

# Financial Econometrics

## Nonlinear time series analysis

Gerald P. Dwyer

Trinity College, Dublin

January 2011

# What does nonlinear mean?

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

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

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

- ARCH models are nonlinear, as are stochastic volatility models
  - Linear in mean function though
- Wold's theorem tells us that any stationary stochastic process has the representation

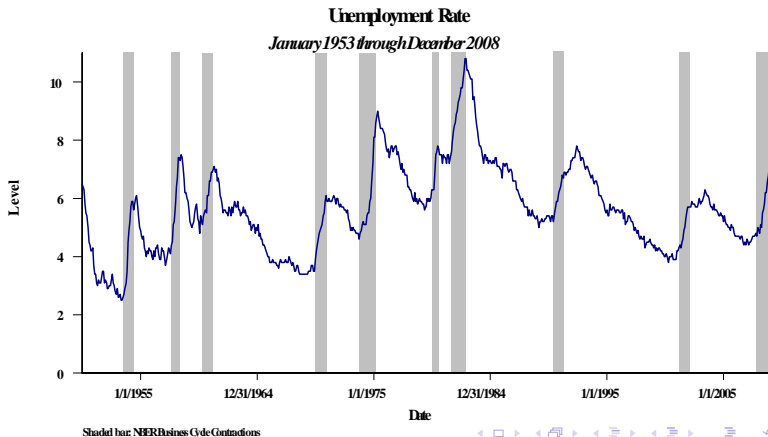
$$y_t = \delta_t + \sum_{i=0}^{\infty} w_i e_{t-i}$$

where  $\delta_t$  is deterministic and the sequence  $\{e_{t-i}\}$  has constant variance and is serially uncorrelated

- This representation does not necessarily capture all of the predictable features of the data
- That is, ARMA models are not the beginning and end of data analysis

# Does nonlinearity matter?

- Unemployment rate in U.S.



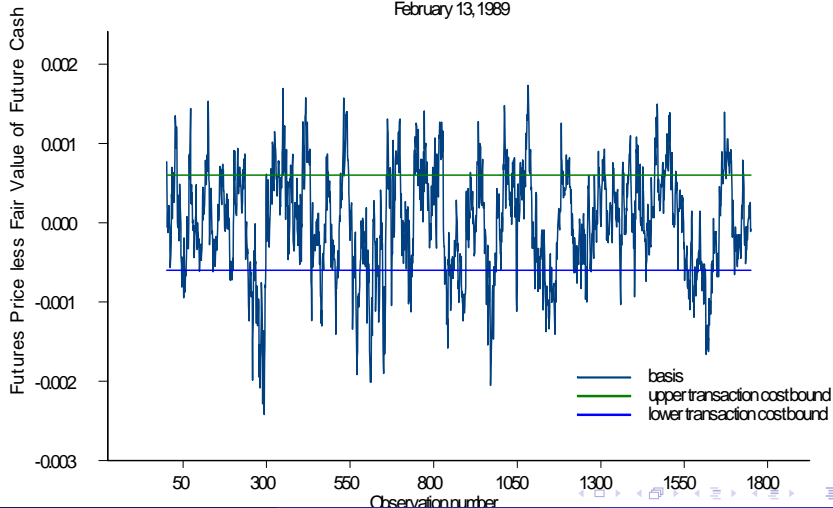
# Does nonlinearity matter for financial economics?

- Not so obvious for returns
  - High and changing volatility
  - Depends on data being examined and question being asked
  - Think of arbitrage between the cash and futures prices of some asset
    - For example, S&P 500 futures in U.S.
    - How futures and cash prices change to become equal likely to depend on how far cash is from futures
    - Arbitrage not worthwhile if there is little difference, arbitrage worthwhile if there is a large difference
    - Suggests possibly faster convergence to futures and cash prices being equal when deviations are bigger
    - Dwyer, Locke and Yu (1996)
    - “Nonlinear Time Series and Financial Applications” (2003) available at [www.jerrydwyer.com](http://www.jerrydwyer.com) summarizes some material

