Consumer Price Index (CPI) time series modeling

Winters Additive vs ARIMA(1,0,3)(0,1,1)

Assessment 3

Time Series & M-V Analysis - STAT8008
Higher Diploma in Data Science and Analytics

30.04.2018 / João Viegas / R00157699

Overview

- Dataset
- Models
  - Winters Additive
  - Arima(1,0,3)(0,1,1)
  - Selection
- Estimation and Forecasting - splitted data (training dataset)
- True Forecast
- Conclusions

Dataset

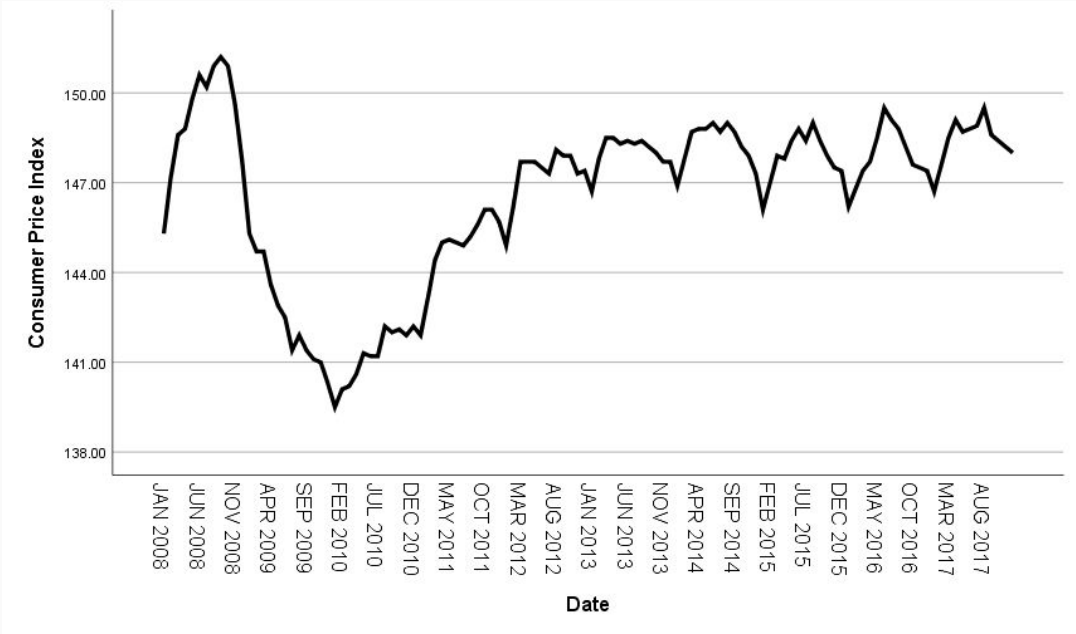
Dataset

Consumer Price Index (CPI)

Ireland

January 2008 to December 2017

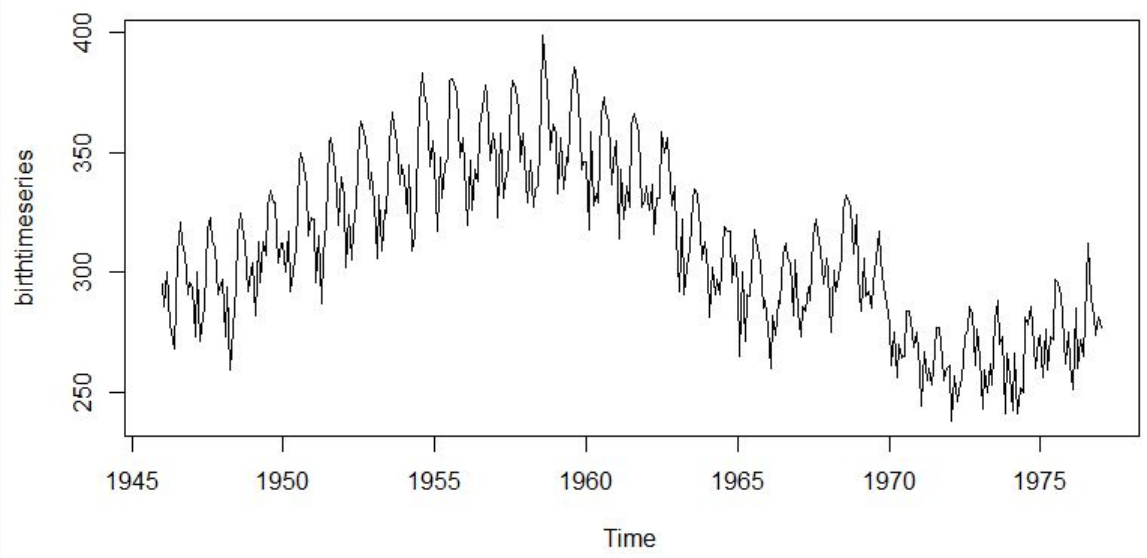
- seasonal;
- cycles difficult to predict;
- 2 phases:
  - runup & dip;
  - recovery to lower growth;



