Employee turnover forecasting for human resource management

based on time series analysis

In some organizations, the lead time of hiring is often long due to responding to human resource requirements associated with technical and security constrains. Thus, the human resource departments in these organizations are pretty interested in forecasting employee turnover since a good prediction of employee turnover could help the organizations to minimize the costs and impacts from the turnover on the operational capabilities and the budget. This study aims to enhance the ability to forecast employee turnover with or without considering the impact of economic indicators. The research utilizes various statistical modelling techniques in time series to identify optimal models for effective employee turnover prediction. More than 11-years of monthly turnover data are used to build and validate the proposed models. Compared with other models, the dynamic regression model with additive trend, seasonality, interventions, and a very important economic indicator is able to predict the turnover efficiently with training R2=0.77 and holdout R2=0.59. The forecasting performance of optimal models in this study confirms that time series modelling approach has a reasonable ability to predict employee turnover for various scenarios.

**Keywords**: human resource management; turnover; time series; forecast

# 1. Introduction

Prediction of employee turnover is a topic that has drawn the attention of researchers and human resource managers because employee turnover cost impacts both the operational capabilities and the budget of an organization. Employee turnover is both costly and disruptive to the functioning of most organizations, and both private firms and governments spend billions of dollars every year managing the issue according to Leonard [1].Furthermore, at crucial times, organizations find themselves short of niche skill-sets and resources which require time and planning to acquire. The lead time for hiring is often long, particularly when special skills are involved, and in some organizations like U.S. national laboratories, due to the security clearance requirements and training, the process could take months. Therefore, a good prediction of employee turnover at firm and departmental levels is essential for effective planning, budgeting, and recruiting in the human resource field.

Human resource planning (HRP) is an ongoing process of systematic planning to achieve optimum use of the human resource pool in an organization. For an organization to execute their tasks efficiently and effectively with the highest quality, they need to ensure that the right people are available at the right places and at the right time [2]. Over the years, organizations have been able to scale up their efforts and success in manufacturing, marketing and financial plans. However, organizations have always struggled to develop sustainable HRP models [3]. The objective of a sustainable HRP model is to ensure the best match between employees and jobs to avoid manpower shortages or surpluses [4]. To achieve this balance employee turnover is an important metric that is often central to organizations workforce planning and strategy.

As summarized in Table 1, some of the previous studies attempt to identify the explanatory predictors of employee turnover. For instance, Bluedorn [5] found that the turnover appears to be related to the individual’s perception of environmental opportunities, routine, age, and the length of service. Balfour and Neff [6] suggested that caseworkers with more education, less experience, and less stake in an organization are more likely to turnover. According to the research conducted by Wright and Cropanzano [7], emotional exhaustion is associated with both job performance and subsequent turnover, but not related to job satisfaction. Morrow, McElroy [8] used employee absenteeism and performance to predict employee turnover. The result from this study shows a positive correlation between absenteeism and voluntary turnover, and a negative correlation between performance ratings and voluntary turnover. The study conducted by Thaden, Jacobs-Priebe [9] indicated that organizational culture may potentially be an important factor for retaining workers in an organization.

Other insights have been gained from some more recent research. For instance, according to Tews, Stafford [10], personal events, professional events, internal work events and constituent attachment were highly related to turnover. Collini, Guidroz [11] found that the interaction between interpersonal respect, mission fulfilment and engagement are statistically significant predictors for turnover in health care. However, diversity climate was not related with turnover. Finally, only Sexton et al (2005) considered outside economic variables, unemployment index and consumer price index, in the forecasting model for employee turnover. However, their final model did not include these two variables. A recent editorial by Ferrara and van Ferrara and van Dijk [12] in the International Journal of Forecasting revealed new interest in forecasting business cycles with some complex methodologies. However, forecasting business cycle turning points is quite difficult, and Hamilton [13] suggests that “the best econometricians can do is probably to nowcast recessions; that is, to recognize a turning point as soon as it occurs, or soon thereafter.” Some type of outside variable might facilitate this situation.

Meanwhile, some other studies try to build turnover prediction models through techniques, such as regression, neural network (NN), and data mining. For example, Ng, Cram [14] used a proportional hazards regression (PHR) to develop turnover prediction model. In the study conducted by Sexton, McMurtrey [15], NN combining with modified genetic algorithm was used to build the prediction model for turnover. Alao and Adeyemo [16] applied decision tree on the employees’ demographical information and personnel records to identify attributes contributing to employee turnover.

In these studies, the data source was acquired either from human resource employment records and demographical information or surveys with time horizons ranging from 1 to 28 years. Most of the data is monthly which is ideal for time series forecasting models.

Although there have been some efforts to predict employee turnover behaviour, no study has been conducted on predicting employee turnover with time series forecasting techniques. In this study, the authors attempt to fill this research gap. The advantage of time series forecasting approach is that it is not necessary to identify the determinants of turnover and is helpful to evaluate the effects of either a planned or unplanned intervention [17].

Thus, this article has four parts. First, there is the introduction which covers the objective of the paper and a literature review. Second, there is a synopsis of tools for finding time series patterns and preparing the data for analysis as well as specific forecasting methods. The methods section is followed by the results and discussion of the study. Finally, there is conclusion of the results, practical implications, and limitations of the paper.

