

국민 국내관광 수요의 모형화 및 예측연구

김형종 · 강현수

Korea
Culture &
Tourism
Institute



한국문화관광연구원
Korea Culture & Tourism Institute

정책연구 2020-12

국민 국내관광 수요의 모형화 및 예측연구

Modeling and Forecasting Korean Domestic Tourism Demand

김형종·강현수



한국문화관광연구원
Korea Culture & Tourism Institute

연구책임

김형종 관광정책연구실 부연구위원

강현수 관광정책연구실 연구원

국민 국내관광 수요의 모형화 및 예측연구



연구개요

1. 서론

1.1. 연구 배경 및 목적

가. 연구 배경

- 한국 경제는 선진 경제(지역) 수준에 도달하여 과거와 같은 고도성장이 어려움
- 점차 둔화하는 우리나라의 경제성장률을 제고하기 위하여 서비스산업의 생산성을 향상시키려는 시도가 이루어지고 있으며(기획재정부, 2020), 서비스업의 핵심인 관광 산업(tourism industry)은 일자리 창출 효과가 높고 지역 경제를 활성화하며, 삶의 질과 행복 추구에도 매우 중요하다는 점이 인식되고 있음(이수일, 2019)
 - 2019년을 기준으로 외국인 인바운드(inbound tourism) 지출규모는 30.4조 원, 국내관광(domestic tourism) 지출규모는 57.7조원으로 나타나 국내관광 규모가 인바운드 관광 대비 약 2배에 달하는 것으로 나타남(김형중, 2020)
 - 또한 지역 균형 발전이나 지역의 부족한 수요를 보충하기 위한 정책적 수단으로 국내관광·지역관광 활성화가 논의되고 있음(관계부처합동, 2019)
- 그러나 크게 증가하고 있는 국내관광 및 그에 대한 대내외적 관심도, 정책적 요청에도 불구하고, 국내관광 수요 예측에 대하여 충분하게 연구하지 못하였으며 정책적 요청에 따라 일시적이고 산발적으로 예측이 수행됨
- 따라서 검증하는 외부의 정책적 요구에 부응하기 위해서는, 국내관광 관련 데이터를 축적하고 예측결과를 지속적으로 향상시키기 위한 지속적인 노력이 필요함

나. 연구 목적

- 본 보고서의 목적은 시의적절하고 정확한 국내관광 예측 수행을 위한 필요 사항과 예측 모형을 발전시키기 위한 방향을 제시하여 정책적 요구 충족에 기여하는데 있음

- 이를 위해서, 첫째 국내관광 수요예측 선행연구 검토를 통해, 국내관광 수요예측에 필요한 사항들과 모형을 점검하고, 둘째, 수 년 이내의 미래를 예측할 수 있는 중장기 예측모형을 제시하며, 셋째 교차 검증을 거친 국내관광 예측 결과를 제시함

〈표 1〉 연구 목적

구분		세부 내용	연구 목적
국내관광 수요의 모형화 및 예측	선행연구 검토	• 국내관광을 예측하기 위하여 필요한 변수, 모형에 대해 문헌조사	<ul style="list-style-type: none"> • 시의적절하고 정확한 국내관광 예측 수행을 위한 필요 사항, • 예측 모형을 발전시키기 위한 방향을 제시하여 • 정책 요구 충족에 기여
	예측모형 제시	• 수 년 이내의 미래를 예측할 수 있는 중장기 예측모형 제시	
	예측결과 제시	• 교차 검증을 거친 국내관광 예측 결과 제시	

1.2. 연구 범위 및 방법

가. 연구 범위

- (공간적 범위) 대한민국, (시간적 범위) 국민 국내관광객 총량 데이터 가용기간
- (내용적 범위) 다음의 표에서 제시

〈표 2〉 연구 범위

연구 범위	내용
공간적 범위	• 대한민국 국내
시간적 범위	• 관광 총량 데이터 가용 기간 전체
내용적 범위	<ul style="list-style-type: none"> • 국내관광 및 예측절차 개념 정리 • 국내관광 수요예측 선행연구 정리 • 국민여행조사 데이터 설명 • 예측 모형 수립 및 예측 모형 검증

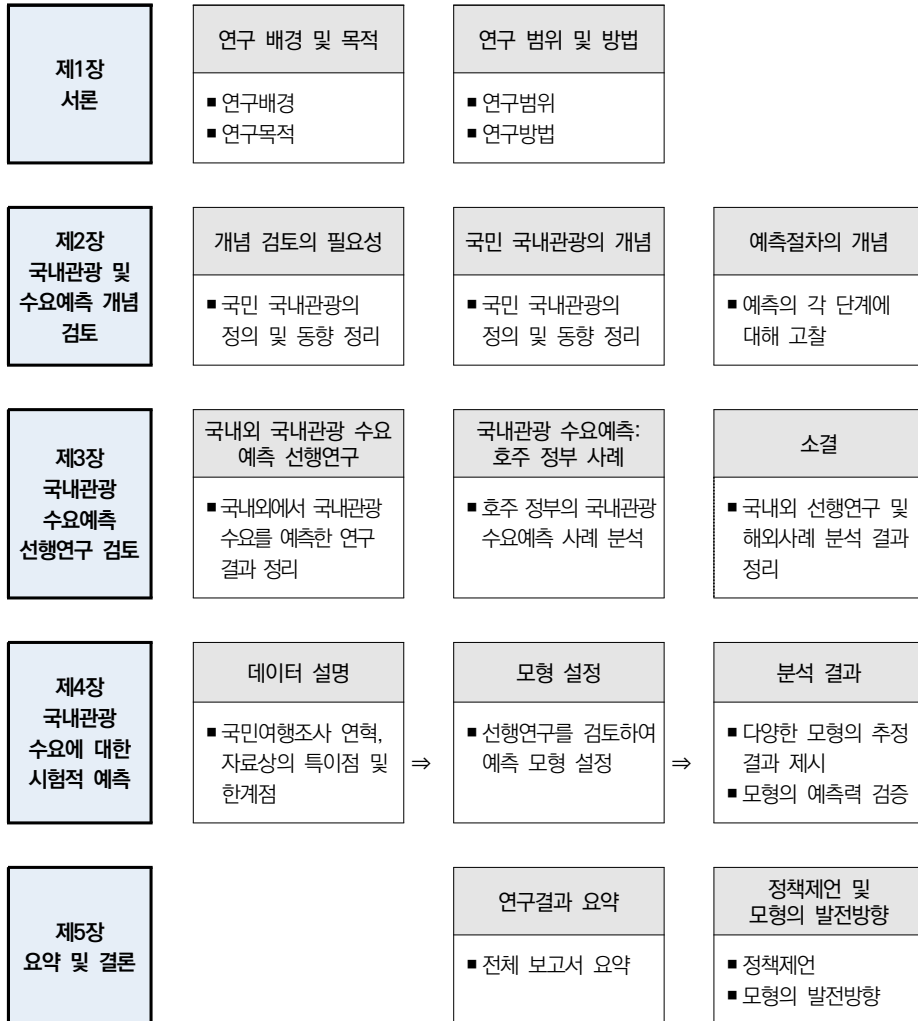
나. 연구 방법

- 연구 방법은 크게 (1) 선행연구 조사, (2) 데이터 정리, (3) 모형 수립 및 추정

1.3. 연구 흐름도

- 본 보고서 전체의 연구 흐름을 다음의 그림으로 정리할 수 있음

[그림 1] 연구 흐름도



2. 국내관광 및 예측절차 개념

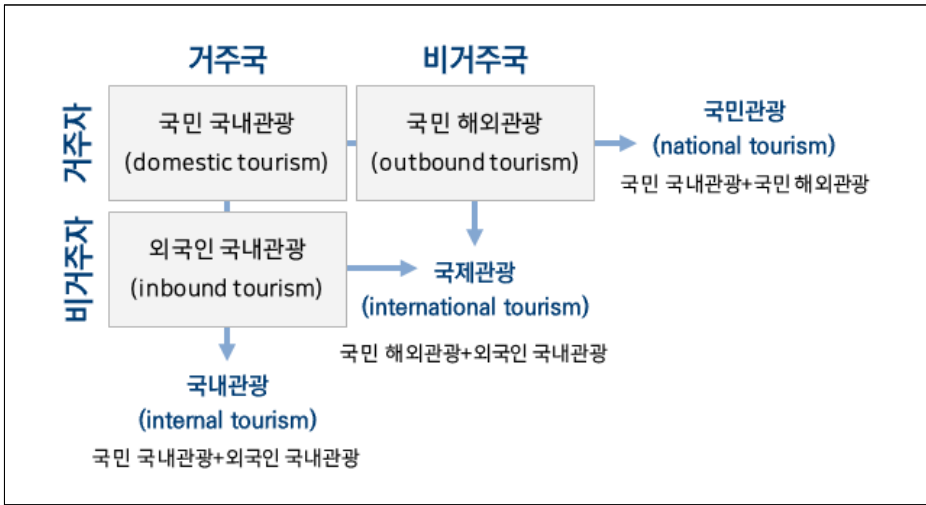
2.1. 개념 검토의 필요성

- 본격적인 연구를 수행하기에 앞서, 사전적으로 두 가지 개념에 대한 검토가 필요
- 첫째, 국민 국내관광에 대한 개념 검토가 필요
 - ‘인바운드 관광(inbound tourism)’을 외국인 국내관광으로 지칭하는 경우도 존재하기 때문에 ‘국민 국내관광’(domestic tourism)이라는 표현을 사용하며, 국민 국내관광이 포괄하는 범위 등에 대한 사전적 논의가 필요
- 둘째, 예측 및 예측절차에 대한 개념 검토가 필요
 - 본 보고서에서 다루는 예측은 통계적 예측임
 - 통계적 예측과 추정은 상당히 밀접한 관련성이 있지만 별개의 개념으로 구분됨
 - 대다수의 실증분석 연구는 예측이 아니라 추정을 수행하기 때문에, 본 보고서의 논지와 상이한 부분이 있고, 일부 개념에 대한 정보가 없으면 오해 소지가 있음
 - 필수적이라 생각되는 예측 및 예측절차에 대한 개념을 검토하여 보고서의 나머지 부분에 대한 이해의 향상을 도모

2.2. 국민 국내관광의 개념

- 국민 국내관광(domestic tourism)은 내국민이 자신의 국가 영토 내에서의 여행으로서 행정구역을 기준으로 자신의 거주지에서 벗어나 여가, 위락, 휴가, 성지순례, 건강, 치료 등의 목적으로 타 지역으로의 이동을 의미함
- 세계관광기구(UNWTO)는 국적과 영토에 따라 관광을 국내관광, 인바운드관광, 아웃바운드관광, 국민관광, 국제관광등 5가지 유형으로 구분하고 있는데 국내관광은 내국민의 국내여행으로 정의됨(이연택, 2015)

[그림 2] 관광 유형 도식화

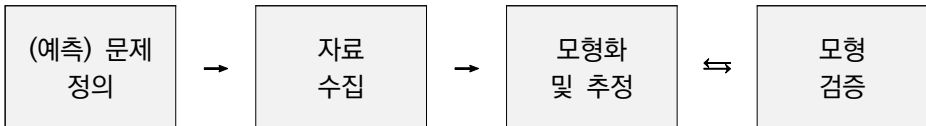


자료: United Nations (2010).

