t_{-i})(p^t_{-i} - \bar{p}^t_{-i})^T
\end{aligned}
\tag{B.8}
$$

B.9 Full conditional PDF of y_0

$$
\begin{aligned}
&\text{f}(\ y_0) \propto \text{f}(\ y_0|\ \text{everything else}) \propto \text{f}(\ y_0,\ \text{everything else})\\
&\quad \propto \text{f}(y_0|\mu_{\text{Int}}1_n, \sigma^2_{\text{Int}}I)\prod_{\text{i=1}}^{n}\text{f}(\ q_{it}\,|\,y_{\text{self}}, y^i_{\text{cross}}, y_0, p)\\
&\quad \propto \frac{1}{|2\pi|^{\frac{n}{2}}|\sigma^2_{\text{Int}}I|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(y_0-\mu_{\text{Int}}1_n)^T(\sigma^2_{\text{Int}}I)^{-1}(y_0-\mu_{\text{Int}}1_n)\}\times\\
&\quad (\prod_{i=1}^{n}\prod_{t=1}^{\text{Time}}\frac{\exp\{-\exp\{e_i^T\,y_{\text{self}}\,e_i^T(p^t-\bar{p}^t)+(p^t_{-i}-\bar{p}^t_{-i})^Ty^i_{cross}+e_i^Ty_0\}\}}{q_{it}!}\times\\
&\quad \exp\{e_i^T\,y_{\text{self}}\,e_i^T(p^t-\bar{p}^t)+(p^t_{-i}-\bar{p}^t_{-i})^Ty^i_{cross}+e_i^Ty_0\}^{q_{it}})\\
&\quad \propto \exp\{-\frac{1}{2}(y_0-\mu_{\text{Int}}1_n)^T(\sigma^2_{\text{Int}}I)^{-1}(y_0-\mu_{\text{Int}}1_n)\}\times\\
&\quad (\prod_{i=1}^{n}\prod_{t=1}^{\text{Time}}\exp\{-\exp\{e_i^T\,y_{\text{self}}\,e_i^T(p^t-\bar{p}^t)+(p^t_{-i}-\bar{p}^t_{-i})^Ty^i_{cross}+e_i^Ty_0\}\}\times\\
&\quad \exp\{e_i^T\,y_{\text{self}}\,e_i^T(p^t-\bar{p}^t)+(p^t_{-i}-\bar{p}^t_{-i})^Ty^i_{cross}+e_i^Ty_0\}^{q_{it}})
\end{aligned}
\tag{B.9}
$$

B.10 Loglikelihood of y_0

$$
\begin{aligned}
l(y_0) &= \log \mathrm{f}(\ y_0) \\
&\propto -\frac{1}{2}(y_0 - \mu_{\text{Int}} 1_n)^T (\sigma^2_{\text{Int}} I)^{-1} (y_0 - \mu_{\text{Int}} 1_n) \\
&- \sum_{i=1}^{n} \sum_{t=1}^{\text{Time}} \exp\{e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0\} \\
&+ \sum_{i=1}^{n} \sum_{t=1}^{\text{Time}} q_{it}[e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0] \\
&= -\frac{1}{2\sigma^2_{\text{Int}}}(y_0 - \mu_{\text{Int}} 1_n)^T (y_0 - \mu_{\text{Int}} 1_n) \\
&- \sum_{i=1}^{n} \sum_{t=1}^{\text{Time}} \exp\{e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0\} \\
&+ \sum_{i=1}^{n} \sum_{t=1}^{\text{Time}} q_{it}[e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0]
\end{aligned}
\tag{B.10}
$$

B.11 Gradient of y_0

$$
\begin{aligned}
d(l(y_0)) &= -\frac{1}{2\sigma^2_{\text{Int}}}[d(y_0)^T(y_0 - \mu_{\text{Int}}1_n) + (y_0 - \mu_{\text{Int}}1_n)^T d(y_0)] \\
&- \sum_{i=1}^{n}\sum_{t=1}^{\text{Time}} d(\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T\, y_0\}) \\
&+ \sum_{i=1}^{n}\sum_{t=1}^{\text{Time}} q_{it}\, d(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T\, y_0) \\
&= -\frac{1}{\sigma^2_{\text{Int}}}(y_0 - \mu_{\text{Int}}1_n)^T d(y_0) \\
&- \sum_{i=1}^{n}\sum_{t=1}^{\text{Time}} \exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T\, y_0\} e_i^T d(y_0) \\
&+ \sum_{i=1}^{n}\sum_{t=1}^{\text{Time}} q_{it}\, e_i^T d(y_0) \\
\nabla\, l(y_0) &= -\frac{1}{\sigma^2_{\text{Int}}}(y_0 - \mu_{\text{Int}}1_n) + \sum_{i=1}^{n}\sum_{t=1}^{\text{Time}} q_{it} e_i \\
&- \sum_{i=1}^{n}\sum_{t=1}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y^T_{\text{self}}\, e_i + (y^i_{cross})^T(p^t_{-i} - \bar{p}^t_{-i}) + y_0^T\, e_i\} e_i
\end{aligned}
\tag{B.11}
$$

