

관광 수요를 위한 결합 예측 모형에 대한 연구

손흥구¹ · 하명호² · 김삼용³

¹중앙대학교 응용통계학과, ²중앙대학교 응용통계학과, ³중앙대학교 응용통계학과

(2012년 2월 27일 접수, 2012년 3월 14일 수정, 2012년 3월 28일 채택)

요약

본 논문은 일별 관광수요 자료를 분석하기 위하여 시계열의 대표적인 3개 모형인 ARIMA, Holt-Winters, AR-GARCH 모형을 적용하였다. 모형의 성능을 비교하기 위해 Armstrong (2001)이 제안한 방법을 이용하여 서로 다른 방법의 예측값을 단순결합과 MSE, SE를 이용한 결합법을 이용하여 정확도 높일 수 있음을 확인하였다.

주요어: ARIMA 모형, Holt-Winters 모형, AR-GARCH 모형, 결합예측.

1. 서론

통계적 수요예측은 다양한 분야에서 자료를 분석하고 예측하는데 널리 쓰이는 기법이다. 특히 관광분야는 급변하는 경제 환경 속에서 중요한 경제활동이기 때문에 이러한 경제 환경 속에서 사회현상을 정확히 설명하고 변화의 크기와 방향을 정확히 파악하는 것이 중요한 과제이다. 현재 정확한 추정과 예측을 위한 연구가 꾸준히 이루어지고 있다. 일례로 허남균 등 (2009)은 항공 수요 예측을 위하여 일변량 시계열 모형과 다변량 시계열 모형을 이용하였으며 윤지성 등 (2010)은 국제항공여객 및 화물의 수요예측을 위하여 계절형 일변량 시계열 모형과 계절형 다변량 시계열 모형을 이용하였다. 관광 수요 예측에서 계절성은 관광 수요의 가장 중요한 특징 중의 하나이고 관광 산업의 많은 부분에서 중요한 영향을 가진다는 것은 널리 알려져 있다. 계절형 예측 모형은 Shen 등 (2009)에 연구에서 여러 계절형 모형을 비교해 본 바가 있다. 이러한 계절형은 그 주기가 하나로만 이루어지지 않고 연별, 달별, 주일별, 일별 주기가 복합적으로 이루어진다. 이러한 이중 계절형 적용은 다른 분야에서도 이루어지고 있는데 Taylor (2003)는 전력수요예측 모형에서 이중 계절형 지수평활법과 이중 계절형 ARIMA 모형을 통해 예측한 바 있으며, Taylor (2010)는 일별과 주별, 연도별 주기를 이용하여 삼중 계절형 지수평활법과 삼중 계절형 ARIMA 모형을 통해 예측한 바 있다.

결합예측(Combining Forecasting)은 여러 가지 예측 방법을 결합함으로써 예측에 정확도를 높일 수 있다고 보였다. Caiado (2010)는 수도사용량 수요예측 모형에 이중 계절형 일변량 시계열 예측모형들의 예측 값을 결합하여 그 효용성을 비교해 본 바 있다. 계절형에 대한 연구를 위하여 계절형이 없는 예측과 계절주기를 하나로만 본 예측과 이중 계절형을 가진 예측을 비교해 보기로 한다. 본 연구는 Holt-Winters, ARIMA, AR-GARCH 일변량 계절시계열 모형을 이용한 예측 방법과 이들을 결합한 모형들의 예측 성능을 비교 분석하고자 한다.

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2010-0010237).

³교신저자: (156-756) 서울 동작구 흑석동 221, 중앙대학교 응용통계학과, 교수. E-mail: sahm@cau.ac.kr

2. 시계열 모형의 소개

2.1. 계절형 지수평활법

지수평활법은 매우 단순하지만 시계열 예측에서 매우 유용한 기법으로 사용이 된다. 지수평활법의 표준 계절조정 방법으로 Holt-Winters의 가법 추세성분과 승법추세성분, 감소가법추세성분, 감소승법추세성분 등을 포함하고 있다. 관광수요예측에서 2개의 계절주기를 가지는 이중계절 Holt-winters 지수평활법 (Taylor, 2003)을 사용하기로 한다. 이에 따라 7일과 364일을 주기로 모형에 적용을 하려고 한다.

