ARIMA FORECASTING & IDENTIFICATION

Dr. Susan Simmons

Institute for Advanced Analytics

Relationship Between AR and MA

- The best part about AR models and MA models is that they are the same thing – approximately.
- In certain situations (stationarity), AR models can be represented as an infinite MA model.
- In certain situations (invertible), MA models can be represented as an infinite AR model.

ARMA Model

- There is nothing to limit both an AR process and an MA process to be in the model simultaneously.
- These “mixed” models are typically used to help reduce the number of parameters needed for good estimation in the model.
- We are going to focus on the most basic model with only one lag of each piece – the ARMA(1,1) model.

$$Y_t = \omega + \phi Y_{t-1} + e_t - \theta e_{t-1}$$

ARMA(1, 1) – ACF, PACF, IACF

- Although these terms can be calculated for the ARMA(1,1) process, they become very complicated.
- There are some important things to note:
  - Characteristics from both are in the correlation functions.
  - All of the functions tail off exponentially as the lags increase.

What does the following models represent?

- ARMA(1,3)
- ARMA(4,2)
- ARMA(1,1)
- ARIMA(1,1,4)
- ARIMA(2,1,1)

MODEL SELECTION

Identification

- There are a couple of different sets of techniques used for model identification for stationary models.
  1. Plotting Patterns – ACF, PACF, IACF
  2. Automatic Selection Techniques (SAS):
     - Minimum Information Criterion – MINIC
     - Smallest Canonical Correlation – SCAN
     - Extended Sample Autocorrelation Function – ESACF
  3. Automatic Selection Techniques (R):
     - auto.arima Function – Similar to MINIC

Recommendation for automatic scans:

- Make sure the series is stationary (i.e. if it has a trend, or random walk, take care of it)
  - In SAS, if you take differences, it will do the scan on the differenced data
  - In SAS, if you have stationarity about the trend line, you should run the automatic procedures on the residuals (i.e. fit the regression line, get residuals and run residuals through PROC ARIMA)

Automatic Selection Techniques (SAS)

- In addition to the ACF, PACF, and IACF, there are 3 automatic methods for identifying models for the data in SAS – MINIC, SCAN, ESACF. The stationarity was already checked…

```
proc arima data=Time.Hurricanes plot(unpack);
     identify var=MeanVMax nlag=12 minic
     P=(0:12) Q=(0:12);
run;
quit;
```

Minimum Information Criterion								
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5	MA 6	MA 7
AR 0	4.58498	4.541439	4.539209	4.515101	4.506568	4.539417	4.508287	4.544625
AR 1	4.601114	4.537332	4.51762	4.46991	4.479599	4.51957	4.515029	4.558396
AR 2	4.611816	4.535499	4.511774	4.497311	4.512292	4.555764	4.550454	4.594881
AR 3	4.53233	4.461663	4.48184	4.523381	4.534056	4.567099	4.580285	4.624875
AR 4	4.553621	4.497145	4.517672	4.559297	4.570437	4.60408	4.617382	4.66192

Minimum Table Value: BIC(3,1) = 4.461663

Automatic Selection Techniques (SAS)

- In addition to the ACF, PACF, and IACF, there are 3 automatic methods for identifying models for the data in SAS – MINIC, SCAN, ESACF.

```
proc arima data=Time.Hurricanes plot(unpack);
     identify var=MeanVMax nlag=12 scan
     P=(0:12) Q=(0:12);
run;
quit;
```

