MSDS 596 Regression & Time Series

Lecture 12 Time Series Diagnostics and Forecasting

Department of Statistics
Rutgers University

Nov 19, 2020

- **Homework 8**: last one, due Monday Nov 30th at 4:30pm.
- **Quiz 2**: Thursday Dec 3 in class. Same rules as Quiz 1.
- **Project report**: due Tuesday Dec 8 at 4:30pm. Please refer to course syllabus and project guideline for formatting and submission requirements.
- **Final evaluation**: Thursday Dec 10, afternoon to evening
    - One-on-one, 8-10 minutes each (order TBD after the break)
    - Two components:
    1. Your mastery of course material (both regression and time series);
    2. Your project. Be prepared to give a 2-minute overview of your project and answer questions about it.
    - Think of it as a “mini interview”.

Schedule

Week	Date	Topic
1	9/3	Intro to linear regression (JF1,2)
2	9/10	Estimation (JF2)
3	9/17	Inference I (JF3)
4	9/24	Inference II (JF3)
5	10/1	Inference and prediction (JF4,5)
6	10/8	Explanation; model diagnostics (JF6-8)
7	10/15	Transformation and model selection (JF9-10)
8	10/22	Shrinkage methods (JF11)
9	10/29	Time series exploratory analysis (CC2,3)
10	11/5	Linear time series: ARIMA models (CC4-5)
11	11/12	Model specification and estimation (CC6,7)
12	11/19	Diagnostics and forecasting (CC8,9)
	11/26	(no class)
13	12/3	Seasonal models (CC10)
14	week of 12/7	Project & final evaluation

Recall - ARIMA model fitting: general strategy

1. **Specification**: decide on reasonable and tentative values for (p, d, q);
    - ACF of MA processes;
    - Partial autocorrelation function (PACF) of AR processes;
    - Extended partial autocorrelation function (EACF) of ARMA processes;
    - Unit-root nonstationarity: Dickey-Fuller test;
    - Selection rules based on AIC and BIC.
2. **Estimation**: estimate parameters in the most efficient way;
3. **Diagnostics**: look critically at the fitted model just obtained to check its adequacy, such as we did for regression and deterministic trend time series models;
    - Residual analysis: covariance of fitted residuals, Ljung-Box test;
    - Model adequacy: causality, invertibility, overfitting, parameter redundancy.

Diagnostics: general philosophy

If a model has been identified and parameters have been estimated, how to decide whether it is adequate?

- No model ever represents the truth absolutely. The best policy is to devise statistical procedures that are sensitive to likely discrepancies, but be prepared to employ models that exhibit slight lack of fit.
    - If there is evidence of serious inadequacy, we shall need to know how the model can be modified in the next iterative cycle.
    - No system of diagnostic checks can ever be comprehensive. It is always possible that unexpected data characteristics could be overlooked.
    - If a model has been fitted to a reasonably large body of data, and thoughtfully devised diagnostic checks are applied to the model and they fail to show serious discrepancies, we should feel more comfortable about using the model.
    - Know the facts as clearly as possible, then use your judgment.

Diagnostics: residual analysis

For ARMA models, residual is defined in AR(∞) form:

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \cdots + e_t$$

where $\pi = \pi(\theta, \phi)$ are (implicit) functions of the ARMA coefficients. The residual is the difference between the *actual* and *predicted* time series, i.e.

$$\hat{e}_t = Y_t - \hat{\pi}_1 Y_{t-1} - \hat{\pi}_2 Y_{t-2} - \cdots ,$$

and should behave like white noise if the model is well-specified.

- zero mean: if not, include a constant term;
- constant variance: if not, consider transforming the time series;
- residual normality: QQ plot, Shapiro-Wilk test;
- residual autocorrelation:
  - ACF, PACF, EACF should look "clean";
  - runs test;
    r_k
  - Ljung-Box test: based on $\hat{r}_k$ (sample ACF of residuals).
- outliers, missing values and imputation.

Diagnostics: residual analysis

Let $\hat{r}_k$ be the **sample ACF of the residuals** of a correctly specified and efficiently estimated model. *Note the distinction between* $\hat{r}_k$ *and* r_k.

If the model is white noise, $\hat{r}_k \sim N(0, 1/n)$ for n large.
If the model is AR(1):

- $Var(\hat{r}_1) \approx \phi^2/n \approx 1/n?$
- for $k \geq 2$, $Var(\hat{r}_k) \approx [1 - (1 - \phi^2)\phi^{2k-2}]/n \approx 1/n$
- $Cor(\hat{r}_1, \hat{r}_k) \approx -\text{sgn}(\phi)(1 - \phi^2)\phi^{2k-2}/[1 - (1 - \phi^2)\phi^{2k-2}]$.

If the model is AR(2):

- $Var(\hat{r}_1) \approx \phi_2^2/n \approx 1/n?$
- $Var(\hat{r}_2) \approx (\phi_2^2 + \phi_1^2(1 + \phi_2)^2)/n$
- for $k \geq 3$, $Var(\hat{r}_k) \approx 1/n$

Ljung-Box test

Instead of looking at residual correlations at individual lags, take their magnitude into account as a group. Define the Ljung-Box test statistic

$$Q_* = n(n+2)\left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \cdots + \frac{\hat{r}_K^2}{n-K}\right).$$

- If the correct ARMA(p, q) model is estimated,

$$Q_* \sim \chi^2_{K-p-q}, \qquad \text{as } n \to \infty.$$

- Incorrect models tend to inflate Q_*.
- The maximum lag K needs to be chosen, such that $\psi_j \approx 0$ for all $j > K$.

