EC4304 Forecasting
Lingjie, November 17, 2021

Conditional mean = trend + seasonal + cycle

$$E(Y_{t+h}|\Omega_t) = T_t + S_t + C_t$$

Forecasting is useful in guiding decisions
Different forecasting methods

1. Guessing
2. Rules of thumb
3. Naive extrapolation
4. Leading indicators
5. Naive/simple model
6. Formal forecasting models

Forecasting steps

1. Create approximate model for $E(Y_{t+h}|\Omega_t)$
2. Estimate parameters from data
3. (alternatively) Non-parametric model for $E(Y_{t+h}|\Omega_t)$

Codes in this cheatsheet is based on STATA

Notations

data frequency	:= time period (e.g. year, month)
in-sample obs	:= $\{Y_t\}_{t=1}^T$
out-of-sample period	:= $\{Y_T, Y_{T+1}, \cdots, Y_{T+h}\}$
forecast horizon	:= h
point forecast	:= $\{\hat{Y}_{t+h\|t}\}$
forecast distribution	:= $F(y)_{t+h\|t}$
forecast density	:= $f(y)_{t+h\|t}$
extrapolative forecast	:= sequence of forecasts $\hat{Y}_{T+1\|T}, \hat{Y}_{T+2\|T}, \cdots \hat{Y}_{T+h\|T}$
fan chart	:= prediction intervals from extrapolative forecast
	:= (common: 50%, 80%)
direct forecast	:= making forecast $\hat{Y}_{T+h\|T}$ directly
information set	:= $\Omega_T = \{(Y_t, X_t)\}_{t=1}^T$
trend	:= long term and smooth variation
seasonal	:= pattern which repeat annually and may be constant or variable
cycle	:= persistent dynamics not captured by trend or seasonal
level	:= actual values
return/growth rate	:= first differenced value

Note

- loss at each time period is different
- no gain from conditioning if information is independent with Y

Forecast reporting

Ideal reporting: interval forecast

Forecast (Predictive) Distribution

Showing the distribution for Y

Unconditional:	$f(y)$; $F(y) = P(Y \le y)$
Conditional:	$f(y\|x)$; $F(y\|x) = P(Y \le y\|x)$

```
* distribution summary
sum, detail
* kernel estimate of density
kdensity
kdensity y if x1==1 & x2==0
* multipl;e densities plot
kdensity y if x1==1 || kdensity y if x2==0
* cumulative distribution estimate
* and save as ydist
cumul y, gen(ydist)
```

Important to know the distribution of impacts.
Conditioning reduces forecast risk.

Point forecast

A point estimate $\hat{Y}$ is a summary of $F(y)$
Possible candidates: mean, median

Optimal point forecast

Ideal choice minimise the expected loss (risk)

- Quadratic: $\hat{Y} = E(Y)$

  estimation: OLS
- Absolute: $\hat{Y} = F^{-1}(0.5)$ (median)

  estimation: Quantile regression

Quadratic risk:

$$\begin{aligned}
R(\hat{Y}) &= E[(Y - \hat{Y})^2] \\
&= E(Y^2) - 2\hat{Y}E(Y) + \hat{Y}^2 \\
FOC: \frac{dR(\hat{Y})}{d\hat{Y}} &= -2E(Y) + 2\hat{Y} = 0 \\
\Rightarrow \hat{Y} &= E(Y)
\end{aligned}$$

Note: $\hat{Y}$ here is realised (constant).

Interval forecast

Intermedia solution to point and distribution forecast

$$C = [\hat{Y}_{lower}, \hat{Y}_{upper}]$$

Forecast interval $100\alpha\%$

$$P(Y \in C) = \alpha \Leftrightarrow Y \in \hat{Y} \pm Z_{\alpha/2} SE(\hat{Y})$$

Note: FI: $\hat{Y}|X = \hat{f}(X) + \epsilon_t$, CI: $E(\hat{Y}|X) = \hat{f}(X)$

Popular choice of α

α	$Z_{\alpha/2}$
0.90	1.64
0.80	1.28
0.68	1.00
0.50	0.67

RMSFE

Root mean squared forecast error

$$\sigma_e = \sqrt{E(Y - \hat{Y})^2} = \sqrt{Var(Y - \hat{Y})}$$

$\because E(Y_t - \hat{Y}_t) = 0$ (unbiased)
MSFE, AR(1)

$$\begin{aligned}
\sigma_e^2 &= E(Y_t - \hat{Y}_t)^2 \\
&\Leftrightarrow Var(Y_t - \hat{Y}_t) \\
&= Var(\beta_0 + \beta_1 + \epsilon_t - \hat{\beta}_0 - \hat{\beta}_1 Y_{t-1}) \\
&= \sigma_\epsilon^2 + Var(\hat{\beta}_0) + Y_{t-1}^2 Var(\hat{\beta}_1) + Cov(\hat{\beta}_0, \hat{\beta}_1)
\end{aligned}$$

useful for normal forecast interval

```
* only works for non-robust regression
predict varname, stdf
```

Quantile intervals

The forecast interval implied a α quantile, $\alpha = F^{-1}(q)$

$$C = \left[\frac{1-\alpha}{2}, \frac{1-(1-\alpha)}{2}\right]$$
$$q(\alpha) = F^{-1}(\alpha) = \inf_y F(y) \geq \alpha$$

Monotonicity rule

For any increasing transformation of Y, the α quantile of the transformation is the α quantile of Y.

$m(Y) = a + bY \Rightarrow q_m(\alpha) = a + bq_Y(\alpha)$
$m(Y) = ln(Y) \Rightarrow q_m(\alpha) = ln(q_Y(\alpha))$
$m(Y) = exp(Y) \Rightarrow q_m(\alpha) = exp(q_Y(\alpha))$

Normal rule

$$Y \sim N(\mu, \sigma^2)$$

$100 \cdot (1-\alpha)$ forecast interval

$$[\mu - \sigma z_{\alpha/2},\ \mu + \sigma z_{\alpha/2}]$$

where $z_{\alpha/2}$ is normal quantile

Log normality

$$ln(Y) \sim N(\mu, \sigma^2)$$

$100 \cdot (1-\alpha)$ forecast interval

$$[\exp(\mu - \sigma z_{\alpha/2}),\ \exp(\mu + \sigma z_{\alpha/2})]$$

Empirical quantile intervals

Estimate quantile directly from large dataset.
Harder to estimate conditional quantile.

```
centile Y, centile(2.5, 97.5)
```

Forecast error and loss function

Forecast error

$$e = Y - \hat{Y}$$

Loss function

Loss function represents the trade-off between errors

$$L(e);\ L(Y, \hat{Y})$$

Three rules

1. $L(0) = 0$
2. $L(e) \geq e,\ \forall e$
3. Non-increasing $e,\ \forall e < 0$, non-decreasing $e,\ \forall e > 0$
   $L(e') \leq L(e), e' < e < 0$
   $L(e) \leq L(e'), e' > e > 0$

Type of loss functions

Different loss function result in different ideals
e.g. bias is desired in asymmetric loss

Symmetric

Penalize positive and negative errors equally

- Quadratic (MSE) $L(e) = e^2$
- Absolute (MAE) $L(e) = |e|$

Asymmetric

Penalize positive and negative errors differently

- Linear-Linear (Linlin)
  Asymmetric version of absolute loss

$$L(e) = \begin{cases} a|e|, & e > 0 \\ b|e|, & e \leq 0 \end{cases}$$

- Linear-exponential (Linex)
  Linear on left if $a > 0$, exponential on the other

$$L(e) = b\left[e^{ae} - ae - 1\right], a \neq 0, b > 0$$

Level-dependent

Error depends on the level of the actual value

- Mean Absolute Percentage (MAPE)
  Unit less, weights error heavily when Y near 0

$$L(e, Y) = \left|\frac{e}{Y}\right|$$

State-dependent

Error depends on the state of the error (near 0 or inf)

- direction-of-change

$$L(Y, \hat{Y}) = \begin{cases} 0, & sign(\Delta Y) = sign(\Delta \hat{Y}) \\ 1, & sign(\Delta Y) \neq sign(\Delta \hat{Y}) \end{cases}$$

QLIKE loss

Error based on Kullback-Leibler divergences.
QLIKE is robust to measurement errors and invariant to unit of measurement.

$$QLIKE = \frac{Y}{\hat{Y}} - \log\left(\frac{Y}{\hat{Y}}\right) - 1$$

Risk (Expected Loss)

Loss after running multiple predictions with different datasets.

$$R(\hat{Y}) = E(L(e)) = E(L(Y - \hat{Y}))$$

Min risk = optimal point forecast (smallest loss on average)

Stationary Time series processes

mean, variance, autocovariance does not depend on time
k-th order:

mean $E(Y_t)$	:	μ
var $Var(Y_t)$	:	σ^2
autocovariance $\gamma(t,k)$	:	$cov(Y_t, Y_{t-k})$
	:	$E[(Y_t - \mu)(Y_{t-k} - \mu)]$
	:	$E(Y_t Y_{t-k}) - E(Y_t)^2$
	:	$E(Y_t Y_{t-k})$ if $E(Y_t) = 0$
autocorrelation $\rho(t,k)$	:	$\frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var(Y_t)Var(Y_{t-k})}}$
	:	$\frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)}$

Stationarity

Covariance stationarity

(weak, wide-sense, second-order) stationarity condition:

$E(Y_t) = \mu$	(mean stationary)
$Var(Y_t) = \sigma^2$	(variance stationary)
$E(Y_t^2) = \mu_2 < \infty$	(finite 2nd moment)
$\gamma(t,k) = \gamma(k)$	(constant autocovariance)
$\rho(k) = \gamma(k)/\sigma^2$	(constant autocorrelation)

for all t and any k

Properties

$$\gamma(k) = Cov(Y_t, Y_{t-k}) = \gamma(-k)$$
$$\gamma(0) = Cov(Y_t, Y_t) = Var(Y_t) = \sigma^2$$
$$|\gamma(k)| \leq \gamma(0) \ \forall k$$
$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \rho(-k) \in [-1, 1]$$
$$\rho(0) = 1$$

Strictly stationary

Condition: joint pdf invariant under time displacement

$$f(Y_{t_1}, Y_{t_2}, \cdots, T_{t_n}) = f(Y_{t_1+k}, Y_{t_2+k}, \cdots, T_{t_n+k})$$

weaker condition: stationarity up to order m
(joint moments up to m exist and stay constant over time)

White noise

white noise process has zero autocorrelation $\rho(k) = 0, k > 0$

$$Y_t = \epsilon_t, \ \epsilon_t \sim (0, \sigma^2)$$
$$\Leftrightarrow Y_t \sim WN(0, \sigma^2)$$

- Can check if ACF plot has no significant $\rho(k)$
- Serially uncorrelated, linearly unforecastable. However, not necessarily iid
- Serially uncorrelated $\neq$ serially independent (ARCH)

$$E(\epsilon_t^2 | \Omega_{t-1}) = \alpha + \beta\epsilon_{t-1}^2$$

- special iid case: Gaussian WN independent due to $\rho(k) = 0$

$$\epsilon_t \sim N(0, \sigma^2)$$
$$E(\epsilon_t^2 | \Omega_{t-1}) = E(\epsilon_t^2) = \sigma^2$$

Ergodicity

Ergodic for the m-th moment: time average converges to ensemble average as T grows large

$$E(Y_t) = \text{plim}_{i \to \infty} \frac{1}{I} \sum_{i=1}^{I} Y_t^{(i)} \text{ (ensemble average)}$$

$$\lim_{t \to \infty} \frac{1}{T} \sum_{t=1}^{T} Y_t = \frac{1}{I} \sum_{i=1}^{I} Y_t^{(i)} \text{ (ergodicity)}$$

Test criteria: $lim_{k \to \infty} \rho(k) = 0$
Ergodicity theorem ensures LLN for time series.