2.3. 예측절차의 개념

- 수요예측의 절차는 (1) 문제 정의, (2) 자료 수집, (3) 모형화 및 추정, (4) 모형 검증(model validation)의 순서로 진행됨

[그림 3] 예측 절차



- 통계적 예측과 추정은 상당한 관련성이 있으나 차이점도 있음
- 우선 일반적으로 (모수적) 추정은 모형에서 정확한 모수의 값이 관심사이나, 예측은 (모수의 값과 관계없이) 정확한 미래의 값이 관심사
- 또한 설명력과 예측력은 구분되는 개념이며, 주어진 데이터에서의 설명력은 매우 높으나 예측력이 없는 경우가 발생하고 그러한 경우를 과적합(overfitting)이라 함
- 과적합을 막기 위하여 추가적으로 모형의 예측력을 검증하는 과정이 필요함

3. 국내관광 수요예측 선행연구 검토

3.1. 우리나라의 관광 수요예측 선행연구

- 국내관광 수요예측 정책연구 보고서로 이강욱(1999), 박상곤(2017a)이 있음
 - 이강욱(1999)은 국내관광 모형에 대한 개념만을 제시
 - 한편, 박상곤(2017a)은 국민여행조사 미시 자료에 기반을 두는 모형을 제시하였으나, 국민여행조사가 크게 개편되어 예측에 활용할 수 없음
- 국내관광 총량에 대한 우리나라의 수요예측 선행연구는 없었으며, 따라서 특정 지역의 관광객 총량에 대한 예측 연구도 검토 대상에 포함
- 국내 선행연구들은 수요예측의 대상을 특정 지역 및 관광지로 한정하고 있을 뿐만 아니라, 해당 지역에 방문한 외국인 국내관광도 수요예측에 포함하고 있어, 국민 국내관광의 총량 예측과는 목적이 상이함
- 그러나 국내 지자체 및 특정 관광 지점에 대한 선행연구 정리는, 본 과제의 국내관광 총량 예측에는 기여가 미미할 수 있음에도, 후속과제 및 정책대응에 있어서는 상당한 기여를 할 수 있을 것으로 판단되어 본 보고서의 내용으로 포함함
- 검토결과 국내의 연구들은 대부분 월별 관광객 수와 같이 시차변수만을 사용하는 시계열 모형으로 나타났으나 최근에는 인과 모형을 함께 검토하고 있음
- 관광 수요예측 관련 해외 연구동향은 크게 정량적인 방법과 정성적인 접근으로 구분하고 있으나 정량적인 연구방법이 다수를 차지하고 있음
- 국내관광(domestic tourism)을 예측한 해외 연구는 호주 및 라스베이거스의 사례를 대상으로 한정적으로 나타남
- 외국의 연구들은 시계열 모형뿐만 아니라 소득, 관광가격, 항공 요금, 숙박 요금을 설명변수로 하는 회귀 모형을 활용하는 것으로 분석됨

〈표 3〉 국내관광 수요예측 선행연구

구분	활용자료	분석방법	분석기간	주기	예측범위
이강욱 (1999)	출입국통계, 국민여행조사	시계열분석, 회귀분석 등	-	-	-
박상곤 (2017a)	소득, 가족구성원 연령	다항 로짓 모형	2011.01- 2016.12	-	-
김상원·이덕순 (2013)	악산시관광객통계 (관광객수)	시계열모형 (원터스 지수평활법)	2005.01- 2012.12	월	2013.01- 2013.12 (1년)
현수영 외 (2015)	제주도입도객통계 (관광객수)	원터스 지수평활법, 계절형 ARIMA	1994.01- 2013.09	월	2013.10- 2014.12 (1년)
김상원·박미선 (2016)	수원시관광객통계 (관광객수), 시간	시계열 모형 (이동평균법, 지수평활법) 인과모형(회귀모형)	2004.01- 2013.12	월	2014.01- 2014.12 (1년)
김상원 (2018)	제주도입도객통계 (관광객수), 시간	시계열 모형(지수평활 모형: 단순, 브라운, 홀트, 원터스 가법, 원터스 승법) 인과 모형(회귀 모형)	2007.01- 2017.12	월	2018.01- 2018.12 (1년)
Witt et al. (1992)	라스베이거스 컨벤션 및 방문자 관리국에서 발행 한 Marketing bulletin과 1990년 Ten Year Summary의 간행물	지수평활법	1973- 1988	월	1989 (1년)
Athanasopoulos and Hyndman (2008)	국민방문객조사(NVS) 방문객 숙박 수	회귀모형, 순수 시계열 모형-단일 오차(또는 혁신) 상태 공간 모형, 외생변수 포함 혁신 상태공간모형	1998.1분기- 2005.2분기	분기	2005- 2014 (10년)
Athanasopoulos et al. (2009)	국민방문객조사(NVS) 방문객 숙박 수	변형 하향식 접근법, 최적 결합 접근법, 지수평활법, 표본외 예측평가	1998.1분기- 2006.4분기	분기	2007- 2008 (1년)

자료: 선행연구를 바탕으로 연구자 정리

3.2. 호주 정부의 국내관광 수요예측 사례

- 호주 무역투자대표부(Austrade) 산하 호주관광연구소(Tourism Research Australia; TRA)는 주기적으로 향후 10년간 국제관광·국내관광 수요예측 실시
- 호주 정부는 관광 수요예측을 위해 공항공사(Airport Coordination Australia), 통계청(Australian Bureau of Statistics), 중앙은행(Reserve Bank of Australia), 호주관광연구소(TRA) 등의 기초 통계자료를 활용함
 - 국민 국내관광 수요예측에 활용되는 시장정보는 세계 및 국내 경제 상황, 항공 수용력 및 항공운임, 국내 숙박시설 공급 및 객실료, 원천 시장에 영향을 미칠 수 있는 중요한 사건 등을 포함
- 호주 국내관광의 관광객 활동과 지출에 대해 다음과 같은 방법으로 수요예측
 - 계량 시계열 모형(Econometric Time Series Model)과 국내외 경제상황, 항공 수용력 및 항공요금, 국내 숙박 공급 및 숙박요금, 주요 이벤트 등에 바탕을 둔 시장정보분석(Market Intelligence)을 함께 고려함
 - 향후 가장 영향을 미칠만한 요소들에 대한 추가적인 통찰력을 얻고자 산업 전망조사(industry sentiment survey)를 실시함
 - 예측의 정확도 제고를 위해 예측결과를 바탕으로 산업계 및 정부 전문가들로 구성된 패널(Tourism Forecasting Reference Panel)을 구성하여 국내관광 예측안에 대한 검토 및 조정기능을 수행함
- 현재 우리나라의 국제관광 수요예측은 호주와 유사한 방식으로 수행되고 있으나, 국내관광은 데이터의 수집, 예측의 주기성, 검토 방식 등에서 모두 부족함
 - 호주는 항공운임, 숙박시설 공급, 객실료와 같은 미시 가격정보를 풍부하게 수집하여 수요예측에 반영하고 있으며, 산업계와 정부 전문가가 초안을 검토

〈표 4〉 호주 국민 국내관광 수요예측 개요

구분	주요 내용
목적	정책 입안자, 기획자, 투자자 등이 의사결정에 활용할 수 있는 기초자료를 제공
예측기간	10년
지역구분	호주 국내 전체, 각 주 및 테리토리(territory)별 예측
유형분류	숙박여행 및 당일여행으로 분류
예측항목	방문객수, 숙박일수, 관광지출

자료: 호주관광연구소(TRA) 홈페이지 내 국내관광 수요예측 내용을 바탕으로 연구자 재정리

4. 국내관광 수요에 대한 시범적 예측

4.1. 시범적 예측의 의의

- 선행연구 조사 결과, 국내관광 예측에 대한 연구는 전 세계적으로도 사례가 희소
- 우리나라에서는 국내관광 총량 예측이 주기적으로 수행되지 않고 있기 때문에, 정부정책을 뒷받침할 수 있는 시의적절한 예측치의 산출에 어떠한 지원이 필요한지, 예측 모형을 발전시키기 위하여 어떠한 점을 보완하고 수정하여야 하는지에 대한 담론 자체가 전무함
- 이에 본 장에서는 국내관광 예측에서의 문제점 파악을 위하여 국내관광 총량의 시범적 예측을 수행

4.2. 국내관광 데이터 설명

- 본 연구에서는 선행연구 분석 및 전문가 의견을 바탕으로 ‘국민여행조사’ 국민여행총량 데이터를 국내관광 수요예측을 위한 기초자료로 활용함
- 국민여행조사의 국민여행 총량 중 대표 지표인 국내여행 횟수를 바탕으로 국민 국내관광 수요예측 모형 구축 및 검증 과정에 활용하고자 하였음
- 국민여행조사는 잦은 변경으로 인하여 시계열이 단절되는 문제가 존재하며, 특히 최근의 조사방식 변경으로 상당한 변화가 발생하였다고 추측됨
- 특히, 최근 조사방식이 변경된 이후로 시계열에서 계절성이 현저한 감소가 관찰되며, 이러한 현상에 대해서 검토와 분석이 필요하다고 보임

4.3. 모형 설정 및 추정 결과

가. 모형 설정

- 본 보고서에서는 국민여행조사 국내여행 횟수를 종속변수로 하고 국내외 선행연구에서 활용된 주요 변수를 고려하여 모형을 설정함

- 모형의 설정은 동일한 설명력을 가지면 모수가 작을수록 좋다는 모수 절약의 원칙을 따르고 변수는 로그값을 취하여 모형에 적용함
- 모형의 설정을 위해 소득, 인구, 공휴일, 계절성 등의 변수를 사용하였으며, 각 변수를 사용하는 모형에 대한 결과를 제시함

나. 추정 결과

- 추정을 통해 여러 모형을 검토한 결과, 국내총생산이나 국내총생산을 총인구로 나눈 값으로 정의되는 일인당소득을 설명변수로 사용하는 모형이 적합하다고 판단
- 인구 역시 관광총량과 상당한 관련성이 존재하는 변수이나, 우리나라의 인구는 앞으로 감소가 예상되는 만큼 예측에 적절한 변수로 간주하기 어려움
- 특히 분기더미를 포함하는 모형과 아닌 모형의 결정계수는 큰 차이가 있었으며, 설명변수 추가보다 계절성의 통제가 설명력 향상에 크게 기여하는 것으로 나타남
- 종속변수의 시차 변수를 추가한 모형에서도 설명력의 향상은 크지 않았으며, 국내관광 시계열의 길이는 예측에 활용하기에 충분하지 않다고 보임
- 현재 주어진 데이터의 수준에서 많은 더미변수의 사용이나 설명변수의 추가는 과적합의 문제를 야기하는 것으로 판단함
- 과적합 문제의 주요 해결책이 데이터 증강(data augmentation) 및 변수의 감소임을 감안하면 현재의 모형에 추가적으로 다양한 가정이나 변수의 추가는 바람직하지 않다고 판단함
- 신뢰할 수 있는 국내관광의 장기 시계열 확보가 국내관광 수요예측의 시의성 및 정확성 해결의 근본적인 열쇠라고 판단함

4.4. 모형 검증 및 예측

- 모형 검증 과정에서도 설명변수의 수가 작은 모형의 예측력이 더 우수하다고 나타났다으며, 과적합 문제를 해결하기 위해서는 장기 시계열의 확보가 시급함을 시사

- 본 보고서에서는 모형 검증을 통해 선정한 예측 결과와 국민여행조사 국내여행 추정치 및 과거 연구원에서 생산한 예측결과를 비교하여 <표 5>와 같이 제시함
- 보고서의 예측모형을 과거 연구원에서 일시적으로 수행하였던 결과와 비교하면, 상대적인 예측력 향상이 존재함을 확인할 수 있음
 - 과거의 예측이 10년 이후를 예측하는 장기 예측인 점, 또한 시계열 기법을 사용하였다는 점에서 상대적으로 큰 오차가 발생한다고 판단됨
- 또한 2018년의 국내여행 수치 증가는, 실제로 국내여행이 증가한 것이 아니라 시계열 단절로 인한 변동임을 시사함
- 따라서 국민여행조사에 시계열 보정이 필요하다고 판단되지만 현재의 결과가 과대추정인지 과거의 결과가 과소추정인지에 대해서는 추가적인 검토가 필요함

〈표 5〉 예측 결과 비교

(단위 : 천회, 천일)

연도	(1) 〈표 4-25〉 1월 예측	(2) 〈표 4-27〉 3월 예측	(3) 〈표 4-28〉 3월 예측	(4) 〈표 4-29〉 3월 예측	(5) 국민여행 조사 (여행횟수)	(6) 국민여행 조사 (여행일수)	(8) 2000년 국내관광 예측	(9) 2011년 국내관광 예측
2002	-	-	-	-	-	-	404,648	-
2005	-	-	-	-	257,790	388,837	507,436	-
2006	-	-	-	-	284,575	416,982	-	-
2007	-	-	-	-	302,861	477,372	-	-
2008	-	-	-	-	245,669	408,027	536,876	-
2009	-	-	-	-	219,586	375,341	-	-
2010	-	-	-	-	168,148	339,608	-	-
2011	-	-	-	-	156,594	286,947	605,968	473,833
2012	-	-	-	-	213,468	365,282	-	492,617
2013	-	-	-	-	231,035	389,220	-	512,220
2014	-	-	-	-	227,100	397,847	-	532,683
2015	-	-	-	-	238,297	406,819	-	554,055
2016	269,055	-	-	-	241,750	412,378	-	576,347
2017	293,721	276,445	-	-	284,966	479,674	-	599,637
2018	317,406	295,451	298,964	-	311,153	556,890	-	623,960
2019	335,502	310,117	314,607	319,311	344,750	585,710	-	649,367

주: 2000년, 2011년 예측은 내부자료 수치이다. 각 예측 값은 해당 연도의 실질 국내총생산을 대입하여 계산하였다. 실제 예측에서는 한국은행이나 한국개발연구원의 경제성장률 전망을 활용할 수 있다.

