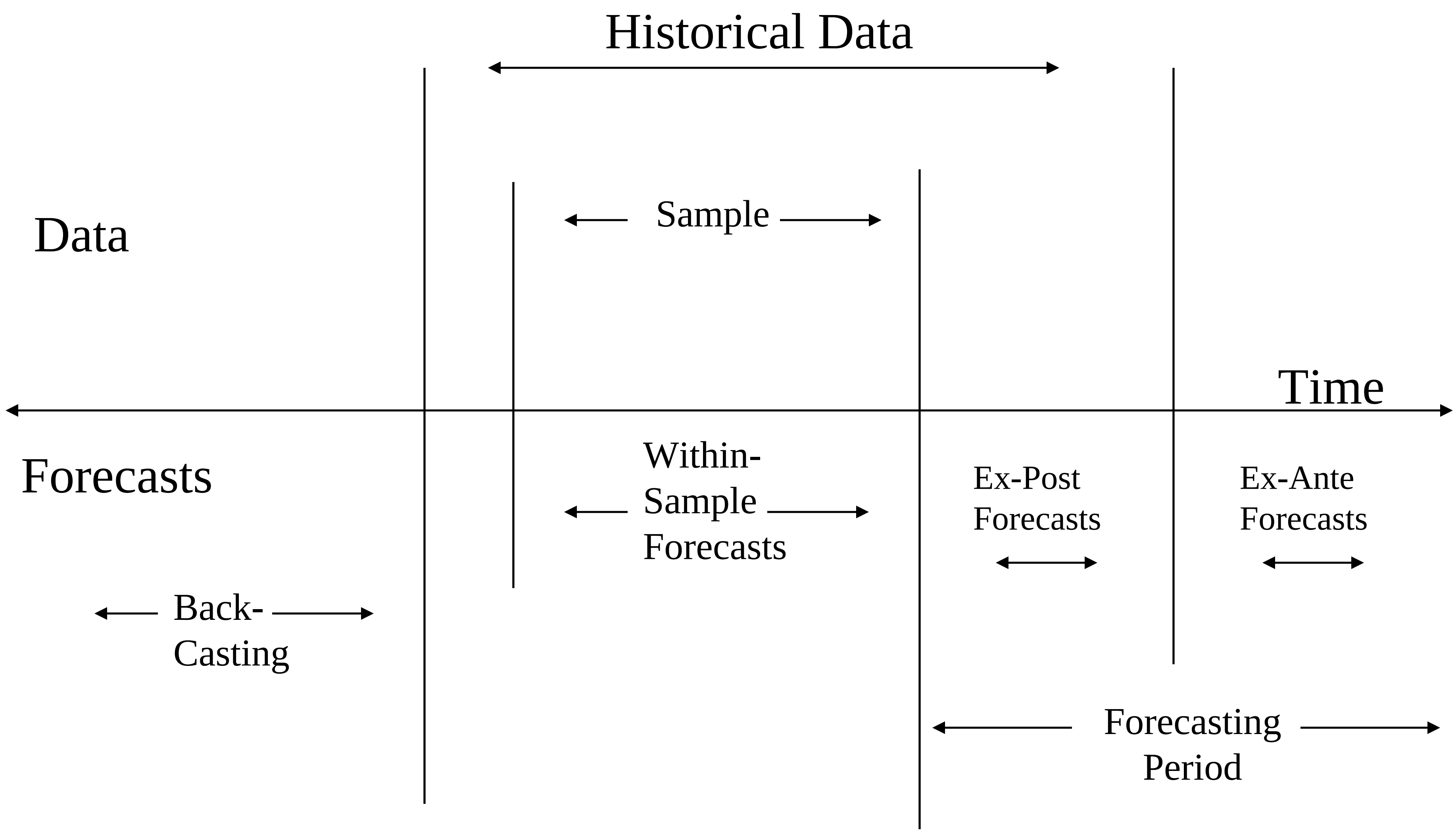
Forecasting Financial Markets

Time Series Analysis

Overview

- Time series data & forecasts
- ARIMA models
- Model diagnosis & testing

Time Series Data & Forecasting



Univariate Time Series Models

- Autoregressive AR(1):
  - $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$
- Moving Average MA(1):
  - $y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1}$
- ε_t = sequence of independent random variables
  - Independent
  - Zero mean
  - Constant variance σ^2

White Noise

- Mean is constant (zero)
  - $E(\varepsilon_t) = \mu \; (0)$
- Variance is constant
  - $Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2$
- Uncorrelated
  - $Cov(\varepsilon_t , \varepsilon_{t-j}) = 0$ for j < > 0 and t
- *Gaussian White Noise*
  - If ε_t is also normally distributed
- *Strict White Noise*
  - ε_t are *independent*

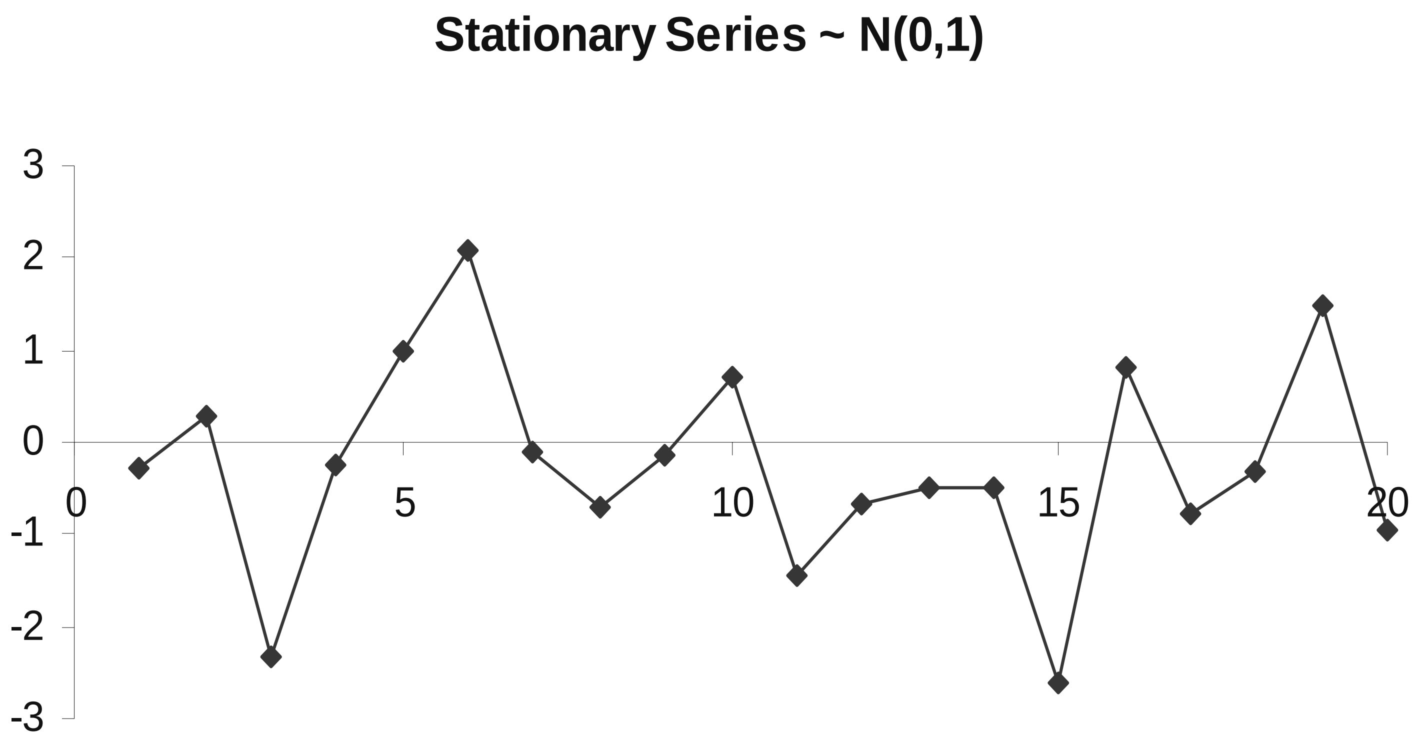
Lag Operator

- $L^m y_t = y_{t-m}$
- So AR(1) process can be represented as:
  - $(1 - \beta L)\, y_t = \varepsilon_t$
- Invertibility
  - An AR(1) process can be represented as MA(∞):
    - If $|\beta| < 1$
  - $y_t = (1 - \beta L)^{-1}\, \varepsilon_t$
  - $y_t = [1 + \beta L + (\beta L)^2 + \ldots]\, \varepsilon_t$
  - $y_t = \varepsilon_t + \beta\, \varepsilon_{t-1} + \beta^2\, \varepsilon_{t-2} + \ldots$

Stationarity

- *Weak* (covariance) stationarity
  - Population moments are time-independent:
    - $E(y_t) = \mu$
    - $Var(y_t) = \sigma^2$
    - $Cov(y_t, y_{t-j}) = \gamma_j$
  - Example: white noise ε_t
- *Strong* stationarity
  - In addition, y_t is *normally distributed*

Stationary Series



Stationarity of AR(1) Process

- AR(1) Process: $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$
- Expected value $E(y_t)$ is time-dependent:

$$E(y_t) = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0$$

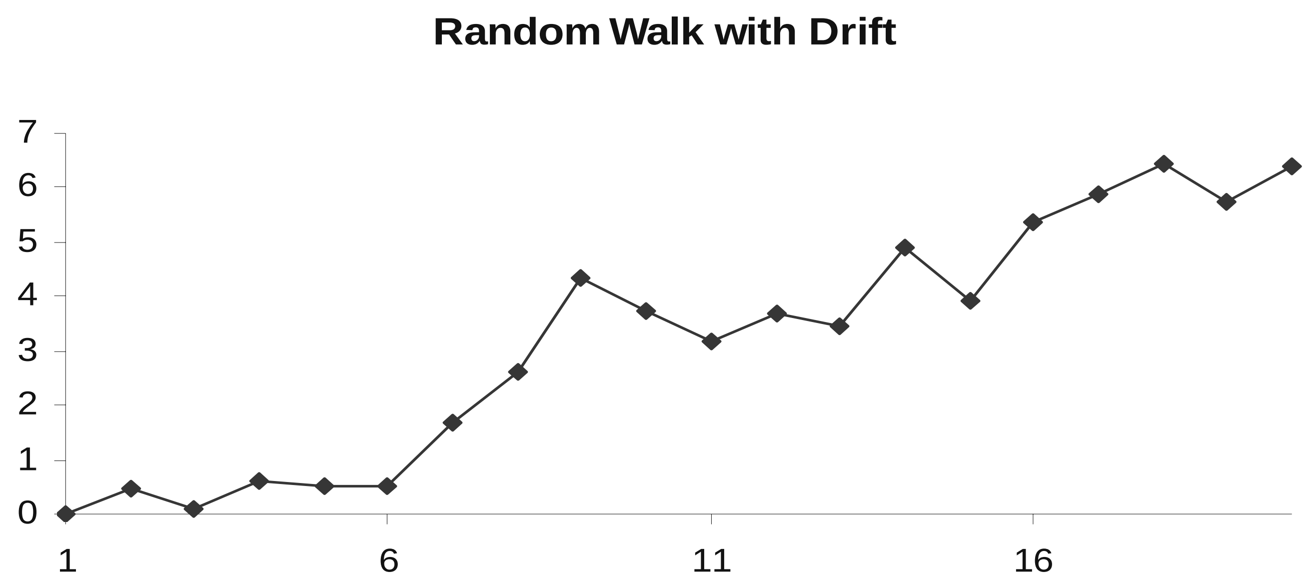
- If $|a_1| < 1$, then as $t \to \infty$, process is stationary
  - Lim $E(y_t) = a_0 / (1 - a_1)$
    - Hence mean of y_t is finite and time independent
  - Also $Var(y_t) = E[\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \ldots)^2]$
    - $= \sigma^2[1 + (a_1)^2 + (a_1)^4 + \ldots] = \sigma^2/[1 - (a_1)^2]$
    - And $Cov(y_t, y_s) = \sigma^2 (a_1)^s /[1 - (a_1)^2]$

Stationarity Considerations

- Sample drawn from recent process may not be stationary
- Hence many econometricians assume process has been continuing for infinite time
- Can be problematic
  - E.G. FX rate changes post Bretton-woods

Random Walk Process

- Random Walk with drift
  - $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$
    - With $a_1 = 1$
    - A *non-stationary* process