# Arbitrage between futures and cash values of S&P 500

Figure 4

S&P500 Futures and Cash and Rough Estimate of Transactions Costs  
February 13, 1989



# Arbitrage between futures and cash

- The logarithm of the basis is

$$b_t = {}_t f_T - p_t$$

where  ${}_t f_T$  is the logarithm of the futures price at  $t$  expiring at  $T$ ,  $p_t$  is the fair value of the cash price at  $t$  including dividends and interest

- A linear characterization might be

$$\begin{aligned} b_t &= \beta b_{t-1} + \varepsilon_t, \quad 0 < \beta < 1 \\ E \varepsilon_t &= 0, \quad E \varepsilon_t^2 = \sigma^2, \quad E \varepsilon_t \varepsilon_s = 0 \quad \forall t \neq s \end{aligned}$$

- Implied behavior when basis is nonzero is to converge at the rate  $\beta$

# Threshold autoregression for basis

- Threshold autoregression

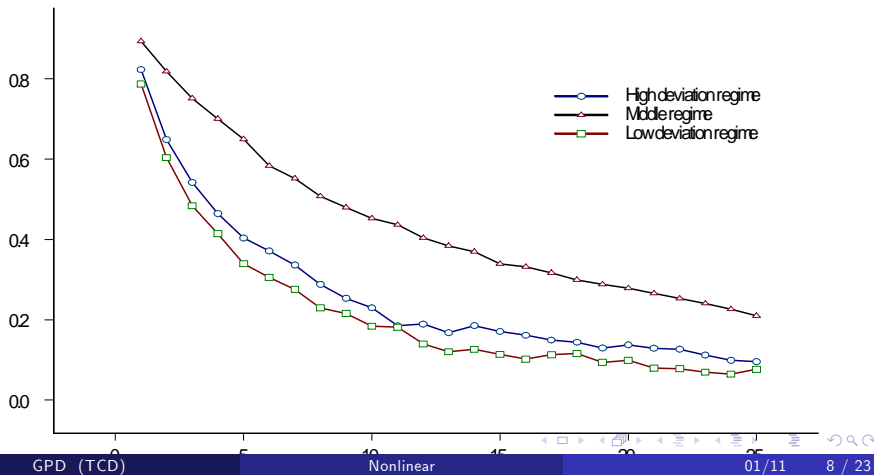
$$b_t = \beta^u b_{t-1} + \varepsilon_t \quad \text{if } c < b_{t-d}$$

$$b_t = \beta^c b_{t-1} + \varepsilon_t \quad \text{if } -c < b_{t-d} < c$$

$$b_t = \beta^l b_{t-1} + \varepsilon_t \quad \text{if } b_{t-d} < -c$$

- $\beta^u$ ,  $\beta^c$ , and  $\beta^l$  allowing for different speeds of convergence
- $c$  divides the deviations of the basis into regions based on the size of the deviation from zero
- $d$  is the delay between the time that the basis deviates from zero by some amount and the change in behavior occurs
- Estimated parameters are  $\beta^u$ ,  $\beta^c$ , and  $\beta^l$ ,  $c$ ,  $d$  and the variance of  $\varepsilon_t$
- In actual application, the equations are more complicated
  - A vector error correction mechanism in the futures and cash prices

# Median impulse response to a unit shock from futures market



# Which nonlinear model?

- There are an infinite number of possible alternative nonlinear models
- Some readable references
  - Bendat, Julius S. 1990. *Nonlinear System Analysis and Identification from Random Data*. New York: John Wiley & Sons.
  - Bendat, Julius S. 1998. *Nonlinear System Techniques and Applications*. New York: John Wiley & Sons, Inc.
  - Priestley, M. B. 1988. *Non-linear and Non-stationary Time Series Analysis*. London: Academic Press.
  - Ramsey, James B. 1990. "Economic and Financial Data as Nonlinear Processes," in *The Stock Market: Bubbles, Volatility, and Chaos*, edited by Gerald P. Dwyer, Jr. and R. W. Hafer, pp. 81-134. Boston: Kluwer Academic Publishers.
  - Tong, Howell. 1990. *Non-linear Time Series: A Dynamical Systems Approach*. Oxford: Clarendon Press.

