

구조변화 통계량을 이용한 적응적 지수평활법 Adaptive Exponential Smoothing Method Based on Structural Change Statistics

김정일, 박대근, 전덕빈, 차경천
한국과학기술원 테크노경영대학원, jungil@kgsms.kaist.ac.kr

Abstract

Exponential smoothing methods do not adapt well to unexpected changes in underlying process. Over the past few decades a number of adaptive smoothing models have been proposed which allow for the continuous adjustment of the smoothing constant value in order to provide a much earlier detection of unexpected changes. However, most of previous studies presented ad hoc procedure of adaptive forecasting without any theoretical background. In this paper, we propose a detection-adaptation procedure applied to simple and Holt's linear method. We derive level and slope change detection statistics based on Bayesian statistical theory and present distribution of the statistics by simulation method. The proposed procedure is compared with previous adaptive forecasting models using simulated data and economic time series data.

1. 서론

지수평활법은 복잡한 예측 모형에 비해 예측 능력이 떨어지지 않으면서 계산량이 적고 수식이 단순하기 때문에 오늘날 산업 현장에서 가장 많이 이용되는 예측 모형 중 하나이다. 그러나 이 방법은 시계열의 내재과정(underlying process)에 급격한 수준의 변화(level change) 또는 기울기(slope)가 발생할 때, 그 변화에 신속히 적응하지 못하는 단점을 가지고 있다. 따라서 지수평활법의 장점인 적용 간편성을 가지면서도 환경 변화에 빠르게 적응할 수 있는 예측 방법의 개발이 요구되었다. 이러한 적응적 지수평활법(adaptive exponential smoothing)은 Trigg and Leach(1967)가 제안한 이후 많은 연구들이 진행되어왔다. Trigg and Leach(1967)는 지수평활모형에서 예측 오차에 따라 평활 계수가 연속적으로 변하는 적응적 예측 모형을 제시하였으며 이후 Whybark(1973)은 예측 오차의 지속적인 모니터링을 통해 기준 수준을 넘을 때만 평활 계수를 변화시키도록 하였다. 단순 지수평활법에 대한 적응적 예측 방법과 더불어 기울기 있는 시계열에 대한 Holt의 선형 방법에 대한 적응적 예측 방법에도 활발한 연구가 진행되어왔다(Gilchrist 1976, Roberts and Reed 1969). 그러나 이들 연구는 구조변화에 대한 이론적 고민 없이 ad hoc한 예측 방법을 제안한 한계를 지니고 있었다.

지수평활법의 개념과는 별도로 전통 통계학적 접근으로 구조변화를 연구한 논문으로 Chang, Chen and Tiao(1988), Tsay(1986) 등이 구조변화시점과 변화유형을 인식한 후 간섭모형을 분석하였으며, Chen and Liu(1993)은 구조변화인식과 모수 추정을

동시에 할 수 있는 방법을 제시하였다.

베이저안 관점에서는 Chernoff and Zacks(1964), Gardner(1969) 등이 constant mean model에서 수준변화를 인식하고자 하였고, Ferreira(1975), Holbert and Broemeling(1977) 등은 베이저안을 이용하여 회귀모형의 변환을 분석하였다.

Jun(1989), Jun과 Oliver(1985)는 Harrison과 Stevens(1976)의 동적 선형 모형(dynamic linear model)에서의 수준 또는 기울기의 급격한 변화 시 그 변화시기 및 변화량의 추정법을 제시하면서 지수이동가중평균법 적용 시 수준변화를 가설 검정하는 통계량을 유도하였다. 이 연구는 베이저안 관점에서 유도된 수준변화 검정통계량이 지수평활법의 평활계수의 함수로 표현될 수 있음을 보임으로써 구조변화 인식 문제를 적응적 지수평활분야에 도입한 최초의 연구가 되었다. 본 논문에서는 Jun(1989)에서 제시한 통계량을 기반으로 수준변화와 기울기 변화에 대한 새로운 적응적 예측 방법을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 Jun(1989)이 제시한 단순 지수평활법에서의 수준변화 감지 통계량을 소개하고 이를 확장하여 Holt의 선형방법의 기울기 변화 감지 통계량을 유도한다. 이들 통계량을 이용한 적응적 예측 절차를 소개한 후 3장에서 시뮬레이션 분석을 통해 논문이 제시하는 예측방법의 우수성을 살펴본다. 4장의 실제 데이터를 통해 모형의 현실 적용성을 살펴본 후 5장에서 결론을 맺는다.

