WHICH METHOD GIVES THE BEST FORECAST FOR LONGITUDINAL BINARY RESPONSE DATA?: A SIMULATION STUDY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YASEMİN ASLAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

SEPTEMBER 2010

Approval of the thesis:

WHICH METHOD GIVES THE BEST FORECAST FOR LONGITUDINAL BINARY RESPONSE DATA?: A SIMULATION STUDY

submitted by **YASEMİN ASLAN** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Canan ÖZGEN
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. H. Öztaş AYHAN
Head of Department, **Statistics Department** _____

Assist. Prof. Dr. Özlem İLK
Supervisor, **Statistics Department, METU** _____

Examining Committee Members:

Assist. Prof. Dr. Kasırga Yıldırak
Department of Economics, Trakya University _____

Assist. Prof. Dr. Özlem İlk
Department of Statistics, METU _____

Assist. Prof. Dr. Ceylan Yozgatlıgil
Department of Statistics, METU _____

Assist. Prof. Dr. Zeynep Işıl Kalaylıoğlu
Department of Statistics, METU _____

Assist. Prof. Dr. Burçak Berna Başbuğ Erkan
Department of Statistics, METU _____

Date: 16.09.2010

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : YASEMİN ASLAN

Signature :

ABSTRACT

WHICH METHOD GIVES THE BEST FORECAST FOR LONGITUDINAL BINARY RESPONSE DATA?: A SIMULATION STUDY

ASLAN, Yasemin
M.S., Department of Statistics
Supervisor: Assist. Prof. Dr. Özlem İLK

September 2010, 68 pages

Panel data, also known as longitudinal data, are composed of repeated measurements taken from the same subject over different time points. Although it is generally used in time series applications, forecasting can also be used in panel data due to its time dimension. However, there is limited number of studies in this area in the literature. In this thesis, forecasting is studied for panel data with binary response because of its increasing importance and increasing fundamental roles. A simulation study is held to compare the efficiency of different methods and to find the one that gives the optimal forecast values. In this simulation, 21 different methods, including naïve and complex ones, are used by the help of R software. It is concluded that transition models and random effects models with no lag of response can be chosen for getting the most accurate forecasts, especially for the first two years of forecasting.

Keywords: Forecasting, logistic regression models, moving and non-moving summary statistics, panel data.

ÖZ

İKİ SONUÇLU UZUNLAMASINA VERİ ÜZERİNE EN İYİ ÖNGÖRÜYÜ HANGİ YÖNTEM VERİR?: BENZETİM ÇALIŞMASI

ASLAN, Yasemin
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Özlem İLK

Eylül 2010, 68 sayfa

Uzunlamasına veri olarak da bilinen panel veri, aynı nesne üzerinden farklı zamanlarda alınan tekrarlı ölçümlerden oluşmaktadır. Öngörü yöntemi genellikle zaman serileri uygulamalarında kullanılsa da, panel verinin zaman boyutuna sahip olmasından dolayı panel veride de kullanılabilinmektedir. Fakat, literatürde bu alanda kısıtlı sayıda çalışma bulunmaktadır. Bu tezde literatürde giderek artan öneme ve artan temel role sahip olan iki sonuçlu panel veri için öngörü çalışılmıştır. Farklı yöntemlerin etkinliklerini karşılaştırmak ve öngörü için en iyi sonucu veren yöntemi bulmak için benzetim çalışması yapılmıştır. R programı kullanılarak gerçekleştirilen bu çalışmada, basit ve karmaşık yöntemlerin bulunduğu 21 farklı yöntem kullanılmıştır. Sonuç olarak içinde açıklanan değişkenin geçmişine dair bilgi bulundurmayan rastgele etkili modelin, özellikle öngörü yapılacak ilk iki yıl için en doğru öngörü değerlerini verdiği söylenebilir.

Anahtar Kelimeler: Öngörü, lojistik regresyon modelleri, hareketli ve hareketsiz özet istatistikleri, panel veri.

To Everyone Who Reads Here

ACKNOWLEDGEMENTS

First of all, I would like to give special thanks to my thesis supervisor, Assist. Prof. Dr. Özlem İlk for her professional support, guidance and encouragements which are invaluable for me during my research. Also, I give my special thanks to her to tolerate my working conditions and read my all drafts with her sensitivity and patience.

I would like to thank to Assist. Prof. Dr. Kasırga Yıldırak, Assist. Prof. Dr. Ceylan Yozgatlıgil, Assist. Prof. Dr. Zeynep Işıl Kalaylıoğlu and Assist. Prof. Dr. Burçak Berna Başbuğ Erkan for their valuable contribution to this thesis.

Furthermore, I would like to thank my mother Hanife Aslan and father Hüseyin Aslan for their sacrificing and endless love which can be replaced by nothing in the world. I also would like to thank my sisters, brother, little nephew and other relatives for their lovely support. They always believe and support me.

Finally, I would like to special thanks to all my friends for their continuous support, love and friendship, and for encouraging me to be an academician.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

CHAPTER 1

INTRODUCTION

Three types of data, which are cross-sectional, panel and time series, are commonly used in empirical analysis. As it is explained in Gujarati (2003), cross-sectional data are formed by the values of one or more variables for several sample units at the same time point while time series data are formed by the values of one or more variables over a period of time. Data on labor force participation rate for women in Turkey for a given year can be given as an example for cross-sectional data whereas data on exchange rate for several months or years can be given for time series. Panel data, also named longitudinal data, can be described as the union of these two types of data. They are composed of repeated measurements taken from the same subject over different time points (Diggle et al., 2002). Innovation expenditures of firms' in different sectors over a specified time period can be given as an example for panel data. The privilege of this data is the ability of measuring the change within firms over time. As it is emphasized in Fitzmaurice et al. (2004) describing the dynamics of change in terms of both time and units is the main aim of the panel data.

This multidimensional property ascribes panel data some advantages and disadvantages. Details of these advantages and disadvantages of longitudinal data can be found in Baltagi (2001), Gujarati (2003) and Hsiao (2003). According to these sources, it can be summarized that panel data increase the accuracy of the parameters by having "more degrees of freedom" and "more sample variability". Furthermore, this more informative data clarify the complicated behavior more properly such as by pooling the data and by "controlling the impact of omitted variables". On the other hand, much more care is needed when analyzing it because of its complexity such as having between and within individual variation and problems due to data collection.

Longitudinal data have an important and increasing role in many fields as it is explained in Chapter 2. This is also true for the longitudinal binary response data, which are used in this thesis. Longitudinal binary response data occur if the response value in longitudinal data takes only two values such as 0 or 1. The article by Sohn and Kim (2007) can be given as an

example for aforementioned case. They benefited from binary response panel data by suggesting a random effects logistic regression model to construct an accurate scorecard to decrease the default rate of funded small and medium enterprises (SMEs) for Korean government. As in this budget-planning example, longitudinal binary response data are helpful for the decisions. This gives an important role for this type of data.

In line with its importance, modeling longitudinal data has a wealth of literature (Fitzmaurice et al. (2009), Diggle et al. (2002), Frees (2004), Gujarati (2003), Yaffee (2003)). Nevertheless, there is a deficiency for discrete longitudinal data which needs a greater care when it is modeled (Fitzmaurice et al. (2009), Liang and Zeger (1986)). İlk (2008) clarifies the reason for this less development as the easiness of the computations in Gaussian models. The author explained that any new model or idea is usually firstly developed in Gaussian models in statistics because the calculations such as derivation of the likelihood are easier than the ones for Non-Gaussian models.

After modeling with the panel data, one can continue to his/her study by making forecasts. Forecasting is to make prediction about the future values after analyzing the past data available and the opinions by using conditional distributions of random variable(s). Forecasting can be seen as prediction because values to be forecasted are also random variables (Bosq and Blanke, 2007). As Plewis asserted in his article, predictions are the probability statements trying to know future by using the information in the past and present (Plewis, 2007).

The ultimate aim for forecasting is decision making and planning in the present (Bosq and Blanke, 2007). Therefore, the uncertainty about the future should be estimated carefully so that optimal decisions can be done (Troccoli et al., 2008). Let's say, one has some money and wants to make a reasonable investment. S/he can buy other currencies according to the forecast values of exchange rates or buy Treasury bond according to the forecast values of interest rates. S/he can make a profit if s/he can make accurate forecasts.

Forecasting is generally used in time series applications; however, it can also be used in longitudinal data since longitudinal data have time dimension (Frees (2004), Pepe et al. (1999)). Moreover, as it is expressed in Hyndman (2010), containing less time dimension with respect to time series data can be an advantage because changes in the data may prevent making accurate forecasts. As a matter of fact, forecast should also be used in panel data because the applications are as important as in time series. For example, modeling the development of a disease by using data on different people over time is important because as Pepe et al. (1999) presented, the early diagnosis of illness and/or development of disease can be done giving the opportunity of taking precautions before being ill if the signs of them are seen. Furthermore, analysis for the predictions of "chronic diseases such as arthritis,

nephritis, diabetes and chronic obstructive pulmonary diseases" can be given as other examples (Stiratelli et al., 1984). Besides these examples in the field of biostatistics, additional examples can be given in econometrics. It can be useful for a firm to analyze its budget and revenue for market entry and exit decisions (Haveman and Nonnemaker, 2000). In addition, it can be useful for budget-planning for governments like in the example of Wisconsin Lottery Sales (Frees, 2004).

Although there are studies making forecast comparisons in time series through different methods on various micro and macro data (e.g., Makridakis and Hibon (2000), Makridakis and Hibon (1979) and Makridakis et al. (1982)), there is no such study in panel data. This comparison is done for panel data with binary response in this thesis through a simulation study. With this simulation, the method that gives the optimal forecast values is found by the help of R program.

In this simulation, 5 continuous variables for 300 subjects on 8 different equally spaced time points are generated using multivariate normal distribution. Next, one of these variables is transformed to binary variable as it is shown in Chapter 3. With these variables, 21 different methods, including naïve and complex ones, are set by using the first 4 years of the data for the binary response variable. Some of these methods depend only on its own history whereas some of them depend only on other variables. Furthermore, some of them include both type of information. According to the type of variables that models include, generalized estimation equations (GEE) or restricted maximum likelihood estimation (REMLE) are used for the estimation of coefficients. The details for these models and estimation techniques can be found in Chapter 4. After modeling the data and measuring the efficiency of the model, forecast is done for the second half of the data using the same formulas which were used in modeling. Forecast results of these methods are compared with the actual values to test the accuracy of forecast by the help of the proportion of correct predictions (PCP), expected percent correctly predicted (ePCP), area under the receiver operating characteristic (AUROC) curve and deviance.

The rest of the chapters are introduced as follows. In Chapter 2, literature review for panel data and forecast is given with the examples of longitudinal binary response data analysis. In the following chapter, which is Chapter 3, data generation process is explained. In Chapter 4, details about the methods used for comparisons, the estimation techniques and the forecast accuracy measures that are used under these methods are given as subchapters in turn. The results of these comparisons can be found in Chapter 5, while conclusions take place in the final chapter, that is, Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1. Panel Data and Binary Response Models

The availability of panel data is increasing in both developed and developing countries. To illustrate, when it is looked at citation databases[1], there were 2,052 studies in 2009 having the topic “panel data” or “longitudinal data” while there were only 729 and 42 studies in the years 2000 and 1990, respectively. The first study was done in 1954 by Fulton. Interested readers are referred to Hsiao (2003) for the history of the panel data.

According to the results, the top area is economics having a ratio of 32%. This is followed by statistics and probability with a ratio of 16%. Other areas have ratios which are less than 10%. Besides these statistics, Fitzmaurice et al. (2004) presented that longitudinal data are used in many subject-matter analyses such as health sciences and epidemiology to analyze the properties and the development of diseases. Yaffee (2003), Song (2007) and Wawro (2002) added that social sciences such as sociology, psychology, educational research, political science and policy analysis, in particular program evaluation, are analyzed with the help of panel data. One can find many examples for panel data in Chapter 2 of Verbeke and Molenberghs (2000), Fitzmaurice et al. (2004), Song (2007) and Ilk (2008).

Panel data vary in the design of missing data, i.e. balanced and unbalanced longitudinal data. When observations can be found for all, every unit and every consecutive time period panel data are named as balanced while it is called unbalanced in the occurrence of some missing data such as in consecutive time points or in individuals (Dougherty, 2007).

Panel data also vary in the types of outcomes of interest. There may be one or even more than one continuous, categorical or binary response variable(s). For example, investigating

[1] Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Conference Proceedings Citation Index-Science (CPCI-S) and Conference Proceedings Citation Index-Social Science & Humanities (CPCI-SSH).

the research and development expenditures of firms in OECD countries between the years 2002 and 2007 is an example for continuous longitudinal data case. Furthermore, the investigation of whether these firms make organizational innovation or not is an example for binary case, while whether they make organizational innovation or process innovation at the same model is an example for multivariate binary case.

Binary response models are the models whose dependent variable is a binary random variable having a Bernoulli probability distribution taking only zero or one (Winkelmann and Boes, 2009). Because this type of data takes places very often in real-life examples, they are important (Winkelmann and Boes, 2009). Actually, many decisions that people make can be thought as discrete choices, such as going to school or not, drinking alcohol or not, buying a house or not, giving money to beggars or not, etc. Moreover, one can transform the other types of responses like multinomial, ordered and continuous to binary response models, although this may cause some loss of information. To see examples, refer to Winkelmann and Boes (2009). Therefore, modeling binary response data makes people's life easier by giving the ability of understanding the causes of their decisions. As it is expressed in Horowitz and Savin (2001), the logit and probit models are the most common approaches for analyzing this type of data. Probit model is used if the cumulative density function (CDF) of the response variable is the cumulative normal distribution function while logit model is used if CDF of the response variable is the cumulative logistic distribution function. Both methods generally give similar results whereas they differ in the extreme tails because of the shape of the distributions. In terms of the estimation of the parameters in both logit and probit models, maximum likelihood can be used as it is presented in Jackman (2007) and Winkelmann and Boes (2009).

2.2. Forecasting

There are two types of forecasting which are called deterministic and probabilistic (Troccoli et al., 2008). They declared that deterministic forecasts have no uncertainty because it chooses the most probable situation for the future, while in probabilistic forecasts, the uncertainty is tried to be estimated by calculating the probability of occurrence of future outcomes. This gives it more reliability and gives the user the chance of making optimal decisions. Probabilistic forecasts can be discussed under two approaches, namely frequentist and Bayesian. In frequentist approach, relative frequency in a large number of trials is considered whereas a prior distribution is used in Bayesian approach. One can make point and interval forecasts in both approaches.

Methods for forecasting can be classified into two groups, namely qualitative and quantitative. While opinions of experts are being used for qualitative forecasting such as in

Delphi method, for quantitative ones, historical data are needed to be able to understand the structure of the data (Dayananda et al., 2002). Moreover, one should choose the method according to data availability and the predictability of the quantity to be forecasted (Hyndman, 2010). When the data for the past values are available and it is thought that the future values depend on the patterns of these values, it is reasonable to apply quantitative forecasting (Hyndman, 2010). Since the past values of the data are available here, quantitative methods are used for forecasting.

Some of the methods for quantitative data modeling for forecasts are moving averages, smoothing and mathematical and statistical models (Sparling). In moving averages method, forecast is done by summing values over a given time period (*t*) and then dividing the result by *t*. Simple moving average, weighted moving average and exponential moving average can be given as examples for this method. Regardless of the response type, it can be applied easily in both time series and panel data. For example, Baadsgaard et al. (2004) used the average of previous disease observations to make clinical forecasts in a herd. Second method, which is called smoothing method, can be described as a kind of weighted moving average allowing the trends. Single exponential smoothing and double exponential smoothing can be given as examples for this method. Moreover, according to Diggle et al. (2002), smooth curves, which have been fitted by kernel estimation, smoothing splines or lowess, estimate the average response value using time as an explanatory variable. They explained that all three methods give quantitatively similar results. Interested readers are referred to Hastie and Tibshirani (1990) and original references therein. Although smooth curves are used in continuous responses commonly, they should not be applied in binary responses because they give only 0 or only 1 for all the future values. The last method, mathematical and statistical models, is linear or non-linear models fitted to data usually by regression methods. Trend lines, log-linear models, Fourier series, simple regression, multiple regression and growth curves can be given as examples for this method. These models are used when the underlying factors which might influence the variable that is being forecasted are known. They can be used for both time series and panel data, and all types of responses. Only the type of the statistical model changes according to available data.

Through these three methods, applicable ones for binary data such as moving approach and statistical models are used for forecasting in this thesis. As it is explained in Chapter 4, some of these methods depend only on response history such as in simple moving average method whereas some of them depend only on independent variables such as in marginal models. Furthermore, some of them depend on both type of information such as in marginalized transition models. As a non-parametric approach, simple moving approach is one of the methods that is used for forecasting. This approach is used in this thesis with three different estimators which are mean, median and mode. Furthermore, as parametric approach autoregressive models, marginal models, transition models, marginalized transition

models and random effects models are used in this thesis. One can choose one of these methods according to the conceptual difference underlying them (Diggle et al. (2002), Gardiner et al. (2009), Fitzmaurice et al. (2004) and Carrière and Bouyer (2002)). According to Gardiner et al. (2009), the selection of models depends not only on statistical diagnosis but also on the goal of the analysis. For example, if one is interested in the average effect of covariates on the response in a population, then marginal models are the choice, which is why marginal models are also called population-average models. On the other hand, if one is interested in subject-specific effects of variables then the random effects models are more appropriate. However, when there is a reason to suspect that unobserved heterogeneity is correlated with explanatory variables then the fixed effect models are more appropriate because the random effects model would yield inconsistent estimates in these cases.

One can follow the steps given in Hyndman (2010) to be able to make a well-established, more efficient quantitative forecast. According to him, these steps are as follows. Firstly one should define the problem carefully to know where the results are going to be used, and talk to everyone who deals with data collection, databases and forecast users. Secondly, s/he should gather statistical data and expert's opinions. Since forecasting regression models are fitted according to current database, it should be collected carefully. Thirdly, making exploratory data analysis is needed by graphing data to see the features of the data like trend, outlier, seasonality or any other pattern or the relationships between the variables. Then, the next step is choosing the appropriate model and fitting it. According to available data source and the relationship between the independent and dependent variables, several models are set to get forecast values. Moreover, as it is written in Tian (2007), for reliable forecasting, it is important that regression method should be fit adequately. The final step is evaluating and comparing these different fitted models. To be able to do these comparisons properly, real data for the forecast values should be available. One can see Hyndman (2010) to get more information about the forecast accuracy methods.

Baltagi (2008) stated that although there is a wealth of literature for forecasting in time series, it is not the case for panel data until recent years. For a brief survey of forecasting with panel data, one can see his study. As well as giving examples, he suggested that more study is needed for this subject. To illustrate some of these examples, Baillie and Baltagi (1999) used four different methods for the error component regression model and mean square error (MSE) for the efficiency of these methods. Moreover, Baltagi (2008) showed the best linear unbiased predictor (BLUP) for the future values of response, and made a comparison between heterogeneous and homogeneous estimators' efficiency. Furthermore, Baltagi and Li (2006) discussed forecasting for spatially auto correlated panel data, and concluded that fixed effect and random effect estimators are the best estimators in terms of root mean square error forecast performance by considering the heterogeneity across states when modeling the liquor demand. However, when these examples are studied, it is seen

that these models are generally for continuous variables having at least 20 time points. Since the time length and response type for panel data in this thesis are different from the one in this econometrics literature, we could not benefit from these studies.

After eliminating these medium time period examples, Rosenberg et al. (2008) can be given as an example in the limited literature. In this study, they did forecasting by using panel data which include 13 years. Pooled cross sectional, fixed effect and mixed effect models are the methods for model fitting. They fitted these models using the first 11 years and did forecasts for the next two years. Next, they calculated the mean absolute error (MAE) and the mean absolute percentage error (MAPE) statistics for the accuracy of forecasts. Since the outcome is continuous here, techniques in these methods are different from what we should use. However, there are some similarities between this thesis and this article. For example, mixed effect models and forecast accuracy techniques are also used in this thesis but they are adapted for binary response.

For the binary response versions, one can see Baadsgaard et al. (2004). In this study, status of the herd health was predicted using the data for 15 pig herds for 12 months. They used two different methods for forecasting namely a naïve approach and a binomial Bayesian state space models approach. Specifically, the first method calculates the averages of previous disease observations as forecast values while the second one uses the time lag variable as random component. Although, it is proposed in Troccoli et al. (2008) that Bayesian approach can be more reliable than the results of frequentist approach, these two methods gave similar results in the study of Baadsgaard et al. (2004). Therefore, we decided to use this naïve method as one of our methods, and not to use Bayesian approach.

Another example of forecasting which used binary panel data is the study done by Horrocks and van Den Heuvely (2009). They predicted the probability of being pregnant for each woman using various methods. These methods were multiple t-tests, mixed models, discriminant analysis and two-stage models for both continuous and binary response. They also used a joint model including linear mixed effects model and generalized linear model, and constructed confidence intervals for the coefficients of these models. As in this study, we used both individual effect models, such as random effects model and population effect models. In addition, we calculated confidence intervals for the accuracy methods, but not for the coefficients.

CHAPTER 3

DATA GENERATION PROCESS

In this thesis, a longitudinal data is generated having a dimension of 5x8x300 by the help of R program. These numbers belong to the number of variables, the number of time points and the number of subjects in turn. That is, there are 5 variables for 300 subjects on 8 different equally spaced time points in the dataset. We generated balanced data because of simplicity and applicability of all models to data. In terms of the variables, 4 of them are independent and continuous and the other one is dependent and transformed as binary. Letting Y_{it} be the response for the *i* th subject at time *t* and X_{itk} be the *k* th covariate, the data scheme can be found in Table 1.

Table 1. Scheme of the generated data

Time Horizon (t) / Subject (i)	1	2	…	8
1	$Y_{1,1}$ $X_{1,1,1}$, $X_{1,1,2}$, $X_{1,1,3}$, $X_{1,1,4}$	$Y_{1,2}$, $X_{1,2,1}$, $X_{1,2,2}$, $X_{1,2,3}$, $X_{1,2,4}$	…	$Y_{1,8}$, $X_{1,8,1}$, $X_{1,8,2}$, $X_{1,8,3}$, $X_{1,8,4}$
2	$Y_{2,1}$ $X_{2,1,1}$, $X_{2,1,2}$, $X_{2,1,3}$, $X_{2,1,4}$	$Y_{2,2}$, $X_{2,2,1}$, $X_{2,2,2}$, $X_{2,2,3}$, $X_{2,2,4}$	…	$Y_{2,8}$, $X_{2,8,1}$, $X_{2,8,2}$, $X_{2,8,3}$, $X_{2,8,4}$
3	$Y_{3,1}$ $X_{3,1,1}$, $X_{3,1,2}$, $X_{3,1,3}$, $X_{3,1,4}$	$Y_{3,2}$ $X_{3,2,1}$, $X_{3,2,2}$, $X_{3,2,3}$, $X_{3,2,4}$	…	$Y_{3,8}$ $X_{3,8,1}$, $X_{3,8,2}$, $X_{3,8,3}$, $X_{3,8,4}$
⋮	⋮	⋮		⋮
300	$Y_{300,1}$ $X_{300,1,1}$, $X_{300,1,2}$, $X_{300,1,3}$, $X_{300,1,4}$	$Y_{300,2}$ $X_{300,2,1}$, $X_{300,2,2}$, $X_{300,2,3}$, $X_{300,2,4}$	…	$Y_{300,8}$ $X_{300,8,1}$, $X_{300,8,2}$, $X_{300,8,3}$, $X_{300,8,4}$

Multivariate normal distribution is used with “rmvnorm” code in R to generate such data. Other codes for data generation process can be found in Appendix A. The sample size and covariance matrix are needed to be able to use “rmvnorm” code. Therefore, rules for the form of covariance/correlation matrix and sample size which are given in the literature are taken into consideration to be able to generate a reasonable longitudinal data set.