# How to choose which nonlinear model

- Let subject matter guide the choice of type of nonlinearity
  - For example, threshold autoregression above when analyzing arbitrage
  - Obviously, you want some familiarity with different nonlinear models to make choice
- A very short selection
  - Threshold autoregression
  - Smooth transition autoregression
  - Bilinear model
  - Markov switching model

# Threshold autoregression

- Threshold autoregressions can be thought of as piecewise linear models

- If you use enough regimes, you probably can characterize almost anything reasonably well
  - That's actually not very comforting because you have to estimate the regimes
- A  $k$ -regime self-exciting threshold autoregression for regime  $j$  is

$$y_t = \varphi_0^j + \varphi_1^j y_{t-1} + \dots + \varphi_p^j y_{t-p} + \varepsilon_t^j \quad \text{if } \gamma_{j-1} \leq r_{t-d} < \gamma_j$$

where

- $j = 1, \dots, k$  is the regime
- $\varphi_i^j$  are parameters in regime  $j$
- $d$  is the delay ( $d > 0$ )
- $\{\varepsilon_t^j\}$  is an iid sequence with zero mean and variance  $\sigma_j^2$
- $\gamma_j$  are the thresholds that determine the regime
- Called “self-exciting” because values of the variable being examined ( $r_t$  here) determine the regime

# Smooth transition autoregression (STAR)

- Discontinuity across regimes not always appealing
- STAR model for two regimes

$$y_t = c_0 + \sum_{i=1}^p \phi_{0,i} y_{t-1} + F\left(\frac{y_{t-d} - \Delta}{s}\right) \left( c_1 + \sum_{i=1}^p \phi_{2,i} y_{t-1} \right) + \varepsilon_t$$

- $d$  is the delay parameter
- $\Delta$  and  $s$  are parameters representing the location and scale that affects model transition (define  $z_t = \frac{y_{t-d} - \Delta}{s}$ )
- $F()$  is a smooth transition function to determine the weight given to

$$c_1 + \sum_{i=1}^p \phi_{2,i} y_{t-1}$$

- $F()$  can be a logistic function or an exponential function or a cumulative distribution function
- $F()$  usually is bounded between zero and one
- Logistic  $F() = \frac{1}{1 + \exp(-\gamma z_t)}$
- Exponential  $F() = 1 - \exp(-z_t^2)$

- Bilinear model

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-1} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^m \sum_{j=1}^s \beta_{ij} y_{t-i} \varepsilon_{t-j} + a_t$$

- Includes ARMA terms and products of lagged values and lagged innovations
- Usually just a few

# Markov switching autoregressive model (MSA)

- Maybe used more in economics (especially macroeconomics) than finance

$$y_t = \begin{cases} c_1 + \sum_{i=1}^p \phi_{i,1} y_{t-1} + \varepsilon_{1t} & \text{if } s_t = 1 \\ c_2 + \sum_{i=1}^p \phi_{i,2} y_{t-1} + \varepsilon_{2t} & \text{if } s_t = 2 \end{cases}$$

- The innovation probabilities  $\{\varepsilon_t^1\}$  and  $\{\varepsilon_t^2\}$  are sequences of iid random variables with mean zero and finite variance independent of each other
- The state  $s_t$  is 1 or 2 for state 1 or 2, which is the usual number
- The states are determined by a first-order Markov chain with transitional probabilities

$$\Pr(s_t = 2 | s_{t-1} = 1) = w_1$$

$$\Pr(s_t = 1 | s_{t-1} = 2) = w_2$$

- $1/w_1$  and  $1/w_2$  are the expected durations of the process to stay in each state given that  $s_t$  is in that state

