学習院大学大学院経済学研究科
博士学位論文

Essays on Statistical Analysis in Revenue Management and Selecting Populations for Dependent Data

(レベニューマネジメントにおける統計的分析 および
従属データの母集団選択問題の研究)

Shizue Sunami

February, 2015

Acknowledgements

In my life, I had never thought that I would be writing a Ph.D. thesis. It was possible only because I was studying under the mentorship of Professor Jun-ichiro Fukuchi. He taught me and led me in the research throughout my years of Master course and Ph.D. course. He has been far more patient with me than anybody can expect when I have difficulties in overcoming my weakness in various areas. He made it possible for me to genuinely enjoy studying and researching. The study and research at Gakushuin are very precious and valuable for my life and I owe such experience entirely to Professor Fukuchi. I would like to thank him sincerely for his kind guidance and support.

In completing the thesis, Professor Katsuto Tanaka and Associate Professor Naoki Nagashima took time to check my thesis and gave me many advices and suggestions in details for which I am very grateful.

The data used in the empirical analysis in Chapter 2 was provided by SCIGNUS. It is often very difficult to get hotels' raw booking and price data. They spared no efforts to collect all the data necessary for the analysis. Without them, the analysis had not been possible. Their kind efforts and generosity are very much appreciated. .

Lastly, I would like to extend my thanks to my 96 year old mother for her love and support.

Table of Contents

Abstract

This thesis is organized in the following manner.

Chapter 1:
In Chapter 1, we describe what revenue management is and statistical methods applied for demand analysis in revenue management. The pick up method to forecast demand, one of the most robust forecast methods for the hotel or airline demand, is discussed in details. The demand is often censored when the resource is limited and cannot be easily changed. The EM algorithm developed by Dempster, Laird, and Rubin (1977) is presented to estimate the true demand for two cases; one where the demand has a normal distribution and the other where it has a Poisson distribution. Further discussed are the regression models to analyze demand. Three regression models reviewed in this chapter are a linear regression model and count regression models, Poisson and negative binomial. A count regression model is often appropriate for hotel or airline demand analysis as the demand analyzed is discrete. The EM algorithm is also presented for linear regression and Poisson regression models for censored data.

Chapter 2:
In Chapter 2, we estimate univariate price response functions for the hotel room demand to determine the hotel price effects on demand. When the parameter estimates for the other hotel prices are significant and positive, we can determine that the hotels are price competitive. Empirical study was made using the booking data of three major hotels. Prices hotels offer are numerous and fluctuate daily. So the issue is which price should be used for the analysis. Since customers are sensitive to low prices, 5 price categories are created from the low prices. First descriptive analysis is made, and then statistical analysis which applies Poisson and negative binomial regression models.

Chapter 3:
In Chapter 3, we investigate the selection procedure that selects, from k populations, the subset that contains the largest mean so that the probability of such selection is larger than or equal to the prespecified value, P^*. This methodology was first developed by Gupta (1963) for independent sequences of random variables. In this chapter, a method for dependent sequences of random variables is investigated and the selection procedure for the Gaussian ARMA processes is presented. The result of a simulation study is provided in order to see finite sample properties of the suggested selection rule.

1 Statistical methods related to revenue management

1.1 Revenue management

1.1.1 What is revenue management ?

Revenue management (RM) was originally developed in the airline industry. In the early 1970's, the airline industry was deregulated and price competition became fierce. The new airlines came in the market and the existing major airlines faced this serious price competition. The new airlines were able to offer low prices as they structured the company so that they could operate flights with low cost. However, the existing airlines could not compete with low prices as they already had high fixed costs for airplanes, salaries etc. Under this circumstance revenue management was developed where capacity, the facilities available and required to offer service (e.g. seats in case of airlines), is controlled instead of prices. Refer to Cross (1998) for further RM history details.

Prices are set for each market segment (a group of people that share one or more characteristics) and various products developed for each market segment are determined to be offered or not to be offered at each point of sale to maximize revenue. For example, leisure passengers can reserve flights far in advance, while business passengers tend to book or change flights very close to departure dates. With these market segments (leisure and business), airlines can sell some of the seats to leisure passengers in advance, keeping just enough seats for business passengers willing to pay higher fares. This results in higher revenue from selling as many seats as possible if the sale of seats is controlled well enough using the fare restrictions (for example, the fares for leisure passengers have an advance booking requirement, restriction on cancellation etc.).

There is no simple definition for revenue management officially or widely accepted. It is often said to be the science to sell the right products to the right customer at the right time for the right price. In Shizue Sunami (2009), revenue management is defined to be "the method to control the capacity based on the demand estimate to maximize revenue in the industries with fixed capacity ". This is a narrowly defined definition of revenue management. Talluri and Van Ryzin (2005) say that revenue management is applied even in the manufacturing industry where price is controlled.

The characteristics of the industries where revenue management is effective are;

1. Capacity is fixed or not easy to change in a short term.

When capacity can be easily changed, the revenue maximization is possible by changing the capacity to meet the demand. However, immediate capacity change is not possible when capacity is fixed (e.g. aircrafts for airlines, guest rooms for hotels etc.) Then, the firms must make do with what they have.

2. Fixed cost is high, while variable cost is low.

In general, variable costs fluctuate in proportion to the goods sold. However, variable cost fluctuation is small in case of the industries such as airlines and hotels as the capacities are fixed and most of the costs are incurred from this fixed capacity and therefore, fixed cost is high.

3. Products for sale are perishable.

Products for sale cannot be inventoried once the sale date passes a certain date. For example, the guest room stay for a certain date no longer has any value after that date.

4. Demand is predictable.

A demand forecast is necessary for a good capacity control to well utilize the capacity.

5. Customer segmentation (market segmentation) is possible.

Different prices are set for the same resource. The availability for the products priced accordingly may be changed based on the demand forecast during the booking process. The prices have different restrictions (advance reservation, restricted duration, time of travel etc. in case of airlines) which deter some market segments from purchasing the products and attract other market segments.

1.1.2 Industries that can apply revenue management

Revenue management is well established in the airline industry as they have developed and applied revenue management roughly for the last 40 years. RM is also well practiced in the hotel industry which has the characteristics mentioned in the section 1.1.1. Although there are a lot of similarities between airlines and hotels, there are some differences. Hotels generate revenue from not only rooms but also banquet rooms, restaurants, health clubs etc. It will be very complex if RM considers all these aspects. Therefore, the hotel RM in principle handles only guest room revenue. Another unique point for the hotel RM is the length of stay. For example, they may accept only the stays over a certain duration during a high-demand period. On the other hand, they may set an upper limit on the duration of stay so that guests staying for a long period and paying a low price would not displace the guests staying for a short period and paying a higher price.

Revenue management is also practiced in other industries such as cruise lines, rent-a-car, retailing, and manufacturing, although it is a rather new practice in some of them. Each industry has their own characteristics different from airlines. Some retailers also have to deal with perishable products. Grocery retailers sell highly perishable products, so do apparel retailers. High-tech retailers have problems with perishable products, because high technology makes high-tech products obsolete very quickly. Although we can say these retailer sell perishable products, they control sale of products by discounts and promotions, not by capacity like airlines. So it is not a classical RM these retailers apply.

1.2 Demand analysis for revenue management

1.2.1 Pickup model

The demand estimate is a critical factor for revenue management. Decisions for price setting, availability of products for sale, i.e. whether to open or close the sale of products, etc. all require demand forecast. Inaccurate forecast results in revenue decline.

Weatherford and Kimes (2003) compared several forecasting methods including pickup methods, linear regression and moving average, using the hotel data. The results show the pickup method and the regression produced the lowest error.

The pickup model is a model often applied to estimate the airline or hotel reservations. Skwarek (1994) and Wickam (1993) provided the details of the model and Fukuchi and Sunami (2005) statistically expressed the model and explained the EM algorithm.

Sunami (2009) estimated the daily room nights sold applying the advanced pickup model for a resort hotel that provided the booking data. The forecast was made 28 days prior to the stay nights for 7 months. Two pickup models were used, one using the one day pickup period and the other one week. (See below for the definition of pickup.) Pickup is assumed to have a Poisson distribution for one day pickup period and a normal distribution for one week pickup period. Censored demand was unconstrained using the EM algorithm explained later in this section. The mean absolute error (MAE) was determined, $\sum_{i=1}^{n} |\hat{Y}_i - Y_i|/n$, where Y_i is the room nights sold on the i stay night and $\hat{Y}_i$ is its estimate. The one week pickup period worked slightly better.

The pickup model and the EM algorithm are next reviewed based on Fukuchi and Sunami (2005), using the case of the hotel FIT (free individual traveler) booking (reservation) estimate.

booking data

Booking data is the time-series room reservation data for a specific arrival day. A hotel keeps a daily record of no. of bookings and cancellations from the day when the first booking is made till the arrival day. Pickup is the incremental booking during a certain interval called a pickup period. The pickup period is to be one week in this section. Let $Y_0(0)$ be the bookings on hand for a specific arrival day on that day, and $Y_0(t)$ be the bookings on the day, t weeks prior. Then the booking data is;

$$Y_0(0), Y_0(1), Y_0(2), \ldots$$

Let $Y_i(t)$ be the bookings for the arrival day, i weeks prior to the subject arrival day on the day, t weeks prior. Then the booking data for the arrival day, i weeks prior, is

$$Y_i(0), Y_i(1), Y_i(2), \ldots$$

The pickup for the period from t weeks prior till $t-1$ weeks prior is

$$X_i(t) := Y_i(t-1) - Y_i(t).$$

additive pick up model

Suppose the booking data for the same day of the week as the arrival day on the subject is available and that K booking data are closed. We are to forecast the bookings on hand for the arrival day on the days, $0, 1, 2, \ldots, \ell-1$ weeks prior;

$$Y_0(0), Y_0(1), Y_0(2), \ldots Y_0(\ell-1).$$

The total number of FIT bookings has an upper limit called booking limit as the total number of rooms, C, is fixed. Deducting the group booking, $G_i(t)$, from the total rooms and then adding the overbooking upper limit, $O_i(t)$, gives the booking limit:

$$U_i(t) = C - G_i(t) + O_i(t). \tag{1.1}$$

It is natural to consider $G_i(t)$ is also a random variable. Therefore,

$$Y_i(t) \leq U_i(t), \ t = 1, 2, \ldots$$

In this paper, the overbooking upper limit is not considered although it is a very important factor. Refer to Chapter 4 in Talluri and Van Ryzin (2004) in regards to how to determine the overbooking upper limit. The booking limit leads to the pickup upper limit. The pickup upper limit on the day t weeks prior,

$$c_i(t) = U_i(t-1) - Y_i(t). \tag{1.2}$$

Now the stochastic model is defined to express the time series change in the bookings on hand.

Let T_i be the day when the first booking is made. T_i is assumed to be constant and dependent on i. Let $D_i(t)$ be the actual incremental demand from t weeks prior to $t-1$, which would be the pickup if no upper limit existed. It is also a random variable.

Pickup is not necessarily the incremental demand, because the daily bookings are limited by the upper limit, $c_i(t)$. We define the model,

$$Y_i(t) = Y_i(t+1) + X_i(t+1), \ t = 0, 1, 2, \ldots, T_i - 1 \tag{1.3}$$

$$Y_i(T_i + 1) = 0. \tag{1.4}$$

$X_i(t)$ is determined to be;

$$X_i(t) = \begin{cases} D_i(t), & \text{when } D_i(t) \leq c_i(t) \\ c_i(t), & \text{when } D_i(t) > c_i(t). \end{cases} \tag{1.5}$$

The equation (1.5) states that the incremental demand is censored when it exceeds the pickup upper limit.

For the random variable $D_i(t), \ t = 1, 2, \ldots, T_i, \ i = 0, 1, 2, \ldots, n$, two assumptions are made;

assumption 1: For all t, $D_0(t), D_1(t), D_2(t), \ldots, D_n(t)$ are independently and identically distributed.

assumption 2: For all i, $D_i(1), D_i(2), \ldots, D_i(T_i)$ are independent, and the marginal distribution for each variable is not necessarily the same.

Let $E(D_i(t))$ be $\mu(t)$. The model expressed in the equations (1.3) through (1.5) is the additive pickup model.

$Y_0(\ell-1)$ can be expressed as;

$$Y_0(\ell-1) = Y_0(\ell) + X_0(\ell). \tag{1.6}$$

Therefore,

$$\hat{Y}_0(\ell-1) = Y_0(\ell) + \min\{\hat{\mu}(\ell), c_0(\ell)\} \tag{1.7}$$

is the natural estimator for $Y_0(\ell-1)$. The equation (1.7) is used sequentially to obtain the estimators of $Y_0(\ell-1), Y_0(\ell-2), \ldots Y_0(0)$;

$$\hat{Y}_0(\ell-1), \hat{Y}_0(\ell-2), \ldots, \hat{Y}_0(0). \tag{1.8}$$

The incremental demand estimator, $\hat{\mu}(t)$, is necessary to forecast the final number of rooms used on a specific arrival day, $Y_0(0)$. There are two additive pickup methods for the forecast. In this section, the incremental demand is assumed not to exceed the pickup upper limit, i.e. $X_i(t) = D_i(t)$. Then, the average of the historical pickups is the unbiased and appropriate estimator for $\mu(t)$.

classical pickup method

The classical pickup method uses only the booking data for the stay nights that arrived. Suppose the lead time is ℓ weeks to estimate $Y_0(0)$ and the booking data for the arrival days up to N weeks prior are available and to be used. The booking data used are those for the arrival days of the same day of the week as that of the arrival day for which a forecast is to be made. The complete booking data;

$$\begin{aligned}
&Y_\ell(0), Y_\ell(1), \ldots, Y_\ell(T_\ell)\\
&Y_{\ell+1}(0), Y_{\ell+1}(1), \ldots, Y_{\ell+1}(T_{\ell+1})\\
&\qquad\qquad \vdots\\
&Y_N(0), Y_N(1), \ldots, Y_N(T_N).
\end{aligned}$$

Then the estimator for $\mu_i(t)$,

$$\hat{\mu}_i(t) = \frac{1}{N-\ell+1}\sum_{i=\ell}^{N} X_i(t). \tag{1.9}$$

advanced pickup method

In addition to the complete booking data, the advance pickup method applies the partial booking data, the data available for the arrival days which have not arrived yet. When estimating the final number of rooms used on a specific arrival day ℓ weeks prior, the partial booking data is;

$$\begin{aligned} &Y_{\ell-1}(1), Y_{\ell-1}(2), \ldots, Y_{\ell-1}(T_{\ell-1}), \\ &Y_{\ell-2}(2), Y_{\ell-2}(3), \ldots, Y_{\ell-2}(T_{\ell-2}), \\ &\qquad\qquad\vdots \\ &Y_1(\ell-1), Y_1(\ell), \ldots, Y_1(T_1). \end{aligned} \tag{1.10}$$

For t such that $1 \leq t \leq \ell$, the estimator for $X_0(t)$,

$$\hat{\mu}(t) = \frac{1}{N-\ell+t} \sum_{i=\ell-t+1}^{N} X_i(t). \tag{1.11}$$

1.2.2 Estimation of the incremental demand from the censored data

When the resource has an upper limit and the observed demand is the same as the upper limit, the actual demand often exceeds it. Then the average of the observed demand is smaller than the average of the actual demand. In case of the pickup method, the pickup estimate would be lower than what it should be if censoring is not taken into consideration.

The maximum likelihood is often used to estimate parameters when the parametric form of the demand probability distribution is given.

First, we define

$$\mathcal{C} = \{i : 1 \leq i \leq n \ x_i = c_i\}$$

$$\mathcal{O} = \{i : 1 \leq i \leq n \ x_i < c_i\},$$

where $x_i (i = 1, 2, \ldots, n)$ is the observed incremental demand and $c_i (i = 1, 2, \ldots, n)$ is its upper limit. Let $f(z_i, \theta)$ be the probability density function of the incremental demand, z_i, where θ is the unknown parameter and $F(z_i, \theta)$ be its cumulative distribution function. Then, the likelihood function is

$$L(\theta) = \prod_{i \in \mathcal{O}} f(x_i, \theta) \prod_{i \in \mathcal{C}} (1 - F(x_i, \theta)), \tag{1.12}$$

given the observed incremental demand, $x_i (i = 1, 2, \ldots, n)$.

It is very difficult to numerically obtain the maximum likelihood estimator for the likelihood function (1.12) and the EM algorithm developed in Dempster et al. (1977) is generally applied in the demand estimate from the censored data. The EM algorithm is now presented for the normal distribution and the Poisson distribution. When the distribution has other parametric forms, the method is the same except that the different formulas for $\mathrm{E}[X|X > a]$ and $V[X|X > a]$ are applied.

EM algorithm-Normal distribution

Suppose the incremental demand, X, has the normal distribution, $N(\mu, \sigma^2)$. Let ϕ be the probability density function of the standard normal distribution and Φ be its cumulative distribution function. Then the likelihood function is

$$L(\mu, \sigma^2) = \prod_{i \in \mathcal{O}} \frac{1}{\sigma} \phi\left(\frac{x_i - \mu}{\sigma}\right) \prod_{i \in \mathcal{C}} \left\{1 - \Phi\left(\frac{x_i - \mu}{\sigma}\right)\right\}. \tag{1.13}$$

We now obtain the MLE for (μ, σ) with the EM algorithm. The 2 formulas required are;

$$\mathrm{E}[X|X > a] = \mu + \frac{\phi(a^*)}{1 - \Phi(a^*)}\sigma, \tag{1.14}$$

$$V[X| > a] = \left[1 + \frac{a^*\phi(a^*)}{1 - \Phi(a^*)} - \left\{\frac{\phi(a^*)}{1 - \Phi(a^*)}\right\}^2\right]\sigma^2, \tag{1.15}$$

where

$$a^* = \frac{a - \mu}{\sigma}.$$

The steps for the EM algorithm are as follows. Note that $a_i (i = 1, 2, \ldots, n)$ is the pickup upper limit.

Step 1 (Initialize μ and σ^2): Obtain the sample average and variance of the uncensored incremental demand to be used as the initial values, $(\mu^{(0)}, (\sigma^2)^{(0)})$. Let $k = 0$.

Step 2: (E step): Set

$$a_i^{*(k)} = \frac{a_i - \mu^{(k)}}{\sigma^{(k)}}.$$

Obtain

$$\mu_{a_i}^{(k)} = \mu^{(k)} + \frac{\phi(a_i^{*(k)})}{1 - \Phi(a_i^{*(k)})}\sigma^{(k)} \tag{1.16}$$

$$v_{a_i}^{(k)} = \left[1 + \frac{a_i^{*(k)}\phi(a_i^{*(k)})}{1 - \Phi(a_i^{*(k))})} - \left\{\frac{\phi(a_i^{*(k)})}{1 - \Phi(a_i^{*(k)})}\right\}^2\right](\sigma^2)^{(k)}. \tag{1.17}$$

Compute

$$S^{(k)} = \sum_{i \in \mathcal{O}} x_i + \sum_{i \in \mathcal{C}} \mu_{a_i}^{(k)}, \tag{1.18}$$

$$S_{11}^{(k)} = \sum_{i \in \mathcal{O}} x_i^2 + \sum_{i \in \mathcal{C}} \left\{(\mu_{a_i}^{(k)})^2 + v_{a_i}^{(k)}\right\}. \tag{1.19}$$

step 3: (M step) Obtain

$$\mu^{(k+1)} = \frac{1}{n}S^{(k)}, \tag{1.20}$$

$$(\sigma^2)^{(k+1)} = \frac{1}{n}S_{11}^{(k)} - (\mu^{(k+1)})^2. \tag{1.21}$$

Repeat the step 2 and step 3 until the parameter estimates converge.