B.12 Hessian of y_0

$$
\begin{aligned}
Hl(y_0) &= \frac{\partial\, \nabla\, l(y_0)}{\partial\, y_0} = -\frac{1}{\sigma^2_{\text{Int}}} I_n \\
&- \sum_{i=1}^{n}\sum_{t=1}^{\text{Time}} \exp\{(p^t - \bar{p}^t)^T e_i\, y^T_{\text{self}}\, e_i + (y^i_{cross})^T(p^t_{-i} - \bar{p}^t_{-i}) + y_0^T\, e_i\} e_i\, e_i^T
\end{aligned}
\tag{B.12}
$$

B.13 Full conditional PDF of μ_{self}

$$
\begin{aligned}
&\text{f}(\mu_{\text{self}}) \propto \text{f}(\mu_{\text{self}}|\text{ everything else}) \propto \text{f}(\ \mu_{\text{self}}, \text{ everything else})\\
&\propto \text{f}(\ \mu_{\text{self}} \,|\, \bar{\mu}_{\text{self}}, \bar{\sigma}^2_{\text{self}})\text{f}(\ \log(-y_{\text{self}})|\mu_{\text{self}} 1_n, \sigma^2_{\text{self}}\, I)\\
&\propto -\frac{1}{\sqrt{2\pi\bar{\sigma}^2_{\text{self}}}} \exp\left(-\frac{(\mu_{\text{self}} - \bar{\mu}_{\text{self}})^2}{2\bar{\sigma}^2_{\text{self}}}\right) \times\\
&\frac{1}{|2\pi|^{\frac{n}{2}}|\sigma^2_{\text{self}} I|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\log(-y_{\text{self}}) - \mu_{\text{self}} 1_n)^T(\sigma^2_{\text{self}} I)^{-1}(\log(-y_{\text{self}}) - \mu_{\text{self}} 1_n)\right\}\\
&\propto \exp\{-\frac{1}{2}[(\mu_{\text{self}} - \bar{\mu}_{\text{self}})^2(\bar{\sigma}^2_{\text{self}})^{-1}\\
&+ \log(-y_{\text{self}}) - \mu_{\text{self}} 1_n)^T(\sigma^2_{\text{self}} I)^{-1}(\log(-y_{\text{self}}) - \mu_{\text{self}} 1_n)]\}\\
&\propto \exp\{-\frac{1}{2}[\mu^2_{\text{self}}(\bar{\sigma}^2_{\text{self}})^{-1} - 2\bar{\mu}_{\text{self}}(\bar{\sigma}^2_{\text{self}})^{-1}\mu_{\text{self}}\\
&+ \mu^2_{\text{self}} 1_n^T(\sigma^2_{\text{self}} I)^{-1} 1_n - 2\log(-y_{\text{self}})^T(\sigma^2_{\text{self}} I)^{-1} 1_n \mu_{\text{self}}]\}
\end{aligned}
$$

which is the same as the pdf of Normal ($\mu_{\mu\text{self}}\,, \sigma^2_{\mu\text{self}}$)

$$
\begin{aligned}
&\mu_{\mu\text{self}} := ((\bar{\sigma}^2_{\text{self}})^{-1} + 1_n^T(\sigma^2_{\text{self}})^{-1} 1_n)^{-1}(\bar{\mu}_{\text{self}}(\bar{\sigma}^2_{\text{self}})^{-1} + 1_n^T(\sigma^2_{\text{self}}\, I)^{-1}\log(-y_{\text{self}}))\\
&\sigma^2_{\mu\text{self}} := ((\bar{\sigma}^2_{\text{self}})^{-1} + 1_n^T(\sigma^2_{\text{self}})^{-1} 1_n)^{-1}
\end{aligned}
\tag{B.13}
$$

B.14 Full conditional PDF of σ^2_{self}

$$
\begin{aligned}
&\mathrm{f}(\ \sigma^2_{\text{self}}) \propto \mathrm{f}(\ \sigma^2_{\text{self}}|\ \text{everything else}) \propto \mathrm{f}(\ \sigma^2_{\text{self}}\ , \text{everything else})\\
&\propto \mathrm{f}(\ \sigma^2_{\text{self}}\,|\,\nu_{0,\text{ self}}, \upsilon_{0,\text{ self}})\mathrm{f}(\ \log(-y_{\text{self}})|\mu_{\text{self}}1_n, \sigma^2_{\text{self}}\,I)\\
&\propto \frac{(\upsilon_{0,\text{ self}})^{\nu_{0,\text{ self}}}}{\Gamma(\nu_{0,\text{ self}})}(\sigma^2_{\text{self}})^{-\nu_{0,\text{ self}}-1}\exp\left\{\frac{-\upsilon_{0,\text{ self}}}{\sigma^2_{\text{self}}}\right\}\times\\
&\frac{1}{|2\pi|^{\frac{n}{2}}|\sigma^2_{\text{self}}I|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\log(-y_{\text{self}})-\mu_{\text{self}}1_n)^T(\sigma^2_{\text{self}}I)^{-1}(\log(-y_{\text{self}})-\mu_{\text{self}}1_n)\right\}\\
&\propto (\sigma^2_{\text{self}})^{-\nu_{0,\text{ self}}-1}\exp\left\{\frac{-\upsilon_{0,\text{ self}}}{\sigma^2_{\text{self}}}\right\}|\sigma^2_{\text{self}}\,I|^{-\frac{1}{2}}\times\\
&\exp\left\{-\frac{1}{2}(\sigma^2_{\text{self}})^{-1}(\log(-y_{\text{self}})-\mu_{\text{self}}1_n)^T(\log(-y_{\text{self}})-\mu_{\text{self}}1_n)\right\}
\end{aligned}
$$

