

Statistical test cheat sheet

→ White test (Breusch - Pagan):

H_0 : Homoskedastic

H_a : Heteroskedastic

→ Breusch - Godfrey (BG) LM test:

H_0 : No AR in variable of orders 1-6

H_a : Some AR in variable of orders 1-6

→ Jarque-Bera (JB) test for normality:

H_0 : normal distribution

H_a : non-normal distribution

→ Lagrange multipliers (LM) for ARCH errors:

H_0 : no AR in squared error terms

H_a : some AR in squared error terms

→ Ljung - Box & Box-Pierce tests:

H_0 : autocorrelations are equal to zero

H_a : at least one AR order is not equal to zero

→ RESET test:

H_0 : correct functional form

H_a : Incorrect functional form.

1

Classical time series info

There are 4 components to time series:

a) Trend ↘ ↙ → General LR A in data (may or may not be linear)

b) Seasonal (< 1 year)

↳ additive ↗ → Fluctuations are constant

↳ multiplicative ↗ → Fluctuations proportionate to level

c) cyclical

↳ Quasi regular fluctuations around some LT trend > 1 year



d) Random (i.e. everything unexplainable)

How to identify components?

↳ Smoothing:

↓
Random is removed
by smoothing

↳ Decomposition

↓
TS broken into
components of
systemic parts

2

CAPM model & extensionsClassic CAPM

$$R_{it} - R_{st} = \alpha + \beta(R_{mt} - R_{st}) + \epsilon$$

Given asset return ↑ Risk free asset return ↑
 excess return on asset ↓

Market return ↑ Risk free asset return ↑
 excess return on mkt ↓

The risk of asset i → $\alpha = E(y) - \beta E(x)$

$$\beta = \frac{\sigma_{xy}}{\sigma_x^2}$$

$\beta > 1$ = aggressive
 $\beta = 1$ = benchmark
 $0 < \beta < 1$ = conservative
 $-1 < \beta < 0$ = imperfect hedge
 $\beta = -1$ = perfect hedge

CAPM extensions

5 factors

$$R_{it} - R_{st} = \alpha - \beta_1(R_{mt} - R_{st}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \epsilon$$

Small minus big
 ↳ diff. b/w avg. return b/w small & big portfolios

High minus low
 ↳ diff. b/w avg. return b/w value & growth ports.

Robust minus weak
 ↳ diff. b/w avg. return on robust & weak portfolios

Cons. minus Agg.
 ↳ Avg. return on conservative ports vs. aggressive ports.

Note: when comparing classical vs. updated, it's likely 3/5 factors will be sig. due to omitted variable issues.

However, if β_i does not change, the F-test is not binding.

2

Linear regression assumptions

TSLR1: estimated from random sample ($T \geq 2$) of statistically independent pair of observations

TSLR2: each random error has zero conditional expected value
 $\hookrightarrow E(\varepsilon_t) = 0$

TSLR3: The conditional variance is homoskedastic
 $\hookrightarrow V(\varepsilon_t) = \sigma^2$

TSLR4: ε_t is serially uncorrelated \Rightarrow random errors are uncorrelated ($t \neq t'$)
 $\hookrightarrow E(\varepsilon_t, \varepsilon_{t'}) = 0$

TSLR5: In the sample (\propto pop.) the independent variable is not constant

TSLR6: the random errors are uncorrelated w/ independent variable
 $\hookrightarrow \text{Cov}(\varepsilon_t, x_t) = E(\varepsilon_t, x_t) = 0$

TSLR7: The random errors are normally distributed.
 $\hookrightarrow \varepsilon_t \sim N(0, \sigma^2)$

TSLR8: The random errors are stationary & weakly dependent.

Notes on Skewness & kurtosis

Skewness

- \hookrightarrow If normally distributed = 0
- \hookrightarrow If positive = right skewed
- \hookrightarrow If negative = left skewed

Kurtosis:

- \hookrightarrow If normally distributed = 3 (mesokurtic)
- \hookrightarrow If leptokurtic > 3 (fat-tailed)
- \hookrightarrow If platykurtic < 3 (thin-tailed)

3

Stationarity, white noise & (P)ACF

- A TS is **weakly stationary** if its mean is constant & autocovariances do not vary over time
 - A TS is **strictly stationary** if its properties are unaffected by a change in the time origin i.e. invariant w.r.t. time.
 $\hookrightarrow \text{Cov}(y_t, y_{t-h}) = E[(y_t - \mu_t)(y_{t-h} - \mu_{t-h})]$
- white noise is the simplest stationary process
- $$\hookrightarrow E(\varepsilon_t) = 0 \rightarrow \text{Var}(\varepsilon_t) = 0 \rightarrow \rho = 0 \quad (k > 0) \rightarrow \varepsilon_t \sim \text{iid}(0, \sigma^2)$$

ACF: Plots the average correlation between TS and previous values for different lag length

PACF: Same as above but measures r/ship to specific lag lengths

ROT for lag length: $s = m(10, \frac{t}{5})$

To test if autocorrelation is present we have two tests

Box-Pierce

$$Q_{BP} = T \sum_{k=1}^s r_k^2$$

Ljung-Box

$$Q_{LB} = T(T+2) \sum_{k=1}^s \frac{r_k^2}{T-k}$$

H_0 : no autocorrelation of order 1-6
 H_a : at least one autocorrelation order $\neq 0$