There are different rules in the literature for the determination of sample size that is the value of N. One of these rules is given in Neter et al. (1996, pp. 330). They expressed that the number of observations should be at least six to ten times higher than the number of variables in the dataset. Because the number of variables that is used in our simulation is 5, 300 is selected as the number of cases. When the conditions for the data in many developing countries and the simplicity of this rule are taken into consideration, this is a reasonable number. In addition, interested readers can also find a bit complicated formulas in Fitzmaurice et al. (2004) and Diggle et al. (2002) for binary longitudinal data. The non-linear link function such as logit and the dependence of the variance on the mean are the reasons for this complication when the response variable is binary.

The formula for the covariance matrix is given in [3.1] (Johnson and Wichern (1998)). We use this formula when constructing the covariance matrix of 5 variables at a single time point.

$$\sum = V^{\frac{1}{2}} * \rho * V^{\frac{1}{2}} \qquad [3.1]$$

$$\text{where } V^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\sigma_{22}} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\sigma_{33}} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{\sigma_{44}} & 0 \\ 0 & 0 & 0 & 0 & \sqrt{\sigma_{55}} \end{bmatrix}$$

$$\text{and } \rho = \begin{bmatrix} 1 & \rho_{Y,X_1} & \rho_{Y,X_2} & \rho_{Y,X_3} & \rho_{Y,X_4} \\ \rho_{X_1,Y} & 1 & \rho_{X_1,X_2} & \rho_{X_1,X_3} & \rho_{X_1,X_4} \\ \rho_{X_2,Y} & \rho_{X_2,X_1} & 1 & \rho_{X_2,X_3} & \rho_{X_2,X_4} \\ \rho_{X_3,Y} & \rho_{X_3,X_1} & \rho_{X_3,X_2} & 1 & \rho_{X_3,X_4} \\ \rho_{X_4,Y} & \rho_{X_4,X_1} & \rho_{X_4,X_2} & \rho_{X_4,X_3} & 1 \end{bmatrix} \text{ where } \rho_{i,k} = \frac{\sigma_{i,k}}{\sqrt{\sigma_{i,i}} * \sqrt{\sigma_{k,k}}}$$

In the matrix of $V^{\frac{1}{2}}$, σ_{11} is the variance for the response variable and is taken as 0.25 to simulate our data. Larger values such as 25 were also tried but there was no big difference at the mean values of the response. For the independent variables, 5, 5, 50 and 500 are used for the variances of X_1, X_2, X_3 and X_4, respectively. That is, X_1 and X_2 have moderate variances, while X_3 and X_4 have large and huge ones, respectively. In addition, 10 was tried

for the variance of X_2 but no changes has occured in the results. Different values, including small and large ones, are given for these variances to be able to create independent variables having different properties. Thus, the effects of these properties in different models can also be analyzed. Moreover, we assume small variances for two of them because all methods have also been tried with only X's having small variances to see whether the small variances can change the difficulties in estimations or not. Furthermore, the same variances are assumed for each time point.

In the matrix of ρ, ρ_{Y,X_1}, ρ_{Y,X_2}, ρ_{Y,X_3} and ρ_{Y,X_4} are taken high to get high correlations between independent and dependent variables whereas other correlations are taken as small to prevent multicollinearity between independent variables. Furthermore, off-diagonal elements of the correlation matrix that correspond to correlations between different time points are taken different from zero to create dependence through repeated measurements. Besides these, Fitzmaurice et al. (2009) asserted that correlation usually decreases with increasing time lag in panel data. Therefore, a pattern such as in the first two parts of Table 2 is used when constructing the whole matrix. In addition to these properties, Fitzmaurice et al. (2004) added other features of the correlation through the repeated measures as being positive in many longitudinal data examples. They also proclaimed that they may approach zero in case there are many years between the measures whereas they may approach one in case time is very close among the repeated measures. Thus, the values are taken as in Table 2.

Table 2. Structure and the values of correlation matrix that are assumed for data generation (with continous response)

<table>
<tr>
<td rowspan="8">Cor ($Y_{t=1}$,$Y_{t=1}$)=1.000
Cor ($Y_{t=1}$,$Y_{t=2}$)=0.985
Cor ($Y_{t=1}$,$Y_{t=3}$)=0.886
Cor ($Y_{t=1}$,$Y_{t=4}$)=0.790
Cor ($Y_{t=1}$,$Y_{t=5}$)=0.679
Cor ($Y_{t=1}$,$Y_{t=6}$)=0.557
Cor ($Y_{t=1}$,$Y_{t=7}$)=0.479
Cor ($Y_{t=1}$,$Y_{t=8}$)=0.353</td>
<td rowspan="8">Cor ($Y_{t=1}$,$X_{t=1}$)=0.985
Cor ($Y_{t=1}$,$X_{t=2}$)=0.886
Cor ($Y_{t=1}$,$X_{t=3}$)=0.790
Cor ($Y_{t=1}$,$X_{t=4}$)=0.679
Cor ($Y_{t=1}$,$X_{t=5}$)=0.557
Cor ($Y_{t=1}$,$X_{t=6}$)=0.479
Cor ($Y_{t=1}$,$X_{t=7}$)=0.353
Cor ($Y_{t=1}$,$X_{t=8}$)=0.275</td>
<td rowspan="8">Cor ($X_{t=1}$,$X_{t=1}$)=1.000
Cor ($X_{t=1}$,$X_{t=2}$)=0.849
Cor ($X_{t=1}$,$X_{t=3}$)=0.753
Cor ($X_{t=1}$,$X_{t=4}$)=0.671
Cor ($X_{t=1}$,$X_{t=5}$)=0.553
Cor ($X_{t=1}$,$X_{t=6}$)=0.465
Cor ($X_{t=1}$,$X_{t=7}$)=0.370
Cor ($X_{t=1}$,$X_{t=8}$)=0.275</td>
<td>Cor ($Y_{t=2}$,$X_{t=1}$)=0.885</td>
</tr>
<tr>
<td rowspan="7">Cor ($X_{t=1,k}$,$X_{t=1,k'}$)=0.25</td>
</tr>
</table>

Table 2 is composed of the untransformed values for the correlation matrix. Averages of 4 covariates are given for the correlations including independent variables. These values are given after trials to prevent multicollinearity and to create an autocorrelation. Here, values are actually chosen higher than we would like to observe because it is seen that the values become smaller after transformation is applied in these trials.

Three real data sets were studied and their correlation structures had been identified to assume reasonable values before these values are assigned. However, since each includes unexpected pattern in the correlation structure such as multicollinearity, none of them was taken solely as the main structure. Mean values of these 3 data sets were also calculated for the correlation structure to swap these problems, but the feature of diminishing correlation with increasing time point could not be captured. Therefore, real data was not used in this thesis.

After having five continuous variables which are generated by using singular value decomposition method, one of these generated variables was transformed to binary variable using the transformation given below:

$$P(y^* = 1) = \frac{\exp(y)}{(1 + \exp(y))}$$

Instead of this transformation, response values could be generated by using a model such as logistic regression. Nonetheless, to be able to make an unbiased comparison between the methods, the transformation is preferred. By this way, the generated response values are model independent.

After the application of this transformation, 0 is given to the response variable if the above probability is smaller than 0.5, whereas 1 is given otherwise. Instead of this cut-off value, mean, mode or a random variate from uniform distribution could be used for the threshold value. However, as it was suggested in Terza (2006), 0.5 is found to be the optimal value in regression for predicting binary response.

Having applied this rule, all calculations in this thesis are done 10,000 times and the average of them is taken when the final decision is claimed. This is done for preventing the bias of the results. The average values for the correlation matrix of 10,000 simulated data are as given in Table 3 while full correlation matrix of it can be found in Appendix B.

Table 3. Structure and the average values of correlation matrix after data generation (with binary response)

Cor ($Y_{t=1}$,$Y_{t=1}$)=1.000	Cor ($Y_{t=1}$,$X_{t=1}$)=0.414	Cor ($X_{t=1}$,$X_{t=1}$)=1.000	Cor ($Y_{t=2}$,$X_{t=1}$)=0.356
Cor ($Y_{t=1}$,$Y_{t=2}$)=0.712	Cor ($Y_{t=1}$,$X_{t=2}$)=0.372	Cor ($X_{t=1}$,$X_{t=2}$)=0.836	
Cor ($Y_{t=1}$,$Y_{t=3}$)=0.625	Cor ($Y_{t=1}$,$X_{t=3}$)=0.330	Cor ($X_{t=1}$,$X_{t=3}$)=0.736	
Cor ($Y_{t=1}$,$Y_{t=4}$)=0.501	Cor ($Y_{t=1}$,$X_{t=4}$)=0.282	Cor ($X_{t=1}$,$X_{t=4}$)=0.652	
Cor ($Y_{t=1}$,$Y_{t=5}$)=0.393	Cor ($Y_{t=1}$,$X_{t=5}$)=0.231	Cor ($X_{t=1}$,$X_{t=5}$)=0.534	Cor ($X_{t=1,k}$,$X_{t=1,k'}$)=0.239
Cor ($Y_{t=1}$,$Y_{t=6}$)=0.344	Cor ($Y_{t=1}$,$X_{t=6}$)=0.197	Cor ($X_{t=1}$,$X_{t=6}$)=0.447	
Cor ($Y_{t=1}$,$Y_{t=7}$)=0.234	Cor ($Y_{t=1}$ $X_{t=7}$)=0.146	Cor ($X_{t=1}$,$X_{t=7}$)=0.354	
Cor ($Y_{t=1}$,$Y_{t=8}$)=0.203	Cor ($Y_{t=1}$,$X_{t=8}$)=0.111	Cor ($X_{t=1}$,$X_{t=8}$)=0.263	

Table 3 is composed of transformed response values and the average values of 10,000 trials. Moreover, averages of 4 covariates are given again for the correlation having covariates. Here, Cramer's V is used for the correlation between binary values while Spearman's correlation is used for the other conditions. The structure of the generated data has a parallelism with the rules given above. The only problem is for the correlation between dependent and independent variables. The values for Cor ($Y_{t=1}$, $X_{t=1}$) and Cor ($Y_{t=2}$, $X_{t=1}$) are smaller than they were expected. Although very high and same values are given for both Cor ($Y_{t=1}$, $X_{t=1}$) and Cor ($Y_{t=1}$, $Y_{t=2}$) as it can be seen in Table 2, the observed correlation between $Y_{t=1}$ and $X_{t=1}$ become smaller than the correlation between $Y_{t=1}$ and $Y_{t=2}$. Various methods have been tried to increase these two problematic values. In these trials, it was seen that the value for the correlation diminishes at around 0.2 level when the transformation given above has been applied. In addition, no solution has been found without causing any multicollinearity between independent variables and/or less autocorrelation within the independent variable. For instance, in one of the trials where independent variables are generated from a binary variable and then transformed to continuous, Cor ($Y_{t=1}$, $X_{t=1}$) and Cor ($Y_{t=1}$, $Y_{t=2}$) were high having the decreasing trend with the increasing time point. However, multicollinearity between independent variables was found to be medium or high, while autocorrelation within the independent variable was small or medium. To continue, when continuous independent variables are generated from different binary variables, the problem of multicollinearity disappeared. Moreover, high autocorrelation is supplied for both dependent and independent variables. Nevertheless, the intended correlation between dependent and independent variables could not be constructed. Although all of these methods were tried, there was no wanted solution.

CHAPTER 4

METHODOLOGY

Both parametric and non-parametric approaches are used in this thesis to compare the forecast values of dependent and independent variables. Some of these methods depend only on response history whereas some of them depend only on independent variables. Furthermore, some of them depend on both type of information. Here, logistic regression is used when modeling the response since Y_i comes from Bernoulli distribution. The detailed information can be found below.

4.1. Methods

As a non-parametric approach, simple moving approach is one of the methods that are used for forecasting. In this thesis, this approach is used with three different estimators, which are mean, median and mode. In addition, marginal models, transition models and subject-specific models are the most common parametric approaches for longitudinal data in the literature (Diggle et al., 2002). Therefore, autoregressive models, marginal models, transition models, marginalized transition models and random effects models are used as parametric approaches in this thesis with a goal of comparing the results. Information on these parametric and non-parametric models is provided in Table 4.

4.1.1. Naïve Methods

Simple, weighted and exponential averages are the most common types for naïve methods as it is written in Roberts (2003). The difference between these is that while more weight is given to the latest observations in weighted and exponential averages, simple average gives equal weight to all observations. In this thesis, only simple moving mean, simple moving median and simple moving mode and non-moving forms of these methods are used as naïve methods because the other ones do not give meaningful forecast values for binary cases. They give only 1 for all future time points if they give 1 for the first forecasted time point or 0

otherwise. These naïve methods calculate the future values for individuals in an easy manner. The formulas where the past values of each individual are used can be found below.

Simple Moving Mean: Y_{it}= mean ($Y_{it-1,}$ $Y_{it-2,}$ $Y_{it-3,}$ Y_{it-4}) for t=5, 6, 7, 8
Simple Non-Moving Mean: Y_{it} = mean ($Y_{i1,}$ $Y_{i2,}$ …,Y_{it-1}) for t=5, 6, 7, 8

Simple Moving Median: Y_{it}= median ($Y_{it-1,}$ $Y_{it-2,}$ $Y_{it-3,}$ Y_{it-4}) for t=5, 6, 7, 8
Simple Non-Moving Median: Y_{it} = median ($Y_{i1,}$ $Y_{i2,}$ …,Y_{it-1}) for t=5, 6, 7, 8

Simple Moving Mode: Y_{it}= mode ($Y_{it-1,}$ $Y_{it-2,}$ $Y_{it-3,}$ Y_{it-4}) for t=5, 6, 7, 8
Simple Non-Moving Mode: Y_{it} = mode ($Y_{i1,}$ $Y_{i2,}$ …,Y_{it-1}) for t=5, 6, 7, 8

Specifically, as it was given in Roberts (2003), the simple or arithmetic average is constructed by summing values over a given time period (*t*) and then dividing the result by *t*. Moreover, the simple moving median and the simple moving mode are constructed by taking the middle value and the most frequent value over the specified time period *t*, respectively. Besides these moving ones, non-moving methods are constructed by using all past data for average, median and mode values as forecasts.

Within these naïve methods, moving averages are the most common method in the literature. As it was expressed in Roberts (2003), although they are simple and easy for the users, they give consistent estimates when defining trend. In the study, which was done by Roberts (2003), moving averages gave better results than complex methods such as multiple regression. Moreover, they also provided better results than weighted or exponential averages in financial studies (Roberts, 2003).

Table 4. Methods that are used for comparing forecast accuracy of binary response variable

No	Method	Properties
1	Simple median	Moving width; cut-off value is taken as 0.5
2		Non-moving width; cut-off value is taken as 0.5
3	Simple mode	Moving width; cut-off value is taken as 0.5
4		Non-moving width; cut-off value is taken as 0.5
5	Simple average	Moving width; cut-off value is taken as 0.5
6		Non-moving width; cut-off value is taken as 0.5
7		Moving width; cut-off value is taken as a random variable distributed uniformly
8		Non-moving width; cut-off value is taken as a random variable distributed uniformly
9		Moving width; cut-off value is taken as the mean value of past values
10		Non-moving width; cut-off value is taken as the mean value of past values
11	AR Models	AR (1) model; cut-off value is taken as 0.5, no intercept
12		AR (2) model; cut-off value is taken as 0.5
13		AR (3) model; cut-off value is taken as 0.5
14	Marginal Models	cut-off value is taken as 0.5
15	Transition Models	TM (1); cut-off value is taken as 0.5
16		TM (2); cut-off value is taken as 0.5
17		TM (3); cut-off value is taken as 0.5
18	Marginalized Transition Models	MTM (1); cut-off value is taken as 0.5
19		MTM (2); cut-off value is taken as 0.5
20	Random effects models	Random intercept model having no lag of response; cut-off value is taken as 0.5
21		Random intercept model having lag of response; cut-off value is taken as 0.5

In this table, the cut-off value is taken only as 0.5 in all methods except simple average case. That is because simple average method was also used as the controlling method. Since it was seen that all three cut-off values resulted in nearly the same forecasting performance and 0.5 is the best one according to Terza (2006), other two cut-off values were not used for the other methods.

4.1.2. Autoregressive (AR) Models

Autoregressive (AR) models and the corresponding Yule Walker Equations are the source of one of the simplest, most transparent and widely used forecasting systems (Pourahmadi, 2001). In this method, response value depends only on its past values. In other words, this method takes the advantage of autocorrelations. These models are titled according to the lag of response in the model. For example, if the response depends only on its one period apart observations, then the model is called AR(1). If there are p lags of response in the model, then it is called AR(p) model. More generally, AR(p) model can be formulized as in [4.1] for a binary response. For simplicity of notation, we suppress the index for subject, that is index *i*:

$$logit\ P(Y_t = 1) = \alpha_0 + \sum_{j=1}^{p} \alpha_j * Y_{t-j} \qquad [4.1]$$

where α_0 is the constant, $\alpha_j's$ are the parameters of the model and Y_{t-j} is the j^{th} lag of the response.

In this thesis, AR(1), AR(2) and AR(3) methods are set because there are only 4 time points. Next, forecast values are calculated by using the same formula as in the other parametric methods because the most commonly used method for point forecasting is to replace unknown parameters with suitable estimators coming from past values (Pourahmadi, 2001). Model fitting and forecasting are done using the formulas given below.

For simplicity, AR (1) method is given as an example. Firstly, $\text{logit}\,\text{P}(Y_t = 1) = \alpha_0 + \alpha_1\, Y_{t-1}$ is used for the estimation of α_0 and α_1 with using data at times t=1, 2, 3, 4. Having gotten the estimates for the parameters, the logit probability of response being 1 is calculated for the forecast value at time period 5. The model $\text{logit}\,\text{P}(\widehat{Y_{t=5}} = 1) = \widehat{\alpha_0} + \widehat{\alpha_1}\, Y_{t=4}$ is used for this purpose. Next, the probability of being 1 for the response at 5th time is computed using the formula provided in [4.2]. Finally, the predicted Y_5 is taken as 1 if the probability is equal or greater than 0.5 whereas it is taken as 0 for other condition. Same methodology is used for the remaining time points, i.e. t= 6, 7, 8.

$$\text{P}(\widehat{Y_{t=5}} = 1) = \frac{\exp\,(\widehat{\alpha_0} + \widehat{\alpha_1}\, Y_{t=4}\,)}{1 + \exp\,(\widehat{\alpha_0} + \widehat{\alpha_1}\, Y_{t=4}\,)} \qquad [4.2]$$

As well as this method which includes the same estimated coefficients for all time points i.e. t=5, 6, 7, 8, we tried forecasting with varying coefficients. That is, only t=1, 2, 3, 4 were used for estimating α_0 and α_1 here, whereas α_0 and α_1 were also calculated for each time point, i.e. t=5, 6, 7, 8, separately adding the data being forecasted. However, since there is not much difference between the outputs of two different approaches, we did not use the latter one in

10,000 trials. Moreover, since there is not much difference between the results of the moving and non-moving approaches in 1 trial and 10.000 trials, there is no loss for not using the varying coefficients method.

4.1.3. Marginal Models

These methods are suitable when the population average inferences are needed (Diggle et al., 2002). In this method, response variable depends only on explanatory variables, that is marginal expectation of response is modeled as a function of independent variables (Diggle et al., 2002). In addition, Fitzmaurice et al. (2004) defined these methods as the methods in which there are neither random effects nor previous responses. Interested readers can find examples for this method in Chapter 8 of Diggle et al. (2002). They specified three assumptions for this method. These assumptions are given below for logistic marginal methods. For simplicity of notation, we suppress the index for subject, that is index *i*.

- $\text{logit}(\mu_t) = log\,\frac{\mu_t}{1-\mu_t} = log\,\frac{P(Y_t=1)}{P(Y_t=0)} = \alpha_0 + \alpha_1 x_t$ where μ_t is the expectation of response
- $\text{Var}(Y_t)=\mu_t * (1-\mu_t)$
- $\text{Cor}(Y_t, Y_{t'})=\varphi$ which means that the correlation is constant for all individuals through all time points

In this thesis, marginal model is fitted by the formula $\text{logit}\,\text{P}(Y_t = 1) = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + \alpha_3 X_{t3} + \alpha_4 X_{t4}$ using the first half of the data, i.e, time points t=1, 2, 3, 4. After estimating the parameters, forecast values are calculated using the formulas given below that are [4.3] and [4.4].

$$\text{P}(\widehat{Y_t = 1}) = \frac{\exp\,(\widehat{\alpha_0} + \widehat{\alpha_1} X_{t1} + \widehat{\alpha_2} X_{t2} + \widehat{\alpha_3} X_{t3} + \widehat{\alpha_4} X_{t4})}{1 + \exp\,(\widehat{\alpha_0} + \widehat{\alpha_1} X_{t1} + \widehat{\alpha_2} X_{t2} + \widehat{\alpha_3} X_{t3} + \widehat{\alpha_4} X_{t4})} \quad for\ t = 5, 6, 7, 8 \qquad [4.3]$$

$$\widehat{Y_t} = 1 \text{ if } \text{P}(\widehat{Y_t = 1}) \geq 0.5 \text{ for t= 5, 6, 7, 8} \qquad [4.4]$$

$$\widehat{Y_t} = 0 \text{ if } \text{P}(\widehat{Y_t = 1}) < 0.5 \text{ for t= 5, 6, 7, 8}$$

Since this method includes the forecast values of independent variables, different naïve and complex methods are also used for continuous independent variables to find the best method. When calculating the forecast values of independent variables, 4 different methods are used. These methods are simple mean average, AR (1), AR (2), and random effects model. Simple mean average is chosen as a naive method while random effects models is chosen as a complex model. Moreover, AR methods are used because of the high

correlation between X_t and X_{t-1}. In this model, lag is taken as 1 or 2 because there are only 4 time points in the dataset.

4.1.4. Transition Models

In transition models (TM), the conditional distribution of the response variable is fit with independent variables and the past values of the response (Diggle et al., 2002). They showed that transition matrix changes through individuals because the probabilities of transition are dependent on independent variables. That is, any individual is evaluated according to its own past values as well as the whole covariates. Moreover, they added that interactions may also be included for predicting. However, for any 2 subjects with same past values and covariates, the estimation and prediction will be the same. These models are also titled according to the lag of response in the model as in AR method. Specifically, formula for TM is provided in [4.5]. For simplicity of notation, we suppress the index for subject, that is index *i*.

$$\begin{aligned} &\text{logit P}(Y_t = 1|Y_{t-1}, Y_{t-2}, \dots, Y_1, X_{t1}, \dots\dots, X_{tk}) \\ &= \alpha_0 + \alpha_1 X_{t1} + \cdots + \alpha_k X_{tk} + \alpha_{k+1} Y_{t-1} + \cdots + \alpha_{k+p} Y_{t-p} \end{aligned} \qquad [4.5]$$

When the formula supplied above is examined, it can be interpreted as that transition models are the marginal models having an exponential autocorrelation function (Diggle et al., 2002). It can be thought that the lag of response enters the model because of this autocorrelation. This model requires balanced data. Furthermore, Diggle et al. (2002) provided the examples for transitional models for binary and count data as Korn and Whittemore (1979), Ware et al. (1988), Wong (1986), Zeger and Qaqish (1988) and Kaufmann (1987).

