WARWICK
Business School

Empirical Finance
Lecture 7: Analysis of long memory and non-stationary processes: part I.

Module Leader: Dr Stuart Fraser

stuart.fraser@wbs.ac.uk

Room D1.18 (Social Studies)

The University of Warwick

Today

- Long Memory Processes (Mills Chp 3.4)
  - Testing for long memory in the forward premium.
- Analysis of Non-stationary Processes: Part I (Brooks Chp 7.1-7.2)
  - Testing for autoregressive unit roots ('unit roots') in economic/finance data.
- Seminar 6: Testing for long memory and unit roots in the real exchange rate.

Long memory processes (Mills Chp 3.4)

Example of a long memory process (Fractional White Noise)

$$y_t = (1-L)^{-d}\varepsilon_t$$

d is a real number - it can take fractional values

Binomial Expansion (see Appendix 1)

$$= \left(1 + dL + \frac{d(d+1)}{2!}L^2 + \frac{d(d+1)(d+2)}{3!}L^3 + \dots\right)\varepsilon_t$$

$$= \sum_{k=0}^{\infty}\psi_k \varepsilon_{t-k}$$

Fractional sum/integral of a white noise process.

The ψ weights (Wold form coefficients) will only decay if d<1

$$\text{If } d = 1 \Rightarrow y_t = \sum_{k=0}^{\infty}\varepsilon_{t-k} \text{ (Random walk/Martingale model)}$$

$\Rightarrow$ Shocks have a permanent effect on the level of y.

The process will display mean reversion for d<1.

Long memory processes

However the process is only covariance (weakly) stationary if d<0.5.

The ACF of FWN is given by:

$$\rho_k = ck^{2d-1}$$

If d<0.5 the ACF decays hyperbolically (slowly) to zero.

- $\Rightarrow$ Possible to have a FWN process which is both mean reverting (d<1) and non-stationary (d≥0.5)!

Compare this with the fast geometric/exponential decay of the ACF for stationary ARMA models.

For example the ACF of an AR(1) process is:

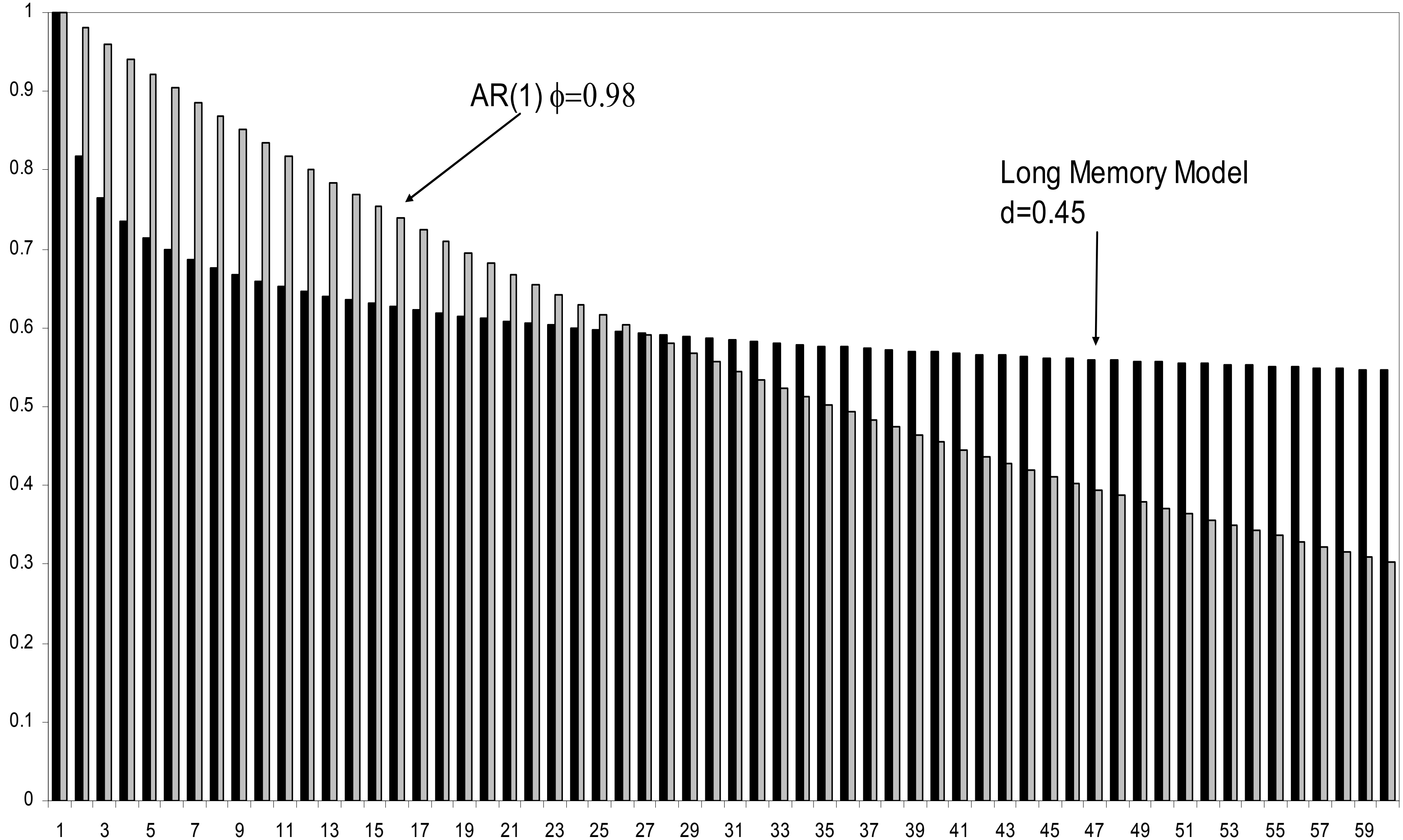
$$\rho_k = \phi^k$$

The stationarity condition is:

$$|\phi| < 1$$

(see lecture 5)

