Time Series Forecasting.
2. ARIMA

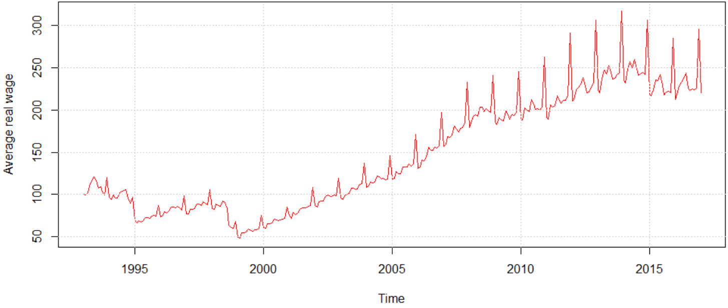
Alexey Romanenko alexromsput@gmail.com

Acknowledgement: Evgeny Riabenko for materials supplied

Содержание

Time Series Forecasting

Time Series: $y_1, \dots, y_T, \dots, \quad y_t \in \mathbb{R},$ — characteristic values measured through constant time intervals

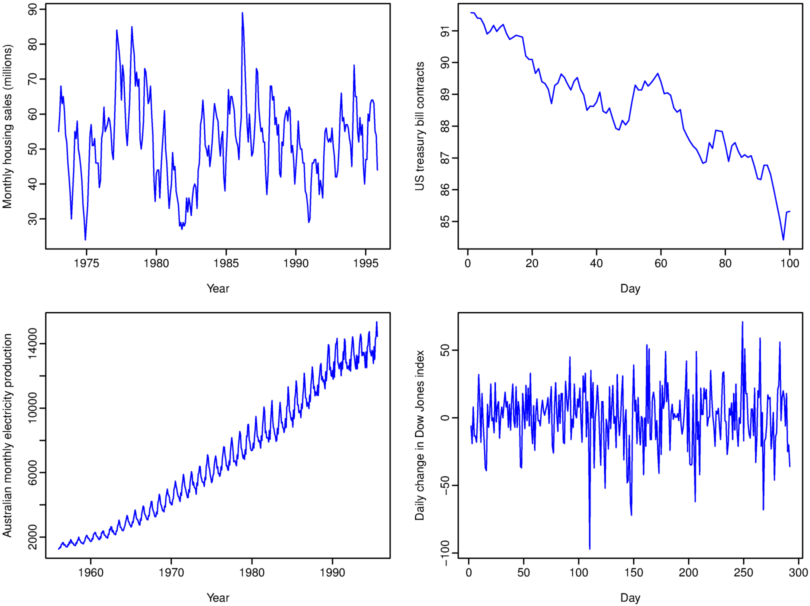


The forecasting task is to find function f_T:

$$y_{T+d} \approx f_T\left(y_T, \dots, y_1, d\right) \equiv \hat{y}_{T+d|T},$$

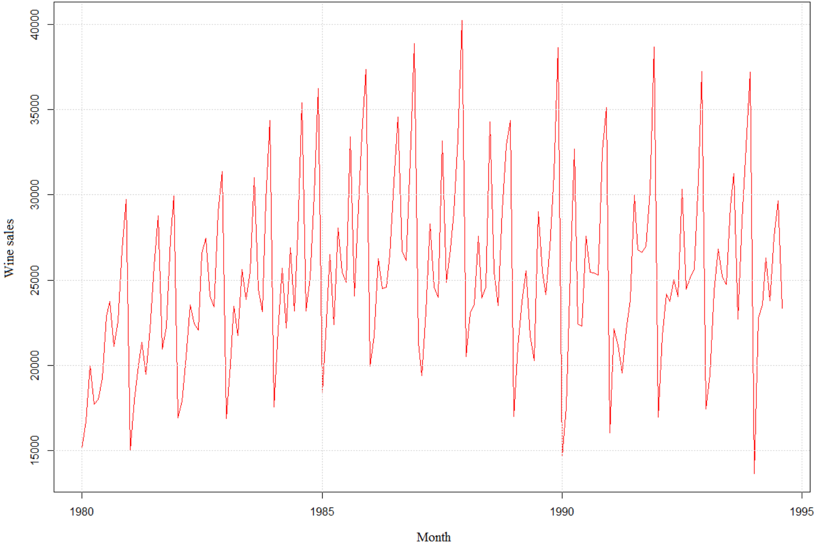
where $d \in \{1, \dots, D\}$ — is forecast delay, D — is forecasting horizon.

Examples of times series



Examples of times series

Wine Sales in Australia:



Autocorrelation (ACF)

Observations of time series are autocorrelated.

Autocorrelation:

$$r_\tau = r_{y_t y_{t+\tau}} = \frac{\sum\limits_{t=1}^{T-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2}, \quad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t.$$

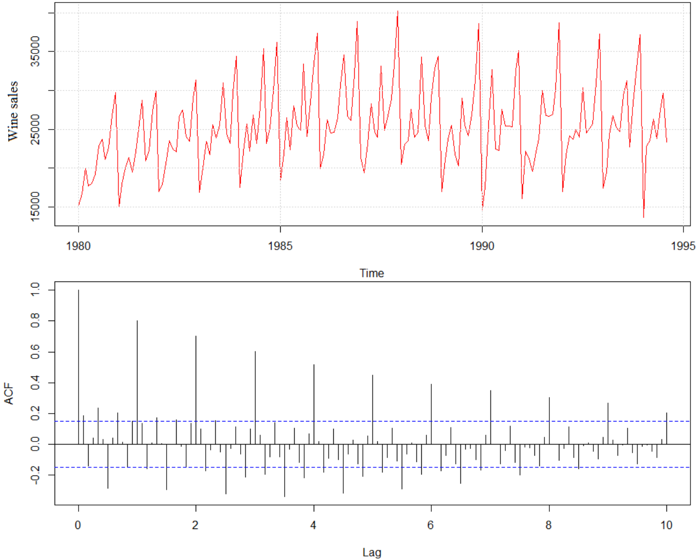
$r_\tau \in [-1, 1]$, τ — autocorrelation lag.

Significance test that the value of autocorrelation is different from zero:

$$\begin{aligned}
\text{time series:}\quad & Y^T = Y_1, \dots, Y_T; \\
\text{null hypotheses:}\quad & H_0\colon r_\tau = 0; \\
\text{alternative:}\quad & H_1\colon r_\tau \neq 0; \\
\text{statistic:}\quad & F\left(Y^T\right) = \tfrac{r_\tau\sqrt{T-\tau-2}}{\sqrt{1-r_\tau^2}}; \\
\text{zero distribution:}\quad & t - distibution\,(T - \tau - 2).
\end{aligned}$$

Autocorrelation (ACF)

Correlogram:



Time Series Components

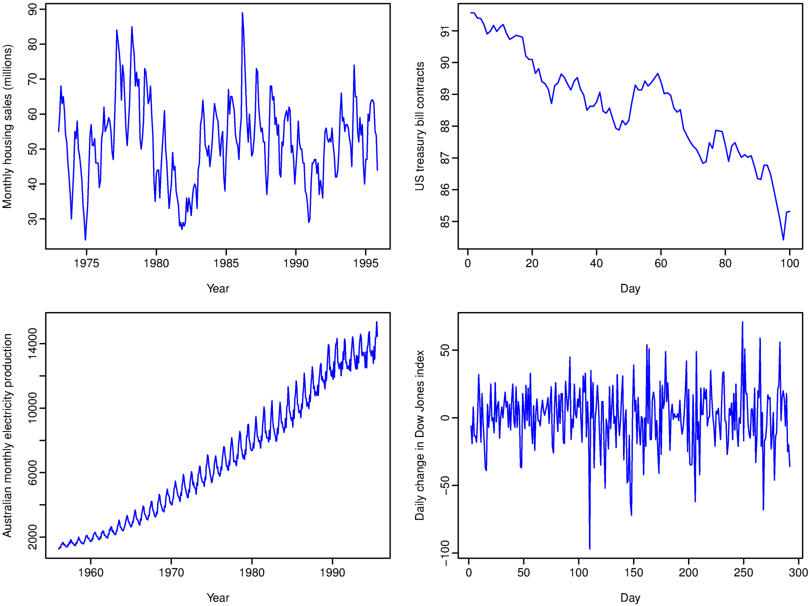
Trend — is a smooth long-term change of time series level.