5. 결론: 정책제언 및 모형의 발전방향

- 결론에서는 전체 결과를 요약함과 동시에 국내관광 수요의 모형화 및 예측 과정에서 발견한 정책제언 및 모형의 발전방향에 대해서 제시함
- 시의성 있고 정확성을 담보하는 국내관광 수요예측 문제에서 가장 큰 걸림돌은 데이터 부족이며, 이를 해결하기 위해서는 현재 실시하고 있는 국민여행조사 표본의 확대와 별개로 과거 시계열에 대한 보정이나 국내관광 지표의 개발이 필요함
- 또한 호주 정부의 수요예측 사례에서 나타나는 것처럼 교통 가격 정보, 숙박 가격 정보를 수요예측에 활용할 수 있도록 문화체육관광부에 해당 정보를 수집할 수 있는 법적 근거를 마련하여 주어야하며, 해당 업무를 수행할 인력의 증원도 필요
- 국내관광과 관련이 있는 외부 기관과의 협력관계 구축 필요
 - 특히 교통과의 협력관계 구축이 국내관광 수요예측에 필수적이라고 보이며, 관련 협회를 통해 숙박 업체의 가격 정보를 수집할 필요가 있음
- 데이터의 품질을 관리하고 연구에 필요한 적기에 활용할 수 있도록 데이터 댐을 구축할 필요성이 있음
 - 데이터 댐은 단순히 데이터만을 수용하고 관리하는 것으로는 부족하며, 세계은행(World Bank)이나 국제통화기금(International Monetary Fund)의 데이터 제공 방식을 참조하여 실제적인 활용도 향상을 도모할 필요가 있음
- 시간가변계수모형(time-varying coefficient model)과 같이 회귀계수의 동적변화를 나타낼 수 있는 모형과 결합하거나 통계적 학습(statistical learning)을 통해 예측을 수행하는 방향으로 발전하는 것이 바람직함
 - 다만, 이러한 발전방향에는 선제적으로 장기 시계열의 확보가 필요
- 매년 축적되는 데이터 변화에 맞춰 장기적으로 예측 결과를 실제 값과 비교하고 개선하는 과정이 필요하며, 이러한 과정을 이해하고 지원하는 관행이 정착될 필요

정책제언 요약

1. 시의성 있고 정확한 국내관광 수요예측을 위해서는 장기 시계열 확보가 필요하다. 이를 위해서 국민 여행조사의 확대·개편과는 별개로 과거의 시계열을 보정하거나 국내관광 지표를 생산하는 사업이 필요하다.
2. 수요예측의 정확도를 향상시키기 위해서는 국내관광과 관련한 추가적인 정보의 수집도 요구된다. 문화체육관광부에서 교통가격, 숙박가격과 관련한 정보를 입수할 수 있도록 법적 근거를 마련함과 동시에 해당 업무를 수행할 인력 증원이 필요하다. 외부의 관련 기관과 협업체계를 구축할 필요성도 있다.
3. 상기의 방식으로 생산하고 수집한 정보를 관리하기 위하여 데이터 댐을 구축하여야 한다. 즉시적인 수요예측 대응을 위해서는 개인이 데이터를 구축하지 않고 데이터 댐에서 완성된 데이터가 제공될 수 있도록 하여야 하며, 장기적으로는 자체적인 분석 및 예측 기능을 탑재하도록 발전하여야 한다.

목차

제1장 서론	1
제1절 연구 배경 및 목적	3
1. 연구 배경	3
2. 연구 목적	5
제2절 연구 범위 및 방법	9
1. 연구 범위	9
2. 연구 방법	10
3. 연구 흐름도	12
 제2장 국내관광 및 수요예측 개념 검토	13
제1절 개념 검토의 필요성	15
제2절 국민 국내관광의 개념	17
제3절 예측절차의 개념	19
1. 예측절차 개요	19
2. 문제 정의	20
3. 자료 수집	21
4. 모형화 및 추정	23
5. 모형 검증	25
 제3장 국내관광 수요예측 선행연구 검토	27
제1절 국내관광 수요예측 선행연구	29
1. 우리나라의 국내관광 수요예측 선행연구	29
2. 외국의 국내관광 수요예측 선행연구	32

제2절 호주 정부의 국내관광 수요예측 사례	37
1. 개요	37
2. 수요예측 데이터	38
3. 수요예측 절차	40
제3절 소결	42
제4장 국내관광 수요에 대한 시범적 예측	45
제1절 시범적 예측의 의의	47
제2절 국내관광 데이터 설명	48
제3절 모형 설정 및 추정 결과	55
1. 모형 설정	55
2. 추정 결과	58
제4절 모형 검증 및 예측	81
1. 표본외 검증	81
2. 예측 결과 비교	85
제5장 요약 및 결론	87
제1절 연구결과 요약	89
제2절 제언	91
참고문헌 /	95
ABSTRACT /	99
부록 /	101
【부록1】 수요예측 모형 설명변수 개요 /	103
【부록2】 국내관광 대리변수 탐색 /	110

표 목차

〈표 1-1〉 연구 목적	8
〈표 1-2〉 연구 범위	9
〈표 2-1〉 관광통계국제권고안(IRTIS 2008) 용어 제개념	18
〈표 2-2〉 국민여행조사 주요 연혁	21
〈표 2-3〉 예측절차별 주요 내용	26
〈표 3-1〉 우리나라의 국내관광 수요예측 선행연구	31
〈표 3-2〉 외국의 국민 국내관광 수요예측 선행연구	32
〈표 3-3〉 모형 계산식	35
〈표 3-4〉 호주 국민 국내관광 수요예측 개요	37
〈표 3-5〉 호주 관광 수요예측 주요 데이터 출처	38
〈표 3-6〉 호주 국가방문자조사(NVS) 개요	39
〈표 4-1〉 분석 대상 기초자료	48
〈표 4-2〉 국민여행조사 개요	49
〈표 4-3〉 2019년 국민여행조사 주요 내용	49
〈표 4-4〉 국민여행조사 연혁 및 변동사항 (2004-2008년)	51
〈표 4-5〉 국민여행조사 연혁 및 변동사항 (2009-2013년)	52
〈표 4-6〉 국민여행조사 연혁 및 변동사항 (2014-2018년)	53
〈표 4-7〉 국민여행조사 분석자료 개요	54
〈표 4-8〉 주요 관광 수요예측 연구에서의 모형 함수 형태(functional form)	56
〈표 4-9〉 기술 통계량	58
〈표 4-10〉 로그 국내총생산을 설명변수로 하는 회귀모형 추정 결과	60
〈표 4-11〉 로그 인구를 설명변수로 하는 회귀모형 추정 결과	64
〈표 4-12〉 로그 일인당 총생산을 설명변수로 하는 회귀모형 추정 결과	66
〈표 4-13〉 로그 공휴일수를 설명변수로 하는 회귀모형 추정 결과	69
〈표 4-14〉 로그 일인당소득을 설명변수로 하는 회귀모형 추정 결과	71
〈표 4-15〉 로그 국내총생산을 설명변수로 하는 습관 지속 모형 추정 결과 (제약: $\psi_s = 0$)	74
〈표 4-16〉 로그 국내총생산을 설명변수로 하는 습관 지속 모형 추정 결과	77

〈표 4-17〉 모형 (1)-(4) 추정결과 및 표본외 검정결과 (분석에 사용한 기간: 2010q1-2016q4)	83
〈표 4-18〉 모형 (1)-(4) 추정결과 및 표본외 검정결과 (분석에 사용한 기간: 2010q1-2017q4)	84
〈표 4-19〉 모형 (1)-(4) 추정결과 및 표본외 검정결과 (분석에 사용한 기간: 2010q1-2018q4)	84
〈표 4-20〉 예측 결과 비교	85

그림 목차

[그림 1-1] 연구 흐름도	12
[그림 2-1] 관광 유형 도식화	17
[그림 2-2] 예측 절차	19
[그림 4-1] 국민여행조사 국민여행 횟수(2010.1분기 — 2019.4분기)	50
[그림 4-2] 국내관광과 <표 4-10> (1)열 모형 추정치 시계열도	61
[그림 4-3] 국내관광과 <표 4-10> (2)열 모형 추정치 시계열도	61
[그림 4-4] 국내관광과 <표 4-10> (3)열 모형 추정치 시계열도	62
[그림 4-5] 국내관광과 <표 4-10>의 (4)열 추정치 시계열도	62
[그림 4-6] 국내관광과 <표 4-11>의 (2)열 추정치 시계열도	64
[그림 4-7] 국내관광과 <표 4-11>의 (4)열 추정치 시계열도	65
[그림 4-8] 국내관광과 <표 4-12>의 (1)열 추정치 시계열도	67
[그림 4-9] 국내관광과 <표 4-12>의 (2)열 추정치 시계열도	67
[그림 4-10] 국내관광과 <표 4-12>의 (3)열 추정치 시계열도	68
[그림 4-11] 국내관광과 <표 4-12>의 (4)열 추정치 시계열도	68
[그림 4-12] 국내관광과 <표 4-14>의 (1)열 추정치 시계열도	72
[그림 4-13] 국내관광과 <표 4-14>의 (2)열 추정치 시계열도	72
[그림 4-14] 국내관광과 <표 4-14>의 (3)열 추정치 시계열도	73
[그림 4-15] 국내관광과 <표 4-14>의 (4)열 추정치 시계열도	73
[그림 4-16] 국내관광과 <표 4-15>의 (1)열 추정치 시계열도	75
[그림 4-17] 국내관광과 <표 4-15>의 (2)열 추정치 시계열도	75
[그림 4-18] 국내관광과 <표 4-15>의 (3)열 추정치 시계열도	76
[그림 4-19] 국내관광과 <표 4-15>의 (4)열 추정치 시계열도	76
[그림 4-20] 국내관광과 <표 4-16>의 (1)열 추정치 시계열도	79
[그림 4-21] 국내관광과 <표 4-16>의 (2)열 추정치 시계열도	79
[그림 4-22] 국내관광과 <표 4-16>의 (3)열 추정치 시계열도	80
[그림 4-23] 국내관광과 <표 4-16>의 (4)열 추정치 시계열도	80

국민 국내관광 수요의 모형화 및 예측연구

제1장

서론

제1절 연구 배경 및 목적

1. 연구 배경

한국은 제조업을 중심으로 선진국에 의해 검증된 경로를 따라 성장하는 전략을 채택하여 고도성장 하여 왔으나, 한국 경제가 선진 경제지역으로 분류되는 수준에 도달함에 따라서 과거와 같은 수준의 경제성장을 달성은 점차 어려워지고 있다.

제조업 노동생산성은 글로벌 금융위기 이후 상당히 둔화된 반면에 서비스업은 소폭 하락에 그쳤으며, 서비스업의 경우 둔화의 주된 원인도 자본장비율 개선세 둔화에 있었으나 제조업은 총요소생산성 둔화에 있었다(김도완·이상협, 2019). 또한 저출산·고령화로 생산가능인구가 줄어들면서 우리나라의 잠재성장률은 지금보다 더욱 낮아질 것으로 전망되고 있다(권지호 외, 2019). 서비스업의 핵심인 관광 산업(tourism industry)은 일자리 창출 효과가 높고 지역 경제를 활성화하며, 삶의 질과 행복 추구에도 매우 중요하다는 점이 인식되고 있다(이수일, 2019).