ACF of AR(1) and FWN processes (geometric vs hyperbolic decay)



Frequency domain/spectral analysis

Auto-covariance Generating Function

$$g_y(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$$

The AGF summarizes the auto-covariances/memory of a time-series: $\gamma_k = \text{cov}(y_t, y_{t-k})$

The AGF is finite if the autocovariances are absolutely summable (roughly this equates to stationary processes).

Population Spectrum

$$\begin{aligned} f_y(\lambda) &= \frac{1}{2\pi} g_y\left(e^{-i\lambda}\right) \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\lambda k} \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k \left[\cos(\lambda k) - i \sin(\lambda k)\right] \\ &= \frac{1}{2\pi} \left\{\gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos(\lambda k)\right\} \end{aligned}$$

Time series are made up of cyclical/periodic components with different frequencies λ:

For example... Seasonal components in a time-series have a high frequency (they repeat over a short period).
Long-run trend components have a low frequency (they repeat over very long periods).

To examine the importance of cyclical components at different frequencies we need to analyze the spectrum of the process.

See Appendix 2 for a derivation of the last line.

Frequency domain/spectral analysis

The population spectrum measures the portion of the variance of y which is attributable to periodic components with frequency λ:

- Analysis of the spectrum is referred to as frequency domain analysis.

Analysis of autocovariances/autocorrelations is referred to as time domain analysis.

The spectrum and autocovariances are simply ‘two-sides of the same coin’:

- The spectrum is just a function of the autocovariances (and vice-versa): they contain the same information (albeit expressed differently).
- Whether you analyze the spectrum or the autocovariances is simply a matter of context.

Frequency domain/spectral analysis

- λ measures the frequency of the periodic components in radians: it can take any value in the range [- π, π] .
  - But the spectrum is symmetric about zero (for real valued time series) so usually we only need to consider the range [0, π] .

The frequencies are calculated as follows:

$$\lambda_j = \frac{2\pi j}{T}, \text{ where}$$

$$j = 1,\ldots,\frac{T}{2} \qquad (T \text{ even})$$

$$j = 1,\ldots,\frac{T-1}{2} \quad (T \text{ odd}).$$

The period of the cycles are given by T/j:

- The shortest cycles repeat every 2 periods (period=T/(T/2)=2).
- The longest cycles repeat every T periods (period=T/1=T)
  - As $T\rightarrow\infty$ these cycles never repeat $\Rightarrow$ long-run trend component.
  - We look at the 'frequency zero' part of the spectrum to analyze the importance of long-run trends in the data.

Spectral examples

1. White noise

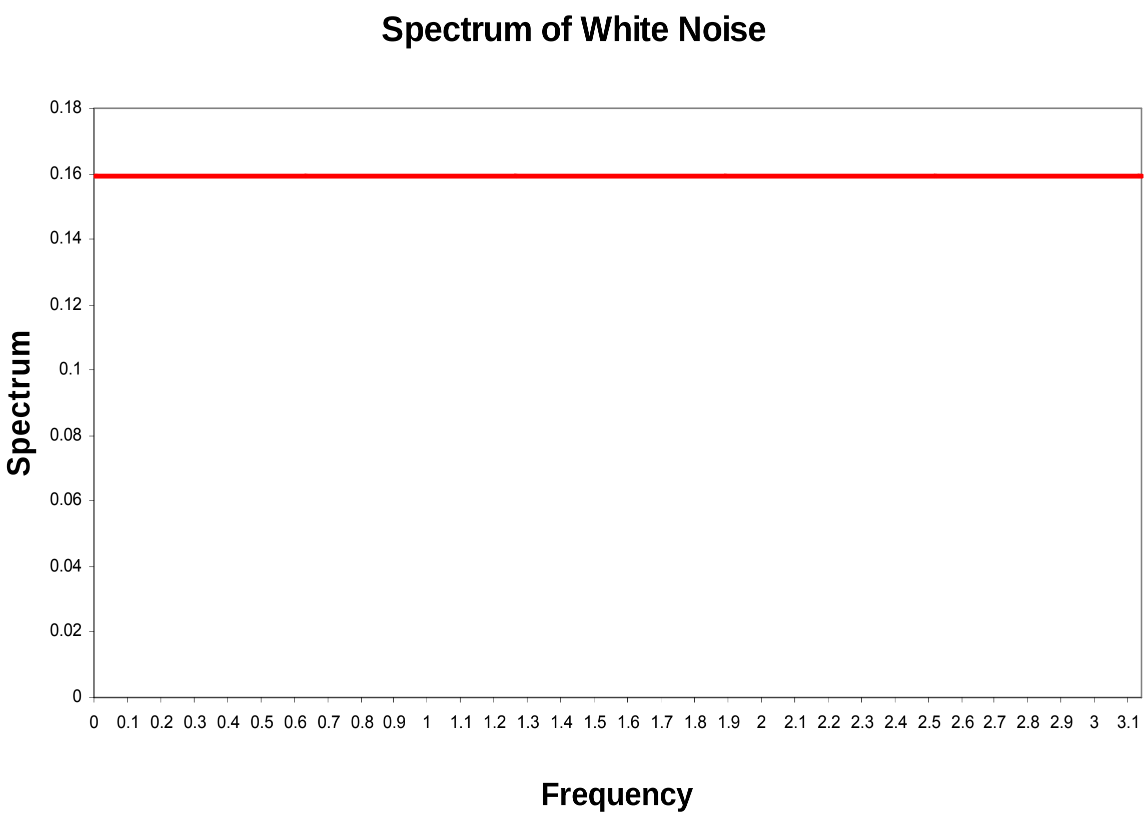
$$\gamma_0 = \sigma^2$$
$$\gamma_k = 0, k > 0$$