Seasonality — is a cyclical change of time series level at a constant period.

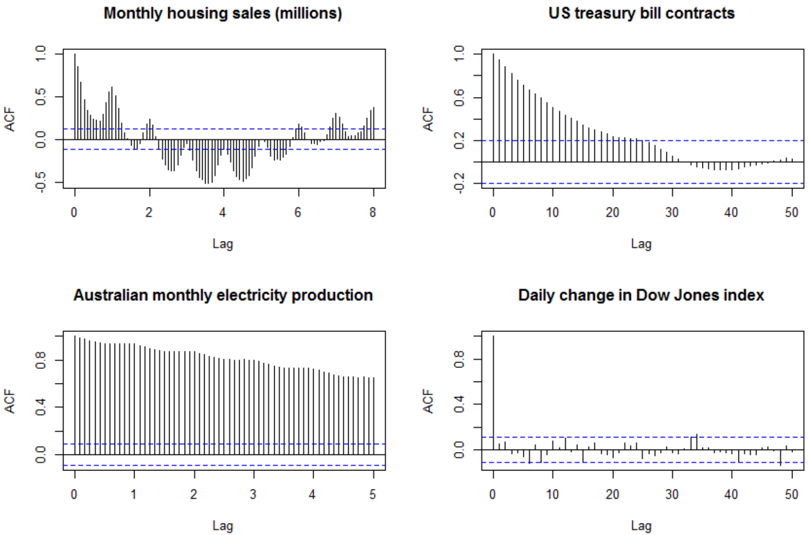
Cycle — is a change of time series level with a changing period (life cycle of a good, economic waves, solar activity periods).

Error — is an unpredictable random component of time series.

Time Series Components



Time Series Components



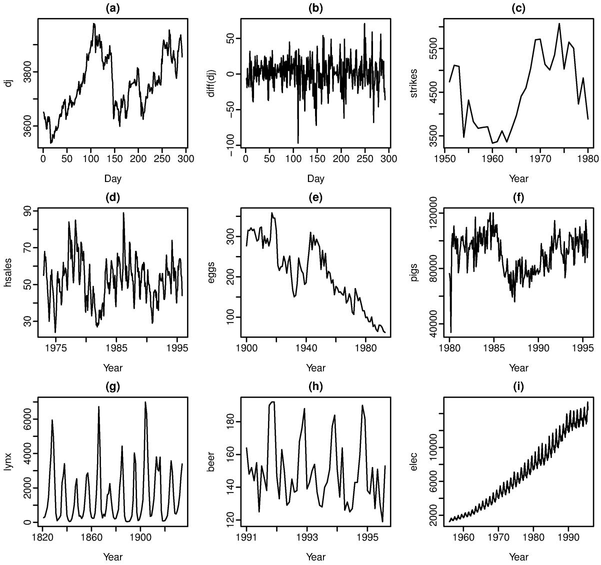
Stationarity

Time series $y_1, \dots, y_T$ is **stationary** if $\forall s$ distribution $y_t, \dots, y_{t+s}$ does not depend on t, i.e. its properties do not depend on time.

Time series with trend or seasonality are not stationary.

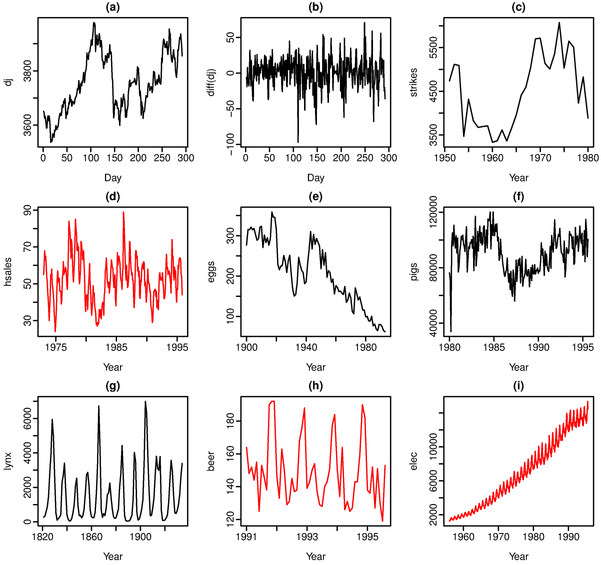
Time series with a-periodical cycles are stationary since it is impossible to predict where the maximums and minimums will be located.

Stationarity



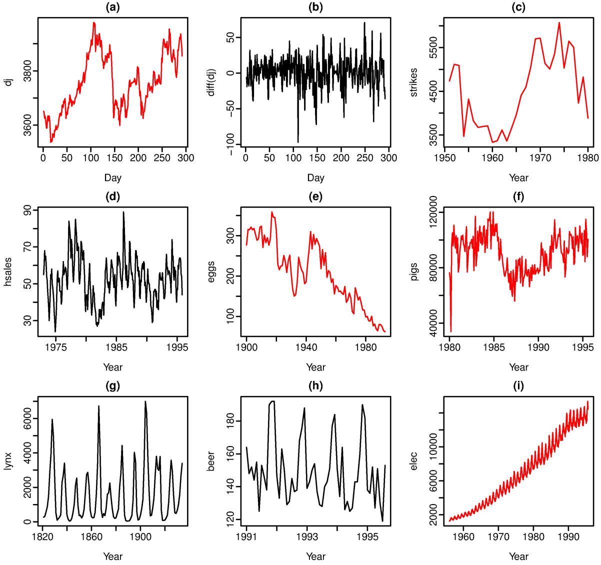
Stationarity

Non-stationary due to seasonality:



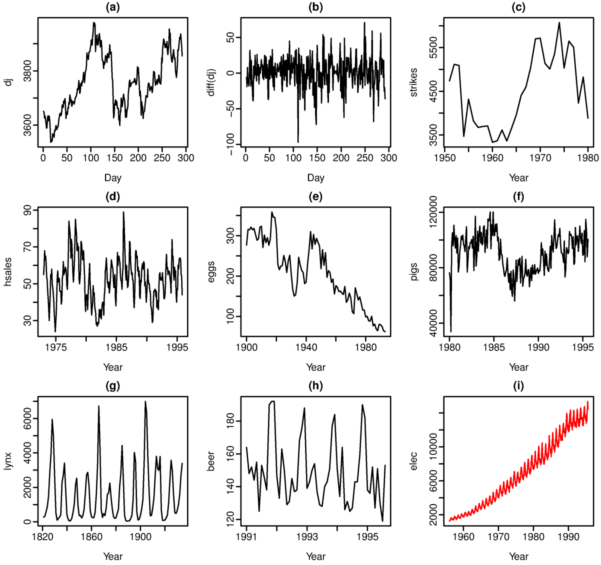
Stationarity

Non-stationary due to trend:



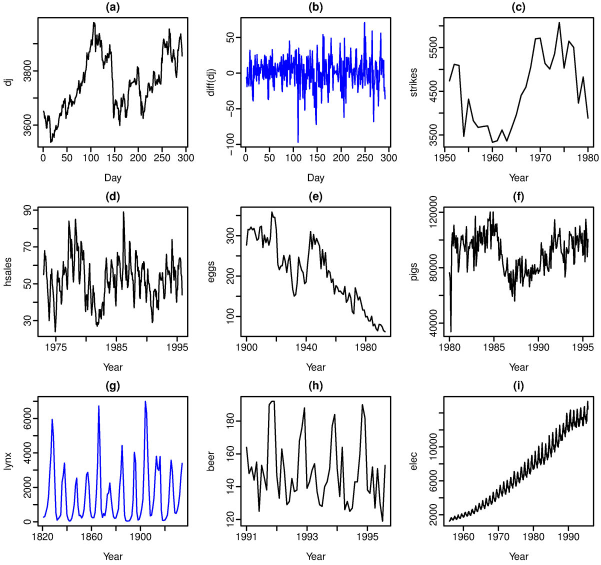
Stationarity

Non-stationary due to changing dispersion:



Stationarity

Stationary:



KPSS (Kwiatkowski-Philips-Schmidt-Shin)

time series of forecast errors: $\varepsilon^T = \varepsilon_1, \ldots, \varepsilon_T$;

null hypotheses: $H_0:$ time series ε^T is stationarity;

alternative: $H_1:$ time series ε^T is described by model of the kind $\varepsilon_t = \alpha\varepsilon_{t-1}$;

statistic: $KPSS\left(\varepsilon^T\right) = \frac{1}{T^2}\sum\limits_{i=1}^{T}\left(\sum\limits_{t=1}^{i}\varepsilon_t\right)^2 \Big/ \lambda^2$;

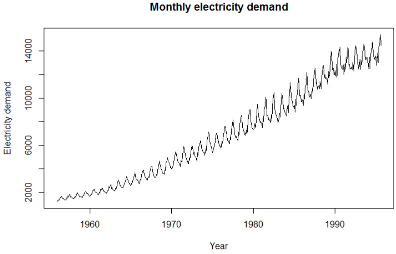
null distribution: as in table.

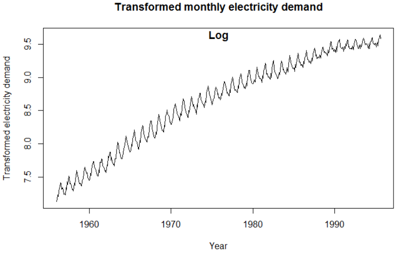
Other tests to check for stationarity: Dickey-Fuller, Phillips-Perron and their many modifications (see Patterson K. *Unit root tests in time series, volume 1: key concepts and problems*. — Palgrave Macmillan, 2011).

Dispersion Stabilization

It is possible to use stabilizing transformation for time series with a monotonously changing dispersion.

Logarithmation is often used:



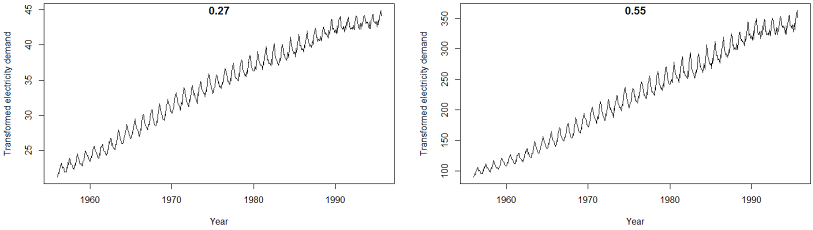


Box-Cox Transformation

Parametric family of transformations that stabilize dispersion:

$$y'_t = \begin{cases} \ln y_t, & \lambda = 0, \\ \left(y_t^\lambda - 1\right)/\lambda, & \lambda \neq 0. \end{cases}$$

Such parameter λ is chosen that dispersion is minimized and model plausibility maximized.



Box-Cox Transformation

After the forecast for the transformed time series is built it should be transformed into forecast of the initial time series:

$$\hat{y}_t = \begin{cases} \exp\left(\hat{y}'_t\right), & \lambda = 0, \\ \left(\lambda \hat{y}'_t + 1\right)^{1/\lambda}, & \lambda \neq 0. \end{cases}$$

- if some $y_t \leq 0$, Box-Cox transformations are impossible (we must add a constant to the time series)
- it often turns out that no transformation at all is needed
- it is possible to round the value of λ in order to simplify interpretation
- as a rule, stabilizing transformation has little influence on the forecast and strong influence on the forecast interval

Differentiation

Time series differentiation — is a shift to pairwise difference of its neighboring values:

$$y_1, \dots, y_T \longrightarrow y'_2, \dots, y'_T,$$

$$y'_t = y_t - y_{t-1}.$$

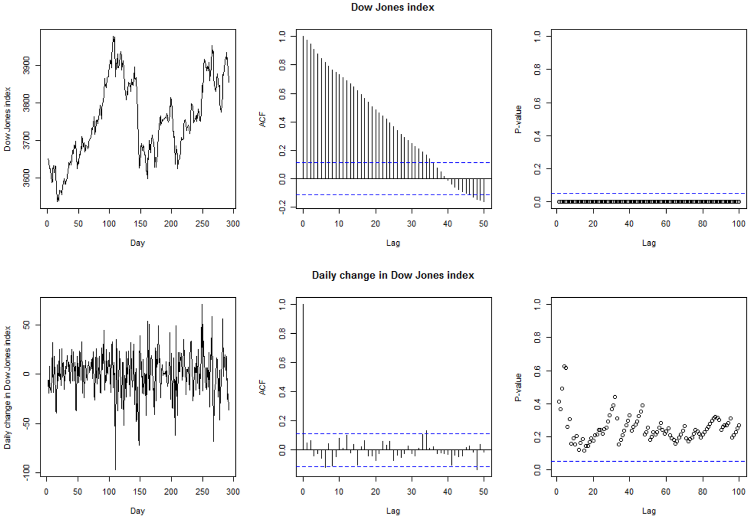
By differentiation it is possible to stabilize the average value of time series and to get rid of trend and seasonality.

Repeated differentiation may be used; for example, for second degree:

$$y_1, \dots, y_T \longrightarrow y'_2, \dots, y'_T \longrightarrow y''_3, \dots, y''_T,$$

$$y''_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_{t-2}.$$

Differentiation



KPSS criterion: for the initial time series $p < 0.01$, for the time series of first differences — $p > 0.1$.

