

# Financial Econometrics

## Linear Time Series Characterizations

Gerald P. Dwyer, Jr.

Trinity College, Dublin

February 2009

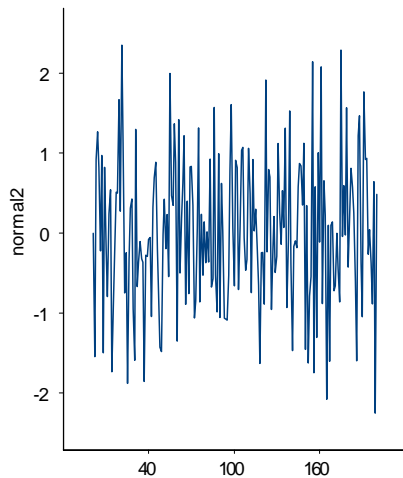
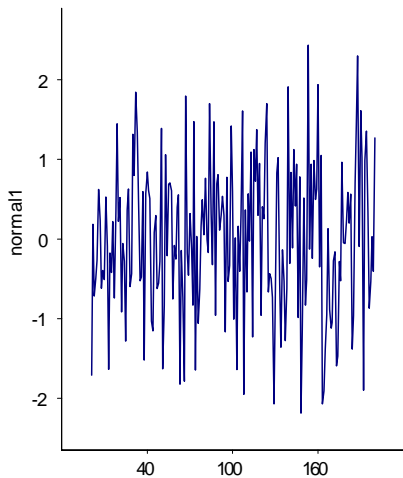
- Analyzing some data, we assume that the future bears some resemblance to the past
  - “Everything changes all the time”
  - If taken seriously, this observation implies we have no idea what will happen next
- We some aspects of a time series do not change over time
- Definition: Stochastic process is a family of random variables ordered by  $t$

For a given  $t$ , the value of  $r_t$  is determined by a probability distribution function e.g.  $r_t \sim N(\mu, \sigma^2)$  for a particular  $t$ , e.g.  $t = \text{January 2005}$   
Or  $r_t \sim N(\mu, \sigma_t^2)$

# Problem of one realization

- We observe one particular realization (actual sequence)
- Can we estimate parameters from one realization?

# Two iid normally distributed sets of values



# Constantly changing series with arbitrary changes

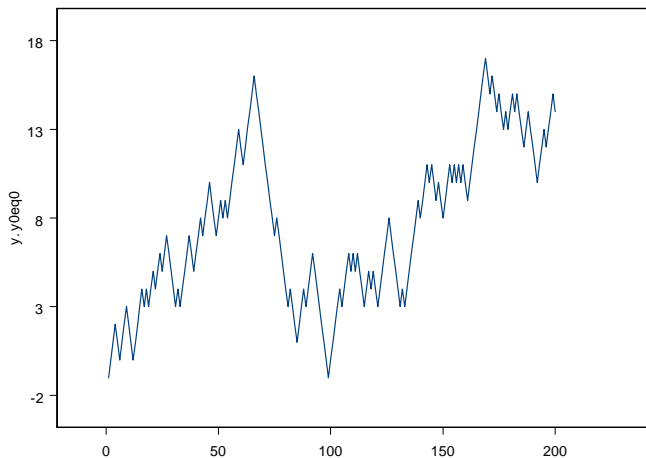
- Suppose have  $r_t \sim \mathbf{N}(\mu_t, \sigma_t)$  where  $\mu_t$  and  $\sigma_t$  are “arbitrary”
  - Can be anything every period
  - Not “random”, determined by some probability law or rule
- We can't estimate  $\mu_t$  or  $\sigma_t$