$$\Rightarrow$$

$$f_\varepsilon(\lambda_j) = \frac{1}{2\pi}\left\{\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k \cos(\lambda_j k)\right\}$$

$$\Rightarrow f_\varepsilon(\lambda_j) = \frac{\sigma^2}{2\pi}$$

The spectrum of white noise is constant – it does not vary with frequency



White noise consists of an infinite number of cyclical components each having equal weight.

A physical example is white light which contains light of all frequencies in equal contribution.

Spectral examples

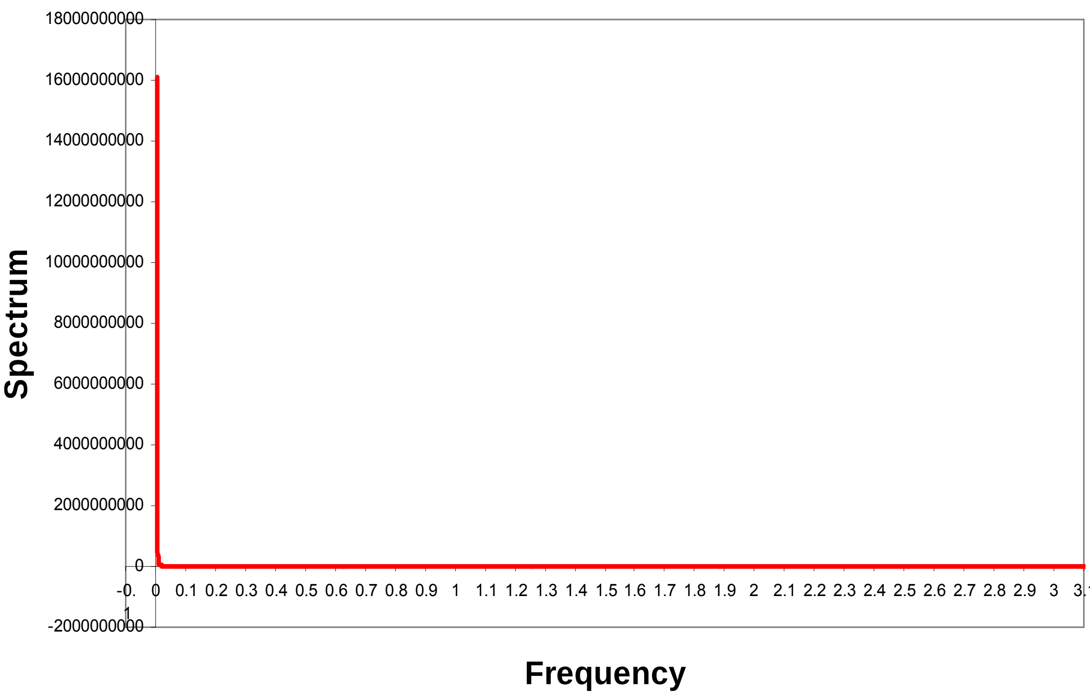
2. Random Walk/Martingale

$$y_t = y_{t-1} + \varepsilon_t$$
$$\Rightarrow y_t = (1-L)^{-1}\varepsilon_t$$

Spectrum of RW/MM (see Appendix 3)

$$f_y(\lambda) = \left|1 - e^{-i\lambda}\right|^{-2} f_\varepsilon(\lambda)$$

Spectrum of a Random Walk (d=1)



As the frequency tends to zero the spectrum tends to infinity:

$$\lim_{\lambda \to 0} f_y(\lambda) = \infty$$

$\Rightarrow$the RW/MM is dominated by its frequency zero (long run trend) component.