Random Walk Process

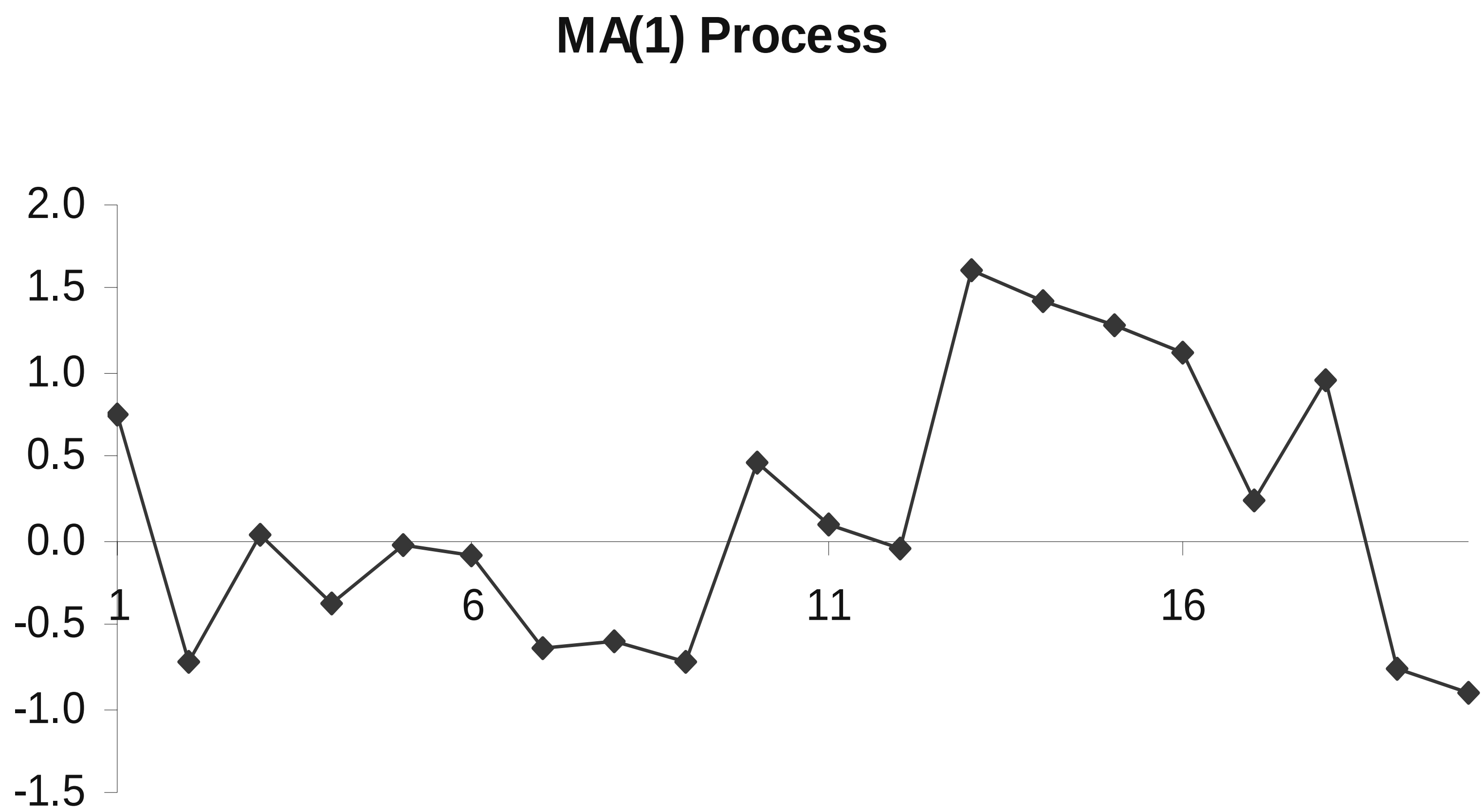
- Random Walk without drift
  - $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$
    - With $a_1 = 1$, $a_0 = 0$
    - $\Delta y_t = \varepsilon_t$ or $y_t = (1- L)^{-1}\varepsilon_t = \varepsilon_t + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} \ldots$
  - Also a *non-stationary* process
    - Variance of y_t gets larger over time
      - Hence not independent of time.

$$Var(y_t) = E\left[\sum_{1}^{n} \varepsilon_t^2 + 2\sum_{t \neq s} \varepsilon_t \varepsilon_s\right] = n\sigma^2$$

Moving Average Process

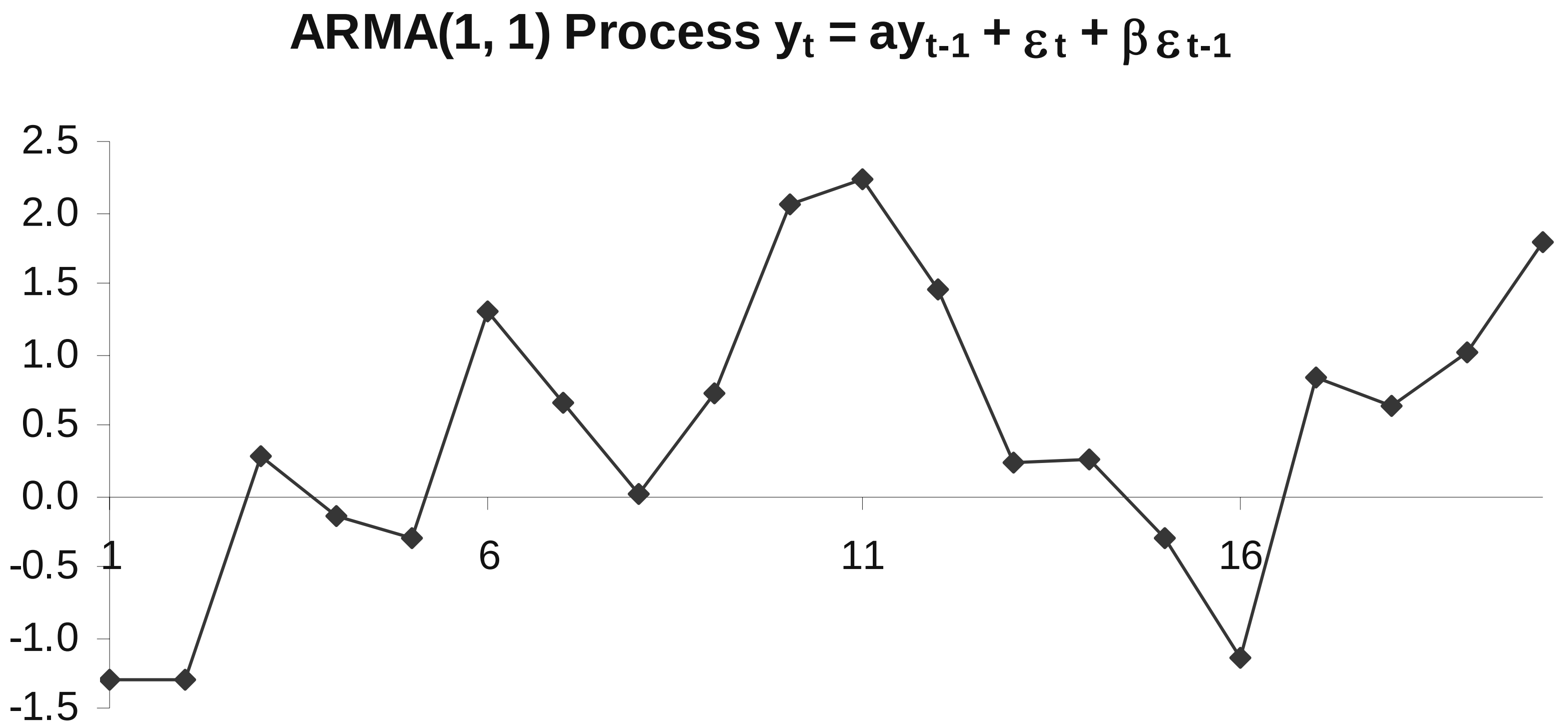
- MA(1) process
  - $y_t = \varepsilon_t + \beta\varepsilon_{t-1} = (1 + \beta L)\varepsilon_t$
- Invertibility: $|\beta| < 1$
  - $(1+\beta L)^{-1} y_t = \varepsilon_t$
  - $y_t = \Sigma(-\beta)^{j} y_{t-j} + \varepsilon_t$
  - So MA(1) process with $|\beta| < 1$ is an infinite autoregressive process
  - Similary an AR(1) process with $|\beta| < 1$ is invertible
    - i.e. can be represented as an infinite MA process

MA(1) Process



ARMA(1, 1) Process

- $y_t = a_1 y_{t-1} + \varepsilon_t + \beta\varepsilon_{t-1}$



General ARMA Process

- Any stationary time series can be approximated by a mixed autoregressive moving average model
- ARMA(p, q)
  - $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$
  - $\Phi(L)\, y_t = \theta(L) \varepsilon_t$
- Φ and Θ are polynomials in the lag operator L
  - $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$
  - $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$

Unit Roots

- Stationarity Condition
  - Roots of $\phi(L)$ must lie outside the unit circle
    - $|x_i| > 1$ for all roots x_i
- Invertibility Condition
  - Roots of $\theta(L)$ must lie outside the unit circle
    - $|z_i| > 1$ for all roots z_i

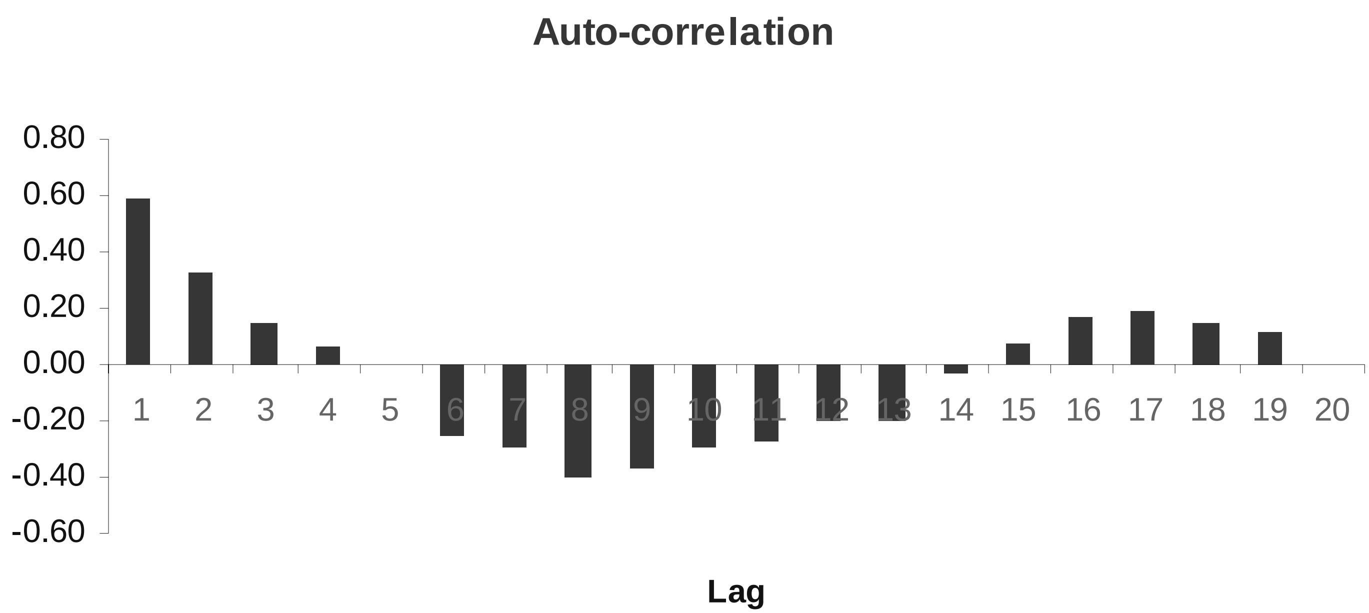
Autocorrelation

- Population autocorrelation between y_t and $y_{t-\tau}$
- $\rho_\tau = \gamma_\tau/\gamma_0$ ($\tau = \pm 1, \pm 2, \ldots$)
  - γ_τ is the *autocovariance* function at lag τ
    - $\gamma_\tau = \mathrm{Cov}(y_t, y_{t-\tau})$
    - $\gamma_0 = \mathrm{Var}(y_t)$
  - $\rho_0 = 1$, by definition
- Sample autocorrelation: $\rho'_\tau = c_\tau/c_0$
  - Where c_τ is the sample autocovariance

$$c_\tau = \frac{1}{n-\tau}\sum_{t=\tau+1}^{n}(y_t - \bar{y})(y_{t-\tau} - \bar{y})$$

Correlogram

- Plot of ACF ρ_τ against τ



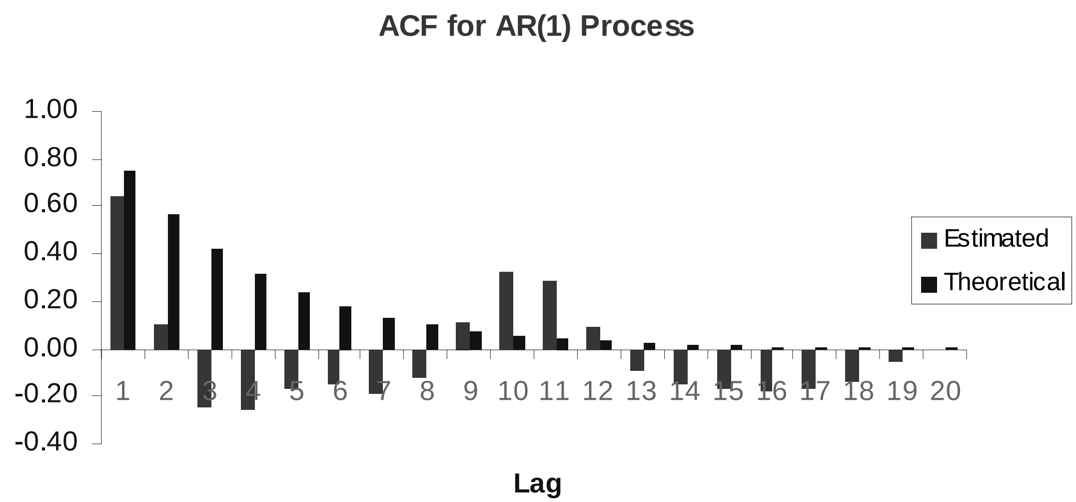
ACF for AR(1) Process

- AR(1) Process: $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$
- Correlation: $\rho_s = (a_1)^s$, $s = 0, 1, \ldots$

