Reviewer: Robert Alan Yaffee
New York University

Stata 10 (Time Series and Forecasting)

StataCorp LP, College Station, TX. USD 1,795 (corporate), USD 985 (educational) for single user **Stata**/SE 10 (exact price varies by version and purchaser status).
http://www.stata.com/

A review of **Stata** versions 9 and 10 time series and forecasting capability

Stata 10 is a powerful, versatile, and flexible statistical package with a wide range of user-friendly and accurate time series analytical and forecasting commands. It was released in June 2007, about two years after **Stata** 9 appeared. Prospective users want to know what features the package offers, while veteran users want to know what is novel about the latest version(s). This review addresses time series and forecasting capability in **Stata** 10 and notes prominent new features in versions 9 and 10, what is still needed and what might be refined. Because all users need to have confidence in the accuracy of the package, this review examines accuracy of the principal time series algorithms. It also briefly describes associated graphics, diagnostics, documentation, help facilities, output, customization, and support. Altogether, these features characterize a very well-designed and well-crafted and powerful time series analysis and forecasting package (Baum 2004).

Filters and smoothers

For each stage of the time series modeling process, Stata provides capability with which to test and model the time series. For preliminary examination of time series and their distributions, **Stata** 10 has flexible and versatile graphical power. For simple, basic series extrapolation, **Stata** has moving average and exponential smoothing capability. It offers simple and customizable, weighted moving averages. It also offers simple and double exponential smoothing. It contains a Holt and Winters two parameter version to accommodate linear trend, as well as three parameter versions to accommodate additive and multiplicative seasonality in the series. For these smoothers, **Stata** allows user-specified or grid search convergence on optimal smoothing constants. Stata contains a variety of Tukey nonlinear smoothers. If the dataset is short and the analyst needs quick, economical, and easy extrapolation, **Stata** provides these smoothers.

Preliminary analysis

For preliminary visual analysis, Stata offers a simple and easy time series line plot. This plot can display single or multiple series under review. These series may be graphed in different colors, line patterns, line thickness, and with or without symbols. The time series plots support multiple titles, footnotes, and captions. The user can have legends and/or annotations. Multiple graphs can be paneled one on top of the other or side by side. Once the user learns the graphics language, he can superimpose one type of graph on the other.

For identification of the time series, Stata provides for the examination of the autocorrelation structure with ASCII and graphical correlograms. For example, the ASCII `corrgram` command on the first difference of the natural log of the consumer price index for all urban consumers on all items (CPIAUCNS) from the Federal Reserve Economic Depository (FRED, http://research.stlouisfed.org/fred2/) neatly generates both correlogram functions:

```
. corrgram dlncpi, lags(25)
                                         -1      0      1 -1      0      1
    LAG      AC       PAC       Q    Prob>Q  [Autocorrelation]   [Partial Autocor]

    1      0.4465    0.4466   207.52  0.0000
    2      0.3544    0.1910   338.42  0.0000
    3      0.3341    0.1581   454.85  0.0000
    4      0.3316    0.1477   569.68  0.0000
    5      0.2614    0.0269   641.08  0.0000
    6      0.2465    0.0432   704.64  0.0000
    7      0.2905    0.1163   792.98  0.0000
    8      0.2665    0.0528    867.4  0.0000
    9      0.2615    0.0584   939.14  0.0000
   10      0.2618    0.0599   1011.1  0.0000
   11      0.3077    0.1133   1110.6  0.0000
   12      0.2766    0.0517   1191.1  0.0000
   13      0.2473    0.0129   1255.6  0.0000
   14      0.2212   -0.0020   1307.2  0.0000
   15      0.1750   -0.0568   1339.5  0.0000
   16      0.1578   -0.0321   1365.8  0.0000
   17      0.1432   -0.0232   1387.5  0.0000
   18      0.1354   -0.0279   1406.9  0.0000
   19      0.1042   -0.0406   1418.4  0.0000
   20      0.1046   -0.0173     1430  0.0000
   21      0.1732    0.0837   1461.8  0.0000
   22      0.0981   -0.0736     1472  0.0000
   23      0.1601    0.0745   1499.3  0.0000
   24      0.1710    0.0510   1530.4  0.0000
   25      0.1162   -0.0523   1544.8  0.0000
```

whereas separate graphical commands can generate functions that can be easily combined by the user as shown in Figure 1.

In addition to the autocorrelation and partial autocorrelation functions, users have the option of using a log periodogram for detection of seasonal periodicity in a model.

To obtain preliminary assessment of nonstationarity, Stata proffers a substantial variety of tests. In addition to the Dickey–Fuller and Augmented Dickey–Fuller (`dfuller`) tests, it offers the Phillips–Perron `pperron`, the Kwiatkowski, Phillips, Schmidt, and Shin `kpss`, and the more powerful Elliott, Rothenberg and Stock `dfgls` tests. Once the series has been transformed to stationarity, Stata offers time series regression or an ARIMA repertoire.
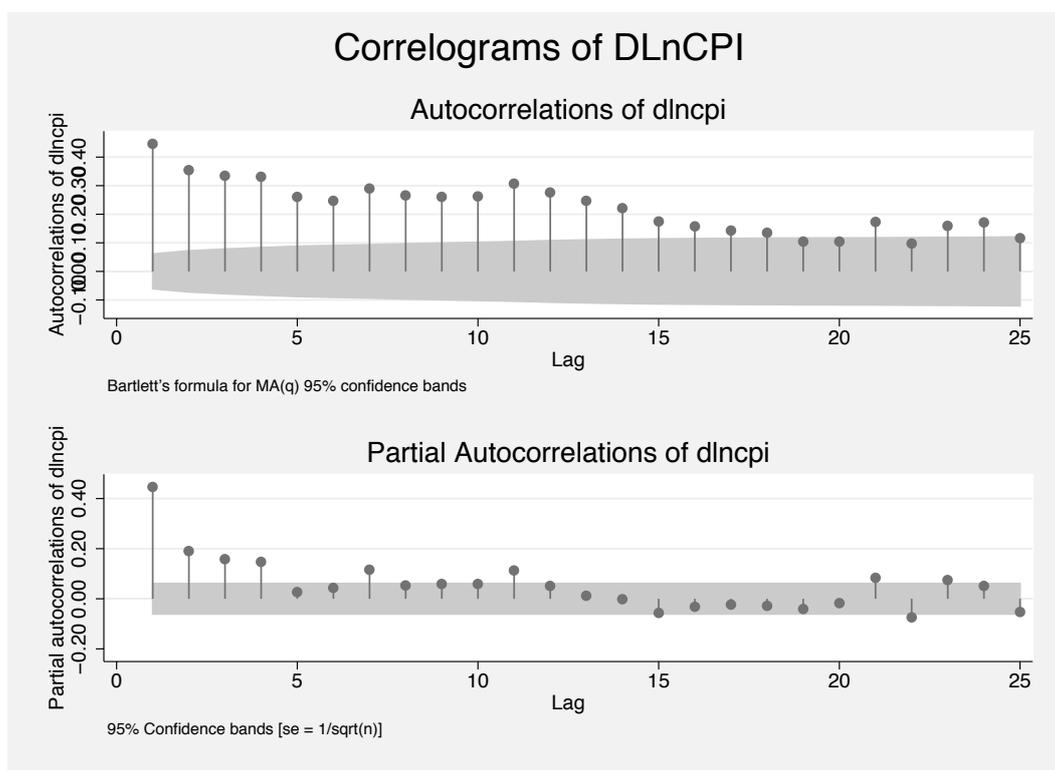
Correlograms of DLnCPI

Autocorrelations of dlncpi

Bartlett's formula for MA(q) 95% confidence bands

Partial Autocorrelations of dlncpi

95% Confidence bands [se = 1/sqrt(n)]

Figure 1: Combined correlograms.

Time series regression models

When a residual analysis of an ordinary least squares regression reveals autocorrelation in the residuals, we can use time series regression models. **Stata**'s `prais` includes Cochrane–Orcutt, Hildreth–Lu and Prais–Winston estimators, while `newey` runs Newey–West regressions. The first two commands can be run with a robust option to asymptotically accommodate heteroskedasticity while the latter can also accommodate autocorrelation. **Stata** 10 allows the use of the `hc2` and `hc3` versions of asymptotically heteroskedastically consistent variance-covariance matrices as well. With the use of White or Newey–West estimators, **Stata** makes semi-robust time series easy.

The new version added new regression postestimation diagnostic tests for time series data. These diagnostics test the need for adjustments for residual serial correlation or ARCH effects. Among the tests included in the post-regression `estat` command are options for the Breusch–Godfrey `estat bgodfrey` test for serial correlation, the Durbin–Watson d (`estat dwatson`), the Durbin–Watson alternative `estat durbinalt` test, and the Lagrange Multiplier test (`estat archlm`) for ARCH effects in the residuals.

ARIMA models

For time series model building, **Stata** features the `arima` command. **Stata** offers additive non-seasonal models as well as multiplicative seasonal modeling capability. Stata also permits the

user to enter time-varying regressors in the ARIMA command to construct dynamic linear models, RegARIMA, or ARIMAX models. The predictors can be indicator or discrete variables used to model the impact of external events. They may also be time-varying predictors that can be used as stochastic regressors in the analysis to form the basis of a dynamic regression model. For ARIMA or RegARIMA models riven with heteroskedastic residuals, `Stata` provides White sandwich variance-covariance estimators most of these programs simply with a `robust` option.

The ARIMA command has its own residual diagnostics. The residuals can be saved with the `predict` *varname*, `residual` command for graphical or objective tests for model misspecification. These tests include the Box-Ljung Q tests `corrgram` or Durbin–Watson `durbina` tests for autocorrelation, the Jarque–Bera `jb`, the Smirnov–Kolmogorov `sktest` test, the Shapiro–Wilk `swilk` test for normality. Bartlett `wntestb` and Box-Ljung white noise `wntestq` tests are also available.

With the `estat ic` postestimation command, the user can save the information criteria (AIC and BIC) of the model just estimated. He can use this command to determine the order of nested models or for a fit comparison of those models. Thus, he can select the best model for forecasting.

`Stata` post-estimation commands make forecasting simple. `Stata` offers either iterative projections (one-step ahead), structural, or dynamic forecasts. For models with explanatory variables, the user may wish to program a conditional forecast. He can generate the forecasts for the explanatory variables first. Stata can generate *ex post* or *ex ante* forecasts and their confidence intervals. The forecast profile can be saved and graphed with ease.

Nonlinear time series analysis

`Stata` boasts a wide range of ARCH and GARCH models for volatility analysis. In fact, `Stata` currently provides more flavors of (G)ARCH models than most other statistical packages with the ARCH `arch` command and its GARCH `garch` option for GARCH models. A later generation of ARCH and GARCH models includes GARCH-in-mean (`archm`), as well as versions designed to model leverage effects of enhanced volatility due to bad news, such as the exponential (`earch`) ARCH and GARCH, threshold (`tarch`) ARCH and (`tgarch`) and GARCH, as well as the Glosten, Jagannathan, and Runkle (1993) GARCH version. The GJR GARCH, formulated as a version of threshold GARCH, was recommended by Engle and Ng (1995) for its ability to model the bad versus good news leverage effects. For long-memory situations, Stata has an asymmetric power (`aparch`, `apgarch`) GARCH. Among the other later generation GARCH models are the nonlinear power (`nparch`) GARCH (for highly persistent volatility), threshold Power (`tparch`) GARCH for different models based on a given threshold, nonlinear power(`narchk`, `nparchk` GARCH) with single shifts, and power (`parch`, `pgarch`) GARCH, among others. `Stata` 9 and 10 permit model formulation with constraints, asymmetric leverage effects, and can graph volatility skews and smiles. `Stata` 9 estimates GARCH with a normal distribution, so that `Stata` 9 users may have had to resort to the normalizing and variance stabilizing (NOVAS) transformation of Dimitris Politis for their GARCH modeling. `Stata` 10, however, also allows GARCH modeling with a *t* or generalized error distribution (GED) to accommodate those fat-tailed leptokurtotic or heteroskedastic financial distributions.

New commands and features

Many analysts work with real-time or tick data that are sampled on intra-day intervals. To facilitate this analysis, Stata 10 has incorporated a clock-time index, so that the time of day can be used as an index in the construction of a time series. For those who might later bin their data into equidistant time intervals, this renders time series of such data possible.

Most time series analysts visually examine their data before they proceed with analysis. Stata 10 continues the newly developed graphics procedures found in Stata 9, but adds point-and-click graphics editing. The Stata 9 time series graphics capability for a wide arrange of time line graphs, interrupted line graphs, lowess plots, graphical correlograms, graphical cross-correlation plots, residual plots, has been fortified with a new point-and-click graphics editor. The user may edit and annotate the graph after it has been generated with the new graphics editor.

With Stata 9, new time series commands included the rolling analysis and recursive least squares, the Box–Jenkins seasonal ARIMA models, the multivariate commands of vector autoregression and vector error correction models as well as the model comparison with postestimation `estat ic` capability. With Stata 10, models could be indexed by time as well as days, weeks, months, quarters, or years.

Both Stata 9 and Stata 10 have the capability of estimating and graphing various types (raw,
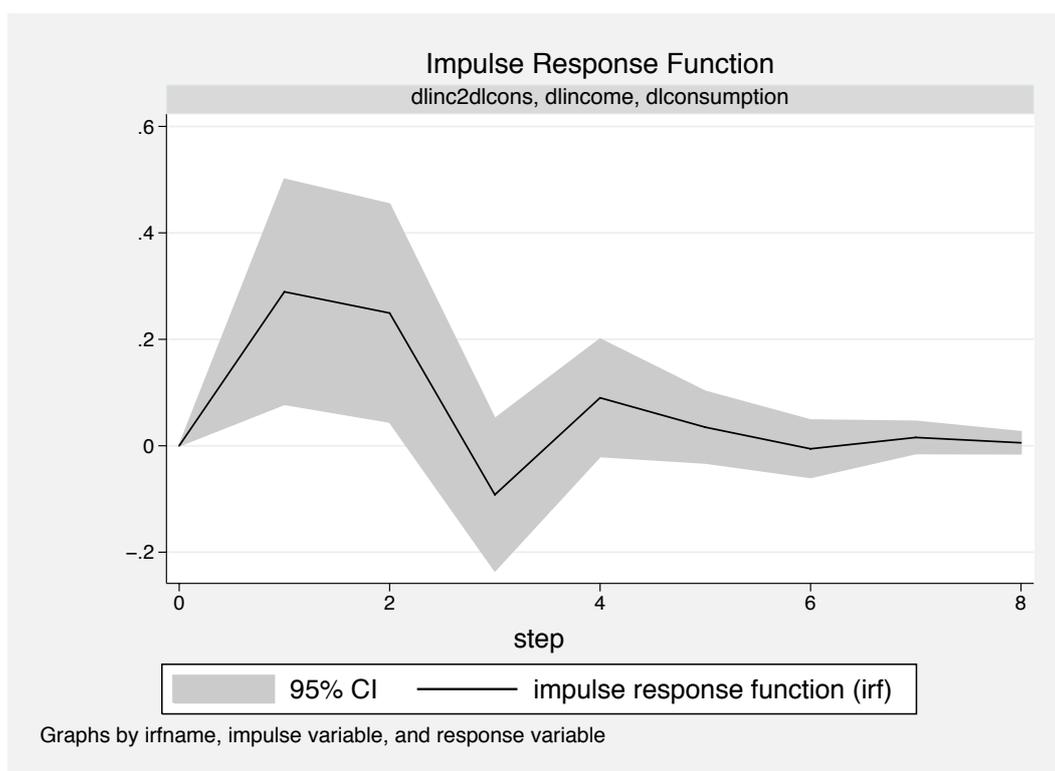

Figure 2: Impulse response function from a univariate dynamic regression model of differenced natural log of consumption as a response to a unit impulse in the differenced natural log of income, using Lütkepohl (1993) Table E1: West German macroeconomic data.

orthogonalized, structural, cumulative) of impulse response functions, dynamic multipliers, and forecast error variance decompositions for both univariate and multivariate vector models. We show an impulse response function of an impulse from the first difference of the natural log of income on the first difference of the natural log of consumption in a univariate dynamic regression, using Lütkepohl (1993), Table E.1. of quarterly West German macroeconomic data in billions of DEM in Figure 2.

Rolling analysis

Rolling analysis is an important new feature. Before the we forecast a series, we want to be sure of parameter stability. Stata 10 offers a new rolling analysis with the `rolling` command, saving the parameters of a model as it is estimated in a window rolling along the time axis. We can plot those parameters over time to see if they are variable or stabilizing. We can use regression, ARIMA, or other models in this command to explore the asymptotic assumptions. The window can be of fixed or expanding size. A recursive least squares analysis option provides a chance to examine the parameters before attempting to forecast with the model. In Figure 3, we examine the parameter instability of the constant and time coefficient of a 60 day windowed rolling regression of the Chicago Board of Options Exchange Volatility Index (VIX) over time before deciding to attempt a forecast.

By generating and overlaying confidence limits for these coefficients, the time variation in the
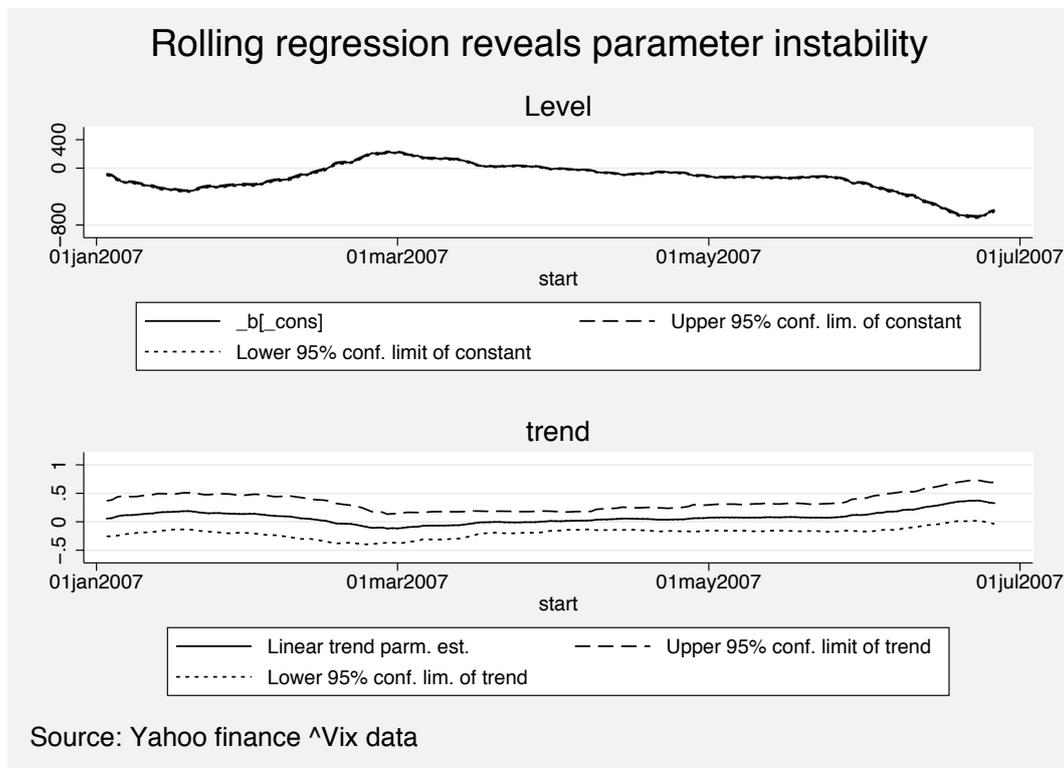

Figure 3: Rolling regression estimates of constant and coefficient in 60-day window.

constant and regression coefficient is shown to be significant over time, providing advance warning to forecasters. By restricting the window to the appropriate size, rolling analysis with confidence intervals can be used for backtesting as well.

Multivariate time series analysis

For multivariate analysis, `Stata` 9 introduced an elegant vector autoregression `var`, structural vector autoregression `svar` and vector error correction `vecm` model capability. For simple reduced form models, `Stata` has a basic command that allows fitting `var` models without constraints. This command generates an output table. The user may also output a set of impulse response functions (the response of an endogenous variable over time to a unit shock to another endogenous variable in the model). When there are multiple endogenous variables, independent impulse response functions may be generated, shown in Figure 4. In addition, forecast error variance decomposition, which reveals the proportional contribution of each variable to the forecast error variance are also easily available (Figure 5).

Most users will need a more elaborate specification. Model specification can be controlled by designating exogenous variables or by constraints on variables and their lags. For `var` and `svar` postestimation, `Stata` has a commendable repertoire of diagnostic options, such as Granger causality tests, Lagrange Multiplier tests for residual autocorrelation, information
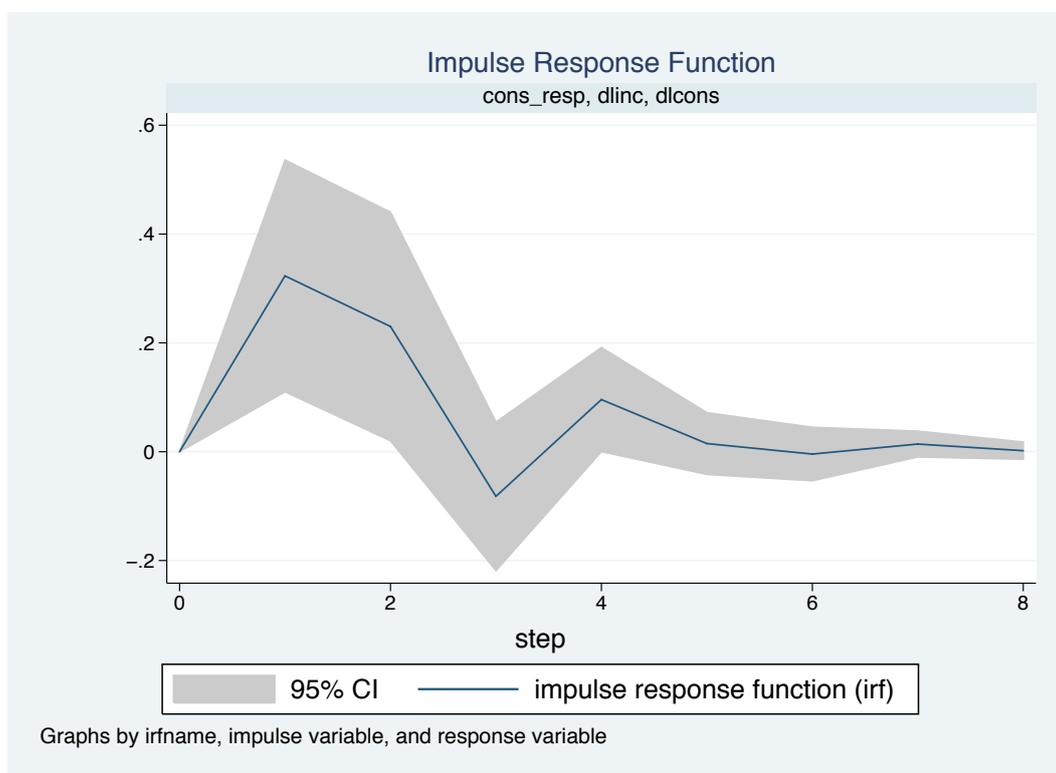

Figure 4: Orthogonalized impulse response functions of differenced natural log of consumption and differenced natural log of income to unit impulses from the other, using Lütkepohl (1993) West German economic data.
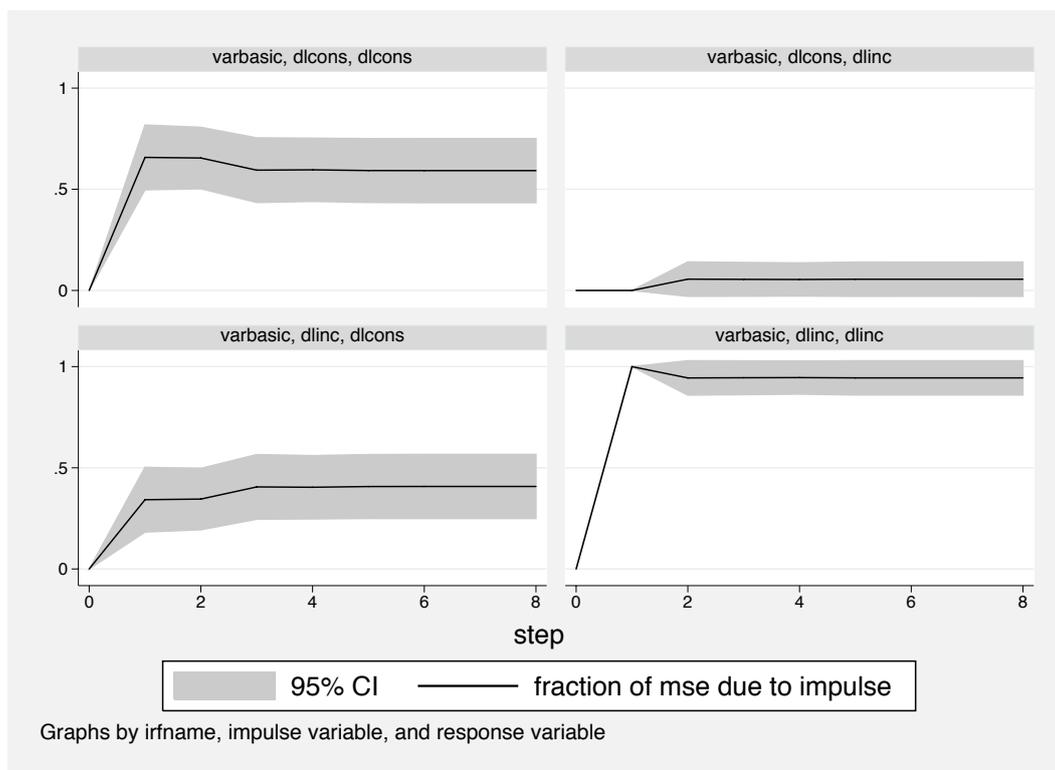
Figure 5:   Forecast error variance decomposition of differenced natural logged consumption and differenced natural log of income, using Lütkepohl (1993) West German economic data.

criteria to test and select optimal lag order, residual normality tests, and eigenvalue stability tests for autoregressive roots, to help diagnose and specify these models. Constraints may be applied as well as deterministic variables in the model. When a stable vector autoregression is reparameterized as a moving average process, it can be modeled as a structural vector autoregression. Orthogonalized impulse response functions, cumulative impulse response functions, forecasts, and their confidence intervals can be graphed while the forecast error variance can be decomposed and tabulated. In this way, a dynamic simultaneous equation model can be analyzed.

Moreover, these processes may be cointegrated. The Stata vec command has the capability to identify the cointegrating rank of the long term parameters with the Johansen trace and maximum eigenvalue tests. A variety of simple options are available for model specification including a variety of restrictions on trends and constants in the levels or in the cointegration relation system. The var or vec commands permit easy autoregressive lag order selection with either standard or Lütkepohl versions of information criteria (AIC, SBC, or HQ). The proper lag order is assessed with Lagrange multiplier tests. Model stability is tested with a simple command to test whether the equation roots are all less than one. Residual normality is easily assessed with Jarque–Bera tests for the parameters. The autoregressive roots are easily analyzed for stability of their eigenvalues. Then impulse response functions and forecast error variance decompositions can be generated. Various models can be tracked with their information criteria and if the user keeps a record of what was in these models, the optimal

Parameters	NIST results	Stata 9.2	SPSS 15	SAS 9.13
R^2	0.995479004577296	0.995479	0.995	0.9955
σ	304.854073561965	304.85407	304.854	304.85407
Constant	-3482258.63459582	-3482258.6	-3482259	-3482259
b1	15.0618722713733	15.061872	15.062	15.06187
b2	-0.0358191792925910	-.03581918	-0.036	-0.03582
b3	-2.02022980381683	-2.0202298	-2.020	-2.02023
b4	-1.03322686717359	-1.0332269	-1.033	-1.03323
b5	-0.0511041056535807	-0.05110411	-0.051	-0.0511
b6	1829.15146461355	1829.1515	1829.151	1829.15146

Table 1: Comparative accuracy of OLS regression model parameters using National Institute of Standards and Technology NIST(a) Longley dataset.

ones can be modeled later.

While Stata 9 introduced the impulse response functions for dynamic models, Stata 10 includes computation of individual and cumulative impulse response functions as well as the capability of bootstrapping their standard errors.

Analytical accuracy

Needless to say, we want to be assured that our software is producing accurate results. I have downloaded data from the National Institute of Standards and Technology (NIST) Web site to test the main time series commands for accuracy. I have tested OLS regression, simple exponential smoothing, nonseasonal and seasonal ARIMA models. To test classical ordinary least squares regression, I used the Longley data (Longley 1967). To test the exponential smoothing, I downloaded sample data from the NIST Web site (National Institute of Standards and Technology 2003). To test ARIMA data, I used the Antuan Negiz aerosol particle size data (Negiz 2003), and to test the seasonal ARIMA model, I used the Box-Jenkins Series G data from the NIST Web site (Box and Jenkins 1976). All of these data are listed on the NIST Web site, and the test results, displayed in four tables, are noteworthy.

In Table 1 we compare the classical ordinary least squares regression results from a NIST analysis of the Longley data, used to assess the accuracy of an electronic computer, to those from Stata 9.2, SPSS 15 (SPSS Inc. 2006), and SAS 9.13 (SAS Institute Inc. 2003), all on a Windows XP/SP2 operating system. The first variable is the dependent while the last variable, year, is treated as x6. We first compare the R^2 and the model error standard deviation and then the parameter estimates and their standard errors. Apart from minor rounding error, these results are essentially identical. Descriptions of the data are found in the references.

In Table 2, we compare the simple exponential smoothing of that of NIST (NIST(b)) with that of Stata. We find that the fit as measured by sum of squared errors (SSE) is actually better for the Stata command.

In the two panels of Table 3, we compare two ARIMA models. Using the nonseasonal Antuan Negiz data measuring particle size after pulverization and drying, we run two different models.

Parameters	NIST	Stata
α	0.5	0.5
SSE	208.94	149.03739

Table 2: Comparison of **Stata** and NIST simple exponential smoothing parameters using NIST(b) sample data.

Parameters	NIST ARIMA$(0,1,1)$	Stata ARIMA$(0,1,1)$
Constant	-0.00511	-0.00511
Constant se	0.011431230	0.01154271
MA(1)	0.393	0.392
MA(1) se	0.03902847	0.02686429
Parameters	NIST ARIMA$(2,1,0)$	Stata ARIMA$(2,1,0)$
Constant	-0.0040182899	-0.0049692
Constant se	0.011972592	0.0121377
AR(1)	-0.4060457	-0.4063475
AR(1) se	0.041885445	0.02851256
AR(2)	-0.1641447	-0.1648906
AR(2) se	0.041836922	0.03925299

Table 3: Comparison of **Stata** and NIST nonseasonal ARIMA model parameters using Antuan Negiz particle size (NIST(c)) data.

In the upper panel, we model an AR(2) model and in the lower panel we model an MA(1) model. Whether the modeling is an AR(2) model or an MA(1) model, the ARIMA results are essentially the same, with the exception that the **Stata** standard errors are a little smaller. In the AR(2) the constant in the NIST AR model is computed from the mean as $Constant = (1 - \phi_1 - \phi_2)\mu$ in order to compare the constants (NIST Web site).

For the seasonal ARIMA, the results are also comparable. There is a difference in the sign of the moving average parameters. **Stata** parameterizes the moving average parameter θ_t in a first order moving average process as $y_t = 1 + \theta e_{t-1}$ whereas the conventional Box–Jenkins parameterization uses a $y_t = 1 - \theta e_{t-1}$ parameterization. To show that the untransformed series are identical, the raw mean, variance, and sample size are given in the lower panel of Table 4. The **Stata** model was run with the `vce(oim)` option. The parameters of the models are almost identical, whereas there is some minor difference in the standard errors. In general, the user can rely on **Stata** for an accurate and reliable time series analysis.

Support, documentation, and other resources

Stata support is superb and virtually full-service. The user can directly access help files in the program or he can obtain competent technical support via e-mail or an 800 telephone number. Integration of **Stata** with the World Wide Web allows the user to obtain assistance from the official Stata support database. From there, we can access *Stata News*, the *Stata Journal*, the *Stata Technical Bulletin* and other Web resources. For training, we can sign up for the NetCourses that **Stata** offers or find other short courses on Stata that are offered.

Parameters	NIST ARIMA$(0,1,1)(0,1,1)_{12}$	Stata ARIMA$(0,1,1)(0,1,1)_{12}$
MA(1)	0.3765	-0.4018323
MA(1) se	0.0811	0.0896442
SMA(1)	0.5677	-0.556934
SMA(1) se	0.0775	0.0731055
raw mean	280.2986	280.2986
raw variance	14391.9170	14391.92
Sample size	144	144

Table 4: Comparison of Stata and NIST seasonal ARIMA model parameters using Box–Jenkins Series G (NIST(d)) monthly total of international airfares from 1949–1960.

The integration of Stata with the World Wide Web makes access of support information easy. The `findit` command obtain lists of official help files, frequently asked questions, examples, relevant *Stata Journal* articles, and resources from Stata and other users available on the World Wide Web. There is a support database on the World Wide Web. Automatic updates over the World Wide Web provide Stata users with the latest programs. The `hsearch` utility builds an index of all of the programs that contain mention of the target word or set of words. *Stata News* notifies users of new developments and upcoming events of interest to members of the Stata community.

Among other Web sources, there is widespread academic support of Stata. Outstanding among these sources is the excellent Web site of the Academic Technical Services at University of California at Los Angeles (UCLA) that maintains an excellent Stata portal at http://www.ats.ucla.edu/stat/stata/, which even features starter kits, classes, seminars, learning modules, movie tutorials, a library of Stata programs for research and teaching, code fragments and tools for generating LaTeX output as well as a variety of important Web links to online sources of support.

For documentation, Stata offers an encyclopedia of documentation, consisting of *Getting Started with Stata*, *Data Management*, *Graphics*, *Users Guide* and *Programming* manuals (StataCorp. 2007). Also included is the multi-volume *Reference Manual*, an alphabetical sorting of statistical command descriptions. Larger modules, such as those of *Time Series*, *Multivariate Statistics* and *Longitudinal/Panel Data* analysis, among others, have their own manual, replete with chapters alphabetically sorted by statistical command. The chapter for each command consists of descriptions of the syntax, their details, common models, options, examples of models, and their output interpretation. Saved statistics, methods, formulae, missing data issues, and references are also described. Following the chapter on the statistical command comes another one devoted to postestimation for the previous command. When connected to the Internet, the user has access to helpfiles on the Stata support Web site as well as a variety of Web resource links to tutorials, books, articles, and other resource materials.

Customization

Another nice feature of Stata is the capability for users to customize the package. Stata provides a matrix language called Mata with which programmers can write their own programs. Stata allows users to modify the programs which they run with the Stata programming lan-

guage. This is a feature that many statistical packages do not permit. These programs are saved in '.ado' files and the user may modify these to suit his needs.

Stata community support is available from a library of programs accessible by a Stata user whose computer is connected to the World Wide Web: the SSC Archive at http://ideas. repec.org/s/boc/bocode.html. We merely type ssc describe*letter* and a list of packages beginning with that letter is displayed. If he wishes to install such a packages, the user simply types: ssc install*packagename* and this is downloaded into the Stata library for our personal use. In this way, users customize Stata's capabilities to fit their needs.

Moreover, users can write their own programs in '.do' files or '.ado' files. These can be activated for submission by typing do filename in the command window. These features allow the users who know the Stata programming language the ability to customize the package to suit their needs.

For example, at the North American Stata Users Group Conference in Boston on 2007-08-14, Ben Jann of ETH Zürich presented a customized estout program for comparative presentation of Stata output (Jann 2007). This program is now available in the SSC Archive, freely available to Stata users for downloading and installation. After installing this package, I was able to display a comparison of three ARIMA models of Intel closing price data, downloaded from http://finance.yahoo.com/.

```
. eststo ar1c: arima intc, ar(1)
. eststo ar12c: arima intc, ar(1/2)
. eststo ar123c: arima intc, ar(1/3)
. esttab, aic bic title(IBM closing prices) mtitle("ar1c" "ar12c" "ar123c")

IBM closing prices
```

	(1) ar1c	(2) ar12c	(3) ar123c
intc			
_cons	32.25***	32.85***	34.41**
	(3.82)	(3.63)	(3.27)
ARMA			
L.ar	0.997***	0.942***	0.963***
	(549.68)	(43.10)	(43.01)
L2.ar		0.0555*	-0.110**
		(2.55)	(-3.23)
L3.ar			0.145***
			(7.28)
sigma			
_cons	0.972***	0.978***	0.976***
	(185.27)	(167.47)	(138.08)
N	1147	1147	1147
AIC	3479.7	3478.7	3465.1
BIC	3494.8	3498.9	3490.3

```
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

The Stata programming language is so familiar to a C programmer that the user can write his own programs with some amount of ease. Moreover, if he knows the C programming language, he can use a plugin option to run the dynamic link library that he might write on his own.

What is needed

Although Stata 10 has a vast variety of time series commands, it, like all other packages, could benefit from the addition of a few more. Perhaps for the vast majority of people, these new suggestions might not be compelling. For some advanced users, however, they might be helpful. We therefore tender some suggestions that some more advanced users might wish included in the Stata repertoire.

At this time, there is no modeling capability for irregularly spaced time series. There are circumstances where the data generating process under study is not equally spaced in time. There are processes with weekend gaps, or trapping gaps during winter snows, etc. where no data were collected. The observations for such series were plagued with missing values at these times. There should be available methods that can handle such irregularly spaced time series. One such method might be an automatic command for binning an irregularly spaced time-series. Another might be Croston's method, usually applied to situations of intermittent demand (Shenstone and Hyndman 2003) and another is the autoregressive conditional duration model (Engle and Russell 1998). Having these capabilities would allow researchers to investigate a larger variety of time series.

The exponential smoothing capability could benefit by the addition of Everette Gardiner Jr.'s damped trend (Gardiner 2005). This method has been found to be very accurate at forecasting under particular circumstances. James W. Taylor has also suggested an exponential smoothing method that can handle interventions, which should prove very useful.

There are processes with strong autocorrelation persistence and slow decay of the autocorrelation function, that can be modeled by long-memory parameterization. Fractionally integrated ARIMA, or ARIFIMA, models are needed to analyze long-memory data generating processes. The size of the d in the p, d, q parameters of the arima ranges between 0.2 and 0.5. Although the SSC Archive contains a rescaled range test and a Geweke and Porter-Hudak test, Stata does not contain a quasi-full-information estimated FARIMA or ARIFIMA command that accommodates stochastic regressors, generates forecasts, while allowing for full estimation of the three p,d,q parameters.

Although Stata has a one of the most powerful GARCH repertoires available from currently available software, it could benefit by adding a few tests and graphs as automatic output, along with some more models for long-memory GARCH capability. To determine whether there are leverage effects, it would be helpful were Stata to include the sign bias test, positive size sign bias test, negative size sign bias test, and the joint test (Engle and Ng 1995). In addition to its currently available asymmetric power GARCH command, it might have integrated GARCH along with fractionally integrated: FIGARCH (Baillie, Bollerslev, and Mikkelsen 1996; Chung 1999) and fractionally integrated exponential GARCH: FIEGARCH (Bollerslev and Mikkelsen 1996) as well as hyperbolic GARCH models (Davidson 2004) to handle long memory processes. The latter would be especially useful in fitting asymmetric models with leverage effects. In addition to the ability to model variance clusters and outliers in univariate GARCH, some multivariate GARCH (MGARCH) models—Engle and Kroner's BEKK (Engle and Kroner 1995), the constant conditional correlation (Tse and Tsui 2002), dynamic conditional correlation along with the dynamic conditional correlation model of (Engle 2002)—should be added. Possible options are discussed by Palandri (2004). Hull and White (1999) have suggested that as the size of the high frequency dataset grows, there may be more need to give greater emphasis to more recent data. They have suggested volatility

updating by historical simulation weighted by adjusting the current with previously observed volatility to give more recent data more weight in estimation of volatility. Although `Stata` has threshold GARCH and nonlinear ARCH with a single shift, stochastic volatility models with multi-jump diffusion might be another helpful addition if there were more than one crisis.

Although stationary dynamic simultaneous equation models can be accommodated by the `var`, `svar`, and `vecm` commands, they deal with nonstationarity through the prior differencing or the incorporation deterministic trends, seasonal dummies, seasonal trigonometric functions, or cointegration. Stata needs more flexibility for dealing with nonstationary processes in their original levels and variances. State space models with the augmented Kalman filter (Durbin and Koopman 2000) could offer a form of structural method of modeling such time series. Because a lot of data are nonstationary and analysts may need to examine them in their levels configuration, such state space models would be a substantial, worthwhile, and welcome addition to `Stata`. Multivariate state space models could permit dynamic factor analysis to be performed. For nonlinear model and non-Gaussian processes, Poisson state space and particle filter commands would be welcome additions.

Many financial analysts want to analyze and forecast the volatility clustering of asset prices or returns, while others would like to model extreme values of a loss series. They may want to use a Black–Scholes calculator to compute the implied volatility to compare it to the historical volatility or realized volatility described by a GARCH model to determine whether asset is under or overvalued. They may want to be able to perform Value-At-Risk (VaR) analysis and and the factors that contribute to it. They need a flexible modeling capability for volatility as well as peaks over thresholds (extreme events). This can be done with a distributional analysis of the losses over time. However, the tails of a normal distribution fail to accurately model such loss distributions. To properly model cumulative losses over time, Frechet or generalized extreme value distributions could be helpful in modeling such tail-loss, allowing us to examine the quantiles of cumulative losses over a some period of time to estimate VaR. Generalized Pareto distributions and functions allowing models to use them would be helpful for risk managers performing a peaks over threshold analysis. Chavez-Demoulin and Embrechts (2004) have suggested smoothing extreme value methods fitted by penalized likelihood to explore the extreme value processes. If the loss distributions are too unusual, they might want the ability to use Markov-chain Monte Carlo (MCMC) estimation of their historical data may necessary. These methods must be adapted so they capturing the volatility clustering within the data. Financial analysts often want to perform portfolio allocation. To do so, they would like the ability to use a mean return-variance Markowitz efficient frontier optimization method to be able to substitute a Beta, Sharpe, or Sortino ratio for the standard deviation as a measure of risk (Chavez-Demoulin and Embrechts 2004; King 2001; Marrison 2002).

There are nonlinear time series, apart from GARCH, that might be very useful. `Stata` 10 needs a test for nonlinearity. Two types of such a test are the Brock, Dechert, and Scheinkman (BDS) or the TAR-F test (Tsay 2001). Other forms of nonlinear time series models should be available. Among those might be the bilinear, threshold autoregressive (TAR), self-exiting threshold autoregressive (SETAR), or exponential autoregressive model. These models would enhance the nonlinear time series capability of `Stata` beyond GARCH.

For time series models with nonnormal residuals, a time series bootstrap could provide empirical standard errors and confidence intervals. For many time series models, the block bootstrap might be helpful. One merely needs to determine the length of the block and the number

of blocks sampled should be a multiple of the block length. For GARCH models, the wild bootstrap, time consuming though it might be, could be helpful.

What needs refinement

Stata by default should automatically construct a log file of the session. We should be prompted for a name and the log file can should be saved with that name. When the next run is executed, the former log file should be backed up automatically if it is replaced by the newer one. A version number can be attached indicating which is the latest version. Responsible replicable research requires that log files be maintained. The log file should automatically be saved by a name given by the user unless he disables this option for the next session.

ARIMA models should automatically output information criteria. These should include the Akaike, Schwartz, Hannan–Quinn, and corrected AIC of Tsay and Hurvich. Moreover, they should also be available for customized user programming.

ARCH and GARCH models testing leverage effects should upon request graphically output the news impact curve rather than require us to program it.

Stata should automatically keep a record of the model parameter estimates and information criteria of each model run in a session. We should be able to review the model coefficients and model fit criteria at will. We should be able to automatically recall the best model for re-analysis or modification at will. When different models are being compared, these criteria can be used by which to compare the models. This model and information criteria should be stored in a file capable of being saved and merged with other such files from other sessions for more global comparisons. The parameterization of the related models should be retrievable once the optimal information criterion is found from these global comparisons.

Stata should also have the capability of out-of-sample forecast evaluation for ARIMA and GARCH models. Using such criteria as mean absolute error, mean absolute percentage error, median absolute percentage error, relative absolute error, and relative absolute percentage error against naive forecasts of the last value carried forward, we should be given the choice of criteria to be applied within a user-defined out-of-sample time period. For GARCH models the user should be able to select one or more options from a slightly different set of criteria. In addition to the aforementioned criteria, we should be able to include the mean logarithmic absolute error and the heteroskedastically adjusted mean square error should be included in his evaluation. Like the information criteria logs, these forecast evaluation logs for each session should be capable of being saved and merged with those from other sessions. In this way, models can be evaluated with respect to their predictive accuracy as well. It would be nice if the associated parameters could be saved so they could be recalled once the optimum model is identified.

Dynamic regression models in Stata are very easy to program. Their impulse response functions are also available. Nevertheless, it would be helpful if we could design a rational polynomial *transfer function* for an ARIMA model. This transfer function is a rational lagged polynomial coefficient of a stochastic regressor. Provision for these transfer functions should be built into the ARIMA command (Pankratz 1991).

Recapitulation

In sum, Stata 10 is a very well-designed, powerful, versatile, accurate, and easily customizable general purpose time series and forecasting package. The variety and capability of the time series commands offered by Stata make this a very substantial offering. With the vast array of commands for estimation and robustness, Stata is one of the best deals for the money. With the superb support offered by Stata, it is no wonder that the user community is rapidly growing around the world.

References

Baillie RT, Bollerslev T, Mikkelsen HO (1996). "Fractionally Integrated Generalized Conditional Heteroskedasticity." *Journal of Econometrics*, **74**, 3–30.

Baum CF (2004). "A Review of Stata 8.1 and its Time Series Capabilities." *International Journal of Forecasting*, **20**, 151–161.

Bollerslev T, Mikkelsen HO (1996). "Modeling and Pricing Long Memory in Stock Market Volatility." *Journal of Econometrics*, **73**, 151–184.

Box GEP, Jenkins G (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, Oakland, CA.

Chavez-Demoulin V, Embrechts P (2004). "Smooth Extremal Models in Finance and Insurance." *Journal of Risk and Insurance*, **71**(2), 183–199.

Chung CF (1999). "Estimating the Fractionally Integrated GARCH Model." *Technical report*, National Taiwan University.

Davidson J (2004). "Moment and Memory Properties of Linear Conditional Heteroskedasticity Models." *Journal of Business & Economic Statistics*, **22**, 16–29.

Durbin J, Koopman SJ (2000). *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford.

Engle RF (2002). "Dynamic Conditional Correlation: A Simple Class of Multivariate GARCH Models." *Journal of Business & Economic Statistics*, **20**, 339–350.

Engle RF, Kroner FK (1995). "Multivariate Simultaneous Generalized ARCH." *Econometric Theory*, **11**, 122–150.

Engle RF, Ng V (1995). *Measuring and Testing the Impact of News on Volatility*, volume ARCH: Selected Readings. Oxford University Press, Oxford.

Engle RF, Russell J (1998). "Autoregressive Conditional Heteroskedastic Duration Models: A New Model for Irregularly Spaced Transaction Data." *Econometrica*, **66**, 1127–1162.

Gardiner E (2005). "Exponential Smoothing: The State of the Art, Part II." URL http://www.bauer.uh.edu/gardner/Exponential%20Smoothing.pdf.

Glosten L, Jagannathan R, Runkle D (1993). "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance*, **48**(5), 1779–1801.

Hull J, White A (1999). "Incorporating Volatility Updating into Historical Simulation Method for Value at Risk." *Journal of Risk*, **1**, 5–19.

Jann B (2007). "From Estimation Output to Document Tables: A Long Way Made Short." URL http://ideas.repec.org/p/boc/asug07/3.html.

King JL (2001). *Operational Risk: Measurement and Modeling.* Wiley, New York.

Longley JW (1967). "An Appraisal of Least Squares Programs for the Electronic Computer from the Viewpoint of the User." *Journal of the American Statistical Association*, **62**, 819–841.

Lütkepohl H (1993). *Introduction to Multiple Time Series Analysis.* Springer-Verlag, New York, 2nd edition.

Marrison C (2002). *The Fundamentals of Risk Measurement.* McGraw-Hill, New York.

National Institute of Standards and Technology (2003). "NIST(b): Data Source for Exponential Smoothing." URL http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm.

Negiz A (2003). "NIST(c): Negiz Data Source for Nonseasonal ARIMA Test." URL http://www.itl.nist.gov/div898/handbook/pmc/section6/pmc621.htm.

Palandri A (2004). "Multivariate GARCH Models: Inference and Evaluation." URL http://www.econ.duke.edu/~get/finlunch/past-semesters/finlunch-documents/2004-02-16b.pdf.

Pankratz A (1991). *Forecasting with Dynamic Regression Models.* Wiley, New York.

SAS Institute Inc (2003). *SAS/STAT Software, Version 9.1.* SAS Institute Inc., Cary, NC. URL http://www.sas.com/.

Shenstone L, Hyndman R (2003). "Stochastic Models Underlying Croston's Method for Intermittent Demand Forecasting." URL http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/2003/wp1-03.pdf.

SPSS Inc (2006). *SPSS for Windows, Release 15.* SPSS Inc., Chicago, IL. URL http://www.spss.com/.

StataCorp (2007). *Stata Statistical Software: Release 10.* StataCorp LP, College Station, TX. URL http://www.stata.com/.

Tsay R (2001). *Nonlinear Time Series Models: Testing and Applications*, volume A Course in Time Series Analysis. Wiley, New York.

Tse YK, Tsui AKC (2002). "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time Varying Correlations." *Journal of Business & Economic Statistics*, **20**, 351–362.

Reviewer:

Robert Alan Yaffee
Shirley M. Ehrenkranz School of Social Work
New York University
New York, NY, United States of America
E-mail: bob.yaffee@nyu.edu