Example: U.S. GNP

Check residuals of the AR(3) model fitted to the U.S. GNP data:

```
Box.test(gnp.ar3$residuals,lag=12,type='Ljung')
## data:  gnp.ar3$residuals
## X-squared = 8.4823, df = 12, p-value = 0.7464
```

Compute p-value using 12 - 3 = 9 degrees of freedom:

```
pv=1-pchisq(8.48, 9); pv
## 0.4865883
```

This can also be achieved by specifying the command with the optional argument: `Box.test(..., fitdf = 3)`.

Diagnostics: residual autocorrelation

If the ACF, PACF, EACF of the fitted residuals are not clean, or Ljung-Box test rejects the white noise assumption, the fitted model may be underspecified or mis-specified.

- If ARMA(p, q) is not satisfactory, increase the order.
  Try ARMA($p+1$, q) or ARMA(p, $q+1$), but **do not increase** both orders at the same time: parameter redundancy.
- If the residual series shows some seasonality, e.g. significant ACF at 12, 24, ... include seasonal terms in the model (SARIMA models).
- If the residual shows an isolated significant ACF at lag k, including a Y_{t-k} or a e_{t-k} term in the model often (but not always) helps.
- Be aware of overfitting.

Diagnostics: causality and invertibility

We require that AR models (or components) be causal, and MA models (or components) be invertible. Therefore, all the roots of the estimated characteristic equations

$$\hat{\phi}(z) = 1 - \hat{\phi}_1 z - \hat{\phi}_2 z^2 - \cdots - \hat{\phi}_p z^p,$$
$$\hat{\theta}(z) = 1 - \hat{\theta}_1 z - \hat{\theta}_2 z^2 - \cdots - \hat{\theta}_q z^p.$$

should be larger than one in modulus.

Example. Check causality of the fitted AR model on the U.S. GNP data:

```
p1=c(1,-gnp.ar3$coef[1:3]) # Characteristic polynomial
roots=polyroot(p1) # Find solutions
roots
## [1]  1.590253+1.063882i -1.920152+0.000000i  1.590253-1.063882i
Mod(roots)
## [1] 1.913308 1.920152 1.913308
```

*Note: pay attention to the sign of coefficients for MA models.

Overfitting as a diagnostic tool

Principle of parsimony: we prefer the least complex explanation for the observation, i.e. the simplest model that provides adequate fit to the data.

Suppose $\{Y_t\}$ is fitted with an AR(p+1) model. We can tell an AR(p) model would suffice, if

- Estimates for $\phi_1, \ldots \phi_p$ from both the AR(p+1) and AR(p) fits are similar in value;
- Estimate for ϕ_{p+1} from the AR(p+1) fit is not statistically different from zero;
- *Note: when reducing model complexity, do not remove all non-significant coefficient at the same time. Remove one, re-estimate the model, then try again.

Diagnostics: parameter redundancy

If $\{Y_t\}$ follows an ARMA(p, q) model

$$\phi(B)Y_t = \theta(B)e_t,$$

then it also follows an ARMA($p+1, q+1$) model:

$$\underbrace{(1 - cB)\phi(B)}_{\phi'(B)}Y_t = \underbrace{(1 - cB)\theta(B)}_{\theta'(B)}e_t$$

(p+1) order poly ---- $\phi'(B)$ $\quad$ $\theta'(B)$ ---- (q+1) order poly.

for any c. In fact, it also follows an ARMA($p+k, q+k$) model

$$\underbrace{c(B)\phi(B)}Y_t = \underbrace{c(B)\theta(B)}e_t$$

where $c(B)$ is any k-th order polynomial.

Diagnostics: parameter redundancy

Parameter redundancy is a type of nonidentifiability in model specification.

- Solution: look for **common (almost equal) roots** to the AR and MA characteristic equations. If one is found, reduce both orders by one.
- If there are several common roots, the model can be further simplified. (Note: if one common root is a complex number, there must be another matching root, which is its conjugate).

Thus when fitting ARMA models,

1. Start with lower order models with small p, q;
2. Increase order as needed, one at a time.

Diagnostics: parameter redundancy

~ AR(1)

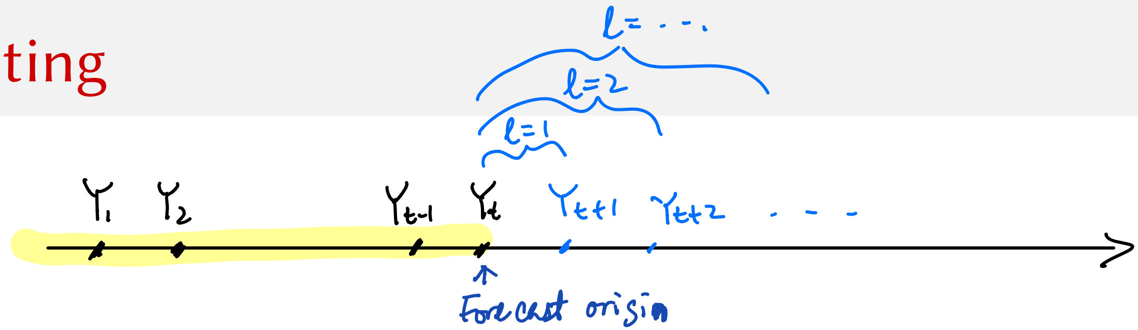
```
> x=arima.sim(model=list(ar=0.7,ma=0),sd=1,n=500)
> out=arima(x,c(2,0,1))
> out
```

ARMA(2,1)

```
Call:
arima(x = x, order = c(2, 0, 1))

Coefficients:
         ar1      ar2      ma1  intercept
      1.6087  -0.6155  -0.9512    -0.2599
s.e.  0.0534   0.0489   0.0330     0.3100

sigma^2 estimated as 0.9644:  log likelihood = -700.84,  aic = 1411.68

> polyroot(c(1,-out$coef[1],-out$coef[2]))
[1] 1.018310+0i 1.595582+0i

> polyroot(c(1,out$coef[3]))
[1] 1.051297+0i
```

MA characteristic poly

Forecasting



Based on observed time series $Y_{1:t} = \{Y_1, \ldots, Y_t\}$, we would like to make predictions of the future observation Y_{t+l}.