Table 1. Summary of previous research on employment turnover forecast.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Authors (Year) | Data Acquisition | Data Horizon | Methods | Software | Economic Indicator | Response Variable | Estimate | Model Evaluation |
| Bluedorn [5] | Employee records and Survey | 1 year | Correlations, multiple regression | N/A | No | Number | Point with intervals | R2=0.22,  Adjusted R2=0.11 |
| Ng et al. [14] | Survey | N/A | Hazard proportional model | BMDP 2L | No | Probability | Point with intervals | Pair t-test |
| Balfour et al. [6] | Employee records | 33 months | Non-linear logistic regression | N/A | No | Probability | Point | Chi-square values |
| Feeley et al. [18] | Survey | 60 months | Social network, logistic regression, correlation | NEGOPY, UCINET | No | Probability | Point | R2=0.23 |
| Wright et al. [7] | Survey | 1 year | Hypothesis test, correlation, logistic regression | N/A | No | N/A | N/A | Correlation r=0.34, P<0.01 |
| Morrow et al. [8] | Demographic information and employee records | 2 years | Logistic regression, correlation | N/A | No | Probability | Point | (-2 log likelihood) chi-square=193.13 |
| Sexton, McMurtrey [15] | Demographic information and employee records | 10 years (yearly) | NN | FORTRAN | Yes | Leave or not | Point | Type I error=0.25% Type II error= 5.83% |
| Hong et al. [19] | Survey | N/A | Logit and probit model | SPSS | No | Probability | N/A | R2=0.5, Quadratic Probability Scores = 0.18 for training and 0.12 for test |
| Nagadevara et al. [20] | Demographic information and employee records | 3 years | NN, logistic regression, classification/regression trees, discriminant analysis | N/A | No | Leave or not | Point | Contingency table |
| Thaden et al. [9] | Survey | 2 years | Multiple regression | N/A | No | Duration | Point with intervals | R2=0.56, P< 0.001 |
| Größler and Zock [21] | Employee records | 360 months | System dynamics | N/A | No | Number | Point | N/A |
| Saradhi et al. [22] | Survey | 2 years | SVMs, random forest, Naïve Bayes classifiers | N/A | No | Probability | Point | True/false positive rate and precision |
| Alao et al. [16] | Employee records | 28 years (yearly) | Decision tree | WEKA See5 | No | Probability | Point | True/false positive rate and precision |
| Tews et al. (2014) | Employee records and Survey | 6 months | Logistic regression | N/A | No | Probability | Point | R2=0.23 |
| Collini et al. (2015) | Survey and turnover rates | 1 year | Correlation and Regression | N/A | No | Turnover rates | Point | No |

# 2. Methods

# *2.1 Data sources and preparation*

The human resource data was provided by a large multipurpose research organization in the U.S. The dataset consists of over 8,000 observations. These observations are active and terminated employees’ with incomplete demographic information from November 2000 to January 2012 including metrics such as payroll category, hired date, termination date, age, years of service, gender, job classification, and department code. The turnover dataset is summarized in the form of monthly data where each field represents the total number of employees leaving the organization. This dataset consists of 135 months spanning from November 2000 to January 2012. For this research, turnover is defined as total number of employees leaving the organization in each month. This definition of turnover is used as a unit of measurement for the turnover prediction.

Several economic indicators have been examined in this study, such as, unemployment rate index, New York Stock Exchange, U.S gasoline price, and U.S monthly composite leading indicator (CLI). However, only CLI published by Organization for Economic Co-operation and Development (OECD) from November 2000 to January 2012 significantly improves the forecasting performance as a predictor of the cyclical component of employee turnover, since, in practice, CLI is used to give an early indication of turning points in the macro-economic cycle. CLI data used for this study was constructed by OECD through aggregating 7 components together. These 7 components include the number of dwellings started, net new orders for durable goods, share prices-NYSE composite, consumer sentiment indicator, the average weekly hours worked by manufacturing workers, the purchasing manager index, and spread of interest rates [23]. The authors have no information on weights of the index construction.

# *Pattern analysis, cross-correlations, and outlier identification*

To model a time series, it is important to look for patterns in the turnover series. First, this pattern analysis is initially applied with a time plot of the series and box plots for the seasons or the months. In this case, the seasonal pattern of the turnover series is tested through Kruskal-Wallis and ANOVA tests (p-value<0.05), which do not correct for any trend in the series. The second stage of the pattern analysis is to use autocorrelation (ACF) and partial autocorrelation (PACF) plots to identify seasonal, autoregressive, and moving average patterns. If there are external variables (as in this case), the third stage of pattern analysis would be to examine the cross-correlations between turnover series (and external variables (CLI () over time). The cross-correlation function (CCF) is used to identify lags of CLI () that might be useful predictors of turnover series ( A longer lag that is strong enables the forecast horizon to be longer when using external variable.

All of the previous stages of the pattern analysis could be contaminated by outliers so it is important to identify outliers before fitting the actual forecasting model. Box plot analysis by seasons can be used to informally flag outliers (this approach tends to over-identify outliers) while ARIMA methods in conjunction with statistical process control (SPC) tend to correctly identify the number of outliers in a series. Conservative ARIMA models ((1,0,1) or (0,1,1)) are used in this control charting. First, the residuals from ARIMA(1,0,1) are divided by the root mean square error to standardize them, and then the outliers are identified with the value greater than ±3 standard deviations from zero by scatterplots of the standardized residuals [24, 25]. Identifying the outliers using SPC and then smoothing them is a way to refine the data for further analysis and to facilitate finding the underlying pattern in the data series. Smoothing outliers in a time series is primary to filter out random noise and other irregularities. When the outlier is identified, it is adjusted to be more similar to its neighboring points [25]. In this study, unusual observations were smoothed using a nonlinear data smoothing method, based on repeated medians (RMD) of a five period span (as shown in equation (1) [26]):

(1)

where is the actual smoothed value at time *t*, is value of response variable at time *t*, and *s* is the number of total seasonal periods which is 12. This smoothing is only done on potential outliers within their season and not across adjacent periods. Sometimes outliers can distort normality, white noise, cross correlations, ACF and PACF, and the predictive performance of the model. Thus, dealing with outliers properly means going back and asking hard questions about these unusual observations from a human resource perspective, and that can truly add understanding and improved forecasting ability. It is always possible that the outliers are not always be unusual events but interventions or change points that need to be accommodated in the model. In this kind of human resource dataset, such abnormalities could be retirement incentives were offered, another company was purchased, a section of the original company was sold off, or the potential outlier might reflect an economic down turn. There are many possibilities; but if there is no over-identification of outliers, a good human resource department should be able to provide quick answers on the unusual situations.

# *2.2 Time series analysis*

In time series forecasting, past observations of the variable being analysed are collected and analysed to develop a model describing the pattern. This modelling approach is particularly useful when little knowledge is available on the data generating process or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables [27]. In this study, basically univariate and multivariate time series methods are used to identify an optimum forecast. The statistical softwares used were Number Cruncher Statistical System (NCSS), SAS, and R. The data is analysed on two partitions of the data: training sample (November 2000 – January 2011) and holdout sample (February 2011- January 2012). For each model, the training sample is used to build the model and the holdout sample is used for the validation of the model because the most recent time series data is considered the most important factor for prediction purposes [28].

