Project Report in Time Series Analysis

Md Kamrul Hasan Khan

1 Introduction:

The Freddie Mac Primary Mortgage Market Survey (PMMS) collects 30-year fixed-rate mortgage data for home purchase every week in five major regions (North East(NE), South East(SE), North Center (NC), South West (SW) and West (W)) in United States since April, 1971. The survey is based on first-lien prime conventional conforming home purchase mortgages with a loan-to-value of 80 percent. In the early 1980s, the mortgage rates topped out because the Federal Reserve was waging a war with inflation which is unlikely today. Threfore, to discard that effect from the data this study analyze the data from January, 1990 and this data is abailable upto December, 2015. Moreover, the current study aims to see the patern of the interest rates at monthly basis instead of weekly. Therefore, the data is first converted into monthly by taking the average of all weeks over the corresponding month and then used in the analysis. Hence, the goal of this project is to build a model to predict the monthly 30-year fixed-rate mortgage in the five major regions in US.

The rest of the report is oganized as follows. In section 2, I describe the 30-year fixed-rate mortgage dataset. Section 3 outlines the modeling options. In Section 4, I presents the ouputs from the data analysis including results from model comparison. Finally, Section 5 summarizes the major features of this work.

2 Data Description:

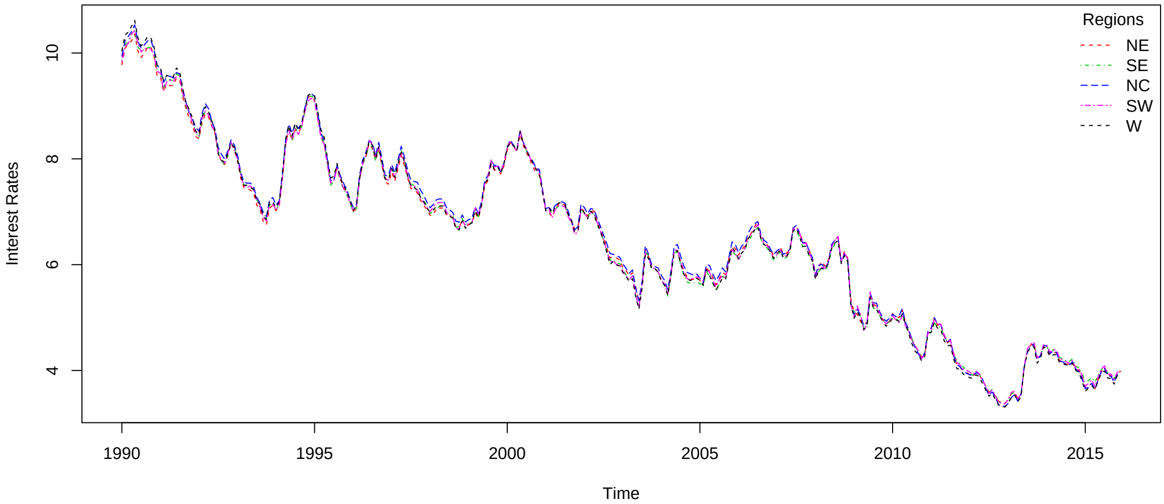
The Federal Home Loan Mortgage Corporation (FHLMC), known as Freddie Mac, is a public government-sponsored enterprise (GSE) and was created in 1970 to expand the secondary market for mortgages in the US. Since April, 1971, it surveys lenders, a mixed of thrifts, credit unions, commercial banks and mortgage lending companies, in five major regions (North East(NE), South East(SE), North Center (NC), South West (SW) and West (W)) in US every week on the rates for 30-year fixed-rate mortgage products. The survey is collected from Monday through Wednesday and the results are published extensively in the media on Thursdays at 10 a.m. ET. This data is available upto December, 2015.