EM algorithm - Poisson distribution

Assume that the incremental demand, X, has a Poisson distribution with the mean, λ. The likelihood function is

$$L(\lambda) = \prod_{i \in \mathcal{O}} p(x_i, \lambda) \prod_{i \in \mathcal{C}} P(x_i, \lambda), \tag{1.22}$$

where

$$p(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \tag{1.23}$$

$$P(c, \lambda) = 1 - \sum_{j=0}^{c} \frac{e^{-\lambda}\lambda^j}{j!}. \tag{1.24}$$

We now obtain the MLE with the EM algorithm using,
Formula 1:

$$\mathrm{E}(X|X > C) = \lambda \left(1 + \frac{p(C, \lambda)}{P(C, \lambda)}\right). \tag{1.25}$$

See the proof for this formula in Appendix A.

The EM algorithm is now provided. Note that $c_i (i = 1, 2, \ldots, n)$ is the upper limit for X_i.
Step 1. (Initialize λ): Obtain the sample average of the uncensored incremental demand, pickup, to be used as the initial value, λ_0, and set $k = 0$.
Step 2. (E step): Obtain

$$\lambda^{(k+1)} = \lambda^{(k)} \left(1 + \frac{p(c_i, \lambda^{(k)})}{P(c_i, \lambda^{(k)})}\right), \tag{1.26}$$

$$w^{(k+1)} = \sum_{i \in \mathcal{O}} x_i + \sum_{i \in \mathcal{C}} \lambda_i^{(k+1)}. \tag{1.27}$$

Step 3. (M step): Compute

$$\lambda^{(k+1)} = \frac{1}{n} w^{(k+1)}. \tag{1.28}$$

Repeat the step 2 and step 3 till the parameter estimate converges.

1.3 Regression models useful for demand analysis

In this section, linear and count regression models are reviewed. As used in Chapter 2, the regression models are useful and practical in analyzing demand. In Revenue Management, the data is often censored as the capacities are limited. Therefore, we discuss such cases for regression models.

1.3.1 Linear regression model

Often used for demand analysis is the linear regression model;

$$y_i = \boldsymbol{x_i'}\boldsymbol{\beta} + \varepsilon_i, \tag{1.29}$$

where the variable $y_i,\ i = 1, 2, \ldots, N$, is a dependent variable, while $x_{ij},\ i = 1, 2, \ldots, N, j = 1, 2, \ldots, J$ are independent variables. $\varepsilon_i \sim \text{NID}(0, \sigma^2),\ i = 1, 2, \ldots, N$.

Then, the estimates of the parameters,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}. \tag{1.30}$$

Weatherford and Kimes (2003) applied the regression model to estimate the hotel final bookings on the arrival day (day 0) using the bookings on hand currently (day ℓ) as an independent variable (e.g., Forecast@Day 0=a+b x Bookings@Dayℓ). In Lee (2011), the linear hotel demand models are suggested using prices, average competitor prices, days prior, length of stay, day of week of arrival date, month of arrival date and Saturday night stay, as the possible explanatory variables. Steiner et al. (2007) estimate the price impact on unit sales with a linear regression model using retail scanner data. In Stefanescu (2009), a multivariate demand model is developed allowing correlated and censored demand; $D(t) = a(t) + W(t) * v + \varepsilon(t)$. The vector $D(t)$ denotes the random demand in period t and v is a random shock; $v \sim (\mathbf{0}, \sum_v)$. The influence of the random shock on the demand for each period t is weighted by the $n \times n$ diagonal matrix $\boldsymbol{W}(t)$. The error terms, $\varepsilon(t)$, are normally distributed, $\varepsilon(t) \sim N(\mathbf{0}, \boldsymbol{I}_n\sigma_e^2)$ and are assumed to be independent across time periods $t = 1, 2, \ldots, T$ and to be independent of the random shock v.

1.3.2 Count regression model

Poisson regression model

A count variable, $Y_i, i = 1, 2, \ldots, N$, is independently distributed and has a Poisson distribution with mean, λ_i,

$$P(Y_i = y) = \frac{e^{-\lambda_i}\lambda_i^y}{y!}. \tag{1.31}$$

The Poisson distribution has the equidispersion property, that is,

$$\text{E}[Y_i] = \lambda_i,$$

$$V[Y_i] = \lambda_i.$$

The Poisson regression model is obtained by setting

$$\text{E}[Y_i] = \lambda_i = \exp(\boldsymbol{x_i'\beta}), \tag{1.32}$$

where $\boldsymbol{x_i}$ is an independent variable, a vector with J elements, i.e., $x_{ij}, j = 1, 2, \ldots, , J$. As $V[Y_i|X_i] = \exp(\boldsymbol{x_i'\beta})$, the Poisson regression is heteroskedastic.

Since the observations, Y_i, are independent, we can use the log likelihood function to estimate $\boldsymbol{\beta}$.

As the log likelihood function is globally concave, the MLE can be obtained by the Newton-Raphson method.

For any model with $\mathrm{E}(Y|X) = \exp[\boldsymbol{x'\beta}]$, differentiation yields

$$\frac{\partial \mathrm{E}[Y|X]}{\partial x_j} = \beta_j \exp(\boldsymbol{x'\beta}), \tag{1.33}$$

where x_j denotes the jth regressor.

The equation (1.33) implies that one unit change in x_j increases $\beta_j \exp(\mathbf{x}'\beta)$ in $\mathrm{E}[Y|X]$. Thus $100\beta_j$ is the percentage increase of the expectation of Y.

The test statistic,

$$z = \frac{\widehat{\beta}_j}{\widehat{se}(\widehat{\beta}_j)}$$

is used for testing : $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$. This statistic will follow a standard normal distribution asymptotically when H_0 is true. If H_0 is false, $|z|$ will tend to be large. We then compare z to the critical value of the standard normal distribution to determine the significance of the parameter estimate, $\hat{\beta}_j$.

See Winkelmann (2010) and Cameron and Trivedi (2005) for the details of the Poisson regression model.

Negative binomial regression model

The negative binomial distribution is one of the probability distributions used to model count data and does not have the property of equidisperion that the Poisson distribution has. It is a mixture of Poisson and gamma distributions. Its derivation is as follows;

Let y have a Poisson distribution with mean, $\tilde{\lambda}$,

$$f(y|\tilde{\lambda}) = \frac{e^{-\tilde{\lambda}}\tilde{\lambda}^y}{y!}. \tag{1.34}$$

$\tilde{\lambda} = \lambda\nu$ and ν has a distribution function $g(\nu) = \nu^{\delta-1}e^{-\nu\delta}\delta^{\delta}/\Gamma(\delta), \nu, \delta > 0$. This is a gamma distribution with $\mathrm{E}[\nu] = 1$ and $V[\nu] = 1/\delta$. Then, the marginal probability density function of $y|\lambda, \alpha$, which is its probability density function, is

$$h[y|\lambda, \alpha] = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda}\right)^{\alpha^{-1}} \left(\frac{\lambda}{\alpha^{-1} + \lambda}\right)^{y} \alpha > 0,\ y = 0, 1, 2, \ldots, \tag{1.35}$$

where $\delta = \alpha^{-1}$. This is called the negative binomial 2 (NB2) model and its first two moments are

$$\mathrm{E}[y|\lambda, \alpha] = \lambda \tag{1.36}$$

and

$$V[y|\lambda, \alpha] = \lambda(1 + \alpha\lambda). \tag{1.37}$$

Let the count variable, $Y_i,\ t = 1, 2, \ldots N$, be independently distributed and have NB2 and $\boldsymbol{x_i}$ be a vector with J elements, i.e., $x_{ij}, j = 1, 2, \ldots, , J$. The negative binomial 2 regression model is defined as;

$$P(Y_i = y) = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i}\right)^{y} \alpha > 0,\ y = 0, 1, 2, \ldots, \tag{1.38}$$

where λ_i is set to be $\exp(\boldsymbol{x_i'\beta})$.

The parameters, $\boldsymbol{\beta}$, are estimated by Maximum Loglikelihood. As $\mathrm{E}(Y|X) = \exp[\boldsymbol{x'\beta}]$, the coefficients are interpreted in the same manner as for the case of Poisson regression model. Differentiation of $\mathrm{E}(Y|X) = \exp[\boldsymbol{x'\beta}]$ yields

$$\frac{\partial \mathrm{E}[Y|X]}{\partial x_j} = \beta_j \exp(\boldsymbol{x'\beta}), \tag{1.39}$$

where x_j denotes the jth regressor.

The equation (1.39) implies that one unit change in x_j increases $\beta_j \exp(\mathbf{x}'\beta)$ in $\mathrm{E}[Y|X]$. Thus $100\beta_j$ is the percentage increase of the expectation of Y.

The test statistic, $z = \widehat{\beta}_j / \widehat{se}(\widehat{\beta}_j)$, is used to test the significance of the parameter estimates as in case of Poisson regression.

See Winkelmann (2010) and Hilbe (2007) for further details of the negative binomial 2 (NB2) regression model.

1.3.3 Linear regression model with censored data

The linear regression model is;

$$y_i = \boldsymbol{x_i'\beta} + \varepsilon_i \quad (i = 1, 2, \ldots, n, n+1, \ldots, n+m), \tag{1.40}$$

where $\varepsilon_i \sim \mathrm{NID}(0, \sigma^2)$, $\boldsymbol{x_i} = (x_{i0}, x_{i1}, \ldots, .x_{iJ})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_J)'$. Then $y_i \sim N(\boldsymbol{x_i'\beta}, \sigma^2)$, giving

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \boldsymbol{x_i'\beta})^2}{2\sigma^2}\right].$$

In case of the complete data, the likelihood function is

$$\begin{aligned} L &= \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \boldsymbol{x_i'\beta})^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[\frac{-\sum_{i=1}^n y_i^2 + 2\sum_{j=0}^J \beta_j \sum_{i=1}^n x_{ij} y_i}{2\sigma^2}\right] \exp\left[\frac{-\sum_{i=1}^n (\boldsymbol{x_i'\beta})^2}{2\sigma^2}\right], \end{aligned} \tag{1.41}$$

where the sufficient statistics are $\sum_{i=1}^n y_i^2$ and $\sum_{i=1}^n x_{ij} y_i$.

Suppose Y_i is censored at c_i;

$$Y_i = \begin{cases} y_i & \text{if } y_i \le c_i \qquad\qquad i = 1, 2, \ldots, n \\ c_i & \text{if } y_i > c_i \quad c_i > 0 \quad i = n+1, n+2, \ldots, n+m. \end{cases} \tag{1.42}$$

We now estimate $\boldsymbol{\beta}$ and σ using the EM algorithm, where two formulas are used;
Formula 2

$$\mathrm{E}[Y|Y > C] = \mu + \frac{\phi(z)}{1 - \Phi(z)}\sigma. \tag{1.43}$$

Formula 3

$$\mathrm{E}[Y^2|Y > C] = \sigma^2 + \mu^2 + \sigma(C + \mu)\left(\frac{\phi(z)}{1 - \Phi(z)}\right), \tag{1.44}$$

where, $Y \sim N(\mu, \sigma^2)$, $z = (C-\mu)/\sigma$, $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$ and $\Phi(C) = \int_{-\infty}^{C} \phi(t)dt$. See the proof for these two formulas in Appendix B.

Let the parameter estimates at the kth iteration be $\boldsymbol{\beta}^{(k)}, \mu_i^{(k)}$ and, $\sigma^{2(k)}$ where $\mu_i^{(k)} = \boldsymbol{x_i}\boldsymbol{\beta}^{(k)}$.

The initial parameters, $(\boldsymbol{\beta}^{(0)}, \mu_i^{(0)}, \sigma^{2(0)})$, are set to be;
$\beta^{(0)} = \boldsymbol{(X'X)}^{-1}\boldsymbol{X'y}$,
$\mu_i^{(0)} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i'}\boldsymbol{\beta}^{(0)}$,
$\sigma^{2(0)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_i^{(0)})^2$,
using only the uncensored data.
E step

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n+m} x_{ij}Y_i\right] &= \sum_{i=1}^{n} x_{ij}y_i + \left[\sum_{i=n+1}^{n+m} x_{ij}\mathrm{E}[Y_i|Y_i \geq c_i, \boldsymbol{\beta}^{(k)}, \sigma^{2(k)}]\right] \\
&= \sum_{i=1}^{n} x_{ij}y_i + \sum_{i=n+1}^{n+m} x_{ij}(\mu_i^{(k)} + \sigma^{(k)}S(z_i^{(k)})) \qquad (1.45)
\end{aligned}$$

where $z_i^{(k)} = \frac{c_i - \mu_i^{(k)}}{\sigma^{(k)}}$ and $S(z_i^{(k)}) = \frac{\phi(z_i^{(k)})}{1-\Phi(z_i^{(k)})}$
The jth element of $\boldsymbol{(X'y)}^{(k)}$ is obtained in (1.45), using Formula 2.

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n+m} Y_i^2\right] &= \sum_{i=1}^{n} y_i^2 + \sum_{i=n+1}^{n+m} \mathrm{E}[y_i^2|Y_i \geq c_i, \boldsymbol{\beta}^{(k)}, \sigma^{2(k)}] \\
&= \sum_{i=1}^{n} y_i^2 + \sum_{i=n+1}^{n+m} (\mu_i^{2(k)} + \sigma^{2(k)} + \sigma^{(k)}(c_i + \mu_i^{(k)})S(z_i^{(k)})). \qquad (1.46)
\end{aligned}$$

M Step
Obtain $\boldsymbol{\beta}^{(k+1)}$, $\mu_i^{(k+1)}$, and $\sigma^{2(k+1)}$, where
$\boldsymbol{\beta^{(k+1)}} = \boldsymbol{(X'X)}^{-1}\boldsymbol{(X'y)}^{(k)}$, $\mu_i^{(k+1)} = \boldsymbol{x_i'}\boldsymbol{\beta}^{(k+1)}$, and

$$\begin{aligned}
(n+m)\sigma^{2(k+1)} &= \sum_{i=1}^{n}(y_i - \mu_i^{(k)})^2 + \sum_{i=n+1}^{n+m} (\mathrm{E}[(Y_i - \mu_i^{(k)})^2|Y_i \geq c_i, \boldsymbol{\beta}^{(k)}, \sigma^{(k)}]) \\
&= \sum_{i=1}^{n}(y_i - \mu_i^{(k)})^2 + \sum_{i=n+1}^{n+m} (\mathrm{E}[Y_i^2|A] - 2\mu_i^{(k)}\mathrm{E}[Y_i|A] + \mu_i^{2(k)}) \\
&= \sum_{i=1}^{n}(y_i - \mu_i^{(k)})^2 + \sum_{i=n+1}^{n+m} (\sigma^{2(k)} + \sigma^{2(k)}z_i^{(k)}S(z_i^{(k)})) \qquad (1.47) \\
&= \sum_{i=1}^{n}(y_i - \mu_i^{(k)})^2 + \sigma^{2(k)}\sum_{i=n+1}^{n+m} (1 + z_i^{(k)}S(z_i^{(k)})), \qquad (1.48)
\end{aligned}$$

$$\sigma^{2(k+1)} = \frac{1}{n+m}\left[\sum_{i=1}^{n}(y_i - \mu_i^{(k)})^2 + \sigma^{2(k)}\sum_{i=n+1}^{n+m} (1 + z_i^{(k)}S(z_i^{(k)}))\right], \qquad (1.49)$$

where $A = (Y_i \geq c_i, \boldsymbol{\beta}^{(k)}, \sigma^{(k)})$. The expression (1.47) is obtained using Formula 2 and Formula 3.

The E step and the M step are iterated until the parameter estimates converge. Refer to Aitkin (1981) for further details.

1.3.4 Poisson regression model with censored data

$Y_i \sim Po(\lambda_i)(i = 1, \ldots, n, n+1, \ldots, n+m)$. Let $\lambda_i = \exp[\boldsymbol{x_i'\beta}]$. Then the Poisson regression model is

$$f(y_i) = \frac{\exp[-\exp[\boldsymbol{x_i'\beta}]](\exp[\boldsymbol{x_i'\beta}])^{y_i}}{y_i!}. \tag{1.50}$$

In case of the complete data, the likelihood function is

$$\begin{aligned} f(\boldsymbol{y}) &= \Pi_{i=1}^n f(y_i) \\ &= \Pi_{i=1}^n \frac{\exp[-\exp[\boldsymbol{x_i'\beta}]](\exp[\boldsymbol{x_i'\beta}])^{y_i}}{y_i!} \\ &= \Pi_{i=1}^n (y_i!)^{-1} \exp\left[\sum_{j=0}^{J} \beta_j \sum_{i=1}^{n} x_{ij} y_i\right] \Pi_{i=1}^n \exp[-\exp[\boldsymbol{x_i'\beta}]], \end{aligned} \tag{1.51}$$

$$\tag{1.52}$$

where the sufficient statistics is $t(\boldsymbol{y}) = \sum_{i=1}^n x_{ij} y_i$.

Its log likelihood function is

$$\mathcal{L}(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}) = \log \Pi_{i=1}^n \left[\frac{\exp[-\exp[\boldsymbol{x_i'\beta}]](\exp[\boldsymbol{x_i'\beta}])^{y_i}}{y_i!}\right] \tag{1.53}$$

$$= \sum_{i=1}^{n} [-\exp[\boldsymbol{x_i'\beta}] + y_i \boldsymbol{x_i'\beta} - \log(y_i!)]. \tag{1.54}$$

The estimate of the parameter, $\boldsymbol{\beta}$, is

$$\hat{\boldsymbol{\beta}} = \arg\max \mathcal{L}(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}) \tag{1.55}$$

which is the solution of

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{i=1}^{n} (y_i - \exp[\boldsymbol{x_i'\beta}])\boldsymbol{x_i} = 0. \tag{1.56}$$

Suppose Y_i is censored at c_i. That is,

$$Y_i = \begin{cases} y_i & \text{if } y_i \le c_i \qquad\qquad i = 1, 2, \ldots, n \\ c_i & \text{if } y_i > c_i \quad c_i > 0 \quad i = n+1, n+2, \ldots, n+m. \end{cases} \tag{1.57}$$

$\boldsymbol{\beta}$ and λ are estimated using the EM algorithm and Formula 1,

$$\mathrm{E}[Y|Y > C] = \lambda\left(1 + \frac{p\,(C, \lambda)}{P(C, \lambda)}\right) \tag{1.58}$$

provided in the subsection 1.2.2. Let the estimates of the parameters be $\boldsymbol{\beta}^{(k)}$ and $\lambda_i^{(k)}$ at the kth iteration. Note that $g(\boldsymbol{\beta}^{(k)})$ is the gradient of the equation(1.54), $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}}$, while $H(\boldsymbol{\beta}^{(k)})$ is its Hessian, $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$.