Here, we set TM(1), TM(2) and TM(3) models for this method. For simplicity, TM(1) is explained as an example. Firstly, TM(1) is fitted using the formula $\text{logit P}(Y_t = 1) = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + \alpha_3 X_{t3} + \alpha_4 X_{t4} + \alpha_5 Y_{t-1}$ when t=2, 3, 4. After the estimation of coefficients, forecast values for binary response are calculated according to the formulas supplied in [4.6] and [4.7] after forecasting independent variables as in marginal models.

$$\widehat{P(Y_t = 1)} = \frac{\exp\left(\widehat{\alpha_0} + \widehat{\alpha_1} X_{t1} + \widehat{\alpha_2} X_{t2} + \widehat{\alpha_3} X_{t3} + \widehat{\alpha_4} X_{t4} + \widehat{\alpha_5} Y_{t-1}\right)}{1 + \exp\left(\widehat{\alpha_0} + \widehat{\alpha_1} X_{t1} + \widehat{\alpha_2} X_{t2} + \widehat{\alpha_3} X_{t3} + \widehat{\alpha_4} X_{t4} + \widehat{\alpha_5} Y_{t-1}\right)} \; for\; t = 5, 6, 7, 8 \qquad [4.6]$$

$$\widehat{Y_t} = 1 \text{ if } \widehat{\text{P}(Y_t = 1)} \geq 0.5 \text{ for t= 5, 6, 7, 8} \qquad [4.7]$$

$$\widehat{Y_t} = 0 \text{ if } \widehat{\text{P}(Y_t = 1)} < 0.5 \text{ for t= 5, 6, 7, 8}$$

4.1.5. Marginalized Transition Model

Marginalized transition models (MTM) were developed by Heagerty (2002) by extending the Azzalini's (1994) first-order Markov models (Comstock and Heagerty). This model estimates the average response for binary data using likelihood – based approach. Therefore, as it is expressed in Comstock and Heagerty, MTM is preferable when the aim is to analyze and make inferences about marginal mean regression. In addition, this model includes information on both the lag of the response directly and explanatory variables indirectly allowing the correlation structure through longitudinal data. The formula where the index for subject, that is index *i*, is suppressed for simplicity of notation, can be found in [4.8] and [4.9].

$$\text{logit P}(Y_t = 1) = \beta_0 + \sum_{k=1}^{p} \beta_k X_{tk} \qquad [4.8]$$

$$\text{logit P}(Y_t = 1) = \Delta_t + \sum_{j=1}^{p} \alpha_j Y_{t-j} \qquad [4.9]$$

where Δ_t is a function of β and X.

Like in Lipsitz et al. (1991) and Fitzmaurice and Laird (1993), Heagerty (1999) suggested using a two step regression models. A marginal logistic regression is set for the average response explained by independent variables in the first model while a conditional model is set for the dependence through longitudinal latent variable. The inference for the coefficients is that β explains the marginal association between dependent and independent variables whereas α gives the effect of serial dependence (Comstock and Heagerty). Having gotten the estimates for parameters after these two steps calculation, the logit probability of response being 1 is calculated for the forecast value as in [4.10].

$$\widehat{P(Y_t = 1)} = \frac{\exp\left(\widehat{\Delta_t} + \sum_{J=1}^{p} \widehat{\alpha_J} Y_{t-J}\right)}{1 + \exp\left(\widehat{\Delta_t} + \sum_{J=1}^{p} \widehat{\alpha_J} Y_{t-J}\right)} \qquad [4.10]$$

Although it can be generalized to a p^{th} order model, there is only up to 2^{nd} order of MTM implemented in R library. Therefore, here, we set MTM(1) and MTM(2). For simplicity, how a MTM(1) is fitted is explained here. The coefficients of MTM(1) which are β_0, β_1, β_2, β_3, β_4, α and Δ are estimated by the help of [4.11] and [4.12] provided below using the first 4 time points for X's and Y. Next, Δ is forecasted using the moving mean approach which gives the best results here compared to AR and random effects models. Finally, the probability of being 1 for the response is computed and classified as in [4.13] and [4.14].

$$\text{logit P}(Y_t = 1) = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \beta_4 X_{t4} \qquad [4.11]$$

$$\text{logit P}(Y_t = 1) = \Delta_t + \alpha Y_{t-1} \qquad [4.12]$$

$$\widehat{P(Y_t = 1)} = \frac{\exp\left(\widehat{\Delta_t} + \hat{\alpha} Y_{t-1}\right)}{1 + \exp\left(\widehat{\Delta_t} + \hat{\alpha} Y_{t-1}\right)} \; for \; t = 5, 6, 7, 8 \qquad [4.13]$$

$$\widehat{Y_t} = 1 \text{ if } \widehat{\text{P}(Y_t = 1)} \geq 0.5 \text{ for t= 5, 6, 7, 8} \qquad [4.14]$$

$$\widehat{Y_t} = 0 \text{ if } \widehat{\text{P}(Y_t = 1)} < 0.5 \text{ for t= 5, 6, 7, 8}$$

4.1.6. Random Effects Models

Fitzmaurice et al. (2004) explained random effects models (REM) as the model having covariates and a vector of random effects. REM is valuable when the aim is individuals' inferences but not the population's (Diggle et al., 2002). They proclaimed that if there is no homogeneity through individuals in terms of their coefficients and if a probability distribution can be defined for this non-homogeneity, then REM is used. Moreover, Torres-Reyna noted that if there is an effect of differences on response through individuals, then REM should be used. Controlling for the individual unobserved heterogeneity as well as allowing time invariant variables can be given as the advantages of this method (Torres-Reyna). The random effects model for binary data is given in [4.15]:

$$\text{logit P}(Y_{it} = 1 \,|\text{X}_{\text{itk}}, \text{U}_\text{i}) = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{itk} + \alpha_{p+1} \text{U}_\text{i} \qquad [4.15]$$

where α_0 is constant, α_k's and α_{p+1} are the coefficients of parameters and U_i is the unobservable variable. U_i's are generally assumed to be coming from normal distribution with zero mean. In this model, coefficients have "subject-specific" interpretations. Interested readers are referred to Chapter 9 of Diggle et al. (2002) for details and examples of REM.

In this thesis, random intercept model having with and without lag of response is fitted using "glmer" code in R. Forecast values are computed as given in [4.17] and [4.18] for REM without lag of response method after data was fitted by [4.16].

$$\text{logit P}(Y_{it} = 1 \,|\text{X}_{\text{itk}}, \text{U}_\text{i}) = \alpha_0 + \alpha_1 X_{it1} + \alpha_2 X_{it2} + \alpha_3 X_{it3} + \alpha_4 X_{,t4} + \alpha_5 U_i \qquad [4.16]$$

$$\widehat{P(Y_{it} = 1)} = \frac{\exp\left(\widehat{\alpha_0} + \widehat{\alpha_1} X_{it1} + \widehat{\alpha_2} X_{it2} + \widehat{\alpha_3} X_{it3} + \widehat{\alpha_4} X_{it4} + \widehat{\alpha_5} U_i\right)}{1 + \exp\left(\widehat{\alpha_0} + \widehat{\alpha_1} X_{it1} + \widehat{\alpha_2} X_{it2} + \widehat{\alpha_3} X_{it3} + \widehat{\alpha_4} X_{it4} + \widehat{\alpha_5} U_i\right)} \qquad [4.17]$$

$$\widehat{Y_{it}} = 1 \text{ if } \widehat{\text{P}(Y_{it} = 1)} \geq 0.5 \text{ for t= 5, 6, 7, 8} \qquad [4.18]$$

$$\widehat{Y_{it}} = 0 \text{ if } \widehat{\text{P}(Y_{it} = 1)} < 0.5 \text{ for t= 5, 6, 7, 8}$$

4.2. Estimation Techniques

The most common estimation techniques that are used in longitudinal data are Generalized Estimating Equation (GEE), Maximum Likelihood Estimation (MLE), Restricted Maximum Likelihood Estimation (REMLE) and Bayesian techniques. Information for the two of these techniques that are used in this thesis, namely GEE and REMLE, is supplied below.

GEE method was first introduced by Liang and Zeger (1986) and Zeger and Liang (1986) for longitudinal data having non-normal response variables such as binary, Poisson, Gamma and negative binomial. By the help of this method, one can get unbiased and more efficient estimates compared to the ones that ignore within correlation. This is because, GEE allows dependence within clusters, although logistic regression does not (Lam, 2007). Therefore, GEE is used widely especially in medical and life sciences such as epidemiology, gerontology, and biology as it is expressed in Ballinger (2004).

The method of Liang's and Zeger's is eventuated in two steps. Firstly, it assumes that there is no dependence between the outcomes, and calculates the naive standard errors according to this assumption using standard generalized linear model techniques. Secondly, it makes an adjustment to these errors by calculating the working correlation matrix, the part that takes the within-subject correlations of responses on dependent variables into account. The statistics which are gotten after this procedure are unbiased and more efficient estimators with respect to ordinary least square ones. In fact, when the variance of response is correctly defined, GEE gives the maximum likelihood score equation for binary data and multivariate Gaussian data (Fitzmaurice and Laird, 1993).

There are three things that one needs to beware when using the GEE technique (Ballinger, 2004). One of them is the distribution of the response variable. In general, users have general information about the distribution of the response variable. Ballinger (2004) stated that distributions coming from exponential family, such as normal, binomial, poisson, negative binomial, and gamma distributions can be defined in GEEs. In the methods for which we used GEE in this thesis, binomial distribution is defined as the distribution of binary response variables. The other issue is the link function that is used with these distributions. Ballinger (2004) stated the link functions as identity link, power link or reciprocal link for normal distribution, while logit link, probit link, power link or reciprocal link can be used for binomial distribution. He added that log link, power link or reciprocal link are used for poisson distribution, whereas gamma and negative binomial distributions need power link or reciprocal link, and power link, respectively. In this thesis, logit link is used when GEE is fitted. Last issue is the structure for the correlation of response variable. For the specification of correlation structure, there are several options, such as exchangeable, autoregressive and

unstructured working correlation as it was presented in Ballinger (2004). See the matrices given below.

$$\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

exchangeable correlation structure

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

first-order autoregressive correlation structure

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}^2 \end{bmatrix}$$

unstructured working correlation

In these matrices, there are 4 time points. For example, in an unstructured working correlation matrix, σ_{12} corresponds to the standard deviation of response at first and second time points whereas σ_{13} is the standard deviation of response at first and third time points. Ballinger (2004) explained that exchangeable correlation structure is used when there are equal correlation within-subject observations meaning that any logical ordering can be found within a cluster. Moreover, autoregressive correlation structure is used when there is within-subject correlation over time having an exponential function of this time. The last one, unstructured working correlation, can be used for the estimation on all types of within-subject correlation. Therefore, we have chosen unstructured one. For the details of structures of working correlations, one can see Lam (2007) and Zorn (2001).

As it was expressed in Ballinger (2004), although the correct form of the correlation of responses is needed to increase the efficiency (Fitzmaurice (1995), Hardin and Hilbe (2003)) especially in which the correlation is high within responses (Diggle et al. (2002), Zorn (2001)), the model is generally robust to errors in the case of moderate amount of misspecification of correlation structure (Liang and Zeger, 1986). Furthermore, to estimate correlation parameters accurately, Park et al. (1998) compared three different methods using Pearson, Anscombe and deviance residuals. They showed through simulation studies that Pearson residual can continue to be used since there is little or no effect on resulting estimates' properties for Poisson and binary outcomes.

Since just marginal mean and covariates are the focus, full distributional assumptions for repeated responses and a very detailed consideration of the dependence structure are not needed in marginal models (Fitzmaurice et al., 2004). Because of this, they reported that GEE is used in marginal models. Moreover, Zorn (2001) used GEE to see the population-averaged effects. He expressed that GEE explain, for marginal methods, how much the average response changes when there is a one-unit increase in the independent variables. Furthermore, for transition models, Diggle et al. (2002) suggested using GEE for the estimation of coefficients if the response variable is discrete.

The other technique, which is Restricted Maximum Likelihood Estimation (REMLE) technique, was first introduced by Patterson and Thomson (1971). This method estimates the variance components in a generalized linear model to prevent the biased estimators of standard maximum likelihood procedure. In fact, it is a particular form of well known maximum likelihood estimation. The difference between them is that nuisance parameters have no effect in REMLE because likelihood function is computed from a transformed set of data while maximum likelihood uses all the information. Interested readers are referred to Cullis and McGilchrist (1990) and Verbyla and Cullis (1990) on the usage of REMLE in longitudinal data.

For the estimation of coefficients in random effects model, one can use conditional maximum likelihood or maximum likelihood (Diggle et al., 2002). However, because maximum likelihood estimators of the variance components are generally biased and sometimes can be negative, restricted maximum likelihood estimation (REMLE) is preferred to overcome these problems (Frees, 2004).

In this thesis, GEE was tried to be used for all methods to make reasonable comparisons by eliminating the estimation technique bias. However, this was not suitable for all cases, such as for MTM and random effects models. Therefore, the estimation of coefficients in AR models, marginal models and transition models are done through GEE as it was suggested by Diggle et al. (2002) and Zorn (2001), while estimation in random effects models are done by using REMLE. Moreover, the starting values for the parameters estimated in MTM is estimated by using GEE, and the estimation is completed by using MLE.

4.3. Accuracy Measures

It is needed to check the model using real future data (Hyndman, 2010) to understand the success of the method. According to Hyndman, instead of checking how the model fits to data, checking how the real future data fits the model is more vital in forecasting. Thus, one should keep some part of the data, and use it after fitting the model with the rest of the data to test how well the model fit for new data. Therefore, we split the data into two parts and kept the second half of the data for checking the accuracy. For model evaluation criteria in case of binary response model, one can find many suggestions in Efron (1978), Jackman (2007), Winkelmann and Boes (2009), Korn and Simon (1991), Bontemps et al. (2009) and Golder. In these articles, most popular ones are the proportion of correct predictions (PCP) and Area under the Receiver Operating Characteristic (AUROC) curve. Moreover, one and the most common way of assessing the predictive power of a model is composing a cross-table of actual outcomes versus forecasted outcomes applying classification to 1 if the predicted probability is equal or greater than 0.5 while 0 if it is less than 0.5 (Jackman, 2007).

That is, one should make such a table given below before calculating the statistics for model accuracy.

		Forecasted		
		0	1	
Observed	0	TN	FP	
	1	FN	TP	Total
			Total	N

Here, TN and FN are true negative and false negative values, respectively. Moreover, FP is false positive while TP is true positive. With the help of this table, we compare the forecast results of the 21 methods with the actual values. PCP, expected PCP, AUROC curve and deviance, which are explained below, are calculated and compared for these methods.

To begin with, PCP is the most widely used method in the literature for measuring the accuracy of binary classification. This proportion measures the percentage of observations predicted truly (Golder). It is the ratio of total number of true negative (TN) and true positive (TP) over the specified sample. The formula is provided in [4.19].

$$PCP = \frac{\text{Number of Correct Predictions}}{\text{N}} * 100 = \frac{\text{TN} + \text{TP}}{\text{N}} * 100 \qquad [4.19]$$

Therefore, the model with higher PCP is the better one. Since it is easy to calculate, it is widely used in the literature; however, there are several problems with PCP. The most important one is that this index is very sensitive to cut-off value ascribing an important role to it as shown in Bontemps et al. (2009) and Golder. Golder stated that Type 1 and Type 2 errors are assumed equally bad because of the dependence on cut-off value, in particular 0.5. Furthermore, Golder added that when there is an increase in the value for cut-off, the chance of making one type of error decreases while other type of error increases. Another problem occurs when the classification is done with respect to cut off value (Golder). To illustrate, $\widehat{p_i}$ = 0.49 and $\widehat{p_i}$ = 0.51 are classified under different categories although they give similar information. Moreover, $\widehat{p_i}$ = 0.51 and $\widehat{p_i}$ = 0.99 are classified under the same category, although the latter value of $\widehat{p_i}$ implies higher probability of event occuring than the former. Thus, the precision of PCP is overstated by this procedure. Discussion on other problems can be found in Golder. In this thesis, this statistics is used for both the first and second half of the data to check the model efficiency and the forecast efficiency.

Since PCP is dependent on cut-off value as well as having precision problems, Golder suggested two other approaches instead of PCP, namely percentage reduction in error (PRE) and expected percent correctly predicted (ePCP). However, PRE has also the same

problems with PCP since it is a function of PCP. Therefore, it is not used in this thesis. Instead, the other alternative method, that is ePCP, is used.

This index was proposed by Herron (1999) to prevent the problem of treating two observations with $\widehat{p_i}$ = 0.51 and $\widehat{p_i}$ = 0.99 as in the same class. The formula for this statistics is provided in [4.20].

$$ePCP = \frac{\sum_{yi=1} \widehat{p_i} + \sum_{yi=0} (1 - \widehat{p_i})}{N} \qquad [4.20]$$

This formula calculates the proportion by summing the estimated probabilities of being 1 when the observed response is 1 and the estimated probabilities of being 0 when the observed response is 0, and then dividing this by the sample size. Higher values are also good for this index. The difference between ePCP and PCP is that, in PCP proportion is calculated by summing the number of observations having same value before and after classification while this operates on probabilities. Therefore, it has an advantage with respect to PCP. For instance, if the $\widehat{p_i}$ is 0.49 when the variable is 1, then it is classified as 0 in PCP while it keeps its value as 0.49 in ePCP. Moreover, let's say the subject has the probability of being 1 as 0.99 when it is in fact observed as 0. It is classified as 1 in PCP whereas it is evaluated as 0.01 in ePCP. Therefore, ePCP solves the precision problem. Furthermore, ePCP should also be preferred for the model fit rather than PCP (Golder).

Beside these point estimations for PCP and ePCP, confidence intervals can be constructed to see the reliability. Therefore, we used the formula given in [4.21] for the approximate confidence intervals.

$$P\left\{\hat{\mathrm{p}} - \mathrm{Z}_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{\hat{\mathrm{p}} * (1-\hat{\mathrm{p}})}{n}} < \mathrm{p} < \hat{\mathrm{p}} + \mathrm{Z}_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{\hat{\mathrm{p}} * (1-\hat{\mathrm{p}})}{n}}\right\} = 0.95 \qquad [4.21]$$

where $\hat{\mathrm{p}}$ is the proportion of successes, $\mathrm{Z}_{\left(1-\frac{\alpha}{2}\right)}$ is the $\left(1-\frac{\alpha}{2}\right)$ percentile of a standard normal distribution, α=0.05 is the significance level and n is the sample size.

To prevent the dependence on classification rule in the indexes like PCP and ePCP, ROC curves are also used in this thesis. A ROC curve, which is composed of specificity and sensitivity, comprises a tradeoff between all possible cut-off points. Thus, AUROC curve is a summary measure of the classification performance. Because it does not depend on any cut-off value, it gives better results than PCP (Bontemps et al., 2009).

Egan (1975) illustrates that ROC is a graphical plot of the sensitivity (the percentage of true positive predictions) versus 1− specificity (the percentage of false positive predictions). Specifically, the y-axis is sensitivity whereas the x-axis is the 1-specificity. Formulas for sensitivity and specificity are provided in [4.22] and [4.23]. One can calculate them using the cross tabulation values.

$$Sensitivity\ (TPR) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad [4.22]$$

$$Specificity\ (SPC) = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \qquad [4.23]$$

The 45 degree line, which divides the area into two parts, shows the tradeoff between sensitivity and 1-specificity when the model has no covariates. The line calculated from the model with covariates is the ROC curve. The more distance between the ROC curve and 45 degree line, the better the model predicts both 0s and 1s. AUROC curve is used for conveying this information. If this value is 1, there is no tradeoff between predicting two values, 1s and 0s, implying that the model predicts everything correctly. When it decreases, the model becomes worse.

Besides these statistics, model checking is also done to find the parsimonious model in this thesis. This is because especially when the forecast accuracies are similar, such checks should be used to choose one of the most parsimonious and most well fitted models. Frees (2004) summarized the goodness of fit statistics and information criteria in his book. He expressed that, for the generalized linear models, deviance statistics can be used for goodness of fit and for providing information on how well the model fits. It can be defined as the log-likelihood ratio of the reduced model compared to the full model. Smaller deviance means that the predicted values and observed values are similar. For Bernoulli distribution, it is defined as the formula given in [4.24].

$$D = -2 * \left\{ \sum_{i=1}^{n} yi * \ln\left({yi}/{\widehat{p_i}}\right) + (1 - yi) * ln\left({1 - yi}/{1 - \widehat{p_i}}\right) \right\} \qquad [4.24]$$

Beside the deviance statistics, R^2 is one of the most commonly used statistics for goodness of fit in linear regression models (Frees, 2004). On the contrary, when the model is nonlinear, there is a possibility of R^2 being in an interval outside [0,1]. Therefore, it is not an appropriate statistics for nonlinear models. As an alternative for R^2 goodness-of-fit summary statistics, marginal R^2 was introduced by Zheng (2000). This extension of R^2 compares how the estimated model and intercept-only model fit. However, it can take value less than zero. Therefore, it is not used in this thesis either. Another alternative to this statistic is Pearson-chi-

square statistics which was the first goodness of fit test defined by Karl Pearson in 1900. It is an optimal test for categorized data (Frees, 2004). The Pearson chi-square which was suggested in Pan (2002) is not used in this thesis since this test requires independence.

Furthermore, choosing one among nested models, likelihood ratio test can be used. When the models are not nested, Akaike's Information Criteria (AIC), which is defined as -2L($\ominus_{MLE}$) + 2*(number of parameters) can be used. Moreover, Bayesian Information Criteria (BIC), -2L($\ominus_{MLE}$) + ln(number of parameters) gives more weight to the number of parameters suggesting more parsimonious model than AIC when the other things are equal (Frees, 2004). On the other hand, as it was emphasized in Dziak and Li (2007) and Pan (2001), in GEE approach the traditional measures of model fit cannot be used because there is not a likelihood structure in GEE. Thus, extending these traditional measures, i.e. Cp, AIC and BIC, is challenging. Therefore, to be able to compare different models for which parameters are estimated by REMLE or GEE, information criteria is not used in this thesis.

For comparing the forecast efficiency of continuous variables, the deviation is used in this thesis. Deviation which measures the difference between the actual and the predicted values is calculated by using the formula given in [4.25].

$$deviation = \frac{\sum_{i=1}^{N}(x_{i,actual} - x_{i,predicted})^2}{N} \qquad [4.25]$$

CHAPTER 5

RESULTS

In this part, the results for both dependent and independent variables after 10,000 trials are explained. Firstly, forecast accuracy results for independent variables that were used in some methods are given. Next, the ones for the response variable are given.

Results can be found in Tables 5 through 8 for the forecast accuracy of independent variables for 10,000 trials, whereas R codes for these calculations can be found in Appendix C. According to these results, AR methods give the smallest deviations in all cases. AR(2) gives smaller values for the time points 5 and 6 while AR(1) gives the smallest values for the time points 7 and 8 for all covariates having different variances. Therefore, AR methods are decided to be used for forecasting continuous independent variables. Moreover, since the chi-square value for 1 degrees of freedom, which is 3.84, is bigger than the difference between the values of AR(1) and AR(2), it is concluded that there is no statistically significant difference between these two models. Therefore, AR(1) model, which is the more parsimonious model, is used for all X's.

