CHARTA LEARNING · LESSON NOTES

Stationary Time Series: AR, MA and ARMA

FRM Part I

1 A story: the dog on a leash

- Leashed dog: wanders but is pulled back to one spot
- Walker with no leash: drifts, spread keeps growing
- Stable series can be described by a few fixed numbers
- That stability is called stationarity

IN THE LESSON

- Picture two walkers in a big park.
- The first has a dog on a leash.
- The dog darts left and right, but the leash always pulls it back toward its owner, so wherever you look, the dog stays near the same spot, with about the same amount of wandering.
- The second walker has no leash and no destination.
- Each step is random, and after an hour nobody can say where he will be.
- The first path is stable, and we can describe it with a few fixed numbers.
- The second keeps changing its own character.
- The first kind of path is called stationary.

2 Covariance stationary: three conditions

- 1. Constant mean: E(y_t) = μ on every date
- 2. Constant, finite variance
- 3. Cov(y_t, y_{t−k}) depends on the gap k only, not on t
- Normality is not required
- Why it matters: sample averages estimate fixed numbers

IN THE LESSON

- A series is covariance stationary when its first two moments do not depend on time.
- There are three conditions.
- First, the mean is constant: the average level is the same on every date.
- Second, the variance is constant and finite: the size of the swings does not change and does not blow up.
- Third, the covariance between today's value and the value k periods ago depends only on the gap k, never on the date.
- Notice what is not required.
- Nothing is said about the shape of the distribution, so normality is not needed.
- And the conditional variance may still change, as long as the unconditional one is constant.
- Stability matters because sample averages then estimate fixed numbers.

3 Series that break the conditions

Which condition does each series break?

Series	Condition broken
Random walk	Variance grows with time
Deterministic trend	Mean moves with time
Explosive, $\|\varphi\| > 1$	Variance not finite
Volatility clustering	Passes: uncond. variance fixed

IN THE LESSON

- The FRM often asks which condition a series breaks, so learn the three typical violators.
- A random walk adds a fresh shock to its last value each period.
- Its variance grows with time, so the second condition fails.
- A series with a deterministic trend, such as a level that rises by the same amount every year, has a mean that moves, so the first condition fails.
- An explosive process has an autoregressive coefficient above one in absolute value, so its swings grow without limit and the variance is not finite.
- Volatility clustering is subtler: the conditional variance moves, yet the unconditional one can stay constant, so such a series can pass.

4 Autocovariance and autocorrelation

$\gamma(k) = \mathrm{Cov}(y_t, y_{t-k})$
$\rho(k) = \gamma(k) \div \gamma(0)$

Autocovariance	Covariance with the past, lag k
Variance	$\gamma(0)$ = variance, so $\rho(0) = 1$
Autocorrelation	Unit-free, between −1 and +1
Symmetry	Symmetric: $\gamma(-k) = \gamma(k)$
Divide by	Always divide by $\gamma(0)$, never $\gamma(1)$

IN THE LESSON

- Now we measure how a stationary series moves with its own past.
- The autocovariance at lag k is the covariance between today's value and the value k periods ago.
- Because the series is stationary it depends only on the lag, so we write it as gamma of k.
- At lag zero it equals the variance, and it is symmetric, so lag minus k gives the same number as lag k.
- The autocorrelation rescales it into a unit free number by dividing by gamma of zero.
- So rho of zero is one, and every autocorrelation lies between minus one and plus one.
- A frequent slip is dividing by the lag one autocovariance, or by some lag specific variance, instead of gamma of zero.

5 Example 1: gamma, rho and a change of units

Illustration: returns in %, $\gamma(0) = 25$ and $\gamma(1) = 10$ (in %²)

Lag-0 autocovariance	= the variance	**25**
Lag-1 autocovariance	$\mathrm{Cov}(y_t, y_{t-1})$	**10**
Autocorrelation $\rho(1)$	$10 \div 25$	**0.40**
Same data in decimals	$0.0010 \div 0.0025$	**0.40**
$\mathrm{Var}(y_t + y_{t-1})$	$25 + 25 + 2\times10$	**70**

IN THE LESSON

- Let's use made up numbers, with returns measured in per cent.
- Suppose the lag zero autocovariance, which is the variance, is twenty five, in squared per cent.
- The lag one autocovariance is ten.
- The autocorrelation at lag one is ten divided by twenty five, which is point four.
- Now switch the same data to decimals.
- Each autocovariance carries the units twice, so it is multiplied by point zero zero zero one: the variance becomes point zero zero two five and the lag one autocovariance point zero zero one zero.
- Dividing gives point four again.
- The autocorrelation does not care about units.
- Last, the variance of a sum of two neighbouring values is twenty five plus twenty five plus twice ten, seventy.

