

관광 수요 예측 모형의 계절효과에 대한 연구

김삼용¹ · 이주형²

¹중앙대학교 통계학과, ²중앙대학교 통계학과

(2010년 11월 접수, 2010년 11월 채택)

요약

본 연구는 관광수요 예측 분야에서 사용되는 계절형 ARIMA 모형과 다변량 계절형 시계열 모형과 오차수정모형의 성능을 비교한 것이다. 본 연구에서는 일본, 중국, 미국, 필리핀에 대한 실제 자료를 이용한 결과 관광 수요에는 계절성이 중요한 역할을 하는 것을 보이고 각 국가별로 예측 정확도를 RMSE를 기준으로 하여 비교하였다.

주요어: 계절형 ARIMA 모형, 다변량 시계열 모형, 오차수정모형, 관광수요.

1. 서론

수요예측은 다양한 분야에서 자료를 분석하고 예측하는데 널리 쓰이는 기법이다. 특히 관광분야는 급변하는 경제환경 속에서 중요한 경제활동이기 때문에 이러한 경제환경 속에서 사회현상을 정확히 설명하고 변화의 크기 또한 정확히 파악하는 것이 중요한 과제이다. 그렇기 때문에 정확한 추정과 예측을 하기 위한 연구가 꾸준히 이루어지고 있다. 일례로 허남균 등 (2009)은 항공 수요 예측을 위하여 일변량 시계열 모형과 다변량 시계열 모형을 이용하였고, 윤지성 등 (2010)은 국제 항공 여객 및 화물의 수요 예측을 위하여 계절형 일변량 시계열 모형과 계절형 다변량 시계열 모형을 이용하였다. Hylleberg 등 (1990)은 계절성을 갖는 자료를 일변량 시계열 모형과 다변량 시계열 모형을 소개하였다. Kulendran과 King (1997)은 미국, 일본, 영국, 뉴질랜드로부터 호주로의 분기별 관광객 수요를 예측하기 위해 일변량 시계열 모형, 다변량 시계열 모형, 회귀모형을 이용하였으며 Lim과 McAleer (2002)은 홍콩, 말레이시아, 싱가포르로부터 호주로의 관광객 수요예측을 위하여 ARIMA모형과 계절형 ARIMA모형을 이용하였다. Kulendran과 Wong (2005)은 비정상 계절형 시계열을 두 가지 ARIMA모형을 이용하여 예측하였으며, Li 등 (2005)은 최근 개발된 계량경제학적 모형의 예측을 비교하여 나타내었고, Wong 등 (2007)은 10개의 국가로부터 홍콩으로의 관광객 수요예측을 위하여 일변량 시계열 모형과 다변량 시계열 모형과 이 모형들의 예측 결합을 비교하여 분석하였다. Shen 등 (2009)은 영국에서 7개의 국가로의 분기별 관광객 수요 예측을 위하여 일변량 시계열 모형과 다변량 시계열 모형을 이용하여 비교 분석하였다. 관광 수요 예측에서 계절성은 관광 수요의 가장 중요한 특징 중의 하나이고 관광 산업의 많은 부분에서 중요한 영향을 가진다는 것이 널리 알려져 있다. 게다가 계절성은 관광 조사부분의 중요한 하나로 인식되어 있다. 따라서 일변량 계절시계열 모형을 이용한 예측 방법과 여러 국가로의 관광 수요를 동시에 수행할 수 있는 다변량 시계열 모형과 오차수정모형을 이용하여 각 국가별로 계절효과의 유무와 모형들의 예측 성능을 비교 분석하고자 한다. Diebold와 Kilian (2000)에서 모형을 선택할 때 단위근 검정을 이용하여 선택하는 것이 유용한 방법이라는 것을 알 수 있기 때문에 단위근 검정을 이용하여 모

¹교신저자: (156-756) 서울시 동작구 흑석동 221, 중앙대학교 통계학과, 교수. E-mail: sahm@cau.ac.kr

을 선택한다. 또한 도출된 모형에서 전체적으로 어느 한 예측 기법이 다른 예측 기법보다 우월하다 해도 변수별로 비교하였을 경우는 우월하다 할 수는 없기 때문에 실제 추정된 모형을 이용한 예측에서 각각의 예측치를 비교 분석하여 어느 변수에서 어떤 시계열 모형이 우수한지를 보고자 한다.