국내관광(domestic tourism)은 2019년 기준 지출규모가 57.7조원으로 외국인 인바운드(inbound tourism) 지출규모 30.4조원의 약 2배에 달하고 있다(김형중, 2020).¹⁾ 또한 지역 균형 발전이나 지역의 부족한 수요를 보충하기 위한 정책적 수단으로 국내관광·지역관광 활성화가 논의되고 있다(관계부처합동, 2019).

그러나 크게 증가하고 있는 국내관광 및 그에 대한 대내외적 관심도, 정책적 요청에도 불구하고 국내관광 수요에 대한 예측 연구는 크게 부족한 상태이다.²⁾ 본 연구원에서 수행한 연구도 예외는 아니어서 국제관광 수요예측 연구(이강욱, 1999; 전

1) 인바운드 지출 규모는 한국은행 국제수지의 운송수입과 일반여행수입 합계액에 '19년 평균환율 1.165.65를 적용하여 계산하였다.

2) 예를 들어, 문화체육관광부의 '관광개발기본계획'이나 '국내관광 정책목표'의 수립에는 지속적인 국내관광 예측치의 제시가 상당한 도움이 될 수 있다. 하지만 아직까지는 문화체육관광부의 요청에 의해 일시적인 예측치를 제공하는데 머무르고 있다.

효재, 2003; 이강욱 외, 2009; 이강욱, 2017)는 수 차례에 걸쳐 수행된 것에 비해, 국내관광 수요예측 연구는 단 두 건(이강욱, 1999; 박상곤, 2017a)으로 수에 있어서도 상대적으로 미미하다.³⁾ 그러나 현수영 외(2015)에서 지적하는 것처럼, 관광수요 예측은 관광정책·관광사업의 목표 설정이나 타당성 검토를 위하여 필요한 기초 작업이며, 프로젝트의 규모가 다대할수록 장기간의 수요 예측을 필요로 한다. 그럼에도 불구하고, 과거에는 국내관광 수요예측에 대하여 직접적으로 연구를 수행하기 보다는, 외부의 요구에 따라 특정 연구자가 국내관광 예측치를 산출하는 방식으로 정책적 요구에 대응하여 온 것이 사실이다.

앞으로는 이러한 방식으로 정책적 요구에 대응하는 일은 점점 어려워질 것으로 예상된다. 소수의 연구자가 정책 요구에 일시적으로 대응하는 방식으로는, 본 연구원 내의 예측 역량 축적이 어려운 동시에 향후 크게 증가할 것으로 예상되는 외부의 정량적 요구를 충족하기도 어렵다고 보이기 때문이다.

따라서 점증하는 외부의 정책적 요구에 부응하기 위해서는, 국내관광 역시 국제관광과 마찬가지로 관련 데이터와 예측 결과를 축적하는 동시에, 예측 모형을 지속적으로 향상시키기 위한 별도의 노력이 필요한 지점에 도달하였다고 생각한다.⁴⁾ 이를 위해 국내관광 수요예측에 대한 연구 결과를 보고서의 형태로 축적하고 수요예측 대내외적으로 공유하는 방안을 검토할 필요성이 제기되고 있으며, 본 보고서에서는 이러한 필요성의 충족을 시도하였다.⁵⁾

3) 이강욱(1999)은 전반적인 관광 수요를 다루면서 일부 지면을 할애하여 개념적인 국내관광 수요를 다루고 있다. 또한 여기서 얘기하는 국내관광 수요예측 연구란 우리나라의 국내관광 총량을 예측하는 연구로, 특정 지역이나 특정 관광지의 수요를 예측하는 연구는 이에 해당하지 않는다.

4) 외국인 인바운드는 국내관광에 비하여 상대적으로 데이터의 축적이나 예측 방식에 대한 연구 결과가 많이 존재하고 있다. 물론 출입국 통계는 국내관광과 달리 추정치 아니며, 일부 밀입국 등을 제외하면 정확하게 측정되기 때문에 기초 데이터 수집이 용이하고 결과물 역시 상대적으로 정확하다는 점은 감안할 필요가 있다.

5) 일시적인 정책 대응과 달리, 보고서 형태로 연구 결과물을 축적하는 것은 상당한 이점이 있다. 우선 상대적으로 많은 부분에 있어서 사전적인 검토가 가능하다. 그리고 어떠한 모형을 어떻게 검토하였는지는 후속 연구에 도움이 될 수 있다. 어떠한 부분을 검토할 필요가 없고, 어떤 부분은 추가적으로 검토할 필요가 있는 지 판단할 수 있기 때문이다. 또한 과거의 예측치를 검토하여 예측 모형을 향상시킬 수도 있다. 또한 동일한 수요예측 방법론을 사용하더라도, 어떠한 데이터를 사용하는가에 따라 예측 결과는 달라질 수 있다. 이는 실제 수요예측을 수행하지 않으면, 사전적으로 알기 어려운 부분이기도 하다.

2. 연구 목적

본 보고서의 목적은 국내관광 수요의 실제적인 모형화와 예측을 통하여 보다 정확하고 시의적절한 국내관광 수요예측의 수행을 위해 요구되는 사항과 후속 예측의 발전방향을 제시하여 후속 예측과 정책적 요구 충족에 도움을 주는 것이다. 본 보고서의 예상 독자는 예측에 숙달한 전문가가 아니라 정책 입안자(policy maker) 또는 정책 입안에 조력하는 자이며, 따라서 본 보고서에서는 국내관광 및 수요예측과 관련하여 지식전달적인 내용도 다소간 포함될 예정이다. 즉시적인 정책적 대응이나 정책 입안자의 시간적인 비용 감소 및 이해 촉진을 위하여 국내관광 수요예측과 관련한 주제들을 압축적으로 집대성할 필요성도 존재하기 때문이다.

보고서의 목적 달성을 위해서는 다음과 같은 세부 목표의 달성이 요구된다. 우선 (1) 기존의 국내외 국내관광(domestic tourism) 수요예측 연구를 정리하고, 이를 토대로 (2) 우리나라의 국내관광 수요예측에 적합한 모형 개발과 그에 필요한 절차들을 제시하는 동시에, (3) 동 모형을 사용한 예측치를 제시할 계획이다.⁶⁾ 연구의 세부 목표를 다음과 같이 보다 구체적으로 설명할 수 있다.

첫째, 국내관광 수요예측의 선행연구를 체계적으로 정리한다. 국제관광 수요의 주요 변인에 대해서는 검토되었으나(이강욱, 2017), 국내관광 수요는 그에 상응하는 연구가 존재하지 않았다. 또한 수요예측 모형 역시 국제관광에서 검토된 바(이강욱, 1999; 전효재, 2003; 이강욱 외, 2009)와 달리 충분한 연구가 이루어지지 못하였다. 이러한 현상은 비단 우리 연구원만의 문제는 아니어서, 국내외의 참고할만한 선행연구는 상당히 부족한 것이 현실이다. 그렇기 때문에 국내의 문헌은 국내 특정 지역에 대한 예측 연구도 일부 문헌조사에서 같이 검토하도록 한다.

그리고 ‘국내관광 수요예측’이라는 용어가 포괄하는 범위는 매우 넓다. 관점에 따라서는 국내 특정 지역이나 시설에 대한 예측 연구도 ‘국내관광 수요예측’ 연구로 볼 수 있다. 하지만 국가의 관광 총량 예측과 이러한 연구들을 동일한 연장선상에서 모두 검토하는 것은 어렵다고 생각한다.

6) 본 연구에서 ‘관광수요’란 관광시장에서 관측된 ‘균형거래량’을 의미하고, 그 척도로 국내여행 횟수를 사용하며, 이는 다른 관광수요 연구에서도 자주 척도로 사용되고 있다(전효재·김하니, 2020; Song et al., 2010).

총량이 아니라 매우 세부적인 시설이 수요예측으로 들어가면 고려하여야 하는 요소가 더욱 많아지게 된다. 예를 들자면 다음과 같은 질문에 어떠한 답변이 나오는지 에 따라서 관광 시설의 수요예측 결과는 크게 달라질 수 있다. 관광시설 근처에 고속도로 톨게이트가 있는가? 고속철도 역사가 있는가? 대도시권과의 거리는 어떤가? 인근에 충분한 숙박시설이 있는가?

그러나 국가 전체적인 차원에서는 세부적인 교통시설이나 대도시권과의 거리는 상대적으로 국내관광 예측에서 중요한 요인이 아닐 수도 있다. 현재 우리나라의 인구는 저출산·고령화라는 용어에서 나타나는 것처럼 자연적인 인구 감소를 걱정하여야 할 시점에 도달하였으며, 이는 인구 총량은 국내관광 총량에 대하여 충분한 설명력을 가지지 못할 가능성이 있음을 시사한다. 현재 국민여행조사상의 국내관광은 지속적인 증가 추세를 보이기 때문이다. 하지만 특정 관광 지점의 관광객을 예측한다면, 주요 대도시권과의 거리는 여전히 상당한 중요성을 가지는 요소라고 생각한다.

이처럼 예측하고자 하는 대상에 따라서 검토하여야 하는 요소가 크게 상이할 수 있기 때문에, 선행연구의 검토는 본 보고서에서 다루고자 하는 명확한 “국내관광”의 개념 범주 내에서 수행되는 것이 타당하다. 이에 대해서는 보고서의 다음 장에서 더 구체적으로 논의한다.

둘째, 본 보고서는 1년 이상 수 년 이내의 중장기 예측 모형 수립을 목표로 한다.⁷⁾ 단기와 장기를 나누는 명확한 기준은 없으나, 경제학에서는 보통 생산에 필요한 모든 투입 요소들이 가변적인 상황을 장기라고 간주하곤 한다. 1년 미만의 기간은 모든 관광 서비스를 생산하는데 필요한 모든 투입 요소들이 가변적이라고 간주하기에는 충분히 길지 않다고 보이고 따라서 단기라고 부를 수 있을 것이다. 단기 예측의 경우는 과거 사용하였던 시계열 모형을 사용하거나, 본 연구원에서 외국인 관광객의 예측에 사용하는 통계적 학습(statistical learning) 기법을 적용할 여지도 존재한다. 그러나 국내관광의 중장기 예측에 대해서는 아직까지 충분하게 연구된 바가 없는 것으로 보인다.⁸⁾

7) 단기와 중기, 장기의 구분은 관점이나 분야마다 상이할 수 있다. 예를 들어, 수 십 년 미래의 수요를 예측하여야 하는 교통 분야에서는 수 년 앞의 예측도 단기이거나 중기일 수 있다.

8) 반면에 국제관광의 중장단기 전망은 여러 번에 걸쳐서 지속적으로 이루어졌다(이강욱, 1999; 전효재, 2003; 이강욱 외, 2009).

예측 모형은 우리가 예측하고자 하는 미래가 얼마나 멀리 떨어져 있는지와 밀접한 관련성이 있다. 보통 추가적인 설명변수 없이 순수하게 시계열적인 방법론만을 사용하는 모형의 경우에는 단기 예측에 더 적합하다고 여겨지고 있다. 또한 한국처럼 경제가 빠르게 성장하고 인구 고령화가 급속하게 일어나는 국가에서 과거의 정보에만 의존하는 시계열 모형은 적절한 미래의 예측을 수행하는데 적합하지 않을 수 있다.⁹⁾ 우리나라의 국내관광이 과거와 유사한 패턴으로 성장하지 않을 가능성이 상당히 높기 때문이다. 따라서 본 보고서는 국내총생산과 같은 경제 변수들을 설명 변수로 고려하는 모형을 수립하고자 한다. 앞서와 마찬가지로 구체적인 내용은 보고서의 다음 장에서 다룰 것이다.

셋째, 수립한 모형을 바탕으로 우리나라의 국내관광 예측치를 제시한다. 보통 모형을 회귀분석하여 얻는 결과물은 예측치가 아니라 추정치이며, 예측은 모형의 추정 결과에 추가적인 절차들을 적용하여 얻게 된다. 만약 정책 입안자가 어떠한 정책을 시행하였을 때의 효과를 원한다면, 이때는 예측이 아니라 추정을 하는 연구가 필요하다.¹⁰⁾ 정책 입안자가 현재 주어진 경제상황이나 전망 아래에서, 내년도에 얼마나 국내관광이 늘어날지 궁금하다면 예측치가 필요한 것이라고 볼 수 있다. 즉 정책 입안자가 필요한 사항에 따라 최종 결과물은 달라진다.

