Financial Econometrics

Nonlinear time series analysis

Gerald P. Dwyer

Trinity College, Dublin

January 2016

Outline

1. Nonlinearity
   - Does nonlinearity matter?
   - Nonlinear models
   - Tests for nonlinearity
   - Forecasting
   - Summary
   - Conclusion

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 δ_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?

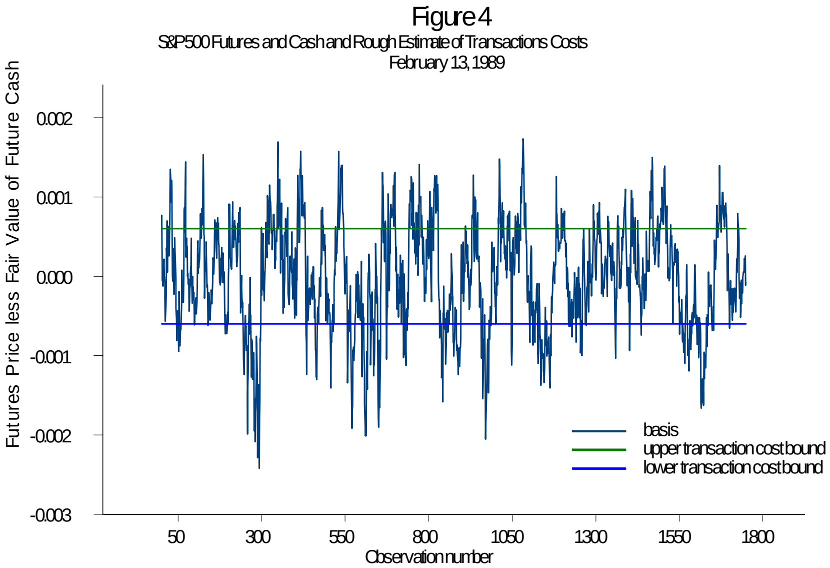
- Unemployment rate in U.S.

Sources: BLS, NBER, Haver Analytics

Does nonlinearity matter for financial economics?

- Unemployment rate is not time reversible
- Not so obvious for returns
  - High and changing volatility
  - Depends on data being examined and question being asked
  - Zumbach has forcefully argued that asset returns are not time reversible
  - In another context, think of arbitrage between the cash and futures prices of some asset
    - For example, S&P 500 futures in U.S. in the 1980s
    - How futures and cash prices change to become equal is 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 summarizes some material

Arbitrage between futures and cash values of S&P 500



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

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

  - Implied behavior when basis is nonzero is to converge at the rate β

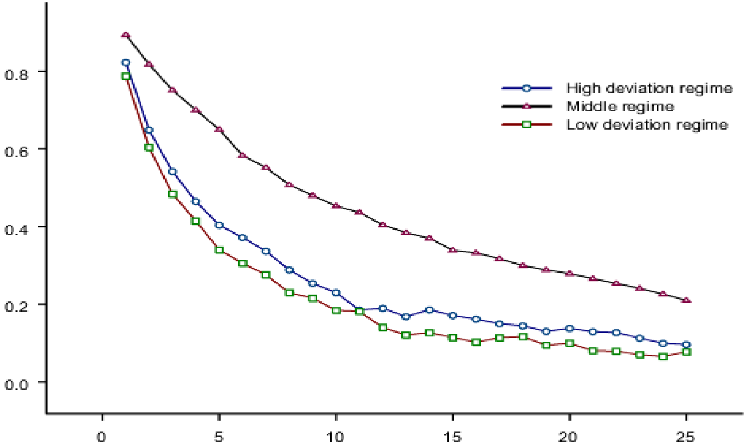
Threshold autoregression for basis

- Threshold autoregression

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

  - β^u, β^c, and β^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 β^u, β^c, and β^l, c, d and the variance of ε_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 relatively 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 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

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

    where
    - $j = 1, \ldots, k$ is the regime
    - φ_i^j are parameters in regime j
    - d is the delay ($d > 0$)
    - $\left\{ \varepsilon_t^j \right\}$ is an iid sequence with zero mean and variance σ_j^2
    - γ_j are the thresholds that determine the regime
  - Called "self-exciting" because values of the variable being examined (y_t here) determine the regime

Smooth transition autoregression (STAR)

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

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

  - d is the delay parameter
  - Δ and s are parameters representing the location and scale that affects model transition (define $z_t = \frac{y_{t-d}-\Delta}{s}$)
  - $\mathrm{F}()$ is a smooth transition function to determine the weight given to $c_1 + \sum_{i=1}^{p} \phi_{2,i} y_{t-1}$
    - $\mathrm{F}()$ can be a logistic function or an exponential function or a cumulative distribution function
    - $\mathrm{F}()$ usually is bounded between zero and one
    - Logistic $\mathrm{F}() = \frac{1}{1+\exp(-\gamma z_t)}$
    - Exponential $\mathrm{F}() = 1 - \exp\left(-z_t^2\right)$