- t: forecast origin
- l: lead time
- Minimum MSE forecasting: use a function of the observed data $h(Y_{1:t})$ to predict Y_{t+l}, while minimizing the mean squared error:

$$E\left((Y_{t+l} - h(Y_{1:t}))^2 \mid Y_{1:t}\right).$$

Minimum MSE Forecasting

The best choice of function h (in terms of minimum MSE) to serve as the predictor for Y_{t+l} is

$$\hat{Y}_t(l) = \underbrace{E(Y_{t+l} \mid Y_{1:t})}, \quad \sim \text{a function of } Y_{1:t}$$

and the forecast error

$$e_t(l) = Y_{t+l} - \hat{Y}_t(l).$$

Forecasting AR(1)

$$(Y_{t+1} - \mu) = \phi(Y_t - \mu) + e_{t+1}$$

The 1-step ahead forecast ($l = 1$) satisfies

$$\begin{aligned} \hat{Y}_t(1) - \mu &= E(Y_{t+1} - \mu \mid Y_{1:t}) \\ &= \phi\{E(Y_t \mid Y_{1:t}) - \mu\} + E(e_{t+1} \mid Y_{1:t}) \\ &= \phi(Y_t - \mu) \\ \Rightarrow \hat{Y}_t(1) &= \mu + \phi(Y_t - \mu). \end{aligned}$$

The forecast error

$$e_t(1) = Y_{t+1} - \hat{Y}_t(1) = e_{t+1},$$

with $E(e_t(1)) = 0$ (unbiased) and $Var(e_t(1)) = \sigma^2$.

Forecasting AR(1)

$$(Y_{t+1} - \mu) = \phi(Y_t - \mu) + e_{t+1}$$

Generally for $l \geq 1$, use the difference equation form to obtain

$$\hat{Y}_t(l) = \mu + \phi\left(\hat{Y}_t(l-1) - \mu\right) = \mu + \underbrace{\phi^l}_{\to 0 \text{ as } l \to \infty}(Y_t - \mu)\,.$$

Therefore, for fixed t, $\hat{Y}_t(l) \to \mu$ as $l \to \infty$. The forecast error

$$\begin{aligned} e_t(l) &= Y_{t+l} - \mu - \phi^l(Y_t - \mu) \\ &= e_{t+l} + \underbrace{\phi}_{\psi_1} e_{t+l-1} + \underbrace{\phi^2}_{\psi_2} e_{t+l-2} + \cdots + \underbrace{\phi^{l-1}}_{\psi_{l-1}} e_{t+1} \\ &\dot{\sim} N\left(0, \sigma^2\left(1 + \psi_1^2 + \cdots + \psi_{l-1}^2\right)\right), \end{aligned}$$

where the ψ's are coefficients from the GLP representation of the AR(1).

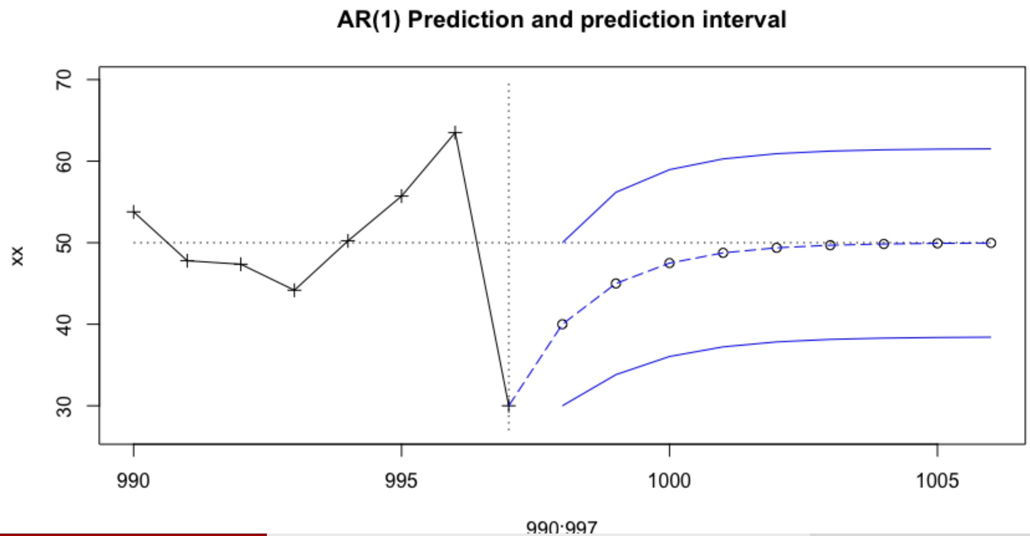
- $E(e_t(l)) = 0$, and
- $Var(e_t(l)) = \sigma^2 \frac{1-\phi^{2l}}{1-\phi^2} \overset{\text{large } l}{\approx} \frac{\sigma^2}{1-\phi^2} = \gamma_0.$

Example: AR(1)

AR(1): $\phi_1 = 0.5$, $\mu = 50$ ($\phi_0 = 25$), $\sigma^2 = 25$ and $Y_t = 30$.

