

관 광 학 부 관 광 경 영 학 과

관광수요예측모델에서 타당성 있는 예측 변수에 대한 실증분석

-ARIMAX 전이함수 모델을 중심으로-

V

권 중 익

I. 서 론(10)	V. 결 론
II. 연구의 이론적 배경	참고문헌
III. 모델 정립	
IV. 실증 분석	

V

I. 서 론(15진)

V

1. 문제제기와 연구목적

V

1) 문제의 제기(10진)

인간은 기본적으로 미래를 예측하고자 하는 욕구를 가지고 있으며, 이런 욕구를 충족하고자 다양한 방법론을 발전시켜왔다. 관광학 분야에서도 타 분야에서 연구·발전시킨 다양한 예측 모형들을 관광 수요 예측 등에 활용해 왔으며, 다양한 연구에서 소기의 성과를 거두어 왔다. 하지만 기본적으로 과거의 데이터를 토대로 미래를 예측한다는 전제 조건상 예측할 수 없는 외생 변수들, 특히 정치적 변수가 수요 변수에 영향을 주는 경우 이런 예측방법들이 크게 빗나가는 한계 또한 보여왔다. 가령 2012년 이명박 전 대통령의 전격적인 독도 방문과 일왕 사과 요구로 촉발된 일본의 험한 분위기는 정치적 충격 외의 별다른 외생 변수가 없었음에도 불구하고 방한 수요 예측에 커다란 오차를 가져 왔다. 기본적으로 예측 모델은 이런 설명할 수 없는 외생 변수들을 더미 변수로 주로 처리해 왔다. 이를 더미 변수로 처리할 경우 모형의 적합성이 개선되고 모수가 과소, 혹은 과대 추정되는 문제를 해결할 수 있으나, 어디까지나 모든 일이 일어난 사후에 해당 사건(Event)을 0 또는 1로 처리하는 방법이라 미래 예측에는 한계가 있다는 문제점이 존재했다.

2) 연구의 목적(10진)

이 연구에서는 <국가 우호도>변수가 방한 관광객수에 유의미한 영향을 미치는 변수인지 회귀모형을 토대로 인과성을 확인하고, 기존 연구들에서 그 타당성을 인정받은 독립변수들인 <환

율>, <산업생산지수>를 토대로 적합한 선형회귀 모형과 투입 계열로 <국가 우호도>변수를 삽입한 <ARIMA-X>모형 예측력 비교를 통하여 어느 모형이 더 예측력이 우수한가 살펴보고, <국가 우호도> 변수가 갑작스런 정치적 이벤트 발생시 이를 예측 모형에 사전 반영할 수 있는 유의미한 선행 지표인지 살펴보고자 한다.

II. 연구의 이론적 배경(15진)

V

1. 모델을 구성하는 개념들에 대한 이론적 연구

V

1) 국가 우호도의 개념

(1)국가 이미지의 개념

국가 이미지란 해당 국가를 연상할 때 떠오르는 심상의 집합들을 의미한다. 초기의 국가 이미지에 대한 연구는 대체로 <제품 원산지>라는 개념으로 적용 되었다.(한국문화관광연구원, 2013). 가령, 국가 이미지에 대한 논의를 촉발시킨 Scholer(1965)는 국가 이미지가 제품 선택에 있어 유의미한 영향을 미친다고 보았으며, 원산지 개념의 차원에서 국가 이미지를 제품 이미지의 선행 변수로서 파악하였다. 한편, 이후의 국가 이미지는 단일 차원의 개념이라기 보다는 특정 국가를 구체적인 하위 구성개념으로 이루어진 다차원적 속성이라는 연구가 다수 제시되었다. 가령, 김유경(2011)은 국민, 관광, 수출, 정부, 투자, 이민, 문화 및 역사라는 국가 정체성 구성 요소를 국가 이미지의 구성 요소로 보았으며, 이진용 외(2013)은 국가 이미지를 국가유산, 글로벌리더십, 국민성, 사회시스템 차원에서 측정하였다.

(2)국가 태도에 대한 개념

국가 태도는 국가에 대한 태도를 의미한다. 여기서 태도란 사물, 대상에 대하여 학습에 의해 선유되는 것으로서, 일관성 있게 호의적, 비호의적, 긍정적, 부정적 반응을 보이려는 학습된 선유 경향을 뜻한다. (이연정 외, 2008). 국가 태도는 국가에 대한 태도로서, 주로 신념 혹은 속성에 대한 인식에 가까운 국가 이미지와 달리 국가에 그 자체에 대해 가지는 감정적인 차원에 가깝다(조윤진, 이유리, 2009). 또한 국가 태도는 학습이 이루어지기 위한 사전 정보들을 필요로 하며, 태도의 형성엔 다양한 정보 출처에 의한 습득된 지식을 필요로 한다.(이연정 외, 2008).

(3) 행동의도

행동 의도란 고객이 특정 대상에 대한 태도를 형성한 후에, 특정 행동을 나타내고자 하는 개인의 의지와 신념으로 정의할 수 있다.(정은지, 2015). Zeithaml(1988)은 행동 의도를 “구매(Purchase)로 보았으며, Purchase에 영향을 미치는 변수들로 지각된 가치와 지각된 품질, 내외

부 변수들과 지각된 가격 등을 꼽았다. 행동 의도는 긍정적 행동의도와 부정적 행동 의도로 구분 가능하다.(Zeithaml & Parasuraman, 1996). 즉 관광객이 관광지에서 긍정적인 서비스를 경험한 후 주변에 구전 전파나 재구매 행동을 하는 것을 긍정적 행동의도로 볼 수 있고, 반대로 부정적인 서비스를 경험 후 부정적 구전의 전파 및 재구매 행위의 중단은 부정적 행동의도로 볼 수 있다.

관광 분야에서 행동 의도는 대개 “구전 의도”와 “재방문 의도” 두 개의 개념을 중심으로 많은 측정이 이루어 졌다.(이충기, 김은정, 이상훈 외) 이는 이 두 개의 차원이 관광 분야에서 주로 관심이 많은 분야인 “구매 행동”을 측정할 수 있는 적절한 종속변수로 여겨지기 때문이다.

(4) 국가 이미지와 국가 태도, 행동 의도간 관계

국가 이미지와 국가 태도의 관계는 여러 형태로 연구가 진행 됐는데, 여러 차원의 하위속성(정치적, 사회적, 경제적)으로 이루어진 국가 이미지가 관광지의 정서적, 인지적 이미지에 영향을 미치고, 이들이 재방문 의도에 영향을 끼친다고 보는 연구결과가 다수 존재한다.(이충기, 이상훈, 외). 또한 정은지(2015)는 중국인 관광객을 토대로 가치 지각이 국가 이미지, 태도에 영향을 미친다고 밝히며, 또한 한국에 대한 국가 태도가 구전이나 재구매 의도 같은 행동 의도에 유의미한 영향을 준다는 것을 밝혀냈다. 또한, 조윤진 외(2009)는 한국에 대한 국가 이미지가 호감적 일수록 국가 태도에 대해 좋은 긍정적인 영향을 미쳤고, 이것이 곧 구매의도로 연결된다는 점을 입증하였다. 이연정(2009)은 한국의 국가 이미지를 구성하는 여러 차원들을 사회문화, 음식문화, 비즈니스 측면으로 구분하고, 이들 하위개념으로 구성된 국가 이미지가 국가태도에 유의미한 영향을 미치며, 이 태도가 곧 구매의도에 영향을 미친다는 것을 밝혀내었다. 또한 관광객이 관광지를 결정할 때, 대안 목록에 존재하는 다양한 관광지들에 대한 구체적인 이미지가 아직 형성되어있지 않은 상황에서 불충분한 정보, 경험을 토대로 방문국 결정을 내리기 때문에(임정우, 2013) 국가 이미지가 특정 국가에 대한 관광 행동을 결정할 때 유의한 영향을 미친다고 할 수 있다.(박인규 외, 2012) 한편, Chi & Qu(2008)은 목적지 이미지가 관광 만족, 충성도에 영향을 미친다는 것을 밝혔다. 또한, 강재정(2003)은 재방문 의도가 구체적인 속성인 관광지 이미지에 의해 결정되는 것이 아닌, 중간에서 이를 매개하는 감정적 이미지가 관광 의도를 결정짓는 변수라고 결론 내었다. 또한 Zenithal(1988)은 상품과 관련된 다양한 속성들이 지각된 품질 및 지각된 가치에 영향을 미치며, 이를 통해 고객이 구매 행동을 결정짓는다고 보았다.