E Step

In case of the complete data,

$$g(\boldsymbol{\beta}^{(k)}) = \sum_{i=1}^{n} (y_i - \exp[\boldsymbol{x_i'}\boldsymbol{\beta}^{(k)}])\boldsymbol{x_i}. \tag{1.59}$$

The jth element of $g_j(\boldsymbol{\beta}^{(k)})$ is

$$g_j(\boldsymbol{\beta}^{(k)}) = \sum_{i=1}^{n} [y_i x_{ij} - \exp[\boldsymbol{x_i'}\boldsymbol{\beta}^{(k)}]x_{ij}]. \tag{1.60}$$

When the m censored data is included,

$$\mathrm{E}[g_j(\boldsymbol{\beta}^{(k)})] = \sum_{i=1}^{n} y_i x_{ij} + \sum_{i=n+1}^{n+m} \lambda_i^{(k)} \left(1 + \frac{p(c_i, \lambda_i^{(k)})}{P(c_i, \lambda_i^{(k)})}\right) x_{ij} - \sum_{i=1}^{n+m} \exp[\boldsymbol{x_i'}\boldsymbol{\beta}^{(k)}]x_{ij}, \tag{1.61}$$

where $\lambda_i^{(k)} = \exp[\boldsymbol{x_i'}\boldsymbol{\beta^{(k)}}]$.

The jth row and hth column element of $H(\boldsymbol{\beta^{(k)}})$ is

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \beta_h} = -\sum_{i=1}^{n+m} \exp[\boldsymbol{x_i'}\boldsymbol{\beta}^{(k)}]x_{ij}x_{ih} \tag{1.62}$$

in case of $j \neq h$, and

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \beta_h} = -\sum_{i=1}^{n+m} \exp[\boldsymbol{x_i'}\boldsymbol{\beta}^{(k)}]x_{ij}^2 \tag{1.63}$$

in case of $j = h$.

M step

Obtain $\boldsymbol{\beta}^{(k+1)}$ and $\lambda_i^{(k+1)}$, using $\mathrm{E}[g(\boldsymbol{\beta}^{(k)})]$ and $H(\boldsymbol{\beta}^{(k)})$.

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} - H^{-1}(\boldsymbol{\beta}^{(k)})\mathrm{E}[g(\boldsymbol{\beta^{(k)}})] \tag{1.64}$$

$$\lambda_i^{(k+1)} = \exp[\boldsymbol{x_i'}\boldsymbol{\beta}^{(k+1)}]. \tag{1.65}$$

The E step and the M step are iterated until the parameter estimates converge.

Appendix A

Formula 1:

$$\mathrm{E}[Y|Y > C] = \lambda\left(1 + \frac{p(C,\lambda)}{P(C,\lambda)}\right), \tag{A.1}$$

where $p(y,\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$, $P(C,\lambda) = 1 - \sum_{y=0}^{C} \frac{e^{-\lambda}\lambda^y}{y!}$.
Proof :

$$\begin{aligned}
\mathrm{E}[Y|Y > C] &= \sum_{y=C+1}^{\infty} \frac{yp(y,\lambda)}{P(C,\lambda)} \\
&= \frac{1}{P(C,\lambda)} \sum_{y=C+1}^{\infty} yp(y,\lambda) \\
&= \frac{1}{P(C,\lambda)} \sum_{y=C+1}^{\infty} \frac{e^{-\lambda}\lambda^y}{(y-1)!} \\
&= \frac{\lambda}{P(C,\lambda)} \sum_{y=C+1}^{\infty} \frac{e^{-\lambda}\lambda^{y-1}}{(y-1)!} \\
&= \frac{\lambda}{P(C,\lambda)} \sum_{w=C}^{\infty} \frac{e^{-\lambda}\lambda^w}{w!} \\
&= \frac{\lambda}{P(C,\lambda)}(P(C,\lambda) + p(C,\lambda)) \\
&= \lambda\left(1 + \frac{p(C,\lambda)}{P(C,\lambda)}\right).
\end{aligned} \tag{A.2}$$

$\square$

Appendix B

Formula 2:

$$\mathrm{E}[Y|Y>C] = \mu + \frac{\phi(z)}{1-\Phi(z)}\sigma, \tag{B.1}$$

where $Y \sim N(\mu, \sigma^2)$.
Proof:
Let $Z \sim N(0.1)$.

$$\begin{aligned}
\mathrm{E}[Z|Z>C] &= \frac{1}{1-\Phi(C)} \int_C^\infty z \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] dz \\
&= \frac{1}{1-\Phi(C)} \left[-\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]\right]_C^\infty \\
&= \frac{1}{1-\Phi(C)} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{C^2}{2}\right] \\
&= \frac{\phi(C)}{1-\Phi(C)},
\end{aligned} \tag{B.2}$$

where $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$ and $\Phi(C) = \int_{-\infty}^C \phi(t)dt$.
Since $Y \sim N(\mu, \sigma^2)$, $Y = \mu + \sigma Z$. Let $z = \frac{C-\mu}{\sigma}$.
$P(Y > C) = P((Y-\mu)/\sigma > (C-\mu)/\sigma) = P(Z > z)$

$$\begin{aligned}
\mathrm{E}[Y|Y>C] &= \mathrm{E}[\mu + \sigma Z | Z > z] \\
&= \mu + \frac{\phi(z)}{1-\Phi(z)}\sigma.
\end{aligned} \tag{B.3}$$

$\square$

Formula 3

$$\mathrm{E}[Y^2|Y > C] = \sigma^2 + \mu^2 + \sigma(C+\mu)\left(\frac{\phi(z)}{1-\Phi(z)}\right), \tag{B.4}$$

where $Y \sim N(\mu, \sigma^2)$.

Proof:

$$\begin{aligned}
\mathrm{E}[Z^2|Z > C] &= \frac{1}{1-\Phi(C)}\int_C^\infty z^2 \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] dz \\
&= \frac{1}{1-\Phi(C)}\left\{\left[-z\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{z^2}{2}\right]\right]_C^\infty + \int_C^\infty \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{z^2}{2}\right] dz\right\} \\
&= \frac{1}{1-\Phi(C)}\left(C\phi(C) + (1-\Phi(C))\right) \\
&= 1 + \frac{C\phi(C)}{1-\Phi(C)}.
\end{aligned} \tag{B.5}$$

Since $Y \sim N(\mu, \sigma^2)$, $Y = \mu + \sigma Z$. Let $z = \frac{C-\mu}{\sigma}$.

$$\begin{aligned}
\mathrm{E}[Y^2|Y > C] &= \mathrm{E}[(\mu + \sigma Z)^2 | Z > z] \\
&= \mathrm{E}[\sigma^2 Z^2 + 2\mu\sigma Z + \mu^2 | Z > z] \\
&= \sigma^2 \mathrm{E}[Z^2|Z > z] + 2\mu\sigma \mathrm{E}[Z|Z > z] + \mu^2 \\
&= \sigma^2\left(1 + \frac{z\phi(z)}{1-\Phi(z)}\right) + 2\mu\sigma\left(\frac{\phi(z)}{1-\Phi(z)}\right) + \mu^2 \\
&= \sigma^2 + \mu^2 + \sigma(z\sigma + 2\mu)\left(\frac{\phi(z)}{1-\Phi(z)}\right) \\
&= \sigma^2 + \mu^2 + \sigma(C - \mu + 2\mu)\left(\frac{\phi(z)}{1-\Phi(z)}\right) \\
&= \sigma^2 + \mu^2 + \sigma(C+\mu)\left(\frac{\phi(z)}{1-\Phi(z)}\right).
\end{aligned} \tag{B.6}$$

□

2 A univariate price response function for hotel demand and determination of the competitive set

In revenue management, it is very important to set the right prices for the products. Hotels must know their own price effects and their competitive hotels' price effects on their demand. In this chapter, we investigate the price effect on pickups using the price response function.

2.1 Methodology

Details of the data used to estimate the price response function for pickup and the count data regression models are provided.

2.1.1 Data

Data required is booking data hotels keep daily, including the pickup of the hotel to be analyzed and the room prices of the subject hotel as well as those of the other hotels seemingly competitive. Pickup is the incremental demand (number of room reservations) for a certain interval. In our case, the interval is to be one day. Let us say, bookings for the night of Oct 31 received on Oct 01 is 1. Then its pickup is said to be 1. The data are taken for 31 nights daily starting on the day, 30 days prior to the specified arrival day.

The data can be summarized as shown in Table 2.1.

reservation date	pickup	no of days prior	rates for subject hotel	rates for competitor
Oct 31	12	0	18000	19000
Oct 30	6	1	17100	18500
Oct 29	4	2	16500	18000
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
Oct 01	1	30	17000	20000

Table 2.1: Data for Oct 31 stay night

In our study, the pickup is the bookings for the July 09 stay received from a specified travel agent who takes reservations over the internet and the room price data is collected from that agent. All the data is collected from Jun 9 through July 9.

In addition to various room types, hotels offer various products to different market segments. Therefore, prices are numerous and fluctuate daily. As customers are sensitive to low prices, seven different prices are considered here; the average of the prices at 5 percentile or below (0.05QAP), that of the prices at 10 percentile or below (0.1QAP), the 5 percentile price (0.05QP), the 10 percentile price (0.1QP), the lowest price (LWSTP), the median price (MEDIANP) and the mean price (MEANP)

Let $p_1, p_2, \ldots, p_n$ be n prices and, $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)}$ be order statistics. Then, its π percentile, $Q(\pi)(0 < \pi < 1)$ is defined as;

$$Q(\pi) = (1-\gamma)p_{(j)} + \gamma p_{(j+1)} \qquad 0 \leq \gamma \leq 1, \tag{2.1}$$

where $j = \lfloor (n-1)\pi + 1 \rfloor$ and $\gamma = (n-1)\pi + 1 - j$. $p_{(k)}$ is $\pi(k)$ percentile where $\pi(k) = \frac{(k-1)}{(n-1)}$.

πQAP, the average of the prices at and below π percentile is defined as ;

$$\pi QAP = \frac{\sum_1^n p_i I(p_i \leq Q(\pi))}{\sum_1^n I(p_i \leq Q(\pi))}.$$

2.1.2 Notation

In this chapter, the arrival day is to be fixed and therefore the suffix for the variables is only t, without i used in Chapter 1.

Let $Y_t, t = 0, 1, 2, \ldots, 30$ be the pickup for the specific arrival day on the day, t days prior to the arrival day. Let $\boldsymbol{x_t} = (x_{t1}, x_{t2}, \ldots, x_{tJ})$ be the independent variable vector. The elements of the vector are a constant, the room price of the hotel to be analyzed, the competitors' room prices on the day, t days prior, and the number of days before arrival denoted by dba. $\boldsymbol{\beta}$ is the parameter vector. We consider 3 hotels including the hotel to be analyzed, thus $J = 5$.

2.1.3 Count regression model

As the pickup values are rather small, the count regression models are applied. One is the Poisson regression model and the other, the negative binomial regression model, discussed in the subsection 1.3.2.

The Poisson regression model is defined as;

$$P(Y_t = y) = \frac{e^{-\mu_t}\mu_t^y}{y!}, \ y = 0, 1, 2, \ldots \tag{2.2}$$

where μ_t is set to be $\exp(\boldsymbol{x_t'\beta})$.

The negative binomial regression model is defined as;

$$P(Y_t = y) = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_t}\right)^{\alpha^{-1}} \left(\frac{\lambda_t}{\alpha^{-1} + \lambda_t}\right)^y \alpha > 0, \ y = 0, 1, 2, \ldots, \tag{2.3}$$

where λ_t is set to be $\exp(\boldsymbol{x_t'\beta})$.

2.2 Analysis

2.2.1 Descriptive analysis

The hotel to be analyzed is referred to as Hotel A, and the other hotels are Hotel B and Hotel C. Hotel A is situated 3.15 km east from Hotel B, and 3.78 km east from Hotel C. Hotel B and Hotel C are 1.25 km away from each other, close to the busy business area.

The number of rooms provided by Hotel A is close to that of Hotel C, while Hotel B provides about 150 rooms less than Hotel A and Hotel C. However, the weighted average of room size of Hotel A is 34.2m^2 and similar to that of Hotel B, 37.8m^2. The weighted average room size of Hotel C is 27.6m^2 and is much smaller than that of Hotel A and Hotel B.

Figure 2.1 is the scatter plot of the pickups of Hotel A which shows its movement for 31 days from June 09 through July 09.

Figure 2.2 is the boxplot of the July 09 prices which indicates that Hotel A and Hotel B price ranges are similar, while Hotel C price range is definitely lower than that of Hotel A.

The seven scatter plots, Figure 2.3 through Figure 2.9, show the movement of seven prices, 0.1QAP, 0.05QAP, 0.1QP, 0.05QP, LWSTP, MEDIANP and MEANP for 31 days from June 09 through July 09. The unit of price is 100 yen. We can see that the prices were very similar between Hotel A and Hotel B for about 2 weeks from June 9. Then Hotel B began to close lower prices whereas Hotel A started to reduce prices, that results in larger difference between the Hotel A price and the Hotel B price. On the other hand, the Hotel C prices, most of the time, stayed lower than the prices of Hotel A and Hotel B.

In terms of location, Hotel A can be competitive with Hotel B as well as Hotel C. However, it is very possible that consumers consider Hotel A and Hotel B equivalent based on the fact that the weighted average of room size and the price range are similar between Hotel A and Hotel B. Therefore, Hotel A can be price competitive with Hotel B.