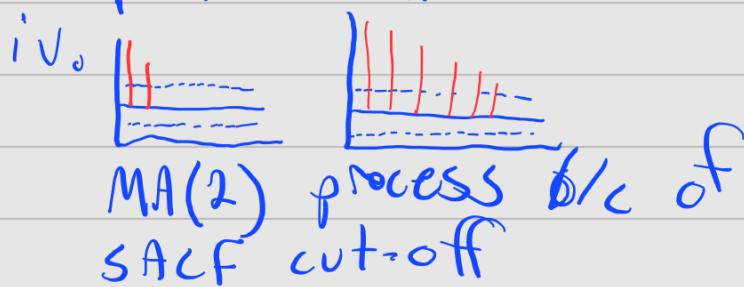
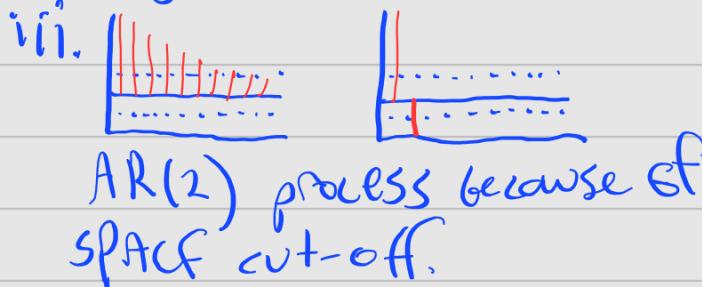
3

Box-Jenkins methodology

Step 1: model selection

- ↳ Plot data & correlograms
- ↳ If non-seasonal, use the following R/T's to determine ARMA(p,q):
 - i. If no sig. ACF or PACF spikes then it's white noise & not an ARMA process

- ii. If ACF decreases slowly & lin. through zero, or a wave-like pattern ~ then not stationary (try differencing)



- v. Both ACF & PACF's converge to zero following some exp. or damped sine wave
 ∴ ARMA(p, q) model.

↳ Need to do more analysis (auto.arima()) for p & q values

Step 2: model estimation

- ↳ Estimate the specified ARIMA model using the Arima() function from forecast package

Step 3: diagnostic checks

- ↳ If in step 1 you selected more than one model use model selection criterion to make a final decision (the smaller the better)

- ↳ The residuals should behave as white noise (i.e. show ACF & PACF for residuals)

- ↳ Check for normality as well (espec. if t is small)

- ↳ If you're satisfied do one final check compare model selection criteria against an overparameterized model (i.e. AR(p+1))

- ↳ If you've chosen the "correct" model then this term should be insignificant.

Key ARMA formulae

MA(1): $y_t = \theta, \epsilon_{t-1}, \epsilon_t \sim WN(0, \sigma^2)$

$$E(y_t) = 0$$

$$Var(y_t) = \sigma^2(1 + \theta^2)$$

every MA(1) but
the first is zero. memory
of MA(1) is only 1 period long

$$y_{T+h} = \epsilon_{T+h} + \theta, \epsilon_{T+h-1}, h > 1$$

$$\hookrightarrow E(y_{T+h}) = E(\epsilon_{T+h}) + \theta E(\epsilon_{T+h-1}) = 0$$

$$c_{T+h} = y_{T+h} - E(y_{T+h})$$

$$\hookrightarrow E(c_{T+h}) = E(\epsilon_{T+h} + \theta \epsilon_{T+h-1}) = 0$$

$$Var(c_{T+h}) = \sigma^2(1 + \theta^2)$$

AR(1): $\varphi_1 y_{t-1} + \epsilon_t, |\varphi| < 1, \epsilon_t \sim WN(0, \sigma^2)$

$$E(y_t) = 0, Var(y_t) = \frac{\sigma^2}{1 - \varphi_1^2}$$

$$y_{T+h} = \varphi_1 y_{T+h-1} + \epsilon_{T+h}, h > 1$$

$$\hookrightarrow E(y_{T+h}) = \varphi_1 E(y_{T+h-1}) + E(\epsilon_{T+h}) = \varphi_1^h y_T$$

$$c_{T+h} = y_{T+h} - E(y_{T+h}) = \varphi_1^{h-1} \epsilon_{T+1} + \varphi_1^{h-2} \epsilon_{T+2} + \dots + \varphi_1 \epsilon_{T+h-1} + \epsilon_{T+h}$$

$$\hookrightarrow E(c_{T+h}) = 0$$

$$Var(c_{T+h}) = \sigma^2 \frac{1 - \varphi_1^{2h}}{1 - \varphi_1^2} \xrightarrow{h \rightarrow \infty} \frac{\sigma^2}{1 - \varphi_1^2}$$

4 Deterministic & stochastic trends Φ_1 = characteristic root

$$y_t = \Phi_1 y_{t-1} + \varepsilon_t, \text{ if:}$$

	$ \Phi_1 < 1$	$ \Phi_1 = 1$	$ \Phi_1 > 1$
$E(y_t)$	0	$\pm y_0$	$\pm \infty$
$\text{Var}(y_t)$	$\frac{\sigma^2}{1-\Phi_1^2}$	$\sigma^2 t$	∞
$\text{Cov}(y_t, y_{t-k})$	$\Phi_1^k \frac{\sigma^2}{1-\Phi_1^2}$	$\sigma^2(t-k)$	$\pm \infty$

∴ Mean, variance & autocovariance time independent
only if $|\Phi_1| < 1$
 ↳ All approach infinity if > 1 . Although mean
is stationary for $|\Phi_1| = 1$, other parts approach infinity

Deterministic trends:

$$y_t = y_{t-1} + \varepsilon_t \rightarrow \text{pure random walk}$$

$$y_t = a_0 + y_{t-1} + \varepsilon_t \rightarrow \text{random walk w/ drift}$$

$$y_t = a_0 + y_{t-1} + d_2 t + \varepsilon_t \rightarrow \text{random walk w/ drift & linear trend}$$

Note: random walks are difference stationary processes

↳ If stationarity achieved after n differences, series
is integrated of order n (i.e. $I(n)$)

4

Unit root tests

When conducting these tests, the structure of the time series determines which "model" we use:

	ADF	KPSS
Model 1 → Pure RW	"none"	"mu"
Model 2 → RW w/ drift	"drift"	"tau"
Model 3 → RW w/ drift & trend	"trend"	fluc. const some trend

Augmented Dickey-Fuller (ADF) test

H_0 : there is a unit root, H_a : there is no unit root
 ↳ $I(1)$ ↳ $\pm(0)$

→ why "augmented"? Normal test assumes E_t is not autocorrelated (bad for AR processes), ADF augments test regression w/ first-differenced lags of y .