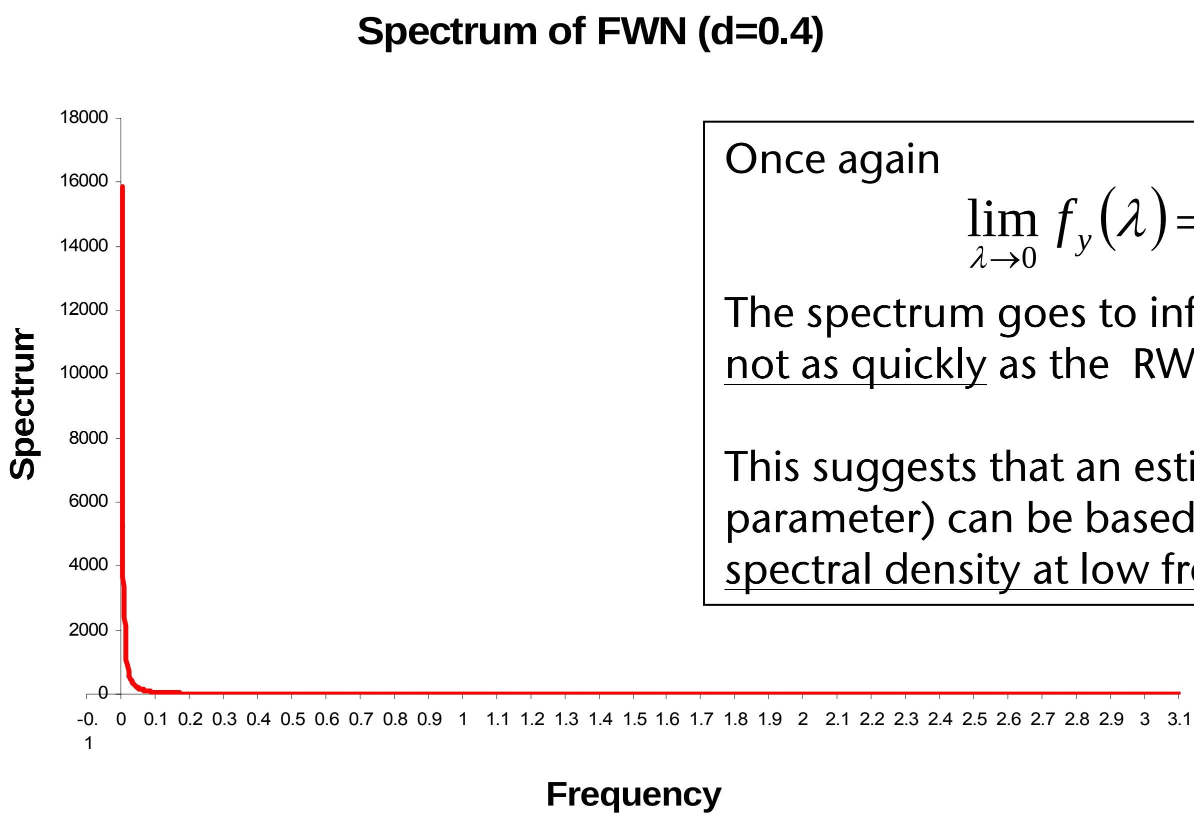
Spectral examples

3. Fractional White Noise:

$$y_t = (1 - L)^{-d} \varepsilon_t$$

Spectrum of FWN

$$f_y(\lambda) = \left|1 - e^{-i\lambda}\right|^{-2d} f_\varepsilon(\lambda)$$



Once again

$$\lim_{\lambda \to 0} f_y(\lambda) = \infty$$

The spectrum goes to infinity at frequency zero but not as quickly as the RW/MM (d=1).

This suggests that an estimator of d (long memory parameter) can be based on the shape/slope of the spectral density at low frequencies.

Testing for long memory (see Mills Chp 3.4)

Geweke and Porter-Hudak (GPH) Estimator

Based on the observation that

$$|1-e^{-i\lambda}|^{-2d} = 4\sin^2(\lambda/2)^{-d}$$

see Appendix 4

$$f_y(\lambda) = |1-e^{-i\lambda}|^{-2d} f_\varepsilon(\lambda)$$

$$\Rightarrow \log f_y(\lambda) = \log f_\varepsilon(\lambda) - d\log[4\sin^2(\lambda/2)]$$

GPH suggested a frequency domain regression

$$\log \hat{f}_y(\lambda_j) = \alpha + \beta \log[4\sin^2(\lambda_j/2)] + v_j$$

Sample/estimated spectrum (estimate this in *Eviews* using 'Spectrum.prg')

$$\hat{d} = -\hat{\beta}$$

Error term

The GPH estimator $\hat{d}$ is consistent and asymptotically normal for d<0.5 (i.e., assuming stationarity)

GPH test for long memory

Need to restrict the frequencies used in estimation to low frequencies – otherwise estimate of *d* will be biased by higher frequency cycles in the series.

Therefore need to choose a cut-off number of frequencies *g(T)* in the GPH regression.

$$\lambda_j = \frac{2\pi j}{T}, \quad j = 1,\ldots,g(T)$$

such that:

$$\lim_{T\to\infty} g(T) = \infty$$

$$\lim_{T\to\infty} g(T)/T = 0$$

⇒ Number of frequencies increases with *T*

Bandwidth = $\lambda_{g(T)} \to 0$

⇒ estimator becomes increasingly 'tuned' to frequency zero (long run component) as *T* increases.

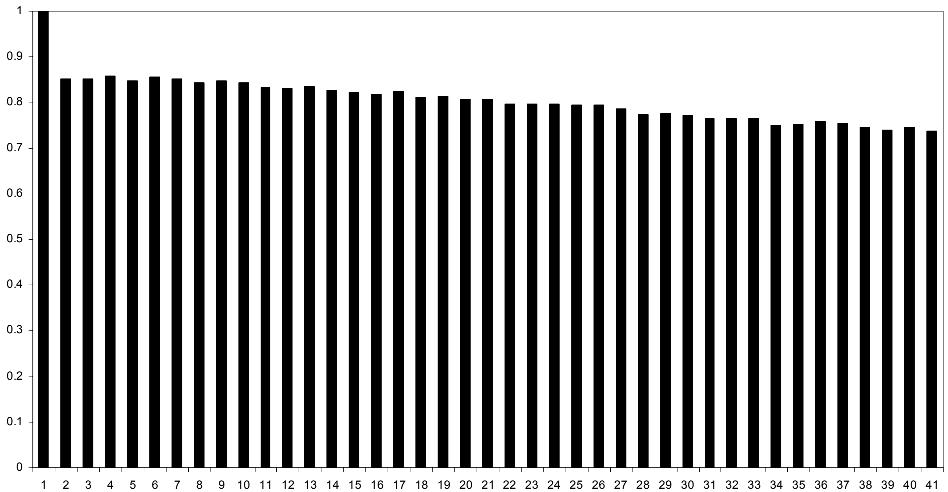
A common choice for *g(T)* is:

$$g(T) = T^{\mu}, \quad 0 < \mu < 1$$

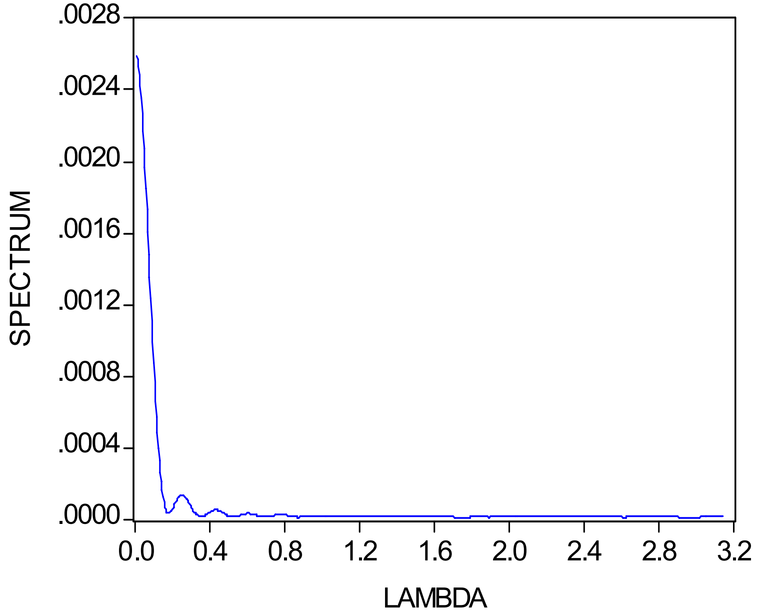
μ=0.5 is typically used.

Application: Testing for long memory in the £/$ forward premium (see Seminar 4)

Sample ACF of the forward premium



Sample spectrum of the forward premium



There is evidence for long memory in the forward premium in both the time domain (sample ACF) and frequency domain (sample spectrum). The analysis in Seminar 4 suggested the presence of a unit root in the forward premium ($\Rightarrow$non-stationary process) - incompatible with finance theory.

GPH estimates of the long memory parameter

Dependent Variable: LOG(SPECTRUM)				
Method: Least Squares				
Sample: 1 34				
Included observations: 34				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-11.0509	0.487439	-22.6713	0
LOG(4*SIN(LAMBDA/2)^2)	-0.63559	0.089053	-7.13723	0
R-squared	0.614179	Mean dependent va		-7.74298
Adjusted R-squared	0.602122	S.D. dependent var		1.395653
S.E. of regression	0.880344	Akaike info criterion		2.640014
Sum squared resid	24.80016	Schwarz criterion		2.7298
Log likelihood	-42.8802	F-statistic		50.94001
Durbin-Watson stat	0.068861	Prob(F-statistic)		0

$g(T) = T^{0.5} = 34 \ (T = 1147)$

$\hat{d} = -\hat{\beta} = 0.636$

Based on a 95% confidence interval: $\hat{d} \pm 1.96\hat{\sigma}_d$ the long memory parameter lies between 0.461 and 0.810 $\Rightarrow$ cannot reject the hypothesis that the forward premium is stationary (d<0.5).

Auto-regressive Fractionally Integrated Moving Average (ARFIMA) processes

More generally a process is ARFIMA if:

$$\phi(L)(1-L)^d y_t = \theta(L)\varepsilon_t$$

The process is stationary if d<0.5 (and all the remaining roots of the AR characteristic polynomial lie outside of the unit circle: see Lecture 5).

The process is invertible if d>-1 (and all the remaining roots of the MA characteristic polynomial lie outside of the unit circle: see Lecture 5).

ARFIMA can model a rich variety of short-run and long-run behaviour of a time-series.

They are now used quite often in empirical finance along with standard ARMA models (see Baillie, 1996, for applications in finance).

Non-stationary processes (Analysis of Price Series)

So far our analysis has involved weakly stationary processes:

- Classical assumption is that d=0 e.g., CLRM and stationary ARMA models.
- More general assumption is $d<0.5$ ⇒stationary long memory models.

What's different about models involving non-stationary processes?

1. Shocks have permanent effects on the levels of non-stationary series.
   - No tendency to revert to mean
   - Series has infinite variance.
   - Non-decay in ACF
2. Test statistics follow non-standard distributions
   - Use of t and F distributions is invalid for inferences.
3. Independent non-stationary series can appear to be related (spurious regression problem):
   - Important to be able to distinguish spurious relationships from meaningful relationships (⇒tests for 'cointegration' see lectures 8/9).

Two types of non-stationarity: TS vs DS

Traditionally (pre-1982) trends in economic/financial data were modelled using a deterministic trend function

$$y_t = f(t) + \varepsilon_t$$

For example...

$f(t) = \alpha + \beta t,$ linear trend

$f(t) = \alpha + \beta t + \gamma t^2,$ quadratic trend.