This study is not analyzing the entire time series. It is analyzing the data from January, 1990 to December, 2015. Here, we want to see the behave of monthly interest rate of 30-year fixed-rate mortgage products in each of the five regions. Therefore, the monthly interest rate data is created from this weekly data and it is done by taking the average of the weekly data within the corresponding month of the corresponding region. Figure 1 depicts the monthly regional mortgage rates during 1990-2015. From the figure it is clear that at every time points the rates are almost similar in all the regions and it falls gradually with some fuctuations over the period.

3 Models for Monthly Regional Mortgage Rates:

Since we are given the monthly mortgage rates from five different regions, the modeling part can be partioned into two broad classes: one is modeling each of the regions independently and the other is modeling them jointly. Here I model the monthly 30-year fixed-rate mortgage data in three different ways: (i) Seasonal Autoregressive Integrated Moving Average (SARIMA) model, (ii) Dynamic Linear Model (DLM) and (iii) Vector Autoregressive (VAR) model, where in SARIMA as well as DLM major regions are modeled independantly and in VAR they are modeled jointly. The models are explained below:

Figure 1: Plots of monthly 30-year fixed-rate mortgage in five regions (North East(NE), South East(SE), North Center (NC), South West (SW) and West (W)) in US during 1990-2015.



i. **Seasonal Autoregressive Integrated Moving Average (SARIMA) model:** Here the five major regions (NE, SE, NC, SW and W) in US are modeled independently. Figure 1 clearly shows that the data are non-stationary as well as there exists trend in the data. Therefore, to make the data stationary, a logarithmic trasformation is performed. Then a SARIMA $(p, d, q) \times (P, D, Q)_s$ model is implemented on this transformed 30-year fixed-rate mortgage where p and q are the lags of AR and MA model, d is the order of differencing, P and Q are seasonal lags of AR and MA model, D is the ordering of seasonal differencing, and s is period of the season.

ii. **Dynamic Linear Model (DLM):** Since these data have a trend, I start with a local linear trend DLM. Then to capture the seasonality a Fourier series with appropriate number of hermonics is augmented with it. And finally, the DLM representation of ARMA model with appropriate lags is used. A similar model is implemented in all regions independently.

iii. **Vector Autoregressive (VAR) model:** Figure 1 shows that the 30-year fixed-rate mortgage in all regions are very similar, which means the data are highly correlated. To use this correlation, I model them jointly with a p-th vector order autoregressive model. Since the data has both trend and seasonality, some deterministic terms such as a constant, a trend as well as centered seasonal dummy variables are included with it. Let $Y_t = \left(y_t^{NC}, y_t^{NE}, y_t^{SE}, y_t^{SW}, y_t^{W}\right)'$ denote an (5×1) vector of mortgage rates in five regions at time t. Hence the model can be written as follows:

$$Y_t = \sum_{j=1}^{p} \Phi_j Y_{t-j} + \Gamma D_t + \epsilon_t, \tag{1}$$

where $\epsilon_t \overset{iid}{\sim} \mathcal{N}_5(0, \Sigma)$, Φ_j are (5×5) j-coefficient matrices of AR(p) model, D_t represents column matrix of all deterministic components and Γ is the parameter matrix of the corresponding deterministics components, D_t.

In all these three choices model selection is performed based on time-series diagnostics as well as information based criterion. The diagnostics is done from the standardized residuals plot, the autocorrelation function of the residuals, and the p-values of Ljung-Box version of the portmanteau test for all lags up to the maximum number of lags for a Portmanteau goodness-of-fit test, and Akaike information criterion (AIC) as well as Bayesian information criterion (BIC) are used as information based criterion.

