The Korean Journal of Applied Statistics
2022, Vol. 35, No. 3, 385–393

DOI: https://doi.org/10.5351/KJAS.2022.35.3.385
Print ISSN1225-066X / Online ISSN 2383-5818

Performance for simple combinations of univariate forecasting models

Seonhong Lee[a], Byeongchan Seong[1,a]

[a]Department of Applied Statistics, Chung-Ang University

Abstract

In this paper, we consider univariate time series models that are well known in the field of forecasting and we study on forecasting performance for their simple combinations. The univariate time series models include exponential smoothing methods and ARIMA (autoregressive integrated moving average) models, their extended models, and non-seasonal and seasonal random walk models, which is frequently used as benchmark models for forecasting. The median and mean are simply used for the combination method, and the data set used for performance evaluation is M3-competition data composed of 3,003 various time series data. As results of evaluating the performance by sMAPE (symmetric mean absolute percentage error) and MASE (mean absolute scaled error), we assure that the simple combinations of the univariate models perform very well in the M3-competition dataset.

Keywords: univariate time series models, exponential smoothing methods, ARIMA, M3-competition

1. 서론

대규모의 시계열 자료들을 모형화하고 예측해야 하는 상황이 일상화되면서 좀더 정확하고 로버스트하며 신뢰성있는 예측 방법의 개발은 중요한 이슈 중 하나이다. 예를 들어, 산업 현장에서 수많은 제품이나 부품들에 대한 재고 관리 및 생산 계획 등은, 정확한 예측 방법이 기업의 이익에 큰 영향을 미치게 된다. 이러한 필요에 발맞춰 최근 여러 시계열 예측 대회들(forecasting competitions)은 우수한 예측 방법론 발전에 일조하고 있다.

예측 방법론의 중심에는 1950년대부터 발전하기 시작한 지수평활 모형(exponential smoothing models; ETS) (Brown, 1959; Holt, 1957)과 1970년대부터 발전을 가속한 ARIMA 모형(autoregressive integrated moving average models) (Box와 Jenkins, 1970) 같은 단변량 시계열 모형이 있다. 이 두 방법론은 다양한 확장된 형태로 발전을 거듭해 오고 있다. 시계열 구성 성분 또는 분해법 토대로 발전된 지수평활 모형은 최근 상태공간 모형(state space models)의 발전에 힘입어 CES 모형(complex exponential smoothing models) (Svetunkov와 Kourentzes, 2018) 및 DOTM 모형(dynamic optimized theta models) (Fiorucci 등, 2016)과 같은 형태로도 발전했으며, 자기상관함수 또는 시차 변수들을 토대로 모형화되는 ARIMA 모형은 신경망 모형(neural network models)과도 융합되었다. 또한, Livera 등 (2011)은 지수평활 모형과 ARIMA 모형을 조합한 형태로 TBATS (trigonometric Box-Cox ARMA trend seasonal) 모형을 개발하였다.

이러한 개별 단변량 시계열 예측 모형들은 각각 장단점들을 모두 가지고 있으며, 전문가들 조차도 상황에 맞게 그 방법론을 선택하여 사용하기 쉽지 않다. 최근 예측 방법론에서는 조합 또는 결합 모형, 앙상블 모형의

This research was supported by the Chung-Ang University Research Scholarship Grants in 2020.
[1] Corresponding author: Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: bcseong@cau.ac.kr

Published 30 June 2022 / journal homepage: http://kjas.or.kr

가치를 높이 평가하고 있다 (Kolassa, 2011). 왜냐하면 조합 결합 모형이 다양한 상황에서 더 높은 예측력을 보여줄 뿐만 아니라 좀더 로버스트한 결과를 보여주고 있기 때문이다. 최근 Petropoulos와 Svetunkov (2020)는 간단한 4개의 단변량 시계열 예측 방법의 결합이 M4- competition에서 우수한 예측력을 보여줌을 입증하였다.

본 연구에서는 Petropoulos와 Svetunkov (2020)의 확장된 형태로서 자주 사용되는 7개의 단변량 예측 모형을 고려하고 이들의 단순 조합이 시계열 예측력에 어떠한 성과를 보여주는지 평가하고자 한다. 본 논문의 구성은 다음과 같다. 2절에서 대표적인 단변량 시계열 예측 모형들을 살펴보고, 3절에서는 그 예측력을 평가한다. 4절에서 결론을 맺는다.