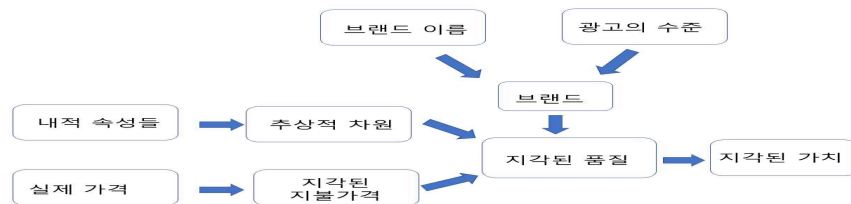


그림 1 - Zenithal(2018)의 지각된 가치 모형

국가 우호도에 대한 이상의 연구결과를 종합해보면, 구체적인 하위 속성들로 이루어진 국가 이미지가 감정적이고 선유적인 감정적 차원인 인 국가 태도에 영향을 미치며, 국가 태도는 구매 행동에 영향을 미친다고 볼 수 있다. 본 연구에서는 궁극적인 구매 행동이라고 할 수 있는 방문관광객 수에 국가 태도가 어떻게 영향을 미치는지 확인해 보고자 한다.

2) 시계열 분석

시계열 분석은 과거의 역사적 데이터를 토대로 미래의 수요를 예측하는 기법으로, 관측된 과거의 자료들을 분석하여 법칙성을 발견, 이를 모형화하고 이 추정된 모형을 토대로 미래 값의 관측하는 분석 기법을 말한다(이우리, 2013). 시계열 분석법은 지수평활법, 추세이동법, ARIMA모형, 확장된 ARIMA모형 등으로 구분할 수 있는데(문화관광연구원, 2006) 이 중 가장 광범위하게 쓰이는 모형은 계절형 ARIMA모형을 비롯한 Box-jenkins(1976)의 ARIMA 모형이다.(김영주,2014)

(1)선형확률과정

확률 과정이란 어떤 특정 시점에서 관측된 변수값 Y_t 가 존재할 때, 이를 전체 표본공간에서 선출된 하나의 표본으로 간주하고, 시간 첨자를 갖는 확률변수들의 집합 $\{Z_t, t=0,1,2,3,\dots\}$ 에서 선출한 표본값 Y_t 를 뽑는 과정을 추정해내는 과정을 말한다.(이우리, 2013). 선형확률과정이란 확률 과정이 선형으로 결합 되어 있는 형태를 나타낸다. 한편, Wold(1932)는 선형확률과정이 서로 독립인 확률과정(백색잡음)들의 결합으로 나타낼 수 있음을 증명하였다. 즉

$$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots + \psi_{t-p} e_{t-p}$$

여기서 e_t 는 백색잡음으로서, ①각각의 항이 서로 독립이고, ②동일한 분포를 가지며, ③0이 아닌 모든 시차 k 에 대하여 공분산 $Cov(e_t, e_{t-k})=0$ 인 항들을 말한다. 즉, Y_t 는 독립적인 여러 백색잡음들의 선형 합으로 나타낼 수 있다.

(2) 시계열의 정상성

그러나 위의 가정이 성립하려면 <시계열이 정상적>이라는 가정을 충족해야 한다. 정상성은 시계열 모형의 성질이 시간의 변화에 따라 변하지 않고 일정하다는 것이다. 즉, 정상성이 만족되려면 시계열 자료가 시간의 변화에 따라 평균, 분산 등에 체계적인 변화가 존재해서는 안된다(김현철, 2005) 정상성엔 강 정상성 조건과 약 정상성 조건이 존재하는 데, 그 중 약 정상성의 조건은 다음과 같다.

①모든 확률변수들의 평균, 즉 $E(Y_t)$ 는 일정하다. 즉 시점 t 에 따라 변하지 않는다.

②모든 확률변수들의 분산은 시점 t에 의존하지 않는다.

③(2-3)Zt와 Zt+k간의 공분산과 상관함수는 시차(lag)에만 의존할 뿐 실제 시점인 t와 t+k에 의존하지 않는다.

(2-1) 시계열의 단위근 검정

시계열 데이터는 정상성을 만족하여야 하며, 정상성을 판단하는 기법은 ADF(Augmented Dickey-Fuller)법이나 PP(Philips-perron)법 등이 개발되어 있다.(이충기, 2008) 정상성을 만족하는 시계열은 상관성이 있는 것으로 알려진 특정 시차를 넘어서는 서로 상관관계를 가져서는 안되며, 완전한 백색잡음이 되어야 한다. 여기서 단위근의 문제가 대두되게 되는데, 단위근이란 AR(p) 모형이 확률보행(Random Walk)이 되도록 하는 모수치를 뜻한다(이우리, 2013). 확률 보행은 각 확률 변수들이 누적적으로 종속 시계열에 영향을 미치는 모습이 한 걸음 한걸음 걸어 가는 비슷하다는 의미에서 붙여진 이름이다. 즉 다음과 같은 자기회귀모형 AR(1) 모형이 존재할 때

$$Z_t = aZ_{t-1} + e_t, \text{단, } e_t \text{는 서로 독립이고 분산이 같은 백색잡음}$$

여기서 만일 a=1이 된다면(즉, 단위근이 된다면) 이 식은 다음과 같이 된다.

$$Z_t = Z_{t-1} + e_t$$

이 식을 Zt-p=Zt-p+et-p 이라는 식을 대입하여 계속 풀어나가게 되면

$$\begin{aligned} &= (Z_{t-2} + e_{t-1}) + e_t \\ &= (Z_{t-3} + e_{t-2} + e_{t-1}) + e_t \\ &\quad \dots \\ &= e_t + \sum_{i=1}^{\infty} e_{t-i} \end{aligned}$$

즉, 이 경우 Zt는 백색잡음들의 선형 합으로 나타나기 때문에, 이 때의 Zt의 분산은 $t\sigma^2$ 가 되어 분산이 특정값으로 수렴되지 않고 무한히 커지게 된다. 이는 다시 말하면 분산이 시차에 따라 값이 커짐을 의미하게 되며, 다시 말하면 분산이 시차에 의존하게 된다.(정효상, 2014) 즉 시차가 지날수록 이는 <분산은 시점 t에 의존하지 않는다>는 시계열의 약정상성 조건을 위배하게 되며, 따라서 추정을 효율적으로 할 수 없게 된다.

(2-2)단위근이 존재할경우 시계열의 정상화

만일 모형이 단위근을 포함하고 있다면 이를 차분하여 변수를 백색화(Whitening)함으로써 해결이 가능하다. 즉

$$Z_t = Z_{t-1} + e_t$$

라는 단위근이 존재하는 모형에서 Zt-1 변수를 좌항으로 이항하면

$$Z_t - Z_{t-1} = e_t$$

로 변환하게 된다. 즉, 변수에서 전 차수의 변수를 뺀 차분변수 ΔZ_t 는 1) 평균이 일정하고 2)분산은 시점 t에 의존하지 않고 일정하다. 단순한 백색잡음(White noise)가 되어 버리며, 백색 잡음은 앞서 살펴본 <약정상성> 조건을 만족하기 때문에 따라서 시계열의 정상조건을 만족할 수 있게 된다.

(2-3)DF(Dickey-Fuller 검증법)

DF(Dickey-Fuller) 검증법은 종속시계열 Yt의 차분값(세모Yt)를 종속변수로, Yt와 지연된 차분변수(세모Yt-1, 세모Yt-2, 세모Yt-3... 세모Yt-p)를 독립변수로 하여 회귀분석을 실시하여, 이 때 나오는 독립변수 Yt의 t값을 이용하여 검증하는 방법이다. 즉

$$\Delta Y_t = Y_t - Y_{t-1}, \text{이라고 할 때,}$$

$$\Delta Y_t = a + \beta Y_{t-1} + e_t$$

이 때, $\beta = 1$ 이라는 귀무가설을 기각하면, $|\beta| < 1$ 이라는 연구가설을 채택할 수 있게 되는데, 이렇게 되면 단위근이 존재하지 않게 됨을 증명할 수 있다. 한편으로, DF 검증은 AR(1)에 대해서만 검증이 가능하다는 약점이 존재하기 때문에 이를 AR(p)과정까지 확장한 Augmented Dickey-Fuller 검증법도 존재한다. ADF 검증법의 식은 다음과 같다.

$$\Delta Y_t = a + \beta t + \theta Y_{t-1} + \delta \Delta Y_{t-2} + \dots + \delta \Delta Y_{t-p-1} + e_t \quad (\text{단, } a \text{는 상수항, } \beta t \text{는 시간적 추세, } \Delta Y_t \text{는 } Y_t - Y_{t-1})$$

(3)자기회귀모형(AutoRegressive Model, AR)

위 가정에서, 연속적으로 진행되는 종속변수의 집단이 존재할 때, 전에 존재하던 변수가 다음 변수에 유의미한 상관관계를 갖고 있다면, 바로 전 변수를 토대로 다음 값을 구해낼 수 있을 것이다. 변수들이 이런 상태에 놓여있는 것 “자기상관”이라고 하며, 자기 상관성을 이용하여 다음 값을 예측하고자 하는 모형이 바로 자기상관모형(AR)이다. 자기상관모형은 다음과 같은 식으로 표현이 가능하다.

$$Y_t = e_t + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} \quad (\text{단, } e_t \text{는 서로 독립인 백색잡음})$$

자기상관모형은 보통 AR(p)로 나타내는데, p는 시차라고 하며 선행변수들의 깊이를 나타내는 수이다. 가령 가장 간단한 AR(1) 모형은 그 깊이가 1차시임을 나타내며, 이라면 바로 전

선행변수만 불러와서 Y_t 를 예측한다. 그 모형은 다음과 같다.

$$Y_t = e_t + a_1 Y_{t-1}$$

즉, 변수 Y_t 는 백색잡음에 바로 전 선행변수를 더한 값으로 나타난다.