which is the same as the pdf of Inverse Gamma($\nu_{0\sigma^2\text{self}}$, $\upsilon_{0\sigma^2\text{self}}$)

$$
\begin{aligned}
&\nu_{0\sigma^2\text{self}} := \nu_{0,\text{ self}} + \frac{n}{2}\\
&\upsilon_{0\sigma^2\text{self}} := \upsilon_{0,\text{ self}} + \frac{1}{2}(\log(-y_{\text{self}})-\mu_{\text{self}}1_n)^T(\log(-y_{\text{self}})-\mu_{\text{self}}1_n)
\end{aligned}
\tag{B.14}
$$

B.15 Full conditional PDF of μ_{cross}

$$
\begin{aligned}
&\text{f}(\mu_{\text{cross}}) \propto \text{f}(\mu_{\text{cross}} |\text{ everything else}) \propto \text{f}(\ \mu_{\text{cross}}, \text{everything else})\\
&\propto \text{f}(\ \mu_{\text{cross}} \,|\, \bar{\mu}_{\text{cross}}, \bar{\sigma}^2_{\text{cross}}) \prod_{\text{i=1}}^{n} \text{f}(y^i_{\text{cross}} | \mu_{\text{cross}} 1_{(n-1)}, \sigma^2_{\text{cross}}\, I)\\
&\propto -\frac{1}{\sqrt{2\pi\bar{\sigma}^2_{\text{cross}}}} \exp\left(-\frac{(\mu_{\text{cross}} - \bar{\mu}_{\text{cross}})^2}{2\bar{\sigma}^2_{\text{cross}}}\right) \times\\
&\prod_{\text{i=1}}^{n} \frac{1}{|2\pi|^{\frac{n-1}{2}} |\sigma^2_{\text{cross}} I|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})^T (\sigma^2_{\text{cross}} I)^{-1} (y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})\right\}\\
&\propto \exp\{-\frac{1}{2}[(\mu_{\text{cross}} - \bar{\mu}_{\text{cross}})^2 (\bar{\sigma}^2_{\text{cross}})^{-1}\\
&+ \sum_{\text{i=1}}^{n} (y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})^T (\sigma^2_{\text{cross}} I)^{-1} (y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})]\}\\
&\propto \exp\{-\frac{1}{2}[\mu^2_{\text{cross}} (\bar{\sigma}^2_{\text{cross}})^{-1} - 2\bar{\mu}_{\text{cross}} (\bar{\sigma}^2_{\text{cross}})^{-1} \mu_{\text{cross}}\\
&+ n\mu^2_{\text{cross}} 1^T_{(n-1)} (\sigma^2_{\text{cross}} I)^{-1} 1_{(n-1)} - 2\mu_{\text{cross}} \sum_{\text{i=1}}^{n} (y^i_{\text{cross}})^T (\sigma^2_{\text{cross}} I)^{-1} 1_{(n-1)}]\}
\end{aligned}
$$

which is the same as the pdf of Normal ($\mu_{\mu\text{cross}}$, $\sigma^2_{\mu\text{cross}}$)

$$
\begin{aligned}
&\mu_{\mu\text{cross}} := ((\bar{\sigma}^2_{\text{cross}})^{-1} + n 1^T_{(n-1)} (\sigma^2_{\text{cross}})^{-1} 1_{(n-1)})^{-1} (\bar{\mu}_{\text{cross}} (\bar{\sigma}^2_{\text{cross}})^{-1} + \sum_{\text{i=1}}^{n} 1^T_{(n-1)} (\sigma^2_{\text{cross}}\, I)^{-1} y^i_{\text{cross}})\\
&\sigma^2_{\mu\text{cross}} := ((\bar{\sigma}^2_{\text{cross}})^{-1} + n 1^T_{(n-1)} (\sigma^2_{\text{cross}})^{-1} 1_{(n-1)})^{-1}
\end{aligned} \tag{B.15}
$$

B.16 Full conditional PDF of σ^2_{cross}

$$
\begin{aligned}
&\text{f}(\ \sigma^2_{\text{cross}}) \propto \text{f}(\ \sigma^2_{\text{cross}} |\ \text{everything else}) \propto \text{f}(\ \sigma^2_{\text{cross}}\ ,\ \text{everything else})\\
&\propto \text{f}(\ \sigma^2_{\text{cross}}\,|\,\nu_{0,\ \text{cross}}, \upsilon_{0,\ \text{cross}}) \prod_{\text{i=1}}^{n} \text{f}(\ y^i_{\text{cross}} | \mu_{\text{cross}} 1_{(n-1)}, \sigma^2_{\text{cross}}\ I)\\
&\propto \frac{(\upsilon_{0,\ \text{cross}})^{\nu_{0,\ \text{cross}}}}{\Gamma(\nu_{0,\ \text{cross}})} (\sigma^2_{\text{cross}})^{-\nu_{0,\ \text{cross}}-1} \exp\left\{\frac{-\upsilon_{0,\ \text{cross}}}{\sigma^2_{\text{cross}}}\right\} \times\\
&\prod_{\text{i=1}}^{n} \frac{1}{|2\pi|^{\frac{n-1}{2}} |\sigma^2_{\text{cross}} I|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})^T (\sigma^2_{\text{cross}} I)^{-1} (y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})\right\}\\
&\propto (\sigma^2_{\text{cross}})^{-\nu_{0,\ \text{cross}}-1} \exp\left\{\frac{-\upsilon_{0,\ \text{cross}}}{\sigma^2_{\text{cross}}}\right\} |\sigma^2_{\text{cross}}\ I|^{-\frac{n}{2}} \times\\
&\exp\left\{-\frac{1}{2}(\sigma^2_{\text{cross}})^{-1} \sum_{\text{i=1}}^{n} (y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})^T (y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})\right\}
\end{aligned}
$$

