Retail-Giant Sales Forecasting

Case Study Submission

Candidate Name:

1. Venkata Nihar Tatavarti
2. Prasanth Rasam
3. Deepak Malhotra

Business Problem

- “Global Mart” is an online store super giant having worldwide operations. It takes orders and delivers across the globe and deals with all the major product categories - consumer, corporate & home office. As Sales/operations manager, finalize the plan for the next 6 months

The aim of Case Study is to

1. Find Top 2 most profitable & consistent segment 21 (i.e., 7 different market segments and in 3 major categories) to forecast the sales and demand for Top 2 segments.

2. Forecast the sales and the demand for the next 6 months, to manage the revenue and inventory

Problem solving methodology

1. Data Understanding & Clean-up
2. Data Preparation Requirements
    1. Segment the whole dataset into the 21 subsets based on the 7 market and the 3 customer segment level
    2. Convert the transaction-level data into a time series, aggregate the 3 attributes - Sales, Quantity & Profit, over the Order Date to arrive at monthly values for these attributes
    3. Arrive at 3 time series for each of the 21 segments (7 market and 3 customer segment)
    4. Find the 2 most profitable and consistently profitable segments, using metric coefficient of variation of the Profit for all 21 market segments
3. Model Building
    1. Forecast the sales and quantity for the next 6 months of Top 2 segments
        1. Smoothen the data to perform classical decomposition
        2. Use classical decomposition & auto ARIMA for forecasting
4. Model Evaluation
    1. Forecast the sales/demand for next 6 months using the Satisfactory Model
    2. Test the accuracy of forecast
        1. Separate out the last 6 months values from the dataset, after aggregating the transaction level data into the monthly data.
        2. Check 6 months forecast using the out-of-sample figures using MAPE

1. Data Understanding & Clean-up

1. 24 Variables are present in Global Superstore.csv with 51290 Observations

2. Data Dictionary has references to 23 Variables. Row ID is not mentioned in Data Dictionary as Observation. This does not have any impact on Problem Solving

3. There are No duplicate Columns

4. The Six Variables required for Time Series Analysis are
    1. Order Date - Date on which the order was placed
    2. Market - Market segment to which the customer belongs
    3. Segment - The market segment to which the product belongs
    4. Sales - Total sales value of the transaction
    5. Quantity - Quantity of the product ordered
    6. Profit - Profit made on the transaction

5. There are no NA Values and Blank columns in the Six selected variables

2. Data Preparation & Requirements

1. Segment the whole dataset into the 21 subsets based on the 7 market and the 3 customer segment level
    1. 7 Markets - Africa, APAC, Canada, EMEA, EU, LATAM and US
    2. 3 Segments - Consumer, Corporate and Home Office

Ref: superStoreData$MarketnSegment

2. Convert the transaction-level data into a time series, aggregate the 3 attributes - Sales, Quantity & Profit, over the Order Date to arrive at monthly values for these attributes

Ref: Dataframes ByMonthView & MarkSegdf

3. Arrive at 3 time series for each of the 21 segments (7 market and 3 customer segment)

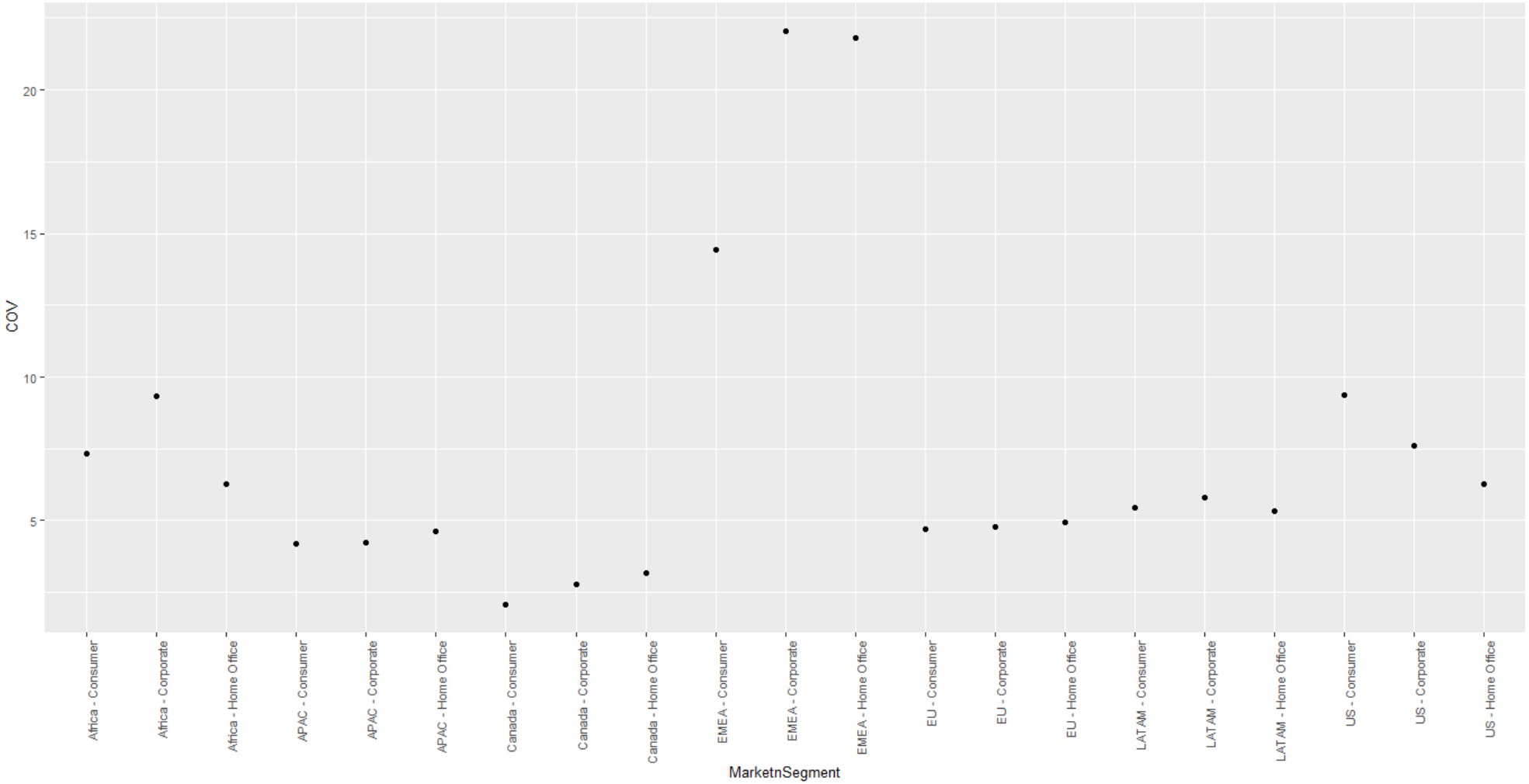
Ref: Plots 1 to 4

4. Find the 2 most profitable and consistently profitable segments, using metric coefficient of variation of the Profit for all 21 market segments

Ref: COV - Coefficient of variation by Market and Segment

2. Data Preparation & Requirements

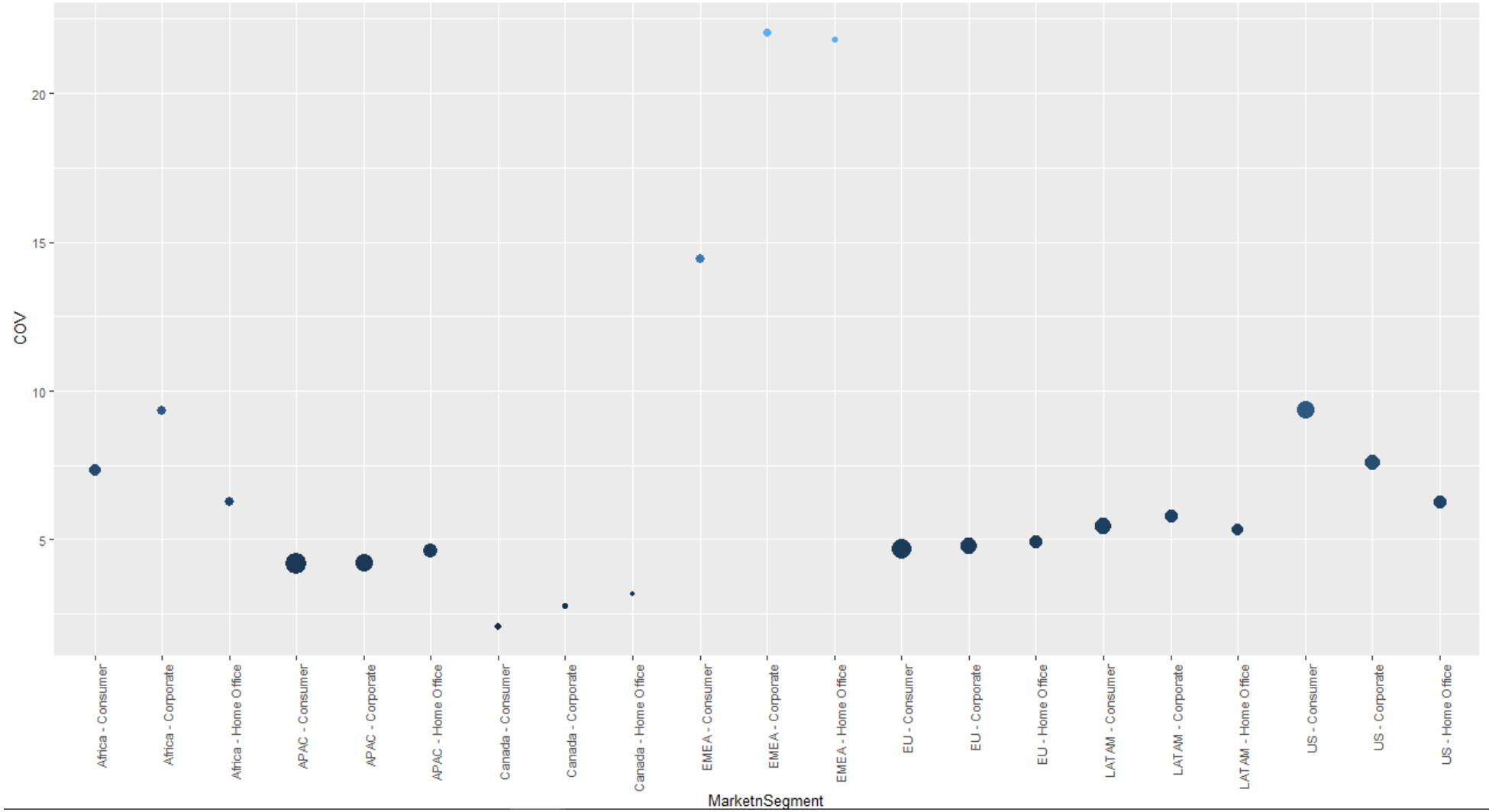
1. Plot 1 - Introduced Profit Coefficient of Variation by Market & Segment
   - COV = Standard Deviation of Profit / Mean of Profit. The lower the value of the coefficient of variation, the more precise the estimate.