Models

Models: Winters Additive

- Increasing / decreasing series
- seasonal pattern of constant magnitude



Models: Winters Additive

Smoothed components:

level: $S(t) = S(t-1) + T(t-1) + \alpha\varepsilon(t)$
trend: $T(t) = T(t-1) + \alpha\gamma\varepsilon(t)$
seasonal: $I(t) = I(t-p) + \delta(1 - \alpha)\varepsilon(t)$

Predicted:

$y(t+m) = S(t) + mT(t) + I(t-p+m)$

Note: δ nonsignificant

Model			Estimate	SE	t	Sig.
Consumer Price Index-Model_1	No Transformation	Alpha (Level)	1	0.091	10.95	0
		Gamma (Trend)	0.102	0.029	3.525	0.001
		Delta (Season)	0.999	9361.466	0	1

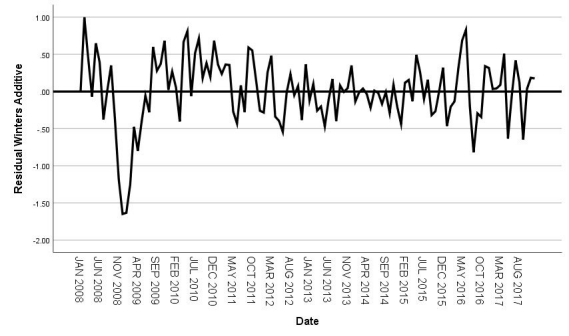
Models: Winters Additive

fit measures

- Stationary R-Squared: substantial explanation;
- Low error values;
- Ljung-Box statistic not significant (no autocorrelation);

residuals sequence plot

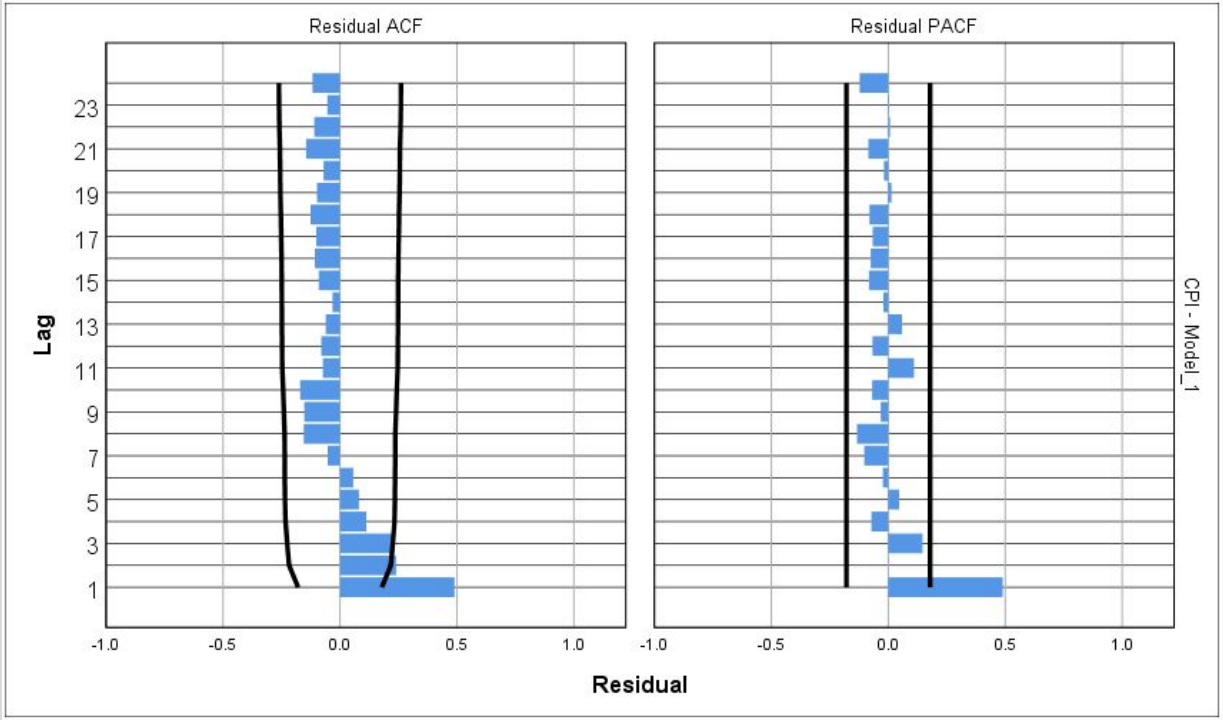
Number of Predictors		0
Model Fit statistics	Stationary R-squared	0.659
	R-squared	0.974
	RMSE	0.454
	MAPE	0.227
	MAE	0.332
	MaxAPE	1.124
	MaxAE	1.649
	Normalized BIC	-1.461
Ljung-Box Q(18)	Statistics	65.146
	DF	15
	Sig.	0
Number of Outliers		0