which is the same as the pdf of Inverse Gamma($\nu_{0\sigma^2\text{cross}}$, $\upsilon_{0\sigma^2\text{cross}}$)

$$
\begin{aligned}
&\nu_{0\sigma^2\text{cross}} := \nu_{0,\ \text{cross}} + \frac{n(n-1)}{2}\\
&\upsilon_{0\sigma^2\text{cross}} := \upsilon_{0,\ \text{cross}} + \frac{1}{2} \sum_{\text{i=1}}^{n} (y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})^T (y^i_{\text{cross}} - \mu_{\text{cross}} 1_{(n-1)})
\end{aligned}
\tag{B.16}
$$

B.17 Full conditional PDF of μ_{Int}

$$
\begin{aligned}
&\text{f}(\mu_{\text{Int}}) \propto \text{f}(\mu_{\text{Int}} | \text{ everything else}) \propto \text{f}(\ \mu_{\text{Int}}, \text{everything else}) \\
&\propto \text{f}(\ \mu_{\text{Int}}\ |\ \bar{\mu}_{\text{Int}}, \bar{\sigma}^2_{\text{Int}}) \text{f}(y_0 | \mu_{\text{Int}} 1_n, \sigma^2_{\text{Int}}\ I) \\
&\propto -\frac{1}{\sqrt{2\pi\bar{\sigma}^2_{\text{Int}}}} \exp\left(-\frac{(\mu_{\text{Int}} - \bar{\mu}_{\text{Int}})^2}{2\bar{\sigma}^2_{\text{Int}}}\right) \times \\
&\frac{1}{|2\pi|^{\frac{n}{2}} |\sigma^2_{\text{Int}} I|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y_0 - \mu_{\text{Int}} 1_n)^T (\sigma^2_{\text{Int}} I)^{-1} (y_0 - \mu_{\text{Int}} 1_n)\right\} \\
&\propto \exp\{-\frac{1}{2}[(\mu_{\text{Int}} - \bar{\mu}_{\text{Int}})^2 (\bar{\sigma}^2_{\text{Int}})^{-1} \\
&+ y_0 - \mu_{\text{Int}} 1_n)^T (\sigma^2_{\text{Int}} I)^{-1} (y_0 - \mu_{\text{Int}} 1_n)]\} \\
&\propto \exp\{-\frac{1}{2}[\mu^2_{\text{Int}} (\bar{\sigma}^2_{\text{Int}})^{-1} - 2\bar{\mu}_{\text{Int}} (\bar{\sigma}^2_{\text{Int}})^{-1} \mu_{\text{Int}} \\
&+ \mu^2_{\text{Int}} 1^T_n (\sigma^2_{\text{Int}} I)^{-1} 1_n - 2y_0^T (\sigma^2_{\text{Int}} I)^{-1} 1_n \mu_{\text{Int}}]\}
\end{aligned}
\tag{B.17}
$$

which is the same as the pdf of Normal ($\mu_{\mu\text{Int}}\ , \sigma^2_{\mu\text{Int}}$)

$\mu_{\mu\text{Int}} := ((\bar{\sigma}^2_{\text{Int}})^{-1} + 1^T_n (\sigma^2_{\text{Int}})^{-1} 1_n)^{-1} (\bar{\mu}_{\text{Int}} (\bar{\sigma}^2_{\text{Int}})^{-1} + 1^T_n (\sigma^2_{\text{Int}}\ I)^{-1} y_0)$

$\sigma^2_{\mu\text{Int}} := ((\bar{\sigma}^2_{\text{Int}})^{-1} + 1^T_n (\sigma^2_{\text{Int}})^{-1} 1_n)^{-1}$