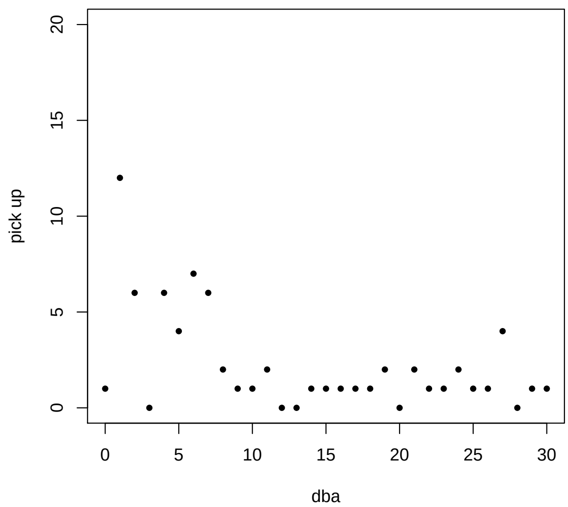

Fig. 2.1: pickups of Hotel A

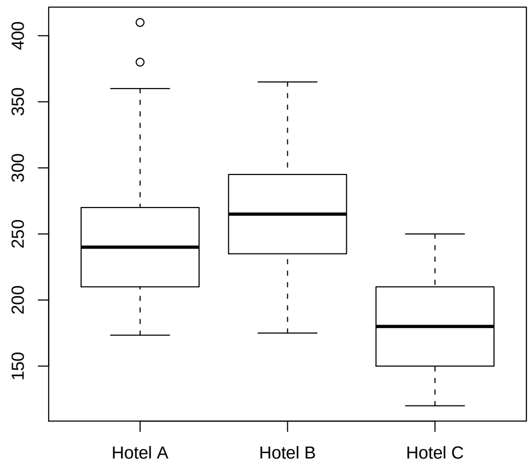

Fig. 2.2: Jul 09 Price

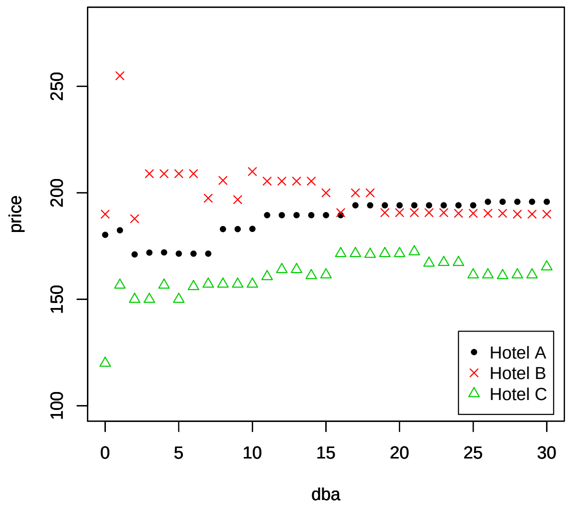

Fig. 2.3: 0.1QAP

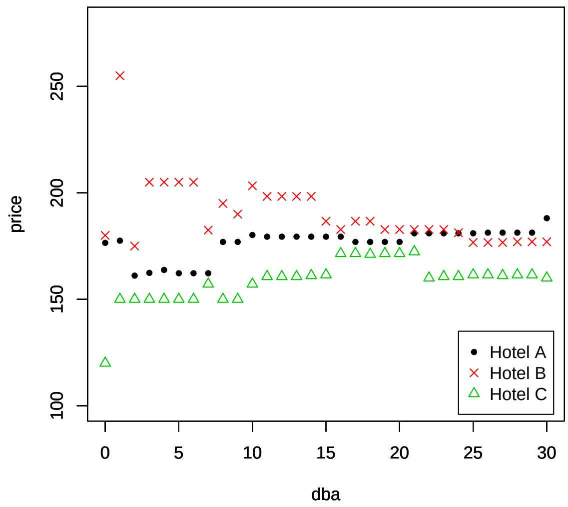

Fig. 2.4: 0.05QAP

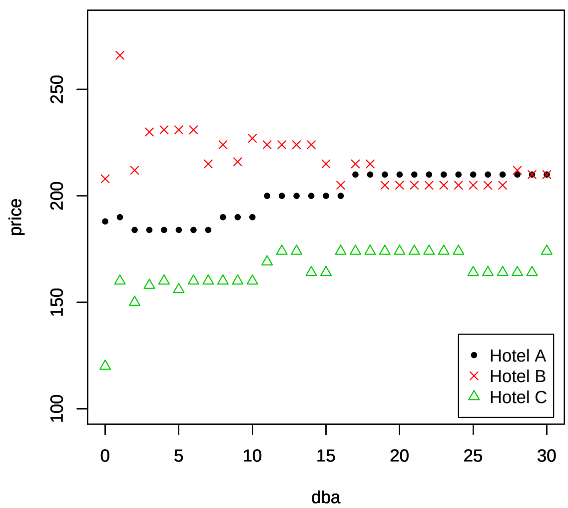

Fig. 2.5: 0.1QP

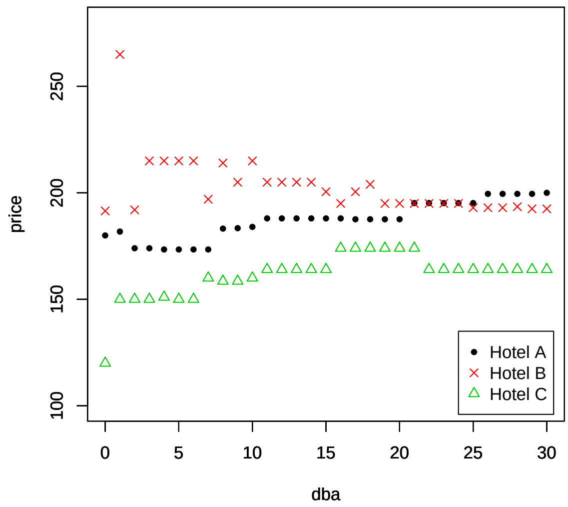

Fig. 2.6: 0.05QP

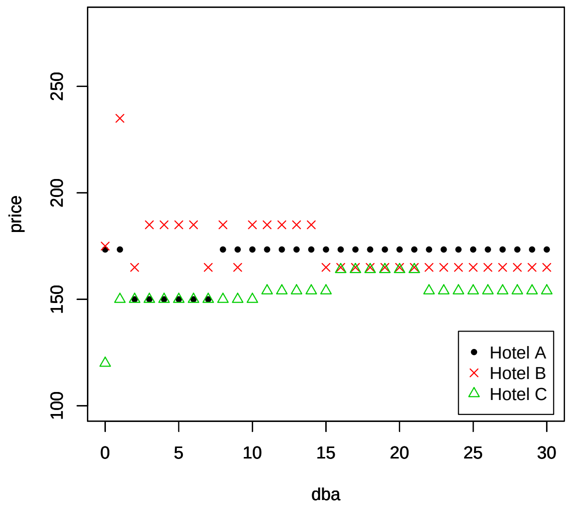

Fig. 2.7: LWSTP

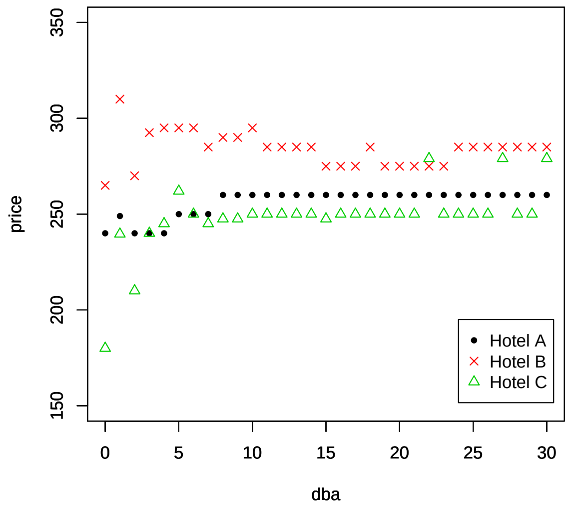

Fig. 2.8: MEDIANP

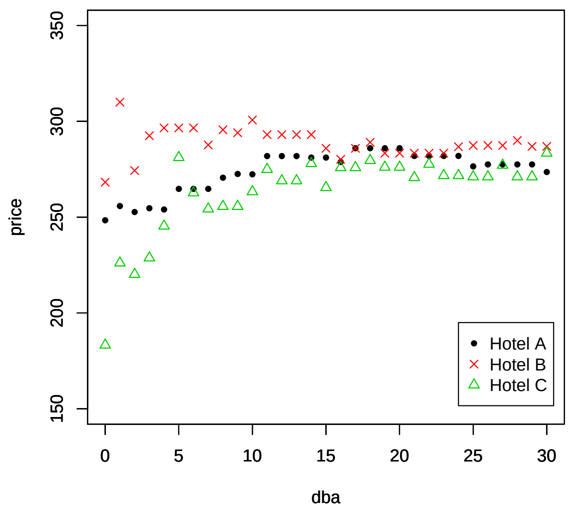

Fig. 2.9: MEANP

2.2.2 Statistical analysis: Count data regression analysis

Both Poisson and negative binomial regression models are estimated with the dependent variable, the pickup of Hotel A and the independent variables, Hotel A price, Hotel B price, Hotel C price, and dba. Parameters are estimated using seven different prices, 0.1QAP, 0.05QAP, 0.1QP, 0.05QP, LWSTP, MEDIANP and MEANP.

The results are summarized in Table 2.2 and Table 2.3.

The parameter estimates for the Hotel A price and for the Hotel B price are significant under the 5% significance level test in both Poisson and negative binomial regression analysis when 0.1QAP, 0.05QAP, 0.05QP and LWSTP are applied and AIC for both models is small. When 0.1QP, MEDIANP and MEANP are applied, many of the parameter estimates of the Hotel A price, Hotel B price and Hotel C price are insignificant. Therefore, both Poisson and negative binomial models using 0.1QAP, 0.05QAP, 0.05QP and LWSTP are appropriate for the price effect analysis. Price parameters explain the price impact on pickup. For example, the price parameter of Hotel A is -0.09195 when the Poisson regression model is applied, using 0.1QAP. This implies that 100 yen increase in the Hotel A price results in about 9.2% decline in the expectation of its pickup, while the Hotel B price increase of 100 yen boosts the expectation of the pickup by about 3.5% as its coefficient is 0.03539. In case of the negative binomial regression, the price parameter for Hotel A is -0.092560 and that of Hotel B is 0.034142 when 0.1QAP is used. Therefore, this also suggests that the expectation of its pickup drops by about 9.3% when the Hotel A price increases by 100 yen, while the Hotel B price increase of 100 yen raises the expectation of the pickup by about 3.4% . The parameter of the Hotel A price is negative. The lower the Hotel A price is, the more its pickup is. On the other hand, the parameter of the Hotel B price is positive. Pickup increases as the Hotel B price rises. Therefore, estimates of the price parameters statistically significant are consistent with the price theory. No parameter estimates of the Hotel C price are significant in any of the regression analysis.

The absolute values of the parameters for the Hotel A price are larger than those for the Hotel B price. This indicates that the Hotel A own price has a larger impact on their pickup.

Thus, Hotel A is considered price competitive with Hotel B, but there is no evidence that Hotel A is price competitive with Hotel C. Further, no parameter estimates of dba are statistically significant in any case.

		parameter				
		constant	Hotel A	Hotel B	Hotel C	dba
0.1QAP	estimate	8.026	-0.09195**	0.03539***	0.012	0.050
	z value	1.733	-3.056	3.370	0.488	1.159
	P value	0.083	0.00224	0.00075	0.625	0.246
residual deviance: 35.853 df: 26				AIC: 112.57		
0.05QAP	estimate	8.365	-0.067860**	0.029260***	-0.011	0.024
	z value	1.868	-2.888	3.377	-0.540	0.663
	P value	0.062	0.003881	0.000733	0.589	0.507
residual deviance: 38.730 df: 26				AIC: 115.45		
0.1QP	estimate	5.083	-0.053425	0.030182**	-0.005	0.031
	z value	0.956	-1.784	2.885	-0.294	0.68
	P value	0.339	0.07436	0.00392	0.769	0.491
residual deviance: 42.905 df: 26				AIC: 119.62		
0.05QP	estimate	13.272*	-0.090913*	0.031492***	-0.019	0.071
	z value	2.067	-2.501	3.549	-1.038	1.378
	P value	0.039	0.012380	0.000387	0.299	0.168
residual deviance:39.343 df: 26				AIC: 116.06		
LWSTP	estimate	2.900	-0.050016**	0.032357***	0.002	0.009
	z value	0.694	-3.033	3.411	0.087	0.282
	P value	0.488	0.002419	0.000648	0.930	0.778
residual deviance: 37.383 df: 26				AIC: 114.1		
MEDIANP	estimate	1.510	-0.045175	0.033854*	0.006	-0.025
	z value	0.243	-1.942	2.117	0.493	-0.909
	P value	0.808	0.0521	0.0342	0.622	-0.909
residual deviance: 41.170 df: 26				AIC: 117.89		
MEANP	estimate	7.610	-0.066640**	0.028144	0.012	-0.010
	z value	1.196	-2.947	1.918	1.317	-0.4105
	P value	0.232	0.00321	0.05514	0.188	0.682
residual deviance: 39.114 df: 26				AIC: 115.83		

Table 2.2: Poisson Regression

Note:
'*' denotes 'significant at 5% level'.
'**' denotes 'significant at 1% level'.
'***' denotes 'significant at 0.1% level'.

		parameter				
		constant	Hotel A	Hotel B	Hotel C	dba
0.1QAP	estimate	8.718539	-0.092560**	0.034142**	0.009887	0.050629
	z value	1.747	-2.928	2.995	0.396	1.148
	P value	0.08064	0.00341	0.00274	0.69192	0.25081
residual deviance: 32.835 df: 26				AIC:114.46		
0.05QAP	estimate	10.28577	-0.07643**	0.02823**	-0.01313	0.03080
	z value	1.908	-2.724	2.648	-0.581	0.764
	P value	0.05643	0.00644	0.00811	0.56152	0.44495
residual deviance: 30.264 df: 26				AIC: 116.19		
0.1QP	estimate	6.256114	-0.055518	0.025931	-0.004099	0.031476
	z value	0.921	-1.533	1.875	-0.207	0.631
	P value	0.3568	0.1253	0.0607	0.8364	0.5283
residual deviance: 32.384 df: 26				AIC: 120.07		
0.05QP	estimate	14.19933	-0.09325*	0.02974**	-0.02042	0.07304
	z value	1.866	-2.222	2.827	-1.011	1.279
	P value	0.06204	0.02627	0.00469	0.31215	0.20074
residual deviance:33.102 df:26				AIC: 117.54		
LWSTP	estimate	3.763774	-0.053557**	0.030846**	0.001991	0.010348
	z value	0.804	-2.856	2.743	0.075	0.316
	P value	0.42141	0.00429	0.00610	0.94027	0.75212
residual deviance: 30.486 df: 26				AIC: 115.26		
MEDIANP	estimate	6.90164	-0.06445*	0.02670	0.01176	-0.02224
	z value	0.829	-2.131	1.277	0.836	-0.680
	P value	0.4069	0.0331	0.2017	0.4032	0.4963
residual deviance: 28.660 df: 26				AIC: 117.15		
MEANP	estimate	9.90837	-0.07203**	0.02229	0.01558	-0.01088
	z value	1.280	-2.723	1.170	1.303	-0.389
	P value	0.20050	0.00647	0.24213	0.19271	0.69729
residual deviance: 30.255 df: 26				AIC: 116.37		

Table 2.3: Negative Binomial Regression

Note:
'*' denotes 'significant at 5% level'.
'**' denotes 'significant at 1% level'.
'***' denotes 'significant at 0.1% level'.

2.3 Conclusions

The price response function for pickup is estimated using the Poisson regression model and the negative binomial regression model. The independent variables are the price of the hotel analyzed, the cross prices (the prices of the other hotels) and dba. The prices used are 0.1QAP, 0.05QAP, 0.1QP, 0.05QP, LWSTP, MEDIANP and MEANP.

None of the dba parameter estimates is significant. It may possibly result from the fact that the data was collected only for 31days, a short period.

The price parameter of Hotel A and that of Hotel B whose weighted average of room size is close to that of Hotel A are significant in case of 0.1QAP, 0.05QAP, 0.05QP and LWSTP. The Hotel A price parameter is negative while the Hotel B price parameter is positive. Thus the signs of the parameter estimates that are significant are consistent with the price theory. However, the Hotel C price parameter estimates are insignificant, whose weighted average of room size is smaller than that of Hotel A. Competitiveness of hotels can be determined by significance of the price parameter estimate and their signs.

In this study, it is neither average daily rate nor hotel size but rather the effects of the hotel's own price and other hotels' prices on pickup that determine hotel competitiveness and therefore, these models reflect the market condition and are appropriate to identify competitive hotels. The hotels are competitive when the price parameter estimates are significant and have correct sings, and not competitive, if not. In our empirical study, it is determined that Hotel A is competitive with Hotel B. On the other hand, Hotel A is not price competitive with Hotel C.

3 Selecting the population with the largest mean for dependent random variables

3.1 Introduction

The selection problem is the problem of selecting the best populations from the k populations. There are two basic approaches, one is the indifference zone approach and the other is the subset selection approach. We consider only the latter in this paper. See, e.g. Gupta and Panchapakesan (1979) for a comprehensive view of two approaches. Consider k normal populations with means, $\mu_1, \ldots, \mu_k$, and a common variance, σ^2. We would like to select the subset of k populations which contains the population with the largest μ_i with probability at least equal to P^* ($k^{-1} < P^* < 1$). Note that the size of the subset is random. This approach was first developed by Gupta (1956, 1965) and later extended to several cases. In this chapter, after a brief explanation of the subset selection approach for independent random variables, we extend the method to Gaussian ARMA processes.

3.2 Selection of the normal populations for independent random variables

In this section, we review the selection for the means of normal populations with independent random variables based on Gupta (1965).

Suppose there are k normal populations, $\Pi_1, \Pi_2, \ldots, \Pi_k$ with means, $\mu_1, \mu_2, \ldots, \mu_k$, respectively and a common variance, σ^2. For each $i = 1, 2, \ldots, k$, let $\{X_{i,1}, X_{i,2}, \ldots, X_{i,n}\}$ be independent random variables from Π_i. We assume that $\{X_{1,j}\}_{j=1}^n, \{X_{2,j}\}_{j=1}^n, \ldots, \{X_{k,j}\}_{j=1}^n$ are independent. Let $M = \{(\mu_1, \mu_2, \ldots, \mu_k) : \mu_i \in \mathbb{R}, \mu_i \neq \mu_j \text{ for some } i \neq j\}$ be the parameter space for means. Let $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ be the ordered means for $\mu_1, \mu_2, \ldots, \mu_k$. Let $\bar{X}_{(i)}$ be the sample mean of the sample from the population with $\mu_{[i]}$. We also let

$$\overline{X}_{[1]} \leq \overline{X}_{[2]} \leq \cdots \leq \overline{X}_{[k]}$$

be the ordered sample means.
Rule R:
Select the ith population, Π_i, when

$$\overline{X}_i \geq \overline{X}_{[k]} - Ds_\nu/\sqrt{n}, \tag{3.1}$$

where D is a constant and $\nu = k(n-1)$ is the degrees of freedom for the unbiased estimates of σ^2, s_ν^2. The random variable, $\nu s_\nu^2/\sigma^2$, has a $\chi^2(\nu)$ distribution. The constant D is to be determined to satisfy the probability requirement for selecting the correct subset.

Let CS stand for Correct Selection of the subset that contains the population with the largest mean. Then,

$$
\begin{aligned}
P(CS|R) &= P(\overline{X}_{(k)} \geq \overline{X}_{[k]} - Ds_\nu/\sqrt{n}) \\
&= P(\overline{X}_{[k]} \leq \overline{X}_{(k)} + Ds_\nu/\sqrt{n}) \\
&= P(\overline{X}_{(j)} - \mu_{[j]} \leq \overline{X}_{(k)} - \mu_{[k]} + \mu_{[k]} - \mu_{[j]} + Ds_\nu/\sqrt{n}\ j = 1, 2, \ldots, k-1) \\
&= \mathrm{E}[P(\overline{X}_{(j)} - \mu_{[j]} \leq \overline{X}_{(k)} - \mu_{[k]} + \mu_{[k]} - \mu_{[j]} + Ds_\nu/\sqrt{n}\ j = 1, 2, \ldots, k-1|\bar{X}_{(k)}, s_\nu)] \\
&= \mathrm{E}\left[P\left(\frac{\overline{X}_{(j)} - \mu_{[j]}}{\sigma/\sqrt{n}} \leq \frac{\overline{X}_{(k)} - \mu_{[k]}}{\sigma/\sqrt{n}} + \frac{\mu_{[k]} - \mu_{[j]}}{\sigma/\sqrt{n}} + D\frac{s_\nu}{\sigma}\ j = 1, 2, \ldots, k-1 \middle| \overline{X}_{(k)}, s_\nu\right)\right] \\
&= \mathrm{E}\left[\prod_{j=1}^{k-1} \Phi\left(\frac{\overline{X}_{(k)} - \mu_{[k]}}{\sigma/\sqrt{n}} + \sqrt{n}\delta_{kj} + D\frac{s_\nu}{\sigma}\right)\middle| \overline{X}_{(k)}, s_\nu\right] \\
&= \int_0^\infty \int_{-\infty}^\infty \prod_{j=1}^{k-1} \Phi(u + \sqrt{n}\delta_{kj} + Dy)\phi(u)q_\nu(y)\ du\ dy, \qquad (3.2)
\end{aligned}
$$

where $q_\nu(y)$ is the density function of $\chi_\nu/\sqrt{\nu}$ and $\delta_{kj} = (\mu_{[k]} - \mu_{[j]})/\sigma$.

Since $\delta_{kj} > 0$ for some $j \neq k$,

$$
\begin{aligned}
\inf_M P(CS|R) &= \inf_M \int_0^\infty \int_{-\infty}^\infty \prod_{j=1}^{k-1} \Phi(u + \sqrt{n}\delta_{kj} + Dy)\phi(u)q_\nu(y)\ du\ dy \qquad (3.3) \\
&= \int_0^\infty \int_{-\infty}^\infty \Phi^{k-1}(u + Dy)\phi(u)q_\nu(y)\ du\ dy. \qquad (3.4)
\end{aligned}
$$

The constant D is determined by;

$$\int_0^\infty \int_{-\infty}^\infty \Phi^{k-1}(u + Dy)\phi(u)q_\nu(y)\ du\ dy = P^*,$$

where P^* is the prespecified probability. Gupta and Sobel (1957) provides the D values for the selected k, n, P^*.

3.3 Selection of the populations: ARMA processes

For each $i = 1, 2 \ldots, k$, let $\{X_{i,t}, t = 1, 2, \ldots, n\}$ be a sample from a Gaussian ARMA process with the mean μ_i. Let $Y_{i,t} = X_{i,t} - \mu_i$. Then $\{Y_{i,t}\}$ satisfies

$$Y_{i,t} = \phi_1 Y_{i,t-1} + \cdots + \phi_p Y_{i,t-p} + Z_{i,t} + \theta_1 Z_{i,t-1} + \cdots + \theta_q Z_{i,t-q}, \quad t = 0, \pm 1, \pm 2, \ldots, \qquad (3.5)$$

where $\{Z_{i,t}\}_{t=-\infty}^{\infty}$ is a sequence of independent normal random variables with the mean 0 and the variance σ_ε^2. Let $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$, and $\phi = (\phi_1, \ldots, \phi_p)', \theta = (\theta_1, \ldots, \theta_q)'$. Define polynomials $\phi(\cdot)$ and $\theta(\cdot)$ by

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

Then model (3.5) can be written by

$$\phi(B)Y_{i,t} = \theta(B)Z_{i,t}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.6}$$

where B is the backward shift operator defined by

$$B^j X_t = X_{t-j}, \quad j = 0, \pm 1, \pm 2, \ldots \tag{3.7}$$

We assume that $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeroes and that $\phi(z) \neq 0$ and $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. Under these assumptions ARMA process (3.5) is causal and invertible (see Brockwell and Davis, 1991). From the causality of the model, it can be written as

$$Y_{i,t} = X_{i,t} - \mu_i = \sum_{j=0}^{\infty} \psi_j Z_{i,t-j}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.8}$$

where $\{\psi_j\}$ is given by the relation

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z), \quad |z| \leq 1. \tag{3.9}$$

The autocovariance function for $\{Y_{i,t}\}$ is given by

$$\gamma(k) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} \tag{3.10}$$

for each $k \in \mathbb{Z}$.