4 Data Analysis:

We implement all three types of models from Section 3 and assess the best model in all these three cases. The detail analysis is explained below:

i. **Seasonal Autoregressive Integrated Moving Average (SARIMA) model:** As mentioned in Section 3 we implement a SARIMA $(p, d, q) \times (P, D, Q)_s$ model on the logarithmic transformed 30-year fixed-rate mortgage in the five regions independently. Since there exists trend in the data, I start with $d = 1$ to eliminate the trend. Then the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the differences of this logarithmic transfomed mortgage rates suggests $p = 1$ and $q = 2$. Although all the diagnostics give a good output, the model can not forecast the seasonal patern due to the lack of seasonal components. Now since this is a monthly data I use $s = 12$ and $D = 1$. Then the ACF and PACF plot suggest $p = 1, q = 2, P = 0$ and $Q = 1$. However, Table 1 shows that a parsimonious model with $p = 1, q = 1, P = 0$ and $Q = 1$ has lower AIC as well as BIC with similar log-likelihood for all the regions. The diagnostics output of SARIMA $(1, 1, 1) \times (0, 1, 1)_{12}$ in Figure 2 also shows that the residuals follow standard normal distribution, the ACF plot of residuals depicts no autocorrelation left and the p-values of Ljung-Box statistic are large for all lags in all regions. Hence we model logarithmic transformed 30-year fixed-rate mortgage for all regions independently by SARIMA $(1, 1, 1) \times (0, 1, 1)_{12}$. Now using this model I predict the mortgage rates for next five years (2016-2020) in each regions. Figure 3 represents the value of the forecasted rates during 2016-2020 with blue curve and the 90% confindence band by green curves.

Table 1: SARIMA model comparison

Regions	Measure	SARIMA $(1, 1, 2) \times (0, 1, 1)_{12}$	SARIMA $(1, 1, 1) \times (0, 1, 1)_{12}$
North East (NE)	LL†	618.791	618.772
	AIC	-1227.583	-1229.544
	BIC	-1208.868	-1214.572
South East (SE)	LL†	615.978	615.807
	AIC	-1221.955	-1223.614
	BIC	-1203.240	-1208.642
North Center (NC)	LL†	608.943	608.660
	AIC	-1207.885	-1209.319
	BIC	-1189.170	-1194.347
South West (SW)	LL†	605.153	604.709
	AIC	-1200.305	-1201.419
	BIC	-1181.590	-1186.447
West (W)	LL†	605.847	605.522
	AIC	-1201.694	-1203.045
	BIC	-1182.979	-1188.073

† LL refers to log-likelihood

ii. **Dynamic Linear Model (DLM):** The number of harmonics as well as the number of lags of ARMA model are selected base on diagnostics and Information based criterions. Here the parameters are estimated using maximum likelihood estimates (MLE). Table 2 shows that for all regions DLM with local linear trend, two harmonics in Fourier series and DLM representation of ARMA(1,1) has lower AIC as well as BIC, and larger log-likelihood compared to the DLM with local linear trend, three harmonics in Fourier series and DLM representation of ARMA(1,1). The diagonstics outputs in Figure 4 also shows that DLM with local linear trend, two harmonics in Fourier series and DLM representation of ARMA(1,1) preforms well in all regions. Therefore, the mortgage rate data is model using this DLM and a five years prediction is performed. Figure 5 represents the 30-year fixed-rate mortgage with the five years (2016-2020) prediction. The blue curve shows the prediction and green curves represents the 90% confidence interval. The figure clearly depicts more uncertainty than the corresponding SARIMA model. Moreover, Table 1 and Table 2 shows that both AIC as well as BIC in DLM model are larger and the log-likelihood is smaller compared to the corresponding SARIMA $(1, 1, 1) \times (0, 1, 1)_{12}$ model in all regions.