풀어서 설명하면, 정책 입안자가 필요한 수치가 정확한 회귀계수인지 또는 정확한 예측치인지의 문제가 된다고 할 수 있다. 정확한 회귀계수를 요구한다면, 자기상관(serial correlation)이나 내생성(endogeneity)의 문제와 같이 통계적 추론(statistical inference)과 관련한 내용적 검토가 필요할 것이다. 정확한 예측치를 요구한다면, 모형이 예측력에 대한 교차 검증(cross validation)이 중요한 문제가 된다. 그러나 예측과 추정을 구분하지 못하는 경우도 드물지 않게 볼 수 있다. 표본 내(in-sample) 결정계수와 예측력을 혼동하는 경우나 예측력에 대한 교차 검증은

9) 예를 들어, 국내관광 시계열을 월별로 나누면 수십여 관측치로 나눌 수 있기 때문에 월별 국내관광 시계열 데이터를 사용하면 예측에 충분한 시계열을 확보할 수 있고, 이를 활용하여 수 년 후를 예측하자는 주장도 생각할 수 있다. 그러나 월별 데이터로 1년 후를 예측한다면 현재로부터 12시점을 모두 예측하여야 하며, 2년 후를 예측한다면 24시점을 예측하여야 한다. 3년 후를 예측한다면 36시점을 모두 예측하여야 할 것이다. 즉, 충분한 데이터는 예측에 상당한 도움을 주지만, 정보가 제한된 상태에서 데이터의 주기만 짧아지는 것이 예측에 도움이 되는지는 논쟁의 여지가 있다.

10) 문화체육관광부의 많은 정책은 가격할인과 관련이 있다. 국내여행의 비용 일부를 보전해주는 사업의 효과가 궁금하다면, 국내여행의 정확한 가격 탄력성을 추정하는 연구가 필요할 것이다.

전혀 되지 않은 상태에서 자기 상관에 대하여 검토하는 경우가 그러한 예시가 될 것이다. 따라서 본 보고서는 단순히 예측한 수치만을 제시하는 것이 아니라 예측의 중간 절차에 대해서도 설명하여 국내관광을 예측하는 후속 연구에도 도움을 주고자 한다. 지금까지 설명한 연구의 목적을 정리하면 <표 1-1>과 같다.

〈표 1-1〉 연구 목적

구분		세부 목표	연구 목적
국내관광 수요의 모형화 및 예측	선행연구 검토	<ul style="list-style-type: none"> • 국내관광을 예측하기 위하여 필요한 변수, 모형에 대해 문헌조사 	<ul style="list-style-type: none"> • 시의적절하고 정확한 국내관광 예측 수행을 위한 필요 사항, • 예측 모형을 발전시키기 위한 방향을 제시하여 • 정책 요구 충족에 기여
	예측모형 제시	<ul style="list-style-type: none"> • 수 년 이내의 미래를 예측할 수 있는 중장기 예측모형 제시 	
	예측결과 제시	<ul style="list-style-type: none"> • 교차 검증을 거친 국내관광 예측 결과 제시 	

제2절 연구 범위 및 방법

1. 연구 범위

연구의 공간적 범위는 대한민국이며 시간적 범위는 국민 국내관광객 총량 데이터가 가용한 기간 전체로 한다. 내용적 범위에는 국내관광(domestic tourism) 수요 예측 문헌조사, 국내관광 예측 절차 정리, 국내관광 데이터 및 국민여행조사 데이터 설명, 예측 모형 수립 및 추정, 예측 모형의 검증이 포함된다. 지금까지 서술한 연구 범위를 정리하면 다음 <표 1-2>와 같다.

<표 1-2> 연구 범위

연구 범위	내용
공간적 범위	<ul style="list-style-type: none">• 대한민국 국내
시간적 범위	<ul style="list-style-type: none">• 관광 총량 데이터 가용 기간 전체
내용적 범위	<ul style="list-style-type: none">• 국내관광 및 예측절차 개념 정리• 국내관광 수요예측 선행연구 정리• 국민여행조사 데이터 설명• 예측 모형 수립 및 예측 모형 검증

다만 국내관광 총량 데이터는 시계열이 짧고 제한되어 있는 점을 감안하여, 시계열 단절의 문제나 국내관광의 대리변수에 대해서도 연구에서 언급하도록 한다. 다만, 완전한 대리변수의 확보 또는 국내관광을 대표하는 지수의 작성은 별도의 연구 주제이므로, 본 보고서에서 완전한 형태의 대리변수나 국내관광 지수를 제시하지는 않는다.

2. 연구 방법

연구 수행 방법은 크게 세 가지로 정리할 수 있다. (1) 전통적인 문헌조사, (2) 데이터 정리, (3) 모형 수립 및 추정이다. 각각의 내용에 대해서 관련 전문가 자문도 수행하였다.

2.1. 문헌조사

우선 문헌조사는 예측 절차나 국내관광 예측을 수행한 연구를 중심으로 기술한다. 또한 이전에 기술한 바와 같이, 실제 국내관광(domestic tourism) 수요를 추정한 연구의 수가 많지 않음을 감안하여, 우리나라의 문헌에 한정하여 주요 관광 지역 및 지점의 예측을 수행한 연구도 포함한다. 다만, 방법론적으로 국내관광 총량의 예측과 유사성이 있는 연구로 한정한다.

2.2. 데이터 정리

국내관광 데이터 정리는 출입국통계에 의해 정확하게 측정되고 있는 국제관광과 달리, 설문조사를 통해 간접적으로 추정되고 있다. 따라서 조사 인원, 조사 방식 등이 시기별로 상이하여, 이를 연혁과 함께 정리한다.¹¹⁾ 최종적인 데이터는 국내총생산(GDP) 등 경제 변수의 발표 주기에 맞추어 분기별 데이터로 가공한다.

2.3. 모형 수립 및 추정

모형 수립 및 추정은 모수적인 함수 관계를 선형 회귀 분석(linear regression)으로 추정하고자 한다. 관광과 관련성이 높은 항공·도로항만 등의 교통 분야에서도 이를 사용하고 있으며, 많은 지침이나 연구에서 시계열 분석 보다는 공신력 있는 기관의 예측치를 활용할 수 있는 설명변수를 사용하여 회귀모형을 수립하도록 권고하고 있다. 설명변수로 주로 고려하는 변수는 국내총생산으로, 소비가 소득의 함수

11) 예를 들어, 2018년 기준으로 국민여행조사는 조사주기가 반기에서 월별로, 월별 조사인원도 2,000명으로 크게 확대하는 등 많은 변경점이 발생하였다.

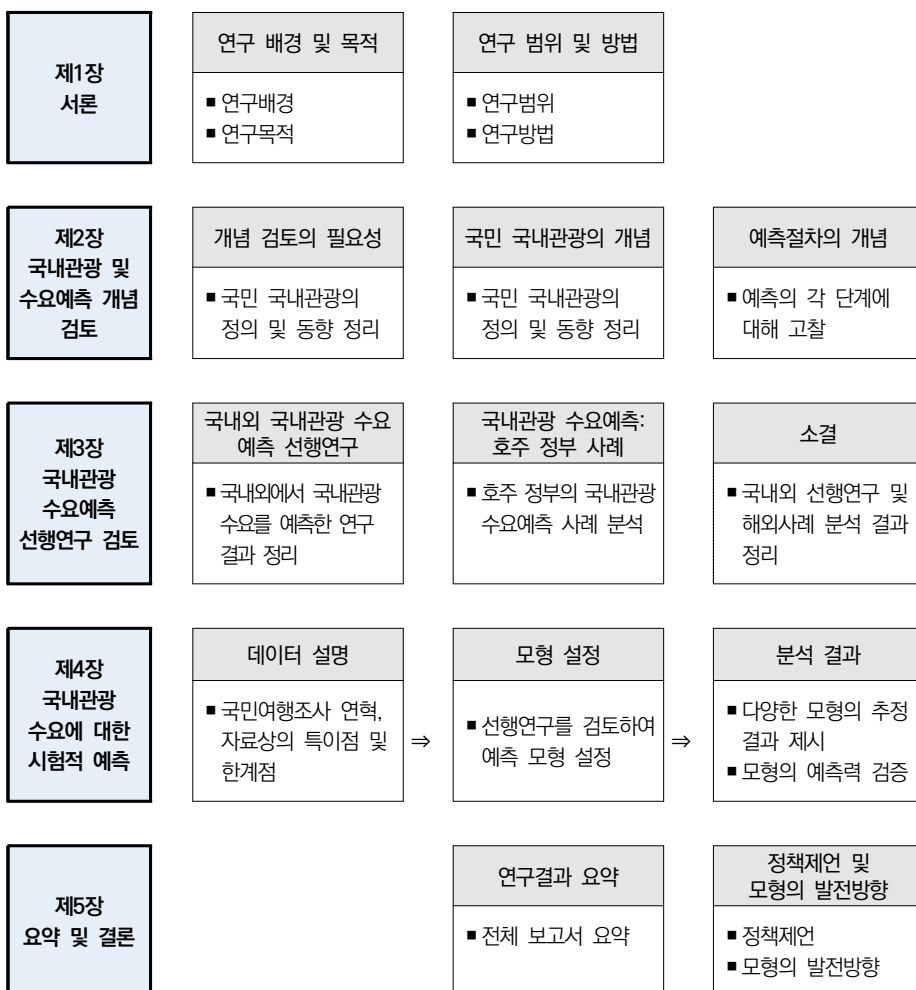
라는 이론적 근거가 명확하며 실무적으로도 널리 사용되는 변수이다.

특히, 한국은행, 한국개발연구원 등 신뢰성 높은 기관에서 예측치를 주기적으로 발표한다는 큰 장점이 있다. 그 외에도 경제 변화를 잘 설명할 수 있는 변수나 그러한 변수들의 교호작용항(interaction term)을 모형에 포함하였을 경우의 변화도 살펴볼 것이다.

3. 연구 흐름도

보고서 전체의 구성을 다음의 연구 흐름도로 정리할 수 있다.

[그림 1-1] 연구 흐름도



국민 국내관광 수요의 모형화 및 예측연구

제2장

국내관광 및
수요예측 개념 검토

제1절 개념 검토의 필요성

연구를 본격적으로 진행하기에 앞서 사전적으로 두 가지 개념을 검토할 필요가 있다. 첫째, 본 보고서의 제목이기도 한 ‘국민 국내관광’에 대한 개념 검토가 필요하다. 국내관광은 일반적으로 우리 국민이 국내 지역을 관광하는 활동으로 이해되나, ‘인바운드 관광(inbound tourism)’을 ‘외국인 국내관광’으로 지칭하는 경우도 존재한다. 따라서 두 개념의 명확한 구분을 위하여 ‘국민 국내관광’이라는 용어를 사용하였으며 이에 대한 구체적인 설명은 다음 절에서 제시하도록 한다.¹²⁾

둘째, 예측 및 예측절차에 대한 개념 검토가 필요하다. 본 보고서에서 다루는 예측은 정성적인 예측에서도 통계적 예측(statistical forecasting)이기 때문에, 통계적인 예측에 대해서 개념적인 지식이 충분하지 못한 경우에는 보고서 전체의 논리나 결과를 이해하는데 상당한 오해나 혼동의 소지가 존재한다. 모형의 설명력과 예측력을 혼동하는 경우가 그러한 사례이다. 하지만 상기 예시의 논리를 이해하기 위해서 ‘설명력’과 ‘예측력’이라는 개념에 대해 알고 있어야 하는 것처럼, 예측 문제의 이해를 위해서는 여러 개념들에 대한 지식이 요구된다.¹³⁾ 따라서 이러한 잠재적 우

12) 원론적으로는 국내관광(domestic tourism)과 국민의 정의를 각각 열거하고 어떠한 관련성이 있는지, 연구 대상을 ‘국민 국내관광’으로 지칭하는 것이 타당한지 검토하여야 할 것이다. 더 멀리 나가면 ‘nation’이 우리 법률상의 국민에 해당하는 개념인지에 대하여도 검토하여야 할 것으로 보인다. 그렇지만 이러한 용어가 관행적으로 사용되어 왔으며, 오랜 기간 관광 분야에 종사하여온 전문가들이 ‘외국인 국내관광’과 ‘국민 국내관광’이라는 용어에 친근감을 느낀다는 점을 감안하여 본 연구도 ‘국민 국내관광’이라는 용어를 사용할 것이다. 그러나 엄격하게는 ‘국민’이라는 표현은 배제하는 것이 더 합당하다고 판단된다. 지금까지는 우리나라가 법률상 국민외의 거주자가 크지 않았기 때문에 국내관광에 ‘국민’이라는 표현을 추가하여도 ‘국내 거주자’와 크게 상이하지 않았고, 따라서 그와 같은 표현이 사용될 수 있는 지점이 존재하였다고 보인다.