6 White noise: the basic building block

- White noise: mean 0, variance σ^2, zero autocorrelation
- Key word: uncorrelated, not independent
- $\text{Var}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3) = 3\sigma^2$, since terms are uncorrelated
- Independent: any function of the draws is unpredictable
- Gaussian: independent and normally distributed

IN THE LESSON

- White noise is the raw material of time series models.
- It is a sequence of shocks, written epsilon, with three properties.
- The mean is zero.
- The variance is a constant, finite number, sigma squared.
- And the autocorrelation is zero at every lag other than zero.
- That is all the definition demands, so the key word is uncorrelated, not independent.
- The variance of a sum of white noise terms is the sum of their variances.
- Independent white noise is stronger: the draws are statistically independent, so even their squares are unpredictable.
- Gaussian white noise adds a normal distribution to each draw.
- For normal variables zero correlation does imply independence, so Gaussian white noise is automatically independent.

7 Three flavors of white noise

Each flavor adds a requirement to the one above

Type	What it demands
White noise	Mean 0, variance σ^2, uncorrelated
Independent white noise	Draws fully independent
Gaussian white noise	Independent and normal
Reverse direction	Does not hold in general
Volatility clustering	White noise, not independent

IN THE LESSON

- Each flavor adds a requirement.
- Plain white noise needs a zero mean, constant variance and no autocorrelation.
- Independent white noise needs fully independent draws.
- Gaussian white noise adds the normal shape.
- Every Gaussian white noise series is independent, and every independent one is white noise, but the reverse does not hold in general.
- The classic example is a return series with volatility clustering.
- Its returns can be uncorrelated, so it is white noise, yet the squared returns are autocorrelated, because big moves follow big moves.
- So the draws are not independent.

8 The AR(1) process

$y_t = \delta + \varphi \cdot y_{t-1} + \varepsilon_t$
Mean $= \delta \div (1 - \varphi)$, Var $= \sigma^2 \div (1 - \varphi^2)$, $\rho(k) = \varphi^k$

AR coefficient, phi	Pull of yesterday, $\lvert\varphi\rvert < 1$
Variance	Variance: $\sigma^2 \div (1 - \varphi^2)$
Autocorrelation	$\rho(k) = \varphi^k$, decays geometrically
Cut-off signature	AR(p): partial ACF cuts off at p

IN THE LESSON

- An autoregressive process regresses a series on its own past.
- The simplest case, the A R one, says today's value equals an intercept, delta, plus phi times yesterday's value, plus a white noise shock with variance sigma squared.
- The process is covariance stationary when the roots of its characteristic equation lie outside the unit circle, which for the A R one just means phi is below one in absolute value.
- Then the mean is delta over one minus phi, the variance is sigma squared over one minus phi squared, and the autocorrelation at lag k is phi to the power k, a smooth geometric decay.

9 Example 2: properties of an AR(1)

Illustration: $y_t = 2 + 0.6 \cdot y_{t-1} + \varepsilon_t$, with $\sigma^2 = 4$

Mean	$2 \div (1 - 0.6)$	**5**
Variance	$4 \div (1 - 0.36)$	**6.25**
ρ(1)	0.6	**0.6**
ρ(2)	0.6^2	**0.36**
ρ(3)	0.6^3	**0.216**

IN THE LESSON

- Take an A R one with intercept two, phi equal to point six, and shock variance four.
- These are illustrative numbers.
- The mean is two divided by one minus point six, which is five.
- The variance is four divided by one minus point three six, so four over point six four, which is six point two five.
- Now the autocorrelations.
- At lag one it is point six.
- At lag two it is point six squared, point three six.
- At lag three it is point six cubed, point two one six.
- If phi were negative, the signs would alternate.

10 Example 3: computing an AR(1) series

Start $y_0 = 8$. Shocks: +1, −2, +0.5. Recursion adds $\delta = 2$ each time

y_1	$2 + 0.6 \times 8 + 1$	**7.8**
y_2	$2 + 0.6 \times 7.8 - 2$	**4.68**
y_3	$2 + 0.6 \times 4.68 + 0.5$	**5.308**
Slip: forgot the intercept	$0.6 \times 8 + 1$	**5.8 (wrong)**

IN THE LESSON

- Now compute the series itself.
- Use the same model, intercept two and phi point six, and start from a value of eight.
- The three shocks are plus one, minus two and plus point five.
- Period one: two plus point six times eight, which is six point eight, plus the shock of one, gives seven point eight.
- Period two: two plus point six times seven point eight, which is six point six eight, minus two, gives four point six eight.
- Period three: two plus point six times four point six eight, which is four point eight zero eight, plus point five, gives five point three zero eight.
- The usual error is forgetting the intercept.
- Without it, period one would wrongly be five point eight.