Remarks

- Requirement for weak stationarity and ergodicity might coincide, but not always
- For stationary Gaussian process, ergodicity for mean and second moment require the condition

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty$$

- $\Rightarrow$ if Y_t is ergodic, $\hat{Y}_{T+h|T} \approx E(Y_t)$ for large h
- series with seasonal and trend component, or NSA (not seasonally adjusted) might not be ergodic

Types of Ergodic series

Geometric decay $\rho(k) \approx c^k, c < 1$	Smooth decline to zero
Negative autocorrelation $\rho(1) < 0$	Erogodic if $\lim_{k \to \infty} \lvert\rho(k)\rvert = 0$ (Alternating sign)
Slow decay $\rho(k) \approx k^{-d}, d > 0$	Power law, long memory process

Estimation

Ergodicity ensures LLN works

$$\hat{\mu} = \frac{1}{T} \sum_{i=1}^{T} Y_i$$
$$\hat{\gamma}(k) = \frac{1}{T} \sum_{t=k+1}^{T} (Y_t - \hat{\mu})(Y_{t-k} - \hat{\mu})$$
$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$$

Note:

- estimation subject to sampling uncertainty
- estimates worsen as k gets large relative to T
- observe general pattern instead of outliers at large k

Confidence bands for autocorrelation

[R] if Y_t is independent white noise, then

$$Var(\hat{\rho}) \approx \frac{1}{T}$$
$$E(\hat{\rho}) \in [-\frac{2}{\sqrt{T}}, \frac{2}{\sqrt{T}}] \text{ (95\% CI)}$$

[STATA] Bartlett's formula:
assume Y_t is $MA(q) \Rightarrow \rho(k) = 0, k > q$

$$Var(\hat{\rho}(k)) \approx \frac{1}{T}(1 + 2\sum_{i=1}^{q} \rho(i)^2), k > q$$

Note:

- If sample autocorrelation fall within 95% CI in R, assume white noise. Else, examine Barlett bands
- Bartlett is point-wise hypothesis $H_0 : \rho(k) = 0$, not joint test
- Points falls outside of Bartlett bands (shaded region) is significantly different from 0

Joint tests for White Noise

Test all $\gamma(k)$ up to m are jointly zero (theory: all $\gamma(k)$)
Ideal m: large but not too large, Diebold suggest $\sqrt{T}$

$$H_0 : \rho(1) = \rho(2) = \cdots = \rho(m) = 0$$
$$\Leftrightarrow H_0 : Y_t \text{ is white noise}$$

Portmanteau tests

$$Q_{BP} = T \sum_{i=1}^{m} \hat{\rho}^2(i) \sim \chi_m^2$$
$$Q_{LB} = T(T+2) \sum_{i=1}^{m} \frac{1}{T-i} \hat{\rho}^2(i) \sim \chi_m^2$$

Box-Pierce (BP), Ljung-Box (LB) Q-statistic
STATA reports LB (better performance in small sample)

```
corrgram
```

Other tests: Lobato (2001), Pena and Rodriguez (2002), Delgado and Velasco (2010, 2011)

Lag operator L

useful way to manipulate lags

$$LY_t = Y_{t-1}$$
$$L^k Y_t = Y_{t-k}$$
$$A(L) = b_0 + b_1 L + b_2 L^2 + \cdots b_k L^k$$

```
gen x1 = L.x
reg rgdp L.rgdp L2.rgdp
reg rgdp L(1/12).rgdp
```

OLS Standard Errors in TS

Standard errors of OLS estimates in time series regression

- White Noise error: robust standard error
- Others: HAC adjusted standard error
  if no ACF plot, use default m

General Variance

Solving OLS estimator

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^T X_t e_t}{\sum_{i=1}^T X_t^2}$$

Asymtotically ($v_t := X_t e_t$)

$$\lim_{n\to\infty} \hat{\beta} = \beta + \frac{\sum_{t=1}^T v_t}{TVar(X_t)}$$

$$\begin{aligned} Var(\hat{\beta}) &= \frac{Var(\sum_{t=1}^T v_t)}{T^2 Var(X_t)^2} \\ &= \frac{\sum_{t=1}^T Var(v_t) + \sum_{t=1}^T Cov(v_t, v_j)}{T^2 Var(X_i)^2} \\ &= \frac{\sum_{i=1}^T \sigma_{v_t}^2 + \sum_{i=1}^T Cov(v_t, v_j)}{T^2 \sigma_X^4} \end{aligned}$$

Classical and Robust standard errors

Assumption

$$Cov(v_t, v_j) = 0 \text{ (independence)}$$

Classical (conditional homoscedasticity)

$$Var(X_t e_t) = Var(X_t) Var(e_t)$$
$$E(e_t^2 | \Omega_{t-1}) = \sigma^2 \text{ (equal var)}$$
$$SE(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}_e^2}{T\hat{\sigma}_X^2}}$$

Robust standard errors (heterscedasticity)

$$SE(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}_{v_t}^2}{T\hat{\sigma}_X^4}}$$

HAC standard errors

Heteroskedasticity and autocorrelation consistent (HAC)

$$Cov(v_t, v_j) \neq 0 \text{ (correlated errors)}$$

Adjustment factor f_T

$$\begin{aligned} Var(\hat{\beta}) &= \frac{Var(v_t)}{TVar(X_i)^2} f_T \\ &= \frac{\sigma_{v_t}^2}{\sigma_X^4} f_T \\ f_T &= \frac{Var(\sum_{t=1}^T v_t)}{TVar(v_t)} \\ \Rightarrow Var(\hat{\beta}) &= \frac{Var(v_t)}{TVar(X_i)^2} \cdot \frac{Var(\sum_{t=1}^T v_t)}{TVar(v_t)} \\ &= \text{original var} \end{aligned}$$

Estimate f_T with sample autocorrelations, and truncate at max significant lag m.

Unweighted and weighted HAC estimator

with a truncation parameter m
Unweighted (can have negative variance)

$$\hat{f} = 1 + 2\sum_{s=1}^m \hat{\rho}(s)$$

Newey-West Weighted (always nonnegative, preferred)

$$\hat{f} = 1 + 2\sum_{s=1}^m \left(\frac{m-s}{m}\right) \hat{\rho}(s)$$

Choice of truncation parameter m

m reflects the autocorrelation structure (ACF plot)

Schwert (max lag)	: $m = 12(T/100)^{1/4}$
Trend/Seasonal (no cycle)	: $m = 1.4T^{1/3}$
Stock and Watson (cycle style)	: $m = 0.75T^{1/3}$
full model (T, S, C), uncorrelated	: $m = 0$

Choice of m for h-step-ahead forecast

Since forecast error is $MA(h-1)$, $m = h - 1$

Model selection

Q: which order for AR(p)?
Fundamental trade-off: estimation error (var) vs model misspecification (bias)

Sequential tests

Test if coefficient for some variables are 0

- sequential t-test
- sequential F-test
  preferred over t-test in presence of high correlation among regressors

Limitation:

- not designed to select best forecast model
- search not comprehensive and outcome is path-dependent
- may end up overparameterization

Information criteria

- AIC: Akaike, minimise Kullback-Leibler distance between model and forecast distribution
- BIC: Schwarz Bayesian, based on highest posterior probability given data

Condition for Information criteria (common mistakes):

1. same number of observations
   (i.e. when comparing models with diff lags, keep the least obs)
2. same dependent variables
   (i.e. compare Y with Y and not $log(Y)$)
3. Assumes conditional homoskedasticity

Bayesian criterion

consider $M_i :=$ model i, $D :=$ data

$$\pi(M_1|D) = \frac{P(D|M_1)\pi(M_1)}{P(D|M_1)\pi(M_1) + P(D|M_2)\pi(M_2)}$$

and $\pi(M_i) = \frac{1}{2}$

Bayesian criterion for AR(p)

Balance fit (RSS) and model complexity (k), has consistency property: select model most likely to be true (also chooses the smaller model)
Assume AR(p) with normal errors and uniform priors

$$\pi(M_1|D) \propto exp\left(-\frac{BIC}{2}\right)$$
$$BIC = T\log\left(\frac{SSR}{T}\right) + k\log(T)$$

where $k :=$ num of estimated coefficients, $T :=$ sample size
Smallest BIC has highest posterior probability
Alternative forms

[STATA]

$$BIC = -2L + k\log(T)$$
$$2L = -T\log(2\pi) + 1) - T\log\left(\frac{SSR}{T}\right)$$

[R]

$$BIC = \log\left(\frac{SSR}{T}\right) + k\frac{\log(T)}{T}$$

Shibata criterion

Minimise forecast risk directly

$$R(\hat{Y}) = E(Y - \hat{Y})^2$$
$$E(SSR) = E(MSFE) - 2\sigma^2 k$$
$$E(MSFE) = T\sigma^2$$

Shibata bias correction criterion

$$S_k = SSR(1 + \frac{2k}{T})$$

Akaike criterion

AIC is an approximately unbiased estimate of the MSFE

$$T\log\left(\frac{S_k}{T}\right) \approx T\log\left(\frac{SSR}{T}\right) + 2k$$

AIC is an approximately unbiased estimate of the Killback-Leibler information criterion (KLIC)