AR(2)라면 두 시차 전의 변수를 포함하여 Y_t 를 예측한다. 앞서 살펴본 식은 AR(2) 수준의 모형이다. 즉

$$Y_t = e_t + a_1 Y_{t-1} + a_2 Y_{t-2}$$

(4) MA(Moving Average)모형

MA모형은 앞서 살펴봤던 Wold(1932)의 모형에서 차수가 필요 이상으로 커지지 않도록 특정 백색잡음에서 계수가 결합된 백색잡음을 빼는 과정을 반복하는 것이다. 즉

$$Y_t = e_t - \psi_1 e_{t-1} - \psi_2 e_{t-2} - \dots - \psi_{t-p} e_{t-p} \quad (\text{단, } e_t \text{는 서로 독립인 백색잡음})$$

MA모형은 백색잡음들의 선형결합으로 이루어졌기 때문에 항상 정상성을 만족한다.

(5) ARIMA 모형

ARIMA(Auto Regressive Integrated Moving Average)는 AR모형과 MA모형을 합친 것에 차분 기능을 추가한 모형이다. 기본적으로 확률과정을 이용한 추정법에선 AR이나 MA를 단독으로 사용하는 경우가 적은 편인데, 이는 두 모형을 단독으로 사용할 경우 추정해야 할 모수가 어디까지 늘어날지 알 수 없기 때문이다.(김현철, 2005) 이는 모형을 만들때는 모수를 최대한 적게 해야한다는 모수 절약의 원칙에 위배되는 것이다.

또한 과거 자신의 값들의 합으로 이루어진 AR(p)모형은 갑작스런 충격이 발생해도 이를 누적적으로 모수 추정에 반영하게 된다. 단순히 과거 값의 연속으로만 이루어져 있기 때문이다. 하지만 과거 값의 잔차(오차)로 이루어진 MA(p)모형이 결합될 경우 $Y_t - p$ 권의 AR항에 누적적으로 영향을 끼치는 과거의 충격변수(Shock Variable)의 영향을 최소화 할 수 있어 예측력을 높이는데 도움이 된다.(Enders, Walter, 2004)

이런 이유로 wold(1932)에 의해 두 모형을 합한 ARMA모형이 제시되었다. 그 이후 Box-jenkins(1976)에 의해 기존 ARMA 모형의 방법론이 집대성되고, 또한 시계열의 안정성을 확보하기 위해 차분(differencing) 개념이 도입된 “Box-jenkins의 ARIMA 모형”이 제시되었다.

(5-1) ARIMA(p,q,r) 차수의 결정

앞서 살펴본 단위근 판별 과정에서 단위근이 존재하는 것이 밝혀지면 차분을 실시한다. 차분이 완료된 시계열은 정상 시계열로 간주하며, 정상화된 시계열을 토대로 ARMA(p,q) 모형

의 차수를 결정하게 된다. 차수를 결정할 땐 보통 자기상관함수(Auto Correction Function)이나 부분자기상관함수(Partial Auto Correction Function)을 이용하는데, 자기상관함수는 시차 0에서 k까지의 상관관계를 구하는 함수식이고, 부분자기상관함수는 시차 0에서 k까지의 상관관계를 구하되 그 사이에 있는 시차 변수들을 통제하는 상관관계를 구하는 함수이다. 자기상관함수(ACF)와 부분자기상관함수(PACF)는 다음의 식으로 구한다.

즉, AR(p) 모형은 ACF모형에선 급격히 혹은 지수적으로 감소하며, PACF에선 특정 시차(p 시차) 이후로 0으로 절단된다.

즉, MA(q) 모형은 ACF모형에선 특정 시차(q) 이후로 0으로 절단되며, PACF에선 급격히 혹은 지수적으로 감소한다.

AR(p) 모형이나 MA(q) 모형의 경우 ACF와 PACF가 급격한 감소 혹은 지수함수적으로 감소하는 양태를 보고 차수를 결정하나, 두 모형이 결합한 ARMA 모형의 경우 이런 방법으로 차수를 결정하는데엔 한계가 있다.(이우리, 2013) 이는 다음의 이유로 그렇다.

①AR모형이나 MA모형이 단독으로 쓰일 때와는 달리 ARMA(p,q) 모형은 아무리 시차가 커진다 하더라도 0으로 절단되는 현상이 잘 발견되지 않는다는 점

②특히 MA(q) 모형을 검증할 땐 ACF에서 자기상관이 존재하는 것으로 알려진 특정 시차 q 이후엔 ACF가 무조건 0으로 절단되어야 하기 때문에(즉 특정 시점 이후로는 상관관계가 통계적으로 유의성을 가지면 안되기에) 이를 검증하는 과정에서 모든 시차에 대하여 통계적으로 유의하지 않음을 검증해야 한다. 이 PACF와는 달리 ACF는 오차가 모든 시차에서 누적되어 유의성이 없는데도 귀무가설을 기각하는(즉, 해당 상관계수가 0이 아닌데도 불구하고 0이라고 간주하고 기각하는) 1종 오류의 확률이 지속적으로 커지기 때문이다.(김현철, 2015, pp102).

(6) ARIMA-X모형

ARIMA-X는 전이함수모형이라고도 하며, 단변량 ARIMA모형에 다중회귀분석 모형의 특성을 결합한 예측 방법을 의미한다. 즉 산출변수의 선행변수인 투입변수와 산출변수의 동적 관련성을 분석하여 예측값을 개선시키기 위해 작성하는 모형이다.(이우리, 2013). 구조식은 다음과 같다.

$$Y_t = v(B)X_{t,t} + \eta_t$$

이 모형은 산출계열 y_t 가 충격반응함수 $v(B)$ 를 계수로 가지는 투입계열 x_t 와 선행필터 $\psi(B)$ 를 계수로 가지는 잡음계열 η_t 의 선형합으로 이루어진 모형이다. x_t 의 계수인 충격반응함수 $v(B)$ 는 다음과 같은 식으로 정리할 수 있다.

$$v(B) = \frac{\omega_b(B)^b}{\delta_b(B)^r} B^b$$

이 때, b는 투입계열 x_t 가 산출계열 y_t 와 최초의 상관관계를 나타내는 시차인 지연 모수를 의미한다. 즉, $\text{cov}(y_t, x_t)$ 를 구할 때, y_t 가 x_{t-b} 사이에서 최초의 상관관계가 b시차에서 나타날 경우 해당 시차를 모수로 취한다. 같은 방식으로 δ_b 와 ω_b 의 모수도 결정하는데, 이 때 산출계열 y_t 와

투입계열 x_t 의 각 시차별 상관계수를 보여주는 <교차상관함수(Cross-Correlation-Function)>를 보고 해당 모수를 결정하게 된다.

한편, 잡음 η_t 는 다음의 과정을 따르는 ARMA모형으로 표현될 수 있다.

$$(1 - \theta(B))\eta_t = (1 - \phi(B))e_t$$

η_t 는 투입계열 x_t 와 산출계열 y_t 의 관계에서 생성되는 잔차(error)로, 산출계열 y_t 를 x_t 가 설명하고 남은(여과된)부분으로 볼 수 있다. 모형이 안정성을 갖기 위해서는 최종적으로 여과된 η_t 가 완전한 백색잡음이 되어야 하는데, 이를 백색잡음으로 만들기 위하여 잡음 η_t 또한 ARMA과정으로 적합하여 모형에 투입한다. 즉

$$\frac{(1 - \theta(B))}{(1 - \phi(B))} \eta_t = e_t$$

η_t 를 한번 더 여과하여 백색잡음인 e_t 로 만들면, 잔차의 i.i.d 조건을 충족할 수 있게 된다.

3) 관광수요 영향 변수의 선정

그간 관광수요를 예측하기 위한 다양한 연구들이 진행되었으며, 타당성을 가지는 변수를 선정하기 위한 다양한 연구들도 함께 진행되어왔다. 이충기(2008)는 환율, 소득, 물가, 이벤트 더미변수(정치적 불안, 사회갈등, 테러리즘, 석유 파동)를 함수식에 포함하면서 이를 VAR 기법으로 분석한 결과 환율과 산업생산지수, 한일간 정치갈등, 국내 사회갈등 변수들이 일본 방한관광객 수요에 영향을 미친다고 보았다. 또한 김성태(2014)는 패널회귀모형기법을 사용하여 환율, 소득이 해외 관광객의 국내 수요에 영향을 미친다고 결론 내었다. 또한 이강욱(2017)은 원/엔의 상승과 전염병의 발생, 정치 및 외교갈등, 자연재난 변수가 방한관광객의 수요에 유의성을 가지는 타당성 있는 변수라고 보았다.