13) 이러한 개념 설명이 필요하지 않다는 의견도 존재한다. 그러나 모든 독자가 보고서에서 다루는 여러 정의나 개념에 대한 지식을 충분히 보유하였다고 가정하여 기술하기보다는, 어느 정도의 개념 설명을 같이 제시하는 편이 보고서의 이해에 더 도움이 된다고 판단한다. 또한 다수의 독자가 그러한 정보를 알고 있는 것이라는 주장도 제기되었으나 이는 너무 주관적인 주장으로 보인다. 만약 여기서 제시하는 개념 설명이 보편지식이기 때문에 필요가 없다면, 본 연구가 수행되어야 하는 배경이 취약해진다. 즉, 통계적 예측이 보편지식이라는 주장과 궤를 같이하므로 너무 과도한 주장으로 보인다.

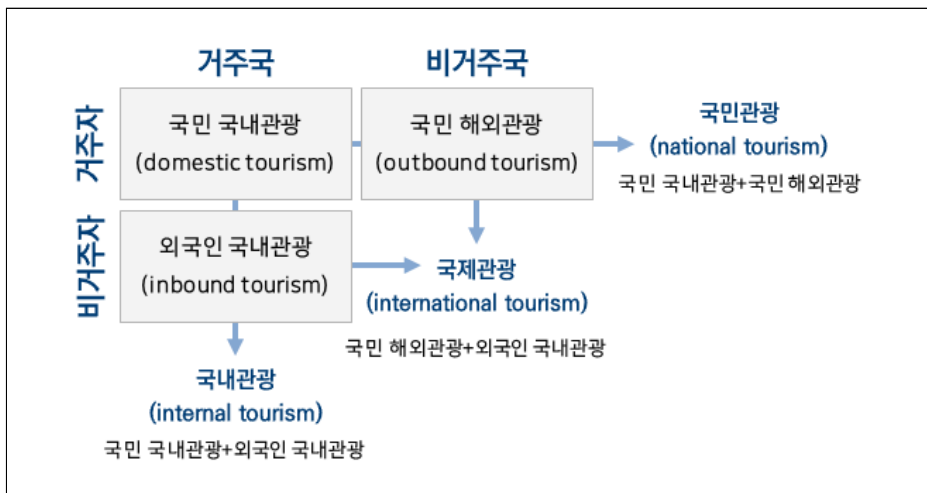
려를 불식시키기 위해서는 별도로 예측 및 예측절차의 개념을 검토할 필요성이 있다.¹⁴⁾ 상기한 필요성에 따라 다음의 각 절에서는 국민 국내관광과 예측 및 예측절차의 개념에 대하여 다루도록 한다.

14) 통계적 예측은 통계적 추정(statistical estimation)과 상당히 밀접한 관련성이 있지만 동일한 개념이 아니며 구분 가능하다. 그렇기 때문에 학술적 또는 실무적으로 실증분석 경험이 상당하다고 하더라도, 그러한 경험으로 직접 예측을 수행하기는 어렵다. 추가적으로 요구되는 지식이 존재하기 때문이다.

제2절 국민 국내관광의 개념

세계관광기구(UNWTO)는 국적과 영토에 따라 관광을 (1) 국내관광(domestic tourism), (2) 인바운드관광(inbound tourism), (3) 아웃바운드관광(outbound tourism), (4) 국민관광(national tourism), (5) 국제관광(international tourism) 등 5가지 유형으로 구분하고 있는데 국내관광은 내국인의 국내여행으로 정의된다(이연택, 2015). 한편, 관광통계국제권고안(IRT S 2008)은 거주국과 여행지를 바탕으로 관광을 6가지 유형으로 구분하고 있는데 그 내용은 [그림 2-1]과 같다.

[그림 2-1] 관광 유형 도식화



자료: United Nations (2010).

관광통계국제권고안(IRT S 2008)에 따르면 국민 국내관광은 내국인 국내여행(domestic travel)의 일부로 해당국가 내에서 이루어지는 거주 방문자(resident visitor)의 활동으로 구성되는데, 일상 거주지와 직장, 학교 등을 정기적으로 통근하거나 통학하는 여행자는 제외된다. 또한, 친지, 쇼핑센터, 종교시설, 의료시설 및 기

타시설 등 상당히 원거리에 위치하였거나 행정구역이 다른 장소를 정기적으로 혹은 빈번히 방문하는 여행자가 제외된다.

〈표 2-1〉 관광통계국제권고안(ITS 2008) 용어 제개념

구분	내용
여행(travel)	여행자(travelers)의 활동을 의미함
방문(visit)	여행(travel)의 일부로 숙박방문자(overnight visitors 또는 tourists)와 당일 방문자(same day visitors 또는 excursionists)를 포함하는 방문자(visitor)의 활동을 의미함
관광(tourism)	방문자(visitor)의 활동을 의미함
여행자(traveler)	일정기간동안 특정 목적을 위하여 서로 다른 지리적 장소 사이를 이동하는 사람을 의미함
방문자(visitor)	사업, 여가 및 기타 개인적 목적 등 특정 목적을 위해 일상환경(usual environment) 밖에 있는 주 목적지로 1년 미만의 여행(trip)을 가는 여행자(traveler)를 의미함
관광자(tourist)	숙박을 포함하고 있는 내외국인 국내방문자 및 내국인 해외방문자를 의미함

자료: United Nations (2010)를 바탕으로 연구자가 정리하였다.

국민여행조사는 국내여행 공간의 범위를 “여행(또는 관광)은 행정구역상 현 거주지(일상생활권)을 벗어나 다른 지역을 다녀온 모든 여행(숙박 및 당일 여행)”으로 정의하고 있다. 다만 행정구역을 벗어난 경우라도 단순 식사, 영화감상, 쇼핑만을 목적으로 하거나, 직장이나 거주지 근처의 산책이나 등산, 취미, 교양생활, 스포츠 등의 일상적이며 규칙적인 여가활동 등은 국내여행에서 제외하고 있어 공간의 범위만이 아니라 활동의 범위에 따라 국민 국내관광을 정의하고 있다. 국민여행조사의 기준은 앞서 살펴본 국제적으로 통용되는 기준이 어느 정도 반영된 결과이다.

본 연구에서도 국민여행조사의 국내여행을 ‘국민 국내관광’으로 사용할 것이다.¹⁵⁾ 따라서 본 연구에서 국민 국내관광(domestic tourism)은 내국민이 자신의 국가 영토 내에서의 여행으로서 행정구역을 기준으로 자신의 거주지에서 벗어나 여가, 위락, 휴가, 성지순례, 건강, 치료 등의 목적으로 타 지역으로의 이동을 의미한다.¹⁶⁾

15) 물론 엄격하게는 관광과 여행은 다르지만, 매우 높은 수준으로 중복되는 동시에 관련성이 있다.

16) 별도로 국민이라는 용어를 추가한 이유는 한국어로는 내국민의 국내관광과 외국인의 국내관광, 즉 인바운드 관광(inbound tourism)이 혼동될 우려가 있다는 연구심의회 의견에 따른 것이다. 물론 국내관광이 반드시 국민으로 이루어지는 것은 아니지만, 국제연합(United Nations, 2010)의 예시에서도 볼 수 있듯, 관행적으로 “국민”이라는 단어와의 결합은 허용되고 있다. 연구에서 사용하는 국민여행조사의 모집단도 최근 ‘국내 거주 가구’에서 ‘국민’으로 변경되었음을 언급하여 둔다.

제3절 예측절차의 개념

1. 예측절차 개요

본 절에서는 예측절차의 개념을 제시한다. 본 보고서의 기술방식이나 논리를 이해하기 위해서 예측절차의 이해는 매우 중요한 역할을 담당하기 때문이다. 독자가 통계적 추정에 상당한 지식을 가지고 있거나 익숙하다면, 상대적으로 예측 이해에 도움이 되는 측면도 존재한다.

‘예측(forecasting)’은 ‘추정(estimation)’에 기반하고 있지만, 추가적인 절차들이 필요하기 때문이다. 그러한 절차의 대표 격으로 통계적 모형 검증(statistical model validation)을 들 수 있다. 모형 검증과 같은 추가적인 절차의 필요성에 명확하게 인식하지 못하는 경우, 표본내 적합도(in-sample fitness)와 표본외(out-of-sample evaluation) 평가의 차이에 대해서 이해하지 못하고, 다양한 방법론의 결정계수만을 비교하여 모형을 선정하는 일이 발생하기도 한다.¹⁷⁾ 이러한 방식으로 선정된 모형은 예측력을 담보하기 어렵다.

[그림 2-2] 예측 절차



그러나 예측절차에 대해 알지 못하면 표본내 적합도에 근거해서 선정한 모형이 예측력이 없을 수도 있다는 사실을 인지하지 못할 수 있으며, 이는 본 보고서에서 제시하는 전체적인 논지나 결과를 이해하는데 장애요인으로 작용할 수도 있다. 따

17) 즉, 모형의 설명력과 예측력은 구분 가능한 개념이고, 실무적으로도 상이한 경우가 많다. 그러나 두 개념이 동일하다고 오해하는 경우는 드물지 않게 발생한다.

라서 본 절에서 수요예측의 절차를 간략하나마 체계적으로 설명하는 것이 보고서의 기술내용과 논리를 이해하는데 중요하게 작용할 것으로 판단한다.

전체 예측절차를 간략하게 그림으로 요약하면 [그림 2-2]와 같다. 다만 자료 갱신 (updating data)은 동일한 과정의 반복이므로 별도로 설명하지 않으며, 모형의 추정 방법론은 회귀 분석(regression)으로 한정한다. 회귀분석은 많은 분석기법을 그 안에 포함할 수 있는 포괄적 용어이며, 본 보고서에서 나중에 제시할 문헌조사에서도 가장 많이 활용되는 방법으로 나타났기 때문이다.¹⁸⁾

2. 문제 정의

문제의 정의를 별도로 서술하는 이유는 이후 우리가 자료를 어떠한 형식과 방법으로 수집하는지에 큰 영향을 주기 때문이다. 1달 후를 예측하는 모형을 위해서 1년 단위로 집계된 데이터를 이용할 것인가? 10년 후를 예측하는 모형을 수립하기 위해 1개월 단위로 집계된 데이터를 이용하는 것은 타당할까? 물음에 대한 답은 예측하고자 하는 바에 따라서 상이할 수 있다.

나중에 추가적으로 설명하겠지만, 우리나라의 국내관광 수요예측에 있어 최대의 난관으로 생각되는 것은 장기 시계열의 부재이다. ‘모형 검증’과 같은 절차를 위해 시계열의 일부를 남겨둘 필요가 있기 때문에, 모형의 추정에 사용할 데이터는 가용한 데이터보다 작은 것이 일반적이다. 이러한 점에 대해서는 나중에 더 자세히 설명하도록 한다.

우선 본 보고서가 수립하고자 하는 예측 모형은 1년 이상 수 년 이내의 예측을 수행하는 모형이다. 만약 데이터 주기를 1년으로 한다면 우리가 사용할 수 있는 기간은 불과 9년 미만이며, 분기 데이터를 사용한다면 36분기 미만을 사용할 수 있다. 그러나 이보다 데이터 주기를 짧게 하는 것이 장기 시계열 부재의 근본적인 해결책은 되지 못한다.