Predictive Least Squares (PLS)/CV

Compute out-of-sample forecasts and associated forecast error

$$e_t = Y_t - \hat{Y}_t$$
$$PLS = \sqrt{\frac{1}{P}\sum_{t=M+1}^{T} e_t^2}$$

$T :=$ total sample
$P :=$ hold-out sample
$M :=$ training sample
Disadvantages

- Tends to overestimate true MSFE
- Tends to over-parsimonious (prefer smaller model)
- very sensitive to choice of P

Out-of-sample (OOS) model update

Extrapolation forecast beyond training data

True vs pseudo OOS

True OOS : made guesses about true unknown future values
pseudo OOS : useful to evaluate models, aka validation/testing data

Pseudo OOS: T : Total observations; N : Training data; P : evaluation data

Produce series of h-step forecasts (note: fixed horizon h), update estimate with additional data as time increase (using the following 3 methods)

Fixed estimation window

- Includes only first N observation (no update)
- Used when estimation costs are high (no real time update possible)
- Not desirable when model is unstable over time

Expanding/recursive estimation window

- Use first N data for estimation
- In next period, include an extra observation to update model
- When DGP is stationary: reduce estimation error over time
- When DGP changes: reduce var increase bias

Rolling estimation window

- Include most recent N observation
- In next period, drop oldest data and include $N + 1$ data
- Fixed sample size of the most recent data
- Used when unsure DGP is stationary and do not want to include outdated data
- Higher parameter uncertainty (var)

Forecast combination

Combine different forecasts to reducing variance
Assume forecast f_1, f_2 that are unbiased and uncorrelated with variance σ_1^2, σ_2^2
Weighted average

$$f = wf_1 + (1-w)f_2$$
$$Var(f) = w^2\sigma_1^2 + (1-w)^2\sigma_2^2$$

Equal weights

Equal weights

$$Var(f) = \frac{1}{4}(\sigma_1^2 + \sigma_2^2)$$

variance increase by 2 but divides by 4

Unequal weights

Solve by minimising $Var(f)$ wrt w

$$w^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}}$$

Note:

- weight on forecast 1 is inversely proportional to its variance
- weights are non-negative and sum to 1
- In reality, true variance is unknown

Bates-Granger

Estimate variance using out-of-sample forecast variances

$$w^* = \frac{\hat{\sigma}_j^{-2}}{\sum_{i=1}^{J}\hat{\sigma}_i^{-2}}$$

Assume uncorrelated forecasts

Granger-Ramanathan combination

Regression method to combine forecasts

$$Y_t = \beta_1 f_{1t} + \cdots + \beta_N f_{N,t} + e_t$$
$$\sum_{i=1}^{N} w_i = 1, w_i \geq 0$$

Note: no intercept

Bayesian model averaging (and AIC)

Based on BIC

$$w_m^* = exp\left(-\frac{BIC_m}{2}\right)$$
$$= exp\left(-\frac{\Delta BIC_m}{2}\right)$$
$$w_m = \frac{w_m^*}{\sum_{m=1}^{M} w_m^*}$$

where $\Delta BIC_m = (BIC_m - BIC^*) :=$ difference between model$m$ and best model. This is to adjust for underflow issue

Note:

- Weighted AIC replaces BIC with AIC
- For prediction interval, compute RMSE of combined forecast and use it as estimate for σ

Forecast evaluation

Evaluating forecast "quality"

Optimal forecast under squared loss

Properties of optimal forecasts under the squared loss (note: properties changes under different loss, for example, in Lin-Lin loss forecast should be biased instead)

- unbiased
- 1-step ahead errors are white noise
- h-step ahead errors are at most $MA(h-1)$
- h-step ahead errors with variance non-decreasing in h, and converging to the unconditional variance of the process

Unbiased

If $e_t = \epsilon_t$, testing $H_0 : \alpha = 0$

$$e_t = \alpha + \epsilon_t$$

e_t : 1-step ahead forecast error
ϵ_t : error in regression (not a value)

Note:

- If serial correlation is present (multi-step ahead error or suboptimal forecast)
- Use MA models and test for $H_0 : \alpha = 0$

1-step forecast errors are white noise

1 step ahead forecast error (optimal)

$$e_{t+1|t} = \epsilon_{t+1}$$

Look for evidence of serial correlation in forecast errors

- examine ACF
  see if autocorrelations are significant
- Exame Ljung-Box statistics
  Test joint tests of autocorrelation

h-step errors are $MA(h-1)$

h-step ahead forecast error (optimal)

$$e_{t+h|t} = \epsilon_{t+h} + b_1\epsilon_{t+h-1} + \cdots + b_{h-1}\epsilon_{t+1}$$

Simple test

- Plot ACF of forecasts errors
  examine whether autocorrelation beyond lag $h-1$ are significant
- Estimate $MA(h-1+q)$ model
  test if parameters beyond lag h-1 (q) are jointly zero

Forecast error variance

Variance increases with forecast horizon h

$$Var(e_{t+h|t}) = (1 + b_1^2 + b_2^2 + \cdots + b_{h-1}^2)\sigma^2$$
$$= \sigma^2 \sum_{i=0}^{h-1} b_i^2$$

$b_0 = 1$

Examine forecast error variances as function of h and observe if they are non-decreasing (or if there are patterns)

Unforecastable errors/MZ regression

Key property of optimal forecast errors: unable to forecast errors

idea: coefficients should be zero in the regression

$$e_{t+h|t} = \alpha + \beta\hat{Y}_{t+h|t} + \epsilon_{t+h}$$
$$\Leftrightarrow Y_{t+h} = \gamma + \theta\hat{Y}_{t+h|t} + e_{t+h|t}$$

$H_0 : \alpha = 0, \beta = 0$ or $H_0 : \gamma = 0, \theta = 1$

actual: Mincer-Zarnowitz regression

$$Y_{t+h} = \alpha + \beta\hat{Y}_{t+h|t} + u_{t+h}$$

Joint test $H_0 : \alpha = 0, \beta = 1$.

Note:

- Reject if there is systematic bias in the forecast
- Use appropriate standard error (robust, HAC)
- R^2 is popular way to compare forecasts from different models

Forecast accuracy

Commonly forecasts are not ideal, we might compare forecasts based on forecast accuracy (bias, risk) instead.

$$\text{Bias} = \frac{1}{P}\sum_{t=1}^{P} e_{t+h|t}$$
$$\text{MAE } (L(e) = |e|) = \frac{1}{P}\sum_{i=1}^{P} |e_{t+h|t}|$$
$$\text{RMSE } (L(e) = e^2) = \sqrt{\frac{1}{P}\sum_{t=1}^{P} e_{t+h|t}^2}$$

$$\text{Percentage error } = \frac{Y_{t+h} - \hat{Y}_{t+h|t}}{Y_{t+h}} = P_{t+h|t}$$
$$\text{MAPE} = \frac{1}{P}\sum_{t=1}^{P} |P_{t+h|t}|$$

Typically report ratio of RMSE to benchmark model

Meese-Rogoff puzzle

Random walk beats economic model - "Exchange rate models of the seventies: do they fit out-of-sample"

forecast risk comparison/DM test

Assumption: d_t is covariance stationary
Test for risk equality for models a and b

$$E(L(e^a_{t+h|t})) = E(L(e^b_{t+h|t}))$$
$$\Leftrightarrow E(d_t) = E(L(e^a_{t+h|t})) - E(L(e^b_{t+h|t})) = 0$$

Diebold Mariano EPA test

$$d_t = L(e^a_{t+h|t}) - L(e^b_{t+h|t})$$
$$DM_{12} = \frac{\bar{d}}{\sigma_{\bar{d}}/\sqrt{P}} \sim^A N(0,1)$$
$$\Leftrightarrow d_t = \mu + \epsilon_t$$

$H_0 : d_t = 0 \Leftrightarrow H_0 : \mu = 0$
Note:

- plot d_t against time to check if d_t is cov stationary
- cov stationary might not hold if using recursive estimation (as forecast error variance reduce over time $\Rightarrow$ not cov stationary)
- $\hat{\sigma}_{\bar{d}} :=$ HAC estimate of standard deviation (examine ACF, Q-stat)
- Test is asymptotic procedure: P has to be large
- Test compares forecasts, not models
- May include conditioning information (like 0/1 recession indicator)
- Estimation errors in both models' parameter might result in the true better performing model performed worse in small sample
- Cannot be applied to nested models with expanding window (AR(p) + expanding window), rolling window with nested model is fine

Finite sample DM test

$$t_{HLN} = (1 - P^{-1}(1-2h)) + P^{-2}h(h-1))^{1/2} t_{DM}$$
$$\sim t(P-1)$$

$h :=$ forecast horizon, $P :=$ number of OOS forecasts

Trend Model & Forecast

1. Specify and estimate trend model
2. Assess model fit/adequacy (selection)
3. construct forecast

Caution on pure trend forecasting

- uncertainty increase with forecast horizon (h)
- inaccurate trend specification result in extreme poor forecast
- long term forecast is poor due to changing trend
- trend generally changes over time

Deterministic vs stochastic trends

Deterministic trend is a nonrandom function of time.

$$T_t = f(t),\ t \in [1, T]$$

Stochastic trend varies randomly with time.