Note that we assume for different i the parameters for ARMA procceses are the same except the mean μ_i. Thus the autocovariances are the same for different i. We also assume that k stationary processes are independent.

Let $M = \{(\mu_1, \ldots, \mu_k) : \mu_i \in \mathbb{R},\ \mu_i \neq \mu_j \text{ for some } i \neq j\}$ be the parameter space for the means. Let

$$\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$$

be the ordered means for $\mu_1, \mu_2, \ldots, \mu_k$. Let $\overline{X}_i$ be the sample mean of $\{X_{i,t}, t = 1, 2, \ldots, n\}$. Let $\overline{X}_{(i)}$ denote the sample mean of the sample that has the mean $\mu_{[i]}$. We also let

$$\overline{X}_{[1]} \leq \overline{X}_{[2]} \leq \cdots \leq \overline{X}_{[k]}$$

be the ordered sample means.

The purpose of the current paper is to investigate a procedure of selecting a subset of indexes that contains the popultation with the largest mean with the prespescified probability. This methodology was first developed by Gupta (1963) for independent sequences of random variables. We investigate a method for dependent sequences of random variables.

The variance of $\sqrt{n}\overline{X}_i$ is given by

$$\sigma_n^2 = \sum_{s=-(n-1)}^{n-1} \left(1 - \frac{|s|}{n}\right) \gamma(s). \tag{3.11}$$

It is known that if $\sum_{s=0}^{\infty} |\gamma(s)| < \infty$,

$$\sigma_n^2 \to \sigma_\infty^2 := \sum_{s=-\infty}^{\infty} \gamma(s) \quad \text{as } n \to \infty.$$

Since the process $\{X_{it}\}_{t\geq 1}$ is Gaussian,

$$\frac{\sqrt{n}(\overline{X}_i - \mu_i)}{\sigma_n} \sim N(0, 1).$$

3.3.1 Subsampling estimator

In order to estimate σ_n^2, we use the subsampling methodology developed by Carlstein (1986), Politis and Romano (1993) among others. Let b_n be such that $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$. We call b_n the subsample size. The subsampling estimator for σ_n^2 based on the sample from j-th population is defined by

$$\hat{\sigma}_{j,b,n}^2 = (n - b + 1)^{-1} \sum_{i=1}^{n-b+1} \left(\frac{1}{\sqrt{b}} \sum_{t=i}^{i+b-1} X_{j,t} - \sqrt{b}\bar{X}_j \right)^2. \tag{3.12}$$

The subsampling estimator is known to be consistent, i.e.

$$\hat{\sigma}_{j,b,n}^2 \xrightarrow{p} \sigma_\infty^2 \quad as \quad n \to \infty$$

for weakly dependent stationary processes under mild conditions. (Refer to Carlstein (1986) and Fukuchi (1999).) The subsampling estimator for σ_n^2 is defined by

$$\hat{\sigma}_{b,n}^2 = \frac{1}{k} \sum_{j=1}^{k} \hat{\sigma}_{j,b,n}^2. \tag{3.13}$$

In the following, we write $\hat{\sigma}_n^2 = \hat{\sigma}_{b,n}^2$ for short.

3.3.2 Selection rule

Consider the selection rule R definded as follows:
R: Select the ith population if and only if

$$\overline{X}_i \geq \overline{X}_{[k]} - \frac{D_n \hat{\sigma}_n}{\sqrt{n}},$$

where D_n is a positive constant to be determined.
Let CS denote the correct selection.

$$\begin{aligned}
P(CS|R) =& P\left(\overline{X}_{(k)} \geq \overline{X}_{[k]} - \frac{D_n\hat{\sigma}_n}{\sqrt{n}}\right) \\
=& P\left(\overline{X}_{[k]} \leq \overline{X}_{(k)} + \frac{D_n\hat{\sigma}_n}{\sqrt{n}}\right) \\
=& P\left(\overline{X}_{(j)} \leq \overline{X}_{(k)} + \frac{D_n\hat{\sigma}_n}{\sqrt{n}},\ j = 1, 2, \ldots, k-1\right) \\
=& P\left(\overline{X}_{(j)} - \mu_{[j]} + \mu_{[j]} \leq \overline{X}_{(k)} - \mu_{[k]} + \mu_{[k]} + \frac{D_n\hat{\sigma}_n}{\sqrt{n}},\ j = 1, \ldots, k-1\right) \\
=& P\left(\overline{X}_{(j)} - \mu_{[j]} \leq \overline{X}_{(k)} - \mu_{[k]} + \mu_{[k]} - \mu_{[j]} + \frac{D_n\hat{\sigma}_n}{\sqrt{n}},\ j = 1, \ldots, k-1\right) \\
=& P\left(\frac{\overline{X}_{(j)} - \mu_{[j]}}{\sigma_n/\sqrt{n}} \leq \frac{\overline{X}_{(k)} - \mu_{[k]}}{\sigma_n/\sqrt{n}} + \frac{\mu_{[k]} - \mu_{[j]}}{\sigma_n/\sqrt{n}} + \frac{D_n\hat{\sigma}_n}{\sigma_n},\ j = 1, \ldots, k-1\right) \\
=& P\left(\frac{\overline{X}_{(j)} - \mu_{[j]}}{\sigma_n/\sqrt{n}} - \frac{\overline{X}_{(k)} - \mu_{[k]}}{\sigma_n/\sqrt{n}} - \frac{D_n\hat{\sigma}_n}{\sigma_n} \leq \frac{\mu_{[k]} - \mu_{[j]}}{\sigma_n/\sqrt{n}},\ j = 1, \ldots, k-1\right). \qquad (3.14)
\end{aligned}$$

Since $\left(\frac{X_{1,1}-\mu_1}{\sigma_\varepsilon}, \frac{X_{1,2}-\mu_1}{\sigma_\varepsilon}, \ldots, \frac{X_{k,n}-\mu_k}{\sigma_\varepsilon}\right)'$ is distributed as the kn-dimensional normal distribution with mean vector 0 and the variance-covariances matrix $diag\{\Lambda, \Lambda, \ldots, \Lambda\}$, where $\Lambda = \{\sigma_\varepsilon^{-2}\gamma(|i-j|)\}_{i,j=1}^n$ and $\gamma(\cdot)$ is given by (3.10).
Thus the distribution of

$$\frac{\overline{X}_{(j)} - \mu_{[j]}}{\sigma_n/\sqrt{n}} - \frac{\overline{X}_{(k)} - \mu_{[k]}}{\sigma_n/\sqrt{n}} - \frac{D_n\hat{\sigma}_n}{\sigma_n} \qquad (3.15)$$

depends on parameters β but not on $(\mu_1, \mu_2, \ldots, \mu_k)'$ and σ_ε^2. See Appendix C for further details. Since $\mu_i \neq \mu_j$ for some $i \neq j$, the infimum of (3.14) over the set M is given by

$$\begin{aligned}
\inf_M P(CS|R) =& P\left(\frac{\overline{X}_{(j)} - \mu_{[j]}}{\sigma_n/\sqrt{n}} - \frac{\overline{X}_{(k)} - \mu_{[k]}}{\sigma_n/\sqrt{n}} - \frac{D_n\hat{\sigma}_n}{\sigma_n} \leq 0,\ j = 1, \ldots, k-1\right) \\
=& P\left(\frac{\overline{X}_{(j)} - \mu_{[j]}}{\hat{\sigma}_n/\sqrt{n}} - \frac{\overline{X}_{(k)} - \mu_{[k]}}{\hat{\sigma}_n/\sqrt{n}} \leq D_n,\ j = 1, \ldots, k-1\right).
\end{aligned}$$

We would like to find the value D_n that satisfies

$$P\left(\frac{\overline{X}_{(j)} - \mu_{[j]}}{\hat{\sigma}_n/\sqrt{n}} - \frac{\overline{X}_{(k)} - \mu_{[k]}}{\hat{\sigma}_n/\sqrt{n}} \leq D_n,\ j = 1, \ldots, k-1\right) = P^*, \qquad (3.16)$$

where P^* is the prespecified probability. It is easy to see that D_n is the P^*-th quantile of the random variable

$$Y_n = \frac{\sqrt{n}\left(\overline{X}_{[k-1]} - \overline{X}_k\right)}{\hat{\sigma}_n}, \tag{3.17}$$

where $\overline{X}_1, \ldots, \overline{X}_{k-1}, \overline{X}_k$ are sample means of k independent samples of the sample size n, each following a Gaussian ARMA processes with the mean 0

$$X_{i,t} = \phi_1 X_{i,t-1} + \cdots + \phi_p X_{i,t-p} + Z_{i,t} + \theta_1 Z_{i,t-1} + \cdots + \theta_q Z_{i,t-q},\ t = 0 \pm 1, \pm 2, \ldots$$

and $\overline{X}_{[k-1]}$ denotes $\max_{1\le i\le k-1} \overline{X}_i$.

3.4 Estimation of D_n

Since the quantile D_n depends on unknown parameters (ϕ, θ), it needs to be estimated.

The first method to obtain approximate value of D_n is the one based on the asymptotic distribution of Y_n.

Since $\sigma_n \to \sigma_\infty > 0$, we have $\hat{\sigma}_n/\sigma_n \xrightarrow{p} 1$ and thus

$$\begin{aligned}
Y_n =& \frac{\sqrt{n}\left(\overline{X}_{[k-1]} - \overline{X}_k\right)}{\hat{\sigma}_n} \\
=& \left(\frac{\max\limits_{1\le i\le k-1} \overline{X}_i}{\sigma_n/\sqrt{n}} - \frac{\overline{X}_k}{\sigma_n/\sqrt{n}} \right) \times \frac{\sigma_n}{\hat{\sigma}_n} \\
\xrightarrow{d}& \max_{1\le i\le k-1} U_i - U \\
=& \max_{1\le i\le k-1} (U_i - U),
\end{aligned}$$

where $U_1, \ldots, U_{k-1}, U$ are independently distributed as the standard normal distribution. The ramdom vector $(U_1 - U, U_2 - U, \ldots, U_{k-1} - U)'$ is distributed as the $(k-1)$-dimensional multivariate normal distribution with the common variance 2 and the common covariance 1. P^*-quantile of $\max_{1\le i\le k-1}(U_i - U)$ is denoted by $\tilde{D}_n(P^*)$.

The P^*-quantile of $\max\limits_{1\le i\le k-1} Z_i$, where $(Z_1, \ldots, Z_{k-1})'$ are $(k-1)$-dimensional normal random variables with the common correlation is given in Gupta, Nagel and Panchapakesan (1973) as Table I for $k = 2(1)11$, $13(2)51$ and $P^* = .75, .90, .95, .975$ and $.99$.

The second method is estimating D_n by plug-in method. The distribution of $\sqrt{n}\left(\overline{X}_{[k-1]} - \overline{X}_k\right)/\hat{\sigma}_n$ depends on the parameter (ϕ, θ). Therefore D_n depends on these parameters and we write $D_n(P^*) = D_n(P^*, \phi, \theta)$. Thus $D_n(P^*, \phi, \theta)$ is estimated by $\hat{D}_n = D_n(P^*, \hat{\phi}, \hat{\theta})$ where $(\hat{\phi}, \hat{\theta})$ is the MLE of (ϕ, θ).

3.4.1 Stationary Gaussian ARMA resampling

The estimate $D_n(P^*, \hat{\phi}, \hat{\theta})$ can be obtained by using the Monte Calro simulation as follows.

1. Given the data $\{X_{i,t}\}_{t=1}^n$, obtain mean adjusted data

$$X'_{i,t} = X_{i,t} - \bar{X}_i. \tag{3.18}$$

2. From $\{X'_{i,t}\}_{t=1}^n$ obtain the MLE $(\hat{\phi}_i, \hat{\theta}_i) = (\hat{\phi}_{1,i}, \ldots, \hat{\phi}_{p,i}, \hat{\theta}_{1,i}, \ldots, \hat{\theta}_{q,i})$.
   Let $\hat{\beta}_n = (\hat{\phi}_n, \hat{\theta}_n)$ be the average of $(\hat{\phi}_i, \hat{\theta}_i)$, that is

$$\hat{\beta}_n = (\hat{\phi}_n, \hat{\theta}_n) = \left(k^{-1} \sum_{i=1}^{k} \hat{\phi}_i, \quad k^{-1} \sum_{i=1}^{k} \hat{\theta}_i \right). \tag{3.19}$$

   Estimation of D_n.

3. Generate k independent samples of size n from ARMA(p, q) process with the parameter $\hat{\beta}_n$ that is,

$$X^*_{i,t} = \hat{\phi}_1 X^*_{i,t-1} + \cdots + \hat{\phi}_p X^*_{i,t-p} + Z^*_{i,t} + \hat{\theta}_1 Z^*_{i,t-1} + \cdots + \hat{\theta}_q Z^*_{i,t-q}, \; t = 0, \pm 1, \pm 2, \ldots, \tag{3.20}$$

   where $\{Z^*_{i,t}\}$ is a sequence of independent standard normal random variables.

4. Obtain subsampling estimator, $\hat{\sigma}_n^{2*}$, of σ_n^2 from the pseudo sample generated by (3.20).

5. Obtain Y_n^* by

$$Y_n^* = \frac{\sqrt{n}\bar{X}^*_{[k-1]} - \bar{X}^*_k}{\hat{\sigma}^*_n}, \tag{3.21}$$

   where $\hat{\sigma}^*_n = \sqrt{\hat{\sigma}^{2*}_n}$

6. Repeat above procedures NR times. (NR=Number of Replications)

7. Obtain $\hat{D}_n(P^*)$ by computing P^*-quantile of Y_n^*.

We call the procedure defined by (1)-(4) the stationary Gaussian ARMA resampling.
The selection rule we propose is the rule R with $\hat{D}_n$:
Selection rule with $\hat{D}_n$. $R(\hat{D}_n)$: Select the ith population if and only if

$$\overline{X}_i \geq \overline{X}_{[k]} - \frac{\hat{D}_n \hat{\sigma}_n}{\sqrt{n}}.$$

3.5 Some results on the stationary Gaussian ARMA resampling for the subsampling estimator

In this section, the consistency of $\hat{D}_n$ is proved. For that purpose, some results on the stationary Gaussian ARMA resampling for the subsampling estimator are given first.

Random variables $\{X_i\}$ are said to be uniformly integrable (u.i.) if there exists n_o such that $\lim_{A\to\infty} sup_{n\geq n_o} \mathrm{E}(|X_n| I(|X_n| \geq A)) = 0$.

Let $\{X_i\}$ be a strong-mixing stochastic process. Let $\boldsymbol{X}_n^i = (X_{i+1}, \ldots, X_{i+n})'$ and let $s_n^i = s_n(\boldsymbol{X}_n^i)$ be a statistic computed from $\boldsymbol{X}_n^i$. It is often of interest to obtain an estimator of the variance of a statistic $s_n^0 = s_n(\boldsymbol{X}_n^0)$. Carlstein(1986) developed the subsampling variance estimator which is consistent under some weak conditions.

Let $t_n^i = n^{1/2}\{s_n^i - E(s_n^0)\}$. Suppose that

$$\mathrm{E}\left(t_n^0\right)^2 = n\mathrm{V}(s_n^0) \to \sigma^2 \in (0, \infty), \text{ as } n \to \infty.$$

Subsampling estimator for σ^2 is defined by

$$\hat{\sigma}_n^2 = N_n^{-1} b_n \sum_{i=0}^{N_n - 1} (s_b^i - \bar{s}_n)^2,$$

where $\bar{s}_n = N_n^{-1}\sum_{i=0}^{N_n-1} s_b^i$. In terms of t_b^i, $\hat{\sigma}_n^2$ can be written as

$$\hat{\sigma}_n^2 = N_n^{-1}\sum_{i=0}^{N_n-1}(t_b^i - \bar{t}_n)^2 = N_n^{-1}\sum_{i=0}^{N_n-1}\left(t_b^i\right)^2 - \left(\bar{t}_n\right)^2,$$

where $\bar{t}_n = N_n^{-1}\sum_{i=0}^{N_n-1} t_b^i$.

Carlstein (1986) proved that the average of a statistic computed from nonoverelapping subsamples converges in L^2 to the limit of its expected value. This result was proved for the overlapping case by Fukuchi (1999). Let $(\Omega, \mathcal{F}, \mathcal{P})$ be the probability space on which random variables $\{X_i\}$ are defined. Let $N_n = n - b_n + 1$ be the number of overlapping subsamples of size b_n in $(X_1, X_2, \ldots, X_n)$. Define

$$\bar{f}_n = N_n^{-1}\sum_{i=0}^{N_n-1} f_{b_n}^i.$$

The following is the result for the overlapping case.

Theorem 1 *Let $\{X_i\}$ be strong-mixing and let $f_n(\boldsymbol{X}_n^i) = f_n^i$ be a statistic. Let $\{b_n : n \geq 1\}$ be such that $b_n \to \infty$ and $b_n/n \to 0$; If*

$$\lim_{n\to\infty} \mathrm{E}(f_n^0) = \psi \in \mathbb{R}$$

and

$$\left\{\left(f_n^0\right)^2\right\} \textit{ are u.i.},$$

then

$$\bar{f}_n \xrightarrow{L^2} \psi \textit{ as } n \to \infty.$$

Proof: See Fukuchi (1999).□

The next result is L^2-consistency of the overlapping subsampling estimator $\hat{\sigma}_n^2$.

Theorem 2 *Let $\{X_i\}$ be strong-mixing. If*

$$\{(t_n^0)^4\} \text{ are u.i.,}$$

then

$$\hat{\sigma}_n^2 \xrightarrow{L^2} \sigma^2 \text{ as } n \to \infty.$$

Proof: The proof is the same as that of Theorem 3 in Carlstein (1986). □

Let C be the causal and invertible region of β, that is,

$$C = \{\beta \in \mathbb{R}^{p+q} : \phi(z)\theta(z) \neq 0, \text{ for } |z| \leq 1, \phi_p \neq 0, \theta_q \neq 0, \text{ and } \phi(\cdot),\ \theta(\cdot) \text{ have no common zeroes } \}.$$

Note that thte set C is bounded and open in $\mathbb{R}^{p+q}$.