The mean is time dependent (⇒non-stationarity) but the variance is constant:

$$\text{var}(y_t) = \text{var}(\varepsilon_t) = \sigma^2$$

This series can be made stationary by regressing y on a trend function

⇒TREND STATIONARY TS (DETERMINISTIC TREND) PROCESS

A random walk with drift also has a trend:

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

$$y_t = \mu + (\mu + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

...

$$y_t = y_0 + \mu t + \sum_{k=0}^{t-1} \varepsilon_{t-k}$$

However the model requires to be first differenced to be made stationary

$$(1-L)y_t = \mu + \varepsilon_t$$

⇒DIFFERENCE STATIONARY DS (STOCHASTIC TREND) PROCESS

The order of integration of a process

A DS process is sometimes called 'I(1)' – 'integrated of order one'.

The '1' in I(1) refers to the number of unit roots in the AR polynomial of the process.

A TS process is I(0) because it has no unit roots in the AR polynomial.

In general a process is I(d) if it has d unit roots in its AR polynomial.

Differencing an I(d) process d times yields a process with no unit roots (an I(0) process).

Differencing an I(d) process d times is therefore <u>sufficient</u> to yield a stationary process (but not <u>necessary</u> because stationarity implies $d<0.5$ <u>not</u> $d=0$).

Testing for (integer) unit roots: Dickey Fuller tests

We've carried out informal tests for non-stationarity based on visual inspection of the ACF:

$\Rightarrow$The ACF does not decay if the series is non-stationary.

The requisite statistical theory for formal testing of AR unit roots was developed by Dickey and Fuller (1979) (DF).

DF took a simple AR(1) model:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

or

$$\Delta y_t = \rho y_{t-1} + \varepsilon_t, \quad \rho = \phi - 1$$

...and derived the distribution for the *t*-test of:

$$H_0 : \rho = 0 \quad \text{(the series is nonstationary} : I(1))$$

versus

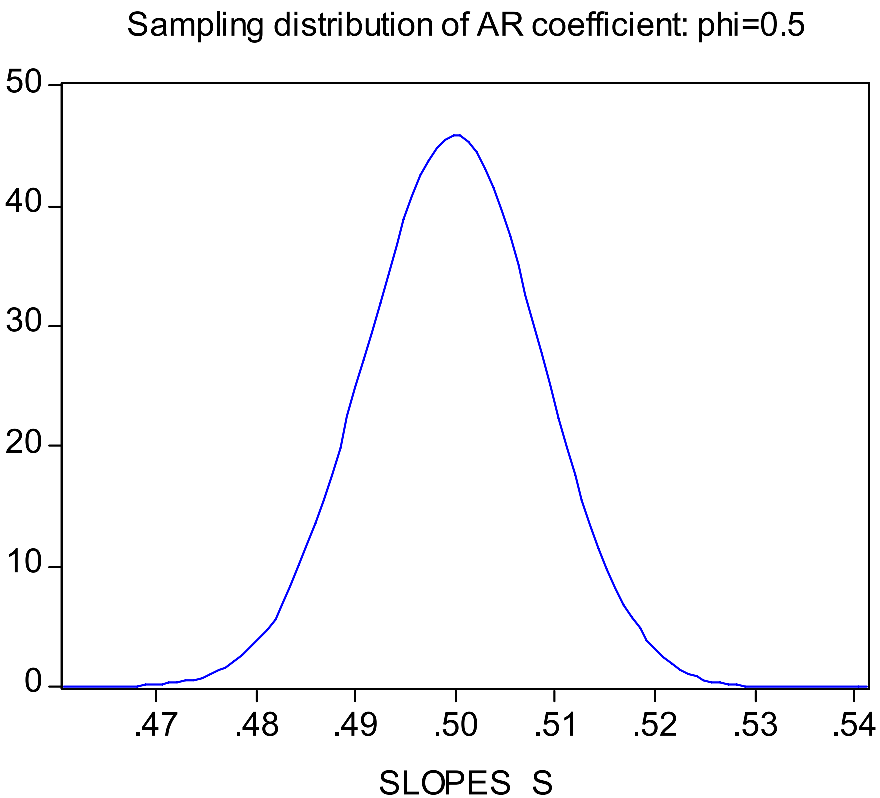
$$H_1 : \rho < 0 \quad \text{(the series is stationary} : I(0))$$

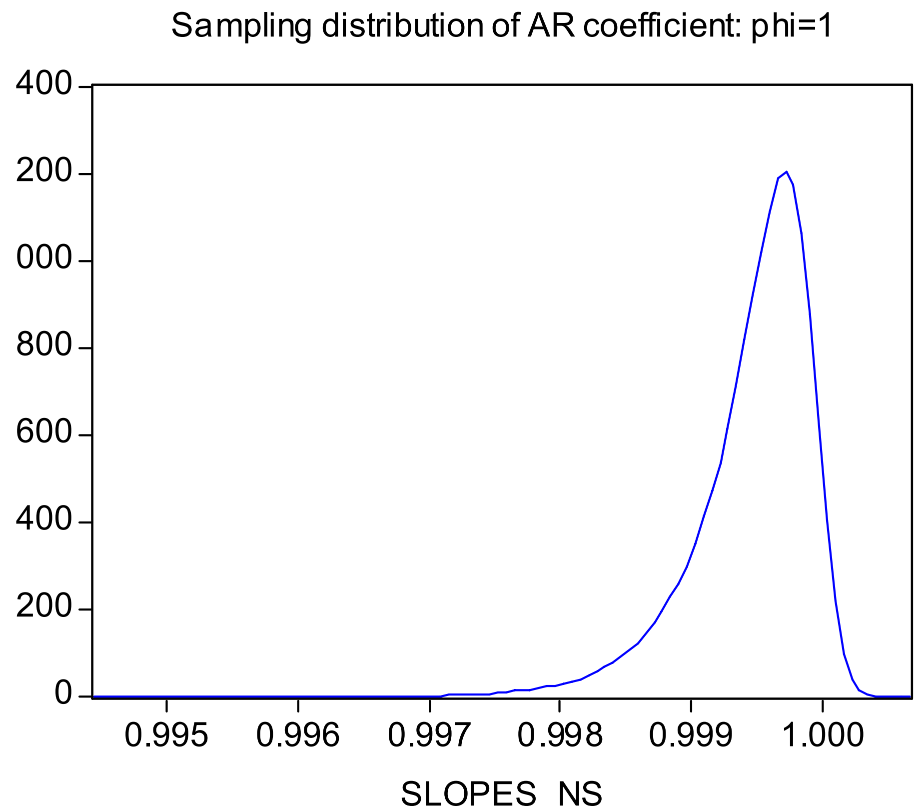
$$DF\ \tau\text{ - test } (tau - test) = \frac{\hat{\rho}}{se(\hat{\rho})}$$