2. 시계열 모형의 소개

2.1. 다변량 시계열 모형

다변량시계열(multivariate time series)은 몇 개의 일변량시계열들로 구성되어 있다. 즉, VAR모형은 어떤 일변량시계열의 현재 움직임과 과거값뿐만 아니라 다른 시계열의 과거값에도 영향을 받아 결정될 때 유용한 모형이다.

이러한 경우, 같은 시점에서 두 개 이상의 시계열들로 구성된 벡터시계열에 관심을 갖고 이 벡터시계열을 구성하는 각 성분 계열들의 자기종속성이나 관련성은 물론이고 서로 다른 성분계열 사이의 연관성이나 교차상관을 파악할 수 있으며 이들 모두를 예측모형에 사용하는 것이 효과적일 수 있다. 특히 VAR모형은 다른 대표적인 다변량시계열모형보다 모수의 추정이나 식별성의 문제가 적고 모형화 하는데 이용되는 간편하고 적용성이 높은 모형이기 때문에 유용하게 쓰이는 모형이다. VAR(p)모형을 정의하면 z_t 가 모든 t 에 대하여 식 (2.1)을 만족한다.

$$z_t = \mu + \Phi_1 z_{t-1} + \dots + \Phi_p z_{t-p} + \varepsilon_t \quad (2.1)$$

이다. 여기서 Φ_i 는 $k \times k$ 행렬이며, μ 는 k -차원 벡터이고 $\{\varepsilon_t\}$ 는 평균벡터가 영벡터이고 공분산행렬이 Σ 인 k 변량백색잡음과정이다. 후진연산자(backward shift operator) B 를 사용하면 $\Phi(B)z_t = \mu + \varepsilon_t$ 로 간단히 표현할 수 있으며 여기서 $\Phi(B) = I - \Phi_1 B - \dots - \Phi_p B^p$ 이다. 또한 여기서 $\Sigma = E(\varepsilon_t \varepsilon_s')$ 은 $t = s$ 일 때 σ^2 이고 다르다면 0이다.

일변량 시계열모형에서 ARIMA모형의 차수를 결정할 때, ACF와 PACF를 보고 판단한 것과 마찬가지로 다변량 시계열모형에서도 차수의 결정을 위해 ACF와 PACF의 패턴을 보고 결정할 수 있으나 이러한 방법은 많은 경험을 필요로 한다. 그러므로 일반적으로 VAR모형에 대한 분산공분산행렬의 추정량($\widehat{\Sigma}_p$)을 이용하여 AIC(Akaike Information Criterion), SBC(Schwarz Bayesian Criterion), HQC(Hannan-Quinn Criterion)를 구한 후 최적의 차수로 결정한다. 일반적으로 AIC기준은 점근적으로 p 에 대한 과추정을 만드는 것으로 알려져 있다. 따라서 본 논문에서는 HQC기준을 사용하였다.

2.2. 오차수정(ECM)모형

Engle과 Granger (1987)에 의해 제안된 오차수정모형(ECM)은 시계열간의 공적분관계를 다른 단기 동적인 관계와 함께 고려한 동적인 모형이다. 단위근을 갖는 시계열 $X_t = (X_{1t}, X_{2t})'$ 가 일반차분으로 정상시계열이 되고 벡터 $\alpha = (1, -\beta')$ 가 존재한다고 할 때 αX_t 가 단위근을 갖지 않는 정상시계열이라 하면 이는 공적분되었다고 하며 α 를 공적분 벡터라 한다. 이 공적분 벡터들로 구성된 행렬 A 가 있는 k 차원의 $\{z_t\}$ 인 벡터시계열이 있다고 할 때 이 모형은

$$\nabla z_t = \Gamma A' z_{t-1} + \sum_{i=1}^p \Gamma_i \nabla z_{t-i} + \varepsilon_t \quad (2.2)$$

로 표현할 수 있다. 여기서 벡터시계열 z_t 에 1개의 공적분 관계가 존재하면 Γ 와 A 는 모두 $k \times l$ 행렬이고 A 는 1개의 공적분 벡터로 구성된 행렬이며 Γ_i 는 $k \times k$ 차원의 계수 행렬이고 ε_t 는 자기상관이 없는 백색잡음인 오차항이다. 여기서 $A' z_{t-1}$ 은 시점 $t-1$ 에서의 불균형오차를 나타내며 Γ 는 공적분 벡터의 영향을 ∇z_t 에 분배하는 조정상수들로 오차수정계수(error correction coefficient)라 한다.