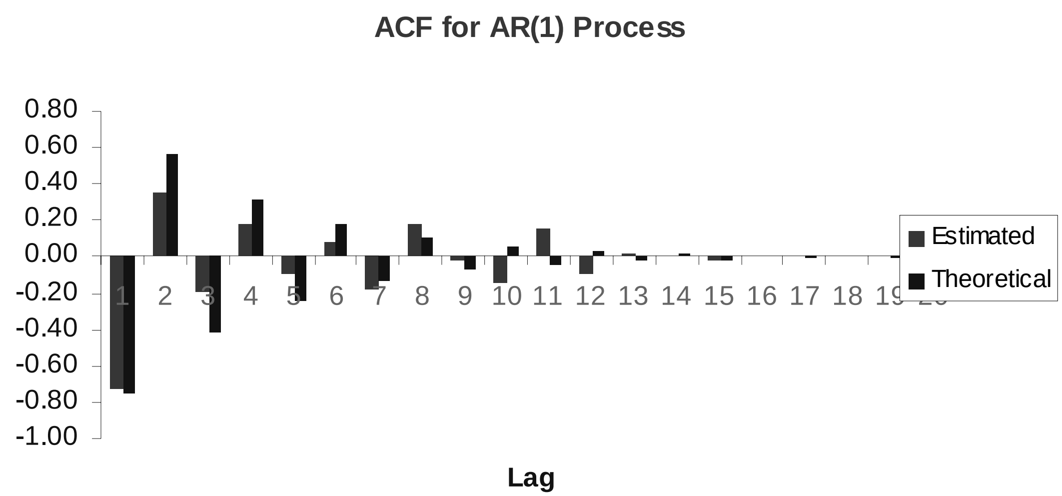
  Since:
  - $\gamma_0 = \sigma^2/[1 - (a_1)^2]$
  - $\gamma_s = \sigma^s (a_1)^s / [1 - (a_1)^2]$

ACF for AR(1) Process

- $a_1 = 0.75$



- $a_1 = -0.75$

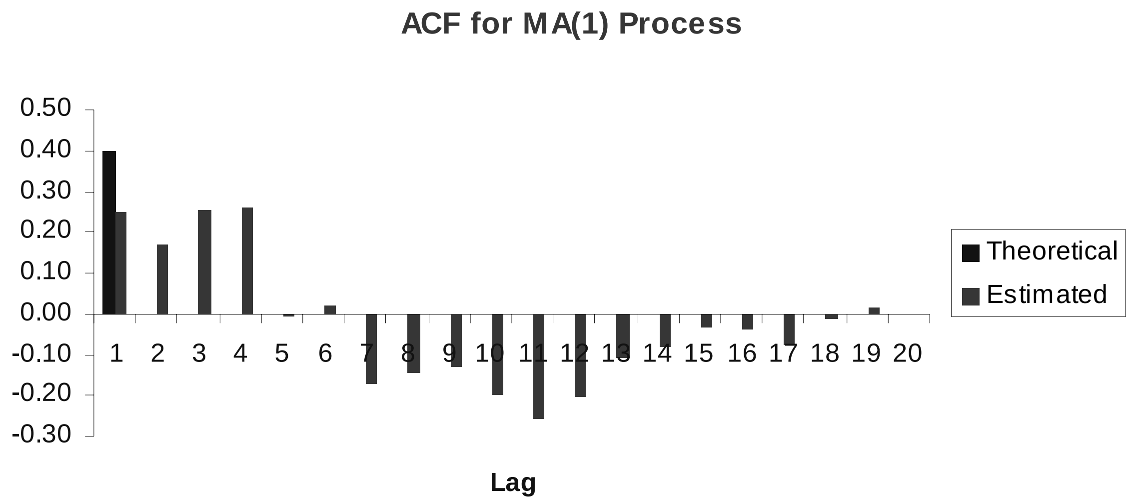


ACF for MA(1) Process

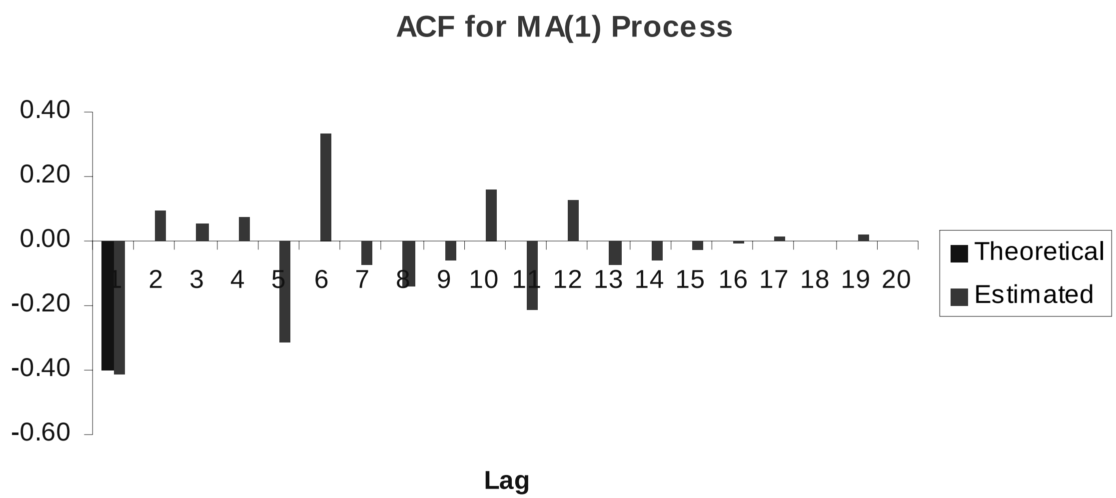
- MA(1) Process: $y_t = \varepsilon_t + \beta\varepsilon_{t-1}$
- *Yule-Walker* Equations
  - $\gamma_0 = Var(y_t) = E(y_t\, y_t) = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_t + \beta\varepsilon_{t-1})]$
    $= (1 + \beta^2)\sigma^2$
  - $\gamma_1 = E(y_t\, y_{t-1}) = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-1} + \beta\varepsilon_{t-2})] = \beta\sigma^2$
  - $\gamma_s = E(y_t\, y_{t-s}) = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-s} + \beta\varepsilon_{t-s-1})] = 0, \; s > 1$
- ACF
  - $\rho_0 = 1$
  - $\rho_1 = \beta / (1 + \beta^2)$
  - $\rho_s = 0$, for $s > 1$

ACF for MA(1) Process

- $\beta = 0.5$



- $\beta = -0.5$



Partial Autocorrelation Function (PACF)

- In AR(1) process y_t and y_{t-2} are correlated
  - Indirectly, through y_{t-1}
  - $\rho_2 = \text{Corr}(y_t, y_{t-2}) = \text{Corr}(y_t, y_{t-1}) * \text{Corr}(y_{t-1}, y_{t-2}) = \rho_1^2$
- Partial autocorrelation between y_t and y_{t-s}
  - Eliminates effects of intervening values y_{t-1} to y_{t-s+1}
  - Effectively doing autoregression of y_t against y_{t-1} to y_{t-s}
    - $y_t = \sum b_i y_{t-i} + \varepsilon_t$

Calculating PACF

- Form series $y^*_t = y_t - \mu$
  - M is mean $e\{y_t\}$
- Form first-order autoregressive equation
- $Y^*_t = \phi_{11} y^*_{t-1} + e_t$
  - e_t is error process which may not be white noise
  - ϕ_{11} is both AC and PAC between y_t and y_{t-1}
- Form second-order autoregressive equation
- $Y^*_t = \phi_{21} y^*_{t-1} + \phi_{22} y^*_{t-2} + e_t$
  - Φ_{22} is PAC between y_t and y_{t-2} , i.E autocorrelation between y_t and y_{t-2} controlling (netting out) effect of y_{t-1}
- Repeat for all additional lags to obtain PACF

PACF by Yule-Walker

- Form PACF from ACF
  - $\phi_{11} = \rho_1$, $\phi_{22} = (\rho_2 - \rho_1^2) / (1 - \rho_1^2)$
- Formula for additional lags s = 3, 4, . . .

$$\phi_{ss} = \frac{\rho_s - \sum_{j=1}^{s-1} \phi_{s-1} \rho_{s-j}}{1 - \sum_{j=1}^{s-1} \phi_{s-1} \rho_j}$$

$$\phi_{sj} = \phi_{s-1,j} - \phi_{ss}\phi_{s-1,s-j} \qquad j = 1, 2, ...s-1$$

PACF for AR and MA Processes

- For AR(p) process
  - No direct correlation between y_t and y_{t-s} for $s > p$
  - Hence $\phi_{ss} = 0$ for $s > p$
  - Good means of indentifying AR(p) type process
- For MA(1) process $y_t = \varepsilon_t + \beta\varepsilon_{t-1} = (1 + \beta L)\varepsilon_t$
  - $y_t = \Sigma(-\beta)^j y_{t-j} + \varepsilon_t$ for $|\beta| < 1$
  - Hence y_t is correlated with all its own lags
  - PACF will *decay geometrically*
    - Direct if $\beta < 0$
    - Alternating if $\beta > 0$

Lab: ARMA(1, 1) Process

- ARMA(1, 1): $y_t = a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$
- Lab:
  - Generate time series
  - Compute theoretical ACF
    - Yule-Walker equations
  - Estimate sample ACF
    - Autocorrel function
  - How does pattern of ACF depend on parameters?

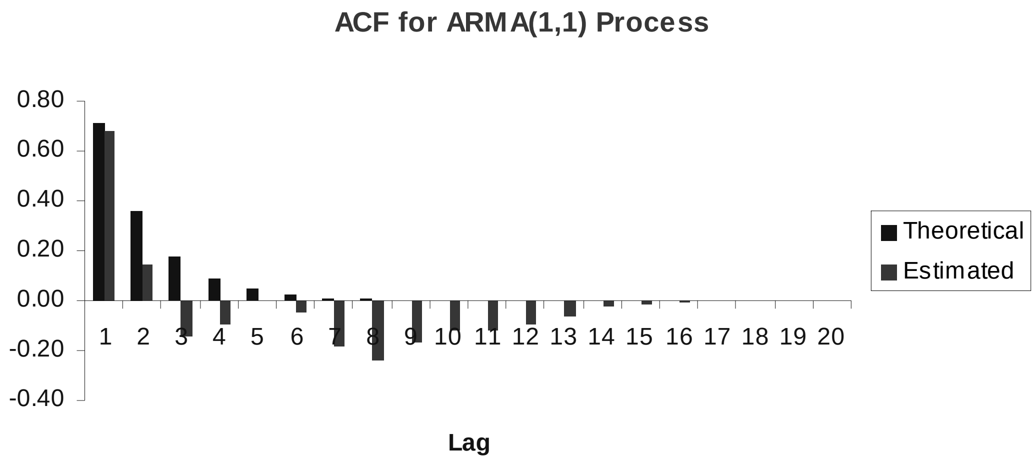
Solution: ARMA(1, 1) Process

- Yule-Walker Equations
  - $\gamma_0 = E(y_t\, y_t) = a_1 E(y_{t-1}\, y_t) + E(\varepsilon_t\, y_t) + \beta_1 E(\varepsilon_{t-1}\, y_t)$
    $= a_1\gamma_1 + \sigma^2 + \beta_1 E[\varepsilon_{t-1}(a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1})]$
    $= a_1\gamma_1 + \sigma^2 + \beta_1(a_1 + \beta_1)\,\sigma^2$
  - $\gamma_1 = E(y_t\, y_{t-1}) = a_1 E(y_{t-1}\, y_{t-1}) + E(\varepsilon_t\, y_{t-1}) + \beta_1 E(\varepsilon_{t-1}\, y_{t-1})$
    $= a_1\,\gamma_0 + \beta_1\,\sigma^2$
  - $\gamma_s = E(y_t\, y_{t-s}) = a_1 E(y_{t-1}\, y_{t-s}) + E(\varepsilon_t\, y_{t-s}) + \beta_1 E(\varepsilon_{t-1}\, y_{t-s})$
    $= a_1\,\gamma_{s-1}$
  - Solution

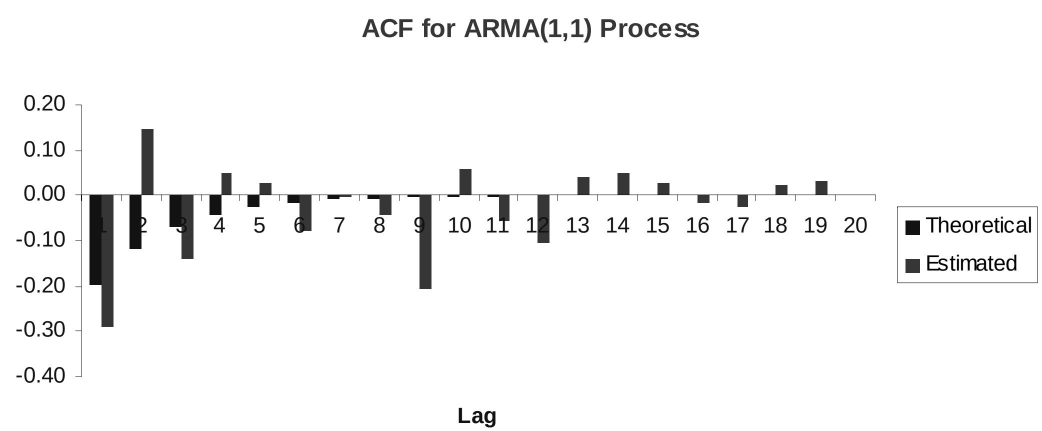
$$\gamma_0 = \frac{1+\beta_1^2 + 2a_1\beta_1}{(1-a_1^2)}\sigma^2 \qquad \gamma_1 = \frac{(1+a_1\beta_1)(a_1+\beta_1)}{(1-a_1^2)}\sigma^2 \qquad \rho_1 = \frac{(1+a_1\beta_1)(a_1+\beta_1)}{(1+\beta_1^2+2a_1\beta_1)}$$