- Kernel regressions

$$y_t = m(y_{t-j}) + \varepsilon_t$$

- $m(y_{t-j})$  is some function of lagged values of variable in the neighborhood of similar values
- Simple example: Suppose the set of lagged values is only the first lag,  $y_{t-1}$ , and there are repeated observations on a value of  $y_{t-1}$ ,  $.01 \pm .001$
- Then  $m(y_{t-1})$  is the average  $y_{t-1}$  when  $y_{t-1} = .01 \pm .001$
- Can get more complicated and allowing for weighting but that's the basic idea

- Neural networks

- Approximating function to arbitrary functions
- Very general but not easy to disentangle into meaningful components

# Tests for nonlinearity

- Many such tests, no single best one in all circumstances
  - Power of test depends on the alternative
  - Probably best to pick tests partly based on what sort of nonlinearity is plausible
- Tests
  - There are many such tests and I mention the tests that may be most common
  - McLeod-Li
  - Bispectral test
  - BDS test
  - Ramsey RESET test
  - Time reversibility test

- McLeod-Li test is based on the Ljung-Box test applied to squared residuals

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

where  $\hat{\rho}_i^2$  is the lag  $i$  autocorrelation of the squared residuals if an ARMA(p,q) model

- Under the null hypothesis,  $Q(m) \sim^A \chi_{m-p-q}^2$
- Similarly, could do Engle regression test for conditional heteroskedasticity
- Possibly most powerful at determining whether conditional heteroskedasticity is important

# Bispectral test

- The bispectral test has a null hypothesis of linearity and normality of the errors of some specified model
- Basic idea is very simple: If a time series is linear and normal, then higher-order moments are zero
  - Use spectral analysis to examine this in detail
  - Beyond the scope of this course to go into details on it
- Test is easy to use
  - Richard Ashley and Douglas Patterson have implemented it in a convenient program available from them
  - Melvin Hinich (1982) introduced the test
- Test works well in practice

- Brock, Dechert and Scheinkman (BDS) proposed a test motivated by chaos theory
- Null hypothesis is that a series is iid
- Basic idea is very simple: If a time series is iid, then the series should be spread evenly over the space of values
  - If the series is not iid, values can cluster around each other
  - Test applies this not only to distance between one residual and other residuals but also distance between a set of values and other sets of values
  - Measures closeness by least upper bound and then counts observations that are at least within a distance  $\delta$  of each other
- Test works well in practice

- Test is due to James Ramsey (1969)
- Introduced as a general specification test for whether a regression is correctly specified
- Two regressions are used to examine the specification
  - Run a regression, say an autoregression on one lagged value  $y_{t-1}$ , and get residuals  $e_t$  and predicted values  $\hat{y}_t$
  - Run a second regression of the first regression's residuals  $e_t$  on the right-hand-side variables in the first regression  $y_{t-1}$  and on powers of the predicted values  $\hat{y}_t$ , that is,  $\hat{y}_t^2, \hat{y}_t^3, \dots$
  - If the first regression is adequate, then the coefficients of the lagged values and the powers of the predicted values should all be zero
  - If the first regression is adequate (series is a linear autoregression with normally distributed innovations), an F-statistic for the second regression has an F distribution with appropriate degrees of freedom

# Time reversibility tests

- Distribution of innovations is invariant to reversal of time indices if the series is normally distributed and the equation is correctly specified
- TR test (Ramsey and Rothman 1996)
- Motivation
  - Consider unemployment rate with gradual decreases, apparently unpredictable changes, and sharp increases in recessions
  - These sharp increases appear in the residuals if the estimated relationship does not predict sharp increases
  - These sharp increases appear as sharp decreases if the time direction is reversed
  - The distribution of the residuals is not invariant to the time direction if the estimated relationship is not adequate
- Seems to work reasonably well in practice

# Unemployment rate and time reversibility

- Unemployment rate in U.S. would have gradual increases and fast decreases if time ran backwards



- Deriving forecasts from a nonlinear model is more complicated than for a linear model
- Two issues
  - The forecasts depend on the initial conditions and the evolution generally cannot be summarized by sets of statistics
  - The existence of different states, e.g. in the threshold model or Markov switching model, increases this dependence on the precise initial state