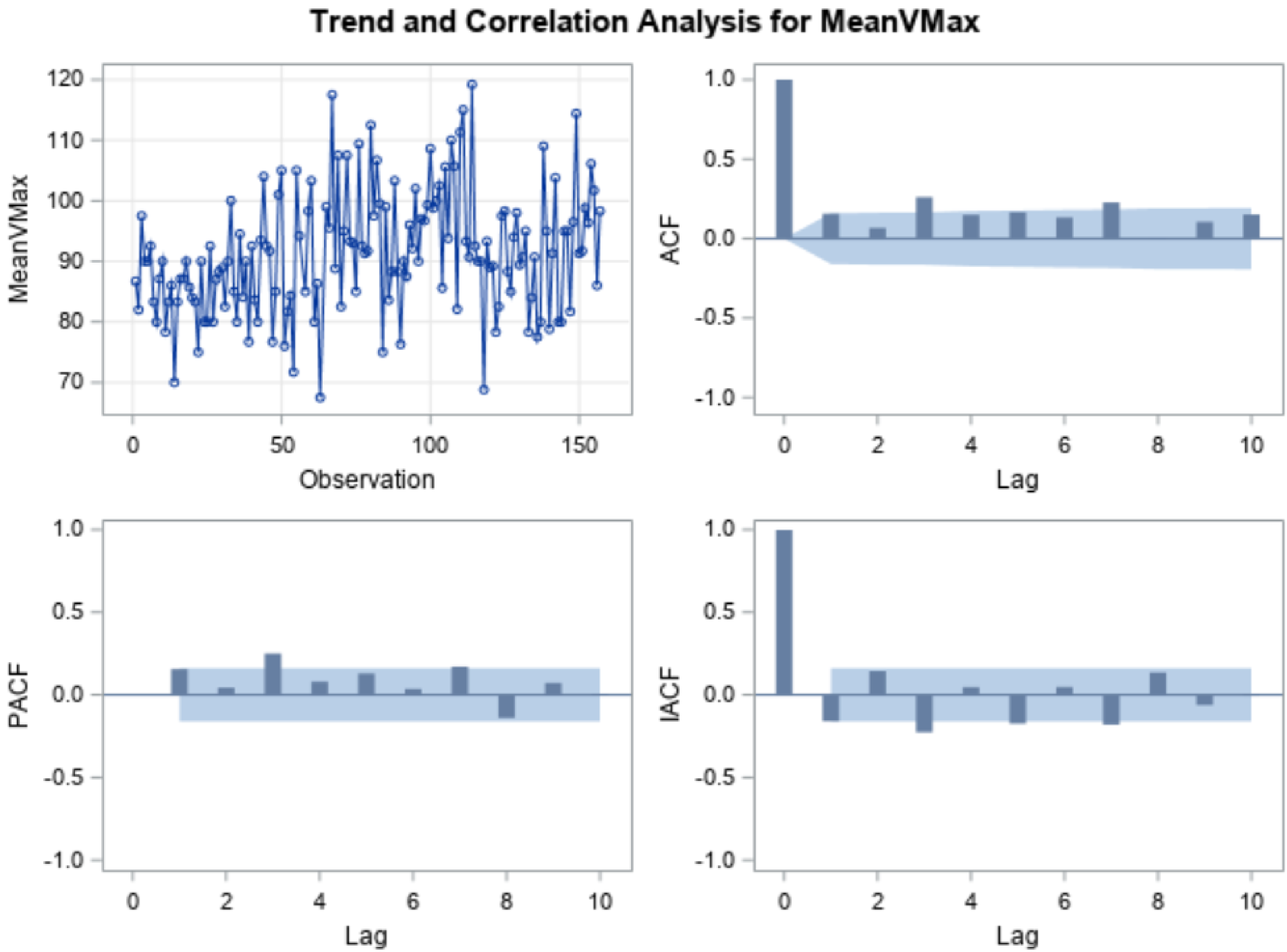
ARMA(p+d,q) Tentative Order Selection Tests	
SCAN	
p+d	q
1	1
7	0
0	7

Automatic Selection Techniques (SAS)

