Journal of Statistics and Management Systems

ISSN: 0972-0510 (Print) 2169-0014 (Online) Journal homepage: https://www.tandfonline.com/loi/tsms20

Statistical methodological review for time-series data

Dewi Rahardja

To cite this article: Dewi Rahardja (2020): Statistical methodological review for time-series data, Journal of Statistics and Management Systems, DOI: 10.1080/09720510.2020.1727618

To link to this article: https://doi.org/10.1080/09720510.2020.1727618

Published online: 26 May 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tsms20

Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
DOI : 10.1080/09720510.2020.1727618

Statistical methodological review for time-series data

Dewi Rahardja
U.S. Department of Defense
Fort Meade
Maryland 20755
U.S.A.

Abstract

Numerous literatures on statistical methods for Time-Series (TS) data have been published. In this paper, a literature of the TS data analysis methods is reviewed. We organize the review based on the basic three-family category of TS models: The Exponential Smoothing Model (ESM) family, the Auto-Regressive Integrated Moving Average (ARIMA) model family, and the Unobserved Component Model (UCM) family. A roadmap is provided in a diagram format to these TS methods which are translatable into nowadays computing statements. Further, the execution of these methods in SAS commands (as one of the most popular nowadays statistical software packages) is also presented. This paper will be very beneficial for practitioners, forecasters, and researchers in diverse fields of study (such as business, management, finance, economics, etc.) to determine which TS data analysis method (along with the corresponding SAS command) are ready to use.

Subject Classification: *(2010) 37M10 – Time Series Analysis; 62M10 – Time series, auto-correlation, regression, etc.; 91B84 – Economic Time Series Analysis.*

Keywords: *Auto-correlation, Forecast, Time Series, Exponential Smoothing Model (ESM), Auto-Regressive Integrated Moving Average (ARIMA), Unobserved Component Model (UCM).*

1. Introduction

Various practitioners including Federal, Industrial, and Academic Forecasters and/or Researchers attend Federal Forecaster Conference (FFC) bi-annually at Washington, D.C. Such forecasters meeting continues

Disclaimer Statement: This research represents the author's own work and opinion. It does not reflect any policy nor represent the official position of the U.S. Department of Defense nor any other U.S. Federal agency.

E-mail: `rahardja@gmail.com`

© Taru Publications

a series of conferences that began three-decade ago in 1988 and have brought wide recognition to the importance of forecasting as a major statistical activity within the Federal Government, among its partner organizations, and numerous practitioners (industrial, academia, and federal). Therefore such conference activities have shown the demands of forecasting needs (on cloud-based data storage, data-driven forecasts, methods availability, software-computing implementation availability, etc.) since the past 30 years, [i.e., how the practitioners implement the existing Forecasting Methods computationally via existing software packages, especially SAS, as the currently most popular data-driven-purposed software (as opposed to simulation-driven-purposed) and commonly used commercial statistical software]. Hence, as Data-Driven practitioners, analysts, researchers or forecasters explore the availability of forecast methods and computing software, there are important questions about the generalizability, specificity, and nowadays computational ability to implement the forecasts or projections. Therefore this article will be helpful for various practitioners who need to make valid and justifiable forecasts and projections in their fields of applications (such as social science, physical science, economic, financial, business, etc.)

In practical applications, when the independent and identically distributed (*iid*) observations assumptions are violated, then the most-common basic **3-type-of-data** methods discussed in Rahardja *et*. *al*. (2016; 2017; 2018) are not suitable to be used for such *non-iid* data or observations. For instance, in the scenarios that the data exhibits temporal (time-related or time-dependent) or spatial (space-related or space-dependent) correlations. This paper discuss the former scenario: when the data exhibits temporal correlations (i.e., observations or data are not independent, or auto-correlated, in time, but are identically distributed), which is literature-wise referred as **Time Series (TS) Data**. Note that several example cases when observations or data are independent-but-not-identically-distributed are discussed in Rahardja (2005), as the fourth case in the article; and another example is in Zhou (2017) article. However, such independent-but-not-identically-distributed data scenario will not be discussed in this paper since it is outside the scope of this TS Data discussion.

Generally, in the scope of TS Data discussion, there are 3 major reasons why the assumptions of *iid* data can be violated: 1) the draw of a data point influences the outcome of a subsequent draw (inter-dependencies or auto-correlated), 2) the distribution changes at some point (non-stationarity), and 3) the data is not generated by a distribution at all (adversarial). This paper focus on (1) and (2), which is the scope of TS Data.