ACF for ARMA(1, 1) Process

- $a_1 = \beta_1 = 0.5$

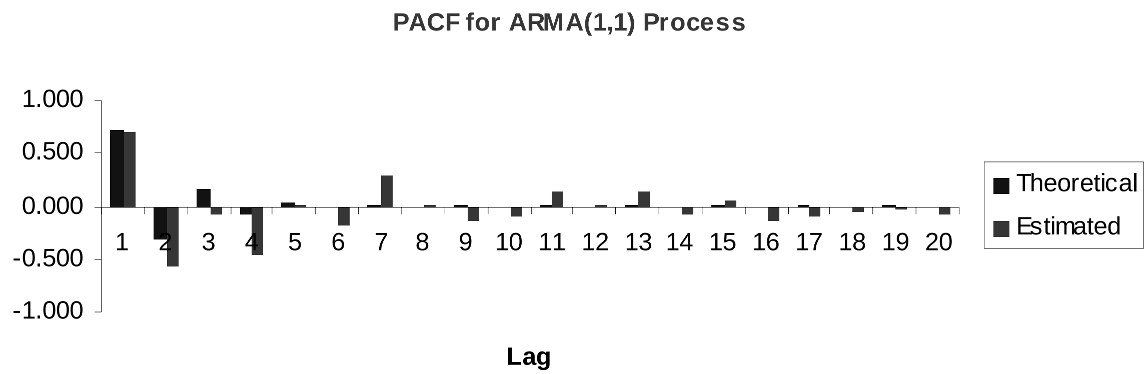


- $a_1 = 0.6$,
  $\beta_1 = -0.95$

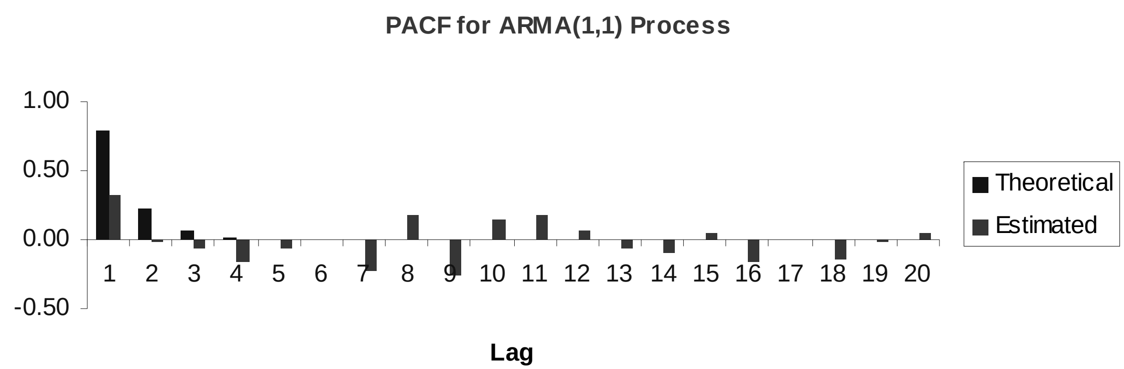


PACF for ARM(1,1) Process

- $a_1 = \beta_1 = 0.5$



- $a_1 = 0.7$,
  $\beta_1 = -0.3$



Properties of ACF and PACF

Process	ACF	PACF
White Noise	All $\rho_s = 0$	All $\phi_{ss} = 0$
AR(1): a > 0	Geometric decay: $\rho_1 = a^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$; s>1
AR(1): a < 0	Oscillating decay: $\rho_1 = a^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$; s>1
MA(1): $\beta > 0$	+ve spike at lag 1. $\rho_0 = 0$ for s > 1	Oscillating decay $\phi_{11} > 0$
MA(1): $\beta < 0$	-ve spike at lag 1. $\rho_0 = 0$ for s > 1	Decay $\phi_{11} > 0$
ARMA(1,1): a < 0	Geometric decay at lag 1 Sign ρ_1 = sign(a+β)	Osc. decay at lag 1 $\phi_{11} = \rho_1$
ARMA(1,1): a > 0	Oscillating decay at lag 1 Sign ρ_1 = sign(a+β)	Geom. decay at lag 1 $\phi_{11} = \rho_1$

Box-Jenkins Methodology

- Phase I - identification
  - Identify appropriate models
- Phase II - estimation & testing
  - Estimate model parameters
  - Check residuals
- Phase III application
  - Use model to forecast

Phase I - Identification

- Data preparation
  - Transform data to stabilise variance
  - Difference data to obtain stationary series
- Model selection
  - Examine data, ACF and PACF to identify potential models

Phase II - Estimation & Testing

- Estimation
  - Estimate model parameters
  - Select best model using suitable criterion
- Diagnostics
  - Check ACF/PACF of residuals
  - Do portmanteau test of residuals
  - Are residuals white noise?
    - If not, return to phase I (model selection)

Model Selection Criteria

- Two objectives
  - Minimize sums of squares of residuals
    - Can always reduce by adding more parameters
  - Parsimony
    - Avoid excess paramterization
      - I.E. Loss of degrees of freedom
    - Better forecasting performance
- Solution
  - Penalize the likelihood for each additional term added to model

Likelihood Function

- Assume $y_t \sim No(\mu, \sigma^2)$
- Likelihood
  - $L = (-n/2)[Ln(2\pi) + Ln(\sigma^2)] - (1/2\sigma^2)\Sigma(y_t - \mu)^2$
  - Maximizing wrt μ, σ^2:
  - MLE Estimates
    - $\mu' = \Sigma y_t / n$
    - $(\sigma')^2 = \Sigma(y_t - \mu')^2 / n$

Likelihood Function in Regression

- Simple Linear Regression: $y_t = \beta x_t + \varepsilon_t$
  - $\varepsilon_t \sim$ IID $No(0, \sigma^2)$
- Likelihood
  - $L = (-n/2)[Ln(2\pi) + Ln(\sigma^2)] - (1/2\sigma^2)\Sigma(y_t - \beta x_t)^2$
- MLE Estimates
  - $(\sigma')^2 = \Sigma(\varepsilon_t)^2 / n$
  - $\beta' = \Sigma(x_t y_t) / \Sigma(x_t)^2$
- Standard Error
  - $\sigma'_\beta = \sigma' / \{\Sigma(x_t - x_{mean})^2\}^{1/2}$

Maximum Likelihood Estimation

- Akaike information criterion (AIC)
  - AIC $= -2Ln(Likelihood) + 2m$
    $\approx nLn(SSE) + 2m$
- Schwartz Bayesian information criterion (BIC)
  - BIC $= -2Ln(Likelihood) + mLn(n)$
    $\approx nLn(SSE) + mLn(n)$
    - L is likelihood function
    - SSE is error sums of squares
    - n is number of observations
    - m is number of model parameters

Using MLE Model Criteria

- When comparing models:
  - N should be kept fixed
    - E.G. With 100 data points estimate an AR(1) and AR(2) using only last 98 points.
  - Use same time period for all models
  - AIC and BIC should be as small as possible
    - What matter is *comparative* value of AIC or BIC
    - BIC has better large sample properties
    - AIC will tend to prefer over-paramaterized models

Sums of Squares

- Sums of Squares
  - Due to Model = SSM

$$SSM = \sum (\hat{y}_t - \bar{y})^2$$

  - Due to Error = SSE

$$SSE = \sum (y_t - \hat{y}_t)^2$$

  - Total Sums of Squares = SST = SSM + SSE

$$SST = \sum (y_t - \bar{y})^2$$

ANOVA and Goodness of Fit

Source of Variation	Sums of Squares	Degrees of Freedo m	Mean Square	F
Regression	SSR	m	MSR=SSR/m	F=MSR/MSE
Error	SSE	n-m-1	MSE-SSE/(n-m)	
Total	SST	n-1		

- F test statistic = MSR/MSE
  - With 1 and n-m-1 degrees of freedom
    - n is #observations, m is # independent variables
    - Large value indicates relationship is statistically significant

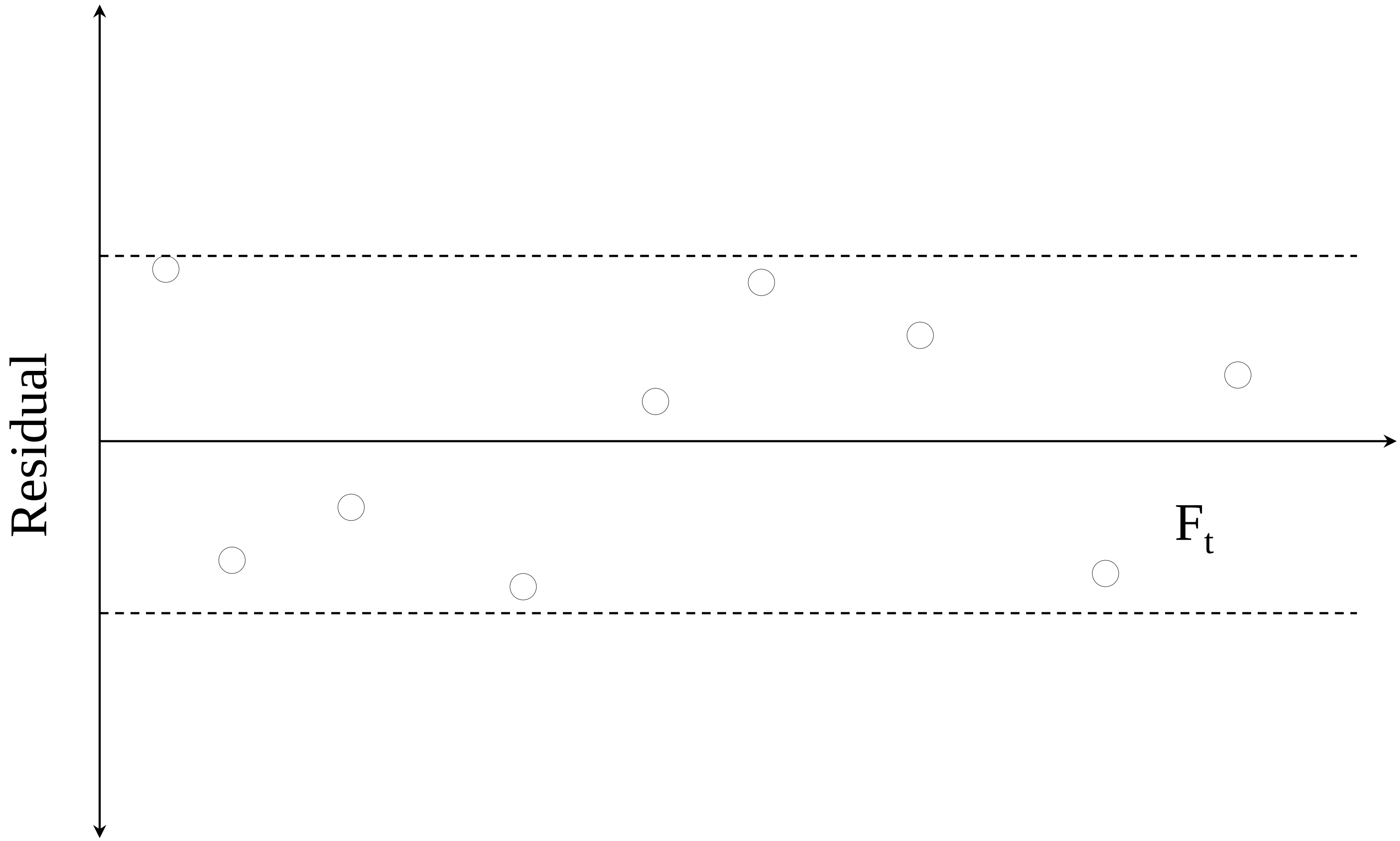
Coefficient of Determination

- $R^2 = SSR/SST$
  - How much of total variation is explained by regression
- Adjusted R^2
  - Adjusted $R^2 = 1 - (1 - R^2)\ (n - 1) / (n - m - 1)$
    - Idea: R^2 can always increase by adding more variables
    - Penalize R^2 for loss of degrees of freedom
    - Useful for comparing models with different # independent variables m

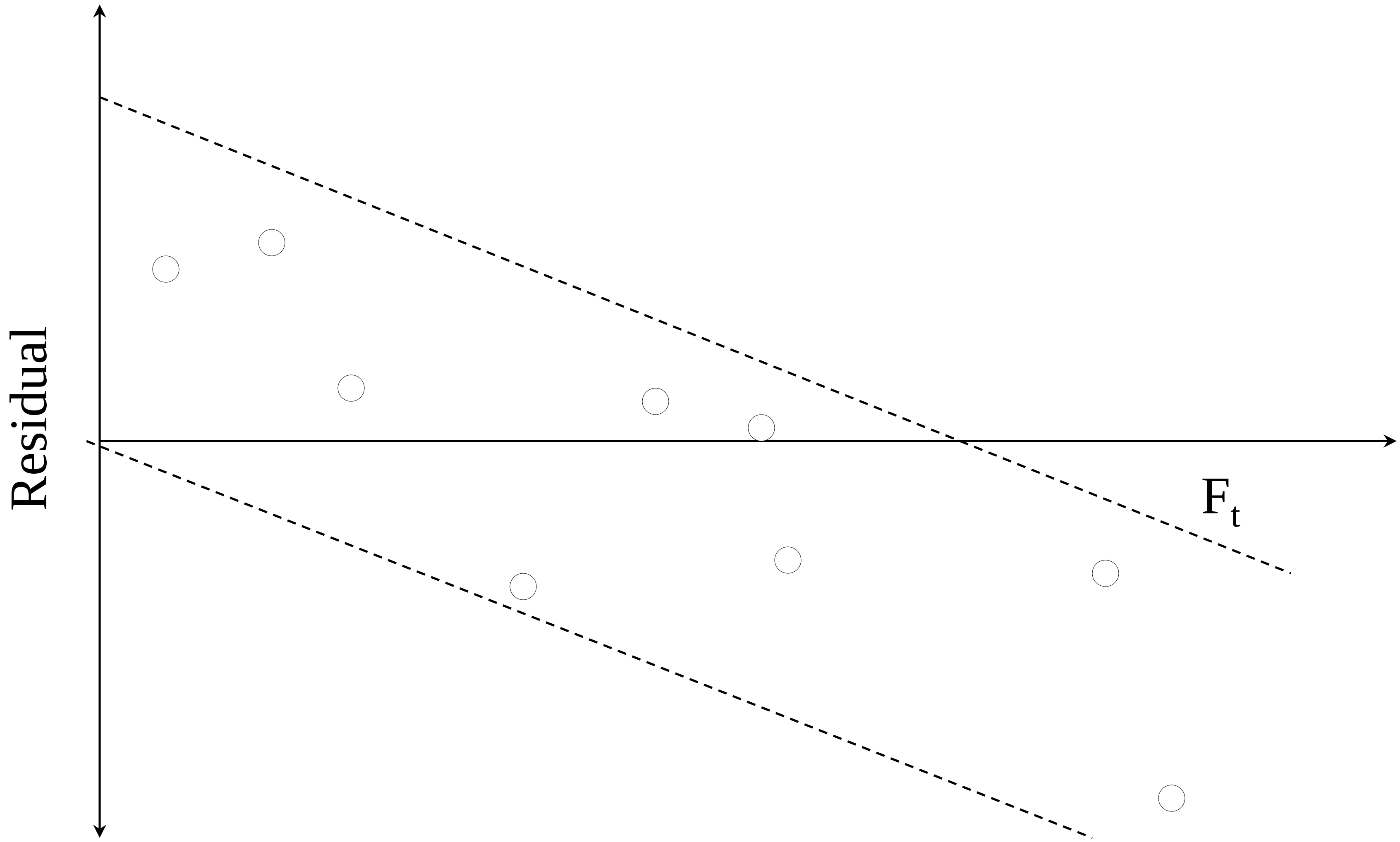
Diagnostic Checking

- You need to check residuals: $e_i = (y_i - f_i)$
  - Residual = actual - forecast
- Residual plot: residual vs. actual
  - Residual plot should be random scatter around zero
  - If not, it implies poor fit, confidence intervals invalid
  - However, estimates are still the best we can achieve, but we can’t say how good they are likely to be.
- Test for:
  - Bias: non-zero mean
  - Heteroscedasticity (non-constant variance)
  - Non-normality of residuals