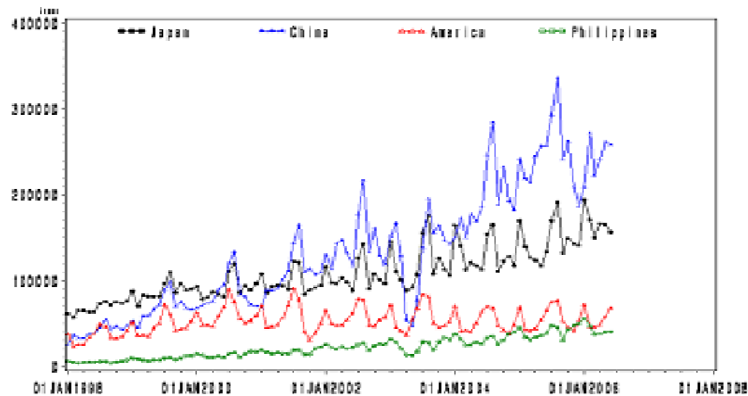


그림 3.1. 일본, 중국, 미국, 필리핀 관광 원자료

식 (2.2)의 p 차 오차수정모형은 z_t 의 수준으로 표시하면 $p + 1$ 차의 벡터자기회귀모형이 된다. 거꾸로 z_t 가 $p + 1$ 차의 벡터자기회귀과정을 따른다면 그들 구성요소들이 단위근을 갖는 시계열로서 공적분 관계를 가진다면 오차수정모형의 식으로 표현할 수 있다. 따라서 Γ_i 는 z_t 의 벡터자기회귀모형 계수의 변형된 형태라 할 수 있다.

3. 관광 자료 분석

본 연구에서 분석 및 예측에 사용된 자료는 1998년 1월부터 2006년 6월까지의 월별 한국인 국제 관광객 수의 자료를 이용하였다. 조사기간 동안의 관광객의 총합을 구하여 1위부터 4위까지 한국인의 주요 관광국인 일본, 중국, 미국, 필리핀으로의 관광객 자료를 이용하였다. 이 자료는 한국관광공사(<http://www.visitkorea.or.kr/>)에서 제공하는 관광통계로부터 얻을 수 있다. 2006년 7월부터 국민출국카드가 폐지되어 국민 해외관광객의 행선지 파악이 어려워졌다. 이에 따라 한국관광공사에서는 각국의 관광청이나 관광관련 기관에서 제공하는 자료를 이용하여 데이터를 구축하고 있으나 다른 몇몇의 국가에서는 우리나라의 해외관광객을 기타로 분류하여 파악하기가 어렵다. 최근 구축된 자료는 2004년 1월부터 구축되어 있으나 현재까지의 자료의 양이 많지 않기 때문에 모형 설정이나 예측에도 어려움이 있다. 따라서 이전에 파악된 자료를 이용하여 1998년 1월부터 2005년 6월까지의 자료를 이용하여 일본, 중국, 미국, 필리핀으로의 국제 관광객 수 각각의 변수들의 일변량 계절 ARIMA모형을 추정하고 모든 변수를 이용한 계절형 VAR모형과 ECM모형을 추정하였다. 마지막으로 추정된 모형으로부터 2005년 7월부터 2006년 6월까지의 1년간의 예측치를 실제 데이터와 비교하여 예측의 정확도를 비교하기로 한다. 또한 각 국가별로 어떠한 모형이 잘 예측하는지를 비교하기로 한다.

3.1. 계절형 ARIMA 모형 적합 결과

그림 3.1을 살펴보면 관광자료가 우선적으로 점점 증가하는 추세를 보이는 것을 알 수 있다. 또한 12개월을 주기로 하는 계절성이 존재하는 것으로 보이며 변동의 폭도 일정하지 않다는 것을 알 수 있다. 따라서 정상성을 만족하는 데이터로 변환하기 위하여 로그변환과 계절차분 및 일반차분을 실시하였다. 그림 3.2는 변환된 자료의 시계열 그림을 나타낸다. 그림에서 계절들은 정상성을 보인다고 할 수 있고 표 3.1에서의 단위근 검정결과도 단위근이 존재한다는 귀무가설을 기각하여, 변환된 자료는 단위근이 존재하지 않는 정상 시계열이라고 볼 수 있다.