COV by Market & Segment	
Market and Segment	COV (Lower the better)
Canada – Consumer	2.093766
Canada – Corporate	2.767642
Canada - Home Office	3.175392
APAC – Consumer	4.206702
APAC – Corporate	4.231301
APAC - Home Office	4.63339
EU – Consumer	4.718084
EU – Corporate	4.776482
EU - Home Office	4.923759
LATAM - Home Office	5.336331
LATAM – Consumer	5.438845
LATAM – Corporate	5.789517
Africa - Home Office	6.264113
US - Home Office	6.280008
Africa – Consumer	7.351006
US – Corporate	7.616929
Africa – Corporate	9.334133
US – Consumer	9.38945
EMEA – Consumer	14.441103
EMEA - Home Office	21.800527
EMEA – Corporate	22.038317

2. Data Preparation & Requirements

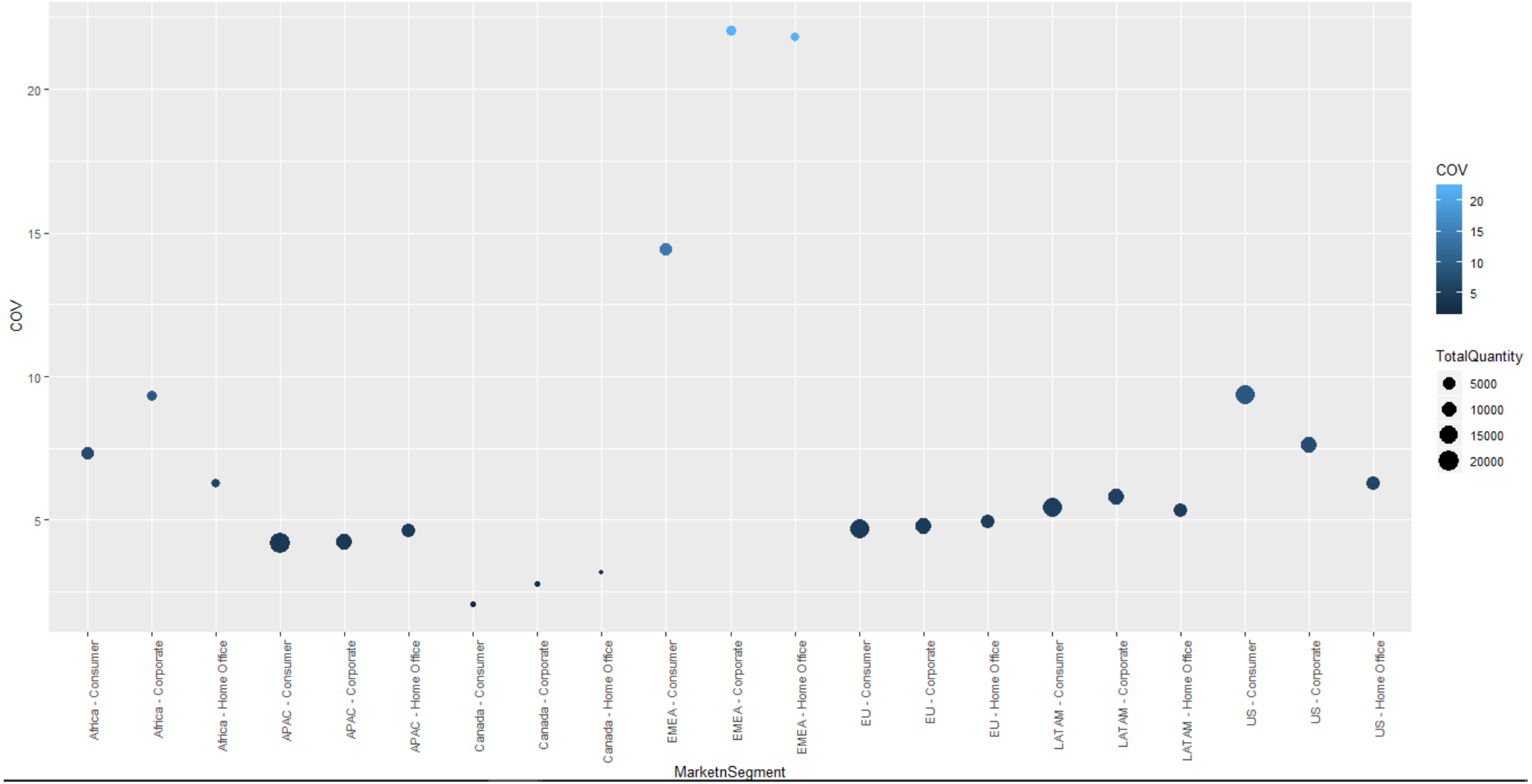
1. Plot 2 - Below chart depicts Profit Coefficient of Variation by Market & Segment
   - Bigger the circle, larger the Profits



Top 5 Market & Segment by Profits	
Market and Segment	Total Profit
APAC – Consumer	222817.56
EU - Consumer	188687.707
US - Consumer	134119.209
APAC - Corporate	129737.235
EU - Corporate	123393.98

2. Data Preparation & Requirements

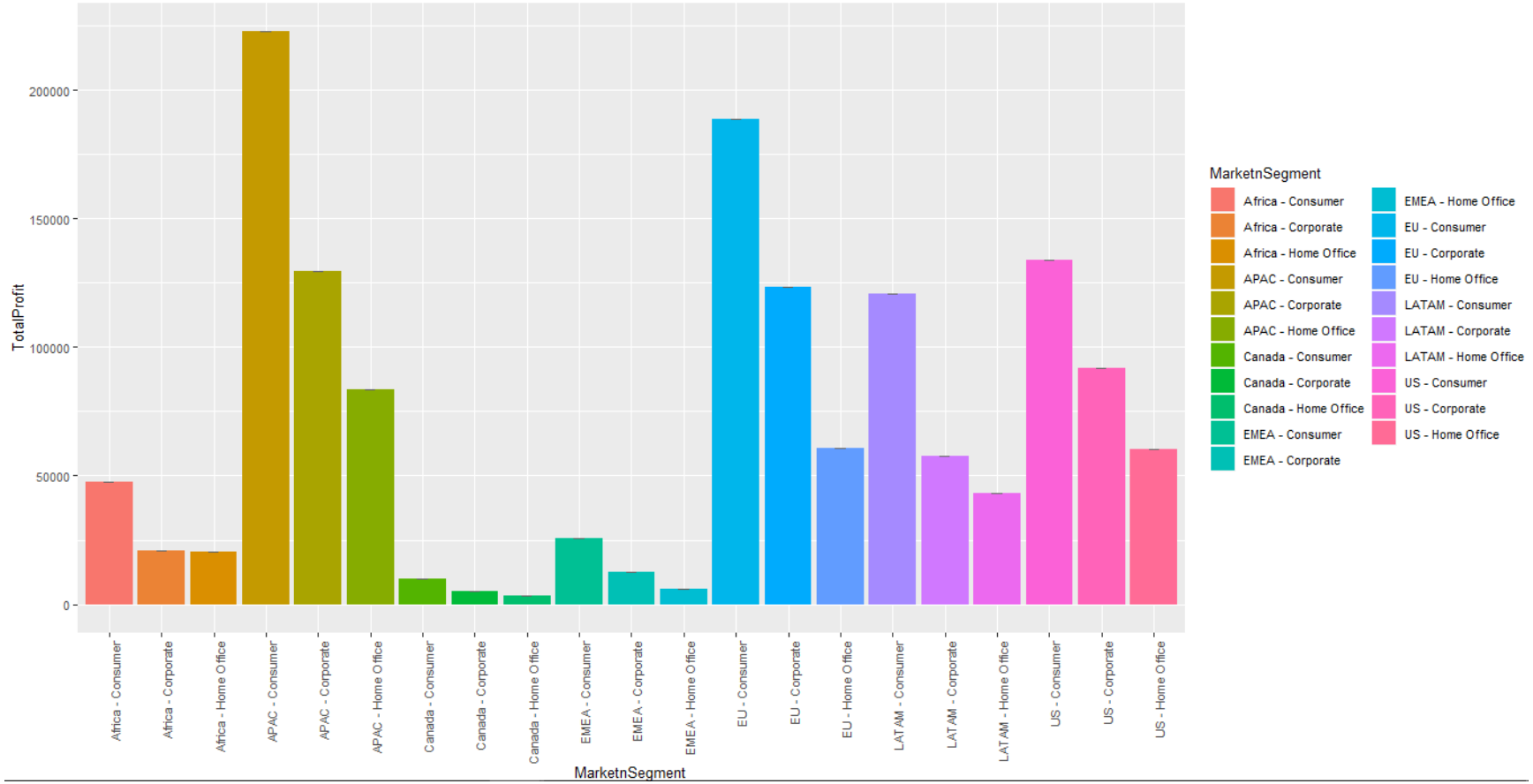
1. Plot 3 - Below chart depicts Profit Coefficient of Variation by Market & Segment
   - Bigger the circle, larger the Quantity



Top 5 Market & Segment by Quantity	
Market and Segment	Total Quantity
APAC – Consumer	21414
LATAM – Consumer	19853
EU – Consumer	19541
US - Consumer	19521
APAC - Corporate	12142

2. Data Preparation & Requirements

1. Plot 4 - Below chart depicts Profits by Market & Segment
   - Bigger the bar, larger the Profit
     - APAC – Consumer, EU – Consumer, US – Consumer & APAC – Corporate have larger Profits
     - Low COV & High Quantity of APAC – Consumer, EU – Consumer, APAC – Corporate
     - High COV & High Quantity of US - Consumer