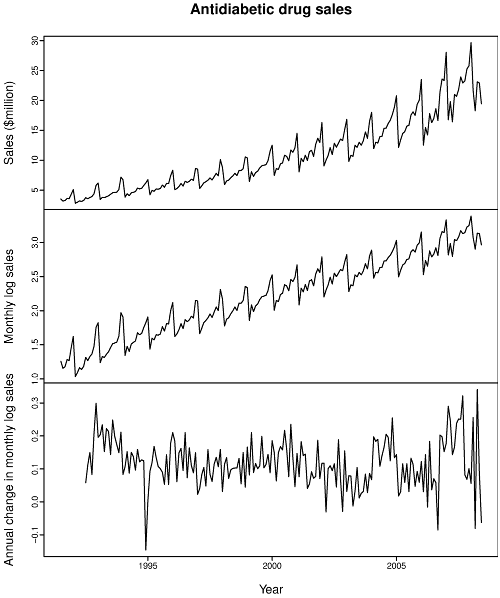
Seasonal Differentiation

Seasonal differentiation of time series — is a shift to pairwise differences of its values in neighboring seasons:

$$y_1, \dots, y_T \;\longrightarrow\; y'_{s+1}, \dots, y'_T,$$

$$y'_t = y_t - y_{t-s}.$$

Seasonal Differentiation



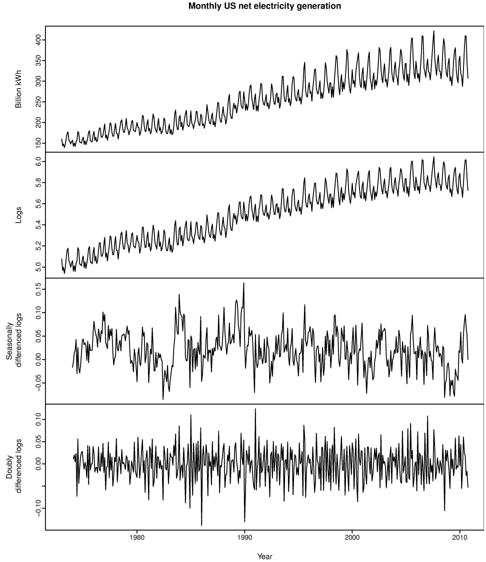
KPSS criterion: for the initial time series $p < 0.01$, for logarithmated — $p < 0.01$, after seasonal differentiation — $p > 0.1$.

Combinated Differentiation

Seasonal and simple differentiation may be applied to the same time series in any order.

If the time series has a clear seasonality profile it is recommended to start with seasonal differentiation — it may be enough to make the time series stationary.

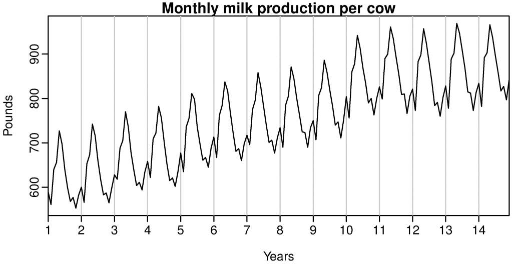
Combinated Differentiation

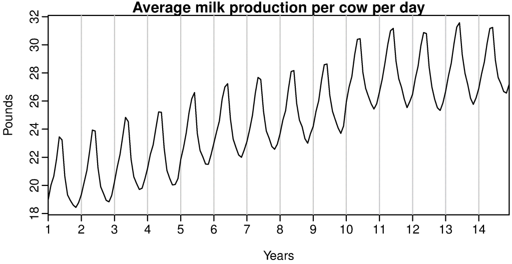


KPSS criterion: for the initial time series $p < 0.01$, for the logarithmated one — $p < 0.01$, after seasonal differentiation — $p = 0.0355$, after one more differentiation — $p > 0.1$.

Additional tricks: calendar effects

Sometimes it is possible to simplify time series structure by accounting for irregularity of observations:





Residuals

Residuals are the difference between fact and forecast:

$$\hat{\varepsilon}_t = y_t - \hat{y}_t.$$

Forecasts $\hat{y}_t$ may be built with a fixed delay:

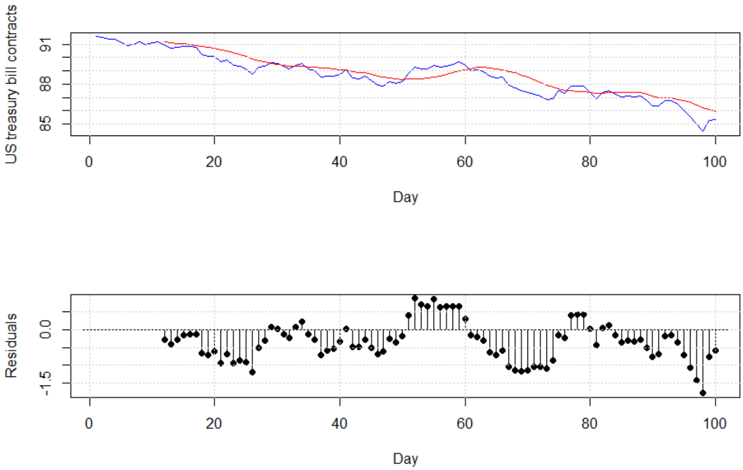
$$\hat{y}_{R+d|R}, \dots, \hat{y}_{T|T-d},$$

or with a fixed end of history at different delays:

$$\hat{y}_{T-D+1|T-D}, \dots, \hat{y}_{T|T-D}.$$

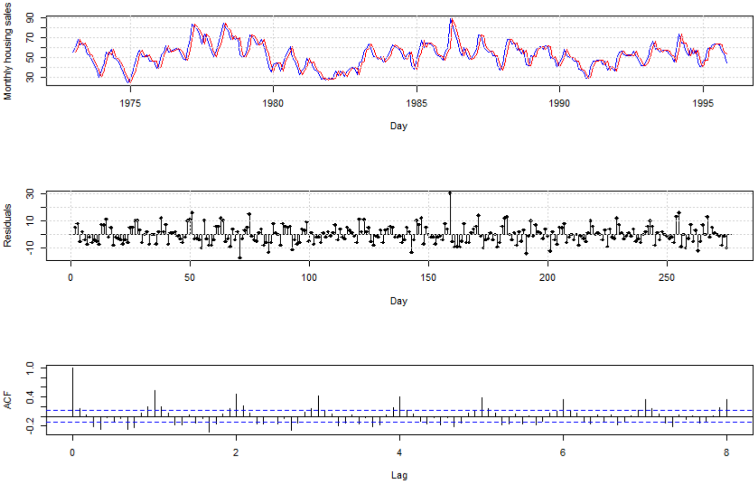
Necessary Characteristics of Forecast Residuals

- Unbiasedness means equality of the average value to zero:



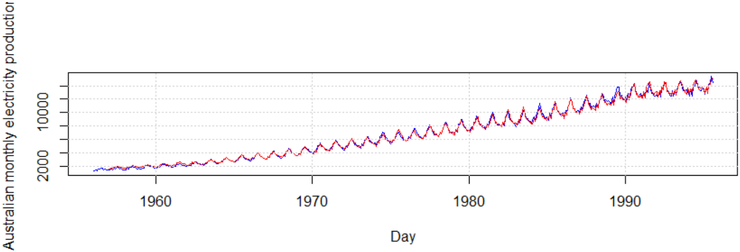
Necessary Characteristics of Forecast Residuals