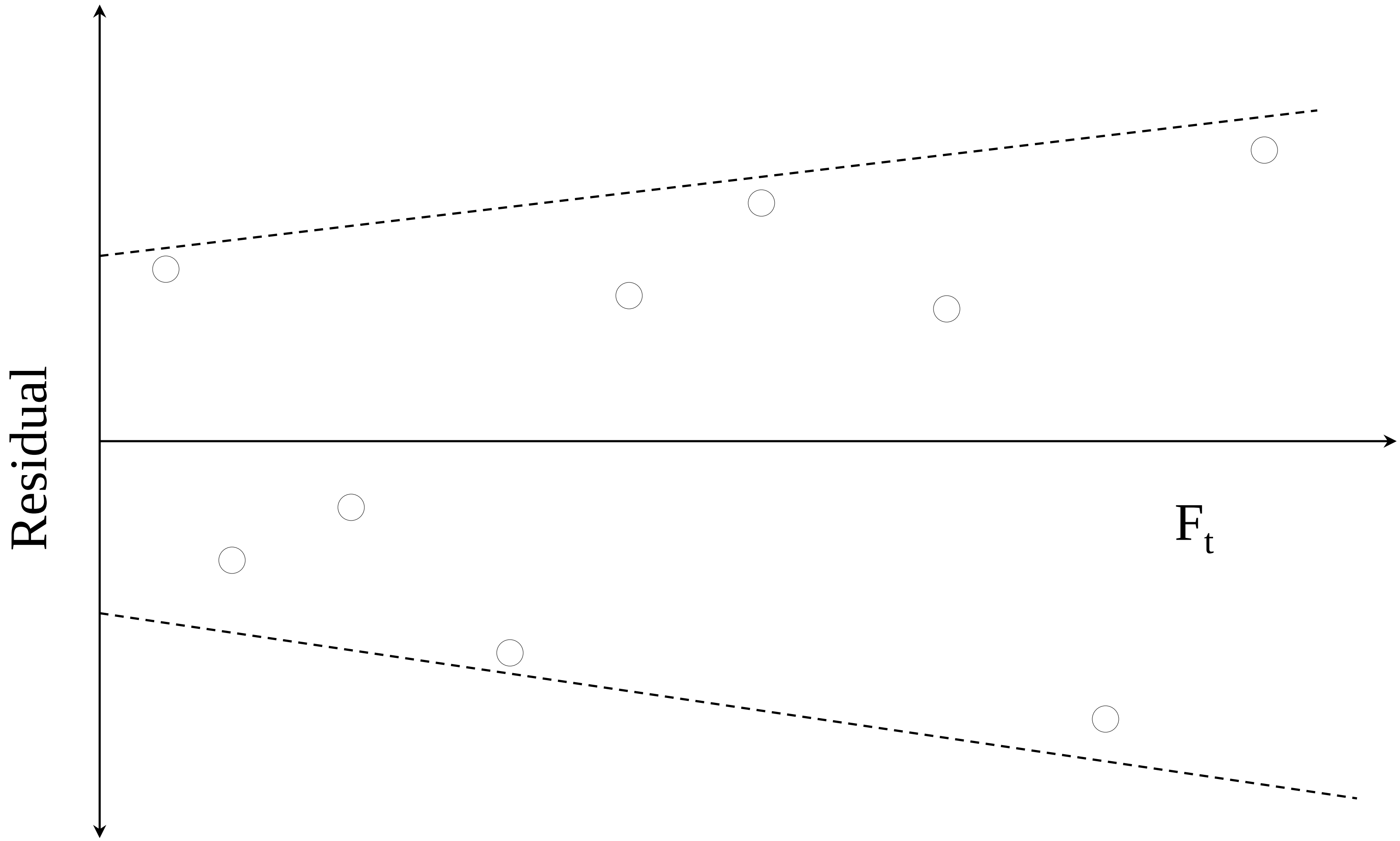
Residual Plot



Residual Plot - Bias

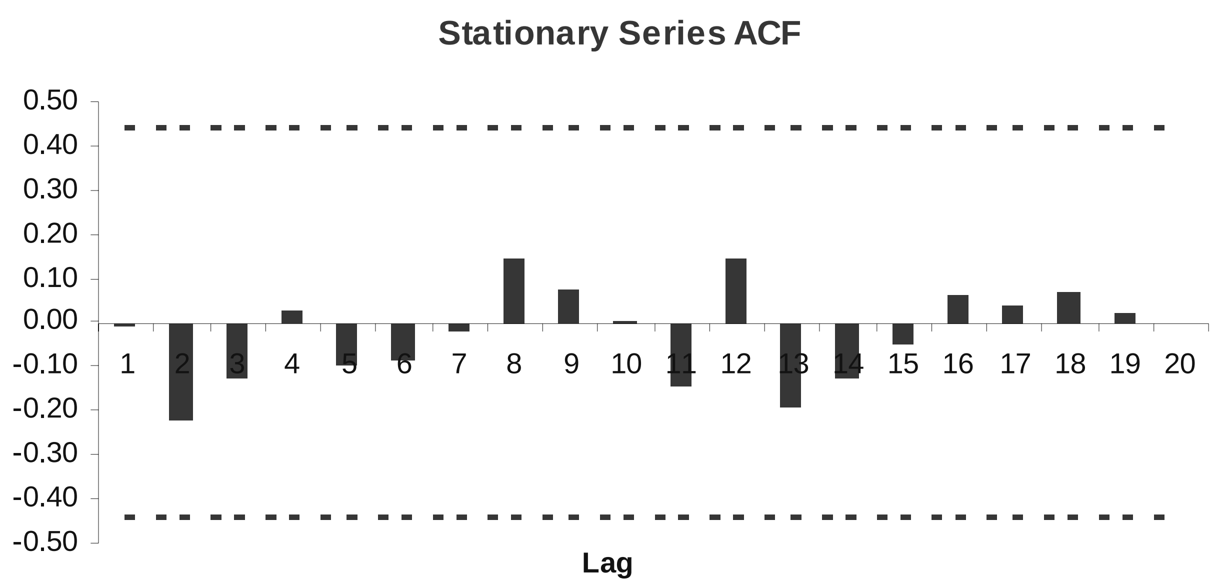


Residual Plot - Heteroscedasticity



Anderson, Bartlett & Quenoille

- ACF and PACF coefficients ~ No(0, 1/n)
  - If data is white noise
  - Hence 95% of coefficients lie in range $\pm 1.96/\sqrt{n}$



Durbin-Watson Test

$$DW = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

- Check for *serial autocorrelation in residuals*
- Range: 0 to 4. DW ≈ 2 for white noise
  - Small values indicate +ve autocorrelation
  - Large values indicate -ve autocorrelation
- NB not valid when model contains lagged values of y_t
  - Use DW-h = (1 - DW/2)√{n/[1-nσ′$_a$]} ~ no(0,1) for large n
    - – Σ′$_a$ is the standard error of the one-period lag coefficient a_1

Portmanteau Tests: Box-Pierce

- Simultaneous tests of ACF coefficients to see if data (residuals) are white noise
- Box-Pierce

$$Q = n\sum_{s=1}^{h}\rho_s^2$$

  - Usually h ≈ 20 is selected
- Used to test autocorrelations of residuals
- If residuals are white noise the Q ~ χ^2(h-m)
  - m is number of model parameters (0 for raw data)

Portmanteau Tests: Ljung-Box

- More accurate for small n

$$Q^* = n(n+2)\sum_{s=1}^{h} \frac{\rho_s^2}{n-s}$$

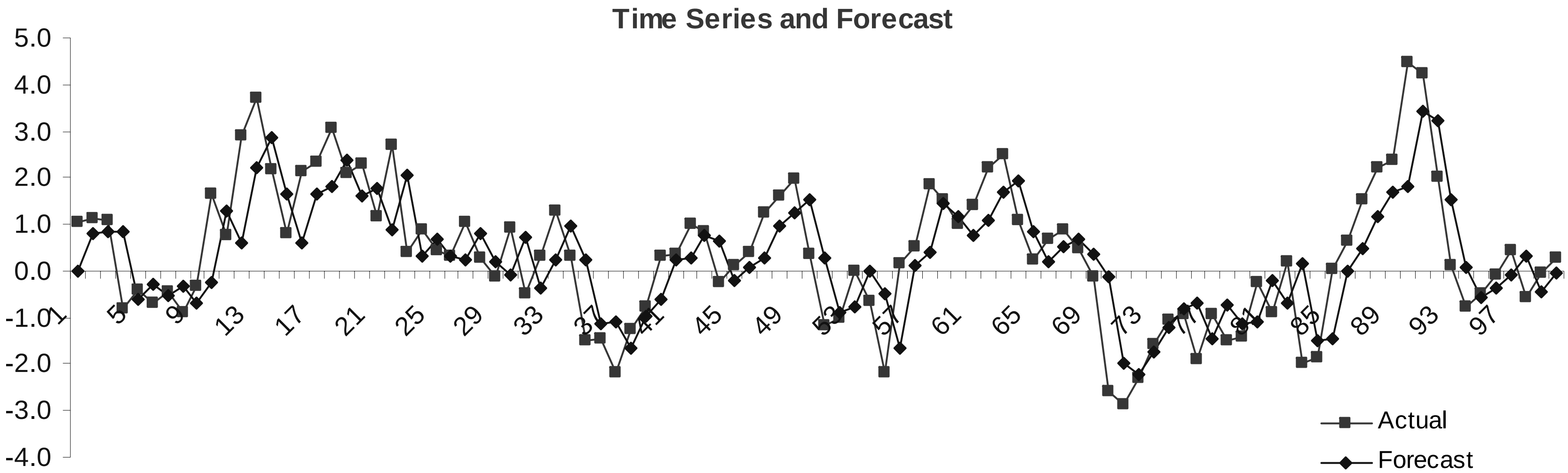
  - If data is white noise then Q* ~ χ^2(h-m)
  - Usual to conclude that data is not white noise if Q exceeds 5% of right hand tail of χ^2 distn.

Tests for Normality

- Error Distribution Moments
  - Skewness: should be ~ 0
  - Kurtosis: should be ~ 3
- Jarque-Bera Test
  - $J\text{-}B = n[\text{Skewness} / 6 + (\text{Kurtosis} - 3)^2 / 24]$
  - $J\text{-}B \sim \chi^2(2)$

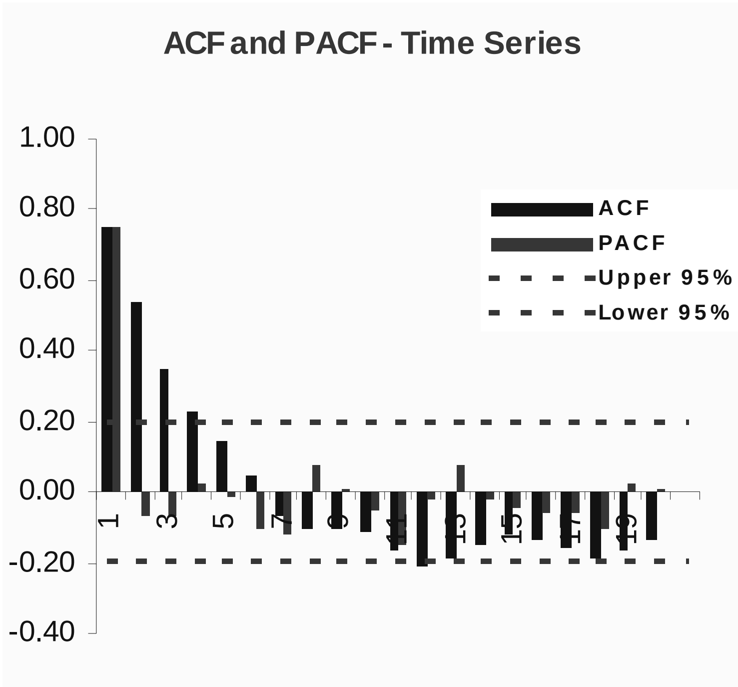
Lab: Box-Jenkins Analysis Test Data Set 1

- Fit ARMA model using
  - Using box Jenkins methodology
    - Compute & examine ACF and PACF
    - Estimate MLE model parameters
    - Check residuals using portmanteau tests
    - How good is your model at forecasting?