원자료에 대하여 로그변환 후 1차 차분을 함에 따라서 변수에 음의 값이 존재함으로 승법 모형과 가법 모형 중 가법 모형을 선택하여 이에 이중 계절성을 적용시킨 이중계절성 Holt- Winters 가법 모형을 선택을 하기로 한다.

이중계절성 Holt-winters 지수평활법은 다음에 의해 정의 된다.

$$\begin{aligned} L_t &= \alpha(Y_t - S_{t-7} - D_{t-364}) + (1 - \alpha)(L_{t-1} + T_{t-1}), \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \\ S_t &= \gamma(Y_t - L_t - D_{t-364}) + (1 - \gamma)S_{t-7}, \\ D_t &= \delta(Y_t - L_t - S_{t-7}) + (1 - \delta)D_{t-364}, \\ F_{t+h} &= L_t + T_t \times h + S_{t+h-7} + D_{t+h-364}, \end{aligned} \quad (2.1)$$

여기서 L 은 평활 수준을 나타내며, T 는 평활가법추세를 의미하며, S 는 7일에 대한 계절지표를 나타내며 D 는 364일에 대한 계절지수를 계절지표를 의미한다. α, β 는 수준과 추세에 대한 모수이고 γ, δ 는 계절평활 모수이며, F_{t+h} 는 h 시점 후에 예측한 값을 의미한다.

식 (2.1)에 초기값은 식 (2.2)를 통하여 구할 수 있다.

$$\begin{aligned} L_7 &= \frac{1}{7} \sum_{t=1}^7 Y_t, \quad L_{364} = \frac{1}{364} \sum_{t=1}^{364} Y_t, \\ T_7 &= \frac{1}{7^2} \left(\sum_{t=8}^{14} Y_t - \sum_{t=1}^7 Y_t \right), \quad T_{364} = \frac{1}{364^2} \left(\sum_{t=365}^{728} Y_t - \sum_{t=1}^{364} Y_t \right), \\ S_1 &= Y_1 - L_7, \dots, S_7 = Y_7 - L_7, \\ D_1 &= Y_1 - L_{364}, \dots, D_{364} = Y_{364} - L_{364}. \end{aligned} \quad (2.2)$$

평활 모수 $\alpha, \beta, \gamma, \delta$ 는 1차 예측단계에서 MSE를 최소화하는 최적값을 적용한다.

2.2. 계절형 ARIMA모형

일변량 시계열 모형 중 수요예측에 사용하는 분석 방법 중에서 가장 널리 이용되는 방법의 하나는 Box와 Jenkins (1994)의 계절형 자기회귀누적이동평균(Autoregressive Integrated Moving Average; ARIMA) 모형이다. 이 모형은 단기예측과 계절형 변동을 포함하는 시계열의 분석에 효과적이다.

Box의 이중 계절형 자기회귀누적의 기본 형태를 보면 아래의 모형과 같다.

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\Pi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}Z_t = \theta_q(B)\Theta_{Q_1}(B^{s_1})\Psi_{Q_2}(B^{s_2})\epsilon_t, \quad (2.3)$$

여기서 c 는 상수항을 말하고, B 는 $B^k Y_t = Y_{t-k}$ 에 연산을 맡는 후진연산자이고, $\phi_p(B)$ 와 $\theta_q(B)$ 는 p 차 AR과 q 차 MA를 의미하며, 다항식을 말하며, $\Phi_{P_1}(B^{s_1})$, $\Pi_{P_2}(B^{s_2})$, $\Theta_{Q_1}(B^{s_1})$, $\Psi_{Q_2}(B^{s_2})$ 는 s_1, s_2 에

따르는 P_1, P_2, Q_1, Q_2 차 계절성 AR과 계절성 MA를 의미한다. d, D_1, D_2 ,는 차분의 차수를 말하며 ϵ_t 는 평균 0과 고정된 분산을 따르는 백색잡음과정이다. 다항식 $\phi_p(B) = 0, \theta_q(B) = 0, \Theta_{Q_1}(B^{s_1}) = 0, \Psi_{Q_2}(B^{s_2}) = 0, \Phi_{P_1}(B^{s_1}) = 0, \Pi_{P_2}(B^{s_2}) = 0$ 의 단위근은 단위원 밖에 위치한다.