11 The MA(1) process

$y_t = \mu + \varepsilon_t + \theta \cdot \varepsilon_{t-1}$
$\text{Var} = \sigma^2(1 + \theta^2)$, $\rho(1) = \theta \div (1 + \theta^2)$

MA coefficient, theta	Weight on yesterday's shock
Variance	Variance: $\sigma^2(1 + \theta^2)$
Autocorrelation	$\rho(1) = \theta \div (1 + \theta^2)$
Lag-one autocovariance	$\gamma(1) = \theta\sigma^2$
Cut-off signature	MA(q): ACF = 0 beyond lag q

IN THE LESSON

- A moving average process is built from current and past shocks rather than past values.
- The M A one says today's value equals a mean, mu, plus today's shock, plus theta times yesterday's shock.
- It is always covariance stationary, whatever theta is.
- Its variance comes from the two shocks, sigma squared times one plus theta squared.
- The lag one autocovariance is theta times sigma squared, because only the shared shock links neighbors.
- Dividing, the lag one autocorrelation is theta over one plus theta squared, and it can never exceed one half in absolute value.
- In general, for a moving average of order q, the autocorrelation is exactly zero beyond lag q, which is the signature that identifies it.

12 Example 4: properties of an MA(1)

Illustration: $y_t = \mu + \varepsilon_t + 0.5\cdot\varepsilon_{t-1}$, with $\sigma^2 = 4$

Variance	4 × (1 + 0.25)	**5**
γ(1)	0.5 × 4	**2**
ρ(1)	0.5 ÷ 1.25	**0.40**
ρ(2)	beyond lag q = 1	**0**
Largest possible ρ(1)	θ = 1: 1 ÷ 2	**0.5**

IN THE LESSON

- Take an M A one with theta equal to point five and shock variance four.
- The variance is four times one plus point two five, which is five.
- The lag one autocovariance is point five times four, which is two.
- So the lag one autocorrelation is two over five, point four, or equivalently point five over one point two five.
- At lag two, the autocorrelation is exactly zero, because two periods apart the series shares no shock.
- That sharp cut off is how an M A shows up in data.
- Finally, the largest the lag one autocorrelation can ever be arises at theta equal to one: one over one plus one, which is one half.
- No M A one can have an autocorrelation above that.

13 Example 5: computing an MA(1) series

Illustration: $\mu = 10$, $\theta = 0.5$. Shocks $\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3 = 1, 2, -1, 0.5$

y_1	10 + 2 + 0.5×1	**12.5**
y_2	10 − 1 + 0.5×2	**10**
y_3	10 + 0.5 + 0.5×(−1)	**10**
Invertibility	needs |θ| < 1	**0.5 yes, 2 no**
ρ(1) if θ = 2 instead	2 ÷ (1+4), same as θ = 0.5	**0.40**

IN THE LESSON

- To compute an M A series, add the mean, the current shock, and theta times the previous shock.
- Let the mean be ten, theta point five, and the shocks in order be one, two, minus one and point five.
- Period one: ten plus two plus point five times one, which gives twelve point five.
- Period two: ten minus one plus point five times two, which gives ten.
- Period three: ten plus point five plus point five times minus one, which gives ten.
- One more subtlety.
- Invertibility, theta below one in absolute value, is a separate condition.
- Theta of two gives the same autocorrelation of point four as theta of point five.

14 The lag operator and the lag polynomial

- L shifts back: $L\cdot y_t = y_{t-1}$, and $L^k\cdot y_t = y_{t-k}$
- AR(1): $(1 - \varphi L)\cdot y_t = \varepsilon_t$, polynomial $1 - \varphi L$
- AR(2): $1 - \varphi_1 L - \varphi_2 L^2$ (coefficients flip sign)
- Characteristic equation: polynomial = 0, $L \to z$
- Stationary: every root $|z| > 1$ ($|z| = 1$: unit root)

IN THE LESSON

- The lag operator, written L, shifts a series back one period.
- Applied to today's value it gives yesterday's, and applying it k times shifts back k periods.
- The A R one becomes one minus phi L, applied to y, equals the shock.
- The lag polynomial is one minus phi L.
- For an A R two it is one minus phi one L minus phi two L squared.
- Coefficients enter with the opposite sign to the equation, which is where most slips occur.
- To judge stationarity, replace L by z and set the polynomial to zero.
- The process is stationary only if every root lies outside the unit circle, meaning its absolute value exceeds one.
- A root of exactly one is a unit root.