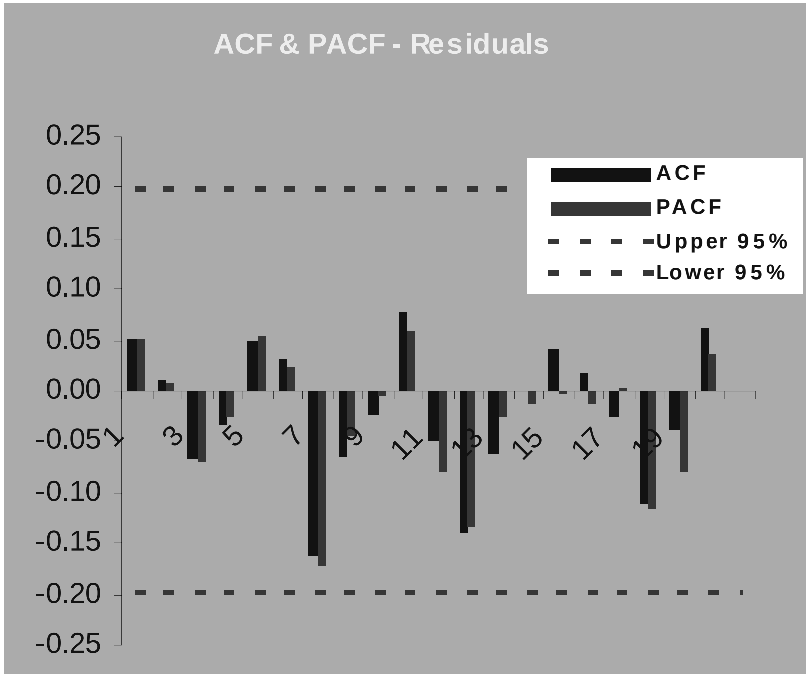
Solution: Box-Jenkins Analysis Test Data Set 1

- ACT and PACF suggest AR(1) Model



Solution: Box-Jenkins Analysis Test Data Set 1

- MLE Estimate: a = 0.766
- Residuals are white noise:



Forecast Function

- E.g. AR(1) process: $y_{t+1} = a_0 + a_1 y_t + \varepsilon_{t+1}$
- Forecast Function
  - $E_t(y_{t+1}) = a_0 + a_1 y_t$
    - $E_t(y_{t+j}) = E_t(y_{t+j} \mid y_t , y_{t-1}, y_{t-2}, \ldots , \varepsilon_t, \varepsilon_{t-1} , \ldots)$
  - $E_t(y_{t+2}) = a_0 + a_1 E_t(y_{t+1}) = a_0 + a_0 a_1 + a_1^2 y_t$
  - $E_t(y_{t+j}) = a_0(1 + a_1 + a_1^2 + \ldots + a_1^{j-1}) + a_1^2 y_t \rightarrow a_0/(1- a_1)$

Forecast Error

- J-step ahead forecast error: $\eta_t(j) = y_{t+j} - E_t(y_{t+j})$
  - $\eta_t(1) = y_{t+1} - E_t(y_{t+1}) = \varepsilon_{t+1}$
  - $\eta_t(2) = y_{t+2} - E_t(y_{t+2}) = \varepsilon_{t+2} + a_1\varepsilon_{t+1}$
  - $\eta_t(j) = \varepsilon_{t+j} + a_1\varepsilon_{t+j-1} + a_1^2\varepsilon_{t+j-2} + \ldots + a_1^{j-1}\varepsilon_{t+1}$
- Forecasts are unbiased
  - $E_t[\eta_t(j)] = E[\varepsilon_{t+j} + a_1\varepsilon_{t+j-1} + a_1^2\varepsilon_{t+j-2} + \ldots + a_1^{j-1}\varepsilon_{t+1}] = 0$

Confidence Intervals

- Forecast Variance
  - $Var[\eta_t(j)] = \sigma^2[1_j + a_1^2 + a_1^4 + \ldots + a_1^{2(j-1)}]$ ⌫ $\sigma^2/(1- a_1^2)$
    - Forecast variance is an increasing function of j
    - In limit, forecast variance converges to variance of $\{y_t\}$
- Confidence Intervals
  - $Var[\eta_t(1)] = \sigma^2$
    - Hence 95% confidence interval for y_{t+1} is $a_0 + a_1 y_t \pm 1.96\sigma$

Non-Stationarity

- Non-stationarity in the mean
  - Differencing often produces stationarity
    - E.g. if y_t is random walk with drift, Δy_t is stationary
  - Differencing Operator: Δ^d
    - Difference y_t d times to yield stationary series $\Delta^d(y_t)$
      - For most economic time series d = 1 or 2 is sufficient
- Non-stationarity in the variance
  - Use power or logarithmic transformation
    - E.g. Stock returns $r_t = Ln(P_{t+1} / P_t)$

ARIMA Models

- ARIMA(p,d,q) models
  - Autoregressive Integrated Moving Average
  - d is the order of the differencing operator required to produce stationarity (in the mean)
- Many economic time series are modeled ARIMA(0,1,1)
  - $\Delta y_t = a_0 + \varepsilon_t + \beta_1 \varepsilon_{t-1}$
  - e.g. GDP, consumption, income

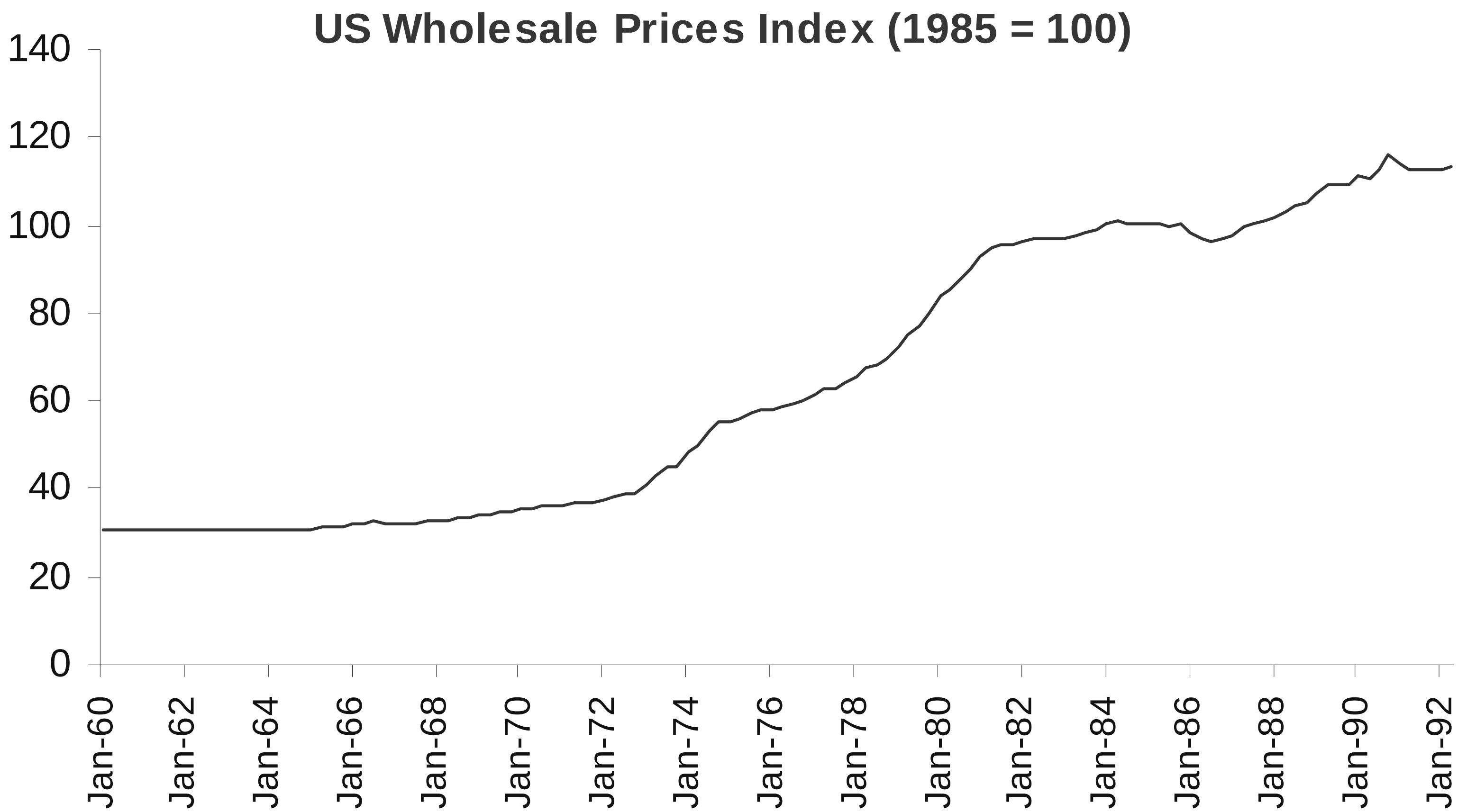
Seasonal Models

- Box-Jenkins technique for seasonal models
  - No different than for non-seasonal
  - Seasonal coefficients of the ACF and PACF appear at lags s, 2s, 3s, . .
- Examples of Seasonal Models
  - Additive
    - $y_t = a_1 y_{t-1} + a_4(y_{t-4}) + \varepsilon_t$
    - $y_t = \varepsilon_t + \beta_4(\varepsilon_{t-4}) + \varepsilon_t$
  - Multiplicative
    - $(1 - a_1L)(1 - a_4L^4)y_t = (1 - \beta_1L)\varepsilon_t$
    - Captures interactive effects in terms e.g. $(a_1a_4\, y_{t-5})$

How to Model Seasonal Data

- Explicitly in model
  - With AR and/or MA terms at lag S
- Seasonal differencing
  - Difference series at lag S to achieve (seasonal) stationarity
    - E.g. for monthly seasonality form $y^*_t = y_t - y_{12}$
    - Model resulting stationary series y^*_t in usual way

Lab: Modelling the US Wholesale Price Index



Lab: US Wholesale Price Index

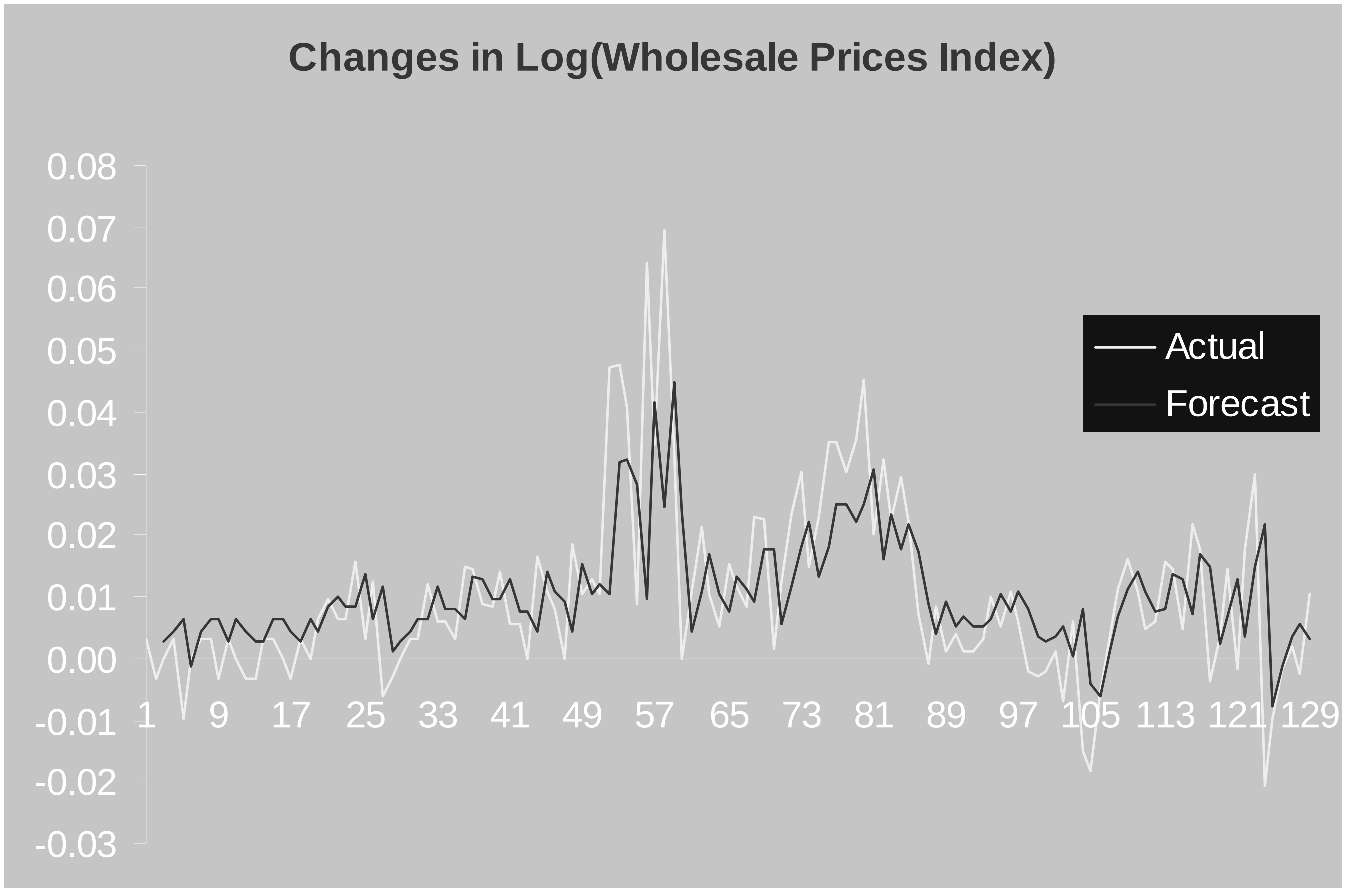
- Data preparation
  - Clearly non-stationary in mean and variance
  - Consider ΔLn(WPI)
- Identification for transformed series
  - Examine transformed series, ACF and PACF
  - Seasonal (quarterly)
- Model estimation & testing
  - AR(2)
  - ARMA(1,1)
  - ARMA(2,1) with seasonal term at lag 4