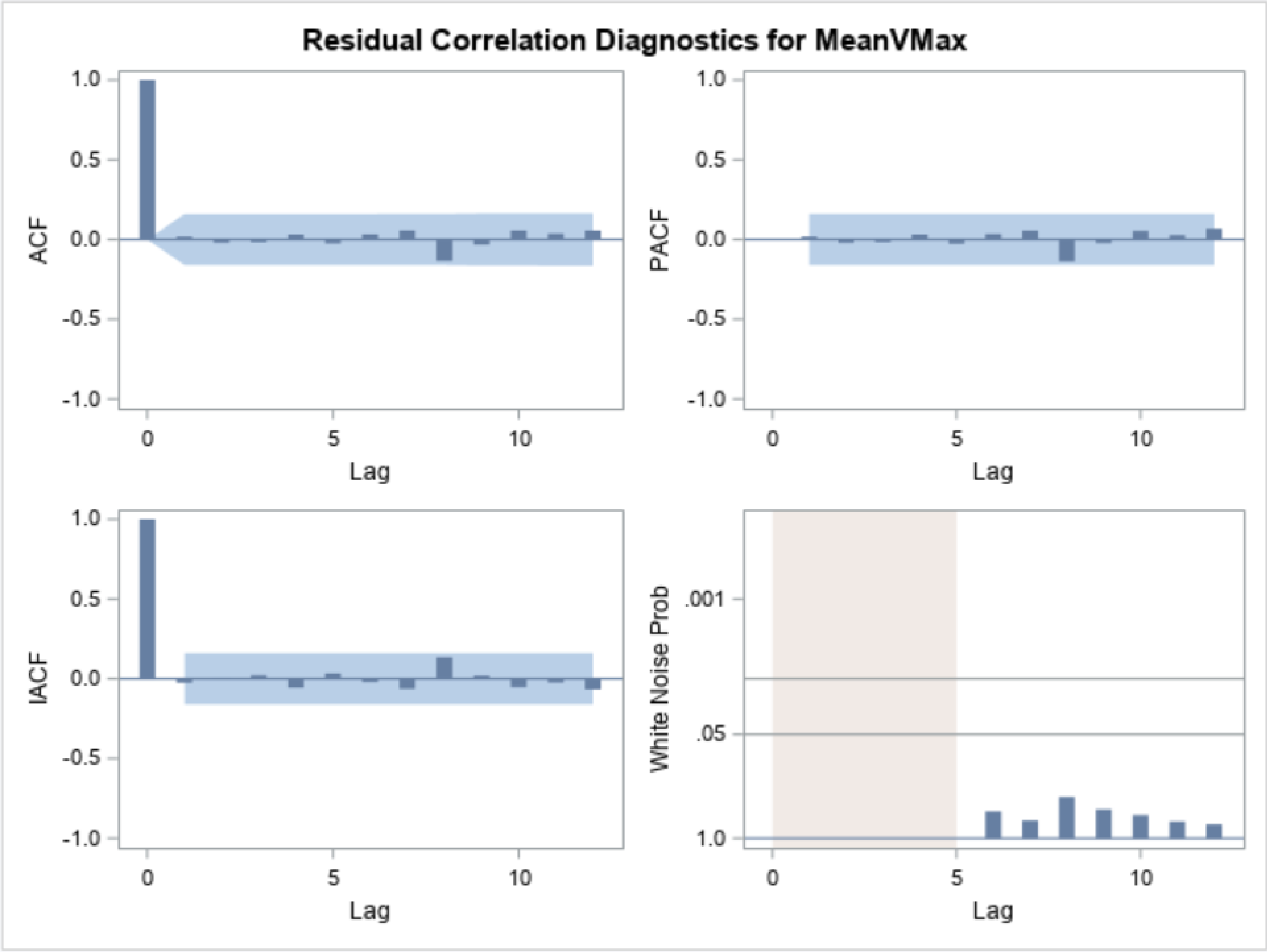
- In addition to the ACF, PACF, and IACF, there are 3 automatic methods for identifying models for the data in SAS – MINIC, SCAN, ESACF.

```
proc arima data=Time.Hurricanes plot(unpack);
     identify var=MeanVMax nlag=12 esacf
     P=(0:12) Q=(0:12);
run;
quit;
```

