Nonlinear Time Series

A Brief Introduction

Gerald P. Dwyer

Clemson University

April 2021

Reading

- This class
  - Tsay and Chen, Chs. 1 and 2
  - Dwyer, "Nonlinear Time Series"
- Next topic: State-space models
  - Tsay and Chen, Chs. 6 and 7

Outline

Introduction

- Nonlinear time series is a very big topic
- Short summary introduction available on class website

Why is nonlinearity interesting? I

- 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, e.g., 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

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

- 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

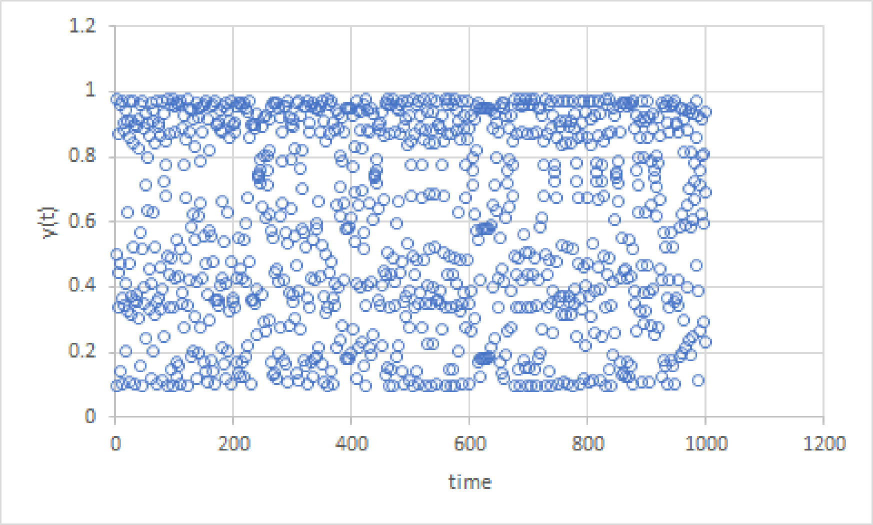
A simple nonlinear equation

- A simple nonlinear equation

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

- This simple equation can have surprisingly complex behavior

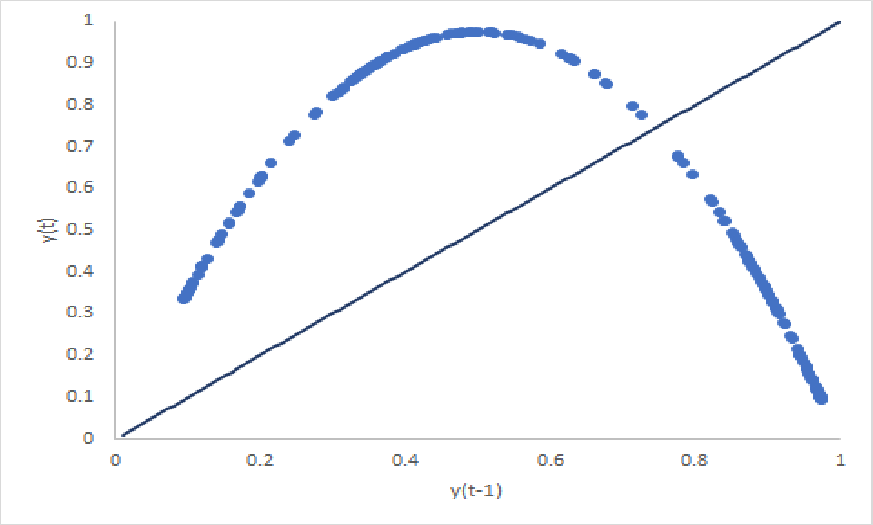
Sequence of values



Phase diagram for this simple nonlinear equation

- And this more interpretable behavior in terms of a phase diagram

Phase diagram for this simple nonlinear equation



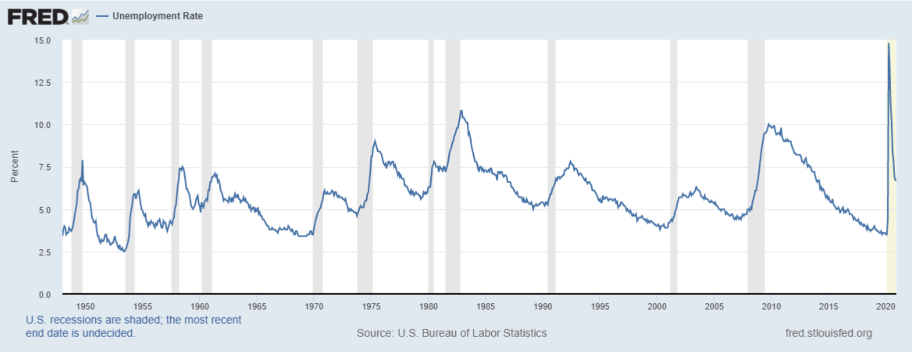
This simple nonlinear equation I

- This equation has "sensitive dependence on initial conditions"
- Subsequent values can be very different for similar initial values
- But the values do not diverge, here with $y_t \in [0, 1]$ for $\lambda \in (0, 4]$

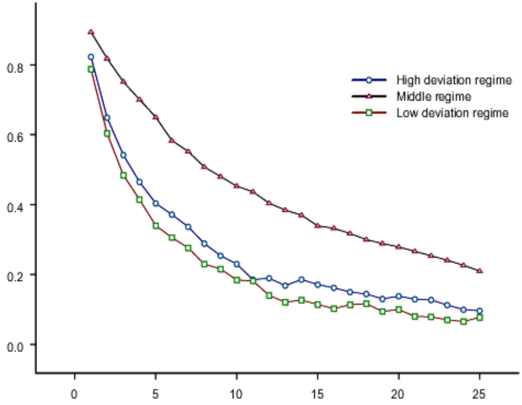