Trend specifications

$$\text{Quadratic: } T_t = \beta_0 + \beta_1 t + \beta_2 t^2$$
$$\text{Exponential: } T_t = \beta_0 \cdot \exp(\beta_1 t)$$
$$\text{log-linear: } \ln(T_t) = \ln(\beta_0) + \beta_1 t$$

Estimation

Quadratic and log-linear: OLS
Exponential: solve

$$\min_{\beta_0, \beta_1} \sum_{t=1}^{T} [Y_t - \beta_0 \cdot exp(\beta_1 t)]^2$$

```
* OLS
reg y t
predict varname, xb
* exp trend model
nl (y = {b0=0.1}*exp({b1}*t)), r
predict varname, yhat
```

Forecasting

$$Model:\ Y_t = \beta_0 + \beta_1 t + \epsilon_t$$
$$Forecast:\ Y_{T+h} = \beta_0 + \beta_1 (T+h) + \epsilon_{T+h}$$
$$Point:\ \hat{Y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 (T+h)$$
$$Interval(Point):\ \hat{Y}_{T+h} \pm \Phi(\alpha/2)\sigma_e$$

$\Phi(\alpha/2)$:= standard normal with $\alpha/2$ quantile
σ_e := root mean squared forecast error
Note: we assume $\epsilon_t \sim N(0, \sigma)$ iid to construct prediction interval

Trend RMSFE

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t$$
$$\sigma_e^2 = \sigma_\epsilon^2 + var(\hat{\beta}_0) + (T+h)^2 var(\hat{\beta}_1) + 2(T+h)cov(\hat{\beta}_0, \hat{\beta}_1)$$

Incorrect trend specification

MSFE increase as T, h increases.
MSFE grow with sample size and time horizon.

$$\text{True: } Y_t = \beta_0 + \beta_1 t + u_t,\ u_t \sim iid(0, \sigma_u^2)$$
$$\text{Misspecified: } Y_t = \beta_0 + u_t, u_t \sim iid(0, \sigma_u^2)$$
$$E[(Y_{T+h} - \beta_0)^2] = \beta_1^2 (T+h)^2 + \sigma_u^2$$

Removing trend remove the need for trend model
(e.g. first difference)

Breaking trend

Changing/ breaking trend: structural change/break

$$t < \tau : Y_t = \beta_0 + \beta_1 t + u_t$$
$$t \geq \tau : Y_t = \alpha_0 + \alpha_1 t + u_t$$

Either estimate each sub-sample separately or use dummy

$$\begin{aligned} Y_t &= (\beta_0 + \beta_1 t) I(t < \tau) + (\alpha_0 + \alpha_1 t) I(t \geq \tau) + u_t \\ &= \beta_0 + \beta_1 t + \beta_2 d_t + \beta_3 t d_t + u_t \\ \beta_2 &= \alpha_0 - \beta_0 \\ \beta_3 &= \alpha_1 - \beta_1 \\ d_t &= I(t \geq \tau) \end{aligned}$$

Continuous break

Want to impose continuous restriction such that

$$\beta_0 + \beta_1 \tau = \alpha_0 + \alpha_1 \tau$$
$$\Leftrightarrow \beta_2 + \beta_3 \tau = 0$$

Using spline technique

$$\begin{aligned} Y_t &= \gamma_0 + \gamma_1 t + \gamma_2 (t - \tau) d_t + u_t \\ &= \gamma_0 + \gamma_1 t + \gamma_2 t^* + u_t \\ t^* &= (t - \tau) d_t \end{aligned}$$

```
gen tstat=
    (time-tq(1974q1))*(time>=tq(1974q1))
```

Deciding break

Break is generally not advised.
Require long data sample (e.g. 10 years) after the breakdate, or economic explanation
Break can be tested with QLR statistic

Seasonality Model & Forecast

Deterministic vs stochastic seasonality

Deterministic seasonal pattern is a repetitive pattern over a calendar year

$$S_t = f(D_{it}), t \in [1, T], i \in s$$

s := seasonal frequency, quarterly (4), monthly (12)
D_{it} := seasonal dummy, time = t, seasonal frequency = i
Stochastic seasonality pattern approximately repeats itself, but evolves over the years.

Seasonal adjustment

Estimate and remove the seasonal component.
Focus on trend and business cycle movement

De-seasonalization

General: subtract seasonal component from original series
Seasonal dummy model: add $E(Y)$ to u_t

Types of seasonality

- Holiday effect
- Trading day effect
- Day of week effect
- Intraday seasonality (hour, time of the day)
- Quarter
- Monthly

Deterministic seasonality

$$Y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \epsilon_t \quad (1)$$

$$\Leftrightarrow = \alpha + \sum_{i=1}^{s-1} \beta_i D_{it} + \epsilon_t \quad (2)$$

D_{it} := seasonal dummies
:= 1 if data in period i
$S_t = \sum_{i=1}^{s} \gamma_i D_{it}$:= seasonality

Interpretation

Model (1) : γ_s = seasonality effect
Model (2) : α = seasonality effect of omitted period
: $\beta_i = \gamma_i - \gamma_s$ differences in $(s-1)$ seasonal components from the omitted period

Error analysis

Examine residuals and ensure no seasonality is present
In case of changing seasonality, residuals will still contain seasonal component

Cycles Model & Forecast

Cycle: persistent dynamic that remains after accounting for trend and seasonality, a stochastic time series process
Note:

- Cycles (C_t) is covariance stationary and ergodic time series process
- Wold representation ($MA(\infty)$) approximate any stationary process by general linear process
- AR, MA, ARMA model aim to provide approximation for Wold representation

Wold's theorem

Let Y_t be

- any mean-zero covariance stationary process
- not containing any deterministic trend or seasonality

We can express any stationary process (incl nonlinear) approximately by the general linear process below:

$$Y_t = B(L)\epsilon_t = \sum_{i=0}^{\infty} b_i \epsilon_{t-i}$$

where

$$b_0 = 1, \quad \sum_{t=0}^{\infty} b_t^2 < \infty$$

$$\epsilon_t = Y_t - E(\hat{Y}_t | Y_{t-s}, s \geq 1) \sim WN(0, \sigma^2)$$

Note:

- Stationary time series processes are constructed as linear function of innovations, or shocks, ϵ_t

  practically, ϵ_t are constructed as 1-step ahead forecast errors with Y_t regress on all available lags $\{Y_{t-k}\}_{k=1}^{T}$
- Further assume ϵ_t is serially independent $E(\epsilon_t^k | \Omega_{t-1}) = E(\epsilon_t^k)$ for PI construction

  assumption is removed at GARCH model

Moments

Let $\Omega_{t-1} := \{\epsilon_{t-1}, \epsilon_{t-2}, \cdots\}$

$$E(Y_t) = E\left(\sum_{i=0}^{\infty} b_i \epsilon_{t-i}\right) = 0$$

$$E(Y_t|\Omega_t) = \sum_{i=1}^{\infty} b_i \epsilon_{t-i}$$

$$Var(Y_t) = Var\left(\sum_{i=0}^{\infty} b_i \epsilon_{t-i}\right) = \sigma^2 \sum_{i=0}^{\infty} b_i^2$$

$$Var(Y_t|\Omega_{t-1}) = E\left\{[Y_t - E(Y_t|\Omega_t)]^2 | \Omega_{t-1}\right\}$$
$$= E(\epsilon_t^2|\Omega_{t-1}) = E(\epsilon_t^2) = \sigma^2$$

Note: White noise are serially uncorrelated

Approximate Wold's infinite order polynomial

Using rational polynomial with $(p+q)$ parameters

$$B(L) \approx \frac{\Theta(L)}{\Phi(L)} = \frac{\sum_{i=0}^{q} \theta_i L^i}{\sum_{j=0}^{p} \phi_j L^j}$$

Box-Jenkins methodology

1. Identify model
2. Estimate parameters
3. Diagnostic check

Identify MA, AR, ARMA with ACF/PACF

ACF : autocorrelation plot $\rho(k) = Corr(Y_t, Y_{t-k})$
PACF : partial autocorrelation plot $p(k) \Rightarrow \phi_k$ in
: $Y_t = \alpha + \sum_{i=1}^{k} \phi_i Y_{t-1}$
Identification:

1. Examine ACF for $MA(q)$, PACF for $AR(p)$

   [remark] if neither ACF/PACF shows clear cut-off, ARMA model might be preferred
2. Use model selection criteria such as AIC/BIC

```
* PACF
pac
* ACF
ac
* both
corrgram
```

Estimation

- MA(q), ARMA(p,q): MLE estimation assuming Gaussianity
- AR(p): OLS estimation

Residual analysis

Diagnostic checking on residuals:

- Cycle is modeled well when residual is white noise

  [method1] Residual plot should show WN

  [method2] ACF plot on residual should show all 0 autocorrelation (WN)

  [method3] Q-test (Ljung-Box Q) $p-$value $> 5\%$ (Note: only observe $M = \sqrt{T}$th p-value)

- Forecast intervals is appropriate when residual follows normal distribution

  [method1] kernel density plot

  [method2] Jarque-Bera test

Cycles: Moving Average (MA) $\Theta(L)\epsilon_t$

$$\begin{aligned} MA(1):\ & Y_t = \epsilon_t + \theta\epsilon_{t-1} = (1+\theta L)\epsilon_t \\ MA(q):\ & Y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q} \\ & = \Theta(L)\epsilon_t = \sum_{i=0}^{q}\theta_i\epsilon_{t-i} \end{aligned}$$

Inversion condition (all MA are stationary):

- $|\theta| < 1$
- all polynomial roots outside of the unit circle
- Express MA processes as lag operations and solve $\Phi(L) = 0$

Moving Average (MA) processes

- $\theta \in (-1,1)$ controls the degree of serial correlation
- $\theta_0 = 1$
- ϵ_t affects Y_t over two periods:

  [1] Contemporaneous impact

  [2] One-period delayed impact

- $MA(q)$ process not forecastable beyond q steps
- 1-period-ahead forecast errors are white noise ϵ_t
- h-period-ahead forecast errors are $MA(h-1) = \sum_{i=0}^{h-1}\theta_i\epsilon_{t-i}$
- Forecast error variance increases with h until $Var(Y_t), h > q$
- $MA(q)$ not often used in persistent economic data

```
* MA(1)
arima rgdp, ma(1)
* MA(p)
arima rgdp, ma(1/p)
```

MA(1)/(q): Mean

MA(1)

$$E(Y_t) = E(\epsilon_t + \theta\epsilon_{t-1}) = 0$$
$$E(Y_t|\Omega_{t-1}) = E(\epsilon_t + \theta\epsilon_{t-1}|\Omega_{t-1}) = \theta\epsilon_{t-1}$$

MA(q)

$$E(Y_t) = E\left(\sum_{i=0}^{q}\theta_i\epsilon_{t-i}\right) = 0$$
$$E(Y_t|\Omega_{t-1}) = \sum_{i=1}^{q}\theta_i\epsilon_{t-i}$$

Note: The optimal forecast error is $Y_t - \hat{Y}_t = \epsilon_t$

MA(1)/(q): Variance

MA(1)

$$Var(Y_t) = Var(\epsilon_t) + \theta^2 Var(\epsilon_{t-1}) = \sigma^2(1+\theta^2)$$
$$Var(Y_t|\Omega_{t-1}) = Var(\epsilon_t|\Omega_{t-1}) + \theta^2 Var(\epsilon_{t-1}|\Omega_{t-1}) = \sigma^2$$

MA(q)

$$Var(Y_t) = Var\left(\sum_{i=0}^{q}\theta_i\epsilon_{t-i}\right) = \sigma^2\sum_{i=0}^{q}\theta_i^2$$
$$Var(Y_t|\Omega_{t-1}) = \sigma^2$$

Note:

- Variance depends on θ: larger coefficient $\Rightarrow$ higher variability
- The conditional variance, the innovation variance and 1-step forecast variance are the same

