ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

# A new approach to forecasting based on exponential smoothing with independent regressors

**Ahmad Farid Osman and Maxwell L. King**

**February 2015**

**Working Paper 02/15**

# A new approach to forecasting based on exponential smoothing with independent regressors

Ahmad Farid Osman[a,*], Maxwell L. King[b]

[a]*Department of Applied Statistics, University of Malaya, Kuala Lumpur 50603, Malaysia*
[b]*Department of Econometrics and Business Statistics, Monash University, Clayton, Vic 3800, Australia*

---

**Abstract**

There is evidence that exponential smoothing methods as well as time varying parameter models perform relatively well in forecasting comparisons. The aim of this paper is to introduce a new forecasting technique by integrating the exponential smoothing model with regressors whose coefficients are time varying. In doing this, we construct an exponential smoothing model with regressors by extending Holt's linear exponential smoothing model. We then translate it into an equivalent state space structure so that the parameters can be estimated via the maximum likelihood estimation procedure. Due to the potential problem in the updating equation for the regressor coefficients when the change in regressor is too small, we propose an alternative structure of the state space model which allows the updating process to be put on hold until sufficient information is available. An empirical study of forecast accuracy shows that the new model performs better than the existing exponential smoothing model as well as the linear regression model.

*Keywords:* State space model, Single source of error, Time varying parameter, Time series, Forecast accuracy

*JEL classification:* C51, C53

---

## 1. Introduction

Exponential smoothing forecasting can be considered as one of the most popular forecasting techniques since the 1950s. This technique is known as a method that is able to produce good forecasting results. There are a number of empirical studies and forecasting competitions that prove the usefulness of the exponential smoothing technique. Makridakis, Chatfield, Hibon, Lawrence, Mills, Ord, & Simmons (1993), for example, performed the "M2-Competition" to determine the post sample accuracy of various forecasting methods. Their results suggest that exponential smoothing methods, particularly damped and single smoothing, were the most accurate.

However, despite some usefulness of the existing class of exponential smoothing models, it also has some limitations in application. One obvious limitation is that it does not provide a procedure to integrate the effect of other variables into the model. It is actually a purely time series class of models in which the modeling procedure only deals with the dynamic movement of a single series. Clearly it would be useful to integrate the exponential smoothing model with regressors

*Corresponding author
*Email addresses:* `faridosman@um.edu.my` (Ahmad Farid Osman), `Max.King@monash.edu` (Maxwell L. King)

as the regression technique has widely been used to model the relationship between variables. Interestingly, the ability of the exponential smoothing technique to update its parameters can be exploited to update the regression coefficients as well. Previous studies have shown that the time varying parameter models perform quite well in forecasting as can be found in Li, Song, & Witt (2005), Li, Wong, Song, & Witt (2006), Song, Liu, & Romilly (1996) and Wolff (1987).

Hyndman, Koehler, Ord, & Snyder (2008) introduced an augmented version of the exponential smoothing model with regressor variables as the first attempt to integrate the exponential smoothing approach with regressors. They proposed the following model specification:

$$y_t = \boldsymbol{w}'\boldsymbol{x}_{t-1} + \boldsymbol{z}_t'\boldsymbol{p}_{t-1} + \varepsilon_t \tag{1a}$$
$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}\varepsilon_t \tag{1b}$$
$$\boldsymbol{p}_t = \boldsymbol{p}_{t-1}, \tag{1c}$$

where $y_t$ denotes the observed value of the dependent variable while $\varepsilon_t$ stands for the error or innovation. The vector $\boldsymbol{x}_t$ represents the state vector that contains information about the level ($\ell_t$), growth ($b_t$) and seasonality ($s_t, s_{t-1}, \ldots, s_{t-m+1}$). The vector $\boldsymbol{z_t}$ contains the regressor variables, $z_{1,t}, z_{2,t}, \ldots, z_{k,t}$ and the vector $\boldsymbol{p_t}$ is a vector of regression coefficients, $p_{1,t}$, $p_{2,t}, \ldots, p_{k,t}$. Since $\boldsymbol{p_t} = \boldsymbol{p_{t-1}}$, this can be defined as a time invariant specification of the effect of the regressors. On the other hand, the vector $\boldsymbol{w}$ gives the magnitude of effect of the previous state vector on the current observed value. Similarly, the matrix $\boldsymbol{F}$ gives the magnitude of effect of the previous state vector on the current state vector and the vector $\boldsymbol{g}$ contains the smoothing parameters which represent the magnitude of effect of the innovation on the current state vector. Alternatively, equation (1) can be rewritten as:

$$y_t = \bar{\boldsymbol{w}}_t'\bar{\boldsymbol{x}}_{t-1} + \varepsilon_t \tag{2a}$$
$$\bar{\boldsymbol{x}}_t = \bar{\boldsymbol{F}}\bar{\boldsymbol{x}}_{t-1} + \bar{\boldsymbol{g}}\varepsilon_t \tag{2b}$$

where

$$\bar{\boldsymbol{x}}_t = \begin{bmatrix}\boldsymbol{x}_t \\ \boldsymbol{p}_t\end{bmatrix}, \quad \bar{\boldsymbol{w}}_t = \begin{bmatrix}\boldsymbol{w} \\ \boldsymbol{z}_t\end{bmatrix}, \quad \bar{\boldsymbol{F}} = \begin{bmatrix}\boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}\end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{g}} = \begin{bmatrix}\boldsymbol{g} \\ \boldsymbol{0}\end{bmatrix}.$$

A detailed explanation of this approach can be found in Hyndman et al. (2008, chapter 9) and an application of it to tourism data can be found in Athanasopoulos & Hyndman (2008).

The purpose of this paper is to introduce another procedure of integrating regressors into the exponential smoothing technique by extending Holt's linear exponential smoothing method (Holt, 1957) which given by:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{3a}$$
$$b_t = \beta^*(L_t - L_{t-1}) + (1-\beta^*)b_{t-1} \tag{3b}$$
$$F_{t+m} = L_t + b_t m, \tag{3c}$$

where $L_t$ is an estimate of the level of the series, $b_t$ denotes the growth in the level of the series, $m$ denotes the number of periods ahead so that $F_{t+m}$ represents the $m$ step ahead forecast value, while $\alpha$ and $\beta^*$ represent smoothing parameters that usually take values between 0 to 1. Note that $\beta^*$ is used here instead of $\beta$ to avoid conflict with the use of the latter term in the state space structure in the following section. Explanation on this method can also be found in Makridakis, Wheelwright, & Hyndman (1998, p.158).

Equation (3a) demonstrates that the estimate of level, $L_t$ is actually a weighted average of the most recent observation, $Y_t$ and the pre-estimate of the level that is given by $(L_{t-1}+b_{t-1})$. Similarly, equation (3b) shows that the estimate of the growth, $b_t$ is actually a weighted average of the "actual growth" that is given by $(L_t - L_{t-1})$ and the most recent estimate of the growth, $b_{t-1}$. To be more precise, the "actual growth" can be expressed as $(L_t - L_{t-1})/\triangle_t$ which reflects that the growth at a certain point of time is equal to the difference between two successive levels divided by the change in time. In other words, it represents the growth for a unit change in time. By a similar argument, the pre-estimate of level in equation (3a) can be written as $(L_{t-1} + b_{t-1}\triangle_t)$, that is the sum of the previous value of level, $L_{t-1}$ and the estimated change in level, $b_{t-1}\triangle_t$.

Therefore, equations (3) are equivalent to:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1}\triangle_t) \tag{4a}$$

$$b_t = \beta^*([L_t - L_{t-1}]/\triangle_t) + (1-\beta^*)b_{t-1} \tag{4b}$$

$$F_{t+m} = L_t + b_t\triangle_{t,t+m}. \tag{4c}$$

$\triangle_t$ in equations (4) denotes the change in time so that $\triangle_t = t - (t-1) = 1$. In contrast, $\triangle_{t,t+m} = (t+m) - t = m$.