2. 주요 단변량 시계열 모형

본 절에서 대표적인 단변량 시계열 예측 모형 및 특징들을 구체적으로 살펴본다.

2.1. 지수평활 모형

지수평활 ETS 모형은 시계열 분해 기법에서 사용되는 시계열 성분들의 조합으로 시계열 자료를 모형화하는 방법이다. 최근 상태공간모형 및 관련 이론의 발전으로 이와 결합하여 다양한 형태의 모형으로 발전하고 있다 (Hyndman과 Athanasopoulos, 2018). 시계열 자료 y_t에 대한 지수평활모형의 형태는 다음과 같다.

$$y_t = f(\ell_t, b_t, s_t, e_t) \quad \text{for } e_t \overset{iid}{\sim} N\left(0, \sigma^2\right),$$

여기서, ℓ_t는 시계열 자료의 단기적 절편 또는 수준을 나타내는 성분(level component)이며 b_t는 단기적 기울기 또는 추세 성분(trend component), s_t는 계절 성분(seasonal component)을 나타낸다. $f(\cdot)$의 형태는 각 성분들의 결합 방법을 결정한다. 각 성분들은 오차항 e_t와 평활상수(smoothing constants) α, β, γ에 의하여 다음과 같은 방정식에 따라서 움직인다.

$$\begin{aligned}
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha e_t, \\
b_t &= b_{t-1} + \beta e_t, \\
s_t &= s_{t-m} + \gamma e_t
\end{aligned}$$

단, m은 계절 주기(seasonal period)를 나타낸다. 대표적인 지수평활모형은 가법 모형과 승법 모형으로서 그 형태는 각각 다음과 같다.

$$\begin{aligned}
y_t &= \ell_{t-1} + b_{t-1} + s_{t-m} + e_t, \\
y_t &= (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + e_t)
\end{aligned}$$

추세가 완화된 모형이 될 경우, b_t는 ϕb_t로 대체된다; 단, $0 < \phi < 1$. 지수평활 모형은 차분없이 비정상 시계열을 그대로 모형화하기 때문에 실용적이다. 근본적으로 박스-젠킨스의 ARIMA 모형의 경우 차분(differencing)을 통한 정상성(stationarity)을 기반으로 모형화 및 예측이 이루어지기 때문에 그 과정이 자연스럽지 않다. 실제 경제 금융 및 각종 사회 분야의 시계열 자료는 어떤 차분을 하더라도 정상성을 완벽하게 만족한다고 할 수 없기 때문이다 (Commandeur와 Koopman, 2007). 지수평활 모형은 시계열을 구성 성분별로 분해하는 역할도 하기 때문에 계절 조정(seasonal adjustment)에도 쉽게 활용할 수 있으며 시계열 자료가 복잡한 계절성을 가질 경우에도 이용 가능한 장점이 있다.

2.2. CES 모형

CES 모형은 ETS 모형의 단점을 극복하기 위한 모형으로서 시계열 구성 성분들 간의 상호 작용이 허용된다. 모형식은 다음과 같다.

$$
\begin{aligned}
y_t &= \ell_{0,t-1} + \ell_{1,t-m} + e_t,\\
\ell_{0,t} &= \ell_{0,t-1} - (1-\alpha_1)c_{0,t-1} + (\alpha_0 - \alpha_1)e_t,\\
c_{0,t} &= \ell_{0,t-1} + (1-\alpha_0)c_{0,t-1} + (\alpha_0 + \alpha_1)e_t,\\
\ell_{1,t} &= \ell_{1,t-m} - (1-\beta_1)c_{1,t-m} + (\beta_0 - \beta_1)e_t,\\
c_{1,t} &= \ell_{1,t-m} + (1-\beta_0)c_{1,t-m} + (\beta_0 + \beta_1)e_t.
\end{aligned}
$$

즉, 4개의 상태변수들 $\ell_{0,t}, \ell_{1,t}, c_{0,t}, c_{1,t}$이 모형식을 구성하고 있으며 관련 모수는 $\alpha_0, \alpha_1, \beta_0, \beta_1$이다. 비계절형 모형인 경우는 위의 식에서, $\ell_{1,t}$와 $c_{1,t}$를 제거하면 된다. CES 모형은 지수평활 모형과 달리 수준과 추세가 따로 구분되어 있지 않고 $\ell_{\cdot,t}$와 $c_{\cdot,t}$는 서로 상호작용하면서 수준과 추세 역할을 동시에 하고 있다. 또한, 지수평활 모형과 달리 계절항의 표준화가 필요치 않다는 장점이 있다. 위의 모형은 ARIMA$(2,0,2m+2)(2,0,0)_m$ 모형 형태로 변환 가능하며 이 경우 정상성 조건은 수반하지 않는다.