Models: Winters Additive

Residuals autocorrelation

- first lag maybe autocorrelated;
- probably effect of initial phase
- not confirmed by Ljung-Box statistic
- short time series



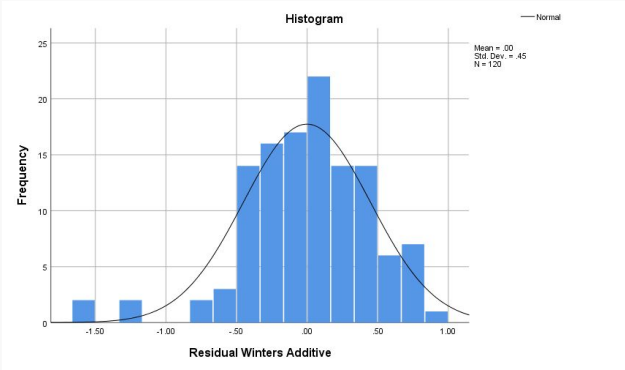
Models: Winters Additive

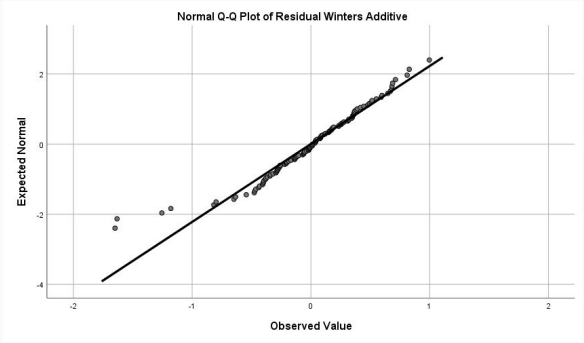
Residuals normality assumption

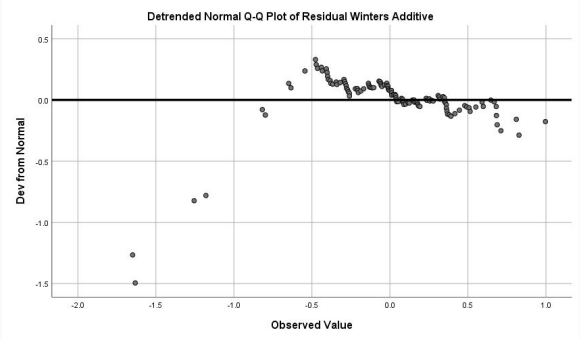
- Kolmogorov: yes
- Shapiro: no
- Histogram: maybe yes
- Normal Q-Q plot: maybe yes
- Detrended Normal Q-Q plot: maybe not

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual Winters Additive	0.07	120	.200*	0.951	120	0

[a] Lilliefors Significance Correction



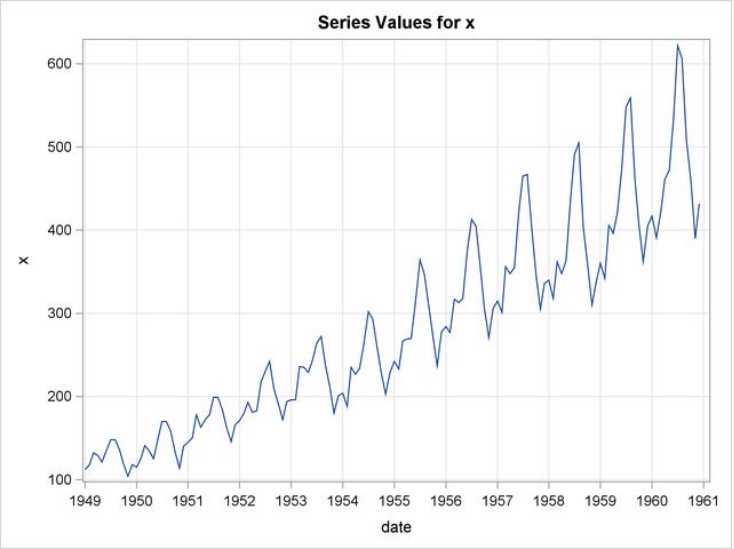




Models: Arima(1,0,3)(0,1,1)

ARIMA(p,d,q)(P,D,Q)