In equations (4), effectively $t$ is a regressor. What would happen if we replaced it with a more general regressor, $z_{1,t}$? In order to integrate the effect of regressors into the exponential smoothing method, we can easily augment equations (4) by replacing $\triangle_t$ with $\triangle_{z_{1,t}}$ where it denotes the change in the regressor, i.e. $\triangle_{z_{1,t}} = z_{1,t} - z_{1,t-1}$ and assuming that $\triangle_{z_{1,t}} \neq 0$. In this case, all of the symbols denote the same items as in equations (3) except for $b_{t-1}$ (with a slight modification to be $b_{1,t-1}$) which is now representing the growth (or coefficient) that is contributed by regressor, $z_{1,t}$ instead of the growth that is contributed by the time trend, $t$. $b_{1,t-1}$ here can also be defined as a growth for a unit change in the regressor, $z_{1,t}$.

Thus, the augmented Holt's linear method with a regressor in place of the time trend can be written as:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{1,t-1}\triangle_{z_{1,t}}) \tag{5a}$$

$$b_{1,t} = \beta_1^*([L_t - L_{t-1}]/\triangle_{z_{1,t}}) + (1-\beta_1^*)b_{1,t-1} \tag{5b}$$

$$F_{t+m} = L_t + b_{1,t}\triangle_{z_{1,t+m}}, \tag{5c}$$

where $\triangle_{z_{1,t+m}} = z_{1,t+m} - z_{1,t}$. The general form of these equations for $k$ regressors can be written as:

$$\begin{aligned}
L_t &= \alpha Y_t + (1-\alpha)(L_{t-1} + \sum_{i=1}^{k} b_{i,t-1}\triangle_{z_{i,t}}) \\
b_{1,t} &= \beta_1^*([L_t - L_{t-1} - \sum_{i=1}^{k} b_{i,t-1}\triangle_{z_{i,t}} + b_{1,t-1}\triangle_{z_{1,t}}]/\triangle_{z_{1,t}}) + (1-\beta_1^*)b_{1,t-1} \\
&\vdots \\
b_{k,t} &= \beta_k^*([L_t - L_{t-1} - \sum_{i=1}^{k} b_{i,t-1}\triangle_{z_{i,t}} + b_{k,t-1}\triangle_{z_{k,t}}]/\triangle_{z_{k,t}}) + (1-\beta_k^*)b_{k,t-1} \\
F_{t+m} &= L_t + \sum_{i=1}^{k} b_{i,t}\triangle_{z_{i,t+m}},
\end{aligned} \tag{6}$$

where $\triangle_{z_{i,t+m}} = z_{i,t+m} - z_{i,t}$, with $z_{i,t}$ denoting the $i^{th}$ regressor.

## 2. State space structure

The new model specified in the previous section can be written in an equivalent state space structure. It is actually a modified state space version of the existing exponential smoothing model structure (without regressors) which can be expressed in the general form of:

$$y_t = \boldsymbol{w}'\boldsymbol{x}_{t-1} + \varepsilon_t \tag{7a}$$

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}\varepsilon_t. \tag{7b}$$

For the augmented model with a regressor as described in equations (5), the corresponding state space structure is given by:

$$y_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \varepsilon_t \tag{8a}$$

$$\ell_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \alpha\varepsilon_t \tag{8b}$$

$$b_{1,t} = b_{1,t-1} + \beta_1\varepsilon_t/\triangle_{z_{1,t}}, \tag{8c}$$

that can be represented in matrix notation as:

$$y_t = \bar{\boldsymbol{w}}_t'\bar{\boldsymbol{x}}_{t-1} + \varepsilon_t \tag{9a}$$

$$\bar{\boldsymbol{x}}_t = \bar{\boldsymbol{F}}_t\bar{\boldsymbol{x}}_{t-1} + \bar{\boldsymbol{g}}_t\varepsilon_t \tag{9b}$$

where

$$\bar{\boldsymbol{x}}_t = \begin{bmatrix} \ell_t \\ b_{1,t} \end{bmatrix}, \quad \bar{\boldsymbol{w}}_t = \begin{bmatrix} 1 \\ \triangle_{z_{1,t}} \end{bmatrix}, \quad \bar{\boldsymbol{F}}_t = \begin{bmatrix} 1 & \triangle_{z_{1,t}} \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{g}}_t = \begin{bmatrix} \alpha \\ \beta_1/\triangle_{z_{1,t}} \end{bmatrix}.$$

We denote this model as ESWR(A,N,N,1,0) representing the combination of ETS(A,N,N) model with one regressor whose coefficient is time varying. ESWR is used as a short form of "Exponential Smoothing with Regressors". The general notation for this model is given by ESWR(E,T,S,V,C) where E represents the type of error (either None, Additive or Multiplicative), T for the type of trend, S for the type of seasonality, V is the number of regressors whose parameter is time varying and C is the number of regressors whose parameter is time invariant.

This state space transformation can be proved by solving equations (8a) for $\varepsilon_t$ and substituting the result into equation (8b) to get:

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}}),$$

which is equivalent to (5a). Furthermore, solving equation (8b) for $\varepsilon_t$, and substituting the result into equation (8c) gives:

$$b_{1,t} = (1-\beta_1/\alpha)b_{1,t-1} + (\beta_1/\alpha)(\ell_t - \ell_{t-1})/\triangle_{z_{1,t}},$$

which is equivalent to equation (5b) with $\beta_1^* = \beta_1/\alpha$ which means that $\alpha \neq 0$. Similar transformation can also be performed for models with multiple regressors.

In the case of models with $k$ regressors which may or may not include a time trend, $t$ as one of the regressors, the state space equations can be written as:

$$
\begin{aligned}
y_t &= \ell_{t-1} + \sum_{i=1}^{k} b_{i,t-1} \triangle_{z_{i,t}} + \varepsilon_t \\
\ell_t &= \ell_{t-1} + \sum_{i=1}^{k} b_{i,t-1} \triangle_{z_{i,t}} + \alpha \varepsilon_t \\
b_{1,t} &= b_{1,t-1} + \beta_1 \varepsilon_t / \triangle_{z_{1,t}} \\
&\vdots \\
b_{k,t} &= b_{k,t-1} + \beta_k \varepsilon_t / \triangle_{z_{k,t}}.
\end{aligned}
\tag{10}
$$

Note that a time trend can be included by setting it as $z_{i,t} = t$, so that $\triangle_{z_{i,t}} = 1$. In above equations, if any of the smoothing coefficients, $\beta_i$ is set to zero, then the corresponding growth coefficient, $b_{i,t}$ is constant or time invariant.

Therefore, for models with multiple regressors, the general state space form can be expressed as equations (9) with the corresponding matrices:

$$
\bar{\boldsymbol{x}}_t = \begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{b}_t \end{bmatrix}, \quad \bar{\boldsymbol{w}}_t = \begin{bmatrix} \boldsymbol{w}_t \\ \triangle_{\boldsymbol{z}_t} \end{bmatrix}, \quad \bar{\boldsymbol{F}}_t = \begin{bmatrix} \boldsymbol{F} & \triangle'_{\boldsymbol{z}_t} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{g}}_t = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{\beta}_{\boldsymbol{z}_t} \end{bmatrix}.
$$

Vector $\triangle_{\boldsymbol{z}_t}$ contains the changes in regressors, $\triangle_{z_{1,t}}, \triangle_{z_{2,t}}, \ldots, \triangle_{z_{k,t}}$ and $\boldsymbol{b}_t$ is a vector of regression coefficients, $b_{1,t}$, $b_{2,t}, \ldots, b_{k,t}$. In addition to that, $\boldsymbol{I}$ is a $k \times k$ identity matrix and $\boldsymbol{\beta}_{\boldsymbol{z}_t}$ is a vector whose elements are $\beta_1/\triangle_{z_{1,t}}, \beta_2/\triangle_{z_{2,t}}, \ldots, \beta_k/\triangle_{z_{k,t}}$.