B.18 Full conditional PDF of σ_{Int}

$$
\begin{aligned}
&\mathrm{f}(\ \sigma^2_{\text{Int}}) \propto \mathrm{f}(\ \sigma^2_{\text{Int}}|\ \text{everything else}) \propto \mathrm{f}(\ \sigma^2_{\text{Int}}\ ,\ \text{everything else})\\
&\propto \mathrm{f}(\ \sigma^2_{\text{Int}}\,|\,\nu_{0,\ \text{Int}}, \upsilon_{0,\ \text{Int}})\mathrm{f}(\ y_0|\mu_{\text{Int}}1_n, \sigma^2_{\text{Int}}\ I)\\
&\propto \frac{(\upsilon_{0,\ \text{Int}})^{\nu_{0,\ \text{Int}}}}{\Gamma(\nu_{0,\ \text{Int}})}(\sigma^2_{\text{Int}})^{-\nu_{0,\ \text{Int}}-1}\exp\left\{\frac{-\upsilon_{0,\ \text{Int}}}{\sigma^2_{\text{Int}}}\right\}\times\\
&\frac{1}{|2\pi|^{\frac{n}{2}}|\sigma^2_{\text{Int}}I|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y_0-\mu_{\text{Int}}1_n)^T(\sigma^2_{\text{Int}}I)^{-1}(y_0-\mu_{\text{Int}}1_n)\right\}\\
&\propto (\sigma^2_{\text{Int}})^{-\nu_{0,\ \text{Int}}-1}\exp\left\{\frac{-\upsilon_{0,\ \text{Int}}}{\sigma^2_{\text{Int}}}\right\}|\sigma^2_{\text{Int}}\ I|^{-\frac{1}{2}}\times\\
&\exp\left\{-\frac{1}{2}(\sigma^2_{\text{Int}})^{-1}(y_0-\mu_{\text{Int}}1_n)^T(y_0-\mu_{\text{Int}}1_n)\right\}
\end{aligned}
\tag{B.18}
$$

which is the same as the pdf of Inverse Gamma($\nu_{0\sigma^2\text{Int}}$, $\upsilon_{0\sigma^2\text{Int}}$)

$\nu_{0\sigma^2\text{Int}} := \nu_{0,\ \text{Int}} + \dfrac{n}{2}$

$\upsilon_{0\sigma^2\text{Int}} := \upsilon_{0,\ \text{Int}} + \dfrac{1}{2}(y_0-\mu_{\text{Int}}1_n)^T(y_0-\mu_{\text{Int}}1_n)$

Appendix C

Calculations and Derivations for Preliminary Model II

C.1 Full conditional PDF of x_{self}

$$
\begin{aligned}
&\text{f}(x_{\text{self}}) \propto \text{f}(x_{\text{self}}|\text{ everything else}) \propto \text{f}(x_{\text{self}}\text{ , everything else})\\
&\propto \text{f}(\text{log(-x}_{\text{self}})|\mu_{\text{xself}} * 1_n \text{ , } \sigma^2_{\text{xself}} * I)\left|\frac{\partial \log(-x_{\text{self}})}{\partial x_{\text{self}}}\right| \prod_{\text{i=1}}^{n} \text{f}(\ q_{it} \mid y_{\text{self}}, y^i_{cross}, y_0, p, x_{\text{self}}, x^i_{\text{cross}}, q_{t-1})\\
&\propto \frac{1}{|2\pi|^{\frac{n}{2}}|\sigma^2_{\text{xself}}I|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n)^T(\sigma^2_{\text{xself}}I)^{-1}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n)\}|\text{diag}(\ x_{\text{self}})^{-1}|\times\\
&(\prod_{i=1}^{n}\prod_{t=1}^{\text{Time}} \frac{\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}\}^{q_{it}}}{q_{it}!}\times\\
&\exp\{-\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}\}\})\\
&\propto \exp\{-\frac{1}{2}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n)^T(\sigma^2_{\text{xself}}I)^{-1}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n)\}(-1)^n \left[\prod_{i=1}^{n} x^i_{\text{self}}\right]^{-1}\times\\
&(\prod_{i=1}^{n}\prod_{t=1}^{\text{Time}} \exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}\}^{q_{it}}\times\\
&\exp\{-\exp\{e_i^T y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}\}\})
\end{aligned}
\tag{C.1}
$$

C.2 Loglikelihood of x_{self}

$$
\begin{aligned}
&l(x_{self}) = \log(f(x_{\text{self}})) \\
&\propto -\frac{1}{2}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n)^T(\sigma^2_{\text{xself}}I)^{-1}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n) - \sum_{\text{i=1}}^{n}\log(-x^i_{\text{self}}) \\
&- \sum_{i=1}^{n}\sum_{\text{t=1}}^{\text{Time}}\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}\} \\
&+ \sum_{\text{i=1}}^{n}\sum_{\text{t=1}}^{\text{Time}} q_{it}[e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}] \\
&= -\frac{1}{2\sigma^2_{\text{xself}}}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n)^T(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n) - (\log(-x_{\text{self}}))^T 1_n \\
&- \sum_{i=1}^{n}\sum_{\text{t=1}}^{\text{Time}}\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}\} \\
&+ \sum_{\text{i=1}}^{n}\sum_{\text{t=1}}^{\text{Time}} q_{it}[e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}]
\end{aligned}
\tag{C.2}
$$