Examples of applications

- 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



Nonlinearity

- In time series, nonlinearity has gone mainstream
- Used in
  - Term struture of interest rates
  - Predicting recessions
  - Many other applications

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
  $$\mathrm{E_{t\text{-}1}}\, y_t = bx_{t-1}$$
  which is a linear function of x_{t-1}
- Problem with this definition is the same as the definition based on derivatives: Can define x to be anything

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**
- We used this to justify focusing on linear representations
  - All that matters are the mean, variance and covariances
  - We can always write this moving-average represenation as an autoregression
- 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

- Recall related but different theorem:
- A linear time series is represented by the same equation as for Wold's Theorem but $\{\varepsilon_t\}$ is independently and identically distributed (iid)
- A series 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

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

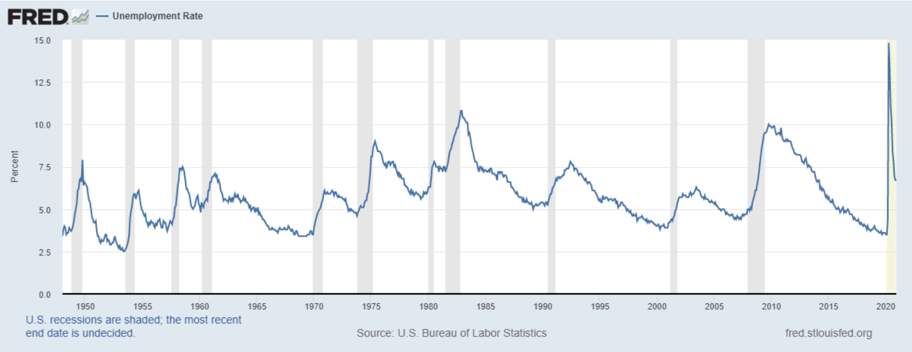
  and $w.$ is a set of coefficients

- 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 changes in moments higher than the first and temporal dependence in those higher-order moments

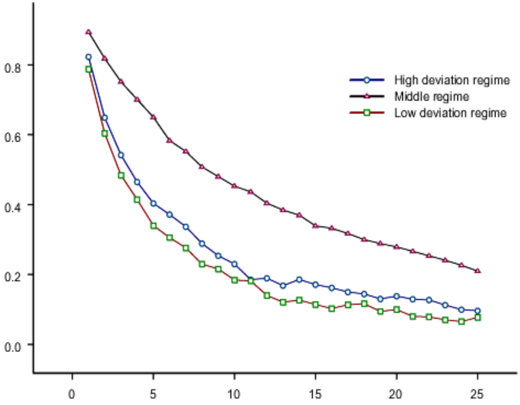
Which nonlinear function?

