INSIGHTFUL SOFTWARE

Modeling Financial Time Series with S-PLUS® and S+FinMetrics™

by Eric Zivot

Analyzing, modeling and predicting financial time series are necessary tasks for developing successful trading strategies. However, these tasks can be daunting. There are many steps to the process, and having the right software tools maximizes the potential for success. S-PLUS, and the add-on module S+FinMetrics, are ideal software tools for statistical modelling of financial time series, and rapid prototyping and backtesting of trading strategies.

The S-PLUS environment is the premier solution for exploratory data analysis and statistical modelling. It has point-and-click functionality to import various data formats, selected statistical functions and graphics. It also has over 4,200 data analysis functions, including the most comprehensive set of robust and modern methods.

However, the real power of S-PLUS for analyzing financial data and building trading strategies comes from its flexible and powerful award-winning object oriented matrix programming language and graphics facilities. S+FinMetrics is an add-on module to S-PLUS with over 500 functions for the econometric modeling and prediction of economic and financial time series that exposes the power of S-PLUS for modeling univariate and multivariate financial time series.

I have been an S-PLUS user for over five years. I use it in my research and my teaching. Many of my Ph.D students who go into the finance industry use S-PLUS as their primary tool for developing, testing and deploying trading strategies. In this article, I will illustrate how one of my students and I recently used S-PLUS and S+FinMetrics in a research project related to modeling and forecasting high frequency intra-day foreign exchange data*.

The main goal of the project was to statistically model the dynamic intraday response of an asset to new information in arbitrage linked markets. Typical arbitrage linked markets are spot and futures markets, spot and forwards markets, and stock exchanges trading identical assets. Assets traded in these markets are tied together by certain arbitrage restrictions (e.g., cost of carry relationships) which makes them well suited for use in trading strategies. In my project, I considered the behavior of the direct Euro/Yen spot rate and the implied spot rate derived from the Dollar/Yen and Euro/Dollar exchange rates.

To illustrate the dynamic response of new information with exchange rate data, Figure 1 (created with S-PLUS) presents one episode of exchange rate movements around 23:50 GMT on August 11, 2003, at which time Japan released the first GDP estimates for the second quarter of 2003. The line with squares depicts the movement of the dollar implied Euro/Yen price and the line with triangles traces the direct Euro/Yen price. The economic recovery data in Japan's GDP release →