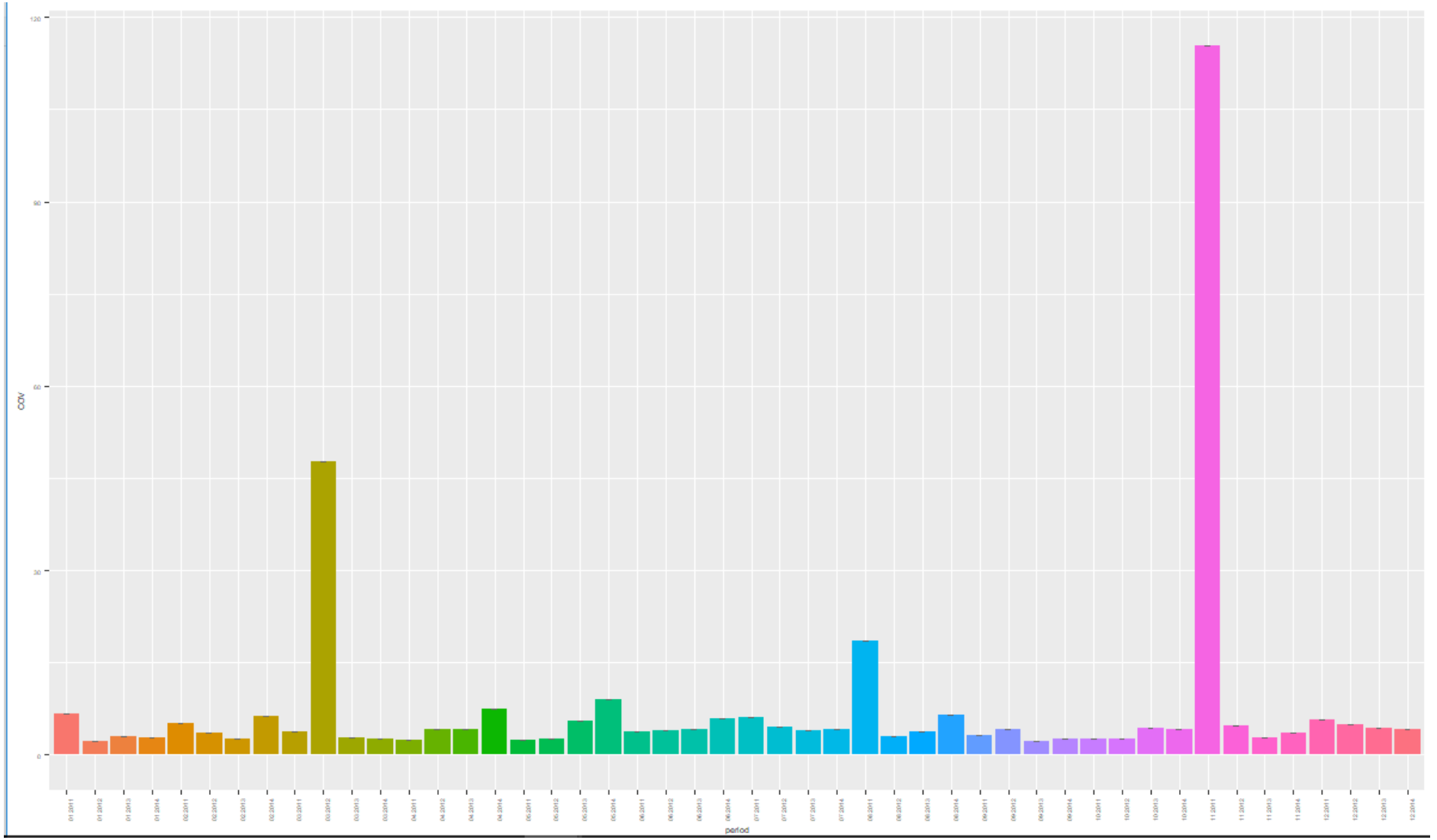
2. Data Preparation & Requirements

APACConsumerPlot

APAC – Consumer Month on Month view of COV

Observation :-

APAC_Consumer is Mostly consistent on Month-on-Month Coefficient of Variation



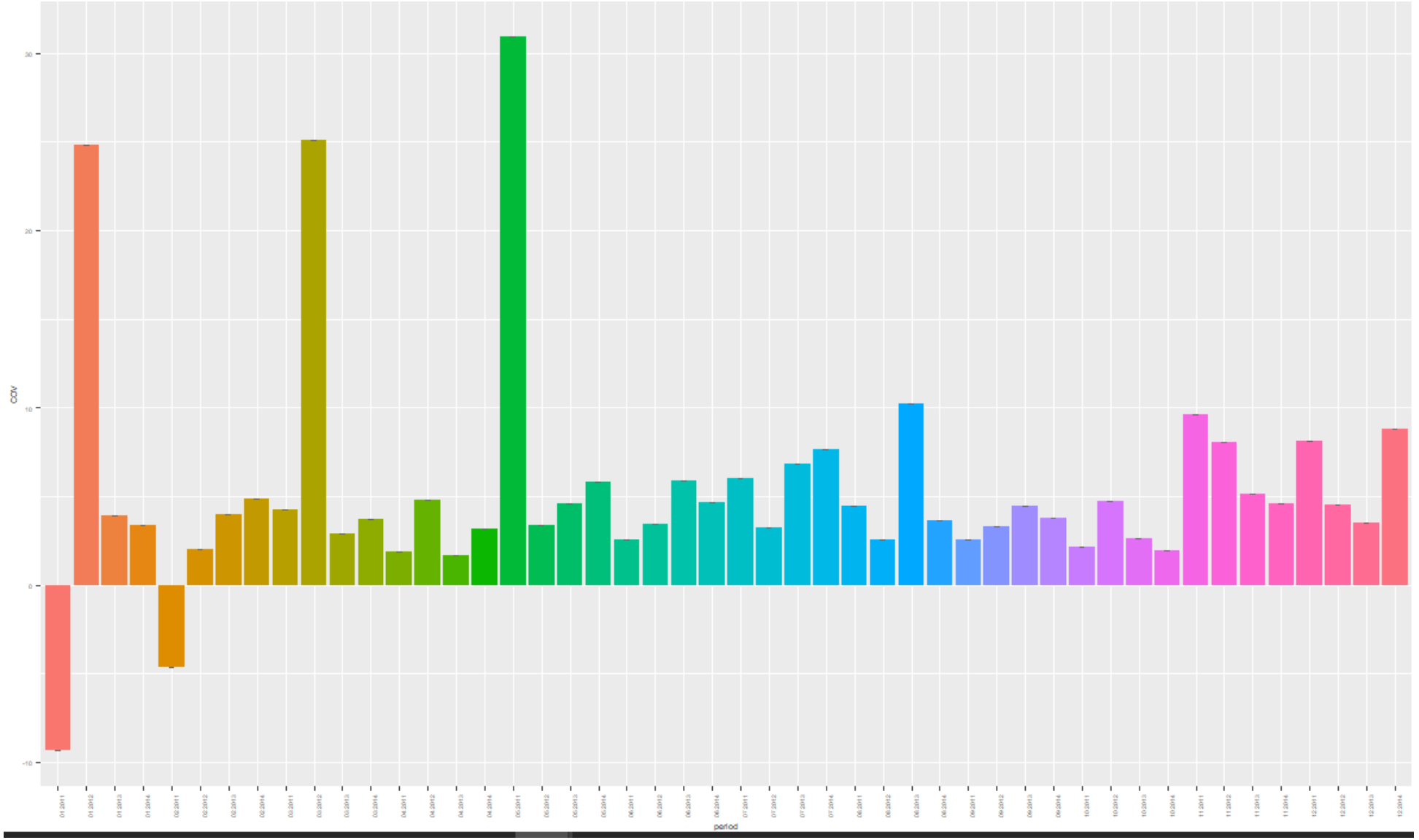
2. Data Preparation & Requirements

EUConsumerPlot

EU – Consumer Month on Month view of COV

Observation :-

EU_Consumer Month-on-Month Coefficient of Variation have spikes but overall Monthly average is consistent barring 5 months



