Key here: after stationary, identify dependence structure (and use for forecasting)

(overshort example)

"White noise" $H_0$: data are just white noise (i.e., no dependence structure)

Let $Z_t$ be the stationary time series after “transforming” (including estimating out time trends and other covariates) the original time series $Y_1, \ldots, Y_n$

$\{Z_t \text{ could be residuals after time-based predictors}\}$

Autocorrelation function (ACF or $\widehat{SACF}$)

- measure linear association between time series observations separated by a lag of $m$ time units:

$$r_m = \frac{\sum_{t=b}^{n-m}(Z_t - \bar{Z})(Z_{t+m} - \bar{Z})}{\sum_{t=b}^{n}(Z_t - \bar{Z})^2}, \quad \bar{Z} = \frac{\sum_{t=b}^{n}Z_t}{n - b + 1}$$

SE of $r_m$ is

$$S_{r_m} = \frac{\sqrt{1 + 2\sum_{l=1}^{m-1}r_l^2}}{\sqrt{n - b + 1}}$$

($b = 1$ unless use differencing)

- call $r_m$ the sample autocorrelation function: $SACF(m)$ or $\widehat{ACF}(m)$

- sometimes used: $t_m = r_m/S_{r_m}$

Autocorrelation plot (or SAC, SACF, moving average of order $q$)

- determine stationarity and identify “MA(q)” dependence structure

- bar-plot $r_m$ vs. $m$ for various lags $m$ (sketch)

- lines often added to represent 2 SE’s (sketch)
  - rough 95% confidence intervals
  - if $r_m$ is more than 2 SE’s away from zero, consider it “significant”
  - compare $|t_m|$ to 2
    - (for lags $m \leq 3$, use 1.6 because “low” lags most important to pick up)

- SAC terminology:
  - “spike”: $r_m$ is “significant”
  - “cuts off”: no “significant” spikes after $r_m$
  - “dies down”: decreases in “steady fashion”
SAC & stationarity:

1. \( Z_t \) stationary if SAC either cuts off fairly quickly or dies down fairly quickly; sometimes dies down in “damped exponential fashion” (sketches)

2. \( Z_t \) nonstationary if SAC dies down extremely slowly (sketch)

3. Ch. 9 of Bowerman & O’Connell: “... if the SAC cuts off fairly quickly, it will often [but not always] do so after a lag \( k \) that is less than or equal to 2.”

MA\( (q) \) dependence structure: moving average process of order \( q \):

- model:
  \[
  Z_t = \delta + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}
  \]
  \( Z_t \): stationary “transformed” time series
  \( \theta_i \): unknown parameters
  \( a_t \): random shocks
  \( \delta \): unknown parameter [only included if \( \bar{Z} \) is statistically different from 0]

  (truly indep. error terms)

- identify using SAC: first \( q \) terms of SAC will be non-zero, then drop to zero (sketch)

Partial Autocorrelation Function (PACF or SPACF)

- autocorrelation of time series observations separated by a lag of \( m \), with the effects of the intervening observations eliminated:

  \[
  r_{m,m} = \begin{cases} 
  r_1 & \text{if } k = 1 \\
  \frac{r_{m} - \sum_{l=1}^{m-1} r_{m-1,l}r_{m-l}}{1 - \sum_{l=1}^{m-1} r_{m-1,l}r_{l}} & \text{if } k \geq 2
  \end{cases}
  \]

  where \( r_m = SACF(m) \)

  and \( r_{m,l} = r_{m-1,l} - r_{m,m}r_{m-1,m-l} \) for \( l = 1, \ldots, m - 1 \)

  SE of \( r_{m,m} \) is \( S_{r_{m,m}} = 1/\sqrt{n - b + 1} \) (\( b = 1 \) unless use differencing)

- call \( r_{m,m} \) the sample partial autocorrelation function: \( SPACF(m) \) or \( \hat{PACF}(m) \)

- sometimes used: \( t_{r_{m,m}} = r_{m,m}/S_{r_{m,m}} \)

- Inverse Autocorrelation Function (IACF) is similar to PACF, and rarely discussed
Partial Autocorrelation Plot (or SPAC)

- bar-plot $r_{m,m}$ vs. $m$ for various lags $m$ (sketch)
- lines often added to represent 2 SE’s; compare $|r_{m,m}|$ to 2; (sketch)
- Use SPAC to identify $AR(p)$ dependence structure:
  - terminology as before: “spike”, “cuts off”, and “dies down” – $r_{m,m}$
  - first $p$ terms of SPAC will be non-zero, then drop to zero (sketch)

*AR(p)* dependence structure: autoregressive process of order $p$:

- recall [special case] first-order autocorrelation: $\varepsilon_t = \phi \varepsilon_{t-1} + a_t$
- generalize to account for error dependence structure where current time series value depends on past values: $\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \ldots + \phi_p \varepsilon_{t-p} + a_t$
- More common representation for AR(p):
  