This test does not follow the usual t-distribution
See next slide

Sampling distributions of AR(1) parameter in two instances

Stationary model	Non-stationary model





Accordingly the DF τ distribution has a fatter left tail than the usual *t*-distribution:
⇒the magnitude of the DF critical values are bigger compared to *t* critical values
For example, a *t*-test with a nominal significance level of 5% would reject a true unit root null more than 5% of the time.

Dickey Fuller Tests

The test is carried out using one of 3 test regressions.

1. $\Delta y_t = \rho y_{t-1} + \varepsilon_t$	Use if the series has a zero mean under H_1.
2. $\Delta y_t = \alpha + \rho y_{t-1} + \varepsilon_t$	Use if the series has a non-zero mean under H_1.
3. $\Delta y_t = \alpha + \beta t + \rho y_{t-1} + \varepsilon_t$	Use if the series is trend stationary under H_1.

Critical values for these tests are reported in Brooks Table A2.7. *Eviews* reports the *p*-values for these tests.

You still need to make an informed decision about which regression to run (see seminar 6):

- Omitting relevant deterministic terms will lead to a test based on the wrong distribution (the test will have the wrong size $\Rightarrow$ can't rely on the *p*-values).
- Including irrelevant terms will reduce the power of the test (less likely to reject the null when it's false: see power problems below).

Dickey Fuller Tests

In principle time series can have more than one unit root

Testing strategy

Starting with y, keep testing successive differences of y until the null of a unit root is rejected:

$$y, \Delta y, \Delta^2 y, \ldots, \Delta^d y$$

If you reject the null for the d^{th} difference then the series has d (integer) unit roots (i.e., the series is d^{th} difference stationary).

In practice economic/financial time series typically have only one unit root (1^{st} difference stationary) and rarely have more than 2 unit roots (2^{nd} difference stationary).

Augmented Dickey Fuller (ADF) Test

If the series is AR(p), p>1, then the test equation needs to be modified.

An ADF test adds in lagged differences of the series to take into account higher order AR terms.

Example: Suppose y is AR(2)$\Rightarrow$

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$
$$\Rightarrow \Delta y_t = (\phi_1 - 1) y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$
$$\Rightarrow \Delta y_t = (\phi_1 + \phi_2 - 1) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t$$

Unit root $\Rightarrow \phi_1 + \phi_2 = 1$

ADF term

The test equation involves one lagged difference term to 'mop up' the higher order dependencies in the series.

In general, if the series is AR(p) (and the alternative is trend stationarity) the ADF regression is

Eviews selects lag length automatically using an information criterion e.g., Schwarz Criterion

$$\Delta y_t = \alpha + \beta t + \rho y_{t-1} + \sum_{j=1}^{p-1} \delta_j \Delta y_{t-j} + \varepsilon_t$$

Problems with unit root tests

ADF (and unit root tests in general) have low power. Unit root tests can be 'tricked' into suggesting there are unit roots (when there are none) if (for example):

- There are deterministic structural breaks in the data.
  - These breaks mimic permanent random shocks which 'fools' the test into implying there is a unit root.
- The data have long memory (see e.g., forward-premium example).
- The AR parameter ϕ is simply close to one (if not equal to one)

With short spans of data shocks which are simply very persistent ('near' unit roots e.g., long memory models or AR with ϕ close to one) can appear permanent.

When testing the long-run behaviour of data choose a sample with a long span (at least 10 years).

- When testing long-run behaviour increasing the sampling frequency of the data won't help if the span is too short.

Conclusions

Baillie (1996) provides a well written review of long memory models with applications in finance:

- His own estimate of the long memory parameter for the £-$ forward premium on a different sample (Jan 1974-December 1991) is d=0.55.
- He also discusses extensions to long memory volatility models (Fractionally Integrated GARCH – FIGARCH).

Testing for unit roots forms an important preliminary analysis when analyzing price series.

Typically we want to go on and test if there is a long run relationship involving the series.

⇒Analysis of Non-stationary Processes: Part II (Testing for cointegration) – next lecture.

References

Baillie (1996), Long memory processes and fractional integration in econometrics, *Journal of Econometrics*, 73, 5-59 (A very good review article on long memory processes – also discusses applications in finance).

Brooks (2002), Introductory econometrics for finance, CUP: Cambridge. Chp 7.1 and 7.2** (Unit root testing)

Dickey and Fuller (1979), Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association, 74, 427-431.