C.3 Gradient of x_{self}

$$d\,l(x_{\text{self}}) = -\frac{1}{2\sigma^2_{\text{xself}}}[d\log(-x_{self})^T(\log(-x_{self}) - \mu_{xself}1_n)+$$
$$(\log(-x_{self}) - \mu_{xself}1_n)^T d\log(-x_{self})] - 1_n^T d\log(-x_{\text{self}})$$
$$-\sum_{i=1}^{n}\sum_{\text{t=1}}^{\text{Time}} d\,(\exp(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 +$$
$$e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}))$$
$$+\sum_{\text{i=1}}^{n}\sum_{\text{t=1}}^{\text{Time}} q_{it}\; d\,(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 +$$
$$e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}})$$
$$= -\frac{1}{2\sigma^2_{\text{xself}}}[d\,x^T_{\text{self}}\text{diag}(\frac{1}{x_{\text{self}}})(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n) + (\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n)^T\text{diag}(\frac{1}{x_{\text{self}}})dx_{\text{self}}]$$
$$-1_n^T\text{diag}(\frac{1}{x_{\text{self}}})dx_{\text{self}}$$
$$-\sum_{i=1}^{n}\sum_{\text{t=1}}^{\text{Time}} d\,(\exp(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 +$$
$$e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}))$$
$$+\sum_{\text{i=1}}^{n}\sum_{\text{t=1}}^{\text{Time}} q_{it}\; d\,(e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 +$$
$$e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}})$$
$$= -\frac{1}{\sigma^2_{\text{xself}}}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n + \sigma^2_{\text{xself}}1_n)^T\text{diag}((x_{\text{self}})^{-1})dx_{\text{self}}$$
$$+(\sum_{i=1}^{n}\sum_{\text{t=1}}^{\text{Time}} q_{it}\; e_i^T q_{t-1}\, e_i^T dx_{\text{self}}$$
$$-\exp(e_i^T y_{\text{self}} e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}})\times$$
$$e_i^T q_{t-1} e_i^T dx_{\text{self}})$$
$$\nabla\, l(x_{\text{self}}) = -\frac{1}{\sigma^2_{\text{xself}}}\text{diag}(x_{\text{self}})^{-1}(\log(-x_{\text{self}}) - \mu_{\text{xself}}1_n + \sigma^2_{\text{xself}}1_n) + (\sum_{\text{i=1}}^{n}\sum_{\text{t=1}}^{\text{Time}} q_{it}\; e_i q^T_{t-1}\, e_i$$
$$-\exp\{(p^t - \bar{p}^t)^T e_i y^T_{\text{self}} e_i + (y^i_{cross})^T(p^t_{-i} - \bar{p}^t_{-i}) + y_0^T e_i + q^T_{t-1} e_i x^T_{\text{self}} e_i + (x^i_{\text{cross}})^T q_{-i,t-1}\} e_i q^T_{t-1} e_i) \tag{C.3}$$

C.4 Hessian of x_{self}

$$
\begin{aligned}
&Hl(x_{\text{self}}) = \frac{\partial \nabla l(x_{\text{self}})}{\partial x_{\text{self}}} = \frac{\partial}{\partial x_{\text{self}}} \frac{\partial f(x_{\text{self}})}{\partial x_{\text{self}}^T} = \frac{\partial^2 f(x_{\text{self}})}{\partial x_{\text{self}} \partial x_{\text{self}}^T} \\
&= \frac{\partial}{\partial x_{\text{self}}} (\text{diag}(x_{\text{self}})^{-1}(\log(-x_{\text{self}}) - \mu_{\text{xself}} 1_n + \sigma^2_{\text{xself}} 1_n) + [\sum_{\text{i=1}}^{n} \sum_{\text{t=1}}^{\text{Time}} q_{it}\, e_i\, q_{t-1}^T\, e_i \\
&- \exp\{(p^t - \bar{p}^t)^T e_i y_{\text{self}}^T e_i + (y^i_{cross})^T (p^t_{-i} - \bar{p}^t_{-i}) + y_0^T e_i + q_{t-1}^T e_i x_{\text{self}}^T e_i + (x^i_{\text{cross}})^T q_{-i,t-1}\} e_i q_{t-1}^T e_i]) \\
&= \frac{\partial}{\partial x_{\text{self}}} (\begin{bmatrix} \frac{1}{x^1_{\text{self}}} & 0 & \cdots & 0 \\ \vdots & & \frac{1}{x^i_{\text{self}}} & \vdots \\ 0 & \cdots & 0 & \frac{1}{x^n_{\text{self}}} \end{bmatrix} \begin{bmatrix} \log(-x^1_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}} \\ \log(-x^2_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}} \\ \vdots \\ \log(-x^n_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}} \end{bmatrix} + [\sum_{\text{i=1}}^{n} \sum_{\text{t=1}}^{\text{Time}} q_{it} e_i\, q_{t-1}^T\, e_i \\
&- \exp\{(p^t - \bar{p}^t)^T e_i y_{\text{self}}^T e_i + (y^i_{cross})^T (p^t_{-i} - \bar{p}^t_{-i}) + y_0^T e_i + q_{t-1}^T e_i x_{\text{self}}^T e_i + (x^i_{\text{cross}})^T q_{-i,t-1}\} e_i q_{t-1}^T e_i]) \\
&= \frac{\partial}{\partial x_{\text{self}}} (\begin{bmatrix} (x^1_{\text{self}})^{-1}(\log(-x^1_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}}) \\ (x^2_{\text{self}})^{-1}(\log(-x^2_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}}) \\ \vdots \\ (x^n_{\text{self}})^{-1}(\log(-x^n_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}}) \end{bmatrix} + [\sum_{\text{i=1}}^{n} \sum_{\text{t=1}}^{\text{Time}} q_{it}\, e_i q_{t-1}^T\, e_i \\
&- \exp\{(p^t - \bar{p}^t)^T e_i y_{\text{self}}^T e_i + (y^i_{cross})^T (p^t_{-i} - \bar{p}^t_{-i}) + y_0^T e_i + q_{t-1}^T e_i x_{\text{self}}^T e_i + (x^i_{\text{cross}})^T q_{-i,t-1}\} e_i q_{t-1}^T e_i]) \\
&= \begin{bmatrix} -(x^1_{\text{self}})^{-2}(\log(-x^1_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}}) + (x^1_{\text{self}})^{-2} & 0 & \cdots & 0 \\ \vdots & & -(x^i_{\text{self}})^{-2}(\log(-x^i_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}}) + (x^i_{\text{self}})^{-2} & \vdots \\ 0 & \cdots & 0 & -(x^n_{\text{self}})^{-2}(\log(-x^n_{\text{self}}) - \mu_{\text{xself}} + \sigma^2_{\text{xself}}) + (x^n_{\text{self}})^{-2} \end{bmatrix} \\
&- \sum_{i=1}^{n} \sum_{\text{t=1}}^{\text{Time}} [\exp\{(p^t - \bar{p}^t)^T e_i\, y_{\text{self}}^T e_i + (y^i_{cross})^T (p^t_{-i} - \bar{p}^t_{-i}) + y_0^T e_i\} \times \\
&\exp\{q_{t-1}^T e_i x_{\text{self}}^T e_i + (x^i_{\text{cross}})^T q_{-i,t-1}\} e_i q_{t-1}^T e_i e_i^T q_{t-1} e_i^T]
\end{aligned}
\tag{C.4}
$$