2. 적응적 지수평활법을 위한 구조변화 통계량

2.1. 단순 지수평활법

Harrison and Stevens(1976)의 동적 선형 모형(Dynamic Linear Model)에서 미지의 시점 M에 수준변화를 나타내는 가변수(dummy variable) Δ 를 추가함으로써 수준변화의 효과를 분석할 수 있다.

$$\begin{aligned} Z_{t+1} &= L_{t+1} + a_{t+1} & t = 0, 1, \dots, n-1 \\ L_{t+1} &= L_t + b_{t+1} & t \neq M \\ &= L_t + \Delta + b_{t+1} & t = M \end{aligned} \quad (1)$$

a_t 과 b_t 은 각각 기대값이 0이고 분산이 각각 σ_a^2 , σ_b^2 이며 서로 무상관인 정규분포를 따른다. L_0, Δ, a_t, b_t 와 구조변화 시점에 대한 확률변수 M은 상호독립으로 가정한다. Harrison(1967)은 단순 지수평활법의 예측식이 수준변화가 없는 $\Delta=0$ 일 때 식 (1)의 최적의 예측식이라는 것을 밝혔다.

Jun(1989)은 미지의 시점 M 에 발생한 수준변화 Δ 를 감지하는 검정 통계량을 유도하였다. Δ 와 M 에 대한 비 정보 사전분포(non-informative prior)하에서 지수평활법의 수준변화를 감지하는 검정통계량은 다음과 같다.

$$S_t^L = e_{t-1}^2 + \frac{(e_{t-2} + \alpha e_{t-1})^2}{1 + \alpha^2} + \dots + \frac{(e_1 + \alpha e_2 + \dots + \alpha^{t-2} e_{t-1})^2}{1 + \alpha^2 + \dots + \alpha^{2t-4}} \quad (2)$$

여기서 e_t 는 지수평활법의 일단계 예측 오차(one-step-ahead forecast error)이고 $1 - \alpha$ 는 단순 지수평활법의 평활계수이다.

2.2 Holt의 선형방법

기울기가 있는 동적 선형모형에서 M 시점에 기울기 변화가 발생하는 상황은 다음 식으로 표현할 수 있다.

$$\begin{aligned} Z_{t+1} &= L_{t+1} + a_{t+1} \\ L_{t+1} &= L_t + S_t + b_{t+1} \\ S_{t+1} &= \begin{cases} S_t + c_{t+1} & t \neq M-1 \\ S_t + \Delta_t + c_{t+1} & t = M-1 \end{cases} \end{aligned} \quad (3)$$

여기서 a_t, b_t, c_t 는 평균이 0이고 분산이 각각 $\sigma_a^2, \sigma_b^2, \sigma_c^2$ 인 serially uncorrelated Gaussian white-noise processes 이다. 그리고 $(L_0, T_0)', \Delta, a_{t+1}, b_{t+1}$ 와 c_{t+1} 는 상호 무상관이다. 위의 동적 선형모형에서 기울기 변화에 대한 가설 검정 통계량을 Jun(1989)의 절차에 따라 유도하면 다음과 같다.

구조변화 없는 경우

Δ_t 가 0인 경우 식 (3)의 내부 수준과 기울기의 사후 확률 분포는 다음과 같이 유도된다.