18) 자기회귀 누적 이동평균(autoregressive integrated moving average; ARIMA), 공적분 회귀분석(cointegration regression) 등과 회귀분석을 다른 기법인양 소개하는 경우도 많다. 사실 회귀분석은 상기한 기법을 포괄할 수 있는 보다 넓은 개념의 용어이다. 예를 들어, 선형추세식을 최소자승법으로 추정했다고 하자. 이 방법은 당연히 회귀분석이다. 비선형 추세식을 로지스틱 회귀분석으로 추정했다면 이 역시 회귀분석이다. 이러한 방법론들은 강한 가정을 사용하는 모수적 회귀분석으로 간주할 수 있다.

우리의 문제는 “수 년 이내”를 예측할 수 있는 모형을 제시하는 것이다. 만약 데이터 주기를 월로 설정하는 경우에 3년 미래를 예측하여야 한다면 현재 시점으로부터 36시점 이후를 예측하여야 한다. 실제 시간으로는 먼 미래가 아닐 수도 있지만, 모형에서 예측하고자 하는 미래의 시점으로는 상당한 미래라고 볼 수도 있다. 따라서 목표 달성에 적합한 데이터 주기는 분기로 보인다.

이처럼 예측하고자 문제가 정의되면 그에 합당한 데이터 역시 요구된다. 예측하기에 충분한 데이터가 주어져 있지 않다면, 데이터의 주기를 나누는 방식으로 근본적인 문제가 해결되지 않는다는 사실을 인지할 필요가 있다.

3. 자료 수집

“수 년 이내”의 예측으로 문제를 정의하였기 때문에, 그러한 문제의 해결에 적합한 데이터를 수집하여야 한다. 그러나 국민여행조사는 조사기관, 조사방법에 잦은 변경이 발생하여 시계열이 단절되는 문제가 있다. <표 2-2>는 국민여행조사가 지금까지 어떻게 바뀌었는지를 간단히 정리하였다.

<표 2-2> 국민여행조사 주요 연혁

연도	조사주관	조사방식 변경 개요	명칭
1976	관광공사	■ 최초 작성	전국민여행동태조사
1993	"	■ 명칭 변경 ■ 작성 주기 변경 (3년→1년)	국민여행실태조사
1998	"	■ 작성 주기 변경 (1년→2년)	"
2004	"	■ 모집단 변경 (13세 이상→15세 이상) ■ 작성 주기 변경 (2년→1년) ■ 연2회, 총12,600명 조사	"
2009	문광연	■ 자기기입 여행기록부 도입 (패널 조사)	"
2010	"	■ 조사원 방문 횟수 (연1회→연2회)	"
2011	"	■ 가계부 항목 분류 (세분류→중분류)	"
2018	"	■ 매월 2,000명, 연12회 조사 ■ 조사원 가구방문 면접 조사	국민여행조사
2019	"	■ 매월 4,000명, 연12회 조사	"

주: 문화체육관광부(2020)를 참조하였으며 "는 상동 기호(ditto mark)이다.

국내관광과 밀접한 연관이 있는 교통 분야에서는 교통 총량이 매우 정확하게 ‘측정’되고 있는 반면에, 매우 안타깝게도 국내관광은 상대적으로 자료가 부족하다.¹⁹⁾ 국내관광 총량을 지속적이고 주기적으로 추정하고 있는 조사로는 국민여행조사가 유일하다.

그러나 국민여행조사는 상당한 수준의 조사 연혁을 가지고 있음에도, <표 2-2>와 같이 조사 주체의 변경과 조사 방식의 변경 등으로 인하여 시계열 단절이 존재한다.²⁰⁾ 우선 2009년 조사를 주관하는 기관이 한국관광공사에서 한국문화관광연구원으로 변경되었고 조사 방법이나 표본에서도 큰 변화가 있었으며, 이에 따라 위탁업체도 한국리서치에서 한국갤럽으로 변경되었다. 2018년에는 조사 방법이 패널 조사에서 횡단면 조사로 다시 변경되는 변화가 있었다. 따라서 조사 방식의 큰 변화 없이 조사된 기간은 2009년부터 2017년까지이다. 다만, 2009년 국민여행조사는 관광여행에 한정하여 월간 국민여행 횟수를 제공하고 있고 우리 국민 전체의 월간 국내여행은 제시하지 않기 때문에, 별도로 가공하여 분기 데이터를 작성하기 어렵다.

따라서 본 연구에서 실제 가용한 분석 기간은 2010-2017년이다. 데이터 주기를 연도로 하면 총 9년에 불과하기 때문에 예측이 아니라 추정조차 어려운 상황이다. 그러나 분석 데이터의 주기를 분기로 구축하면 36분기의 정보가 데이터에 포함되며, 경험칙(rule of thumb)에 의하면 예측이 불가능하지는 않다. 따라서 데이터의 주기는 분기로 설정하며, 이 경우 12기간 미래는 3년 후가 되기 때문에 우리의 목표 달성에도 적합하다고 보인다.

설명변수로는 각 분기별 국내총생산(gross domestic product; GDP), 각 분기별 공휴일 수와 같이 국민의 국내여행과 밀접한 연관성이 있으며 공신력이 있는 기관에서 예측치를 발표하거나 미래의 데이터를 쉽게 계산할 수 있는 변수들을 데이터에 포함하려고 노력하였다. 변수의 설명력이 매우 높더라도, 신뢰할 수 없는 예측

19) 국내 항공 교통은 전수가 측정되고 있으며 국내 도로 교통은 최소한 고속도로 톨게이트를 지나가는 차량이 전수 측정된다. 그러나 국내관광은 측정이 아니라 조사에 의해 전체 총량을 추정해야 한다. 수요예측 과정은 많은 점에서 요리 과정과 비슷하다. 분석가나 주방장의 실력이나 성의와 관련 없이 주어진 재료의 한계를 넘을 수 없다. 주어진 재료가 쌀 뿐이라면 나올 수 있는 결과물은 밥일 것이다. 재료가 1인분만 주어졌다면 2인분을 만들 수 있는 방법은 없다.

20) 또한 한국관광공사에서 조사한 과거의 표본 자료의 존재가 불분명하기 때문에 보정을 통한 시계열 연장의 가능성도 불분명한 상황이다. 즉, 시계열 연장은 매우 불분명한 상황이며 분석에 가용한 자료는 제한되어 있다.

치를 구하기 어려운 변수라면 실무적으로는 쓸모가 떨어지기 때문이다.²¹⁾

본 보고서에서 사용하는 자료는 국가에서 공인하였거나 그에 준하는 자료를 분기별로 단순 집계하거나 결합하기 때문에, 데이터 정제외의 데이터 전처리나 로우데이터(raw data)의 처리에 대해서는 언급하지 않는다.

4. 모형화 및 추정

모형화(modeling)를 간단히 표현하면 논리나 이론에 맞추어 여러 변수 사이의 양적 관계(quantitative relationship)를 수립하는 것이다. 예를 들어, 변수 y_t 를 변수 x_t 가 설명하는 모형을 생각해보자. 즉, y_t 와 x_t 의 양적 관계를 수리적으로 표현하는 것이다. 아무런 정보가 없는 상태에서 이를 수식으로 표현하면 $y_t = f(x_t) + \epsilon_t$ 으로 쓸 수 있으며, 여기서 함수 $f(\cdot)$ 는 두 변수의 관계를 표현하는 임의의 함수이고 ϵ_t 는 백색 잡음이다. 그러나 우리가 y_t 와 x_t 에 대하여 더 많은 정보를 알고 있다고 하면, 즉 함수의 구체적인 형태가 $f(x_t) = \beta x_t$ 임을 안다면, 모형을 보다 명시적으로 쓸 수 있다. 이처럼 모수(parameter)로 표현하는 모형은 모수적 회귀분석(parametric regression)으로 추정하며, 본 보고서 역시 모수적 모형을 설정하고 추정한다.

또한 여기서 우리가 설정하고자 하는 모형은 예측 모형이라는 사실도 중요하다. 구속력이 있는 사항은 아니지만 예측에 사용되는 변수들은 공신력 있는 기관에서 발표하는 값을 차용하는 것이 일반적이다. 따라서 국내관광 총량의 예측 모형 설정에서 가장 명확하게 고려할 수 있는 것은 구매력을 대표할 수 있는 실질 국내총생산(real GDP), 국내관광 총량과 비례 관계일 것이 명확한 인구와 같은 변수들이다.²²⁾ 다시 설명하자면, x_t 의 값으로는 공신력 있는 기관에서 발표·예측하는 값을 사용하는 것이 일반적이고 바람직하다고 간주되고 있다.

21) 많은 연구에서 지적하는 바와 같이(조아라, 김형중, 2019;), 소셜 데이터나 검색 지수는 관광에 상당히 높은 설명력을 가지고 있다. 하지만 공신력 기관에서 발표하는 수치가 아니며 예측치 또한 없기 때문에 매우 가까운 미래 예측이나 현재 상태를 추정하는 외에는 실용적인 쓰임새가 떨어진다.

22) 실질 국내총생산의 예측치는 한국은행이나 한국개발연구원 등의 연구기관 예측치를 활용할 수 있고, 인구의 경우는 통계청 인구전망을 활용할 수 있다.

수립된 모형의 추정(estimation)에 회귀분석을 사용한다는 사실은 이미 기술한 바 있다. 추정은 모형 $y_t = \beta x_t + \epsilon_t$ 에서 β 의 값을 계산하는 것이다.²³⁾ 만약 β 의 값을 알고 있다면 추정은 불필요한 과정이 될 것이다. 그러나 많은 경우에 β 는 알려져 있지 않다. x_t 가 소득이고 y_t 가 국내관광이라고 하자. 아마 소득이 많으면 국내 관광을 더 많이 할 가능성이 높다. 부정하기 어려운 명제이다. 그러나 얼마나 더 많이 할까에 대한 답을 제시하기 위해서는 β 의 값을 알아야 한다. β 의 값을 구하는 과정이 추정이며, 추정을 통해 구한 β 의 값은 $\hat{\beta}$ 로 표현하고 추정치라 부른다. 많은 통계적 검정(statistical testing)은 예측의 정확성이 아니라 추정된 계수나 통계적 추론의 정확성을 검정하며, 그렇기 때문에 그러한 통계적 검정의 여부가 예측력을 향상시키지는 않는다.

예측(forecasting)은 미래의 y_{t+1} 을 구하는 것이다. 미래의 값을 구하는 기법이 꼭 통계적일 필요는 없다. 예측은 통계적인 예측과 정성적인 예측을 모두 포괄하기 때문이다. 다만 여기서는 통계적인 예측에 한정하여 설명을 진행할 것이다. 통계적 예측(statistical forecasting)을 통해 예측치를 구하기 위해서는, 모형의 예측력을 검증하는 절차도 필요하다. 모형의 예측력은 결정계수(coefficient of determination)로 대표되는 모형의 설명력과는 개념적으로 구분된다. 높은 결정계수가 예측력을 보장하지 않기 때문에, 모형의 예측력을 검증하는 추가적인 절차가 필요하다. 이러한 절차를 모형 검증이라고 한다.

모형 검증을 위해서는 모든 데이터를 사용하여 추정하지 않고 일부 데이터를 남겨둘 필요가 있다. 이는 모형 검증에서 다시 설명하며, 여기서는 예측을 위한 모형 추정에서 모든 데이터를 다 사용하지 않는 방식이 일반적이라는 사실을 주지한다.

23) 보통은 오차항 ϵ_t 에 대한 추가적인 가정이 주어진 상태에서 β 의 값을 계산한다.

