Nonlinear Time Series

A Brief Introduction

Gerald P. Dwyer

Clemson University

April 2016

Outline

1. Nonlinear Time Series
   - Introduction
   - Definition of linearity
   - Nonlinear functions
   - Which function to estimate?
   - Summary

Introduction

- Nonlinear time series is a very big topic
- Short summary introduction available on class website
- Chapter 7 in Enders is an accessible introduction

Why is nonlinearity interesting?

- Nonlinear functions are capable of representing much more varied responses of outputs to inputs
- The linear equation

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

  has quite limited behavior in response to a change in ε_t
- The nonlinear equation

$$y_t = \alpha + \beta f(y_{t-1}) + \varepsilon_t$$

  has behavior that depends on the function $f(y_{t-1})$ which may be quite complicated
- Even a deterministic equation can have quite complicated responses to initial conditions
  - "Chaos" is a possible outcome of the simple deterministic equation

$$y_t = \beta y_{t-1}(1 - y_{t-1})$$

  - The series y_t may never settle down to an equilibrium level or cycle of any finite order

Nonlinearity and chaos

- "Chaos" is a possible outcome of the simple equation

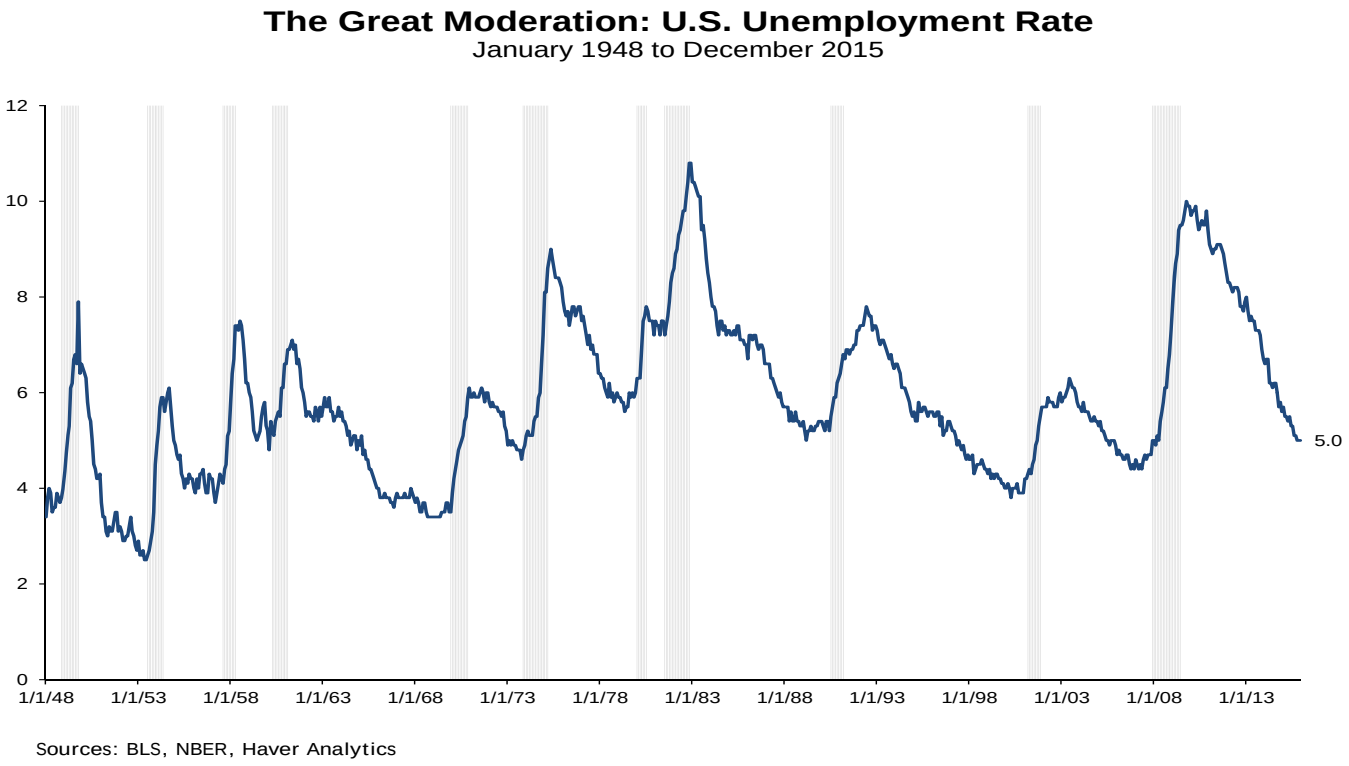
$$y_t = \beta y_{t-1}(1 - y_{t-1})$$

- Point is not that chaos in itself is interesting for economics or finance
- Point is that intuition associated with linear equations may not carry over to nonlinear equartions
- Even seemingly simple nonlinear equations can have quite complicated responses to innovations

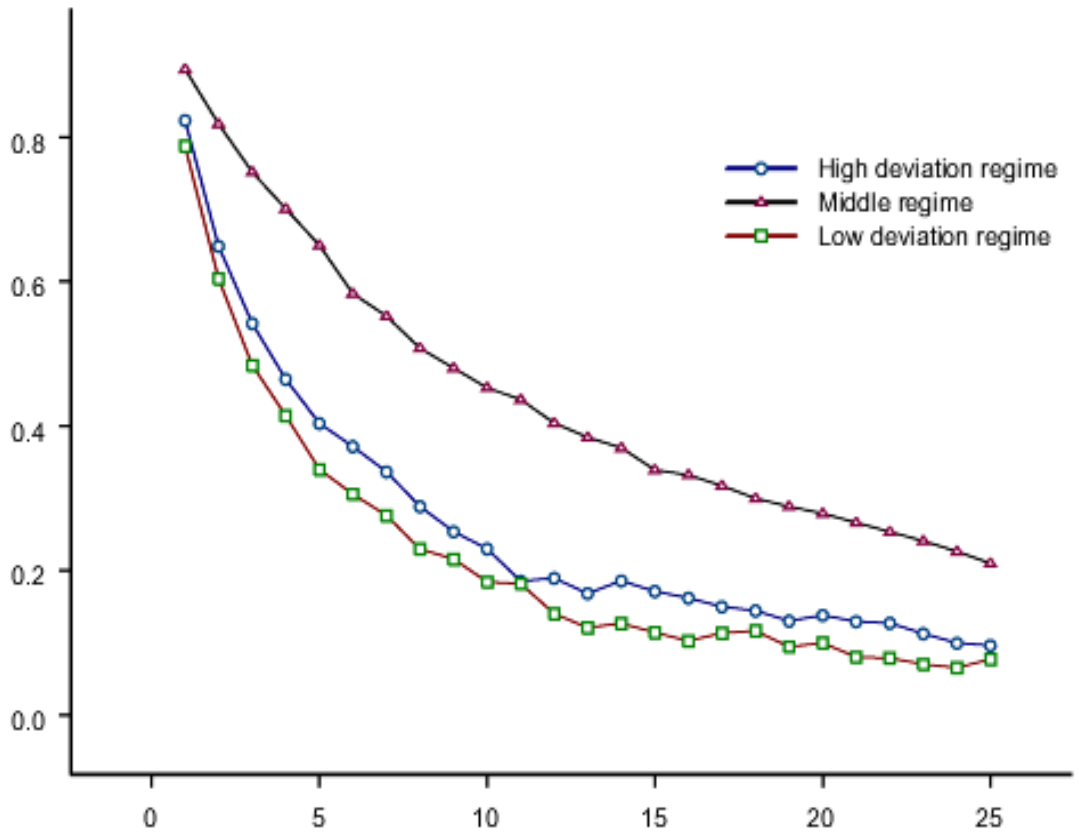
Examples of applications

- Unemployment rate has asymmetric behavior
  - Fast increases and slow decreases
- S&P 500 futures and cash
  - Sometimes arbitrage profitable and sometimes not

