

Modelling Financial Time Series with S-PLUS

Eric Zivot and Jiahui Wang

May 21, 2002

Contents

References	1
1 S and S-PLUS	3
1.1 Introduction	3
1.2 S Objects	4
1.2.1 Assignment	4
1.2.2 Class	5
1.2.3 Method	8
1.3 Modeling Functions in S+FinMetrics	10
1.3.1 Formula Specification	10
1.3.2 Method	13
References	15
2 Time Series Specification, Manipulation and Visualization in S-PLUS	17
2.1 Introduction	17
2.2 The Specification of “ timeSeries ” Objects in S-PLUS	17
2.2.1 Basic Manipulations	20
2.2.2 S-PLUS “ timeDate ” Objects	21
2.2.3 Creating Common “ timeDate ” Sequences	26
2.2.4 Miscellaneous Time and Date Functions	30
2.2.5 Creating “ timeSeries ” Objects	31
2.2.6 Aggregating and disaggregating time series	32

2.2.7	Merging Time Series	40
2.2.8	Dealing with Missing Values using the S+FinMetrics Function interpNA	41
2.3	Time Series Manipulation in S-PLUS	42
2.3.1	Creating lags and differences	42
2.3.2	Return Definitions	45
2.3.3	Computing Asset Returns Using the S+FinMetrics Function getReturns	48
2.4	Visualizing Time Series in S-PLUS	50
2.4.1	Plotting “ timeSeries ” Using the S-PLUS Generic plot function	50
2.4.2	Plotting “ timeSeries ” Using the S+FinMetrics Trel- lis Plotting Functions	53
	References	57
3	Time Series Concepts	59
3.1	Introduction	59
3.2	Univariate Time Series	60
3.2.1	Stationary and Ergodic Time Series	60
3.2.2	Linear Processes and ARMA Models	66
3.2.3	Autoregressive Models	68
3.2.4	Moving Average Models	72
3.2.5	ARMA(p,q) Models	76
3.2.6	Estimation of ARMA Models and Forecasting	78
3.2.7	Martingales and Martingale Difference Sequences	85
3.2.8	Long-run Variance	87
3.3	Univariate Nonstationary Time Series	90
3.4	Long Memory Time Series	94
3.5	Multivariate Time Series	97
3.5.1	Stationary and Ergodic Multivariate Time Series	98
3.5.2	Multivariate Wold Representation	102
3.5.3	Long Run Variance	103
	References	107
4	Unit Root Tests	109
4.1	Introduction	109
4.2	Testing for Nonstationarity and Stationarity	110
4.3	Autoregressive Unit Root Tests	111
4.3.1	Simulating the DF and Normalized Bias Distributions	113
4.3.2	Trend Cases	115
4.3.3	Dickey-Fuller Unit Root Tests	118
4.3.4	Phillips-Perron Unit Root Tests	124
4.3.5	Some Problems with Unit Root Tests	126

4.4	Stationarity Tests	127
4.4.1	Simulating the KPSS Distributions	128
4.4.2	Testing for Stationarity Using the S+FinMetrics Function <code>stationaryTest</code>	129
	References	131
5	Modeling Extreme Values	133
5.1	Introduction	133
5.2	Modeling Maxima and Worst Cases	134
5.2.1	The Fisher-Tippet Theorem and the Generalized Extreme Value Distribution	135
5.2.2	Estimation of the GEV distribution	139
5.2.3	Return Level	145
5.3	Modeling Extremes Over High Thresholds	149
5.3.1	The Limiting Distribution of Extremes Over High Thresholds and the Generalized Pareto Distribution	150
5.3.2	Estimating the GPD by Maximum Likelihood	155
5.3.3	Estimating the Tails of the Loss Distribution	156
5.3.4	Risk Measures	160
5.4	Hill's Non-parametric Estimator of Tail Index	163
5.4.1	Hill Tail and Quantile Estimation	165
	References	169
6	Time Series Regression Modeling	171
6.1	Introduction	171
6.2	Time Series Regression Model	172
6.2.1	Least Squares Estimation	173
6.2.2	Goodness of Fit	173
6.2.3	Hypothesis Testing	174
6.2.4	Residual Diagnostics	175
6.3	Time Series Regression Using the S+FinMetrics Function <code>OLS</code>	175
6.4	Dynamic Regression	190
6.4.1	Distributed Lags and Polynomial Distributed Lags	195
6.4.2	Polynomial Distributed Lag Models	196
6.5	Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation	198
6.5.1	The Eicker-White Heteroskedasticity Consistent (HC) Covariance Matrix Estimate	198
6.5.2	Testing for heteroskedasticity	200
6.5.3	The Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) Covariance Matrix Estimate	203
6.6	Recursive Least Squares Estimation	207
6.6.1	CUSUM and CUSUMSQ Tests for Parameter Stability	207