For TS data, the *outcome* (or *response* or *dependent*) variable of interest is not independent but identically distributed, auto-correlated series of equal-length time (such as hourly, daily, weekly, monthly, yearly, decade, etc.) observations. Forecasting Analysis is another commonly used terminology for TS data analysis.

Traditional (standard / classical) statistical methods which rely on *iid* assumptions, for examples, the **continuous** data analyses methods discussed in Rahardja (2017), the **categorical** data analyses methods discussed in Rahardja, *et. al.* (2016), and the **time-to-event (TTE)** or Survival/Failure/Reliability Analysis methods discussed in Rahardja and Wu (2018) depend on the *iid* assumptions and hence, are not suited to be able to include the auto-correlated data assumption as the outcome (or response) in the TS scenario. The analysis methodology must correctly account for such auto-correlated observations, in time. Hence such data sets can only be analyzed via TS methods. To date, there is no current up-to-date literature that concisely and methodically summarize the review of TS methods for the auto-correlated data (or TS) type of response variable (or outcome measure); and along with how to implement them in nowadays (up-to-date) computing software, such as SAS. Note that such up-to-date SAS implementability are not always available via SAS online documentation. Often, some SAS procedures are only learnable via paid onsite SAS training courses. Even so, a lot of time, the instructor may not always be able to answer all questions since typically they are not those who wrote the subroutines and sometimes could not find internal documentations / references. Hence often, users or practitioners need to resort to research papers or any of the above ways, as the author has been and is therefore prescribing the results here so that others do not need to reinvent the wheels.

In this paper, we present the basic three-family category of TS models along with the implementation procedures (PROC's) availability in SAS software, the most commonly used Statistical Software. The review includes the basic three-family category of TS models: the Exponential Smoothing Model (ESM) family, the Auto-Regressive Integrated Moving Average (ARIMA) model family, and the Unobserved Component Model (UCM) family. Sypsas (1989) and De Gooijer and Hyndman (2006), among many other older literature review, only discussed ESM family and ARIMA family models, and briefly mentioned Basic Structural Model (BSM) as another representation of UCM family models; but not detailed discussion nor translation to implement the model via statistical software (such as SAS) procedure (PROC) for the computing doability / implementation. Hence in our line of (statistical) practice, we often find many of both statistician and non-statistician practitioners, forecasters, and researchers get confused/mixed-up about the method, stuck in the model, but not sure what software-computing procedure to translate / implement / execute. To close such translation gap, this article will be a very practical and useful basic guidance / roadmap to both statisticians and non-statisticians in various fields of study.

2. Three-family TS Models

Fundamentally, there are basic three-family category of TS models: the Exponential Smoothing Model (ESM) family, the Auto-Regressive Integrated Moving Average (ARIMA) model family, and the Unobserved Component Model (UCM) family. Most of the ESM can also be expressed in terms of ARIMA model, as well.

First, we begin with the **Exponential Smoothing Models (ESM)** as the first family of TS Model. The basic idea of ESM is to fit forecasts of future values as weighted averages of past observations. Recent observations carry more weight in determining forecasts than observations in the distant past. A "smoothing" method implies this weighted-averages approach. The adjective "exponential" derives from the fact that some of the ESMs not only have weights that diminish with time but they do so in an exponential way. Listed in Table 1 are the best-known version of ESM (which are also translated into SAS computing software). The term ARIMA, Autoregressive Integrated Moving Average is explained in the next paragraph; however it will be used briefly in this ESM paragraph to include the ARIMA-equivalence model expression. In Table 1, as we can see, the Simple (Single) ESM can also be expressed as ARIMA (0, 1, 1) model; the Double ESM or Brown (1959 and 1962) ESM can also be expressed as ARIMA (0, 2, 2) model; the Linear or Holt (1957) ESM can also be expressed as ARIMA (0, 1, 2) model; the Damped Trend ESM discussed in Gardner and McKenzie (1985) can also be expressed as ARIMA (1, 1, 2) model and this method "dampens" the trend from the Linear (Holt) ESM to a flat line sometime in the future; the Add-Seasonal or Additive Seasonal or Seasonal ESM can be expressed as ARIMA (0, 1, 1) $(0, 1, 1)_m$ where m is the number of periods per season, or ARIMA (p, 0, 0) (0, 1, 1) + c where p = 1, 2, 3, and c is a constant; the Mult-Seasonal or Multiplicative Seasonal ESM has no ARIMA Equivalence; the Winters

(1960) or Winters Multiplicative ESM has no ARIMA Equivalence; and the Add-Winters or Winters Additive ESM can also be expressed as ARIMA (0, 1, [p+1])(0, 1, 0)$_p$. Both ESM and their ARIMA equivalence will produce the approximately the same time series projections. Fomby (2008) describes these ESM models in more details, based on the model written in SAS Manual (1995), for the first 4 ESM in Table 1. Hyndman and Athanasopoulos (2018) generalized the mathematical-statistics formulas for the last 4 ESM discussed in Table 1.