Unemployment rate



Impulse response functions of adjusted basis for S&P 500 to a futures shock



Nonlinear

- *Not* linear suggests that we know what *linear* means
- A natural definition of linear

$$\frac{\mathrm{d}\,y}{\mathrm{d}\,x} = \beta$$

  where β is a constant
- Equivalently

$$\frac{\mathrm{d}\,y^2}{\mathrm{d}\,x^2} = 0$$

- But write this as

$$y = \beta f(x)$$

  where $f(x)$ is a continuous function of x
- Then

$$\frac{\mathrm{d}\,y}{\mathrm{d}\,f(x)} = \beta$$

Linearity with discontinuous functions

- The observation that y is a linear function of $f(x)$ may seem like a trivial point, but it is common to note that $f(x) = ln(x)$ implies that y is a linear equation of $ln(x)$
- More generally, whatever the continuous function $f(x)$, the derivative with respect to $f(x)$ can be written as a constant
- This function may itself have other parameters of course besides β, especially if x is a vector
- Argument does not work so clearly for discontinuous functions $f(x)$
- But then the definition of linearity is not so obvious either
- Perhaps biggest problem for a statistical analysis: Not obvious what statistical approach is suggested by this sort of definition

Related statistical definition of linearity

- A definition of linearity with some appeal:
- The conditional expectation of y_t (with no deterministic component) is a linear function of conditioning variables
  - For example,
  $$y_t = bx_{t-1} + \varepsilon_t$$
  - which implies
  $$\mathrm{E}_{\text{t-1}}\, y_t = bx_{t-1}$$
  which is a linear function of x_{t-1}
- Problem with this definition is the same as the definition based on derivatives

More suggestive definition of linearity

- Recall Wold's theorem:
- There exists a representation of every covariance stationary time series as

$$y_t = d_t + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \qquad \sum_{i=0}^{\infty} \psi_i^2 < \infty$$

  where the sequence of innovations $\{\varepsilon_t\}$ is **uncorrelated**
- A linear time series is represented by the same equation as for Wold's Theorem but $\{\varepsilon_t\}$ is independently and identically distributed (iid)

Advantage of this statistical definition of linearity

- A function is linear if there exists a representation of the series $\{y_t\}$ as

$$y_t = d_t + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \qquad \sum_{i=0}^{\infty} \psi_i^2 < \infty$$

  where the sequence of innovations $\{\varepsilon_t\}$ is **iid**
- The innovations must be independent and identically distributed

What is *not* linear?

- Volterra function

$$y_t = d_t + \sum_{i=0}^{\infty} w_i \eta_{t-i} + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} w_{ij}\eta_{t-i}\eta_{t-j} + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} w_{ijk}\eta_{t-i}\eta_{t-j}\eta_{t-k}$$
$$+ \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} w_{ijkl}\eta_{t-i}\eta_{t-j}\eta_{t-k}\eta_{t-l} + \ldots.$$