Solution: US Wholesale Price Index

- Best model is seasonal ARMA[1, (1,4)]
  - $y_t = 0.0025 + 0.7700 y_{t-1} + \varepsilon_t - 0.4246\varepsilon_{t-1} + 0.3120\varepsilon_{t-4}$

Model	a_0	a_1	a_2	β_1	β_4	AIC	BIC	Adj. R^2
AR(1)	0.0013	0.0738				-497.3	-494.4	33.3%
	0.04%	0.00%						
AR(2)	0.0035	0.4423	0.2345			-502.3	-496.6	36.4%
	0.52%	0.00%	0.46%					
ARMA(1,(1,4))	0.0025	0.7700		-0.4246	0.3120	-511.0	-502.6	42.7%
	5.96%	0.03%		3.48%	0.07%			
ARMA(2,(1,4))	0.0025	0.7969	-0.0238	-0.4411	0.3132	-509.0	-497.8	42.3%
	6.25%	0.02%	43.38%	2.98%	0.06%			

Solution: US Wholesale Price Index



Regression Models

- Linear models of form:
- $Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + \ldots + b_m X_{mt} + \varepsilon_t$
  - $\{\varepsilon_t\}$ is strict white noise process
  - X_i are independent, explanatory variables
    - May or may not be causal

Example: Regression Model for Excess Equity Returns

- Pesaran & Timmermann (1974)

$$Y_t = \beta_0 + \beta_1 YSP_{t-1} + \beta_2 PI12_{t-2} + \beta_3 DI11_{t-1} + \beta_4 DIP12_{t-2} + \varepsilon_t$$

- **Y_t** is excess return on S&P500 over the 1-month T-Bill rate.
- **YSP** is the dividend yield, defined as:

12-month average dividend / month-end S&P500 Index value

- **PI12** is the rate of change of the 12-month moving average of the producer price index:

  PI12 = Ln{PPI12 / PPI12(-12)}
- **DI11** is the change in the 1-month T-Bill rate
- **DIP12** is the rate of change of the 12-month moving average of the index of industrial production
  - **DIPI12 = Ln{IP12 / IP12(-12)}**

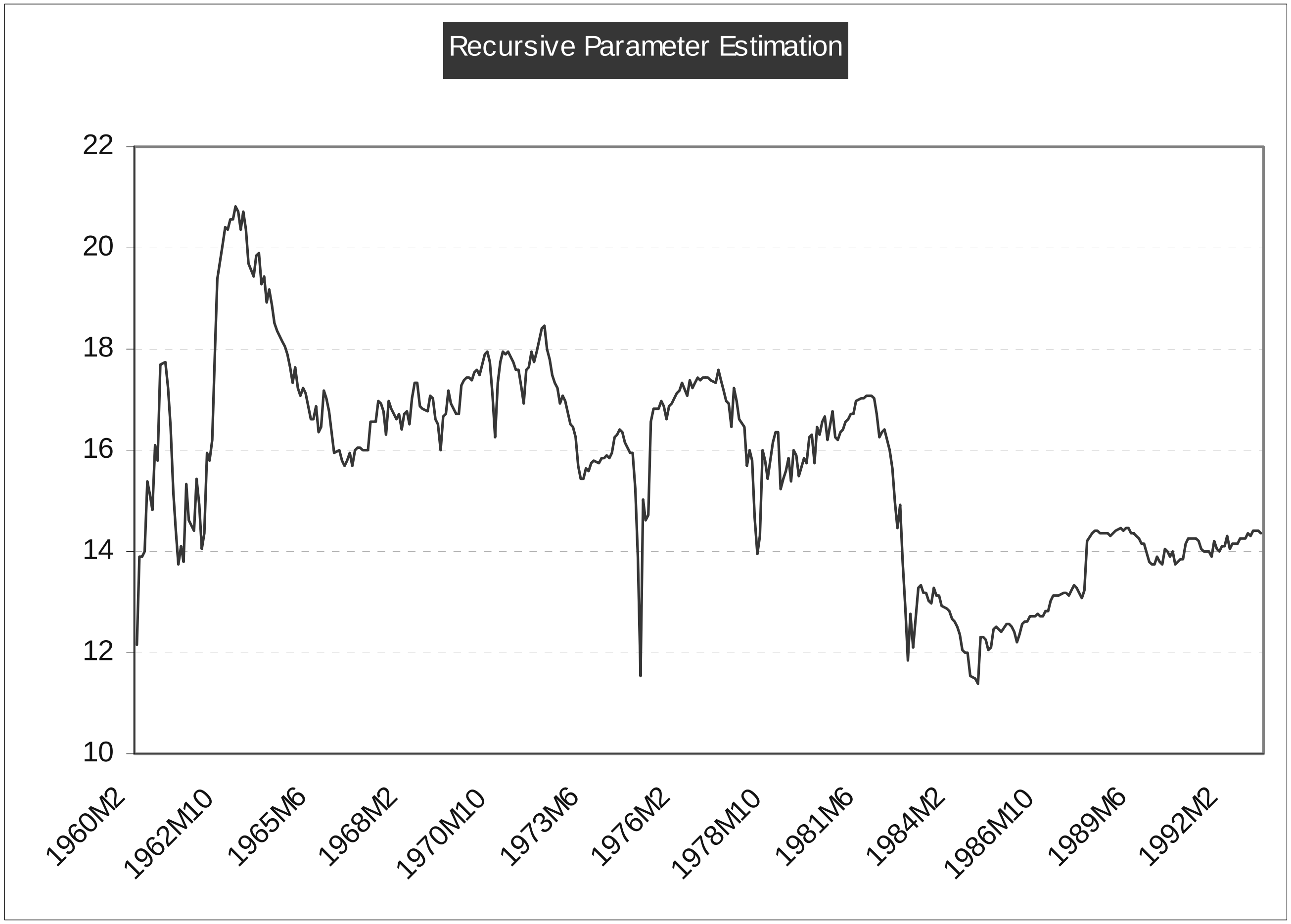
Regression Methods

- Standard Method
  - Use all data
  - Problem: data dependent; structural change
- Stepwise
  - Forward: start with minimal model, add variables
  - Backward: start with full model and eliminate variables
  - Estimate contribution of individual variables
- Rolling/ Recursive
  - Re-estimate regression over overlapping, successive fixed-length periods
  - Re-estimate regression after adding each new period's data
  - Useful for ex-ante estimation & out of sample forecasting

Lab: Recursive Regression Prediction of Excess Equity Returns

- Replicate part of Pesaran & Timmermann study
- Monthly SP500 excess returns 1954 – 1992
- Use recursive regression & ex-ante variables
- Examine forecasting performance
- Develop trading system

Recursive Parameter Estimates

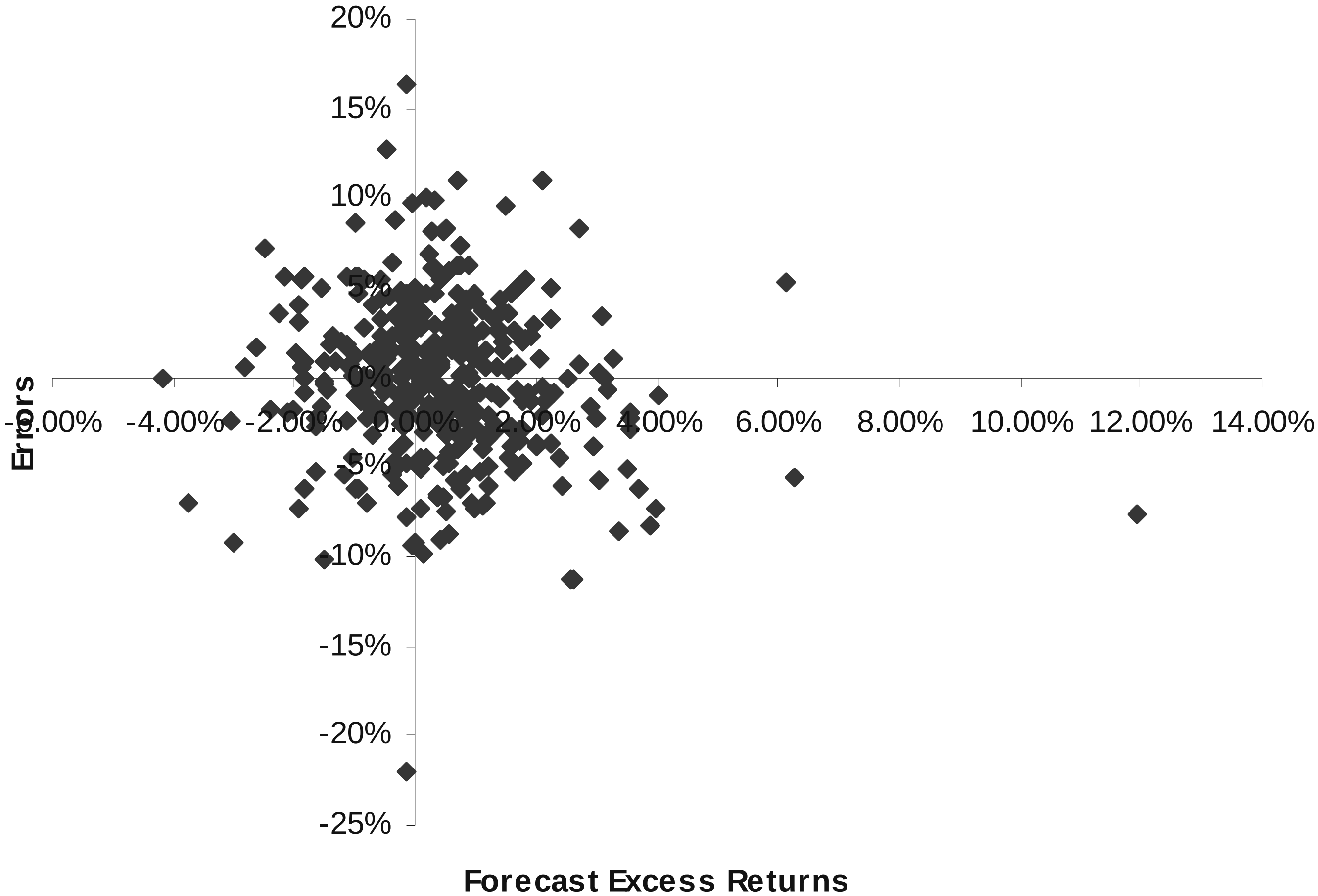


Parameter Estimates & ANOVA

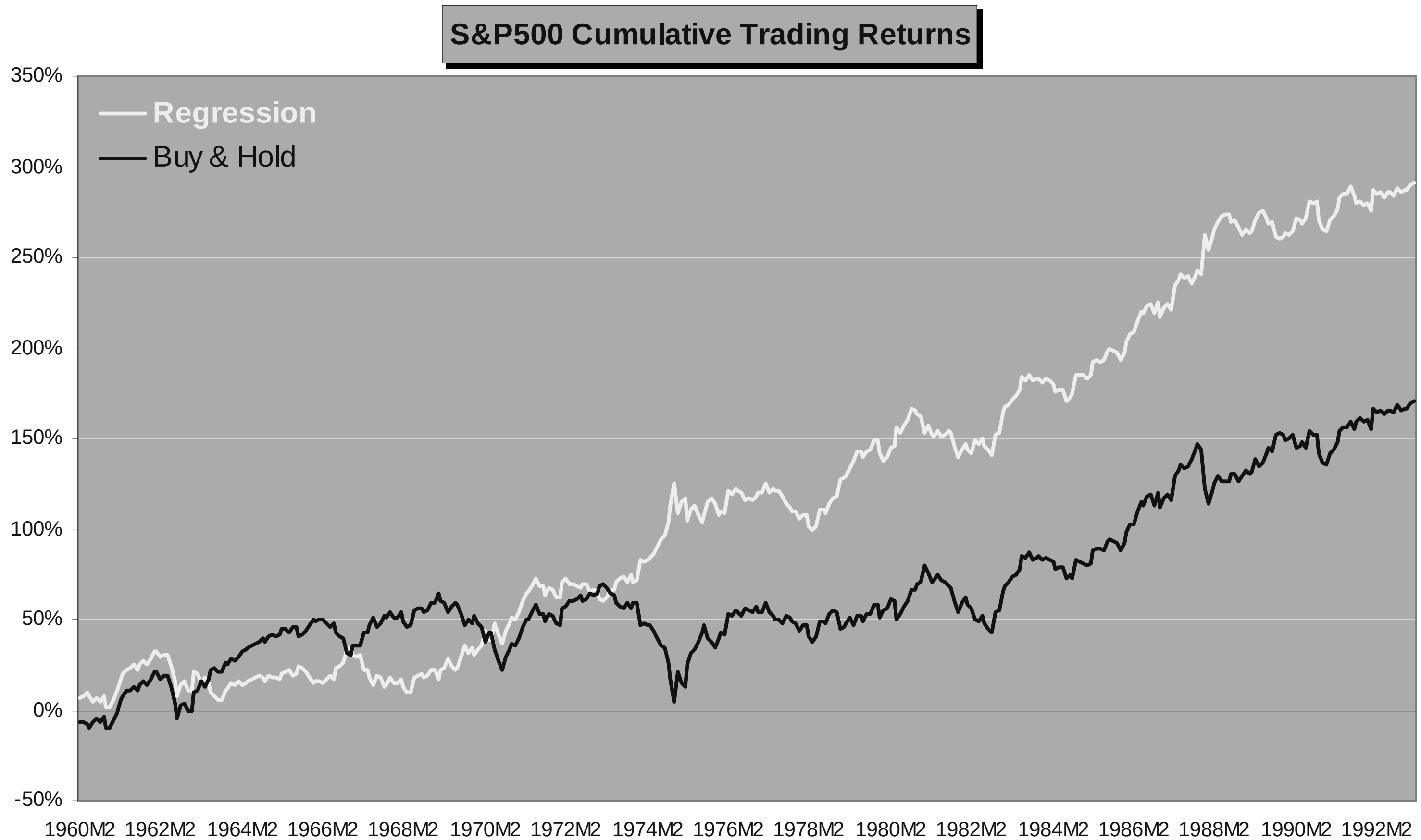
PARAMETERS	-0.024	14.338	-0.280	-0.007	-0.159
SE	0.010	3.424	0.065	0.003	0.040
t-statistic	-2.442	4.188	-4.321	-2.763	-3.941
Prob	1.497%	0.003%	0.002%	0.595%	0.009%