2.3. 계절형 AR-GARCH 모형

실제적으로 시계열 예측 및 모델링을 하는 데에 있어, 이분산성이 많이 발생한다. 이에 대하여 Engle (1982)과 Bollerslev (1986)는 이분산 문제점을 다루는 방법으로 ARCH와 GARCH 모형을 제안했다. GARCH 모형은 분산에 대한 모형을 적용함으로써 모형의 추정 및 예측을 가능하게 한다. AR(k)-GARCH(p, q) 모형은 다음과 같은 형태로 제안된다.

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_k y_{t-k} + \epsilon_t, \\ \epsilon_t &= e_t \sqrt{h_t}, \\ h_t &= \alpha_0 + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^q \beta_j h_{t-j}. \end{aligned} \tag{2.4}$$

이는, $e_t \sim iid N(0, \sigma^2), \alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1$ 을 따른다. p 는 ARCH 부분의 차수이고 q 는 GARCH 부분의 차수이다.

계절요소에 의한 영향을 예측하기 위하여 우리는 다음과 같은 형태로 정의를 하기로 한다.

$$\begin{aligned} \phi_p(B)\Phi_{p_1}(B^{s_1})\Pi_{p_2}(B^{s_2})Y_t &= \epsilon_t, \quad \epsilon_t = e_t \sqrt{h_t}, \quad e_t \sim iid N(0, \sigma^2) \\ h_t &= \alpha_0 + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^q \beta_j h_{t-j} + \sum_{j=1}^{p_1} \alpha_{js_1} \epsilon_{t-is_1}^2 \\ &+ \sum_{j=1}^{q_1} \beta_{js_1} h_{t-js_1} + \sum_{j=1}^{p_2} \alpha_{js_2} \epsilon_{t-is_2}^2 + \sum_{j=1}^{q_2} \beta_{js_2} h_{t-js_2}, \end{aligned} \tag{2.5}$$

여기서 s_1, s_2 는 t 시점에서 계절주기를 의미하며 AIC(Akaike information criterion)가 최소인 모형을 최적의 모형으로 고려한다.

2.4. 결합 예측 방법

Armstrong (2001)은 여러 예측 방법을 결합함으로써 예측의 정확도를 높일 수 있음을 보였다. 본 연구에서는 Holt-Winters(HW), ARIMA(A), GARCH(G) 모형을 이용한 결합한 예측 방법을 식 (2.6)과 같이 정의한다.

$$F_t = \frac{F_t^{HW} + F_t^A + F_t^G}{3} \tag{2.6}$$

F^{HW}, F^A, F^G 은 각각 Holt-Winters, ARIMA, GARCH 모형을 이용한 예측 값이고, F 는 이를 단순 결합한 값이다. 더 나은 결합 예측을 하기 위하여 MSE(Mean Squared Error)와 SE(Squared Error)에 비율에 따라서 결합을 한다. 식 (2.7)과 (2.8)은 각 모형의 MSE값과 SE값을 고려하여 결합 예측하는 방법으로 $MSE^{(HW)}, MSE^{(A)}, MSE^{(G)}$ 는 Holt-Winters, ARIMA, GARCH 모형에 대응되는 시계열 모형을 이용한 예측된 MSE이고, MSE의 합은 $M = MSE^{(HW)} + MSE^{(A)} + MSE^{(G)}$ 로 정의된다. $SE^{(HW)}, SE^{(A)}, SE^{(G)}$ 는 Holt-Winters, ARIMA, GARCH 모형에 대응되는 시계열 모형을 이용한 예

측된 SE이고, 예측된 SE는 $SE = (Y_t - F_t)^2$ 와 같이 정의되며, SE의 합은 $S_t = SE_t^{(HW)} + SE_t^{(A)} + SE_t^{(G)}$ 로 정의한다.

$$F_t^{MSE} = \frac{(M - MSE^{(HW)}) F_t^{(HW)} + (M - MSE^{(A)}) F_t^{(A)} + (M - MSE^{(G)}) F_t^{(G)}}{2M}, \quad (2.7)$$

$$F_t^{SE} = \frac{(S_t - SE_t^{(HW)}) F_t^{(HW)} + (S_t - SE_t^{(A)}) F_t^{(A)} + (S_t - SE_t^{(G)}) F_t^{(G)}}{2S_t}. \quad (2.8)$$

2.5. 예측평가방법

t 시점의 실제 관측을 Y_t , 예측값을 F_t 라고 지정을 한다. MSE(Mean Squared Error) 통계량은 사후 관측값을 얻은 구간 $t = m + 1, m + 2, \dots, n$ 에서 다음의 식으로 지정이 된다.