$$\begin{pmatrix} L_t \\ S_t \end{pmatrix} / z_1, \dots, z_t \sim BVN \left(\begin{pmatrix} \hat{\mu}_t \\ \hat{\beta}_t \end{pmatrix}, \begin{pmatrix} \hat{p}_t & \hat{\omega}_t \\ \hat{\omega}_t & \hat{\eta}_t \end{pmatrix} \right) \quad (4a)$$

$$\begin{aligned} \hat{\mu}_t &= \hat{\mu}_{t-1} + \hat{\beta}_{t-1} + k_{t-1}^L (Z_t - \hat{\mu}_{t-1} - \hat{\beta}_{t-1}) \\ \hat{\beta}_t &= \hat{\beta}_{t-1} + k_{t-1}^S (Z_t - \hat{\mu}_{t-1} - \hat{\beta}_{t-1}) \\ \hat{p}_t &= \sigma_a^2 k_{t-1}^L \\ \hat{\omega}_t &= \sigma_a^2 k_{t-1}^S \\ \hat{\eta}_t &= \hat{\eta}_{t-1} + \sigma_c^2 - (\hat{\omega}_{t-1} + \hat{\eta}_{t-1}) k_{t-1}^S \end{aligned} \quad (4b)$$

$$\begin{aligned} k_t^L &= \frac{\hat{p}_t + 2\hat{\omega}_t + \hat{\eta}_t + \sigma_b^2}{\hat{p}_t + 2\hat{\omega}_t + \hat{\eta}_t + \sigma_a^2 + \sigma_b^2} \\ k_t^S &= \frac{\hat{\omega}_t + \hat{\eta}_t}{\hat{p}_t + 2\hat{\omega}_t + \hat{\eta}_t + \sigma_a^2 + \sigma_b^2} \end{aligned} \quad (4c)$$

여기서 k_t^L 와 k_t^S 는 각각 수준과 기울기의 Kalman

gain 이다. 또한 최소자승 예측치 f_t 와 v_t 의 분산 v_t 는 다음과 같이 주어진다.

$$\begin{aligned} f_t &= \hat{\mu}_t + \hat{\beta}_t & t = 0, 1, \dots, n \\ v_t &= \hat{p}_t + 2\hat{\omega}_t + \hat{\eta}_t + \sigma_a^2 + \sigma_b^2 & t = 0, 1, \dots, n \end{aligned} \quad (5)$$

평형상태에서의 Kalman gain, $k^L = \lim_{i \rightarrow \infty} k_i^L$ 와 $k^S = \lim_{i \rightarrow \infty} k_i^S$ 는 각각 식 Holt의 선형방법의 모수 α 와 $\alpha\beta$ 와 같다.

구조변화 시점을 알 때

구조변화가 시점 m 에 발생한 것을 알 때 $m+1$ 시점은 구조변화 효과가 관측되는 첫번째 시점이 된다. 따라서 m 시점 이전의 Z_{t+1} 에 대한 조건부 분포는 $\Delta = 0$ 인 경우와 동일하다.

$$Z_{t+1} / z_1, \dots, z_t \sim N(f_t, v_t) \quad t = 1, \dots, m-1 \quad (6)$$

구조변화 시점과 크기를 알 때, 시점 이후의 조건부 분포는 다음과 같다.

$$Z_{t+1} / z_1, \dots, z_t; m; \Delta \sim N(f_t + \gamma_t \Delta, v_t) \quad t = m, \dots, n \quad (7a)$$

$$\gamma_i = \begin{cases} 0 & i \leq m-1 \\ 1 & i = m \\ 2 - k_i^L - k_i^S & i = m+1 \\ (2 - k_i^L - k_i^S) \gamma_{i-1} - (1 - k_i^L) \gamma_{i-2} & i > m+1 \end{cases} \quad (7b)$$

여기서 $L_0, S_0, \{a_t\}, \{b_t\}$ 와 $\{c_t\}$ 가 Gaussian 분포를 따르므로 z_1, \dots, z_t 조건하에서 Δ 의 분포 역시 Gaussian 분포를 따른다. multivariate normal analysis를 통해 Δ 의 사후 확률 분포를 구하면 식 (15)와 같다.

$$\begin{aligned} \Delta | z_1, \dots, z_t; m &\sim N(\lambda_t, q_t) & t = 1, \dots, n \\ q_t &= q_0 & t = 1, \dots, m \\ &= \frac{1}{(1/q_0) + \sum_{i=m}^{t-1} (\gamma_i^2 / v_i)} & t = m+1, \dots, n \\ \lambda_t &= \lambda_0 & t = 1, \dots, m \\ &= q_t \left\{ \frac{\lambda_0}{q_0} + \sum_{i=m}^{t-1} \frac{\gamma_i}{v_i} e_i \right\} & t = m+1, \dots, n \end{aligned} \quad (8)$$