- p/P: autoregressive process order
- d/D: differencing order (to become stationary)
- q/Q: moving average process order

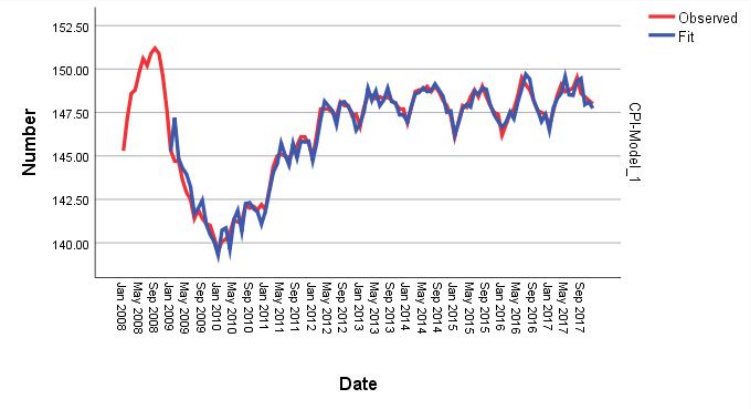


$$Y_t = f_1 Y_{t-1} + f_2 Y_{t-2} + \cdots + f_p Y_{t-p} + \epsilon_t - b_1 \epsilon_{t-1} - b_2 \epsilon_{t-2} - \cdots - b_q \epsilon_{t-q}$$

Models: Arima(1,0,3)(0,1,1)

- all the parameters, non-seasonal and seasonal are significant;
- fit values follow the observed ones closely;

ARIMA (1,0,3)(0,1,1) Model Parameters				Estimate	SE	t	Sig.
Consumer Price Index	No Transformation	AR	Lag 1	0.964	0.037	25.869	0
		MA	Lag 1	-0.591	0.093	-6.322	0
			Lag 2	-0.248	0.112	-2.216	0.029
			Lag 3	-0.272	0.103	-2.635	0.01
		Seasonal Difference		1			
		MA, Seasonal	Lag 1	0.89	0.224	3.973	0



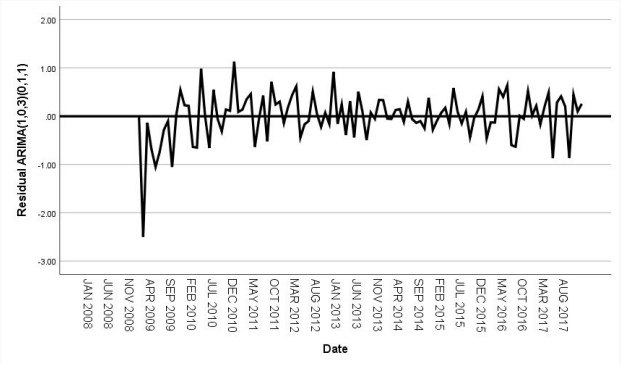
Models: Arima(1,0,3)(0,1,1)

fit measures

- Stationary R-Squared: substantial explanation;
- Low error values;
- Normalized BIC > Winters => not as good
- Ljung-Box statistic on significance limit;