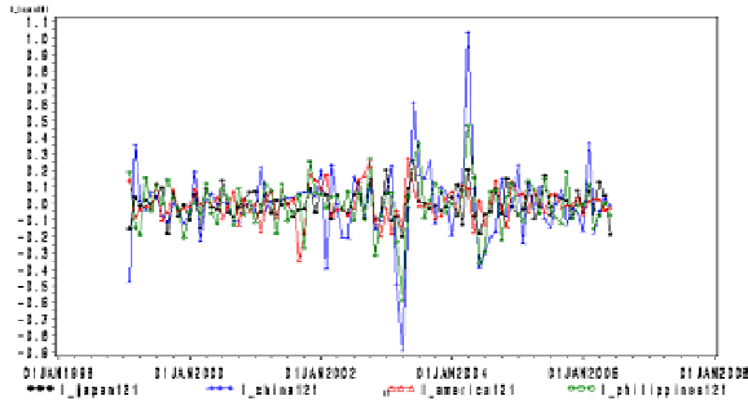


그림 3.2. 변환된 일본, 중국, 미국, 필리핀 관광 자료

표 3.1. 로그변환한 자료의 단위근 검정

변수	Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
일본	Zero Mean	2	-214.684	0.0001	-6.43	<.0001		
	Single Mean	2	-217.089	0.0001	-6.39	0.0001	20.46	0.0010
	Trend	2	-223.202	0.0001	-6.38	<.0001	20.39	0.0010
중국	Zero Mean	2	-173.632	0.0001	-6.25	<.0001		
	Single Mean	2	-173.896	0.0001	-6.21	0.0001	19.26	0.0010
	Trend	2	-176.297	0.0001	-6.18	<.0001	19.09	0.0010
미국	Zero Mean	2	-138.702	0.0001	-5.87	<.0001		
	Single Mean	2	-140.517	0.0001	-5.85	0.0001	17.12	0.0010
	Trend	2	-143.882	0.0001	-5.84	<.0001	17.07	0.0010
필리핀	Zero Mean	2	-351.592	0.0001	-7.12	<.0001		
	Single Mean	2	-353.087	0.0001	-7.07	0.0001	25.01	0.0010
	Trend	2	-358.972	0.0001	-7.03	<.0001	24.73	0.0010

변환된 해외관광객 자료를 모형에 적합하면 표 3.2와 같은 추정계수의 결과를 얻을 수 있다. 적합한 모형을 이용하여 잔차분석에 근거하여 모형을 진단하였고 추정된 계수들을 이용하여 적합한 모형은 식 (3.1)과 같다.

$$\begin{aligned}
 (1 - B)(1 - B^{12}) \log \text{일본}_t &= (1 - 0.42979B)(1 - 0.29310B^{12})\varepsilon_t, \\
 (1 - B)(1 - B^{12}) \log \text{중국}_t &= (1 - 0.25476B^2)(1 - 0.76176B^{12})\varepsilon_t, \\
 (1 - B)(1 - B^{12}) \log \text{미국}_t &= (1 - 0.72581B^{12})\varepsilon_t, \\
 (1 - B)(1 - B^{12}) \log \text{필리핀}_t &= (1 - 0.31622B^2)(1 - 0.74817B^{12})\varepsilon_t.
 \end{aligned} \tag{3.1}$$

3.2. 계절형 VAR 모형 적합 결과

위에서 살펴보았듯이 그림 3.1에서 4개의 국가의 원자료의 시계열도에서 모두 정상성을 만족한다고 볼 수 없으며 단위근 검정결과 단위근이 모두 존재하는 비정상시계열이다. 그림 3.2의 변환된 자료의 시계열도와 표 3.1을 보았을 때 변환된 자료는 단위근이 존재하지 않는 정상 시계열 자료라고 할 수 있다.

표 3.2. 각 계절 ARIMA모형의 적합결과

Variable	Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
일본	MA1,1	0.42979	0.10513	4.09	0.0001	1
	MA2,1	0.29310	0.11352	2.58	0.0118	12
중국	MA1,1	0.25476	0.11203	2.27	0.0258	2
	MA2,1	0.76176	0.08562	8.90	<.0001	12
미국	MA1,1	0.72581	0.09065	8.01	<.0001	12
필리핀	MA1,1	0.31622	0.11059	2.86	0.0055	2
	MA2,1	0.74817	0.08413	8.89	<.0001	12

표 3.3. 변환된 한국인의 4개 주요 관광국으로의 관광객 자료의 Johansen 공적분 검정

$H_0 : \text{Rank} = r$	$H_1 : \text{Rank} > r$	Eigenvalue	Trace	5% Critical Value
0	0	0.7495	288.9904	39.71
1	1	0.6813	183.7951	24.08
2	2	0.5144	96.8925	12.21
3	3	0.4245	41.9923	4.14