### 2.2.1 Univariate methods (without external variables)

These univariate time series analysis models use months for seasonality and trend as predictors of turnover. The univariate models without external variables which were used in this study include time series regression, decomposition, Winter’s Exponential Smoothing (WES), and Box-Jenkins Autoregressive Integrated Moving Averages (ARIMA). One critical reason for selecting these kinds of time series models was to avoid any restriction on the forecasting horizon, i.e., how far in the future one could predict.

### Time series regression

The univariate time series regression model uses trend and seasonality as two predictors. The additive time series regression model with intercept, trend, monthly seasonality, and error terms takes the form as shown in equation (2):

(2)

where is the value of response variable at time *t*, is the coefficients estimated by regression, is a continuous variable representing trend with value from 1 to *n*, and is dummy variable representing seasonal periods. In addition, the additive regression models with interventions: pulse or steps are also considered in the analysis due to the downsize policy in certain time points in our case, which are denoted by dummy variable in the regression function. The multiplicative time series regression model with intercept, trend, monthly seasonality, and error terms is also considered as shown in equation (3):

(3)

where is natural log transformation of. For these regression models, the significance of the model and the variables is examined by p-value at 0.05 significant level, lack of collinearity, good holdout performance, and valid regression assumptions.

### Decomposition

Decomposition time series methods attempt to separate the series into four components: trend, cycle, seasonality, and irregularity. Multiplicative decomposition methods can be described globally as equation (4). The decomposition model can be used assuming no cyclical variation or the cyclical variation can be extracted and fit a model to enhance the forecasting. Of course, there can be quite complex decomposition models, but the classical multiplicative decomposition model in NCSS was used here for this turnover series for an easier application.

(4)

### Winters exponential smoothing (WES)

WES models work well on series that either have seasonality or seasonality and trend. The models can be additive or multiplicative as well, but the preferred option tends to have additive trend and either additive or multiplicative seasonality. Multiplicative trend in these kinds of time series models tends to over or under forecast for future values. Of course, dampened models can help avoid this future forecast problem, but one must be careful as that the forecasts will be short forecast horizon [29]. The robustness and accuracy of Winter’s exponential smoothing methods has led to their widespread use in applications where a large number of series necessitates an automated procedure [30, 31]. WES model is easy to interpret and is easily understood by management or even those who lack any technical training in forecasting.

1. *Autoregressive Integrated Moving Average (ARIMA)*

The ARIMA models were introduced by the statisticians George Box and Gwilym Jenkins [32]. The general form of ARIMA models is ARIMA(p,d,q); where *p*, *d*, and *q* are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. In addition, ARIMA models can handle seasonality, and their forms would be as follows: ARIMA(p,d,q)(P,D,Q). Thus, the seasonal aspect of a series could have an autoregressive, differencing, or moving average patterns as well. ARIMA methods are popular to some forecasters because they provide a wide class of models for univariate time series forecasting [33].

### 2.2.2 Univariate methods (with external variables)

Univariate models that incorporated an external variable (CLI) as a predictor of cyclical component of turnover series were also examined. Basically, this included dynamic regression and more complex decomposition models.

1. *Dynamic regression with external variable*

The dynamic regression model describes how the forecasting output is linearly related to current and past values of one or more input series. There are two crucial assumptions for dynamic regression model. First, the observations of the input series are assumed to occur at equally spaced time intervals. Second, the input series are not affected by the output [34]. Dynamic regression models allow one to include external variables, interventions, and transfer functions. In this study, the external variables (CLI and interventions) are incorporated into the dynamic regression model. Equation (5) can be a simple representation of the model with trend, seasonal dummy variable, and interventions,

(5)

where is dummy variable representing pulse and step periods, and is dummy variable representing pulse periods. Also,, , and are the change point coefficients estimated by regression and refers to the delayed rise or fall in the forecast variable.

1. *Decomposition with external variable*

In this more complex decomposition model, CLI and its lag terms are used as a predictor of cyclical component in the decomposition model. This research applies two approaches to obtain a decomposition model: one is the automatic decomposition in NCSS with the cyclical variable (CLI) incorporated and the other is a multiplicative decomposition model built by the product of the best ARIMA and cyclical factor (CLI).

### 2.2.3 Nonlinear and multivariate methods

1. *State space model*

A state space model consists of an observation equation as shown in equation (6) and a Markovian transition equation as shown in equation (7),

(6)

(7)

where is a data vector; is a unknown state-vector, is a state vector relating the observation data to the state vector , and is a state transition matrix. In addition, are random error matrices which independently and identically follow a multinomial distribution. State space models are modelled using R “stats”, “dlm”, and “forecast” packages.

1. *Vector autoregressive model*

Vector autoregressive model (VAR) is an econometric model for multivariate time series analysis. It is an extension of the univariate autoregressive model, i.e., each variable is represented as a linear function of its lags and lags of the other variables. VAR models are often used to describe and forecast financial and economic time series. A VAR model consists of a set of variables (also called endogenous variables) , denoted as vector. A pth order VAR model is

(8)

where there are p ( coefficient matrices. Also, is a unobservable white noise vector process where expectation value of this vector is zero and time invariant covariance matrix . The VAR model is modelled by R “vars” package.

### 2.2.4 Model evaluation

A good forecasting model should be evaluated on predictive ability, goodness of fit using the R2 value, mean absolute percentage error (MAPE), mean absolute error (MAE) or other fit diagnostics, normality tests on residuals, and a white noise test on those same residuals to make sure that no time series pattern is left. The best fitting model is generally selected based on higher R2 value in the holdout data, lower MAPE, normally distributed residuals, and a passed white noise test. Of course, there are more recent measures of forecasting accuracy [35], but the authors tried to keep it simple.

1. *Pseudo R2*

For evaluation of the time series methods, the pseudo R2 for training and holdout data is calculated as standard criteria to test the goodness of model fitting taking the form in equation (9):

(9)

1. *Mean absolute percentage error*

MAPE is another measure of accuracy of the time series model fitting methods as shown in equation (10) [36, 37]. This criterion is used to compare the model performance for the specific dataset by using time series methods, since it measures relative performance [38].

(10)

1. *Normality test*

A good time series model should have normally distributed residuals. In this paper, normality of residuals is evaluated by two powerful normality tests: Shapiro-Wilk [39] and D'Agostino Omnibus normality test [40] in NCSS.