III. 연구의 이론적 배경(15진)

이상의 선행 연구를 바탕으로 다음과 같은 모델을 정립하고자 한다.

<연구모형 1>

H11 : 국가 우호도의 변화는 방한 관광객수의 변화에 유의미한 영향을 미칠 것이다.



구체적으로, ①국가 태도 변수를 포함한 선형 회귀모형과, ②ARIMA-X 전이함수 모형을 모두 적합을 시도해보고, 국가 태도 변수가 관광객 예측 모델을 수립할 때 타당성이 있는 변수인지 확인해 보고자 한다.

만일 국가 우호도가 관광객이 관광지를 결정하는 데 유의한 변수라면 국가 우호도의 변화에 따라 방한 관광객수 또한 인과적으로 늘어나거나 줄어들어야 한다. 따라서 양 변수간 회귀분석을 실시하여 계수가 유의함을 갖는지 검증을 실시해보고자 한다.

H12 : <환율>, <산업생산지수>를 독립변수로 적합한 선형회귀모형은 타당성을 가질 것이다.

다양한 연구에서 그 타당성을 인정받은 독립변수들인 <환율>과 <산업생산지수> 변수를 토대로 적합한 선형회귀모형은 ①각 변수들의 계수가 타당성을 가지는지 여부 ②전체 모형이 타당성을 가지는지 여부를 토대로 그 타당성을 본 연구에서 재검증 하고자 한다.

<연구모형 2>

H21 : 일본인 방한관광객수를 산출계열로, <국가 우호도> 변수를 투입 계열로 적합한 ARIMA_X(0,0,0)(P,Q,R,S) 모형은 타당성을 가질 것이다.

H22 : 일본인 방한관광객수를 산출계열로, <환율> 변수를 투입계열로 적합한 ARIMA-X(0,0,0)(P,Q,R,S) 모형은 타당성을 가질 것이다.

H21의 경우 기존 연구에서 이미 여러차례 그 타당성이 증명된 변수들인 “환율”변수와 “산업생산지수” 변수 중 대부분의 연구에서 타당성을 인정받은 “환율”변수만 뽑아, 전이함수 모형에서도 “환율”이 유의한 변수인지 확인해 보고자 한다. 즉, 적합한 전이함수의 교차상관분석함수(CCF)와 자기상관함수(ACF) 및 편자기상관함수(PACF)의 양태를 보고 이 모형이 적합한지 판단하고자 한다.

또한, 국가 우호도 변수를 투입계열로 마찬가지로 전이함수 모형의 적합을 시도해서, 국가 우호도 변수가 전이함수 모형에서 타당한 변수인지 또한 확인을 해보고자 한다.

IV. 실증 분석

1. 분석에 사용된 데이터

1) 분석에 사용된 데이터들은 다음과 같다.

	기간	출처	비고
방한관광객수	1984~2017 연간 데이터	한국관광공사	
환율	1984~2017 연간 데이터	한국은행	엔화의 대원화 환율
산업활동지수	1984~2017 연간 데이터	일본 경제산업성	계절차분된 자료
국가 우호도	1984~2017 연간 데이터	일본 내각부	“친근감을 느낀다”(5점) + “친근감을 느끼는 편(4점) 이용

2. <연구모형 1>의 검증

1) 총론

국가 우호도와 방한 관광객수 변수가 유의성을 갖는지 회귀분석을 실시하고자 한다. 프로그래밍 언어인 Python을 이용했으며, Python에서 통계 및 수리 분석 툴로 널리 이용되고 있는 <statsmodels> 라이브러리의 statsmodels.OLS 패키지를 임포트하여 회귀분석을 실시하였다. 연구가설에 더하여, 차후 적합할 전이함수와의 예측력 비교를 위하여 1984년부터 2016년까지의 데이터를 Train Set로 삼아 회귀모형을 적합하고, 2017년의 데이터를 Test Set로 삼아 전이함수모형과 회귀모형의 예측력 비교 연구를 추가적으로 진행하기로 한다.

	year	PAX	IAA	Fri_ab	Fri_almost	Cur
23	2011	3289051	99.28	20.3	41.9	1391.31
24	2012	3518792	100.75	9.7	29.4	1413.14
25	2013	2747750	101.98	8.4	32.3	1123.41
26	2014	2280434	102.18	6.9	24.6	996.19
27	2015	1837782	102.48	6.0	27.0	934.56
28	2016	2297893	102.98	9.1	29.0	1068.17
29	2017	2311447	104.79	8.3	29.3	1008.52

모형 적합에 앞서, 데이터의 구조를 먼저 살펴보았다. 확인 결과, PAX변수는 100만 단위인 반면 다른 변수들은 1~1000자리대 변수로서, 독립변수와 종속변수의 단위가 크게 차이날뿐더러, 독립변수들 사이에서도 단위가 크게 차이나는 것으로 나타났다. 이를 정규화하기 위하여, 모든 변수들에 대하여 log변환을 취해주었다.

	PAX	IAA	Fri_ab	Fri_almost	Cur
2011	15.006110	4.597944	3.010621	3.735286	7.238001
2012	15.073628	4.612642	2.272126	3.380995	7.253569
2013	14.826293	4.624777	2.128232	3.475067	7.024124
2014	14.639876	4.626736	1.931521	3.202746	6.903938
2015	14.424070	4.629668	1.791759	3.295837	6.840076
2016	14.647503	4.634535	2.208274	3.367296	6.973702
2017	14.653384	4.651958	2.116256	3.377588	6.916239

로그변환을 취한 결과, 모든 변수들의 자리수가 1~2단위로 정규화 되었다.

2) 연구가설 H11의 검증

OLS Regression Results						
Dep. Variable:	PAX	R-squared:	0.982			
Model:	OLS	Adj. R-squared:	0.982			
Method:	Least Squares	F-statistic:	1561.			
Date:	Wed, 14 Nov 2018	Prob (F-statistic):	4.21e-26			
Time:	12:19:13	Log-Likelihood:	-60.212			
No. Observations:	29	AIC:	122.4			
Df Residuals:	28	BIC:	123.8			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Fri_ab	6.0364	0.153	39.505	0.000	5.723	6.349
Omnibus:	0.940	Durbin-Watson:				0.730
Prob(Omnibus):	0.625	Jarque-Bera (JB):				0.840
Skew:	-0.158	Prob(JB):				0.657
Kurtosis:	2.229	Cond. No.				1.00

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

제공되는 변수 중 방한관광객수의 증감에 영향을 미칠것으로 기대되는 “친근감을 느낀다(5점) 변수를 모형 적합을 실시한 결과, t검증치는 39.505, 이 때의 p-value는 0.000으로 회귀계수 6.0364가 사실은 0일 귀무가설이 참일 확률이 0%에 수렴하는 것으로 나타났다. 따라서, 회귀모형에서 <친근감을 느낀다>의 증감은 <일본인 방한관광객수>와 공변의 관계에 놓여 있다고 추정할 수 있다.

OLS Regression Results					
Dep. Variable:	PAX	R-squared:	0.998		
Model:	OLS	Adj. R-squared:	0.998		
Method:	Least Squares	F-statistic:	1.479e+04		
Date:	Wed, 14 Nov 2018	Prob (F-statistic):	1.11e-39		
Time:	14:31:07	Log Likelihood:	-27.835		
No. Observations:	29	AIC:	57.67		
Df Residuals:	28	BIC:	59.04		
Df Model:	1				
Covariance Type:	nonrobust				
	coef	std err	t	P> t	[0.025 0.975]
Fr1almost	4.0780	0.034	121.808	0.000	4.009 4.147
Omnibus:	3.622		Durbin-Watson:	0.860	
Prob(Omnibus):	0.163		Jarque-Bera (JB):	2.751	
Skew:	0.754		Prob(JB):	0.253	
Kurtosis:	3.007		Cond. No.	1.00	

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

"친근감을 느끼는 편(4점)" 변수도 마찬가지로 회귀모형을 통해 그 타당성을 추정했다. 그 결과 "친근감을 느낀다(5점)" 변수와 마찬가지로 타당성을 가지는 것으로 추정이 가능하다. t검증치는 121로 회귀계수가 0이라는 귀무가설이 참일 확률이 0%에 수렴하는 것으로 나타났다.

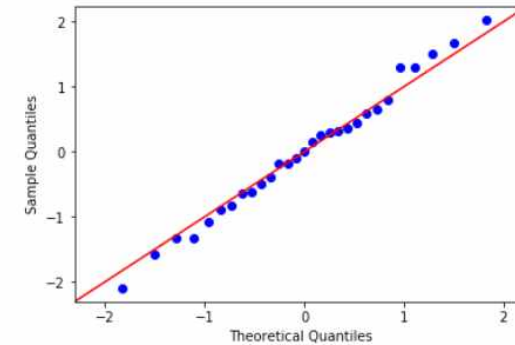
3) 연구가설 H12의 검증

기존에 타당성을 인정받은 변수인 "환율"과 "산업생산지수" 변수를 투입하여 회귀모형의 타당성을 검증하고자 한다.