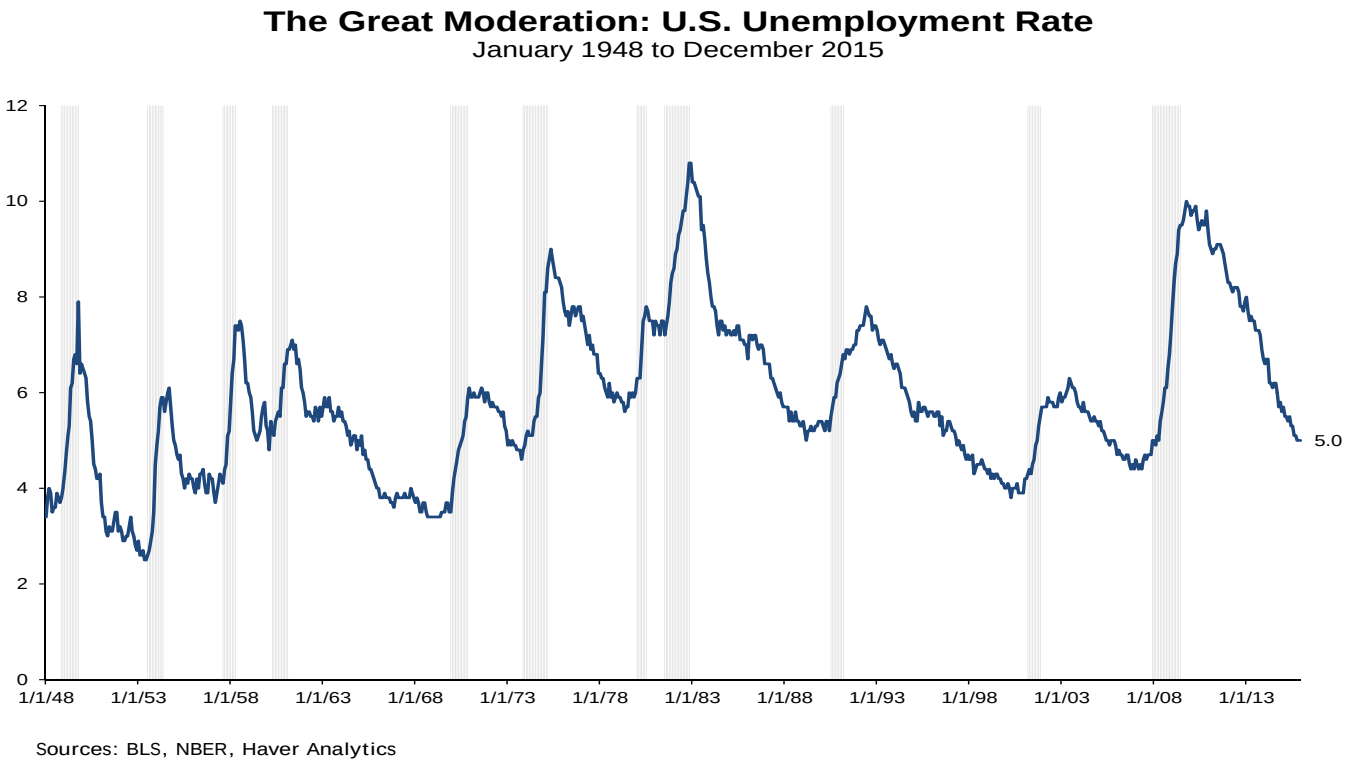
$\eta_t \backsim$ iid

- For example, a nonlinear time series might well have dependence of innovations over time in higher moments in the infinite-moving average, which is inconsistent with linearity
- Nonlinearity is all about moments higher than the second and temporal dependence in those higher-order moments

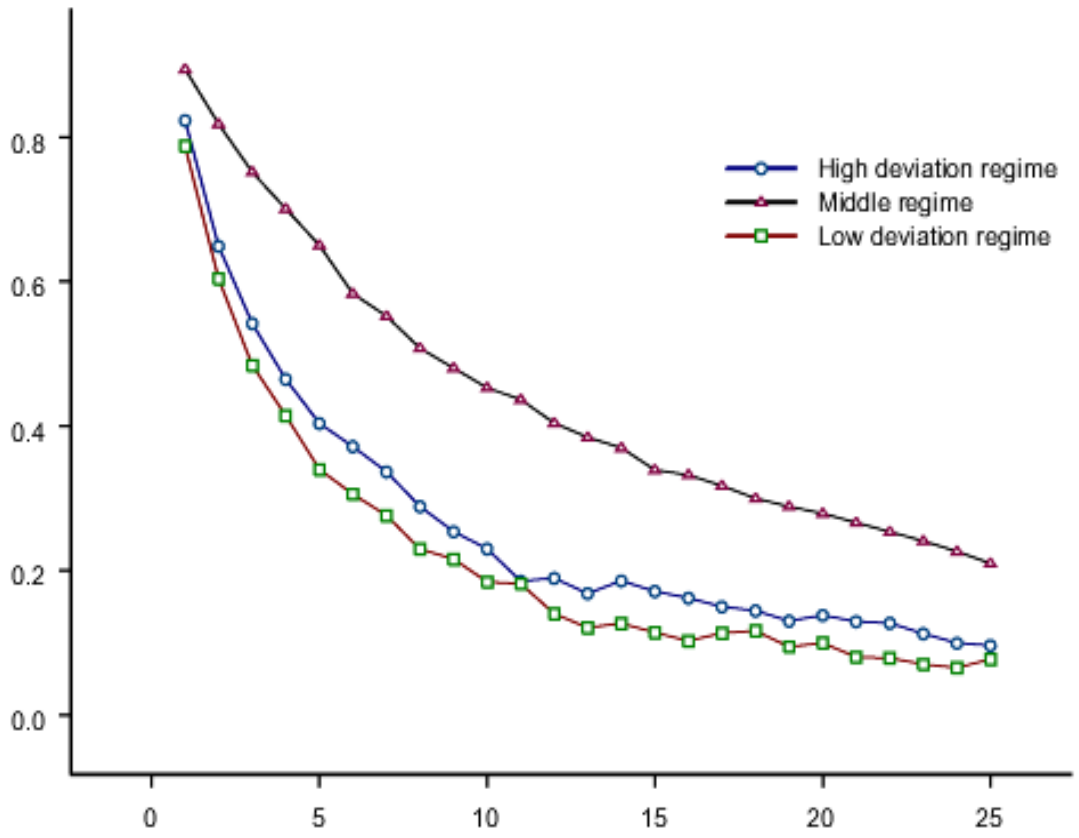
Which nonlinear function?