- Suppose have  $r_t = r_{t-1} + \varepsilon_t$

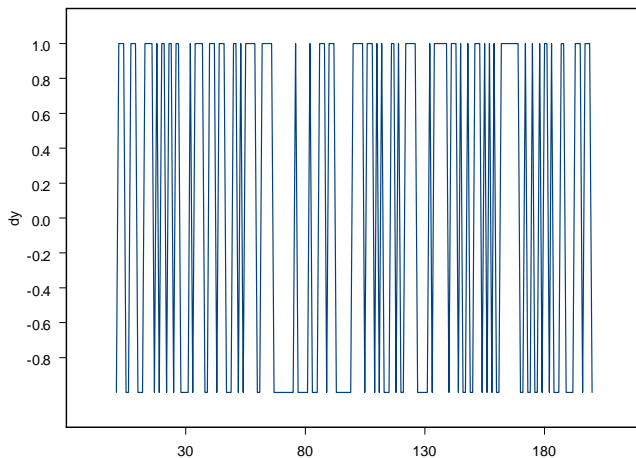
$$\varepsilon_t = \left\{ \begin{array}{l} 1 \text{ with probability } \frac{1}{2} \\ -1 \text{ with probability } \frac{1}{2} \end{array} \right\}$$

- Observe one realization
- Not “arbitrary”; it is determined by some probability law or rule

# One realization from this process



# First difference of values



# Can estimate parameter for this model

- Obvious estimator is fraction of plus ones for probability
  - Consistent
  - Average over time converges to parameter

- Always can write a joint cumulative distribution function (cdf) for  $T$  observations

$$F(r_1, r_2, \dots, r_T)$$

- What can we assume to be able to estimate parameters?
- **Strictly stationary:** Definition is

$$\begin{aligned} F(r_t) &= F(r_{t+k}) \quad \forall t \text{ and } k \\ F(r_t, r_{t+1}) &= F(r_{t+k}, r_{t+1+k}) \\ &\dots \end{aligned}$$

- Moments need not be finite
- **Covariance stationary:** Definition is first and second moments are constant and finite
  - Interested in first and second moments in linear time-series analysis
  - Want them constant
  - Not just mean and variance, but also covariances over time
  - Also called weakly stationary

# Covariance stationary

- Mean

$$E[r_t] = \mu$$

where  $\mu$  is the mean

- Second moments

$$\text{Var}[r_t] = \sigma^2$$

$$\text{Cov}[r_t, r_{t-k}] = \gamma_k$$

- Mean and second moments are constant and finite

# Examples of series that are covariance stationary or not

- First example

$$r_t = \varepsilon_t, \quad \mathbb{E} \varepsilon_t = 0, \quad \mathbb{E} \varepsilon_t \varepsilon_s = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}$$

- Second example

$$r_t = \varepsilon_t, \quad \mathbb{E} \varepsilon_t = 0, \quad \mathbb{E} \varepsilon_t \varepsilon_s = \begin{cases} \sigma^2 t, & t = s \\ 0, & t \neq s \end{cases}$$

- Third example

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \quad \mathbb{E} \varepsilon_t = 0, \quad \mathbb{E} \varepsilon_t \varepsilon_s = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}, \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty$$

- 

$$\mathbb{E} r_t = \mu$$

$$\mathbb{E} (r_t - \mu)^2 = \mathbb{E} \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$$

- A time series is **linear** if its evolution can be summarized as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

where the sequence  $\{\varepsilon_{t-i}\}$  is **independent and identically distributed**

- It is a normalization to set  $\psi_0 = 1$  ( $\psi_0$  can be chosen arbitrarily so we might as well pick a number which simplifies things, like unity)
- Autocorrelations tell us much about behavior of a covariance stationary series

$$\rho_k = \frac{\text{Cov}[r_t, r_{t-k}]}{\text{SD}[r_t] \text{SD}[r_{t-k}]} = \frac{\text{Cov}[r_t, r_{t-k}]}{\text{Var}[r_t]}$$

- Estimator

$$\hat{\rho}_k = \frac{\sum (r_t - \bar{r})(r_{t-k} - \bar{r})}{\sum (r_t - \bar{r})^2}$$