$\hat{Y}_t(1) = 50 + 0.5 \times (30 - 50) = 40$, prediction s.e. $= \sqrt{Var(e_t(1))} = \sigma = 5$.
$\hat{Y}_t(2) = 50 + 0.5 \times (40 - 50) = 45$, prediction s.e. $= \sqrt{\sigma^2(1 + \phi_1^2)} = 5.59$.

l	1	2	3	4	5	6	7	8	9
$\hat{Y}_t(l)$	40	45	47.5	48.8	49.4	49.7	49.8	49.9	50.0
$s.e.$	5.00	5.59	5.73	5.76	5.77	5.77	5.77	5.77	5.77

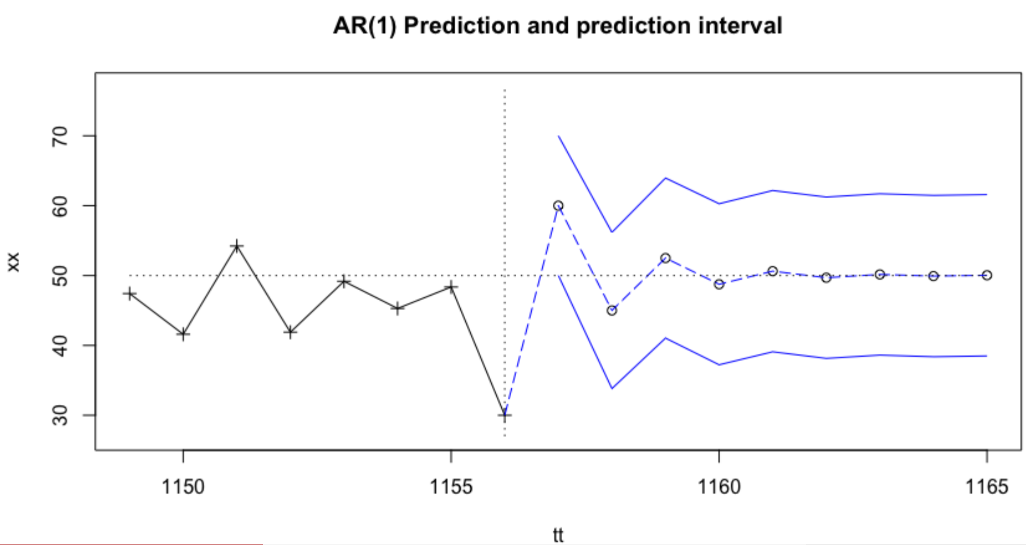


Forecasting AR(1): example

AR(1): $\phi_1 = -0.5$, $\mu = 50$ ($\phi_0 = 75$), $\sigma^2 = 25$ and $Y_t = 30$.

$\hat{Y}_t(1) = 50 - 0.5 \times (30 - 50) = 60$, prediction s.e. $= \sigma = 5$.
$\hat{Y}_t(2) = 50 - 0.5 \times (60 - 50) = 45$, prediction s.e. $= \sqrt{\sigma^2(1 + \phi_1^2)} = 5.59$.

l	1	2	3	4	5	6	7	8	9
$\hat{Y}_t(l)$	60	45	52.5	48.8	50.6	49.7	50.2	49.9	50.0
$s.e.$	5.00	5.59	5.73	5.76	5.77	5.77	5.77	5.77	5.77



Forecasting MA(1)

$$Y_{t+1} = \mu + e_{t+1} - \theta e_t$$

- The 1-step ahead forecast satisfies

$$\hat{Y}_t(1) = E\left(\mu + e_{t+1} - \theta e_t \mid Y_{1:t}\right) = \mu - \theta \underbrace{E\left(e_t \mid Y_{1:t}\right)}_{=e_t,\ \text{if invertible}} = \mu - \theta e_t.$$