15 Example 6: judging stationarity by roots

Illustration: roots of $1 - \varphi_1 z - \varphi_2 z^2 = 0$

Case 1: φ = 0.5 and 0.3	$1 - 0.5z - 0.3z^2 = 0$	**1.17, −2.84**
Both outside the circle?	$\|1.17\| > 1$, $\|-2.84\| > 1$	**Stationary**
Case 2: φ = 0.7 and 0.4	$1 - 0.7z - 0.4z^2 = 0$	**0.93, −2.68**
One root inside the circle	$\|0.93\| < 1$	**Not stationary**

IN THE LESSON

- Take two illustrative A R twos.
- The first has phi one equal to point five and phi two equal to point three.
- Its characteristic equation is one minus point five z minus point three z squared, equal to zero.
- Solving gives roots of about one point one seven and minus two point eight four.
- Both have absolute value above one, so they lie outside the unit circle, and the process is stationary.
- The second has coefficients point seven and point four.
- The equation is one minus point seven z minus point four z squared, equal to zero, with roots near point nine three and minus two point six eight.
- The first root has absolute value below one, so it sits inside the circle, and the process is not stationary.

16 Example 7: mean reversion and its level

Illustration: AR(1) with δ = 2, φ = 0.6. AR(2) with δ = 1.5, φ = 0.5, 0.3

AR(1) level	$2 \div (1 - 0.6)$	**5**
AR(2) level	$1.5 \div (1 - 0.5 - 0.3)$	**7.5**
Gap when starting at 8	$8 - 5$	**3**
Gap one period later	0.6×3	**1.8**
Gap two periods later	0.6×1.8	**1.08**

IN THE LESSON

- Mean reversion means a stationary series is pulled back toward a fixed long run level, so gaps fade over time.
- For an A R one, that level is the unconditional mean, delta over one minus phi.
- With intercept two and phi point six, the level is five, not two.
- For an A R two, use the signed coefficients: intercept one point five, phi one point five and phi two point three give one point five over point two, which is seven point five.
- Now the gap.
- A series starting at eight is three above its mean of five.
- Each period the gap shrinks by the factor phi: one point eight, then one point zero eight.
- A larger positive phi means slower reversion.

17 ARMA: autoregressive plus moving average

- ARMA(1,1): $y_t = \delta + \varphi \cdot y_{t-1} + \varepsilon_t + \theta \cdot \varepsilon_{t-1}$
- Stationarity: only the AR roots matter
- Mean = $\delta \div (1 - \varphi)$; θ does not enter the mean
- ACF, PACF (partial ACF) both decay, no clean cut-off
- If $\varphi + \theta = 0$, the series is white noise around its mean

IN THE LESSON

- An A R M A with orders p and q combines both ideas: it has p lags of the series and q lags of the shock.
- The A R M A one one is today's value equals an intercept, plus phi times yesterday's value, plus today's shock, plus theta times yesterday's shock.
- Stationarity depends only on the autoregressive part, whose roots must lie outside the unit circle.
- The moving average part never threatens it.
- The mean ignores theta and equals delta over one minus phi.
- Both autocorrelations and partial autocorrelations decay gradually, with no clean cut off.
- If phi plus theta is zero, the series is white noise around its mean.

18 Example 8: ARMA(1,1) autocorrelation

Illustration: $\delta = 1$, $\varphi = 0.5$, $\theta = 0.4$

Mean	$1 \div (1 - 0.5)$	**2**
Numerator	$(1 + 0.2)(0.5 + 0.4)$	**1.08**
Denominator	$1 + 0.4 + 0.16$	**1.56**
ρ(1)	$1.08 \div 1.56$	**0.692**
ρ(2)	0.5×0.692	**0.346**

IN THE LESSON

- Let's work an A R M A one one with intercept one, phi point five and theta point four.
- These are illustrative figures.
- The mean is one over one minus point five, which is two.
- Now the lag one autocorrelation, using the formula one plus phi theta, times phi plus theta, all over one plus two phi theta plus theta squared.
- The phi theta product is point two.
- The numerator is one point two times point nine, which is one point zero eight.
- The denominator is one plus point four plus point one six, which is one point five six.
- So rho one is one point zero eight over one point five six, about point six nine two.
- After lag one, the autocorrelation decays at rate phi, so lag two is point five times point six nine two, which is about point three four six.