- 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 are used less

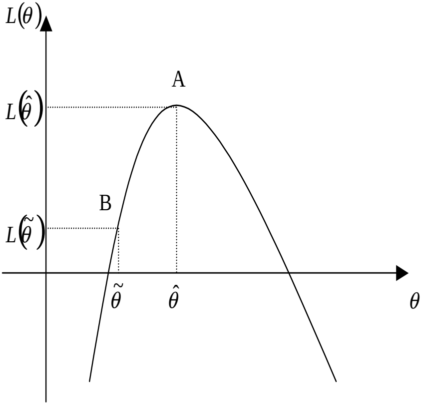
General classes of functions

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

Tests for nonlinearity

- 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
- Also, it is possible to overfit data
- Likelihood Ratio, Wald and Lagrange Multiplier tests can produce different results

Likelihood ratio tests and Wald and Lagrange multiplier tests



Tests for nonlinearity

- Two general classes of tests
  - Tests for nonlinearity in general
  - Tests for specific nonlinearities or nonlinear functions

Tests for nonlinearity in general

- BDS test
- Ramsey Reset test
- Time Reversibility test
- Hinich bispectral test

Tests for specific types of nonlinearity

- Tests for ARCH and GARCH
  - McLeod-Li can be interpreted as a test for nonlinearity in general
  - Dependence in higher moments
- Test for TAR

BDS test

- The BDS test tests whether a series is independent and identically distributed
- If the series is linearly filtered for autocorrelation, it is a test whether third and higher cross-moments are zero
- It is motivated by the correlation integral from chaos theory
- It looks at the number of sets of data points within a certain distance to see whether the probability of being within a certain distance is constant across observations
- Essentially looking for a pattern other than IID for sets of observations

Ramset RESET test I

- Ramsey's RESET test (Ramsey 1969) is a general specification test
  - RESET is REgression Specification Error Test
- An expected nonzero value of an equation's error term is produced by OLS with
  - Incorrect functional form
  - Omitted variables
  - Correlation of a right-hand-side variable and the error term
- Null hypothesis is zero mean of errors and alternative is a nonzero mean error
  - Both normally distributed
- Original equation is

$$y_t = \boldsymbol{X}_t'\boldsymbol{\beta} + \varepsilon_t$$

- An augmented equation is

$$y_t = \boldsymbol{X}_t'\boldsymbol{\beta} + \boldsymbol{Z}_t'\gamma + \varepsilon_t$$

Ramset RESET test II

- Instead of using this equation directly, use the residuals e_t from the linear regression in

$$e_t = \boldsymbol{X'}_t \boldsymbol{\alpha_1} + \boldsymbol{Z'}_t \boldsymbol{\alpha_2} + \eta_t$$

- and test whether $\boldsymbol{\alpha_1} = \boldsymbol{\alpha_2} = 0$ using a standard F-test
- Especially for tests for nonlinearity, using powers of variables in $\boldsymbol{X}'_t$ in $\boldsymbol{Z}'_t$ is an obvious way to proceed
- The number of powers of variables in $\boldsymbol{X}'_t$ would be limited by multicollinearity
- In the case of an autoregression, it can be useful to use products of lagged values in addition to squares (Tsay 1986)
    - For example, if $\boldsymbol{X}'_t = [y_{t-1}, y_{t-2}]$, then $\boldsymbol{Z}'_t = [y_{t-1}^2, y_{t-1}y_{t-2}, y_{t-2}^2]$

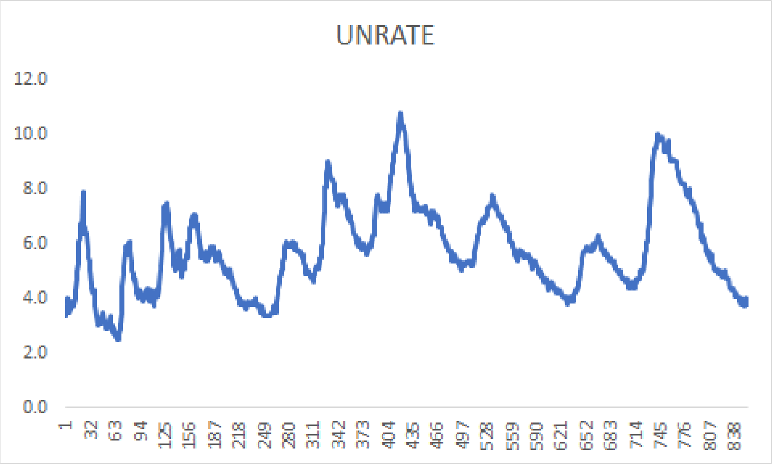
Tests for Threshold Autoregression

- Tests whether a threshold autoregression fits better than a linear equation
- Tsay discusses the issues