MA(1): Autocovariance

MA(1)

$$\begin{aligned} \gamma(1) &= E(Y_tY_{t-1}) = E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})] \\ &= E(\epsilon_t\epsilon_{t-1}) + \theta E(\epsilon_{t-1}^2) + \theta E(\epsilon_t\epsilon_{t-2}) + \theta^2 E(\epsilon_{t-1}\epsilon_{t-2}) \\ &= \rho_\epsilon(1) + \theta E(\epsilon_{t-1}^2) + \theta\rho_\epsilon(2) + \theta^2\rho_\epsilon(1) \\ &= \theta\sigma^2 \\ \gamma(k) &= E(Y_tY_{t-k}) = E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-k} + \theta\epsilon_{t-k-1})] \\ &= E(\epsilon_t\epsilon_{t-k}) + \theta E(\epsilon_{t-1}\epsilon_{t-k}) + \theta E(\epsilon_t\epsilon_{t-k-1}) \\ &+ \theta^2 E(\epsilon_{t-1}\epsilon_{t-k-1}) \\ &= \rho_\epsilon(k) + \theta\rho_\epsilon(k-1) + \theta\rho_\epsilon(k+1) + \theta^2\rho_\epsilon(k) \\ &= 0 \end{aligned}$$

Note: For WN, $\rho(k) = 0, k > 0 \Rightarrow$ autocovaraince function is zero for all $k > 1$

MA(1): Autocorrelation

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta\sigma^2}{\sigma^2(1+\theta^2)} = \frac{\theta}{1+\theta^2}$$

Note:

- Since $\gamma(k) = 0, k > 1$, $\rho(k) = 0, k > 1$.
  Process has very short memory (1 period)
- $Sign(\theta)$ determines sign of the first autocorrelation
- For invertible MA(1): $\theta \in [-1,1] \Rightarrow \rho(1) \in [-0.5, 0.5]$

MA(1): Autoregressive representation

Using lag operator

$$Y_t = \epsilon_t + \theta\epsilon_{t-1} = (1+\theta L)\epsilon_t$$
$$(1+\theta L)^{-1}Y_t = \epsilon_t, |\theta| < 1$$

When MA process is invertible, we can represent MA(1) in terms of current period shock and lags of Y_t

$$\begin{aligned} Y_t &= \epsilon_t + \theta\epsilon_{t-1} \\ \Leftrightarrow \epsilon_t &= Y_t - \theta\epsilon_{t-1} \\ &= Y_t - \theta(Y_{t-1} - \theta\epsilon_{t-2}) \\ \Rightarrow Y_t &= \epsilon_t - \theta Y_{t-1} + \theta^2 Y_{t-2} - \theta^3 Y_{t-3} + \cdots \\ &= -\sum_{i=1}^{\infty}(-\theta)^i Y_{t-i} + \epsilon_t \end{aligned}$$

MA(1)/(p): Invertible MA processes

Series converge if $|\theta| < 1$ (and inversion exists)

$$\begin{aligned} \epsilon_t &= (1-(-\theta L))^{-1}Y_t \\ &= (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \cdots)Y_t \\ &= \Theta(L)Y_t \end{aligned}$$

To ensure invertibility:

- first q autocorrelations are nonzero, above q are zero
- all q roots of the polynomial outside the unit circle

MA(1)/(p): optimal forecast

1-period-ahead optimal forecast

$$MA(1): \hat{Y}_{T+1|T} = E_T(\epsilon_{T+1} + \theta\epsilon_T) = \theta\epsilon_T$$
$$MA(q): \hat{Y}_{T+1|T} = \sum_{i=1}^{q}\theta_i\epsilon_{T-q+1}$$

h-periods-ahead optimal forecast, $h \le q$

$$MA(q): \hat{Y}_{T+h|T} = E_T\left(\sum_{i=0}^{q}\theta_i\epsilon_{T+h-i}\right) = \sum_{i=h}^{q}\theta_i\epsilon_{T+h-i}$$

h-periods-ahead optimal forecast, $h > q$

$$MA(1): \hat{Y}_{T+h|T} = E_T(\epsilon_{T+h} + \theta\epsilon_{T+h-1}) = 0$$
$$MA(q): \hat{Y}_{T+h|T} = E_T\left(\sum_{i=0}^{q}\theta_i\epsilon_{T+h-i}\right) = 0$$

For $h > q$, optimal forecast is unconditional mean $(= 0)$

MA estimation with recursive forecast

Problem: in reality ϵ_t is not observed
Solution: construct ϵ_t recursively assuming $\epsilon_0 = 0$
$\epsilon_1 = Y_1 - \theta\epsilon_0 = Y_1, \epsilon_2 = Y_2 - \theta\epsilon_1, \cdots, \epsilon_T = Y_T - \theta\epsilon_{T-1}$
Done automatically in software by using $\hat{\theta}$ from MLE

MA(1)/(q): forecast errors

1-period-ahead forecast error

$$\begin{aligned} MA(1): e_{T+1|T} &= Y_{T+1} - \hat{Y}_{T+1} \\ &= \epsilon_{T+1} + \theta\epsilon_T - \theta\epsilon_T \\ &= \epsilon_{T+1} \\ MA(q): e_{T+1|T} &= \epsilon_{T+1} \end{aligned}$$

h-period-ahead, $h \le q$

$$\begin{aligned} MA(q): e_{T+h|T} &= \sum_{i=0}^{h-1}\theta_i\epsilon_{T+h-i} \\ &= MA(h-1) \end{aligned}$$

h-period-ahead, $h > q$

$$MA(1): e_{T+h|T} = \epsilon_{T+h} + \theta\epsilon_{T+h-1}$$
$$MA(q): e_{T+h|T} = \sum_{i=0}^{q}\theta_i\epsilon_{T+h-i}$$

MA(1)/(p): forecast error variance

$$\begin{aligned} MA(1), h = 1: Var(e_{T+1|T}) &= Var(\epsilon_{T+1}) = \sigma^2 \\ MA(1), h > 1: Var(e_{T+h|T}) &= Var(\epsilon_{T+h} + \theta\epsilon_{T+h-1}) \\ &= \sigma^2(1+\theta^2) \\ MA(q), h \le q: Var(e_{T+h|T}) &= \sigma^2\left(1 + \sum_{i=0}^{h-1}\theta_i^2\right) \\ MA(q), h > q: Var(e_{T+h|T}) &= \sigma^2\left(1 + \sum_{i=0}^{q}\theta_i^2\right) \end{aligned}$$

For $h \le q: Var(e_{T+h|T}) \le Var(Y_t)$
For $h > q: Var(e_{T+h|T}) = Var(Y_t)$

Cycles: Autoregressive (AR) $\frac{1}{\Phi(L)}\epsilon_t$

$$\begin{aligned} AR(1)\ Y_i &= \phi Y_{t-1} + \epsilon_t = (1-\phi L)^{-1}\epsilon_t \\ AR(p)\ Y_i &= \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t \\ &= \Phi(L)^{-1}\epsilon_t = \left(\sum_{i=0}^{p}\phi_p L^p\right)^{-1}\epsilon_t \end{aligned}$$

Stationary condition (all AR are invertible):

- all roots of lag polynomial outside of unit circle.
- sum of $AR(p)$ coefficient < 1 (note: not abs)

Autoregressive (AR) processes

$\epsilon_t \sim WN(0, \sigma^2)$

AR(1)

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \epsilon_t \\ \Rightarrow \epsilon_t &= (1-\phi L)Y_t \\ \Rightarrow Y_t &= (1-\phi L)^{-1}\epsilon_t \end{aligned}$$

AR(p)

$$\begin{aligned} Y_t &= \epsilon_t + \sum_{i=1}^{p}\phi_i Y_{t-i} = \epsilon_t + \left(\sum_{i=1}^{p}\phi_i L^i\right)Y_t \\ \Rightarrow \epsilon_t &= \left(1 - \sum_{i=1}^{p}\phi_i L^i\right)Y_t = \Phi(L)Y_t \\ \Rightarrow Y_t &= \Phi(L)^{-1}\epsilon_t \end{aligned}$$

Note: ϕ determines if Y_t, Y_{t-1} are positively/negatively correlated

AR(1): Inversion

Rewrite process as $MA(\infty)$

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \epsilon_t = \epsilon_t + \phi(\phi Y_{t-2} + \epsilon_{t-1}) \\ &= \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots \\ &= \sum_{i=0}^{\infty}\phi^i\epsilon_{t-i} \end{aligned}$$

Require $|\phi| < 1$ for inversion and stationarity
AR(1) is infinite MA (Wold) with one free parameter

AR(1): Mean

$$E(Y_t) = E\left(\sum_{i=0}^{\infty}\phi^i\epsilon_{t-1}\right) = 0$$
$$E(Y_t|\Omega_{t-1}) = E(\phi Y_{t-1} + \epsilon_t|\Omega_{t-1}) = \phi Y_{t-1}$$

AR(1): Variance

$$Var(Y_t) = Var\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-1}\right) = \sum_{i=0}^{\infty} \phi^{2i}\sigma^2$$
$$= \frac{\sigma^2}{1-\phi^2} \text{ (by series convergence)}$$
$$Var(Y_t|\Omega_{t-1}) = Var(\phi Y_{t-1} + \epsilon_t|\Omega_{t-1}) = \sigma^2 \text{ (assume iid)}$$

Trick: $Var(Y_t) = Var(Y_{t-1})$ (variance stationarity)

$$\begin{aligned} Var(Y_t) &= Var(\phi Y_{t-1} + \epsilon_t) \\ &= \phi^2 Var(Y_{t-1}) + \sigma^2 \\ &= \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

If $\phi = 1$, Var is infinite

Random Walk/ Unit Root

When $\phi = 1$, AR(1) is known as random walk or unit root process

$$\begin{aligned} Y_t &= Y_{t-1} + \epsilon_t \\ &= Y_0 + \sum_{i=0}^{t-1} \epsilon_{t-i} \\ \Delta Y_t &= Y_t - Y_{t-1} = \epsilon_t \end{aligned}$$

Infinite memory: shocks have permanent effects.

Wonders without mean reversion.