  $Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t$

  - $\phi_i$ are unknown parameters; random shock $a_t$ iid $N(0, \sigma^2)$
  - $\delta = \mu(1 - \phi_1 - \ldots - \phi_p)$; $\mu = E[Z_t]$
  - $Z_t$ are “residuals” $\Rightarrow \mu \equiv 0$ $\Rightarrow$ common to assume $\delta = 0$

- special case: Random Walk Model
  - $Z_t = Z_{t-1} + a_t$
  - $AR(1)$ is a discrete time continuous Markov Chain
    (probability at time $t$ depends only on state at time $t-1$)

(GE investment example) $H_0 \neq 0, \mu, \sigma^2$

Some convenient notation: backshift operator

$B Z_t = Z_{t-1}$

$B^2 Z_t = BBZ_t = BZ_{t-1} = Z_{t-2}$

AR(p): $Z_t = \delta + \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + a_t$

$= \delta + \phi_1 BZ_t + \ldots + \phi_p B^p Z_t + a_t$

$= \delta + (\phi_1 B + \ldots + \phi_p B^p)Z_t + a_t$

$\Rightarrow (1 - \phi_1 B - \ldots - \phi_p B^p)Z_t - \delta = a_t$

MA(q): $Z_t = \delta + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$

$= \delta + a_t - \theta_1 Ba_t - \ldots - \theta_q B^q a_t$

$= \delta + (1 - \theta_1 B - \ldots - \theta_q B^q) a_t$

$\Rightarrow (1 - \theta_1 B - \ldots - \theta_q B^q)^{-1}(Z_t - \delta) = a_t$
Interpreting and representing common dependence structures:

- **AR(p)** – “Autoregressive” Process:
  - current & future values (of \( Z_t \)) depend on previous \( p \) time series values (\( Z_t \))
  - \( (1 - \phi_1 B - \ldots - \phi_p B^p)Z_t - \delta = a_t \)

- **MA(q)** – “Moving” Average Process:
  - current & future values (of \( Z_t \)) depend on previous \( q \) random shocks (\( a_t \))
    (less intuitive)
  - \( (1 - \theta_1 B - \ldots - \theta_q B^q)^{-1}(Z_t - \delta) = a_t \)

(Gas price example) \( (H07.12d, \, Ex. \, 3) \)

**ARMA(p,q)** dependence structure: mixed autoregressive-moving average model

- If AR and MA alone aren’t enough, what about a “composite” (AR and MA) model?
  \[
  Z_t = \delta + \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}
  \]

  - in backshift notation, ARMA(p,q):
    \[
    (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)Z_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)a_t
    \]
    \[
    \Rightarrow (1 - \theta_1 B - \ldots - \theta_q B^q)^{-1}[(1 - \phi_1 B - \ldots - \phi_p B^p)Z_t - \delta] = a_t
    \]

- Invertibility – model assumption (in addition to stationarity)
  - intuitively, “weights” (\( \phi_l \) & \( \theta_l \)) on past observations decrease for larger \( l \)

---

**Common Dependence Structures for Stationary Time Series**

<table>
<thead>
<tr>
<th>SAC (ACF)</th>
<th>SPAC (PACF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA(1)</td>
<td>cuts off after lag 1</td>
</tr>
<tr>
<td>MA(2)</td>
<td>cuts off after lag 2</td>
</tr>
<tr>
<td>AR(1)</td>
<td>dies down in damped exponential decay</td>
</tr>
<tr>
<td>AR(2)</td>
<td>dies down, in mixture of damped exp. decay &amp; sine waves</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>dies down in damped exp. decay</td>
</tr>
</tbody>
</table>
Several approaches exist to estimate $\phi_l$'s, $\theta_l$'s, and $\beta_j$'s, and deal with initial lag

- **ULS (unconditional least squares): MA(q) & AR(p)**
  - also called nonlinear least squares
  - minimize SS error

- **YW (Yule-Walker): AR(p)**
  - generalized least squares using OLS residuals to estimate covariance across observations

ARIMA(p,d,q) dependence structure: Autoregressive Integrated Moving Average Model

- a very flexible family of models $\Rightarrow$ useful prediction

Recall differencing:

- First difference: $Z_t = Y_t - Y_{t-1}$, $t = 2, \ldots, n$
- Second difference: $W_t = Z_t - Z_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$, $t = 3, \ldots, n$
- Pros: help make time series more stationary ("stubborn" trends)
- Cons: can destroy cyclic behavior (harder to forecast)
- Useful when transformations and addition of time-related predictors (low-order polynomial, trigonometric, dummy) do not make time series stationary

After differencing, AR and MA dependence structures may exist: ARIMA(p, d, q)

- p : AR(p) – value at time $t$ depends on previous $p$ values
- d : # of differences (need to take $d^{th}$ difference to make stationary)
- q : MA(q) – value at time $t$ depends on previous $q$ random shocks

use SAC and SPAC to select $p$ and $q$ – but how to select $d$?