Table 2: DLM comparison

Regions	Measure	Local linear trend + two hermonics in Fourier Series + DLM representation of ARMA(1, 1)	Local linear trend + three hermonics in Fourier Series + DLM representation of ARMA(1, 1)
North East (NE)	LL†	316.844	293.419
	AIC	-621.688	-574.838
	BIC	-599.230	-552.380
South East (SE)	LL†	311.891	288.795
	AIC	-611.781	-565.590
	BIC	-589.323	-543.132
North Center (NC)	LL†	308.619	285.351
	AIC	-605.237	-558.702
	BIC	-582.779	-536.244
South West (SW)	LL†	305.762	282.197
	AIC	-599.525	-552.395
	BIC	-577.067	-529.937
West (W)	LL†	305.262	281.833
	AIC	-598.523	-551.666
	BIC	-576.065	-529.207

† LL refers to log-likelihood

iii. **Vector Autoregressive (VAR) model:** Here the number of lags in AR and in seasonal periodic components are chosen based on information based criterions.

Table 3 represents that the VAR model with a constant, a trend component, a first order AR and an annual lag has lower AIC and BIC compared to the VAR model without any constant and trend components. The log-likelihood also increases after adding a constant and a trend in the model. Hence the data is modeled with a constant, a trend component, a first order AR and an annual lag, and forecasting is performed for January, 2016 to December, 2020 in all regions. Figure 6 depicts the 30-year fixed-rate mortgage as well as the forecasted rates during 2016-2020 for all regions. The blue curve represents the prediction and the 90% confidence band is shown with the green curves.

Table 3: VAR model comparison

Measure	A first oreder AR + an annual lag	A constant + a trend + a first oreder AR + an annual lag
LL†	2904.293	2937.888
AIC	-5492.586	-5539.777
BIC	-4901.191	-4910.952

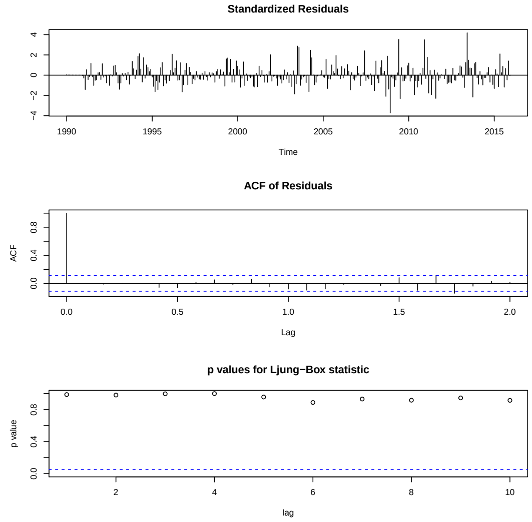
† LL refers to log-likelihood

5 Conclusion:

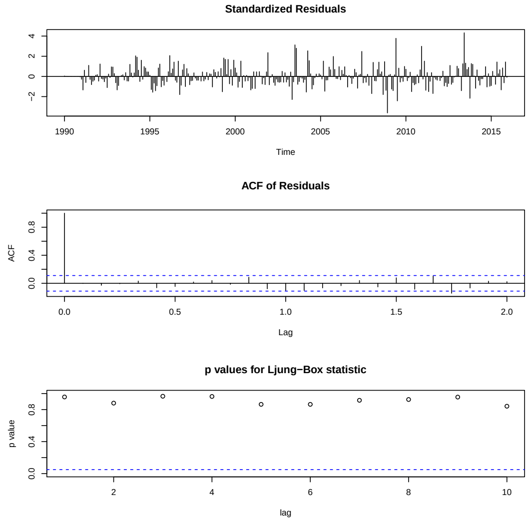
I have implemented three different models on monthly 30-year fixed-rate mortgage data and using the corresponding model the mortgage interest rates are forecasted for next five years (2016-2020). Although the model predictions are similar in all three choices in all five regions, the 90% confidence interval is significantly large in the DLM even with some negative values in the lower bound which is unrealistic. On the other hand, the uncertainty interval in VAR model increases slower than the corresponding value in SARIMA model over the time which may be because of using the correlation between different region's mortgage rates.