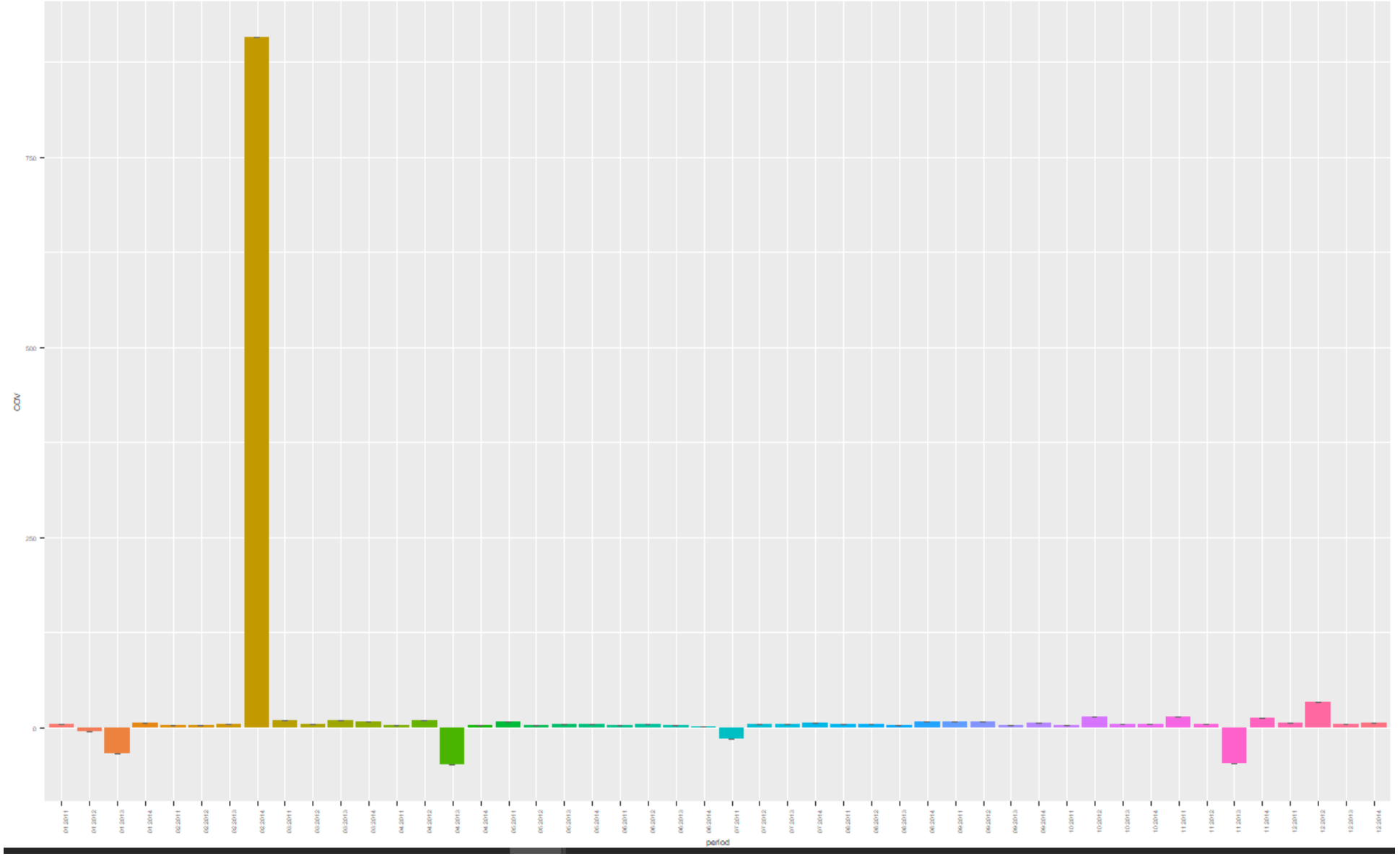
2. Data Preparation & Requirements

USConsumerPlot

US – Consumer Month on Month view of COV

Observation :-

US_Consumer is Mostly inconsistent on Month-on-Month Coefficient of Variation



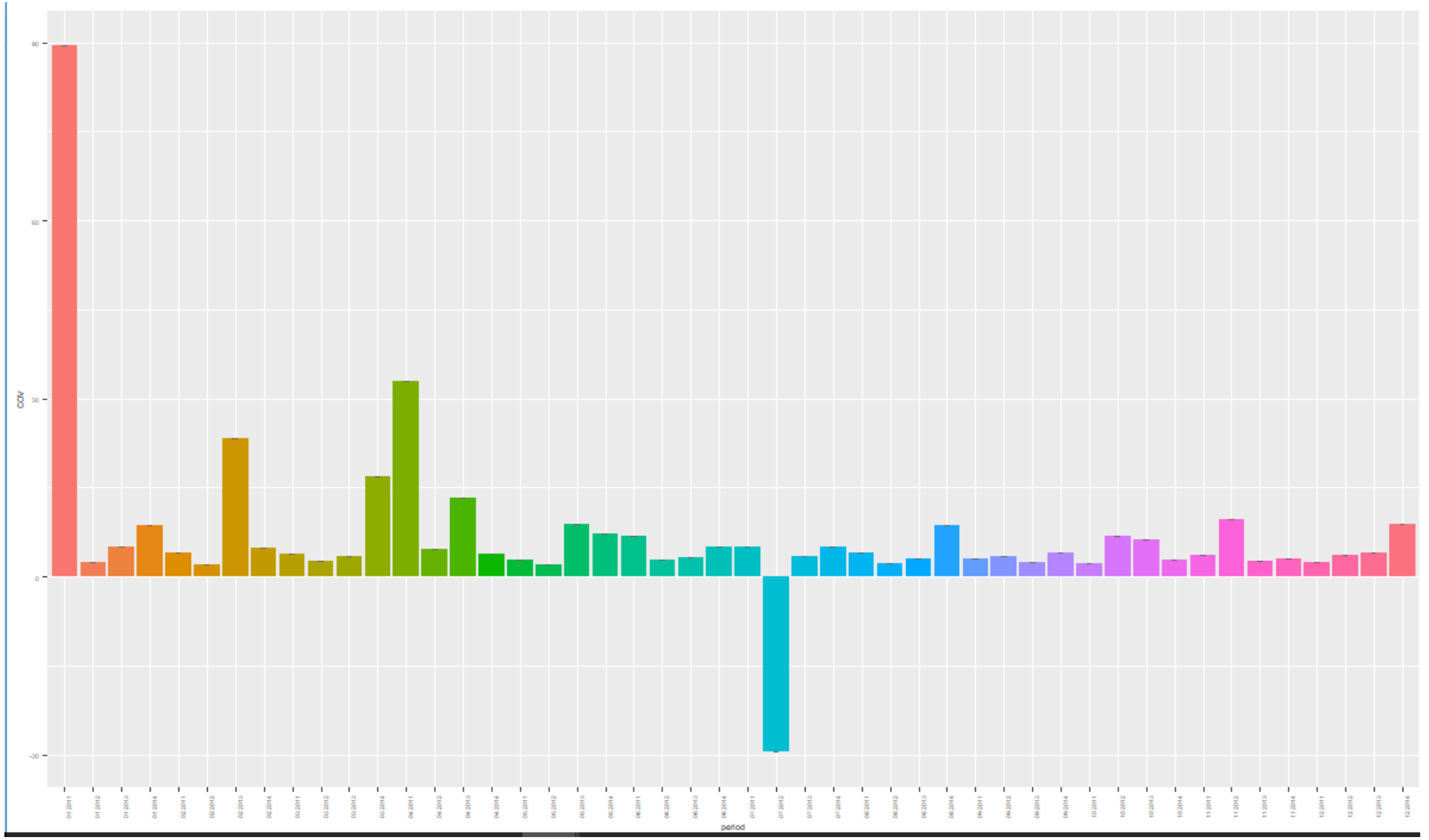
2. Data Preparation & Requirements

APACCorporatePlot

APAC – Corporate Month on Month view of COV

Observation :-

APAC_Corporate is Mostly inconsistent on Month-on-Month Coefficient of Variation with huge variation



2. Data Preparation & Requirements

Across 21 Market Segments, Top 2 Market Segments based on Higher Profits, Higher Quantity and Low COV are

- APAC – Consumer (Ref : SalesAPACCon & QtyAPACCon)
- EU – Consumer (Ref : SalesEUCon & QtyEUCon)

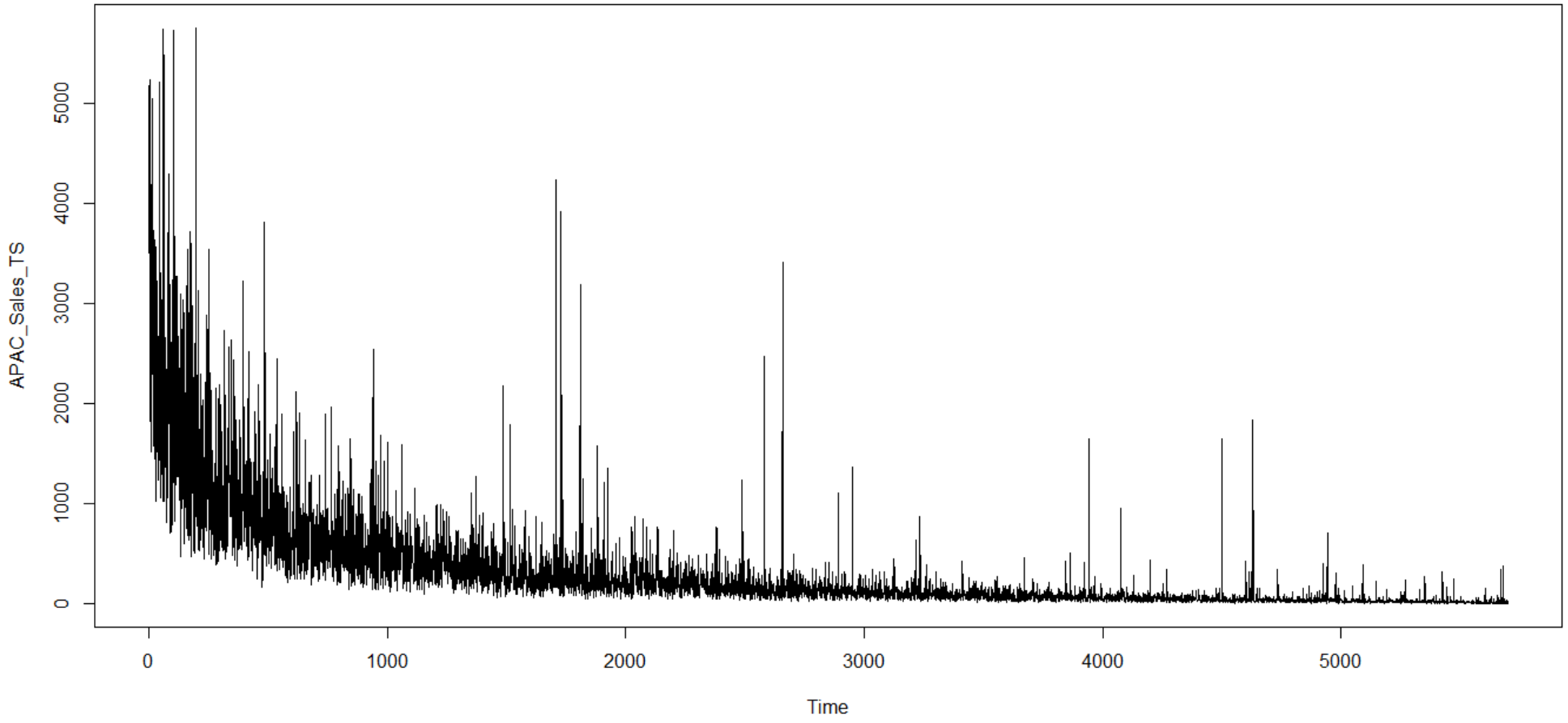
Top 7 Market Segments View and Top 2 Selection Criteria									
Market Segment	Total Sales	Profit Percentage	Total Quantity	Total Profit	COV	Rank Quantity	Rank Profit	COV	Top 2 Segments
APAC - Consumer	1816753.7	12.3	21414	222817.6	3.2	1	1	Low COV	✓
EU - Consumer	1529716.2	12.3	19541	188687.7	2.8	3	2	Low COV	✓
US - Consumer	1161401.3	11.5	19521	134119.2	21.8	4	3	High COV	✗
APAC - Corporate	1078466.3	12.0	12142	129737.2	2.1	5	4	Low COV	✗
EU - Corporate	920008.3	13.4	11635	123394.0	22.0	6	5	High COV	✗
LATAM - Consumer	1133847.0	10.6	19853	120632.9	6.3	2	6	High COV	✗
US - Corporate	706146.4	13.0	11608	91979.1	9.3	7	7	High COV	✗