2.3. ARIMA 모형

ARIMA 모형은 시계열 자료의 종속 구조를 나타내는 자기상관함수(autocorrelation function)를 바탕으로 시계열 자료를 모형화하는 방법이다. 지수평활 모형이 비정상성 모형인 반면에, ARIMA 모형은 정상성을 기반으로 한다. 따라서, 정상화를 위한 차분이 중요한 역할을 한다. ARIMA 모형은 주로 승법형태로 사용되며, 다음과 같은 모형식을 가지고 있다.

$$\phi_p(B)\Phi_P(B^m)(1-B)^d(1-B^m)^D y_t = c + \theta_q(B)\Theta_Q(B^m)e_t$$

여기서, $B^k y_t = y_{t-k}$로서 시차 연산자(lag operator)이고 d는 비계절차분의 차수, D는 계절차분의 차수이며, 시차 다항식(lag polynomial)은 다음과 같이 정의된다.

$$
\begin{aligned}
\phi_p(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p, \quad \Phi_P(B^m) = 1 - \Phi_1 B^m - \cdots - \Phi_P B^{Pm},\\
\theta_q(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q, \quad \Theta_Q(B^m) = 1 - \Theta_1 B^m - \cdots - \Theta_Q B^{Qm}.
\end{aligned}
$$

시차 다항식 $\phi_p(B)$ 및 $\Phi_P(B^m)$은 비계절 및 계절 자기회귀항 과거 시차들의 선형결합을 나타내며, $\theta_q(B)$ 및 $\Theta_Q(B^m)$은 비계절 및 계절 이동평균항 과거 시차들의 선형결합을 나타낸다; 단, p와 P는 각각 비계절 및 계절 자기회귀항의 차수를, q와 Q는 각각 비계절 및 계절 이동평균항의 차수를 나타낸다. ARIMA 모형은 다른 어떤 모형보다도 정교한 이론적 체계를 가지고 있으며, Box와 Jenkins (1970) 이후 이분산성(heteroscedasticity) 모형 및 다변량 모형으로의 체계적 발전도 거듭해 왔다. 금융 시장에서의 가격 모형 및 거시 경제학적 시각에서의 균형 이론 등에 지대한 공헌을 해 온 것이다.

2.4. DOTM 모형

DOTM 모형은 Assimakopoulos와 Nikolopoulos (2000)가 제안한 Theta 예측 방법의 확장으로서 M3-competition (Makridakis와 Hibon, 2000)에서 가장 우수한 예측력을 보인 모형이다. 기본적으로 Theta 방법은 승법 계절 분해법으로 계절조정 이후 모형화가 이루어진다. 우선, 시계열 자료는 계절 주기에 대응하는 시차에서 자기상관함수로 계절성 유무를 검정한다. 계절성이 있을 경우 승법 분해법을 이용하여 계절 지수가 계산되며

원자료를 계절 지수로 나누어 계절조정 시계열을 만든다. 계절조정된 시계열을 y_t라고 할 때 모형식은 다음과 같다.

$$
\begin{aligned}
y_t &= \mu_t + e_t,\\
\mu_t &= \ell_{t-1} + \left(1 - \frac{1}{\theta}\right)\left[(1-\alpha)^{t-1}A_{t-1} + \left(\frac{1-(1-\alpha)^t}{\alpha}\right)B_{t-1}\right],\\
\ell_t &= \alpha y_t + (1-\alpha)\ell_{t-1},\\
A_t &= \bar{y}_t - \frac{t+1}{2}B_t,\ B_t = \frac{1}{t+1}\left[(t-2)B_{t-1} + \frac{6}{t}(y_t - \bar{y}_{t-1})\right], \quad \bar{y}_t = \frac{1}{t}[(t-1)\bar{y}_{t-1} + y_t].
\end{aligned}
$$

