Time Series Analysis

Emanuel Nussli

June 2021

1 Descriptive Techniques

Seasonal: at a yearly lag
Cyclical: at a non-yearly lag

Transformations

Stabilize the variance, make the seasonality additive, make the data (more) normally distributed with the Box-Cox transformation for example:

$$\mathrm{y}_t = \begin{cases} (x_t^\lambda - 1)/\lambda & \forall\ \lambda \neq 0 \\ \log x_t & \forall\ \lambda = 0 \end{cases}$$

Linear Filtering or Moving Average

Convert $\{X_t\}_t$ into another series $\{Y_t\}_t$ by the linear operation

$$y_t = Sm(x_t) = \sum_{r=-q}^{s} a_r x_{t+r} \tag{1}$$

with $\sum_{r=-q}^{q} a_r = 1$ which can only be computed for $t = (q+1), ..., T-q$ which is a problem for forecasting.

Exponential Smoothing

Designed to solve the end-point problem from above by using an asymmetric filter of the form

$$Sm(x_t) = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j x_{t-j} \text{ with } \alpha \in (0,1) \tag{2}$$

which ends at x_1 in practice and is thus finite. The weights decrease geometrically with j. We can calculate the residuals by subtracting the moving average:

$$Res(x_t) = x_t - Sm(x_t) = \sum_{r=-q}^{s} b_r x_{t+r} \tag{3}$$

which is also a linear filter with $b_0 = 1 - a_0$ and $b_r = -a_r\ \forall\ r \neq 0$ and $\sum b_r = 0$. These properties can be derived by writing down $Res(x_t)$ and writing out the sum and collecting terms.

Sequential Filtering

Filtering a filtered series again: (i) filter $\{x_t\}_t$ with weights $\{a_r\}$ to obtain $\{y_t\}_t$ and then (ii) filter $\{y_t\}_t$ with weights $\{b_j\}$ to obtain $\{z_t\}_z$.

Convolution: obtain weights $\{c_k\}$ directly from $\{x_t\}_t$ Example: $y_t = 1/3x_{t-1} + 1/3x_t + 1/3x_{t+1}$ and $z_t = 1/3y_{t-1} + 1/3y_t + 1/3y_{t+1} \implies z_t = 1/3(1/3x_{t-2} + 1/3x_{t-1} + 1/3x_t) + 1/3(1/3x_{t-1} + 1/3x_t + 1/3x_{t+1}) + 1/3(1/3x_t + 1/3x_{t+1} + 1/3x_{t+2})$ which sums up to 1, which should be checked for correctness. Formal description:

$$z_t = \sum_{j\in J} b_j y_{t+j} = \sum_j b_j \sum_{r\in R} a_r x_{t+j+r} = \sum_{k\in K} c_k x_{t+k} \text{ with } c_k = \sum_{r\in R} a_r b_{k-r} \tag{4}$$

Differencing

Special case of linear filtering:

$$y_t = \nabla^n x_t = \nabla(\nabla^{n-1} x_t) \text{ with } \nabla x_t = x_t - x_{t-1} \tag{5}$$

First-order differencing is often enough for non-seasonal data and forecasts based on ∇X_t often work better than forecasts based on $\{X_t\}_{t=1}^T$.

Kernel Smoothing

Also a moving average type smoother such that

$$Sm(x_t)\sum_{i=1}^{T} w_t(i)x_i \text{ where } w_t(i) = K(t-i/b)/\sum_{j=1}^{T} K(t-j/b) \tag{6}$$

with $K(x)$ being a kernel function.

Seasonality

Three common models:

$$X_t = m_t + S_t + \epsilon_t \text{ effects sum to zero} \tag{7}$$
$$X_t = m_t S_t + \epsilon_t \text{ effects multiply to one} \tag{8}$$
$$X_t = m_t S_t \epsilon_t \text{ effects multiply to one} \tag{9}$$

Estimating seasonal effects: (i) **without trend**: season minus mean for additive case over the total corresponding year and season divided by mean over the total corresponding year for the multiplicative case (ii) **with trend** $x_t - Sm(x_t)$ or $x_t/Sm(x_t)$. If the seasonality is assumed to be constant, one averages over the years.

Autocorrelation and Correlogram

Compute correlation for pairs $(x_1, x_2), ..., (x_{T-1}, x_T)$ by

$$r_1 = \frac{\sum_{t=1}^{T-1}(x_t - \bar{x}_{(1)}(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\sum_{t=1}^{T-1}(x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{T-1}(x_{t+1} - \bar{x}_{(2)})^2}} \tag{10}$$

with $\bar{x}_{(1)} = \sum_{t=1}^{T-1} x_t/(T-1)$ and $\bar{x}_{(2)} = \sum_{t=1}^{T-1} x_{t+1}/(T-1)$ which are almost the same so we use $\bar{x}$ instead, i.e $\bar{x} = \sum_{t=1}^{T}(x_t - \bar{x})^2$, which can be generalized to lag k such that we have

$$r_k = \frac{\sum_{t=1}^{T-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \tag{11}$$

Correlogram
Plot r_k against k, with $k = 0, ..., M$ with $M << T$
$r_k \dot{\sim} \mathcal{N}(0, 1/T)$ so we can expect 95% of all r_k to lie within $1.96/\sqrt{T}$

2 Some Time Series Models

Strict Stationarity the joint distribution of $\{X_{t1}, ..., .X_{tn}\}$ is the same as the joint distribution of $\{X_{t1+k}, ..., X_{tn+k}\}$ $\forall$ $t_1, ..., t_n, k$
Weak Stationarity

- $\mathbb{E}(X_t) = \mu \ \forall \ t$
- $\text{Cov}(X_t, X_{t+k}) = \gamma(k) \ \forall t, k$
- if $k = 0$, this implies $Var(X_t) = \gamma(0) = \sigma^2 \ \forall \ t$

Properties of the ACF

Let $\{X_t\}_t$ be stationary, then:

- $(k) = \rho(-k)$
- $|\rho(k)| \leq 1$ and $|\rho(0)| = 1$
- The ac.f does not uniquely identify the probability mechanism P

Moving Average Process

$\{Z_t\}_t$ is a purely random process and then X_t is a $MA(q)$ if:

$$X_t = \sum_{i=0}^{q} \theta_i Z_{t-i} \tag{12}$$

and hence

$$\gamma(k) = \begin{cases} 0 \ \forall \ k > q \\ \sigma_z^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k} \ k = 0, 1, ..., q \\ \gamma(-k) \ k < 0 \end{cases} \tag{13}$$

which is always second order stationary as long as $\{Z_t\}_t$ are strictly stationary. If the $\{Z_t\}_t$ are iid normal, then $\{X_t\}_t$ is a strictly stationary normal process. If the $\{Z_t\}_t$ are strictly stationary, so is $\{X_t\}_t$.