→ Process:

Step 1: choose relevant model from above

Step 2: Conduct test, if tau test stat is smaller than a given critical value, you can reject H_0 at that sig. level.

Step 3: Check the autocorrelation of residuals. If there's many/patterned ACF lags then try increasing lags (take caution if you can't eliminate)

Step 4: Conduct same test on differenced series, to make sure the variable isn't $I(2)$. Test only checks for a single unit root.

KPSS test

H_0 : there is no unit root H_a : there is a unit root

→ Only diff. = step 3: If test stat is smaller than a given critical value you can maintain H_0 at that level.

4

Asset price bubbles & tests

$$P_t(1+R) = E[P_{t+1} + D_{t+1}] = \sum_{i=1}^h \beta^i E_t(D_{t+i}) + \beta^h E_t(P_{t+h})$$

asset price
 constant) cond. expec.
 r.f. IR info) value using available info
 Dividend payment

PV of asset + Bubble (" β^h ")

IF no bubble, $B_t = 0$, otherwise:

$$\beta_t = \beta E_t(P_{t+1}) \therefore E_t(B_{t+h}) = (1+R)^h \beta_t = \text{exponential for } h > 1$$

tests

→ no bubble ↗ bubble

→ Right-tail ADF = $H_0: \phi = 1, H_a: \phi_1 > 1$

↳ issue: when bubble collapses during sample, low prob. of rejecting H_0

→ Supremum ADF (SADF) = same H_0 & H_a as above

↳ Based on recursive calcs of ADF test stats with an expanding sample window. Test stats are calculated from progressive sub-samples, SADF = the largest (supremum) of them.

→ Generalised SADF (GSADF) = H_0 : same, H_a : mult. period. collapsing bubbles

↳ The sample window is gradually increased from $t=0$ one at a time & SADF is performed at each starting point. → GSADF is their supremum.

↳ If any test rejects H_0 , you can use date stamping to start/end point of bubbles.

↳ start point is the first time where test stat > crit value → ends at first failure.

5

Forecasting theory & evaluation

Period Errors

1	e_1	ex-post in sample forecast
2	e_2	
\vdots	\vdots	
t^*	e_{t^*}	
t^*+1	e_{t^*+1}	
\vdots	\vdots	
T	e_T	ex-post out of sample forecast
T+1	-	ex-ante forecast
\vdots	\vdots	
$T+h$	-	

Mean Error:

$$ME = \frac{1}{T-t^*} \sum_{t=t^*+1}^T e_t$$

+ very simple, preserves UoM
& shows if avg. Forecast is above (>0) or below (<0) true values

- assigns same weight to each error, small or large, & pos/neg errors might offset each other

Note: ME, MAE & RMSE are

all sensitive to UoM - no good for comparing TS w/ diff. UoMs or intervals.
↳ Find! three are relative measures

$$\hookrightarrow P = \frac{C_t}{Y_t} = \frac{Y_t - S_t}{Y_t} \times 100\% \quad f = \text{forecast} \quad g = \text{observation}$$

Median absolute error:

$$MAE = \frac{1}{T-t^*} \sum_{t=t^*+1}^T |e_t|$$

+ easy to interpret, preserves UoM & overcome offsetting ME issue.

- same as above but offsetting

Root mean squared errors:

$$RMSE = \sqrt{\frac{1}{T-t^*} \sum_{t=t^*+1}^T e^2}$$

+ preserves UoM no offsetting pos/neg errors, large errors are penalised

Mean per cent error:

$$MPE = \frac{1}{T-t^*} \sum_{t=t^*+1}^T P_t$$

Mean absolute per cent error:

$$MAPE = \frac{1}{T-t^*} \sum_{t=t^*+1}^T |P_t|$$

+ same as above but relative measures

- If any $g_t = 0$ they are undefined
∴ only use if measurement scale = ratio

Mean absolute scaled errors:

$$MASE = \frac{1}{T-t} \sum_{t=t+1}^T \frac{|e_t|}{MAE_{NF}}, \quad MAE_{NF} = \frac{1}{t} \sum_{t=1}^t |y_t - y_{t-1}|$$

↳ scales f/cast errors to the naive f/cast over training period.

↳ Rules:

- ↳ MASE is never negative
- ↳ MASE = 0 if & only if $e_t = 0$, i.e. each out-of-sample is perf.
- ↳ MASE < 1 if model performs better than NF
- ↳ MASE = 1 if model performs like NF
- ↳ MASE > 1 if model performs worse than NF

Point f/cast → single number → can't verify confidence/reliability

Interval f/cast → range of numbers → size of range indicates confidence

Density f/cast → estimate of prob. dist → i.e. annual GDP = $N(1.80, 2.03)$

External regressors / exogenous predictors

→ Providing an external regressor to ARMA model does a linear regression w/ ARMA errors

- ↳ Doing this might improve sample regression eqn
- ↳ To verify if worthwhile, do statistical test on suitability of linear regression, compare resid's etc.

6

(G)ARCH models

- (G)ARCH models are employed b/c sometimes having a constant variance limits interpretation for some very volatile series. They drop this assumption by modelling ϵ_t as an ARMA process.
 - ↳ If it's reasonable to adopt a GARCH model, the squared correlograms will look very AR, indicating the series has a substantial dependence in the volatility of the returns.