The second TS family is **ARIMA** model. ARIMA stands **Autoregressive Integrated Moving Average**. ARIMA models are the general TS forecasting models which decompose or filter the time series attributes in terms of *parameters* (p, d, q and P, D, Q), using **Spectral Decomposition algorithm**. ARIMA expresses values of a dependent time series with a combination of its auto-regressive component (p), differencing (d), moving average component (q), and other *predictor(s)* time series. Hence, a generic ARIMA model consists of an autoregressive (AR) term, a moving-average term (MA), and for a time series that requires differencing, an integrated (I) element. Furthermore, a comprehensive ARIMA model can have both non-seasonal and seasonal components, each characterized by three parameters, (p, d, q) for non-seasonal and $(P, D, Q)_s$ for seasonal. As such, an ARIMA model is usually denoted / abbreviated by ARIMA (p, d, q) $(P, D, Q)_s$, where p, d, q, P, D, Q are non-negative integers and s is the length of a season (for example, s = 12 month for a season of 12-month term; and s = 4 month for a season of quarterly data).

An ARIMA model can also include one or more predictors. Theoretically, the significance of the predictors needs to be statistically tested prior to fitting a forecast. Practically (interpretation-wise), it is also important to point out that predictors should only be included in a model when there is sensible reason to believe that the predictor influences the outcome. Hence, for an inclusion of any predictor into an ARIMA model, both perspectives (theoretical and practical) must concur.

The third TS family is the **Unobserved Components Model (UCM)**. The UCMs are also called **Structural Models** in the time series literature, for example, see the book by Harvey (1989). In SAS software, UCM decomposes or filters the response series into "components" using **Kalman-filtering algorithm** (instead of expressing the model equations as "parameters" as in the ARIMA scenario) in an convenient additive way, such as trend (intercept and slope), seasonality, cycles, auto-regressive components, a regression term or effects involving the lags of the dependent variables, due to predictors series, and the Irregular

Component, also called Disturbance Term, typically is assumed to be Gaussian White Noise with its corresponding variance. The sum of the regression terms (involved as predictors) includes contribution of regression variables with *fixed* regression coefficients.

For the notation of UCM, the fully specified UCM is written as

$$y_t = \mu_t + \gamma_t + \psi_t + r_t + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{j=1}^{m} \beta_j x_{jt} + \varepsilon_t . \tag{1}$$

In equation (1), the left-hand-side component y_t represents the response or time series to be modeled or forecast. On the right-hand-side of equation (1), the first component μ_t represents the trend component – which can be broken down further into two components: an intercept (or level) and a slope (or trend); the second component γ_t represents the seasonality component; the third component ψ_t represents the cyclical component; the fourth component r_t represents the autoregressive component; the fifth component represents the *i*-th (auto) regression term involving the lags of the dependent variables; the sixth component represents the *j*-th predictors or regressors term; and ε_t the irregular component (or white noise error term). All these components are assumed to be **unobserved** (**unknown** parameters) and must be **estimated** given the time series data on y_t and x_{jt}, hence give the title unobserved components model (UCM). On additional note, equation (1) allows the inclusion of the autoregressive regression terms $\sum_{i=1}^{p} \phi_i \, y_{t-i}$ and the explanatory regression terms $\sum_{j=1}^{m} \beta_j x_{jt}$, the former representing the *momentum* of the time series as it relates to its past observations and the latter representing the causal factors that one is willing to suppose affects the time series in question.

The components in the UCM model are supposed to capture the salient features of the series that are useful in explaining and predicting its behavior. These components are then sequentially tested to determine whether they are **deterministic (fixed)** or **stochastic (random)** and for inclusion/exclusion to the model.