Invertibility: A MA process is not always invertible which is desirable because invertibility $\implies$ unique MA process for a given a.c.f. A MA process is invertible if it can be written as $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ with $\sum_{j=0}^{\infty} |\pi_j| < \infty$.
Writing the MA process via the backshift operator B:

$$X_t = \left(\sum_{i=0}^{q} \theta_i B^i \right) Z_t = \theta(B) Z_t \text{ with } \theta(B) \text{ being a polynomial of order } q \tag{14}$$

The MA process is invertible if the roots of $\theta(x) = 0$ lie outside the unit circle.

Autoregressive Processes

$\{Z_z\}_t$ is again a purely random process and $\{X_t\}_t$ is called a autoregressive process of order p if:

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + Z_t \tag{15}$$

and hence (via iterative substitution)

$$X_t = \sum_{j=0}^{\infty} \alpha^j Z_{t-j} \tag{16}$$

meaning that $\{X_t\}_t$ can be expressed an an $MA(\infty)$ process (given $\sum |\alpha|^j$) converges ($|\alpha| < 1$ for AR(1)). We call that duality of MA and AR.

AR(1) Process

We have $Var(X_t) = \sigma_x^2 = \frac{\sigma_z^2}{1-\alpha^2}$ from the MA rep. and $\gamma(k) = \mathbb{E}(X_t X_{t+k}) = \alpha^k \sigma_x^2$ and hence $\rho(k) = \alpha|k| \ \forall \ k$ for the a.c.f. We have $\rho(k) = \alpha^{|k|} \ \forall \ k$. If $|\alpha| = 1$, we have $|\rho(k)| = 1 \ \forall k$.

AR(p) Process

- Any $AR(p)$ process can be written as an $MA(\infty)$ process but there are no easy formulas for the MA coefficients when $p \geq 2$
- any AR is by definition invertible

Stationarity: let $X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + Z_t = \alpha_1 B X_t + ... + \alpha_p B^p X_t + Z_t \iff X_t(1 - \alpha_1 B ... - \alpha_p B^p) \iff Z_t = X_t \alpha(B)$ and we often dont call it $\alpha(B)$ but $\alpha(x)$ if we are interested in its roots. If the roots of $\alpha(x)$ lie outside the unit circle, our AR process is stationary.

Yule-Walker Equations: We use those equations to find the a.c.f in the stationary case via:

$$\rho(k) = \sum_{i=1}^{p} \alpha_i \rho(k-i) \ \forall k > 0 \tag{17}$$

The set of Yule-Walker equations has the general solutions

$$\rho(k) = \sum_{i=1}^{p} A_i \pi_i^{|k|} \tag{18}$$

where $\{\pi_i\}$ are the roots of the auxiliary equation given by

$$y^p - \sum_{i=1}^{p} \alpha_i y^{p-i} = 0 \tag{19}$$

which is derived by ierating through the Yule-Walker equations. It is more common writing the auxiliary equation in matrix notation. A_i are chosen such that $\rho(0) = 1 \implies \sum A_i = 1$ and the first $(p-1)$ Yule Walker equations provide $(p-1)$ further restrictions on A_i. Another equivalent condition for stationarity is that the roots of the auxiliary equation all satisfy $|\pi_i| < 1$.

3 Inequalities for AR(2)-processes Stationary if

- $\alpha_1 + \alpha_2 < 1$
- $\alpha_1 - \alpha_2 > -1$
- $a_2 > -1$

if π_i are real $\implies$ the a.c.f decreases exponentially and if they are complex $\implies$ the a.c.f is a damped sinusoidal wave.

Processes with non-zero mean

- mean $\mu : \phi(B)(X_t - \mu) = Z_t$
- mean $\frac{v}{\phi(1)} : \phi(B)X_t = v + Z_t$

Mixed ARMA Models

$ARMA(p, q)$ is given by

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + Z_t + \sum_{i=1}^{q} \theta_i Z_{t-i} \tag{20}$$

$$\phi(B)X_t = \theta(B)Z_t \tag{21}$$

with $\phi(B) = 1 - \sum_{i=1}^{p} \alpha_i B^i$ and $\theta(B) = 1 + \sum_{i=1}^{q} \theta_i B^i$. For stationarity, the AR has to be stationary (MA is always stationary) and for invertibility the MA part has to be invertible (the AR part is always invertible).

Parsimony: Its often better to use a smaller ARMA rather than a large MA or a large AR.

Transforming an ARMA into a pure AR or a pure MA:
Pure MA:

$$X_t = \psi(B)Z_t \text{ where } \psi(B) = 1 + \sum_{i \geq 1} \psi_i B^i \text{ with } \psi(B) = \frac{\theta(B)}{\phi(B)} \tag{22}$$

Pure AR:

$$\pi(B)X_t = Z_t \text{ where } \pi(B) = 1 - \sum_{1 \geq 1} \pi_i B^i \text{ where } \pi(B) = \frac{\phi(B)}{\theta(B)} \tag{23}$$

such that $\psi(B)\pi(B) = 1$. How do we get the weights for the pure models?

- directly by division
- equating powers of B such as
  - $\pi(B)\theta(B) = \phi(B)$
  - $\psi(B)\phi(B) = \theta(B)$

Integrated ARMA Models

Let $W_t = \nabla^d X_t = (1 - B)^d X_t$ for some $d = 1, 2, ...$ Then use an ARMA on W_t such that we have

$$\phi(B)W_t = \theta(B)Z_t \iff \phi(B)(1 - B)^d X_t = \theta(B)Z_t \tag{24}$$

which is an $ARIMA(p, d, q)$. A series $\{X_t\}_t$ is called an $I(d)$ process if the d-th order difference series $\{(1 - B)^d X_t\}$ is stationary which makes an ARIMA a special case of an $I(d)$ process. An ARIMA(p,d,q) with $d > 0$ is clearly non-stationary since the AR operator $\phi(B)(1 - B)^d$ has d roots inside the unit circle.

General Linear Process

$\{Z_z\}_t$ is a purely process and $\{X_t\}_t$ a general linear process if

$$X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i} \tag{25}$$

which is stationary if $\sum_{i=0}^{\infty} |\psi_i| < \infty \longrightarrow$ that is nothing else than a $MA(\infty)$ process.