Note: differencing ΔY_t gives white noise

AR(1): Autocovariance

Since $E(Y_t) = 0, E(\epsilon_t Y_{t-k}) = E(\epsilon_t)E(Y_{t-k})$

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \epsilon_t \\ E(Y_t Y_{t-k}) &= E(\phi Y_{t-1} Y_{t-k}) + E(\epsilon_t Y_{t-k}) \\ &\Rightarrow \gamma(k) = \phi\gamma(k-1) \end{aligned}$$

Yule-Walker equation:

recursively work out $\gamma(k)$ with known $\gamma(0) = Var(Y_0)$

$$\begin{aligned} \gamma(1) &= \phi\gamma(0) = \phi\frac{\sigma^2}{1-\phi^2} \\ \gamma(2) &= \phi\gamma(1) = \phi^2\frac{\sigma^2}{1-\phi^2} \\ &\vdots \\ \gamma(k) &= \phi\gamma(k-1) = \phi^k\frac{\sigma^2}{1-\phi^2} \end{aligned}$$

Yule-Walker trick

Multiple both side by Y_{t-k}

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \epsilon_t \\ Y_t Y_{t-k} &= \phi Y_{t-1} Y_{t-k} + \epsilon_t Y_{t-k} \end{aligned}$$

AR(1): Autocorrelation

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi^k, k \geq 0$$

Note: AR(1) autocorrelations exhibit geometric decay
Rate of decay $\propto \frac{1}{\phi}$. Therefore, ϕ describes persistence in ts.

AR(1): forecast without intercept

1-period-ahead

$$\hat{Y}_{T+1|T} = E_t(\phi Y_t + \epsilon_{t+1}) = \phi Y_t$$

h-period-ahead

$$\begin{aligned} \hat{Y}_{T+h|T} &= E_T(Y_{T+h}) \\ &= \phi^h Y_T \\ &\Leftrightarrow \phi\hat{Y}_{T+h-1} \end{aligned}$$

Optimal h-step-ahead forecast derived from chain rule

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \epsilon_t \\ &= \epsilon_t + \phi(\phi Y_{t-2} + \epsilon_{t-1}) \\ &= \phi^2 Y_{t-2} + \epsilon_t + \phi\epsilon_{t-1} \\ &\vdots \\ &= \phi^h Y_T + \epsilon_{T+h} + \phi\epsilon_{T+h-1} + \cdots + \phi^{h-1}\epsilon_{T-h+1} \\ &= \phi^h Y_T + \sum_{i=0}^{h-1} \phi^i \epsilon_{T+h-i} \end{aligned}$$

Note:

- Forecast can be obtained through OLS
- Chain rule of forecasting:
  $\hat{Y}_{T+h|T} = \phi^h Y_T = \phi\hat{Y}_{T+h-1|T}$

AR(1): forecasting with intercept (1-period)

$$\hat{Y}_{T+1|T} = E(\alpha + \phi Y_T + \epsilon_{T+1}) = \alpha + \phi Y_T$$

Constant in ARIMA Y, AR(1)

Note: constant in STATA ARMA is $E(Y_t)$ and not α

$$\begin{aligned} E(Y_t) &= \alpha + \phi E(Y_{t-1}) \\ &= \frac{\alpha}{1-\phi} \end{aligned}$$

When using OLS, constant $= \alpha$

(1-step) forecast error

1-step-ahead:

$$\begin{aligned} AR(1): \ e_{T+1|T} &= (\alpha + \phi Y_T + \epsilon_{T+1}) - (\alpha + \phi Y_T) \\ &= \epsilon_{T+1} \end{aligned}$$

(1-step) forecast error variance

1-step-ahead:

$$AR(1): \ Var(e_{T+1|T}) = \sigma^2$$

(1-step) Forecast intervals

1-step ahead (assume $\epsilon_{T+1} \sim N(0, \sigma^2)$)

$$AR(1): \hat{Y}_{T+1|T} \pm \hat{\sigma} \times z_\alpha$$

where

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t^2$$

AR(1): forecasting with intercept (h-step)

(h-step) Plug-in method: Estimation

Forecast as function of parameters (back substitution)
2-step-ahead

$$\begin{aligned} \hat{Y}_{T+2|T} &= E_T\left[\alpha + \phi\left(\alpha + \phi Y_T + \epsilon_{T+1}\right) + \epsilon_{T+2}\right] \\ &= E_T\left[(1+\phi)\alpha + \phi^2 Y_T + \epsilon_{T+2} + \phi\epsilon_{T+1}\right] \\ &= (1+\phi)\alpha + \phi^2 Y_T \end{aligned}$$

h-step

$$\hat{Y}_{T+h|T} = (1+\hat{\phi}+\hat{\phi}^2+\cdots+\hat{\phi}^{h-1})\hat{\alpha}+\hat{\phi}^h Y_T$$

derived using back substitution

$$\begin{aligned}
Y_t &= \alpha + \phi Y_{t-1} + \epsilon_t \\
&= \alpha + \phi(\alpha + \phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&\vdots \\
&= (1+\phi+\phi^2+\cdots+\phi^{h-1})\alpha + \phi^h Y_{t-h} + u_t \\
u_t &= \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots + \phi^{h-1}\epsilon_{t-h+1} \\
&\sim MA(h-1)
\end{aligned}$$

Note:

- Simple but cumbersome for multi-step forecast

(h-step) Plug-in method: Forecast Variance

Since

$$\begin{aligned}
Y_{T+2|T} &= (1+\phi)\alpha + \phi^2 Y_T + \epsilon_{T+2} + \phi\epsilon_{T+1} \\
Var(\epsilon_{T+2}) &= Var(\epsilon_{T+1})
\end{aligned}$$

Therefore

$$\hat{\sigma}_u = \sqrt{(1+\hat{\phi}^2)\hat{\sigma}^2}$$

Note:

- Hard to generalize beyond AR(1) models
- Require result from AR(1) regression to get estimates

(h-step) Iterated method: Estimation

Use chain-rule to compute 1-step then 2-step forecast
2-period-ahead

$$\begin{aligned}
E_T(Y_{T+2}) &= E_T(\alpha + \phi Y_{T+1} + \epsilon_{T+2}) \\
&= \alpha + \phi E_T(Y_{T+1}) \\
\hat{Y}_{T+1|T} &= \hat{\alpha} + \hat{\phi} Y_T \\
\hat{Y}_{T+2|T} &= \hat{\alpha} + \hat{\phi}\hat{Y}_{T+1|T}
\end{aligned}$$

h-period-ahead

$$\hat{Y}_{T+h|T} = \hat{\alpha} + \hat{\phi}\hat{Y}_{T+h-1|T}$$

- Convenient in linear models, does not work for non-linear models
- less useful when other covariates are used
- More efficient but prone to bias

(h-step) Iterated method: Forecast Variance

Require simulation, 3 ways:

1. errors: $\epsilon_t \sim N(0, \hat{\sigma}^2)$, assume normal error
2. residuals: draw from actual data
3. betas: include parameter uncertainty

```
*errors (betas), residuals (betas)
forecast solve, simulate(errors,
    statistic(stddev,
        predix(sd_)) reps(1000))
```

(h-step) Direct method: Estimation

Estimate h-step regression function
2-period-ahead

$$\begin{aligned}
Y_{T+2} &= (1+\phi)\alpha + \phi^2 Y_T + \epsilon_T + \phi\epsilon_{T+1} \\
&= \alpha^* + \phi^* Y_T + u_T \\
\alpha^* &= (1+\phi)\alpha \\
\phi^* &= \phi^2 \\
u_T &= \epsilon_T + \phi\epsilon_{T+1} \sim MA(h-1)
\end{aligned}$$

h-period-ahead (need h regressions to forecast 1 to h steps)

$$\hat{Y}_{T+h|T} = \hat{\alpha}^* + \hat{\phi}^* Y_T$$

Note:

- Can be estimated directly by OLS
- Minimizes parameter directly
  (different result from iterated and plug-in method)
- error term is not white noise
  (but still uncorrelated with the regressor)
- Can only produce i-period-ahead with $Y_T \sim Y_{T-i}$
- More robust to misspecification (theoretical literature agrees)

(h-step) Direct method: Forecast Variance

Using regression RMSE

$$\hat{\sigma}_u = \sqrt{\frac{1}{T}\sum_{i=1}^{T}\hat{u}_i^2}$$

Remember to adjust for parameter uncertainty as well.
In STATA: predict shat, stdf

AR(p) with intercept

Process model, $\epsilon_t \sim WN(0, \sigma^2)$

$$\begin{aligned}
Y_t &= \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t \\
\Leftrightarrow &(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t = \alpha + \epsilon_t
\end{aligned}$$

Necessary condition for stationarity:

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1$$

Alternative expressions (ADF test regression):

$$\begin{aligned}
Y_t &= \alpha + \gamma_1 Y_{t-1} + \gamma_2 \Delta Y_{t-1} + \cdots + \gamma_p \Delta Y_{t-p+1} + \epsilon_t \\
\Rightarrow \Delta Y_t &= \alpha + (\gamma_1 - 1) Y_{t-1} + \gamma_2 \Delta Y_{t-1} + \cdots + \gamma_p \Delta Y_{t-p+1} + \epsilon_t \\
(\gamma_1 - 1) &= \phi_1 + \phi_2 + \cdots + \phi_n
\end{aligned}$$

AR(p): estimation and forecasting

OLS:

$$Y_t = \hat{\alpha} + \hat{\phi}_1 Y_{t-1} + \hat{\phi} Y_{t-2} + \cdots + \hat{\phi}_p Y_{t-p} + \hat{\epsilon}_t$$

Iterated forecasts:

$$\hat{Y}_{T+h|T} = \hat{\alpha} + \hat{\phi}_1 \hat{Y}_{T+h-1|T} + \hat{\phi}_2 \hat{Y}_{T+h-2|T} + \cdots + \hat{\phi}_p \hat{Y}_{T+h-p|T}$$

Direct forecasts:

$$\begin{aligned}
Y_t &= \hat{\alpha}^* + \hat{\phi}_1^* Y_{t-h} + \hat{\phi}_2^* Y_{t-h-1} + \cdots + \hat{\phi}_p^* Y_{t-h-p+1} + \hat{u}_t \\
\hat{Y}_{T+h|T} &= \hat{\alpha}^* + \hat{\phi}_1^* Y_t + \hat{\phi}_2^* Y_{t-1} + \cdots + \hat{\phi}_p^* Y_{t-p+1}
\end{aligned}$$

ARMA(p, q) processes

Combining AR and MA model.
Use low order, max ARMA(2, 2)
Note: ARMA process has non zero component:
$Cov(Y_{t-1}, Y_{t-2}) \neq 0$
ARMA(1, 1)

$$\begin{aligned}
Y_t = \phi Y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}, \epsilon_t \sim WN(0, \sigma^2) \\
\text{require: } |\phi| < 1 \text{ stationarity }, |\theta| < 1 \text{ invertibility}
\end{aligned}$$