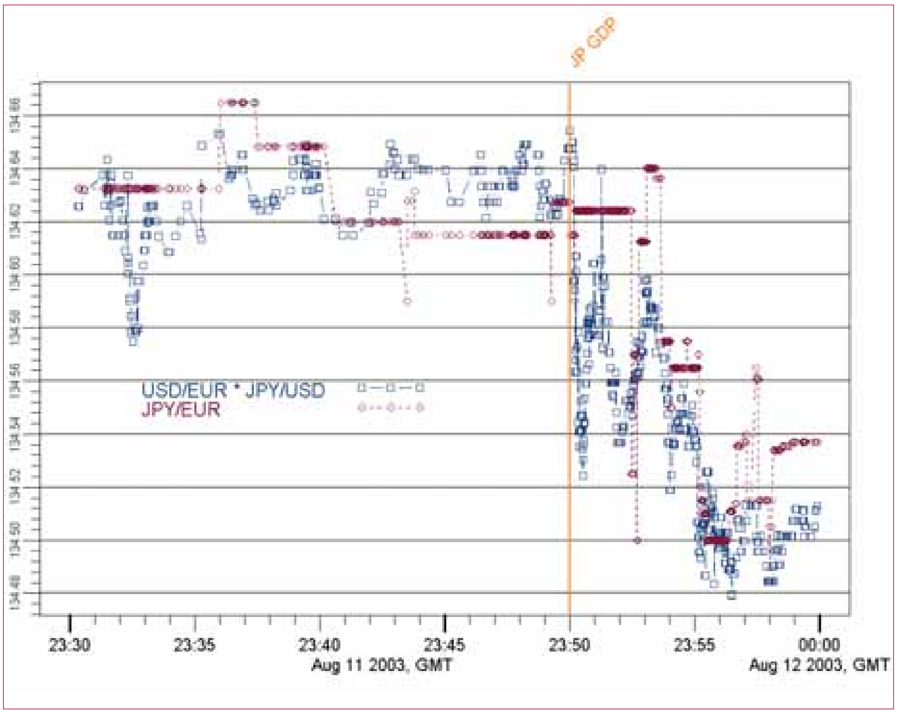

Figure 1.

caused appreciation of the Yen from 134.63 Euro/Yen to 134.50 Euro/Yen. Both the direct and implied Euro/Yen rates moved toward the new price level following the announcement, with the direct rate lagging behind the dollar implied rate by about three minutes. The figure clearly indicates an incidence of how new macroeconomic information about the Euro/Yen price is first transmitted through the dollar rates and then through the direct rates.

The data preparation and statistical modeling of the intra-day exchange rate data to describe the dynamic transmission of new information involved the following steps: (1) data cleaning and variable construction; (2) estimation and evaluation of custom statistical models; (3) creation of specialized graphics for displaying results.

Most statistical software packages are not well suited for the above analysis. In particular, they lack: flexible time and date handling facilities and graphics for irregularly spaced data; functionality for data cleaning; proper statistical models; custom programming ability; and capability for handling very large data sets. S-PLUS and S+FinMetrics, however, contain all of the tools required for the project.

Data cleaning and variable construction

The raw exchange rate data (bid and ask quotes) were captured from an electronic data feed and were recorded with time-stamps up to the second from midnight GMT. The data feed is not always reliable and so the raw data often contain errors and must be cleaned prior to analysis. The time stamps for different exchange rates are generally not the same, and for the statistical analysis, all exchange rates must be aligned to the same time clock.

Additionally, the foreign exchange market is a global twenty-four hour market with three major trading centers (Tokyo, London and New York), which exhibits complicated intra-day seasonal patterns as trading passes from one market to another. Furthermore, the market is highly active except for the weekend period Friday 22:00 GMT to Sunday 22:00 GMT. Observations over the weekend period behave differently and should be modelled separately.

S-PLUS contains a powerful and flexible set of tools for representing, manipulating and graphing irregularly spaced time series data. As S-PLUS is an object-oriented language, it has a special class for representing general time series data. For example, the first five quotes for the Euro/Dollar exchange rate on March 11, 2001 are represented as:

```
> eurusd.ts[1:5,]
          Positions    Bid      Ask     Institution
3/11/01   22:01:35    0.9326   0.9330     ONEC
3/11/01   22:01:37    0.9326   0.9331     AREX
3/11/01   22:09:34    0.9326   0.9331     NWHK
3/11/01   22:09:36    0.9327   0.9332     AREX
3/11/01   22:11:08    0.9322   0.9327     NWHK
```

The time series class allows for flexible date arithmetic and manipulation. Important information such as time zone and daylight savings time conventions are also easily incorporated into the data.

Once data are represented as a time series object, then built-in S-PLUS functions can be used to perform a variety of tasks such as aligning the irregularly spaced data to a regularly spaced time clock (e.g. every 5 seconds), interpolating missing values, aggregating the data over fixed or moving windows, and computing technical analysis indicators (see Box 1).

In preparing the data for the project, I found that it was convenient to create a library of new S-PLUS functions to perform specialized tasks. This was easily done using the S-PLUS programming language. The resulting library of functions can be shared with other S-PLUS users.

Estimation and evaluation of custom statistical models

The statistical modeling of the system of exchange rate data involved a number of sophisticated econometric models including vector autoregressive models, cointegrated vector error correction models and linear state space models. Fitting, forecasting and simulation functions for all of these models are available in S+FinMetrics (see Box 2).

Creation of specialized graphics

The dynamic response of exchange rate prices to new information was captured by simulating the arrival of news and then graphically tracing out the future response of the exchange rates from the estimated statistical model. The model is constructed such that that the arrival of news that leads to a one basis point increase in the fundamental value of the Euro/Yen exchange rate should eventually increase both the direct and Dollar implied Euro/Yen rates by one basis point.

The construction of the dynamic response path involved some custom programming using the state space modeling and Kalman filtering tools in S+FinMetrics as well as the S-PLUS Trellis graphics functions. Figure 3 illustrates the end result for the Euro/Yen exchange rate during the Tokyo trading hours. The solid blue line represents the estimated response of a one basis point increase in fundamental value, and the dotted red lines represent 95 per cent confidence band constructed from a bootstrapping procedure. The upper panel shows the dynamic response of the Euro/Yen exchange rate implied by the US Dollar rate to new information, and the lower panel shows the dynamic response of the direct Euro/Yen rate to the same new information.

The figure clearly shows that new

Examples of S-PLUS functions

Aligning the irregularly spaced tick-by-tick data to a regularly spaced 5 second time clock can be done with the S-PLUS align function:

```
> eurusd.5sec.ts = align(eurusd.ts, by="seconds", k.by=5,
                          how="nearest")
> eurusd.5sec.ts[1:5, ]
          Positions    Bid      Ask    Institution
3/11/01   22:01:35   0.9326   0.9330      ONEC
3/11/01   22:01:40   0.9326   0.9331      AREX
3/11/01   22:01:45   0.9326   0.9331      AREX
3/11/01   22:01:50   0.9326   0.9331      AREX
3/11/01   22:01:55   0.9326   0.9331      AREX
```

The argument "`how=nearest`" is used to specify the tick nearest to the 5 second clock as the observation at that time.

Intra-day seasonal patterns may be captured using built-in spline and non-parametric smoothing functions. Outliers and other data errors may be effectively identified using the unique collection of robust statistical methods. Custom time series graphics, like the one shown in Figure 1, are easily produced. For example, the S-PLUS code to produce Figure 1 is:

```
>plot(exchdata.ts[timeEvent("8/12/2003 23:30:00",
+                            "8/12/2003 23:59:59"), ],
+       reference.grid=F,
+       plot.args=list(type="b", pch=c(0, 5)))
> legend(0.2, 134.56, legend=c("USD/EUR * JPY/USD",
+                               "JPY/USD"))
```

The `plot` function is a generic function which invokes the appropriate plotting method for the time series object `exchdata.ts`.

Box 1.

Examples of S+FinMetrics commands

Fitting a single lag vector error correction model with a pre-specified cointegrating vector and a restricted trend to a bivariate system of exchange rates may be done with the single command

```
vecm.fit = VECM(exchdata.ts,
                coint.vec=as.matrix(c(1,-1,0)),
                 lags=1, trend="rc")
```

Figure 2 shows the h-step ahead predictions of log-returns, with 95 per cent confidence bands, from the fitted vector error correction model computed using the commands

```
vecm.fcst = predict(vecm.fit,
n.prdict=12)
plot(vecm.fcst, xold=diff(exchdata.ts),
n.old=12)
```

Additional statistical and graphical diagnostics are readily produced to evaluate the fit of these models using the generic summary and plot methods

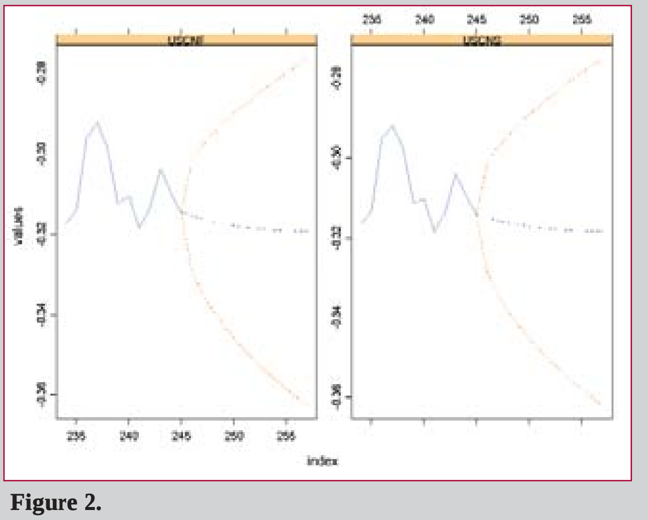

Figure 2.

Box 2.

→

"THE [SOFTWARE] CLEARLY SHOWS THAT NEW INFORMATION IS INCORPORATED MORE QUICKLY AND EFFICIENTLY INTO THE DOLLAR IMPLIED RATE THAN INTO THE DIRECT RATE."

- ERIC ZIVOT

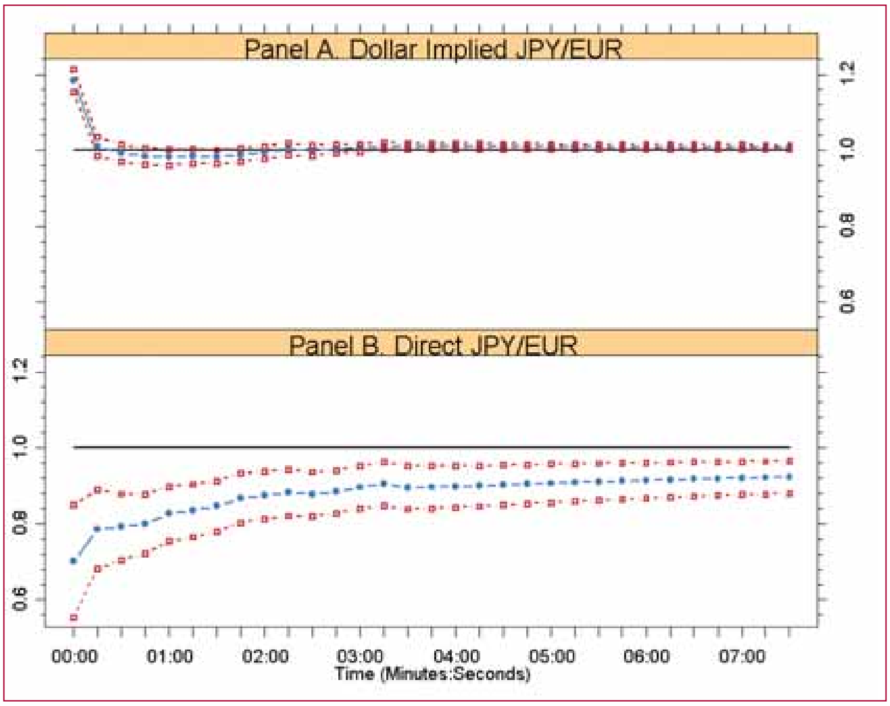

Figure 3.

information is incorporated more quickly and efficiently into the Dollar implied rate than into the direct rate. The Dollar implied rate fully incorporates the information in about two and one half minutes whereas the direct rate takes more than seven minutes.

The figure clearly shows that new information is incorporated more quickly and efficiently into the Dollar implied rate than into the direct rate. The Dollar implied rate fully incorporates the information in about two and one half minutes whereas the direct rate takes more than seven minutes.

Summary

S-PLUS and S+FinMetrics are ideal tools for modeling financial time series and for developing, evaluating and deploying quantitative trading strategies.

Eric Zivot is Associate Professor and Gary Waterman Distinguished Scholar in the Economics Department at the University of Washington. He is co-author of the book Modeling Financial Time Series Using S-PLUS.

* See "The Dynamics of Price Discovery" by Bincheng Yan and Eric Zivot, available at http://faculty.washington.edu/ezivot/ezresearch.htm