Unmasking the Theta method

Rob J Hyndman[1], Baki Billah[1]

27 September 2001

Abstract: The "Theta method" of forecasting performed particularly well in the M3-competition and is therefore of interest to forecast practitioners. The original description of the method given by Assimakopoulos and Nikolopoulos (2000) involves several pages of algebraic manipulation. We show that the method can be expressed much more simply and that the forecasts obtained are equivalent to simple exponential smoothing with drift.

Keywords: exponential smoothing, forecasting competitions, state space models.

1 Introduction

The "Theta method" of forecasting was introduced by Assimakopoulos and Nikolopoulos (2000), hereafter referred to as A&N. Their description of the method is complicated, potentially confusing, and involves several pages of algebra. However, the method performed particularly well in the M3-competition (Makridakis & Hibon, 2000), and is therefore of interest to forecast practitioners.

We examine the Theta method and show that it can be expressed much more simply than in A&N; furthermore we show that the forecasts obtained are equivalent to simple exponential smoothing (SES) with drift. Using this equivalence, we derive appropriate prediction intervals for the method based on a state space model underlying SES with drift. Finally, we show that SES with drift can produce better forecasts than the Theta method if the parameters are optimized using a maximum likelihood approach.

Section 2 reproduces the main results from A&N using a different (and much simpler) notation. We obtain an explicit expression for point forecasts in Section 3 and show that these are equivalent to the point forecasts from SES with drift. In Section 4, we describe a state space model with equivalent forecasts, thus enabling the computation of prediction intervals and likelihood estimates. Finally, in Section 5 we compare the Theta method with fully optimized SES with drift by applying both methods to the annual data from the M3-competition. The Appendix contains a list of equivalencies in the notation in A&N and in this paper.

[1]Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia. Correspondence to Rob.Hyndman@buseco.monash.edu.au.

2 Theta Method

Let $\{X_1, \dots, X_n\}$ denote the observed univariate time series. From this series A&N construct a new series $\{Y_{1,\theta}, \dots, Y_{n,\theta}\}$ such that for $t = 3, \dots, n$

$$Y''_{t,\theta} = \theta X''_t \qquad (1)$$

where X''_t denotes the second difference of X_t and $Y''_{t,\theta}$ denotes the second difference of $Y_{t,\theta}$. We note that (1) is a second-order difference equation and has the solution (see Kelley, 2000)

$$Y_{t,\theta} = a_\theta + b_\theta(t-1) + \theta X_t \qquad (2)$$

where a_θ and b_θ are constants and $t = 1, \dots, n$. Thus $Y_{t,\theta}$ is equivalent to a linear function of X_t with a linear trend added. A&N call $Y_{t,\theta}$ a "theta line".

For a fixed θ, A&N find the values of $Y_{1,\theta}$ and $Y_{2,\theta} - Y_{1,\theta}$ which minimize the sum of squared differences

$$\sum_{i=1}^{t} [X_t - Y_{t,\theta}]^2 = \sum_{i=1}^{t} \left[(1-\theta)X_t - a_\theta - b_\theta(t-1)\right]^2.$$

This is equivalent to minimizing the above sum of squares with respect to a_θ and b_θ. Thus, it is a simple regression of $(1-\theta)X_t$ against time $t-1$. Therefore the solution is

$$\begin{aligned} \hat{b}_\theta &= \frac{6(1-\theta)}{n^2-1}\left(\frac{2}{n}\sum_{t=1}^{n} tX_t - (n+1)\bar{X}\right) \\ \text{and} \quad \hat{a}_\theta &= (1-\theta)\bar{X} - \hat{b}_\theta(n-1)/2. \end{aligned}$$

The equivalent results are derived in more than two pages of algebra by A&N (spanning their Appendices A and B). Note, in particular, that when $\theta = 0$, $\hat{a}_0$ and $\hat{b}_0$ are the parameters for the linear trend fitted to the original series $\{X_t\}$. We also observe that $b_\theta/(1-\theta)$ is independent of θ (which is the main result given in A&N Appendix A).

The mean value of the new series is

$$\bar{Y}_\theta = \hat{a}_\theta + \hat{b}_\theta(n-1)/2 + \theta\bar{X} = \bar{X},$$

the same result A&N derived (in their Appendix A) in one-third of a page using their notation, but note that A&N give the result erroneously as $\bar{Y}_i = \bar{X}_i$. Furthermore, it is easy to see that $\frac{1}{2}[Y_{t,\theta} + Y_{t,2-\theta}] = X_t$ since $\hat{a}_\theta + \hat{a}_{2-\theta} = 0$ and $\hat{b}_\theta + \hat{b}_{2-\theta} = 0$; this result takes up about half a page in Appendix B of A&N.