- There is an arbitrarily large number of nonlinear functions
- Which to choose?
- Two criteria suggested by literature
  - Statistical tests of the data for various characteristics
  - Application-specific knowledge
- Unemployment rate
- S&P 500 futures and cash

Unemployment rate



Impulse response functions of adjusted basis for S&P 500 to a futures shock



General classes of functions

- Parametric functions
- Nonparameteric functions
  - Will not discuss them
  - Not that they are less interesting, but we have only so much time and they probably are used less

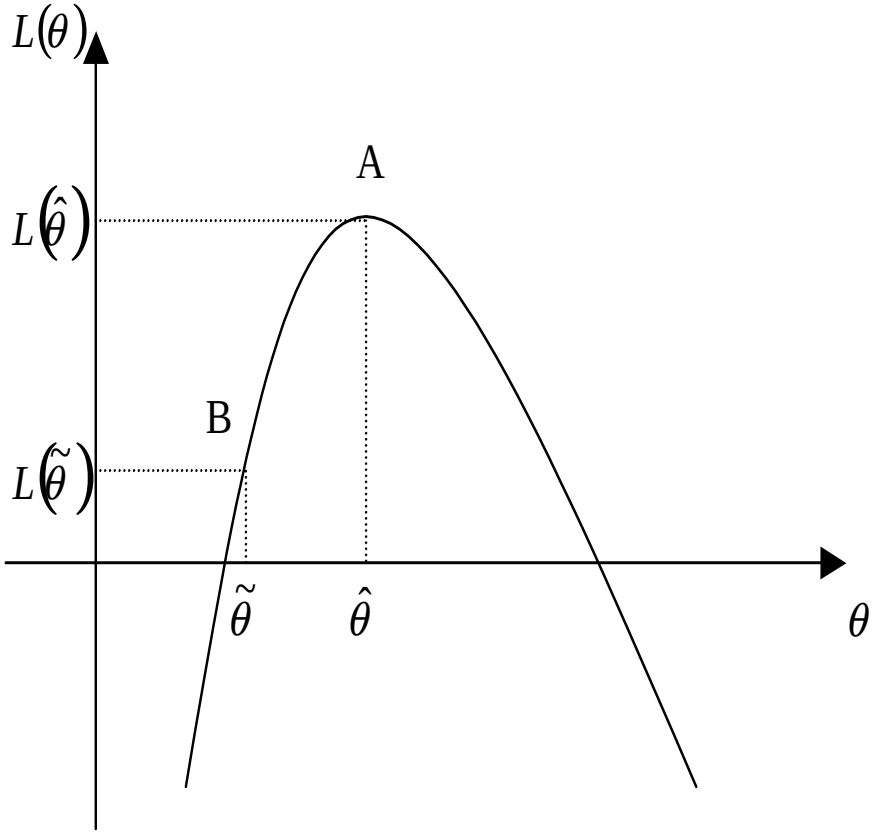
General classes of functions

- General classes of parametric functions
- Nonlinear functions
- Locally linear functions