The process for estimating a model w/ GARCH errors:

- i. Estimate the mean eqn (mult. regress. or ARIMA) for y_t
- ii. Take the squared residuals, ϵ_t^2 , and estimate variance eqn

Once you specify the mean eqn you can formally test for conditional heteroskedasticity using the Lagrange Multiplier (LM) test.

H_0 : no ARCH effects of order 1, ..., q. H_a : some ARCH effects of same order.

Finalising this hypothesis test is a two-step process:

- i. Estimate ϵ_t^2 w/ OLS: $\epsilon_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2 + \gamma_t$
 - ↳ No ARCH effects of order 1 to q if insig.
- ii. Take R^2 from regression to compute LM statistic
 - ↳ $LM = TR^2$ $\rightarrow T = \text{usable sample size}$
 - ↳ Under H_0 , LM converges to a chi-square dist. w/ $df = q$

→ Reject H_0 if LM is sufficiently large (use test to check if GARCH specif. is sufficient)

6

Interpreting GARCH model output

. Titles refer to their actual printout titles.

Optimal parameters

. Provides info about estimates & significance of mean & variance equation terms

was some autocorrelation, but not sufficiently material to merit a b/c of abs. value of sig. spikes.

Robust standard errors

. Same as above, but std. errors are robust against violations of the normal distribution assumption

↳ IF inconsistent w/ above, implies this ass. is imp. for interp.

Weighted Ljung-Box test on standardised squared residuals

. Same test as previously but for squared resids

↳ Rejecting H_0 here indicates your variance eqn is not correctly specified

Weighted ARCH-LM tests

. H_0 : there are no ARCH effects remaining in the error term for lag p .

. H_a : there are ARCH " " "

. Also indicates if variance eqn suff.

Information criteria

. Only useful for comparing diff. model outputs - the smaller the better.

Weighted Ljung-Box test on standardised residuals

. H_0 : there is no autocorrelation in the mean eqn residuals of order 1 to 2.

. H_a : there is autocorrelation " " "

. The outcome of this shouldn't be a surprise b/c you'd know from the correlograms. It might be when you set mean eqn there

Nyblom stability test

. tests for structural change in the data generating process, jointly & individually.

↳ To do this compare the test stat to crit value, if greater than α then you can reject that given test at the $\alpha\%$ significance level.

- helpful to compare mean vs. var outcomes

Sign bias tests

- Tests on if negative & positive shocks have diff. values on future volatility
↳ "leverage effect"
- H_0 : no leverage effect
- H_a : leverage effect

Pearson goodness of fit

- compare the empirical dist. of standardised residuals with the chosen cond. dist. of E_ϵ (specified in spec call - usually normal)
- "groups refers to bins"
- H_0 : empirical dist reflects chosen
- H_a : empirical dist. does not reflect chosen.

7

Extensions to GARCH models

GARCH extensions tend to play around w/ a few key assumptions:

i. the impact of shocks on volatility depend on magnitudes, not their sign

↳ i.e. h_t depends on ε_{t-1}^2 , not ε_{t-1} .

ii. For positive cond. variances, all α_i and β_j coefficients are restricted to be non-negative & their sum must be less than 1. this is because stationarity requires:

$$\sum \alpha_i + \sum \beta_j < 1$$

Integrated GARCH (IGARCH)

Relaxes assumption ii, instead stipulating coeffs must sum to 1.

↳ implying a unit root in the conditional variance. Therefore any shock to the conditional variance has a persistent effect, likely driven by some thick-tail distribution.

In this model β_1 is not estimated b/c $\beta_1 = 1 - \alpha_1$

↳ You should check this is the case: $H_0: \alpha_1 + \beta_1 = 1$ on GARCH model.

Threshold GARCH (TGARCH)

→ allows for asymmetry where, e.g., bad news ($\varepsilon_{t<0}$) has a larger impact than 'good news' ($\varepsilon_{t>0}$)

$$y = \mu_t + \varepsilon_t, h = \eta_0 \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=2}^q \eta_i d_{t-i} \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$$

$d=1$ if $\varepsilon_{t-1} < 0$. α_i & β_j satisfy other std. reg → η is meant to be pos.

∴ A significantly positive η implies negative shock are more impactful than positive.

Exponential GARCH (EGARCH)

→ The opposite of TGARCH

$$y_t = \mu_t + \varepsilon_t, \ln h_t = \alpha_0 + \sum \alpha_i \frac{\varepsilon_{t-i}}{\sqrt{h_{t-i}}} + \gamma \frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} + \sum \beta_j \ln h_{t-j}$$

$$E_{t-1}(\ln h_t) = \alpha_0 + \alpha_1 \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \gamma \frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} + \beta_1 \ln h_{t-1}$$

↳ The effect of an $\varepsilon_{t-1} > 0$ shock on expect log volatility is γ , +.

↳ For $\varepsilon_{t-1} < 0$, it is $-\gamma$.

→ The former is smaller than the latter if $\alpha_1 < 0$

→ No need to sign restrict β_j as logs are always positive

GARCH-in-mean (GARCH-M)

→ Allows the mean of $\{y_t\}$ to depend on conditional var.

↳ Useful for modelling asset mkt where risk-averse agents require risk premium. → higher avg. returns for holding a risky asset.

If y_t is return for risky asset:

$$y_t = \underbrace{\beta + \delta h_t}_{\text{exp. risk prem}} + \varepsilon_t, \delta > 0, h_t = \alpha_0 + \sum \alpha_i \varepsilon_{t-i}^2 + \sum \beta_j h_{t-j}$$

is constant if cond. var.
is constant → otherwise
 $E(y_t)$ is increasing func. of h

$\underbrace{\sum \beta_j h_{t-j}}_{\text{std. GARCH process}}$

8

High frequency data

How to measure the volatility of high frequency data?