The forecast error

$$e_t(1) = Y_{t+1} - \hat{Y}_t(1) = e_{t+1}.$$

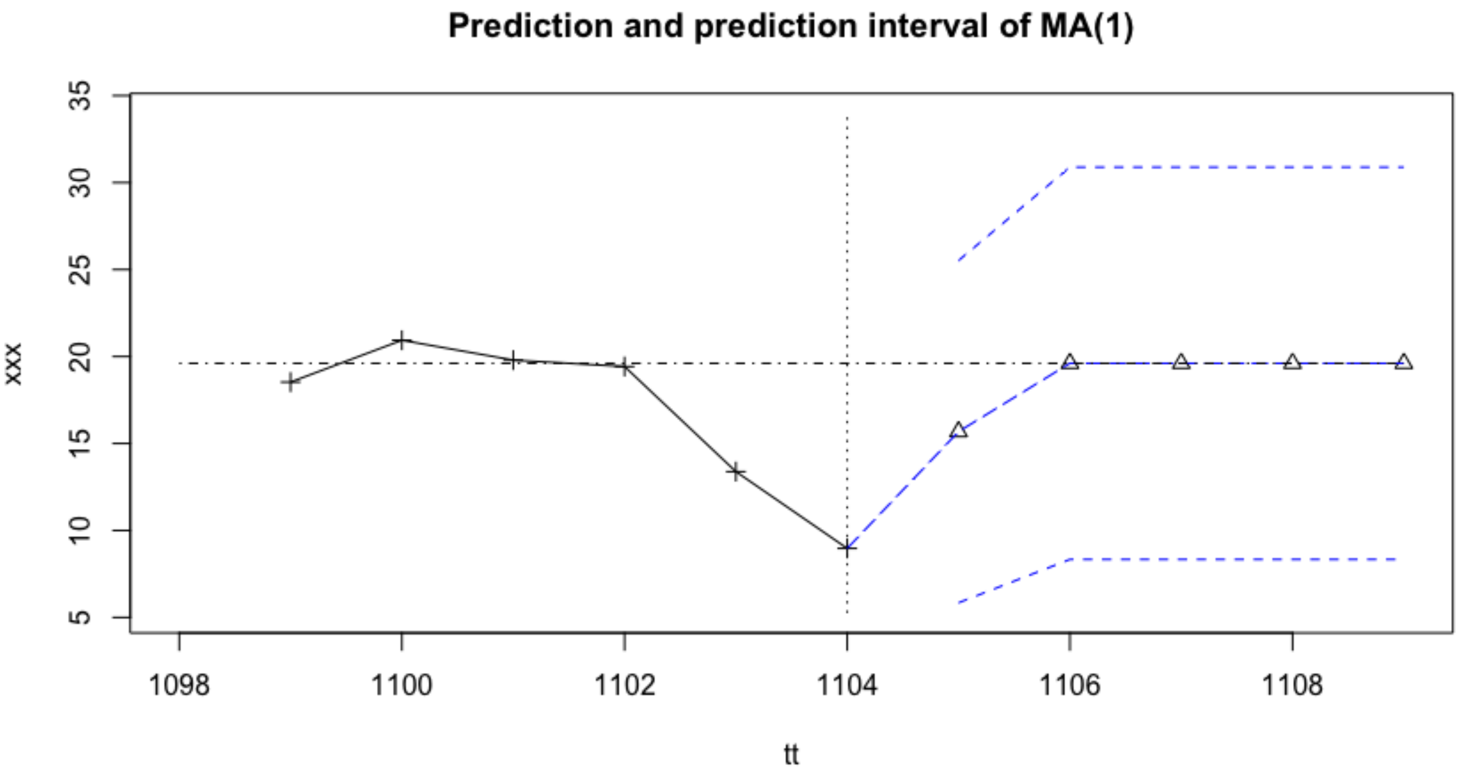
- For $l > 1$,

$$\hat{Y}_t(l) = \mu + \underbrace{E\left(e_{t+l} \mid Y_{1:t}\right)}_{=0} - \theta \underbrace{E\left(e_{t+l-1} \mid Y_{1:t}\right)}_{=0} = \mu,$$

The forecast error

$$e_t(l) = e_{t+l} - \theta e_{t+l-1} \dot{\sim} N\left(0, \sigma^2\left(1 + \theta^2\right)\right).$$

Forecasting MA(1): example



Forecasting ARMA(p, q)

$$Y_{t+l} = \phi_1 Y_{t+l-1} \cdots + \phi_p Y_{t+l-p} + \theta_0 + e_{t+l} - \theta_1 e_{t+l-1} - \cdots - \theta_q e_{t+l-q}.$$

Write the forecast in difference equation form

$$\begin{aligned} \hat{Y}_t(l) &= \phi_1 \hat{Y}_t(l-1) + \cdots + \phi_p \hat{Y}_t(l-p) + \theta_0 - \\ &\quad \theta_1 E\left(e_{t+l-1} \mid Y_{1:t}\right) - \cdots - \theta_q E\left(e_{t+l-q} \mid Y_{1:t}\right), \end{aligned}$$

where

$$E\left(e_{t+j} \mid Y_{1:t}\right) = \begin{cases} 0 & j > 0 \quad (e_{t+1}, e_{t+2}, \ldots) \\ e_{t+j} & j \leq 0 \quad (e_t, e_{t-1}, \ldots). \end{cases}$$

Note: $\hat{Y}_t(j)$ is a true forecast when $j > 0$; otherwise $\hat{Y}_t(j) = Y_{t+j}$ when $j \leq 0$.

Forecasting ARMA(p, q)

Example. ARMA(1, 1): $Y_{t+l} = \phi_1 Y_{t+l-1} + \theta_0 + e_{t+l} - \theta_1 e_{t+l-1}$.

Difference equation form gives

$$\begin{aligned} \hat{Y}_t(1) &= \phi Y_t + \theta_0 - \theta e_t \\ \hat{Y}_t(l) &= \phi \hat{Y}_t(l-1) + \theta_0, \quad l \geq 2 \end{aligned}$$

Written alternatively,

$$\hat{Y}_t(l) = \mu + \phi^l (Y_t - \mu) - \phi^{l-1} e_t, \quad l \geq 1.$$

Forecasting ARMA(p, q)

$$Y_{t+l} = \phi_1 Y_{t+l-1} \cdots + \phi_p Y_{t+l-p} + \theta_0 + e_{t+l} - \theta_1 e_{t+l-1} - \cdots - \theta_q e_{t+l-q}.$$

Generally for ARMA(p, q):

- Short lead time ($l = 1, \ldots, q$): $\hat{Y}_t(l)$ involves the noise terms $\{e_{t-(q-1)}, \ldots, e_t\}$;
- Long lead time ($l > q$): the AR component of the difference equation takes over, and forecasts behave like AR forecasts:

$$\underbrace{\hat{Y}_t\,(l) - \mu}_{} = \phi_1 \underbrace{\left(\hat{Y}_t\,(l-1) - \mu\right)}_{} + \cdots + \phi_p \underbrace{\left(\hat{Y}_t\,(l-p) - \mu\right)}_{},$$

  reminiscent of the Yule-Walker equations for ARMA(p, q) ACF.
- Forecast error expressed in terms of the GLP coefficients:

$$e_t\,(l) = e_{t+l} + \psi_1 e_{t+l-1} + \psi_2 e_{t+l-2} + \cdots + \psi_{l-1} e_{t+1}$$

  In fact, the above is true for all ARIMA processes.
- $\hat{Y}_t(l) \to \mu$ as $l \to \infty$ for stationary ARMA processes.