1. *White noise test*

The white noise test is performed on the residuals to check for any undetected time series pattern still remaining in the residuals and not being accounted for by the model. Independent residuals, random scatter of those residuals, or no time series pattern in residuals [41] is always preferred when choosing a best model. In practice, the Q-statistic (also called Box-Pierce statistic or Ljung-Box statistic) is used as an objective diagnostic measure of white noise for a time series to compare whether the autocorrelations from residuals and white noise are statistically significantly different. This test statistic is illustrated in equation (11):

(11)

where *k* is selected from the lesser of two seasonal cycles, or one-fourth of the observations, or 24 when two seasonal cycles is much greater than 24. In most cases, if a model is lacking in white noise, it means this model is deficient and has to be rectified [29].

# 3. Results and discussion

# *3.1 Patterns of employ turnover and outlier identification*

The time based turnover series was restructured in a format so that it could be analysed to observe patterns like trend or seasonality and understand any inherent time series in the data for further investigation in that direction. As shown in Figure 1 (a), there is a clear presence of seasonality pattern in the series. The box plot for turnover series from ANOVA test confirms this seasonality pattern. In addition, there is a decreasing trend from January to November as shown in Figure 1 (b), and then the trend line rises up in December. The points with year label of 2008 in Figure 1 (b) are considered as outliers, since they are beyond the upper whiskers. After all ARIMA models were tested, three outliers appeared in SPC chart (as shown in Figure 2); and two of these three were also flagged by the box plot analysis. The outliers identified by box plots and SPC include the turnover numbers for December 2001, March 2008, April 2008, May 2008, June 2008, August 2008, and September 2008. December 2001 and March 2008 to July 2008 are treated as temporary pulse and steps while August 2008 is treated as pulse with gradual decay in the dynamic regression with intervention methods. In other methods, these outliers are smoothed to soften their impacts, which are univariate regression without intervention, decomposition, exponential smoothing, ARIMA, state space, and vector autoregressive methods. Whenever there are outliers, one needs to investigate the cause. For instance, December 2001 is a 9/11 lag impact on job hiring with stronger background checking and increased retiring/hiring. The outliers in 2008 reflect the downsize policy issued in January 2008 with three months response time window to accommodate a voluntary reduction in workforce from the organization.

C:\Users\julia\Dropbox\Timeseries paper\figure\FIG1_TSBOXPLOT_FINAL_2.tif

Figure 1. (a) Monthly turnover series plotted over time and (b) box plot of turnover data.

Figure 2. SPC chart for standardized residuals.

The ACF and PACF plots for the turnover series are shown in Figure 3 (a, b). The pattern of unsmoothed data in the ACF and PACF hints at ARIMA(1,0,1) with some type of seasonality, but the seasonal pattern is not obvious.



Figure 3. (a) Autocorrelation and (b) partial autocorrelation plots.

# *3.2 Cross-correlations*

The cross-correlation is applied between turnover series and CLI series to identify significant lag correlations. The cross-correlations between first differences for turnover and the CLI series were examined [29], and a “pre-whitening” process for the two series was used to identify the cross-correlation patterns [32, 42, 43]. Based on all these calculations, CCFs from Lag 0 to Lag 8 were found to be statistically significant, which indicates that the turnover series has statistically significant correlation with CLI and its 8 lags. The CLI and its 8 lags are incorporated into the dynamic regression, decomposition, and ARIMA model respectively when using the cyclical variable for forecasting.

# *3.3 Forecasting results and comparisons*

Forecasting evaluations for the time series models are provided in the Appendix Table A. 1, Table A. 2, Table A. 3, and Table A. 4. Based on the evaluation statistics, eight univariate models are selected because of an acceptable R2 value for the training and holdout data as well as their residual statistics that are the optimum among the other models (as shown in Table 2). On average, the holdout R2 value of these eight univariate models is 0.51 (range from 0.40 to 0.59).

### 3.3.1 Univariate methods (without external variables)

The regression model with additive trend and seasonality has the highest holdout R2 (0.57) among the univariate models without external variables, indicating the model’s ability to explain 57% of the total variation of the holdout sample. It is statistically significant (p-value<0.05) for the model and parameters. However, the residuals are not normally distributed, and the model does not pass the white noise test.

Table 2. Statistics for selected univariate time series models.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Method | # | Model | Pred.1 R2 | Holdout R2 | MAPE | Normality2 | WN3 |
| Univariate without external variable | U1 | Regression with additive trend and seasonality | 0.51 | 0.57 | 26.15 | No | No |
| U2 | Regression with additive trend, seasonality and intervention4 | 0.72 | 0.52 | 22.84 | Yes | No |
| U3 | Decomposition | 0.65 | 0.54 | 17.97 | Yes | Yes |
| U4 | WES with additive trend and seasonality | 0.52 | 0.52 | 20.65 | Yes | Yes |
| U5 | ARIMA(1,0,1)(0,1,1) | 0.47 | 0.40 | 22.89 | Yes | Yes |
| Univariate with external variable | V1 | Dynamic regression using lag7 CLI as predictor4 | 0.77 | 0.59 | 19.91 | Yes | Yes |
| V2 | Decomposition using lag1 CLI as cycle | 0.65 | 0.55 | 17.97 | Yes | Yes |
| V3 | ARIMA combining with lag1 CLI as cycle | 0.37 | 0.41 | 22.73 | Yes | Yes |

Notes:

1. Pred. R2 is prediction R2 value for training data.
2. Normality is residuals’ normality test.
3. WN is white noise test.
4. The data is unsmoothed (or outliers are unadjusted) so as to take advantage of time series models that can accommodate interventions.

The regression model with additive trend, seasonality, and interventions (pulse and step) performs well with a training R2 of 0.72 and a holdout R2 of 0.52, indicating the model’s ability to explain 72% of the total variation of turnover for the training data and 52% of the total variation of the holdout sample. It is statistically significant (p-value<0.05) for the model and parameters. The model has normally distributed residuals, but it does not have white noise. However, this regression model can capture the spike in December 2001 and sharp fluctuations from March 2008 to August 2008 (as shown in the Figure 4).

Figure 4. Regression with interventions forecast vs. actual turnover number plot.