The UCM considered in SAS software can be thought of a special case of more general models, the (linear) Gaussian State Space Model (GSSM). The state space formulation of a UCM has many computational advantages, such as convenient algorithms which are quite robust (in terms of computational convergence or difficulties/failures). Another major advantages of the UCM in equation (1) is its *interpretability* and the SAS PROC UCM provides very nice graphical representations of this decomposition.

Table 1

Listing of Time Series (TS) Family Models with the appropriate Model Equations and the corresponding SAS command.

Time Series (TS) Family Models	Model Equation	SAS Software Type & Default Estimator	SAS (Manual Coding) Command
ESM Family	ESM = Time Trend Term + Seasonality Term + Error Term; (with Different time-varying Smoothing Weights for each term: ***level*** smoothing weight, ***trend*** smoothing weight, and ***seasonal*** smoothing weight)		
SIMPLE (Single) ESM [Default]; or equivalently, **ARIMA (011)**	Level Trend Model: $y_t = \mu_t + a_t$ Smoothing Weights Formula: $Z_t = \omega y_t + (1- \omega) Z_{t-1}$ where, y_t = the observed time series up to t-th time; μ_t = the time-varying mean (level) term; **ω = smoothing constant;** $0 < \omega < 1$; Z_t = forecast value which is smoothed value at t; t = time index at current value; This model should be used when TS data has <ins>**No Trend**</ins> and <ins>**No Seasonality**</ins>.	• SAS Manual Coding: MLE • SAS-ETS: MLE • SAS-FS: CLS only (i.e., MLE option is N/A). N/A = Not Available. • Additionally, **transformed versions** of these models are also available: ❖ **NONE**: for no transformation. This is the default. ❖ **LOG**: for logarithmic transformation. ❖ **SQRT**: for square root transformation. ❖ **LOGISTIC**: for logistic transformation ❖ **BOXCOX(*n*)**: Box-Cox.	**PROC ESM** with / **model=SIMPLE** option

Contd...

DOUBLE (Brown) ESM; or equivalently, **ARIMA (022)**	Time Trend Model: $y_t = \mu_t + \beta_t t + a_t$ where, β_t = time-varying trend (slope) term; 1 Smoothing Weight Formula: $Z_t^{[2]} = \omega y_t + (1- \omega) Z_{t-1}^{[2]}$ This model should be used when TS data has **Trend** but **No Seasonality**.		**PROC ESM** with / **model = DOUBLE** option
LINEAR (Holt) ESM; or equivalently, **ARIMA (012)**	Time Trend Model: $Z_t = \mu_t + \beta_t t + a_t$ 2 Smoothing Weights Formula: $\mu_t = \alpha y_t (1-\alpha)(\mu_{t-1} + \beta_{t+1});$ $\beta_t = \omega(\mu_t - \mu_{t-1}) + (1-\omega) \beta_{t+1};$ where, α **= another smoothing constant; 0<** α **<1;** This model should be used when TS data has **Trend** (2 smoothing weights) but **No Seasonality**.		**PROC ESM** with / **model=LINEAR** option
DAMPTREND ESM; [when $\phi = 1$ then it is equivalent to Linear (Holt) ESM] or equivalently, **ARIMA (112)**	Time Trend Model: $Z_t = \mu_t + [\phi + \phi^2 + \ldots + \phi^h] \beta_t t + a_t$ 2 Smoothing Weights Formula: $\mu_t = \alpha y_t (1-\alpha)(\mu_{t-1} + \phi \beta_{t+1});$ $\beta_t = \omega(\mu_t - \mu_{t-1}) + (1-\omega) \phi \beta_{t+1};$ where, ϕ = damping parameter; $0 < \phi < 1$; h = h-step horizon This model should be used when TS data has **Trend** (2 smoothing weights & a damping parameter) but **No Seasonality**.		**PROC ESM** with / **model = DAMPTREND**

Contd...