Range volatility: diff. b/w highest & lowest log of variable
↳ simple & easy to calculate but not very insightful

Realised Volatility:

M = Sample size for day
 $\hookrightarrow 1\text{-sec freq.} = 23400$
 $(6.5 \times 60 \times 60)$

$\hookrightarrow 5\text{-sec freq.} = 4680$
 $(23400/5)$
 $\hookrightarrow 1\text{-min freq.} = 390$
 $(23400/60)$

$\hookrightarrow 5\text{-min freq.} = 78$
 $(23400 / (60 \times 5))$

$$RV(M) = M \times \text{Var}(r) \approx \sum_{i=1}^M r_i^2$$

↳ RV is the sum of the squared log returns during that day

$$\text{Var}(r) = \frac{1}{m} \sum_{i=1}^m (r_i - \bar{r})^2 \approx \frac{1}{m} \sum_{i=1}^m r_i^2$$

mechanism for biasing
volatility estimates

Microstructure noise

- Unintuitively, as M increases, RV estimates become more volatile as $RV(M)$ converges towards Integrated volatility.
- Additional movements in price at higher frequencies caused by trading, not market movements are microstructure noise
↳ makes RV very unstable, $P_t = \text{fundamental price} + \epsilon_t$
- ↳ 5-10 min frequencies generally considered unaffected by micro. noise

Bias: Power volatility estimates

- Jumps (policy, new CEO, new data) can also bias volatility estimates as one-off movements in price caused by ext. shock.
- ↳ If jumps occur, RV estimate represents sum of two terms due to market jumps, respectively.
- A jump-robust estimator of RV must detect persistent dynamics of daily volatility of asset returns

→ The Gipper volatility estimator does this, based off AR of absolute log returns

↳ Ratt oracle: jumps are infrequent ∴ if $|r_i|$ is large it's likely offset by $|r_{i-1}|$ (which is expected to be small)

$$BV(M) = \frac{\pi}{2} \sum_{i=2}^m |r_i| \times |r_{i-1}|$$

9

Vector Autoregression models

VAR models are just a system of AR models estimated together.

↳ A weakness of univariate models is the implied unidirectional relationship, which is unrealistic.

Steps for forecasting VAR models:

1. Test for stationarity

↳ If using non-stationary variables, OLS estimation might cause spurious regression

2. Determine the optimal lag length for the system

↳ This is key b/c otherwise output will be misleading.

↳ To do this, experiment w/ diff. lag lengths, optimising for model selection criterion, ensuring residuals are uncorrelated

⇒ Start w/ smallest lag suggested by criterion, check residuals

↳ Continue increasing lags until residuals are uncorrelated

↳ Breusch-Godfrey LM test: H_0 : no AR of orders 1-t, H_a : AR " " "

3. Estimate the model w/ number of lags from previous steps

↳ Interpret output

i. Check if all roots of the characteristic polynomial are $< 1 \Leftrightarrow$ stable

ii. Check the F-test for the model for significance

↳ Don't interpret coeffs. VAR models are very often overparameterised w/ little economic content

4. Forecast & reconstruct

↳ Because you're forecasting stationary variables you probably want the relevant level, so use `cumsum()`

↳ e.g. `UNR_eq = ts(UNR[length(UNR)] + cumsum(DUNR_eq), start...)`

9

Granger Causality

Granger causality means precedence: One TS variable consistently & predictably changes before another does.

Formally:

Two-way/feedback Granger causality requires two corresponding one-way causal relationships ($Z \rightarrow Y$)

↳ Z is Granger causal to Y ($Z \rightarrow Y$) only if y_{t+1} can be predicted better when the info. set contains **current/lagged** Z .

$$\text{cond. f/cast error variance} \left\{ \sigma^2(e_{t+1} | \Omega_t) \right\} < \sigma^2(e_{t+1} | \Omega_t - \{Z_{t-1}\}) \quad Q = \text{f/cast error}$$

denoted as $Z-Y$

↳ A variable is **instantaneously** Granger causal if y_{t+1} can be better predicted when the info set contains **current/lagged/future** Z .

$$\sigma^2(e_{t+1} | \Omega_t + Z_{t+1}) < \sigma^2(e_{t+1} | \Omega_t) \quad \text{this is a mutual relationship, i.e. } Z-Y \text{ implies } Y-Z.$$

Notes:

1. Granger causality definitions refer to one period ahead. For bi-variate r/ships this is sufficient, but for multivariate cases it might be insufficient. $\rightarrow X$ might not cause Y_{t+1} , but it causes Z which in turn cause Y_{t+2} . \rightarrow Implying a two-period ahead causal r/ship
2. Granger causality b/w two stationary variables can be tested for in individual eqns. Though VAR models are typically used to test in all directions.
3. Granger causality can depend on data frequency. Daily data probably highly AR unlike annual data.
4. In a multivariate system, a variable is endogenous if the other variables jointly cause it - otherwise it's exogenous

Testing for Granger Causality

Two approaches: F-test or Wald chi-square test on all lags of a variable(s) jointly.

↳ H_0 : all lags have zero coefficients, H_a : some lag(s) has (have) non zero coeffs

↳ The F-test requires normally distributed error terms, but in large samples is equivalent to the Wald test because:

$$\lim_{m_2 \rightarrow \infty} m_1 F \sim \chi^2_{m_1}$$

Tips for interpreting the printout:

→ The final test for each variable ([var.] <= ALL) indicates the endogeneity of the variable
↳ Either test detail causality amongst specific variables

→ Sometimes the p-values of the F-test and Wald chi-square test will disagree. In this case you need to be cognizant of the F-test requiring normally distributed error terms.

