“Developments in Statistical Computing for Time Series Analysis” by Ian McLeod

Department of Statistical and Actuarial Sciences

November 13, 2012. Systems Design Engineering, University of Waterloo

http://www.stats.uwo.ca/faculty/aim/2012/UWSeminar/

Mathematica slideshow also available as PDF and CDF (Computable Document Format)

QPE and Statistical Software

- QPE
  - free: R, python
  - $$$: *Mathematica*, MatLab, etc
- Menu style
  - free: MHTS, JMulti, Weka
  - $$$: SAS, SPSS, etc

MHTS Package

ARIMA(p, d, q) models are a type of linear time series model,

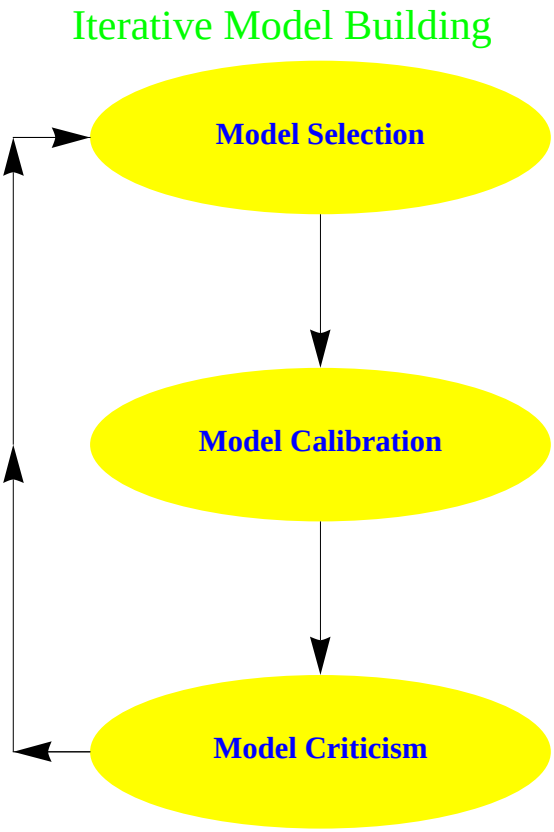
$$z_t - \phi_1 z_{t-1} - \ldots - \phi_p z_{t-p} = \zeta + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q},\ \ a_t \sim \text{IID}\left(0, \sigma_a^2\right)$$

Extensions: regression, transfer-function-noise, seasonal models, intervention analysis
estimation, diagnostic checking, forecasting and simulation, Granger causality, trend analysis, and spectral analysis.

Time Series Modelling of Water Resources and Environmental Systems. Keith Hipel & Ian McLeod

McLeod-Hipel Time Series Package

Box-Jenkins Iterative Model Building



R Programming Language

CRAN: **http://cran.r-project.org/**

http://probability.ca/cran/

Although a relatively new QPE, built on solid foundations. Generally very high quality. Elementary to advanced statistics.

Rich ecosystems: refereed journal (jss), reproducible research, books, blogs, consulting

Interface with SPSS, Excel, *Mathematica* (Version 9), Weka, C, Java, Fortran, etc.

Built-in support for parallel and cluster computing. 64 bit versions.

Time Series Analysis with R

"Time Series Analysis with R" (2012), McLeod, Yu and Krougly

`http://www.stats.uwo.ca/faculty/aim/tsar/default.htm`

- **Time Series Analysis Textbooks with R Packages**
  - Venables and Ripley, 2002
  - Lutkepohl and Kratzig, 2004
  - Cryer and Chan, 2008
  - Chan, 2010
  - Tsay, 2010
  - Shumway and Stoffer, 2011

Time Series Plots

- lynx series, Beveridge wheat prices, CO_2
- IBM returns
- annual temperatures
- Airline Passengers
- Creatine Clearances