표 3.4. 계절형 VAR(1)모형에 대한 잔차의 백색잡음 검정

Variable	Durbin Watson	Normality		ARCH	
		Chi-Square	Pr > ChiSq	F Value	Pr > F
일본	2.04105	0.75	0.6876	0.08	0.7818
중국	1.8303	99.47	<.0001	0.09	0.7691
미국	1.99185	2.16	0.3402	0.36	0.5485
필리핀	2.08511	10.6	0.0050	0.03	0.8635

표 3.5. VAR(1)의 추정계수

Lag	Variable	일본	중국	미국	필리핀
1	일본	-0.35165	0.07133	0.19541	-0.07480
	중국	-0.06397	-0.00224	0.54245	0.21770
	미국	0.04993	-0.04329	0.06267	-0.04064
	필리핀	0.19537	0.19589	0.23996	-0.25850

계절 다변량 시계열분석에서도 정상성을 만족하는 자료를 이용해야 하며 따라서 일변량 시계열분석과 마찬가지로 로그변환과 일차차분, 계절차분으로 변환한 자료를 이용한다.

변환된 자료를 계절형 VAR모형에 적합하기 전에 4개의 국가 변수들 사이에 선형결합인 공적분 관계가 존재하는지 검정해야 한다. 만약 공적분 관계가 존재한다면 4개의 국가 변수의 다변량 시계열자료는 계절형 VAR모형에 적합할 수 없고 오차수정모형에 적용한다.

변환된 자료를 이용하여 공적분 관계가 존재하는지를 살펴보기 위해 실시한 Johansen (1988)의 대각합 통계량은 표 3.3과 같다. 검정결과 4개의 계열간에 공적분이 존재하지 않는다는 것을 알 수 있다. 그러므로 정상성을 만족하고 공적분 관계도 존재하지 않는 이 계열들을 계절형 VAR모형에 적합하고자 한다. 적합하기 위해서는 계절형 VAR모형의 차수 p 를 결정해야 하며 최적 차수를 결정하기 위해서 HQC 기준을 고려하여 이 값을 최소로 하는 모형을 선택하고자 한다. 여기서 HQC는 식 (3.2)와 같이 구한다.

$$HQC(p) = \log \left(\left| \widehat{\Sigma}_p \right| + \frac{2k^2 p \log(\log(n))}{n} \right) \quad (3.2)$$

표 3.6. 로그변환한 자료의 단위근 검정

변수	Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
일본	Zero Mean	2	0.0590	0.6941	0.83	0.8878		
	Single Mean	2	-10.1719	0.1184	-2.43	0.1377	3.34	0.2302
	Trend	2	-102.9260	0.0001	-5.62	<.0001	15.93	0.0010
중국	Zero Mean	2	0.1678	0.7197	1.12	0.9316		
	Single Mean	2	-6.6168	0.2901	-1.95	0.3077	2.67	0.3993
	Trend	2	-48.6398	0.0003	-4.40	0.0037	9.72	0.0010
미국	Zero Mean	2	0.0549	0.6931	0.30	0.7717		
	Single Mean	2	-82.6592	0.0008	-5.57	<.0001	15.64	0.0010
	Trend	2	-114.5930	0.0001	-6.01	<.0001	18.15	0.0010
필리핀	Zero Mean	2	0.2024	0.7281	1.56	0.9701		
	Single Mean	2	-3.3362	0.6095	-1.58	0.4884	2.65	0.4023
	Trend	2	-29.5503	0.0051	-3.52	0.0431	6.39	0.0550

표 3.7. 로그변환한 자료의 단위근 검정

$H_0 : \text{Rank} = r$	$H_1 : \text{Rank} > r$	Eigenvalue	Trace	5% Critical Value
0	0	0.4294	68.3786	47.21
1	1	0.1416	20.6902	29.38
2	2	0.0805	7.7105	15.34
3	3	0.0068	0.5769	3.84

HQC값이 가장 작은 차수는 $p = 1$ 인 경우이고 변환된 각 계열들 모두는 그림 3.2를 보았을 때 평균이 0이라고 할 수 있기 때문에 상수항이 없는 계절형 VAR(1)모형에 적합한다. 포트맨토 검정결과 모두 자기상관이 존재하지 않는 것으로 나타났으며 또한 잔차들이 이분산성이 존재하지 않는다는 결과를 표 3.4에서 얻을 수 있다. 따라서 계절형 VAR(1)모형이 적합하다고 할 수 있다. 계절형 VAR(1)모형에 변환된 자료를 적합시켜 얻은 추정계수들과 모형식은 표 3.5와 식 (3.3)과 같이 나타난다.