The model structure explained above can be categorized as a state space model with time varying regression parameters. It easily can be transformed to the corresponding model with time invariant regression parameters by setting $\beta_i = 0$. Given below is a corresponding model with a time invariant regressor parameter based on equations (8) which is denoted as ESWR(A,N,N,0,1).

$$y_t = \ell_{t-1} + b_{1,t-1} \triangle_{z_{1,t}} + \varepsilon_t \tag{11a}$$

$$\ell_t = \ell_{t-1} + b_{1,t-1} \triangle_{z_{1,t}} + \alpha \varepsilon_t \tag{11b}$$

$$b_{1,t} = b_{1,t-1}, \tag{11c}$$

so that in matrix notation the model is:

$$y_t = \bar{\boldsymbol{w}}'_t \bar{\boldsymbol{x}}_{t-1} + \varepsilon_t \tag{12a}$$

$$\bar{\boldsymbol{x}}_t = \bar{\boldsymbol{F}}_t \bar{\boldsymbol{x}}_{t-1} + \bar{\boldsymbol{g}} \varepsilon_t \tag{12b}$$

where

$$
\bar{\boldsymbol{x}}_t = \begin{bmatrix} \ell_t \\ b_{1,t} \end{bmatrix}, \quad \bar{\boldsymbol{w}}_t = \begin{bmatrix} 1 \\ \triangle_{z_{1,t}} \end{bmatrix}, \quad \bar{\boldsymbol{F}}_t = \begin{bmatrix} 1 & \triangle_{z_{1,t}} \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{g}} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}.
$$

In this case, the matrix notation is similar to that for equations (9) but with a slight adjustment in the vector $\bar{\boldsymbol{g}}$ which now is time invariant.

## 3. Models with a seasonal component

In regard to the original exponential smoothing method, Winters (1960) introduced an extended version of Holt's linear method that allows the integration of a seasonal component into it as given by:

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (13a)$$
$$b_t = \beta^*(L_t - L_{t-1}) + (1-\beta^*)b_{t-1} \quad (13b)$$
$$S_t = \gamma^*(Y_t - L_t) + (1-\gamma^*)S_{t-s} \quad (13c)$$
$$F_{t+m} = L_t + b_t m + S_{t-s+m}, \quad (13d)$$

where $S_t$ represents the seasonal component while $s$ represents the periodicity of the seasonality.

This concept can directly be adopted into Holt's linear method augmented with regressors. Considering an exponential smoothing model with one regressor as defined by equations (5), we can integrate the seasonal effect by rewriting it as:

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{1,t-1}\triangle_{z_{1,t}}) \quad (14a)$$
$$b_{1,t} = \beta_1^*([L_t - L_{t-1}]/\triangle_{z_{1,t}}) + (1-\beta_1^*)b_{1,t-1} \quad (14b)$$
$$S_t = \gamma^*(Y_t - L_t) + (1-\gamma^*)S_{t-s} \quad (14c)$$
$$F_{t+m} = L_t + b_{1,t}\triangle_{z_{1,t+m}} + S_{t-s+m}. \quad (14d)$$

Similarly, we just need a slight modification to the corresponding state space structure of equations (8) which now can be expressed as:

$$y_t = \ell_{t-1} + s_{t-m} + b_{1,t-1}\triangle_{z_{1,t}} + \varepsilon_t \quad (15a)$$
$$\ell_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \alpha\varepsilon_t \quad (15b)$$
$$s_t = s_{t-m} + \gamma\varepsilon_t \quad (15c)$$
$$b_{1,t} = b_{1,t-1} + \beta_1\varepsilon_t/\triangle_{z_{1,t}}. \quad (15d)$$

$s_t$ in the above specification represents the seasonal component while $m$ represents the length of seasonality. ESWR(A,N,A,1,0) is used to denote this model.

To prove the above state space model is equivalent to equations (14), we need first to solve equation (15a) for $\varepsilon_t$ and substitute the result into equation (15b) to give:

$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}}),$$

which is equivalent to (14a). Then, solving equation (15b) for $\varepsilon_t$, and substituting the result into equation (15d) to give:

$$b_{1,t} = (1 - \beta_1/\alpha)b_{1,t-1} + (\beta_1/\alpha)(\ell_t - \ell_{t-1})/\triangle_{z_{1,t}},$$

which is equivalent to equation (14b) with $\beta_1^* = \beta_1/\alpha$. In addition to that, we also need to solve equation (15a) for $\varepsilon_t$, equation (15b) for $b_{1,t-1}\triangle_{z_{1,t}}$ and equation (15c) for $\varepsilon_t$, and substitute them into equation (15c) to give:

$$s_t = [\gamma/(1-\alpha)]\,(y_t - \ell_t) + [1 - \gamma/(1-\alpha)]\,s_{t-m},$$

which is equivalent to equation (14c) with $\gamma^* = \gamma/(1-\alpha)$.

The matrix representation of equations (15) is given by:

$$\bar{\boldsymbol{w}}_t' = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \triangle_{z_{1,t}} \end{bmatrix}, \; \bar{\boldsymbol{x}}_t = \begin{bmatrix} \ell_t \\ s_t \\ s_{t-1} \\ \vdots \\ s_{t-m+1} \\ b_{1,t} \end{bmatrix}, \; \bar{\boldsymbol{g}}_t = \begin{bmatrix} \alpha \\ \gamma \\ 0 \\ \vdots \\ 0 \\ \beta_1/\triangle_{z_{1,t}} \end{bmatrix}, \text{ and } \bar{\boldsymbol{F}}_t = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & \triangle_{z_{1,t}} \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}.$$

For models with $k$ regressors whose parameters are time varying, the state space equations are written as:

$$\begin{aligned} y_t &= \ell_{t-1} + s_{t-m} + \sum_{i=1}^{k} b_{i,t-1}\triangle_{z_{i,t}} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + \sum_{i=1}^{k} b_{i,t-1}\triangle_{z_{i,t}} + \alpha\varepsilon_t \\ s_t &= s_{t-m} + \gamma\varepsilon_t \\ b_{1,t} &= b_{1,t-1} + \beta_1\varepsilon_t/\triangle_{z_{1,t}} \\ &\vdots \\ b_{k,t} &= b_{k,t-1} + \beta_k\varepsilon_t/\triangle_{z_{k,t}}. \end{aligned} \tag{16}$$

The general state space forms for the above time varying and time invariant parameter of regressors models are exactly the same as equations (9) with the corresponding matrix notation:

$$\bar{\boldsymbol{x}}_t = \begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{b}_t \end{bmatrix}, \quad \bar{\boldsymbol{w}}_t = \begin{bmatrix} \boldsymbol{w}_t \\ \triangle_{\boldsymbol{z}_t} \end{bmatrix}, \quad \bar{\boldsymbol{F}}_t = \begin{bmatrix} \boldsymbol{F} & \triangle_{\boldsymbol{z}_t}^{+} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{g}}_t = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{\beta}_{\boldsymbol{z}_t} \end{bmatrix}.$$

In this case, $\triangle_{\boldsymbol{z}_t}^{+}$ is a matrix of dimension $(1+m)\times k$ whose first row contains the vector of changes in regressors, $\triangle_{\boldsymbol{z}_t}'$, while the remaining rows are the zero matrix, $\boldsymbol{0}$.