ADD SEASONAL or SEASONAL (Additive Seasonal) ESM;	Time Seasonal Model: $Z_t = \mu_t + s_{t+h-m(k+1)} + a_t$		**PROC ESM** with / **model = SEASONAL**
TRIPLE (Holt-Winters) ESM;	where,		
or equivalently, **ARIMA (011) (011)$_m$** where m is the number of periods per season;	$\mu_t = \alpha(y_t - s_{t\text{-m}})(1\text{-}\alpha)(\mu_{t\text{-}1} + \beta_{t\text{-}1})$; $s_t = \gamma(y_t - \mu_{t\text{-}1} - \beta_{t\text{-}1}) + (1\text{-}\gamma)\, s_{t\text{-}m}$; k = the integer part of $(h{-}1)/m$, which ensures that the estimates of the seasonal indices used for forecasting come from the final year of the sample;		
or equivalently, **ARIMA (01q) (011)** with q=1 or 2;	$Z_{t+h-m(k+1)}$ = the time-varying seasonal term for the m season in the year; and a_t = white noise error term (and sometimes denoted by ε_t).		
or equivalently, **ARIMA (p00) (011)+c** with p=1,2,3, and c is a constant.	This model should be used when TS has **<u>No Trend</u>** but **<u>Seasonality</u>**.		
MULTSEASONAL (Multiplicative Seasonal) ESM;	Time Seasonal Model: $Z_t = (\mu_t)\, s_{t+h-m(k+1)} + a_t$		**PROC ESM** with / **model = MULTSEASONAL**
Hybrid Model hence No ARIMA Equivalence.	$\mu_t = \alpha(y_t / s_{t\text{-m}}) + (1\text{-}\alpha)(\mu_{t\text{-}1} + \beta_{t\text{-}1})$; $s_t = \gamma[y_t / (\mu_{t\text{-}1} - \beta_{t\text{-}1})] + (1\text{-}\gamma)\, s_{t\text{-}m}$; This model should be used when TS has **<u>No Trend</u>** but **<u>Seasonality</u>**.		
WINTERS (Winters Multiplicative) ESM;	Time Trend & Time Seasonal Model:		**PROC ESM** with / **model = WINTERS** option
No ARIMA Equivalence.	$Z_t = (\mu_t + h\, \beta_t\, t)\, s_{t+h-m(k+1)} + a_t$		

Contd...

	$\mu_t = \alpha(y_t / s_{t\text{-m}}) +$ $(1\text{-}\alpha)(\mu_{t\text{-}1} + \beta_{t\text{-}1});$ $\beta_t = \omega(\mu_t - \mu_{t\text{-}1}) + (1\text{-}\omega)\ \beta_{t\text{-}1};$ $s_t = \gamma[y_t / (\mu_{t\text{-}1} - \beta_{t\text{-}1})] +$ $(1\text{-}\gamma)\ s_{t\text{-}m};$ This model should be used when the TS data has **trend** and **seasonality**.		
ADDWINTERS (Winters Additive) ESM; or equivalently, **ARIMA (01[p+1])(010)$_p$**	Time Trend & Time Seasonal Model: $Z_t = \mu_t + h\ \beta_t\, t + s_{t+h-m(k+1)} + a_t$ where, $\mu_t = \alpha(y_t - s_{t\text{-m}})(1\text{-}\alpha)(\mu_{t\text{-}1} + \beta_{t\text{-}1});$ $\beta_t = \omega(\mu_t - \mu_{t\text{-}1}) + (1\text{-}\omega)\ \beta_{t\text{-}1};$ $s_t = \gamma(y_t - \mu_{t\text{-}1} - \beta_{t\text{-}1}) + (1\text{-}\gamma)\ s_{t\text{-}m};$ k = the integer part of $(h\text{–}1)/m$, which ensures that the estimates of the seasonal indices used for forecasting come from the final year of the sample; $Z_{t+h-m(k+1)}$ = the time-varying seasonal term for the m season in the year; and a_t = white noise error term (and sometimes denoted by ε_t). This model should be used when the TS data has **trend** and **seasonality**.		**PROC ESM** with / **model = ADDWINTERS**
ARIMA Family			
Generic / General / Basic ARIMA model: ARIMA(pdq)(PDQ) where,	$W_t = \mu + \dfrac{\theta(B)}{\phi(B)} a_t$ where,	• SAS Manual Coding: MLE • SAS-ETS: MLE • SAS-FS: CLS (but MLE option is available)	**PROC ARIMA** with the following statements: **identify** **estimate** **forecast**

Contd...