3 Fitting Time Series Models

Sample ACVF

$$c_k = \frac{\sum_{t=1}^{T-1}(x_t - \bar{x})(x_{t+k} - \bar{x})}{T} \tag{26}$$

which is an asymptotically unbiased: $\lim_{T\to\infty} \mathbb{E}(c_k) = \gamma(k)$ estimator (the sample autocorrelation) also and

$$\mathrm{Cov}(c_k, c_m) \approx \frac{\sum_{r=-\infty}^{\infty} (\gamma(r)\gamma(r+m-k) + \gamma(r+m)\gamma(r-k))}{T} \tag{27}$$

That tells us that sucessive values of c_k may be (highly) correlated.

Correlogram and Modelling

We also have (under weak conditions: iid series) that $\forall k \neq 0\ r_k \dot{\sim} \mathcal{N}\left(-\frac{1}{T}, \frac{1}{T}\right)$ which we can use to interpret the correlogram as we know by checking which values r_k fall outside the band $-1/T \pm 1.96/\sqrt{T} \approx \pm 2/\sqrt{T}$. The values outside are deemed significant but we expect (*Multiple Testing*) $0.05 \times \#\{r_k \text{plotted}, k > 0\}$ to be outside of the band, even for an i.i.d sequence.

Inference for the Mean

$\bar{X} = T^{-1}\sum_{t=1}^{T} X_t$ and $Var(\bar{X}) = \frac{\sigma^2}{T}\underbrace{\left(1 + 2\sum_{k=1}^{T-1}(1 - \frac{k}{T})\rho(k)\right)}_{\text{new term}}$ which cannot by estimated by plugging in the estimated versions as there are as many components as observations. Depending on $\rho(k)$ is it possible that

- $\mathrm{Var}(\bar{X}) > \sigma^2/T$
- $\mathrm{Var}(\bar{X}) < \sigma^2/T$

while typically the first one is the case, making our estimator less precise than in the iid setting.

Fitting an AR process

For a centered process:

$$(\hat{\mu}, \hat{\alpha_1}, ..., \hat{\alpha_p}) = \underset{(\mu,\boldsymbol{\alpha})}{\mathrm{argmin}}\ SS = \sum_{t=p+1}^{T} \left(x_t - \mu - \alpha_1(x_{t-1} - \mu) - ... - \alpha_p(x_{t-p} - \mu)\right)^2 \tag{28}$$

which yields the following estimators for the **AR(1)** process:

$$\hat{\mu} = \frac{\bar{x}_{(1)} - \hat{\alpha_1}\bar{x}_{(2)}}{1 - \hat{\alpha_1}} \text{ and } \hat{\alpha_1} = \frac{\sum_{t=2}^{T}(x_{t-1} - \hat{\mu})(x_t - \hat{\mu})}{\sum_{t=2}^{T}(x_{t-1} - \hat{\mu})^2} \tag{29}$$

which can be simplyfied using $\hat{\mu} = \bar{x}$ which is equivalent to regressing $(x_t - \bar{x})$ on $(x_{t-1} - \bar{x})$ by OLS. $\hat{\alpha_1}$ can be further simplyfied by replacing the denominator with $\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2$ and that further by $\sum_{t=1}^{T}(x_t - \bar{x})^2$ which shows that $\hat{\alpha_1} \approx r_1$ which makes sense as they both estimate the same quantity.

Standard Errors for Inference: $asy.SE(\bar{x}) = \sqrt{\frac{s^2}{T}\frac{1+\hat{\alpha_1}}{1-\hat{\alpha_1}}}$ and $asy.SE(\hat{\alpha_1}) = \sqrt{\frac{(1-\hat{\alpha_1}^2)}{T}}$. The test statistic under the null for $\hat{\alpha}_1$ is hence $\frac{\hat{\alpha}_1 - 0}{\frac{1}{\sqrt{T}}} = \sqrt{T}\hat{\alpha}_1$ which makes the two sided test equivalent to the correlogram check.

AR(2) process: $\hat{\mu} \approx \bar{x}$ and $\hat{\alpha}_1 \approx \frac{r_1(1-r_2)}{(1-r_1^2)}$ and $\hat{\alpha}_2 \approx \frac{(r_2 - r_1^2)}{(1-r_1^2)}$ with $\hat{\alpha}_2$ being the **partial autocorrelation coefficient** of order two which measures the linear dependence between X_t and X_{t-2} that is not accounted for by $\rho(1)$.
AR(p) process There are three methods:

- exact: least squares
- approximate: regress $(x_t - \bar{x})$ on $\{(x_{t-1} - \bar{x}), ..., (x_{t-p} - \bar{x})\}$ by OLS with standard software
- approximate:
  - substitute the sample ac coefficients into the first p Yule-Walker equations
  - p equations in p unknowns
  - solve for coefficients

where the approximate methods work well for large $T(e.g \geq 50)$ but especially the Yule-Walker approach is not so good for small samples.

Partial AC Function

Consider stationary time series $\{X_t\}_t$ where the coefficient π_k in

$$X_t = v + \pi_1 X_{t-1} + ... + \pi_k X_{t-k} + \epsilon_t \tag{30}$$

is the **partial ac coefficient** of order k which measures the left over correlation between X_t and X_{t-k} that is not accounted for by $\rho_1, ..., \rho_{k-1}$. $\pi_1 = \rho_1$ but $\pi_k \neq \rho_k \; \forall \; k > 1$. The π_k can be estimated via sample multiple regressions that are estimated *iteratively*. For example: $\pi_2 : X_t = v + \pi_1 X_{t-1} + \boldsymbol{\pi_2} X_{t-2} + \epsilon_t$ with $\pi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$.

Using ACF and PACF for model selection:

- $\pi_k = 0$ for $k > p$ in a AR(p) process (sample p.ac.f) $\longrightarrow$ sample p.ac.f can help determine the order of an AR process
- $\rho_k = 0$ for $k > q$ in a MA(q) process (sample ac.f) $\longrightarrow$ sample ac.f can help determine the order of a MA process

Fitting an MA process

MA(1):
(Bad) Idea: solve $r_1 = \frac{\hat{\theta}_1}{1 + \hat{\theta}_1^2}$ and take soultion that satifies $|\hat{\theta}_1| < 1$ which leads to inefficient estimates as the equations are quadratic instead of linear.