구조변화 시점을 모를 때

미지의 구조변화 시점 M 에 대한 사전 확률 분포가 $p(M)$ 으로 주어질 때 M 과 Δ 의 사후확률 분포는 다음과 같이 유도된다.

$$p(M/z_n, \dots, z_1) \propto p(M) \sqrt{q_n} \exp\left(-\frac{\lambda_n^2}{2q_n}\right) \quad M=1, \dots, n-1$$

$$f(\Delta/z_n, \dots, z_1) \propto \sum_{M=1}^{n-1} p(M/z_n, \dots, z_1) \Phi(\Delta/z_n, \dots, z_1; M) \quad (9)$$

Jun(1989)에 따라 각 시점에 대해 $\Delta=0$ 여부를 가설 검정하는 통계량은 다음과 같고

$$\frac{1}{2} \sum_{M=1}^{n-1} p(M) \frac{\lambda_n^2}{q_n} \quad (10)$$

여기에 구조변화 시점과 크기에 대한 diffuse prior를 가정하면 Holt의 선형 방법에서 기울기 변화를 감지할 수 있는 통계량을 다음과 같이 유도할 수 있다.

$$S_t^S = \gamma_1 e_{t-1}^2 + \frac{(\gamma_1 e_{t-2} + \gamma_2 e_{t-1})^2}{1 + \gamma_2^2} + \dots + \frac{(\sum_{i=1}^{t-1} \gamma_i e_i)^2}{1 + \gamma_2^2 + \dots + \gamma_{t-1}^2} \quad (11a)$$

여기서

$$\gamma_i = \begin{cases} 1 & i=1 \\ 2 - \alpha - \alpha\beta & i=2 \\ (2 - \alpha - \alpha\beta)\gamma_{i-1} - (1 - \alpha)\gamma_{i-2} & i > 2 \end{cases} \quad (11b)$$

식 (11)에서 제시된 통계량은 Holt의 선형방법에서 계산되는 오차항과 평활 계수만으로 이루어진 함수이다. 통계량의 성질을 살펴보면, 구조변화가 발생하지 않을 경우 0으로 수렴하고 구조변화가 지속적으로 발생할 경우 큰 값을 가지게 된다.

2.3 통계량의 분포 및 예측 절차

식(2)와 (11)에서 유도된 통계량의 귀무 분포(null distribution)는 간단한 형태가 아니다. Jun(1989)은 이 식(2)의 극한 분포(limiting distribution)가 정규분포를 따름을 밝혔지만 가설 검정에 필요한 임계치(critical values)의 계산은 수치적분(numerical integration)이 필요한 작업으로 실제적인 적용에 어려움이 따른다. 따라서 본 논문에서는 시뮬레이션 분석을 통해 각 통계량의 경험적(empirical) 분포를 유도하고 이를 근거로 가설 검정하는 절차를 제안한다. 먼저 S_t^L 의 분포 유도를 위해 ARIMA(0,1,1)을 따르는 길이가 10인 시계열 10000개를 생성하여 각 시계열의 수준변화 통계량을 계산한 후 계산된 통계량의 분포를 계산한다. 통계량의 분포는 구조변화의 발생여부에 관심이 있는 모니터링 구간의 길이에 따라 다르므로 가설 검정을 위한 임계치도 달라진다. MA계수가 0.6인 경우 모니터링 구간을 2부터 10까지 변화시키면서 통계량의 분포를 그려보면 다음과 같다.

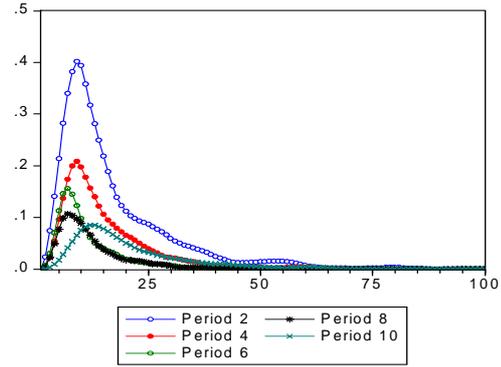


그림 1. 구조변화 통계량의 시뮬레이션 분포

유도된 통계량을 이용한 적응적 예측 절차는 다음과 같다.