residuals sequence plot

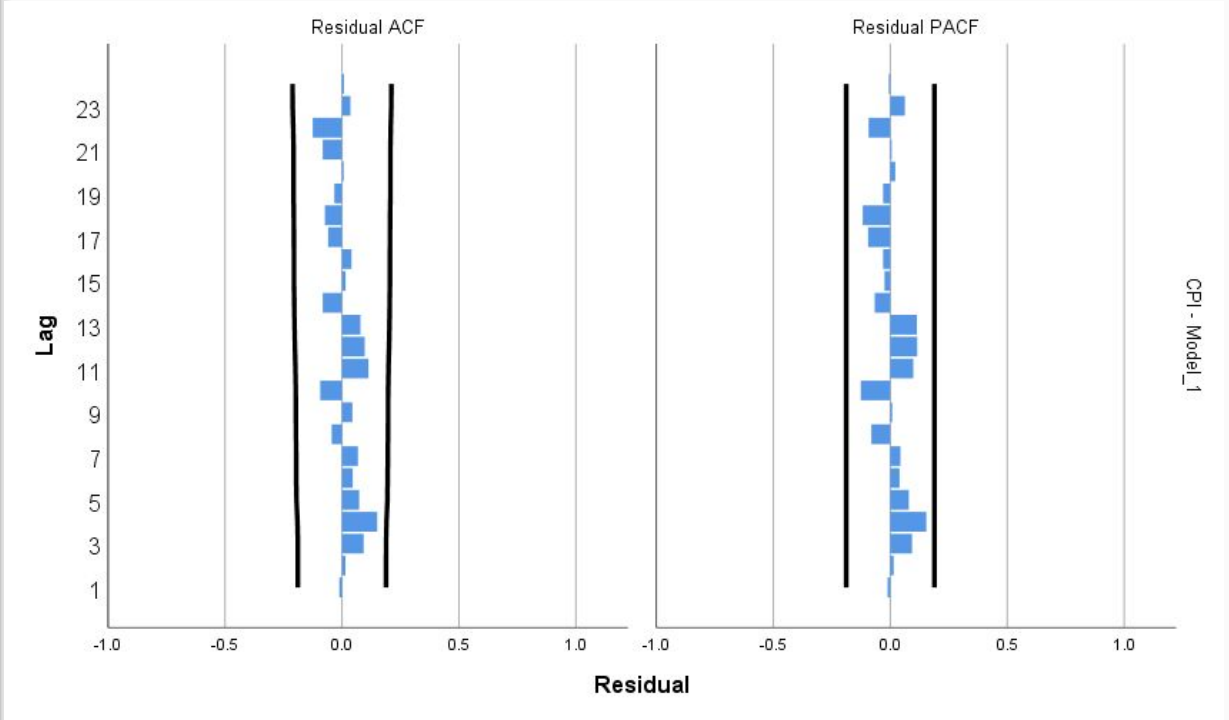
Number of Predictors		0
Model Fit statistics	Stationary R-squared	0.975
	R-squared	0.969
	RMSE	0.492
	MAPE	0.237
	MAE	0.345
	MaxAPE	1.728
	MaxAE	2.5
	Normalized BIC	-1.202
Ljung-Box Q(18)	Statistics	12.342
	DF	13
	Sig.	0.5
Number of Outliers		0



Models: Arima(1,0,3)(0,1,1)

Residuals autocorrelation

- No autocorrelation in any lag



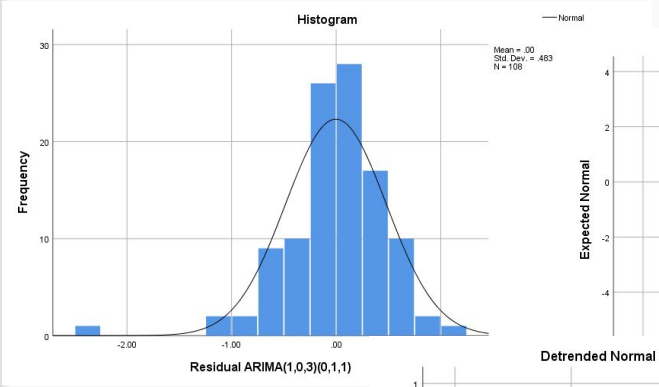
Models: Arima(1,0,3)(0,1,1)

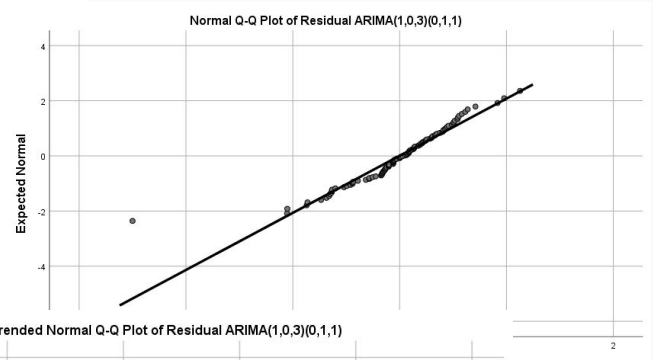
Residuals normality assumption

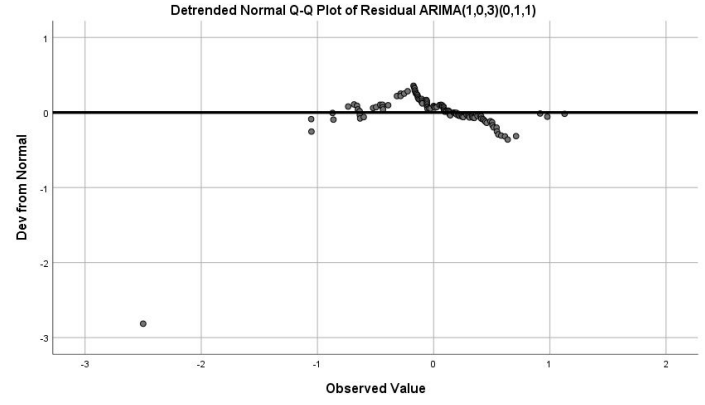
- Kolmogorov: no
- Shapiro: no
- Histogram: maybe yes
- Normal Q-Q plot: maybe yes
- Detrended Normal Q-Q plot: maybe no