$$\begin{aligned}
 & \begin{bmatrix} \nabla_{12}\nabla(\log(\text{일본}_t)) \\ \nabla_{12}\nabla(\log(\text{중국}_t)) \\ \nabla_{12}\nabla(\log(\text{미국}_t)) \\ \nabla_{12}\nabla(\log(\text{필리핀}_t)) \end{bmatrix} \\
 &= \begin{bmatrix} -0.35165 & 0.07133 & 0.19541 & -0.07480 \\ -0.06397 & -0.00224 & 0.54245 & 0.21770 \\ 0.04993 & -0.04329 & 0.06267 & -0.04064 \\ 0.19537 & 0.19589 & 0.23996 & -0.25850 \end{bmatrix} \begin{bmatrix} \nabla_{12}\nabla(\log(\text{일본}_{t-1})) \\ \nabla_{12}\nabla(\log(\text{중국}_{t-1})) \\ \nabla_{12}\nabla(\log(\text{미국}_{t-1})) \\ \nabla_{12}\nabla(\log(\text{필리핀}_{t-1})) \end{bmatrix} + \begin{bmatrix} \varepsilon_{\text{일본}_t} \\ \varepsilon_{\text{중국}_t} \\ \varepsilon_{\text{미국}_t} \\ \varepsilon_{\text{필리핀}_t} \end{bmatrix}. \quad (3.3)
 \end{aligned}$$

3.3. ECM 모형 적합 결과

분석할 로그변환된 데이터들이 비정상시계열인 것은 그림 3.3에서 확인할 수 있으며 단위근을 갖는다는 것을 단위근 검정 결과인 표 3.6에서 확인할 수 있다. 또한 각 시계열들간의 공적분 관계를 가지고 있는지를 검정하기 위해 Johansen의 대각합 통계량을 이용한 결과는 표 3.7에 나타나 있다. 검정 결과 4개의 시계열 사이에 rank = 1의 공적분 관계가 존재함을 확인할 수 있다. 공적분 관계가 존재하므로 4개의 시계열들을 오차수정모형에 적합시켜 보았다.