Time Series Analysis

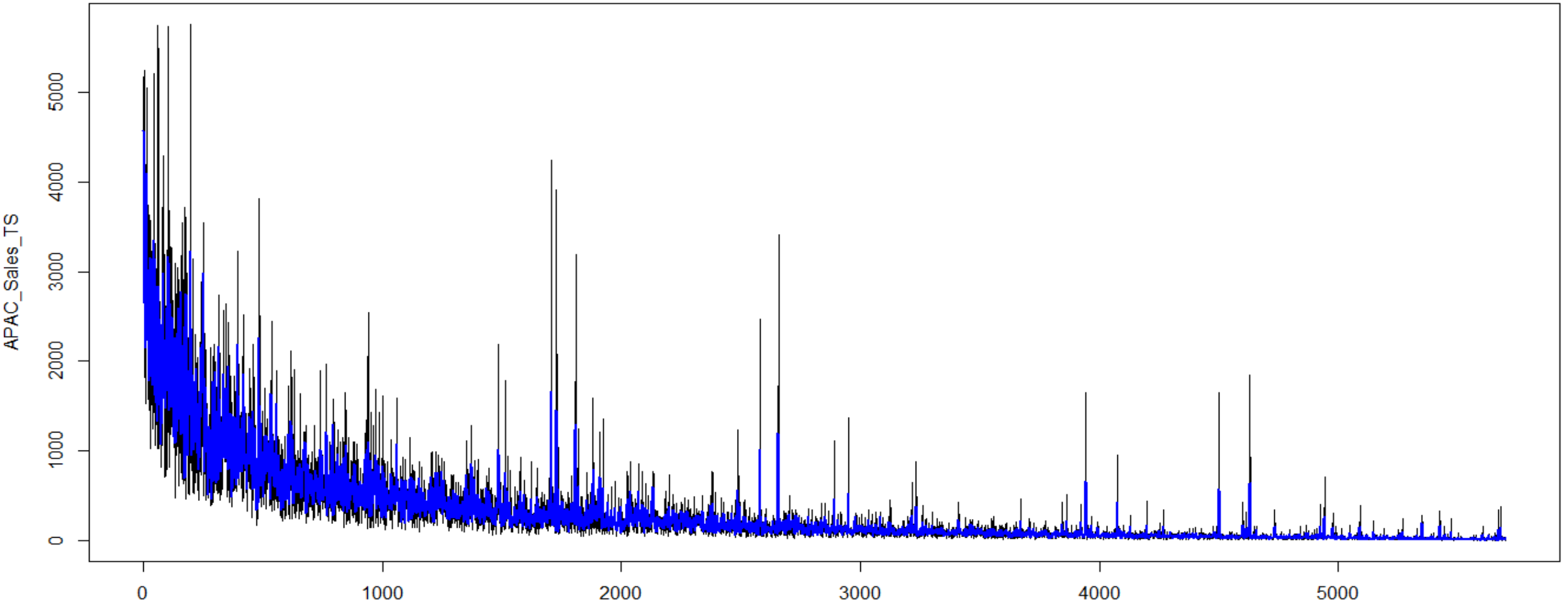
- In subsequent slides Time series analysis has been done for APAC Sales, APAC Quantity , EU Sales and EU Quantity
- First modeling was done at Decomposition models by predicting Global, Local and Manual ARIMA
- After doing the Manual modeling, evaluation was done by plotting ACF and PACF
- Also done Auto ARIMA for all the 4 required attributes
- By checking MAPE for all the modls Auto ARIMA is best suited for Predictions

Time series Analysis for APAC-Sales

There is downward trend in the sales , also there are spikes in sales



Smoothening APAC Sales curve

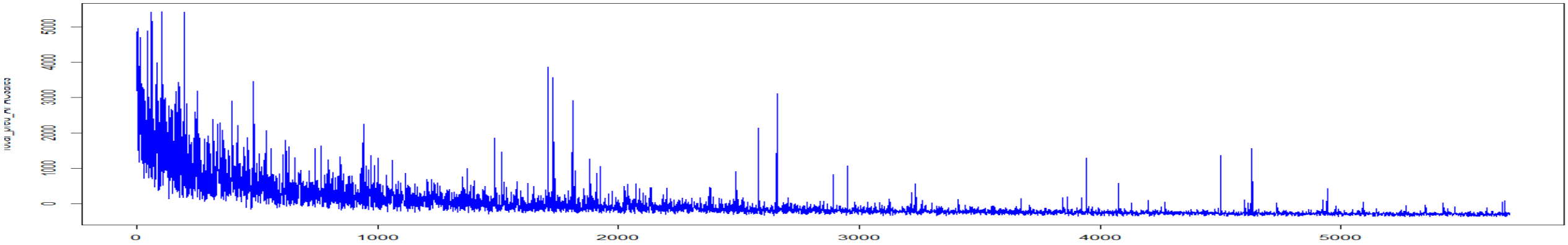


Global Prediction

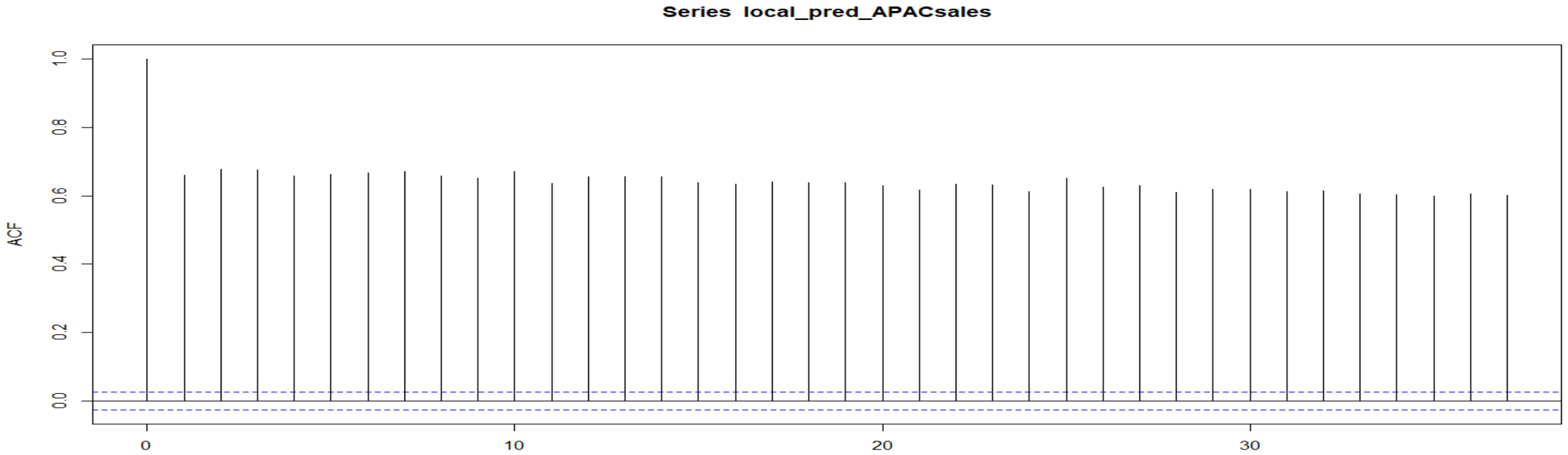
```
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Residual standard error: 478.9 on 5687 degrees of freedom
```

```
lmfit_APACsales <- lm(smootheddf_Apacsales$Sales  ~
sin(0.5*smootheddf_Apacsales$Order_Date ) * poly( smootheddf_Apacsales$Order_Date
,3) + cos(0.5*smootheddf_Apacsales$Order_Date) *
poly(smootheddf_Apacsales$Order_Date,3)
        + smootheddf_Apacsales$Order_Date, data=smootheddf_Apacsales)
```