The decomposition model without external variables is considered as the best univariate model without external variables because this model has a reasonably high training R2 value (0.65), a good holdout R2 value (0.54), and low MAPE (17.97). The residuals of this model are normally distributed and have a white noise pattern. Figure 5 shows the predicted turnover based on the decomposition model and the actual turnover number for holdout dataset. This plot validates the holdout performance of the decomposition model as it is able to mimic the changes in trend and seasonality of the turnover and prediction is close to the actual turnover numbers. However, it does seem to under-forecast for the 6 months June through November.

Figure 5. Decomposition validation forecast vs. actual turnover number with prediction intervals.

### 3.3.2 Univariate methods (with external variables)

According to the cross-correlation analysis results, CLI and its 8 lags are applied in dynamic regression, decomposition model, and ARIMA(1,0,1)(0,1,1) respectively as external variable to forecast turnover number. The dynamic regression model with additive trend, seasonality, interventions (pulse and step with gradual decay), and lag7 of CLI is the best model among all models, since it has highest training and holdout R2 value (0.77 and 0.59), normal residuals, and white noise. The dynamic regression is globally statistically significant for the model and individually significant for the parameters (p-value<0.05). Figure 6 shows the predicted and actual turnover number plots for holdout dataset from the dynamic regression model. Although there are under forecasts from July to September as well, the differences between forecasting values and actual values have become much tighter. Compared with top rated univariate methods without external variables, the performance of dynamic regression model is much improved after using CLI as an outside cyclical variable.

Figure 6. Dynamic regression with lag7 CLI forecast vs. actual turnover number for holdout dataset.

### 3.3.3 Nonlinear and multivariate methods (with external variables)

State space models with various combinations of errors, trend, seasonality, or an exogenous variable (CLI here) and VAR models of bivariate time series (turnover and CLI) are employed to forecast turnover number (as shown in Table 3). The exponential smoothing state space model with multiplicative error, no trend, and multiplicative seasonality has the highest holdout R2 value (0.43) among all state space models, which is automatically selected by R forecast package from 27 exponential smoothing state space models. The residuals have white noise and are normally distributed. However, the training and holdout R2 value (0.53 and 0.43) are relatively lower than the univariate regression models.

Table 3 shows that the VAR models perform better than state space models. The VAR (4, constant, trend, and seasonality) is considered as the best among all these nonlinear and multivariate model with highest training and holdout R2 value (0.62 and 0.64). The residuals are normally distributed and have white noise. However, variables in this model: lag terms of turnover and CLI are not statistically significant, indicating the model is over fitted. Only VAR (1, constant, trend, and seasonality) has significant lag 1 term of both turnover and CLI. This model also has higher training and holdout R2 value (0.57 and 0.47) and white noise. However, its residuals are not normally distributed.

The bivariate time series (turnover and CLI) does not have multivariate moving average pattern, which is not statistically significant. Therefore, the vector moving average (VMA) and vector autoregressive moving average (VARMA) methods are not employed to forecast. The volatility models (Garch (1,1) and stochastic volatility model) and nonlinear models (nonlinear autoregressive model and nonlinear threshold autoregressive model) have also been considered. However, all these models have lower training and holdout R2 values and higher MAPE values. Their statistics are not provided in the Table 3 due to their poor performance.

Table 3. Statistics for selected nonlinear and multivariate models.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Method | # | Model | Pred. R2 | Holdout R2 | MAPE | Normality | WN |
| State Space | S1 | Trend, Seasonality | 0.27 | 0.39 | 24.80 | No | Yes |
| S2 | Trend, Slope, Seasonality | 0.23 | 0.32 | 28.58 | Yes | Yes |
| S3 | Trend, Slope, Seasonality, CLI as regressor | 0.13 | 0.32 | 28.89 | Yes | Yes |
| S4 | Exponential Smoothing (M,N,M)1 | 0.51 | 0.43 | 21.14 | Yes | Yes |
| S5 | Structural TS (Level, Slope) | 0.46 | -0.16 | 22.82 | No | No |
| S6 | Structural TS (Level, Slope, Seasonality) | 0.51 | 0.04 | 21.47 | Yes | Yes |
| Vector Autoregressive2 | VA1 | VAR(5, Constant, Seasonality) | 0.62 | 0.49 | 18.20 | No | Yes |
| VA2 | VAR(5, Trend, Seasonality) | 0.63 | 0.52 | 18.37 | Yes | Yes |
| VA3 | VAR(1, Constant, Trend, Seasonality) | 0.52 | 0.47 | 21.08 | No | Yes |
| VA4 | VAR(2, Constant, Trend, Seasonality) | 0.59 | 0.44 | 19.63 | Yes | Yes |
| VA5 | VAR(3, Constant, Trend, Seasonality) | 0.61 | 0.51 | 19.06 | Yes | Yes |
| VA6 | VAR(4, Constant, Trend, Seasonality) | 0.62 | 0.64 | 18.44 | Yes | Yes |
| VA7 | VAR(5, Constant, Trend, Seasonality) | 0.65 | 0.61 | 17.98 | Yes | Yes |

Note:

1. M, N, M is multiplicative errors, no trend, multiplicative seasonality, respectively.
2. VAR(p) is pth order of lag term.

Even more non-linear and multivariate models were considered, but the intent of this research was not to go fishing for a best model but to find a simplistic model that could be used by human resource management (HRM). It should be noted that a combination model might have been the better than our chosen dynamic regression model, but the authors were trying to keep it as simple as possible for HRM.