## 4. Pitfalls and solutions

There are some issues that arise in the use of the state space structure described by equations (10) and (16) where the growth equation for the $i^{th}$ regressor is expressed as:

$$b_{i,t} = b_{i,t-1} + \beta_i\varepsilon_t/\triangle_{z_{i,t}}. \tag{17}$$

As $\triangle_{z_{i,t}}$ denotes the change in the $i^{th}$ regressor, there are two possible problems that can occur. The first one is when $\triangle_{z_{i,t}} = 0$. In this case, there is no change or difference between the two successive values of regressor ($z_{i,t} = z_{i,t-1}$). As a result, $b_{i,t}$ is undefined. Another related problem is when $\triangle_{z_{i,t}}$ is close to zero. This could blow up the value of $b_{i,t}$ (which is an element of vector

$\bar{\boldsymbol{x}}_t$). Consequently, the following fitted value, $\hat{y}_{t+1}$ as well as the residual, $\varepsilon_{t+1}$ will experience an "extreme change". This is due to the fact that the expected value of equation (9a) and the corresponding residual are recursively calculated based on the following equations.

$$\hat{y}_t = \bar{\boldsymbol{w}}_t'\bar{\boldsymbol{x}}_{t-1} \tag{18a}$$
$$\varepsilon_t = y_t - \hat{y}_t \tag{18b}$$
$$\bar{\boldsymbol{x}}_t = \bar{\boldsymbol{F}}_t\bar{\boldsymbol{x}}_{t-1} + \bar{\boldsymbol{g}}_t\varepsilon_t \tag{18c}$$

The "huge" $\varepsilon_{t+1}$ produced in equation (18b) will then make the updating process for $\ell_{t+1}$ and $b_{i,t+1}$ in equation (18c) a little erratic. This could give a prolonged effect for several steps of the following updating process.

Fortunately, the above problems can be solved by introducing a switching procedure for equation (17). When there is no change in the regressor values or when the change is too small (close to zero), then we can assume that there is no or very little information in the data about how $b_{i,t}$ might have changed. Thus, it is appropriate not to update the growth coefficient at that time period. In other words, when $|\triangle_{z_{i,t}}|$ is equal to zero or close to zero, equation (17) is replaced by:

$$b_{i,t} = b_{i,t-1}. \tag{19}$$

To do this, we need to fix a lower boundary for $|\triangle_{z_{i,t}}|$ as a guide as to when we will use equation (17) or switch to equation (19) for the growth equation. By taking a model with one regressor defined by equations (8) as an example, we can now rewrite the state space equations as:

$$y_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \varepsilon_t \tag{20a}$$
$$\ell_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \alpha\varepsilon_t \tag{20b}$$
$$b_{1,t} = \begin{cases} b_{1,t-1} + \beta_1(\varepsilon^+_{1,t-1} + \varepsilon_t)/\triangle^*_{z_{1,t}}, & |\triangle^*_{z_{1,t}}| \geq L_{b_1} \\ b_{1,t-1}, & |\triangle^*_{z_{1,t}}| < L_{b_1} \end{cases} \tag{20c}$$
$$\varepsilon^+_{1,t} = \begin{cases} 0, & |\triangle^*_{z_{1,t}}| \geq L_{b_1} \\ \varepsilon^+_{1,t-1} + \varepsilon_t, & |\triangle^*_{z_{1,t}}| < L_{b_1} \end{cases} \tag{20d}$$

where

$$\triangle^*_{z_{1,t}} = z_{1,t} - z^*_{1,t-1} \text{ and } z^*_{1,t} = \begin{cases} z_{1,t}, & |\triangle^*_{z_{1,t}}| \geq L_{b_1} \\ z^*_{1,t-1}, & |\triangle^*_{z_{1,t}}| < L_{b_1}. \end{cases} \tag{21}$$

In this case

$$\bar{\boldsymbol{x}}_t = \begin{bmatrix} \ell_t \\ b_{1,t} \\ \varepsilon^+_{1,t} \end{bmatrix}, \quad \bar{\boldsymbol{w}}_t = \begin{bmatrix} 1 \\ \triangle_{z_{1,t}} \\ 0 \end{bmatrix},$$

$$\bar{\boldsymbol{F}}_t = \begin{bmatrix} 1 & \triangle_{z_{1,t}} & 0 \\ 0 & 1 & \beta_1/\triangle^*_{z_{1,t}} \\ 0 & 0 & 0 \end{bmatrix} \text{ if } |\triangle^*_{z_{1,t}}| \geq L_{b_1} \quad \text{or} \quad \bar{\boldsymbol{F}}^*_t = \begin{bmatrix} 1 & \triangle_{z_{1,t}} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ if } |\triangle^*_{z_{1,t}}| < L_{b_1}$$

and

$$\bar{\boldsymbol{g}}_t = \begin{bmatrix} \alpha \\ \beta_1/\triangle^*_{z_{1,t}} \\ 0 \end{bmatrix} \text{ if } |\triangle^*_{z_{1,t}}| \geq L_{b_1} \quad \text{or} \quad \bar{\boldsymbol{g}}^*_t = \begin{bmatrix} \alpha \\ 0 \\ 1 \end{bmatrix} \text{ if } |\triangle^*_{z_{1,t}}| < L_{b_1}.$$

$L_{b_1}$ in the above equations represents the lower boundary for $|\triangle^*_{z_{1,t}}|$.

It must be stressed here that even though we skip the updating process for the growth equation when $|\triangle^*_{z_{1,t}}| < L_{b_1}$, at that specific time we actually put the updating process on hold and bring forward the relevant information to be used in the following step of the updating process. In this case, the dummy regressor, $z^*_{1,t}$ is playing a role of keeping the information of $z_{1,t-i}$ that hasn't been used for the updating process. Similarly, the dummy error, $\varepsilon^+_{1,t}$ is playing a role of keeping the information of $\varepsilon_{t-i}$ that hasn't been used to update the growth coefficient.

As an example, let's say at $t = 5$, $|\triangle^*_{z_{1,t}}| > L_{b_1}$. Thus, we perform the updating process for the growth equation and set $z^*_{1,t} = z_{1,t}$. Then at $t = 6$, $|\triangle^*_{z_{1,t}}| < L_{b_1}$ so that we are not going to update the growth equation and simply set $b_{1,t} = b_{1,t-1}$ and $z^*_{1,t} = z^*_{1,t-1} = z_{1,t-1}$. After that at $t = 7$, $|\triangle^*_{z_{1,t}}| > L_{b_1}$ which requires us to update the growth equation. In this case, it can be shown that at $t = 7$:

$$y_t = \ell_{t-1} + b_{1,t-1}(z_{1,t} - z_{1,t-1}) + \varepsilon_t \tag{22a}$$

$$\ell_t = \ell_{t-1} + b_{1,t-2}(z_{1,t} - z_{1,t-1}) + \alpha\varepsilon_t \tag{22b}$$

$$b_{1,t} = b_{1,t-2} + \beta_1(\varepsilon_t + \varepsilon_{t-1})/(z_{1,t} - z_{1,t-2}). \tag{22c}$$

To get the equivalent exponential smoothing model for the above equations, we need first to solve equations (22a) and (22b) for $\varepsilon_t$ and substitute the results into equations (22b) and (22c) which yield:

$$\ell_t = \alpha y_t + (1 - \alpha)[\ell_{t-1} + b_{1,t-2}(z_{1,t} - z_{1,t-1})]$$

and

$$b_{1,t} = (1 - \beta_1/\alpha)b_{1,t-2} + (\beta_1/\alpha)[(\ell_t - \ell_{t-2})/(z_{1,t} - z_{1,t-2})].$$

Thus, the information collected at the previous two steps, $(t - 2)$ is used instead of the previous step, $(t - 1)$ to update the growth coefficient at the current step, $t$.