Mills (1999), The econometric modelling of financial time series, CUP: Cambridge. Chp 3.4** (Long memory processes)

**Key references

Appendix 1

The binomial expansion of $(1-L)^d$ for any real *d*>–1 is given by

$$(1-L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \ldots$$

For *d*= –1

$$(1-L)^d = 1 + L + L^2 + L^3 + \ldots$$

This is a non-convergent sum of an infinite geometric series (used e.g., in the MA(∞) form of a random walk).

For *d*=1

$$(1-L)^d = 1 - L$$

which is simply the first difference operator.

Appendix 2 (for your information only - not examinable)

For a stationary process $\gamma_k = \gamma_{-k}$

$$\Rightarrow \sum_{k=-\infty}^{\infty} \gamma_k \left[\cos(\lambda k) - i\sin(\lambda k)\right]$$

$$= \gamma_0\left[\cos(0) - i\sin(0)\right] + \sum_{k=1}^{\infty} \gamma_k \left[\cos(\lambda k) + \cos(-\lambda k) - i\sin(\lambda k) - i\sin(-\lambda k)\right]$$

$$= \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos(\lambda k)$$

Using:

$\cos(0) = 1$

$\sin(0) = 0$

$\sin(-\lambda k) = -\sin(\lambda k)$

$\cos(-\lambda k) = \cos(\lambda k)$

Appendix 3: Spectrum of an AR(1) process (for your information only – not examinable)

For an AR(1) process $\gamma_k = \phi^k \sigma^2 \left(1 + \phi^2 + \phi^4 + \ldots\right) = \phi^k \sigma^2 / \left(1 - \phi^2\right)$ (see lecture 5). Also $\gamma_k = \gamma_{-k}$. Therefore the Autocovariance Generating Function is given by:

$$
\begin{aligned}
g_y(z) &= \frac{\sigma^2}{1-\phi^2} \sum_{k=-\infty}^{\infty} \phi^{|k|} z^k \\
&= \frac{\sigma^2}{1-\phi^2} \left(\sum_{k=-\infty}^{0} \phi^{|k|} z^k + \sum_{k=0}^{\infty} \phi^k z^k - 1 \right) \\
&= \frac{\sigma^2}{1-\phi^2} \left(\frac{1}{1-\phi z^{-1}} + \frac{1}{1-\phi z} - 1 \right) \\
&= \frac{\sigma^2}{1-\phi^2} \left\{ \frac{1-\phi z + \left(1-\phi z^{-1}\right) - \left(1-\phi z\right)\left(1-\phi z^{-1}\right)}{\left(1-\phi z\right)\left(1-\phi z^{-1}\right)} \right\} \\
&= \frac{\sigma^2}{1-\phi^2} \frac{1-\phi^2}{\left(1-\phi z\right)\left(1-\phi z^{-1}\right)} \\
&= \frac{\sigma^2}{\left(1-\phi z\right)\left(1-\phi z^{-1}\right)}
\end{aligned}
$$

Accordingly the spectrum $f_y(\lambda) = (2\pi)^{-1} g_y\left(e^{-i\lambda}\right)$ is given by

$$
\begin{aligned}
f_y(\lambda) &= \frac{1}{2\pi} \frac{\sigma^2}{\left(1-\phi e^{-i\lambda}\right)\left(1-\phi e^{i\lambda}\right)} \\
&= \left|1-\phi e^{-i\lambda}\right|^{-2} f_\varepsilon(\lambda)
\end{aligned}
$$

The last line follows since the modulus of a complex number $|h - iv| = \sqrt{h^2 + v^2}$ (by Pythagoras' Theorem) so $|h - iv|^2 = h^2 + v^2 = (h - iv)(h + iv)$. Also $f_\varepsilon(\lambda) = \sigma^2 / 2\pi$.

Appendix 4 (not examinable)

In Appendix 3 it was shown

$$\begin{aligned}\left|1 - e^{-i\lambda}\right|^2 &= \left(1 - e^{-i\lambda}\right)\left(1 - e^{i\lambda}\right)\\ &= 2 - \left(e^{-i\lambda} + e^{i\lambda}\right)\end{aligned}$$

We can expand this quantity using three trigonometric identities

1. $e^{\pm i\lambda} \equiv \cos\lambda \pm i\sin\lambda$
2. $\cos(2\lambda) \equiv \cos^2\lambda - \sin^2\lambda$
3. $\sin^2\lambda + \cos^2\lambda \equiv 1$

Using the first identity

$$e^{-i\lambda} + e^{i\lambda} = 2\cos\lambda$$

Using the second identity

$$\cos\lambda = \cos^2(\lambda/2) - \sin^2(\lambda/2)$$

Using the third identity

$$\cos^2(\lambda/2) = 1 - \sin^2(\lambda/2)$$

Therefore (using each of these results in turn)

$$\begin{aligned}\left|1 - e^{-i\lambda}\right|^2 &= 2 - 2\cos\lambda\\ &= 2 - 2\left(\cos^2(\lambda/2) - \sin^2(\lambda/2)\right)\\ &= 2 - 2\left(1 - 2\sin^2(\lambda/2)\right)\\ &= 4\sin^2(\lambda/2)\end{aligned}$$

Finally

$$\begin{aligned}\left|1 - e^{-i\lambda}\right|^{-2d} &= \left(\left|1 - e^{-i\lambda}\right|^2\right)^{-d}\\ &= \left[4\sin^2(\lambda/2)\right]^{-d}\end{aligned}$$