***Time Series Analysis of Yokohama Household Consumption by Autoregressive Analysis and Yokohama Weather Data from 2000 to 2016.***

***by***

***School of Economics, Keio University***

***Taishiro Yamada(2018)***

***嘉治佐保子研究会***

***-Abstract-***

This paper aims to investigate the impact of weather against household total consumption. I study Yokohama, Japan weather from 2000 to 2016 to analyze Yokohama total consumption. My main purpose is to study the relationship between weather and household consumption.

Data on weather in Japan can be found in many periods on Japan Meteorological Agency Database. However, data on consumption in Yokohama does only exist from 2000 to the previous year. Thus, in this paper analysis is mainly focused on 2000s.

In chapter one, I provide a general introduction of weather and consumption data. In chapter two, I study consumption time series data, using autoregressive model of autoregressive moving average model(ARMA), autoregressive integrated moving average model(ARIMA). Scoring the accuracy of these models with root mean square error provides that ARIMA model was the most accurate. In chapter three, weather variables are taken into account by conducting vector autoregressive model(VAR) and granger causality model. In the last chapter, I compare the results of the results of chapter two and three, concluding that weather has a little contribute to prediction of the household consumption. Main factor that decides future consumption is past household consumption.

-Acknowledgement-

I would like to thank to Professor Sahoko Kaji, for your helpful comments and suggestions made in the course of your seminar of writing this paper. To finish this thesis, I have ever received a lot of advice from the participants who I have studied with as well. I appreciate all of your help.

January 2016

Taishiro Yamada

-Contents-

[Chapter 1 5](¥l)

[1.1 Introduction 5](¥l)

[1.2 Overview of data 5](¥l)

[1.3 Tests for regression analysis 9](¥l)

[1.4 Different series 10](¥l)

[Chapter 2 13](¥l)

[2.1 Estimation with autoregressive moving average model. 13](¥l)

[2.2 Scoring the accuracy of ARMA 14](¥l)

[2.3 Estimation with autoregressive moving integrated average model. 15](¥l)

[2.4 Comparing the accuracy of the models 17](¥l)

[Chapter 3 18](¥l)

[3.1 Selection of variables for VAR model 18](¥l)

[3.2 Estimation with VAR model 20](¥l)

[Chapter4 22](¥l)

[*-*References*-* 22](¥l)

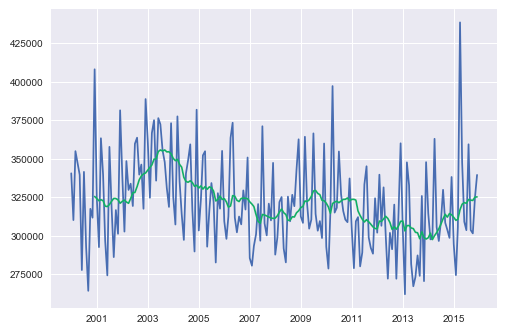
# Chapter 1

## 1.1 Introduction

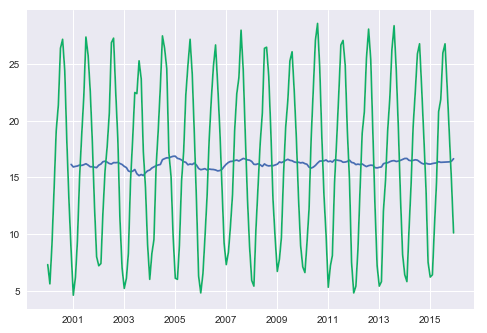
As factors of transition in household consumption, transition in propensity to consume is sometimes taken up. However, weather is sometimes also recognized as factors of transition in that. Therefore, in this thesis, I try quantitative analysis on the relationship between household consumption and weather. In this thesis, I employ data of Yokohama prefecture in Japan from 2000 to 2016 since household data in consumption does not exist before 2000. There are 5 variables summarized as weather data, including monthly average of daily average temperature, monthly sum of precipitation, monthly sum of sunshine hours, monthly average of daily average wind speed and monthly average of daily average cloud cover.