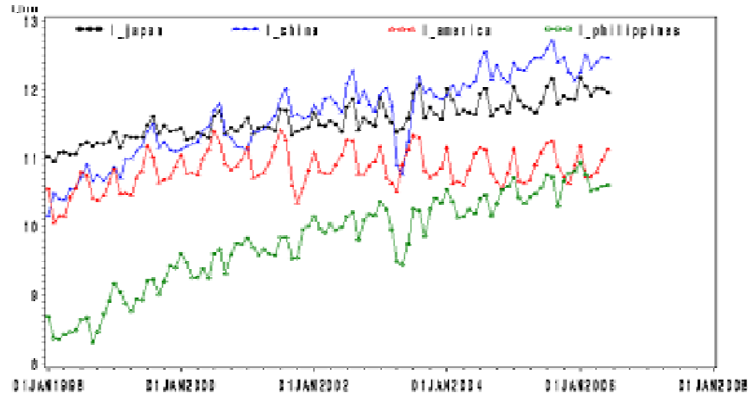


그림 3.3. 로그변환된 일본, 중국, 미국, 필리핀 관광 자료

$$\begin{aligned}
 \begin{bmatrix} \nabla \log(\text{일본}_t) \\ \nabla \log(\text{중국}_t) \\ \nabla \log(\text{미국}_t) \\ \nabla \log(\text{필리핀}_t) \end{bmatrix} &= \begin{bmatrix} -1.08728 \\ 1.67130 \\ 1.95945 \\ -6.23018 \end{bmatrix} + \begin{bmatrix} 0.05888 & 0.06024 & 0.04816 & -0.07974 \\ -0.08547 & -0.08744 & -0.06991 & 0.11575 \\ -0.10057 & -0.10289 & -0.08226 & 0.13621 \\ 0.32947 & 0.33707 & 0.26949 & -0.44622 \end{bmatrix} \begin{bmatrix} \log(\text{일본}_{t-1}) \\ \log(\text{중국}_{t-1}) \\ \log(\text{미국}_{t-1}) \\ \log(\text{필리핀}_{t-1}) \end{bmatrix} \\
 &+ \begin{bmatrix} -0.88444 & -0.00765 & 0.00148 & 0.14706 \\ -0.75866 & -0.00448 & 0.41432 & 0.16466 \\ -0.56144 & 0.17563 & -0.06825 & -0.13536 \\ -0.32947 & -0.08596 & -0.08440 & 0.23470 \end{bmatrix} \begin{bmatrix} \nabla(\log(\text{일본}_{t-1})) \\ \nabla(\log(\text{중국}_{t-1})) \\ \nabla(\log(\text{미국}_{t-1})) \\ \nabla(\log(\text{필리핀}_{t-1})) \end{bmatrix} \\
 &+ \begin{bmatrix} -1.08906 & 0.01224 & 0.42921 & -0.16969 \\ -0.43564 & -0.00918 & 0.52250 & -0.42102 \\ -0.70528 & 0.10607 & 0.16402 & -0.28940 \\ -0.65564 & -0.24426 & 0.09623 & -0.21346 \end{bmatrix} \begin{bmatrix} \nabla(\log(\text{일본}_{t-2})) \\ \nabla(\log(\text{중국}_{t-2})) \\ \nabla(\log(\text{미국}_{t-2})) \\ \nabla(\log(\text{필리핀}_{t-2})) \end{bmatrix} \\
 &+ \begin{bmatrix} 0.37743 & -0.14556 & 0.00534 & 0.14289 \\ -0.10966 & -0.13248 & 0.22923 & -0.06920 \\ -0.38579 & 0.24852 & -0.22195 & -0.37221 \\ -0.02782 & -0.28520 & -0.25734 & 0.17805 \end{bmatrix} \begin{bmatrix} \nabla(\log(\text{일본}_{t-3})) \\ \nabla(\log(\text{중국}_{t-3})) \\ \nabla(\log(\text{미국}_{t-3})) \\ \nabla(\log(\text{필리핀}_{t-3})) \end{bmatrix} \\
 &+ \begin{bmatrix} -0.27926 & -0.16922 & 0.08959 & -0.25665 \\ -0.09022 & -0.20431 & 0.09402 & -0.00115 \\ 0.03998 & -0.13686 & -0.15792 & -0.42099 \\ 0.24241 & -0.27965 & -0.17285 & -0.19165 \end{bmatrix} \begin{bmatrix} \nabla(\log(\text{일본}_{t-4})) \\ \nabla(\log(\text{중국}_{t-4})) \\ \nabla(\log(\text{미국}_{t-4})) \\ \nabla(\log(\text{필리핀}_{t-4})) \end{bmatrix} + \begin{bmatrix} \varepsilon_{\text{일본}_t} \\ \varepsilon_{\text{중국}_t} \\ \varepsilon_{\text{미국}_t} \\ \varepsilon_{\text{필리핀}_t} \end{bmatrix} \tag{3.4}
 \end{aligned}$$

ECM(p)모형에서 차수 p 를 결정하기 위해 HQC기준을 이용하여 구한 값은 표 3.8에 나타나 있으며 HQC의 값을 최소로 하는 $p = 5$ 임을 알 수 있다. 포트맨토 검정결과 모두 자기상관이 존재하지 않는 것으로 나타났다. 또한 잔차들이 이분산성이 존재하지 않는다는 결과를 표 3.9에서 얻을 수 있다. 따라서 ECM(5)모형에 로그변환된 자료를 적합시켜 얻은 추정계수들을 적합시킨 모형은 식 (3.4)와 같이 표현할 수 있다.

표 3.8. ECM모형의 차수결정

lag	MA 0
AR 0	-11.2718
AR 1	-15.1807
AR 2	-15.2704
AR 3	-15.4708
AR 4	-15.7651
AR 5	-15.9214

표 3.9. 계절형 VAR(1)모형에 대한 잔차의 백색잡음 검정

Variable	Durbin Watson	Normality		ARCH	
		Chi-Square	Pr > ChiSq	F Value	Pr > F
일본	1.95496	0.72	0.6994	1.42	0.2366
중국	2.00786	141.69	<.0001	0.89	0.3470
미국	1.87295	5.89	0.0526	2.18	0.1434
필리핀	1.97352	8.42	0.0148	0.56	0.4565

표 3.10. 예측력 비교 결과

목적지	계절형 ARIMA	계절형 VAR	ECM
전체	94528.38	73900.31	55990.22
일본	13695.20	15568.19	27865.81
중국	93135.86	71798.08	47031.83
미국	2699.74	3861.97	10127.47
필리핀	8153.45	7000.59	6621.45

3.4. 모형에 대한 예측력 비교

2005년 7월부터 2006년 6월까지 한국인의 국제관광에 대한 자료를 이용하여 위의 모형들로 예측된 값들을 비교하여 RMSE(Root Mean Square Error)가 작은 우수한 모형을 선택하고자 한다. 비교는 모형별 비교와 국가별 비교로 나누어 계산하였다. 표 3.10의 결과는 모형별, 국가별 RMSE를 구한 값이며 값이 작을수록 예측 정확도가 좋다고 할 수 있다.