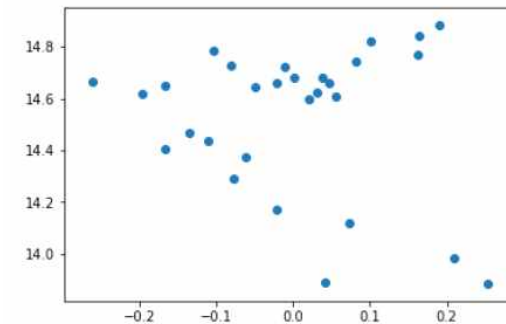
OLS Regression Results					
Dep. Variable:	PAX	R-squared:	1.000		
Model:	OLS	Adj. R-squared:	1.000		
Method:	Least Squares	F-statistic:	1.829e+05		
Date:	Wed, 14 Nov 2018	Prob (F-statistic):	1.65e-56		
Time:	14:59:12	Log Likelihood:	19.187		
No. Observations:	29	AIC:	-34.37		
Df Residuals:	27	BIC:	-31.64		
Df Model:	2				
Covariance Type:	nonrobust				
	coef	std err	t	P> t	[0.025 0.975]
IAA	2.0905	0.136	15.398	0.000	1.812 2.369
Cur	0.7226	0.092	7.873	0.000	0.534 0.911
Omnibus:	0.146		Durbin-Watson:	0.933	
Prob(Omnibus):	0.930		Jarque-Bera (JB):	0.363	
Skew:	0.034		Prob(JB):	0.834	
Kurtosis:	2.456		Cond. No.	56.0	

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

지수(IAA)에 대해서는 t검증치가 15.398, 환율(Cur)의 경우 7.873으로 각 계수들이 0이라는 귀무가설을 채택할 확률이 0에 수렴하는 것으로 나타났다.



회귀식의 잔차들이 정규분포를 따르는지 확인하기 위해 잔차들을 대상으로 정규분포와의 유사성을 확인할 수 있는 Q-Q-plot을 그려보았다. 그 결과 각 관측치의 분위수가 이상적인 정규분포의 분위수와 45도 각도에서 대부분 만나는 것으로 나타났다.



회귀식들의 잔차와 예측값들이 무작위적으로 분포하는지 확인하기 위하여 잔차들과 예측값의 산점도를 그려보았다. 그 결과 관측치와 잔차 사이에는 두 가지의 경향성이 존재하는 것으로 보인다. 가로로 길게 늘어져있는 상단의 산점도는 해당 관측치들이 예측치-잔차간에 상관성이 존재하지 않는다는 것을 의미한다. 하지만 좌상단에서 우하단으로 대각선으로 그 어지는 산점도는 이 케이스들의 경우 잔차와 예측치 간 음의 상관관계가 존재함을 보여준다.

한편, 1986년부터 2016년까지의 데이터를 토대로 적합한 회귀모형으로 2017년까지의 테스트 데이터를 토대로 예측값을 생산하였다.

	예측값	실값	오차
2017	2476863	2311447	7%

3. <ARIMA-X 모형>의 수립

1)총론

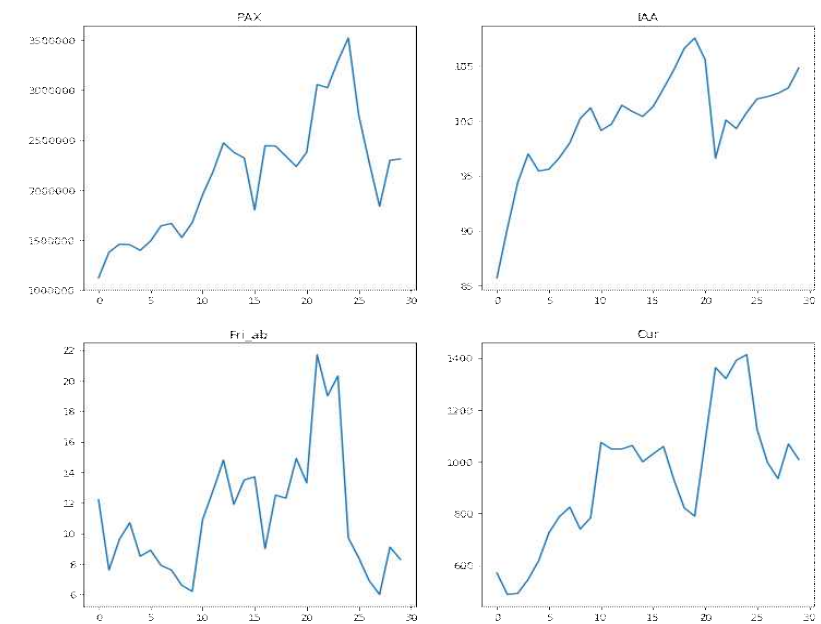
앞서 살펴본 선형회귀모형들이 모두 잘 적합된 것으로 보이지만, 사실 모두 치명적인 오류의 가능성이 상존한다. 관측치가 29개, 변수가 2개인 선형회귀모형의 Dirbin-watson 검증의 기각역은 1.07~1.339로, H12모형의 Dirbin_watson 검증치인 0.933은 이 기각역을 하회하여 잔차들의 자기상관성이 존재하는 것으로 나타났다. 일반적으로, 잔차들의 자기상관성이 존재하는 경우는 t검증치를 გადა 추정하여 귀무가설을 잘못 기각하는 2종 오류 가능성이 커진다(이충기, 2008).

선형회귀모형에 자기상관성이 존재하는 경우, 차선택으로 시계열 분석을 고려해볼 수 있다.(이충기,2008). 관계의 타당성을 좀 더 정밀하게 추정하기 위해, “방한 관광객수”를 출력시계열로 삼고, 환율과 국가우호도를 각각 투입 시계열로 하여 ARIMA-X 전이함수 모형을 수립하고자 한다.

2) 단위근 존재 여부의 검증

본격적으로 시계열 분석에 들어가기에 앞서, 해당 시계열에 단위근이 존재하는지 여부를 확인하였다. 시계열에 단위근이 존재할 경우 시계열의 공분산은 시간이 아닌 시차에만 의존한다는 이차정상성 조건을 위배하기 때문에 효율적인 추정이 불가능해진다. 따라서 단위근이 존재할 경우 현재 시점의 데이터에서 전 시점의 데이터를 빼서 시계열을 whitenoise로 만드는 과정이 반드시 필요하다. 이를 차분(differentiate)라고 한다.

시계열이 ① Y_t 의 평균이 일정하고 ② Y_t 의 분산이 특정 상수이며 ③ Y_t 와 Y_{t+k} 의 공분산이 시점 t와 무관하게 모두 같을 시 충족되는 이차정상성을 만족하기 위해서는 시계열에 결정적이나 확률적인 추세가 존재하면 안된다. 우선, 이를 시각적으로 확인하기 위하여 각 시계열 데이터의 추세선을 확인하였다.



방한관광객수(PAX)의 경우 평균이 지속적으로 상승하고, 분산도 시간이 갈수록 커지는 것으로 보인다. 또한, 약 10개월을 시차로 계절성도 있는 것으로 보여지나, 시계열 자체의 데이터 포인트가 29개에 불과한 관계로 그 추세를 뚜렷히 확인하기에는 다소 어렵다.

IAA도 마찬가지로 평균이 변하는 추세변동인 존재하는 것으로 보이며, cur의 경우도 평균이 변화하는 추세변동이 존재하는 것으로 보인다. 이들은 대부분 경제와 관련된 변수들로, 경제적인 변동과 밀접한 연관성이 존재하는 것으로 보인다.

한편, 국가 우호도를 나타내는 Fri_ab의 경우 분산이 시간에 따라 변화하기는 하나 평균은 대체적으로 일정한 것으로 보인다. 이는 Fri_ab는 태도를 나타는 정성적인 변수로서 경제적 변동과 큰 상관이 존재하지 않기 때문으로 보인다.

실제로, 통계적으로도 이들 시계열이 불안정한 추이를 보이는지 확인하기 위하여 단위근 검정을 실시하였다. 시계열에 단위근이 존재하는지 여부를 확인하기 위하여 Augmented Dickey-Fuller 검정법을 사용하였다. 해당 검정법을 사용하기 위해 E-views 10 Student Lite 버전을 사용하였다.