ARMA(p+d,q) Tentative Order Selection Tests	
ESACF	
p+d	q
2	2
6	2
8	3
9	3
0	7
1	7
12	5


Trend and Correlation Analysis for MeanVMax
MeanVMax
Observation
ACF
PACF
IACF
Lag
Lag
Lag

Correlations of Parameter Estimates						
Parameter	MU	MA1,1	MA1,2	MA1,3	AR1,1	AR1,2
MU	1.000	0.025	0.073	0.030	0.029	0.055
MA1,1	0.025	1.000	-0.448	-0.649	0.886	-0.703
MA1,2	0.073	-0.448	1.000	0.510	-0.486	0.845
MA1,3	0.030	-0.649	0.510	1.000	-0.442	0.598
AR1,1	0.029	0.886	-0.486	-0.442	1.000	-0.790
AR1,2	0.055	-0.703	0.845	0.598	-0.790	1.000


Residual Correlation Diagnostics for MeanVMax
ACF
PACF
IACF
White Noise Prob
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
5
10
Lag

Automatic Selection Techniques (R)

- In addition to the ACF, PACF, and IACF, there are 2 automatic methods for identifying models for the data in R – auto.arima, EACF.

```
Hurricane.ts2<-Hurricane.ts%>% na_interpolation(option =
"spline")
auto.arima(Hurricane.ts2)
```

Series: Hurricane.ts2
ARIMA(0,1,1)

Coefficients:

```
          ma1
      -0.9050
s.e.   0.0429
```

sigma^2 estimated as 95.59: log likelihood=-577.39
AIC=1158.79   AICc=1158.87   BIC=1164.89

```
Hurricane.model<-Arima(Hurricane.ts2,order=c(2,0,3))
summary(Hurricane.model)
```

```
Series: Hurricane.ts2
ARIMA(2,0,3) with non-zero mean

Coefficients:
         ar1     ar2      ma1      ma2     ma3     mean
      0.2387  0.6373  -0.1160  -0.7093  0.1592  91.2544
s.e.  0.1904  0.1921   0.2031   0.1439  0.1088   1.9170
```

FORECASTING

Estimation Methods – CLS

- Conditional Least Squares estimators are the following:
  - Generally inferior to MLE for small samples
  - More computationally efficient than MLE
  - Are the DEFAULT in PROC ARIMA
- “Conditional” least squares comes from the fact that estimation of the parameter estimates is *conditioned* on unobserved past values being equal to the sample mean.