# Tests for autocorrelation equal to zero

- $\rho_k = 0$
- If  $r_t$  is IID (which implies  $\rho_k = 0$ ) and  $\mathbb{E} r_t^2 < \infty$ , then asymptotically  $\hat{\rho}_k \sim N(0, 1/T)$  and the test statistic is based on the t-ratio

$$t = \sqrt{T} \hat{\rho}_k$$

- If  $r_t = \mu + \sum_{i=0}^q \psi_i \varepsilon_{t-i}$  and  $\varepsilon_t \sim N(0, \sigma^2)$ , then

$$\hat{\rho}_l \sim N\left(0, \left(1 + 2 \sum_{i=1}^q \rho_i^2\right) / T\right) \text{ for } l > q$$

- Portmanteau test – Ljung-Box test on  $m$  autocorrelations

$$Q(m) = T(T+2) \sum_{l=1}^m \frac{\hat{\rho}_l^2}{T-l}$$

where  $Q(m) \sim \chi_m^2$  and reject null hypothesis that first  $m$  autocorrelations are zero if p-value less than value  $\alpha$

# Box-Jenkins time series analysis

## Overall estimation strategy

- Basic idea is to estimate process generating series from data (suppress mean)

$$r_t = \sum_{j=1}^p \varphi_j r_{t-j} + \sum_{i=1}^q \psi_i \varepsilon_{t-i} + \varepsilon_t$$

- ARMA(p,q) model
- Suppose that  $E \varepsilon_t = 0, E \varepsilon_t \varepsilon_s = 0 \forall t \neq s$
- What model to fit?
  - Compare values of moments to observed values
- Steps
  - 1 Reduce to stationarity – trend or first difference
  - 2 Identification – compare observed second moments to those from various simple processes and pick best
  - 3 Estimate process
  - 4 Check diagnostics to see whether estimated process is adequate
  - 5 If diagnostics indicate problems, go back to 2; else done

# Partial autocorrelation function

- Help for identifying process
- Partial autocorrelation function

$$\text{Corr} \left[ r_t, r_{t-k} \mid r_{t-1}, r_{t-2}, \dots, r_{t-(k-1)} \right] = \phi_{kk}$$

- Coefficient of the  $k$ th lagged value in autoregression with  $k$  lagged values
  - Regress  $r_t$  on  $r_{t-1}, r_{t-2}, \dots, r_{t-k}$  and examine statistical significance of last coefficient
  - t-ratio of last coefficient in regression
- Use sequence of last coefficients to evaluate how many to include

# Example of covariance stationary AR(1)

$$r_t = \varphi_0 + \varphi_1 r_{t-1} + \varepsilon_t, \quad |\varphi_1| < 1, \quad \mathbb{E} \varepsilon_t = 0, \quad \mathbb{E} \varepsilon_t \varepsilon_s = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}$$

- Properties of autocorrelation function

$$\mathbb{E} r_t = \varphi_0 + \varphi_1 \mathbb{E} r_{t-1} + \mathbb{E} \varepsilon_t = \varphi_0 + \varphi_1 \mathbb{E} r_{t-1}$$

$$\mathbb{E} r_t = \frac{\varphi_0}{1 - \varphi_1} \equiv \mu$$

$$\text{Var} [r_t] = \mathbb{E} (r_t - \mathbb{E} r_t)^2 = \frac{\sigma^2}{1 - \varphi_1^2} \equiv \gamma_0$$

$$\text{Cov} [r_t, r_{t-1}] = \varphi_1 \gamma_0 \equiv \gamma_1$$

$$\text{Cov} [r_t, r_{t-k}] = \varphi_k \gamma_{k-1}$$

$$\text{Corr} [r_t, r_{t-k}] = \varphi_1^k$$

as  $\text{Cov} [r_t, r_{t-k}] / \text{Var} [r_t] = \varphi_1 \gamma_{k-1} / \gamma_0 = \varphi_1^2 \gamma_{k-2} / \gamma_0 = \dots = \varphi_1^{k-1} \varphi_{k-(k-1)} \gamma_0 / \gamma_0 = \varphi_1^k$