## 1.2 Overview of data

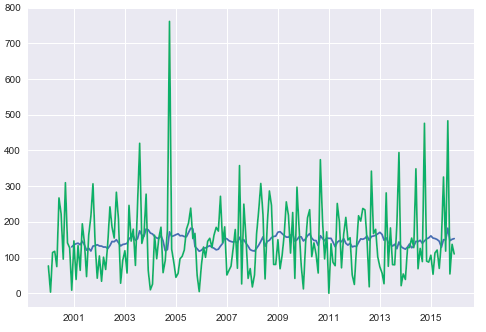
Although consumption in household sector in Yokohama prefecture has a trend, it does not have a drastic transition. Figure 1.1 illustrates monthly consumption in household sector in Yokohama prefecture from 2000 to 2016. Some of weather in Yokohama does not have a trend but some does. From figure1.2 to figure 1.6 shows each phenomena change from 2000 to 2016. Focusing on figure1.1, this shows that Yokohama prefecture household consumption is under unit root process. Weather data represented by figure 1.2, 1.3, 1.4, 1.5, 1.6 is mixture of unit root process and stationary process. From figure 1.2 and 1.3, temperature and rain are apparently under stationary process because moving average and variance has uniformity over the years. In contrast, figure 1.5 illustrates that wind speed is under unit root process since moving average has a trend. Additionally, variance is not uniform. Therefore, before conducting regression analysis to study the relation between weather and household consumption, unit root and cointegration should be tested. In the next section, I test all the variables with unit root test and conduct cointegration test to the variables which are defined as unit root.



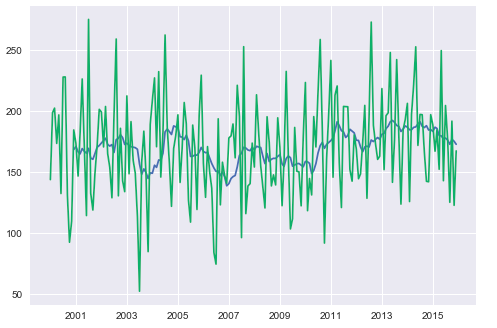
**Figure1.1** Yokohama prefecture monthly total household consumption from 2000 to 2016 and that of moving average.



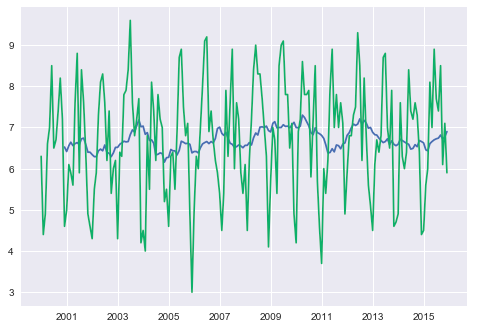
**Figure1.2** Yokohama prefecture monthly of daily average temperature from 2000 to 2016 and that of moving average.



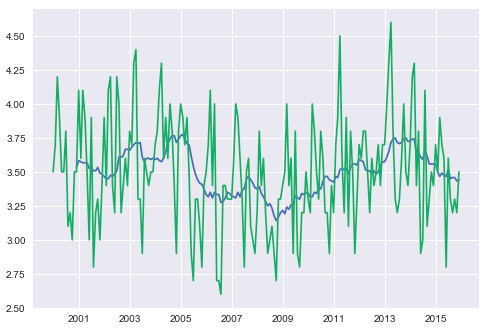
**Figure1.3** Yokohama prefecture monthly precipitation from 2000 to 2016 and that of moving average. Y axis’s unit is mm.



**Figure1.4** Yokohama prefecture monthly sum of sunshine hours from 2000 to 2016 and that of moving average. Y axis’s unit is hours.



**Figure1.5** Yokohama prefecture monthly average of cloud cover from 2000 to 2016. Definition of cloud cover is the rate of how much cloud covers all the skies. If cloud covers all, the index of cloud cover is 10. If cloud does not cover the sky at all, then the index is 1.



**Figure1.6** Yokohama prefecture monthly average of wind speed from 2000 to 2016 and that of moving average. Unit is mm per second.