Table 5. The results of the average forecast accuracy values through deviation for $X_{k=1}$ after 10,000 trials

Time / Model	t=5	t=6	t=7	t=8
AR (1)	1.40	2.17	2.77	3.51
AR (2)	1.39	2.16	2.80	3.53
Simple mean average	2.07	2.98	3.88	4.82
Random effect	2.14	2.86	3.64	4.59

Table 6. The results of the average forecast accuracy values through deviation for $X_{k=2}$ after 10,000 trials

Method \ Time	t=5	t=6	t=7	t=8
AR (1)	1.40	2.18	2.77	3.51
AR (2)	1.39	2.16	2.80	3.53
Simple mean average	2.07	2.97	3.88	4.82
Random effect	2.14	2.85	3.64	4.59

Table 7. The results of the average forecast accuracy values through deviation for $X_{k=3}$ after 10,000 trials

Method \ Time	t=5	t=6	t=7	t=8
AR (1)	14.01	21.75	27.66	34.94
AR (2)	13.86	21.64	27.94	35.18
Simple mean average	20.70	29.76	38.75	48.09
Random effect	21.38	28.56	36.33	45.75

Table 8. The results of the average forecast accuracy values through deviation for $X_{k=4}$ after 10,000 trials

Method \ Time	t=5	t=6	t=7	t=8
AR (1)	139.77	217.06	276.18	349.90
AR (2)	138.28	215.93	279.04	352.20
Simple mean average	206.64	297.55	387.59	481.89
Random effect	213.46	285.64	363.33	458.51

From these tables, it is seen that the values are high for the variables $X_{k=3}$ and $X_{k=4}$. These are the variables generated with high variances. In an attempt to be able to decrease these quantities, standardization method was used. That is, the mean of the values are subtracted from the observations and divided by the variance values for each year. The forecast accuracy values for independent variables became tiny when standardization was used; however, no difference was found for the forecast values of response variable. Therefore, standardization was not used in the methods for forecasting the response variable.

Table 9 is supplied again to remind the names of the methods. The results of the comparison of forecast performance of these methods can be found through Table 10 and Table 13 while R codes for methods and forecasting binary data and forecast accuracy measures are provided at Appendix D and Appendix E, respectively.

Table 9. Methods that are used in this thesis

No	Method
1	Moving simple median
2	Non-moving simple median
3	Moving simple mode
4	Non-moving simple mode
5	Moving simple average; cut-off value is taken as 0.5
6	Non-moving simple average; cut-off value is taken as 0.5
7	Moving width simple average; cut-off value is taken as a random variable distributed uniformly
8	Non-moving simple average; cut-off value is taken as a random variable distributed uniformly
9	Moving simple average; cut-off value is taken as the mean value of past values
10	Non-moving simple average; cut-off value is taken as the mean value of past values
11	AR (1) model
12	AR (2) model
13	AR (3) model
14	Marginal Models
15	TM (1)
16	TM (2)
17	TM (3)
18	MTM (1)
19	MTM (2)
20	Random intercept model with no lag of response
21	Random intercept model with lag of response

Table 10 shows the values for proportion of correct predictions (PCP) and confidence intervals for these proportions. Here, because the estimation cannot be done in naïve models for the first 4 years, the values are absent for these years. According to the PCP results, although all of the methods give good results, random effects models are the most suitable ones for fitting the model, i.e. for t=1,2,3,4. Specifically, random effects models with no lag of response is the best method while the marginal models are the worst ones in this case. Correct prediction proportions range between 79 % and 95 %. Furthermore, when the forecast efficiency is compared, all of the models have again good values although the ratio decreases when it is compared to model fit performance. However, this decrease is not the case for all methods. For instance, forecast performance and model fit performance are almost the same for some methods such as AR, marginal models and transition models especially for t=5. Besides, ratios are generally higher than or around 80 % for the first two years of forecasting implying that short term forecasting can be done. The only exception is the marginalized transition models. In forecasting, AR methods take the place from random effects models supplying the highest ratio. However, random effects models with no lag of response also has high values. AR (1) model gives the best ratios as 85.9 % and 81.0 % for the first two years of forecasting with an interval of 81.9-89.8 % and of 76.5-85.4 %, respectively. One interesting result, here, is that naïve methods also give a performance around 80 % with a confidence interval of 73-84 % at a 0.05 significance level for the first forecasting year. This performance passes the performance of some of the complex methods such as marginalized transition models, random effects models with a lag of response and marginal models. Another interesting result is that marginalized transition models presents the lowest ratio for forecasting although it gives good values when fitting the model. This is because the calculated values for Δ's do not have a specific trend and/or it has outliers. Therefore, forecasting Δ's is challenging. In order to overcome this, two approaches were tried: standardized X's were used with the aim of converging Δ's, and non-linear estimation techniques were used to estimate Δ's directly. However, no success was achieved in any methods.

Table 10. Values and confidence intervals for the proportion of correct predictions (PCP)

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
1	NA	0.791 (0.745,0.837)	0.740 (0.690,0.789)	0.690 (0.638,0.743)	0.652 (0.598,0.705)	0.718 (0.693,0.744)
2	NA	0.791 (0.745,0.837)	0.740 (0.690,0.789)	0.690 (0.638,0.743)	0.652 (0.598,0.705)	0.718 (0.693,0.744)
3	NA	0.791 (0.745,0.837)	0.740 (0.690,0.789)	0.690 (0.638,0.743)	0.651 (0.598,0.705)	0.718 0.693,0.744)
4	NA	0.791 (0.745,0.837)	0.740 (0.690,0.789)	0.690 (0.638,0.743)	0.651 (0.598,0.705)	0.718 (0.693,0.744)
5	NA	0.791 (0.745,0.837)	0.740 (0.690,0.789)	0.690 (0.638,0.743)	0.652 (0.598,0.705)	0.718 (0.693,0.744)
6	NA	0.791 (0.745,0.837)	0.740 (0.690,0.789)	0.690 (0.638,0.743)	0.652 (0.598,0.705)	0.718 (0.693,0.744)
7	NA	0.778 (0.731,0.825)	0.747 (0.698,0.796)	0.708 (0.656,0.759)	0.667 (0.614,0.720)	0.725 (0.700,0.750)
8	NA	0.779 (0.732,0.826)	0.732 (0.682,0.782)	0.684 (0.631,0.736)	0.646 (0.592,0.700)	0.710 (0.684,0.736)
9	NA	0.789 (0.743,0.835)	0.743 (0.694,0.792)	0.696 (0.644,0.748)	0.656 (0.602,0.710)	0.721 (0.696,0.746)
10	NA	0.789 (0.743,0.835)	0.738 (0.689,0.788)	0.689 (0.637,0.741)	0.650 (0.597,0.704)	0.717 (0.691,0.742)
11	0.857 (0.834,0.880)	0.859 (0.819,0.898)	0.810 (0.765,0.854)	0.750 (0.701,0.799)	0.696 (0.644,0.748)	0.779 (0.755,0.802)
12	0.858 (0.830,0.886)	0.857 (0.818,0.897)	0.809 (0.764,0.853)	0.749 (0.700,0.798)	0.696 (0.644,0.748)	0.778 (0.754,0.801)
13	0.859 (0.819,0.898)	0.856 (0.817,0.896)	0.808 (0.764,0.852)	0.748 (0.699,0.797)	0.696 (0.644,0.748)	0.777 (0.753,0.800)
14	0.789 (0.766,0.812)	0.749 (0.700,0.798)	0.716 (0.665,0.767)	0.681 (0.628,0.733)	0.646 (0.592,0.700)	0.698 (0.672,0.724)
15	0.864 (0.842,0.886)	0.846 (0.805,0.887)	0.798 (0.753,0.843)	0.742 (0.692,0.791)	0.690 (0.638,0.743)	0.769 (0.745,0.793)
16	0.869 (0.842,0.896)	0.843 (0.802,0.884)	0.792 (0.746,0.838)	0.734 (0.684,0.784)	0.689 (0.637,0.741)	0.764 (0.740,0.788)
17	0.874 (0.837,0.912)	0.848 (0.807,0.888)	0.797 (0.752,0.843)	0.739 (0.689,0.788)	0.691 (0.639,0.743)	0.769 (0.745,0.793)
18	0.849 (0.829,0.869)	0.536 (0.480,0.592)	0.539 (0.483,0.595)	0.534 (0.477,0.590)	0.524 (0.468,0.581)	0.533 (0.505,0.561)
19	0.853 (0.833,0.873)	0.560 (0.504,0.616)	0.564 (0.508,0.620)	0.553 (0.496,0.609)	0.545 (0.488,0.601)	0.555 (0.527,0.583)
20	0.950 (0.938,0.962)	0.839 (0.797,0.880)	0.785 (0.738,0.831)	0.726 (0.675,0.776)	0.677 (0.625,0.730)	0.757 (0.732,0.781)
21	0.895 (0.875,0.915)	0.720 (0.670,0.771)	0.678 (0.626,0.731)	0.640 (0.586,0.694)	0.602 (0.547,0.657)	0.660 (0.633,0.687)

Table 11 presents the values and confidence intervals for expected percent correctly predicted (ePCP). Here, because the estimation cannot be done in naïve models for the first 4 years, the values are absent for these years. According to the ePCP results, all of the models give worse values when compared to PCP outputs except for the naïve methods. PCP and ePCP give similar values for naïve methods. Although the proportions are different in terms of these two criteria, the ranking is almost the same for model fitting. Random effects models and transition models are the first two methods while marginal models are in the last place. AR methods are in the fourth place in ePCP ranking after marginalized transition models whereas it is in the third place ranking before marginalized transition model in PCP. On the other hand, the situation for forecasting is a bit different. AR methods are in the third place while it is in the first in PCP. Transition models are in the second place in both criteria while marginalized transition models gives the worst values in both cases. In addition, naïve models are in a better position than AR, random effects models with response as a lag, marginal models and marginalized transition models. Thus, it can be expressed that random effects models are the most suitable ones for both fitting the model and forecasting new values according to this criteria as well. Specifically, random effects model having no lag of response is the best one whereas marginal models and marginalized transition models are the worst ones for modeling data and forecasting, respectively. In addition, for the best method, ratios are 80.9 % and 75.9 % for the first two coming years with an interval of 76.5-85.3 % and of 71.0-80.8%, respectively, implying that short term forecasting can be done.

Results of the area under the receiver operating characteristic (AUROC) curve, which is the criteria not depending on any cut-off value, are presented at Table 12. Here, because the estimation cannot be done in naïve models, the values are absent. When we look at these quantities, it is seen that there is an increase for the accuracy in all models with respect to other criteria, which are PCP and ePCP. Here, random effects models are the most suitable ones for fitting the model as in the other criteria. The AUROC curve is 0.99 meaning that the best model fits quite well. In addition, for model fitting, transition models and marginalized transition models follow it with a big difference than other models contrary to other criteria. When the forecast accuracies are compared, random effects models, transition models and AR methods are in the first three places in turn as in the ePCP. Marginalized transition models are the worst one again with a worse quantity for forecasting efficiency.

Table 11. Values and confidence intervals for expected percent correctly predicted (ePCP)

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
1	NA	0.791 (0.745,0.837)	0.750 (0.701,0.799)	0.699 (0.647,0.751)	0.652 (0.598,0.706)	0.723 (0.698,0.748)
2	NA	0.791 (0.745,0.837)	0.740 (0.690,0.790)	0.690 (0.638,0.742)	0.652 (0.598,0.706)	0.718 (0.693,0.743)
3	NA	0.791 (0.745,0.837)	0.740 (0.690,0.790)	0.690 (0.638,0.742)	0.651 (0.598,0.705)	0.718 (0.693,0.743)
4	NA	0.791 (0.745,0.837)	0.740 (0.690,0.790)	0.690 (0.638,0.742)	0.651 (0.598,0.705)	0.718 (0.693,0.743)
5	NA	0.778 (0.732,0.824)	0.749 (0.700,0.798)	0.707 (0.656,0.758)	0.663 (0.609,0.717)	0.724 (0.699,0.749)
6	NA	0.778 (0.732,0.824)	0.734 (0.684,0.784)	0.686 (0.633,0.739)	0.649 (0.595,0.703)	0.712 (0.686,0.738)
7	NA	0.778 (0.732,0.824)	0.747 (0.698,0.796)	0.708 (0.656,0.760)	0.667 (0.614,0.720)	0.725 (0.700,0.750)
8	NA	0.778 (0.732,0.824)	0.732 (0.682,0.782)	0.683 (0.631,0.735)	0.646 (0.592,0.700)	0.710 (0.684,0.736)
9	NA	0.778 (0.732,0.824)	0.749 (0.700,0.798)	0.708 (0.656,0.760)	0.665 (0.611,0.719)	0.725 (0.700,0.750)
10	NA	0.778 (0.732,0.824)	0.733 (0.683,0.783)	0.685 (0.633,0.737)	0.648 (0.594,0.702)	0.711 (0.686,0.736)
11	0.781 (0.754,0.808)	0.782 (0.735,0.829)	0.743 (0.694,0.792)	0.697 (0.645,0.749)	0.654 (0.600,0.708)	0.719 (0.693,0.745)
12	0.769 (0.735,0.803)	0.768 (0.720,0.816)	0.741 (0.691,0.791)	0.694 (0.642,0.746)	0.652 (0.599,0.705)	0.714 (0.688,0.740)
13	0.772 (0.725,0.819)	0.769 (0.721,0.817)	0.740 (0.691,0.789)	0.693 (0.641,0.745)	0.652 (0.599,0.705)	0.714 (0.688,0.740)
14	0.700 (0.674,0.726)	0.665 (0.611,0.719)	0.635 (0.580,0.690)	0.606 (0.551,0.661)	0.580 (0.524,0.636)	0.621 (0.594,0.648)
15	0.797 (0.770,0.824)	0.775 (0.728,0.822)	0.731 (0.681,0.781)	0.687 (0.634,0.740)	0.646 (0.592,0.700)	0.710 (0.684,0.736)
16	0.808 (0.776,0.840)	0.784 (0.737,0.831)	0.745 (0.696,0.794)	0.695 (0.643,0.747)	0.655 (0.602,0.708)	0.720 (0.694,0.746)
17	0.817 (0.773,0.861)	0.792 (0.746,0.838)	0.752 (0.703,0.801)	0.701 (0.649,0.753)	0.660 (0.606,0.714)	0.726 (0.701,0.751)
18	0.781 (0.757,0.805)	0.393 (0.337,0.449)	0.408 (0.352,0.464)	0.424 (0.368,0.480)	0.435 (0.379,0.491)	0.415 (0.387,0.443)
19	0.787 (0.764,0.810)	0.386 (0.331,0.441)	0.403 (0.348,0.458)	0.417 (0.361,0.473)	0.428 (0.372,0.484)	0.408 (0.381,0.435)
20	0.902 (0.885,0.919)	0.809 (0.765,0.853)	0.759 (0.710,0.808)	0.705 (0.653,0.757)	0.660 (0.607,0.713)	0.733 (0.708,0.758)
21	0.842 (0.819,0.865)	0.687 (0.635,0.739)	0.652 (0.598,0.705)	0.620 (0.566,0.674)	0.590 (0.534,0.646)	0.637 (0.610,0.664)

Table 12. Values for area under the receiver operating characteristic (AUROC) curve

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
1	NA	0.819	0.763	0.706	0.652	0.733
2	NA	0.819	0.740	0.690	0.652	0.723
3	NA	0.791	0.740	0.690	0.651	0.718
4	NA	0.791	0.740	0.690	0.651	0.718
5	NA	0.861	0.808	0.750	0.695	0.774
6	NA	0.861	0.802	0.739	0.691	0.772
7	NA	0.861	0.808	0.751	0.698	0.777
8	NA	0.861	0.802	0.738	0.690	0.773
9	NA	0.861	0.808	0.751	0.695	0.774
10	NA	0.861	0.802	0.739	0.691	0.772
11	0.857	0.859	0.810	0.750	0.696	0.778
12	0.885	0.885	0.810	0.749	0.696	0.782
13	0.888	0.885	0.810	0.749	0.696	0.782
14	0.875	0.832	0.792	0.749	0.703	0.772
15	0.935	0.898	0.833	0.779	0.726	0.812
16	0.941	0.905	0.845	0.775	0.725	0.815
17	0.944	0.905	0.843	0.777	0.726	0.815
18	0.927	0.206	0.246	0.290	0.330	0.268
19	0.931	0.217	0.262	0.307	0.344	0.283
20	0.990	0.905	0.849	0.785	0.729	0.821
21	0.963	0.811	0.761	0.714	0.664	0.740

In Table 13, values for deviance are provided for only non-naïve methods. This is because there is no likelihood function for naïve methods. According to the results, TM(3) is the best method, while AR(3) and random effects models without lag of response is second and third one, respectively. The worst method is marginal model followed by the marginalized transition models with one lag of response. Furthermore, when we compare the significance of the difference between TM(3) and TM(2), it is seen that there is a significant difference at 0.05 significance level. This is because, 190.75 which is the difference between 370.94 and 180.19 is bigger than the tabulated chi-square value with 1 degrees of freedom which is 3.84. Therefore, TM(3) is statistically better than TM(2) at 0.05 significance level. Therefore, although it seems to be reasonable to use TM(1) since there is not much difference between TM(3), TM(2) and TM(1) in accuracy measures such as PCP, ePCP and AUROC, when the difference for the deviance values are tested, it is seen that TM(3) is statistically different from TM(2).

Table 13. Values for deviance

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
1	NA	NA	NA	NA	NA	NA
2	NA	NA	NA	NA	NA	NA
3	NA	NA	NA	NA	NA	NA
4	NA	NA	NA	NA	NA	NA
5	NA	NA	NA	NA	NA	NA
6	NA	NA	NA	NA	NA	NA
7	NA	NA	NA	NA	NA	NA
8	NA	NA	NA	NA	NA	NA
9	NA	NA	NA	NA	NA	NA
10	NA	NA	NA	NA	NA	NA
11	-747.71	-248.85	-311.38	-387.42	-456.44	-1404.10
12	-462.89	-235.05	-310.11	-386.08	-453.22	-1384.46
13	-229.31	-237.16	-313.62	-389.75	-455.88	-1396.41
14	-1057.75	-301.60	-330.35	-357.46	-381.48	-1370.89
15	-584.69	-241.74	-320.79	-383.46	-439.48	-1385.46
16	-370.94	-237.93	-325.94	-419.24	-478.16	-1461.27
17	-180.19	-241.63	-336.75	-427.02	-490.28	-1495.68
18	-822.22	-4266.80	-4056.85	-3800.63	-3633.92	-15758.21
19	-798.80	-3662.39	-3430.06	-3307.58	-3137.62	-13537.64
20	-325.67	-284.68	-395.36	-509.71	-593.98	-1783.62
21	-440.19	-338.75	-390.28	-435.51	-479.80	-1644.34

Tables 14 through 17 give same indexes calculated when only X_1 and X_2 are used as covariates in the models for forecasting. This is done to understand the effect of assumed variances on results. Here, the results for other methods have not been given again since there were no changes for them; however, they were used for all comparisons. There are small changes for the accuracy proportions between the methods with for all X's and only X_1 and X_2. The results can be summarized in general as follows. Random effects models without lag of response give the best result for PCP, ePCP and AUROC for model fitting. In addition, transition models are the best and random effects models without lag of response is the second method for forecasting in terms of PCP, AUROC and deviance, while in terms of ePCP, random effects models without lag of response is the best and transition model is the second method. For all indexes, marginal models are the worst one for fitting the model while marginalized transition models is the worst one for forecasting.

Table 14. Values and confidence intervals for proportion of correct predictions (PCP) with only $X_{k=1}$ and $X_{k=2}$ as covariates

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
14	0.720 (0.695,0.746)	0.694 (0.642,0.746)	0.669 (0.616,0.723)	0.644 (0.589,0.698)	0.616 (0.561,0.671)	0.656 (0.629,0.683)
15	0.858 (0.836,0.881)	0.857 (0.817,0.896)	0.808 (0.764,0.853)	0.749 (0.700,0.798)	0.695 (0.643,0.747)	0.777 (0.754,0.801)
16	0.861 (0.833,0.888)	0.848 (0.808,0.888)	0.797 (0.752,0.842)	0.738 (0.688,0.788)	0.691 (0.639,0.743)	0.769 (0.745,0.792)
17	0.865 (0.826,0.903)	0.851 (0.811,0.891)	0.801 (0.756,0.846)	0.742 (0.692,0.791)	0.693 (0.641,0.745)	0.772 (0.748,0.796)
18	0.824 (0.803,0.846)	0.526 (0.470,0.582)	0.527 (0.470,0.583)	0.521 (0.465,0.578)	0.514 (0.458,0.571)	0.522 (0.494,0.550)
19	0.828 (0.807,0.850)	0.537 (0.481,0.593)	0.538 (0.482,0.594)	0.530 (0.473,0.586)	0.525 (0.468,0.581)	0.532 (0.504,0.561)
20	0.934 (0.920,0.948)	0.827 (0.784,0.869)	0.772 (0.724,0.819)	0.715 (0.664,0.766)	0.669 (0.616,0.722)	0.746 (0.721,0.770)
21	0.864 (0.841,0.886)	0.558 (0.502,0.614)	0.537 (0.481,0.593)	0.522 (0.466,0.578)	0.512 (0.456,0.568)	0.532 (0.504,0.560)

Table 15. Values and confidence intervals for expected percent correctly predicted (ePCP) with only $X_{k=1}$ and $X_{k=2}$ as covariates

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
14	0.621 (0.594,0.648)	0.599 (0.543,0.655)	0.579 (0.523,0.635)	0.561 (0.505,0.617)	0.545 (0.489,0.601)	0.571 (0.543,0.599)
15	0.784 (0.757,0.811)	0.775 (0.728,0.822)	0.736 (0.686,0.786)	0.691 (0.638,0.744)	0.649 (0.595,0.703)	0.713 (0.687,0.739)
16	0.787 (0.755,0.819)	0.775 (0.728,0.822)	0.740 (0.691,0.789)	0.691 (0.638,0.744)	0.652 (0.599,0.705)	0.715 (0.689,0.741)
17	0.796 (0.750,0.842)	0.781 (0.735,0.827)	0.745 (0.696,0.794)	0.696 (0.644,0.748)	0.656 (0.602,0.710)	0.720 (0.694,0.746)
18	0.741 (0.716,0.766)	0.396 (0.341,0.451)	0.412 (0.356,0.468)	0.430 (0.374,0.486)	0.442 (0.386,0.498)	0.420 (0.392,0.448)
19	0.749 (0.724,0.774)	0.392 (0.337,0.447)	0.412 (0.356,0.468)	0.427 (0.371,0.483)	0.441 (0.385,0.497)	0.418 (0.390,0.446)
20	0.877 (0.858,0.896)	0.789 (0.743,0.835)	0.741 (0.692,0.790)	0.690 (0.637, 0.743)	0.649 (0.595,0.703)	0.717 (0.692,0.742)
21	0.793 (0.766,0.820)	0.578 (0.522,0.634)	0.561 (0.505,0.617)	0.547 (0.491,0.603)	0.534 (0.478,0.590)	0.555 (0.527,0.583)

Table 16. Values for area under the receiver operating characteristic (AUROC) curve with only $X_{k=1}$ and $X_{k=2}$ as covariates

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
14	0.797	0.764	0.733	0.700	0.664	0.717
15	0.915	0.889	0.836	0.778	0.723	0.809
16	0.924	0.901	0.837	0.765	0.715	0.805
17	0.927	0.899	0.834	0.769	0.717	0.806
18	0.899	0.209	0.243	0.286	0.325	0.266
19	0.905	0.217	0.256	0.300	0.337	0.278
20	0.985	0.893	0.836	0.770	0.715	0.806
21	0.928	0.666	0.641	0.619	0.595	0.631

Table 17. Values for deviance with only $X_{k=1}$ and $X_{k=2}$ as covariates

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
14	-1307.48	-346.47	-362.61	-377.73	-391.23	-1478.04
15	-655.78	-242.84	-306.32	-375.21	-436.91	-1361.27
16	-414.39	-236.25	-317.90	-406.39	-465.46	-1425.99
17	-202.93	-239.06	-324.97	-410.46	-473.01	-1447.49
18	-963.77	-4511.83	-4444.43	-4198.01	-4017.95	-17172.22
19	-936.83	-4220.54	-4068.16	-3925.89	-3719.78	-15934.37
20	-409.74	-277.71	-373.72	-474.34	-548.97	-1674.53
21	-608.23	-476.09	-501.55	-523.48	-544.36	-2045.48

Tables 18 through 21 give the results for the methods including fixed covariates to eliminate the effect of forecast error in independent variables. For this, we assumed that the all four X's do not change within time. That is, we take the first year observation for all other time points. When the X's are taken as fixed, it seems that the results get worse in general. The results can be summarized in general as follows. Random effects models without lag of response give the best result for PCP, ePCP and AUROC for model fitting whereas marginal models are the last one in terms of all indexes. Furthermore, transition models and AR models are the best models in terms of all indexes while random effects models without lag of response follows them for forecasting. Marginalized transition model is the worst one in terms of ePCP, AUROC and deviance while it shares the last place with random effects model with lag of response having a small difference in terms of PCP.