Figure 2: Diagnostics outputs of SARIMA $(1,1,1) \times (0,1,1)_{12}$

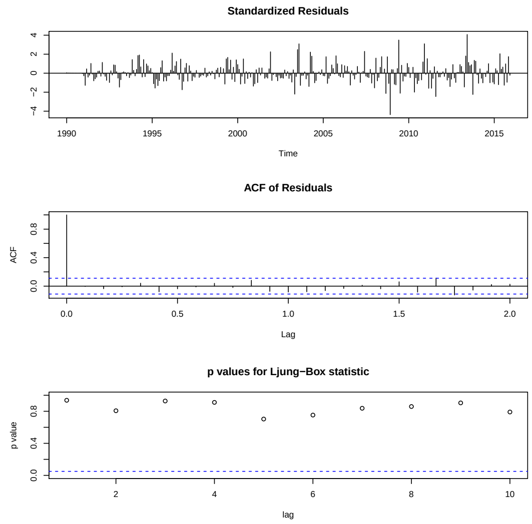
North East (NE)



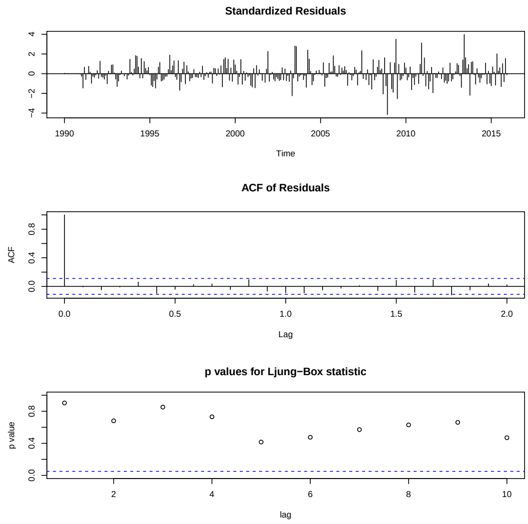
South East (SE)



North Center (NC)



South West (SW)



West (W)

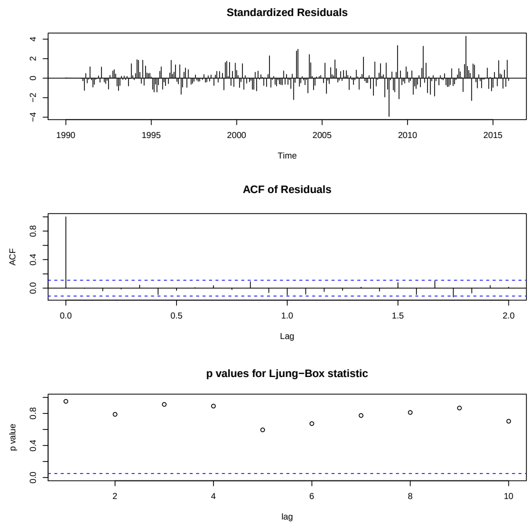


Figure 3: Five years (2016-2020) forecast plots of monthly 30-year fixed-rate mortgage in five regions (North East(NE), South East(SE), North Center (NC), South West (SW) and West (W)) in US using SARIMA model where the blue curve dipicts the prediction and the green curve represents the 90% confindence band.