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual ARIMA(1,0,3)(0,1,1)	0.13	108	0	0.921	108	0

[a] Lilliefors Significance Correction







Models: Selection

	Winters Additive	ARIMA(1,0,3)(0,1,1)
model parameters	Delta parameter, season related, is nonsignificant	all parameters significant
observed vs Fit values	good	good
residuals	less accurate in the beginning with not that mush randomness, variance increases in the last phase	less accurate in the beginning with not that mush randomness, variance increases in the last phase
fit and measure statistics	Similar errors (RMSE,MAE,etc)	Similar errors (RMSE,MAE,etc)
	Better BIC (lower)	
	R-square stationary=0.659	better R-square stationary(for non stationary data) =0.975
	Ljung-Box non significant	Ljung-Box on the limit of significance
ACF and PACF	some autocorrelation in the 1st phase	no autocorrelation
normality of residuals	assuming normal	assuming normal

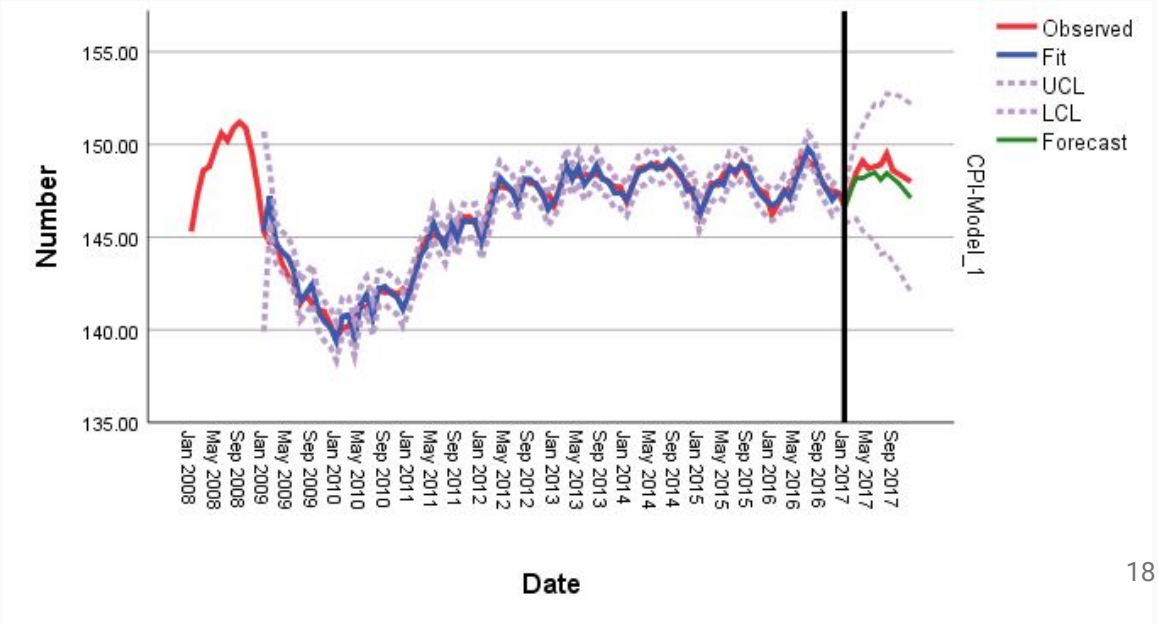
Estimation and Forecasting

splitted data (training dataset)

Estimation and Forecasting

- Estimation: Jan 2008 to Dec 2016;
- Training: Jan 2017 to Dec 2017;
- Seasonal MA parameter not significant;
- fit values follow the observed ones closely;

				Estimate	SE	t	Sig.
Consumer Price Index	No Transformation	AR	Lag 1	0.964	0.043	22.158	0
		MA	Lag 1	-0.679	0.096	-7.087	0
			Lag 2	-0.338	0.119	-2.846	0.005
			Lag 3	-0.338	0.106	-3.197	0.002
		Seasonal Difference		1			
		MA, Seasonal	Lag 1	0.948	0.612	1.548	0.125



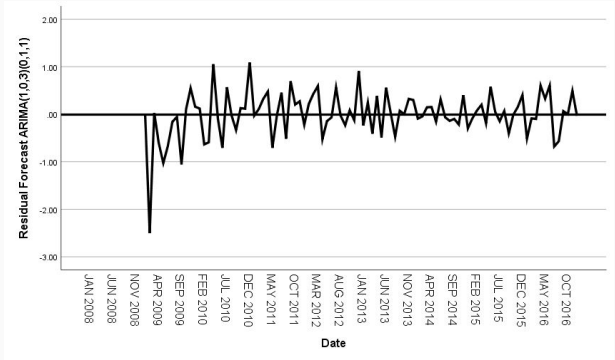
Estimation and Forecasting

fit measures

- generally slightly better;
- Low error values;
- Normalized BIC > Selection phase => not as good
- Ljung-Box statistic not significant;