↳ You need to therefore conduct a `normality.test()` on the VAR system.

↳ If the JB test has a sufficiently small p-value, reject H_0 that errors are normally distributed and rely on the Wald chi-square test.

10

Structural Vector Autoregressive models

There are two major diff's b/w VAR & SVAR models

1. The error terms in an SVAR are uncorrelated with each other whereas for VAR models this might not be the case.
2. A VAR model can be estimated eqn by eqn by OLS because each eqn has the same predetermined variables on the RHS.
 ↳ Due to contemporaneous relationships among endogenous variables, SVAR models must be recovered from corresponding VAR if possible.

Under what conditions can we recover the SVAR?

$$\begin{aligned} y_t &= a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + \epsilon_{1t} \\ &\text{VAR} \end{aligned}$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + \epsilon_{2t}$$

$$\begin{aligned} y_t &= b_0 - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \\ &\text{SVAR} \end{aligned}$$

$$z_t = b_2 - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt}$$

- On top of coeffs, VAR system estimates variances of ϵ terms & their covariances → ∴ 9 terms
- The SVAR system only estimates variances of ϵ → ∴ 10 terms
 ↳ ∵ SVAR is unidentifiable unless we restrict one variable

The Cholesky decomposition is the simplest way to restrict:

- Sometimes it's reasonable to assume r/ships are asymmetric b/c y_t has a contemp. effect on z but not vice versa (i.e. $b_{12}=0$)
- Making this assumption reduces the number of params to 9!
- However this changes the r/ship b/w ϵ_t & ϵ_{t-1}

$$U_t = B^{-1}\epsilon_t = \begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} = \begin{bmatrix} \epsilon_{yt} \\ -b_{21}\epsilon_{yt} + \epsilon_{zt} \end{bmatrix}$$

↑ note: not b_{12} !

→ Therefore, you're able to estimate the SVAR with OLS. However, you need to be cognizant of this new $\varepsilon_t \nprec u_t$ relationship.

$$\left\{ \begin{array}{l} \hat{\varepsilon}_{y_t} = \hat{u}_{y_t}, \hat{\varepsilon}_1 = \hat{b}_2 \hat{\varepsilon}_{y_t} + \hat{u}_{2t} = \hat{b}_{21} \hat{u}_{1t} + \hat{u}_{2t} \end{array} \right.$$

Notes:

→ Cholesky decomposition assumes y_t "precedes" z_t . Therefore, it's only useful when theory suggests it's appropriate to do so.
↳ You must check how robust results are to alternative ordering.

→ The importance of ordering depends on the correlation b/w the error terms in the VAR system.

↳ The weaker the correlation b/w u_{nt} terms the less important ordering is.

↳ If you have several endog. variables, corrs will not be low & ∴ impractical to experiment w/ all orders.

→ Recursive SVAR models can be too restrictive in terms of economic theory

↳ Restrictions can be very arbitrary as you begin developing large systems.

10

Impulse response analysis

Objective is to trace the effects of structural innovations or shocks ($\varepsilon_{t,i}$) on the entire time path of LHS variables.

To do this we use the VMA representation of the VAR model. → VAR generates forecasts, VMA for studying dyn. props of system & calculate forecast errors.

$$\begin{aligned} y_t &= \alpha_{10} + \alpha_{11} y_{t-1} + \alpha_{12} z_{t-1} + \alpha_{12} \\ z_t &= \alpha_{20} + \alpha_{21} y_t + \alpha_{22} z_{t-1} + \alpha_{22} \end{aligned}$$

$$\Rightarrow \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \sum_{i=0}^{\infty} (\text{AL})^i \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix} + \sum_{i=0}^{\infty} (\text{AL})^i \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

with $\mu = \begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix}$ and $\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \frac{1}{1 - 6_{12} 6_{21}} \begin{bmatrix} 1 & -6_{12} \\ -6_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix} + \frac{1}{1 - 6_{12} 6_{21}} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} 1 & -6_{12} \\ -6_{21} & 1 \end{bmatrix} = \begin{bmatrix} \varphi_{11}(i) & \varphi_{12}(i) \\ \varphi_{12}(i) & \varphi_{22}(i) \end{bmatrix}$$

∴ We get the final form of the VMA(∞) rep. of VAR(1) system.

$$x = \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix} + \sum_{i=0}^{\infty} \underbrace{\begin{bmatrix} \varphi_{11}(i) & \varphi_{12}(i) \\ \varphi_{21}(i) & \varphi_{22}(i) \end{bmatrix}}_{\Phi_i} \begin{bmatrix} \varepsilon_{y,t-i} \\ \varepsilon_{z,t-i} \end{bmatrix} = \mu + \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i}$$

The elements of the $\Phi_i (i=0, 1, 2, \dots)$ matrices measure the effect of the ε_{zt} shocks on current & future values of $\{y_t\} \times \{z_t\}$ to $\{\varphi_{11}(i)\}, \{\varphi_{12}(i)\}, \{\varphi_{21}(i)\}, \{\varphi_{22}(i)\}$ ($i=0, 1, 2, \dots$) sets of coefficients are the impulse response functions.

The first element of ϕ_0 are called the **impact multipliers**, whereas the sum of n elements are the **intermediate multipliers**

Before doing this you need to make sure the VAR system is stationary:

$$\text{e.g. } \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \rightarrow I - A_1 L = \begin{bmatrix} 1 - 0.8L & -0.2L \\ -0.4L & 1 - 0.1L \end{bmatrix}$$

Determine inverse characteristic equation:

$$|I - A_1 L| = (1 - 6.8L)(1 - 0.1L) - (0.02L)(-0.4L) = 1 - 0.9L = 0$$

\therefore Single root, $L = 10\%$, is outside unit circle. \therefore VAR(1) system is stable & $\{y_t\}$, $\{z_t\}$ processes are stationary.