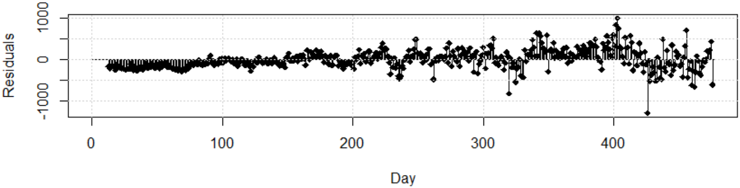
- No autocorrelation means absence of the unaccounted dependency on previous observations:



Necessary Characteristics of Forecast Residuals

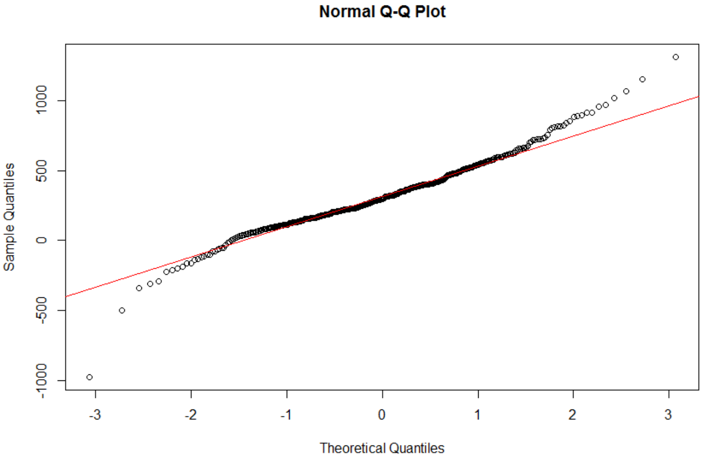
- Stationarity means absence of dependency on time:





Desirable Characteristics of Forecast Residuals

- Normality:



Check of Residual Characteristics

- Unbiasedness — Student or Wilcoxon.
- Stationarity — visual analysis, KPSS.
- No autocorrelation — correlogram, Ljung-Box Q-test.
- Normality — q-q plot, Shapiro-Wilk test.

Ljung-Box Q-test

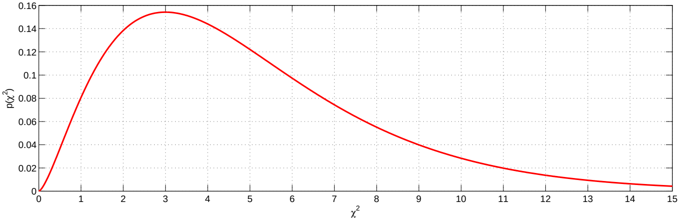
time series of forecast errors: $\varepsilon^T = \varepsilon_1, \dots, \varepsilon_T$;

null hypotheses: $H_0\colon r_1 = \dots = r_L = 0$;

alternative: $H_1\colon H_0$ is not true;

statistic: $Q\left(\varepsilon^T\right) = T\left(T+2\right) \sum\limits_{\tau=1}^{L} \frac{r_\tau^2}{T-\tau}$;

zero distribution: χ^2_{L-K}, K — the number of parameters



Autoregression

$$AR(p): \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

where y_t — is a stationary time series with zero average, $\phi_1, \ldots, \phi_p$ — are constants ($\phi_p \neq 0$), ε_t — is gaussian white noise with zero average and constant dispersion σ_ε^2.

If the average equals μ the model looks like

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

where $\alpha = \mu\left(1 - \phi_1 - \cdots - \phi_p\right)$.

Another way to note:

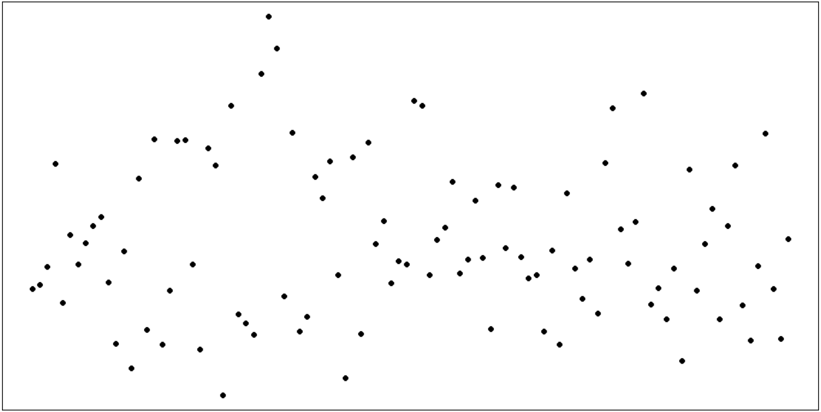
$$\phi\left(B\right)\left(y_t - \mu\right) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)\left(y_t - \mu\right) = \varepsilon_t,$$

where B — is difference operator $\left(B y_t = y_{t-1}\right)$.

Linear combination p of consecutive time series members results in white noise.

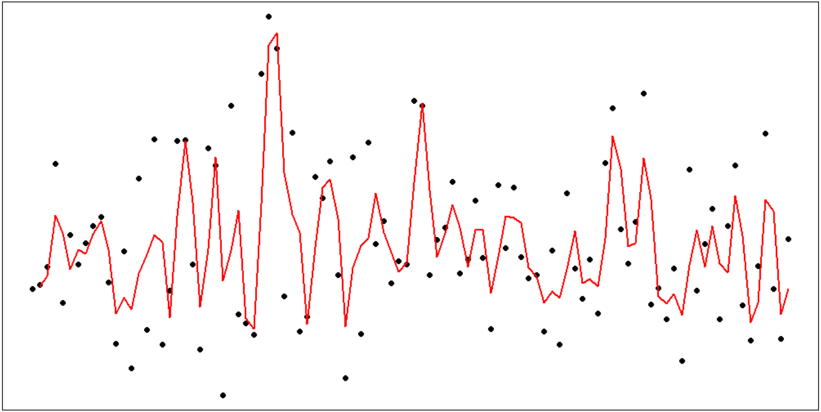
Moving Average

Let us have an independent equally distributed in time noise ε_t:



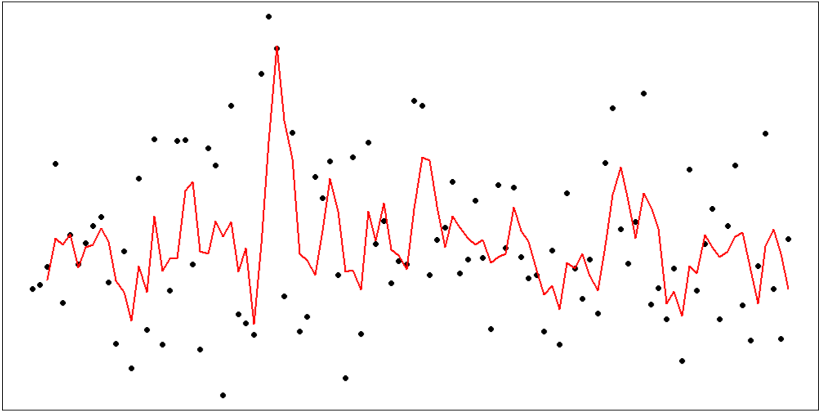
Moving Average

The average of two neighboring points:



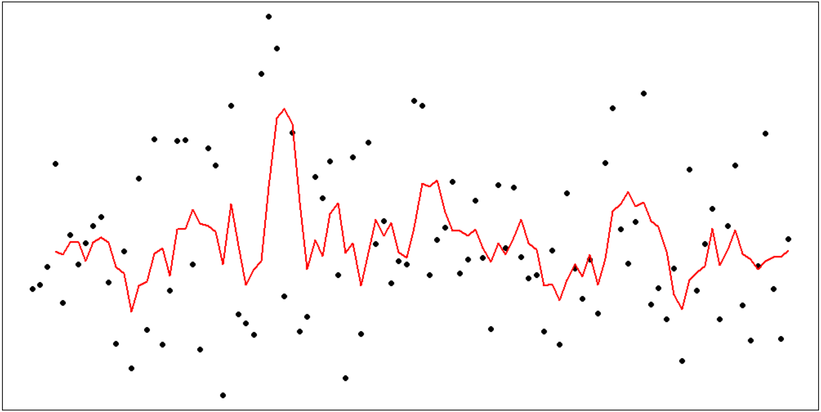
Moving Average

The average of three neighboring points:



Moving Average

The average of four neighboring points:



Moving Average

$$MA(q): \quad y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

where y_t — is a stationary time series with zero average, $\theta_1, \ldots, \theta_q$ — are constants ($\theta_q \neq 0$), ε_t — is gaussian white noise with zero average and constant dispersion σ_ε^2.

If the average equals μ the model looks like

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}.$$

Another way to note:

$$y_t - \mu = \theta(B) \varepsilon_t = \left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right) \varepsilon_t,$$

where B — is difference operator.

Linear combination q of consecutive components of white noise ε_t gives an element of the time series.

ARMA (Autoregressive moving average)

$$ARMA(p,q): \quad y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

where y_t — is a stationary time series with zero average, $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ — are constants ($\phi_p \neq 0$, $\theta_q \neq 0$), ε_t — is gaussian white noise with zero average and constant dispersion σ_ε^2.

If the average equals μ the model looks like

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

where $\alpha = \mu \left(1 - \phi_1 - \dots - \phi_p\right)$.

Another way to note:

$$\phi(B)(y_t - \mu) = \theta(B)\varepsilon_t.$$

Argumentation of ARMA model

Theorem (Wold, 1938)

Every covariance-stationary (WSS) time series y_t can be written as the sum of two time series, one deterministic and one stochastic, formaly:

$$y_t = \theta\left(B\right)\varepsilon_t + \eta_t$$

where η_t is a deterministic time series, such as one represented by a sine wave.

Definition

Covariance-stationary (or weak-sense stationarity, wide-sense stationarity, WSS) random processes only require that 1st moment (i.e. the mean) and autocovariance do not vary with respect to time:

$$\mathrm{E}[y_t] = m_y(t) = m_t(t+\tau) \text{ for all } \tau \in \mathbb{R}$$

and

$$\mathrm{E}[(y(t_1) - m_y(t_1))(y(t_2) - m_y(t_2))] = C_y(t_1, t_2) = C_y(t_1 + (-t_2), t_2 + (-t_2)) \\ = C_y(t_1 - t_2, 0).$$

ARIMA (Autoregressive integrated moving average)

The time series is described by $ARIMA(p, d, q)$, if the time series of its differences

$$\nabla^d y_t = (1 - B)^d y_t$$

is described by $ARMA(p, q)$:

$$\phi(B) \nabla^d y_t = \theta(B) \varepsilon_t.$$

Seasonal ARMA/ARIMA

$ARMA(p,q) \times (P,Q)_s$:

$$\Phi_P\left(B^s\right)\phi\left(B\right)\left(y_t - \mu\right) = \Theta_Q\left(B^s\right)\theta\left(B\right)\varepsilon_t,$$

where

$$\Phi_P\left(B^s\right) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps},$$
$$\Theta_Q\left(B^s\right) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}.$$

SARIMA:

$$\Phi_P\left(B^s\right)\phi\left(B\right)\nabla_s^D\nabla^d(y_t - \mu) = \Theta_Q\left(B^s\right)\theta\left(B\right)\varepsilon_t.$$

Equivalence to some ES models

ARIMA contains all ES models with linear trend and additive seasonality

- ARIMA(p=0,d=1,q=1) is equivalent to Simple ES with

$$(1-L)y_t = (1-\phi_1 L)\varepsilon_t$$

$$\phi_1 = 1-\alpha$$

Proof:

$$y_t - y_{t-1} = \varepsilon_t - \phi_1\varepsilon_{t-1} = y_t - \hat{y}_t - (1-\alpha)\cdot(y_{t-1}-\hat{y}_{t-1})$$

$$\hat{y}_t = y_{t-1} - y_{t-1} + \alpha y_{t-1} + (1-\alpha)\cdot\hat{y}_{t-1} = \hat{y}_{t-1} + \alpha\cdot e_{t-1}$$

- ARIMA(p=0,d=2, q=2) is equivalent to Holt (linear trend) with:

$$(1-L)^2 Y_t = (1-\phi_1 L - \phi_2 L^2)\varepsilon_t$$

$$\phi_1 = 2-\alpha-\alpha\beta,\ \ \phi_2 = \alpha - 1$$

Equivalence to some ES models

- damped-trend linear exponential smoothing is the ARIMA(1,1,2) model

$$(1-\phi B)(1-B)Y_t = (1-\theta_1 B - \theta_2 B^2)\epsilon_t$$

$$\theta_1 = 1+\phi-\alpha-\alpha\beta\phi,\ \theta_2 = (\alpha-1)\phi$$

  ϕ — coefficient of damped trend;

- seasonal exponential smoothing is the ARIMA(0,1,p+1)$(0,1,0)_p$ model

$$(1-B)(1-B^p)Y_t = (1-\theta_1 B - \theta_2 B^p - \theta_3 B^{p+1})\epsilon_t$$

$$\theta_1 = 1-\alpha$$

$$\theta_2 = 1-\gamma(1-\alpha)$$

$$\theta_3 = (1-\alpha)(\gamma-1)$$

Equivalence to some ES models

- ARIMA$(0,1,p+1)(0,1,0)_p$ is equivalent to additive seasonality ES model with:

$$(1-B)(1-B^p)Y_t = [1-\sum_{i=1}^{p+1}\theta_i B^i]\epsilon_t$$

$$\theta_j = \begin{cases} 1-\alpha-\alpha\beta & j=1 \\ -\alpha\beta & 2 \leq j \leq p-1 \\ 1-\alpha\beta-\gamma(1-\alpha) & j=p \\ (1-\alpha)(\gamma-1) & j=p+1 \end{cases}$$

d, D

- The degrees of differentiation are chosen so that the time series becomes stationary
- Once more: if the time series is seasonal, seasonal differentiation should be applied first
- The fewer times we differentiate the less will be dispersion of the final forecast

q, Q, p, P

- Hyperparameters cannot be chosen using ML: L is always taken into account with their gtime seriesth
- Informational criteria may be used to compare models of different q, Q, p, P
- Initial approximations may be chosen using autocorrelations

Partial Autocorrelation Function (PACF)