VAR and SVAR Models

- often environmental and economic time series are multivariate
- y_t is a k-dimensional time series
- $y_t = \delta\, d_t + \Phi_1\, y_{t-1} + \ldots + \Phi_p\, y_{t-p} + e_t$ (VAR)
- $A\, y_t = \delta\, d_t + \Phi_1\, y_{t-1} + \ldots + \Phi_p\, y_{t-p} + B\, e_t$ (SVAR)
- **Mahdi & McLeod, Improved Multivariate Portmanteau Test, JTSA, March 2012.**

Kalman Filter

- includes ARMA, multivariate ARMA and many other models
- estimation, optimal smoothing, interpolation and forecasting

Nile riverflow example: random walk with additive noise:

- $y_t = \theta_t + v_t,\ \ v_t \sim \text{NID}(0,\ \ V)$
- $\theta_t = \theta_{t-1} + w_t,\ \ w_t \sim \text{NID}(0,\ \ W)$

Unit Root Testing: The Null Hypothesis

Consider the model $z_t = c + \phi\, z_{t-1} + a_t,\ \ a_t \sim \text{NID}\left(0,\ \sigma_a^2\right).\ \ \mathrm{H}_0 : \phi = 1$ vs $\mathrm{H}_a : \phi < 1$.

If $\nabla^d z_t$ is stationary for $d = 1, 2, \ldots$ it is called difference-stationary.

More general H_0 states that z_t is difference-stationary and alternate is that z_t is stationary.

Large research literature including research monographs.

Advanced probability theory used to derive asymptotic distributions but practical answer provided by using Response Surface Regression methods using large-scale simulations to obtain more exact p-values.

Model Building Approach to Unit Roots

$$\nabla z_t = c_t + \tau z_{t-1} + \sum_{i=1}^{p} \phi_i \nabla z_{t-i} + a_t$$

$$c_t = \begin{cases} 0 & \text{none} \\ \beta_0 & \text{drift} \\ \beta_0 + \beta_1 t & \text{trend} \end{cases}$$

$H_0 : \tau = 0$ and the alternative is $H_a : \tau < 0$.

Co-integration

http://nobelprize.org/nobel_prizes/economics/laureates/2003/granger-lecture.html

Looks for deep relationships between econometric/financial series that are assumed to be difference-stationary.

x_t and y_t are co-integrated if for some $\beta \neq 0$, $y_t - \beta x_t$ is stationary.

- consumption and income
- wages and prices
- short and long term interest rates

VECH Model and Test for Co-integration

y_t is k-dimensional time series, $t = 1, 2, \ldots$ then VEC (vector-error-correction) model,

$\nabla y_t = \Pi y_t + \Gamma_1 \nabla y_{t-1} + \ldots + \Gamma_p \nabla y_{t-p} + a_t$, $a_t \sim \text{NID}(0, \Sigma_a)$

co-integration exists if $0 < \text{rank}\,\Pi < k$

model building procedure used to construct suitable VEC model

Nonlinear time series

- generalized linear models and generalized additive models: **Car Fatalities IA**
- multi-adaptive regression splines (MARS)
- neural nets
- support vector machines

GARCH/ARMA

GARCH(r, s) provides a model for the innovations, $a_t \sim \text{NID}\left(0,\ \sigma_t^2\right)$

$\sigma_t^2 = \alpha_0 + \alpha_1\, a_{t-1}^2 + \ldots + \alpha_r\, a_{t-r}^2 + \beta_1\, \sigma_{t-1}^2 + \ldots + \beta_s\, \sigma_{t-s}^2$

Theorem: $a_t \sim \text{GARCH}(r, s) \iff a_t^2 \sim \text{ARMA}(r, s)$.

Simulation of GARCH(1,1) with $\alpha_0 = 10^{-6}$, $\alpha_1 = 0.2$, $\beta_1 = 0.7$ with GED error distribution with skewness coefficient 1.25 and shape parameter 4.8. **R-Script for Simulated GARCH.**

Fitting ARMA/GARCH model to U.S. GNP deflator for 1947 to 2010 (quarterly). Let $r_t = \log(z_t) - \log(z_{t-1})$ then the final model is given by

$r_1 = 0.103 + 0.369\, r_{t-1} + 0.223\, r_{t-2} + 0.248\, r_{t-3} + \epsilon_t$

$\sigma_t^2 = 0.004 + 0.269\, \epsilon_{t-1}^2 + 0.396\, r_{t-1} + 0.223\, r_{t-2} + 0.248\, r_{t-3}$

R-Script for GNP Deflator Example.