# Example of AR(1)

## Partial autocorrelation function

$$r_t = \varphi_0 + \varphi_1 r_{t-1} + \varepsilon_t, \quad |\varphi_1| < 1, \quad E\varepsilon_t = 0, \quad E\varepsilon_t \varepsilon_s = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}$$

- Properties of partial autocorrelation function

$$r_t = \varphi_0 + \varphi_{11} r_{t-1} + \varepsilon_t$$

$$\varphi_{11} = \varphi_1$$

$$r_t = \varphi_0 + \varphi_{21} r_{t-1} + \varphi_{22} r_{t-2} + \varepsilon_t$$

$$\varphi_{22} = 0$$

$$r_t = \varphi_0 + \varphi_{21} r_{t-1} + \varphi_{22} r_{t-2} + \dots + \varphi_{kk} r_{t-k} + \varepsilon_t$$

$$\varphi_{kk} = 0$$

- In general, AR( $k$ ) has first  $k$  nonzero partial autocorrelations are nonzero and the rest are zero

# Determination of lag length

- How many lags to include?
  - Sequential regression strategy
  - Use t-ratio on last estimated coefficient in regression longer than plausible
    - 1 Select lag length  $k$  greater than plausible
    - 2 Estimate regression of lag length  $k$
    - 3 If last coefficient statistically significant using t-ratio, stop; otherwise reduce lag length ( $k$ ) by one and go to 2

- Akaike information criterion

$$\text{AIC} = -\frac{2}{T} \ln (\text{max of likelihood}) + \frac{2}{T} (\text{number of parameters})$$

- For normal distribution of errors and  $k$  lags,

$$\text{AIC}(k) = -\frac{2}{T} \hat{\sigma}_k^2 + \frac{2}{T} k$$

- (Schwarz) Bayesian information criterion (for normal distribution of errors and  $k$  lags)

$$\text{BIC}(k) = \ln \hat{\sigma}_k^2 + \ln T/k$$

- One-step-ahead forecast

- At time period  $h$  and want to forecast this period from AR(1)  
 $r_t = \varphi_0 + \varphi_{11}r_{t-1} + \varepsilon_t$  and know  $r_h$
- Forecast  $r_{h+1}$  by

$$\hat{\varphi}_0 + \hat{\varphi}_1 r_h \equiv \hat{r}_h(1)$$

- Next period – two-step-ahead forecast

$$\hat{r}_h(2) = \hat{\varphi}_0 + \hat{\varphi}_1 \hat{r}_h(1) = \hat{\varphi}_0 + \hat{\varphi}_1 (\hat{\varphi}_0 + \hat{\varphi}_1 r_h) = \hat{\varphi}_0 + \hat{\varphi}_1 \hat{\varphi}_0 + \hat{\varphi}_1^2 r_h$$

- Successive substitution gives forecasts for future periods, k-step-ahead forecast error

$$\hat{r}_h(k) = \hat{\varphi}_0 + \hat{\varphi}_1 \hat{r}_h(k-1) = \dots = \hat{\varphi}_0 + \hat{\varphi}_1 \hat{\varphi}_0 + \dots + \hat{\varphi}_1^{k-1} \hat{\varphi}_0 + \hat{\varphi}_1^k r_h$$

# Forecast errors

- One-step-ahead forecast error

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = r_{h+1} - \hat{\varphi}_0 + \hat{\varphi}_1 r_h$$

Ignoring that parameters are estimated

$$\text{Var}[e_h(1)] = \text{Var}[\varepsilon] = \sigma^2$$

- Two-step-ahead forecast error

$$e_h(2) = r_{h+k} - \hat{r}_h(k) = r_{h+k} - \hat{\varphi}_0 + \hat{\varphi}_1 \hat{\varphi}_0 + \hat{\varphi}_1^2 r_h$$