Feasible Idea:

1. consider tuple $(\hat{\mu}, \hat{\theta}_1)$
2. let $\tilde{z_0} = 0$: recursively get residuals from $\tilde{z} = x_t - \hat{\mu} - \hat{\theta}_1 \tilde{z_{t-1}}$ where we always know $\tilde{z}_{t-1}$
3. this yields $SS = \sum_{t=1}^{T} \tilde{z}_t^2$
4. $(\hat{\mu}, \hat{\theta}_1) = \underset{(\mu, \theta_1)}{\operatorname{argmin}} \; SS$
5. this can further be refined via backforecasting the value $\tilde{z}_0$ which is not necessary unless T is small or $|\theta_1|$ is close to 1.

MA(q) process:
Exactly analogous to the case of $q = 1$.

Fitting an ARMA process

Similar to MA models (conditional least squares) or using Maxiumum Likelihood estimates.

Model Selection

	AR(p)	MA(q)	ARMA(p, q)
ac.f.	Tails off	Cuts off after lag q	Tails off
p.ac.f.	Cuts off after lag p	Tails off	Tails off

Fitting an ARIMA Model

Get a stationary model by differencing sufficiently often and then fit an ARMA to the differenced series.

Fitting a SARIMA

A multiplicative seasonal ARIMA, which we call $SARIMA(p,d,q) \times (P,D,Q)_S$ is defined by:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D X_t = \theta_q(B)\Theta_Q(B^s)Z_t \tag{31}$$

example: lets see how a $SARIMA(1,1,1) \times (1,1,1)_4$ looks like:

$$\underbrace{(1-\phi B)}_{\text{non-s. AR}} \times \underbrace{(1-\Phi B^4)}_{\text{s. AR}} \times \underbrace{(1-B)}_{\text{non-s. I}} \times \underbrace{(1-B^4)}_{\text{s.I}} X_t = \underbrace{(1+\theta B)}_{\text{non-s. MA}} \times \underbrace{(1+\Theta B^4}_{\text{s. MA}} Z_t \tag{32}$$

Model Selection for SARIMA Models

Table 2: Behavior of the ac.f. and p.ac.f. for stationary and invertible SARMA processes

	AR(P)$_s$	MA(Q)$_s$	ARMA(P, Q)$_s$
ac.f.	Tails off at lags ks	Cuts off after lag Qs	Tails off at lags ks
p.ac.f.	Cuts off after lag Ps	Tails off at lags ks	Tails off at lags ks

Remarks:

- The values of both the ac.f. and the p.ac.f. at large non-seasonal lags $h \neq ks$, for $k = 1, 2, \ldots$, are zero
- Of course, the values do not have to be zero for small lags, such as $h = 1, 2$

Residual Analysis

A good model should produce a residual series $\{\hat{z}_t\}_{t=1}^T$ the behaves like a purely random series. One can look at the time plot, the sample a.c.f and the sample p.a.c.f of the residual series. The fitted values are the one-step ahead forecasts which can be computed directly for AR models and successively for moedls with an MA part.

Standardized Residuals: Make all the $\hat{z}_t$ have the same variance as as the z_t have the same variance by computing standardized residuals:

$$\hat{z}_{t,stand} = \frac{\hat{z}_t}{SE(\hat{z}_t - z_t)} \tag{33}$$

We prefer them when looking at the time plot or at QQ-plots.

Testing for left over autocorrelation: The $\pm 2/\sqrt{T}$ bands are conservative for residuals, which is why we check combined tests:

- Portmanteau test combines the first K sample a.c's: $Q = T\sum_{k=1}^K r_{\hat{z},k}^2$
- Ljung-Box-Pierce makes a small sample adjustment: $\tilde{Q} = T(T+2)\sum_{k=1}^K \frac{r_{\hat{z},k}^2}{T-k}$
- Durbin-Watson test: $d = \frac{\sum(\hat{z}_t - \hat{z}_{t-1})^2}{\sum \hat{z}_t^2} \approx 2(1 - r_{\hat{z},1})$ which reduces the test to a test of the first order autocorrelation coefficient.

while $Q, \tilde{Q} \dot{\sim} \chi^2_{K-p-q}$. These tests can have low power though.

Remarks on Model Building

Remark: $T \geq 50$ at the very least for ARIMA models.

Model Selection Criteria

Negative log-likelihood is equal to SS if the innovations are normally distributed.

- AIC: $AIC = -2\text{log-likelihood} + 2r$
- Bias corrected AIC: $AIC_C = -2\text{log-likelihood} + \frac{2rT}{T-r-1}$ with the latter term converging to 1 for $T \to \infty$
- BIC (SIC): $BIC = -2\text{log-likelihood} + r \log T$

the BIC penalized larger models more than the AIC (for $T \geq \exp(2)$). r are the number of freely estimated parameters (see slides).

4 Forecasting

Univariate (naive; projection) Methods

Extrapolation of Trend Curves

(i) fit a global trend (polynomial, Gompertz etc) (ii) fit ARMA model to residuals (iii) $\hat{x}_T(h)$ = estimated trend at time $T+h$ + forecasted residual $\hat{z}_T(h)$

Simple Exponential Smoothing

Based on simple updating equations based on the idea of $\hat{x}_T(1) = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j X_{T-j}$ or in recurrence form:

$$\hat{x}_T(1) = \alpha x_T + (1-\alpha)\hat{x}_{T-1}(1) = \alpha\left(x_T - \hat{x}_{T-1}(1)\right) + \hat{x}_{T-1}(1) = \alpha e_T + \hat{x}_{T-1}(1) \tag{34}$$

which we can compute recursively. The SES is an optimal forecasting method for a number of models, including the $ARIMA(0,1,1)$ which explains why SES appears to be such a robust method. How to choose α (low numbers mean that a lot of past x_t get included). We choose the value such that

$$\hat{\alpha} = \underset{\alpha}{\text{argmin}} SS = \sum_{t=2}^{T} e_t^2 \tag{35}$$

while the SS is quite flat around the minimum quite often so the choice of α does not matter that much.

ES with Trend: Holt Procedure (Two-parameter version of ES)

Note that $\hat{x}_t$ can be interpreted as the local mean level and the SES could therefore be rewritten as $L_t = \alpha x_t + (1-\alpha)L_{t-1}$ and if including a trend term T_t we get

$$L_t = \alpha x_t + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{36}$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \tag{37}$$

which we call the updating equations in recurrence form and we have $\hat{x}_t(h) = L_t + hT_t$ and the two parameters (α, γ) are also found by minimizing SS.