Partial autocorrelation of a stationary time series y_t — is autocorrelation of autoregression residuals of the previous order:

$$\phi_{hh} = \begin{cases} r\left(y_{t+1}, y_t\right), & h = 1, \\ r\left(y_{t+h} - \hat{y}_{t+h}, y_t - \hat{y}_t\right), & h \geq 2, \end{cases}$$

where $\hat{y}_{t+h}$ и $\hat{y}_t$ — are predictions of regressions y_{t+h} and y_t by $y_{t+1}, y_{t+2}, \dots, y_{t+h-1}$:

$$\hat{y}_t = \beta_1 y_{t+1} + \beta_2 y_{t+2} + \dots + \beta_{h-1} y_{t+h-1},$$
$$\hat{y}_{t+h} = \beta_1 y_{t+h-1} + \beta_2 y_{t+h-2} + \dots + \beta_{h-1} y_{t+1}.$$

Behavior of ACF and PACF for different AR and MA components

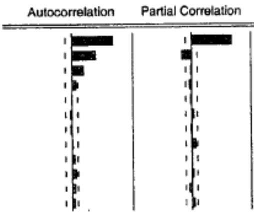
Lag	AC	PAC	Q-Stat	Prob
1	0.539	0.539	116.40	0.000
2	0.319	0.041	157.37	0.000
3	0.190	0.004	171.91	0.000
4	0.092	-0.029	175.35	0.000
5	0.014	-0.044	175.43	0.000
6	0.012	0.033	175.50	0.000
7	-0.013	-0.026	175.56	0.000
8	0.025	0.059	175.81	0.000
9	0.042	0.018	176.52	0.000
10	0.069	0.042	178.47	0.000
11	0.027	-0.051	178.78	0.000
12	0.036	0.028	179.32	0.000

Рис. 11.9. AR(1). $Y_t = 0.5Y_{t-1} + \varepsilon_t$. Корень $\mu = 2$

Lag	AC	PAC	Q-Stat	Prob
1	-0.500	-0.500	100.19	0.000
2	0.281	0.041	131.88	0.000
3	-0.125	0.041	138.15	0.000
4	0.104	0.063	142.49	0.000
5	-0.106	-0.049	147.01	0.000
6	0.090	0.009	150.33	0.000
7	-0.096	-0.043	154.11	0.000
8	0.080	0.011	156.70	0.000
9	-0.068	-0.010	158.57	0.000
10	0.103	0.074	162.91	0.000
11	-0.081	0.009	165.60	0.000
12	0.063	-0.002	167.23	0.000

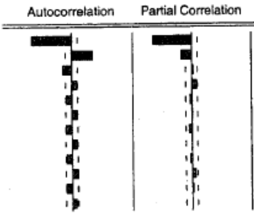
Рис. 11.10. AR(1). $Y_t = -0.5Y_{t-1} + \varepsilon_t$. Корень $\mu = -2$

Behavior of ACF and PACF for different AR and MA components



	AC	PAC	Q-Stat	Prob
1	0.700	0.700	196.54	0.000
2	0.403	-0.171	261.80	0.000
3	0.203	-0.016	278.34	0.000
4	0.072	-0.037	280.46	0.000
5	-0.006	-0.023	280.47	0.000
6	-0.021	0.035	280.64	0.000
7	-0.022	-0.016	280.84	0.000
8	0.017	0.071	280.95	0.000
9	0.049	0.008	281.93	0.000
10	0.071	0.025	283.99	0.000
11	0.051	-0.043	285.05	0.000
12	0.048	0.045	286.00	0.000

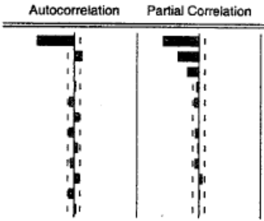
Рис. 11.11. AR(2). $Y_t = 0.8Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$.
Корни $\mu_1 = 2 + i, \mu_2 = 2 - i$



	AC	PAC	Q-Stat	Prob
1	-0.670	-0.670	179.75	0.000
2	0.353	-0.173	229.82	0.000
3	-0.147	0.028	238.48	0.000
4	0.087	0.083	241.55	0.000
5	-0.088	-0.032	244.67	0.000
6	0.090	0.009	247.99	0.000
7	-0.097	-0.042	251.78	0.000
8	0.088	0.007	254.96	0.000
9	-0.086	-0.030	257.98	0.000
10	0.106	0.062	262.57	0.000
11	-0.092	0.029	266.04	0.000
12	0.071	0.010	268.12	0.000

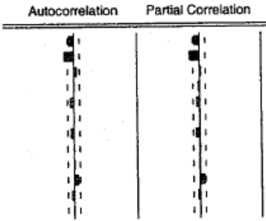
Рис. 11.12. AR(2). $Y_t = -0.8Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$.
Корни $\mu_1 = -2 + i, \mu_2 = -2 - i$

Behavior of ACF and PACF for different AR and MA components



	AC	PAC	Q-Stat	Prob
1	-0.593	-0.593	140.88	0.000
2	0.124	-0.351	147.01	0.000
3	0.004	-0.185	147.02	0.000
4	0.026	-0.034	147.29	0.000
5	-0.069	-0.068	149.21	0.000
6	0.076	0.003	151.55	0.000
7	-0.074	-0.050	153.79	0.000
8	0.056	-0.014	155.06	0.000
9	-0.055	-0.058	156.32	0.000
10	0.088	0.050	159.47	0.000
11	-0.077	0.024	161.89	0.000
12	0.035	0.010	162.40	0.000

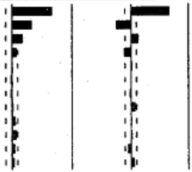
Рис. 11.16. MA(2). $Y_t = \varepsilon_t - 0.9\varepsilon_{t-1} + 0.2\varepsilon_{t-2}$.



	AC	PAC	Q-Stat	Prob
1	-0.074	-0.074	2.1884	0.139
2	-0.151	-0.158	11.407	0.003
3	0.048	0.024	12.338	0.006
4	0.008	-0.010	12.365	0.015
5	-0.052	-0.042	13.451	0.020
6	0.016	0.008	13.559	0.035
7	-0.043	-0.057	14.316	0.046
8	0.009	0.008	14.352	0.073
9	0.015	0.000	14.438	0.108
10	0.067	0.075	16.307	0.091
11	-0.040	-0.027	16.974	0.109
12	-0.002	0.009	16.975	0.151

Рис. 11.17. MA(2). $Y_t = \varepsilon_t - 0.1\varepsilon_{t-1} - 0.2\varepsilon_{t-2}$.