For the case of the model with multiple regressors, the modified state space equations are written as:

$$y_t = \ell_{t-1} + \sum_{i=1}^{k} b_{i,t-1}\triangle_{z_{i,t}} + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + \sum_{i=1}^{k} b_{i,t-1}\triangle_{z_{i,t}} + \alpha\varepsilon_t$$

$$b_{1,t} = \begin{cases} b_{1,t-1} + \beta_1(\varepsilon^+_{1,t-1} + \varepsilon_t)/\triangle^*_{z_{1,t}}, & |\triangle^*_{z_{1,t}}| \geq L_{b_1} \\ b_{1,t-1}, & |\triangle^*_{z_{1,t}}| < L_{b_1} \end{cases}$$

$$\vdots$$

$$b_{k,t} = \begin{cases} b_{k,t-1} + \beta_k(\varepsilon^+_{k,t-1} + \varepsilon_t)/\triangle^*_{z_{k,t}}, & |\triangle^*_{z_{k,t}}| \geq L_{b_k} \\ b_{k,t-1}, & |\triangle^*_{z_{k,t}}| < L_{b_k} \end{cases}$$

$$\varepsilon^{+}_{1,t} = \begin{cases} 0, & |\triangle^{*}_{z_{1,t}}| \geq L_{b_1} \\ \varepsilon^{+}_{1,t-1} + \varepsilon_t, & |\triangle^{*}_{z_{1,t}}| < L_{b_1} \end{cases}$$
$$\vdots \tag{23}$$
$$\varepsilon^{+}_{k,t} = \begin{cases} 0, & |\triangle^{*}_{z_{k,t}}| \geq L_{b_k} \\ \varepsilon^{+}_{k,t-1} + \varepsilon_t, & |\triangle^{*}_{z_{k,t}}| < L_{b_k} \end{cases}$$

where

$$\triangle^{*}_{z_{i,t}} = z_{i,t} - z^{*}_{i,t-1} \text{ and } z^{*}_{i,t} = \begin{cases} z_{i,t}, & |\triangle^{*}_{z_{i,t}}| \geq L_{b_i} \\ z^{*}_{i,t-1}, & |\triangle^{*}_{z_{i,t}}| < L_{b_i}. \end{cases} \tag{24}$$

This remedial procedure requires us to amend the matrix notation in equations (9) which now can be written as:

$$\bar{\boldsymbol{x}}_t = \begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{b}_t \\ \boldsymbol{\varepsilon}^{+}_t \end{bmatrix}, \quad \bar{\boldsymbol{w}}_t = \begin{bmatrix} \boldsymbol{w}_t \\ \triangle_{\boldsymbol{z_t}} \\ \boldsymbol{0} \end{bmatrix}, \quad \bar{\boldsymbol{F}}_t = \begin{bmatrix} \boldsymbol{F} & \triangle^{+}_{\boldsymbol{z}_t} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{d\beta}^{*}_{\boldsymbol{z}^{*}_t} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}^{*} \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{g}}_t = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{\beta}^{*}_{\boldsymbol{z}^{*}_t} \\ \boldsymbol{1}^{*} \end{bmatrix}.$$

Matrix $\boldsymbol{\varepsilon}^{+}_t$ in the above notation contains dummy errors, $\varepsilon^{+}_{1,t}, \varepsilon^{+}_{2,t}, \ldots, \varepsilon^{+}_{k,t}$, while the matrix $\boldsymbol{\beta}^{*}_{\boldsymbol{z}^{*}_t}$ is a $k \times 1$ vector with each of the $i^{th}$ element being either $\beta_i/\triangle^{*}_{z_{i,t}}$ if $|\triangle^{*}_{z_{i,t}}| \geq L_{b_i}$ or 0 otherwise. Similarly, $\boldsymbol{1}^{*}$ is a $k \times 1$ vector of ones with the $i^{th}$ element set to zero if $|\triangle^{*}_{z_{i,t}}| \geq L_{b_i}$. In addition to that, $\boldsymbol{d\beta}^{*}_{\boldsymbol{z}^{*}_t}$ is a $k \times k$ diagonal matrix with the $i^{th}$ diagonal element being either $\beta_i/\triangle^{*}_{z_{i,t}}$ if $|\triangle^{*}_{z_{i,t}}| \geq L_{b_i}$ or 0 otherwise. $\boldsymbol{I}^{*}$ is a $k \times k$ identity matrix with the $i^{th}$ diagonal element set to zero if $|\triangle^{*}_{z_{i,t}}| \geq L_{b_i}$.

Regarding the lower boundary, simply setting any small number for the boundary, $L_{b_i}$ in practice could eliminate the possible "extreme change" between the two successive $b_{i,t}$ values. However, an "excessive" value of the boundary may cause most of the $b_{i,t}$ to be unchangeable. In order to restore the flexibility of $b_{i,t}$ in updating its value, we need to find a value of the boundary at the optimal point, that is as small as possible but sufficient to avoid the "extreme change" behavior.

Treating the boundary, $L_{b_i}$ as a parameter to be estimated in the likelihood estimation procedure proved not to be a practical approach to find an optimal boundary. This is because the nature of maximum likelihood estimation in some cases led to an excessive value of boundary which reduces the flexibility of $b_{i,t}$ from being time varying. More than that, we need an approach to find a boundary that takes into account both in-sample and out-of-sample observations of the regressor so that there will be no extreme change in $b_{i,t}$ for out-of-sample forecasts.

One can set any suitable value of $L_{b_i}$ for a regressor series, $z_{i,t}$. However, it would be tedious to try the suitability of each possible value of $L_{b_i}$ especially when dealing with a long period of observations and huge number of regressors. Thus, it is appropriate to have a procedure that easily can be used to select an appropriate boundary, $L_{b_i}$ for any choice of regressors. Appropriate boundary here means that we just need the smallest possible value of $L_{b_i}$ that minimizes the number of $b_{i,t}$'s that are not to be updated or not to have a freedom to change.

Based on the above arguments, we can now define the appropriate boundary as a small value that is sufficient enough to eliminate the presence of extreme changes in $b_{i,t}$. Since we are interested in eliminating extreme change behavior, we could find an optimal boundary by adopting a procedure introduced by Tukey (1977) to flag outliers. The basic idea here is to eliminate any extreme effect of $\beta_i/\triangle_{z_{i,t}}$ that could cause an extreme change in $b_{i,t}$. To do this, we need to treat $\beta_i/|\triangle_{z_{i,t}}|$ just like ordinary observations and calculate its dispersion summary including the first quartile ($Q_1$), the second quartile ($Q_2$) which also known as the median, and the third quartile ($Q_3$) where $Q_1 = \frac{1}{4}(n+1)^{th}$ smallest value, $Q_2 = \frac{1}{2}(n+1)^{th}$ smallest value, and $Q_3 = \frac{3}{4}(n+1)^{th}$ smallest value where $n$ is the number of observations. Note that, $|\triangle_{z_{i,t}}|$ is used here instead of $\triangle_{z_{i,t}}$ as we wish to have a single value of the lower boundary for both positive and negative sides of $\triangle_{z_{i,t}}$. According to Tukey (1977), any points smaller than $Q_1 - 1.5(Q_3 - Q_1)$ which is known as a lower inner fence ($LIF$) and greater than $Q_3 + 1.5(Q_3 - Q_1)$ which is known as an upper inner fence ($UIF$) are considered to be outliers. However, our main concern here is to eliminate the right hand side outliers as they are the source of extreme change in $b_{i,t}$. Thus, in order to avoid extreme change in $b_{i,t}$, all of $\beta_i/|\triangle_{z_{i,t}}|$ must satisfy the following condition:

$$\beta_i/|\triangle_{z_{i,t}}| \leq UIF_{\beta_i/|\triangle_{z_{i,t}}|}. \tag{25}$$

In other words, we need to find a value of boundary that eliminates all the right hand side outliers in $\beta_i/|\triangle_{z_{i,t}}|$. Since different values of $\beta_i$ do not change the dispersion summary, we can assign any value for it. Here we just simply assign $1/|\triangle_{z_{i,t}}|$. Due to the fact that the "explosion" in coefficient only occurs when the difference in regressors takes a small value or is close to zero, based on our experience, we feel it is appropriate to set a maximum value of 0.5 for the boundary, $L_{b_i}$.

The following procedure explains how to find the "optimal" boundary, $L_{b_i}$.

1. Calculate $|\triangle_{z_{i,t}}|$ for each regressor series, $z_{i,t}$, where $\triangle_{z_{i,t}} = z_{i,t} - z_{i,t-1}$.
2. Remove $|\triangle_{z_{i,t}}| = 0$ and calculate the dispersion summary of $1/|\triangle_{z_{i,t}}|$ and find $UIF_{1/|\triangle_{z_{i,t}}|}$.
3. Get the lower boundary by calculating $L_{b_i} = 1/UIF_{1/|\triangle_{z_{i,t}}|}$ or set it at 0.5, which ever is smaller.