Forecast error decomposition

Useful for calculating the proportion of movements attributable a variable's own shocks versus other variables.

recall the $MA(oo)$ representation from previous note:

$$x_t = \mu + \sum_{i=0}^{\infty} \phi_i \cdot \epsilon_{t-i} \rightarrow E(x) = \mu + \sum_{i=0}^{\infty} \phi_i \cdot \epsilon_{t+h-i}$$

$$\therefore x_{t+h} - E(x_{t+h}) = \sum_{i=0}^{h-1} \phi_i \cdot \epsilon_{t+h-i} \rightarrow \text{The } h\text{-period ahead forecast errors}$$

$$1 = \frac{\sigma_y^2}{\sigma_y^2(u)} \sum_{i=0}^{h-1} \phi_{11}^2(i) + \frac{\sigma_y^2}{\sigma_y^2(h)} \sum_{i=0}^{h-1} \phi_{12}^2(i)$$

prop. of $\sigma_y^2(h)$ that
is due to shocks in
the $\{\epsilon_t\}$ sequence prop. of $\sigma_y^2(h)$ that
is due to shocks
in the $\{\epsilon_{t+h}\}$ sequence

notes:

- In practice, forecast error typically driven by own shocks in SR & other variables in LR
- Just like impulse response analysis, variance decomps reqs the estimation of some matrix which requires restriction. we can again rely on the cholesky decomposition.
- If the VAR system is stable and h is increasing, variance decomp should be approaching some constant.

11

Cointegration

Allows you to construct regressions using non-stationary variables if they have stationary linear combination - a cointegrated r/ship. These relationships are often referred to as equilibrium r/ships.

- ↳ Non-stationary ($I(1)$) variables might have a stationary ($I(0)$) linear combination. Specifically, variables x_1 & x_2 are said to be $CI(d, b)$ if:
 - i. Each is integrated of order d
 - ii. They have at least one non-trivial linear combination that is integrated of order $(d-b)$, where $d \geq b > 0$.

Cointegration rules:

1. If x_{1t} & x_{2t} are $CI(1,1)$, then so are $x_{1,t-i}$ & $x_{2,t-i}$ for any i .
2. Up to a scalar, cointegrate variables share the same stoch. trends.
3. Two $I(1)$ variables can have at most 1 lin. ind. coint. vecs.
 - ↳ In general, the **cointegration rank (r)**, or number of linearly independent coint. vecs cannot exceed integration order
4. Sample sizes need to be relatively large otherwise OLS estimations of β cointegration vector will be biased

For a bivariate system, solve so the dependent variable is E_t , the RHS is your error correction term.

- ↳ Its β is referred to as the **speed of adjustment coefficient**.
- ↳ At least one must be different from zero
- ↳ Given $\beta_i > 0$, stability requires $-2 < \alpha_1$ and $\alpha_2 < 2$

↳ This is because if the r/ship is cointegrated, regressing one variable over another will produce white noise residuals

The basic bivariate VECM for $I(1,1)$ is:

$$\Delta y = \alpha_1 (y_{t-1} - \beta_1 z_{t-1}) + \varepsilon_{1t}$$

$$\Delta z = \alpha_2 (y_{t-1} - \beta_1 z_{t-1}) + \varepsilon_{2t}$$

It is denoted a VECM(0)
because it does not
have lagged differences

It has several key features:

i. This VECM(0) is equivalent to the following VAR(1) model:

$$y_t = (1 + \kappa_1) y_{t-1} - \alpha_1 \beta_1 z_{t-1} + \varepsilon_{1t}$$

$$z_t = \alpha_2 y_{t-1} + (1 - \alpha_1 \beta_1) z_{t-1} + \varepsilon_{2t}$$

A level VAR(1),
but a restricted one
(3 ind. pars. not 9)

→ It can be shown every VEC(m) model has an equivalent VAR(p+1) representation

ii. As y & z are $I(1)$ & no lag is necessary,
 $\varepsilon_{1t}, \varepsilon_{2t}, \Delta y_{t-1}$ & Δz_{t-1} are all stationary

iii. This parameterisation allows for two different types of dynamics

- ↳ Adjustment to LR equilibrium via the lagged EC term
- ↳ Additional SR dynamics captured by lagged first diff's (AR distributed lags)

iv. One of the speed of adjustment coeffs can be zero
e.g.: if $\alpha_1 < 0$ & $\alpha_2 = 0$ equilibrium occurs thru y , not z
↳ Vice versa: if $\alpha_1 = 0$ & $\alpha_2 < 0$.

II

Cointegration tests

The objective of cointegration testing is to determine if variables that have stochastic trends share a **common stochastic trend**. There are two broad approaches:

Engle-Granger (EG) test Or we want WN residuals, use "Model 1" of the ADF test.
→ Effectively an (A)DF (unit root) test on the residuals, as white noise residuals imply cointegration.

$$\Delta e = \alpha_e + \sum_{i=1}^p \beta_i \Delta e_{t-i+1} + \varepsilon_t$$

$$H_0: \alpha_e = 0$$

$$H_a: -2 < \alpha_e < 2$$

e_t has a unit root, so ε_t is likely $I(1)$ & y_t, z_t are not $C(1,1)$

e_t does not have a unit root
∴ ε_t is likely $I(0)$ & y_t, z_t are $C(1,1)$

- When testing, test both orderings of $y_t \times z_t$, if at least 1 null hypothesis is rejected we can conclude cointegration.
- If two variables are $C(1,1)$, the relationship is best captured by a VECM
- Note that EG tests are only useful for bivariate systems, and cannot be used to detect the **number of cointegrating relationships** – the cointegration rank.