General classes of functions

- Tests for nonlinearity
- May be useful for deciding whether the characteristic is important
  - Simple example: Test of an autoregression's residuals for ARCH
- On the other hand, it may not be that expensive to estimate the function and decide whether or not nonlinearity is important
- On the flip side, the nonlinear maximization used to estimate nonlinear models may not converge to a maximum if nonlinearity is unimportant
- Likelihood Ratio, Wald and Lagrange Multiplier tests can produce different results

Likelihood ratio tests and Wald and Lagrange multiplier tests



Examples of some nonlinear functions in the literature I

- Threshold autoregressions

$$y_t = \beta^u y_{t-1} + \varepsilon_t^u \text{ if } x_{t-d} \geq c$$
$$y_t = \beta^l y_{t-1} + \varepsilon_t^l \text{ if } x_{t-d} < c$$

  where error terms here and below are iid

- A nonlinear model suggested early in the literature is the bilinear function

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

- Smooth transition autoregression

$$y_t = \beta y_{t-1} + \gamma F(y_{t-1}) y_{t-1} + \varepsilon_t$$

  where $F(y_{t-1})$ is a continuous function of y_{t-1}

Examples of some nonlinear functions in the literature II

- Possible smooth transition functions $F(.)$ that can be used include the logistic function
$$F(y_{t-1}) = \frac{1}{1 - \exp(-y_{t-1}^2)}.$$
- Exponential
$$F(y_{t-1}) = \exp(-\delta y_{t-1}^2)$$

- ARCH models are nonlinear models on our definition
  - ARCH does not directly affect the mean function but the variance is not constant and there is temporal dependence in moments beyond the second

Markov-switching model I

- The Markov-switching model has states and there is some probability of switching from one state to another
- We will do an example with two states, $s = 1$ and $s = 2$.
- A simple version of a Markov-switching model has constant probabilities of switching between states

$$p(s_t = 1 | s_{t-1} = 1) = p_{11} \qquad p(s_t = 1 | s_{t-1} = 2) = p_{12}$$
$$p(s_t = 2 | s_{t-1} = 1) = 1 - p_{11} \qquad p(s_t = 2 | s_{t-1} = 2) = 1 - p_{12}$$

- Different regressions apply in the different regimes

$$y_t = \beta_1 y_{t-1} + \varepsilon_{t,1} \qquad \mathrm{E}\,\varepsilon_{t,1}^2 = \sigma_1^2$$
$$y_t = \beta_2 y_{t-1} + \varepsilon_{t,2} \qquad \mathrm{E}\,\varepsilon_{t,2}^2 = \sigma_2^2$$

- How estimate probabilities that are not constant?

Markov-switching model II

- For each observation, estimate the posterior probability of being in state 1 or state 2
- Use Bayes rule for state $m = 1, 2$

$$p(s_t = m | y_t, \mathrm{I}_{t-1}) = \frac{p(y_t | s_t = m, \mathrm{I}_{t-1}) p(s_t = m | \mathrm{I}_{t-1})}{p(y_t | \mathrm{I}_{t-1})}$$

  where I_{t-1} is information available through period $t - 1$
- Somewhat limiting specification because the effect of the lagged variable is independent of the state in which it occurred

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

Which function?

- Economic theory may suggest plausible behavior and then estimate appropriate function
- Possibly look at properties of data and estimate appropriate function

Test between functions

- Estimation can be nontrivial and tests between alternative nonlinear models are done rarely
- How would one do such a test?