Estimation Methods – MLE

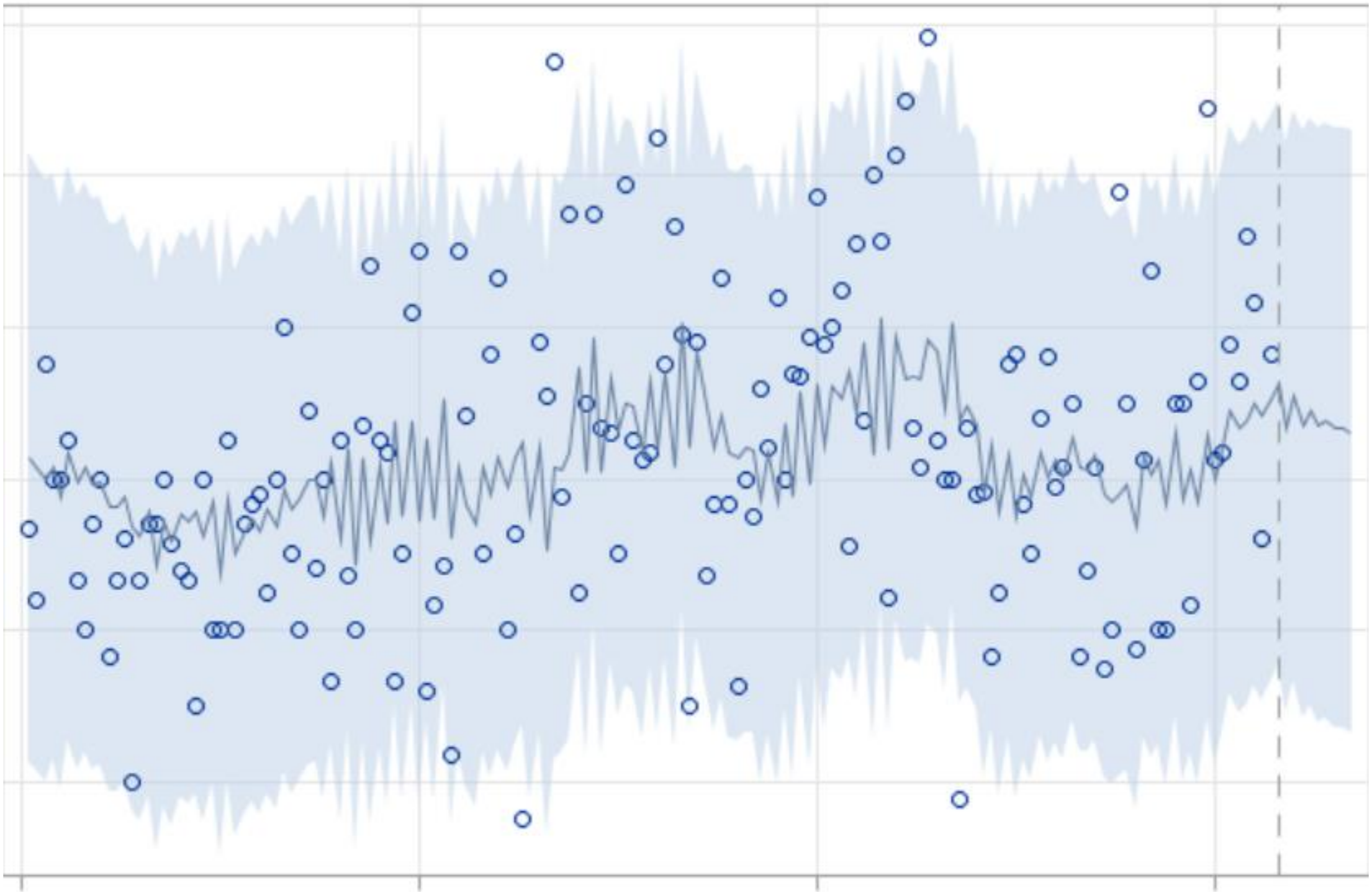
- Maximum Likelihood estimators are the following:
  - Superior to other estimates, especially in small samples.
  - Least efficient computationally, therefore not made the default when PROC ARIMA was created in 1980.
  - To help keep upward compatibility in SAS, the defaults have stayed the same in PROC ARIMA.
  - Method of choice by most forecasting professionals as well as SAS.

Optimization Algorithms

- CLS and ML algorithms are not guaranteed to find an optimal solution.
- Problems:
  - Local Maxima/Minima
  - Ridges (no improvement in any direction, but stopping rule not satisfied)
  - Stability Problems
  - Others