19 When to use which model, and its signature

Read the sample ACF and PACF to shortlist candidates

Model	Use and signature
AR(p)	PACF cuts off after lag p
MA(q)	ACF cuts off after lag q
ARMA(p, q)	Both ACF and PACF decay
Typical AR uses	Rates, spreads, volatility

IN THE LESSON

- Autoregressive models suit persistent, mean reverting series such as interest rates, credit spreads or volatility.
- Moving average models suit series where a shock fades after a fixed number of periods.
- A R M A mixes both, often with fewer parameters than a long pure A R or pure M A would need.
- The sample patterns guide the choice.
- An A R of order p has a partial autocorrelation that cuts off after lag p, while its autocorrelation decays.
- An M A of order q has an autocorrelation that cuts off after lag q, while its partial autocorrelation decays.
- An A R M A shows decay in both.
- Patterns only suggest candidates.

20 Choosing a model: AIC and BIC

AIC = ln(σ²) + 2k ÷ T
BIC = ln(σ²) + k·ln(T) ÷ T

Residual variance	Fit: lower variance is better
Coefficients, k	Estimated coefficients
Observations, T	Sample size
Penalty	BIC heavier if T ≥ 8

IN THE LESSON

- To choose among candidates, estimate each one and record its residual variance from T observations with k estimated coefficients.
- Then compute an information criterion.
- The Akaike criterion is the log of the residual variance plus two k over T.
- The Schwarz criterion, also called the Bayesian criterion, is the log of the residual variance plus k times the log of T, over T.
- Lower is better.
- The first term rewards fit, and the second penalizes extra parameters.
- The Bayesian penalty is heavier whenever the log of T exceeds two, which is true once T is eight or more, so it favors smaller models, while Akaike tends to pick larger ones.
- Choosing by smallest residual variance always favors the biggest model.

21 Example 9: AIC and BIC disagree

Illustration: T = 100, ln T = 4.605. Values shown as AIC | BIC

AR(1): σ² 1.60, k = 1	ln σ² = 0.4700	**0.4900|0.5161**
ARMA(1,1): σ² 1.50, k = 2	ln σ² = 0.4055	**0.4455|0.4976**
ARMA(2,2): σ² 1.44, k = 4	ln σ² = 0.3646	**0.4446|0.5488**
AIC picks	smallest AIC	**ARMA(2,2)**
BIC picks	smallest BIC	**ARMA(1,1)**

IN THE LESSON

- Suppose we have one hundred observations, so the log of T is four point six zero five.
- Three candidate models are fitted, and the numbers are illustrative.
- The A R one has residual variance one point six and one coefficient.
- The A R M A one one has variance one point five and two coefficients.
- The A R M A two two has variance one point four four and four coefficients.
- The criteria come out as follows.
- For the A R one, Akaike is point four nine zero zero and Bayesian is point five one six one.
- For the A R M A one one, point four four five five and point four nine seven six.
- For the A R M A two two, point four four four six and point five four eight eight.
- Akaike picks the largest model, narrowly.
- The Bayesian criterion picks the A R M A one one.

22 Example 10: the sample autocorrelation

Illustration: data 1, 2, 3, 5, 4. Mean 3, deviations −2, −1, 0, 2, 1

Sum of squares (all T)	4 + 1 + 0 + 4 + 1	**10**
Lag-1 products	2 + 0 + 0 + 2	**4**
Sample ρ(1)	4 ÷ 10	**0.40**
Lag-2 products	0 − 2 + 0	**−2**
Sample ρ(2)	−2 ÷ 10	**−0.20**

IN THE LESSON

- Sample statistics estimate the theoretical functions from data.
- The sample autocorrelation at lag k sums the products of deviations from the mean for the available pairs, and divides by the sum of squared deviations over all observations.
- Take five illustrative observations: one, two, three, five and four.
- The mean is three, so the deviations are minus two, minus one, zero, two and one.
- The squares add up to ten.
- At lag one, the products are two, zero, zero and two, adding to four, so the autocorrelation is four over ten, point four.
- At lag two, the products are zero, minus two and zero, adding to minus two, so the autocorrelation is minus point two.

23 Example 11: partial autocorrelation and the band

Illustration: ρ(1) = 0.6, ρ(2) = 0.5, T = 100

PACF at lag 1	same as ρ(1)	**0.60**
PACF at lag 2	(0.5 − 0.36) ÷ (1 − 0.36)	**0.219**
Standard error	1 ÷ √100	**0.10**
5% band	±1.96 × 0.10	**±0.196**
Sample value 0.30	outside the band	**Significant**

IN THE LESSON

- The partial autocorrelation at lag k is the correlation between today and k periods ago after removing the effect of the lags in between.
- At lag one the two measures coincide.
- At lag two it equals rho two minus rho one squared, over one minus rho one squared.
- Suppose rho one is point six and rho two is point five.
- Then point five minus point three six is point one four, divided by point six four, giving about point two one nine.
- For significance, under white noise the standard error is one over the square root of T.
- With one hundred observations it is point one, so the band is plus or minus point one nine six.
- A sample value of point three is outside, hence significant.