- and again ignoring estimation of parameters

$$\text{Var}[e_h(2)] = (1 + \varphi_1^2) \sigma^2$$

- Note that forecast error two steps ahead is greater
- Can proceed similarly for further ahead in future and variance of forecast error continues to increase
- As forecast horizon goes to infinity, forecast goes to mean  $\hat{\mu} = \frac{\hat{\varphi}_0}{1 - \hat{\varphi}_1}$

# Moving-average models

- In practice, easier to identify and harder to estimate
- MA(1)

$$r_t = \psi_0 + \psi_1 \varepsilon_{t-1} + \varepsilon_t, \quad E \varepsilon_t = 0, \quad E \varepsilon_t \varepsilon_s = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}$$

- Autocorrelations

$$\text{Cov}[r_t, r_{t-1}] = \psi_1 \sigma^2$$

$$\text{Cov}[r_t, r_{t-k}] = 0 \text{ for all } k > 1$$

$$\text{Var}[r_t] = (1 + \psi_1^2) \sigma^2$$

$$\text{Corr}[r_t, r_{t-1}] = \frac{\psi_1}{1 + \psi_1^2}$$

- First autocorrelation is nonzero and the rest of the autocorrelations are zero
- Exponential decay of partial autocorrelations – same as autocorrelations for MA(1)
- Two strategies for estimating this model with sample from  $t = 1$  to  $T$

# Forecasts with MA(1)

- Forecast one step ahead at  $h$  with  $e_h$  (the estimated error in  $h$ ) known
  - Because know  $e_h = r_h - \hat{\psi}_0 - \hat{\psi}_1 e_{h-1}$

$$\hat{r}_h(1) = \hat{\psi}_0 + \hat{\psi}_1 e_h$$

- Forecast two steps ahead

$$\hat{r}_h(2) = \hat{\psi}_0 + \hat{\psi}_1 e_{h+1}$$

but don't know  $e_{h+1}$  and therefore forecast is just mean  $\hat{r}_h(2) = \hat{\psi}_0$

- In general, with MA( $q$ ), forecasts different than mean for  $q$  periods and then forecast is just mean

- Model such as

$$r_t = \sum_{j=1}^p \varphi_j r_{t-j} + \sum_{i=1}^q \psi_i \varepsilon_{t-i} + \varepsilon_t$$

- In general, not useful beyond  $p=2$  or  $q=2$ 
  - Also, hard to estimate high-order MA models
  - High-order AR models often work reasonably well although error estimating parameters becomes an issue
- MA versus AR or both
  - As with the AR(1), an AR(p) has partial autocorrelations that fall to zero after lag p
  - As with the MA(1), an MA(q) has autocorrelations that fall to zero after q

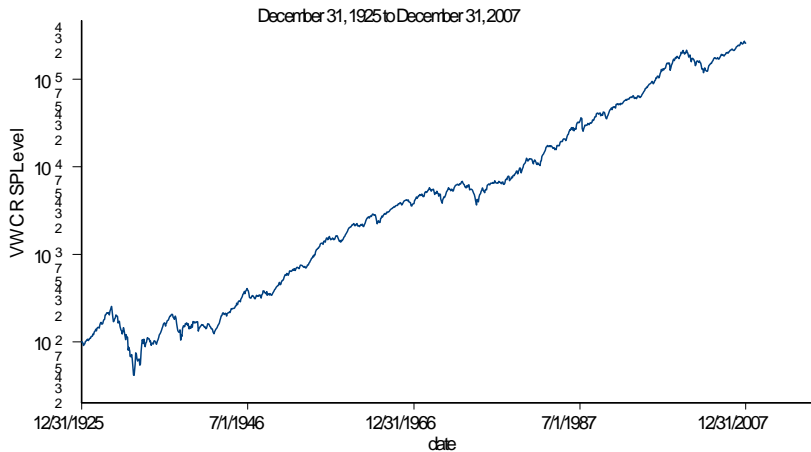
# Nonstationarity in mean

- Two types of nonstationarity in mean usually considered
  - Unit-root
  - Trend
- Unit root nonstationary (difference stationary)
  - $r_t - r_{t-1}$  is stationary
- Trend stationary
  - $r_t - \beta t$  is stationary
- Huge literature on unit roots
  - Some literature on unit roots versus trend stationary
  - More in macroeconomics than finance

# S&P 500 Monthly



# Value-weighted CRSP level with dividends reinvested Monthly



# What can we infer from graph?

- No evidence of a constant mean or level of stock prices in 82 years
- Not likely to have a stationary representation of level
- Generally goes up but not always
- Unit root nonstationarity or trend?