4. 결론

본 연구에서는 기존의 계절 ARIMA모형과 계절형 VAR모형, ECM모형의 예측정확도를 실제자료와 비교해 보았다. 또한 각 국가별로 어느 모형이 좋은 예측력을 가지는지도 살펴보았다. 일변량 계절 ARIMA모형은 한국인의 일본과 미국관광에 대한 예측을 하는데 적절하며 ECM모형은 중국과 태국관광에 대한 예측을 하는데 적절한 모형이다. 또한 전체 국가에 대해 모형별로 비교하였을 때는 ECM모형이 계절형 VAR모형보다 우수하고 계절형 VAR모형은 일변량 계절 ARIMA모형보다 우수한 예측력을 가지는 것을 알 수 있었다. 이로 ECM모형이 가장 좋은 모형이라 할 수 있으나 국가별로 비교하였을 때는 어느 한 모형이 다른 모형보다 예측이 우수하다고 할 수 없다. 그렇기 때문에 미래의 예측에서 전체적으로 가장 우수한 모형을 사용하는 것이 아니라 변수별로 우수한 모형을 사용하는 것이 적절하다.

또한 이러한 관광수요예측에 대한 자료들은 모두 계절성을 가지며 계절성을 갖는 자료들은 자료의 변환과 모형 적합을 통하여 계절효과를 예측할 수 있다. 또한 향후의 연구에서는 더 다양한 모형들을 이용하며 최신의 자료들이 더 축적되어 앞으로의 미래에 대한 예측을 할 수 있는 연구가 진행되어야 할 것이다.

참고문헌

- 윤지성, 허남균, 김삼용, 허희영 (2010). 계절형 다변량 시계열 모형을 이용한 국제항공 여객 및 화물 수요예측에 관한 연구, <한국통계학회논문집>, **17**, 473-481.
- 허남균, 정재윤, 김삼용 (2009). 다변량 시계열 모형을 이용한 항공 수요 예측 연구, <응용통계연구>, **22**, 1007-1017.
- Diebold, F. X. and Kilian, L. (2000). Unit root tests are useful for selecting forecasting models, *Journal of Business and Economic Statistics*, **18**, 265-273.
- Engle, R. F. and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation, and testing, *Econometrica*, **50**, 251-76.
- Hylleberg, S., Engle, R. F., Granger, C. W. J. and Yoo, B. S. (1990). Seasonal integration and cointegration, *Journal of Econometrics*, **44**, 215-238.
- Johansen, S. (1988). Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control*, **12**, 231-54.
- Kulendran, N. and King, M. (1997). Forecasting international quarterly tourism flows using error correction and time series models, *International Journal of Forecasting*, **13**, 319-327.
- Kulendran, N. and Wong, K. K. F. (2005). Modelling seasonality in tourism forecasting, *Journal of Travel Research*, **44**, 163-170.
- Li, G., Song, H. and Witt, S. F. (2005). Recent development in econometric modelling and forecasting, *Journal of Travel Research*, **44**, 82-99.
- Lim, C. and McAleer, M. (2002). Time series forecasts of international travel demand for Australia, *Tourism Management*, **23**, 389-396.
- 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.
- Wong, K. K. F., Song, H., Witt, S. F. and Wu, D. C. (2007). Tourism forecasting: To combine or not to combine?, *Tourism Management*, **28**, 1068-1078.

A Study on the Seasonal Effects of the Tourism Demand Forecasting Models

Sahm Kim¹ · Ju-Hyoung Lee²

¹Department of Statistics, Chung-Ang University; ²Department of Statistics, Chung-Ang University

(Received November 2010; accepted November 2010)

Abstract

In this paper, we compared the performance of the several time series models for tourism demand forecasting. We showed that seasonal effects in the data(Japan, China, USA, and Philippines) exist in the tourism data and the forecasting accuracies are compared by the RMSE criterion

Keywords: Seasonal ARIMA model, VAR model, ECM model, tourism demand.

¹Corresponding author: Professor, Department of Statistics, Chung-Ang University, 221 Heukseok-Dong, Dongjack-Gu, Seoul 156-756, Korea. E-mail: sahm@cau.ac.kr