p = order of Auto Regressive (AR) component; q = order of Moving Average (MA) component; d = differencing; P = seasonality order of Auto Regressive (AR) component; Q = seasonality order of Moving Average (MA) component; D = seasonality differencing.	t = indexes time; W_t = the response series Y_t or a difference of the response series; μ = the mean term; B = the backshift operator; that is $B\ X_t = X_{t\text{-}1}$; $\phi(B)$ = the AR operator, represented as a polynomial in the backshift operator: $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$; $\theta(B)$ = the MA operator, represented as a polynomial in the backshift operator: $\phi(B) = 1 - \phi_1 B - \ldots - \phi_q B^q$; a_t = the independent disturbance, also called the random error and sometimes denoted by ε_t and often called white noise error term.		
ARIMA with predictors:			
ARIMA(pdq)(PDQ) + Predictor_*i* where i is non-negative integer.	$W_t = \mu + \dfrac{\theta(B)}{\phi(B)} a_t$ + Predictor_*i* where, Predictor_*i* = the i-th predictor or regressor.		**PROC ARIMA** with the following option: **crosscorr** = (pred); under the **identify** statement. Also, **input**=((1)pred); under the **estimate** statement.
UCM Family			
Generic or General **UCM model**	See Equation (1). **NOTE**: Equation (1) has already included the possibility of predictor or regressor components.	• SAS Manual Coding: MLE • SAS-ETS: Not Available • SAS-FS: CLS only (i.e., MLE option is N/A).	**PROC UCM**

3. Roadmap

Figure 1 provides the roadmap for practitioners, forecasters, and researchers to optimize a suitable (family-type) model for their TS (outcome measure or response variable) data analysis. In the Figure 1, the roadmap method is provided by the 3 category of family-type models: ESM, ARIMA, or UCM family.

First, for the family type of **ESM**, the roadmap will lead to eight (8) available candidate/initial/baseline models from PROC ESM [either by SAS manual (2013) coding, automatic SAS-ETS (2017), or automatic SAS-FS (2015)]. Note that only SAS manual coding and the automatic SAS-ETS have MLE (Maximum Likelihood Estimation) as the default and the default can be changed to CLS (Conditional Least Squares) option; for the SAS-FS, the default is CLS and there is no option for MLE because one of the eight ESM model (Multiplicative-Winters ESM) has no pure likelihood formulation; hence, MLEs cannot uniformly (and hence no once-for-all for automatic MLE option) be made available for all ESM models at SAS-FS. The Multiplicative-Winters ESM model is actually a hybrid model with a multiplicative trend/seasonal specification and an additive error specification. It has no likelihood formulation that fits in the usual Levenberg-Marquardt (1944 and 1963) formulation which converts a maximum likelihood problem into a least squares optimization problem. Note that such explanation are not available in the SAS online documentation and can only be obtained via direct question to the SAS Instructor whom researched the SAS internal documentation for quite long time.

Second, for the family type of **ARIMA** model, we use PROC ARIMA, and the roadmap directs to whether the SAS-computing option or pathway is manual coding, automatic coding via SAS-ETS candidate/initial/baseline models of 21 repository models (or 42 with their log transformation), or automatic coding via SAS Forecast Studio (SAS-FS) candidate/initial/baseline models of 28 model repository (or 56 with their log transformation). For the ARIMA-family modeling, the manual coding can utilize the X11 procedure to obtain five (5) candidate models; while the FORECAST procedure can further be used to obtain another twelve (12) different candidate models. Next, the TIMESERIES procedure can be used to obtain a quick TS plot [to examine any trend or patterns, visually] while the AUTOREG procedure is available to give a formal hypothesis testing result (in terms of ADF and/or PP test) to examine whether there is differencing or whether unit root exist and is not rejected. For ARIMA models, all these 3 SAS software options can be changed from MLE to CLS and vice versa. However, we note that when the option comes to select model selection criteria, some likelihood-based criteria such as (as opposed to distance-based) Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), have different formulas; i.e., the formulas at SAS-FS are different than SAS-ETS and/or SAS manual coding because they are coded by different subroutines with different coders and at different time. Therefore, it is not advisable to compare them because the comparison will not be fair / under same formulas. However those distanced-based model selection criteria such as Mean Absolute Percent Error (MAPE), Root Mean Squared Error (RMSE), etc. have the same formulas. Hence, their outputs (from SAS-ETS, SAS-FS, and SAS manual coding) are sensible to be compared.

Third, for the family type of **UCM** model, the roadmap will lead to UCM procedure. For PROC UCM, the SAS manual coding default is MLE; while the SAS-FS default is CLS (and with no MLE option); and the automatic SAS-ETS does not include UCM as part of the 21 (or 42 with their log transformation) repository model.