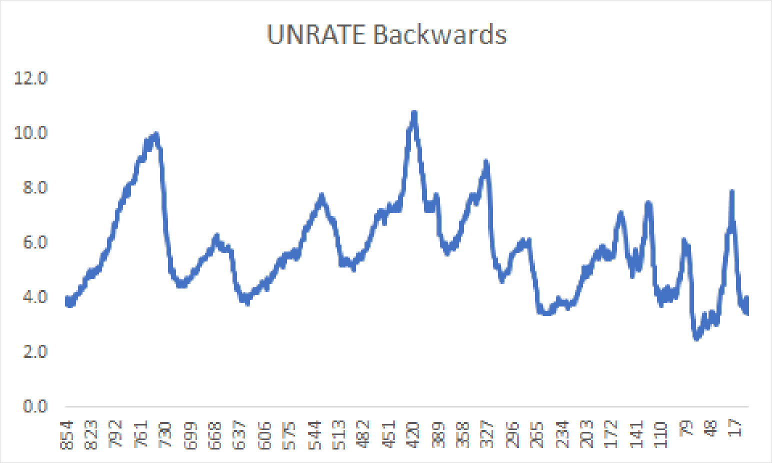
Time reversibility test

- Linear time series with normally distributed innovations are time reversible
- A time reversible time series is a series $\{y_t\}$ that is the same whether time runs forwards or backwards
- A series that appears to be time *irreversible*: the U.S. unemployment rate

U.S. unemployment rate forward



U.S. unemployment rate backward



How test time reversibility?

- How test time reversibility?
- Use moments above the second (above ACF) and test whether these moments are zero
- Most common tests
  - Ramsey and Rothman (1996) – TR test
  - Ashley, Hinich and Patterson (1986 and later)– bispectral test

Examples of some nonlinear functions in the literature I

- Threshold autoregressions

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

  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 first

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

$$
\begin{aligned}
p(s_t = 1 | s_{t-1} = 1) &= p_{11} & p(s_t = 1 | s_{t-1} = 2) &= p_{12} \\
p(s_t = 2 | s_{t-1} = 1) &= 1 - p_{11} & p(s_t = 2 | s_{t-1} = 2) &= 1 - p_{12}
\end{aligned}
$$

- Different regressions apply in states 1 and 2

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

- If probabilities are not constant, how can they be estimated?
- For each observation, estimate the posterior probability of being in state 1 or state 2

Markov-switching model II

- 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 hypotheses, which may be 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 (Markov Chain Monte Carlo)

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 (Markov Chain Monte Carlo)
- *This sounds difficult but there are packages for many statistical programs to do this in contexts in which such comparisons have been useful*

Bayesian model averaging I

- An alternative is to suppose that both models are important and allow for that
- "Bayesian model averaging" is a Bayesian way to average over models
- Rather than using H_0 and H_1 which is suggestive of hypothesis tests, I use M_1 and M_2 for models 1 and 2
- We will take models 1 and 2 to be mutually exclusive and exhaustive
- We are assessing the relative importance of the two models but not necessarily ruling out a combination of the two
- The set of all parameters in both models is ϕ
- Some parameters might be restricted in the two models

$$p(\phi) = p(\phi|y, M_1)p(M_1|y) + p(\phi|y, M_2)p(M_2|y)$$

- where $p(M_1|y)$ is the probability of model 1 after observing some data y,
- $p(M_2|y)$ is the probability of model 2 after observing some data y,

Bayesian model averaging II

- $p(\phi|y, M_1)$ is the probability distribution (pdf) of the parameters ϕ conditional on the data and model 1
- $p(\phi|y, M_2)$ is the pdf of the parameters ϕ conditional on the data and model 2
- ϕ is the set of all parameter values in the models
- $p(\phi)$ is the posterior pdf of all parameter values in the models
- Models 1 and 2 might be regression equations including some of the same variables and some different variables
- Models 1 and 2 might two nonlinear functions including possibly some of the same variables and possibly some different variables
- Bayesian model averaging has been used to assess the importance of various theories and variables in growth regressions
- Koop in Chapter 11 of *Bayesian Econometrics* has a very nice discussion of this
    - The chapter is available on Canvas

Summary

- Nonlinear time series analysis is useful in a wide range of contexts
- To some extent, 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 variation in second and higher moments

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 or combining them