APPENDIX B

Forecasting Software

Good forecasting software is essential for both forecasting practitioners and students. The history of forecasting is to a certain extent written by those who have developed user friendly software packages early in the dissemination of a new method. For example, the early development of Box-Jenkins ARIMA software (Chapter 6) ensured rapid adoption while the paucity of state space software (Chapter 5) limited its adoption until the development of open source software in R (Hyndman and Khandakar, 2008; Hyndman, 2017). Kusters, McCullough and Bell (2006) have looked at the history of forecasting software and its progress as it has been affected by computer technology. They draw a distinction between research-oriented software and software for operational organizational planning. The former requires a more sophisticated analysis of data while the latter often relies on simple methods but needs to deal with many time series simultaneously. "Business software" also must be standardized so that it can be implemented consistently across companies or different units within a company. A further requirement is that it is integrated (or at least compatible) with other corporate systems such as ERP (Enterprise Resource Planning). In this book, we have considered both types of problem (in Chapter 12). As a consequence of these diverse requirements we have chosen to use a multiplicity of software products rather than focus on a single product. There are two further reasons for this decision: (i) no single software product includes all the features that a user might want, different users have different requirements; and (ii) different universities and organizations have different software available to them. So, we hope the book stands alone in supporting a range of excellent software products.

Since the first edition, *R* has become more widely available: *R* is a free open source and flexible language dedicated to statistical computing. It has a library of functions that are constantly being enhanced with useful features, including many that focus on forecasting. We have therefore chosen to illustrate some of the examples in the book with *R*, but more importantly, we provide in Appendix C; Forecasting in *R*: Tutorial and Examples.

Software for forecasting comes in a variety of forms, from the ubiquitous Excel though to specialist software incorporating a single method. In surveys of practitioners, *Excel* is usually named as the most commonly used forecasting package (Sanders and Manrodt, 2003; Weller and Crone, 2012) although it is best thought of a data handling and analysis program, rather than providing a forecasting package, as its forecasting routines remain limited, even in its latest 2016 incarnation. In addition, its statistical features have weaknesses (McCullough, 2000, 2006 and McCullough and Heiser, 2008). On the book's website, we provide *Excel* macros for simple exponential smoothing and Holt-Winters trend and seasonal smoothing as well as trend curve modeling.

Software is usefully broken down into a number of different categories that reflect both the specialization for which it is designed and the focus of its application. The first category of forecasting system we have used is the general purpose statistical package. We have used three of the most widely used in this text, *Minitab*, *SAS*, and *SPSS*. (In addition, *Systat*, *Stata*, and *Statgraphics* fall into this category.) These packages include a wide range of statistical techniques supplemented by forecasting routines. Sometimes there is more specialist add-in (costing more) which enhances the basic system. Typically, the base general purpose packages do not include the full range of forecasting methods and the additional features that a more advanced user might require.

The second software category we consider is the specialist product which focuses on a single technique (although some simple additional techniques may be included). These include *Autobox* (which is an automatic Box-Jenkins ARIMA modeling and forecasting package, see Chapter 6), various specialist neural network packages (such as *Alyuda*), state space modeling (in *STAMP*, a part of the *OxMetrics* range). In addition, there are the seasonal decomposition methods discussed in Section 4.5, *X-13-ARIMA* and *Tramo-Seats*. Within the area of specialist software, the particular area of data mining (now often called predictive analytics) deserves extra attention. This software comprises a range of methods suitable for predicting individual consumer behavior (see the discussion of classification methods in Chapter 10). Typically, methods include logistic regression, predictive classification including CHAID and cross-sectional neural nets. The general purpose statistical packages such as *SAS* and *SPSS* include these methods, but sometimes at an additional purchase cost. We have used *SAS* (in Section 12.5) to give an example of how all these methods can be used together to predict loan default. *Weka* provides free, well-regarded software: other alternatives are *XLMiner*, *NeuroIntelligence*, and *Neurosolutions* (and may be used as *Excel* add-ins).

There is also a range of software primarily designed to support forecasting, for both academics but importantly, practitioners. The difference between these packages and the earlier categories is that they give more attention to measuring forecast accuracy in ways which are intuitive and accord with the key concepts of keeping a hold-out sample distinct from the in-sample fitting period, and second, may permit easy incorporation of rolling origin forecasting (see Tashman, 2000 for a full explanation of how these are best carried out). *ForecastPro* is a good example of the use of hold-out samples in that various error measures are calculated for both fixed and rolling origins. There are many packages that fall into this category which offer varying degrees of sophistication: they may focus on a particular method (e.g. *Autobox*) or may be more general (e.g. *ForecastPro*). While these are well-validated examples, others may not conform to the principles we have laid out in this book — so buyer beware must be the maxim.

Econometric packages such as *EViews*, *PcGive*, *RATS*, and *TSP* all include a wide range of regression based modeling routines. Other econometric packages are more specialist and focus on a limited set of routines, e.g. *LINDEP*, which includes a range of models suitable for classification, including the logit, a method we considered in Chapter 10.

Finally, there are programming languages such as *R*, which include a wide range of preprogrammed routines that support time series analysis, econometrics and forecasting. (*Python* is a general purpose language and does not offer the forecasting functionality of *R*.) As part of the *R* suite of programs, smoothing methods are available in packages *forecast* and *smooth*. Some of the more advanced methods and techniques discussed in the textbook, such as intermittent demand method (*tsintermittent*) and multiple temporal aggregation (*MAPA* and *thief*), as well as VAR modeling (*MTS*), amongst others are also

available. The reader is referred to *https://cran.r-project.org/view=TimeSeries* for available packages in *R* related to time series analysis, modeling and forecasting. Other programming languages also provide tool boxes which offer various time series techniques and forecasting models, e.g., *MATLAB*. *Gauss* and *OxMetrics* provide programming languages and pre-programmed routines but are more focused on econometric applications. *SAS* and other general-purpose statistical packages also provide programming languages which, like *R*, can be used to incorporate their various statistical and forecasting methods into more complex analyses.

In Table B1 we summarize a number of package types on the dimensions which we regard as important and note some associated software as examples. Undoubtedly, we have omitted some that are widely available and useful. An additional, if still incomplete resource is the review by Yurkiewicz, (2016). The headings we use speak for themselves. But there are important dimensions we have excluded from the table — whether the packages can easily import and export data and output (preferably direct into *Excel*). And second, how easy the packages are to learn and to use. The former issue of import/export used to be highly problematic but now all the packages we have used have these features. The second question of ease of learning and ease of use is another matter — but it is a subjective issue that would require a full review to justify. We will note however that not all the packages we have used can be regarded as easy to use for novices (like us)!

One category of packages is not discussed here despite its importance: forecasting packages to support supply chain demand forecasting (see Section 12.2). Such all-embracing packages include *SAP*, *Oracle*, and *JDA* as well as packages focused on particular sectors of the economy such as retail. These packages cost many thousands of dollars and their forecasting components (despite being forecasting packages in part) used to be overly limited: now some companies such as *SAP* have included many advanced modeling features in their extended range of program suites including automatic regression modeling — the base suite may however still rely on exponential smoothing. But if they are so expensive and often limited why do companies use them? The answer is straightforward — the packages include data handling facilities which are extremely flexible and meet the needs of most companies working in the supply chain. In particular, they will link to the ERP and financial systems. They also permit multiple user access. We have offered advice in section 12.2 to aid practitioners in selecting a package. There are many pitfalls, not least to assume that all such packages are well-designed from a forecasting perspective. Some are truly awful!

In summary, this appendix does not attempt to give a complete overview of the many commercial and free forecasting oriented packages that are available. Regular reviews of packages have been produced over the years by Yurkiewicz (2016). However, the dimensions we have used should give the reader some insight into what is available. They should also supply the practicing forecaster looking to review their software implementation a checklist of features to look for in a benchmarking exercise.

Table B1 Features of Selected Forecasting Software

		General purpose statistics		Forecasting	Specific technique	Econometric	Opensource	
Example software	**Excel**	**Basic versions: e.g. Minitab/ SAS/SPSS**	**Extended Versions: e.g. SAS, Stata**	**e.g. Forecast Pro, ForecastX, Peer Forecaster, Prophet, Smart Forecast**	**e.g Autobox**	**e.g. PcGive, Eviews**	**e.g. R/ MatLab**	**Note that these characteristics are illustrative only.**
	B	B		B, C	C	B	B	
Notes	(1)		(1, 4)			(6)	(7)	(1) We only include features standard in the software except where noted (2) Estimation of parameters only in Excel2016 (3) AR plus regression (4) SAS includes Enterprise Miner and Forecast Server (5) Includes Advanced Statistics module (6) PcGive Includes Stamp and Oxmetrics (7) Including only standard toolboxes (8) This can always be programmed easily **Code** B Used in this book C Participated in M3 Competition
Forecasting								
Allows selection of estimation and hold-out sample	✓		✓		✓	✓	✓	
Fixed horizon evaluations		✓	✓	✓	✓	✓	✓	
Rolling origin evaluations			?	✓	✓	✓	Programmable	
Multiple error measures	✓	Limited	✓	✓	✓	Programmable	✓	
Saving forecasts and forecast errors for further analysis	✓	✓	✓	✓	✓	✓	✓	
Model combination (8)								
Batch forecasting for multiple series			✓	✓	✓		✓	
Hierarchical forecasting			✓	✓				
Judgmental adjustment of statistical forecasts			✓	✓				
Exploratory analysis and Graphics								
Time Series Plot	✓	✓	✓	✓	✓	✓	✓	
Multiple scatterplot		✓	✓				✓	
Autocorrelation Function		✓	✓	✓	✓	✓	✓	
Partial Autocorrelation Function		✓	✓	✓	✓	✓	✓	
Stationarity tests			✓			✓	✓	
Simple Moving Averages	✓	✓	✓	✓			✓	
Transformations	✓	✓	✓	✓	✓	✓	✓	
Smoothing models and State Space								
Simple Exponential Smoothing	(2)	✓	✓	✓	✓		✓	
Linear Exponential Smoothing	(2)	✓	✓	✓	✓		✓	
Holt-Winters Seasonal	(2)	Limited	✓	✓	✓		✓	
Damped Trend			✓	✓	✓		✓	
Methods for intermittent demand			✓	✓			✓	
State space formulation and estimation			✓			✓	✓	
Automatic Model Selection			✓	✓	✓		✓	
Prediction intervals		✓	✓	✓	✓	✓	✓	
Seasonal decomposition		Limited	✓			✓	✓	
Census X-12, X-13			✓				✓	

(continues)

Table B1 Features of Selected Forecasting Software *(continued)*

		General purpose statistics		Forecasting	Specific technique	Econometric	Opensource	
Example software	**Excel**	**Basic versions: e.g. Minitab/ SAS/SPSS**	**Extended Versions: e.g. SAS, Stata**	**e.g. Forecast Pro, ForecastX, Peer Forecaster, Prophet, Smart Forecast**	**e.g Autobox**	**e.g. PcGive, Eviews**	**e.g. R/ MatLab**	**Note that these characteristics are illustrative only.**
ARIMA and ARCH/Garch								
Regular and Seasonal ARIMA		✓	✓	✓	✓	✓	✓	(1) We only include features standard in the software except where noted
Intervention Analysis & ARMAX			✓	(3)	✓	✓	✓	(2) Estimation of parameters only in Excel2016
Automatic Model Selection			✓	✓	✓	✓	✓	(3) AR plus regression
ARCH/GARCH			✓			✓	✓	(4) SAS includes Enterprise Miner and Forecast Server
Regression modelling								(5) Includes Advanced Statistics module
Multiple linear Regression	✓	✓	✓	✓	✓	✓	✓	(6) PcGive Includes Stamp and Oxmetrics
Model selection (e.g. forward selection)		✓	✓		✓		✓	(7) Including only standard toolboxes
Residuals Plot	✓	✓	✓	✓	✓	✓	✓	(8) This can always be programmed easily
Outlier/ Leverage identification		✓	✓	Limited	✓	✓	✓	
Multicollinearity statistics		✓	✓			✓	✓	**Code**
General autocorrelation tests of residuals		✓		✓	✓	✓	✓	B Used in this book
Extended range of tests for mis-specification		✓	✓				✓	C Participated in M3 Competition
Logistic Regression		✓	✓				✓	
Time-varying parameter regression			✓				✓	
Multiple equation models and VAR			✓			✓	✓	
Prediction intervals		✓	✓	✓	✓	✓	✓	
Computationally intensive methods								
Data selection - training, validation, test		Limited	✓				✓	
Classification and Regression Trees		Limited	✓				✓	
Neural Nets - cross sectional		Limited	✓				✓	
Advanced features – e.g., multiple initialisation, activation function alternatives			✓				✓	
Other								
Growth and Diffusion Curves (Bass etc)		Non-linear LS				Non-linear LS	Non-linear LS	
System features								
Programmable	✓	Limited	✓			✓	✓	
Web-based			✓				Some	
ERP linked			✓	✓	✓		✓	

References

Hyndman, R.J. (2017). Forecast: Forecasting functions for time series and linear models. *R* package version 8.1, *http://github.com/robjhyndman/forecast*.

Hyndman, R.J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for *R*. *Journal of Statistical Software*, 26 (3), 1–22.

Kusters, U., McCullough, B.D., & Bell, M. (2006). Forecasting software: Past, present and future. *International Journal of Forecasting*, 22 (3), 599–615.

McCullough, B. D. (2000). Is it safe to assume that software is accurate? *International Journal of Forecasting*, 16, 349–357.

McCullough, B.D. (2006). The unreliability of Excel's statistical procedures. *Foresight*, (Issue 3: February), 44–45.

McCullough, B. D. & Heiser, D. A. (2008). On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics and Data Analysis*, 52, 4570–4578.

Tashman, L. (2000). Out-of-sample tests of forecasting accuracy: analysis and review. *International Journal of Forecasting*, 19, 437–450.

Weller, M., and Crone, S. F. (2012). Supply Chain Forecasting: Best Practices and Benchmarking Study. Lancaster Centre for Forecasting White Paper: Available on: *http://goo.gl/MPbAjz*.

Yurkiewicz. J. (2016). Forecasting: What can you predict for me? (Software Survey), *ORMS Today*, June 2016, vol. 43:3. Available online.