위의 모형식은 상태공간 모형으로 표현할 수 있으며 각 모수들은 $\ell_0 \in \mathbb{R}, \alpha \in (0,1), \theta \in [1,\infty)$의 범위를 가지고, 각 상태들의 초기값은 $A_0 = B_0 = B_1 = \bar{y}_0 = 0$과 같이 가정된다. y_t의 조건부 평균을 나타내는 μ_t는 단기적 추세를 나타내는 ℓ_{t-1}와 장기적 추세를 나타내는 나머지 항으로 구성된다. 단기적 추세는 SES(simple exponential smoothing) 모형에서와 동일한 형태를 가지고 있으며, 장기적 추세를 구성하는 A_t 및 B_t는 각각 시간 추세$(1,2,\ldots,t)$에 대한 y_t의 회귀 절편 및 기울기를 나타낸다. 특히, $\theta = 1$인 경우 DOTM 모형은 SES 모형으로 축소되며, $\theta > 1$일 경우 장기적 추세를 설명하는 성분이 추가된 확장된 SES 모형으로 볼 수 있다. $\theta = 2$인 경우를 보통 표준적 형태로 간주하여 DSTM(dynamic standard Theta model) 모형으로 부른다.

2.5. TBATS 모형

TBATS 모형은 ETS 모형을 확장한 것으로서 이분산성을 안정화하기 위하여 박스-콕스 변환을 도입하고, 단기 움직임을 잘 설명하기 위하여 오차항에 ARMA 모형을 도입하였으며 계절성은 푸리에(Fourier) 타입의 삼각함수$(s_{j,t}, s^*_{j,t})$를 이용한다. 모형식은 다음과 같이 표현된다.

$$
y_t^{(\omega)} = \begin{cases} \dfrac{(y_t^{\omega} - 1)}{\omega}, & \text{if } \omega \neq 0,\\ \log(y_t), & \text{if } \omega = 0, \end{cases} \qquad y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + d_t,
$$

$$
\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t, \quad b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t, \quad s_t = \sum_{j=1}^{k} s_{j,t},
$$

$$
\begin{aligned}
d_t &= \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + e_t,\\
s_{j,t} &= s_{j,t-1}\cos\lambda_j + s^*_{j,t-1}\sin\lambda_j + \gamma_1 d_t,\\
s^*_{j,t} &= s_{j,t-1}\sin\lambda_j + s^*_{j,t-1}\cos\lambda_j + \gamma_2 d_t.
\end{aligned}
$$

단, k는 계절을 나타내기 위한 푸리에 계절항 $s_{j,t}$의 개수를 나타내며 $\lambda_j = 2\pi jt/m$이다. ω는 박스-콕스 변환의 차수를 나타낸다. 최근 시계열 자료들은 더 복잡해진 형태를 가지고 있으며 고주파수화 됨에 따라서 2개 이상의 계절성을 가지는 경우가 많다. TBATS 모형은 이러한 경우의 모형화 및 예측에 강점을 가지며 다중 계절성(multiple seasonality) 및 정수값이 아닌 계절 주기(non-interger periods)까지 모형화할 수 있다는 장점이 있다.

2.6. 신경망 모형

시계열 예측을 위해서 자주 사용되는 신경망(neural network) 모형은, 1개의 은닉층(a single hidden layer)을 가지는 순전파(feed-forward) 모형이 대표적이다 (Hyndman과 Athanasopoulos, 2018). 직관적으로 이해하기

쉽고 ARIMA 모형과 비교 및 대응이 용이하기 때문이다. 이 모형은 보통 NNAR(·)로 표시하며, 입력 변수는 AR 모형과 유사하게 시계열 자료의 과거값들이 사용된다. 보통 활성화 함수(activation function)로서 다음과 같은 시그모이드(sigmoid) 함수를 이용한다.

$$s(z) = \frac{1}{1 + e^{-z}}$$

따라서, ARIMA 모형과 달리 과거값들의 비선형적 관계를 이용하고 있으며 정상성에 대한 제약이 없다는 특징이 있다. 비계절형 모형 NNAR(p, k)의 경우, 1-시차부터 p-시차까지의 과거 시계열 자료가 입력되며 k개 노드가 은닉층에 사용된다. 계절형 모형 NNAR$(p, P, k)_m$의 경우, 입력 변수는 ARIMA$(p, 0, 0)(P, 0, 0)_m$과 동일하게 다음과 같은 값들이 입력값으로 사용된다.

$$(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, y_{t-m}, y_{t-2m}, \ldots, y_{t-Pm})$$

은닉층 구성을 위해서 사용되는 초기 가중치는 랜덤하게 결정되기 때문에 신경망 모형에 의한 예측 결과는 시행 때마다 달라진다는 단점이 존재한다. 따라서, 적당한 반복을 통하여 그 결과를 평균화하는 것이 합리적이다.