## 5. Empirical study

The aim of this section is to compare the forecast performance of the proposed new model with regressors to the state space version of the existing exponential smoothing models without regressors as well as to a linear regression model. By doing this, we can examine the forecast performance of the new model that is constructed based on both dynamic movement and the effect of regressors as compared to the methods that only rely on either dynamic movement or the effect of regressors. In this empirical study, an experiment is performed using some simulated data.

### *5.1. Models under consideration*

The three existing state space structures of the exponential smoothing models (without regressor) considered here are the local level model: ETS(A,N,N), local trend model: ETS(A,A,N) and damped trend model: ETS(A,A$_\text{d}$,N) as described by the following equations.

Model 1: ETS(A,N,N).

$$y_t = \ell_{t-1} + \varepsilon_t \tag{26a}$$
$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t, \tag{26b}$$

whose matrix notation is given by:

$$\boldsymbol{x}_t = \left[\ell_t\right], \ \boldsymbol{w} = \left[1\right], \ \boldsymbol{F} = \left[1\right], \text{ and } \boldsymbol{g} = \left[\alpha\right].$$

Model 2: ETS(A,A,N).

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t \tag{27a}$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \tag{27b}$$
$$b_t = b_{t-1} + \beta\varepsilon_t. \tag{27c}$$

Thus, the matrix notation is given by:

$$\boldsymbol{x}_t = \begin{bmatrix}\ell_t \\ b_t\end{bmatrix}, \ \boldsymbol{w} = \begin{bmatrix}1 \\ 1\end{bmatrix}, \ \boldsymbol{F} = \begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}, \text{ and } \boldsymbol{g} = \begin{bmatrix}\alpha \\ \beta\end{bmatrix};$$

Model 3: ETS(A,$\mathrm{A_d}$,N).

$$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t \tag{28a}$$
$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t \tag{28b}$$
$$b_t = \phi b_{t-1} + \beta\varepsilon_t, \tag{28c}$$

which can be represented in matrix notation by:

$$\boldsymbol{x}_t = \begin{bmatrix}\ell_t \\ b_t\end{bmatrix}, \ \boldsymbol{w} = \begin{bmatrix}1 \\ \phi\end{bmatrix}, \ \boldsymbol{F} = \begin{bmatrix}1 & \phi \\ 0 & \phi\end{bmatrix}, \text{ and } \boldsymbol{g} = \begin{bmatrix}\alpha \\ \beta\end{bmatrix}.$$

Another model included in this analysis is simple linear regression model whose regressor coefficient is time invariant as represented by the following equations. For this empirical study, only one regressor is included into all types of models with regressors.

Model 4: Linear regression (Reg)

$$y_t = b_0 + b_1 z_{1,t} + \varepsilon_t \tag{29}$$

where $b_0$ is an intercept coefficient, $b_1$ is a slope coefficient and $\varepsilon_t$ is Gaussian white noise.

The remaining three models under consideration are state space models with one regressor. Two of them are the new approach of Holt's linear exponential smoothing method augmented with regressor (ESWR) as described by equations (11) and (20). Another model is the exponential smoothing models with a time invariant regressor parameter (ETSX) that proposed by Hyndman et al. (2008) as described by equations (1).

Model 5: ETSX(A,N,N)

$$y_t = \ell_{t-1} + z_{1,t} p_{1,t-1} + \varepsilon_t \tag{30a}$$
$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t \tag{30b}$$
$$p_{1,t} = p_{1,t-1}, \tag{30c}$$

whose matrix notation is given by:

$$\bar{\boldsymbol{x}}_t = \begin{bmatrix} \ell_t \\ p_{1,t} \end{bmatrix}, \ \bar{\boldsymbol{w}}_t = \begin{bmatrix} 1 \\ z_{1,t} \end{bmatrix}, \ \bar{\boldsymbol{F}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \text{and} \ \bar{\boldsymbol{g}} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}.$$

Model 6: ESWR(A,N,N,0,1)

$$y_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \varepsilon_t \tag{31a}$$
$$\ell_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \alpha\varepsilon_t \tag{31b}$$
$$b_{1,t} = b_{1,t-1}, \tag{31c}$$

Model 7: ESWR(A,N,N,1,0)

$$y_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \varepsilon_t \tag{32a}$$
$$\ell_t = \ell_{t-1} + b_{1,t-1}\triangle_{z_{1,t}} + \alpha\varepsilon_t \tag{32b}$$
$$b_{1,t} = \begin{cases} b_{1,t-1} + \beta_1(\varepsilon^+_{1,t-1} + \varepsilon_t)/\triangle^*_{z_{1,t}}, & |\triangle^*_{z_{1,t}}| \geq L_{b_1} \\ b_{1,t-1}, & |\triangle^*_{z_{1,t}}| < L_{b_1} \end{cases} \tag{32c}$$
$$\varepsilon^+_{1,t} = \begin{cases} 0, & |\triangle^*_{z_{1,t}}| \geq L_{b_1} \\ \varepsilon^+_{1,t-1} + \varepsilon_t, & |\triangle^*_{z_{1,t}}| < L_{b_1}, \end{cases} \tag{32d}$$

### 5.2. *Application to simulated data*

Twelve different simulated series are used in this study in which each of them represents a different characteristic of series so that we can examine the forecast performance of each model in different"environments". Each of these simulated series consist 288 observations that divided into two parts of in-sample period (168 observations) for estimation process and out-of-sample period (120 observations) for forecast analysis.

#### 5.2.1. *Simulation procedure*

All twelve simulated series of $y_t$ were obtained by using the linear regression relationship given by:

$$y_t = \beta_0 + \beta_{1,t}z_{1,t} + \varepsilon_t$$

where $\varepsilon_t \sim \text{NID}(\mu_{\varepsilon_t}, \sigma^2_{\varepsilon_t})$ for the first nine series and $\varepsilon_t$ follows a random walk process for the last three series. However, the way the regressor, $z_{1,t}$ and the coefficient, $\beta_{1,t}$ are generated is different for each of the corresponding $y_t$ series.

There are three different ways of generating the coefficient, $\beta_{1,t}$ based on the following equations: (i)

$$\beta_{1,t} = \mu_{\beta_1,t} + \eta_t,$$

where

$$\mu_{\beta_1,t} = \begin{cases} 0.5, & d_t < 0.005 \\ 0.6, & d_t > 0.995 \\ \mu_{\beta_1,t-1}, & 0.005 \leq d_t \leq 0.995, \end{cases}$$

with $\eta_t \sim \text{NID}(0, 0.001^2)$ and $d_t \sim \text{U}(0, 1)$.
(ii)

$$\beta_{1,t} = 0.4 + (t-1)(0.001) + \eta_t,$$

where $\eta_t \sim \text{NID}(0, 0.001^2)$.
(iii)

$$\beta_{1,t} = 0.5 + 0.05 \sin[1 + (t-1)(0.05)] + \eta_t,$$

where $\eta_t \sim \text{NID}(0, 0.001^2)$.

Then, each of these coefficients are applied to four different regressors, $z_{1,t}$ resulting in the total number of twelve simulated series of $y_t$. Below is an explanation of how each simulated series is generated.
(i) Simulations 1-3.
Regressor, $z_{1,t}$ is generated based on the following equation:

$$z_{1,t} = 1 + (t-1)(0.025) + \epsilon_t,$$

where $\epsilon_t \sim \text{NID}(0, 0.01^2)$. Then the $y_t$ series is simulated according to the following linear equation:

$$y_t = 100 + \beta_{1,t} z_{1,t} + \varepsilon_t,$$

where $\varepsilon_t \sim \text{NID}(0, 0.01^2)$.

This is actually a typical regressor that has a stationary differenced series. Thus, it is also common to have this kind series of $y_t$. In addition to that, the rationale of having this kind of series is to investigate whether Holt's linear method augmented with a regressor is able to work well when dealing with small values of the differenced regressor.
(ii) Simulations 4-6.
Regressor, $z_{1,t}$ is generated based on the following equation:

$$z_{1,t} = 100 + 50 \cos\left[1 + (t-1)(0.025)\right].$$

Then the $y_t$ series is simulated according to the following linear equation:

$$y_{1,t} = 100 + \beta_{1,t} z_{1,t} + \varepsilon_t,$$

where $\varepsilon_t \sim \text{NID}(0, 0.01^2)$.