C.5 Full conditional PDF of $x^i_{\text{cross}}\ \forall i$

$$
\begin{aligned}
&\mathrm{f}(\ x^i_{\text{cross}}) \propto \mathrm{f}(x^i_{\text{cross}}|\text{everything else}) \propto \mathrm{f}(\ x^i_{\text{cross}}, \text{everything else})\\
&\propto \mathrm{f}(\ x^i_{\text{cross}}|\mu_{\text{xcross}}1_{(n-1)}, \sigma^2_{\text{xcross}}I)\mathrm{f}(q_{it}|\ y_{\text{self}}, y^i_{\text{cross}}, y_0, p, x_{\text{self}}, x^i_{\text{cross}}, q_{t-1})\\
&\propto \frac{1}{|2\pi|^{\frac{n-1}{2}}|\sigma^2_{\text{xcross}}I|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(x^i_{\text{cross}}-\mu_{\text{xcross}}1_{(n-1)})^T(\sigma^2_{\text{xcross}}I)^{-1}(x^i_{\text{cross}}-\mu_{\text{xcross}}1_{(n-1)})\}\times\\
&(\prod_{\text{t=1}}^{\text{Time}}\exp\{-\exp(e_i^T y_{\text{self}}e_i^T(p^t-\bar{p}^t)+(p^t_{-i}-\bar{p}^t_{-i})^T y^i_{\text{cross}}+\\
&e_i^T y_0 + e_i^T x_{\text{self}}e_i^T q_{t-1}+(q_{-i,t-1})^T x^i_{\text{cross}})\}\times\\
&\frac{\exp\{e_i^T\ y_{\text{self}}\ e_i^T(p^t-\bar{p}^t)+(p^t_{-i}-\bar{p}^t_{-i})^T y^i_{cross}+e_i^T y_0+e_i^T x_{\text{self}}e_i^T q_{t-1}+(q_{-i,t-1})^T x^i_{\text{cross}}\}^{q_{it}}}{q_{it}!})\\
&\propto \exp\{-\frac{1}{2}(x^i_{\text{cross}}-\mu_{\text{xcross}}1_{(n-1)})^T(\sigma^2_{\text{xcross}}I)^{-1}(x^i_{\text{cross}}-\mu_{\text{xcross}}1_{(n-1)})\}\times\\
&(\prod_{\text{t=1}}^{\text{Time}}\exp\{-\exp(e_i^T y_{\text{self}}e_i^T(p^t-\bar{p}^t)+(p^t_{-i}-\bar{p}^t_{-i})^T y^i_{\text{cross}}+\\
&e_i^T y_0 + e_i^T x_{\text{self}}e_i^T q_{t-1}+(q_{-i,t-1})^T x^i_{\text{cross}})\}\times\\
&\exp\{e_i^T\ y_{\text{self}}\ e_i^T(p^t-\bar{p}^t)+(p^t_{-i}-\bar{p}^t_{-i})^T y^i_{cross}+e_i^T y_0+e_i^T x_{\text{self}}e_i^T q_{t-1}+(q_{-i,t-1})^T x^i_{\text{cross}}\}^{q_{it}})
\end{aligned}
\tag{C.5}
$$

C.6 Loglikelihood of $x^i_{\textbf{cross}}\ \forall\, i$

$$
\begin{aligned}
&l(x^i_{\text{cross}}) = \log \mathrm{f}(x^i_{\text{cross}}) \\
&\propto -\frac{1}{2}(x^i_{\text{cross}} - \mu_{\text{xcross}} 1_{(n-1)})^T (\sigma^2_{\text{xcross}} I)^{-1} (x^i_{\text{cross}} - \mu_{\text{xcross}} 1_{(n-1)}) \\
&- \sum_{\text{t=1}}^{\text{Time}} \exp(e_i^T y_{\text{self}} e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}) \\
&+ \sum_{\text{t=1}}^{\text{Time}} q_{it} [e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}] \\
&= -\frac{1}{2\sigma^2_{\text{xcross}}} (x^i_{\text{cross}} - \mu_{\text{xcross}} 1_{(n-1)})^T (x^i_{\text{cross}} - \mu_{\text{xcross}} 1_{(n-1)}) \\
&- \sum_{\text{t=1}}^{\text{Time}} \exp(e_i^T y_{\text{self}} e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{\text{cross}} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}) \\
&+ \sum_{\text{t=1}}^{\text{Time}} q_{it} [e_i^T\, y_{\text{self}}\, e_i^T (p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}]
\end{aligned}
\tag{C.6}
$$