Threshold Autoregression (TAR)

Tsay (2010) fits a two-regime TAR model to the monthly U.S. unemployment data, z_t, January 1948 to March 2009 ($n = 735$).

$w_t = 0.983\, w_{t-2} + 0.158\, w_{t-3} + 0.0118\, w_{t-4} - 0.180\, w_{t-12} + a_{1,t}$, if $w_{t-1} \le 0.01$

$w_t = 0.421\, w_{t-2} + 0.239\, w_{t-3} - 0.127\, w_{t-12} + a_{2,t}$, if $w_{t-1} > 0.01$

where $w_t = \nabla z_t$. This model fits better than SARIMA. **See R graphics plot.**

Wavelet Methods for Time Series

Best books are Percival & Walden (2000), *Wavelet Methods for Time Series Analysis* and an applied version of this book by Gencay et al. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Several packages available in R. I have implemented WMTSA in *Mathematica* for my research and teaching.

- Wavelet variance, denoising, smoothing, multi-resolution analysis.
- Recent research uses wavelet methods for forecasting riverflow.
- **R-Script:** denoised mean Annual Nile Riverflow

Stochastic Differential Equations

- SDE's used in financial mathematics and many other areas
- in math finance simulation of SDE's used to get answers
- recent book by Iacus (2009) and R package provides estimation and simulation
- $d\,x(t) = \left(5 - 11\,x(t) + 6\,x(t)^2 - x(t)^3\right) d\!t + W(t)$
- simulation using Shoji-Ozaki algorithm: **R-Script**

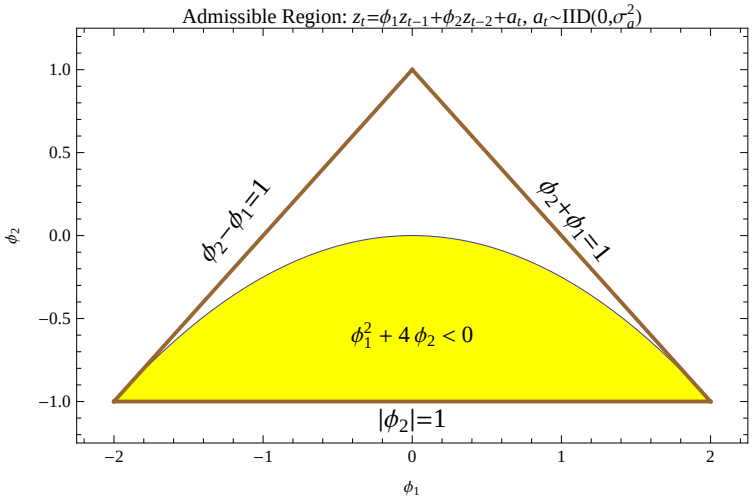
Mathematica

- high level advanced programming language
- superb graphics, numerics, symbolics, technical typesetting
- curated datasets, interface to Wolfram|Alpha, C and R

Mathematica Demonstrations

- **ARIMA Explorer**
- **Mathematica Demonstrations Project**

Visualizing the Admissible Region



AR(4) is a more complex 4-dimensional region.

Symbolic Computing and Time Series

- Derivation of asymptotic bias in Yule-Walker, LS and Burg
- Derivation of asymptotic distribution of a new unit root test
- **Derivation of closed form expression of the SDF of FNG**

WolframAlpha Pro

- The free version is amazing:
  - replaces need for mathematical tables and integrals
  - does many symbolic computations
  - natural language processing interface to the world’s computational knowledge
  - available on mobile devices: Android, iOS
- Pro version has many more amazing features:
  - can output data in csv format
  - input data for automated statistical analysis

Sonification

"Listening to uncertainty: Information that sings", Ethan Brown & Nick Bearman (2012). *Significance*, October 2012.