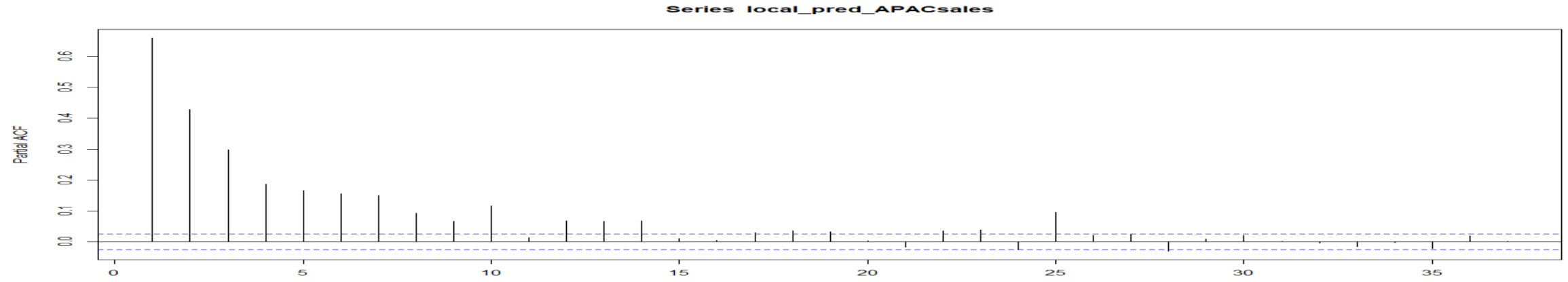
Local Predictions



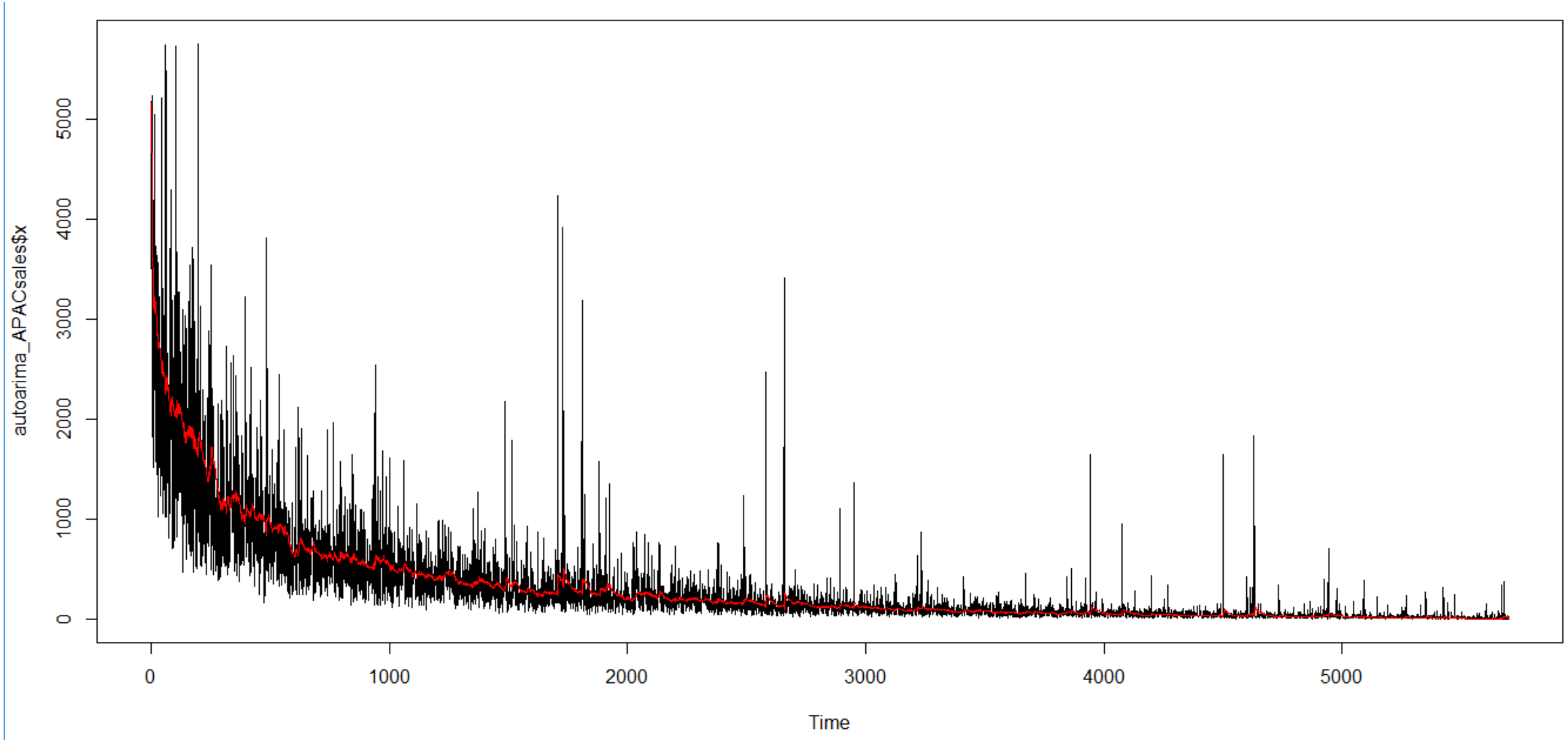
ACF an PACF



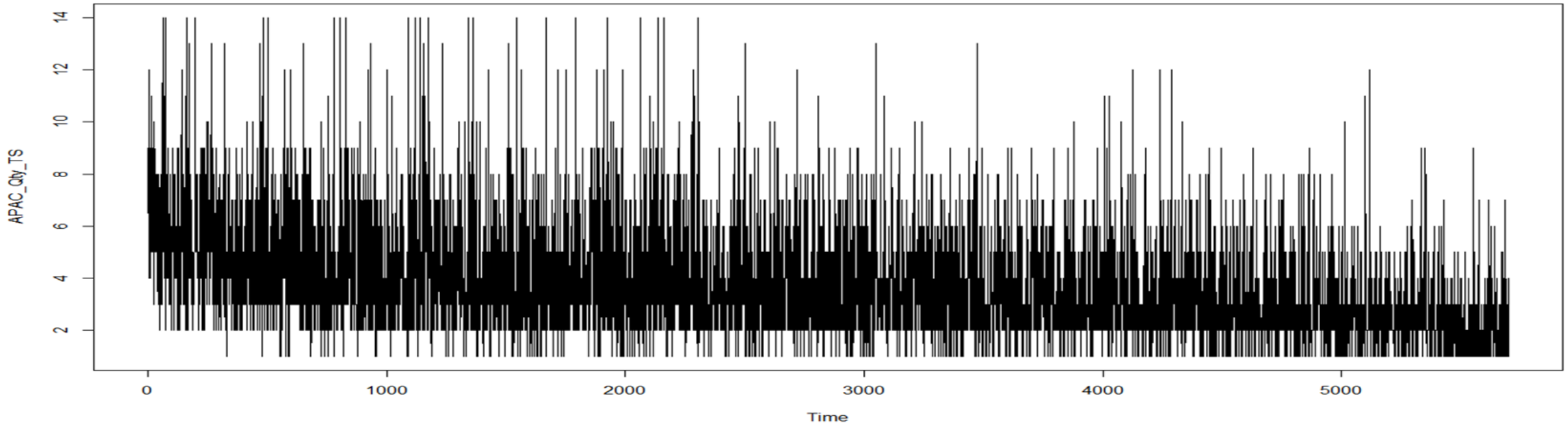
For ACF and PACF the historical data is in optimal value