C.7 Gradient of x^i_{cross} $\forall\, i$

$$
\begin{aligned}
d(l(x^i_{\text{cross}})) &= -\frac{1}{2\sigma^2_{\text{xcross}}}[d(x^i_{\text{cross}})^T(x^i_{\text{cross}} - \mu_{\text{xcross}}1_{(n-1)}) + (x^i_{\text{cross}} - \mu_{\text{xcross}}1_{(n-1)})^T d(x^i_{\text{cross}})] \\
&- \sum_{\text{t=1}}^{\text{Time}} d[\exp\{e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}\}] \\
&+ \sum_{\text{t=1}}^{\text{Time}} q_{it}\, d[e_i^T\, y_{\text{self}}\, e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0 + e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}] \\
&= -\frac{1}{\sigma^2_{\text{xcross}}}(x^i_{\text{cross}} - \mu_{\text{xcross}}1_{(n-1)})^T d(x^i_{\text{cross}}) + [\sum_{\text{t=1}}^{\text{Time}} q_{it}\,(q_{-i,t-1})^T d(x^i_{cross}) \\
&- (\exp\{e_i^T y_{\text{self}} e_i^T(p^t - \bar{p}^t) + (p^t_{-i} - \bar{p}^t_{-i})^T y^i_{cross} + e_i^T y_0\}\times \\
&\exp\{e_i^T x_{\text{self}} e_i^T q_{t-1} + (q_{-i,t-1})^T x^i_{\text{cross}}\}(q_{-i,t-1})^T d(x^i_{cross}))] \\
\nabla\, l(x^i_{\text{cross}}) &= -\frac{1}{\sigma^2_{\text{xcross}}}(x^i_{\text{cross}} - \mu_{\text{xcross}}1_{(n-1)}) + [\sum_{\text{t=1}}^{\text{Time}} q_{it}\,(q_{-i,t-1}) \\
&- \exp\{(p^t - \bar{p}^t)^T e_i\, y^T_{\text{self}}\, e_i + (y^i_{cross})^T(p^t_{-i} - \bar{p}^t_{-i}) + \\
&y_0^T e_i + q^T_{t-1} e_i x^T_{\text{self}} e_i + (x^i_{\text{cross}})^T q_{-i,t-1}\}(q_{-i,t-1})]
\end{aligned}
\tag{C.7}
$$

C.8 Hessian of x^i_{cross} $\forall\, i$

$$
\begin{aligned}
Hl(x^i_{\text{cross}}) &= \frac{\partial\, \nabla\, l(x^i_{\text{cross}})}{\partial\, x^i_{\text{cross}}} = -\frac{1}{\sigma^2_{\text{xcross}}} I_{(n-1)} \\
&- \sum_{\text{t=1}}^{\text{Time}} [\exp\{(p^t - \bar{p}^t)^T e_i\, y^T_{\text{self}}\, e_i + (y^i_{cross})^T(p^t_{-i} - \bar{p}^t_{-i}) + y_0^T e_i\}\times \\
&\exp\{q^T_{t-1} e_i x^T_{\text{self}} e_i + (x^i_{\text{cross}})^T q_{-i,t-1}\}(q_{-i,t-1})(q_{-i,t-1})^T]
\end{aligned}
\tag{C.8}
$$

Bibliography

[1] R. C. Blattberg and S. A. Neslin, "Sales promotion: The long and the short of it," *Marketing Letters*, vol. 1, pp. 81–97, 1989.

[2] R. C. Blattberg and K. J. Wisniewski, "Price-induced patterns of competition," *Marketing Science*, vol. 8, no. 4, pp. 291–309, 1989.

[3] V. Kumar and R. P. Leone, "Measuring the effect of retail store promotions on brand and store substitution," *Journal of Marketing Research*, vol. 25, no. 2, pp. 178–185, 1988.

[4] A. Musalem, M. Olivares, S. Borle, H. Che, C. T. Conlon, K. Girotra, S. Gupta, K. Misra, J. H. Mortimer, G. Vulcano, and F. Zheng, "A review of choice modeling in the marketing-operations management interface," *Kelley School of Business Research Paper*, pp. 17–60, 2017.

[5] A. G. Kök and M. L. Fisher, "Demand estimation and assortment optimization under substitution: Methodology and application," vol. 55, no. 6, pp. 1001–1021, 2007.

[6] P. E. Rossi, G. M. Allenby, and R. McCulloch, *Bayesian Statistics and Marketing*. John Wiley and Sons, Ltd, 2005.

[7] A. Zellner and P. E. Rossi, "Bayesian analysis of dichotomous quantal response models," *Journal of Econometrics*, vol. 25, pp. 365–394, 1984.