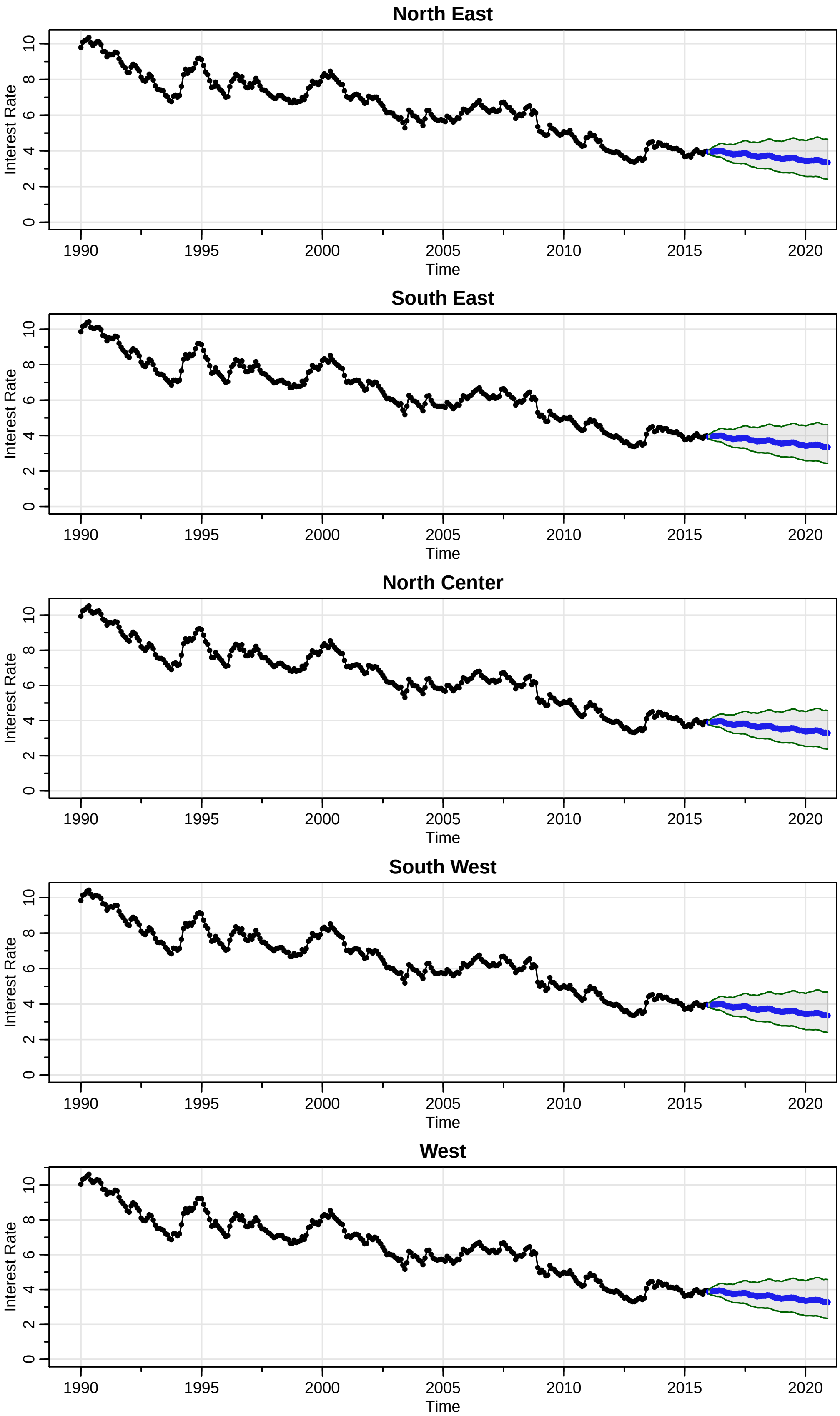


Figure 4: Diagnostics outputs of DLM with local linear trend, two harmonics in Fourier series and DLM representation of ARMA(1,1)