Null Hypothesis: PAX has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=7)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.966144	0.2992
Test critical values:		
1% level	-3.679322	
5% level	-2.967767	
10% level	-2.622989	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PAX)
Method: Least Squares
Date: 11/05/18 Time: 23:55
Sample (adjusted): 1989 2017
Included observations: 29 after adjustments

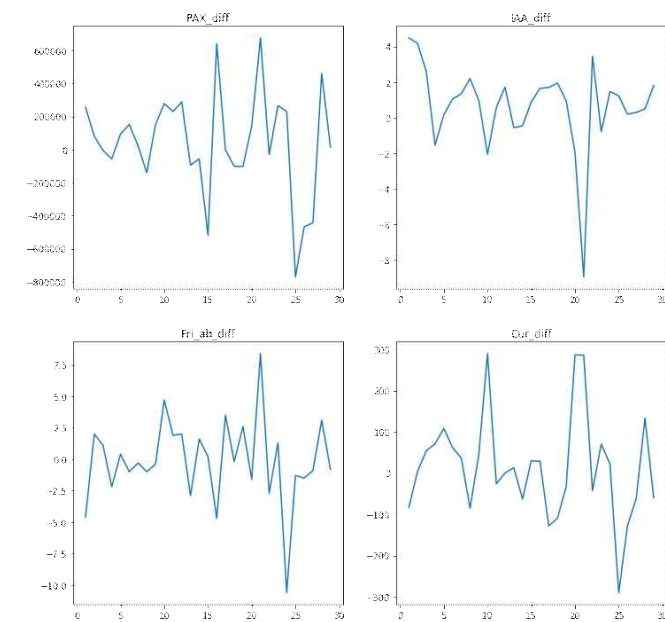
Variable	Coefficient	Std. Error	t-Statistic	Prob.
PAX(-1)	-0.183872	0.093519	-1.966144	0.0596
C	432903.2	207179.8	2.089505	0.0462
R-squared	0.125243	Mean dependent var	40941.31	
Adjusted R-squared	0.092845	S.D. dependent var	318879.0	
S.E. of regression	303715.3	Akaike info criterion	28.15204	
Sum squared resid	2.49E+12	Schwarz criterion	28.24634	
Log likelihood	-406.2046	Hannan-Quinn criter.	28.18157	
F-statistic	3.865721	Durbin-Watson stat	1.798803	

우선 산출계열로 활용할 방한관광객수(PAX) 변수의 단위근 존재 여부를 확인하였다. 확인 결과, $\Delta Y_t = aY_{t-1} + \phi \Delta Y_{t-1} + \phi \Delta Y_{t-2} \dots$ 에서 우변의 맨 처음항 aY_{t-1} 에서 a 가 1이라는 영가설을 채택할 확률이 0.2992로 나타났다. 이는 다시 말하면 a 는 1일 확률이 29% 존재한다는 뜻이며, 단위근이 존재할 가능성이 충분하다는 것을 의미한다. 이를 토대로 방한관광객수 변수는 차분이 필요하다는 것을 확인할 수 있다.

기타 투입계열로 활용할 예측변수들도 마찬가지로 단위근 검증을 실시하였으며, 그 결과는 다음과 같다.

	t-검증치	채택 확률	적정 차분	비고
PAX	-1.97	0.2992	I(1)	p = 0.005
IAA	-3.43	0.0178	I(0)	단위근 존재X
CUR	-2.43	0.1425	I(1)	p = 0.0042
FRI_ab	-2.31	0.1732	I(1)	p = 0.0000
FRI_almost	-2.52	0.1206	I(1)	p = 0.0000

확인 결과, 산업생산지수(IAA)를 제외한 대부분의 변수에서 단위근이 존재했다. 이들 변수들에 1차 차분을 실시한 결과, 대부분 단위근이 존재하지 않는 것으로 나타나서 1차분시 안정화 될 수 있음을 확인하였다.



차분을 실시한 결과 대부분의 시계열에서 추세가 제거되었고, 분산도 시간에 따라 일정한 것으로 보인다. 시계열이 정상화 되었기 때문에 이제 이 시계열을 토대로 전이함수 모형을 수립하기로 한다.

3)모형의 적합

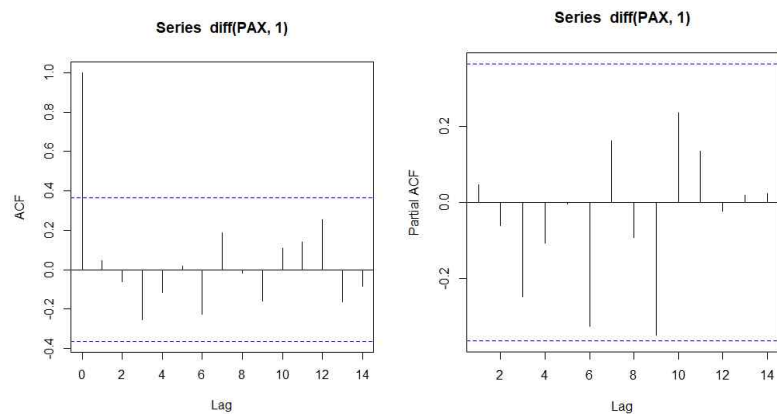
본 모형을 수립하는 가장 큰 이유는 FRI_ab 변수로 PAX_diff를 설명할 수 있으나 없나를 보는 것이다. 따라서, PAX_diff를 산출 계열로, FRI_ab를 투입계열로 놓고 전이함수 모형을 수립하였을 때 ①계수가 유의성이 있고 ②잔차들이 무작위적인 백색잡음을 보인다면 잘 적합한 모형으로 판단할 수 있고, 따라서 FRI_ab가 PAX_diff를 잘 설명하는 변수라고 할 수 있다.

전이함수 모형을 적합하기 위하여, 우선 산출계열이 되는 PAX_diff의 ARMA 모형을 적합하기로 한다. 산출 계열의 ARMA모형을 적합하는 이유는 투입계열의 시계열 구조를 파악하기 위함이다. 산출계열 Y_t 의 시계열 구조(ARMA)를 적합하여 추정된 모수 θ 를 구하는데, 산출계열이 투입계열의 좋은 설명변수라고 가정한다면, 투입계열을 산출계열은 비슷한 시계열 구조를 가진다고 할 수 있다. 따라서 산출계열의 모수로 투입계열을 “여과”할 경우, 투입계열은 무작위적인 백색잡음만을 잔차로 갖게 되며 여과된 잔차들을 토대로 교차상관계수를

구할 경우, 교차상관함수의 정의에 따라 모수를 효율적으로 추정할 수 있게 된다.

백색잡음이 된 투입계열의 잔차(=Xt의 백색잡음)과 마찬가지로 백색잡음이 된 산출계열의 잔차(=Yt의 백색잡음)의 상관계수(교차상관계수,CCF)를 구할 경우, 산출계열과 투입계열의 공분산 Pxy를 이용해 충격반응가중함수 v(B)의 각 차수를 잠정 추정할 수 있다. 구체적으로는 이들 백색잡음들의 CCF가 어느 지점에서 유의성을 갖고, 어느 지점에서 절단되는지를 판단하여 충격반응가중함수 v(B)를 추정한다.

PAX_diff를 ARMA모형으로 적합하기 위하여 통계 프로그래밍 언어인 <R>의 “forecast”라이브러리를 사용하였다.



차분된 PAX 변수를 토대로 자기상관계수(ACF) 그래프와 편자기상관계수(PACF) 그래프를 그렸다.

ACF그래프에서는자기상관계수에서 자기 자신과의 상관관계를 의미하는 0차시를 제외하고는 그 어떤 유의미한 상관관계도 없는 것으로 나타나, MA 모형의 차수의 경우 0차시로 잠정 추정이 가능한 것으로 보인다. 또한, ACF에서 상관계수들이 지수함수적으로 감소하는 양태를 보이지 않는 것으로 미루어보아, 이 모형의 확률변수간 상관성이 존재하는지 확인할 수 없다.

PACF그래프에서는 3,6,9시차에서 특 튀어나온 형태를 확인할 수 있다. 비록 유의미한 상관관계가 존재한다고 여겨지는 채택수준인 Pr = 0.4를 넘는 상관성은 보이지 않아 단순 백색잡음으로 취급할 수도 있으나, 이를 유의미한 상관관계로 본다면 AR모형에서 계절성을 제

외하는 계절 차분이 추가로 필요할 수도 있다는 점을 시사한다.

잠정적으로 추정컨대 PAX 변수의 시계열 구조는 ARIMA(0,1,0)으로 추정되어 백색잡음으로 보이나, 연구의 진행을 위하여 임의적으로 이 구조를 ARIMA(1,1,0)으로 간주하여 모형을 적합하도록 한다.

```
a <- arima(PAX, order=c(1,1,0))
tsdisplay(residuals(a), lag.max=15, main='PAX ARIMA(1,1,0)')
Box.test(resid(a),type="Ljung",lag=20,fitdf=1)
> Box.test(resid(a),type="Ljung",lag=20,fitdf=1)
```

Box-Ljung test

data: resid(a)
X-squared = 15.706, df = 19, p-value = 0.6768

Box-Ljung test가 검증하는 가설은 다음과 같다.