Tests of non-nested hypotheses

- Non-nested hypotheses
- Have two alternative functions

$$y_t = f(x_t) + \varepsilon_{0t}$$
$$y_t = g(z_t) + \varepsilon_{1t}$$

- Have two hypotheses with the two equations, H_0 and H_1
- They are not nested, in which case one would be a subset of the other
  - Neither x_t nor z_t is a subset of the other
  - Neither $f(.)$ nor $g(.)$ is a special case of the other

Davidson-MacKinnon's (1981) J Test

- Non-nested hypotheses, H_0 and H_1

$$y_t = f(x_t) + \varepsilon_{0t}$$
$$y_t = g(z_t) + \varepsilon_{1t}$$

- An equation that encompasses both models and in which both are nested:

$$y_t = (1-\alpha)f(x_t) + \alpha g(z_t) + \varepsilon_{ct}$$

  where ε_{ct} is the error term from the composite equation
- Want to test $\alpha = 0$
- How do that? In general, can't just estimate simultaneously or run a regression on fitted values from both estimated separately
- α generally is not identified – just a redundant scale parameter for coefficients

Execution of Davidson-MacKinnon's (1981) J Test

- Nested equation

$$y_t = (1-\alpha)f(x_t) + \alpha g(z_t) + \varepsilon_t^c$$

- How test $\alpha = 0$?
- Estimate

$$y_t = g(z_t) + \varepsilon_t^0$$

- Now estimate

$$y_t = \bar{f}(x_t) + \alpha \widehat{g(z_t)} + \varepsilon_t^T$$

  where $\bar{f}(x_t) = f(x_t)$ under the null hypothesis that $\alpha = 0$
  $\widehat{g(z_t)}$ is the set of predicted values based on the estimate of $g(z_t)$ and ε_t^T indicates the residuals from the test equation
- Now test $\alpha = 0$ using an asymptotic normal distribution for the estimated α compared to to its standard error

Problems with J Test

- Have outlined a test of H_0 $\alpha = 0$ in

$$y_t = (1-\alpha)\bar{f}(x_t) + \alpha g(z_t) + \varepsilon_t^c$$

- But there are two nonlinear models

$$y_t = f(x_t) + \varepsilon_t^0$$
$$y_t = g(z_t) + \varepsilon_t^1$$

- What happens if I reverse the test and test whether $f(x_t)$ is redundant with $g(z_t)$ in equation?
- If $\alpha = 0$ is not rejected, it is possible that I will not be able to reject $(1-\alpha) = 0$
- It is possible that I will reject both
- The test does not provide a way of comparing the two hypotheses
- Not used a lot in empirical analysis

Bottom line for comparing nonlinear models

- It is very seldom that alternative nonlinear models are compared
- It is common to run specification tests to determine whether there are glaring divergences between the data and the estimated model

An informative way of comparing hypotheses

- Bayesian analysis makes it easy to compare two non-nested hypotheses and examine what the data say about them
- Always a consistent conclusion
- Generally will not indicate that one is informative and the other is useless
- Start from prior probabilities of the two mutually exclusive and exhaustive hypotheses, $p(\mathrm{H}_0)$ and $p(\mathrm{H}_1)$
- The posterior probability of H_0 conditional on y is $p(\mathrm{H}_0 \,|y)$
- The posterior probability of H_0 is related to the prior probability by Bayes rule,

$$p(\mathrm{H}_0 \,|y) = \frac{p(y|\,\mathrm{H}_0)p(\mathrm{H}_0)}{p(y)}$$

- Similarly, the posterior probability of H_1 is related to the prior probability by Bayes rule,

$$p(\mathrm{H}_1 \,|y) = \frac{p(y|\,\mathrm{H}_1)p(\mathrm{H}_1)}{p(y)}$$

Bayesian comparison of hypotheses

- The prior odds ratio of H_0 relative to H_1 is just

$$\frac{p(\mathrm{H}_0)}{p(\mathrm{H}_1)}$$

- The posterior probability of H_0 is

$$p(\mathrm{H}_0 \,|y) = \frac{p(y|\,\mathrm{H}_0)p(\mathrm{H}_0)}{p(y)}$$

- The posterior probability of H_1 is

$$p(\mathrm{H}_1 \,|y) = \frac{p(y|\,\mathrm{H}_1)p(\mathrm{H}_1)}{p(y)}$$

- The posterior odds ratio of H_0 relative to H_1 is

$$\frac{p(\mathrm{H}_0 \,|y)}{p(\mathrm{H}_1 \,|y)} = \frac{p(y|\,\mathrm{H}_0)}{p(y|\,\mathrm{H}_1)}\frac{p(\mathrm{H}_0)}{p(\mathrm{H}_1)}$$