24 Box-Pierce and Ljung-Box Q statistics

$Q_{BP} = T \times \Sigma(k=1..m)\ \hat{\rho}_k^2$
$Q_{LB} = T(T+2) \times \Sigma(k=1..m)\ \hat{\rho}_k^2 \div (T-k)$

Sample autocorrelation	Squared first, then summed
Lags tested, m	Lags 1 to m tested together
Observations, T	Sample size
Chi-squared test	Null: χ^2 with df = m

IN THE LESSON

- Both tests ask whether the first m autocorrelations of a series, usually model residuals, are jointly zero.
- The Box Pierce statistic is T times the sum of the squared sample autocorrelations from lag one to lag m.
- The Ljung Box statistic reweights each term.
- Each squared autocorrelation is multiplied by T times T plus two, divided by T minus k, instead of just T.
- That suits small samples better, so Ljung Box is preferred.
- Under the null of no autocorrelation, each follows a chi squared with m degrees of freedom.
- When testing A R M A residuals, reduce that by the number of estimated A R M A parameters.
- A large Q rejects the null.

25 Example 12: testing the residuals

Illustration: T = 100, residual $\hat{\rho}$ at lags 1–3 = 0.25, 0.15, 0.10

Σ ρ^2	0.0625 + 0.0225 + 0.01	**0.095**
Box-Pierce	100 × 0.095	**9.5**
Ljung-Box	100×102×Σ ρ^2÷(100−k)	**9.83**
5% critical, df = 3	chi-squared table	**7.815**
ARMA(1,1) residuals	df = 3 − 2 = 1	**crit 3.841**

IN THE LESSON

- Suppose we have one hundred residuals and the first three sample autocorrelations are point two five, point one five and point one.
- These are illustrative numbers.
- Square them: point zero six two five, point zero two two five and point zero one.
- They sum to point zero nine five.
- The Box Pierce statistic is one hundred times that, nine point five.
- For Ljung Box, each term is divided by one hundred minus its lag, and the sum is multiplied by one hundred times one hundred and two.
- The result is about nine point eight three, slightly larger.
- The critical value for a chi squared with three degrees of freedom at five per cent is seven point eight one five, so we reject.
- But if these were residuals from an A R M A one one, with two estimated parameters, there is only one degree of freedom, with critical value three point eight four one.

26 Forecasting from ARMA models

- Future shocks → 0; past residuals kept as estimated
- Future y values → their own forecasts
- ARMA(1,1) one step: δ + φ·y_T + θ·e_T
- Two steps: δ + φ·(one-step forecast), MA term gone
- MA(q) forecast = the mean beyond q steps

IN THE LESSON

- Forecasts from A R M A models are conditional expectations given everything known at time T.
- The rules are simple.
- Replace each future shock by its expected value, zero.
- Keep known past shocks, the residuals, at their estimated values.
- And replace future values of the series by their own forecasts.
- For an A R M A one one, the one step forecast is the intercept plus phi times the last value plus theta times the last residual.
- At two steps, the shock at time T plus one is unknown and set to zero, so the moving average term drops out, and the forecast is the intercept plus phi times the one step forecast.
- A pure M A forecast equals the mean beyond q steps, while an A R forecast keeps feeding on earlier forecasts.