**4. Conclusions**

In this paper, various time series forecasting models for employee turnover prediction are tested and optimal models for turnover forecasts are identified. The model in this paper actually performs better than those accessed in the literature review as a result of the external variable. Although VAR (4, constant, trend, and seasonality) has the highest holdout R2, normally distributed residuals and white noise, dynamic regression model is concluded as the best forecasting model. There are several reasons why univariate methods are selected. In most cases, multivariate models help in generating more accurate model fit when compared to univariate models. However, univariate models are preferred as they are able to negate several drawbacks of multivariate models. For example, univariate models have less parameter uncertainty and less chance for outliers and errors due to their design simplicity. In most cases, univariate models are relatively easier to develop, interpret and get concrete conclusions. Multivariate models due to their complexity are more susceptible to misspecification. Besides, the explanatory variables in univariate models have to be determined accurately before forecasting the depending variable. Errors in forecasting the explanatory variable for a multivariate model may significantly affect the accuracy of the forecasts of the dependent variable when compared with an equivalent univariate model [44]. Apart from the benefits listed for univariate models, the dynamic regression model has several additional advantages. For example, an ARIMA error term which has autocorrelation pattern can be included in the model. Dynamic regression model is able to handle lagged regressors and various types of seasonalities. In addition, dynamic regression model can handle interventions or change points effectively, since these interventions or change points, such as holidays, promotions, new policy and so forth, are often common to the time series data. Thus, dynamic regression model could be used to forecast turnover for most of organizations of any size. However, to implement dynamic regression modelling, at least 5-years of monthly employee turnover data is preferred to make an accurate forecast. If the horizon of the dataset is less than 5 years, a special decomposition model [45] could be considered as a substitute. Even though the forecasting horizon provided by dynamic regression model is relatively short, this is not a big issue for human resource departments, since most of human resource departments are only interested in a short term, such as three months, forecast. Therefore, if the human resource (HR) departmentin an organization is not well familiar with forecasting techniques, dynamic regression model could be a good option for a preliminary turnover forecast once a CLI can be identified as in this paper.

It is worth mentioning that an external variable like CLI in this paper does help in forecasting turnover since it does anticipate cyclical turning points. Incorporating such an external variable in the model is very helpful to get a good forecast when the HR department has a small and unreliable data set. Incorporating external variables, such as CLI, may help the whole forecasting process. If an external variable such as CLI is not available, a decomposition model could be considered as the first choice rather than dynamic regression model. In practice, some software such as NCSS or MINITAB have an embedded decomposition macro, which the HR departments could easily run given they knew how to estimate the cyclical variation.

# *4.1 Practical implications*

1. According to our findings, employee turnover forecast, in practice, could be handled easily. We suggest that HR departments could use a univariate linear regression model for the preliminary forecast and the accuracy of forecast is acceptable. However, there are some types of interventions that regression cannot handle, such as pulse or steps with exponential decay or growth. Dynamic regression is likely to be used as an alternative for forecasting in such cases.
2. In this study, statistical analysis packages SAS and NCSS were used for the forecast. However, for some organizations, the HR departments are unwilling to spend extra funding on software purchase. Under this circumstance, Microsoft Excel could be a good alternative for the time series forecast without extra investment, because there have been some open source time series forecasting packages designed to run under Excel environment in the market. The forecasting models such as naïve models, moving average, exponential smoothing, decomposition, regression, and ARIMA model have been included into these packages [46].
3. Another option these days is to use R for the dynamic regression [47]. The pseudo R code for dynamic regression and test on residuals (normality and white noise test) is provided below.

## Install R package

library (dynlm)

library (normwhn.test)

## “Turnover” is turnover data.

## “Trend” is a continuous variable with value from 1 to n.

## “Seasonality” is dummy variable representing seasonal periods.

## “X” is the lag term of CLI.

## “Intervention” is external variable impact: yes=1and no=0.

## Build model ## Load dataset

Turnover <- read.xls (“turnover\_data\_in\_excel.xls”)

## IMPORTANT: Some R codes in dynamic regression may not handle fancy intervention analysis.

Turnover.Model <- dynlm (Turnover ~ Trend + Seasonality + L(CLI, X) + Intervention)

## Model summary

summary ( Turnover.Model)

## Calculate residuals

Residual <- resid (Turnover.Model)

## Normality test on residuals

normality.test1 (Residual)

## White noise test on residuals

whitenoise.test (Residual)

# *4.2 Limitations and future research*

This study is limited to forecasting total turnovers in an organization. It may be possible to apply time series forecasting models to forecast turnovers in different categories like retirement or voluntary resignations. Although this study incorporated CLI as an external factor and the accuracy of forecasts is well improved, the external factors affecting turnover can be much beyond the scope of CLI, as local and cyclical economic fluctuations strongly influence the propensity of employees to quit ([Abelson and Baysinger, 1984](#_ENREF_1)) as well.

# References

[1] B. Leonard, *Turnover at the top*, HR Magazine. (2001), pp. 46-52.

[2] C. Khoong, *An integrated system framework and analysis methodology for manpower*

*planning*, International Journal of Manpower. 17 (1996), pp. 26-46.

[3] H.G. Heneman, D.P. Schwab, J.A. Fossum, and L.D. Dyer, *Personnel/Human Resource Management*, 4th ed., Irwin Homewood, Illinois, 1993.

[4] M. Čambál, A. Holková, and Z. Lenhardtová, *Basics of the Management*, Trnava:

AlumniPress, 2011.

[5] A.C. Bluedorn, *A unified model of turnover from organizations*, Human Relations. 35 (1982), pp. 135-153.

[6] D.L. Balfour, and D.M. Neff, *Predicting and managing turnover in human service*

*agencies: A case-study of an organization in crisis,* Public Personnel Management. 22

(1993), pp. 473-486.

[7] T.A. Wright, and R. Cropanzano, *Emotional exhaustion as a predictor of job*

*performance and voluntary turnover*, J. Appl. Psychol. 83 (1998), pp. 486-493.

[8] P.C. Morrow, J.C. McElroy, K.S. Laczniak, and J.B. Fenton, *Using absenteeism and*

*performance to predict employee turnover: Early detection through company records*,

Journal of Vocational Behavior. 55 (1999), pp. 358-374.

[9] E. Thaden, L. Jacobs-Priebe, and S. Evans, *Understanding attrition and predicting*

*employment durations of former staff in a public social service organization*, Journal of

Social Work. 10 (2010), pp. 407-435.

[10] M.J. Tews, K. Stafford and J.W. Michel, *Life happens and people matter: Critical*

*events, constituent attachment, and turnover among part-time hospitality employees*,

International Journal of Hospitality Management. 38 (2014), pp. 99-105.

[11] S.A. Collini, A.M. Guidroz and L.M. Perez, *Turnover in health care: the mediating*

*effects of employee engagement*, Journal of Nursing Management. 23 (2015), pp. 169-

178.

[12] L. Ferrara and D. van Dijk, *Forecasting the business cycle*, International Journal of

Forecasting. 30 (2014), pp. 517-519.

[13] J.D. Hamilton, *Calling recessions in real time*, International Journal of Forecasting. 27

(2011), pp. 1006-1026.

[14] S.H. Ng, F. Cram, and L. Jenkins, *A proportional hazards regression analysis of employee turnover among nurses in New-Zealand,* Human Relations. 44 (1991), pp. 1313-1330.