Forecasts from the Theta method are obtained by a weighted average of forecasts of $Y_{t,\theta}$ for different values of θ. However, A&N only explain how to get forecasts for $\theta = 0$ and $\theta = 2$, the set-up they used in the M3-competition. In this case, the h-step ahead forecast based on observations $X_1, \dots, X_n$ is given by

$$\hat{X}_n(h) = \tfrac{1}{2}[\hat{Y}_{n,0}(h) + \hat{Y}_{n,2}(h)]$$

where $\hat{Y}_{n,0}(h)$ is obtained by extrapolating the linear part of (2) and $\hat{Y}_{n,2}(h)$ is obtained using SES on the series $\{Y_{t,2}\}$. Hence,

$$\hat{Y}_{n,0}(h) = \hat{a}_0 + \hat{b}_0(n+h-1) \tag{3}$$

and (see Makridakis, Wheelwright & Hyndman, 1998, p.149)

$$\hat{Y}_{n,2}(h) = \alpha \sum_{i=0}^{n-1} (1-\alpha)^i Y_{n-i,2} + (1-\alpha)^n Y_{1,2} \tag{4}$$

where α is the smoothing parameter for the SES. (Note that SES forecasts are equivalent for all h.)

So far we have simply shown how A&N's results can be replicated much more easily using our notation. The rest of our paper gives new results concerning this forecasting methodology.

3 Point forecasts

The above results can be combined to obtain a simple expression for the forecasts $\hat{X}_n(h)$. From (4) we obtain

$$\begin{aligned}\hat{Y}_{n,2}(h) &= \alpha \sum_{i=0}^{n-1} (1-\alpha)^i \left[\hat{a}_2 + \hat{b}_2(n-i-1) + 2X_{n-i}\right] + (1-\alpha)^n(\hat{a}_2 + 2X_1) \\ &= \hat{a}_2 + \hat{b}_2\left[n - \frac{1}{\alpha} + \frac{(1-\alpha)^n}{\alpha}\right] + 2\tilde{X}_n(h)\end{aligned}$$

where $\tilde{X}_n(h)$ is the SES forecast of the series $\{X_t\}$. Noting that $\hat{a}_2 = -\hat{a}_0$ and $\hat{b}_2 = -\hat{b}_0$, we obtain

$$\hat{X}_n(h) = \tilde{X}_n(h) + \tfrac{1}{2}\hat{b}_0\left(h - 1 + \frac{1}{\alpha} - \frac{(1-\alpha)^n}{\alpha}\right). \tag{5}$$

For large n, this can be written as

$$\hat{X}_n(h) = \tilde{X}_n(h) + \tfrac{1}{2}\hat{b}_0(h - 1 + 1/\alpha).$$

Thus it is SES with an added trend plus a constant, where the slope of the trend is half that of the fitted trend line through the original time series $\{X_t\}$.

4 Underlying stochastic models

A&N do not give an underlying stochastic model for their forecasting method. However, it is possible to find such a model using a state space approach. We initialize the model by setting $X_1 = \ell_1$ and then for $t = 2, 3, \ldots$, let

$$X_t = \ell_{t-1} + b + \varepsilon_t \tag{6}$$

$$\text{and} \quad \ell_t = \ell_{t-1} + b + \alpha\varepsilon_t \tag{7}$$

where $\{\varepsilon_t\}$ is Gaussian white noise with mean zero and variance σ^2.

Then X_t follows a state space model which gives forecasts equivalent to SES with drift. This is a special case of Holt's method with the smoothing parameter for the slope set to zero. Note that X_t can also be written as

$$X_t = X_{t-1} + b + (\alpha - 1)\varepsilon_{t-1} + \varepsilon_t,$$

that is an ARIMA(0,1,1) process with drift (see Box, Jenkins & Reinsel, 1994, Appendix A4.2).

Now, the point forecasts for $\{X_t\}$ are given by (see, e.g., Hyndman et al., 2001)

$$\hat{X}_n(h) = \ell_n + hb.$$

Further, we note that

$$\hat{X}_n(1) \mid X_1, \ldots, X_t \quad = \quad X_t + b + (\alpha - 1)\varepsilon_t \tag{8}$$

$$\text{and that} \quad \varepsilon_t \quad = \quad X_t - X_{t-1} - b + (1 - \alpha)\varepsilon_{t-1}. \tag{9}$$

By repeatedly substituting (9) into (8), we obtain

$$\begin{aligned}
\hat{X}_n(1) \mid X_1, \ldots, X_n &= \alpha \sum_{i=0}^{n-1} (1-\alpha)^i X_{n-i} + (1-\alpha)^n X_1 + \frac{b}{\alpha}[1 - (1-\alpha)^n] + (1-\alpha)^n \varepsilon_1 \\
&= \tilde{X}_n(1) + \frac{b}{\alpha}[1 - (1-\alpha)^n]
\end{aligned}$$

where $\tilde{X}_n(1)$ is the SES forecast since $\varepsilon_1 = 0$ (which follows from the initialization $X_1 = \ell_1$). Similarly, the h-step ahead forecast is:

$$\hat{X}_n(h) \mid X_1, \ldots, X_n \quad = \quad \tilde{X}_n(1) + b\left[h - 1 + \tfrac{1}{\alpha} - \tfrac{(1-\alpha)^n}{\alpha}\right].$$

Thus, we obtain identical point forecasts as for the Theta method (5) if $b = \hat{b}_0/2$.