Table 18. Values and confidence intervals for proportion of correct predictions (PCP) with fixed X's

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
14	0.737 (0.712,0.762)	0.648 (0.595,0.702)	0.626 (0.571,0.681)	0.592 (0.536,0.647)	0.571 (0.515,0.627)	0.609 (0.582,0.637)
15	0.857 (0.835,0.880)	0.858 (0.819,0.897)	0.809 (0.765,0.853)	0.750 (0.701,0.798)	0.696 (0.644,0.748)	0.778 (0.755,0.802)
16	0.857 (0.829,0.885)	0.854 (0.814,0.894)	0.805 (0.760,0.849)	0.745 (0.696,0.794)	0.694 (0.642,0.746)	0.774 (0.751,0.798)
17	0.860 (0.821,0.899)	0.855 (0.815,0.894)	0.806 (0.762,0.851)	0.746 (0.696,0.795)	0.695 (0.643,0.747)	0.775 (0.752,0.799)
18	0.839 (0.818,0.860)	0.538 (0.482,0.594)	0.532 (0.475,0.588)	0.523 (0.467,0.580)	0.517 (0.460,0.573)	0.527 (0.499,0.556)
19	0.837 (0.816,0.858)	0.546 (0.490,0.602)	0.539 (0.482,0.595)	0.530 (0.474,0.587)	0.525 (0.469,0.582)	0.535 (0.507,0.563)
20	0.896 (0.879,0.913)	0.778 (0.731,0.825)	0.729 (0.679,0.780)	0.681 (0.628,0.734)	0.643 (0.589,0.697)	0.708 (0.682,0.734)
21	0.858 (0.835,0.881)	0.526 (0.470,0.583)	0.523 (0.466,0.579)	0.517 (0.461,0.574)	0.514 (0.457,0.570)	0.520 (0.492,0.548)

Table 19. Values and confidence intervals for expected percent correctly predicted (ePCP) with fixed X's

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
14	0.653 (0.626,0.680)	0.600 (0.544,0.655)	0.585 (0.529,0.641)	0.563 (0.506,0.619)	0.549 (0.492,0.605)	0.574 (0.546,0.602)
15	0.772 (0.745,0.800)	0.763 (0.715,0.811)	0.727 (0.677,0.777)	0.682 (0.629,0.735)	0.643 (0.588,0.697)	0.704 (0.678,0.730)
16	0.771 (0.738,0.805)	0.766 (0.718,0.814)	0.736 (0.686,0.786)	0.689 (0.636,0.741)	0.649 (0.595,0.703)	0.710 (0.684,0.736)
17	0.777 (0.730,0.824)	0.770 (0.722,0.818)	0.739 (0.689,0.788)	0.690 (0.637,0.742)	0.650 (0.596,0.704)	0.712 (0.687,0.738)
18	0.747 (0.722,0.771)	0.408 (0.352,0.464)	0.417 (0.361,0.473)	0.434 (0.378,0.490)	0.445 (0.389,0.501)	0.426 (0.398,0.454)
19	0.751 (0.727,0.775)	0.401 (0.345,0.456)	0.417 (0.361,0.472)	0.432 (0.376,0.488)	0.446 (0.390,0.502)	0.424 (0.396,0.452)
20	0.839 (0.818,0.860)	0.751 (0.702,0.800)	0.710 (0.658,0.761)	0.665 (0.611,0.718)	0.631 (0.577,0.686)	0.689 (0.663,0.715)
21	0.777 (0.750,0.804)	0.553 (0.497,0.609)	0.545 (0.489,0.601)	0.535 (0.479,0.591)	0.527 (0.471,0.583)	0.540 (0.512,0.568)

Table 20. Values for area under the receiver operating characteristic (AUROC) curve with fixed X's

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
14	0.814	0.704	0.674	0.628	0.599	0.651
15	0.892	0.872	0.819	0.751	0.697	0.784
16	0.901	0.886	0.818	0.744	0.692	0.783
17	0.903	0.884	0.811	0.744	0.693	0.781
18	0.889	0.269	0.283	0.325	0.362	0.310
19	0.893	0.263	0.288	0.333	0.368	0.313
20	0.972	0.859	0.802	0.736	0.688	0.771
21	0.914	0.637	0.616	0.589	0.569	0.603

Table 21. Values for deviance with fixed X's

Time / Method	t=1,2,3,4	t=5	t=6	t=7	t=8	t=5,6,7,8
14	-1255.72	-399.10	-423.67	-462.02	-486.30	-1771.09
15	-711.15	-248.22	-308.28	-382.33	-445.86	-1384.69
16	-455.73	-238.63	-312.84	-395.53	-460.41	-1407.40
17	-223.83	-242.43	-322.16	-402.77	-467.75	-1435.12
18	-986.07	-3695.33	-3873.97	-3644.60	-3472.82	-14686.73
19	-972.13	-4121.62	-4130.84	-3971.85	-3764.24	-15988.55
20	-538.29	-322.15	-418.98	-532.61	-616.33	-1890.06
21	-658.38	-499.18	-515.87	-536.00	-552.16	-2103.21

CHAPTER 6

CONCLUSION AND DISCUSSION

Although panel data has some advantages compared to other types of data, because of its complexity it is difficult to analyze. Moreover, when the response variable is binary, the number of studies are more limited compared to Gaussian models. On the other hand, considering its increasing importance and increasing fundamental roles in the literature, further efforts are needed in this area. Furthermore, forecasting, another important term, is not generally used in longitudinal data especially for binary response. Therefore, in this thesis, forecasting in longitudinal binary response data is studied. For this purpose, a simulation study is done for panel data with binary response. R programming language is used for generating a data set having high autocorrelation within variables, medium correlation between independent and dependent variables and low correlation between independent variables. 4 continuous and one binary response variables for 300 subjects on 8 different equally spaced time points are generated. 21 different methods, including naïve and complex ones, are set for modeling and forecasting for the binary response variable with these variables.

Although Makridakis and Hibon (2000) stated that the complex models do not mean more accurate forecasts, transition models and random effects models, which are one of the most complex models, give the best output in this thesis. On the other hand, naïve methods have a better position than some complex methods such as marginal models and marginalized transition models. Furthermore, Makridakis and Hibon (2000) commented that different accuracy measures change the ranking of the performance of the model. This is true for some methods in this thesis, as well. For example, random effects models is in the first place for the model fitting in all criteria except deviance. Specifically, random effects models without lag of response is the best method for model fitting. Transition models follow it in the second place. AR and marginalized transition models have a changing place as in the third and fourth place according to the criteria. Marginal models are in the last place in all cases. For the ranking of forecast accuracy, random effects models having no response are in the first place in terms of ePCP and AUROC. However, AR is in the first place for PCP and

deviance. Transition models are in the second place in all cases while it shares the first place with random effects models having no response in terms of AUROC and with AR models in terms of deviance. Marginalized transition models are in the worst place with almost all criteria. Although some changes appear in the ranking, transition models and random effects models having no lag of response and are the best models in most situations.

In addition, when we compare the results for methods which includes all covariates with the ones for only X_1 and X_2 covariates to understand the effect of variances, it can be stated that the variances of the covariates do not have big effects on the ranking in terms of both modeling and forecasting. In general the ranking is the same with smaller accuracy values. Moreover, when the results of methods having fixed X's are compared to eliminate the effects of forecast errors for covariates, it can be summarized that random effects models having no lag of response is the best model for model fitting while transition models and AR models are the best models for forecasting. Random effects models follow them as in third place for all indexes. Marginal models and marginalized transition models are the worst ones for model fitting and forecasting, respectively. Furthermore, when we use the real 8 values of X's, which is not meaningful in real life examples, for forecasting to eliminate the effects of forecast errors for covariates again, it can be stated that all the accuracy number increased but the ranking did not change. Since the values decreases and increases when fixed and real X's are used respectively, it can be said that models take some power from the changes in X's when the X's are time dependent. Here, another interesting result is that marginalized transition models give the worst results for forecasting although it gives better results for model fitting in all trials. We believe, this is because the calculated values for Δ's do not have a specific trend.

Furthermore, it was expressed in Carrière and Bouyer (2002) that random models are more suitable than marginal models despite of their complexity. Moreover, the simulation results done by Baillie and Baltagi (1999) validated that individual effects are important for forecasting. As a final remark, it can be concluded that for forecasting new values, transition models and random effects models having no lag of response are the best models in almost most of the situations although some changes appears in the ranking.

These models deserve this achievement by taking the each individual's effect into account in the model. However, to be able to generalize this result, these methods should be tried under different conditions. For this purpose, 27 different combinations, such as high autocorrelation within dependent variable, less autocorrelation in independent variables, high correlation between dependent and independent variables, were tried to be generated in order to understand which method gives the best results under different correlation structures. In other words, we aimed to investigate whether the value for correlation between

the variables has an effect on the results or not. However, except a few ones, needed correlation structures could not be constituted. Therefore, it is decided to analyze one of these conditions which is high autocorrelation, no multicollinearity and medium correlation between dependent and independent variables. However, the other few conditions could also be simulated as a future work.

One also needs to notice that the frequency of probability of the response being 1 is almost 0.5 here. That is, data is generated so that $P(Y=1) \cong 0.5$ and $P(Y=0) \cong 0.5$. This is not a rare disease type of event but can be considered as a case-control situation. Therefore, results can be used for case-control studies or similar situations.

Furthermore, Fitzmaurice et al. (2009) asserted that in longitudinal data, correlation usually decreases with increasing time lag and variance is instable across time. However, we use the same variance across time in this thesis to avoid the complexity. That is, if the variance is taken unstable, then different parameters might be needed in models for different time points. Since it is a disadvantage for simulation, we prefer to use stable variances.

Finally, it is observed that panel data does not have the same meaning in different disciplines in the literature. For example, in econometric literature, a data can be called panel when it has 20 time points although it is not the case in biostatistics. This difference occurs also for the models used in panel data. For instance, fixed effects, random effects and dynamic models are common terms in econometrics literature. However, all fixed effect models cannot be used in panel data in biostatistics because the coefficients cannot be estimated efficiently. Instead, marginal models which are one of the possible fixed effect models having the constant intercept and slope coefficients for each subject can be used. Furthermore, when the lag of response is included in the model, it is called dynamic model in econometrics literature while it is called transition model in biostatistics. For the estimations in dynamic models, vector autoregression (VAR) and autoregressive moving average (ARMA) analysis are used but they cannot be used with less time points, i.e. with transition models. Therefore, the studies in econometric literature could not be used here although the name is the same. If this study will be enlarged by increasing the number of time points as 10 or 15 as a future work, then the econometric literature can also be discussed and compared. However, we preferred to use less time points regarding the question that how forecasting can be done especially for the situations in biostatistics or medical sciences in Turkey.

REFERENCES

Azzalini, A. (1994), “Logistic Regression for Autocorrelated Data with Application to Repeated Measures”, *Biometrika*, 81(4), 767-775.

Baadsgaard, N.P., Højsgaard, S., Gröhn, Y.T. and Schukken, Y.H. (2004), “Forecasting clinical disease in pigs: comparing a naive and a Bayesian approach”, *Preventive Veterinary Medicine*, 64, 85–100.

Baillie R.T. and Baltagi B.H. (1999), Prediction from the regression model with one-way error components. In *Analysis of Panels and Limited Dependent Variable Models*, Hsiao C, Lahiri K, Lee LF, Pesaran H (eds). Cambridge University Press: Cambridge, 255–267.

Ballinger, G.A. (2004), “Using Generalized Estimating Equations for Longitudinal Data Analysis”, *Organizational Research Methods,* 7, 127-150.

Baltagi, B.H. (2001), *Econometric Analysis of Panel Data*, 2nd ed., West Sussex: John Wiley & Sons, Ltd.

Baltagi, B.H. and Li, D. (2006), *"Prediction in the Panel Data Model with Spatial Correlation: The Case of Liquor",* Center for Policy Research Working Papers 84, Center for Policy Research, Maxwell School, Syracuse University.

Baltagi, B.H. (2008), “Forecasting with Panel Data”, *Journal of Forecasting,* 27, 153–173.

Bontemps, C., Racine, J.S. and Simioni, M. (2009), “Nonparametric vs parametric binary choice models: An empirical investigation”, Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2009 AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin.

Bosq, D. and Blanke D. (2007), *Inference and prediction in large dimensions*, John Wiley & Sons Ltd., West Sussex England.

Carrière I. and Bouyer J. *(*2002), “Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disability among older persons”, *BMC Medical Research Methodology,* 2:15.

Comstock B. A. and Heagerty P.J., *Introduction to mtm: an R package for marginalized transition models*, Retrieved from http://faculty.washington.edu/heagerty/Software/LDA/MTM/MTMhelp.pdf, 01/05/2010.

Cullis, B.R. and McGilchrist, C.A. (1990), “A model for the analysis of growth data from designed experiments”, *Biometrics,* 46, 131-142.

Dayananda D., Irons, R., Harrison, S., Herbohn, J. and Rowland P. (2002), *Capital Budgeting: Financial Appraisal of Investment Projects,* Cambridge University Press

Diggle, P.J., Heagerty, P.J., Liang, K.Y., and Zeger, S.L. (2002), *Analysis of Longitudinal Data*, 2nd ed., Oxford: Oxford University Press.

Dziak, J. J. and Li, R. (2007), *An overview on variable selection for longitudinal data*, In D. Hong (Ed.), Singapore: World Sciences Publisher.

Dougherty, C. (2007), *Introduction to Econometrics,* 3rd ed., Oxford: Oxford University Press.

Egan, J. P. (1975), "*Signal Detection Theory and ROC Analysis*, Vol. 195, New York: Academic Press.

Efron, B. (1978), "Regression and ANOVA with zero-one data:Measures of residual variation", *Journal of American Statistical Association,* 73, 112-121.

Fitzmaurice, G. M. and Laird, N. M. (1993), "A likelihoodbased method for analysing longitudinal binary responses", *Biometrika,* 80, 141-151.

Fitzmaurice, G. M. (1995), "A caveat concerning independence estimating equations with multivariate binary data", *Biometrics*, *51*, 309-317.

Fitzmaurice, G.M, Laird, N.M. and Ware, J.H. (2004), *Applied Longitudinal Analysis*, New Jersey: John Wiley & Sons, Inc.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenbergs, G. (2009), *Longitudinal Data Analysis,* New York: Taylor & Francis Group, LLC.

Frees, E.W. (2004), *Longitudinal and Panel Data Analysis and Applications in the Social Sciences,* New York: Cambridge University Press.

Fulton, J.T. (1954), "Longitudinal Data on Eruption and Attack of the Permanent Teeth", *Journal of Dental Research*, 33(1), 65-79.

Gardiner, J.C., Luo, Z. and Roman, L.A. (2009), "Fixed effects, random effects and GEE: What are the differences?" *Statisistics in Medicine,* 28, 221–239.

Golder, M., *Binary Response Models,* Retrieved from http://homepages.nyu.edu/~mrg217/binaryresponse.pdf, 01/05/2009

Gujarati, D. N. (2003), *Basic Econometrics,* 4th ed., New York: McGraw-Hill.

Hardin, J.W., and Hilbe, J. M. (2003), "*Generalized estimating equations",* Boca Raton, FL: Chapman and Hall/CRC Press.

Hastie, T.J. and Tibshirani, R.J. (1990), *Generalized additive models.* Chapman and Hall, New York.

Haveman H.A. and Nonnemaker L. (2000), "Competition in multiple geographic markets: the impact on growth and market entry", *Administrative Science Quarterly*, 45(2), 232-267.

Heagerty, P.J. (1999), "Marginally specified logistic-normal models for longitudinal binary data", *Biometrics*, 55, 688-698.

Heagerty, P.J. (2002), "Marginalized Transition models and likelihood inference for longitudinal categorical data", *Biometrics,* 58, 342-351.

Herron, M. (1999), "Postestimation Uncertainty in Limited Dependent Variable Models", *Political Analysis,* 8, 83-98.

Horowitz, J.L. and Savin, N.E. (2001), "Binary Response Models: Logits, Probits and Semiparametrics*", Journal of Economic Perspectives*, 15(4), 43–56.

Horrocks, J. and van Den Heuvely, M.J. (2009) *"Prediction of Pregnancy: A Joint Model for Longitudinal and Binary Data"*, Bayesian Analysis, 4(3), 523-538.

Hsiao, C. (2003), *Analysis of Panel Data,* 2nd ed., Cambridge: Cambridge University Press.

Hyndman, R. J. (2010), *Forecasting overview*, In: International Encyclopedia of Statistical Science, Springer.

Ilk, O. (2008), *Multivariate Longitudinal Data Analysis: Models for Binary Response and Exploratory Tools for Binary and Continuous Response,* Verlag Dr. Muller (VDM).

Jackman, S. (2007), *Models for Binary Outcomes and Proportions,* Retrieved from *http://jackman.stanford.edu/classes/350C/07/binary1.pdf*, 01/02/2009.

Johnson R. A. and Wichern D. W. (1998), *Applied Multivariate Statistical Analysis,* 4th ed., Prentice-Hall, Inc.

Kaufmann, H. (1987), "Regression models for nonstationary categorical time series: asymptotic estimation theory", *Annals of Statistics,* 15, 863-871.

Korn, E.L. and Whittemore, A.S. (1979), "Methods for analyzing panel studies of acute health effects of air pollution", *Biometrics,* 35, 795-802.

Korn E. L. and Simon R. (1991), "Explained residual variation, explained risk and goodness of fit", *The American Statistician,* 45(3), 201-206.

Lam P. (2007), *logit.gee: General Estimating Equation for Logistic Regression*, Retrieved from http://cran.r-project.org/web/packages/Zelig/vignettes/logit.gee.pdf, 01/01/2009.

Liang, K.Y and Zeger, S.L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models", *Biometrics*, 73(1), 13-22.

Lipsitz, S., Laird, N. and Harrington, D. (1991), "Generalized Estimating Equations for Correlated Binary Data: Using Odds Ratios as a Measure of Association", *Biometrika*, 78, 153-160.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1982), "The accuracy of extrapolation (time series) methods: results of a forecasting competition", *Journal of Forecasting,* 1, 111–153.

Makridakis, S., and Hibon, M. (1979), "Accuracy of forecast ing: an empirical investigation", *Journal of the Royal Statistical Society"* 142, 97–145.

Makridakis, S. and Hibon M. (2000), "The M3-Competition: results, conclusions and implications", *International Journal of Forecasting*, 16, 451–476.

Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman W. (1996), *Applied Linear Statistical Models,* 4th ed., Irwin Publication.

Pan, W. (2001), "Akaike's Information Criterion in Generalized Estimating Equations", *Biometrics*, 57(1), 120-125.

Pan, W. (2002), "Goodness-of-fit Tests for GEE with Correlated Binary Data", *Scandinavian Journal of Statistics,* 29, 101-110.

Park, T., Davis, C.S. and Lid, N. (1998), "Alternative Gee estimation procedures for discrete longitudinal data", *Computational Statistics & Data Analysis,* 28, 243-256.

Patterson, H.D. and Thompson, R. (1971), “Recovery of inter-block information when block sizes are unequal”, *Biometrika*, 58, 545-554.

Pepe, M.S., Heagerty, P. and Whitaker, R. (1999), “Prediction Using Partly Conditional Time-Varying Coefficients Regression Models” *Biometrics,* 55, 944-950.

Plewis, I. (2007), *Predictions, explanations and causal effects from longitudinal data*, London: Institute of Education, University of London.

Pourahmadi M. (2001), *Foundation of time series analysis and prediction theory*, Willey series in probability and statistics, John Wiley and Sons, Inc.

Roberts, P. (2003), “Moving Averages: The Heart of Trend Analysis”, *The London Bullion Market Association*, *Alchemist,* 33, 12-14.

Rosenberg, M.A., Frees, E.W., Sun, J. and Johnson, P.H. and Robinson, J.M. (2008), “*Predictive Modeling with Longitudinal Data: A Case Study of Wisconsin Nursing Homes” Retrieved from http://www.soa.org/library/journals/north-american-actuarial-journal/2007/july/naaj0703-3.pdf*, 01/05/2010.

Sohn, S.Y. and Kim, H.S. (2007), “Random effects logistic regression model for default prediction of technology credit guarantee fund*”, European Journal of Operational Research,* 183, 472–478.

Song, P.X.-K. (2007), *Correlated data analysis: Modeling, Analytics, and Applications, New York:* Springer Science+Business Media, LLC.

Sparling, D., *Forecasting,* Retrieved from http://www.uoguelph.ca/~dsparlin/forecast.htm, 01/02/2009.

Stiratelli, R., Laird, N. and Ware, J.H. (1984), “Random-Effects Models for Serial Observations with Binary Response”, *Biometrics*, 40(4), 961-971.

Terza J.V. (2006), “Optimal discrete prediction in parametric binary response models”, *Economics Letters*, 91, 72–75.

Tian, L. (2007), “Model evaluation based on the sampling distribution of estimated absolute prediction error” *Biometrika*, 94(2), 297–311.

Torres-Reyna, O., *Panel Data Analysis Fixed & Random Effects (ver. 3.0)* [Power Point slides] Retrieved from http://dss.princeton.edu/training/Panel101.pdf, 01/04/2009.

Troccoli A., Harrison M., Anderson D. L.T. and Mason S. J. (2008), “Seasonal Climate: Forecasting and Managing Risk”, *IV. Earth and Environmental Sciences,* 82, 241-264.

Verbeke, G. and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer.

Verbyla, A.P. and Cullis, B.R. (1990), “Modeling in repeated measures experiments”, *Applied Statistics,* 39, 341-356.

Ware, J.H., Lipsitz, S. and Speizer, F.E. (1988), “Issues in the analysis of repeated categorical outcomes”, *Statistics in Medicine,* 7, 95-107.

Wawro G. (2002), “Estimating Dynamic Panel Data Models in Political Science”, *Political Analysis*, 10, 25-48.

Winkelmann, R. and Boes, S. (2009), *Analysis of micro data*, 2nd ed., Springer.

Wong, W.H. (1986), “Theory of partial likelihood”, *Annals of Statistics,* 14, 88-123.
Yaffee, R.A. (2003), A *Primer for Panel Data Analysis,* Retrieved from http://www.nyu.edu/its/statistics/Docs/pda.pdf, 01/05/2009.

Zeger, S. L. and Liang, K.-Y. (1986), “Longitudinal data analysis for discrete and continuous outcomes”, *Biometrics*, *42*, 121-130.

Zeger, S.L. and Qaqish, B. (1988), “Markov regression models for time series: a quasi-likelihood approach”, *Biometrics,* 44, 1019-1031.

Zheng, B. (2000), “Summarizing the goodness of fit on generalized linear models for longitudinal data” *Statistics in Medicine*, *19*, 1265-1275.

Zorn, C. (2001), “Generalized Estimating Equation Models for Correlated Data: A Review with Applications”, *American Journal of Political Science*, 45, 470–490.