6.6.2	Computing Recursive Least Squares Estimates Using the <code>S+FinMetrics</code> function RLS	208
	References	213
7	Univariate GARCH Modeling	215
7.1	Introduction	215
7.2	The Basic ARCH Model	216
7.2.1	Testing for ARCH Effects	220
7.3	The GARCH Model and Its Properties	221
7.3.1	ARMA Representation of GARCH Model	222
7.3.2	GARCH Model and Stylized Facts	222
7.4	GARCH Modeling Using <code>S+FinMetrics</code>	224
7.4.1	GARCH Model Estimation	224
7.4.2	GARCH Model Diagnostics	227
7.5	GARCH Model Extensions	232
7.5.1	Asymmetric Leverage Effects and News Impact	233
7.5.2	Two Components Model	239
7.5.3	GARCH-in-the-Mean Model	242
7.5.4	ARMA Terms and Exogenous Variables in Condi- tional Mean Equation	244
7.5.5	Exogenous Explanatory Variables in the Conditional Variance Equation	247
7.5.6	Non-Gaussian Error Distributions	248
7.6	GARCH Model Selection and Comparison	251
7.6.1	Constrained GARCH Estimation	253
7.7	GARCH Model Prediction	253
7.8	GARCH Model Simulation	256
7.9	Conclusion	258
	References	261
8	Long Memory Time Series Modeling	265
8.1	Introduction	265
8.2	Long Memory Time Series	266
8.3	Statistical Tests for Long Memory	270
8.3.1	R/S Statistic	270
8.3.2	GPH Test	272
8.4	Estimation of Long Memory Parameter	274
8.4.1	R/S Analysis	274
8.4.2	Periodogram Method	276
8.4.3	Whittle's Method	277
8.5	Estimation of FARIMA and SEMIFAR Models	278
8.5.1	Fractional ARIMA Models	279
8.5.2	SEMIFAR Model	287

8.6	Long Memory GARCH Models	290
8.6.1	FIGARCH and FIEGARCH Models	290
8.6.2	Estimation of Long Memory GARCH Models	292
8.6.3	Custom Estimation of Long Memory GARCH Models	295
8.7	Prediction from Long Memory Models	298
8.7.1	Prediction from FARIMA/SEMIFAR Models	299
8.7.2	Prediction from FIGARCH/FIEGARCH Models	302
References		305
9	Rolling Analysis of Time Series	309
9.1	Introduction	309
9.2	Rolling Descriptive Statistics	310
9.2.1	Univariate Statistics	310
9.2.2	Bivariate Statistics	317
9.2.3	Exponentially Weighted Moving Averages	319
9.2.4	Moving Average Methods for Irregularly Spaced High Frequency Data	323
9.2.5	Rolling Analysis of Miscellaneous Functions	330
9.3	Technical Analysis Indicators	333
9.3.1	Price Indicators	334
9.3.2	Momentum Indicators and Oscillators	334
9.3.3	Volatility Indicators	336
9.3.4	Volume Indicators	337
9.4	Rolling Regression	338
9.4.1	Estimating Rolling Regressions Using the S+FinMetrics Function rollOLS	339
9.4.2	Rolling Predictions and Backtesting	345
9.5	Rolling Analysis of General Models Using the FinMetrics Function roll	354
References		357
10	Systems of Regression Equations	359
10.1	Introduction	359
10.2	Systems of Regression Equations	360
10.3	Linear Seeming Unrelated Regressions	362
10.3.1	Estimation	362
10.3.2	Analysis of SUR Models with the S+FinMetrics Func- tion SUR	365
10.4	Nonlinear Seemingly Unrelated Regression Models	372
10.4.1	Analysis of Nonlinear SUR Models with the S+FinMetrics Function NLSUR	373
References		381