- 1단계: 과거 관측치를 통해 평활계수를 추정한다.
- 2단계: 평활 계수를 이용하여 1단계 예측치를 계산하고 모니터링 구간에 대한 예측 오차를 계산한다. 모니터링 구간은 마지막으로 감지된 구조변화 시점으로부터 현재 시점까지이고, 구조변화가 발생한 적 없는 경우에는 임의의 큰 구간으로 정한다. 본 논문에서는 5 시점으로 정한다.
- 3단계: 매 시점마다 식 (11)를 이용하여 통계량을 계산하고 모니터링 구간에 대한 분포와 임계치를 계산한 후 통계량과 비교한다. 통계량이 임계치를 넘을 경우 구조변화가 발생하였다고 결론 내리고 평활 계수를 1로 변경하여 구조변화에 신속히 적응하도록 한다.

3. 시뮬레이션 분석

제안모형의 예측 능력을 평가하기 위해 시뮬레이션 분석을 수행하였다. 구조변화 통계량을 이용한 적응적 단순 지수평활법을 평가하기 위해 먼저 ARIMA(0,1,1), $(1-B)Z_t = (1-\theta B)e_t$ 을 따르는 길이가 100인 시계열을 생성하고, 70번째 시점에 크기가 3인 수준변화를 발생시킨 후 41번째에서 100번째 시점의 1단계 예측오차를 구하였다. 다양한 평활 계수에 대한 예측결과를 살펴보기 위해 θ 를 0.1부터 0.9로 0.2씩 증가시키면서 각 θ 당 300개씩의 시계열을 생성하고 예측모형별로 가장 적은 MSE를 산출한 회수를 기록하였다. θ 가 0.9인 경우를 제외하고 모든 범위에서 제안방법의 예측 성능이 우수했다.

θ	SES	Trigg & Leach	Whybark	Proposed
0.9		146	35	119
0.7	67	72	54	111
0.5	131	20	30	151
0.3	163	3	13	184
0.1	173	2	3	197

표1. 수준변화 시계열에 대한 시뮬레이션 결과

구조변화 통계량을 이용한 적응적 Holt의 선형방법의 예측 능력을 비교하기 위해서 ARIMA(0,2,2), $(1-B)^2 Z_t = (1-\theta_1 B)(1-\theta_2 B)e_t$ 를 따르는 시계열을 생성하였고 역시 70번째 시점에

크기가 3인 기울기 변화를 주었다. 표 2는 각 θ_1 값에 대해 θ_2 를 -0.9에서 0.9 사이를 0.2씩 변화시키면서 300개씩의 시계열을 생성하고 1단계 MSE를 구한 후 예측 모형별로 가장 적은 MSE를 산출한 횟수를 기록한 것이다. θ_1 의 모든 구간에서 제안 방법이 우수한 예측 능력을 보였다.

θ_1	Holt's Linear	Robert & Reed	Gilchrist	Proposed
-0.9	200	970	218	1634
-0.7	731	451	113	1582
-0.5	968	265	4	1468
-0.3	978	199	0	1313
-0.1	843	152	0	1172
0.1	631	96	0	1014
0.3	377	104	0	791
0.5	215	69	0	620
0.7	89	35	0	478
0.9	0	0	0	300