H0 : 각 확률변수들의 자기상관은 존재하지 않는다(자기상관이 0이다)

H1 : 각 확률변수들의 자기상관성이 존재한다(자기상관이 0이 아니다)

Box-ljung 검증법은 확률변수들이 상관성을 가지는지 여부를 판단하는 방법으로, chi-squared 분포를 사용한다. 이를 잔차들이 서로 독립성을 가지는지 확인하기 위하여 시계열의 자기상관성이 0이라는 귀무가설을 토대로 가설을 검증하는 검증법인 box-ljung 검증법을 잔차들의 상관성을 보기 위해 적용한 결과, 자기상관성이 0이라는 귀무가설을 채택할 확률이 0.6768로 귀무가설을 기각하지 못하는 것으로 나타났다. 이는 다시 정리하면 잔차들의 상관성이 0일 확률이 높다는 이야기로, ARIMA(1,1,0)으로 적합한 후 여과된 잔차들이 서로 독립성을 유지하고 있다는 점을 유추할 수 있다. 따라서, 본 모형은 잘 적합했다고 판단할 수 있다.

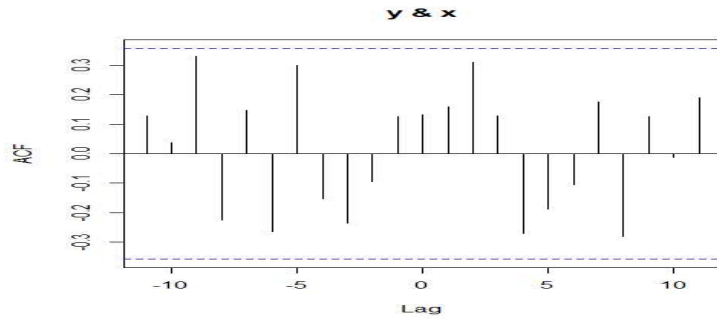
ARIMA(1,1,0)으로 PAX_diff 변수를 적합한 결과, 추정된 산출계열의 ARIMA식은 다음과

$$\nabla Y_t = 0.0316 \nabla Y_{t-1} + \epsilon_Y$$

이를 투입계열인 Fri_ab에 적용하면 다음과 같은 구조로 추정할 수 있다.

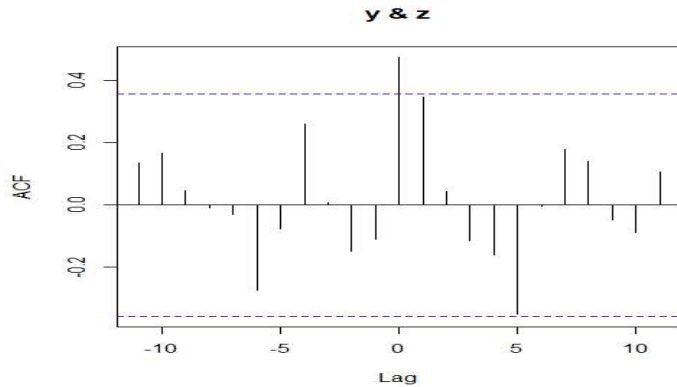
$$\nabla X_t = 0.0316 \nabla X_{t-1} + \epsilon_X$$

이를 토대로 자기상관성을 제거하고 평균0, 분산1의 백색잡음으로 여과한 산출계열, 투입계열의 잔차인 ϵ_Y 와 ϵ_X 의 교차상관계수를 토대로 충격반응함수의 잠재적인 차수를 결정하기로 한다.



CCF 교차상관그래프를 그린 결과, 상관관계가 있다고 볼 수 있는 상관계수 0.3을 넘는 시차가 존재하지 않는 것으로 나타났다. 이는 두 계열이 서로 종속적인 관계가 아니라 독립적인 관계임을 의미하며, 따라서 방한관광객과 국가 우호도 사이에는 인과성을 가진 관계가 존재하지 않으며, 전이함수모형을 사용할 이유가 없다.(시계열 분석 강의 11장)

기준에 타당성을 인정받은 다른 변수와의 관계를 보기 위해, PAX_diff와 IAA, Cur의 관계가 전이함수 모형을 수립하기에 충분한 상관관계가 존재하는지 확인해 보았다.



PAX_diff와 Cur의 교차상관그래프를 그려본 결과, 0차시에서 최초의 유의한 상관관계가 나타나 지연모수는 0차시로 추정되었으며, 1차시와 5차시에 유의미한 상관관계가 나타난 것으로 보인다. 이를 토대로 잠정 추정한 충격가중함수의 구조는 다음과 같다.

$$v(B) = \frac{\omega_0(B)}{1 - \delta_1(B)} B^0$$

$$Y_t = \frac{\omega_0(B)}{1 - \delta_1(B)} B^0 X_t = \omega_0 X_t - \delta_1 Y_{t-1}$$

혹은, 1차시의 상관계수를 지수함수적으로 줄어드는 상관계수가 아닌, X_t 가 가장 큰 상관관계를 가지는 0이 아닌 최초의 지점으로 본다면 다음의 구조식도 추정이 가능하다.

$$v(B) = \frac{\omega_1(B)}{\delta_0(B)} B^0$$

$$Y_t = \frac{\omega_0(B)}{1 - \delta_1(B)} B^0 X_t = \omega_1 X_{t-1}$$

```
Call:
lm(formula = pax_diff ~ 0 + cur_diff + pax_diffd1, data = su2)

Residuals:
    Min       1Q   Median       3Q      Max
-0.269760 -0.060268  0.005196  0.053820  0.303797

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
cur_diff      0.63116    0.16681   3.784  0.00082 ***
pax_diffd1    0.06838    0.15176   0.451  0.65603
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1134 on 26 degrees of freedom
Multiple R-squared:  0.3557,    Adjusted R-squared:  0.3062
F-statistic: 7.178 on 2 and 26 DF,  p-value: 0.003294
```

첫 번째 잠정 추정모형의 경우, $\delta_1 Y_{t-1}$ 인 Pax_diffd1 변수가 유의성을 가지지 않는 것으로 나타났다. 모형이 불안정하므로, 1안은 폐기하고 2안으로 모형을 다시 적합하였다.

```
Call:
lm(formula = pax_diff ~ 0 + cur_diffd1, data = su1)

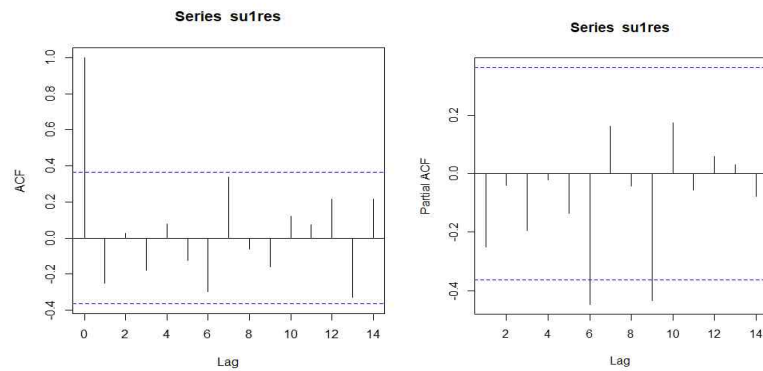
Residuals:
    Min       1Q   Median       3Q      Max
-0.253747 -0.057231 -0.000496  0.103384  0.291968

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
cur_diffd1    0.4086    0.1828   2.236  0.0338 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1274 on 27 degrees of freedom
Multiple R-squared:  0.1562,    Adjusted R-squared:  0.125
F-statistic: 4.999 on 1 and 27 DF,  p-value: 0.03383
```

두 번째 잠정 추정모형의 경우, $\omega_1 X_{t-1}$ 인 cur_diffd1 변수가 유의성을 가지며, F통계량에 기반한 p-value도 0.05 미만으로 안정된 모형으로 볼 수 있다. 따라서, 두 번째 모형을 토대로 모형을 계속 적합하기로 한다.

잠정 추정된 구조식을 토대로 임시 적합을 실시한 후, 여과된 잡음 η_t 를 구한다. 이 때 η_t 는 가적화된 모형에서 도출된 잔차들을 의미한다. 잔차들의 ARMA구조를 파악하여 평균은 0이고 표준편차는 1인 백색잡음으로 최종 여과시키는 것이 모델 적합의 마지막 단계가 된다.



잠재 적합한 모델의 잔차를 토대로 ARIMA 모델을 적합하기 위하여 ACF와 PACF 그래프를 그려보았다. 그 결과, PACF의 경우 6차시와 8차시에서 유의미한 관계가 존재하는 것으로 나타났다. ACF를 지수함수적으로 줄어드는 경향으로 볼 수 있다면, 잡음 η_t 는 대략 ARMA(6,0) 과정을 따르는 확률모형이라고 볼 수 있다. 적합한 모형은 다음과 같다.