Here, the regressor, $z_{1,t}$ as well as its differenced series are moving around with a cycle process and the differenced series having repeated small (close to zero) values.
(iii) Simulations 7-9.
Regressor, $z_{1,t}$ is generated based on the following equation:

$$z_{1,t} = 0.5 + z_{1,t-1} + (t-1)(0.001) + \epsilon_t,$$

where $z_{1,1} = 1 + \epsilon_1$ and $\epsilon_t \sim \text{NID}(0, 0.01^2)$. Then the $y_t$ series is simulated according to the linear equation:

$$y_{1,t} = 100 + \beta_{1,t} z_{1,t} + \varepsilon_t,$$

where $\varepsilon_t \sim \text{NID}(0, 0.01^2)$.

In this case, the regressor $z_{1,t}$ is set to have rapid growth over time so that the differenced series is non-stationary. The goal of the experiment here is to see whether any proposed model is able to deal with this kind of regressor whose differenced series consists of only positive values.
(iv) Simulations 10-12.
Regressor, $z_{1,t}$ is generated based on the following equation:

$$z_{1,t} = 1 + (t-1)(0.025) + \epsilon_t,$$

where $\epsilon_t \sim \text{NID}(0, 0.01^2)$. Then the $y_t$ series is simulated according to the following linear equation:

$$y_t = 100 + \beta_{1,t} z_{1,t} + \varepsilon_t,$$

with $\varepsilon_t = \varepsilon_{t-1} + \nu_t$, where $\nu_t \sim \text{NID}(0, 0.01^2)$.
The last three series are similar to the first three series, however, $\varepsilon_t$ here is generated via a random walk process.

*5.2.2. Empirical results*

The evaluation of forecast performance is done using five accuracy measures, namely Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE). MASE is calculated based on formula given by Hyndman & Koehler (2006). Evaluation is done based on 1-step ahead forecasts for 120 out-of-sample observations. The results are shown in Tables 1-5.

Clearly, the new approach of Holt's linear model augmented with a time varying parameter of the regressor appears to be the best in term of overall performance where it managed to outperform all other models in four out of five accuracy measures. For the case of Root Mean Squared Error, even though it appears to be the third best, however, it is not so much different to the first and the second one. Obviously MSE and RMSE are related in the sense that one is a monotonic transformation of the other. However, when averaged over the 12 different generating processes, they do produce slightly different rankings.

One of our major concerns in using the proposed new model with time varying parameter of regressor (ESWR) is that whether small values of the differenced regressor could cause the updating process for the growth term to be unstable. Fortunately, empirical results for simulated series 1-6 and 10-12 have shown that this approach has performed quite well. This also suggests that the proposed switching procedure outlined in section (4) is adequate to handle the possible explosion problem in the growth component.

The performance of time invariant parameter models are also not too bad with the model proposed by Hyndman et al. (2008) performing slightly better than its counterpart of the augmented Holt's linear model. The linear regression model on the other hand, performs relatively poorly as it turns out to be the worst in all of the five accuracy measures. Among the three existing exponential smoothing models without a regressor, both local trend model, ETS(A,A,N) and local damped trend model ETS(A,$\text{A}_\text{d}$,N) perform quite well as they managed to secure the second and third place in most cases.

Table 1: Forecast accuracy measure (MSE)

| S | ETS: ANN | ETS: AAN | ETS: $AA_dN$ | Reg | ETSX: ANN | ESWR: ANN0,1 | ESWR: ANN1,0 |
|---|---|---|---|---|---|---|---|
| 1 | 0.0012 | 0.0010 | 0.0010 | 0.1956 | 0.0008 | 0.0008 | 0.0008 |
| 2 | 0.0007 | 0.0004 | 0.0004 | 0.0036 | 0.0004 | 0.0004 | 0.0004 |
| 3 | 0.0017 | 0.0007 | 0.0007 | 0.3768 | 0.0014 | 0.0014 | 0.0009 |
| 4 | 1.7851 | 1.5468 | 1.5481 | 170.9224 | 1.4757 | 1.4762 | 1.4762 |
| 5 | 0.1202 | 0.0238 | 0.0238 | 3.8857 | 0.0178 | 0.0178 | 0.0194 |
| 6 | 0.1544 | 0.0435 | 0.0436 | 86.7575 | 0.0979 | 0.0986 | 0.0567 |
| 7 | 1.6060 | 1.0865 | 1.0882 | 206.0043 | 1.0898 | 1.0904 | 1.0904 |
| 8 | 0.1334 | 0.0273 | 0.0272 | 1.7677 | 0.0307 | 0.0307 | 0.0356 |
| 9 | 0.2580 | 0.0407 | 0.0407 | 31.6546 | 0.1040 | 0.1040 | 0.0411 |
| 10 | 0.0008 | 0.0006 | 0.0008 | 0.1914 | 0.0005 | 0.0005 | 0.0005 |
| 11 | 0.0008 | 0.0005 | 0.0005 | 0.0026 | 0.0005 | 0.0005 | 0.0005 |
| 12 | 0.0014 | 0.0006 | 0.0006 | 0.3558 | 0.0012 | 0.0012 | 0.0007 |
| Av | 0.3386 | 0.2310 | 0.2313 | 41.8432 | 0.2351 | 0.2352 | 0.2269 |
| R | 6 | 2 | 3 | 7 | 4 | 5 | 1 |

Notes: S=Simulated series, Av=Average, R=Rank.

Table 2: Forecast accuracy measure (MAE)

| S | ETS: ANN | ETS: AAN | ETS: $AA_dN$ | Reg | ETSX: ANN | ESWR: ANN0,1 | ESWR: ANN1,0 |
|---|---|---|---|---|---|---|---|
| 1 | 0.0270 | 0.0247 | 0.0247 | 0.3642 | 0.0232 | 0.0232 | 0.0232 |
| 2 | 0.0219 | 0.0165 | 0.0155 | 0.0521 | 0.0168 | 0.0169 | 0.0167 |
| 3 | 0.0334 | 0.0221 | 0.0221 | 0.5242 | 0.0308 | 0.0308 | 0.0243 |
| 4 | 0.5476 | 0.3940 | 0.3727 | 12.5393 | 0.2979 | 0.2986 | 0.2985 |
| 5 | 0.2930 | 0.1242 | 0.1241 | 1.9439 | 0.1055 | 0.1055 | 0.1081 |
| 6 | 0.3238 | 0.1646 | 0.1646 | 7.3136 | 0.2570 | 0.2577 | 0.1840 |
| 7 | 0.5052 | 0.2865 | 0.2891 | 13.2322 | 0.2669 | 0.2672 | 0.2672 |
| 8 | 0.2861 | 0.1189 | 0.1189 | 1.1633 | 0.1333 | 0.1333 | 0.1491 |
| 9 | 0.4057 | 0.1543 | 0.1549 | 4.7396 | 0.2568 | 0.2568 | 0.1534 |
| 10 | 0.0237 | 0.0204 | 0.0233 | 0.4043 | 0.0186 | 0.0186 | 0.0186 |
| 11 | 0.0219 | 0.0181 | 0.0181 | 0.0436 | 0.0171 | 0.0171 | 0.0171 |
| 12 | 0.0305 | 0.0203 | 0.0202 | 0.4975 | 0.0277 | 0.0277 | 0.0208 |
| Av | 0.2100 | 0.1137 | 0.1124 | 3.5682 | 0.1210 | 0.1211 | 0.1068 |
| R | 6 | 3 | 2 | 7 | 4 | 5 | 1 |

Notes: S=Simulated series, Av=Average, R=Rank.