Theorem 3 states that if $\{X_i\}$ is a stationary Gaussian ARMA process, the L^2 convergence of $\hat{\sigma}_n^2$ is shown to be uniform in a neighborhood of ARMA parameter $\beta \in C$.

Theorem 3 *Let $\{X_i\}$ be a stationary Gaussian ARMA process. Let $\{b_n : n \geq 1\}$ be such that $b_n \to \infty$ and $b_n/n \to 0$. Let β_0 be a vector such that $\beta_0 \in C$.*

Assume the following.
Assumption 1: $\{(t_n^0)^4\}$ are u.i..
Assumption 2: There exists $\delta > 0$ such that

$$(i) \quad \sup_{\beta \in \overline{O}_\delta(\beta_0)} \left| \mathrm{E}_\beta (t_n^0)^2 - \sigma^2(\beta) \right| \to 0,$$

$$(ii) \quad \sup_{\beta \in \overline{O}_\delta(\beta_0)} V_\beta \left(N_n^{-1} \sum_{i=0}^{N_n - 1} (t_b^i)^2 \right) \to 0,$$

$$(iii) \quad \sup_{\beta \in \overline{O}_\delta(\beta_0)} \mathrm{E}_\beta (\bar{t}_n)^4 \to 0.$$

where $O_\delta(\beta_0)$ is the δ-ball centered at β_0 and $\overline{O}_\delta(\beta_0)$ is its closure. Then

$$\sup_{\beta \in \overline{O}_\delta(\beta_0)} \mathrm{E}_\beta \left(\hat{\sigma}_n^2 - \sigma^2(\beta) \right)^2 \to 0 \text{ as } n \to \infty.$$

Proof:

$$\begin{aligned}
\mathrm{E}_\beta \left(\hat{\sigma}_n^2 - \sigma^2(\beta) \right)^2 &= \mathrm{E}_\beta \left(N_n^{-1} \sum_{i=0}^{N_n - 1} (t_b^i)^2 - \sigma^2(\beta) - (\bar{t}_n)^2 \right)^2 \\
&\leq 2 \left\{ \mathrm{E}_\beta \left(N_n^{-1} \sum_{i=0}^{N_n - 1} (t_b^i)^2 - \sigma^2(\beta) \right)^2 + \mathrm{E}_\beta (\bar{t}_n)^4 \right\} \\
&= 2 \left[V_\beta \left(N_n^{-1} \sum_{i=0}^{N_n - 1} (t_b^i)^2 \right) + \{\mathrm{E}_\beta (t_b^0)^2 - \sigma^2(\beta)\}^2 + \mathrm{E}_\beta (\bar{t}_n)^4 \right].
\end{aligned}$$

Taking the supremum over the set $O_\delta(\beta_0)$ of the both sides of the above inequality completes the proof. $\square$

Theorem 4 is a stationary Gaussian ARMA resampling version of Theorem 2.

Theorem 4 *Let $\{X_i\}$ be a stationary Gaussian ARMA process with the parameter $\beta_0 \in C$ and let $\hat{\beta}_n$ be the maximum likelihood estimator of β_0. Let $\{b_n : n \geq 1\}$ be such that $b_n \to \infty$ and $b_n/n \to 0$.*

Then under the assumptions 1 and 2 of Theorem 3 and the stationary Gaussian ARMA resampling, for almost every $\omega \in \Omega$,

$$\hat{\sigma}_n^2(\omega) \xrightarrow{L^2} \sigma^2(\beta_0) \text{ as } n \to \infty.$$

Proof: Since $\hat{\beta}_n \to \beta_0$ with probability 1, there exists $\tilde{\Omega} \subset \Omega$ such that $P(\tilde{\Omega}) = 1$ and $\hat{\beta}_n(\omega) \to \beta_0$ for all $\omega \in \tilde{\Omega}$.

Let $\delta > 0$ be such that $\overline{O}_\delta(\beta_0) \subset C$. For any $\omega \in \tilde{\Omega}$, there exists $n_0(\omega) \geq 1$ such that $|\hat{\beta}_n(\omega) - \beta_0| < \delta$ for any $n \geq n_0(\omega)$. For this δ and ω

$$\mathrm{E}_{\hat{\beta}_n(\omega)}\left[\left(\hat{\sigma}_n^2(\omega) - \sigma^2(\beta_0)\right)^2\right] \leq \sup_{\beta \in \overline{O}_\delta(\beta_0)} \mathrm{E}_\beta\left[\left(\hat{\sigma}_n^2(\beta) - \sigma^2(\beta)\right)^2\right] \to 0$$

from Theorem 3. $\square$

For the case of $s_n^0 = s_n(\boldsymbol{X}_n^0) = \bar{X}_n$, we will show that Assumption 2 of Theorem 3 holds for the stationary Gaussian AR(p) process. Let C_1 be the causal region of autoregressive parameters ϕ, i.e.,

$$C_1 = \{\phi \in \mathbb{R}^p : \phi(z) \neq 0, \text{ for } |z| \leq 1, \phi_p \neq 0\}.$$

Suppose that $\{X_i\}$ follows the AR(p) procsess $X_i = \phi_1 X_{i-1} + \cdots + \phi_p X_{i-p} + \varepsilon_i$ $i = 0, \pm 1, \pm 2, \ldots$, where $\phi \in C_1$ and $\{\varepsilon_i\}$ is a sequence of independently and normally distributed random variables with the mean zero and the variance σ_ε^2. Let ξ_i, $i = 1, \ldots, k$ be distinct zeroes of the polynomial $\phi(z)$ and r_i, $i = 1, \ldots, k$ be their respective multiplicities.
Assumption 2 (i)
Then autocovariance function of the process $\{X_i\}$ is given by

$$\gamma_s = \sum_{i=1}^{k} \sum_{j=0}^{r_i} c_{ij} s^j \xi_i^{-s}, \quad s \geq 0, \tag{3.22}$$

where the constants c_{ij} are uniquely determined by the equations

$$\gamma_k - \phi_1 \gamma_{k-1} - \cdots - \phi_p \gamma_{k-p} = \sigma_\varepsilon^2, \quad 0 \leq k \leq p, \tag{3.23}$$

(Brockwell and Davis, 1991, p. 93). After replacing $\gamma_0, \gamma_1 \ldots, \gamma_p$ in (3.23) by (3.22), constants c_{ij}'s are obtained by solving equations (3.23). Therefore for each $i = 1, \ldots, k$ and $j = 1, \ldots, r_i$,

c_{ij} is a continuous function of ϕ. The zeroes ξ_i are also continuous functions of ϕ. Let $\phi_0 \in C_1$ and $\delta > 0$ be such that $\overline{O}_\delta(\phi_0) \subset C_1$. Then it follows that for $i = 1, \ldots, k$ and $j = 1, \ldots, r_i$,

$$\sup_{\phi \in \overline{O}_\delta(\phi_0)} |\xi_i(\phi)| < \infty,$$
$$\sup_{\phi \in \overline{O}_\delta(\phi_0)} |c_{ij}(\phi)| < \infty.$$

We have

$$\begin{aligned} \left|\mathrm{E}(t_n^0)^2 - \sigma^2(\phi)\right| &= \left| \sum_{s=-n}^{n} \left(1 - \frac{|s|}{n}\right) \gamma_s - \sum_{s=-\infty}^{\infty} \gamma_s \right| \\ &\leq \sum_{s=-n}^{n} |s||\gamma_s| + 2 \left| \sum_{s=n+1}^{\infty} \gamma_s \right|. \end{aligned} \tag{3.24}$$

There exists $\tilde{\phi} \in \overline{O}_\delta(\phi_0)$ such that

$$\sup_{\phi \in \overline{O}_\delta(\phi_0)} |\xi_i(\phi)| = \left|\xi_i(\tilde{\phi})\right|.$$

It follows that

$$\begin{aligned} \sup_{\phi \in \overline{O}_\delta(\phi_0)} \frac{1}{n} \left| \sum_{s=-n}^{n} |s|^j \xi_i(\phi)^{-s} \right| &\leq \frac{1}{n} \sum_{s=-n}^{n} |s|^j \sup_{\phi \in \overline{O}_\delta(\phi_0)} |\xi_i(\phi)|^{-s} \\ &= \frac{1}{n} \sum_{s=-n}^{n} |s|^j \left|\xi_i(\tilde{\phi})\right|^{-s} \to 0. \end{aligned}$$

Thus

$$\sup_{\phi \in \overline{O}_\delta(\phi_0)} \left| \sum_{s=-n}^{n} |s| \gamma_s(\phi) \right| \to 0.$$

Similarly it follows that

$$\sup_{\phi \in \overline{O}_\delta(\phi_0)} \left| \sum_{s=n+1}^{\infty} \gamma_s(\phi) \right| \to 0.$$

Assumption 2 (ii)
Let $Y_i = (t_b^{i-1})^2$ then $N_n^{-1} \sum_{i=0}^{N_n - 1} (t_b^i)^2 = N_n^{-1} \sum_{i=1}^{N_n} Y_i = \bar{Y}_{N_n}$. Thus it is enough to show that $\mathrm{V}_\phi(\bar{Y}_{N_n})$ converges to zero uniformly over the set $\overline{O}_\delta(\phi_0)$. Let $\gamma_Y(s) = \mathrm{Cov}(Y_1, Y_{1+s})$. Then

$$\begin{aligned} \gamma_Y(s) &= \mathrm{E}(Y_1 - \sigma_b^2)(Y_{1+s} - \sigma_b^2) \\ &= \mathrm{E}(t_b^0)^2 (t_b^s)^2 - \sigma_b^4 \\ &= 2\{\mathrm{Cov}(t_b^0, t_b^s)\}^2. \end{aligned}$$

The last equality holds since (t_b^0, t_b^s) is distributed as a multivariate normal distribution.

Since the process $\{X_i\}$ is causal, it has the MA(∞) representation

$$X_i = \sum_{k=0}^{\infty} \psi_k \varepsilon_{i-k},$$

where

$$\psi_k = \sum_{i=1}^{k} \sum_{j=0}^{r_i - 1} \alpha_{ij} k^j \xi_i^{-k}, \quad k \geq 0. \tag{3.25}$$

The constants α_{ij} are uniquely determined from equations

$$\psi_0 = 1, \tag{3.26}$$

$$\psi_k - \phi_1 \psi_{k-1} - \cdots - \phi_k \psi_0 = 0, \quad 1 \leq k \leq p, \tag{3.27}$$

(Brockwell and Davis, 1991, p. 92). For each i, j, α_{ij} is a continuous function of ϕ in C_1 and thus

$$\sup_{\phi \in \overline{O}_\delta(\phi_0)} |\alpha_{ij}(\phi)| < \infty.$$

We will show that Assumption 2(ii) and (iii) only in the case when the polynomial $\phi(z)$ has the zero ξ of multiplicity p in order to avoid a notational complication. Then

$$\psi_k = \sum_{j=0}^{p-1} \alpha_j k^j \xi^{-k}, \quad k \geq 0. \tag{3.28}$$

The proofs for other cases are similar.

The moving sum $\sum_{j=1}^{b} X_{i+j}$ has MA(∞) representation

$$\sum_{j=1}^{b} X_{i+j} = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{i+b-j},$$

where

$$\varphi_j = \begin{cases} 1 + \psi_1 + \cdots + \psi_j = \sum_{k=0}^{\infty} \psi_k - a_j, & 0 \leq j \leq b-1, \\ \psi_{j-b+1} + \psi_{j-b+2} + \cdots + \psi_j, & b \leq j, \end{cases} \tag{3.29}$$

where $a_j = \sum_{k=j+1}^{\infty} \psi_k$. Let

$$K = \sum_{i=0}^{p-1} |\alpha_i|,$$

$$f(\xi, b) = \sum_{i=1}^{b} i^{p-1} \xi^{-(i-1)}, \quad f(\xi, \infty) = \sum_{i=1}^{\infty} i^{p-1} \xi^{-(i-1)}.$$

Then we have

$$
\begin{aligned}
|\psi_i + \psi_{i+1} + \cdots + \psi_{i+b-1}| &\le K f(\xi, b) i^{p-1} \xi^{-i}, \\
\sum_{i=1}^{\infty} (\psi_i + \psi_{i+1} + \cdots + \psi_{i+b-1})^2 &\le K^2 f(\xi, b)^2 \sum_{i=1}^{\infty} i^{2(p-1)} \xi^{-2i} \\
&\le K^2 f(\xi, \infty)^2 \sum_{i=1}^{\infty} i^{2(p-1)} \xi^{-2i}. \qquad (3.30)
\end{aligned}
$$

The variance and autocovariance of $\sum_{i=1}^{b} X_{i+j}$ is given by

$$
\sigma_\varepsilon^{-2} \mathrm{V}(X_{i+j}) = \sum_{j=0}^{\infty} \varphi_j^2 = b \left(\sum_{k=0}^{\infty} \psi_k \right)^2 + A(b, \xi) + B(b, \xi),
$$

where

$$
\begin{aligned}
A(b, \xi) &= -2 \left(\sum_{k=0}^{\infty} \psi_k \right) \sum_{j=0}^{b-1} a_j + \sum_{j=0}^{b-1} a_j^2, \\
B(b, \xi) &= \sum_{i=1}^{\infty} (\psi_i + \psi_{i+1} + \cdots + \psi_{i+b-1})^2.
\end{aligned}
$$

For $1 \le s \le b-1$,

$$
\begin{aligned}
\sigma_\varepsilon^{-2} \mathrm{Cov}(X_{i+j}, X_{i+s+j}) &= \sum_{j=0}^{\infty} \varphi_j \varphi_{j+s} \\
&= \sum_{j=0}^{b-s-1} \varphi_j \varphi_{j+s} + \sum_{j=b-s}^{b-1} \varphi_j \varphi_{j+s} + \sum_{j=b}^{\infty} \varphi_j \varphi_{j+s} \\
&= (b-s) \left(\sum_{k=0}^{\infty} \psi_k \right)^2 + C(b, s, \xi) + D(b, s, \xi) + E(b, s, \xi),
\end{aligned}
$$

where

$$
\begin{aligned}
C(b, s, \xi) &= \sum_{j=0}^{b-s-1} a_j a_{j+s} - \left(\sum_{k=0}^{\infty} \psi_k \right) \sum_{j=0}^{b-s-1} (a_j + a_{j+s}), \\
D(b, s, \xi) &= \left(\sum_{k=0}^{\infty} \psi_k \right) \sum_{i=1}^{s} (\psi_i + \psi_{i+1} + \cdots + \psi_{i+b-1}) - \sum_{i=1}^{s} a_{b-s-1+i} (\psi_i + \psi_{i+1} + \cdots + \psi_{i+b-1}), \\
E(b, s, \xi) &= \sum_{i=1}^{\infty} (\psi_i + \psi_{i+1} + \cdots + \psi_{i+b-1}) (\psi_{i+s} + \psi_{i+s+1} + \cdots + \psi_{i+s+b-1}).
\end{aligned}
$$

For $b \le s$,

$$
\begin{aligned}
\sigma_\varepsilon^{-2} \mathrm{Cov}(X_{i+j}, X_{i+s+j}) &= \sum_{j=0}^{b-1} \varphi_j \varphi_{j+s} + \sum_{j=b}^{\infty} \varphi_j \varphi_{j+s} \\
&= F(b, s, \xi) + E(b, s, \xi),
\end{aligned}
$$

where

$$F(b,s,\xi) = \left(\sum_{k=0}^{\infty} \psi_k\right) \sum_{j=0}^{b-1} \left(\psi_{j+s-b+1} + \cdots + \psi_{j+s}\right) - \sum_{j=0}^{b-1} a_j \left(\psi_{j+s-b+1} + \cdots + \psi_{j+s}\right).$$

Thus it follows that

$$\sigma_\varepsilon^{-2} \mathrm{Cov}(t_b^0, t_b^s) = \begin{cases} \left(\sum_{k=0}^{\infty} \psi_k\right)^2 + \frac{1}{b}\{A(b,\xi) + B(b,\xi)\}, & s = 0, \\ \left(1 - \frac{s}{b}\right)\left(\sum_{k=0}^{\infty} \psi_k\right)^2 + \frac{1}{b}\{C(b,s,\xi) + D(b,s,\xi) + E(b,s,\xi)\}, & 1 \le s \le b-1, \\ \frac{1}{b}\{F(b,s,\xi) + E(b,s,\xi)\}, & b \le s. \end{cases} \tag{3.31}$$

Since $\gamma_{Y,b} = 2\left\{\mathrm{Cov}(t_b^0, t_b^s)\right\}^2$,

$$\sigma_\varepsilon^{-4}\gamma_{Y,b} =$$

$$\begin{cases} 2\left(\sum_{k=0}^{\infty} \psi_k\right)^4 + \frac{1}{b}\left(\sum_{k=0}^{\infty} \psi_k\right)^2 \{A(b,\xi) + B(b,\xi)\} + \frac{2}{b^2}\{A(b,\xi) + B(b,\xi)\}^2, & s = 0, \\ 2\left(1 - \frac{s}{b}\right)^2 \left(\sum_{k=0}^{\infty} \psi_k\right)^4 + \frac{4}{b}\left(1 - \frac{s}{b}\right)\left(\sum_{k=0}^{\infty} \psi_k\right)^2 \{C(b,s,\xi) + D(b,s,\xi) + E(b,s,\xi)\} & \\ + \frac{4}{b^2}\{C(b,s,\xi) + D(b,s,\xi) + E(b,s,\xi)\}^2, & 1 \le s \le b-1, \\ \frac{2}{b^2}\{F(b,s,\xi) + E(b,s,\xi)\}^2, & b \le s. \end{cases}$$

We have

$$\mathrm{V}(\bar{Y}_N) = N^{-1} \sum_{s=-(b-1)}^{b-1} \left(1 - \frac{|s|}{N}\right) \gamma_Y(s) + N^{-1} \sum_{b \le |s| \le N} \left(1 - \frac{|s|}{N}\right) \gamma_Y(s) = V_{1,n} + V_{2,n}(\text{say}).$$

Then

$$\begin{aligned} |V_{1,n}| &\le N^{-1} \sum_{s=-(b-1)}^{b-1} |\gamma_Y(s)| \\ &= N^{-1} V(Y_1) + N^{-1} \sum_{s=1}^{b-1} |\gamma_Y(s)| \\ &\le N^{-1}(4b-2)\left(\sum_{k=0}^{\infty} \psi_k\right)^4 + N^{-1} R_1(b,\xi) + N^{-1} 2 \sum_{s=1}^{b-1} R_2(b,s,\xi), \end{aligned}$$

where

$$R_1(b,\xi) = 2\left(\sum_{k=0}^{\infty}\psi_k\right)^4 + \frac{1}{b}\left(\sum_{k=0}^{\infty}\psi_k\right)^2 \{A(b,\xi)+B(b,\xi)\} + \frac{2}{b^2}\{A(b,\xi)+B(b,\xi)\}^2,$$

$$R_2(b,s,\xi) = 2\left(1-\frac{s}{b}\right)^2\left(\sum_{k=0}^{\infty}\psi_k\right)^4 + \frac{4}{b}\left(1-\frac{s}{b}\right)\left(\sum_{k=0}^{\infty}\psi_k\right)^2 \{C(b,s,\xi)+D(b,s,\xi)+E(b,s,\xi)\}$$
$$+\frac{4}{b^2}\{C(b,s,\xi)+D(b,s,\xi)+E(b,s,\xi)\}^2.$$

From the relation (3.9), we see that

$$\sum_{k=0}^{\infty}\psi_k = \{\phi(1)\}^{-1}. \tag{3.32}$$

Thus

$$\sup_{\phi\in\overline{O}_\delta(\phi_0)}\left(\sum_{k=0}^{\infty}\psi_k\right)^4 = \sup_{\phi\in\overline{O}_\delta(\phi_0)}\{\phi(1)\}^{-4} < \infty. \tag{3.33}$$

Since

$$\left|\sum_{j=0}^{b-1}a_j\right| \le K\left|\sum_{j=0}^{b-1}\sum_{k=j+1}^{\infty}k^{p-1}\xi^{-k}\right|,$$

it follows that

$$\sup_{\phi\in\overline{O}_\delta(\phi_0)}\left|\sum_{j=0}^{b-1}a_j\right| \le \sup_{\phi\in\overline{O}_\delta(\phi_0)}K(\phi)\times\sum_{j=0}^{b-1}\sup_{\phi\in\overline{O}_\delta(\phi_0)}\left|\sum_{k=j+1}^{\infty}k^{p-1}\xi(\phi)^{-k}\right| < \infty.$$

Thus $\sup_{\phi\in\overline{O}_\delta(\phi_0)}|A(b,\xi)| < \infty$. From the inequality (3.30), we have $\sup_{\phi\in\overline{O}_\delta(\phi_0)}|B(b,\xi)| < \infty$. Thus

$$N^{-1}\sup_{\phi\in\overline{O}_\delta(\phi_0)}|R_1(b,\xi)| \to 0. \tag{3.34}$$

$$\left|b^{-1}\sum_{s=1}^{b-1}\sum_{j=0}^{b-s-1}a_j\right| \le Kb^{-1}\sum_{s=1}^{b-1}\sum_{j=0}^{b-s-1}\sum_{k=j+1}^{\infty}k^{p-1}\xi^{-k}$$
$$\le K\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}k^{p-1}\xi^{-k}$$

By the similar arguments, we can prove that

$$N^{-1}\sup_{\phi\in\overline{O}_\delta(\phi_0)}\left|\sum_{s=1}^{b-1}R_2(b,s,\xi)\right| \to 0, \tag{3.35}$$

and thus $\sup_{\phi\in\overline{O}_\delta(\phi_0)}|V_{1,n}|\to 0$.