2.7. 랜덤워크 모형

랜덤워크 모형은 효율적 시장 가설(efficient market hypothesis)을 토대로 가장 최근의 시계열 값을 이용하여 미래를 예측하는 모형이다. 비계절형일 경우는 가장 최근값이, 계절형일 경우는 동일한 계절에서 가장 최근값이 이용된다. 비계절형 및 계절형 랜덤워크(nonseasonal and seasonal random walk) 모형은 각각 다음과 같은 모형식으로 표현된다.

$$y_t = y_{t-1} + e_t, \qquad y_t = y_{t-m} + e_t.$$

이 두가지 모형은 예측 시계열 모형에 대한 벤치마크 모형으로 자주 사용된다. 비록 랜덤워크 모형이 단순하게 보이지만, 이보다 좋은 예측력을 항상 유지하기는 쉽지 않다. 효율적 시장 가설에 대한 논란이 금융 시장에서 여전히 존재하는 이유이다. 이후 비계절형 랜덤워크 모형은 Naive, 계절형은 Snaive로 표시하기로 한다.

3. 결합 모형의 성능 평가

본 절에서는 2절에서 설명된 모형들을 결합해서 예측에 사용하는 방법을 설명하고 예측력의 성능을 M3 자료를 이용해서 평가한다.

3.1. 결합 모형

본 연구에서 고려하는 결합 모형은 단순하지만 중앙값 또는 평균이다. 본 연구의 초점은 쉽게 계산할 수 있는 조합 방법이기 때문이다. Lichtendahl 등 (2013)도 사분위수 및 평균과 같은 단순한 방법이 결합 모형으로서 좋은 예측력을 보인다고 주장하였다. 예측 성능 평가를 위해 사용하는 시계열 자료는 M3 자료로서 음이 아닌 값들로 구성되어 있기 때문에, 결합을 위한 평균 또는 중앙값의 결과값이 음수인 경우 0으로 대체한다. 이와 같이 조합된 방법을 표시하기 위하여, 중앙값 결합은 SCUM: median으로, 평균 결합은 SCUM: mean으로 표시하기로 한다; 단, SCUM은 simple combination of univariate models의 줄임말이다.

Table 1: Forecasting models and the corresponding R functions

Model	R package	Function
ETS	*forecast*	ets()
CES	*smooth*	auto.ces()
ARIMA	*forecast*	auto.arima()
DOTM	*forecTheta*	dotm()
TBATS	*forecast*	tbats()
NNAR	*forecast*	nnetar()
Naive	*forecast*	naive()
Snaive	*forecast*	snaive()

Table 2: M3-Competition dataset

Frequency	Forecasting horizon(h)	Frequency	Number of time series
Yearly	6	1	645
Quarterly	8	4	756
Monthly	18	12	1428
Other	8	1	174
Total	-	-	3003

3.2. 예측 성능 평가

2절에서 설명된 모형들에 의한 예측치를 R에서 실제로 계산하기 위하여 Table 1과 같이 R 패키지인 forecast, smooth 및 forecTheta 내의 해당 함수들 및 forecast 패키지의 forecast() 함수를 사용하였다. M3 시계열 자료는 R 패키지 Mcomp를 통해서 사용하였으며 Table 2와 같이 비계절형(Yearly, Other)과 계절형(Quarterly, Monthly) 시계열이 혼합되어 있다. 또한, 검증 시계열의 길이 또는 예측 기간(forecast horizon)도 해당 범주의 시계열마다 다르다.