ES with Trend and Seasonality: Holt-Winters Procedure

I_t is the seasonality where it is important whether the seasonality is additive or multiplicative. Multiplicative case:

$$L_t = \alpha(x_t / I_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{38}$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \tag{39}$$
$$I_t = \delta(x_t / L_t) + (1-\delta)I_{t-s} \tag{40}$$

such that we have $\hat{x}_t(h) = (L_t + hT_t)I_{t-s+h}$

Box-Jenkins Procedures

Idea:

1. plug current and past X_t and z_t into the model
2. set future z_t equal to their expected value 0
3. replace future x_t by their previous predictions
4. the forecast X_{T+1} is then the conditional expectation of X_{T+1} given all current and past information

Using the actual model

see slides 5 page 25

Using the MA representation

let $X_{T+h} = Z_{T+h} + \psi_1 Z_{T+h-1} + ...$ such that $\hat{x}_T(h) = \sum_{j=0}^{\infty} \psi_{h+j} Z_{T-j}$ as the future innovations cannot be included. We then have that the h-step forecast error is $e(h) = Z_{T+h} + \psi_1 Z_{T+h-1} + ... - \sum_{j=0}^{\infty} \psi_{h+j} z_{T-j} = Z_{T+h} + \psi_1 Z_{T+h-1} + ... + \psi_{h-1} Z_{T+1}$ and $\text{Var}(e(h)) = (1 + \psi_1^2 + ... + \psi_{h-1}^2)\sigma_z^2$

Using the AR representation

let $X_{T+h} = \pi_1 X_{T+h-1} + ... + Z_{T+h}$ such that the prediction can be computed recursively by

$$\hat{x}_T(h) = \pi_1 \hat{x}_T(h-1) + ... + \pi_{h-1} \hat{x}_T(1) + \sum_{j=0}^{\infty} \pi_{h+j} x_{T-j} \tag{41}$$

Prediction Intervals

$$\hat{x}_T(h) \pm z_{1-\alpha/2} \sqrt{\widehat{Var}[e_T(h)]} \tag{42}$$

which assumes that the point forecast is (nearly unbiased) and that the forecast error follows a normal distribution. How do we estimate $\sqrt{\widehat{\text{Var}}[e_T(h)]}$?

- ES: determined by smoothing parameters and the Variance of the one-step-ahead forecast error σ_ϵ^2 which for SES is given by: $\text{Var}[e_T(h)] = [1 + (h-1)\alpha^2]\sigma_\epsilon^2$ with $\hat{\sigma}_\epsilon^2 = (T-1)^{-1} \sum_{t=2}^{T} e_t^2$. That can be generalized for the other SE approaches.
- ARMA models: using the MA(∞) representation, we have $(1 + \psi_1^2 + ...)\sigma_Z^2$ where we just use the estimates for the coefficients and the variance.
- General forecasting model: fit model to data, compute in sample forecast errors at step h and then take the sample variance of these in-sample forecast errors

Problem: Forecast intervals generally to narrow as the estimator for the variance might be biased downwards or because the distribution of the forecast errors is not normal (fat tails)

Multivariate Methods

Not necessarily better.

5 Bivariate Processes

We make T observations of two variables at unit time intervals over the same period and denote those observations $(x_1, y_1), ..., (x_T, y_T)$ and assume that they are finite realizations of the discrete-time bivariate stochastic process (X_t, Y_t). We define the cross-covariance function ccv.f by

$$\text{Cov}(X_t, Y_{t+k}) = \mathbb{E}[(X_t - \mu_X)(Y_{t+k} - \mu_Y)] = \gamma_{XY}(k) \tag{43}$$

which only depends on k due to stationarity. R defines it as $\text{Cov}(X_{t+k}, Y_t) = \gamma_{XY}^\star(k)$ while we have $\gamma_{XY}(k) = \gamma_{XY}^\star(-k)$. Note that the ccv.f is not even but we have $\gamma_{XY}(k) = \gamma_{YX}(-k)$. **Cross Correlation:** $\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sqrt{\gamma_X(0)\gamma_Y(0)}}$ and the properties hold:

- $\rho_{XY}(k) = \rho_{YX}(-k)$
- $|\rho_{XY}(k)| \leq 1$
- $\rho_{XY}(0)$ is usually not 0

These quantities are estimated via their sample functions (see slides). These estimators are asymptotically unbiased and consistent but: *However, it can also be shown that estimators at neighboring lags are themselves autocorrelated. Furthermore it can be shown that the variances of sample cc.'s depend on the ac.f.'s of the two components. In general, the variances will be inflated.* That means that uncorrelated series can give rise to large cc. coefficients which are spurious in that they solely arise from ac.'s within the two series. For two uncorrelated series, one being white noise, it can be shown that

$$\mathbb{E}[r_{XY}(k)] \approx 0 \tag{44}$$

$$\text{Var}[r_{XY}(k)] \approx \frac{1}{T} \tag{45}$$

Assessing Significance of CC.F

Checking if X_t and Y_t are linearly dependent at certain lags: Fit ARMA models to both time series and obtain the residuals, compute the sample cc.f from the two residual series and watch out for coefficients that fall outside of the $\pm 2/\sqrt{T}$ bands.

6 Multivariate Time Series

let $\{\boldsymbol{X_t}\} = \{(X_{1t}, X_{2t}, ..., X_{mt})^T\}$ and mean function $\boldsymbol{\mu_t} = \boldsymbol{\mu(t)} = \mathbb{E}(\boldsymbol{X_t}) \in \mathbb{R}^m$ and variance function $\Sigma_t = \Sigma(t) = \text{Var}(\boldsymbol{X_t}) \in \mathbb{R}^{m\times m}$ where $\Sigma_{ij} = \text{Cov}(X_{it}, X_{jt})$ and $\Sigma_{ii} = \text{Var}(X_{it})$. We further have the cross-covariance function given by $\Gamma_{t,t+k} = \Gamma(t, t+k) = \Gamma(\boldsymbol{X_t}, \boldsymbol{X_{t+k}}) \in \mathbb{R}^{m\times m}$ where $\Gamma_{ij}(k) = \gamma_{ij}(k) = \mathbb{E}[(X_{it} - \mu_i)(X_{j,t+k} - \mu_j)]$ if we have stationary time series. Stationarity:

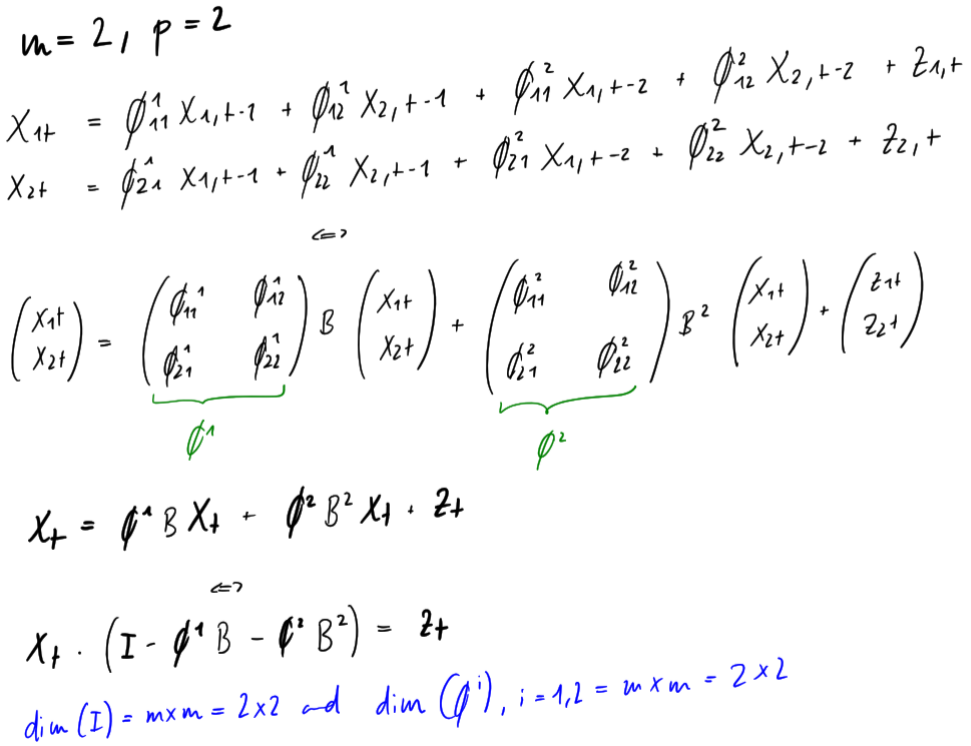
- $\boldsymbol{\mu(t)} = \boldsymbol{\mu}\ \forall t$
- $\Sigma(t) = \Sigma\ \forall t$
- $\Gamma(t, t+k) = \Gamma(k)\ \forall t$ which we call the covariance matrix function with $\Gamma(-k) = \Gamma(k)^T$

Correlation matrix function $P(k)$ with $\rho_{ij}(k) = \gamma_{ij}(k)/\sigma_i\sigma_j$ with $\sigma_i^2 = \gamma_{ii}(0) = \text{Var}(X_{it})$. We estimate these via their sample counterparts.

VAR Models

$$\boldsymbol{X_t} = \Phi\boldsymbol{X_{t-1}} + \boldsymbol{Z_t} \Longleftrightarrow (\boldsymbol{I} - \Phi B)\boldsymbol{X_t} = \boldsymbol{Z_t} \tag{46}$$

example: let $m = 2$ be the number of components and $p = 2$ the order of the VAR model, i.e the number of lags.



meaning a $VAR(p)$ ca be written compactly as $\Phi(B)\boldsymbol{X_t} = \boldsymbol{Z_t}$ with $\Phi(B)$ being the polynomial $I - \Phi^1 B - ... - \Phi^p B^p$ where $^{\{1,...,p\}}$ index the coefficient matrices and are not to be confused for powers.

Conditions for Stationarity

The roots of $\det(\Phi(x)) = |I - \phi_1 x - \phi_2 x^2 - ... - \phi_p x^p| = 0$ must all lie outside the unit circle. **Equivalent:** The roots of $\det(\phi - \lambda I)$ which are the eigenvalues must lay inside the complex unit circle.

Non-zero mean

1. centered variables: $\Phi(B)(\boldsymbol{X_t} - \boldsymbol{\mu}) = \boldsymbol{Z_t}$
2. intercept in autoregression: $\Phi(B)\boldsymbol{X_t} = \boldsymbol{v} + \boldsymbol{Z_t}$

while the second approach is more common here.

Estimation

OLS per equation or all equations jointly using maximum likelihood methods.

Model Selection

AIC, AIC_C or BIC while the BIC tends to work better for VARs.

Model Diagnostics

For each component: compute residual series and check sample ac.f and p.ac.f
Combined over all components to test for left-over serial correlation: check slides.

Impulse Response Functions

Lets look at the VMA(∞)-representation of the VAR(p) model:

$$\boldsymbol{X_t} = \boldsymbol{\Theta(B) Z_t} = \sum_{k=0}^{\infty} \boldsymbol{\Theta_k Z_{t-k}} \text{ with } \boldsymbol{\Theta_k} \in \mathbb{R}^{m \times m} \tag{47}$$

such that the function $k \to \boldsymbol{\Theta_{k,ij}}$ is the impulse response function from X_j on X_i. That is not sensible as $\boldsymbol{Z_t}$ does not have to be diagnoal which means that the c.p interpretation does not hold. We use the orthogonal version for that reason:

$$\boldsymbol{X_t} = \tilde{\boldsymbol{\Theta}}(B) \boldsymbol{V_t} = \sum_{k=0}^{\infty} \tilde{\boldsymbol{\Theta}} \boldsymbol{V_{t-k}} \tag{48}$$

which gives us the orthogonal IR. We use the cholesky decomposition of $\sum_Z = PP^T$ and that $\boldsymbol{V_t} = P^{-1}\boldsymbol{Z_t}$ and $\tilde{\boldsymbol{\Theta}}_{\boldsymbol{0}} = \boldsymbol{\Theta_0}$ and $\tilde{\boldsymbol{\Theta}}_{\boldsymbol{k}} = \boldsymbol{\Theta_k} P$. We prefer the orthogonal versions in applied work as they allow a c.p interpretation.
*Forecasting Multivariate models not always better than the univariate ones, especially as there are many more parameters to estimate, making the model less precise, which is not a problem for fitting the observed data, but for forecasting (overfitting on training set).

Adding a Deterministic Time Trend

$$x_{1t} = v_1 + \delta_{1t} + \phi_{11} x_{1,t-11} + \phi_{12} x_{2,t-1} + z_{1t} \tag{49}$$
$$x_{2t} = v_1 + \delta_{2t} + \phi_{21} x_{1,t-11} + \phi_{22} x_{2,t-1} + z_{2t} \tag{50}$$

or generalize and add a polynomial trend.

Co-Integration

We have e.g two non-stationary $I(d)$ time series but a linear combination of the two is $I(b)$ with $b < d$. We then call them **co-integrated of order d-b**.

7 Intervention Analysis

We limit the impacts to steps and spikes (impulse).