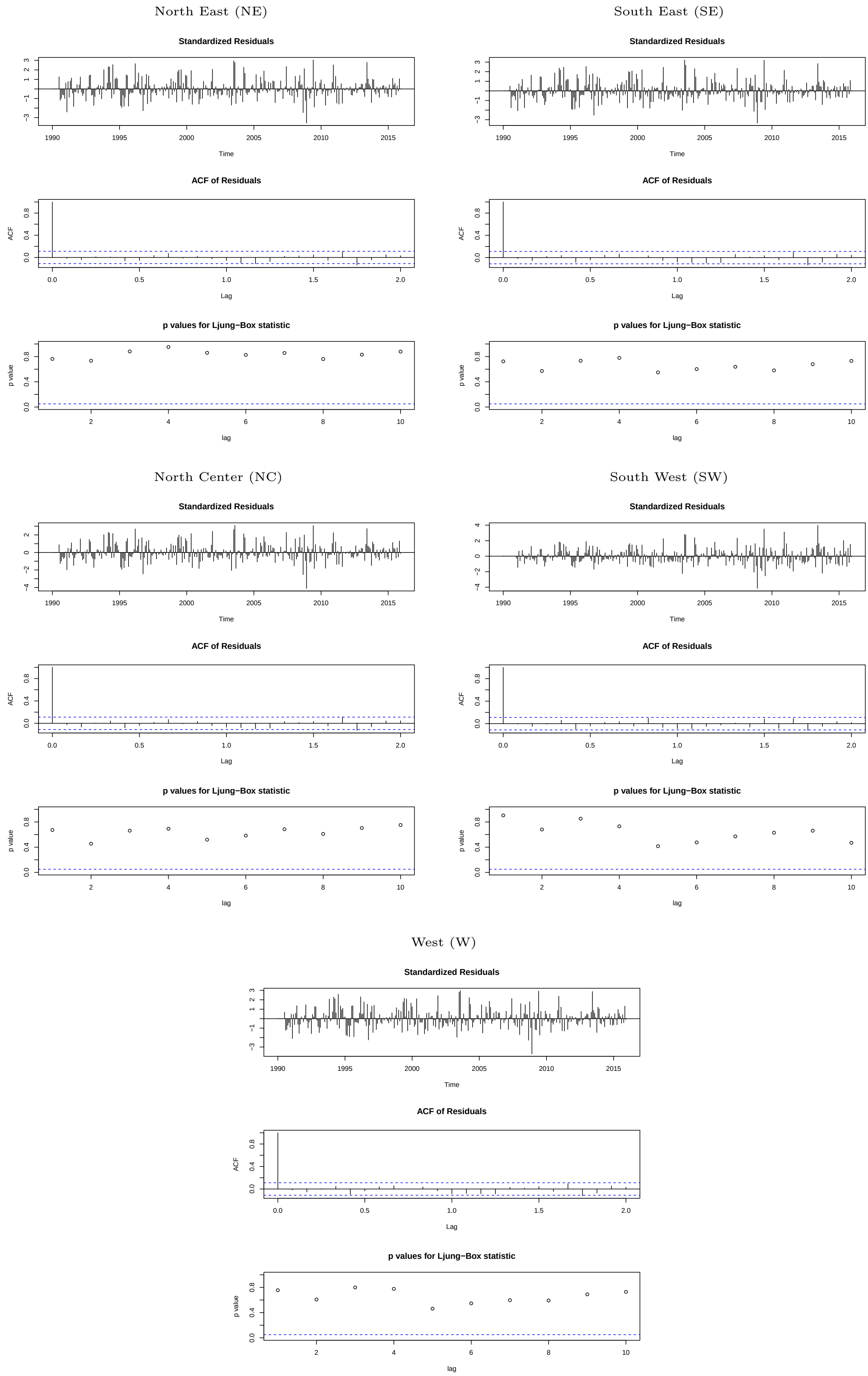


Figure 5: Monthly 30-year fixed-rate mortgage data plot with five years (2016-2020) forecast in five regions (North East(NE), South East(SE), North Center (NC), South West (SW) and West (W)) in US using DLM where the predictions are repesented with the blue curve and the 90% confindence interval is shown by green curve.