- usually look at plots of time series
- choose lowest $d$ to make stationary (also SAC)

Recall backshift notation:

- $d = 1$ : $Z_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)Y_t$
- general $d$: $Z_t = (1 - B)^dY_t$

ARIMA – Forecasting & Goodness of Fit (note order in which SAS does this)

\[
(1 - B)^dY_t = \beta_0 + \beta_1 X_{t,1} + \ldots + \beta_{k-1} X_{t,k-1} + (1 - \phi_1 B - \ldots - \phi_p B^p)^{-1}(1 - \theta_1 B - \ldots - \theta_q B^q)\ a_t
\]

$\ a_t $ iid $N(0, \sigma^2)$

Independence (given $p$, $d$, and $q$, SAS estimates $\beta_j$'s, $\phi_l$'s, and $\theta_l$'s)
ARIMA(p,d,q) model rewritten, with \( t = 1, \ldots, n \):

\[
Y_t = g_1(Y_1, \ldots, Y_{t-1}) + g_2(X_{t,1}, \ldots, X_{t,k}) + g_3(a_1, \ldots, a_t)
\]

where

\( g_1 = \) linear combination (LC) of previous observations (Differencing)

\( g_2 = \) LC of predictors at time \( t \), in terms of parameters \( \beta_j \) (Linear Model)

\( g_3 = \) function of random shocks in terms of parameters \( \phi_l \) & \( \theta_l \) (AR & MA dependence structures)

“fit model” → estimates & standard errors for \( \beta_j \)’s, \( \phi_l \)’s, & \( \theta_l \)’s

Diagnostics for “goodness of fit”

- Predicted values (point forecast from Box-Jenkins model, even for times \( t > n \)):

\[
\hat{Y}_t = \hat{g}_1(Y_1, \ldots, Y_{t-1}) + \hat{g}_2(X_{t,1}, \ldots, X_{t,k}) + \hat{g}_3(\hat{a}_1, \ldots, \hat{a}_t)
\]

Estimate \( Y_t \) with \( \hat{Y}_t \) if no obs. at time \( l > n \)

- Numerical diagnostics – what are \( S \) and \( Q^* \) like for a “better” model?

  - Standard Error – measure of “overall fit”

    * in SAS: Std Error Estimate

    \[
    S = \frac{\sqrt{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}}{n - n_p}, \quad n_p = \# \text{ parameters in model}
    \]

  - Ljung-Box statistic – diagnostic checking

    * in SAS: lag 6 \( \chi^2 \) for Autocorrelation Check of Residuals

    * Residuals reflect model assumptions

    * Check “adequacy” of overall Box-Jenkins model (for these data)

    \[
    Q^* = n'(n' + 2) \sum_{m=1}^{M} (n' - m)^{-1} r_m^2(\hat{a}) \quad (\text{this is the Ljung-Box statistic})
    \]

    \[
    n' = n - d, \quad d = \text{degree of differencing}
    \]

    \[
    r_m(\hat{a}) = \text{RSAC: sample autocorrelation of residuals at lag } m
    \]

    \[
    M = \text{somewhat arbitrary number of lags to consider, usually multiple of 6;}
    \]

    basic idea: look at “local” dependence among residuals in first \( M \) sample autocorrelations

    * Under \( H_0 \): model is adequate, \( Q^* \sim \chi_{M-n_p}^2 \) (sketch)
• Graphical assessment of model adequacy – what dependence structure remains?
  – look at Residual Sample Autocorrelation (RSAC) & Residual Sample Partial Autocorrelation (RSPAC) plots

Interval Forecasting

• recall: Estimate \( \pm (\text{CriticalValue}) \times (\text{StandardError}) \)
  Note: get CriticalValue from the sampling distribution (sketch)

• get point forecast \( \hat{Y}_t = \hat{Y}_{n+\tau} \) (most interested in \( \tau > 0 \))

• \( SE_{n+\tau} \): standard error of \( \hat{Y}_{n+\tau} \)
  – depends on \( S \)
  – usually, smaller \( S \) \( \Rightarrow \) smaller \( SE_{n+\tau} \)

• \( \hat{Y}_{n+\tau} \pm \left( t_{n-n_p}(1 - \alpha/2) \right) \times SE_{n+\tau} \)

• Based on historical data and the selected Box-Jenkins model, we are \( (1 - \alpha)100\% \) confident that the “true” value of \( Y \) at time \( n + \tau \) will be inside this interval.