APPENDIX A

R CODES FOR DATA GENERATION PROCESS

```
# setting sample size
n=300

# introducing the correlation matrix that is given by the user
cor=matrix(scan("multinormal_xyuksek_yyuksek_xyyuksek_5.txt"),ncol=40,byrow=T)

# setting variance values and calculating sigma to be able to generate data from multivariate
normal distribution
k=c(0.25,5,5,50,500,0.25,5,5,50,500,0.25,5,5,50,500,0.25,5,5,50,500,0.25,5,5,50,500,0.25,5
,5,50,500,0.25,5,5,50,500,0.25,5,5,50,500)
k=sqrt(k)
sd=diag(k,40)
sigma=sd%*%cor%*%sd

# generating data from multivariate normal distrubition
data=rmvnorm(n, mean = rep(0, nrow(sigma)), sigma=sigma,method="svd")

# taking values of response variable and transforming them to binary
b=cbind(data[,1],data[,6],data[,11],data[,16],data[,21],data[,26],data[,31],data[,36])
y=(exp(b)/(1+exp(b)))
yhat=ifelse(y>=0.5,1,0)

# taking values of independent variables
x=cbind(data[,2:5],data[,7:10],data[,12:15],data[,17:20],data[,22:25],data[,27:30],data[,32:35],
data[,37:40])

# data which is composed of binary response and continuous independent variables
datanew=cbind(yhat[,1],data[,2:5],yhat[,2],data[,7:10],yhat[,3],data[,12:15],yhat[,4],data[,17:2
0],yhat[,5],data[,22:25],yhat[,6],data[,27:30],yhat[,7],data[,32:35],yhat[,8],data[,37:40])

# taking the upper triangle values of correlation matrix
cor_new=cor(datanew, use="pairwise.complete.obs", method="spearman")
cor_new[lower.tri(cor_new)]=NA
cor_neww=c(cor_new)
cor_newww=na.exclude(cor_neww)[1:820]
cor_xy=rbind(cor_xy,cor_newww)
```

APPENDIX B

Table 22. Average Correlation Matrix[2]

	Y_1	X_{11}	X_{12}	X_{13}	X_{14}	Y_2	X_{21}	X_{22}	X_{23}	X_{24}
Y_1	1.000	0.399	0.400	0.428	0.430	0.712	0.361	0.360	0.383	0.385
X_{11}		1.000	0.240	0.238	0.238	0.356	0.836	0.211	0.209	0.210
X_{12}			1.000	0.238	0.238	0.356	0.211	0.836	0.210	0.209
X_{13}				1.000	0.239	0.381	0.210	0.209	0.835	0.210
X_{14}					1.000	0.383	0.210	0.210	0.210	0.836
Y_2						1.000	0.401	0.400	0.426	0.429
X_{21}							1.000	0.239	0.238	0.239
X_{22}								1.000	0.238	0.238
X_{23}									1.000	0.238
X_{24}										1.000

[2] X_{12} stands for the second covariate at time 1

Table 22. Continued

	Y_3	X_{31}	X_{32}	X_{33}	X_{34}	Y_4	X_{41}	X_{42}	X_{43}	X_{44}
Y_1	0.625	0.318	0.318	0.340	0.342	0.501	0.272	0.273	0.291	0.292
X_{11}	0.314	0.736	0.173	0.171	0.172	0.269	0.652	0.144	0.142	0.143
X_{12}	0.313	0.173	0.736	0.171	0.171	0.269	0.144	0.652	0.143	0.143
X_{13}	0.338	0.172	0.171	0.736	0.171	0.289	0.143	0.142	0.652	0.142
X_{14}	0.341	0.172	0.172	0.172	0.737	0.292	0.143	0.143	0.143	0.652
Y_2	0.717	0.358	0.358	0.381	0.383	0.621	0.317	0.317	0.339	0.340
X_{21}	0.358	0.836	0.211	0.210	0.210	0.317	0.736	0.173	0.171	0.172
X_{22}	0.358	0.210	0.836	0.209	0.209	0.317	0.173	0.736	0.171	0.171
X_{23}	0.381	0.210	0.210	0.835	0.210	0.339	0.171	0.171	0.736	0.171
X_{24}	0.383	0.210	0.210	0.210	0.836	0.341	0.172	0.172	0.171	0.736
Y_3	1.000	0.399	0.399	0.426	0.428	0.717	0.357	0.357	0.381	0.383
X_{31}		1.000	0.240	0.238	0.239	0.357	0.836	0.211	0.210	0.210
X_{32}			1.000	0.238	0.238	0.358	0.211	0.836	0.210	0.210
X_{33}				1.000	0.239	0.381	0.209	0.210	0.836	0.210
X_{34}					1.000	0.383	0.210	0.210	0.210	0.836
Y_4						1.000	0.399	0.400	0.426	0.429
X_{41}							1.000	0.240	0.238	0.239
X_{42}								1.000	0.238	0.239
X_{43}									1.000	0.238
X_{44}										1.000
Y_5										
X_{51}										
X_{52}										
X_{53}										
X_{54}										
Y_6										
X_{61}										
X_{62}										
X_{63}										
X_{64}										
Y_7										
X_{71}										
X_{72}										
X_{73}										
X_{74}										
Y_8										
X_{81}										
X_{82}										
X_{83}										
X_{84}										

Table 22. Continued

	Y_5	X_{51}	X_{52}	X_{53}	X_{54}	Y_6	X_{61}	X_{62}	X_{63}	X_{64}
Y_1	0.393	0.223	0.224	0.238	0.239	0.344	0.190	0.190	0.203	0.204
X_{11}	0.219	0.534	0.114	0.113	0.114	0.187	0.447	0.095	0.094	0.095
X_{12}	0.220	0.115	0.534	0.114	0.114	0.187	0.096	0.447	0.095	0.094
X_{13}	0.236	0.114	0.114	0.534	0.114	0.202	0.095	0.094	0.447	0.095
X_{14}	0.238	0.115	0.115	0.114	0.534	0.205	0.095	0.096	0.095	0.447
Y_2	0.500	0.271	0.271	0.289	0.291	0.394	0.222	0.222	0.236	0.237
X_{21}	0.271	0.652	0.143	0.142	0.143	0.222	0.534	0.114	0.114	0.114
X_{22}	0.272	0.144	0.652	0.142	0.142	0.222	0.115	0.535	0.113	0.113
X_{23}	0.290	0.143	0.142	0.651	0.143	0.237	0.114	0.114	0.533	0.114
X_{24}	0.291	0.143	0.143	0.142	0.651	0.238	0.114	0.114	0.113	0.533
Y_3	0.620	0.317	0.317	0.338	0.340	0.501	0.271	0.271	0.289	0.291
X_{31}	0.316	0.736	0.172	0.171	0.172	0.270	0.652	0.143	0.142	0.143
X_{32}	0.316	0.172	0.736	0.171	0.171	0.271	0.144	0.652	0.142	0.142
X_{33}	0.338	0.171	0.171	0.735	0.172	0.289	0.143	0.143	0.651	0.143
X_{34}	0.340	0.172	0.172	0.171	0.736	0.291	0.143	0.143	0.142	0.651
Y_4	0.717	0.357	0.357	0.381	0.384	0.619	0.316	0.316	0.338	0.340
X_{41}	0.357	0.835	0.210	0.209	0.210	0.316	0.736	0.172	0.171	0.172
X_{42}	0.358	0.211	0.835	0.209	0.210	0.317	0.172	0.736	0.171	0.171
X_{43}	0.381	0.209	0.209	0.835	0.210	0.338	0.171	0.171	0.735	0.172
X_{44}	0.383	0.210	0.210	0.209	0.836	0.340	0.171	0.172	0.171	0.736
Y_5	1.000	0.399	0.400	0.426	0.428	0.717	0.357	0.357	0.381	0.383
X_{51}		1.000	0.239	0.238	0.239	0.357	0.835	0.210	0.209	0.210
X_{52}			1.000	0.238	0.238	0.357	0.210	0.836	0.209	0.209
X_{53}				1.000	0.238	0.381	0.210	0.209	0.835	0.210
X_{54}					1.000	0.382	0.210	0.210	0.209	0.835
Y_6						1.000	0.398	0.399	0.426	0.428
X_{61}							1.000	0.239	0.238	0.238
X_{62}								1.000	0.238	0.238
X_{63}									1.000	0.238
X_{64}										1.000
Y_7										
X_{71}										
X_{72}										
X_{73}										
X_{74}										
Y_8										
X_{81}										
X_{82}										
X_{83}										
X_{84}										

Table 22. Continued

	Y_7	X_{71}	X_{72}	X_{73}	X_{74}	Y_8	X_{81}	X_{82}	X_{83}	X_{84}
Y_1	0.234	0.142	0.141	0.150	0.150	0.203	0.106	0.106	0.116	0.117
X_{11}	0.138	0.354	0.066	0.066	0.067	0.106	0.263	0.037	0.037	0.038
X_{12}	0.138	0.067	0.354	0.067	0.066	0.106	0.038	0.262	0.038	0.037
X_{13}	0.149	0.066	0.066	0.354	0.067	0.116	0.038	0.037	0.263	0.038
X_{14}	0.150	0.067	0.067	0.066	0.354	0.118	0.039	0.038	0.038	0.263
Y_2	0.340	0.191	0.190	0.203	0.204	0.234	0.139	0.138	0.149	0.149
X_{21}	0.190	0.447	0.095	0.095	0.095	0.142	0.354	0.066	0.066	0.067
X_{22}	0.190	0.096	0.447	0.095	0.094	0.142	0.067	0.354	0.066	0.066
X_{23}	0.203	0.095	0.095	0.446	0.096	0.151	0.067	0.066	0.354	0.067
X_{24}	0.204	0.096	0.095	0.094	0.446	0.151	0.067	0.066	0.066	0.354
Y_3	0.395	0.222	0.222	0.237	0.238	0.344	0.188	0.187	0.202	0.204
X_{31}	0.221	0.534	0.114	0.114	0.114	0.190	0.447	0.095	0.095	0.095
X_{32}	0.222	0.115	0.534	0.114	0.114	0.190	0.096	0.447	0.095	0.095
X_{33}	0.237	0.114	0.115	0.534	0.115	0.204	0.095	0.095	0.447	0.096
X_{34}	0.238	0.115	0.114	0.114	0.533	0.205	0.096	0.095	0.095	0.447
Y_4	0.501	0.272	0.271	0.290	0.291	0.393	0.220	0.220	0.236	0.238
X_{41}	0.271	0.652	0.143	0.142	0.143	0.223	0.534	0.114	0.114	0.115
X_{42}	0.272	0.144	0.652	0.143	0.143	0.223	0.115	0.534	0.113	0.114
X_{43}	0.289	0.143	0.143	0.651	0.144	0.238	0.114	0.114	0.534	0.115
X_{44}	0.291	0.143	0.143	0.142	0.651	0.239	0.115	0.114	0.114	0.534
Y_5	0.621	0.316	0.317	0.338	0.340	0.501	0.269	0.269	0.289	0.291
X_{51}	0.316	0.736	0.171	0.171	0.171	0.272	0.651	0.143	0.142	0.143
X_{52}	0.317	0.172	0.736	0.171	0.171	0.272	0.143	0.652	0.142	0.143
X_{53}	0.338	0.171	0.171	0.735	0.172	0.290	0.143	0.142	0.652	0.143
X_{54}	0.340	0.172	0.171	0.171	0.735	0.292	0.143	0.142	0.142	0.651
Y_6	0.718	0.357	0.357	0.381	0.383	0.624	0.314	0.313	0.338	0.339
X_{61}	0.357	0.836	0.210	0.210	0.210	0.318	0.736	0.172	0.171	0.172
X_{62}	0.358	0.211	0.835	0.210	0.210	0.318	0.172	0.736	0.171	0.171
X_{63}	0.381	0.209	0.208	0.836	0.210	0.339	0.171	0.170	0.736	0.171
X_{64}	0.383	0.210	0.209	0.210	0.835	0.342	0.172	0.171	0.171	0.735
Y_7	1.000	0.400	0.400	0.426	0.429	0.712	0.356	0.355	0.381	0.383
X_{71}		1.000	0.239	0.238	0.239	0.360	0.836	0.211	0.210	0.211
X_{72}			1.000	0.238	0.238	0.360	0.210	0.836	0.209	0.210
X_{73}				1.000	0.239	0.383	0.210	0.209	0.836	0.210
X_{74}					1.000	0.385	0.210	0.210	0.210	0.835
Y_8						1.000	0.400	0.399	0.427	0.431
X_{81}							1.000	0.240	0.239	0.240
X_{82}								1.000	0.238	0.239
X_{83}									1.000	0.238
X_{84}										1.000

APPENDIX C

R CODES FOR FORECASTING INDEPENDENT RANDOM VARIABLES

```
################ AR1 for x1
id=rep(1:300,each=3)
x1_m1=matrix(rbind(x[,1],x[,5],x[,9]),ncol=1)
x1_m2=matrix(rbind(x[,5],x[,9],x[,13]),ncol=1)
gee_1=gee(x1_m2~x1_m1,id, family="gaussian", corstr="unstructured" ,scale.fix = T )
gee_1_c=gee_1$coef
x1_5_p=gee_1_c[1]+(gee_1_c[2]*x[,13])
x1_6_p=gee_1_c[1]+(gee_1_c[2]*x1_5_p)
x1_7_p=gee_1_c[1]+(gee_1_c[2]*x1_6_p)
x1_8_p=gee_1_c[1]+(gee_1_c[2]*x1_7_p)
dev_x1_1=((sum((x1_5_p-x[,17])*(x1_5_p-x[,17])))/n)
dev_x1_2=((sum((x1_6_p-x[,21])*(x1_6_p-x[,21])))/n)
dev_x1_3=((sum((x1_7_p-x[,25])*(x1_7_p-x[,25])))/n)
dev_x1_4=((sum((x1_8_p-x[,29])*(x1_8_p-x[,29])))/n)
dev_x1=c(dev_x1_1,dev_x1_2,dev_x1_3,dev_x1_4)
################ AR2 for x1
id=rep(1:300,each=2)
x1_m1=matrix(rbind(x[,1],x[,5]),ncol=1)
x1_m2=matrix(rbind(x[,5],x[,9]),ncol=1)
x1_m3=matrix(rbind(x[,9],x[,13]),ncol=1)
gee_2=gee(x1_m3~x1_m2+x1_m1,id, family="gaussian", corstr="unstructured" ,scale.fix = T
)
gee_2_c=gee_2$coef
x1_5_p=gee_2_c[1]+(gee_2_c[2]*x[,13])+(gee_2_c[3]*x[,9])
x1_6_p=gee_2_c[1]+(gee_2_c[2]*x1_5_p)+(gee_2_c[3]*x[,13])
x1_7_p=gee_2_c[1]+(gee_2_c[2]*x1_6_p)+(gee_2_c[3]*x1_5_p)
x1_8_p=gee_2_c[1]+(gee_2_c[2]*x1_7_p)+(gee_2_c[3]*x1_6_p)
dev_x1_1=((sum((x1_5_p-x[,17])*(x1_5_p-x[,17])))/n)
dev_x1_2=((sum((x1_6_p-x[,21])*(x1_6_p-x[,21])))/n)
dev_x1_3=((sum((x1_7_p-x[,25])*(x1_7_p-x[,25])))/n)
dev_x1_4=((sum((x1_8_p-x[,29])*(x1_8_p-x[,29])))/n)
dev_x1=c(dev_x1_1,dev_x1_2,dev_x1_3,dev_x1_4)
################ random effect for x1
id=rep(1:300,each=4)
time=rep(1:4,300)
x1_m1=matrix(rbind(x[,1],x[,5],x[,9],x[,13]),ncol=1)
glmer=glmer( x1_m1~time + (1 |id),family = gaussian)
rand=ranef(glmer)
fixed=fixef(glmer)
time2=rep(5:8,each=300)
```

```
x1_5_p=(fixed[1])+((fixed[2])*time2[1:300])+rand$id
x1_6_p=(fixed[1])+((fixed[2])*time2[301:600])+rand$id
x1_7_p=(fixed[1])+((fixed[2])*time2[601:900])+rand$id
x1_8_p=(fixed[1])+((fixed[2])*time2[901:1200])+rand$id
dev_x1_1=((sum((x1_5_p-x[,17])*(x1_5_p-x[,17])))/n)
dev_x1_2=((sum((x1_6_p-x[,21])*(x1_6_p-x[,21])))/n)
dev_x1_3=((sum((x1_7_p-x[,25])*(x1_7_p-x[,25])))/n)
dev_x1_4=((sum((x1_8_p-x[,29])*(x1_8_p-x[,29])))/n)
dev_x1=c(dev_x1_1,dev_x1_2,dev_x1_3,dev_x1_4)
###################################### moving mean for x1
anew=matrix(rep(0,n),ncol=1)
for (i in 1:n)
        {anew[i]=mean(cbind(x[i,1], x[i,5], x[i,9], x[i,13]), na.rm=T)}
x1_5_p=anew
anew=matrix(rep(0,n),ncol=1)
for (i in 1:n)
        {anew[i]=mean(cbind(x[i,5], x[i,9], x[i,13], x1_5_p[i]), na.rm=T)}
x1_6_p=anew
anew=matrix(rep(0,n),ncol=1)
for (i in 1:n)
        {anew[i]=mean(cbind(x[i,9], x[i,13],x1_5_p[i], x1_6_p[i]), na.rm=T)}
x1_7_p=anew
anew=matrix(rep(0,n),ncol=1)
for (i in 1:n)
        {anew[i]=mean(cbind(x[i,13],x1_5_p[i], x1_6_p[i], x1_7_p[i]), na.rm=T)}
x1_8_p=anew
dev_x1_1=((sum((x1_5_p-x[,17])*(x1_5_p-x[,17])))/n)
dev_x1_2=((sum((x1_6_p-x[,21])*(x1_6_p-x[,21])))/n)
dev_x1_3=((sum((x1_7_p-x[,25])*(x1_7_p-x[,25])))/n)
dev_x1_4=((sum((x1_8_p-x[,29])*(x1_8_p-x[,29])))/n)
dev_x1=c(dev_x1_1,dev_x1_2,dev_x1_3,dev_x1_4)
```