Table 3: Forecast accuracy measure (RMSE)

| S | ETS: ANN | ETS: AAN | ETS: $AA_dN$ | Reg | ETSX: ANN | ESWR: ANN0,1 | ESWR: ANN1,0 |
|---|---|---|---|---|---|---|---|
| 1 | 0.0344 | 0.0311 | 0.0311 | 0.4423 | 0.0291 | 0.0291 | 0.0291 |
| 2 | 0.0264 | 0.0203 | 0.0191 | 0.0602 | 0.0206 | 0.0206 | 0.0208 |
| 3 | 0.0411 | 0.0270 | 0.0269 | 0.6139 | 0.0378 | 0.0378 | 0.0307 |
| 4 | 1.3361 | 1.2437 | 1.2442 | 13.0737 | 1.2148 | 1.2150 | 1.2150 |
| 5 | 0.3467 | 0.1542 | 0.1542 | 1.9712 | 0.1334 | 0.1333 | 0.1391 |
| 6 | 0.3929 | 0.2085 | 0.2087 | 9.3144 | 0.3128 | 0.3140 | 0.2382 |
| 7 | 1.2673 | 1.0424 | 1.0432 | 14.3528 | 1.0440 | 1.0442 | 1.0442 |
| 8 | 0.3653 | 0.1653 | 0.1648 | 1.3296 | 0.1751 | 0.1751 | 0.1888 |
| 9 | 0.5079 | 0.2019 | 0.2017 | 5.6262 | 0.3225 | 0.3225 | 0.2027 |
| 10 | 0.0286 | 0.0252 | 0.0283 | 0.4375 | 0.0230 | 0.0230 | 0.0230 |
| 11 | 0.0274 | 0.0223 | 0.0223 | 0.0509 | 0.0214 | 0.0214 | 0.0214 |
| 12 | 0.0379 | 0.0249 | 0.0248 | 0.5965 | 0.0341 | 0.0341 | 0.0266 |
| Av | 0.3677 | 0.2639 | 0.2641 | 3.9891 | 0.2807 | 0.2808 | 0.2650 |
| R | 6 | 1 | 2 | 7 | 4 | 5 | 3 |

Notes: S=Simulated series, Av=Average, R=Rank.

Table 4: Forecast accuracy measure (MAPE)

| S | ETS: ANN | ETS: AAN | ETS: $AA_dN$ | Reg | ETSX: ANN | ESWR: ANN0,1 | ESWR: ANN1,0 |
|---|---|---|---|---|---|---|---|
| 1 | 0.0247 | 0.0226 | 0.0226 | 0.3319 | 0.0212 | 0.0212 | 0.0212 |
| 2 | 0.0205 | 0.0155 | 0.0146 | 0.0488 | 0.0158 | 0.0158 | 0.0156 |
| 3 | 0.0310 | 0.0205 | 0.0204 | 0.4880 | 0.0285 | 0.0285 | 0.0225 |
| 4 | 0.3153 | 0.2193 | 0.2102 | 6.9469 | 0.1691 | 0.1695 | 0.1694 |
| 5 | 0.1902 | 0.0798 | 0.0797 | 1.2545 | 0.0680 | 0.0680 | 0.0694 |
| 6 | 0.1989 | 0.0991 | 0.0991 | 4.3755 | 0.1548 | 0.1552 | 0.1105 |
| 7 | 0.3048 | 0.1739 | 0.1755 | 7.5870 | 0.1612 | 0.1614 | 0.1614 |
| 8 | 0.1777 | 0.0735 | 0.0735 | 0.7039 | 0.0821 | 0.0821 | 0.0916 |
| 9 | 0.2319 | 0.0900 | 0.0904 | 2.9235 | 0.1486 | 0.1486 | 0.0895 |
| 10 | 0.0217 | 0.0186 | 0.0213 | 0.3686 | 0.0170 | 0.0170 | 0.0170 |
| 11 | 0.0205 | 0.0170 | 0.0170 | 0.0409 | 0.0161 | 0.0161 | 0.0161 |
| 12 | 0.0282 | 0.0188 | 0.0187 | 0.4628 | 0.0256 | 0.0256 | 0.0192 |
| Av | 0.1304 | 0.0707 | 0.0702 | 2.1277 | 0.0757 | 0.0758 | 0.0670 |
| R | 6 | 3 | 2 | 7 | 4 | 5 | 1 |

Notes: S=Simulated series, Av=Average, R=Rank.

Table 5: Forecast accuracy measure (MASE)

| S | ETS: ANN | ETS: AAN | ETS: $AA_dN$ | Reg | ETSX: ANN | ESWR: ANN0,1 | ESWR: ANN1,0 |
|---|---|---|---|---|---|---|---|
| 1 | 0.9023 | 0.8269 | 0.8255 | 12.1742 | 0.7744 | 0.7742 | 0.7742 |
| 2 | 1.1049 | 0.8351 | 0.7853 | 2.6337 | 0.8512 | 0.8533 | 0.8416 |
| 3 | 1.2456 | 0.8241 | 0.8220 | 19.5346 | 1.1459 | 1.1459 | 0.9053 |
| 4 | 0.9791 | 0.7044 | 0.6664 | 22.4178 | 0.5327 | 0.5339 | 0.5337 |
| 5 | 0.7961 | 0.3375 | 0.3373 | 5.2823 | 0.2868 | 0.2867 | 0.2939 |
| 6 | 0.6363 | 0.3234 | 0.3235 | 14.3718 | 0.5051 | 0.5063 | 0.3615 |
| 7 | 1.4408 | 0.8171 | 0.8246 | 37.7383 | 0.7612 | 0.7620 | 0.7620 |
| 8 | 1.1610 | 0.4825 | 0.4827 | 4.7214 | 0.5411 | 0.5411 | 0.6052 |
| 9 | 1.3704 | 0.5211 | 0.5232 | 16.0098 | 0.8673 | 0.8673 | 0.5181 |
| 10 | 0.5741 | 0.4934 | 0.5649 | 9.7846 | 0.4495 | 0.4496 | 0.4496 |
| 11 | 1.1822 | 0.9784 | 0.9781 | 2.3578 | 0.9260 | 0.9256 | 0.9256 |
| 12 | 1.2106 | 0.8079 | 0.8023 | 19.7561 | 1.0990 | 1.0990 | 0.8254 |
| Av | 1.0503 | 0.6626 | 0.6613 | 13.8985 | 0.7283 | 0.7287 | 0.6497 |
| R | 6 | 3 | 2 | 7 | 4 | 5 | 1 |

Notes: S=Simulated series, Av=Average, R=Rank.

## 6. Conclusions

To sum up, the proposed Holt's linear exponential smoothing model augmented with a regressor whose coefficient is time varying was found to be the best forecasting procedure for the empirical analysis. This appears to be because of its ability to capture the dynamic movement of the regressor coefficient. Similar models with a time invariant regressor parameter are also found to perform quite well.


## Acknowledgements

We would like to acknowledge Ralph D. Snyder for helping us in constructing the state space version of the proposed model in this paper. We are also want to thank Rob J. Hyndman for giving us valuable suggestions in setting up this new model.


## References


Athanasopoulos, G. & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. *Tourism Management*, *29*(1), 19–31.

Holt, C. C. (1957). Forecasting seasonal and trends by exponentially weighted moving averages. *Office of Naval Research, Research Memorandum, No. 52.*

Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679–688.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach.* Berlin: Springer-Verlag.

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, *44*(1), 82–99.

Li, G., Wong, K. K. F., Song, H., & Witt, S. F. (2006). Tourism demand forecasting: a time varying parameter error correction model. *Journal of Travel Research*, *45*(2), 175–185.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simmons, L. F. (1993). The M2-Competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*(1), 5–22.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: methods and application*. New York, USA: John Wiley & Sons, 3rd edn.

Song, H., Liu, X., & Romilly, P. (1996). A time varying parameter approach to the Chinese aggregate consumption function. *Economics of Planning*, *29*(3), 185–203.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*, 324–342.

Wolff, C. C. P. (1987). Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models. *Journal of Business and Economic Statistics*, *5*(1), 87–97.