ANOVA					
R^2	8.6%				
Correl	20.7%				
F	10.82	DF	461.00	Prob	0.000%

Residuals


20%
15%
10%
5%
0%
-5%
-10%
-15%
-20%
-25%
-4.00%
-2.00%
0.00%
2.00%
4.00%
6.00%
8.00%
10.00%
12.00%
14.00%
Errors
Forecast Excess Returns

Trading System Performance



Random Walk Model

- Special case of AR(1) with $a_0 = 0$ and $a_1 = 1$
  - $y_t = y_{t-1} + \varepsilon_t$
  - $y_t = y_0 + \Sigma\varepsilon_i$ for $i = 1, \ldots, t$
- Mean is Constant
  - $E(y_t) = E(y_0) + E(\Sigma\varepsilon_i) = y_0$
- Conditional Mean = y_t
  - $y_{t+s} = y_t + \Sigma\varepsilon_{t+i}$ for $i = 1, \ldots, s$
  - $E_t(y_{t+s}) = y_t + E_t(\Sigma\varepsilon_{t+i}) = y_t$

Shocks and Random Walks

- Series is permanently affected by shocks
  - ε_t has non-decaying effect on $\{y_t\}$
- Variance is *time-dependant*
  - $Var(y_t) = Var(\Sigma\varepsilon_t) = t\sigma^2$
    - Hence non-stationary
- Covariance
  - $E[(y_t - y_0)(y_{t-s} - y_0) = E[(\Sigma\varepsilon_i)(\varepsilon_{t-s} + \varepsilon_{t-s-1} + \ldots \varepsilon_1)]$

$$= E[(\varepsilon_{t-s})^2 + \ldots + (\varepsilon_1)^2]$$

$$\gamma_{t-s} = (t - s)\sigma^2$$

Correlation of Random Walk Process

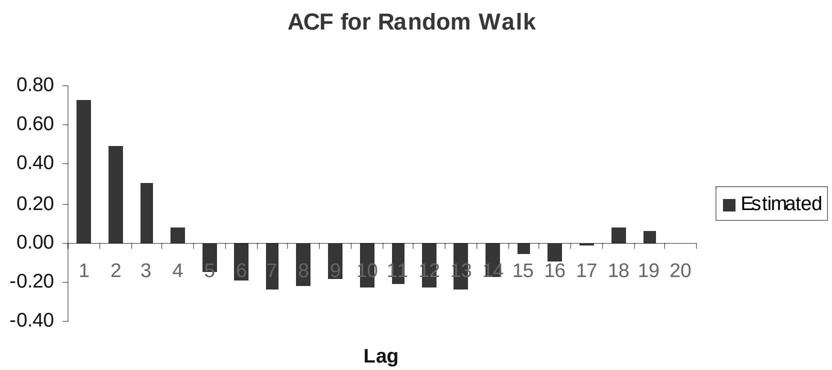
- Correlation: $\rho_s = [(t-s)/t]^{1/2}$
  - For small s, $(t-s)/t \approx 1$
  - As s increases, ρ_s will decay very slightly
- Identification Problem
  - Can't use ACF to distinguish between a unit root process ($a_1 = 1$) and one in which a_1 is close to 1
    - Will mimic an AR(1) process with a near unit root

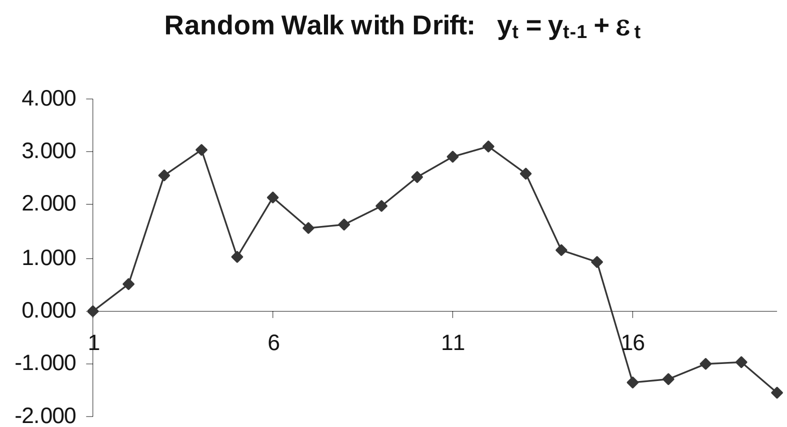
Testing for Random Walk

- AR process $y_t = a_1 y_{t-1} + \varepsilon_t$
- Hypothesis test $a_1 = 0$
  - Can use t-test
    - OLS estimate of a_1 is efficient
    - Because $|a_1| < 1$ and $\{\varepsilon_t\}$ is white noise
- Hypothesis test $a_1 = 1$; can't use t-test
  - $\{y_t\}$ is non-stationary: $y_t = \Sigma\varepsilon_i$
  - Variance becomes infinitely large
  - OLS estimate of a_1 will be *biased* below true value
    - $a_1 \sim \rho_1 = [(t-1)/t]^{1/2} < 1$

Random Walk Example

- Appears stationary
  - ACF decays to zero





Dickey-Fuller Methodology

- Use Monte-Carlo
  - Generate 10,000 unit root processes $\{y_t\}$
  - Estimate parameter a_1
  - Estimate confidence levels:
    - 90% of estimates are less than 2.58 SE from 1
    - 95% of estimates are less than 2.89 SE from 1
    - 99% of estimates are less than 3.51 SE from 1
  - Test Example
    - Suppose we have series for which estimated value of parameter a_1 is 2.95 SE < 1
    - Reject hypothesis of unit root at 5% level

Dickey-Fuller Tests

- Unit Root Process: $y_t = a_1 y_{t-1} + \varepsilon_t$
- Equivalent form
  - $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$
    - $\gamma = 1 - a_1$
- Test: $\gamma = 0$
  - Equivalent to testing $a_1 = 1$
- Other unit root regression models
  - $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$
  - $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$

Dickey-Fuller Test Procedure

- Test Procedure
  - Estimate γ using OLS
  - Compute t-statistic
    - Divide OLS estimate by SE
  - Compare t-statistic with appropriate critical value in Dickey-Fuller tables
  - Critical value depends on
    - sample size
    - form of model
    - confidence level

Critical Values

Model	Hypothesis	Test Statistic	95% and 99% critical values
$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$	$\gamma = 0$	τ_τ	-3.45 & -4.04
	$\gamma = a_2 = 0$	ϕ_3	6.49 & 8.73
	$a_0 = \gamma = a_2 = 0$	ϕ_2	4.88 & 6.50
$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ_μ	-2.89 & -3.51
	$a_0 = \gamma = 0$	ϕ_1	4.71 & 6.70
$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ	-1.95 & -2.60

Joint Tests

- Used to test joint hypotheses e.g. $a_0 = \gamma = 0$
- Constructed like ordinary F-test

$$\phi_i = \frac{\left[RSS(restricted) - RSS(unrestricted)\right]/r}{RSS(unrestricted)/(T-k)}$$

  - RSS(restricted) = error sums of squares from restricted model
  - RSS(unrestricted) = error sums of squares from unrestricted model
  - r = # restrictions
  - T = # observations
  - k = # parameters in unrestricted model

Extensions of Dickey-Fuller

- AR(p) Process
  - $y_t = a_0 + a_1 y_{t-1} + \ldots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p} + \varepsilon_t$
  - Add and subtract $a_p y_{t-p+1}$
    - $y_t = a_0 + a_1 y_{t-1} + \ldots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} - a_p \Delta y_{t-p+1} + \varepsilon_t$
  - Add and subtract $(a_{p-1} + a_p) y_{t-p+2}$
    - $y_t = a_0 + a_1 y_{t-1} + \ldots -(a_{p-1} + a_p) \Delta y_{t-p+2} - a_p \Delta y_{t-p+1} + \varepsilon_t$

General Form of AR(p) Process

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t$$

$$\gamma = -\left[1 - \sum_{i=1}^{p} a_i\right] \quad b_i = \sum_{j=i}^{p} a_i$$

- If $\gamma = 0$, equation has unit root (since all in differences)
- Hence can use same Dickey-Fuller statistic
  - No intercept or trend: τ
  - Intercept, no trend: τ_μ
  - Intercept and Trend: τ_τ

Problems With Dickey-Fuller

- How to handle MA terms
  - Invertibility: MA model $\rightarrow$ AR(∞) model
  - Said & Dickey: ARIMA(p,1,q) $\approx$ ARIMA(n, 1, 0)
    - $N \leq T^{1/3}$
- Require order of AR(p) process to estimate γ
  - Start with long lag and pare down model using standard t-tests

Tests for Multiple Unit Roots

- Dickey & Pantula
  - Perform DF tests on successive differences
- E.g. 2 unit roots suspected
  - Form $\Delta^2 y_t = a_0 + \beta_1 \Delta y_{t-1} + \varepsilon_t$
  - Use DF τ statistic to test $\beta_1 = 0$
  - If β_1 differs from zero then test for single unit root
  - Form $\Delta^2 y_t = a_0 + \beta_1 \Delta y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$
  - Test null hypothesis: $\beta_1 = 0$ using DF
    - If rejected, conclude $\{y_t\}$ is stationary

Phillips-Perron Tests

- Phillips-Perron generalizes DF to cover:
  - Serially correlated errors and non-constant variance
  - Models: $y_t = a_0 + a_1 y_{t-1} + a_2 t + \mu_t$
    - Test $a_1 = 0$ using standard DF critical values and statistic:

$$t_{a1}\left(\frac{\tilde{S}}{\tilde{\sigma}_{T\omega}}\right) - \left(\frac{T^3}{4\sqrt{3}D_X^{1/2}\tilde{\sigma}_{T\omega}}\right)\left(\tilde{\sigma}_{T\omega}^2 - \tilde{S}^2\right)$$

    - $D_X = \det(X^T X)$, the determinant of the regressor matrix X
    - S is the standard error of the regression
    - ω is the # of estimated correlations

$$\tilde{\sigma}_{T\omega}^2 = \frac{1}{T}\sum_{i=1}^{T} u_i^2 + \frac{2}{T}\sum_{s=1}^{T}\sum_{t=s+1}^{T} u_t u_{t-s}$$

Problems in Testing for Unit Roots

- Low power of unit root tests
  - Can't distinguish between unit root and near unit root process
  - Too often indicate that process contains unit root
- Tests are conditional on model form
  - Tests for unit roots depend on presence of deterministic regressors
  - Test for deterministic regressors depend on presence of unit roots

Unit Roots In FX Markets

- Purchasing power parity
  - Currency depreciates by difference between domestic & foreign inflation rates
- PPP model
  - $E_t = p_t - p^*_t + d_t$
    - E_t is log of dollar price of foreign exchange
    - p_t is log of US price levels
    - p^*_t is log of foreign price levels
    - d_t represents deviation from PPP in period t
- Testing PPP
  - Reject if series $\{d_t\}$ is non-stationary

Real Exchange Rates

- Real exchange rates
  - Define $r_t \equiv e_t + p^*_t - p_t$
  - PPP holds if $\{r_t\}$ is stationary
- Create series using:
  - $r_t = Ln(S_t \times WPI^{JP}_t / WPI^{US}_t)$
    - S_t is the spot yen fx rate at time t
    - WPI^{JP}_t is the Japanese whole price index at time t (Feb 1973 = 100)
    - WPI^{US}_t is the US whole price index at time t

Lab: Testing Purchasing Power Parity

- Worksheet: PPP
  - Series of real Yen FX rates 1973-89
- Dickey Fuller Test
  - Form series $\Delta r_t = a_0 + \gamma r_{t-1} + \varepsilon_t$
  - Estimate parameters using max. likelihood
  - Do T-Test
  - D-F test with critical value of -2.88

Solution: Purchasing Power Parity

	MLE	SE	t	p
a_0	0.038	0.0203	1.881	6.14%
γ	-0.031	0.0173	-1.820	7.03%
m	1			
n	202			

Max Likelihood	
AIC	-291.35
BIC	-288.04
DW	2.03
R^2	1.6%
Adj. R^2	1.1%

ANOVA	DF	SS	MS	F	p
Model	1	0.0039	0.00388	3.31	7.03%
Error	200	0.2340	0.00117		
Total	201	0.2379			

Portmanteau Tests	Q(24)	p
Box-Pierce	26.83	26.32%
Ljung-Box	29.10	17.69%

- T-Test: H_0: $\gamma = 0$
  - Could reject at the 93% confidence level
    - Conclude series is stationary and PPP holds
- Dickey-Fuller
  - Can't reject unit root hypothesis at 95% level

Summary: Time Series Analysis

- Simple methods
  - Exponential smoothing, etc.
    - Simple, low cost, often effective
    - Limitations
      - Query out of sample performance
      - Underlying model not articulated
- ARIMA models
  - Staple of econometricians
  - Models articulated and testable
  - Limitations
    - Estimation is non-trivial
    - Problems with (near) random processes