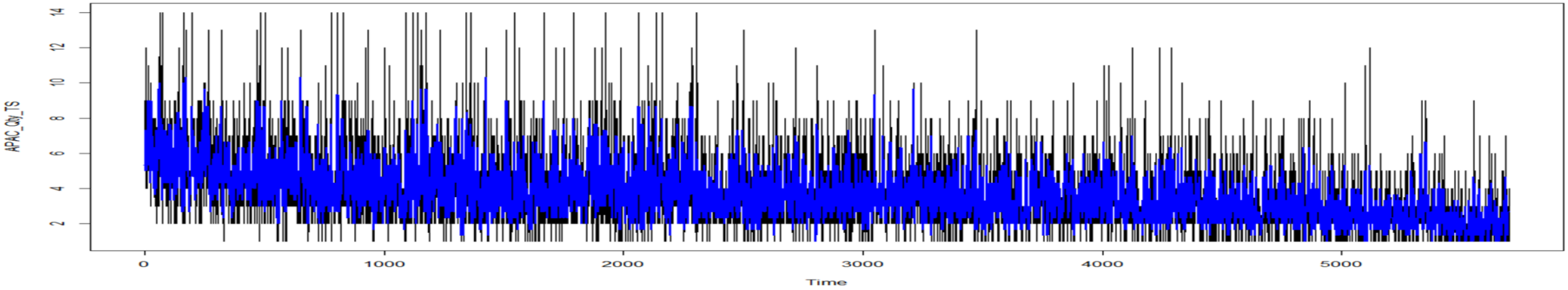


Auto ARIMA –APAC Sales

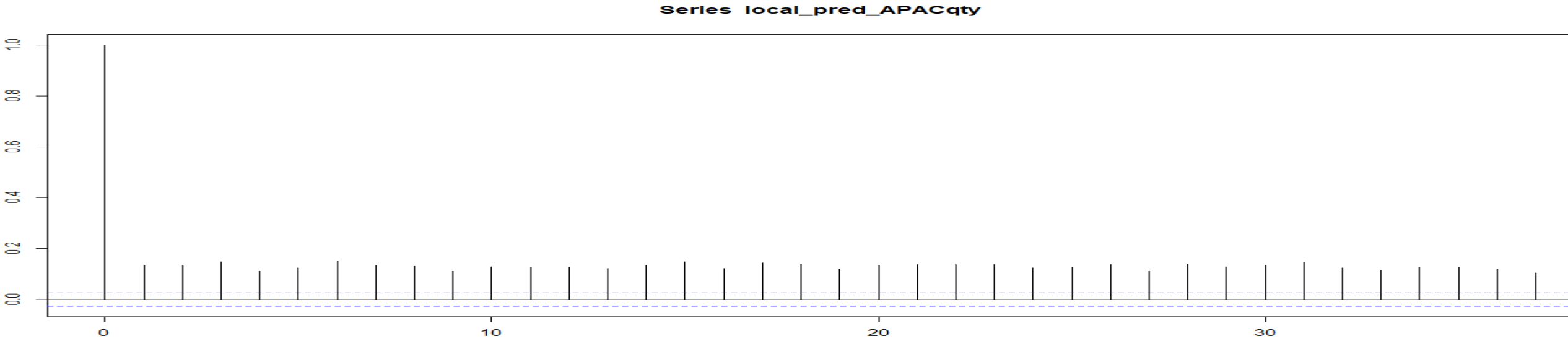


Time Series plots- APAC Quantity- with and with out smoothing

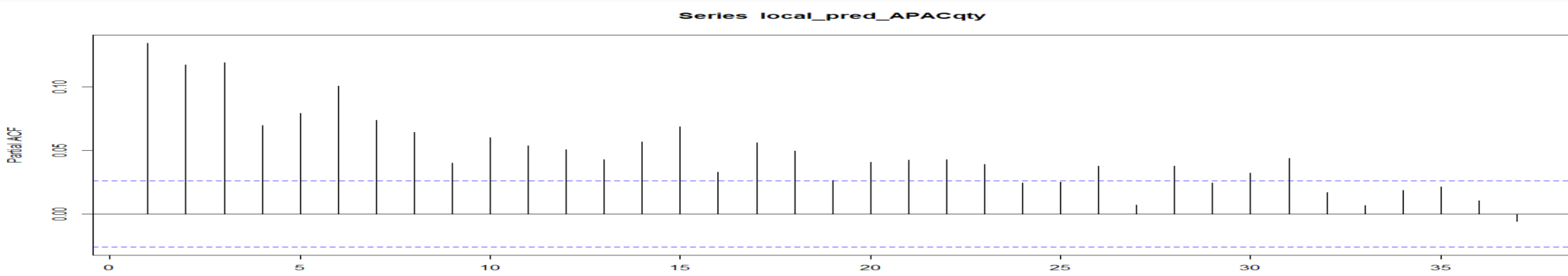




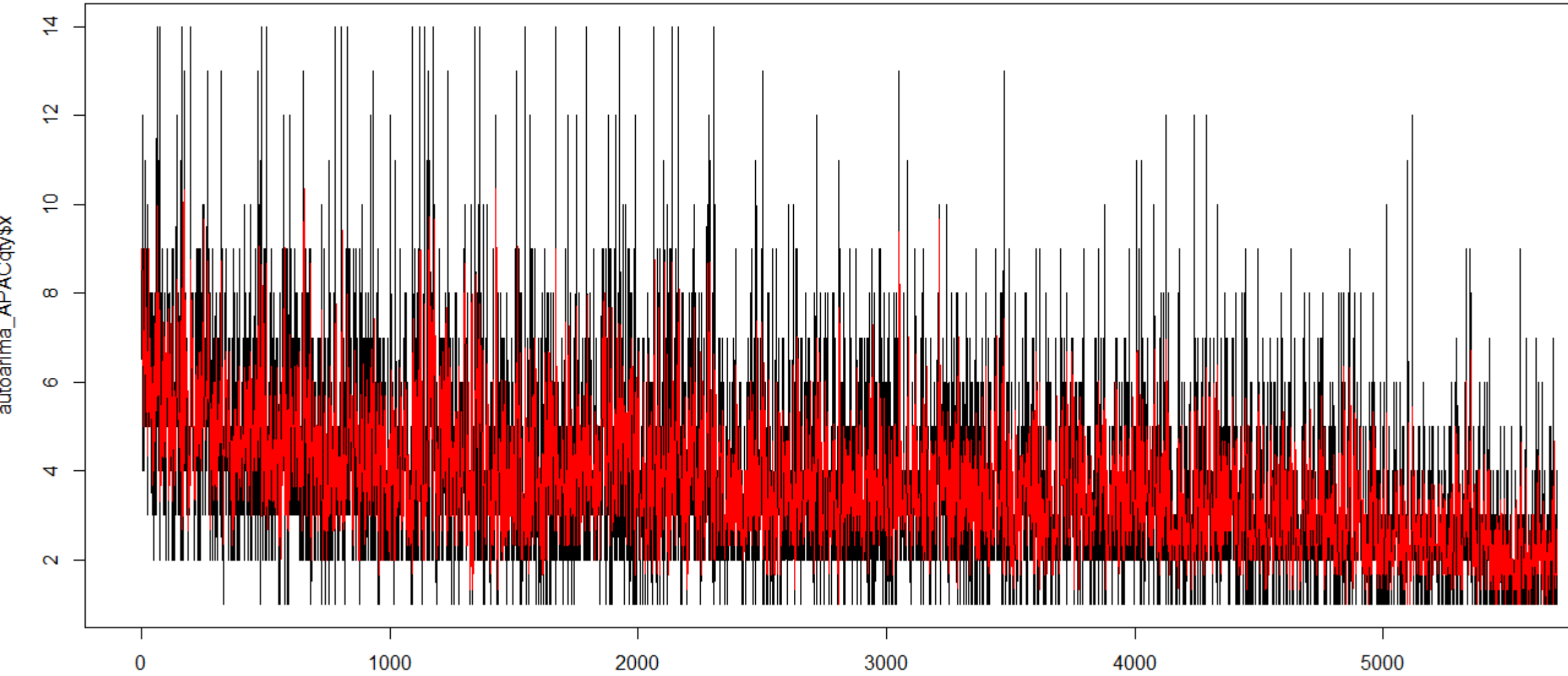
ACF and PACF for APAC Quantity



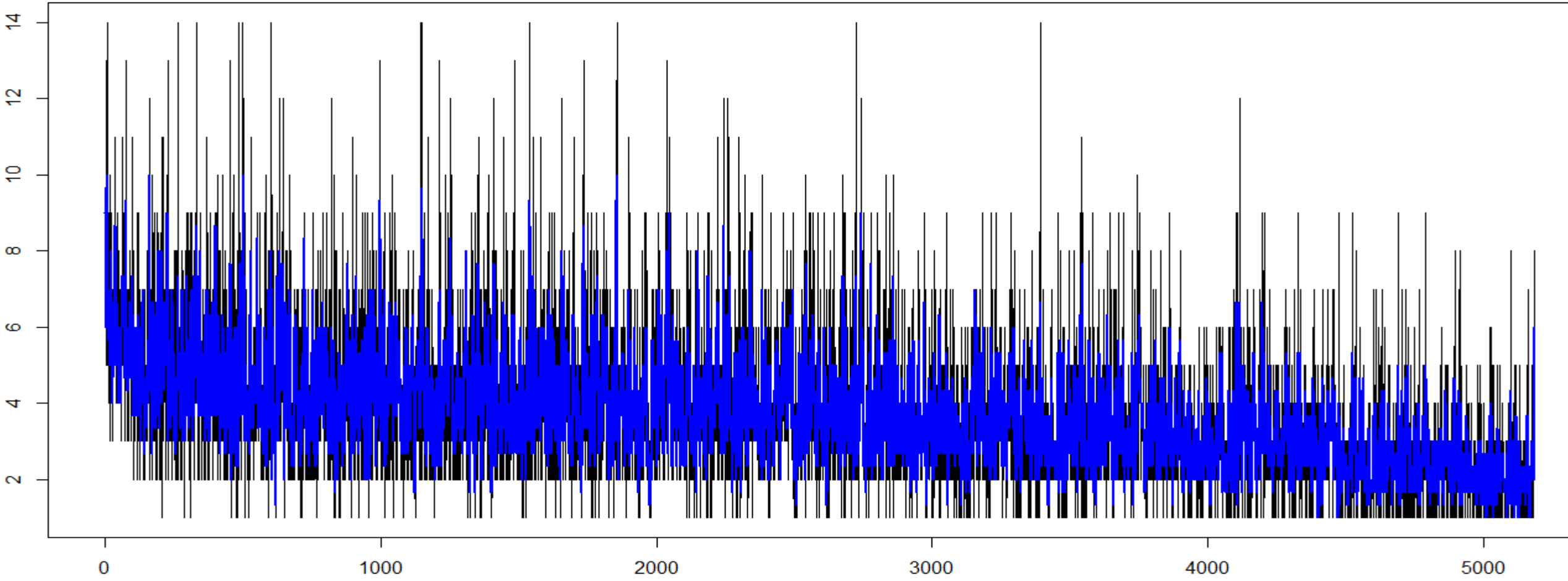
For ACF and PACF the historical data is in optimal value



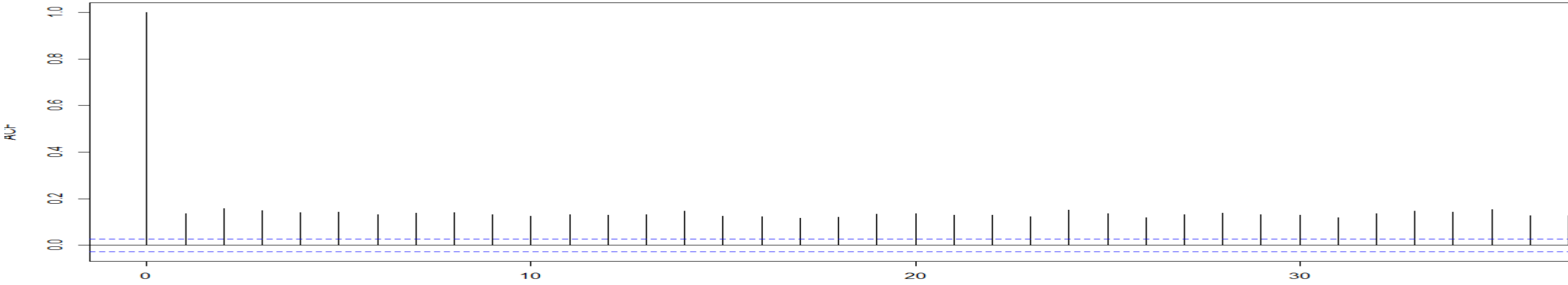
Auto ARIMA- APAC Qty



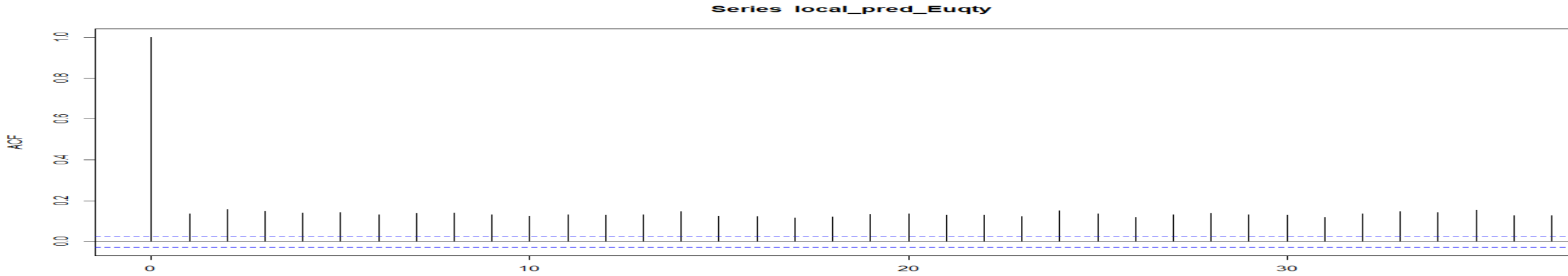
Time Series for EU Quantity – With and without smoothing