Impulse Impact

$$X_t = \begin{cases} \tilde{X}_t \ \forall \ t < t^\star \\ \tilde{X}_t + \delta \ \forall \ t = t^\star \end{cases}$$

where δ denotes the one-time shock. We model that by adding an impulse dummy $D^\star$ to the model, e.g:

$$D_t = \begin{cases} 0 \ \forall \ t \neq t^\star \\ 1 \ \forall \ t = t^\star \end{cases}$$

Step Impact

$$X_t = \begin{cases} \tilde{X}_t \ \forall \ t < t^\star \\ \tilde{X}_t + \delta \ \forall \ t \geq t^\star \end{cases}$$

which we model by

$$D_t = \begin{cases} 0 \ \forall \ t \leq t^\star \\ 1 \ \forall \ t \geq t^\star \end{cases}$$

We can also use that idea to deal with outliers. We treat them as an impulse impact and therefore implicitely remove the outliers. In the **Multivariate Case** we just add a dummy to each component-equation where the dummies are typically of the same nature, although thats not strictly necessary.

8 Nonlinear models

TAR Models

Used to model rises and falls at different speeds and are structured as follows: You are in one of several AR(p) regimes depending on the value of X_{t-d} with d being a known time lag. If we chose $p = 1, d = 1$, we get:

$$X_t = \begin{cases} v_1 + \alpha_1 X_{t-1} + \sigma_1 Z_t \text{ if } X_{t-1} < r \\ v_2 + \alpha_2 X_{t-1} + \sigma_2 Z_t \text{ if } X_{t-1} \geq r \end{cases}$$

GARCH Models

ARCH(1) Models

We are looking at the residual series $\{Y_t\}_t$ derived from $\{X_t\}_t$ by removing trend, seasonalities and linear short term correlation effects may also have been removed. We then define

$$Y_t = \sigma_t \epsilon_t \tag{51}$$

with $\{\epsilon_t\}_t$ being mean-zero purely random with $\sigma_\epsilon = 1$ and σ_t being the conditional standard deviation. We model σ_t by an ARCH(1) as follows

$$\sigma_t^2 = \gamma + \alpha y_{t-1}^2 \text{ with } \gamma > 0 \text{ and } \alpha \ \in [0, 1) \tag{52}$$

which means that conditional on y_{t-1} the conditional variance σ_t is deterministic. $\{Y_t\}_t$ is stationary as we have: $\mathbb{E}(Y_t) = \mathbb{E}((\gamma + \alpha Y_{t-1}^2)\epsilon_t) = 0$ by the fact that ϵ_t is purely random. We have $\text{Var}(Y_t) = \mathbb{E}(Y_t^2) - \mathbb{E}(Y_t)^2 = \mathbb{E}(Y_t^2) = \mathbb{E}[(\sigma_t \epsilon_t)] = \mathbb{E}[(\gamma + \alpha Y_{t-1}^2)\epsilon_t^2] = \mathbb{E}([\gamma + \alpha Y_{t-1}^2]^2)\mathbb{E}[\epsilon_t^2]$ with the latter being 1 per definition. We then have $\mathbb{E}[Y_t^2] = \text{Var}(Y_t)$ which we can solve to get $\sigma_Y^2 = \frac{\gamma}{1-\alpha}$ which means we have unconditional stationarity.

GARCH Models

let $\sigma_t^2 = \gamma + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$ with $\gamma > 0, \alpha, \beta \geq 0$ and $\alpha + \beta < 1$ We now also regress σ_t^2 on its past while still remaining non-stochastic, conditional on Y_{t-1} and σ_{t-1}. The process is again unconditionally stationary with $\sigma_Y^2 = \frac{\gamma}{1-\alpha-\beta}$ Properties:

- GARCH models create volatility clustering
- $\{Y_t\}_t$ is uncorrelated while $\{Y_t^2\}_t$ is positively correlated
- GARCH models generate fat tails with the tails being fatter than the ones of the innovations $\{\epsilon_t\}_t$

We estimate GARCH models via maximum likelihood using psuedo maximum likelihood.

Incorporating a Leverage Effect

Negative returns lead to larger 'shocks' to financial markets compared to positive returns: we call that the leverage effect.The GARCH model cannot cpature that as the return of yesterday only enters squared. We can use leverage GARCH or asymmetric GARCH models for that.

9 Inference for Time Series

Inference for the Mean

we are in the case of a stationary univariate time series: We have as usual the natural estimator for the mean given by $\bar{X} = T^{-1}\sum_{t=1}^{T} X_t$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{T}\left(1 + 2\sum_{t=1}^{T-1}(1-\frac{k}{T})\rho_k\right) = T^{-1}\left(\gamma_0 + 2\sum_{t=1}^{T-1}(1-\frac{k}{T})\gamma_k\right)$ does not work if we just plug in s^2 for σ^2 and r_k for ρ_k and c_k for γ_k. Because depending on $\rho(\cdot)$ the following can happen:

- $\text{Var}(\bar{X}) > \frac{\sigma^2}{T}$ (For AR(1) Case: $\alpha > 0$)
- $\text{Var}(\bar{X}) < \frac{\sigma^2}{T}$ (For AR(1) Case: $\alpha < 0$)

For AR(1) processes, we have $\text{Var}(\bar{X}) \approx \frac{\sigma^2}{T}\frac{1+\alpha}{1-\alpha}$. If $\alpha > 0$, the time series estimator is bigger than the conventional one and the inference based on the standard estimator would be too liberal (anti-conservative). In reality, the case with the positive coefficient is dominant.

Limiting Variance and CLT

Lets define the following:

$$\sigma^2_\infty = \lim_{T\to\infty} \text{Var}(\sqrt{T}\bar{X}) = \gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k \tag{53}$$

Using the CLT under some regularity conditions, we have asymptotically that

$$\bar{X} \dot{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2_\infty}{T}\right) \tag{54}$$

meaning we need a reliable estimator for σ^2_∞ to make inference using $SE(\bar{X}) = \frac{\hat{\sigma^2}}{\sqrt{T}}$. The problem is that we cannot just plug c_k for γ_k as the estimates c_k are very unreliable. We instead use $\hat{\sigma}^2_\infty = c_0 + 2\sum_{k=1}^{T-1} w_k c_k$ with w_k being a kernel weight funcion with the properties $w_k \in [-1, 1]$ and $\lim_{k\to\infty} w_k = 0$ such that the unreliable estimates c_k for large k get down-weighted to zero.
What needs to be fulfilled:

- T cannot be small and the slower γ_k dies down to zero, the larger T has to be
- if important contributions c_k are downweighted, the resulting estimator $\hat{\sigma}^2_\infty$ can be biased towards 0.