ARMA(p, q)

$$\begin{aligned}
Y_t &= \epsilon_t + \sum_{i=1}^{p}\phi_i Y_{t-i} + \sum_{j=1}^{q}\theta_j\epsilon_{t-j} \\
\Leftrightarrow \Phi(L) Y_t &= \Theta(L)\epsilon_t \\
\Rightarrow Y_t &= \frac{\Theta(L)}{\Phi(L)}\epsilon_t
\end{aligned}$$

Require: all roots of AR/MA polynomial outside the unit circle for stationarity/invertibility

Combining components

Recall:

$$Y_t = T_t + S_t + C_t$$

Trick:

1. Lag the first equation
2. Multiply lagged equation with ϕ
3. Subtract from original equation

Trend + Cycle model

Model:

$$Y_t = T_t + C_t$$

Supposed:

$$\begin{aligned} C_t &= \phi C_{t-1} + \epsilon_t \\ &\sim AR(1) \end{aligned}$$

constant trend

Supposed:

$$\begin{aligned} T_t &= \mu \\ \Rightarrow Y_t &= \mu + C_t \\ \Rightarrow Y_t - \phi Y_{t-1} &= \mu + C_t - \phi(\mu + C_{t-1}) \\ &= (1-\phi)\mu + C_t - \phi C_{t-1} \\ \Rightarrow Y_t &= (1-\phi)\mu + \phi Y_{t-1} + \epsilon_t \\ &\sim AR(1) \end{aligned}$$

linear trend

Supposed:

$$\begin{aligned} T_t &= \mu_1 + \mu_2 t \\ C_t &= \phi C_{t-1} + \epsilon_t \\ \Rightarrow Y_t - \phi Y_{t-1} &= \mu_1 + \mu_2 t + C_t - \phi(\mu_1 + \mu_2(t-1) + C_{t-1}) \\ &= (1-\phi)\mu_1 + \phi\mu_2 + (1-\phi)\mu_2 t + C_t - \phi C_{t-1} \\ \Rightarrow Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 t + \epsilon_t \\ &\sim AR(1) \end{aligned}$$

Trend + AR cycle

1. Constant or Linear time trend:
   regression on trend variable + p lags of Y_t,
   where $C_t \sim AR(p)$
2. Quadratic trend:
   same way (algebra messier)
3. Exponential trend:
   logged series with linear trend

Forecasting is same as AR(p), with trend components

$$Y_t = \alpha + \gamma t + \beta_1 Y_{t-h} + \cdots + \beta_p Y_{t-h-p+1} + \epsilon_t$$

Issue with omitted trend

Issue: $\hat{\beta} \approx 1$ (unit coefficient) on the lag due to misspecification (differ from true β)

$$\begin{aligned} \text{True: } Y_t &= \alpha + \gamma t + \beta Y_{t-1} + \epsilon_t \\ \text{Misspecified: } Y_t &= \hat{\alpha} + \hat{\beta} Y_{t-1} + \hat{\epsilon}_t \end{aligned}$$

For example, true model: $Y_t = \mu_1 + \mu_2 t$

$$\begin{aligned} \text{Estimated: } Y_t &= \mu_2 + Y_{t-1} \\ &= \mu_2 + (\mu_1 + \mu_2(t-1)) \\ \Rightarrow Y_t &= \hat{\alpha} + \hat{\beta} Y_{t-1} + \hat{\epsilon}_t \end{aligned}$$

where $\hat{\alpha} = \mu_2, \hat{\beta} = 1$ (wrong estimation)
Therefore, consider using growth rate instead

Seasonal + Cycle model

Model:

$$Y_t = S_t + C_t$$

Supposed:

$$\begin{aligned} C_t &= \phi C_{t-1} + \epsilon_t \\ &\sim AR(1) \\ S_t &= \sum_{i=1}^{s} \gamma_i D_{it} \\ D_{it} &= I(t = i) \end{aligned}$$

Estimation

$C_t \sim AR(1)$

$$\begin{aligned} Y_t - \phi Y_{t-1} &= S_t + C_t - \phi(S_{t-1} + C_{t-1}) \\ &\Rightarrow Y_t = \phi Y_{t-1} + S_t - \phi S_{t-1} + \epsilon_t \end{aligned}$$

Lagged seasonal dummy is redundant as it perfect collinear with current seasonal dummy

$$\Leftrightarrow Y_t = \phi Y_{t-1} + S_t + \epsilon_t$$

Final estimation

$$\begin{aligned} Y_t &= \alpha_0 + \sum_{t=1}^{s-1} \alpha_1 D_{it} + \beta Y_{t-1} + \epsilon_t \\ &\Leftrightarrow \sum_{i=1}^{s} \alpha_i D_{it} + \beta Y_{t-1} + \epsilon_t \end{aligned}$$

$C_t \sim AR(p)$

$$Y_t = \alpha_0 + \sum_{t=1}^{s-1} \alpha_i D_{it} + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \epsilon_t$$

Trend + Seasonal + Cycle model

Full model:

$$\begin{aligned} Y_t &= T_t + S_t + C_t \\ T_t &= \mu_1 + \mu_2 t \\ S_t &= \alpha_0 + \sum_{i=1}^{s-1} \alpha_i D_{it} \\ C_t &= \phi_1 C_{t-1} + \cdots + \phi_p C_{t-p} + \epsilon_t \end{aligned}$$

Finally

$$Y_t = \alpha_0 + \sum_{i=1}^{s-1} \alpha_i D_{it} + \gamma t + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \epsilon_t$$

Forecasting with regression models

$$Y_t = \alpha + \beta X_t + e_t$$

When conditional mean of Y_t depends on present period X_t, we run into forecasting the right hand side variable problem.
Solution:

- Assume future value of X (scenario analysis)
- Build a model to forecast X

Scenario/contingency analysis

Assume $X_{T+h} = X^*_{T+h}$ based on business assumption

Forecast models for Y and X

First predict X_t, then sub in Y_t

$$\hat{Y}_{T+h|T} = \alpha + \beta X_{T+h}$$
$$\hat{X}_{T+h} = \gamma + \phi X_T$$

Direct forecasts

Combine model for X_t and Y_t

$$\begin{aligned}\hat{Y}_{T+h|T} &= \alpha + \beta(\gamma + \phi X_T) \\ &= \mu + \theta X_T\end{aligned}$$

helps to estimate standard error correctly

Distributed lag models

The general idea of direct forecasts

$$\begin{aligned}Y_t &= \mu + \beta_1 X_{t-1} + \cdots + \beta_k X_{t-k} + e_t \\ &= \mu + B(L)X_{t-1} + e_t\end{aligned}$$

Interpretation of β (under suitable assumption):

- β_k is dynamic multipliers at lag k
- sum of coefficients $B(1)$ is the long-run dynamic multiplier

ADL models

Distributed lag models + AR(p) = autoregressive distributed lag model

$$\begin{aligned}Y_t =& \mu + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} \\ &+ \beta_1 X_{t-1} + \cdots + \beta_k X_{t-k} + e_t \\ A(L)Y_t =& \mu + B(L)X_{t-1} + e_t\end{aligned}$$

h-step ahead forecast

$$\begin{aligned}Y_{t|t-h} =& \mu + \alpha_1 Y_{t-h} + \cdots + \alpha_p Y_{t-p-h+1} \\ &+ \beta_1 X_{t-h} + \cdots + \beta_k X_{t-k-h+1} + e_t\end{aligned}$$

Predictive (Granger) causality

Variable X affects the forecast for Y if (some of) the true coefficients on lags of X in the ADL models are non-zero

- Does not mean causality in the usual cause-and-effect sense. True causality could be the reverse.
- Testing: H_0 : all lags of X jointly = 0 (note: HAC errors with Stock-Watson default lag choice)
- Eaiser to reject H_0 in small in-sample
- In-sample might not work out-of-sample

Volatility modelling

Adjust for the white noise to be non i.i.d
Mean model:

$$Y_t = \mu + \epsilon_t$$
$$\epsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$$

Key insight:

- squared error could potentially be forecastable

$$Var(Y_t|\Omega_{t-1}) = E(\epsilon_t^2|\Omega_{t-1}) = \sigma_t^2$$

- time series is still covariance stationary

Law of Iterated Expectation (LIE)

key trick for computing mean, var in this section

$$\begin{aligned}E(X) &= E(E(X|\Omega)) \\ E(\epsilon_t^2) &= \sigma^2 \\ E(\epsilon_t^2) &= E(E(\epsilon_t^2|\Omega_t)) = E(\sigma_t^2) \\ &= E(model) \\ &= \sigma^2\end{aligned}$$

Conditional Variance

If squared white noise is forecastable, then conditional variance is time varying and serially correlated

- error term is unforecastable (assumed)

$$E(\epsilon_t|\Omega_{t-1}) = 0$$

- conditional var of Y_t is time varying (previously assumed)

$$\begin{aligned}Var(Y_t|\Omega_{t-1}) &= E([Y_t - E(Y_t|\Omega_{t-1})]^2|\Omega_{t-1}) \\ &= E(\epsilon_t^2|\Omega_{t-1}) \\ &= \sigma_t^2 \\ &\neq E(\epsilon_t^2) = \sigma^2\end{aligned}$$

- conditional distribution of $Y_t - E(Y_t) = \epsilon_t$

$$\epsilon_t|\Omega_{t-1} \sim (0, \sigma_t^2)$$

Intuitively, high Volatility tend to be followed by more high volatility days

ARCH(1)

Model conditional var (σ_t^2) with autoregressive dynamics with squared mean-zero series (ϵ_t^2) as a proxy for volatility

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2$$
$$\omega > 0,\ 0 \le \alpha < 1$$

- Constant variance case when $\alpha = 0 \Rightarrow E(\epsilon_t^2|\Omega_{t-1}) = \sigma_t^2 = \sigma^2$
- Spot ARCH by looking at ACF of squared white noise

Unconditional Variance

Solve for σ^2

$$\begin{aligned}\sigma^2 &= E(\epsilon_t^2) = E(E(\epsilon_t^2|\Omega_{t-1})) \\ &\text{by LIE} \\ &= E(\sigma_t^2) \\ &\text{sub in ARCH(1)} \\ &= E(\omega + \alpha\epsilon_{t-1}^2) = \omega + \alpha E(\epsilon_{t-1}^2) \\ &= \omega + \alpha\sigma^2 \\ \Rightarrow \sigma^2 &= \frac{\omega}{(1-\alpha)}\end{aligned}$$

Solve for σ_t^2 in ARCH(1)

$$\begin{aligned}\omega &= \sigma^2(1-\alpha) \\ \Rightarrow \sigma_t^2 &= \sigma^2(1-\alpha) + \alpha\epsilon_{t-1}^2 \\ &= \sigma^2 + \alpha(\epsilon_{t-1}^2 - \sigma^2)\end{aligned}$$