# Why is a unit root usually assumed in financial time series?

Returns series at least

- Suppose that mean log return is constant

$$E r_t = \mu$$

- A simple model of stock returns consistent with this is

$$r_t = \mu + \varepsilon_t$$

where  $\varepsilon_t$  is white noise – serially uncorrelated with constant-variance

- This implies

$$p_t = \mu + p_{t-1} + \varepsilon_t$$

where  $p_t$  is the natural logarithm of the level of the index

- This is a random walk with drift, where the drift is  $\mu$

- Also

$$E [p_{t+1} | p_{t-1}] = \mu + p_t$$

and

$$\hat{r}_t(1) = \mu$$

- Knowing past price does not help to predict return.

# Suppose that returns were trend stationary

- This can be consistent with a simple representation such as

$$p_t = \alpha + \beta t + \gamma p_{t-1} + \varepsilon_t$$

- where  $t$  is a variable measuring time and  $\beta$  is its coefficient
- With this representation, returns are

$$r_t = \alpha + \beta t + (\gamma - 1) p_{t-1} + \varepsilon_t$$

- If  $\varepsilon_t$  is white noise, then

$$\hat{r}_t(1) = \alpha + \beta(t+1) + (\gamma - 1) p_t$$

- Has implications many do not believe
  - Implies mean reverting component of level because  $(\gamma - 1) < 0$
- Both random walk and trend representation imply predictable increases over time
- Nothing in economics or finance says this is impossible
  - There are reasons to think it's improbable
  - It's just generally not believed to be true

# Unit root tests

## Dickey-Fuller test

- Dickey-Fuller test with no drift or trend
  - Constant level versus random walk

$$r_t = \varepsilon_t$$

$$r_t = \alpha + (\gamma - 1) p_{t-1} + \varepsilon_t$$

- Test whether  $(\gamma - 1) = 0$
- Table of test statistics because finite-sample distribution very different than normal or other standard ones

# Unit root tests

## Dickey-Fuller test

- Dickey-Fuller test with drift or trend
  - Trend in level versus random walk with drift

$$r_t = \mu + \varepsilon_t$$

$$r_t = \alpha + \beta t + (\gamma - 1) p_{t-1} + \varepsilon_t$$

- Test whether  $(\gamma - 1) = 0$
- Table of test statistics because distribution very different than normal or other standard ones and different than with no trend

# Unit root tests

## Augmented Dickey-Fuller test

- Augmented Dickey-Fuller test with drift or trend
  - Trend in level versus random walk with drift

$$r_t = \mu + \sum_{i=1}^k \pi_i r_{t-i} + \varepsilon_t$$

$$r_t = \alpha + \beta t + (\gamma - 1) p_{t-1} + \sum_{i=1}^k \pi'_i r_{t-i} + \varepsilon_t$$

- Test whether  $(\gamma - 1) = 0$
- Table of test statistics because distribution very different than normal or other standard ones

# Unit root tests

## Phillips-Perron test

- Phillips-Perron test probably industry standard today
- Same test statistics but different tables
- Trend in level versus random walk with drift

$$r_t = \mu + \varepsilon_t$$

$$r_t = \alpha + \beta t + (\gamma - 1) p_{t-1} + \varepsilon_t$$

- Test whether  $(\gamma - 1) = 0$
- Error term generally not serially uncorrelated
- Table of test statistics because distribution very different than normal or other standard ones