For $V_{2,n}$,

$$
\begin{aligned}
|V_{2,n}| &\le N^{-1}\sum_{b\le|s|\le N}|\gamma_Y(s)| \\
&\le 8b^{-2}N^{-1}\sum_{s=b}^{N}\{F(b,s,\xi)^2+E(b,s,\xi)^2\}.
\end{aligned}
$$

We have

$$
\begin{aligned}
F(b,s,\xi)^2 \le 2&\left\{\left(\sum_{k=0}^{\infty}\psi_k\right)^2\left(\sum_{j=0}^{b-1}(\psi_{j+s-b+1}+\cdots+\psi_{j+s})\right)^2\right. \\
&+\left.\left(\sum_{j=0}^{b-1}a_j\left(\psi_{j+s-b+1}+\cdots+\psi_{j+s}\right)\right)^2\right\} = F_1^2(b,s,\xi)+F_2^2(b,s,\xi)(\text{say}).
\end{aligned}
$$

Since

$$
\begin{aligned}
b^{-2}N^{-1}\sum_{s=b}^{N}F_1^2(b,s,\xi)^2 &\le \left(\sum_{k=0}^{\infty}\psi_k\right)^2K^2f(\xi,b)^2b^{-2}N^{-1}\sum_{s=b}^{N}\left(\sum_{i=s-b+1}^{s}i^{p-1}\xi^{-i}\right)^2 \\
&\le \left(\sum_{k=0}^{\infty}\psi_k\right)^2K^2f(\xi,b)^2b^{-2}N^{-1}\sum_{s=1}^{N}\left(\sum_{i=1}^{s}i^{p-1}\xi^{-i}\right)^2,
\end{aligned}
$$

it follows $\sup_{\phi\in\overline{O}_\delta(\phi_0)}b^{-2}N^{-1}\sum_{s=b}^{N}F_1^2(b,s,\xi)^2\to 0$. Similarly $\sup_{\phi\in\overline{O}_\delta(\phi_0)}b^{-2}N^{-1}\sum_{s=b}^{N}F_2^2(b,s,\xi)^2\to 0$. Thus $\sup_{\phi\in\overline{O}_\delta(\phi_0)}|V_{2,n}|\to 0$.

Assumption 2 (iii)

In order to show Assumption 2 (iii) holds, we write $\mathrm{E}(\bar{t}_n)^4$ as

$$
\begin{aligned}
\mathrm{E}(\bar{t}_n)^4 &= N_n^{-4}3\left\{\mathrm{E}\left(\sum_{i=0}^{N_n-1}t_b^i\right)^2\right\}^2 \\
&= N_n^{-2}3\left\{\mathrm{E}\left(\frac{1}{\sqrt{N_n}}\sum_{i=0}^{N_n-1}t_b^i\right)^2\right\}^2.
\end{aligned}
$$

We have

$$
\begin{aligned}
\mathrm{E}\left(\frac{1}{\sqrt{N_n}}\sum_{i=0}^{N_n-1}t_b^i\right)^2 &= \sum_{s=-N_n}^{N_n}\left(1-\frac{|s|}{N_n}\right)\mathrm{Cov}(t_b^0,t_b^s) \\
&= \sum_{s=-(b-1)}^{b-1}\left(1-\frac{|s|}{N_n}\right)\mathrm{Cov}(t_b^0,t_b^s)+\sum_{b\le|s|\le N_n}\left(1-\frac{|s|}{N_n}\right)\mathrm{Cov}(t_b^0,t_b^s) \\
&= V_{3,n}+V_{4,n}\ (\text{say}).
\end{aligned}
$$

Thus

$$\mathrm{E}(\bar{t}_n)^4 \le 6N_n^{-2}\left\{(V_{3,n})^2 + (V_{4,n})^2\right\}. \tag{3.36}$$

By the same argument of showing (3.34) and (3.35), it is shown that $N^{-1}\sup_{\phi\in\overline{O}_\delta(\phi_0)}\{(V_{3,n})\}\to 0$ and $N^{-1}\sup_{\phi\in\overline{O}_\delta(\phi_0)}\{(V_{4,n})\}\to 0$, which completes the proof.

Theorem 5 states that if $\{X_i\}$ is a causal Gaussian AR(p) process $\hat{D}_n - D_n$ converges to zero with probability 1. Let G be the distribution function of $\max_{1\le i\le k-1}(U_i - U)$, where $U_1,\ldots,U_{k-1},U$ are independently distributed as standard normal distribution. Given a sample $(X_{i,1},X_{i,2},\ldots,X_{i,n})$, $i=1,\ldots,k$, let $(X^*_{i,1},X^*_{i,2},\ldots,X^*_{i,n})$, $i=1,\ldots,k$ be the sample generated by the stationary Gaussian ARMA resampling and Y_n^* be defined by

$$Y_n^* = \frac{\sqrt{n}\left(\overline{X}^*_{[k-1]} - \overline{X}^*_k\right)}{\hat{\sigma}^*_n}.$$

Define

$$G_n(y) := P\left(Y_n^* \le y \mid \hat{\phi}_n\right),$$

where $\hat{\phi}_n$ is the maximum likelihood estimator of ϕ. Note that for each $y\in\mathbb{R}$, $G_n(y)$ is a random variable.

Theorem 5 *Let $\{X_i\}$ be a causal Gaussian AR(p) process with the parameter ϕ_0. Then for any $y\in\mathbb{R}$,*

$$G_n(y)\to G(y)\ \text{as } n\to\infty\quad \text{with probability } 1$$

and for any $y\in(0,1)$,

$$G_n^{-1}(y)\to G^{-1}(y)\ \text{as } n\to\infty.\quad \text{with probability } 1 \tag{3.37}$$

Proof: Let

$$\sigma_n^{2*} = \mathrm{V}\left(\sqrt{n}\bar{X}^*|X_1,\ldots,X_n\right) = \sum_{s=-n}^{n}\left(1-\frac{|s|}{n}\right)\gamma_s(\hat{\phi}_n).$$

There exists $\tilde{\Omega}\subset\Omega$ such that $P(\tilde{\Omega})=1$ and $\hat{\phi}_n(\omega)\to\phi_0$ for all $\omega\in\tilde{\Omega}$. For any $\omega\in\tilde{\Omega}$ and $\delta>0$, there exists $n_0(\omega)$ such that $\hat{\phi}_n\in O_\delta(\phi_0)$ for $n\ge n_0$. Since

$$\sigma_n^{2*} - \sigma_n^2 = \sum_{s=-n}^{n}\left(1-\frac{|s|}{n}\right)\left\{\gamma_s(\hat{\phi}_n) - \gamma_s(\phi_0)\right\},$$

it follows that for any $\omega \in \tilde{\Omega}$, $\delta > 0$ and $n \geq n_0(\omega)$,

$$
\begin{aligned}
\left|\sigma_n^{2*} - \sigma_n^2\right| &\leq \sum_{s=-n}^{n} \left|\gamma_s(\hat{\phi}_n) - \gamma_s(\phi_0)\right| \\
&\leq \sum_{s=-n}^{n} \sup_{\phi \in O_\delta(\phi_0)} |\gamma_s(\phi) - \gamma_s(\phi_0)| \\
&\leq \sum_{s=-\infty}^{\infty} \sup_{\phi \in O_\delta(\phi_0)} |\gamma_s(\phi)| + \sum_{s=-\infty}^{\infty} |\gamma_s(\phi_0)| \\
&\leq \sum_{s=-\infty}^{\infty} \sup_{\phi \in O_\delta(\phi_0)} \left|\sum_{i=1}^{k} \sum_{j=0}^{r_i} c_{ij} s^j \xi_i(\phi)^{-s}\right| + \sum_{s=-\infty}^{\infty} |\gamma_s(\phi_0)| < \infty.
\end{aligned}
$$

Since $\gamma_s(\hat{\phi}_n) - \gamma_s(\phi_0) \to 0$ for each $s \geq 1$, it follows from the domonated convergence theorem that $\sigma_n^{2*} - \sigma_n^2 \to 0$ with probability 1.

Since $\sigma_n^*/\sigma_n \to 1$ with probability 1 and $\sigma_n/\hat{\sigma}_n^* \to 1$ in probability conditional on $\hat{\phi}_n$, it follows from the standard argument (see Rao (1973), section 2c) that for each $y \in \mathbb{R}$ and $\omega \in \tilde{\Omega}$,

$$
\begin{aligned}
P\left(Y_n^* \leq y \middle| \hat{\phi}_n\right) &= P\left(\frac{\sqrt{n}\left(\overline{X}_{[k-1]}^* - \overline{X}_k^*\right)}{\hat{\sigma}_n^*} \leq y \middle| \hat{\phi}_n\right) \\
&= P\left(\frac{\sqrt{n}\left(\overline{X}_{[k-1]}^* - \overline{X}_k^*\right)}{\sigma_n^*} \times \frac{\sigma_n^*}{\sigma_n} \times \frac{\sigma_n}{\hat{\sigma}_n^*} \leq y \middle| \hat{\phi}_n\right) \\
&\to G(y) \quad \text{as } n \to \infty.
\end{aligned}
$$

The convergence (3.37) follows from the standard result of the equivalence of the weak convergence of the distribution functions and convergence of the quantile functions (see, for example, van der Vaart (1998), chapter 21). □

3.6 Simulation study

Simulation study was made on the MA(2) process and the AR(1) process to see finite sample properties of the suggested selection rule. The section 3.6.1 and the section 3.6.2 discuss the methods for MA(2) and AR(1) respectively. The results are shown in the table 3.1 for MA(2) and the table 3.2 for AR(1).

3.6.1 Simulation steps for MA(2) process

Computing the true D_n

Let k be the number of populations, n be the sample size of each population and b be the subsample size.

1. Generate k independent samples from MA(2) process with the parameters, $\theta = (\theta_1,\ \theta_2)$, that is,

$$X_{i,t} = Z_{i,t} + \theta_1 Z_{i,t-1} + \theta_2 Z_{i,t-2}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.38}$$

where $\{Z_{i,t}\}$ is a sequence of independent standard normal random variables. (need to generate n observations for each population separately as dependency arises if kn observations are generated consecutively.)
2. Obtain the subsumpling estimator, $\hat{\sigma}_n^2$, of σ_n^2 from the samples generated from (3.38).
3. Obtain Y_n.
4. Repeat the steps, 1,2, and 3 NR times. (NR=Number of Replications)
5. Obtain $\hat{D}_n(P^*)$ by computing P^*-quantile of Y_n.

Computing $\hat{D}_n$

Let k be the number of populations, n be the sample size of each population and b be the subsample size.

1. Generate samples from populations.
Generate k independent samples from the MA(2) process with different means, $\{\mu_i, i = 1, 2, \ldots, k\}$, and a common parameters, $\boldsymbol{\theta} = (\theta_1,\ \theta_2)$, that is

$$X_{i,t} = \mu_i + Z_{i,t} + \theta_1 Z_{i,t-1} + \theta_2 Z_{i,t-2}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.39}$$

where $\{Z_{i,t}\}$ is a sequence of independent standard normal random variables.

2. Estimate the parameters.
Obtain the average, $\bar{X}_i$, for each sample.
Deduct the sample average from each sample.

$$X'_{i,t} = X_{i,t} - \bar{X}_i \tag{3.40}$$

Obtain the MLE, $\hat{\boldsymbol{\theta}}_i = (\hat{\theta}_{1,i}, \hat{\theta}_{2,i})$ from each sample, $\{X'_{i,t}\}_{t=1}^n$.
Then the estimate of θ is the average of $\hat{\theta}_i$:

$$\hat{\theta}_n = \left(k^{-1} \sum_{i=1}^{k} \hat{\theta}_{1,i}, \quad k^{-1} \sum_{i=1}^{k} \hat{\theta}_{2,i} \right). \tag{3.41}$$

3.Estimate D_n
1. Generate k independent samples using MA(2) process with the parameter, $\hat{\boldsymbol{\theta}}$, that is,

$$X^*_{i,t} = Z^*_{i,t} + \hat{\theta}_1 Z^*_{i,t-1} + \hat{\theta}_2 Z^*_{i,t-2}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.42}$$

where $\{Z^*_{i,t}\}$ is a sequence of independent standard normal random variables.
2. Obtain the subsampling estimator $\hat{\sigma}^{2*}_n$ of σ^2_n from the samples generated by the equation (3.42).
3. Obtain Y^*_n by (3.21).
4. Repeat the steps 1, 2 and 3 NR times. (NR=Number of Replications)
5.Obtain $\hat{D}_n(P^*)$ by comuting P^*-quantile of Y^*_n.

MSE for $\hat{D}_n$ and probability of correct selection

Mean Squared Error (MSE) for $\hat{D}_n$ is defined as;

$$\text{MSE} = \text{E}(\hat{D}_n - D_n)^2. \tag{3.43}$$

As discussed in the subsection 3.3.2,

$$\text{PCS} = P(CS) = P\left(\bar{X}_{(k)} \geq \bar{X}_{[k]} - \frac{\hat{D}_n \hat{\sigma}_n}{\sqrt{n}}\right). \tag{3.44}$$

MSE and PCS are obtained by simulation as follows.

Generate samples NS times following the item 1 of "Computing $\hat{D}_n$" in this subsection. Let $\hat{D}_{n,i}$ be the $\hat{D}_n$ obtained from the ith sample. Then MSE is computed by

$$\frac{1}{NS}\sum_{i=1}^{NS}(\hat{D}_{n,i} - D_n)^2. \tag{3.45}$$

PCS is computed by

$$\frac{1}{NS}\sum_{i=1}^{NS} I\left(\bar{X}_{(k),i} \geq \bar{X}_{[k],i} - \frac{\hat{D}_{n,i}\hat{\sigma}_{n,i}}{\sqrt{n}}\right), \tag{3.46}$$

where $\bar{X}_{(k),i}$, $\bar{X}_{[k],i}$, $\hat{D}_{n,i}$, and $\hat{\sigma}_{n,i}$ are $\bar{X}_{(k)}$, $\bar{X}_{[k]}$, $\hat{D}_n$, and $\hat{\sigma}_n$, for the ith sample respectively and I is the indicator function.

3.6.2 Simulation steps for AR (1) process

Computing the true D_n

Let k be the number of populations, n be the sample size of each population and b be the subsample size.

1. Generate k independent samples from AR(1) process with the parameter, ϕ, that is,

$$X_{i,t} = \phi X_{i,t-1} + Z_{i,t}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.47}$$

where $\{Z_{it}\}$ is a sequence of independent standard normal random variables. .
2. Obtain the subsampling estimator, $\hat{\sigma}_n^2$, of σ_n^2 from the samples generated from the equation (3.47).
3. Obtain Y_n.
4. Repeat the steps, 1,2, and 3 NR times. (NR=Number of Replications)
5. Obtain $D_n(P^*)$ by computing P^*-quantile of Y_n.

Computing $\hat{D}_n$

Let k be the number of populations, n be the sample size of each population and b be the subsample size.

1. Generate samples from pupulations.
Generate k independent samples from the AR(1) process with different means, $\{\mu_i, i = 1, 2, \ldots, k\}$, and a common parameter, ϕ. Let

$$Y_{i,t} = X_{i,t} - \mu_i.$$

Then, Y_{it} is from the AR(1) process;

$$Y_{i,t} = \phi Y_{i,t-1} + Z_{i,t}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.48}$$

where $\{Z_{i,t}\}$ is a sequence of independent standard normal random variables.
2. Estimate the parameters.
Obtain the average, $\bar{X}_i$, for each sample.
Deduct the sample average from each sample.

$$X'_{i,t} = X_{i,t} - \bar{X}_i. \tag{3.49}$$

Obtain the MLE, $\hat{\phi}_i$, from each sample, $\{X'_{i,t}\}_{t=1}^n$.
Then the estimate of ϕ is the average of $\hat{\phi}_i, i = 1, 2, \ldots, k$:

$$\hat{\phi}_n = k^{-1} \sum_{i=1}^{k} \hat{\phi}_i. \tag{3.50}$$

3.Estimate D_n
1. Generate k independent samples using AR(1) process with the parameter, $\hat{\phi}$, that is,

$$X^*_{i,t} = \hat{\phi} X^*_{i,t-1} + Z^*_{i,t}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.51}$$

where $\{Z^*_{it}\}$ is a sequence of independent standard normal random variables.
2. Obtain the subsampling estimator $\hat{\sigma}_n^{2*}$ of σ_n^2 from the samples generated by (3.51).
3. Obtain Y_n^* by (3.21)
4. Repeat the steps 1, 2 and 3 NR times. (NR=Number of Replications)
5. Obtain $\hat{D}_n(P^*)$ by computing P^*-quantile of Y_n^* .