Comparison of models in a Bayesian analysis

- The comparison of models is given by

$$\frac{p(\mathrm{H}_0 \,|y)}{p(\mathrm{H}_1 \,|y)} = \frac{p(y|\,\mathrm{H}_0)}{p(y|\,\mathrm{H}_1)} \frac{p(\mathrm{H}_0)}{p(\mathrm{H}_1)}$$

- The ratio

$$\frac{p(y|\,\mathrm{H}_0)}{p(y|\,\mathrm{H}_1)}$$

  is called the "Bayes Factor" and is the relative likelihood of the models
- Note: the relative likelihood is not evaluated only at the maximum
- An application would have to have a prior distribution for the underlying parameters in the functions $f(x)$ and $g(z)$ and they would be integrated out numerically

Comparison of models in a Bayesian analysis

- The comparison of models is given by

$$\frac{p(\mathrm{H}_0 \,|y)}{p(\mathrm{H}_1 \,|y)} = \frac{p(y|\,\mathrm{H}_0)}{p(y|\,\mathrm{H}_1)} \frac{p(\mathrm{H}_0)}{p(\mathrm{H}_1)}$$

- The ratio

$$\frac{p(y|\,\mathrm{H}_0)}{p(y|\,\mathrm{H}_1)}$$

  is called the "Bayes Factor" and is the relative likelihood of the models
- Note: the relative likelihood is not evaluated only at the maximum
- An application would have to have a prior distribution for the underlying parameters in the functions $f(x)$ and $g(z)$ and they would be integrated out numerically
- *This sounds really hard but there are packages for many statistical programs to do this in contexts in which such comparisons have been useful*

Summary

- Nonlinear time series analysis is useful in a wide range of contexts
- To some degree, it is mainstream econometrics
- Macroeconomics
  - Survey of econometrics in *Modelling Nonlinear Economic Time Series* by Teräsvirta, Tjøstheim and Granger (2010)
  - Any dynamic stochastic general equilibrium is highly nonlinear
  - Common to estimate the Markov-switching model to estimate the probability of a recession
- Finance
  - Summary of applications and econometrics in *Financial Econometrics* by Gourieroux and Jasiak (2001)
  - Models of volatility
  - Asset pricing diffusion models
  - Term structure of interest rates
  - Market micro-structure

Summary

- A function is linear if there exists a representation of the series $\{y_t\}$ as

$$y_t = d_t + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \qquad \sum_{i=0}^{\infty} \psi_i^2 < \infty$$

  where the sequence of innovations $\{\varepsilon_t\}$ is **iid**
- The Volterra function (or expansion) represents all functions with

$$y_t = d_t + \sum_{i=0}^{\infty} w_i \eta_{t-i} + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} w_{ij} \eta_{t-i}\eta_{t-j} + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} w_{ijk} \eta_{t-i}\eta_{t-j}\eta_{t-k}$$
$$+ \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} w_{ijkl} \eta_{t-i}\eta_{t-j}\eta_{t-k}\eta_{t-l} + \ldots.$$

  $\eta_t \backsim$ iid

- Nonlinear time-series functions reflect dependence in moments above the first two (mean, variances and covariances)

Summary

- There are many different nonlinear functions
  - Deciding what nonlinear function to fit is important for informative empirical work
  - There are no firm guidelines for picking the right function or functions
  - The best thing is choosing a model implied by economic analysis which seems consistent with interesting aspects of the data and yields interesting conclusions
- Bayesian analysis holds the most promise as a statistical way of resolving choices among models