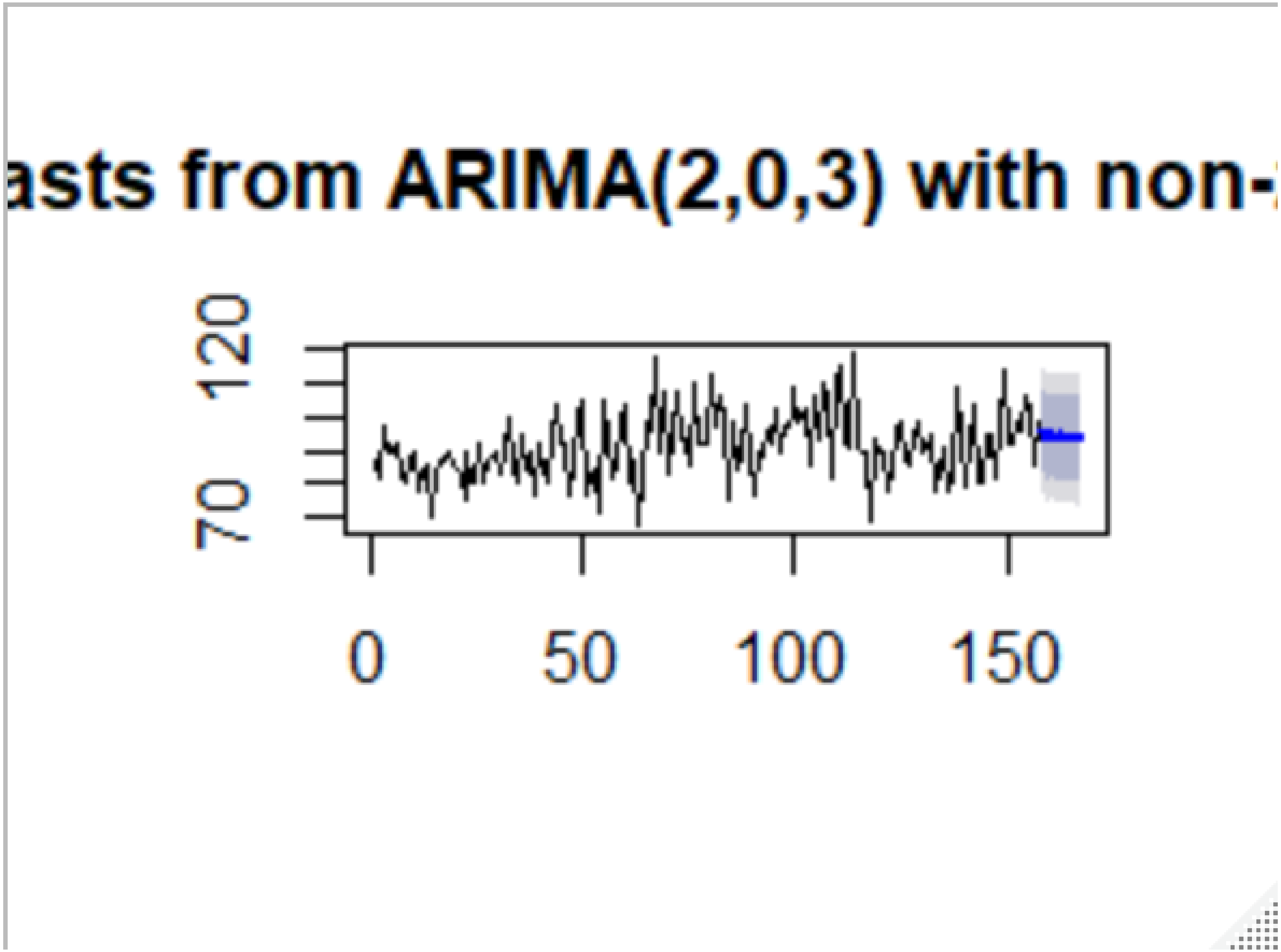
Forecasting – SAS

```
proc arima data=Time.Hurricanes plot=all;
          identify var=MeanVMax nlag=10 ;
          estimate  p=___ q=___ method=ML;
          forecast lead=10;
run;
quit;
```



Forecasting – R

```
forecast(Hurricane.model, h = 10)
plot(forecast(Hurricane.model, h = 10))
```


asts from ARIMA(2,0,3) with non-
120
70
0
50
100
150