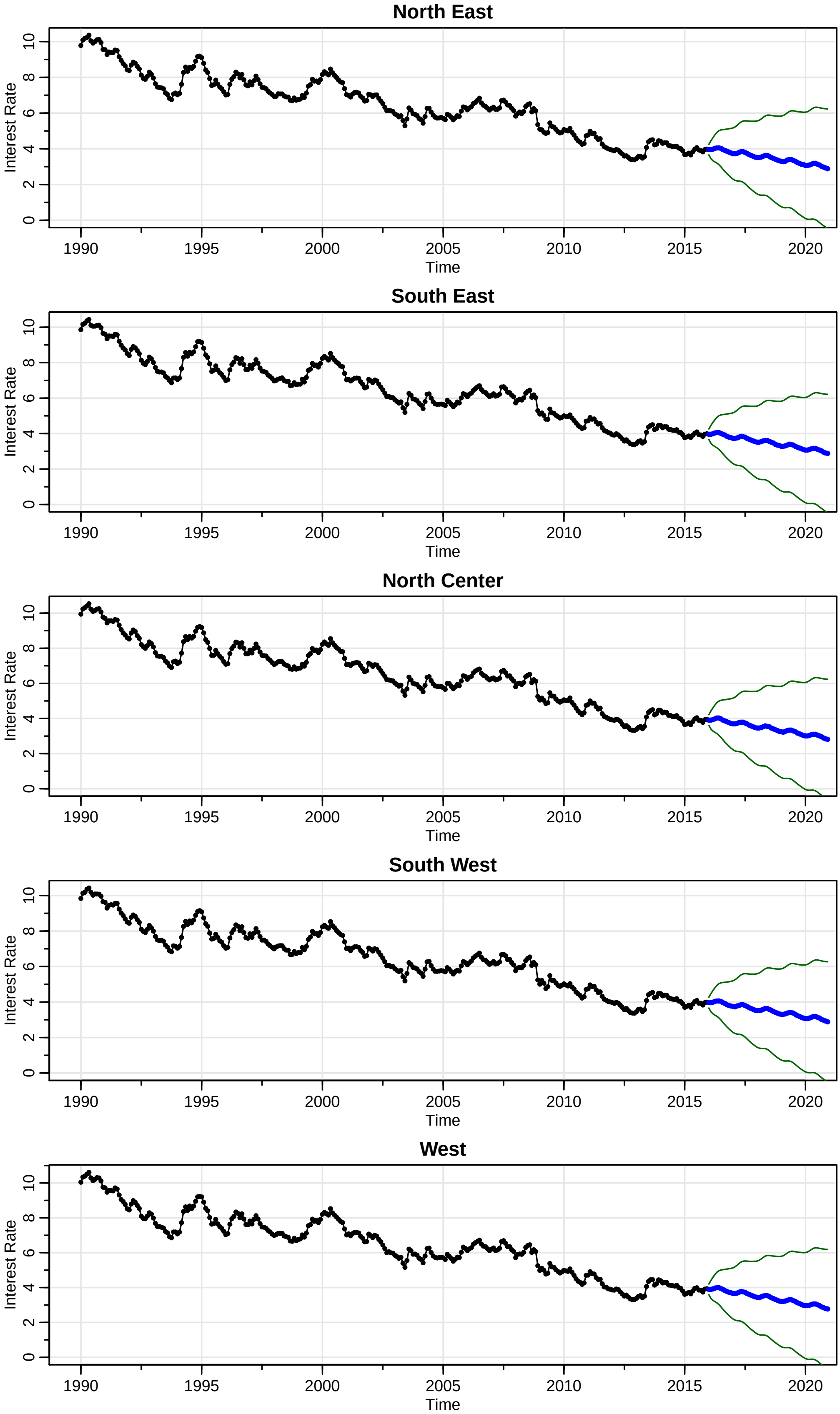


Figure 6: Plots of monthly 30-year fixed-rate mortgage with the predictions during 2016-2020 in five regions (North East(NE), South East(SE), North Center (NC), South West (SW) and West (W)) in US using VAR model where the blue curve dipicts the forecasted values and the green curve shows the 90% confindence band.