residuals sequence plot

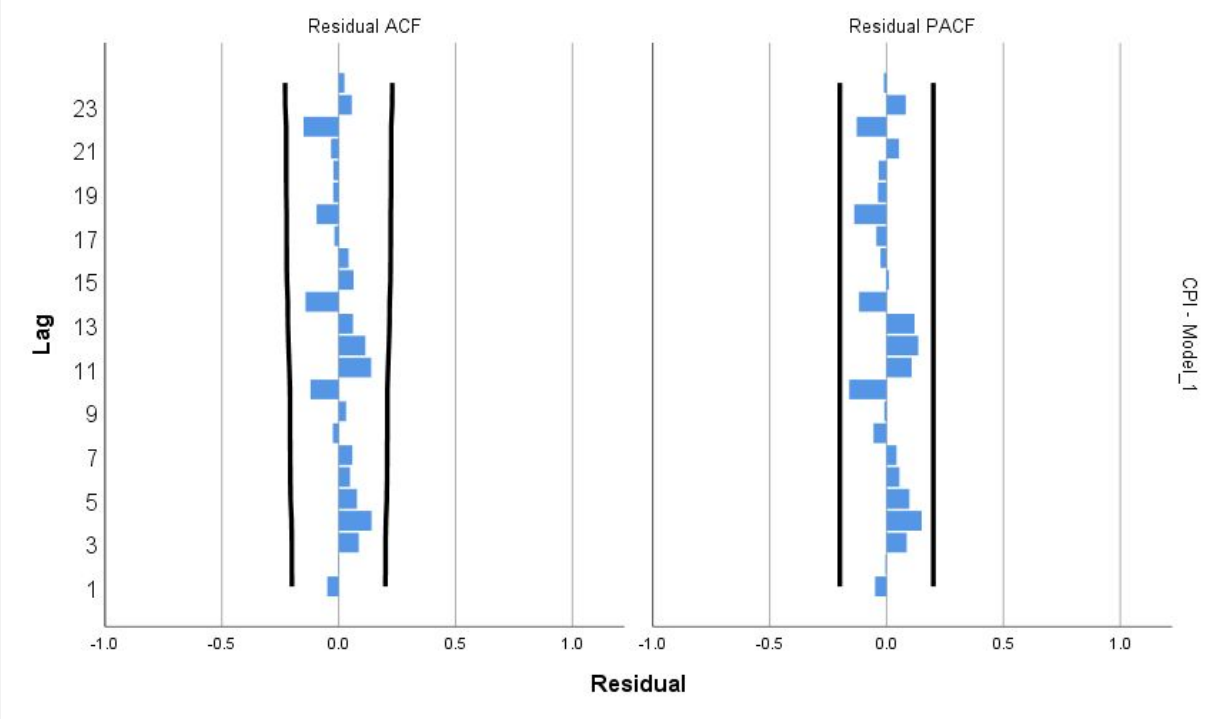
Number of Predictors		0	Selection Phase
Model Fit statistics	Stationary R-squared	0.977	0.975
	R-squared	0.97	0.969
	RMSE	0.498	0.492
	MAPE	0.232	0.237
	MAE	0.338	0.345
	MaxAPE	1.728	1.728
	MaxAE	2.5	2.5
	Normalized BIC	-1.156	-1.202
Ljung-Box Q(18)	Statistics	14.144	12.342
	DF	13	13
	Sig.	0.364	0.5
Number of Outliers		0	0