APPENDIX D

R CODES FOR METHODS AND FORECASTING

```
# general rule for notation of y: yhat?pre_??= prediction of y where ?=time point and
??=number of model
# general rule for notation of pre: pre_?= prediction values of y time point after 4 where
?=model number
########################## 1 moving median, cut-off=0.5
anew_1=apply(cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4]),1,median,na.rm=T)
yhat5pre_1=ifelse(anew_1>=0.5,1,0)
anew_2=apply(cbind(yhat[,2], yhat[,3], yhat[,4], yhat5pre_1),1,median,na.rm=T)
yhat6pre_1=ifelse(anew_2>=0.5,1,0)
anew_3=apply(cbind(yhat[,3], yhat[,4],yhat5pre_1, yhat6pre_1),1,median,na.rm=T)
yhat7pre_1=ifelse(anew_3>=0.5,1,0)
anew_4=apply(cbind(yhat[,4],yhat5pre_1, yhat6pre_1, yhat7pre_1),1,median,na.rm=T)
yhat8pre_1=ifelse(anew_4>=0.5,1,0)
############## 2 non-moving median cut-off=0.5
anew_1=apply(cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4]),1,median, na.rm=T)
yhat5pre_2=ifelse(anew_1>=0.5,1,0)
anew_2=apply(cbind(yhat[,1],yhat[,2], yhat[,3], yhat[,4], yhat5pre_2),1,median, na.rm=T)
yhat6pre_2=ifelse(anew_2>=0.5,1,0)
anew_3=apply(cbind(yhat[,1],yhat[,2],yhat[,3], yhat[,4],yhat5pre_2, yhat6pre_2),1,median,
na.rm=T)
yhat7pre_2=ifelse(anew_3>=0.5,1,0)
anew_4=apply(cbind(yhat[,1],yhat[,2],yhat[,3],yhat[,4],yhat5pre_2, yhat6pre_2,
yhat7pre_2),1,median, na.rm=T)
yhat8pre_2=ifelse(anew_4>=0.5,1,0)
############## 3 moving mode, cut-off=0.5
anew_1=matrix(rep(0,n),ncol=1)
aneww=cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4])
for (i in 1:n)      {anew_1[i]=as.numeric(names(sort(-table(aneww[i,])))[1])}
yhat5pre_3=anew_1
anew_2=matrix(rep(0,n),ncol=1)
aneww=cbind(yhat[,2], yhat[,3], yhat[,4], yhat5pre_3)
for (i in 1:n)      {anew_2[i]=as.numeric(names(sort(-table(aneww[i,])))[1])}
yhat6pre_3=anew_2
anew_3=matrix(rep(0,n),ncol=1)
aneww=cbind(yhat[,3], yhat[,4],yhat5pre_3, yhat6pre_3)
for (i in 1:n)      {anew_3[i]=as.numeric(names(sort(-table(aneww[i,])))[1])}
yhat7pre_3=anew_3
anew_4=matrix(rep(0,n),ncol=1)
aneww=cbind(yhat[,4],yhat5pre_3, yhat6pre_3, yhat7pre_3)
for (i in 1:n)      {anew_4[i]=as.numeric(names(sort(-table(aneww[i,])))[1])}
yhat8pre_3=anew_4
```

```
############## 4 non-moving mode cut-off=0.5
anew_1=matrix(rep(0,n),ncol=1)
aneww=cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4])
for (i in 1:n)      {anew_1[i]=as.numeric(names(sort(-table(aneww[i,])))[1])}
yhat5pre_4=anew_1
anew_2=matrix(rep(0,n),ncol=1)
aneww=cbind(yhat[,1],yhat[,2], yhat[,3], yhat[,4], yhat5pre_4)
for (i in 1:n)      {anew_2[i]=as.numeric(names(sort(-table(aneww[i,])))[1])}
yhat6pre_4=anew_2
anew_3=matrix(rep(0,n),ncol=1)
aneww=cbind(yhat[,1],yhat[,2],yhat[,3], yhat[,4],yhat5pre_4, yhat6pre_4)
for (i in 1:n)      {anew_3[i]=as.numeric(names(sort(-table(aneww[i,])))[1])}
yhat7pre_4=anew_3
anew_4=matrix(rep(0,n),ncol=1)
aneww=cbind(yhat[,1],yhat[,2],yhat[,3],yhat[,4],yhat5pre_4, yhat6pre_4, yhat7pre_4)
for (i in 1:n)      {anew_4[i]=as.numeric(names(sort(-table(aneww[i,])))[1])}
yhat8pre_4=anew_4
########################## 5 moving mean, cut-off=0.5
anew_1=apply(cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4]),1,mean, na.rm=T)
yhat5pre_5=ifelse(anew_1>=0.5,1,0)
anew_2=apply(cbind(yhat[,2], yhat[,3], yhat[,4], yhat5pre_5),1,mean, na.rm=T)
yhat6pre_5=ifelse(anew_2>=0.5,1,0)
anew_3=apply(cbind(yhat[,3], yhat[,4],yhat5pre_5, yhat6pre_5),1,mean, na.rm=T)
yhat7pre_5=ifelse(anew_3>=0.5,1,0)
anew_4=apply(cbind(yhat[,4],yhat5pre_5, yhat6pre_5, yhat7pre_5),1,mean, na.rm=T)
yhat8pre_5=ifelse(anew_4>=0.5,1,0)
######################## 6 non-moving mean cut-off=0.5
anew_1=apply(cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4]),1,mean, na.rm=T)
yhat5pre_6=ifelse(anew_1>=0.5,1,0)
anew_2=apply(cbind(yhat[,1],yhat[,2], yhat[,3], yhat[,4], yhat5pre_6),1,mean, na.rm=T)
yhat6pre_6=ifelse(anew_2>=0.5,1,0)
anew_3=apply(cbind(yhat[,1],yhat[,2],yhat[,3], yhat[,4],yhat5pre_6, yhat6pre_6),1,mean,
na.rm=T)
yhat7pre_6=ifelse(anew_3>=0.5,1,0)
anew_4=apply(cbind(yhat[,1],yhat[,2],yhat[,3],yhat[,4],yhat5pre_6, yhat6pre_6,
yhat7pre_6),1,mean, na.rm=T)
yhat8pre_6=ifelse(anew_4>=0.5,1,0)
######################## 7 moving mean cut-off=u
anew_1=apply(cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4]),1,mean, na.rm=T)
u_1=runif(n, min=0, max=1)
yhat5pre_7=ifelse(anew_1>=u_1,1,0)
anew_2=apply(cbind(yhat[,2], yhat[,3], yhat[,4], yhat5pre_7),1,mean, na.rm=T)
u_2=runif(n, min=0, max=1)
yhat6pre_7=ifelse(anew_2>=u_2,1,0)
anew_3=apply(cbind(yhat[,3], yhat[,4],yhat5pre_7, yhat6pre_7),1,mean, na.rm=T)
u_3=runif(n, min=0, max=1)
yhat7pre_7=ifelse(anew_3>=u_3,1,0)
anew_4=apply(cbind(yhat[,4],yhat5pre_7, yhat6pre_7, yhat7pre_7),1,mean, na.rm=T)
u_4=runif(n, min=0, max=1)
yhat8pre_7=ifelse(anew_4>=u_4,1,0)
######################## 8 non-moving mean cut-off=u
anew_1=apply(cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4]),1,mean, na.rm=T)
u_1=runif(n, min=0, max=1)
yhat5pre_8=ifelse(anew_1>=u_1,1,0)
anew_2=apply(cbind(yhat[,1],yhat[,2], yhat[,3], yhat[,4], yhat5pre_8),1,mean, na.rm=T)
u_2=runif(n, min=0, max=1)
yhat6pre_8=ifelse(anew_2>=u_2,1,0)
anew_3=apply(cbind(yhat[,1],yhat[,2],yhat[,3], yhat[,4],yhat5pre_8, yhat6pre_8),1,mean,
na.rm=T)
```

```
u_3=runif(n, min=0, max=1)
yhat7pre_8=ifelse(anew_3>=u_3,1,0)
anew_4=apply(cbind(yhat[,1],yhat[,2],yhat[,3],yhat[,4],yhat5pre_8, yhat6pre_8,
yhat7pre_8),1,mean, na.rm=T)
u_4=runif(n, min=0, max=1)
yhat8pre_8=ifelse(anew_4>=u_4,1,0)
######################## 9 moving mean cut-off=mean
cutoff_1=mean(yhat5obs)
cutoff_2=mean(yhat6obs)
cutoff_3=mean(yhat7obs)
cutoff_4=mean(yhat8obs)
anew_1=apply(cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4]),1,mean, na.rm=T)
yhat5pre_9=ifelse(anew_1>=cutoff_1,1,0)
anew_2=apply(cbind(yhat[,2], yhat[,3], yhat[,4], yhat5pre_9),1,mean, na.rm=T)
yhat6pre_9=ifelse(anew_2>=cutoff_2,1,0)
anew_3=apply(cbind(yhat[,3], yhat[,4],yhat5pre_9, yhat6pre_9),1,mean, na.rm=T)
yhat7pre_9=ifelse(anew_3>=cutoff_3,1,0)
anew_4=apply(cbind(yhat[,4],yhat5pre_9, yhat6pre_9, yhat7pre_9),1,mean, na.rm=T)
yhat8pre_9=ifelse(anew_4>=cutoff_4,1,0)
######################## 10 non-moving mean cut-off=mean
cutoff_1=mean(yhat5obs)
cutoff_2=mean(yhat6obs)
cutoff_3=mean(yhat7obs)
cutoff_4=mean(yhat8obs)
anew_1=apply(cbind(yhat[,1], yhat[,2], yhat[,3], yhat[,4]),1,mean, na.rm=T)
yhat5pre_10=ifelse(anew_1>=cutoff_1,1,0)
anew_2=apply(cbind(yhat[,1],yhat[,2], yhat[,3], yhat[,4], yhat5pre_10),1,mean, na.rm=T)
yhat6pre_10=ifelse(anew_2>=cutoff_2,1,0)
anew_3=apply(cbind(yhat[,1],yhat[,2],yhat[,3], yhat[,4],yhat5pre_10, yhat6pre_10),1,mean,
na.rm=T)
yhat7pre_10=ifelse(anew_3>=cutoff_3,1,0)
anew_4=apply(cbind(yhat[,1],yhat[,2],yhat[,3],yhat[,4],yhat5pre_10, yhat6pre_10,
yhat7pre_10),1,mean, na.rm=T)
yhat8pre_10=ifelse(anew_4>=cutoff_4,1,0)
################## 12 AR1 for y with same coefficients
id_12=rep(1:300,each=3)
y_12_m1=matrix(rbind(yhat[,1],yhat[,2],yhat[,3]),ncol=1)
y_12_m2=matrix(rbind(yhat[,2],yhat[,3],yhat[,4]),ncol=1)
gee_12=gee(y_12_m2~y_12_m1,id_12, family="binomial", corstr="unstructured"
,maxiter=100 )
cc_12=gee_12$coef
# calculation of y when time=2
b_2_12=cc_12[1]+(cc_12[2]*yhat[,1])
yhat2pre_12_0=(exp(b_2_12)/(1+exp(b_2_12)))
yhat2pre_12=ifelse(yhat2pre_12_0>=0.5,1,0)
# calculation of y when time=3
b_3_12=cc_12[1]+(cc_12[2]*yhat[,2])
yhat3pre_12_0=(exp(b_3_12)/(1+exp(b_3_12)))
yhat3pre_12=ifelse(yhat3pre_12_0>=0.5,1,0)
b_4_12=cc_12[1]+(cc_12[2]*yhat[,3])
yhat4pre_12_0=(exp(b_4_12)/(1+exp(b_4_12)))
yhat4pre_12=ifelse(yhat4pre_12_0>=0.5,1,0)
b_5_12=cc_12[1]+(cc_12[2]*yhat[,4])
yhat5pre_12_0=(exp(b_5_12)/(1+exp(b_5_12)))
yhat5pre_12=ifelse(yhat5pre_12_0>=0.5,1,0)
b_6_12=cc_12[1]+(cc_12[2]*yhat5pre_12)
yhat6pre_12_0=(exp(b_6_12)/(1+exp(b_6_12)))
yhat6pre_12=ifelse(yhat6pre_12_0>=0.5,1,0)
b_7_12=cc_12[1]+(cc_12[2]*yhat6pre_12)
```

```
yhat7pre_12_0=(exp(b_7_12)/(1+exp(b_7_12)))
yhat7pre_12=ifelse(yhat7pre_12_0>=0.5,1,0)
b_8_12=cc_12[1]+(cc_12[2]*yhat7pre_12)
yhat8pre_12_0=(exp(b_8_12)/(1+exp(b_8_12)))
yhat8pre_12=ifelse(yhat8pre_12_0>=0.5,1,0)
################### 14 AR2 for y with same coefficient
id_14=rep(1:300,each=2)
y_14_m1=matrix(rbind(yhat[,3],yhat[,4]),ncol=1)
y_14_m2=matrix(rbind(yhat[,1],yhat[,2]),ncol=1)
y_14_m3=matrix(rbind(yhat[,2],yhat[,3]),ncol=1)
gee_14=gee(y_14_m1~y_14_m3+y_14_m2,id_14, family="binomial", corstr="unstructured"
,maxiter=100 )
cc_14=gee_14$coef
b_3_14=cc_14[1]+(cc_14[2]*yhat[,2])+(cc_14[3]*yhat[,1])
yhat3pre_14_0=(exp(b_3_14)/(1+exp(b_3_14)))
yhat3pre_14=ifelse(yhat3pre_14_0>=0.5,1,0)
b_4_14=cc_14[1]+(cc_14[2]*yhat[,3])+(cc_14[3]*yhat[,2])
yhat4pre_14_0=(exp(b_4_14)/(1+exp(b_4_14)))
yhat4pre_14=ifelse(yhat4pre_14_0>=0.5,1,0)
b_5_14=cc_14[1]+(cc_14[2]*yhat[,4])+(cc_14[3]*yhat[,3])
yhat5pre_14_0=(exp(b_5_14)/(1+exp(b_5_14)))
yhat5pre_14=ifelse(yhat5pre_14_0>=0.5,1,0)
b_6_14=cc_14[1]+(cc_14[2]*yhat5pre_14)+(cc_14[3]*yhat[,4])
yhat6pre_14_0=(exp(b_6_14)/(1+exp(b_6_14)))
yhat6pre_14=ifelse(yhat6pre_14_0>=0.5,1,0)
b_7_14=cc_14[1]+(cc_14[2]*yhat6pre_14)+(cc_14[3]*yhat5pre_14)
yhat7pre_14_0=(exp(b_7_14)/(1+exp(b_7_14)))
yhat7pre_14=ifelse(yhat7pre_14_0>=0.5,1,0)
b_8_14=cc_14[1]+(cc_14[2]*yhat7pre_14)+(cc_14[3]*yhat6pre_14)
yhat8pre_14_0=(exp(b_8_14)/(1+exp(b_8_14)))
yhat8pre_14=ifelse(yhat8pre_14_0>=0.5,1,0)
######################################## 16 AR3 for y with same coefficients
id_16=rep(1:300,each=1)
y_16_m1=matrix(rbind(yhat[,4]),ncol=1)
y_16_m2=matrix(rbind(yhat[,3]),ncol=1)
y_16_m3=matrix(rbind(yhat[,2]),ncol=1)
y_16_m4=matrix(rbind(yhat[,1]),ncol=1)
gee_16=gee(y_16_m1~y_16_m2+y_16_m3+y_16_m4,id_16, family="binomial",
corstr="unstructured" ,maxiter=100 )
cc_16=gee_16$coef
b_4_16=cc_16[1]+(cc_16[2]*yhat[,3])+(cc_16[3]*yhat[,2])+(cc_16[4]*yhat[,1])
yhat4pre_16_0=(exp(b_4_16)/(1+exp(b_4_16)))
yhat4pre_16=ifelse(yhat4pre_16_0>=0.5,1,0)
b_5_16=cc_16[1]+(cc_16[2]*yhat[,4])+(cc_16[3]*yhat[,3])+(cc_16[4]*yhat[,2])
yhat5pre_16_0=(exp(b_5_16)/(1+exp(b_5_16)))
yhat5pre_16=ifelse(yhat5pre_16_0>=0.5,1,0)
b_6_16=cc_16[1]+(cc_16[2]*yhat5pre_16)+(cc_16[3]*yhat[,4])+(cc_16[4]*yhat[,3])
yhat6pre_16_0=(exp(b_6_16)/(1+exp(b_6_16)))
yhat6pre_16=ifelse(yhat6pre_16_0>=0.5,1,0)
b_7_16=cc_16[1]+(cc_16[2]*yhat6pre_16)+(cc_16[3]*yhat5pre_16)+(cc_16[4]*yhat[,4])
yhat7pre_16_0=(exp(b_7_16)/(1+exp(b_7_16)))
yhat7pre_16=ifelse(yhat7pre_16_0>=0.5,1,0)
b_8_16=cc_16[1]+(cc_16[2]*yhat7pre_16)+(cc_16[3]*yhat6pre_16)+(cc_16[4]*yhat5pre_16)
yhat8pre_16_0=(exp(b_8_16)/(1+exp(b_8_16)))
yhat8pre_16=ifelse(yhat8pre_16_0>=0.5,1,0)
################################### 17 Marginal Models
id_17=rep(1:300,each=4)
y_17=matrix(rbind(yhat[,1],yhat[,2],yhat[,3],yhat[,4]),ncol=1)
x1_17=matrix(rbind(x[,1],x[,5],x[,9],x[,13]),ncol=1)
```

```
x2_17=matrix(rbind(x[,2],x[,6],x[,10],x[,14]),ncol=1)
x3_17=matrix(rbind(x[,3],x[,7],x[,11],x[,15]),ncol=1)
x4_17=matrix(rbind(x[,4],x[,8],x[,12],x[,16]),ncol=1)
data_17=matrix(cbind(id_17,y_17,x1_17,x2_17,x3_17,x4_17),ncol=6)
gee_17=gee( y_17 ~ x1_17+x2_17+x3_17+x4_17,id_17, family="binomial",
corstr="unstructured" ,maxiter=100 )
gee_17_c=gee_17$coef
fitted_y_17_0=gee_17$fit
fitted_y_17=ifelse(fitted_y_17_0>=0.5,1,0)
y_17=gee_17$y
b_5_17=gee_17_c[1]+(gee_17_c[2]*pre_x[,1])+(gee_17_c[3]*pre_x[,2])+(gee_17_c[4]*pre_x[
,3])+(gee_17_c[5]*pre_x[,4])
yhat5pre_17_0=(exp(b_5_17)/(1+exp(b_5_17)))
yhat5pre_17=ifelse(yhat5pre_17_0>=0.5,1,0)
b_6_17=gee_17_c[1]+(gee_17_c[2]*pre_x[,5])+(gee_17_c[3]*pre_x[,6])+(gee_17_c[4]*pre_x[
,7])+(gee_17_c[5]*pre_x[,8])
yhat6pre_17_0=(exp(b_6_17)/(1+exp(b_6_17)))
yhat6pre_17=ifelse(yhat6pre_17_0>=0.5,1,0)
b_7_17=gee_17_c[1]+(gee_17_c[2]*pre_x[,9])+(gee_17_c[3]*pre_x[,10])+(gee_17_c[4]*pre_
x[,11])+(gee_17_c[5]*pre_x[,12])
yhat7pre_17_0=(exp(b_7_17)/(1+exp(b_7_17)))
yhat7pre_17=ifelse(yhat7pre_17_0>=0.5,1,0)
b_8_17=gee_17_c[1]+(gee_17_c[2]*pre_x[,13])+(gee_17_c[3]*pre_x[,14])+(gee_17_c[4]*pre
_x[,15])+(gee_17_c[5]*pre_x[,16])
yhat8pre_17_0=(exp(b_8_17)/(1+exp(b_8_17)))
yhat8pre_17=ifelse(yhat8pre_17_0>=0.5,1,0)
################################### 19 Transition Models (1)
id_19=rep(1:300,each=3)
y_19_m1=matrix(rbind(yhat[,2],yhat[,3],yhat[,4]),ncol=1)
y_19_m2=matrix(rbind(yhat[,1],yhat[,2],yhat[,3]),ncol=1)
x1_19=matrix(rbind(x[,5],x[,9],x[,13]),ncol=1)
x2_19=matrix(rbind(x[,6],x[,10],x[,14]),ncol=1)
x3_19=matrix(rbind(x[,7],x[,11],x[,15]),ncol=1)
x4_19=matrix(rbind(x[,8],x[,12],x[,16]),ncol=1)
data_19=matrix(cbind(id_19,y_19_m1,x1_19,x2_19,x3_19,x4_19,y_19_m2),ncol=7)
gee_19=gee( y_19_m1 ~ x1_19+x2_19+x3_19+x4_19+y_19_m2, id_19, family="binomial",
corstr="unstructured" ,maxiter=100 )
gee_19_c=gee_19$coef
fitted_y_19_0=gee_19$fit
fitted_y_19=ifelse(fitted_y_19_0>=0.5,1,0)
y_19=gee_19$y
b_5_19=gee_19_c[1]+(gee_19_c[2]*pre_x[,1])+(gee_19_c[3]*pre_x[,2])+(gee_19_c[4]*pre_x[
,3])+(gee_19_c[5]*pre_x[,4])+(gee_19_c[6]*yhat[,4])
yhat5pre_19_0=(exp(b_5_19)/(1+exp(b_5_19)))
yhat5pre_19=ifelse(yhat5pre_19_0>=0.5,1,0)
b_6_19=gee_19_c[1]+(gee_19_c[2]*pre_x[,5])+(gee_19_c[3]*pre_x[,6])+(gee_19_c[4]*pre_x[
,7])+(gee_19_c[5]*pre_x[,8])+(gee_19_c[6]*yhat5pre_19)
yhat6pre_19_0=(exp(b_6_19)/(1+exp(b_6_19)))
yhat6pre_19=ifelse(yhat6pre_19_0>=0.5,1,0)
b_7_19=gee_19_c[1]+(gee_19_c[2]*pre_x[,9])+(gee_19_c[3]*pre_x[,10])+(gee_19_c[4]*pre_
x[,11])+(gee_19_c[5]*pre_x[,12])+(gee_19_c[6]*yhat6pre_19)
yhat7pre_19_0=(exp(b_7_19)/(1+exp(b_7_19)))
yhat7pre_19=ifelse(yhat7pre_19_0>=0.5,1,0)
b_8_19=gee_19_c[1]+(gee_19_c[2]*pre_x[,13])+(gee_19_c[3]*pre_x[,14])+(gee_19_c[4]*pre
_x[,15])+(gee_19_c[5]*pre_x[,16])+(gee_19_c[6]*yhat7pre_19)
yhat8pre_19_0=(exp(b_8_19)/(1+exp(b_8_19)))
yhat8pre_19=ifelse(yhat8pre_19_0>=0.5,1,0)
```

```
################################### 20 Transition Models (2)
id_20=rep(1:300,each=2)
y_20_m1=matrix(rbind(yhat[,3],yhat[,4]),ncol=1)
y_20_m2=matrix(rbind(yhat[,2],yhat[,3]),ncol=1)
y_20_m3=matrix(rbind(yhat[,1],yhat[,2]),ncol=1)
x1_20=matrix(rbind(x[,9],x[,13]),ncol=1)
x2_20=matrix(rbind(x[,10],x[,14]),ncol=1)
x3_20=matrix(rbind(x[,11],x[,15]),ncol=1)
x4_20=matrix(rbind(x[,12],x[,16]),ncol=1)
data_20=matrix(cbind(id_20,y_20_m1,x1_20,x2_20,x3_20,x4_20,y_20_m2,y_20_m3),ncol=
8)
gee_20=gee( y_20_m1 ~ x1_20+x2_20+x3_20+x4_20+y_20_m2+y_20_m3,id_20,
family="binomial", corstr="unstructured" ,maxiter=100 )
gee_20_c=gee_20$coef
fitted_y_20_0=gee_20$fit
fitted_y_20=ifelse(fitted_y_20_0>=0.5,1,0)
y_20=gee_20$y
b_5_20=gee_20_c[1]+(gee_20_c[2]*pre_x[,1])+(gee_20_c[3]*pre_x[,2])+(gee_20_c[4]*pre_x[
,3])+(gee_20_c[5]*pre_x[,4])+(gee_20_c[6]*yhat[,4])+(gee_20_c[7]*yhat[,3])
yhat5pre_20_0=(exp(b_5_20)/(1+exp(b_5_20)))
yhat5pre_20=ifelse(yhat5pre_20_0>=0.5,1,0)
b_6_20=gee_20_c[1]+(gee_20_c[2]*pre_x[,5])+(gee_20_c[3]*pre_x[,6])+(gee_20_c[4]*pre_x[
,7])+(gee_20_c[5]*pre_x[,8])+(gee_20_c[6]*yhat5pre_20)+(gee_20_c[7]*yhat[,4])
yhat6pre_20_0=(exp(b_6_20)/(1+exp(b_6_20)))
yhat6pre_20=ifelse(yhat6pre_20_0>=0.5,1,0)
b_7_20=gee_20_c[1]+(gee_20_c[2]*pre_x[,9])+(gee_20_c[3]*pre_x[,10])+(gee_20_c[4]*pre_
x[,11])+(gee_20_c[5]*pre_x[,12])+(gee_20_c[6]*yhat6pre_20)+(gee_20_c[7]*yhat5pre_20)
yhat7pre_20_0=(exp(b_7_20)/(1+exp(b_7_20)))
yhat7pre_20=ifelse(yhat7pre_20_0>=0.5,1,0)
b_8_20=gee_20_c[1]+(gee_20_c[2]*pre_x[,13])+(gee_20_c[3]*pre_x[,14])+(gee_20_c[4]*pre
_x[,15])+(gee_20_c[5]*pre_x[,16])+(gee_20_c[6]*yhat7pre_20)+(gee_20_c[7]*yhat6pre_20)
yhat8pre_20_0=(exp(b_8_20)/(1+exp(b_8_20)))
yhat8pre_20=ifelse(yhat8pre_20_0>=0.5,1,0)
############################################### 21  Transition Models (3)
id_21=rep(1:300,each=1)
y_21_m1=matrix(rbind(yhat[,4]),ncol=1)
y_21_m2=matrix(rbind(yhat[,3]),ncol=1)
y_21_m3=matrix(rbind(yhat[,2]),ncol=1)
y_21_m4=matrix(rbind(yhat[,1]),ncol=1)
x1_21=matrix(rbind(x[,13]),ncol=1)
x2_21=matrix(rbind(x[,14]),ncol=1)
x3_21=matrix(rbind(x[,15]),ncol=1)
x4_21=matrix(rbind(x[,16]),ncol=1)
data_21=matrix(cbind(id_21,y_21_m1,x1_21,x2_21,x3_21,x4_21,y_21_m2,y_21_m3,y_21_
m4),ncol=9)
gee_21=gee( y_21_m1 ~ x1_21+x2_21+x3_21+x4_21+y_21_m2+y_21_m3+y_21_m4,
id_21, family="binomial", corstr="unstructured" ,maxiter=100 )
gee_21_c=gee_21$coef
fitted_y_21_0=gee_21$fit
fitted_y_21=ifelse(fitted_y_21_0>=0.5,1,0)
y_21=gee_21$y
b_5_21=gee_21_c[1]+(gee_21_c[2]*pre_x[,1])+(gee_21_c[3]*pre_x[,2])+(gee_21_c[4]*pre_x[
,3])+(gee_21_c[5]*pre_x[,4])+(gee_21_c[6]*yhat[,4])+(gee_21_c[7]*yhat[,3])+(gee_21_c[8]*y
hat[,2])
yhat5pre_21_0=(exp(b_5_21)/(1+exp(b_5_21)))
yhat5pre_21=ifelse(yhat5pre_21_0>=0.5,1,0)
b_6_21=gee_21_c[1]+(gee_21_c[2]*pre_x[,5])+(gee_21_c[3]*pre_x[,6])+(gee_21_c[4]*pre_x[
,7])+(gee_21_c[5]*pre_x[,8])+(gee_21_c[6]*yhat5pre_21)+(gee_21_c[7]*yhat[,4])+(gee_21_c
[8]*yhat[,3])
```

```
yhat6pre_21_0=(exp(b_6_21)/(1+exp(b_6_21)))
yhat6pre_21=ifelse(yhat6pre_21_0>=0.5,1,0)
b_7_21=gee_21_c[1]+(gee_21_c[2]*pre_x[,9])+(gee_21_c[3]*pre_x[,10])+(gee_21_c[4]*pre_
x[,11])+(gee_21_c[5]*pre_x[,12])+(gee_21_c[6]*yhat6pre_21)+(gee_21_c[7]*yhat5pre_21)+(
gee_21_c[8]*yhat[,4])
yhat7pre_21_0=(exp(b_7_21)/(1+exp(b_7_21)))
yhat7pre_21=ifelse(yhat7pre_21_0>=0.5,1,0)
b_8_21=gee_21_c[1]+(gee_21_c[2]*pre_x[,13])+(gee_21_c[3]*pre_x[,14])+(gee_21_c[4]*pre
_x[,15])+(gee_21_c[5]*pre_x[,16])+(gee_21_c[6]*yhat7pre_21)+(gee_21_c[7]*yhat6pre_21)+
(gee_21_c[8]*yhat5pre_21)
yhat8pre_21_0=(exp(b_8_21)/(1+exp(b_8_21)))
yhat8pre_21=ifelse(yhat8pre_21_0>=0.5,1,0)
################################### 22 MTM (1) Marginalized Transition Models
id_22=rep(1:300,each=4)
y_22_m1=matrix(rbind(yhat[,1],yhat[,2],yhat[,3],yhat[,4]),ncol=1)
x1_22=matrix(rbind(x[,1],x[,5],x[,9],x[,13]),ncol=1)
x2_22=matrix(rbind(x[,2],x[,6],x[,10],x[,14]),ncol=1)
x3_22=matrix(rbind(x[,3],x[,7],x[,11],x[,15]),ncol=1)
x4_22=matrix(rbind(x[,4],x[,8],x[,12],x[,16]),ncol=1)
replicate=rep(-99,n)
y_22_m2=matrix(rbind(replicate,yhat[,1],yhat[,2],yhat[,3]),ncol=1)
data_22=matrix(cbind(id_22,y_22_m1,x1_22,x2_22,x3_22,x4_22,y_22_m2),ncol=7)
gee_22=gee( y_22_m1 ~ x1_22+x2_22+x3_22+x4_22, id_22, family="binomial",
corstr="unstructured", maxiter=100 )
gee_22_c=gee_22$coef
  source("C:/Users/hp/Desktop/tez_kod_100107/MTM/mtm_lag1.q")
  dyn.load("C:/Users/hp/Desktop/tez_kod_100107/MTM/BFKDISS__C.dll")
x_model = model.matrix( ~x1_22+x2_22+x3_22+x4_22 )
z = rep(1,nrow(data_22))
alpha = 1.0
beta = gee_22$coef
lag1mtm<- mtm.lag1( id=id_22,  y=y_22_m1, x=x_model, z=z, beta=beta, alpha=alpha,
tol=1e-4)
print(lag1mtm)
coef=lag1mtm$beta
alpha=lag1mtm$alpha
logli=lag1mtm$logL
pi4=lag1mtm$muT[seq(4,1200,4)]
yi4=y_22_m1[seq(4,1200,4)]
pi3=lag1mtm$muT[seq(3,1199,4)]
yi3=y_22_m1[seq(3,1199,4)]
pi2=lag1mtm$muT[seq(2,1198,4)]
yi2=y_22_m1[seq(2,1198,4)]
pi1=lag1mtm$muT[seq(1,1197,4)]
yi1=y_22_m1[seq(1,1197,4)]
di4=(log(pi4)/log(1-pi4))-(alpha*yi3)
di3=(log(pi3)/log(1-pi3))-(alpha*yi2)
di2=(log(pi2)/log(1-pi2))-(alpha*yi1)
######## forecasting di
di5=apply(cbind(di4, di3, di2),1,mean, na.rm=T)
di6=apply(cbind(di5,di4,di3),1,mean, na.rm=T)
di7=apply(cbind(di6,di5,di4),1,mean, na.rm=T)
di8=apply(cbind(di7,di6,di5),1,mean, na.rm=T)
b_5_22=di5+(alpha*yhat[,4])
yhat5pre_22_0=(exp(b_5_22)/(1+exp(b_5_22)))
yhat5pre_22=ifelse(yhat5pre_22_0>=0.5,1,0)
b_6_22=di6+(alpha*yhat5pre_22)
yhat6pre_22_0=(exp(b_6_22)/(1+exp(b_6_22)))
yhat6pre_22=ifelse(yhat6pre_22_0>=0.5,1,0)
```

```
b_7_22=di7+(alpha*yhat6pre_22)
yhat7pre_22_0=(exp(b_7_22)/(1+exp(b_7_22)))
yhat7pre_22=ifelse(yhat7pre_22_0>=0.5,1,0)
b_8_22=di8+(alpha*yhat7pre_22)
yhat8pre_22_0=(exp(b_8_22)/(1+exp(b_8_22)))
yhat8pre_22=ifelse(yhat8pre_22_0>=0.5,1,0)
################################### 23 MTM (2)
replicate_2=rep(-99,n)
y_22_m3=matrix(rbind(replicate,replicate_2,yhat[,1],yhat[,2]),ncol=1)
data_23=matrix(cbind(id_22,y_22_m1,x1_22,x2_22,x3_22,x4_22,y_22_m2,y_22_m3),ncol=
8)
x_model_2 = model.matrix( ~x1_22+x2_22+x3_22+x4_22 )
beta=gee_22_c
  source("C:/Users/hp/Desktop/tez_kod_100107/MTM/mtm_lag2.q")
z1 = matrix(c(rep(1,nrow(data_23))),ncol=1)
z2=matrix(c(rep(1,nrow(data_23))),ncol=1)
alpha1= 2.4
alpha2= 1.3
 lag2mtm = mtm.lag2( id=id_22,  y=y_22_m1, x=x_model_2, z1=z1,z2=z2, beta=beta,
alpha1=alpha1, alpha2=alpha2,  tol=1e-4,Fisher="FALSE")
print(lag2mtm)
names(lag2mtm)
coef=lag2mtm$beta
alpha1=lag2mtm$alpha1
alpha2=lag2mtm$alpha2
pi4=lag2mtm$muT[seq(4,1200,4)]
yi4=y_22_m1[seq(4,1200,4)]
pi3=lag2mtm$muT[seq(3,1199,4)]
yi3=y_22_m1[seq(3,1199,4)]
pi2=lag2mtm$muT[seq(2,1198,4)]
yi2=y_22_m1[seq(2,1198,4)]
pi1=lag2mtm$muT[seq(1,1197,4)]
yi1=y_22_m1[seq(1,1197,4)]
di4=(log(pi4)/log(1-pi4))-(alpha1*yi3)-(alpha2*yi2)
di3=(log(pi3)/log(1-pi3))-(alpha1*yi2)-(alpha2*yi1)
di5=apply(cbind(di4, di3),1,mean, na.rm=T)
di6=apply(cbind(di5,di4),1,mean,na.rm=T)
di7=apply(cbind(di6,di5),1,mean,na.rm=T)
di8=apply(cbind(di7,di6),1,mean,na.rm=T)
b_5_23=di5+(alpha1[1]*yhat[,4])+(alpha2[1]*yhat[,3])
yhat5pre_23_0=(exp(b_5_23)/(1+exp(b_5_23)))
yhat5pre_23=ifelse(yhat5pre_23_0>=0.5,1,0)
b_6_23=di6+(alpha1[1]*yhat5pre_23)+(alpha2[1]*yhat[,4])
yhat6pre_23_0=(exp(b_6_23)/(1+exp(b_6_23)))
yhat6pre_23=ifelse(yhat6pre_23_0>=0.5,1,0)
b_7_23=di7+(alpha1[1]*yhat6pre_23)+(alpha2[1]*yhat5pre_23)
yhat7pre_23_0=(exp(b_7_23)/(1+exp(b_7_23)))
yhat7pre_23=ifelse(yhat7pre_23_0>=0.5,1,0)
b_8_23=di8+(alpha1[1]*yhat7pre_23)+(alpha2[1]*yhat6pre_23)
yhat8pre_23_0=(exp(b_8_23)/(1+exp(b_8_23)))
yhat8pre_23=ifelse(yhat8pre_23_0>=0.5,1,0)
################################ 24 Random effects models having no lag of response
id_24=rep(1:300,each=4)
y_24=matrix(rbind(yhat[,1],yhat[,2],yhat[,3],yhat[,4]),ncol=1)
x1_24=matrix(rbind(x[,1],x[,5],x[,9],x[,13]),ncol=1)
x2_24=matrix(rbind(x[,2],x[,6],x[,10],x[,14]),ncol=1)
x3_24=matrix(rbind(x[,3],x[,7],x[,11],x[,15]),ncol=1)
x4_24=matrix(rbind(x[,4],x[,8],x[,12],x[,16]),ncol=1)
data_24=matrix(cbind(id_24,y_24,x1_24,x2_24,x3_24,x4_24),ncol=6)
```

```
glmer_24=glmer( y_24 ~ x1_24+x2_24+x3_24+x4_24 + (1 | id_24),family = binomial)
ranef=ranef(glmer_24)
fixef=fixef(glmer_24)
fitted_y_24_0=fitted(glmer_24)
fitted_y_24=ifelse(fitted_y_24_0>=0.5,1,0)
b_524=fixef[1]+(fixef[2]*pre_x[,1])+(fixef[3]*pre_x[,2])+(fixef[4]*pre_x[,3])+(fixef[5]*pre_x[,4])+
ranef$id_24
yhat5pre_24_0=(exp(b_524)/(1+exp(b_524)))
yhat5pre_24=ifelse(yhat5pre_24_0>=0.5,1,0)
b_624=fixef[1]+(fixef[2]*pre_x[,5])+(fixef[3]*pre_x[,6])+(fixef[4]*pre_x[,7])+(fixef[5]*pre_x[,8])+
ranef$id_24
yhat6pre_24_0=(exp(b_624)/(1+exp(b_624)))
yhat6pre_24=ifelse(yhat6pre_24_0>=0.5,1,0)
b_724=fixef[1]+(fixef[2]*pre_x[,9])+(fixef[3]*pre_x[,10])+(fixef[4]*pre_x[,11])+(fixef[5]*pre_x[,1
2])+ranef$id_24
yhat7pre_24_0=(exp(b_724)/(1+exp(b_724)))
yhat7pre_24=ifelse(yhat7pre_24_0>=0.5,1,0)
b_824=fixef[1]+(fixef[2]*pre_x[,13])+(fixef[3]*pre_x[,14])+(fixef[4]*pre_x[,15])+(fixef[5]*pre_x[,
16])+ranef$id_24
yhat8pre_24_0=(exp(b_824)/(1+exp(b_824)))
yhat8pre_24=ifelse(yhat8pre_24_0>=0.5,1,0)
################################### 25 Random effects models  with lag of response
id_25=rep(1:300,each=3)
y_25_m1=matrix(rbind(yhat[,2],yhat[,3],yhat[,4]),ncol=1)
y_25_m2=matrix(rbind(yhat[,1],yhat[,2],yhat[,3]),ncol=1)
x1_25=matrix(rbind(x[,5],x[,9],x[,13]),ncol=1)
x2_25=matrix(rbind(x[,6],x[,10],x[,14]),ncol=1)
x3_25=matrix(rbind(x[,7],x[,11],x[,15]),ncol=1)
x4_25=matrix(rbind(x[,8],x[,12],x[,16]),ncol=1)
data_25=matrix(cbind(id_25,y_25_m1,x1_25,x2_25,x3_25,x4_25,y_25_m2),ncol=7)
glmer_25=glmer( y_25_m1 ~ x1_25+x2_25+x3_25+x4_25+y_25_m2 + (1 | id_25),family =
binomial)
ranef=ranef(glmer_25)
fixef=fixef(glmer_25)
fitted_y_25_0=fitted(glmer_25)
fitted_y_25=ifelse(fitted_y_25_0>=0.5,1,0)
b_5_25=fixef[1]+(fixef[2]*pre_x[,1])+(fixef[3]*pre_x[,2])+(fixef[4]*pre_x[,3])+(fixef[5]*pre_x[,4])
++(fixef[6]*y_25_m2)+ranef$id_25
yhat5pre_25_0=(exp(b_5_25)/(1+exp(b_5_25)))
yhat5pre_25=ifelse(yhat5pre_25_0>=0.5,1,0)
b_6_25=fixef[1]+(fixef[2]*pre_x[,5])+(fixef[3]*pre_x[,6])+(fixef[4]*pre_x[,7])+(fixef[5]*pre_x[,8])
+(fixef[6]*y_25_m2)+ranef$id_25
yhat6pre_25_0=(exp(b_6_25)/(1+exp(b_6_25)))
yhat6pre_25=ifelse(yhat6pre_25_0>=0.5,1,0)
b_7_25=fixef[1]+(fixef[2]*pre_x[,9])+(fixef[3]*pre_x[,10])+(fixef[4]*pre_x[,11])+(fixef[5]*pre_x[,
12])+(fixef[6]*y_25_m2)+ranef$id_25
yhat7pre_25_0=(exp(b_7_25)/(1+exp(b_7_25)))
yhat7pre_25=ifelse(yhat7pre_25_0>=0.5,1,0)
b_8_25=fixef[1]+(fixef[2]*pre_x[,13])+(fixef[3]*pre_x[,14])+(fixef[4]*pre_x[,15])+(fixef[5]*pre_x[
,16])+(fixef[6]*y_25_m2)+ranef$id_25
yhat8pre_25_0=(exp(b_8_25)/(1+exp(b_8_25)))
yhat8pre_25=ifelse(yhat8pre_25_0>=0.5,1,0)
```