MSE for $\hat{D}_n$ and probability of correct selection (PCS)

Mean Squared Error (MSE) for $\hat{D}_n$ is defined as;

$$\mathrm{MSE} = \mathrm{E}(\hat{D}_n - D_n)^2. \tag{3.52}$$

As discussed in the subsection 3.3.2,

$$\mathrm{PCS} = P(CS) = P\left(\bar{X}_{(k)} \geq \bar{X}_{[k]} - \frac{\hat{D}_n \hat{\sigma}_n}{\sqrt{n}}\right). \tag{3.53}$$

MSE and PCS are obtained by simulation.

Generate samples NS times following the item 1 of "Computing $\hat{D}_n$"in this subsection . Let $\hat{D}_{n,i}$ be the $\hat{D}_n$ generated from the ith sample. Then MSE is computed by

$$\frac{1}{NS}\sum_{i=1}^{NS}(\hat{D}_{n,i} - D_n)^2. \tag{3.54}$$

PCS is computed by

$$\frac{1}{NS}\sum_{i=1}^{NS} I\left(\bar{X}_{(k),i} \geq \bar{X}_{[k],i} - \frac{\hat{D}_{n,i}\hat{\sigma}_{n,i}}{\sqrt{n}}\right), \tag{3.55}$$

where $\bar{X}_{(k),i}$, $\bar{X}_{[k],i}$, $\hat{D}_{n,i}$, and $\hat{\sigma}_{n,i}$ are $\bar{X}_{(k)}$, $\bar{X}_{[k]}$, $\hat{D}_n$, and $\hat{\sigma}_n$, for the ith sample respectively and I is the indicator function.

3.6.3 Simulation result

In this section, the result of a simulation study is given in order to see finite sample properties of the suggested selection rule.

A relevant measure for the performance of selection rule is the expected size of the selected subset. Let S be the size of the subset selected by the rule $R(\hat{D}_n)$. Since S is a random variable, we are interested in the expectation of S, $\mathrm{E}(S)$. We would like $\mathrm{E}(S)$ to be small. An approximate value of $\mathrm{E}(S)$ is computed by

$$\frac{1}{NS}\sum_{i=1}^{NS} S_i, \tag{3.56}$$

where S_i is the number of elements in the set $\left\{1 \leq j \leq k : \bar{X}_{j,i} \geq \bar{X}_{[k],i} - \frac{\hat{D}_{n,i}\hat{\sigma}_{n,i}}{\sqrt{n}}\right\}$.

Notation:
k: number of populations

b: subsample size to obtain $\hat{\sigma}_n$
NR: number of replication to obtain Y_n distribution
NS: number of replication to obtain MSE, PCS and E(S)
PCS: probability for selecting the correct subset containing the sample with the largest mean.

We created three data sets each from MA(2) process and from AR(1) process. The samples for 5 populations are independent. The means are set to be $(1, 2, 3, 4, 5)$, $(1.47, 1.48, 1.49, 1.49, 1.50)$ and $(1, 1, 1, 1, 3)$ respectively for MA(2) and $(1, 2, 3, 4, 5)$, $(1, 1.2, 1.4, 1.4, 1.5)$ and $(1, 1, 1, 1, 3)$ for AR(1). The parameters for MA(2) are -0.5 and -0.2. The parameter for AR(1) is 0.8. We ran simulations with three different sample size; one: $n = 50$, $b = 10$, another $n = 100$, $b = 20$, and the other, $n = 200$, $b = 40$. NS is set to be 300. The probability for selecting the subset which contains the population with the largest mean, P^*, is to be 0.95 and 0.99. Note that the true D_n is estimated under $NR = 10,000$ for a better estimate.

The results are shown in the Table 3.1 and the Table 3.2.

In case of MA(2) with the mean, 1, 2, 3, 4, 5, and also with the mean, 1, 1, 1, 1, 3, all PCS of $R(\hat{D}_n)$, PCS of $R(D_n)$, E(S) of $R(\hat{D}_n)$ and E(S) of $R(D_n)$are 1. That is to say, only the population with the largest mean was selected when using D_n and $\hat{D}_n$ and no incorrect selection was made.

However, when there is only a small difference among the means, the subset selected contains the population whose mean value is not the largest. E(S) is close to 5 and most of PCS is less than 1 in both cases of MA(2) with the mean, 1.47, 1.48, 1.49, 1.49, 1.50, and AR(1) with the mean, 1, 1.2, 1,4, 1.4, 1.5.

In case of AR(1) with the mean, 1, 2, 3, 4, 5, and also with the mean, 1, 1, 1, 1, 3, PCS of $R(\hat{D}_n)$ and PCS of $R(D_n)$ are 1 except the case where P^* is 0.95, $n = 100$ and $b = 20$, while E(S) of $R(\hat{D}_n)$ and E(S) of $R(D_n)$ are more than 1. That is to say, for the most of these cases, the subset selected contains not only the population with the largest mean, but also some populations with the mean that is not the largest. The E(S) is smaller when the sample size is larger.

PCS of $R(\hat{D}_n)$ and PCS of $R(D_n)$ are similar. E(S) of $R(\hat{D}_n)$ is smaller than E(S) of $R(D_n)$, that is not what is expected, although they are similar. This remains to be further investigated.

In case of the mean (1, 2, 3, 4, 5) and (1, 1, 1, 1. 3), compared with the result from AR(1) with the parameter, 0.8, the MA(2) with the parameter, -0.5 and -0.2, resulted in smaller E(S) and the same PCS, 1, in most of the cases. In addition, D_n for AR(1) with the larger parameter is larger than that for MA(2) with the smaller parameters.

parameter:$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$					1,2,3,4,5					
parameter:θ_1, θ_2					-0.5, -0.2					
n	b	NR	NS	D_n	P^*	MSE	PCS of $R(\hat{D}_n)$	PCS of $R(D_n)$	E(S) of $R(\hat{D}_n)$	E(S) of $R(D_n)$
50	10	1000	300	2.223116	0.95	0.1757988	1	1	1	1
				2.875941	0.99	0.3195195	1	1	1	1
100	20	1000	300	2.538334	0.95	0.08408475	1	1	1	1
				3.243584	0.99	0.13970983	1	1	1	1
200	40	1000	300	2.857209	0.95	0.02685977	1	1	1	1
				3.723132	0.99	0.06446194	1	1	1	1
parameter:$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$					1.47,1.48,1.49,1.49,1.50					
parameter:θ_1, θ_2					-0.5, -0.2					
n	b	NR	NS	D_n	P^*	MSE	PCS of $R(\hat{D}_n)$	PCS of $R(D_n)$	E(S) of $R(\hat{D}_n)$	E(S) of $R(D_n)$
50	10	1000	300	2.223116	0.95	0.1757988	0.94	0.9633333	4.426667	4.723333
				2.875941	0.99	0.3195195	0.9766667	1	4.746667	4.95
100	20	1000	300	2.538334	0.95	0.08408475	0.953333	0.9733333	4.446667	4.616667
				3.243584	0.99	0.13970983	0.99	0.9933333	4.806667	4.886667
200	40	1000	300	2.857209	0.95	0.02685977	0.9733333	0.98	4.373333	4.40
				3.723132	0.99	0.06446194	0.9866667	0.9966667	4.763333	4.83
parameter:$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$					1,1,1,1,3					
parameter:θ_1, θ_2					-0.5, -0.2					
n	b	NR	NS	D_n	P^*	MSE	PCS of $R(\hat{D}_n)$	PCS of $R(D_n)$	E(S) of $R(\hat{D}_n)$	E(S) of $R(D_n)$
50	10	1000	300	2.223116	0.95	0.1757988	1	1	1	1
				2.875941	0.99	0.3195195	1	1	1	1
100	20	1000	300	2.538334	0.95	0.08408475	1	1	1	1
				3.243584	0.99	0.13970983	1	1	1	1
200	40	1000	300	2.857209	0.95	0.02685977	1	1	1	1
				3.723132	0.99	0.06446194	1	1	1	1

Table 3.1: MA(2) process

parameter:$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$					1, 2, 3, 4, 5					
parameter:ϕ					0.8					
n	b	NR	NS	D_n	P^*	MSE	PCS of $R(\hat{D}_n)$	PCS of $R(D_n)$	E(S) of $R(\hat{D}_n)$	E(S) of $R(D_n)$
50	10	1000	300	4.866158	0.95	0.2169513	1	1	2.566667	2.713333
				6.624203	0.99	0.6684435	1	1	2.5	2.55
100	20	1000	300	4.198213	0.95	0.04616113	0.9933333	0.993333	2.026667	2.043333
				5.517788	0.99	0.1014397	1	1	4.093333	4.38
200	40	1000	300	3.84575	0.95	0.01547903	1	1	1.666667	1.67667
				5.072887	0.99	0.05217417	1	1	1.936667	1.97667
parameter:$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$					1, 1.2, 1.4, 1.4, 1.5					
parameter:ϕ					0.8					
n	b	NR	NS	D_n	P^*	MSE	PCS of $R(\hat{D}_n)$	PCS of $R(D_n)$	E(S) of $R(\hat{D}_n)$	E(S) of $R(D_n)$
50	10	1000	300	4.866158	0.95	0.2169513	0.9533333	0.96	4.576667	4.723333
				6.624203	0.99	0.6684435	0.99	0.993333	4.88	4.936667
100	20	1000	300	4.198213	0.95	0.04616113	0.9533333	0.9633333	4.653333	4.703333
				5.517788	0.99	0.1014397	0.99	0.99	4.903333	4.926667
200	40	1000	300	3.84575	0.95	0.01547903	0.9533333	0.97	4.553333	4.56
				5.072887	0.99	0.05217417	0.99	0.9933333	4.873333	4.896667
parameter:$\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$					1, 1, 1, 1, 3					
parameter:ϕ					0.8					
n	b	NR	NS	D_n	P^*	MSE	PCS of $R(\hat{D}_n)$	PCS of $R(D_n)$	E(S) of $R(\hat{D}_n)$	E(S) of $R(D_n)$
50	10	1000	300	4.866158	0.95	0.2169513	1	1	3.103333	3.316667
				6.624203	0.99	0.6684435	1	1	3.906667	4.286667
100	20	1000	300	4.198213	0.95	0.04616113	1	1	2.026667	2.06
				5.517788	0.99	0.1014397	1	1	3.043333	3.193333
200	40	1000	300	3.84575	0.95	0.01547903	1	1	1.226667	1.243333
				5.072887	0.99	0.05217417	1	1	1.68	1.743333

Table 3.2: AR(1) process

3.7 Conclusion

We developed the selection rule, R, to select the subset that contains the population with the largest means when the samples are from the ARMA process. The rule is;
R: Select the ith population if and only if

$$\overline{X}_i \geq \overline{X}_{[k]} - \frac{\hat{D}_n \hat{\sigma}_n}{\sqrt{n}}, \tag{3.57}$$

where $\hat{D}_n$ is estimated from the sample. $\hat{D}_n$ is obtained by using the Monte Carlo Simulation. The size of the subset selected, S, should be as small as possible.

We made a simulation study using different figures for the parameters, μ, θ and ϕ to create samples with different properties. The simulation study reveals that the selection rule $R(\hat{D}_n)$ gives smaller E(S), when the difference among $\mu_i, i = 1, 2, \ldots, k$, is large and larger E($S$) when the difference is small. From the equation (3.10) and the equation (3.11), the larger parameters are, the larger σ_n is. D_n is dependent on the parameters, θ and ϕ and the simulation result shows that D_n for AR(1) with the larger parameter is larger than that of MA(2) with the smaller parameters. Then the selection rule $R(\hat{D}_n)$ gives the larger range for populations to be selected when the parameter is large, resulting in large E(S). As mentioned in the subsection 3.6.3, the simulation results support this statement.

Subsampling method is used to estimate σ_n^2. We must determine what the subsample size, b, should be. We determined b by testing different values in our simulation study. Therefore, the size of b remains to be studied. In addition, the E(S) of $R(D_n)$ is larger than the E(S) of $R(\hat{D}_n)$ in case where they are not equal to 1. It should not be, although they are similar. This also remains to be investigated.

Appendix C

The proof of the nondependency of the distribution of (3.15) on $(\mu_1, \mu_2, \ldots, \mu_k)', \sigma_\varepsilon^2$:
It suffices to show that the distribution of $\hat{\sigma_n}/\sigma_n$ is free of $(\mu_1, \mu_2, \ldots, \mu_k)', \sigma_\varepsilon^2$.

$$
\begin{aligned}
\frac{\sigma_n}{\sigma_\varepsilon} &= \frac{1}{\sigma_\varepsilon}\sqrt{\sum_{s=-(n-1)}^{n-1}\left(1-\frac{|s|}{n}\right)\gamma(s)} \\
&= \sqrt{\sum_{s=-(n-1)}^{n-1}\left(1-\frac{|s|}{n}\right)\frac{\gamma(s)}{\sigma_\varepsilon^2}} \\
&= \sqrt{\sum_{s=-(n-1)}^{n-1}\left(1-\frac{|s|}{n}\right)\sum_{j=0}^{\infty}\psi_j\psi_{j+|k|}}.
\end{aligned}
$$

Therefore $\frac{\sigma_n}{\sigma_\varepsilon}$ is free of $(\mu_1, \mu_2, \ldots, \mu_k)', \sigma_\varepsilon^2$.
Since

$$
\frac{\hat{\sigma}_{b,n}^2}{\sigma_\varepsilon^2} = \frac{1}{k}\cdot\frac{1}{n-b+1}\sum_{j=1}^{k}\sum_{i=1}^{n-b+1}\left\{\frac{1}{\sqrt{b}}\sum_{t=i}^{i+b-1}\frac{\left(X_{j,t}-\bar{X}_j\right)}{\sigma_\varepsilon}\right\}^2 \tag{C.1}
$$

and

$$
\begin{aligned}
\frac{(X_{j,t}-\bar{X}_j)}{\sigma_\varepsilon} &= \frac{1}{\sigma_\varepsilon}\left\{(X_{j,t}-\mu_j)-(\bar{X}_j-\mu_j)\right\} \\
&= \left(\frac{X_{j,t}-\mu_j}{\sigma_\varepsilon}\right) - \frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_{j,t}-\mu_j}{\sigma_\varepsilon}\right),
\end{aligned}
$$

$\hat{\sigma}_{b,n}^2/\sigma_\varepsilon^2$ is a function of the kn-dimensional vector,

$$
\left(\frac{X_{11}-\mu_1}{\sigma_\varepsilon}, \frac{X_{12}-\mu_1}{\sigma_\varepsilon}, \ldots, \frac{X_{kn}-\mu_k}{\sigma_\varepsilon}\right)'.
$$

Therefore, the distribution of $\frac{\hat{\sigma}_n}{\sigma_n}$ does not depend on $(\mu_1, \mu_2, \ldots, \mu_k)'$ and σ_ε^2. $\square$

Reference

[1] Aitkin, M. (1981) "A Note on the regression analysis of censored data". *Technometrics.* Vol. 23, No. 2, 161-163.

[2] Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods.* Second edition. Springer.

[3] Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data.* Cambridge University Press, Cambridge.

[4] Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics.* Cambridge University Press, Cambridge.

[5] Carlstein, E. (1986). "The use of subseries methods for estimating the variance of a general statistic from a stationary time series". *Ann. Statist..* 14, 1171-1179.

[6] Dempster, A. P., Laird, N.M. and Rubin D.B. (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society.* Series B , 39, 1-38.

[7] Fukuchi, J. I., (1999). "Subsampling and model selection in time series analysis". *Biometrika.* Vol. 86, No.3, 591-604.

[8] Fukuchi, J. and Sunami, S (2005). "Statistical methods for revenue management"(in Japanese). *Gakushuin economic papers.* 42(1), 33-45. Gakushuin University, Tokyo.

[9] Gupta, S. S., (1956). *On a decision rule for a problem in ranking means* Mimeo. Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina.

[10] Gupta, S. S. (1965) "On some multiple decision (Selection and Ranking)", *Technometrics.* 225-245.

[11] Gupta, S. S., Nagel, K. and Panchapakesan (1973). "On the order statistics from equally correlated normal random variables". *Biometrika.* Vol. 60, No.2, 403-413.

[12] Gupta, S. S., Nagel, K. and Panchapakesan (1983). "The distribution of the maximum of N equally correlated normal standardised random variables" *H. Kres, Statistical Tables for multivariate analysis.* Springer-Verlag, New York.

[13] Gupta, S. S. and Panchapakesan (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations.* John Wiley & Sons.

[14] Gupta, S. S., and Sobel, M. (1957) "On a statistic which arises in selection and ranking problems", *The Annals of Mathematical Statistics.* Vol. 28, 957-967.

[15] Hilbe, J. M. (2007). *Negative Binomial regression.* Cambridge University Press, New York.

[16] Lee, S. (2011). *Study of demand models and price optimization performance.* Ph. D Thesis, Georgia Institute of Technology.

[17] Politis, D. N., and Romano, J. P. (1993). "On the sample variance of linear statistics derived from mixing sequences". *Stochastic Processes and Their Applications.* Vol. 45, issue 1, 155-167.

[18] Politis, D. N., Romano, J. P. and Wolf, M (1999). *Subsampling.* Springer, New York.

[19] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* Second edition. John Wiley, New York.

[20] Skwarek, D. K. (1994). *Competitive impacts of yield management system components: Forecasting and sell-up models.* Master thesis, MIT, Boston.

[21] Stefanescu, C. (2005). "Forecasting models for correlated demand." London Business School, London.

[22] Sunami, S. (2009) "Hotel revenue management and demand forecast"(in Japanese). Master Thesis, Gakushuin University.

[23] Steiner, W. J., Brezger, A., Christiane, B. (2007). "Flexible estimation of price response functions using retail scanner data." *Journal of Retailing and Consumer Services* 14, 383-393

[24] Talluri, K. T. and Van Ryzin, G. J. (2004). *The Theory and Practice of Revenue Management.* Kluwer Academic Publishers, Boston.

[25] van der Vaart (2000). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

[26] Weatherford L. R. Kimes, S. E. (2003). "A comparison of forecasting methods for the hotel revenue management" *International Journal of Forecasting.* 19, 401-415

[27] Wickham, R. R. (1993). *Evaluation of forecasting techniques for short-term demand of air transportation* Master Thesis, MIT, Boston.

[28] Winkelmann, R. (2010). *Econometric Analysis of Count Data.* Springer, Verlag, Berlin Heidelberg.