이에 따라 MSE와 SE에 의한 결합 예측은 각각 아래에 식 (2.9), (2.10)으로 나타난다.

$$MSE = \frac{1}{n - m} \sum_{t=m+1}^n (Y_t - F_t)^2. \quad (2.9)$$

이 통계량은 m 개의 모형을 추정하는데 사용된 데이터 자료를 통하여 설정한 모델의 예측의 정확도를 평가하기 위하여 사용을 하게 된다. 이에 대하여 추가적으로 RMSE(Root Mean Square Error) 통계량은 아래와 같이 나타난다.

$$SE = \sqrt{\frac{1}{n - m} \sum_{t=m+1}^n (Y_t - F_t)^2}. \quad (2.10)$$

3. 자료 분석

본 연구에서 사용한 자료는 2003년 7월 1일 부터 2010년 12월 31일까지로 2741개의 일별 입출국자료에서 출국자수를 이용하여 분석을 하였다. 여기서 2004년과 2008년은 2월이 29일까지 있는데 29일을 없앨 경우 주간 시계열 예측이 영향을 받을 수 있음에 따라 제거하지 않고 연간 계절형을 364일로 정하기로 하여, 계절 성분을 주간을 7일로 연간 계절형을 364일에 기반을 두고 예측을 하려고 한다. 이때 모형을 추정하는데 사용된 자료로 2003년 7월 1일 부터 2010년 6월 30일까지로 2557개 자료를 이용하였고, 평가 자료로 2010년 7월 1일 부터 2010년 12월 31일까지로 184개 자료를 이용하였다. 예측력을 비교하기 위하여 2010년 7월 1일부터 1달간의 자료에 1주 간격으로 예측력에 대하여 RMSE에 따라서 평가해 보고 6개월간에 예측자료를 이용하여 1달 간격으로 증가시키며 예측력을 RMSE에 따라서 평가해 보도록 한다.

그림 3.1에서 보면 원자료와 로그변환과 일반차분을 한 자료의 시계열도표와 ACF, PACF를 나타낸 것이다. 이에 따라서 원자료의 시계열 도표가 일정한 주기를 가지고 같은 패턴이 반복 되고 있음을 알 수 있고 ACF가 lag에 따라서 느리게 감소를 하고 있으며 7일을 주기로 반복적으로 높아짐을 볼 수 있다. 또한 PACF를 보면 7을 주기로 반복적인 패턴을 볼 수 있다. 이에 따라서 7일에 대한 주기가 나타난다고 할 수 있다. 또한 364일을 주기로 시계열도표가 어떠한 패턴이 나타남에 따라서 7일과 364일에 대한 주기를 모두 가지고 있다고 할 수 있다. 본 연구에서는 로그변환과 일반차분을 통하여 정상성을 만족시킨 자료에 ARIMA 모형과 ARCH 모형, Holt-Winters 모형을 적용시킨다.

이중계절형 ARIMA 모형은 $ARIMA(p, d, q) \times (P_1, D_1, Q_1)_{s_1} \times (P_2, D_2, Q_2)_{s_2}$ 와 같이 표기된다. 모형