Note that here, the SAS default for the Base SAS 9.4 (SAS Institute Inc. 2013) manual coding and the automatic SAS-ETS version 14.3 (SAS Institute Inc. 2017) are using Maximum Likelihood Estimation (MLE) approach; while the automatic SAS-FS version 14.1 (SAS Institute Inc. 2015) is using the Conditional Least Square (CLS) approach. Typically, MLE is the best estimator and most desired due to its 5 nice asymptotic properties (which are Asymptotically or By-Large-Sample: Consistent, Normal, Unbiased, Efficient estimators and Invariant to transformations); however not every likelihood functions are twice differentiable nor converge (to its limiting distribution) nor has rapid-enough convergence; hence nice-and-easy (closed form) MLE are not always obtainable. Therefore, a CLS estimator can be a tangible alternative feasible-solution since it is a distance-based approach (not likelihood-based). However, although CLS may not have the same issues as MLE but it is not issue-free, either. CLS can have other computational issues such as existence of zero-cell matrices which causes non-invertible matrices due to division by zero entries, etc. Additionally, CLS may not guarantee to have those 5 nice asymptotic properties of MLE. Hence, CLS might not always be a better estimate than MLE; but they both can serve as reasonably-assured valid (i.e., sound and justifiable) and feasible estimators. In deciding which estimation methods to use, how many iterations of computing to stabilize, model selection criteria to use, whether or not to add/reduce the p, d, q, P, D, Q, pulse, etc., a statistical judgement from a subject-matter-expert (SME) is necessary and highly recommended.

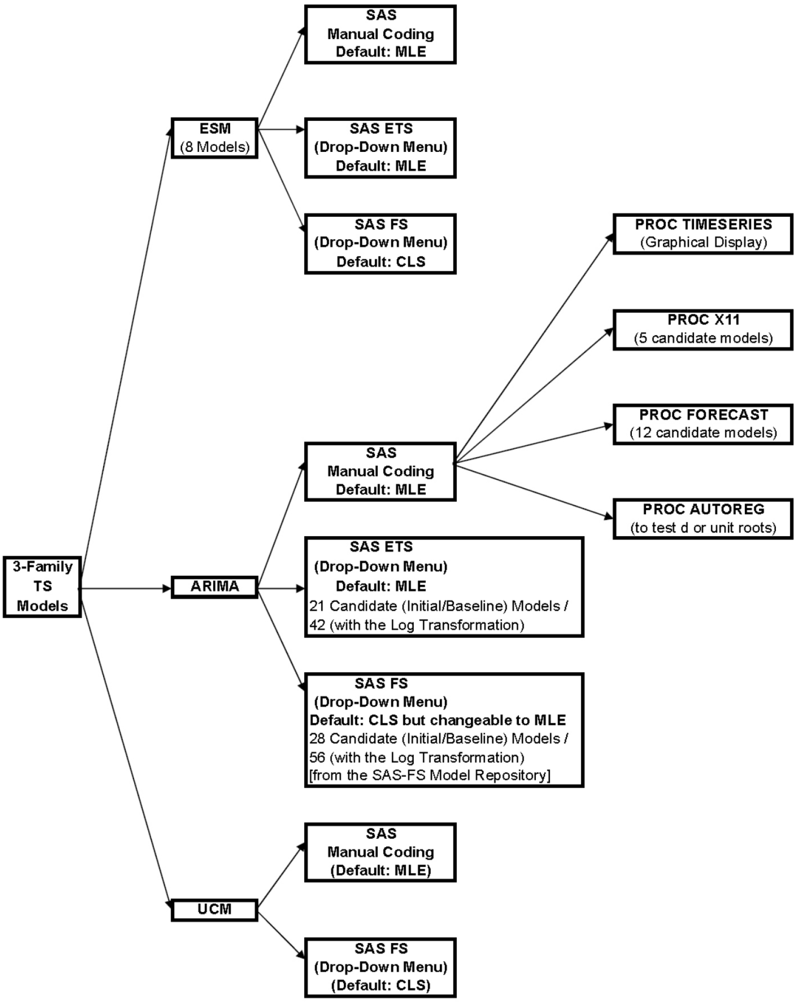

Figure 1

A Roadmap for the 3-Family Time-Series Models for the corresponding SAS PROC.

Next, the corresponding SAS procedures to the suitable statistical method directed from Figure 1 can be found in Table 1 prescription.

4. Summary

TS data (or response or outcome measure) is very common in real-data applications such as sociology, physical sciences, economics, financial data, etc. The analysis of such TS data (outcome measure or response variable) has a long history, condensed or summarized in the basic three-family category of TS models: ESM, ARIMA, and UCM. In this paper, we provide a review of the basic three-family category of TS statistical models and procedures that are practical/implementable in the recent literature for such TS data or auto-correlated observations (outcome measure or response variable), i.e., non-independent but identically distributed observations.