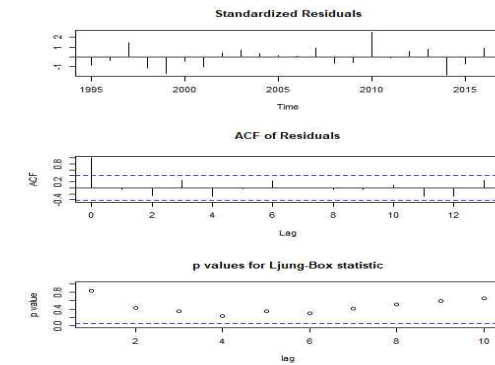
$$Y_t = \frac{\omega_0(B)}{1 - \delta_1(B)} B^0 X_t + (\theta B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4 - \theta_5 B^5 - \theta_6 B^6) \eta_t + e$$

이를 투입계열 X_t 의 잔차와 잡음 η_t 의 잔차로 정리하기 위하여, 앞서 진행하였던 투입계열 사전 백색화 $(1-0.0316B)X_t$ 와, $1 - \theta B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4 - \theta_5 B^5 - \theta_6 B^6 \eta_t$ 를 이용하여 다시 적으면

$$\frac{(1-0.0316)}{(1-\theta B - \theta_2 B^2 - \dots - \theta_6 B^6)} Y_t = \frac{\omega_0(B)(1-0.0316)}{1 - \delta_1(B)(1-\theta B - \theta_2 B^2 - \dots - \theta_6 B^6)} B^0 X_t + \frac{(1-0.0316)}{1-\theta B - \theta_2 B^2 - \dots - \theta_6 B^6} e_t$$

$(1-0.0316B)X_t$ 는 X_t 의 백색잡음이 되고, $1 - \theta B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4 - \theta_5 B^5 - \theta_6 B^6 \eta_t$ 는 η_t 의 백색잡음이 되므로, Y_t 는 서로 독립적인 두 백색잡음(X_t , η_t)의 선형합으로 간단해진다. 따라서, 이를 이용하여 Y_t 의 예측치를 생산할 수 있게 된다.

4) 모형의 검증



적합한 모형의 잔차검증을 실시하였다. 잔차들의 Ljung_Box 검증결과 P-value는 대부분 기각수준인 0.05 이상으로 나타나서, 잔차들이 상호 독립이라는 귀무가설을 기각하지 못하는 것으로 나타났다. 잔차들의 ACF를 살펴보면, 모든 차수에서 상관계수가 0.3 이하로 나타나 잔차들이 서로 독립인 것으로 나타났다. 이를 토대로, 잔차들의 상호 독립성이 만족되며, 모형은 대체로 잘 적합한 것으로 판단할 수 있다.

5) 모형의 예측

2017년 방한관광객수의 예측을 위해 1차시 분의 예측치가 필요하다. 따라서 Cur_diff와 적합한 모형의 잔차들인 su1res의 1차시분의 예측치를 생산하였다.

```
Time Series:
Start = 1
End = 1
Frequency = 1
cur_diff_fore  su1resd1  su1resd2  su1resd3  su1resd4  su1resd5  su1resd6
1      0.1074537  0.02293577  0.1234897  0.2504675  -0.164929  -0.08928723  -0.2539258
```

적합한 모형과 Cur_diff, su1res의 2017년의 예측치를 토대로 방한관광객수의 2017년 예측치를 생산한 결과 다음과 같이 나왔다.

```
$pred
Time Series:
Start = 2017
End = 2017
Frequency = 1
[1] 0.0366038
```

이 중, 2017년의 예측치에 해당하는 0.0366038는 사실 로그 차분된 데이터의 예측치로서,

$$\nabla X_{t+1} = X_{t+1} - X_t$$

이다. 따라서 원래 예측치로 다시 복원하기 위해서는 0.0366038의 데이터에 바로 전 시차의 데이터인 2016년 방한관광객수의 로그변환 수치를 더해줄 필요가 있다.

$$\nabla X_{t+1} + X_t = X_{t+1}$$

로그변환한 수치를 다시 지수변환하여 원값으로 돌린 후, 이 예측값을 실값과 대조한 결과는 다음과 같다.

	예측값	실값	오차
2017	2383563	2311447	3%

V. 결 론

1. 연구가설 H11의 경우, 국가 우호도는 “친근감을 느낀다(5점)” 과 “친근감을 느끼는 편(4점)” 모두 방한관광객수와 유의성이 존재하는 것으로 나타났다.

2. 연구가설 H12의 경우, “산업생산지수”와 “환율” 변수 모두 방한관광객수와 유의성이 있는 것으로 나타났다.

3. 연구가설 H21의 경우, 국가 우호도를 투입계열로 적합한 ARIMAX 모형의 경우 타당성을 가지지 못하는 것으로 나타났다. 이는 사전 백색화를 실시한 산출계열(방한관광객수), 투입계열(국가우호도)의 교차상관계수 검증 결과 두 개의 계열이 서로 상관성이 없는 독립 계열로 나타났기 때문이다.

4. 연구가설 H22의 경우, 환율을 투입계열로 적합한 ARIMAX 모형의 경우 모형이 타당성을 가지는 것으로 나타났다. 이는 사전 백색화를 실시한 산출계열(방한관광객수), 투입계열(환율)의 교차상관계수가 0차시와 1차시에서 유의한 상관관계를 도출하였으며, 최종 적합된 모형의 잔차들이 서로 자기상관을 가지지 않는 완전한 백색잡음으로 나타나 투입계열 X_t 와 잡음 η_t 가 정보를 잘 여과한 것으로 나타났기 때문이다.

5. 한편, 연구가설과는 별개로 “환율”과 “산업생산지수”를 토대로 적합한 선형회귀모형과 “환율” 변수를 토대로 적합한 ARIMAX모형의 예측력을 비교한 결과, 선형회귀모형은 오차가 7%이고, ARIMA-X 모형의 경우 오차가 3%로 나타나 ARIMAX 모형이 선형회귀 모형보다 예측력이 더 우수한 것으로 나타났다.

참고문헌(15진)

▽

이우리, “시계열분석과 예측”, 도서출판 탐진, 2013.

김현철, “시계열자료의 분석과 예측”, 교육과학사, 2005.

정효상, “전이함수모형을 이용한 한국인 입국자 수의 분석”, 중앙대학교 통계대학원, 석사학위논문, 2014.

김영주, “계절 ARIMA 모형을 활용한 자가항공 수요예측”, 관광연구논총 제26권 제1호(통권 33), 한양대학교 관광연구소, 2014.

한국관광문화연구원, “관광수요예측 가이드라인 설정 연구”, 한국관광문화연구원, 2006

민경창 외, “계절성 ARIMA 모형을 이용한 항공화물 수요 예측 : 인천국제공항발 유럽항공 노선을 중심으로”, Journal of korea transportation vol33 no3, 인하대학교 물류전문대학원, 2013.

윤지성 외, “계절형 다변량 시계열 모형을 이용한 국제항공 여객 및 화물 수요예측에 관한 연구”, 한국통계학회논문집, 중앙대학교 통계학과&항공대학교 경영학과, 2010.

김성태, “관광수요예측에 대한 실증연구 : 패널데이터 분석 기법을 중심으로”, 관광레저연구 제26권 제1호 pp.115-129, 한국관광레저학회, 2014.

이충기 외, “일본인 관광객의 방한 수요결정요인에 관한 연구 : 계량경제모형을 중심으로”, 관광레저연구 제18권 제4호,

이강욱 외, “국제관광수요 영향요인 분석”, 한국문화관광연구원, 2017.

이충기 외, “한국관광에 대한 일본 관광객의 지각된 가치가 방문객 만족도와 재방문의사에 미치는 영향 연구”, 관광레저연구 제20권 제4호 pp69-85, 한국관광레저학회, 2008.

전형규 외, “관광지의 특성, 관광객의 지각된 가치, 만족, 행동의도간의 관계 연구”, 관광연구 제25권 제3호 pp.237-258, 대한관광경영학회, 2010.

이진용 외, “한국의 국가 브랜드 이미지 차별화 전략에 대한 연구 : 설문응답자의 소속 국가 특성 및 자기범주화에 따른 한-중-일 국가브랜드이미지 인식 비교를 중심으로”, 광고학연구 제24권 2호, 한국광고학회, 2013.

이상훈 외, “국가이미지가 관광지 이미지와 재방문 의도에 미치는 영향”, 관광레저연구 제28권 5호, 한국관광레저학회, 2017.

정은지, “방한 중국인 관광객의 기대-지각 가치가 국가 이미지와 태도 및 행동의도에 미치는 영향에 관한 연구”, 경희대학교 석사학위논문, 2017.

유승훈, “항만투자자과 경제성장의 인과관계에 대한 소고”, 해양정책연구 제 19권 1호, 한국 해양수산개발원, 2000.

Valarie, A, Zenithal, “Consumer Perception of Price, Quality, and value A means-end Model and Synthesis of Evidence”, Journal of marketing vol.52, 1988.

R.F. Engle and C.W.J. Granger, “Cointegration and Error Correction: Representation, Estimation and Testing”, Econometrica, Vol. 55, 1987, pp. 251 ~ 276.