Behavior of ACF and PACF for different AR and MA components



	AC	PAC	Q-Stat	Prob
1	0.687	0.687	189.53	0.000
2	0.336	-0.259	235.01	0.000
3	0.177	0.129	247.59	0.000
4	0.071	-0.107	249.63	0.000
5	0.003	0.018	249.63	0.000
6	-0.013	0.004	249.70	0.000
7	-0.015	-0.009	249.79	0.000
8	0.016	0.067	249.90	0.000
9	0.051	0.009	250.97	0.000
10	0.066	0.023	252.76	0.000
11	0.042	-0.044	253.50	0.000
12	0.039	0.057	254.13	0.000

Рис. 11.20. ARMA(1,1). $Y_t = 0.4Y_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$.
Корни $\mu_{AR} = 2, \mu_{MA} = -2$

	AC	PAC	Q-Stat	Prob
1	-0.662	-0.662	175.75	0.000
2	0.281	-0.279	207.58	0.000
3	-0.114	-0.117	212.83	0.000
4	0.082	0.021	215.55	0.000
5	-0.095	-0.047	219.20	0.000
6	0.095	0.008	222.85	0.000
7	-0.095	-0.047	226.53	0.000
8	0.082	-0.006	229.26	0.000
9	-0.079	-0.046	231.83	0.000
10	0.102	0.054	236.10	0.000
11	-0.091	0.029	239.50	0.000
12	0.063	0.014	241.12	0.000

Рис. 11.21. ARMA(1,1). $Y_t = -0.4Y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$.
Корни $\mu_{AR} = -2, \mu_{MA} = 2$

q, Q, p, P

- Model $ARIMA(p, d, 0)$: ACF dumps exponentially or is sinusoidal, PACF is significantly different from zero at lag p
- Model $ARIMA(0, d, q)$: PACF dumps exponentially or is sinusoidal, ACF is significantly different from zero at lag q

$\Rightarrow$ initial approximation for p, q, P, Q:

- q: the number of the last lag $\tau < S$ at which ACF was significant
- $Q * S$: the number of the last seasonal lag at which ACF was significant
- p: the number of the last lag $\tau < S$ at which PACF was significant
- $P * S$: the number of the last seasonal lag at which PACF was significant

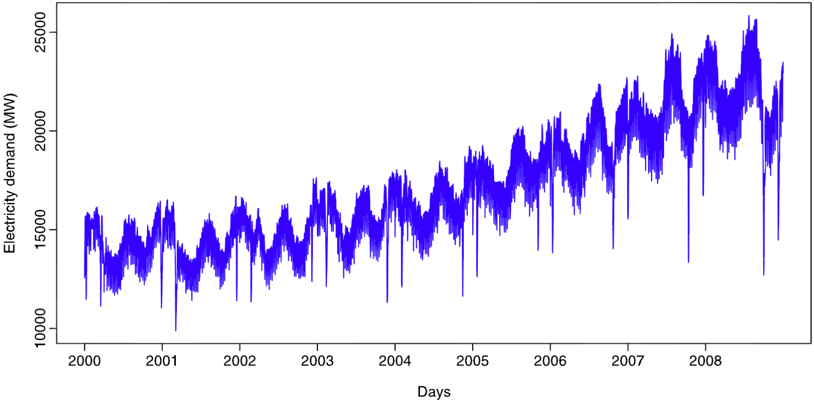
Model Parameters Estimation

1. Check stationarity of parameters, if there is non-stationarity, shift to differences. For the sake of easier interpretation the difference operator should also be applied to parameters.
2. A regression is built for the time series of differences in supposition that errors are described by a model of initial approximation (as a rule it is either $AR(2)$ or $SARMA(2,0) \times (1,0)_s$).
3. A suitable model $ARMA\,(p_1, q_1)$ for residuals of regression $\hat{z}_t$ is selected.
4. Regression is rebuilt in supposition that the errors are described by model $ARMA\,(p_1, q_1)$.
5. Residuals $\hat{\varepsilon}_t$ are analyzed.

Formal check of parameters significance is highly important for the sub-task of regression, in order to select parameters it is neessary to compare the values of models AIC to all subsets x_j.

Example: `https://www.otexts.org/fpp/9/1`

Electricity Consumption in Turkey



- weekly seasonality
- yearly seasonality
- holidays according to islamic calendar (the year is about 11 days shorter than according to Gregorian calendar)

ARIMAX

The effects of floating holidays, short-term promotions and other irregular events with a known date may be modeled with regARIMA:

$$\Phi_P\left(B^s\right)\phi\left(B\right)\nabla_s^D\nabla^d z_t = \Theta_Q\left(B^s\right)\theta\left(B\right)\varepsilon_t$$

$$+$$

$$y_t = \sum_{j=1}^{k}\beta_j x_{jt} + z_t$$

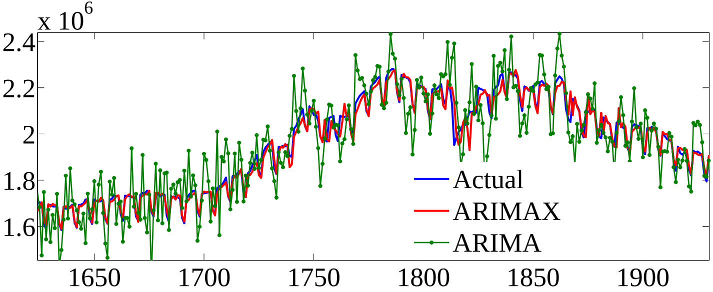
$$=$$

$$\Phi_P\left(B^s\right)\phi\left(B\right)\nabla_s^D\nabla^d\left(y_t - \sum_{j=1}^{k}\beta_j x_{jt}\right) = \Theta_Q\left(B^s\right)\theta\left(B\right)\varepsilon_t.$$

Модель ARIMAX

y_t — is non-stationary,
X_t — vector of regressors from $\mathbb{R}^N$, is known for step $t+d$ before forecasting;
ARIMAX(p,q,d):

$$z_t = \mu + \sum_{n=1}^{N} \frac{v_n(B)}{u_n(B)} X_{n,t} + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$$



Scheme of TS forecasting with ARIMA

1. The graph of time series is built, outliers are identified.
2. Dispersion is stabilized through transformation if needed.
3. If the time series is non-stationary the differentiation degree is chosen.
4. ACF/PACF are analyzed in order to understand whether AR(p)/MA(q) may be used.
5. Candidate models are trained, their AIC/AICc is compared.
6. Unbiasedness, stationarity and non-autocorrelation of the residuals of the obtained model are tested; if the tests fail model modifications are reviewed.
7. In the final model we replace t with $T + h$, future observations with their forecasts, future errors with zeros, previous errors with residuals.

Specification of Confidence Interval

If residuals of the model are normal and stationary forecast intervals are specified theoretically.
For example, the forecast interval for a forecast of the next time point is — $\hat{y}_{T+1|T} \pm 1.96\hat{\sigma}_\varepsilon$.

If normality or stationarity is not fulfilled forecast intervals are simulated.

Conclusion

PRO ARIMA models:

- have strong theoretical argumentation for stationary TS
- can be applied to time series for trends and seasonality
- allow to take into account independent variables

CONS:

- do not work for time series with missing values
- finding of internal coefficients α, ϕ, θ is complicated
- it is not easy to find p,q,d, P, Q, D you need look at ACF, PACF
- ARIMA is based in assumption of iid from Normal distribution:
  - it's not true for all time series
  - it can not be checked for short time series

Literature

Магнус Я.Р. и др. Эконометрика. Начальный Курс М.: Дело, 2007, http://math.isu.ru/ru/chairs/me/files/books/magnus.pdf

Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice. — OTexts, https://www.otexts.org/book/fpp

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003, http://www.arshinov74.ru/files/files/3.pdf

Jonathan D. Cryer, Kung-Sik Chan Time Series Analysis With Applications in R. Second Edition. Springer, 2008

Unit roots tests. https://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf