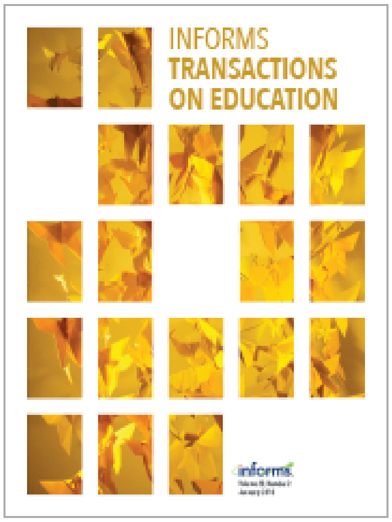
This article was downloaded by: [159.226.100.198] On: 15 July 2019, At: 20:22
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

SCOTS: The Searchable Collection of Time Series

Constance H. McLaren, Bruce J. McLaren

To cite this article:
Constance H. McLaren, Bruce J. McLaren (2018) SCOTS: The Searchable Collection of Time Series. INFORMS Transactions on Education ():. https://doi.org/10.1287/ited.2017.0188

Full terms and conditions of use: https://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2017, The Author(s)

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

informs
http://pubsonline.informs.org/journal/ited/

INFORMS TRANSACTIONS ON EDUCATION
Vol. 19, No. 1, September 2018, pp. 12–22
ISSN 1532-0545 (online)

SCOTS: The Searchable Collection of Time Series

Constance H. McLaren,[a] Bruce J. McLaren[a]

[a] Scott College of Business, Indiana State University, Terre Haute, Indiana 47809

Contact: Constance.mclaren@indstate.edu, http://orcid.org/0000-0003-4577-4061 (CHM); bruce.mclaren@indstate.edu (BJM)

Received: December 26, 2016
Revised: May 15, 2017; July 16, 2017
Accepted: July 17, 2017
Published Online in Articles in Advance: December 12, 2017

https://doi.org/10.1287/ited.2017.0188

Copyright: © 2017 The Author(s)

Abstract. Instructors searching for interesting time series data to use for class illustrations or assignments can draw from the data files that accompany their text books. They can also search for data from annual reports, trade organizations, government entities, and other public sources. However, the amount of data accompanying textbooks is limited, and determining what public data might be available and then finding it is not simple, making it time consuming to find new examples for large classes or multiple offerings of a course. The search becomes more complicated for the instructor who needs to find data with specific types of trend or seasonality to illustrate concepts for students. This article presents the SCOTS master table, a searchable Excel file that provides information about and links to a curated collection of real time series data files stored in Excel format. Instructors can easily filter the list to find a time series with the kinds of features they want and download the Excel file to use in class. Many of the data files in this collection have been used for class assignments, and examples of assignment questions are included in the article.

Open Access Statement: This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License. You are free to download this work and share with others commercially or noncommercially, but cannot change in any way, and you must attribute this work as "*INFORMS Transaction on Education*. Copyright © 2017 The Author(s). https://doi.org/10.1287/ited.2017.0188, used under a Creative Commons Attribution License: https://creativecommons.org/licenses/by-nd/4.0/."

Supplemental Material: The supplement file is available at http://pubsonline.informs.org/doi/suppl/10.1287/ited.2017.0188.

Keywords: forecasting • data • trend • seasonality

1. Introduction

Those who teach forecasting concepts, whether as part of an analytics class or a dedicated class in forecasting, soon find that locating useful forecasting data can be difficult. Although many time series are available, the issue for instructors is how to efficiently find those that are suitable for their purposes. Textbooks include time series data files for illustration or homework problems, and thousands of time series are published in annual reports and by government agencies and industry organizations. Statistical collections are also available, such as the files used for the M competitions (see https://forecasters.org/resources/time-series-data/m-competition/, accessed October 11, 2017) and those at the Machine Learning Repository (http://archive.ics.uci.edu/ml/, accessed October 11, 2017), Nonetheless, even with these resources we have found that it is challenging and time consuming to find the right time series to illustrate specific concepts or to supply enough fresh, real data for homework, projects, and exams. For example

- You are introducing a unit on forecasting and would like to have students view a time series plot to discover the standard features of a time series, i.e., trend, seasonality, cycles, and randomness. A time series with linear trend, quarterly seasonality, little evidence of cyclical behavior, and a small amount of randomness would be a good place to start.
- You are emphasizing seasonal fluctuations and would like to show quarterly and monthly seasonality to help students understand that the number of seasons is the number of observations between repetitions.
- In a unit on trend, you would like to show a time series that has nonlinear trend but does not exhibit seasonality.
- For a class project you would like to assign each team a different trend-seasonal time series, but you would like all the project data sets to be of similar complexity.

Finding the time series to fit these qualifications, particularly those that reflect real and meaningful data, takes time that an instructor may be unable to allocate. The Searchable Collection of Time Series (SCOTS) master table (included with this article) provides links to download files from 262 time series that can be filtered to find particular types of trend, seasonality, and cyclical behavior.

The need for such a time series collection became evident when a colleague said, "You teach forecasting. I need some new examples to introduce seasonality tomorrow. Do you have any good ones?" The answer to that question was yes. Through years of teaching business statistics, operations management, and a forecasting elective course, we have found and collected a wide variety of data sets for class assignments. The SCOTS master table gathers this collection in a way that allows easy filtering to help instructors find exactly the real time series data they need. Each of the time series Excel files it references is available to ITE readers for download by following the link in the SCOTS master table. It may be beneficial for readers to have the table open as they follow this article.

1.1. Source

The collection began as we accumulated time series that were particularly applicable for class examples. Many of these were chosen from files in the Time Series Data Library, established by Professor Rob Hyndman (http://data.is/TSDLdemo, accessed October 11, 2017). These files, many of which dealt with data from Australian sources, were organized by application and included files on areas such as production, agriculture, hydrology, and tourism. Files that we found to be particularly intriguing for our undergraduate and graduate business majors were collected and saved to use in class. The Time Series Data Library is now hosted at Datamarket.com (https://datamarket.com/data/list/?q=provider:tsdl, accessed October 11, 2017) and files are still grouped into similar application categories. DataMarket allows a user to search by tag (e.g., "environment") and to narrow the search to monthly, yearly or quarterly data. However, we have found that it is still a time-consuming process of trial and error to locate time series with the desired trend and/or seasonality. A DataMarket time series designated as monthly, for example, does not necessarily have monthly seasonality. SCOTS contains approximately 200 of the original 800 Time Series Data Library files.

A second source of SCOTS time series data comes from observations we have collected for use in class and those collected by our students. Projects in our forecasting elective course require students to collect and analyze their own time series data. Students record some real measure at regular intervals, write about what they see in the time series, recommend appropriate forecasting methods, and then in some cases conduct a tournament to see which standard forecasting methods performed well for both in-sample and hold-out observations. Over 50 of these student-generated time series are accessible from the SCOTS master file.

Finally, SCOTS includes examples of sources of publicly available time series that we have found to be particularly interesting to students. The U.S. National Park Service (https://irma.nps.gov/Stats/, accessed October 11, 2017) publishes monthly visitor traffic information, and SCOTS includes the traffic count for one entrance of one park as an example. Movie box office data is published by http://www.the-numbers.com, and SCOTS includes the daily and weekend box office receipts for *Star Wars VII* as examples. The U.S. Bureau of Transportation Statistics (http://www.rita.dot.gov/bts/, accessed October 11, 2017) lets users configure queries to create a wide variety of tables; we have included time series for domestic and international passenger enplanements as examples. Instructors could access these sites to find data for other parks, movies or transportation modes that are appropriate for their students or geographic locale. We have found that these three public online sources also lend themselves to class project assignments that require students to search for and analyze a time series that is interesting to them. Assignment questions used for the national park data appear in the appendix.

1.2. Outline

The article begins with justification for the use of real data for instruction and then describes how the characteristics of each time series were determined. Instructions, with examples, are provided for using the SCOTS master table to extract those time series that fit an instructor's needs. The reader is then given examples of class activities and assignments that incorporate the time series data. By providing this curated collection, SCOTS enables instructors to spend their time planning how to use the time series rather than carrying out an extensive search for one that is appropriate.

2. Background

2.1. The GAISE Recommendations for the Use of Real Data

Publication of the original Guidelines for Assessment and Instruction in Statistics Education (GAISE) College (Aliaga et al. 2005) and Pre-K–12 (Franklin et al. 2007) Reports laid the foundation for changes in statistics education that would make instruction "more modern, engaging, and authentic" (Zieffler et al. 2012, p. 2). Since that time, instructors and researchers have welcomed the reports' conclusions. Both publications and classroom practices have embraced the reports' recommendations, including the use of real data for instruction. Authors have called for the GAISE recommendations to be followed for analytics education in pre-K-12 and post-secondary settings in the United States and abroad (Groth 2008, Metz 2010, Bargagliotti 2012, Chen et al. 2012, Woodard 2012, Franklin et al. 2013, Wall et al. 2014), and in areas as diverse as nursing (Hayat 2014), science (Watson 2014), medicine (Nowacki 2015), chemistry (Pienta 2013) and public affairs (Ozturk 2012).

Ongoing work on the Statistics Teaching Inventory (STI) (Zieffler et al. 2012), an instrument to understand instructors' incorporation of the GAISE recommendations, includes survey questions about the use of real data in the classroom in its Teaching Practice and

Teaching Beliefs sections. For an excellent overview of curriculum developments since the initial GAISE reports were published, see Horton (2015) and Horton and Hardin (2015).

In 2012, a committee was established to consider updating the GAISE College Report, noting that, among other developments, changes in technology, classroom delivery methods, and data science necessitated a review of the report's recommendations. In its draft report (Carver et al. 2016) the committee continues to support the use of real data with context and purpose, active learning, and the use of technology to explore concepts and analyze data. It notes that "technology should also aid faculty and students by facilitating access to real (and often large) data sets, fostering active learning, and embedding assessment into course activities" (Carver et al. 2016, p. 20). It seems only fitting, therefore, that we make it easier for instructors to access real, meaningful data for students to use in time series modeling.

2.2. Determination of Time Series Features

The SCOTS master table (see the supplemental material) is an Excel file that provides a link to each of the 262 time series files and summary information about the trend, seasonality, and cycles that may be present. Figure 1 shows an abbreviated view of the SCOTS master table.

When the link in a particular row is clicked, the actual time series file opens in Excel, as illustrated in Figures 2 and 3 for the quarterly chocolate (AusChocQ) time series data file. The first tab for each of the time series files provides summary information and the second has the time series data. By viewing the time series plot on the first tab, an instructor can take a quick look at the behavior of the time series; the additional information on the tab will help the instructor decide whether this time series is appropriate to incorporate in class materials. If it is, the instructor can access the observations on the second tab and save the data as an xlsx or csv file or copy it to whatever application is being used.

To make determinations for the presence and type of trend, seasonality, and cycles in each of the 262 time series, we relied on visual evidence and statistical tests. The plot for each time series was reviewed as a first indicator of the features that might be present.

2.2.1. Trend. If present, trend is characterized in SCOTS as linear, nonlinear or piecewise. In general, the determination of trend was made after a visual inspection. Although we considered testing for the significance of the slope of a simple linear trend line, we discarded that idea because, as is so often the case with real data, many of the time series had changes in behavior over their observations. We concluded that verbal descriptors rather than statistical measurements were more appropriate to describe the overall trend. Figure 4 shows plots from four of the time series to illustrate the four classifications for trend.

2.2.2. Seasonality. As was the case with trend, seasonality was often visible from the time series plot. When seasonality is present, the number shown in the seasonality description is the number of seasons. However, when a visual inspection did not reveal a regular repeating pattern for those daily, monthly or quarterly time series where logic might predict such a pattern, we used Minitab to run statistical tests for the significance of autocorrelation and partial autocorrelation before drawing a conclusion. As an example, consider the Steps15 time series, the daily steps during 2015 recorded by a Fitbit user. One might expect this daily data to show significant autocorrelation at the seasonal lags (multiples of 7), but that was not the case. As Figure 5 indicates, although there were some significant autocorrelations and partial autocorrelations (indicated by spikes that extend beyond the red critical values), there was no indication of seasonality by day of the week for this time series.

2.2.3. Cycles. Cycles, those irregular length periods of similar movement, appear in many of the time series. Although cycles generally cover multiple years and reflect the business cycle, we indicated the presence of cycles if the time series had at least two different sequences of similar behavior. For example, the Crest time series shown in Figure 6 has a significant jump between the first and second halves of its 276 observations.

2.2.4. Notes. The Notes field in the SCOTS master file provides information about suggested uses for many of the time series. An instructor who wishes to teach classical seasonal decomposition, for example, might filter the trend and seasonality columns to locate time series

Figure 1. The SCOTS Master Table Gives Links and Information About Each Time Series

File N	Filename	Link	Period	Obser	Source	Short description	Variables	Trend	Seasonality	Cycles	Notes	StartYear	EndYear
1	AccidentalDeaths	http://	Monthly	72	Hyndman	Accidental deaths i	1	No	12	No		1973	1978
2	Accidents	http://	Monthly	192	Hyndman	Number of deaths a	2	No	12	Yes	Two scale	1969	1984
3	Active	http://	Daily	69	McLaren	Active minutes per	1	No	7	No	Contains	2016	2016
4	Ad&Sales	http://	Monthly	36	Hyndman	Advertising and sale	2	No	12	No			
5	AircraftMiles	http://	Monthly	96	Hyndman	U.S. airlines: month	1	Linear	12	No		1963	1970
6	AirPassMiles	http://	Monthly	216	Hyndman	Monthly U.S air pas	1	Nonlinear	12	No		1960	1977
7	Applications	http://	Weekday	38	McLaren	Applications receive	1	No	No	No		2009	2009

Figure 2. The First Tab of Each File Provides Information About the Time Series

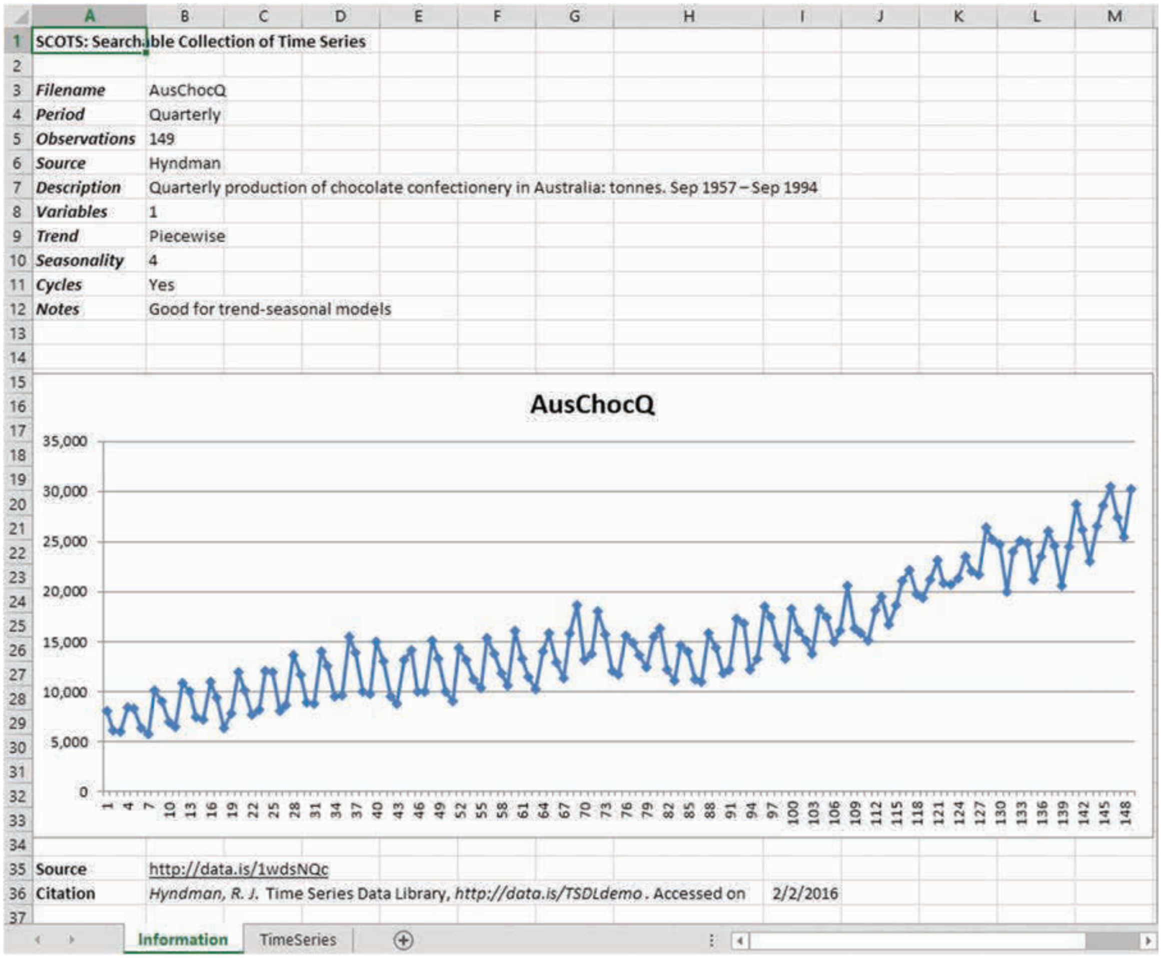


Figure 3. The Time Series Data Appears on the Second Tab of Each File

	A	B	C
1	Observation	Production	Quarter
2	1	7992	1957 Q3
3	2	6114	1957 Q4
4	3	5965	1958 Q1
5	4	8460	1958 Q2
6	5	8323	1958 Q3
7	6	6333	1958 Q4
8	7	5675	1959 Q1
9	8	10090	1959 Q2
10	9	9035	1959 Q3
11	10	6976	1959 Q4

with linear trend and the desired number of seasons, yielding a wide list of possibilities. The instructor could then filter on the Notes field to find time series that have been class tested or would be particularly appropriate for student exploration. Time series with missing values, negative values or values of zero are also noted. Instructors who use time series with values that are negative or zero will need to provide additional instruction about calculation of the mean absolute percentage error (MAPE).

2.2.5. StartYear and EndYear. To make it easier for instructors to sort and find recent time series or to determine how many years are included in the time series, columns with the starting and ending years of the observation are included where available. Entries are blank where the date is unavailable. Note that several of the time series include vast amounts of historical information; the limberpine time series dates from the year 837, and the wages time series from the year 1260.

3. Using SCOTS

The SCOTS master table lists each of the time series data files along with its link and characteristics. As shown by the excerpt in Figure 1, each row represents one time series data file. Primary fields in the SCOTS table are filename, clickable link to download the Excel data file, period of the time series (daily, monthly, quarterly, annual, etc.), number of observations, short description, number of variables in the time series (1–5), seasonality, trend, cycles, notes, and the starting and ending years. Source is given as Hyndman

Figure 4. There Are Four Different Classifications of Trend

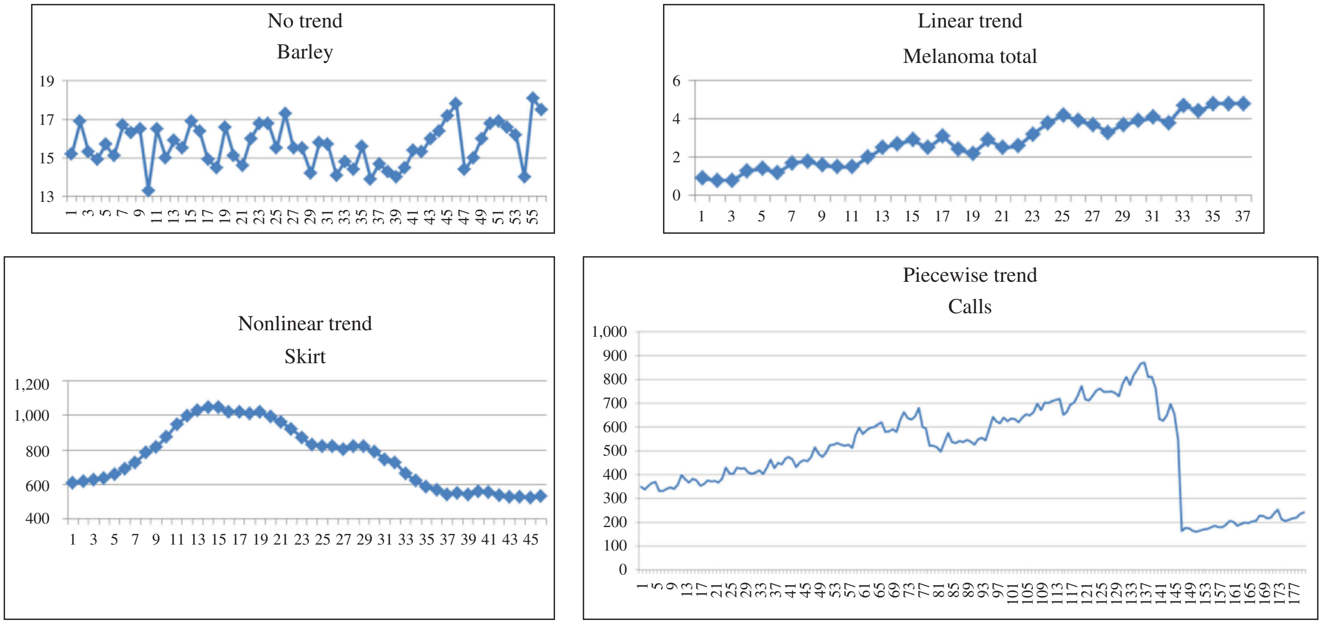


Figure 5. Autocorrelation Was Used to Test for Seasonality

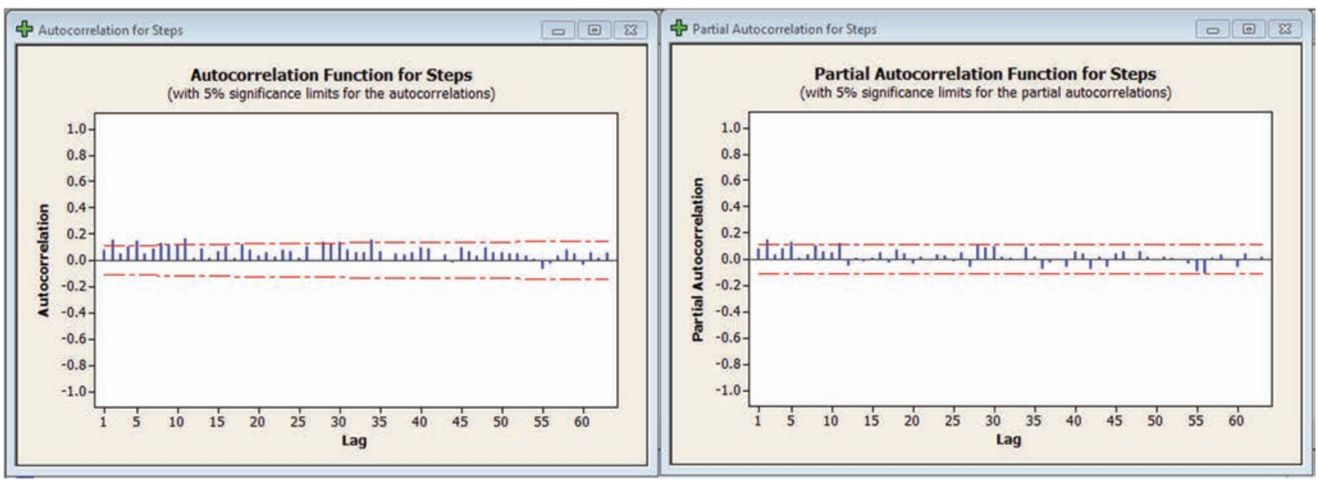


Figure 6. The Crest Time Series Has at Least Two Cycles

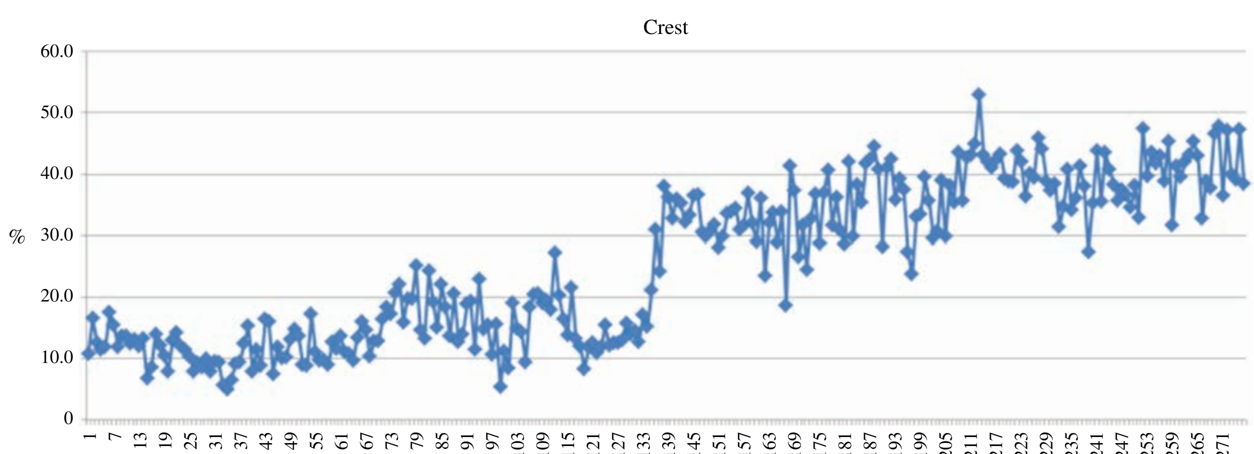
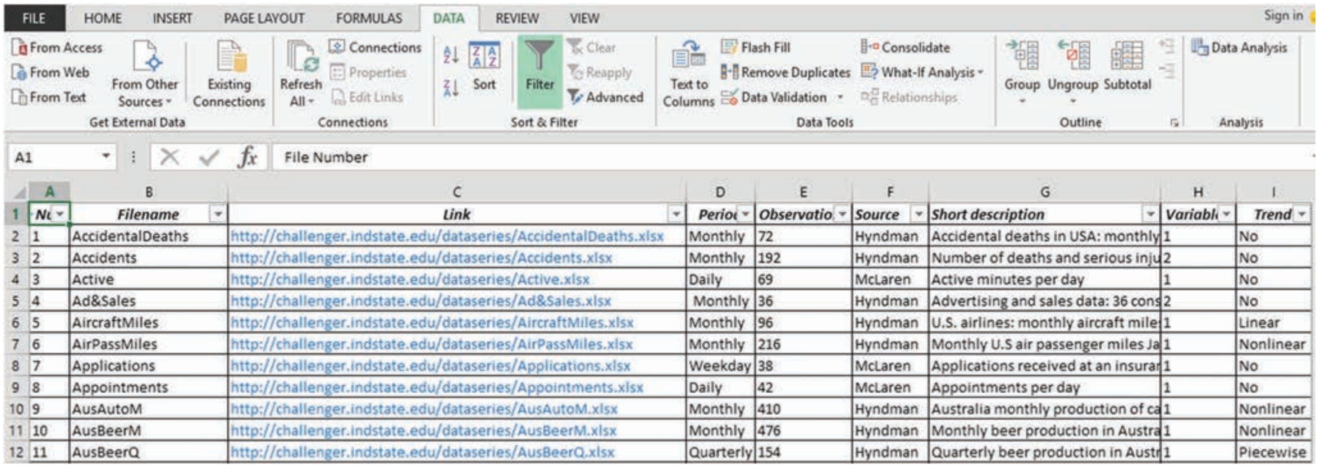


Figure 7. Use Excel Filters to Find Appropriate Time Series



	A	B	C	D	E	F	G	H	I
1	Nu	Filename	Link	Period	Observatio	Source	Short description	Variabl	Trend
2	1	AccidentalDeaths	http://challenger.indstate.edu/dataseries/AccidentalDeaths.xlsx	Monthly	72	Hyndman	Accidental deaths in USA: monthly	1	No
3	2	Accidents	http://challenger.indstate.edu/dataseries/Accidents.xlsx	Monthly	192	Hyndman	Number of deaths and serious inju	2	No
4	3	Active	http://challenger.indstate.edu/dataseries/Active.xlsx	Daily	69	McLaren	Active minutes per day	1	No
5	4	Ad&Sales	http://challenger.indstate.edu/dataseries/Ad&Sales.xlsx	Monthly	36	Hyndman	Advertising and sales data: 36 cons	2	No
6	5	AircraftMiles	http://challenger.indstate.edu/dataseries/AircraftMiles.xlsx	Monthly	96	Hyndman	U.S. airlines: monthly aircraft mile	1	Linear
7	6	AirPassMiles	http://challenger.indstate.edu/dataseries/AirPassMiles.xlsx	Monthly	216	Hyndman	Monthly U.S air passenger miles Ja	1	Nonlinear
8	7	Applications	http://challenger.indstate.edu/dataseries/Applications.xlsx	Weekday	38	McLaren	Applications received at an insurar	1	No
9	8	Appointments	http://challenger.indstate.edu/dataseries/Appointments.xlsx	Daily	42	McLaren	Appointments per day	1	No
10	9	AusAutoM	http://challenger.indstate.edu/dataseries/AusAutoM.xlsx	Monthly	410	Hyndman	Australia monthly production of ca	1	Nonlinear
11	10	AusBeerM	http://challenger.indstate.edu/dataseries/AusBeerM.xlsx	Monthly	476	Hyndman	Monthly beer production in Austra	1	Nonlinear
12	11	AusBeerQ	http://challenger.indstate.edu/dataseries/AusBeerQ.xlsx	Quarterly	154	Hyndman	Quarterly beer production in Austr	1	Piecewise

for the Time Series Data Library, McLaren for author and student-generated time series, or Public for publicly available data.

3.1. Instructions

The SCOTS master table is stored as an Excel workbook file in xlsx file format to retain the clickable hyperlink cell to each time series data file. For maximum compatibility open the SCOTS table in Microsoft Excel version 2007 or later. Note that each file's link adds the filename to the base url. Row 1 contains column headings that may be used to filter or sort the set of files. To filter this table on desired time series characteristics, place the cursor in cell A1. In the Data ribbon click the Filter button in the Sort and Filter group. Each column heading should now have an arrow at its right side as shown in Figure 7.

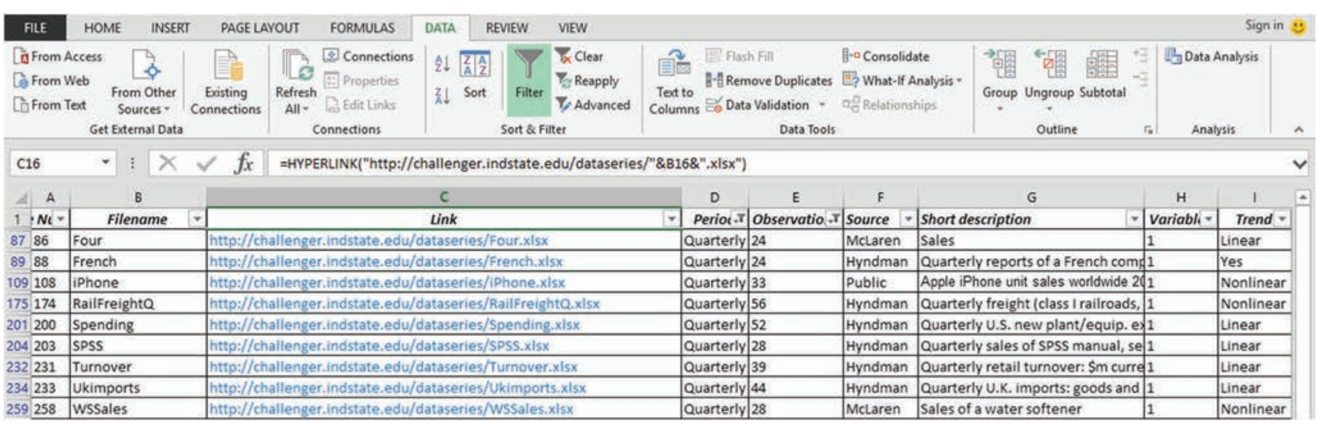
The SCOTS table is ready to filter on desired time series characteristics (criteria). Filtering the table will select the rows containing the time series data files that match the specified characteristics. From the filtered list click on a row's link cell in column C to open and download the time series data file to your computer. All rows in the SCOTS table remain; filtering temporarily hides rows that do not match the criteria set. You can filter on a single field or combine criteria. Keep in mind that multiple column filters are cumulative so each filter will apply unless you remove that filter.

3.2. Examples of Filtering the SCOTS Table

To see all the time series with quarterly data, for example, first click the filter arrow in the Period column in the SCOTS table. In the dropdown menu uncheck the Select All button to remove the check marks and then click the box in front of Quarterly. The result of this action is a list of the 28 time series that are based on quarterly observations. Click on the link cell in column C for any row; the Excel file for that time series will open on your computer. For example, if you click on the AusIronQ link in cell C32 that time series Excel file will open.

Columns with numerical data can be filtered by using Number Filters. For example, to further filter the set of quarterly series to find those that have fewer than 60 observations, click the filter arrow in the Observations column and then click Number Filters. Select the Is Less Than choice and enter 60. Click OK and the nine rows for quarterly time series that have fewer than 60 observations will appear. Figure 8 shows the results of the quarterly time series with fewer than 60 observations.

Figure 8. Filtering Can Be Done for the Number of Observations



	A	B	C	D	E	F	G	H	I
1	Nu	Filename	Link	Period	Observatio	Source	Short description	Variabl	Trend
87	86	Four	http://challenger.indstate.edu/dataseries/Four.xlsx	Quarterly	24	McLaren	Sales	1	Linear
89	88	French	http://challenger.indstate.edu/dataseries/French.xlsx	Quarterly	24	Hyndman	Quarterly reports of a French comp	1	Yes
109	108	iPhone	http://challenger.indstate.edu/dataseries/iPhone.xlsx	Quarterly	33	Public	Apple iPhone unit sales worldwide 2(	1	Nonlinear
175	174	RailFreightQ	http://challenger.indstate.edu/dataseries/RailFreightQ.xlsx	Quarterly	56	Hyndman	Quarterly freight (class I railroads,	1	Nonlinear
201	200	Spending	http://challenger.indstate.edu/dataseries/Spending.xlsx	Quarterly	52	Hyndman	Quarterly U.S. new plant/equip. e)	1	Linear
204	203	SPSS	http://challenger.indstate.edu/dataseries/SPSS.xlsx	Quarterly	28	Hyndman	Quarterly sales of SPSS manual, se	1	Linear
232	231	Turnover	http://challenger.indstate.edu/dataseries/Turnover.xlsx	Quarterly	39	Hyndman	Quarterly retail turnover: $m curre	1	Linear
234	233	Ukimports	http://challenger.indstate.edu/dataseries/Ukimports.xlsx	Quarterly	44	Hyndman	Quarterly U.K. imports: goods and	1	Linear
259	258	WSSales	http://challenger.indstate.edu/dataseries/WSSales.xlsx	Quarterly	28	McLaren	Sales of a water softener	1	Nonlinear

To find a set of time series that are well suited for teaching a particular forecasting method, instructors can specify the type of trend, the type of seasonality, and the presence or absence of cycles. To find a time series that works well for double exponential smoothing, for example, the trend filter could be set for linear and nonlinear trend, the seasonality filter set for No, and the cycles filter set for No. The result of this action is a list of 38 files that satisfy all of these characteristics.

To remove a single filter, click that column's filter arrow and click the Clear Filter choice in the dropdown menu. To clear all filters, go to the Data ribbon in Excel and click the Clear button in the Sort and Filter group. All 262 rows reappear. Do not forget to clear filters that are no longer relevant.

4. Class Examples

We have typically used the files in SCOTS for four basic purposes: introduction to time series data and its features; measuring forecasting error; selection of an appropriate forecasting method; and student assessment. The following sections follow the sequential approach we use in teaching forecasting using real data.

4.1. Introduction to Time Series Data and Its Features

Most forecasting chapters in business statistics and management science textbooks (see, for example, Anderson et al. 2014, Groebner et al. 2014, Berenson et al. 2015, Ragsdale 2015, Winston and Albright 2015, and Anderson et al. 2016) and introductory chapters in dedicated forecasting texts (Makridakis et al. 1998, Wilson and Keating 2008, Hanke and Wichern 2010, Ord and Fildes 2013) list four features of time series data, i.e., trend, seasonality, cycles, and randomness. We have found that a good way to introduce these concepts to students is to provide a time series and have them construct an index or line plot. Individual review or group discussion can then be used to discover the features present in the plot and tell the story of the data.

Although specialty software could be used, we have the students open the time series in Excel, highlight the values, go to the Insert menu choice, and select the Line chart in the Charts group. This reinforces the idea that the time series is a sequence; we prefer this approach to having students create a scatter plot. However, it is important to remind students that they should plot only the values; students sometimes plot the observation number as one series and the time series values as a second series on the same set of axes and then marvel at the linear trend in their first series.

SCOTS contains links to a number of time series files that we have used to begin the forecasting unit. We sometimes provide the data without an explanation of its source and then ask the students to speculate about which of several scenarios the data might fit. We frequently use theAusChocQ data to introduce time series analysis. After the students plot the data, which clearly shows increased chocolate sales around holidays such as Valentine's Day, Halloween, and Christmas, we might ask whether it appears to be monthly or daily or quarterly data and whether it is more likely to represent sales of chocolate or of school supplies or of loaves of bread. They seem to enjoy this detective work.

Daily movie box office data provides a good illustration of nonlinear trend and seasonality (McLaren and DePaolo 2009). It is easy for students to think about the initial popularity of a movie and how attendance varies by day of the week; we sometimes leave off the day and let them think about what days of the week match the pattern. To generate interest, we often choose a movie from their childhood and remind them that the admission their family paid is reflected in the data.

We also find it useful to include files with no trend or seasonality in this initial exploration. Typically, we have students examine a handful of time series with different patterns to begin to understand how to recognize which features are present.

4.2. Measuring Forecast Error

Students are sometimes frustrated to realize that forecasting is an inexact science and that although an appropriate method can be selected, forecasters are still making predictions for an unknown future. We usually introduce forecast error early in the unit so that they can begin to understand how practitioners evaluate models. Our basic classes teach traditional forecasting methods that can be done with Excel formulas. Our more advanced students use the time series applications in Minitab and other specialty software. The SCOTS data files can be used in any application that can import Excel files. However, for the purposes of this article we will assume analysis of the files in Excel.

In business statistics and management science classes, we limit error analysis to three measures: mean absolute error (MAE or MAD), mean squared error (MSE), and MAPE. Because we want students to be proficient with Excel, we ask them to calculate (and save) these three error measures for each forecasting method applied to the time series so that they can compare the performance results among methods. To begin, we have them build their first forecast using the naïve method, where the forecast for time period t is the actual value from time period $t-1$. If you ask students how forecasts are made in their workplaces, you will almost certainly find at least one student who says "we just do what we did last time" giving credence to the naïve method. Others, particularly those who work in the restaurant industry, will say that they assume this Friday will be like last Friday, giving you a chance to introduce the concept of a seasonal naïve forecast, where the forecast for time period t is the actual value from the most recent occurrence of the same season.

We suggest that for error calculation practice, you use a time series with a limited number of observations. This does not necessarily have to be a short time series; you can select a subset of a longer one. It is simply helpful for students to have all of their observations visible in a worksheet without repeatedly having to scroll up or down. We also suggest that for practice, you include at least one time series with features that make the naïve method unattractive. If your students are familiar with simple linear regression, an interesting activity is to have them develop a linear trend forecast for a nonlinear time series. Looking at positive and negative errors helps them to understand over- and under-prediction.

4.3. Appropriate Selection of Forecasting Method

Being able to prepare for class by filtering SCOTS and searching for a time series with specific features will give you materials to help students understand which methods are appropriate for which kind of data. We emphasize this point by having students first identify the features in a time series and then select a class of appropriate methods. Those that are included in our learning objectives and that can be easily modeled in Excel are listed in Table 1.

As new methods are introduced in class, we find it useful for students to complete (or instructors to provide) a table that summarizes which methods are appropriate for which features of a time series. Table 2 shows a possible design using the numbering of methods from Table 1.

An instructor who is planning a lesson on a particular method can find that numbered method in Table 2 and then use the SCOTS filters to locate time series data files that illustrate the features (row and column headings) for that method.

Table 1. Many Introductory Texts Include These Time Series Forecasting Methods

1.	Naïve
2.	Seasonal Naïve
3.	Moving Average
4.	Simple Exponential Smoothing
5.	Double Exponential Smoothing (Holt's Method)
6.	Winters' Method (additive or multiplicative)
7.	Classical Seasonal Decomposition (additive or multiplicative)
8.	Linear Trend
9.	Nonlinear Trend

Table 2. Time Series Features Can Be Matched with Appropriate Forecasting Methods

	No trend	Linear trend	Nonlinear trend
No seasonality	1, 3, 4	5, 8	5, 9
Additive seasonality	2	6, 7	6
Multiplicative seasonality	6, 7	6, 7	6

4.4. Student Assessment

After students have completed their study of forecasting methods, we find it is useful to assess their understanding of which kinds of forecasting methods are appropriate for which time series features. Figure 9 shows examples of test questions we have used for this kind of assessment.

Ensuring security for large or multiple sections (over one or successive semesters) makes it necessary for instructors to have multiple data sets for homework, quizzes, and exams. Good practice indicates that the degree of complexity should not change significantly for versions of the same assignment. Although algorithmic or simulated data has a role in such situations, the very factors that make real data attractive for class examples or projects also make it attractive for exams and other assessment activities. We have found it particularly helpful to use subsets of quarterly data from large time series to create multiple versions of an exam. We have used the beer, bricks, cars, cement, and chocolate quarterly time series as sources for assignments, selecting subsets of 36 observations for each version of the assignment.

5. Conclusion

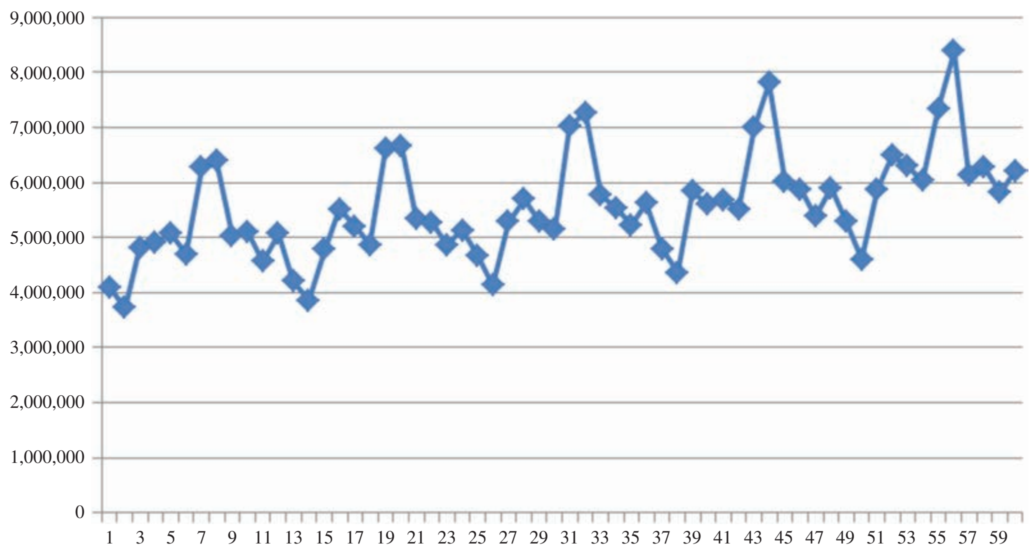
In teaching a mix of required and elective courses that incorporate time series data, we have found that students are much more engaged when a time series reflects measurements that match their experience. Whether this data is provided by the textbook, sourced from SCOTS or elsewhere, or collected by students, the true benefit comes when students obtain appropriate real data with all of its inherent complexities.

When time permits, we have found that having students collect their own data after they have practiced with data provided by the instructor is a worthwhile experience because it reinforces the fact that time series data measures the same quantity at regular intervals. We typically have the students gather data during the first half of the semester and then submit a report on their findings, providing a time series plot, a description of any trend or seasonality, and recommendations for applicable forecasting methods.

However, instructors who have students collect their own time series still need examples to use in class for the introduction of concepts, illustration of particular features, and student assignments. The 262 time series in SCOTS provide a rich resource to augment the files bundled with texts and enable the instructor to quickly and easily find appropriate data, whether the time series is used in its entirety or broken into subsets. Beyond providing the actual set of time series data files, the major purpose of this searchable collection is

Figure 9. Exam Questions Can Match Time Series Features with Appropriate Methods

1. For which of these time series would simple exponential smoothing be a good choice?
   a. A time series with seasonality, such as monthly sales of flip flops
   b. A time series with trend, such as the price of the same car model over the last 20 years
   c. A time series with randomness, such as the time it takes every day for someone to commute from our town to the state capital
   d. A time series with cycles, such as the daily price of a barrel of oil
2. To develop a forecasting model for this time series



   a. Additive decomposition would be a good choice and a trend model would be a poor choice
   b. A naïve model would be a good choice and SES would be a poor choice
   c. Seasonal naïve would be a better choice than additive decomposition
   d. Nonlinear trend would be better choice than multiple regression with seasonal indicators

the curation of the files and the notes for their use with students at all levels.

Acknowledgments
We gratefully acknowledge the help and expertise provided by Dr. Yihua Bai, Academic Programmer Analyst, and her staff for developing the procedure to host the 262 individual SCOTS data files. We also thank those students whose project data has been shared in this collection.

Appendix. Homework Assignments Using National Park Traffic Time Series

Assignment 1 (Suitable for a General Analytics Class at the High School, Undergraduate or Graduate Level)
Many government agencies collect time series data, and the National Park Service is no exception. For this exercise go to https://irma.nps.gov/Stats/. You will be examining data for a park of your choice. You can use a major park such as Yosemite or a local historical site or anything in between. Locate your facility by clicking on its state or type the name directly into the search box. Choose a monthly or a quarterly time series for your location. Copy and paste the data into your Excel file, including the name of the park and a description of the time series. Create a time series plot of the data using Excel (or use Minitab and paste the result into your Excel file). Write a paragraph describing what you see based on the four characteristics of a time series. Speculate on reasons for any patterns.

Note to the Instructor: *This data does not appear in chronological order and some students do not notice this. You may want to include the same sort of instructions that appear in Assignment* 2. *However, we have found that students who ignore the dates learn a valuable lesson when their assignment is returned.*

Assignment 2 (Suitable for a More Advanced Class in Forecasting)
This problem uses the seasonal time series for the total traffic count at the Shark Valley gate of Everglades National Park in Florida. Find this data at https://irma.nps.gov/Stats/SSRSReports/Park Specific Reports/Traffic Counts?Park=EVER. Scroll down to find the Shark Valley data. For this exercise you will need to practice manipulating data, something that is often necessary in messy real situations.

The data from 1991 to the present is arranged with a column for each month and with the years arranged so that the most recent is on top. For time series analysis, we need to arrange this so that there is a single column of values with the oldest values listed first. One of the skills you need to learn is how to manipulate data using features in Minitab, Excel, and other packages. There are several ways that you can accomplish this; one way is to

- Copy the Shark Valley data from the website, including the header row for the months and the column that displays the years, and paste into Excel. Delete the row that lists the park name.

- Using the year column, sort the data from oldest to newest, keeping the row data for each year together.
- Select all rows and columns of the sorted data (including the column of year labels and the row of month names), choose copy, and use the Excel Paste Transpose command to place the table in a new location. You should now see the years in columns and the months in rows.
- Finish by putting all the data into one column.
    - In Excel, repeatedly cut and paste the columns to create one long column.
    - If you have access to Minitab, paste the results into Minitab. Go to Data > Stack, and stack all of the year columns into a single column. Keep the year with each observation as an identifier. Create a column that identifies the month of each observation.

Cut the observations from 2014 to the present from the time series and store it in a separate place. You will use this as a holdout sample.

1. Develop these models for the within sample data and in each case, forecast for the holdout periods.

    (a) An additive seasonal decomposition model with linear trend.

    (b) A multiplicative seasonal decomposition model with linear trend.

    (c) A Holt-Winters model.

    a. If you are using Minitab, try at least five different combinations of smoothing constants to get a feel for performance before you settle on the parameters you want to use to create the forecasts.

    b. If you are using direct calculations in Excel, try at least five different combinations of the smoothing constants OR try to optimize the three smoothing constants using Solver. Be sure to specify which of the error measures you are optimizing if you use Solver.

Calculate the MAE, the MSE, and the MAPE for the holdout sample for each of the three forecast methods. To do this you will need to compare the forecast values to the values you held out and calculate the error measures directly. Based on these measurements, which of the three methods would you recommend for forecasting this time series?

2. Discuss the selection and viability of the trend portion of your model. Did trend exist? If so, do you support using simple linear regression (a straight line) to forecast trend from the deseasonalized values, or would applying double exponential smoothing (following a curve) to the deseasonalized values work better? You do not need to do any calculations for this question.

References

Aliaga M, Cobb G, Cuff C, Garfield J, Gould R, Lock R, Moore T, Rossman A, Stephenson B, Utts J, Velleman P, Witmer J. (2005) *Guidelines for Assessment and Instruction in Statistics Education* (*GAISE*) *College Report* (American Statistical Association, Alexandria, VA). Accessed October 11, 2017, http://www.amstat.org/education/gaise/.

Anderson D, Sweeney D, Williams T, Camm J, Cochran J (2014) *Statistics for Business and Economics*, 13th ed. Revised (South-Western, Mason, OH).

Anderson D, Sweeney D, Williams T, Camm J, Cochran J, Fry M, Ohlmann J (2016) *An Introduction to Management Science: Quantitative Approaches to Decision Making*, 14th ed. (Cengage Learning, Boston).

Bargagliotti A (2012) How well do the NSF funded elementary mathematics curricula align with the GAISE report recommendations? *J. Statist. Ed.* 20(3). Accessed October 11, 2017, http://ww2.amstat.org/publications/jse/v20n3/bargagliotti.pdf.

Berenson M, Levine D, Szabat K (2015) *Basic Business Statistics*, 13th ed. (Pearson Prentice Hall, Upper Saddle River, NJ).

Carver R, Everson M, Gabrosek J, Rowell G, Horton N, Lock R, Mocko M, Rossman A, Velleman P, Witmer J, Wood B (2016) Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report DRAFT. Accessed October 11, 2017, http://www.amstat.org/education/gaise/collegeupdate/GAISE2016_DRAFT.pdf.

Chen X, Zhou J, Gong F (2012) Teaching reform of college statistics course in USA—An introduction to GAISE and some related suggestions for educating college students in China. *J. Tianjin Univ. Tech. Ed.*2012(1):63–67.

Franklin C (2013) Common core state standards and the future of teacher preparation in statistics. *Guest Editorial, Math. Educator* 22(2):3–10.

Franklin C, Kader G, Mewborn D, Moreno J, Peck R, Perry M, Schaeffer R (2007) *Guidelines for Assessment and Instruction in Statistics Education* (*GAISE*) *Pre-K–12 Report* (American Statistical Association, Alexandria, VA). Accessed October 11, 2017, http://www.amstat.org/education/gaise/.

Groebner D, Shannon P, Fry P (2014) *Business Statistics*, 9th ed. (Pearson Prentice Hall, Upper Saddle River, NJ).

Groth R (2008) Assessing teachers' discourse about the pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE). *Statist. Ed. Res. J.* 7(1):16–39.

Hanke J, Wichern D (2009) *Business Forecasting*, 9th ed. (Pearson Prentice Hall, Upper Saddle River, NJ).

Hayat M (2014) Guidelines for Assessment and Instruction in Statistics Education (GAISE): Extending GAISE into nursing education. *J. Nursing Ed.* 53(4):192–198.

Horton N (2015) Challenges and opportunities for statistics and statistics education: Looking back, looking forward. *Amer. Statist.* 69(2):138–145.

Horton N, Hardin J (2015) Teaching the next generation of statistics students to "Think With Data": Special issue on statistics and the undergraduate curriculum. *Amer. Statistician* 69(4):259–265.

Makridakis S, Wheelwright S, Hyndman R (1998) *Forecasting Methods and Applications*, 3rd ed. (John Wiley & Sons, New York).

McLaren C, DePaolo C (2009) Movie data. *J. Statist. Ed.* 17(1). Accessed October 11, 2017, http://dx.doi.org/10.1080/10691898.2009.11889512.

Metz M (2010) Using GAISE and NCTM standards as frameworks for teaching probability and statistics to pre-service elementary and middle school mathematics teachers. *J. Statist. Ed.* 18(3). Accessed October 11, 2017, http://dx.doi.org/10.1080/10691898.2010.11889585.

Nowacki A (2015) Teaching statistics from the operating table: Minimally invasive and maximally educational. *J. Statist. Ed.* 23(1). Accessed October 11, 2017, http://dx.doi.org/10.1080/10691898.2015.11889726.

Ord K, Fildes R (2013) *Principles of Business Forecasting* (South-Western, Mason, OH).

Ozturk A (2012) A computer-assisted instruction in teaching abstract statistics to public affairs undergraduates. *J. Political Sci. Ed.* 8(3):251–257.

Pienta N (2013) Taking a cue from statistics education. *J. Chemical Ed.* 90(5):525–526.

Ragsdale C (2015) *Spreadsheet Modeling and Decision Analysis*, 7th ed. (Cengage Learning, Boston).

Wall A, Carver R, Pearl D, Martin N, Kuiper S, Posner M, Erickson P, Liao S, Albert J, Weber J (2014) Using fun in the statistics classroom: An exploratory study of college instructors' hesitations and motivations. *J. Statist. Ed.* 21(1). Accessed October 11, 2017, http://ww2.amstat.org/publications/jse/v20n3/bargagliotti.pdf.

Watson J (2014) Curriculum expectations for teaching science and statistics. ICOTS9 (2014) Invited Paper. Accessed October 11, 2017, http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_1A1_WATSON.pdf.

Wilson J, Keating B (2008) *Business Forecasting*, 6th ed. (McGraw Hill Irwin, New York).

Winston W, Albright S (2015) *Practical Management Science*, 5th ed. (Cengage Learning, Boston).

Woodard R, McGowan H (2012) Redesigning a large introductory course to incorporate the GAISE guidelines. *J. Statist. Ed.* 20(3). Accessed October 11, 2017, http://dx.doi.org/10.1080/10691898.2012.11889650.

Zieffler A, Park J, Garfield J, delMas R, Bjornsdottir A (2012) The statistics teaching inventory: A survey on statistics teachers' classroom practices and beliefs. *J. Statist. Ed.* 20(1). Accessed October 11, 2017, http:/dx.doi.org/10.1080/10691898.2012.11889632.