Forecasting ARIMA processes: random walk with drift

$$Y_{t+1} = Y_t + \theta_0 + e_{t+1}$$

- What kind of ARIMA process is this? ARIMA(0, 1, 0) = I(1)
- $\hat{Y}_t(1) = E\left(Y_t + \theta_0 + e_{t+1} \mid Y_{1:t}\right) = Y_t + \theta_0;$
- For $l \geq 1$,

$$\hat{Y}_t(l) = \hat{Y}_t(l-1) + \theta_0 = Y_t + \theta_0 \cdot l.$$

  Thus if $\theta_0 \neq 0$, $\hat{Y}_t(l)$ is linear in l with slope θ_0: be careful when including constant terms in nonstationary models!
- Forecast error: for $l \geq 1$,

$$e_t\left(l\right) = e_{t+1} + e_{t+2} + \cdots + e_{t+l},$$

  since the GLP representation of the random walk is $\psi_j = 1$ for all $j \geq 0$.
- $Var\left(e_t(l)\right) = \sigma^2 \sum_{j=0}^{l-1} \psi_j^2 = l \cdot \sigma^2$, which goes to ∞ as $l \to \infty$.

Forecasting ARIMA processes

Suppose $\{Y_t\}$ is ARIMA with GLP representation $Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$. It turns out that Y_t can always be written as a truncated linear process

$$Y_{t+l} = C_t(l) + I_t(l),$$

where for $l > 1$,

- $C_t(l)$ is some function of $Y_t, Y_{t-1}, \ldots$. In particular, if Y_t is invertible and t is large, $C_t(l)$ is a function of $Y_{1:t}$;
- $I_t(l) = e_{t+l} + \psi_1 e_{t+l-1} + \psi_2 e_{t+l-2} + \cdots + \psi_{l-1} e_{t+1}$, and is free of $Y_{1:t}$.

We have that the forecasts and forecast errors

- $\hat{Y}_t(l) = C_t(l)$, and
- $e_t(l) = I_t(l)$.

Forecasting ARIMA processes

Note that the forecast error has

$$E\left(e_t(l)\right) = 0 \quad \text{and} \quad Var\left(e_t(l)\right) = \sum_{j=0}^{l-1} \psi_j^2.$$

If Y_t is stationary, $Var\left(e_t(l)\right) \approx Var\left(Y_t\right) = \gamma_0$ for large l;

- e.g. AR(1): $\psi_j = \phi^j$;
- e.g. MA(1): $\psi_1 = -\theta$ and $\psi_j = 0$ for $j > 1$;

If nonstationary, $Var\left(e_t(l)\right) \to \infty$ as $l \to \infty$.

- e.g. Random walk/I(1): $e_t = \nabla Y_t$. $\psi_j = 1$ for all j;
- e.g. IMA(1, 1): $\nabla Y_t = e_t - \theta e_{t-1}$, $\psi_j = 1 - \theta$ for all $j \geq 1$;
- e.g. ARI(1, 1) $\nabla Y_t = \nabla Y_{t-1} + e_t$, for $|\phi| < 1$
  - $\psi_j = (1 - \phi^{j+1})(1 - \phi)$ for all $j \geq 1$;
- e.g. IMA(2, 2): $\nabla^2 Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$
  - $\psi_j = 1 + \theta_2 + (1 - \theta_1 - \theta_2)j$ for all $j \geq 1$.

Forecasting ARIMA processes: general strategy

General strategy to forecasting ARIMA(p,d,q) processes is to write the process as a nonstationary ARMA(p+d, q) process, and obtain $\hat{Y}_t(l)$ using forecasting methods for stationary ARMA processes.

This strategy is equivalent to forecasting the stationary differenced series $W_t = \nabla^d Y_t$, then "undo" the difference.

Forecasting ARIMA processes: general strategy

Example: IMA(1, 1): $\nabla Y_t = e_t - \theta e_{t-1}$.

- Treat as nonstationary $ARMA(1,1)$:

$$\hat{Y}_t(1) = Y_t - \theta e_t$$
$$\hat{Y}_t(l+1) = \hat{Y}_t(l), \quad l \geq 2$$

- Consider $W_t \sim MA(1)$:

$$\hat{W}_t(1) = -\theta e_t$$
$$\hat{W}_t(l) = 0, \quad l \geq 2$$

and

$\hat{Y}_t(0)$

$$\hat{W}_t(1) = \hat{Y}_t(1) - Y_t$$
$$\hat{W}_t(l) = \hat{Y}_t(l) - \hat{Y}_t(l-1)$$

gives the same result.

Time series regression

- Time series regression: suppose the model for Y_t is such that

$$Y_t = \beta_0 + \beta_1 X_{t1} + \cdots + \beta_k X_{tk} + e_t, \; 1 \leq t \leq T;$$
$$e_t \sim \text{ARIMA}(p, d, q).$$

- R implementation. For example, if $e_t \sim \text{ARIMA}(1, 1, 1)$:

```
fit = arima(y, order=c(1,1,1), xreg=XMat)
```

where `XMat` is the $T \times (k+1)$ matrix with t-th row

$$1, x_{t1}, \ldots, x_{tk}.$$

Forecasting log transformed series

Let $Z_t = \log(Y_t)$ be the modeled series.