[15] R.S. Sexton, S. McMurtrey, J.O. Michalopoulos, and A.M. Smith, *Employee turnover: A neural network solution*, Computers & Operations Research. 32 (2005), pp. 2635-2651.

[16] D. Alao and A. Adeyemo, *Analyzing employee attrition using decision tree algorithms,* Computing, Information Systems, Development Informatics and Allied Research Journal. 4 (2013), pp. 17-28.

[17] W.F. Velicer and J.L. Fava, *Time Series Analysis*, Handbook of Psychology, 2003.

[18] T.H. Feeley and G.A. Barnett, *Predicting employee turnover from communication networks*, Human Communication Research. 23 (1997), pp. 370-387.

[19] W.-C. Hong, S.-Y. Wei, and Y.-F. Chen, *A comparative test of two employee turnover*

*prediction models*, Int. J. Mgmt. 24 (2007), pp. 216-229.

[20] V. Nagadevara, V. Srinivasan, and R. Valk, *Establishing a link between employee turnover and withdrawal behaviours: Application of data mining techniques,* Research & Practice in Human Resource Management. 16 (2008), pp. 81-99.

[21] A. Größler and A. Zock, *Supporting long-term workforce planning with a dynamic*

*aging chain model: A case study from the service industry*, Human Resource

Management. 49 (2010), pp. 829-848.

[22] V.V. Saradhi and G.K. Palshikar, *Employee churn prediction*, Expert Systems with

Applications. 38 (2011), pp. 1999-2006.

[23] OECD’s Library, *Composite Leading Indicators(MEI)* (2013). Available at

[http://stats.oecd.org/index.aspx?queryid=6617](http://stats.oecd.org/index.aspx?queryid=6617%20%20%20%20%20%20%20%20)

[24] L.C. Alwan and H.V. Roberts, *Time-series modeling for statistical process control*,

Journal of Business & Economic Statistics. 6 (1988), pp. 87-95.

[25] J. Grznar, D.E. Booth and P. Sebastian, *A robust smoothing approach to statistical*

*process control*, Journal of Chemical Information and Computer Sciences. 37 (1997),

pp. 241-248.

[26] P.F. Velleman, *Definition and comparison of robust nonlinear data smoothing*

*algorithms*, Journal of the American Statistical Association. 75 (1980), pp. 609-615.

[27] G.P. Zhang, *Time series forecasting using a hybrid ARIMA and neural network model*,

Neurocomputing. 50 (2003), pp. 159-175.

[28] C. Bergmeir and J.M. Benítez, *On the use of cross-validation for time series predictor*

*evaluation*, Information Sciences. 191 (2012), pp. 192-213.

[29] S.A. DeLurgio, *Forecasting Principles and Applications*, 1st ed., McGraw-Hill/Irwin, New York, 1998.

[30] P.R. Winters, *Forecasting sales by exponentially weighted moving averages*,

Management Science. 6 (1960), pp. 324-342.

[31] J.W. Taylor, *Short-term electricity demand forecasting using double seasonal*

*exponential smoothing*, Journal of the Operational Research Society. 54 (2003), pp. 799-

805.

[32] G.E. Box, G.M. Jenkins, and G.C. Reinsel, *Time Series Analysis: Forecasting and Control,* Holden Day, San Francisco, 1970.

[33] A.C. Harvey and P. Todd, *Forecasting economic time series with structural and Box-*

*Jenkins Models: A case study*, Journal of Business & Economic Statistics. 1 (1983), pp.

299-307.

[34] A. Pankratz, *Forecasting with Dynamic Regression Models,* John Wiley & Sons, 2012.

[35] J.G. De Gooijer and R.J. Hyndman, *25 Years of time series forecasting*, International

Journal of Forecasting. 22 (2006), pp. 443-473.

[36] J.E. Hanke, A.G. Reitsch, and D.W. Wichern, *Business Forecasting*, Prentice Hall Upper Saddle River, New Jersey, 1998.

[37] B.L. Bowerman, R.T. O'Connell, and A.B. Koehler, *Forecasting, Time Series, and Regression: An Applied Approach*, South-Western Pub., 2005.

[38] F.-L. Chu, *Forecasting tourist arrivals: nonlinear sine wave or ARIMA?*, Journal of

Travel Research. 36 (1998), pp. 79-84.

[39] S.S. Shapiro and M.B. Wilk, *An analysis of variance test for normality(complete samples)*, JSTOR, 1964.

[40] R.B. D'agostino, A. Belanger and R.B. D'Agostino Jr, *A suggestion for using powerful*

*and informative tests of normality*, The American Statistician. 44 (1990), pp. 316-321.

[41] J. Weisent, W. Seaver, A. Odoi and B. Rohrbach, *Comparison of three time-series*

*models for predicting campylobacteriosis risk*, Epidemiology and Infection. 138 (2010),

pp. 898-906.

[42] C. Bowie and D. Prothero, *Finding causes of seasonal diseases using time series*

*analysis*, International Journal of Epidemiology. 10 (1981), pp. 87-92.

[43] Department of Sciences at Pennsylvania State University, *Pre-whitening as an aid to interpreting the CCF* (2014). Available at

<https://onlinecourses.science.psu.edu/stat510/?q=node/75.>

[44] C. Chatfield, *Time-Series Forecasting*, CRC Press, 2000.

[45] P.T. Ittig, *A seasonal index for business*, Decision Sciences. 28 (1997), pp. 335-355.

[46] H.E. Warren, *Forecasting time series within Excel* (2008). Available at <http://web.calstatela.edu/faculty/hwarren/a503/forecast%20time%20series%20within%20Excel.htm>.

[47] R.J. Hyndman, *CRAN task view: Time series analysis* (2014). Available at

<http://cran.r-project.org/web/views/TimeSeries.html>.