27 Example 13: an ARMA(1,1) forecast

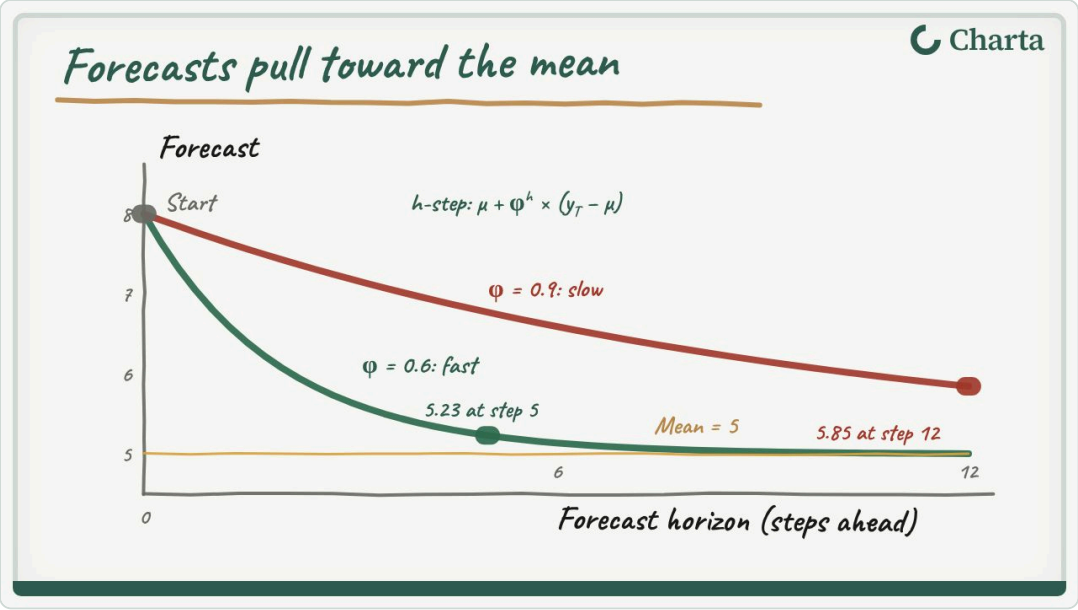
Illustration: δ = 1, φ = 0.5, θ = 0.4. Last value 3, last residual 0.5

1 step ahead	1 + 0.5×3 + 0.4×0.5	**2.7**
2 steps ahead	1 + 0.5×2.7	**2.35**
3 steps ahead	1 + 0.5×2.35	**2.175**
Long-run mean	1 ÷ (1 − 0.5)	**2**
One-step error variance	only the new shock	**σ^2**

IN THE LESSON

- Take the A R M A one one with intercept one, phi point five and theta point four.
- The mean is two.
- Suppose the last observed value is three and the last residual is point five.
- The one step forecast is one plus point five times three plus point four times point five, which is one plus one point five plus point two, giving two point seven.
- For two steps, the unknown shock is set to zero, so the forecast is one plus point five times two point seven, which is two point three five.
- For three steps it is one plus point five times two point three five, which is two point one seven five.
- And the one step forecast error is just the unforecastable shock, so its variance is sigma squared.

28 Forecasts pull toward the mean



IN THE LESSON

- For a stationary A R one, the h step forecast is the mean plus phi to the power h times today's gap from the mean.
- The gap shrinks geometrically, so forecasts converge on the long run mean as the horizon grows.
- Here the mean is five and the series starts at eight, a gap of three.
- The lower curve has phi point six, so it reaches the mean quickly: after five steps it is about five point two three.
- The upper curve has phi point nine, a very persistent series, and it is still at five point eight five after twelve steps.
- The more persistent the series, the slower the convergence.

29 Example 14: error variance at long horizons

Var(h) = σ²(1 + φ² + … + φ^{2h−2}). Here φ = 0.6, σ² = 4

h = 1	4	**4**
h = 2	4 × (1 + 0.36)	**5.44**
h = 3	4 × (1 + 0.36 + 0.1296)	**5.9584**
Limit as h grows	4 ÷ (1 − 0.36)	**6.25**
Std dev at h = 3	wrong 2×√3; right √5.9584	**3.46 vs 2.44**

IN THE LESSON

- While the forecast settles on the mean, the forecast error variance rises toward the unconditional variance.
- For an A R one, the h step error variance is sigma squared times one plus phi squared, and so on, up to phi to the power two h minus two.
- Take phi point six and sigma squared four.
- One step ahead the variance is four.
- Two steps: four times one point three six, which is five point four four.
- Three steps: four times one point four eight nine six, which is five point nine five eight four.
- The limit is four over point six four, which is six point two five.
- It stops growing, unlike a random walk.
- A frequent error is to scale the one step standard deviation of two by the square root of three, giving three point four six.
- The true figure is two point four four.

30 Seasonality in an ARMA model

(1 − φL)(1 − ΦL^s)·y_t = ε_t
y_t = φ·y_{t−1} + Φ·y_{t−s} − φΦ·y_{t−s−1} + ε_t

AR coefficient, phi	Lag-1 coefficient; L shifts back
Seasonal coefficient, Phi	Seasonal AR (MA: shock ε_{t−s})
Season length, s	12 for monthly, 4 for quarterly
Extra lag, s + 1	Coefficient is −φ × Φ
Cut-off signature	ACF spikes at lags s, 2s: seasonal

IN THE LESSON

- Seasonality is a regular pattern at a fixed period s, such as twelve for monthly data or four for quarterly data.
- A covariance stationary A R M A model can capture it by adding terms at the seasonal lag.
- A seasonal autoregression includes the value s periods ago with its own coefficient, and a seasonal moving average includes the shock s periods ago.
- In the multiplicative form, the polynomials multiply: one minus phi L, times one minus capital phi L to the s, applied to the series, equals the shock.
- Expanded, the model has a lag at one, a lag at s, and an extra lag at s plus one, whose coefficient is minus the product of the two.
- Spikes in the sample autocorrelation at lags s and two s point to seasonality.