- Jensen's Inequality:

$$E\left(Y_{t+l} \mid Y_{1:t}\right) \geq \exp\left[E\left(Z_{t+l} \mid Y_{1:t}\right)\right],$$

  so the naíve forecast $\exp\left[\hat{Z}_t\left(l\right)\right]$ is **not** minimum MSE forecast for Y_t.
- What is minimum MSE forecast, then? If $Z_{t+1} \mid Z_t \sim Normal$,

$$E\left(Y_{t+l} \mid Y_{1:t}\right) = E\left(\exp\left(Z_{t+l}\right) \mid Z_{1:t}\right) = \exp\left(\mu_z + \frac{\sigma_z^2}{2}\right),$$

  where

$$\begin{aligned}\mu_z &= E\left(Z_{t+l} \mid Z_{1:t}\right) = \hat{Z}_t\left(l\right)\\ \sigma_z^2 &= Var\left(Z_{t+l} \mid Z_{1:t}\right) = Var\left(e_t\left(l\right)\right).\end{aligned}$$

Forecasting log transformed series

For skewed and heavy-tailed distributions, we want to consider the forecast that minimizes the **Mean Absolute Error (MAE)**:

$$E \left\| Y_{t+l} - h\left(Y_{1:t}\right) \right\|$$

If $Z_t \sim Normal$:

- $Y_t = \exp(Z_t) \sim logNormal$, which is right-skewed and heavy tailed.
- $\hat{Z}_t\left(l\right)$ is both MSE and MAE optimal for $\{Z_t\}$;
- $\exp\left[\hat{Z}_t\left(l\right)\right]$ is MAE optimal for $\{Y_t\}$, because the median is invariant to monotone transformations.

R implementation of ARIMA: constant term

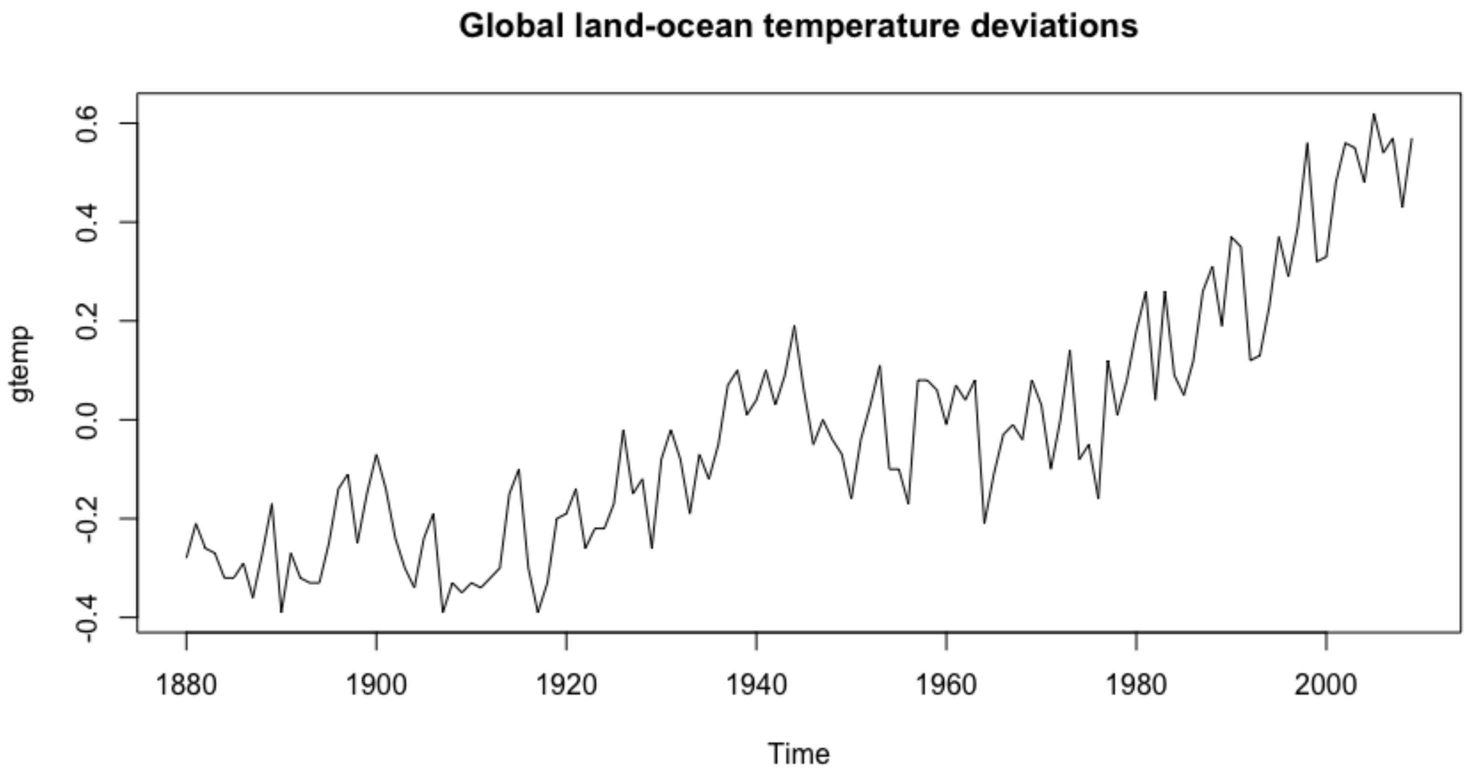
- Note that a constant term is NOT included in the `arima()` fit, if the ARIMA model involves differencing.
- For example, if $\{Y_t\}$ has a drift, and the intended model is

$$\phi(B)(\nabla Y_t) = \phi_0 + \theta(B)e_t,$$

  the `arima()` fit will not include the ϕ_0 term.
- The remedy is to use the argument: `xreg=1:length(x)`. That is, R is fitting the model

$$\phi(B)[\nabla(Y_t - \beta t)] = \theta(B)e_t \iff \phi(B)(\nabla Y_t - \beta) = \theta(B)e_t.$$