Conditional variance is a combination of unconditional variance and deviation of the squared error from average error value

Estimate ARCH(1) as AR(1)

Model:

$$\sigma_t^2 = E(\epsilon_t^2|\Omega_{t-1}) = \omega + \alpha\epsilon_{t-1}^2$$
$$v_t := \epsilon_t^2 - \sigma_t^2 \text{ (WN)}$$
$$\epsilon_t^2 - \sigma_t^2 + \sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + v_t$$

AR(1) Regression:

$$\epsilon_t^2 = \omega + \alpha\epsilon_{t-1}^2 + v_t$$

Estimate ARCH(1) (with regression parameter estimates):

$$\begin{aligned}\hat{\sigma}_t^2 &= \hat{\omega} + \hat{\alpha}\hat{\epsilon}_{t-1}^2 \\ &= \hat{\omega} + \hat{\alpha}(Y_{t-1} - \hat{\mu})^2\end{aligned}$$

Forecast (1-step ahead):

$$\hat{\sigma}_{t+1|t}^2 = \hat{\omega} + \hat{\alpha}(Y_t - \hat{\mu})^2$$

Forecast interval

Adjust the forecast interval by including estimated variance (varying across time)

$$\hat{Y}_{t+1|t} \pm Z_{\alpha/2}\hat{\sigma}_{t+1|t}$$

ARCH(p)

model

$$\begin{aligned}Y_t &= B(L)\epsilon_t \\ \sigma_t^2 &= \omega + A(L)\epsilon_t^2\end{aligned}$$

where

$$\omega > 0,\ A(L) = \sum_{i=1}^{p} \alpha_i L^i$$

$$\alpha_i \geq 0\ \forall\ i,\ \sum_{i=1}^{p} \alpha_i < 1$$

Note:

- Y_t can be any stationary ARMA model
- large lags (> 10) are usually required for ARCH

Detecting ARCH effects

1. Model conditional mean $Y_t = \beta_0 + \epsilon_t$
2. Check for serial correlation in squared residuals: ACF, Ljung-Box stats

Formal test: Engle's LM test for ARCH effects

$$\epsilon_t^2 = \beta_0 + \sum_{i=1}^{m} \beta_i\epsilon_{t-i}^2 + u_t$$
$$H_0 = \beta_1 = \cdots = \beta_m = 0$$

ARCH(p) order selection

- check PACF of squared residuals ϵ_t^2 from mean model
- AIC/BIC for model selection
- check if ARCH effect is captured well with standardized return

$$\epsilon_t^2|\Omega_{t-1} \sim N(0, \sigma_t^2)$$
$$\Rightarrow \frac{\epsilon_t^2}{\sigma_t^2}\bigg|\Omega_{t-1} \sim N(0, 1)$$

Generalized ARCH(1, 1)

Assume model

$$Y_t = \epsilon_t,\ \epsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$$

GARCH(1, 1):

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$
$$\omega > 0,\ \alpha \geq 0,\ \beta \geq 0,\ \alpha + \beta < 1$$

where variance is a function of all past lags (ARCH(∞))

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \sum_{j=0}^{\infty} \beta^j(\omega + \alpha\epsilon_{t-1-j}^2)\end{aligned}$$

Unconditional Variance

Using Law of Iterated Expectation

$$\begin{aligned}E(\epsilon_t^2) &= E(E(\epsilon_t^2|\Omega_{t-1})) \\ &= E(\sigma_t^2) = E(\omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2) \\ &= \omega + \alpha\sigma^2 + \beta\sigma^2 \\ &= \sigma^2 \\ \Rightarrow \sigma^2 &= \frac{\omega}{1 - \alpha - \beta}\end{aligned}$$

Estimation GARCH(1, 1) as ARMA(1, 1)

Model:

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ v_t &:= \epsilon_t^2 - \sigma_t^2 \text{ (WN)} \\ \epsilon_t^2 - \sigma_t^2 + \sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + v_t \\ \epsilon_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \beta\epsilon_{t-1}^2 - \beta\epsilon_{t-1}^2 + v_t \\ \epsilon_t^2 &= \omega + (\alpha + \beta)\epsilon_{t-1}^2 + \beta(\sigma_{t-1}^2 - \epsilon_{t-1}^2) + v_t\end{aligned}$$

ARMA(1, 1) Regression:

$$\epsilon_t^2 = \omega + (\alpha + \beta)\epsilon_{t-1}^2 - \beta v_{t-1} + v_t$$

Forecast

Forecast (1-step ahead)

$$\begin{aligned}\hat{\sigma}_{t+1|t}^2 &= \hat{\omega} + \hat{\alpha}\hat{\epsilon}_t^2 + \hat{\beta}\hat{\sigma}_t^2 \\ \hat{\epsilon}_t^2 &= (Y_t - \hat{Y}_{t-1})^2 \text{ (squared error)} \\ \hat{\sigma}_t^2 &= \hat{\omega} + \hat{\alpha}\hat{\epsilon}_{t-1}^2 + \hat{\beta}\hat{\sigma}_{t-1}^2 \text{ (fit var iteratively)}\end{aligned}$$

Forecast (2-step ahead)

$$\begin{aligned}\hat{\sigma}_{t+2|t}^2 &= \omega + \alpha E(\epsilon_{t+1}^2|\Omega_t) + \beta\hat{\sigma}_{t+1|t}^2 \\ &= \omega + \alpha\hat{\sigma}_{t+1|t}^2 + \beta\hat{\sigma}_{t+1|t}^2 \\ &= \omega + (\alpha + \beta)\hat{\sigma}_{t+1|t}^2\end{aligned}$$

Forecast (3-step ahead)

$$\begin{aligned}\hat{\sigma}_{t+3|t}^2 &= \omega + \alpha E(\epsilon_{t+2}^2|\Omega_t) + \beta\hat{\sigma}_{t+2|t}^2 \\ &= \omega + \alpha\hat{\sigma}_{t+2|t}^2 + \beta\hat{\sigma}_{t+2|t}^2 \\ &= \omega + (\alpha + \beta)\hat{\sigma}_{t+2|t}^2 \\ &= \omega + (\alpha + \beta)(\omega + (\alpha + \beta)\hat{\sigma}_{t+1|t}^2) \\ &= \omega + (\alpha + \beta)\omega + (\alpha + \beta)^2\hat{\sigma}_{t+1|t}^2\end{aligned}$$

Converging into unconditional variance

h-step forecast is unconditional variance

Forecast error (h-step ahead)

$$\epsilon_{t+h} - E(\epsilon_{t+h}|\Omega_t) = \epsilon_{t+h}$$

Conditional variance

$$\begin{aligned}&E\left[(\epsilon_{t+h} - E(\epsilon_{t+h}|\Omega_t))^2\right] = E(\epsilon_{t+h}^2|\Omega_t) \\ &= \omega\left(\sum_{i=0}^{h-2} \{\alpha(1) + \beta(1)\}^i\right) + (\alpha(1) + \beta(1))^{h-1}\sigma_{t+1}^2\end{aligned}$$

Consider limits

$$\lim_{h\to\infty}\sum_{i=0}^{h-2}\{\alpha(1)+\beta(1)\}^i = \frac{1}{1-\alpha(1)-\beta(1)}$$
$$\lim_{h\to\infty}(\alpha(1)+\beta(1))^{h-1} = 0$$

Therefore

$$\lim_{h\to\infty} E(\epsilon_{t+h}|\Omega_t) = \frac{\omega}{1-\alpha(1)-\beta(1)}$$

optimal forecast converges to the unconditional variance

Generalized ARCH(p, q)

Assume model:

$$Y_t = \epsilon_t,\ \epsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$$

GARCH(p, q):

$$\sigma_t^2 = \omega + \alpha(L)\epsilon_t^2 + \beta(L)\sigma_t^2$$
$$\alpha(L) = \sum_{i=1}^{p}\alpha_i L^i,\ \beta(L) = \sum_{j=1}^{q}\beta_j L^j$$
$$\omega > 0,\ \alpha_i \geq 0,\ \beta_j \geq 0,\ \sum_{i=1}^{p}\alpha + \sum_{j=1}^{q}\beta < 1$$

Note:

- GARCH(p,q) nests ARCH(p) and iid Gaussian WN
- Generally never consider p, q > 2

Limitation and extensions

- Require less parameters than ARCH, works well in practice
- Captures volatility clustering and leptokurtosis (fatter tails)
- Cannot capture asymmetric effect on volatility (leverage effect)

Asymmetric GARCH: Threshold GARCH

Corrects for leverage effect

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \gamma\epsilon_{t-1}^2 I(\epsilon_{t-1} < 0)$$

$I(\epsilon_{t-1} < 0) = 1$ when last period shock was negative

Asymmetric GARCH: Exponential GARCH

Corrects for leverage effect

$$\log(\sigma_t^2) = \omega + \beta\log(\sigma_{t-1}^2) + \alpha\left|\frac{\epsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma\frac{\epsilon_{t-1}}{\sigma_{t-1}}$$

$|\epsilon_{t-1}/\sigma_{t-1}|$ measures absolute magnitude of shock
$\epsilon_{t-1}/\sigma_{t-1}$ measures sign of shock

GARCH in mean: GARCH-M

Expected return (mean) to be positively correlated with volatility $\Rightarrow$ add σ_t^2 to Y_t model

$$Y_t = \beta_0 + \beta_1\sigma_t^2 + \epsilon_t \text{ (added } \sigma_t^2)$$
$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\sigma_{t-1}^2 \text{ (GARCH)}$$

Model the risk return relationship in financial assets

Volatility ground truth

True daily volatility is not observed
Approximation by

- squared daily return (tradition, not recommended)
- 5min high frequency data (new standard)
  realised variance (RV) approximate integrated volatility (IV)

$$RV_t = \sum_{i=1}^{M} r_{t,i}^2$$

Note that only some loss functions (e.g. MSE, QLIKE) are robust to measurement errors and invariant to unit of measurement (assuming proxy is unbiased)

Heterogeneous Autoregressive (HAR)

Uses RV as forecast variable (but avoids long lags of using daily AR model with multi-period realised variance)

$$RV_{t,t+h} = \frac{1}{h}(RV_{t,t+1} + RV_{t,t+2}, \cdots + RV_{t,t+h})$$
$$RV_{t+1} = \alpha + \beta_D RV_t + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \epsilon_{t+1}$$

Includes daily lag, 5-day average, and 22-day average

Robust Regression

Correct for leverage points by weighing observations (lesser weights for large leverage points)