11	Vector Autoregressive Models for Multivariate Time Series	383
11.1	Introduction	383
11.2	The Stationary Vector Autoregression Model	384
11.2.1	Estimation	386
11.2.2	Inference on Coefficients	388
11.2.3	Lag Length Selection	388
11.2.4	Estimating VAR Models Using the S+FinMetrics Function VAR	388
11.3	Forecasting	396
11.3.1	Traditional Forecasting Algorithm	396
11.3.2	Simulation-Based Forecasting	400
11.4	Structural Analysis	404
11.4.1	Granger Causality	405
11.4.2	Impulse Response Functions	407
11.4.3	Forecast Error Variance Decompositions	411
11.5	An Extended Example	415
11.6	Bayesian Vector Autoregression	422
11.6.1	An Example of a Bayesian VAR Model	422
11.6.2	Conditional Forecasts	425
	References	427
12	Cointegration	429
12.1	Introduction	429
12.2	Spurious Regression and Cointegration	430
12.2.1	Spurious Regression	430
12.2.2	Cointegration	433
12.2.3	Cointegration and Common Trends	435
12.2.4	Simulating Cointegrated Systems	435
12.2.5	Cointegration and Error Correction Models	440
12.3	Residual-Based Tests for Cointegration	441
12.3.1	Testing for Cointegration when the Cointegrating Vec- tor is Pre-specified	442
12.3.2	Testing for Cointegration when the Cointegrating Vec- tor is Estimated	444
12.4	Regression-Based Estimates of Cointegrating Vectors and Error Correction Models	448
12.4.1	Least Square Estimator	448
12.4.2	Stock and Watson's Efficient Lead/Lag Estimator	448
12.4.3	Estimating Error Correction Models by Least Squares	452
12.5	VAR models and cointegration	453
12.5.1	The Cointegrated VAR	453
12.5.2	Johansen's Methodology for Modeling Cointegration	456
12.5.3	Specification of Deterministic Terms	456

12.5.4	Likelihood Ratio Tests for the Number of Cointegrating Vectors	459
12.5.5	Testing for the Number of Cointegrating Vectors Using the <code>S+FinMetrics</code> Function <code>coint</code>	460
12.5.6	Maximum Likelihood Estimation of the Cointegrated VECM	462
12.5.7	Maximum Likelihood Estimation of the Cointegrated VECM Using the <code>S+FinMetrics</code> Function <code>VECM</code>	463
12.5.8	Forecasting from the VECM	467
12.6	Appendix: Maximum Likelihood Estimation of a Cointegrated VECM	470
	References	473
13	Multivariate GARCH Modeling	475
13.1	Introduction	475
13.2	Exponentially Weighted Covariance Estimate	476
13.3	Diagonal VEC Model	480
13.4	Multivariate GARCH Modeling in <code>FinMetrics</code>	481
13.4.1	Multivariate GARCH Model Estimation	481
13.4.2	Multivariate GARCH Model Diagnostics	483
13.5	Multivariate GARCH Model Extensions	490
13.5.1	Matrix-Diagonal Models	490
13.5.2	BEKK Models	491
13.5.3	Univariate GARCH-based Models	494
13.5.4	ARMA Terms and Exogenous Variables	498
13.5.5	Multivariate Conditional t-Distribution	501
13.6	Multivariate GARCH Prediction	503
13.7	Custom Estimation of GARCH Models	506
13.7.1	GARCH Model Objects	506
13.7.2	Revision of GARCH Model Estimation	508
13.8	Multivariate GARCH Model Simulation	509
	References	513
14	State Space Models	515
14.1	Introduction	515
14.2	State Space Representation	516
14.2.1	Initial Conditions	517
14.2.2	State Space Representation in <code>S+FinMetrics/SsfPack</code>	517
14.2.3	Missing Values	523
14.2.4	<code>S+FinMetrics/SsfPack</code> Functions for Specifying the State Space Form for Some Common Time Series Models	524
14.2.5	Simulating observations from the State Space Model	536

14.3	Algorithms	538
14.3.1	Kalman Filter	538
14.3.2	Kalman Smoother	539
14.3.3	Smoothed State and Response Estimates	540
14.3.4	Smoothed Disturbance Estimates	540
14.3.5	Forecasting	540
14.3.6	S+FinMetrics/SsfPack Implementation of State Space Modeling Algorithms	540
14.4	Estimation of State Space Models	549
14.4.1	Prediction Error Decomposition of Log-Likelihood	549
14.4.2	Fitting State Space Models Using the S+FinMetrics/SsfPack Function SsfFit	550
14.5	Simulation Smoothing	555
	References	559
15	Factor Models for Asset Returns	561
15.1	Introduction	561
15.2	Factor Model Specification	562
15.3	Macroeconomic Factor Models for Returns	563
15.3.1	Sharpe's Single Index Model	564
15.3.2	The General Multifactor Model	569
15.4	Fundamental Factor Model	572
15.4.1	BARRA-type Single Factor Model	573
15.4.2	BARRA-type Industry Factor Model	574
15.5	Statistical Factor Models for Returns	582
15.5.1	Factor Analysis	582
15.5.2	Principal Components	589
15.5.3	Asymptotic Principal Components	597
15.5.4	Determining the Number of Factors	602
	References	607
16	Term Structure of Interest Rates	609
16.1	Introduction	609
16.2	Discount, Spot and Forward Rates	610
16.2.1	Definitions and Rate Conversion	610
16.2.2	Rate Conversion in S+FinMetrics	611
16.3	Quadratic and Cubic Spline Interpolation	612
16.4	Smoothing Spline Interpolation	616
16.5	Nelson-Siegel Function	620
16.6	Conclusion	624
	References	627

17 Robust Change Detection	629
17.1 Introduction	629
17.2 REGARIMA Models	630
17.3 Robust Fitting of REGARIMA Models	631
17.4 Prediction Using REGARIMA Models	636
17.5 Controlling Robust Fitting of REGARIMA Models	637
17.5.1 Adding Seasonal Effects	637
17.5.2 Controlling Outlier Detection	639
17.5.3 Iterating the Procedure	641
17.6 Algorithms of Filtered τ -Estimation	643
17.6.1 Classical Maximum Likelihood Estimates	644
17.6.2 Filtered τ -Estimates	645
References	647
Index	649

Preface

What is the book and why was it written?

This book is a guide to analyzing and modeling financial time series using **S-PLUS** and **S+FinMetrics**. It is a unique blend of econometric theory, financial models, data analysis, and statistical programming. It serves as a user's guide for Insightful's **S+FinMetrics** module of statistical functions for financial time series analysis and financial econometrics as well as a general reference for models used in applied financial econometrics. The format of the chapters in the book is to give a reasonably complete description of a statistical model and how it works followed by illustrations of how to analyze the model using **S-PLUS** and the functions in **S+FinMetrics**. In this way, the book stands alone as an introduction to financial time series analysis as well as a user's guide for **S+FinMetrics**. It also highlights the general analysis of time series data using the new time series objects in **S-PLUS 6**.

Intended audience

This book is written for a wide audience of individuals who work, do research or study in the areas of empirical finance and financial econometrics. The field of financial econometrics has exploded over the last decade, and this book represents an integration of theory, methods and examples using the **S-PLUS** modeling language to facilitate the practice of financial econometrics. This audience includes researchers and practitioners in the finance industry, academic researchers in economics and finance, and advanced MBA and graduate students in economics and finance. Researchers

and practitioners in the finance industry who already use **S-PLUS** and desire more functionality for analyzing and modeling financial data will find this text useful. It is also appropriate for financial analysts who may not be familiar with **S-PLUS** but who desire an integrated and open statistical modeling and programming environment for the analysis of financial data. This guide is useful for academic researchers interested in empirical finance and financial econometrics. Finally, this book may be used as a textbook or a textbook companion for advanced MBA and graduate level courses in empirical finance and financial econometrics.

Audience background

It is assumed that the reader has a basic familiarity with **S-PLUS** at the level of Krause and Olson (2002) and a background in mathematical statistics at the level of Hogg and Craig (1994), is comfortable with linear algebra and linear regression, and has been exposed to basic time series concepts as presented in Harvey (1993) or Franses (1998). Most importantly, the book assumes that the reader is interested in modeling and analyzing financial time series.

Overview of the book

The chapters in the book cover univariate and multivariate models for analyzing financial time series using **S-PLUS** and the functions in **S+FinMetrics**. Chapter one gives a general overview of the use of **S-PLUS 6** and highlights certain aspects of the language for statistical modeling. Chapter two introduces the new time series objects in **S-PLUS 6** and illustrates the specification, manipulation and visualization of these objects. Chapter three surveys time series concepts used throughout the book. Chapters four through eight cover a variety of topics in the modeling of univariate financial time series, including testing for unit roots, extreme value theory, time series regression models, GARCH models of volatility, and long memory models. Chapter nine introduces rolling analyses of time series models and covers related topics such as technical analysis of financial time series and moving average methods for high frequency data. Chapters ten through fifteen cover models for the analysis of multivariate financial time series. Topics include systems of regression equations, classical and Bayesian vector autoregressive models, cointegration, factor models, multivariate GARCH models, and state space models. Chapter 16 covers aspects of modeling time series arising from fixed income financial assets. Chapter 17, written by Victor Yohai and Jiahui Wang, describes robust REGARIMA models that allow for structural change.

What is S+FinMetrics?

S+FinMetrics is an **S-PLUS** module for the econometric modeling and prediction of economic and financial time series. With some 600 functions, version 1.0 of **S+FinMetrics** offers the following functionality:

- Easy-to-use Trellis plots for multivariate time series
- Time series manipulations such as missing value interpolation, disaggregation, differences, distributed lags and polynomial distributed lags
- Rolling sample statistics such as variance, maximum, and minimum
- Moving average operators for both regularly spaced and irregularly spaced time series
- Common technical analysis measures and indicators
- Statistical tests for normality, autocorrelation, heteroskedasticity, multicollinearity, GARCH effects, and long memory
- Extreme value theory models based on generalized extreme value and generalized Pareto distributions as well as copulas
- Autoregressive distributed lag regression models
- White and Newey-West corrections for heteroskedasticity and serial correlation
- Robust estimation of REG-ARIMA models and robust detection of level shifts, trend breaks, and outliers
- Rolling and recursive regression
- Generic rolling models for back-testing
- Long memory fractional ARIMA and SEMIFAR models
- Univariate GARCH models including long memory FIGARCH models
- Multivariate GARCH models
- Linear and nonlinear systems of regression equations
- Classical and Bayesian vector autoregression models
- Tests for unit roots and cointegration
- Vector error correction models
- State space models and efficient estimation, prediction, smoothing, and simulation using the Kalman filter

- Statistical multifactor models for large data sets based on asymptotic principal components
- Term structure interpolation

S+FinMetrics incorporates functions from **S+GARCH**, the **EVIS** library of functions for modeling extreme values created by Alexander McNeil, the **EVANESCA** library of functions for modeling extreme values utilizing copulas created by Rene Carmona, and the *SsfPack* C library of state space modeling functions created by Siem Jan Koopman. **S+GARCH** was originally developed by Zhuanxin Ding, Hong-Ye Gao, Doug Martin, Jiahui Wang and Yihui Zhan. The **S+FinMetrics** function `arima.rob` was written by Ana Bianco, Marta Garcia Ben, Elena Martinez and Victor Yohai. The **S+FinMetrics** long memory modeling functions **FAR**, **FARIMA**, **SEMIFAR** and `fgarch` were developed by Jan Beran, Andrew Bruce, Don Percival, Alan Gibbs and Jiahui Wang and supported by NSF grant DMI-9801614 to Insightful Corporation (formerly MathSoft, Inc.). Hu McColloch kindly provided the term structure data included with **S+FinMetrics**, and James MacKinnon provided data sets for the response surface critical values for the Dickey-Fuller and Phillips-Ouliaris distributions.

Contact information and website

The authors are responsible for all of the material in the book except the material on robust change detection, which was written by Victor Yohai. Eric Zivot is primarily responsible for chapters 2-6, 9-12 and 14-15 and Jiahui Wang is primarily responsible for chapters 1, 7-8, 13, and 16. The authors may be contacted by electronic mail at

`ezivot@u.washington.edu`
`jwang@insightful.com`

and welcome feedback and suggestions for improvements to the contents of the book. The website for the book is located on the Insightful Corporation website at

<http://www.insightful.com/support/finmetrics10>

Acknowledgements.

This book would not have been written without the support and encouragement from Insightful. The idea for the **S+FinMetrics** module was conceived by Douglas Martin and the authors. The development of **S+FinMetrics** was completed at Insightful by Jiahui Wang, Quan Wen and Hui Huang

with help from many people. In particular, Jan Beran wrote many of the long memory functions while acting as a consultant to Insightful. Siem Jan Koopman helped to incorporate the *SsfPack* functions into S-PLUS and to write the chapter on state space models. Alexander McNeil and Rene Carmona graciously provided background material and S-PLUS examples for the material in the chapter on modeling extreme values. A number of people were helpful in proofreading the book and testing the software. Particular thanks go to Andrew Bruce, Chuck Curry, Zhuanxin Ding, Ruud Koning, Steve McKinney, David Weitzel, Quan Wen and Bingcheng Yan.

Topographical conventions

This book obeys the following typographic conventions:

- The *italic* font is used for emphasis, and also for user-supplied variables within UNIX, DOS and S-PLUS commands.
- The **typewriter** font is used for S-PLUS functions, the output of S-PLUS functions and examples of S-PLUS sessions.
- S-PLUS objects of a specified class are expressed in **typewriter font** enclosed in quotations “”. For example, the S-PLUS function `timeSeries` creates objects of class “`timeSeries`”.

Displayed S-PLUS commands are shown with the prompt character `>`. For example

```
> summary(ols.fit)
```

S-PLUS commands that require more than one line of input are displayed with the continuation prompt indicated by `+` or `Continue string:.` The S-PLUS output and plots in this book were generated from `S+FinMetrics` Version 1.0 and S-PLUS Version 6.0 release 2 for Windows. The S-PLUS output and “`timeSeries`” objects were generated with the options settings

```
options(width=60)
options(time.zone='GMT')
```

In some cases, parts of long output from S-PLUS functions is omitted and these lines are indicated by

```
...
```

Some of the output has been hand edited to avoid line overflow.

References

- [1] FRANCES, P.H. (1998). *Time Series Models for Business and Economic Forecasting*. Cambridge University Press, Cambridge.
- [2] HARVEY, A.C. (1993). *Time Series Models, Second Edition*. MIT Press, Cambridge.
- [3] HOGG, R.V. AND A.T. CRAIG (1994). *Introduction to Mathematical Statistics, Fifth Edition*. Prentice Hall, New York.
- [4] KRAUSE, A. AND M. OLSON (2002). *The Basics of S and S-PLUS, Third Edition*. Springer-Verlag, New York.