Estimation and Forecasting

Residuals autocorrelation

- No autocorrelation in any lag



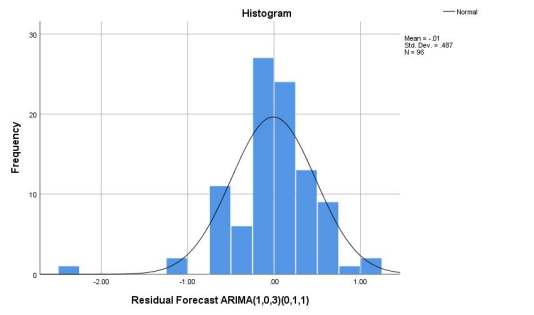
Estimation and Forecasting

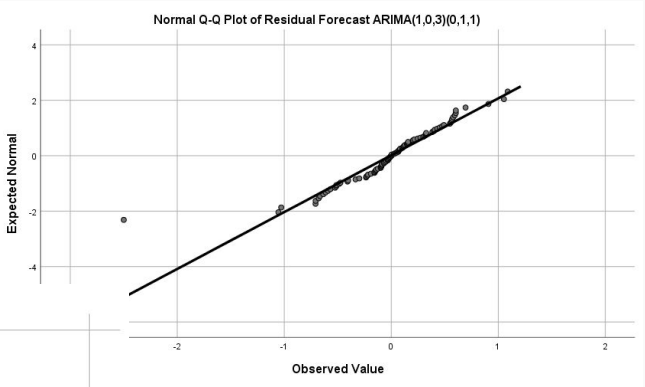
Residuals normality assumption

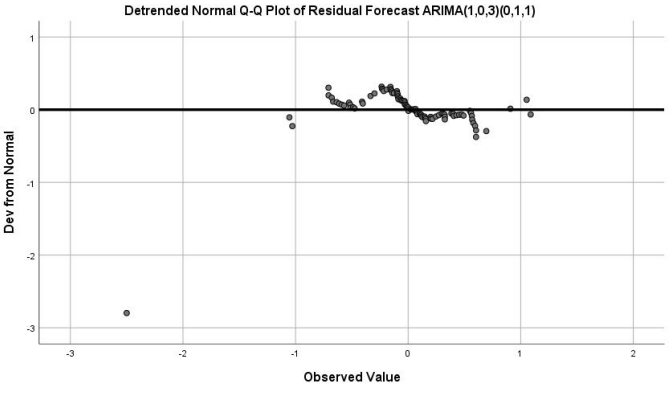
- Kolmogorov: no
- Shapiro: no
- Histogram: maybe yes
- Normal Q-Q plot: maybe yes
- Detrended Normal Q-Q plot: maybe yes

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual Forecast ARIMA(1,0,3)(0,1,1)	0.12	96	0.002	0.913	96	0

[a] Lilliefors Significance Correction



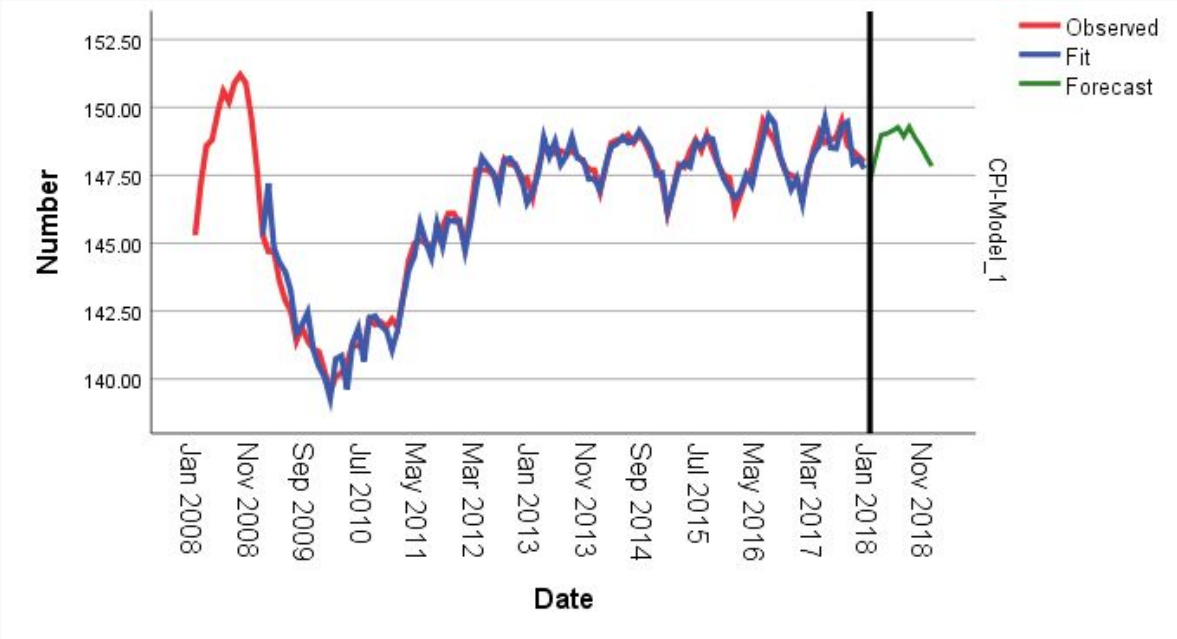




True Forecast

True Forecast

- Estimation: Jan 2008 to Dec 2017;
- Forecasting: Jan 2018 to Dec 2018;
- Model parameters are the same as in Selection phase;
- replicates the seasonality and slight upbeat trend seen in the latest stages of the observed data;



Conclusions

- models struggle with 1st phase;
- Cyclical component hard to predict;
- models adapt reasonably well to inflections;
- Don’t predict too much into the future;
- Incorporate errors asap;

thank you
obrigado