General SAS code for ARIMA\((p,d,q)\), \( Y \) in terms of \( X_1, \ldots, X_{k-1} \):

```sas
proc arima data = a1;
   identify var = \( Y \) \( (d) \) crosscorr = (\( X_1, \ldots, X_{k-1} \)) ;
   estimate p = p q = q input = (\( X_1, \ldots, X_{k-1} \)) method = uls plot;
   forecast lead = L alpha = \( a \) noprint out = fout;
run;
```

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d, p, q )</td>
<td>differencing, AR, &amp; MA settings (as before)</td>
</tr>
<tr>
<td>plot</td>
<td>adds RSAC &amp; RSPAC plots</td>
</tr>
<tr>
<td>( L )</td>
<td># times after last observed to forecast</td>
</tr>
<tr>
<td>( a )</td>
<td>set confidence limit; ( a = .10 ) ( \Rightarrow ) 90% conf. limits</td>
</tr>
<tr>
<td>noprint</td>
<td>optional, suppresses output</td>
</tr>
<tr>
<td>out = fout</td>
<td>optional, sends forecast data to fout data set</td>
</tr>
</tbody>
</table>

AR model
\[
\begin{align*}
Z_t &= \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t \\
Z_t &= \delta + \phi_1 Z_{t-1} + \phi_3 Z_{t-3} + a_t
\end{align*}
\]

- \( p = 2 \) \( \text{MA} \) model
\[
\begin{align*}
Z_t &= \delta + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} \\
Z_t &= \delta + a_t - \theta_1 a_{t-1} - \theta_3 a_{t-3} - \theta_7 a_{t-7}
\end{align*}
\]

- \( q = 3 \)

Differencing
\[
\begin{align*}
\text{First: } Z_t &= Y_t - Y_{t-1} = (1 - B)Y_t \\
\text{Second: } Z_t &= (1 - B)(1 - B)Y_t \\
\text{“Lagged”: } Z_t &= Y_t - Y_{t-7}
\end{align*}
\]

- \( d = 1 \)

For \( a = .10 \), data set fout will contain variables \( Y \), forecast, std, 190, u90, residual

- what about time or \( X_1, \ldots, X_{k-1} \)? No – must massage or all “by hand”
How to get forecast values?

- easiest to use backshift notation
- forecasting equation for ARIMA(1,1,1) with one covariate:

\[
(1 - B)Y_t = (\beta_0 + \beta_1 X_t) + (1 - \phi_1 B)^{-1} (1 - \theta_1 B) a_t
\]
\[
(1 - \phi_1 B) (1 - B)Y_t = (1 - \phi_1 B) (\beta_0 + \beta_1 X_t) + (1 - \theta_1 B) a_t
\]
\[
(1 - (1 + \phi_1) B + \phi_1 B^2) Y_t = \beta_0 (1 - \phi_1) + \beta_1 (1 - \phi_1 B) X_t + (1 - \theta_1 B) a_t
\]
\[
Y_t - (1 + \phi_1) Y_{t-1} + \phi_1 Y_{t-2} = \beta_0 (1 - \phi_1) + \beta_1 (X_t - \phi_1 X_{t-1}) + a_t - \theta_1 a_{t-1}
\]
\[
\hat{Y}_t = \hat{\beta}_0 (1 - \hat{\phi}_1) + \hat{\beta}_1 (X_t - \hat{\phi}_1 X_{t-1})
\]
\[
+ \hat{a}_t - \hat{\theta}_1 \hat{a}_{t-1}
\]
\[
+ (1 + \hat{\phi}_1) Y_{t-1} - \hat{\phi}_1 Y_{t-2}
\]

- Note: \( \hat{a}_t = 0, \hat{a}_l = Y_l - \hat{Y}_l \) for \( l < t \), and \( \hat{a}_l = 0 \) for \( l > n \)

Summary of Box-Jenkins Models - choosing a “good” model (choice of \( p, d, \) & \( q \))

- for useful forecasts, meet model assumptions
  - stationarity – transform \( Y \), add predictors, differencing (\( d \))
  
- account for what influences \( Y \):
  - “obvious” effects: effects of predictors
  - less “obvious”: dependence structure (SAC & SPAC together)
    * previous values (autoregressive, \( p \))
    * previous errors (moving average, \( q \))
  
- assess model adequacy
  - \( \text{ACF} \) & \( \text{PACF} \) of residuals
    - \( \text{RSAC} \) & \( \text{RSPAC} \) die down quickly – should have “nothing” left
    - small standard error (\( S \)) & small Ljung-Box statistic (\( Q^* \))
  
- forecast (point & interval) with “adequate” model
  - want “tight” forecasting intervals
  - how far into “future”? \( t = n + \tau, \tau > 0 \)
  - good summary / comparison plot: forecast with confidence limits (sketch)

Now – a case study