# Appendices

Table A. 1. Time series univariate models for turnover data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Variables | Pred. R2 | Holdout R2 | MAPE | Normality | WN |
| Regression1 | Additive T+S | 0.51 | 0.57 | 26.15 | No | No |
| Additive S | 0.45 | 0.09 | 22.32 | No | No |
| Additive T+S+T\*S | 0.58 | 0.22 | 23.41 | No | No |
| Additive T+S+Intervention2 | 0.72 | 0.52 | 22.84 | Yes | No |
| Multiplicative T.S | 0.34 | 0.55 | 25.42 | Yes | No |
| Multiplicative S | 0.33 | 0.50 | 25.58 | Yes | No |
| Multiplicative T.S (T\*S) | 0.30 | 0.26 | 29.75 | Yes | No |
| Decomposition1 | T\*C\*S | 0.65 | 0.54 | 17.97 | Yes | Yes |
| WES1 | Additive T+ Additive S | 0.52 | 0.52 | 20.65 | Yes | Yes |
| Additive T+ Multiplicative S | 0.54 | 0.46 | 20.17 | Yes | Yes |
| Multiplicative T+ Additive S | 0.52 | 0.40 | 19.82 | Yes | Yes |
| Multiplicative T+ Multiplicative S | 0.52 | 0.39 | 19.77 | Yes | Yes |
| ARIMA | ARIMA(1,0,0)(2,1,0) | 0.39 | 0.33 | 22.94 | No | Yes |
| ARIMA(1,0,0)(2,0,0) | 0.38 | 0.38 | 23.96 | Yes | Yes |
| ARIMA(0,1,1)(2,1,0) | 0.45 | 0.14 | 22.96 | Yes | Yes |
| ARIMA(0,1,1)(2,0,1) | 0.42 | 0.17 | 22.45 | Yes | Yes |
| ARIMA(0,0,1)(1,0,1) | 0.35 | 0.30 | 25.35 | Yes | Yes |
| ARIMA(0,0,1)(1,0,2) | 0.39 | 0.33 | 22.96 | Yes | No |
| ARIMA(1,0,0)(0,1,1) | 0.40 | 0.48 | 25.20 | Yes | Yes |
| ARIMA(1,0,1)(2,0,0) | 0.43 | 0.26 | 21.82 | Yes | Yes |
| ARIMA(1,0,1)(1,0,2) | 0.44 | 0.20 | 21.55 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) | 0.47 | 0.40 | 22.89 | Yes | Yes |
| ARIMA(2,1,0)(0,1,1) | 0.41 | 0.26 | 25.24 | Yes | Yes |
| ARIMA(0,0,1)(2,0,0) | 0.36 | 0.37 | 25.00 | Yes | Yes |
| ARIMA(1,1,1)(1,0,0) | 0.35 | 0.04 | 24.20 | No | Yes |
| ARIMA(1,0,1)(1,1,0) | 0.39 | 0.32 | 22.45 | Yes | Yes |
| ARIMA(1,0,1)(1,0,1) | 0.41 | 0.24 | 22.18 | No | Yes |
| ARIMA(0,1,1)(2,0,0) | 0.40 | 0.16 | 22.60 | Yes | Yes |
| ARIMA(1,1,1)(2,0,0) | 0.41 | 0.20 | 22.26 | Yes | Yes |
| ARIMA(1,0,1)(1,0,1) | 0.35 | 0.04 | 23.87 | No | Yes |

Note: 1. In Regression, Decomposition and Exponential Smoothing models: T denotes Trend, S denotes seasonality, C denotes cycle, and T\*S denotes an interaction term between T and S.

2. The data is unsmoothed (or outliers are unadjusted) so as to take advantage of time series models that can accommodate interventions.

Table A. 2. Dynamic regression model with CLI and its lags as cyclical factor.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Pred. R2 | Holdout R2 | MAPE | Normality | WN |
| Dynamic regression & CLI | 0.72 | 0.56 | 22.56 | Yes | No |
| Dynamic regression & lag1 CLI | 0.72 | 0.54 | 22.73 | Yes | No |
| Dynamic regression & lag2 CLI | 0.74 | 0.55 | 21.87 | Yes | No |
| Dynamic regression & lag3 CLI | 0.74 | 0.55 | 21.50 | Yes | Yes |
| Dynamic regression & lag4 CLI | 0.74 | 0.56 | 21.35 | Yes | Yes |
| Dynamic regression & lag5 CLI | 0.75 | 0.57 | 21.32 | Yes | Yes |
| Dynamic regression & lag6 CLI | 0.76 | 0.58 | 20.73 | Yes | Yes |
| Dynamic regression & lag7 CLI | 0.77 | 0.59 | 19.91 | Yes | Yes |
| Dynamic regression & lag8 CLI | 0.77 | 0.59 | 20.05 | Yes | Yes |

Table A. 3. Decomposition model with CLI and its lags as cyclical factor.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Pred. R2 | Holdout R2 | MAPE | Normality | WN |
| Decomposition & CLI | 0.65 | 0.55 | 17.97 | Yes | Yes |
| Decomposition & lag1 CLI | 0.65 | 0.55 | 17.97 | Yes | Yes |
| Decomposition & lag2 CLI | 0.65 | 0.55 | 17.97 | Yes | Yes |
| Decomposition & lag3 CLI | 0.65 | 0.55 | 17.97 | Yes | Yes |
| Decomposition & lag4 CLI | 0.65 | 0.54 | 17.97 | Yes | Yes |
| Decomposition & lag5 CLI | 0.65 | 0.54 | 17.97 | Yes | Yes |
| Decomposition & lag6 CLI | 0.65 | 0.53 | 17.97 | Yes | Yes |
| Decomposition & lag7 CLI | 0.65 | 0.53 | 17.97 | Yes | Yes |
| Decomposition & lag8 CLI | 0.65 | 0.53 | 17.97 | Yes | Yes |

Table A. 4. ARIMA model with CLI and its lags as cyclical factor.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Pred. R2 | Holdout R2 | MAPE | Normality | WN |
| ARIMA(1,0,1)(0,1,1) & CLI | 0.37 | 0.41 | 22.77 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) & lag1 CLI | 0.37 | 0.41 | 22.73 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) & lag2 CLI | 0.37 | 0.41 | 22.78 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) & lag3 CLI | 0.37 | 0.41 | 22.85 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) & lag4 CLI | 0.36 | 0.40 | 22.91 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) & lag5 CLI | 0.36 | 0.40 | 22.98 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) & lag6 CLI | 0.36 | 0.40 | 23.04 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) & lag7 CLI | 0.36 | 0.40 | 23.09 | Yes | Yes |
| ARIMA(1,0,1)(0,1,1) & lag8 CLI | 0.36 | 0.40 | 23.12 | Yes | Yes |