APPENDIX E

R CODES FOR ESTIMATING THE FORECASTING ACCURACY OF BINARY RESPONSE VARIABLE

```
# cross tabulation of actual values versus predicted ones for the first half of the data
tab_17_1234=table(fitted_y_17,y_17)
# prediction error rate for the first half of the data
per_17_1234=((tab_17_1234[1,2]+tab_17_1234[2,1])/(tab_17_1234[1,1]+tab_17_1234[1,2]+
tab_17_1234[2,1]+tab_17_1234[2,2]))
# cross tabulation and prediction error rates for time points 5, 6, 7, 8 (second half of the data)
tab_17_5=table(yhat5pre_17,yhat5obs)
per_17_5=ifelse(length(tab_17_5)==4,((tab_17_5[1,2]+tab_17_5[2,1])/(tab_17_5[1,1]+tab_1
7_5[1,2]+tab_17_5[2,1]+tab_17_5[2,2])),(ifelse(length(tab_17_5)==2&&yhat5pre_17[1]==1,(t
ab_17_5[1]/(tab_17_5[1]+tab_17_5[2])),(tab_17_5[2]/(tab_17_5[1]+tab_17_5[2])))))
tab_17_6=table(yhat6pre_17,yhat6obs)
per_17_6=ifelse(length(tab_17_6)==4,((tab_17_6[1,2]+tab_17_6[2,1])/(tab_17_6[1,1]+tab_1
7_6[1,2]+tab_17_6[2,1]+tab_17_6[2,2])),(ifelse(length(tab_17_6)==2&&yhat6pre_17[1]==1,(t
ab_17_6[1]/(tab_17_6[1]+tab_17_6[2])),(tab_17_6[2]/(tab_17_6[1]+tab_17_6[2])))))
tab_17_7=table(yhat7pre_17,yhat7obs)
per_17_7=ifelse(length(tab_17_7)==4,((tab_17_7[1,2]+tab_17_7[2,1])/(tab_17_7[1,1]+tab_1
7_7[1,2]+tab_17_7[2,1]+tab_17_7[2,2])),(ifelse(length(tab_17_7)==2&&yhat7pre_17[1]==1,(t
ab_17_7[1]/(tab_17_7[1]+tab_17_7[2])),(tab_17_7[2]/(tab_17_7[1]+tab_17_7[2])))))
tab_17_8=table(yhat8pre_17,yhat8obs)
per_17_8=ifelse(length(tab_17_8)==4,((tab_17_8[1,2]+tab_17_8[2,1])/(tab_17_8[1,1]+tab_1
7_8[1,2]+tab_17_8[2,1]+tab_17_8[2,2])),(ifelse(length(tab_17_8)==2&&yhat8pre_17[1]==1,(t
ab_17_8[1]/(tab_17_8[1]+tab_17_8[2])),(tab_17_8[2]/(tab_17_8[1]+tab_17_8[2])))))
tab_17_5678=table(pre_17,obs_y_5678)
per_17_5678=ifelse(length(tab_17_5678)==4,((tab_17_5678[1,2]+tab_17_5678[2,1])/(tab_1
7_5678[1,1]+tab_17_5678[1,2]+tab_17_5678[2,1]+tab_17_5678[2,2])),(ifelse(length(tab_17_
5678)==2&&yhat5pre_17[1]==1,(tab_17_5678[1]/(tab_17_5678[1]+tab_17_5678[2])),(tab_17
_5678[2]/(tab_17_5678[1]+tab_17_5678[2])))))
per_17=cbind(per_17_1234,per_17_5,per_17_6,per_17_7,per_17_8,per_17_5678)
ccr_17=1-per_17
#confidence intervals at 0.05 signifance level for correctly classified ratio
z=1.96
ci_ccr_17_1234=c((1-per_17_1234)-(z*sqrt((per_17_1234*(1-per_17_1234)/(4*n)))),(1-
per_17_1234)+(z*sqrt((per_17_1234*(1-per_17_1234)/(4*n)))))
ci_ccr_17_5=c((1-per_17_5)-(z*sqrt((per_17_5*(1-per_17_5)/n))),(1-
per_17_5)+(z*sqrt((per_17_5*(1-per_17_5)/n))))
ci_ccr_17_6=c((1-per_17_6)-(z*sqrt((per_17_6*(1-per_17_6)/n))),(1-
per_17_6)+(z*sqrt((per_17_6*(1-per_17_6)/n))))
ci_ccr_17_7=c((1-per_17_7)-(z*sqrt((per_17_7*(1-per_17_7)/n))),(1-
per_17_7)+(z*sqrt((per_17_7*(1-per_17_7)/n))))
```

```
ci_ccr_17_8=c((1-per_17_8)-(z*sqrt((per_17_8*(1-per_17_8)/n))),(1-
per_17_8)+(z*sqrt((per_17_8*(1-per_17_8)/n))))
ci_ccr_17_5678=c((1-per_17_5678)-(z*sqrt((per_17_5678*(1-per_17_5678)/(4*n)))),(1-
per_17_5678)+(z*sqrt((per_17_5678*(1-per_17_5678)/(4*n)))))
ci_ccr_17=c(ci_ccr_17_1234,ci_ccr_17_5,ci_ccr_17_6,ci_ccr_17_7,ci_ccr_17_8,ci_ccr_17_
5678)
# expected percent correctly predicted and confidence interval for it
comb_17_1234=matrix(cbind(y_17,fitted_y_17_0),ncol=2)
comb_17_1234_0=comb_17_1234[comb_17_1234[,1]==0,]
comb_17_1234_1=comb_17_1234[comb_17_1234[,1]==1,]
comb_17_5=cbind(yhat5obs,yhat5pre_17_0)
comb_17_6=cbind(yhat6obs,yhat6pre_17_0)
comb_17_7=cbind(yhat7obs,yhat7pre_17_0)
comb_17_8=cbind(yhat8obs,yhat8pre_17_0)
comb_17_5_0=comb_17_5[comb_17_5[,1]==0,]
comb_17_6_0=comb_17_6[comb_17_6[,1]==0,]
comb_17_7_0=comb_17_7[comb_17_7[,1]==0,]
comb_17_8_0=comb_17_8[comb_17_8[,1]==0,]
comb_17_5_1=comb_17_5[comb_17_5[,1]==1,]
comb_17_6_1=comb_17_6[comb_17_6[,1]==1,]
comb_17_7_1=comb_17_7[comb_17_7[,1]==1,]
comb_17_8_1=comb_17_8[comb_17_8[,1]==1,]
comb_17_5678_0=matrix(rbind(comb_17_5_0,comb_17_6_0,comb_17_7_0,comb_17_8_0),
ncol=2)
comb_17_5678_1=matrix(rbind(comb_17_5_1,comb_17_6_1,comb_17_7_1,comb_17_8_1),
ncol=2)
epcp_17_1234=(sum(1-comb_17_1234_0[,2])+sum(comb_17_1234_1[,2]))/(4*n)
epcp_17_5=(sum(1-comb_17_5_0[,2])+sum(comb_17_5_1[,2]))/n
epcp_17_6=(sum(1-comb_17_6_0[,2])+sum(comb_17_6_1[,2]))/n
epcp_17_7=(sum(1-comb_17_7_0[,2])+sum(comb_17_7_1[,2]))/n
epcp_17_8=(sum(1-comb_17_8_0[,2])+sum(comb_17_8_1[,2]))/n
epcp_17_5678=(sum(1-comb_17_5678_0[,2])+sum(comb_17_5678_1[,2]))/(4*n)
ci_epcp_17_1234=c((epcp_17_1234)-(z*sqrt((epcp_17_1234*(1-
epcp_17_1234)/(4*n)))),(epcp_17_1234)+(z*sqrt((epcp_17_1234*(1-epcp_17_1234)/(4*n)))))
ci_epcp_17_5=c((epcp_17_5)-(z*sqrt((epcp_17_5*(1-
epcp_17_5)/n))),(epcp_17_5)+(z*sqrt((epcp_17_5*(1-epcp_17_5)/n))))
ci_epcp_17_6=c((epcp_17_6)-(z*sqrt((epcp_17_6*(1-
epcp_17_6)/n))),(epcp_17_6)+(z*sqrt((epcp_17_6*(1-epcp_17_6)/n))))
ci_epcp_17_7=c((epcp_17_7)-(z*sqrt((epcp_17_7*(1-
epcp_17_7)/n))),(epcp_17_7)+(z*sqrt((epcp_17_7*(1-epcp_17_7)/n))))
ci_epcp_17_8=c((epcp_17_8)-(z*sqrt((epcp_17_8*(1-
epcp_17_8)/n))),(epcp_17_8)+(z*sqrt((epcp_17_8*(1-epcp_17_8)/n))))
ci_epcp_17_5678=c((epcp_17_5678)-(z*sqrt((epcp_17_5678*(1-
epcp_17_5678)/(4*n)))),(epcp_17_5678)+(z*sqrt((epcp_17_5678*(1-epcp_17_5678)/(4*n)))))
epcp_17=cbind(epcp_17_1234,epcp_17_5,epcp_17_6,epcp_17_7,epcp_17_8,epcp_17_567
8)
ci_epcp_17=c(ci_epcp_17_1234,ci_epcp_17_5,ci_epcp_17_6,ci_epcp_17_7,ci_epcp_17_8,
ci_epcp_17_5678)
# area under Receiver Operating Characteristic curve
auroc_17_1234=somers2(fitted_y_17_0,y_17)["C"]
auroc_17_5=somers2(yhat5pre_17_0,yhat5obs)["C"]
auroc_17_6=somers2(yhat6pre_17_0,yhat6obs)["C"]
auroc_17_7=somers2(yhat7pre_17_0,yhat7obs)["C"]
auroc_17_8=somers2(yhat8pre_17_0,yhat8obs)["C"]
auroc_17_5678=somers2(cbind(yhat5pre_17_0, yhat6pre_17_0, yhat7pre_17_0,
yhat8pre_17_0), obs_y_5678) ["C"]
auroc_17=cbind(auroc_17_1234,auroc_17_5,auroc_17_6,auroc_17_7,auroc_17_8,auroc_1
7_5678)
# deviance
```

```
dev_17_1234=-2*(sum(log(1/comb_17_1234_1[,2]))+sum(log(1/(1-comb_17_1234_0[,2]))))
dev_17_5=-2*(sum(log(1/comb_17_5_1[,2]))+sum(log(1/(1-comb_17_5_0[,2]))))
dev_17_6=-2*(sum(log(1/comb_17_6_1[,2]))+sum(log(1/(1-comb_17_6_0[,2]))))
dev_17_7=-2*(sum(log(1/comb_17_7_1[,2]))+sum(log(1/(1-comb_17_7_0[,2]))))
dev_17_8=-2*(sum(log(1/comb_17_8_1[,2]))+sum(log(1/(1-comb_17_8_0[,2]))))
dev_17_5678=-2*(sum(log(1/comb_17_5678_1[,2]))+sum(log(1/(1-comb_17_5678_0[,2]))))
dev_17=cbind(dev_17_1234,dev_17_5,dev_17_6,dev_17_7,dev_17_8,dev_17_5678)
```