31 Example 15: a quarterly seasonal AR

Illustration: quarterly, s = 4, φ = 0.5, Φ = 0.4

Lag 1 coefficient	φ	**0.5**
Lag 4 coefficient	Φ	**0.4**
Lag 5 coefficient	−φ × Φ = −0.5 × 0.4	**−0.2**
Stationary?	|0.5| < 1 and |0.4| < 1	**Yes**

IN THE LESSON

- Take quarterly data, so s equals four, with illustrative coefficients phi equal to point five and capital phi equal to point four.
- Multiply out the polynomials: one minus point five L, times one minus point four L to the fourth.
- The expanded model has three lag coefficients.
- The lag one coefficient is point five.
- The lag four coefficient is point four.
- And a lag five term appears, with coefficient minus the product, minus point five times point four, which is minus point two.
- Stationarity needs the seasonal coefficient to be below one in absolute value, as well as phi, and here both are, so the model is stationary.

32 Four mistakes to avoid

- Intercept ≠ mean: use δ ÷ (1 − φ)
- Uncorrelated ≠ independent, unless Gaussian
- Smallest σ² picks the biggest model: use AIC or BIC
- Do not scale the one-step sd by √h: variance levels off

IN THE LESSON

- Mistake one: treating the intercept as the long run level.
- The mean of an autoregression is delta over one minus phi, and it equals delta only when phi is zero.
- Mistake two: confusing uncorrelated with independent.
- White noise only needs zero autocorrelation, and only the Gaussian form guarantees independence, so volatility clustering can hide in white noise.
- Mistake three: choosing a model by the smallest residual variance.
- That always favors the biggest model, so use Akaike or Bayesian criteria with their penalties.
- Mistake four: scaling the one step standard deviation by the square root of the horizon.
- For a stationary model, the error variance levels off at the unconditional variance, and the forecast settles on the mean.

33 Recap: the main ideas

- Stationary: fixed mean, finite variance, γ(k) only on k
- ρ(k) = γ(k) ÷ γ(0); white noise is uncorrelated
- AR(1): mean δ÷(1−φ), ρ(k) = φ^k; PACF cuts off at p
- MA: always stationary; ACF cuts off after q
- Roots outside; AIC/BIC; Ljung-Box; forecasts → mean

IN THE LESSON

- Let's pull it together.
- Covariance stationary means constant mean, constant finite variance, and covariance that depends only on the lag.
- The autocorrelation is the autocovariance divided by the variance.
- White noise is uncorrelated, and Gaussian white noise is also independent.
- An A R one has mean delta over one minus phi, geometric autocorrelation, and a partial autocorrelation that cuts off.
- An M A is always stationary and its autocorrelation cuts off after q.
- Stationarity means characteristic roots outside the unit circle.
- Information criteria choose the model, and the Ljung Box test checks the residuals.
- Forecasts replace future shocks with zero and converge on the mean, while seasonal terms enter at lag s.

34 Try three quick questions

1. AR(1): δ = 3, φ = 0.5, σ^2 = 3. Mean and variance?
   Answer: Mean = 3 ÷ 0.5 = 6. Variance = 3 ÷ 0.75 = 4.
2. MA(1) with θ = 0.6. Lag-1 autocorrelation?
   Answer: 0.6 ÷ (1 + 0.36) = 0.6 ÷ 1.36 ≈ 0.44.
3. AR(1): mean 5, φ = 0.6, last value 8. Two-step forecast?
   Answer: 5 + 0.6^2 × (8 − 5) = 5 + 1.08 = 6.08.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- An A R one has intercept three, phi point five and shock variance three.
- What are its mean and variance?
- The mean is three over point five, which is six, and the variance is three over point seven five, which is four.
- Question two.
- An M A one has theta point six.
- What is its lag one autocorrelation?
- Point six over one point three six, about point four four.
- Question three.
- An A R one has mean five and phi point six, and the last value is eight.
- What is the two step forecast?
- Five plus point three six times three, which is six point zero eight.