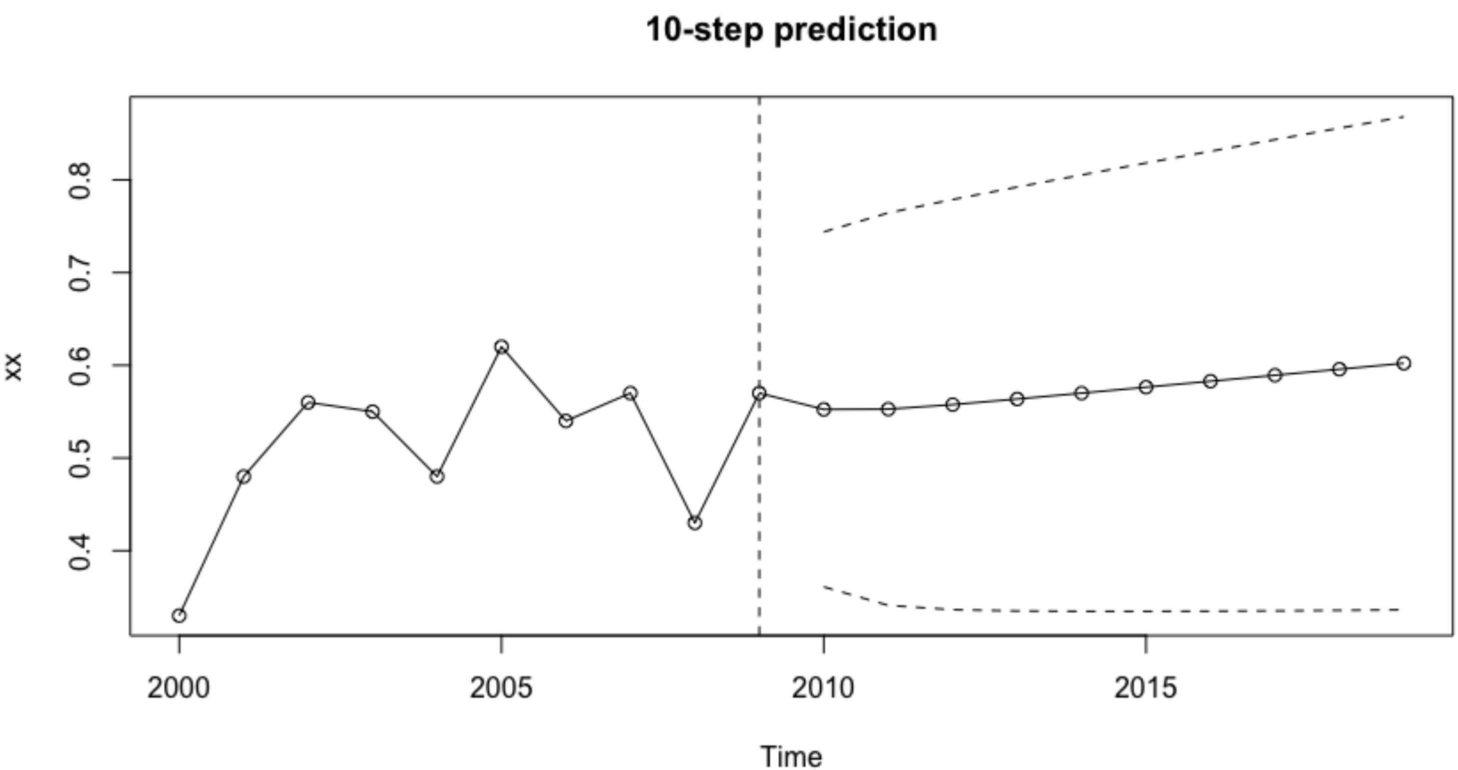
Example: global mean land-ocean temperature deviations



Example: global mean land-ocean temperature deviations

```
library(astsa); data(gtemp);
l=10; nobs = length(gtemp);

fit1 = arima(gtemp, order=c(1,1,1))  ## incorrect
Coefficients:
         ar1      ma1
      0.2256  -0.7158
s.e.  0.1235   0.0792
sigma^2 estimated as 0.009539:
log likelihood = 116.83,  aic = -227.65

fit2 = arima(gtemp, order=c(1,1,1), xreg=1:nobs)  ## correct
Coefficients:
         ar1      ma1  1:nobs
      0.2570  -0.7854  0.0064
s.e.  0.1177   0.0707  0.0025
sigma^2 estimated as 0.009163:
log likelihood = 119.34,  aic = -230.68

fore2 = predict(fit2, l, newxreg=(nobs+1):(nobs+l))
```

Example: global mean land-ocean temperature deviations



Prediction limits for ARIMA forecasts

If $\{e_t\} \sim N(0, \sigma^2)$ i.i.d., then

$$e_t(l) \sim N(0, \sigma^2 \sum_{j=0}^{l-1} \psi_j^2).$$

Can use the estimates $\hat{\sigma}^2$ and $\hat{\psi}_j$ to construct prediction intervals.

Example. AR(1) process: $Var\left(e_t(l)\right) = \sigma^2 \left(1 - \phi^{2l}\right) / \left(1 - \phi^2\right)$.

- $\hat{Var}\left(e_t(1)\right) = \hat{\sigma}^2$;
- $\hat{Var}\left(e_t(2)\right) = \hat{\sigma}^2 \left(1 - \hat{\phi}^4\right) / \left(1 - \hat{\phi}^2\right)$;
- $\hat{Var}\left(e_t(l)\right) \approx \hat{\sigma}^2 / \left(1 - \hat{\phi}^2\right)$ for l large.

See R markdown for additional illustrations.