## 1.3 Tests for regression analysis

From figure 1.1 to 1.6, necessity for unit root test and cointegration test becomes clear. Accordingly, I test household consumption and average temperature, precipitation, sunshine hours, cloud cover and wind speed. This time, I adopt Augmented Dickey–Fuller test(ADF test)（Dickey, Fuller 1979） to all data with a way that constant is only considered (trend is ignored) and autolag is decided by Akaike’s information criterion(AIC). Table1.1 provides the results of adfuller-dickey test. This shows that household consumption and wind speed are unit root process because p-value is less than 0.05, meaning these two null hypothesis fail to be rejected. The other null hypothesis succeeded to be rejected in terms of p-value. Therefore, cointegration test to data has to be done as some of data has unit root process.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Household Consumption | Average Temperature | Sunshine Hours | Precipitation | Cloud Cover | Wind Speed |
| ADF  statistic | -1.8142038948570183 | -3.2900637663836467 | -5.616589192480479 | -7.969310934484765 | -2.8600582212346004 | -1.9690874586506262 |
| Pvalue | 0.3734330038095893 | 0.015328508913554114 | 1.170606324354971e-06 | 2.8091012202364616e-12 | 0.05019373726663235 | 0.3002894367029275 |

**Table1.1** Results of ADF test to Yokohama household consumption and weather data from 2000 to 2016

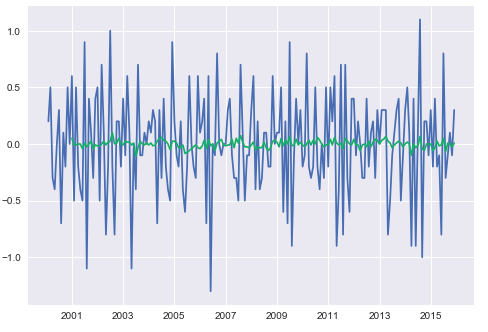
From the results of ADF test, household consumption, cloud cover and wind speed are judged as unit root since p-values are more than 0.05. Thus, I study cointegration of household consumption with cloud quantity and wind speed by using cointegration test advocated by Johansen (1991, 1995). Table 1.2 shows the results of cointegration test. Table 1.2 illustrates that cloud cover and wind speed do not have any cointegration with household consumption since p-values of cloud cover and wind speed are not significant, leading to the conclusion that these variables are not in the relationship of cointegration with household consumption. Thus, original series of weather data and household consumption must not be analyzed by regression. Therefore, in next section, I take different series out of original series of weather and household consumption data having unit root process for regression analysis.

|  |  |  |
| --- | --- | --- |
|  | Cloud Cover | Wind Speed |
| Coint-t | -1.656160653380096 | -1.731214228680023 |
| P-value | 0.6966142169811066 | 0.6624925441074415 |

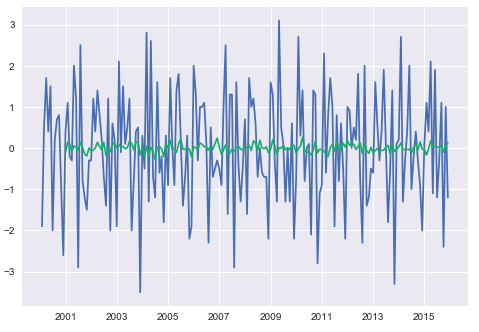
**Table1.2** Results of cointegration test of household consumption with unit root data.

## 1.4 Different series

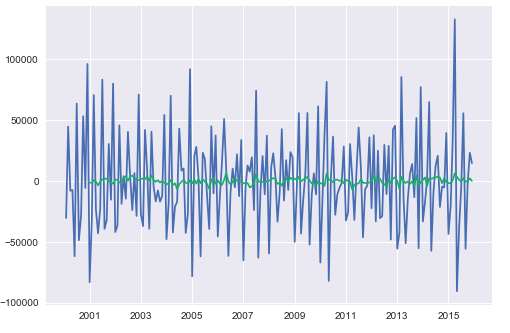
Tests provided in the section 1.3 proves that original series of household consumption, wind speed and cloud cover are unit root and do not have cointegration. Hence, I take different series out of these data. Figure 1.7, 1,8 and 1.9 are the different series. By taking difference series, wind speed, cloud cover and household consumption come to be stationary. Variance and mean of moving average of different series become more stationary compared to variance and mean of that of original series.



**Figure 1.7** Different series of wind speed and moving average of that.



**Figure 1.8** Different series of cloud cover and moving average of that.



**Figure 1.9** Different series of household consumption and moving average of that.

Again, I conduct ADF test to different series of unit root variables. The results of ADF test, represented by Table 1.3, shows these variable become stationary. Method of ADF test takes over the last method represented in Table 1.1. Since p-value of each variable is less than 0.05 and ADF statistics of each variable is enough big, these variables change stationary. Here, all the data preprocess ends as all variables come to be stationary. In chapter 2, I study autoregressive model applied to household consumption for estimating goodness of fit with the preprocessed household consumption.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Household consumption | Cloud Cover | Wind Speed |
| ADF  statistics | -8.00045229539132 | -10.686799428742827 | -12.10730820681971 |
| P-value | 2.3416464404906063e-12 | 3.814864619840895e-19 | 1.9554655283788218e-22 |

**Table 1.3**

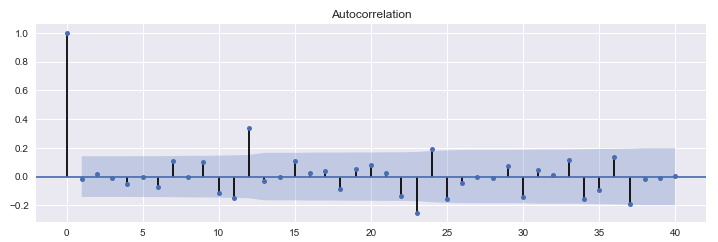
# Chapter 2

## 2.1 Estimation with autoregressive moving average model.

In chapter 2, I study the relationship between the past household consumption and the current household consumption by autoregressive model. Two models are adopted for achieving this. First model is autoregressive moving average (ARMA) model in which the value at the time point t is a linear function of the past Y values ​​, ..., and the past white noise , , ..., , denoted as

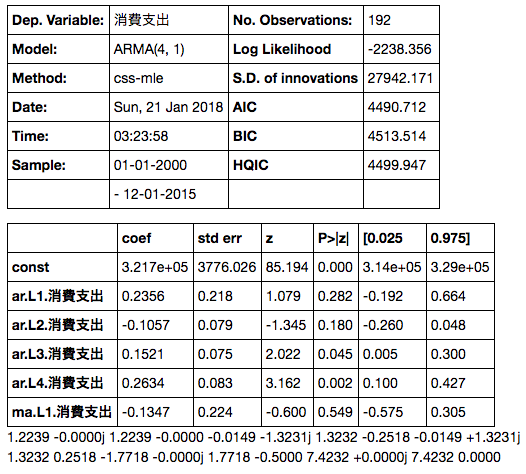
**(2.1)**

is monthly household consumption in each month in 2016. Explanatory variables are past monthly household consumption from 2000 to 2015. The results of ARMA model is below figure 2.1. Figure 2.1 shows that residual error still has seasonal periodicity. In every twelve months, autocorrelation becomes high. This decreases the accuracy of the model because periodicity in residual error means model does not ensure stationarity and does not consider a trend completely.



**Figure2.1** Residual error of ARMA model

In fact, looking at the summary of ARMA model represented by figure 2.2, AR model’s and MA model’s lag operator’s p-value is bigger than 0.05. Thus, this ARMA model is not significant. Therefore, other models should be studied.



**Figure2.2** Summary of ARMA model

## 2.2 Scoring the accuracy of ARMA

To test the generalized performance of this model, root mean square error(RMSE) is measured.

**(2.2)**

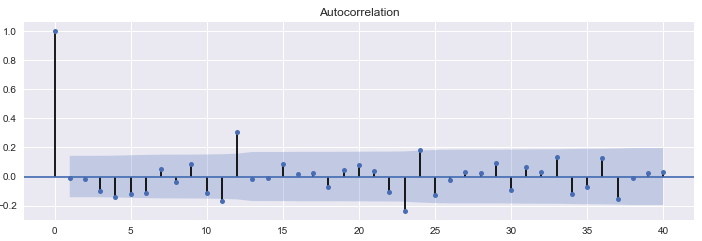
In this model, is mean of household consumption in 2016 estimated by ARMA model with data from 2000 to 2015 whereas is real value of household consumption in 2016. Table 2.1 is obtained by substituting these variables into the equation(2.2). This RMSE is compared to other models’ RMSE in the later section of this chapter.

|  |  |
| --- | --- |
|  | Root Mean Square Error |
| ARMA | 38929.07006164997 |

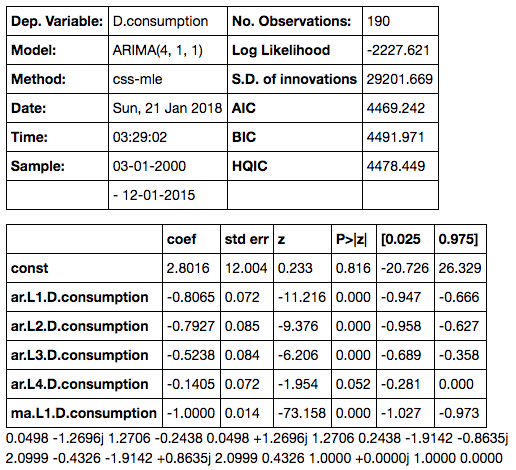
**Table 2.1** Root mean square error of ARMA model applied to household consumption. Test data is in 2016 as train data is from 2000 to 2015.

## 2.3 Estimation with autoregressive moving integrated average model.

From figure 2.1, it comes out that residual error still has a trend. In every twelve months, autocorrelation goes up. To make model more accurate, residual error of seasonal periodicity should be less. Therefore, to decrease a trend, autoregressive integrated moving average model is employed which is one of ARMA model fit to difference series. Figure 2.3 represents residual error of ARIMA model.



**Figure 2.3** Residual error of ARIMA model



**Figure 2.4**Summary of ARIMA model result

Figure 2.3 shows that a trend still exists, which is autocorrelation becomes high every 12 months. However, figure 2.4 illustrates that p-value of each lagged consumption is significant. Thus, though there is still possibility that generalized performance can be more accurate by omitting seasonal periodicity, model would become better than ARMA model. To test the generalized performance of this model, root mean square error(RMSE) is measured with equation defined in (2.2).

In the next section, I compare RMSE of ARMA model with that of ARIMA.

|  |  |
| --- | --- |
|  | Root Mean Square Error |
| ARIMA | 37611.87509611298 |

**Table 2.2** ARIMA models’ root mean square from January 2016 to December 2016

## 2.4 Comparing the accuracy of the models

Comparing root mean square error of ARMA with ARIMA, table 2.3 concludes that ARIMA model is more accurate than ARMA model. However, ARIMA model still has a seasonal residual error represented by figure 2.3. To raise the model accuracy, a seasonal periodicity must be removed. However, in this paper, I do not study model considering seasonal periodicity as there is no way to conduct multivariate regression considering seasonal periodicity.

|  |  |
| --- | --- |
|  | Root Mean Square Error |
| ARMA | 38929.07006164997 |
| ARIMA | 37611.87509611298 |

**Table 2.3** Comparison between ARMA and ARIMA models’ root mean square from January 2016 to December 2016

In the next chapter, I compare the result of ARIMA and that of the model which includes weather variable dealt as exogenous variable in ARIMA to study weather variable can improve the generalized performance. This is because if the RMSE of model including weather variable becomes less than that of ARIMA, it becomes clear that weather does affect household consumption.

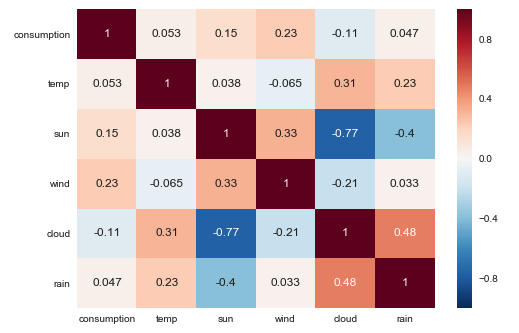
# Chapter 3

## 3.1 Selection of variables for VAR model

To estimate household consumption, I use vector autoregressive model (VAR) denoted as

**(3.1)**

where is constant expressing n columns vector whereas *Ai* (i = 1, 2, . . ., p) is coefficient and is disturbance. To use this model, I study correlation of weather variables to avoid multiple collinearity. Figure 3.1 is a heat map of correlation coefficients of weather variables illustrating that cloud cover has strong correlation with sunshine hours. Additionally, cloud cover also has a strong correlation with precipitation. Therefore, cloud cover must be dropped out of VAR model’s explanatory variables.



**Figure 3.1** correlation matrix of weather variables.

Figure 3.2, 3.3, 3.4 and 3.5 show results of granger-causality test representing relationship between consumption and each weather variable. Granger-causality test is denoted as

**(3.2)**

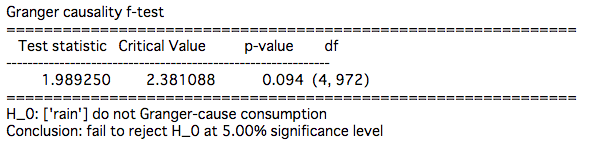
where is residual square of ordinary least square estimated by (3.3)

**(3.3)**

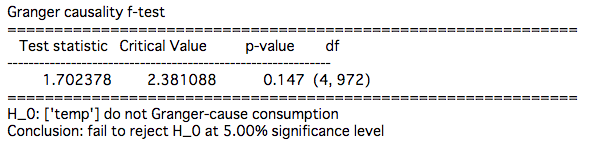
and is residual square of ordinary least square estimated by (3.4)

**(3.4)**

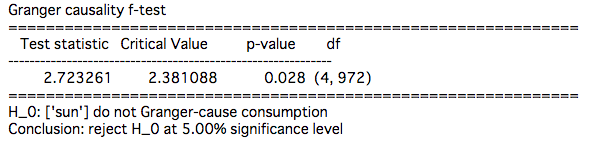
This test can be used only to data which is stationary and do not have cointegration because equation (3.3) and (3.4) can only be established to stationary data. In chapter 1 and 2, tests to check unit root and cointegration are passed. Therefore, granger-causality test can be applied. Granger-causality’s null hypothesis is each weather variable does not grander-cause consumption. From the figures from 3.2 to 3.5, it comes clear that wind speed has the weakest effect on household consumption since p-value is the biggest amongst the figures. Therefore, in the next section, I estimate VAR model with consumption, temperature, sunshine hours and precipitation and score the accuracy.



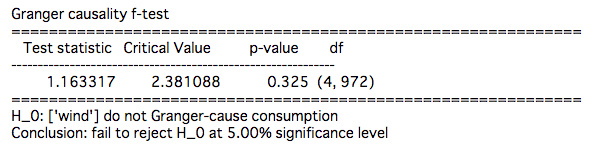
**Figure 3.2** Granger-causality test to precipitation.



**Figure 3.3** Granger-causality test to temperature



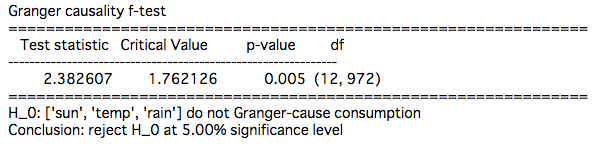
**Figure 3.4** Granger-causality test to sunshine hours

****

**Figure 3.5** Granger-causality test to wind speed

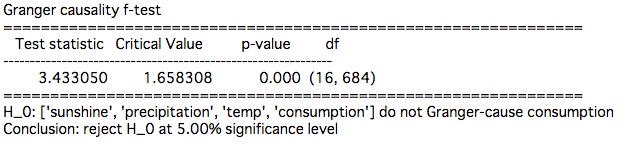
## 3.2 Estimation with VAR model

To study the relationship between weather and household consumption, VAR model is applied. The result is illustrated in figure 3.6. This shows that



**Figure 3.6** Granger-causality test to significant weather variables

Since p-value is less than 0.05, null hypothesis that sunshine hours, temperature and precipitation do not have Granger-causality with consumption is rejected. By including past household consumption, the accuracy of the model increases.



**Figure 3.7** Granger-causality test to weather variables and past consumption

I estimate 2016 household consumption with this model as done in chapter2 and score the accuracy of the model by RMSE. Table 3.1 shows the RMSE of VAR.

|  |  |
| --- | --- |
|  | Root Mean Square Error |
| VAR | 37812.38912206353 |

**Table 3.1** Root mean square error of VAR.

In the next Chapter, I compare the results of univariate autoregressive model and multivariate autoregressive model in terms of RMSE.