Johansen (J) test

- Tests the **number of endogenous variables (n)**, the **number of different stochastic trends (k)** they have, and the **cointegration rank (r)**

→ Suppose $n=3$ variables, which are pure random walks and are $I(1)$ with a stochastic trend.

$$y_t = \sum_i \varepsilon_{1,t-i} \quad z_t = \sum_i \varepsilon_{2,t-i} \quad v_t = \sum_i \varepsilon_{3,t-i}$$

when $n=2$ there are two possibilities:

i. The two stochastic trends are different, $k=2$.

↳ y & z are not $CI(1,1)$ → $r=0=2-2=n-k$

ii. The two stochastic trends are the same (up to a scalar)

↳ y & z are $CI(1,1)$ → $r=1=2-1=n-k$

when $n=3$ there are now three possibilities:

i. The three stochastic trends are different, $k=3$

↳ y , z & v are not $CI(1,1)$ → $r=0=3-3=n-k$

ii. There are two different stochastic trends

↳ y & z (or any combo) are $CI(1,1)$, but none are cointegrated w/ v individually.

↳ However together they are cointegrated, the third element of the cointegration vector is zero,

↳ y , z & v are $CI(1,1)$ → $r=1=3-2=n-k$

iii. There is only one stochastic trend, $k=1$

↳ All three variables have the same stochastic trend which drives the whole system → Two different scalars.

↳ y , z & v are $CI(1,1)$ → $r=2=3-1=n-k$

In general, for $n \geq 2$ of $I(1)$ variables, each can have their own stochastic trend. The cointegration rank must be between 1 and $n-1$. The value of r has useful modelling info:

i. If $r=0$ variables are not CI & ∴ no error correction in the system. The appropriate model is a first difference VAR.

- ii. If $r=n$, there are n linearly independent stationary combos of the variables ($k=0$). The appropriate model is a **level VAR**.
- iii. If $1 \leq r \leq n-1$, there are n independent cointegration relations and $k=n-r$ different stochastic trends. The appropriate model is a **VECM**.

It can be shown r is equal to the number of non-zero characteristic roots, also called the eigenvalues, of your matrix.

↳ The J test estimates if the number of eigenvalues (r_0) found are statistically significant.

There are two kinds of J-tests:

i. Trace test

$$H_0: CI \text{ rank is } \leq r_0$$

$$H_a: CI \text{ rank is } > r_0$$

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i)$$

ii. Maximum eigenvalue test

$$H_0: CI \text{ rank is } r_0$$

$$H_a: CI \text{ rank is } r_0 + 1$$

$$\begin{aligned} \lambda_{\text{eigen}}(r+1) &= -T \ln(1 - \hat{\lambda}_{r+1}) \\ &= \lambda_{\text{trace}}(r) - \lambda_{\text{trace}}(r+1) \end{aligned}$$

Perform these tests sequentially, moving from $r_0=0$ to $r_0=n-1$ until we first fail to reject H_0

12

Estimating VECMs

General approach:

- i. Test for stationarity \rightarrow If variables are $I(1)$ they might be CI
- ii. Confirm CI w/ tests \rightarrow Using the EG or J methodologies
- iii. Estimate if variables are CI \rightarrow Using EC or J methodologies

This note walks through (ii) & (iii), assuming (i) applies

Engle-Granger (EG) method

GC residuals
should be w.w. \leftarrow "type I" in
the printout

- \rightarrow EG test is a kind of ADF test which always uses "Model 1"
- \rightarrow When you are running the tests, do it twice swapping dependents.
- \rightarrow There are several steps to actually estimating the VECM:
 - i. estimate the LR r/sip b/w variables: $\ln(y - x)$
 - ii. Save the residuals from this regression
 - iii. Take the first difference of the dependent variable
 - iv. Take the lag of the residuals from the LR equilibrium egn
 - v. To get the first VECM egn, regress (iii) on (iv)
 $\Delta y = \beta_0 + \beta_1 (\text{lagged error})$
 - vi. Regress the original independent variable on (iii)

Johansen (J) method

- \rightarrow The tricky part about J tests is figuring out how to treat deterministic terms in the VECM. There are three possibilities, and three different function arguments to consider:
 - "none": no determ. term in EC/a constant outside EC
 \hookrightarrow Appropriate for linearly trending series (assuming all trends = statistic!)
 - "constant": a constant in EC/no determ. term outside EC.
 \hookrightarrow Appropriate only if no variables appear to have sustained increase/decrease tendency

- "trend": a trend variable but no constant in EC/a constant outside EC.
↳ Reasonable when there's some LR linear growth which CI relation doesn't capture

→ Once you've dealt with deterministic trends you need to determine lag length for the test, to do this use the normal method for determining VAR model lag-lengths.

→ Then conduct the "trace" J test:

- Evaluate $r=0$ first & then $r \leq 1$
- If $r=0$ is rejected, but $r \leq 1$ is maintained, suggests $r=1$.
- If $r=1$, only consider the first column of the "Eigenvectors" part of the output to determine EC term: $y = \beta_0 + \beta_1 x + \xi$, $\xi = y - \beta_0 - \beta_1 x$
- The last part of the printout shows the speed of adjustment coefficients. → If $r=1$ refer to first column only.

→ Then conduct the "maximum eigenvalue" J test:

- The only thing that changes is the second part of the printout → use the same approach as before: evaluate $r=0$ & the $r \leq 1$ sequentially.

→ To estimate the VECM simply pass the object you conducted the test on (to `cajols()`) and specify the cointegration rank.