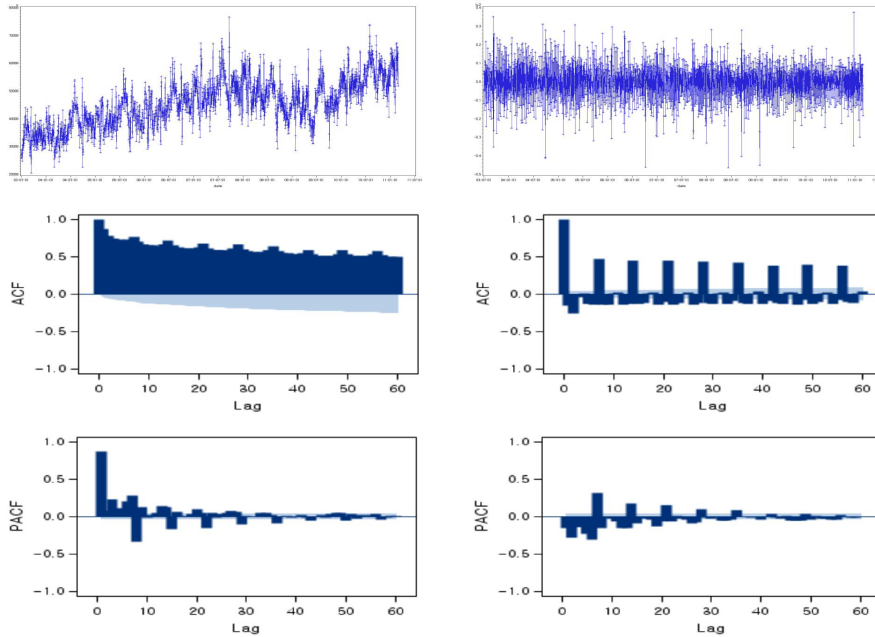


그림 3.1. 원자료와 로그화 후 1차 차분한 자료의 시계열도표와 ACF, PACF

표 3.1. 관광수요예측에 추정 ARIMA 모형

ARIMA(1, 0, 3) × (1, 0, 1) ₇ × (0, 1, 1) ₃₆₄				Residual ACF		Residual PACF	
parameter	lag	Estimate	Standard error	lag	Estimate	lag	Estimate
<i>c</i>	0	-0.00005	-0.00000	1	0.00001	1	0.00187
<i>θ</i> ₁	1	0.88501	0.12224	2	-0.00002	2	-0.00326
<i>θ</i> ₂	2	0.19421	0.09546	3	-0.00012	3	-0.02413
<i>θ</i> ₃	3	-0.09101	0.10345	4	0.00006	4	0.01244
<i>Θ</i> ₁	7	0.85591	0.11019	5	0.00018	5	0.03732
<i>Ψ</i> ₁	364	0.78493	0.16025	6	-0.00006	6	-0.01356
<i>φ</i> ₁	1	0.64776	0.14261	7	0.00015	7	0.03218
<i>Φ</i> ₁	7	0.95523	0.03399	8	-0.00013	8	-0.02583
				9	0.00006	9	0.01059
				10	0.00000	10	0.00079
				11	-0.00003	11	-0.00657
				12	0.00009	12	0.01785

은 모형의 잔차 적합 검정에 만족하고 AIC(Akaike information criterion)값을 기준으로 최소인 모형을 선택하였으며, 모형이 너무 복잡해지지 않게 하기 위하여 lag3까지 고려한다.

예측된 모형은 ARIMA(1, 0, 3) × (1, 0, 1)₇ × (0, 1, 1)₃₆₄와 같이 나타났다. 모수추정량 및 ACF와 PACF는 표 3.1과 같이 나타난다. AR-GARCH 모형도 AIC가 최소인 모형을 선택한다. GARCH 모형만을 적용 시에 적용이 되지 않음에 따라서 변수간의 상관성을 보정하기 위하여 AR(2)(1)₇(1)₃₆₄를 적용시킨 AR(2)(1)₇(1)₃₆₄-GARCH(3, 0)(1, 0)₇(1, 1)₃₆₄ 모형을 선택한다. Holt-Winters 모형은 최적화된 가중치를 적용하여 모형을 결정한다.

표 3.2. 단순계절형 모형과 이중 계절형 모형에 비교

RMSE	$S = 7$			$S = 7,364$			RW
	ARIMA	HW	GARCH	ARIMA	HW	GARCH	
1MONTH	0.040099	0.046528	0.043152	0.041624	0.041556	0.036799	0.064075
2MONTH	0.040825	0.048703	0.051347	0.041386	0.042268	0.040761	0.072427
3MONTH	0.057013	0.07031	0.064561	0.063578	0.059855	0.059336	0.080431
4MONTH	0.052228	0.077284	0.063782	0.05754	0.055416	0.053653	0.078609
5MONTH	0.050278	0.085522	0.068508	0.053696	0.053795	0.052928	0.08284
6MONTH	0.06763	0.107941	0.079441	0.064501	0.07042	0.066941	0.091179