예측 성능 평가를 위하여 다음과 같은 2가지 측도를 사용하였다. 먼저, 시계열 자료 y_t의 예측 원점 T와 예측 시점 $1, 2, \ldots, h$에 대하여 다음과 같은 sMAPE(symmetric mean absolute percentage error)를 사용한다.

$$\text{sMAPE} = \frac{2}{h} \sum_{j=1}^{h} \frac{\left|y_{T+j} - \hat{y}_{T+j}\right|}{|y_{T+j}| + |\hat{y}_{T+j}|} \tag{3.1}$$

단, $j = 1, \ldots, h$에 대하여 $\hat{y}_{T+j}$은 검증 시계열 y_{T+j}에 대한 예측값을 나타낸다. sMAPE는 MAPE가 실제값과 예측값과의 대소 관계에 따라 과대 추정하는 예측값에 패널티를 더 많이 부여하는 문제를 해결한다(Makridakis, 1993). 다만, 분모에 예측값이 포함되기 때문에 이 측도가 예측값에 의존하는 단점이 있다.

두 번째로 다음과 같은 MASE(mean absolute scaled error) (Hyndman과 Koehler, 2006)를 사용한다.

$$\text{MASE} = \frac{1}{h} \sum_{j=1}^{h} \frac{|y_{T+j} - \hat{y}_{T+j}|}{\frac{1}{T-m} \sum_{j=m+1}^{T} |y_j - y_{j-m}|} \tag{3.2}$$

단, m은 계절 주기를 나타내며 비계절형일 경우 $m = 1$로 고려한다. MASE는 분자인 예측 오차들의 절대값 평균을 분모인 훈련 샘플의 m차 차분에 대한 절대값들의 평균으로 나눈 것이다. 즉, MASE는 랜덤워크 모형에 의한 예측에 대한 상대적 예측 오차 비율로 해석할 수 있다.

Table 3과 4는 각각 sMAPE와 MASE를 이용하여 고려된 8개 단변량 모형 및 결합 모형의 예측력을 계산 및 비교한 결과를 나타낸다. 표에서 가장 우수한 예측력을 보인 경우는 진하게, 두 번째로 우수한 경우는

Table 3: Results for all methods using the sMAPE

Methods	Yearly	Quarterly	Monthly	Other	All
ETS	17.00	9.68	14.14	4.37	13.07
CES	17.28	9.47	15.09	4.70	13.54
ARIMA	17.10	10.01	14.96	4.51	13.57
DOTM	15.94	9.28	13.74	4.58	12.56
TBATS	17.37	10.22	13.84	**4.35**	13.14
NNAR	20.63	11.81	16.36	5.68	15.51
Naive	17.88	11.32	18.18	6.30	15.70
Snaive	na	11.07	17.23	na	na
SCUM: median	**15.73**	**8.98**	**13.37**	**4.35**	**12.25**
SCUM: mean	15.92	9.10	13.53	4.37	12.40

Table 4: Results for all methods using the MASE

Methods	Yearly	Quarterly	Monthly	Other	All
ETS	2.86	1.17	0.86	**1.81**	1.43
CES	3.14	1.12	0.86	2.18	1.49
ARIMA	2.96	1.19	0.87	1.84	1.45
DOTM	**2.59**	1.12	0.85	1.94	**1.36**
TBATS	3.13	1.26	0.86	1.85	1.50
NNAR	3.75	1.48	1.08	2.72	1.85
Naive	3.17	1.46	1.17	3.09	1.79
Snaive	na	1.43	1.15	na	na
SCUM: median	2.73	**1.06**	**0.83**	1.86	**1.36**
SCUM: mean	2.74	1.08	0.85	1.93	1.38

밑줄로 표시하였다.

8개 단변량 모형들을 중앙값으로 결합한 SCUM: median이 거의 모든 경우에 가장 우수한 예측력을 보였다. 특히, sMAPE로 측정한 경우는 항상 가장 우수한 예측력을 유지하고 있다. 평균으로 결합 경우인 SCUM: mean은 중앙값 결합보다 예측력이 떨어지지만, 시계열 자료에 사분기 및 월과 같이 계절성이 뚜렷한 경우에는 우수한 예측력을 나타내었다. DOTM은 단일 방법으로도 우수한 예측력을 보였다. 거의 모든 경우에 상위 3위 내에 포함되고 있다. 전형적인 계절성 형태인 사분기 및 월별 시계열 자료에서 중앙값 또는 평균 결합 모형의 우수성은 더욱 뚜렷하게 나타났다. 반면, 계절성이 없는 Other 영역에서는 결합 모형의 우수성이 약해지는 경향이 나타났다.