# Chapter4

## 4.1 Comparison of root mean square.

In this chapter, the results of univariate and multivariate models and conclude the relationship between weather and household consumption are conpared. Table 4.1 is results of RMSE. This shows that ARIMA has the smallest RMSE whereas ARMA has the biggest. Comparing ARMA and VAR, VAR has better model in terms of RMSE.

|  |  |
| --- | --- |
|  | Root Mean Square Error |
| ARMA | 38929.07006164997 |
| ARIMA | 37611.87509611298 |
| VAR | 37812.38912206353 |

**Table 4.1 Root mean square of autoregressive models.**

## 4.2 Conclusion

Two conclusion can be derived. First conclusion is that weather has an effect on household consumption to some extent. Looking at table 4.1 VAR is better than ARMA. ARMA considers autoregressive and moving average whereas VAR considers autoregressive and other variables. In other words, the difference of the results between these two models is derived from moving average and other variables. If the accuracy of ARMA is higher than VAR, this means moving average of household consumption can more accurately explain the change than weather. Conversely, if the accuracy of the VAR is higher than ARMA, this means weather has more effective to household consumption than moving average. Additionally, figure 3.6 reinforce this conclusion. Granger-causality test in figure 3.6 assuming that sunshine hours, precipitation and average temperature are cause of transition in household consumption is passed, which means that sunshine hours and temperature and precipitation granger-cause household consumption. Therefore, it can be concluded that sunshine hours, precipitation and temperature contributes to the change in household consumption.

Second conclusion is that accuracy of the model can be still increased by considering seasonal periodicity. Figure 2.1 and 2.3 proves that seasonal periodicity cannot be extracted from ARMA and ARIMA. Since VAR model is the extension of AR model which is the base of ARMA and ARIMA, VAR model also fail to extract seasonal periodicity. Thus, by introducing seasonal autoregressive integrated moving average model (SARIMA), considering seasonal periodicity, accuracy of the model will increase. Relationship between weather and household consumption will be more accurately explainable by considering seasonal periodicity.

I can conclude that sunshine hours, temperature and precipitation have some impact on consumption. However, by considering seasonal periodicity relationship between them will be more clarified.

# ***-***References***-***

*-Papers and Articles-*

* *Yoshinori Kawasaki and Genjiro Kitagawa (2017), Iwanami Data Science vol6, pp4-81*
* *Kitagawa, G. (1987) Non-Gaussian State Space Modeling of Nonstationary Time Series, Journal of the American Statistical Association Vol82, 1032-1063*
* *Kitagawa, G. (1987) A self-organizing state-space model. Journal of the American Statistical Association, Vol.93 1203-1215*
* *Kitagawa, G(2005) 時系列解析入門, Iwanami*
* Green. (1994) “Econometric Analysis,” 5th ed., Pearson (2003)
* Hamilton, J.D. “Time Series Analysis”. Princeton
* MacKinnon, J.G. (1994). “Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics* 12, 167-76.
* MacKinnon, J.G. (2010). “Critical Values for Cointegration Tests.” Queen’s University, Dept of Economics, Working Papers. Available
  + Found from <http://ideas.repec.org/p/qed/wpaper/1227.html>
* Durbin, James, and Siem Jan Koopman. (2012). Time Series Analysis by State Space Methods: Second Edition. Oxford University Press.

*-Data Source-*

* *Temperature, sunshine hours, precipitation, wind speed and cloud cover from Japan Meteorological Agency.*
  + *Found from* [*http://www.data.jma.go.jp/gmd/risk/obsdl/index.php*](http://www.data.jma.go.jp/gmd/risk/obsdl/index.php)
  + *Visited in December in 2017*
* *Yokohama consumption and its breakdown from Yokohama statistical portal site*
  + *Found from* [*http://www.city.yokohama.lg.jp/ex/stat/toukeisho/new/index3.html#12*](http://www.city.yokohama.lg.jp/ex/stat/toukeisho/new/index3.html#12)
  + *Visited in December in 2017*
* *Consumer Price Index from Statistic Bureau*
  + *Found from* [*http://www.stat.go.jp/data/cpi/*](http://www.stat.go.jp/data/cpi/)
  + *Visited in December in 2017*