표 3.3. 단순계절형 결합 모형과 이중 계절형 결합 모형에 비교

RMSE	$S = 7$			$S = 7,364$			RW
	단순 결합	MSE기반 결합	SE 기반 결합	단순 결합	MSE기반 결합	SE 기반 결합	
1WEEK	0.031652	0.031165	0.031165	0.034724	0.034045	0.03317	0.076907
2WEEK	0.034794	0.034011	0.034011	0.035056	0.034097	0.032514	0.075301
3WEEK	0.030744	0.03026	0.03026	0.030726	0.029906	0.028094	0.071174
4WEEK	0.032951	0.032726	0.032726	0.030818	0.03015	0.026728	0.067311

표 3.4. 단순 모형과 결합 모형의 비교

RMSE	ARIMA	HW	GARCH	ARIMA- HW	ARIMA- GARCH	HW- GARCH	단순 결합	MSE기반 결합	SE 기반 결합	RW
1MONTH	0.041624	0.041556	0.036799	0.034422	0.035597	0.035194	0.033249	0.032542	0.027855	0.064075
2MONTH	0.041386	0.042268	0.040761	0.034444	0.038515	0.035657	0.034325	0.034019	0.029697	0.072427
3MONTH	0.063578	0.059855	0.059336	0.056224	0.057831	0.055395	0.054925	0.054628	0.0495	0.080431
4MONTH	0.05754	0.055416	0.053653	0.050638	0.051964	0.049998	0.049212	0.048958	0.043832	0.078609
5MONTH	0.053696	0.053795	0.052928	0.047989	0.0494	0.049105	0.047187	0.046993	0.041939	0.08284
6MONTH	0.064501	0.07042	0.066941	0.062688	0.062632	0.065589	0.062377	0.062167	0.058175	0.091179

3.1. 자료 분석 결과

본 연구에서 사용한 모형은 계절주기를 주별로 7일 하나만 사용한 단순 계절형 ARIMA, Holt-Winters, AR-GARCH 모형과 계절주기로 주별과 연별로 7일과 364일의 2개의 계절주기를 이용 이용한 이중 계절형 ARIMA, Holt-Winters, AR-GARCH 모형을 사용하였다. 추가적으로 RW 모형을 사용한 예측값을 기준으로 다른 모형들과 비교를 하였다. 이에 대하여 1주에서 4주까지의 주별간격으로 비교를 하였고, 1개월부터 6개월까지 한달 단위의 자료의 예측값을 통하여 RMSE를 비교하였다.

6개월간의 자료를 통하여 7일을 계절주기로 사용한 계절형 모형과 7일과 364일의 2개의 계절주기를 사용한 이중계절형 ARIMA, Holt-Winters, AR-GARCH 모형을 비교해 본다. 표 3.2에 결과를 보면 ARIMA를 제외한 HW와 GARCH 모형에서 2개 모형에서 이중계절형 모형이 더 좋은 결과를 보였다. ARIMA에 경우는 단순 계절형 모형이 높게 나오지만 그 차이가 매우 작음으로 전체적으로 이중 계절형 모형이 더 나은 결과를 보임을 알 수 있다.

1달간의 자료를 주별로 증가시키며 7일을 주기로 한 결합모형과 7일과 364일의 2개의 주기를 가지고 계절형 결합모형을 비교해 본 결과 표 3.3에 따라서 2주까지는 오히려 계절 주기를 7만으로 한 것이 단기 예측으로 더 높게 나타났다. 그러나 2주 이후부터는 7일과 364일을 동시에 고려한 이중 계절형 모델들의 결합예측이 더 좋게 나타난다. 이는 2주까지는 7일 주기에 대한 고려만으로 예측이 가능하지만 그 이상에 예측에 대하여는 1년 주기인 364일 주기도 고려해야만 하는 것으로 생각 할 수 있다.