표2. 기울기변화 시계열에 대한 시뮬레이션 결과

4. 실증분석

실증분석 자료로는 한국 산업생산지수와 상용근로자수 데이터가 이용되었다. 산업생산지수 데이터는 IMF시기에 수준변화가 발생한 I(1) 시계열이며, 상용근로자수 데이터는 기울기 변화가 발생한 I(2) 시계열이다. 산업생산지수의 경우 1991년 2분기부터 1997년 2분기까지의 자료로 추정된 $\alpha=0.886$ 이었다. 1997년 3분기부터 2005년 4분기까지의 1단계 MSE를 비교한 결과 단순지수평활법이 2,641,044, Trigg & Leach 방법이 2,976,117, Whybark 방법이 3,619,280, 수준변화 통계량을 이용한 적응적 단순지수평활법이 2,587,095로 제안방법이 가장 우수했다. 상용근로자 데이터의 경우 1990년 1월부터 1994년 12월까지의 자료로 추정된 $\alpha=0.736$, $\beta=2.273$ 이었다. 1995년 1월부터 2005년 12월까지의 1단계 MSE를 비교해본 결과 Holt방법이 192.34, SAFT 방법이 247.36, Gilchrist가 670.29, 수준변화 통계량을 이용한 적응적 Holt 선형방법이 179.89 로 역시 제안방법이 가장 우수한 예측결과를 보였다.

5. 결론

지수평활법은 간단하면서도 뛰어난 예측력을 보여 실제 현장에서 많이 이용되는 예측 방법이다. 그러나 시계열에 구조변화 발생 시 적절한 예측값을 제공하지 못하는 단점을 가지고 있기 때문에 불확실성이 많은 데이터의 예측에는 문제가 있다.

본 연구에서는 Jun(1992)이 제안한 구조변화 통계량을 이용한 수준변화에 대한 적응적 지수평활법과 이를 확장한 기울기 변화에 대한 적응적 지수평활법을 제안하였다. 다양한 환경에서 생성된 시뮬레이션 데이터를 통해 제안 방법의 예측력을 테스트하였고 더 나아가 실제 데이터를 이용한 실증분석을 통해 제안 방법이 기존의 적응적

예측 방법들에 비해 우수함을 입증하였다. 이러한 결과들은 구조변화가 빈번한 시계열의 예측에 제안 방법이 유용하게 적용될 수 있음을 시사한다.

참고문헌

- Chang, I., G. C. Tiao and C. Chen, "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193-204, 1988
- Chen, C. and L. M. Liu, "Forecasting Time Series with Outliers," *Journal of Forecasting*, 12, 13-15, 1993.
- Chernoff, H. and S. Zacks, "Estimation the Current Mean of a Normal Distribution Which is Subjected to Changes over Time," *Annals of Mathematical Statistics*, 35, 999-1089, 1964.
- Gilchrist, W. *Statistical Forecasting*, London: John Wiley and Sons, 1976.
- Ferreira, R. E., "A Bayesian Analysis of a Switching Regression Model: Known Number of Regimes," *Journal of the American Statistical Association*, 3, 84-97, 1975.
- Harrison, P. J., "Exponential Smoothing and Short-Term Sales Forecasting," *Management Science(A)*, 13(11), 821-42, 1967.
- Holbert, D. and L. D. Broemeling, "Bayesian Inference Related to Shifting Sequences and Two Phase Regression," *Communications in Statistics*, A6, 265-275, 1997.
- Harrison, P. J. and C. F. Stevens, "Bayesian Forecasting," *Journal of the Royal Statistical Society, Series B*, 38, 205-247, 1976.
- Jun, D. B. and R. M. Oliver, "Bayesian Forecasts Following a Major Level Changes in Exponential Smoothing," *Journal of Forecasting*, 4, 293-302, 1985.
- Jun, D. B., "On Detecting and Estimating a Major Level of Slope Change in General Exponential Smoothing," *Journal of Forecasting*, 8(1), 55-64, 1989.
- Jun, D. B., "A Design of Adaptive Exponential Smoothing Using A Change Detection Statistic," *Journal of the Operations Research Society of Japan*, Vol. 35, No. 2, June 1992.
- Robert, S. D. and Reed, R., "The Development of a Self-Adaptive Forecasting Technique," *AIIE transactions*, 1, 4, 1969.
- Trigg, D. W. and A. G. Leach, "Exponential Smoothing with an Adaptive Response Rate," *Operations Research Quarterly*, 18, pp. 132-141, 1967.
- Tsay, R. S., "Time Series Model Specification in the Presence of Outliers," *Journal of the American Statistical Association*, 81, 132-141, 1986.
- Whybark, D. W., "A Comparison of Adaptive Forecasting Techniques," *Logistics Transp. Rev.* 9, 13-26, 1973.