Bilinear model

- 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

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

  - The innovation $\{\varepsilon_t^1\}$ and $\{\varepsilon_t^1\}$ 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 (usually just 2 states)
  - 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

Hamilton Markov-switching model I

- The Hamilton model – which can be called the Markov Switching Autoregressive Model – allows for serially correlated deviations between the actual value and the predicted value and effects of being in a different regime
- Let
$$\mu_t(s_t) = \beta_{s_t} y_{t-1}$$
- which implies
$$y_t = \mu_t(s_t) + \varepsilon_{t,s_t}$$
- In addition to these dynamics, suppose that
$$[1 - \sum_{i=0}^{p} \rho_i(s_t)\, \mathrm{L}^{\mathrm{i}}][y_t - \mu_t(s_t)] = \varepsilon_{t,s_t}$$

Hamilton Markov-switching model II

- For $p = 1$, this implies

$$\begin{aligned} y_t &= \mu_t(s_t) + \rho_1(s_t)[y_{t-1} - \mu_{t-1}(s_{t-1})] + \varepsilon_{t,m} \\ &= \beta_{s_t} y_{t-1} + \rho_1(s_t)[y_{t-1} - \mu_{t-1}(s_{t-1})] + \varepsilon_{t,m} \end{aligned}$$

- This specification implies that past errors in past states $y_{t-1} - \mu_{t-1}(s_{t-1})$ affect the current dynamics
- This model commonly is used to estimate the probability of two states of the economy, which might be called “recession” and “not recession”
- The model then is used to predict the probability of a recession

Other nonlinear models

- Kernel regressions
$$y_t = m\left(y_{t-j}\right) + \varepsilon_t$$
    - $m\left(y_{t-j}\right)$ 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 $m\left(y_{t-1}\right)$ 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

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

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

  where $\widehat{\rho}_i^2$ is the lag i autocorrelation of the squared residuals if an ARMA(p,q) model
- Under the null hypothesis, $Q(m) \sim^A \chi^2_{m-p-q}$
- Similarly, could do Engle regression test for conditional heteroskedasticity
- Possibly most powerful at determining whether conditional heteroskedasticity is important

Bispectal 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

BDS test

- 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 δ of each other
- Test works well in practice

RESET test

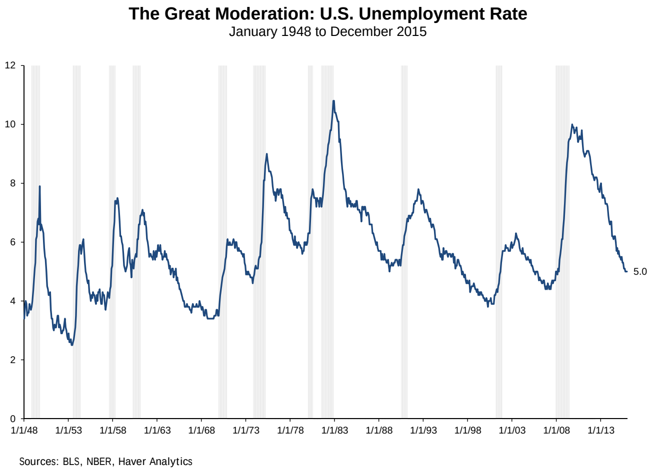
- Test is due to James Ramsey (1969)
- Introduced as a general specification test for whether a regression is correctly specified
- Two regressions are used to examine the specification
  - Run a regression, say an autoregression on one lagged value y_{t-1}, and get residuals e_t and predicted values $\widehat{y}_t$
  - 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 $\widehat{y}_t$, that is, $\widehat{y}_t^2, \widehat{y}_t^3, \ldots$
  - 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, 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 increases if time ran backwards



Forecasting with nonlinear models

- 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
- It is difficult to say what is "typical"

Summary

- While ARMA and ARCH models are informative, they are not the last word in time-series analysis
- Many time series have characteristics that are not reflected in ARMA and straightforward ARCH and GARCH models
  - Time reversibility
  - Nonlinear relationships between prices and returns
- There are an infinite number of possible nonlinear models
- Let the subject matter determine which nonlinear models are likely to be informative
- There are a variety of tests for nonlinearity
  - Let the form of the nonlinearity expected determine what test to use
  - Or just estimate the nonlinear model and examine whether it has a better fit than a linear model by a well-defined test

Conclusion

- Nonlinear models can be informative but are complicated to estimate
- More judgment involved than in estimating a linear regression
  - Judgment can be more important when nonlinearity is important