표 3.5. 월별측정자료와 일별측정자료 예측 비교

RMSE	월별측정자료			일별측정자료		
	ARIMA	HW	GARCH	ARIMA	HW	GARCH
1MONTH	0.05615	0.055592	0.056192	0.041624	0.041556	0.036799
2MONTH	0.059209	0.059773	0.05926	0.041386	0.042268	0.040761
3MONTH	0.069554	0.069868	0.069809	0.063578	0.059855	0.059336
4MONTH	0.067628	0.067897	0.067785	0.05754	0.055416	0.053653
5MONTH	0.071425	0.071711	0.071547	0.053696	0.053795	0.052928
6MONTH	0.081506	0.081809	0.081589	0.064501	0.07042	0.066941

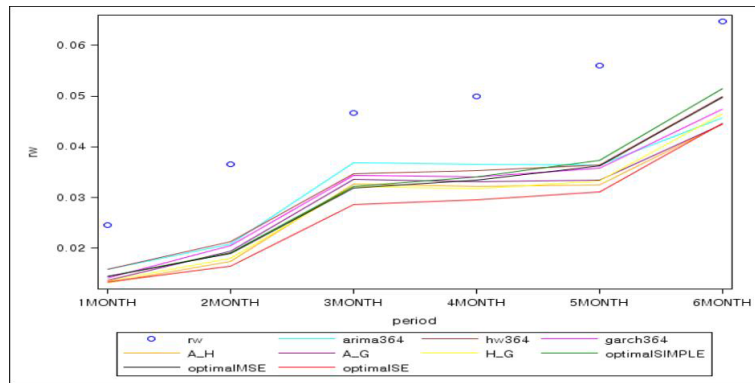


그림 3.2. 모형들의 RMSE비교

단순 결합 예측과 MSE기반에 결합 예측, SE 기반에 결합 예측 중에서는 SE기반에 결합 예측이 가장 좋은 예측 결과를 나타내고 있다. 이에 따라서 SE기반 예측이 가장 최적에 예측으로 보인다. 이에 대하여 그림 3.2와 표 3.4에 이중 계절형 모형들의 예측값의 RMSE를 비교해 보면 Holt-Winters 모형과 ARIMA, AR-GARCH 모형에 대한 비교에서 AR-GARCH 모형이 가장 나은 결과를 보이고 있다. 그러나 6번째 달에 불규칙한 변동에 대하여 ARIMA 모형이 가장 적게 변함을 볼 수가 있다.

모형들의 결합을 2개 모형씩 결합한 것과 3개 모형을 최적화 시켜 결합한 것들을 보면 모든 결합모형들이 결합을 하기 이전에 모형보다 나은 결과를 보였고, 특히 2개 모형에 결합보다는 3개 모형의 결합이 더 좋은 결과를 보였다. 특히 SE기반에 결합예측이 가장 최적의 모형임을 알 수가 있었다. 단순 결합예측보다는 MSE와 SE를 기반으로 한 최적화 결합예측이 더 나은 것으로 나타났다.

표 3.5에서 일별자료가 아닌 월별추정을 이용하여 일별 추정에 적용을 해 본 결과를 보면 월별자료를 통한 예측과 같은 경우에는 계산에 있어서 편할 수는 있으나 그 정확도가 현저하게 떨어지고 일별 자료에 변동을 고려하지 못하여 세세한 변동을 예측하지 못하여 좋지 못한 결과를 나타낼 수 있었다. 이에 따라 일별 자료를 수집을 하여 일별자료를 통한 예측을 하는 것이 더 정확한 예측을 하게 한다는 것을 알 수 있었다.

4. 결론

본 연구는 출국자수에 대한 시계열 모형 적용 및 분석이 관건이다. 시계열 예측 모형을 만들 때 단순 계절형 모형을 사용할 것인지 이중 계절형 모형을 사용할 것인지에 대하여 생각해 보았으며, 모형의 예측값의 결합을 통하여 더 우수한 예측값을 얻을 수 있다는 것을 알 수 있었다. 그리고 일별 수집 자료와 월

별 수집 자료를 통하여 어느 것이 더 우수한 예측을 할 수 있는지 비교하였다.

이에 대하여 여행사업과 같이 단기간에 미래가 아닌 몇 달 이후에 여행자 수를 예측하여 대비하여야 하는 사업과 같은 경우에는 이중 계절형 모형을 사용하는 것이 옳다는 결론을 내렸다. 2주까지에 단기간 예측에서는 단순 계절형 모형이 더 나은 결과를 보였지만 3주 이후의 예측에서 이중 계절형 모형이 더 나은 결과를 보임에 따라 이중 계절형 모형이 여행 사업에서 필요하다고 생각이 된다. 또한 한 개 모형만을 사용하는 것이 아닌 모형들의 예측값을 결합하여 예측하는 것이 더 나은 결과를 가졌다. 2개에 모형을 결합하는 것보다 3개의 모형을 결합하는 것이 더 좋은 결과를 나타내었고 특히 SE를 기반으로 한 결합은 다른 모형들보다 특히 좋은 결과를 나타내었다.