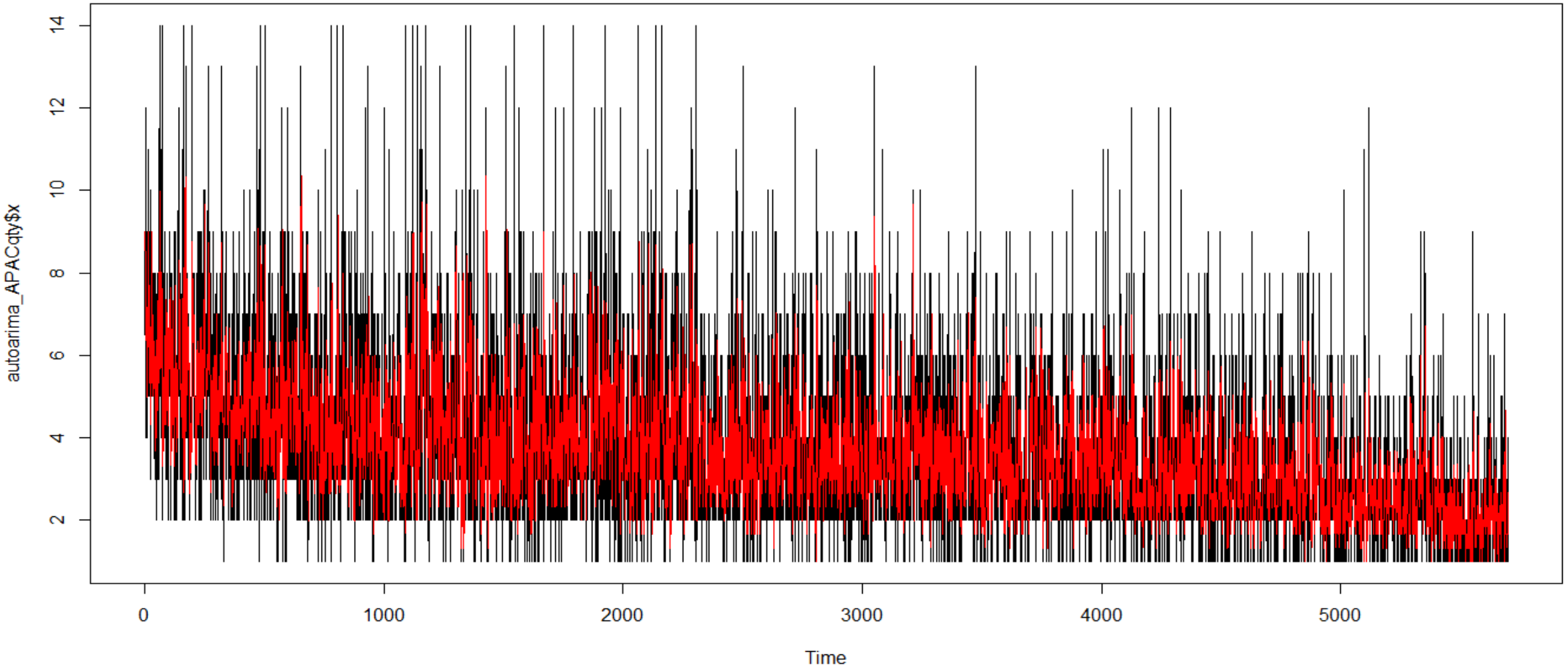
ACF and PACF- EU Quantity



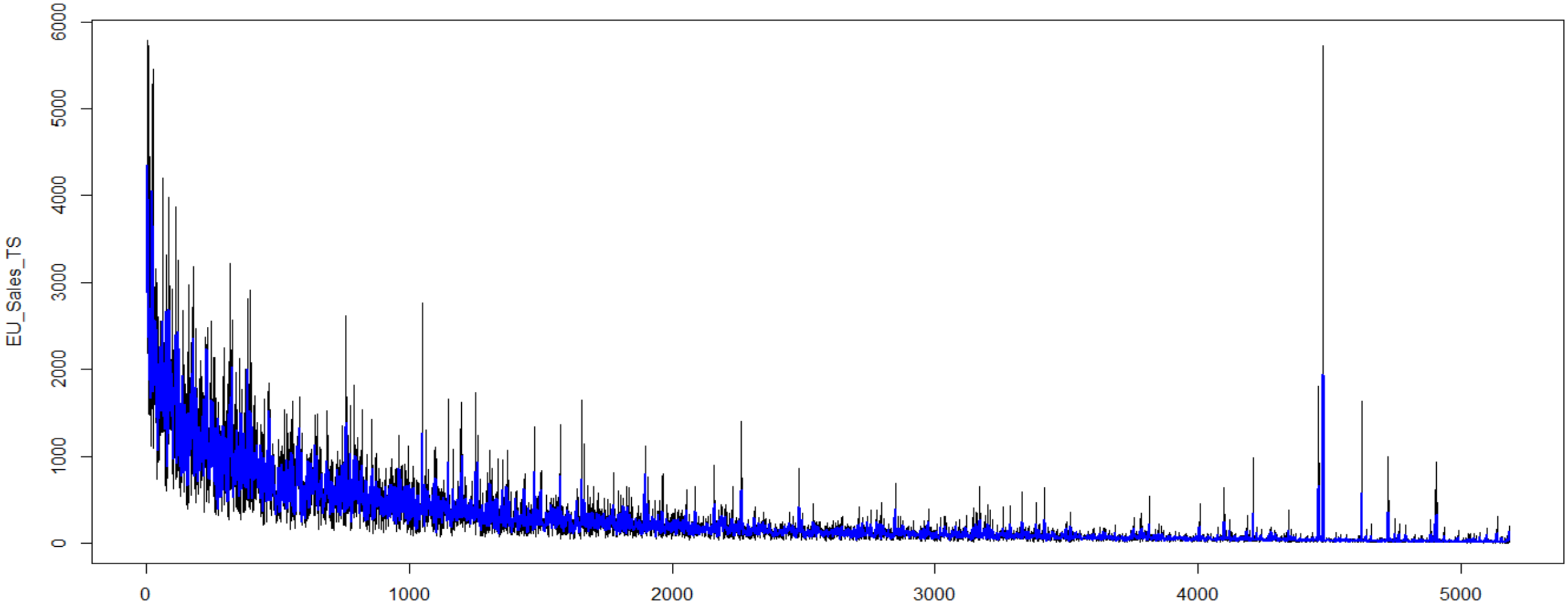
For ACF and PACF the historical data is in optimal value



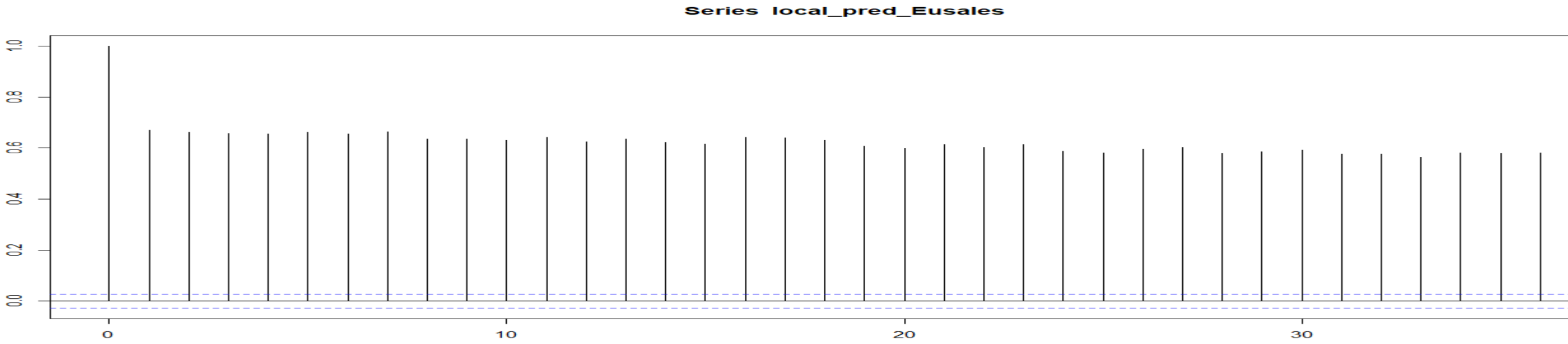
Auto ARIMA- EQ Quantity

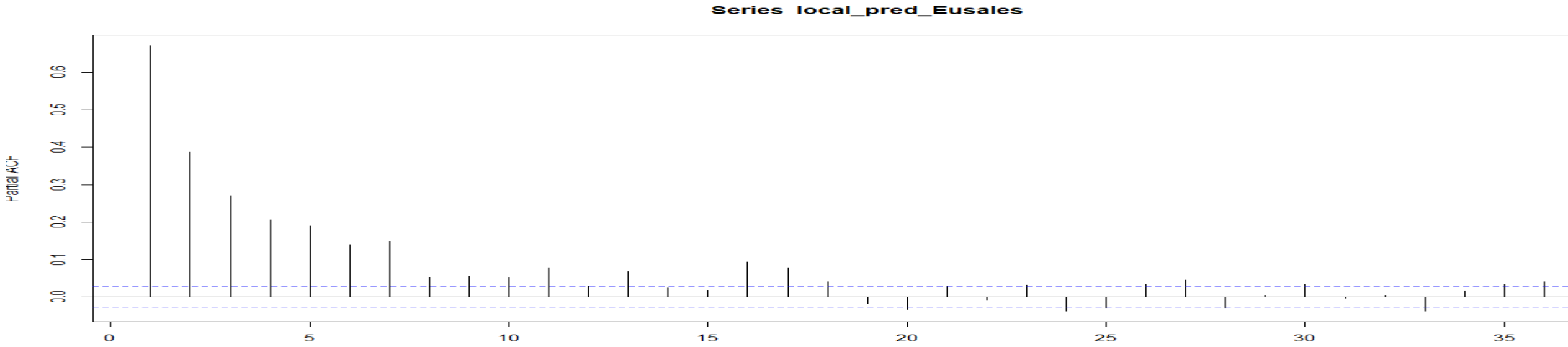


Time Series – EU Sales with and without smoothing

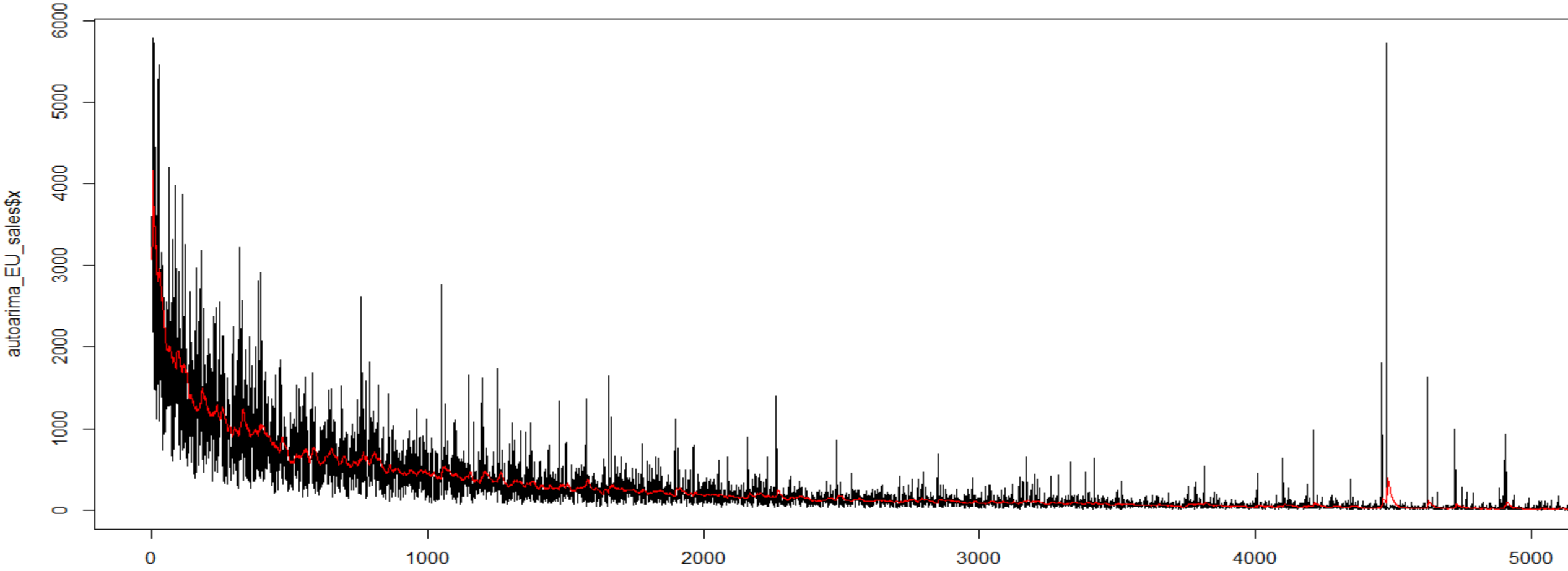


ACF and PACF- EU Sales





Auto ARIMA- EU Sales



Forecasting- for 6 months

- As specified EU- Customer and APAC- Customer are more profitable for stores
- By auto ARIMA process the Predictions are follows::
- Sales units are in Lacs and Quantity units are in Thousands

APAC Sales Next6 months	EU Sales Next6 months	APAC Qty Next6 months	EU Qty Next6 months
17.39977	23.77783	2.683024	3.261499
16.46559	21.75368	2.241486	3.460744
15.8324	25.38164	1.959972	3.500517
15.40814	24.85152	2.300049	3.710197
14.90011	28.01747	2.16877	4.01198
14.42566	27.4482	27.4482	4.261811

THANK YOU