General Weight Functions

$$\omega_k = K\left(\frac{k}{S}\right) \tag{55}$$

with K being a kernel with

- $K(-x) = K(x)$
- $K(0) = 1$
- $\lim_{x\to\infty} K(x) = 0$
- typically $K(x) = 0 \;\forall\; x > 0$
- S is the bandwidth which depends on $\gamma(\cdot), K(\cdot)$ and T

Multivariate Regression

This is the equivalence of the above but for dimensions $d \geq 2$. Under nonrestrictive regularity conditions, we have:

$$\sqrt{T}\left(\hat{\beta}_T - \beta\right) \xrightarrow{d} \mathcal{N}(0, \Sigma) \tag{56}$$

with Σ being a p$\times p$ matrix that is positive definite. $\sum$ has to bestimated consistently to make inference.

$$\Sigma = \lim_{T\to\infty}\left(T^{-1}\sum_{t=1}^{T} X_t X_t'\right)^{-1} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_s X_s (\epsilon_t X_t)'\right]\left(T^{-1}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \tag{57}$$

where we only have to estimate $J_\infty = \lim_{T\to\infty} J_T$ with $J_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_s X_s (\epsilon_t X_t)'\right]$. If we define $V_t = \epsilon_t X_t$ and note that $\mathbb{E}(V_t) = 0$ we can rewrite J_T as follows:

$$J_T = \sum_{k=-T+1}^{T-1}(1 - \frac{|k|}{T})\Gamma_V(k) \tag{58}$$

with $\Gamma_V(k)$ being the covariance matrix function of the process $\{V_t\}_t$. Using the kernel method again, we get

$$\hat{J}_\infty = \frac{T}{T-p} \sum_{k=-T+1}^{T-1} K\left(\frac{k}{S}\right) C_{\hat{V}}(k) \tag{59}$$

with $T/(T-p)$ being a finite sample bias correction as we have to estimate β first to get the residuals. We use a data-dependent method to pick a good bandwidth as a function of the data. Finally, we can estimate the limiting covariance matrix as

$$\hat{\Sigma}_T = \left(T^{-1}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \hat{J}_\infty \left(T^{-1}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \tag{60}$$

That estimator is called a HAC estimator, which means its heteroskedasticity and auto-correlation consistent.

Inference for Parameters

Let $\theta = a'\beta$ with $a \in \mathbb{R}^p$. We then have $SE(\hat{\theta}) = \sqrt{\frac{a'\hat{\Sigma}_T a}{T}}$. We can use the quantiles of a normal distribution to construct confidence intervals.

General F Test

Check Slides 10

10 Unit Roots and Co-Integration

Testing for a Unit Root

AR(1) Case

Let $X_t = v + \alpha X_{t-1} + Z_t$ where we want to test for:

$$H_0: \text{ unit root vs. } H_1: \text{ stationary}$$

where we look at the regression $\nabla X_t = v + (\alpha - 1)X_{t-1} + Z_t$ where $\alpha - 1 = \delta$ so we can test:

$$H_0 : \delta = 0 \text{ vs. } H_1 : \delta < 0$$

The limiting distribution of the test statistic is not $\mathcal{N}(0,1)$ though as we have a nonstationary time series under H_0. We use the non-standard distribution derived by Dickey and Fuller.

General AR Case: ADF Test

Let $X_t = v1 + \alpha_1 X_{t-1} + ... + \alpha_{t-p-1} X_{t-p-1} + Z_t$ where we want to test for:

$$H_0: \text{ at least one unit root vs. } H_1: \text{ stationary}$$

which we do by looking at the following regression $\nabla X_t = v + \delta X_{t-1} + \gamma_1 \nabla X_{t-1} + ... + \gamma_p \nabla X_{t-p} + Z_t$ and testing $H_0 : \delta = 0$ vs. $H_1 : \delta < 0$. Run an OLS regression for the ∇X_t regression and derive the t-statistics for δ and use the critical values provided by Dickey and Fuller. As we do not know the order p, we apply the AIC criterion to the regression and check which value p minimizes it.

General Case

For a general time series, having at least one unit roots corresponds to the series being integrated of order one, i.e $\{X_t\}_t \sim I(1)$. As we know that any $I(d)$ series can be approximated by a suitable AR process for $d = 1, 2$, we simply use the ADF test from above again. We then also use the AIC criterion for p or some prior beliefs.

Deterministic Time Trend

If we think that a series is stationary around a deterministic time trend, we test the ADF but include βt and proceed as known. The critical values change though (slides 10/47).

Testing for Co-Integration

Two $I(1)$ series are co-integrated if a linear combination of them is $I(0)$.We have $\{X_t\}_t$ and $\{Y_t\}_t$ who are both $I(1)$ but $U_t = Y_t - \theta X_t$ is $I(0)$ but not necessarily mean-zero. We call θ the co-integrating coefficient that quantifies the long-run relationship. Let $\alpha = \mathbb{E}[Y_t - \theta X_t]$. As the series are co-integrated, the deviations from the equilibrium are stationary, meaning they should hoover around their equilibrium. If they are not co-integrated, their deviations (even if $\mathbb{E}[Y_t - \alpha - \theta X_t] = 0$) $U_t = Y_t - \alpha -_t$ are not stationary, but rather $I(1)$ meaning the series wander arbitrarily far from equilibrium over time. 3 cases:

- Case 1: both α and θ are known
- Case 2: θ is known, α not: estimate $\alpha = \bar{X}_t$. Compute $U_t = Y_t - \theta X_t$ and apply the **ADF** test to the $\{U_t\}_t$ series (do not include a time trend). H_0 states that they are no cointegrated.
- Case 3: Both α, β are unknown: regress Y_t on X_t plus a constant. More general: Estimate $Y_t = \alpha + \theta X_t + \eta_t$ via OLS (no HAC inference), then compute $\hat{U}_t = Y_t - \hat{\theta} X_t$ and then apply the **ADF** to the $\{U_t\}_t$ series.

As we once use $\hat{U}_t$ and once U_t, the critical values of the ADF test must be adjusted. The critical values also depend on whether there is a drift; see slides.

VEC Models

We have two $I(1)$ series and we want in principle a VAR. If the two series are co-integrated, we can add the past **error-correction term** as a further regressor to get a potentially more informative model, which works because the EC term is also stationary. More generally:

- model ∇X_t with $\nabla X_{t-1}, ...,$ and ∇Y_{t-1} and also the past EC term$Y_{t-1} - \theta X_{t-1}$ where we use $\hat{\theta}$ if θ is not known.