일별 수집 자료와 월별 수집 자료를 이용한 분석 비교에서는 일별자료가 월별자료에 대한 예측에 비해 훨씬 높은 능력을 보였다. 편리성을 위하여 월별 예측 자료로 수집이 되고 있는데 정확한 예측을 위하여 일별 자료로 수집을 하여야 될 것으로 요구된다.

본 연구에서 예측된 모형을 출국자수 예측에 적용하여 미래 예상 출국자 수를 고려할 수 있는데 이를 각 나라별 출국자수 연구로 발전시킨다면 각 항공사들의 항공편 편성에 도움을 줄 수 있을 것이며, 그 외에 여행관련 회사들의 사전 사업구성에 도움을 줄 수 있을 것으로 보인다. 또한 이를 출국자가 아닌 입국자에 대한 분석을 통하여 인원이 많이 몰리는 때와 적게 몰리는 때를 고려한다면 국내 관광 사업에 도움을 줄 수 있을 것으로 보인다.

이중 계절형 모형은 관광 시계열 예측뿐만 아니라 다른 분야에도 적용될 수 있는 분야가 많다고 생각이 된다. 특히 인터넷트래픽 예측과 같은 분야는 5분 단위 이하로 자료가 수집이 되고 있는데 이는 일별 주기와 주별 주기를 동시에 가지고 있음에 따라 적용이 가능할 것이다. 또한 에너지 사용량과 같은 경우는 일별 주기와 주별 주기, 연도별 주기를 가지고 있다고 생각되는데 이에 따라서 이중계절형 모형을 확장시킨 삼중 계절형 모형에 대한 적용을 고려할 필요가 있다고 생각되며, 이에 대한 연구가 계속 되어야 할 것이다.

참고문헌

- 윤지성, 허남균, 김삼용, 허희영 (2010). 계절형 다변량 시계열 모형을 이용한 국제항공 여객 및 화물 수요예측에 관한 연구, <한국통계학회 논문집>, **17**, 473-481.
- 허남균, 정재윤, 김삼용 (2009). 다변량 시계열 모형을 이용한 항공 수요 예측 연구, <응용통계연구>, **22**, 1007-1077.
- Armstrong, J. (2001). Combining forecasting, *International Series in Operations Research and Management Science*, 417-440.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.
- Box, G. E. P. and Jenkins, G. (1994). *Time Series Analysis: Forecasting and Control*, Prentice Hall.
- Caiado, J. (2010). Performance of combined double seasonal univariate time series models for forecasting water demand, *Journal of Hydrologic Engineering*, **15**, 215-222.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica: Journal of the Econometric Society*, 987-1007.
- Shen, S., Li, G. and Song, H. (2009). Effect of seasonality treatment on the forecasting performance of tourism demand models, *Tourism Economics*, **15**, 693-708.
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing, *Journal of the Operational Research Society*, **54**, 799-805.
- Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting, *European Journal of Operational Research*, **204**, 139-152.

A Study on the Tourism Combining Demand Forecasting Models for the Tourism in Korea

H. G. Son¹ · M. H. Ha² · S. Kim³

¹Department of Applied Statistics, Chung-Ang University

²Department of Applied Statistics, Chung-Ang University

³Department of Applied Statistics, Chung-Ang University

(Received February 27, 2012; Revised March 14, 2012; Accepted March 28, 2012)

Abstract

This paper applies forecasting models such as ARIMA, Holt-Winters and AR-GARCH models to analyze daily tourism data in Korea. To evaluate the performance of the models, we need single and double seasonal models that compare the RMSE and SE for a better accuracy of the forecasting models based on Armstrong (2001).

Keywords: ARIMA model, Holt-Winters model, AR-GARCH model, combining forecasting.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2010-0010237).

³Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, Dongack-Gu, Seoul 156-756, Korea. E-mail: sahm@cau.ac.kr