Furthermore, the state space model enables us to obtain maximum likelihood estimates of b and prediction intervals provided the model assumptions are satisfied. For example, it can be shown that 95% prediction intervals for h period ahead forecasts are given by

$$\hat{X}_n(h) \pm 1.96\sigma\sqrt{(h-1)\alpha^2 + 1}.$$

Equivalent results are obtained using the ARIMA(0,1,1) model with drift. (As with most time series forecasts, this formula does not include the variation due to estimation error and will therefore give intervals which are too narrow.)

5 Application to Annual M3 Competition Data

The preceding analysis suggests we may be able to obtain better forecasts if we optimize the value of b rather than setting it equal to $\hat{b}_0/2$. To evaluate this idea, we apply the

Table 1: *Average SMAPE for the annual M3 competition data*

Methods	Forecasting Horizons						Average	
	1	2	3	4	5	6	1 to 4	1 to 6
(1) A&N Theta method	8.0	12.2	16.7	19.2	21.7	23.6	14.02	16.90
(2) Recalculated Theta method	8.2	12.3	16.4	18.6	21.2	23.0	13.89	16.62
(3) SES with drift	7.9	12.1	18.6	17.2	20.6	22.9	13.95	16.55

model to the 645 annual series from the M3 competition (Makridakis and Hibon, 2000). We computed forecasts up to 6 steps ahead and then we computed the symmetric mean absolute percentage error (SMAPE) as in Makridakis and Hibon (2000). The results are presented in Table 1.

Table 1 shows the average SMAPE for: (1) the original A&N forecasts (as tabulated by Makridakis & Hibon, 2000, Table 13); (2) our implementation of the Theta method as described in A&N; and (3) SES with drift forecasts based on the state space model (6) and (7). Our results in (2) differ slightly from those of A&N, probably because we initialized the SES of $Y_t(2)$ differently, and possibly also because we estimated the smoothing parameter α differently (we used a likelihood approach). The SES with drift forecasts based on the state space model (i.e., Method (3) in Table 1) generally perform better than the Theta method forecasts because the drift parameter b was optimized.

In method (3), we treated ℓ_1 as an unknown parameter and then optimized the likelihood of the state space model (6) and (7) over the parameters ℓ_1, b and α as described in Ord, Koehler and Snyder (1997). For estimating the SES method in method (2), we use the same likelihood method (except b was set to 0). To initialize the optimizations, we use the procedure outlined in Hyndman et al. (2001). The value of α was constrained to lie between 0.1 and 0.99.

The SES with drift method performed relatively badly for series N0529 (which contains a large level shift), particularly for forecasting horizon 3. If series N0529 is omitted, the state space SMAPE for $h = 3$ becomes 16.3.

6 Conclusion

We have demonstrated that the Theta method proposed by A&N is simply a special case of SES with drift where the drift parameter is half the slope of the linear trend fitted to the data. We have also demonstrated that prediction intervals and likelihood-based estimation of the parameters can be obtained using a state space model.

References

ASSIMAKOPOULOS, V. and K. NIKOLOPOULOS (2000) The theta model: a decomposition approach to forecasting. *International Journal of Forecasting* **16**, 521–530.

BOX, G.E.P., G.M. JENKINS, and G.C. REINSEL (1994) *Time series analysis: forecasting and control,* 3rd ed., Prectice Hall: New Jersey.

HYNDMAN, R.J., A.B. KOEHLER, R.D. SNYDER, and S. GROSE (2001) A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting*, to appear.

KELLEY, W.G. (2000) *Difference equations : an introduction with applications*, 2nd ed., Academic: London.

MAKRIDAKIS, S., and M. HIBON (2000) The M3-Competitions: results, conclusions and implications, *International Journal of Forecasting*, **16**, 451–476.

MAKRIDAKIS, S., S.C. WHEELWRIGHT, and R.J. HYNDMAN (1998) *Forecasting Methods and Applications*, 3rd edition, John Wiley & Sons: New York.

ORD, J.K., A.B. KOEHLER and R.D. SNYDER (1997) Estimation and prediction for a class of dynamic nonlinear statistical models, *J. Amer. Statist. Assoc.*, **92**, 1621–1629.

Appendix: Notation

Relationships between the notation of this paper and A&N are given in the following table.

Hyndman & Billah	Assimakopoulos & Nikolopoulos
X_t	X_t
X''_t	X''_{data} or X''_t
$Y_{t,\theta}$	Y_t
$Y''_{t,\theta}$	X''_{new}
a_θ	$Y_1 - \theta X_1$
b_θ	$Y_2 - Y_1 - \theta(X_2 - X_1)$
$b_\theta/(1-\theta)$	b_θ or b_n