Additionally we also provide with the TS models, the corresponding SAS procedure, for the statistical computing translations/implementation, the most commonly used professional statistical software for data analysis. The optimum model obtained via automation (SAS-ETS and/or SAS-FS) can be enhanced either manually (best recommended for high accuracy) or via the available drop-down menu (from SAS-ETS and SAS-FS). However there are pros and cons. SAS manual coding are harder to learn (difficulty is tolerated for accuracy); while SAS automation or drop-down menu are easier to learn (accuracy is sacrificed for user friendliness).

In summary, this paper will be helpful for the practitioners, forecasters, and researchers in the various fields of study to determine the appropriate method for their data, according to the provided roadmap in Figure 1.

References

[1] Brown, R.G. (1959). Statistical Forecasting for Inventory control, McGraw-Hill: New York, NY.

[2] Brown, R.G. (1962). *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice Hall: New Jersey, NJ.

[3] De Gooijer, J.G., Hyndman, R.J. (2006). 25 years of Time Series Forecasting. *International Journal of Forecasting*, Vol. 22, No. 3, pp. 443–473.

[4] Fomby, T.B. (2008). Exponential Smoothing Models. *Class Notes Version 6*. Department of Economics. Southern Methodist University, Dallas, TX, June 2008.

[5] Gardner, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, *31*(10), 1237–1246. https://doi.org/10.1287/mnsc.31.10.1237

[6] Harvey, A.C. (1989). Forecasting, Structural Time Series Models, and the Kalman Filter. Cambridge University Press.

[7] Holt, C. E. (1957). *Forecasting seasonals and trends by exponentially weighted averages* (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh, USA. https://doi.org/10.1016/j.ijforecast.2003.09.015

[8] Hyndman, R.J., and Athanasopoulos, G. (2018). *Forecasting Principles and Practices*. Monash University, Australia. https://otexts.org/fpp2/

[9] Levenberg, Kenneth (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, Vol. 2, pp. 164–168.

[10] Marquardt, Donald (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, Vol. 11, No. 2, pp. 431–441. doi:10.1137/0111030

[11] Rahardja, D. (2005). X-Charts versus X/MR Chart Combinations: IID Cases & Non-IID Cases. *Quality Engineering*, Vol. 17, No. 2, pp. 189–196.

[12] Rahardja, D., Yang, Y., and Zhang, Z. (2016). A Comprehensive Review of the Two-Sample Independent or Paired Binary Data – with or without Stratum Effects. *Journal of Modern Applied Statistical Methods*, Vol. 15, No. 2, pp. 215–223.

[13] Rahardja, D. (2017). A Review of the Multiple-Sample Tests for the Continuous-Data Type. *Journal of Modern Applied Statistical Methods*, Vol. 16, No. 1, pp. 127–136.

[14] Rahardja, D., and Wu, H. (2018). Statistical methodological review for time-to-event data. *Journal of Statistics and Management Systems*, Vol. 21, No. 1, pp. 189–199.

[15] SAS Institute Inc. (1995). SAS Manual *SAS/ETS Software: Time Series Forecasting System*, Version 6, First Edition, pp. 225–235. Cary, NC: SAS Institute Inc.

[16] SAS Institute Inc. (2013). Base SAS® 9.4 Procedures Guide: Statistical Procedures, Second Edition. Cary, NC: SAS Institute Inc.

[17] SAS Institute Inc. (2015). SAS® Forecast Studio 14.1: User's Guide. Cary, NC: SAS Institute Inc.

[18] SAS Institute Inc. (2017). SAS/ETS® 14.3 User's Guide. Cary, NC: SAS Institute Inc.

[19] Sypsas, P.T. (1989). Identifying Patterns in Multiple Time Series Data. *Journal of Information and Optimization Sciences* Vol.10, No. 3, pp. 471–494.

[20] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6, 324–342. https://doi.org/10.1287/mnsc.6.3.324

[21] Zhou, M. (2017). Consistency of the Kaplan-Meier Estimator when Data are Independent Non-Identically Distributed. *Technical Report, Department of Statistics, University of Kentucky,* Lexington, KY 40506-0027, file property created 11 September 2017, pp. 1–15. www.ms.uky.edu/~mai/sta709/Tech351-1.pdf

Received April, 2019