4. 결론

최근 다양한 분야 및 상황에서 신뢰성있고 로버스트한 예측 방법의 필요성이 증가하고 있다. 본 연구에서는 입수하기가 쉽고 비교적 널리 알려진 단변량 시계열 모형들만을 이용하여 예측력을 증가시킬 수 있는 단순 조합 방법에 대해서 연구하였다. 예측 정확성 측면 뿐만 아니라 계산의 복잡도 및 계산 시간도 짧아서 그 방법의 우수성을 높이 평가할 수 있다. 향후, 시계열 자료의 계절성 유무 및 다중 계절성과 같은 시계열 특성과 시계열 자료의 길이와 예측 기간 등과 같은 예측 상황을 적절히 고려할 수 있는 예측 조합으로 발전시킬 수 있겠다.

References

Assimakopoulos V and Nikolopoulos K (2000). The Theta model: a decomposition approach to forecasting, *International Journal of Forecasting*, **16**(4), 521–530.

Box GEP and Jenkins GM (1970). *Time series analysis: forecasting and control*, Holden-Day, San Francisco.

Brown RG (1959). *Statistical forecasting for inventory control*, McGraw/Hill.

Commandeur JJF and Koopman SJ (2007). *Introduction to State Space Time Series Analysis*, Oxford University Press.

Fiorucci JA, Pellegrini TR, Louzada F, Petropoulos F, and Koehler AB (2016). Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting*, **32**(4), 1151–1161.

Holt CE (1957). *Forecasting seasonals and trends by exponentially weighted averages*, O.N.R. Memorandum No. 52. Carnegie Institute of Technology, Pittsburgh USA.

Hyndman RJ and Athanasopoulos G (2018). *Forecasting: principles and practice*, 2nd Ed., OTexts.

Hyndman RJ and Koehler AB (2006). Another look at measures of forecast accuracy, *International Journal of Forecasting*, **22**(4), 679–688.

Kolassa S (2011). Combining exponential smoothing forecasts using Akaike weights, *International Journal of Forecasting*, **27**, 238–251.

Lichtendahl KC, Grushka-Cockayne Y, and Winkler RL (2013). Is it better to average probabilities or quantiles? *Management Science*, **59**(7), 1594–1611.

Livera AMD, Hyndman RJ, and Snyder RD (2011). Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association*, **106**(496), 1513–1527.

Makridakis S (1993). Accuracy measures: theoretical and practical concerns, *International Journal of Forecasting*, **9**(4), 527–529.

Makridakis S and Hibon M (2000). The M3-Competition: results, conclusions and implications, *International Journal of Forecasting*, **16**(4), 451–476.

Petropoulos F and Svetunkov I (2020). A simple combination of univariate models, *International Journal of Forecasting*, **36**(1), 110–115.

Svetunkov I and Kourentzes N (2018). Complex exponential smoothing for seasonal time series, *Working Paper of Department of Management Science*, Lancaster University, 2018:1, 1–20.

Received February 22, 2022; Revised February 24, 2022; Accepted February 25, 2022

단변량 시계열 모형들의 단순 결합의 예측 성능

이선홍[a], 성병찬[1,a]

[a]중앙대학교 응용통계학과

요 약

본 논문에서는 시계열 예측 분야에서 잘 알려져 있는 단변량 시계열 모형들을 이용하여, 그들의 단순 조합이 어떤 예측력을 보여주는지 연구한다. 고려된 단변량 시계열 모형으로는, 지수평활 및 ARIMA(autoregressive integrated moving average) 모형들과 그들의 확장된 형태인 모형들 그리고 예측의 벤치마크 모형으로 자주 사용되는 비계절 및 계절 랜덤워크 모형이다. 단순 조합의 방법은 중앙값과 평균을 이용하였으며, 검증을 위하여 사용된 데이터셋은 3,003개의 시계열 자료로 구성된 M3-competition 자료이다. 예측 성능을 sMAPE(symmetric mean absolute percentage error)와 MASE(mean absolute scaled error)로 평가한 결과, 단변량 시계열 모형들의 단순 조합이 아주 우수한 예측력을 가지고 있음을 확인하였다.

주요용어: 단변량 예측 모형, 지수평활법, ARIMA, M3-competition

이 논문은 2020년도 중앙대학교 연구장학기금 지원에 의한 것임.

[1]교신저자: (06974) 서울시 동작구 흑석로 84, 중앙대학교 경영경제대학 응용통계학과. E-mail: bcseong@cau.ac.kr