5. 모형 검증

마지막으로 가장 중요한 절차인 모형 검증(model validation)에 대해서 설명한다. 여러 가지 방법이 있지만 가장 널리 사용되는 방법은 전체 데이터에서 일부만 추정(estimation)에 사용하고 나머지 데이터는 검증에 사용하는 방법이다. 이를 시계열 계량경제학에서는 보통 ‘표본외 검정’(out-of-sample test)이라는 용어를 사용하고 있지만, 인공지능 및 통계적 학습(statistical learning)의 인기에 힘입어 최근에는 ‘교차검증’(cross validation)이라는 용어가 더 널리 사용되고 익숙할 것으로 생각한다.²⁴⁾

회귀분석은 주어진 데이터에 가장 잘 맞는 선을 산출하는 방법이다.²⁵⁾ 따라서 표본내(in-sample)에서는 실제 값과 추정 값이 잘 맞는다. 하지만 그게 미래에 대한 예측력을 보장하는 것은 아니기 때문에 모형 검증이 필요한 것이다. 예측 모형의 추정결과에서 결정계수가 매우 높은 값을 가진다고 하여도, 실제 예측력은 기대보다 낮을 수 있다.²⁶⁾ 모형에서 예측된 값과 추정에 사용하지 않은 검증 데이터(validation data)를 비교하면, 모형의 예측력을 판단할 수 있다.²⁷⁾ 이러한 과정을 모형 검증이라고 부른다.

예측절차의 주요 내용, 특히 모형화 및 추정과 모형 검증의 주요 내용은 보고서의 나머지 부분을 이해하기 위해 필요하기 때문에 별도로 기억할 필요가 있다. 예를 들어, 보고서에서 결정계수가 더 높은 모형이 있는데, 그보다 낮은 결정계수를 보이

24) 둘은 개념적인 측면에서는 동일하지만 시계열 데이터의 경우에는 관행적으로 ‘표본외 검정’이라는 표현이 오랫동안 사용되어 왔으며 표현상으로도 더 적절한 면이 있다. 왜냐하면 시계열 데이터는 완전히 임의로 검정 데이터를 추출할 수 없기 때문이다. 언제나 과거의 데이터로 미래를 예측하는 형태가 되어야 한다. 미래의 데이터로 과거를 예측한다는 모형은 성립 자체가 되지 않기 때문이다. 아들이 아버지를 닮은 것이 아니라 아버지가 아들을 닮은게 아니기 때문이다. 이러한 이유로 시계열 데이터에서 말하는 ‘표본외 검정’은 교차검증을 모두 포괄하지는 못한다. 시계열 데이터에서 행해지는 특수한 형태의 교차검증이 표본외 검정이라고 생각하면 정확할 것이다.

25) 정확하게는 이차원 이상의 경우는 선이 아니라 초평면(hyper plane)이다.

26) 학습 데이터(training data)에서는 매우 높은 설명력을 가지지만, 실제 예측력이 떨어지는 문제를 과적합(overfitting)이라고 한다. 결정계수는 1에 매우 가깝지만 예측치와 실제 실현된 값의 괴리가 큰 경우가 바로 모형이 과적합된 경우이다. 예를 들어, 결정계수가 0.999인 모형을 수립하여도 바로 내일의 값조차 매우 큰 오차로 틀린다면, 그러한 모형으로 수행한 예측의 신뢰도는 매우 낮은 것이다.

27) 조금 더 설명하면, 기계 학습에서는 주어진 데이터 내에서 학습 데이터와 검증(validation) 데이터를 구분하여 학습할 수 있고, 검정(test)을 위한 데이터는 추정에 아예 활용하지 않는 점에서 표본외 검정과 교차검증이 완전히 동일한 의미를 내포하지는 않는다.

는 모형에서 예측치를 산출하는지에 대해서 의문이 발생할 수 있다. 그러한 논리를 이해하기 위해서 지금까지 설명한 예측과 추정의 차이, 설명력과 예측력의 차이, 모형 검증에 대한 정보가 필요하다. 예측절차의 주요 내용을 간략히 정리하자면 다음의 <표 2-3>과 같다.

<표 2-3> 예측절차별 주요 내용

예측절차	주요 내용	비고
문제 정의	<ul style="list-style-type: none"> ■ 예측하고자 하는 대상 구체화 ■ 목표로 하는 예측 기간 등 	-
자료 수집	<ul style="list-style-type: none"> ■ 예측 목표를 달성할 수 있는 데이터 수집 ■ 데이터 전처리 	-
모형화 및 추정	<ul style="list-style-type: none"> ■ 주어진 데이터에서 변수간의 양적 관계 추정 	<ul style="list-style-type: none"> ■ 예측을 위해서 가용한 데이터의 일부를 모형 검증을 위해 보존
모형 검증	<ul style="list-style-type: none"> ■ 검증 데이터(validation)를 사용하여 모형의 예측력 검증 	<ul style="list-style-type: none"> ■ 모형의 설명력이 모형의 예측력으로 바로 이어지는 것은 아님

국민 국내관광 수요의 모형화 및 예측연구

제3장

국내관광 수요예측
선행연구 검토

제1절 국내관광 수요예측 선행연구

1. 우리나라의 국내관광 수요예측 선행연구

본 절에서는 국내관광(domestic tourism) 수요예측 문헌을 정리하여 크게는 우리나라와 해외의 연구, 작게는 본 연구원의 정책보고서, 국내문헌, 해외문헌, 그리고 해외사례의 순서로 정리하여 제시한다.

국내관광 수요예측과 관련한 정책연구 보고서로는 이강욱(1999), 박상곤(2017a)이 있다. 이강욱(1999)은 1991년 1분기부터 1999년 2분기까지의 대미환율, 국내총생산 등의 거시경제 변수와 계절 더미를 설명변수로 설정하고 외래객수, 출국자수, 여행수지를 예측하는 모형을 수립하였으며, 동 모형을 사용하여 2000-2002년 간의 출입국객 및 여행수지를 실제로 예측하였다. 다만 국내관광 수요에 대해서는 직접적인 예측 모형이나 예측 결과를 제시하지 못하고 개념적인 예측 방향에 대해서만 언급하고 있다.

한편, 박상곤(2017a)은 국민 국내여행을 대상으로 관광 수요예측에 대한 이론 및 선행연구 검토와 계량경제학적 분석을 실시하였으며, 기술통계량 분석, 시계열 분석 등 기존 수요예측 방법론 외에 ‘여행’으로 정의되는 상품의 반복적 구매에 대한 미시적인 접근을 바탕으로 모형 구성하였다. 또한, 여행 주체들의 효용극대화의 전제 하에 미시경제학에서 이용되는 다항 로짓(multinomial logit) 모형을 이용하여 관광여행의 시기와 빈도 결정, 여행지와 여행기간의 결합 결정을 모형화 하였다. 주요 변수는 논리성과 데이터 가용성을 고려하여 소득, 가족구성원 연령 등을 고려하였다.

검토결과 기존 본원의 선행연구(이강욱, 1999; 박상곤, 2017a)가 본 연구와 유사한 주제를 다루고 있기는 하지만, 이강욱(1999)은 분석 결과 없이 국내관광 수요예측 모형에 대해서만 소략하게 설명하고 있고, 박상곤(2017a)은 수요 추정을 국민

여행조사 개편 이전의 데이터 구조에 크게 의존하는 형태로 수행하여, 국민여행조사가 크게 개편된 2018년 이후에는 적용할 수 없다는 한계점이 존재한다. 따라서 본 연구는 실제 사용할 수 있는 모형을 제시한다는 점에서 선행연구와 비교하여 상당한 차별성이 있다.

한편, 국내관광 총량에 대해 예측을 수행한 국내문헌은 매우 부족하기 때문에, 국내의 지자체나 특정 관광 지점의 수요에 대하여 예측한 연구(김상원·이덕순, 2013; 현수영 외, 2015; 김상원·박미선, 2016; 김상원, 2018)를 추가적으로 조사하였다.

김상원·이덕순(2013)은 익산시 관광객 시계열 자료를 기초로 익산시 관광객 수요를 예측하였다. 2005년 1월부터 2012년 12월까지 총 96개월의 관측치에 대하여 홀트-윈터스 지수평활법을 적용하여, 2013년 1월부터 2013년 12월의 익산시 관광객을 예측하였다.

현수영 외(2015)는 1994년 1월부터 2013년 9월 사이의 제주도 입도객 통계자료를 이용하여, 데이터 상에서 2003년 3월과 2010년 3월에 구조적 변환점이 존재함을 확인하였다. 홀트-윈터스 지수평활법과 계절형 자기회귀 누적 이동평균법(seasonal auto-regressive integrated moving average; SARIMA)을 적용한 예측 결과를 검토한 결과, 구조적 변화가 발생한 2003년 이후의 데이터만을 사용하여 'SARIMA'를 사용한 예측 모형이 제주도 수요예측에 가장 적합하였음을 보고하였다.

김상원·박미선(2016)은 004년 1월부터 2013년 12월까지의 수원시 관광객 시계열 자료를 기초로 다양한 모형의 예측 결과를 비교하였으며 이동평균법과 윈터스 가법 지수평활법이 가장 높은 예측력을 보여주었음을 보고하였다. 김상원(2018)은 2007년 1월부터 2017년 12월까지의 제주도 입도객 데이터를 다양한 모형의 예측 결과를 비교하여 윈터스 가법 지수평활법의 예측력이 가장 높다고 보고하였다.

본 연구에서 검토한 모든 선행연구는 1년 이내의 예측치를 보고하였으며, 일부 예외를 제외하면 모두 시계열 기법을 사용한 예측 모형이 가장 적합하였다고 보고하고 있다.²⁸⁾ 시계열 모형은 실무적으로 단기 예측에 주로 활용됨을 감안한다면, 본 연구에서 목표로 하는 국내관광 중장기 예측 모형에 대한 연구는 아직까지 국내

28) 일부 예외적인 연구의 결과도 어떠한 설명변수를 어떠한 형태로 사용하였는지 정확하게 보고하지 않은 사례가 많아, 실제적인 활용도가 낮다고 보인다.

에서 충분히 수행되지 않았던 것으로 보인다.²⁹⁾ 국내관광 수요 예측에 사용된 시계열 모형들은, 이미 널리 알려졌으므로 여기서는 별도로 소개하지 않는다. 지금까지의 논의를 정리하면 다음의 <표 3-1>과 같다.

<표 3-1> 우리나라의 국내관광 수요예측 선행연구

구분	활용자료(변수)	분석방법	분석기간	주기	예측범위
이강욱(1999)	출입국통계, 국민여행조사	시계열분석, 회귀분석 등	-	-	-
박상곤(2017a)	소득, 가족구성원 연령	다항 로짓 모형	2011.01-2016.12	-	-
김상원·이덕순(2013)	익산시관광객통계 (관광객수)	시계열모형 (윈터스 지수평활법)	2005.01-2012.12	월	2013.01-2013.12 (1년)
현수영 외(2015)	제주도입도객통계 (관광객수)	윈터스 지수평활법, 계절형 ARIMA	1994.01-2013.09	월	2013.10-2014.12 (1년)
김상원·박미선(2016)	수원시관광객통계 (관광객수), 시간	시계열 모형(이동평균법, 지수평활법) 인과모형(회귀모형)	2004.01-2013.12	월	2014.01-2014.12 (1년)
김상원(2018)	제주도입도객통계 (관광객수), 시간	시계열 모형(지수평활 모형: 단순, 브라운, 홀트, 윈터스 가법, 윈터스 승법), 인과 모형(회귀 모형)	2007.01-2017.12	월	2018.01-2018.12 (1년)

자료: 선행연구를 바탕으로 연구자 정리

29) 위 선행연구들은 분석의 대상을 특정 지역 및 관광지로 한정하고 있을 뿐만 아니라 국민 관광객 외에 외국인 관광객을 분석 대상에 포함하고 있어, 본 연구의 대상인 대한민국 전체의 국내관광 총량 예측과는 상이한 점이 있음을 밝힌다.

2. 외국의 국내관광 수요예측 선행연구

2.1. 개요

해외의 관광수요 연구동향은 정량적인 연구방법이 다수를 차지하고 있다. Song and Li (2008)는 2000년 이후 2007년까지 수행된 실증연구 121건을 분석하였는데, 2건을 제외한 나머지 연구에서 모두 정량적인 연구방법이 적용되었다. 특히, 예측의 정확도에 있어서 다른 모형 보다 지속적으로 우위에 있는 모형은 없음이 밝혀졌는데, 예측 결과의 조합, 정성·정량 접근법, 관광주기 및 계절성 분석 등을 통하여 예측 정확도가 향상될 수 있음을 주장하였다.

〈표 3-2〉 외국의 국민 국내관광 수요예측 선행연구

구분	활용자료	분석방법	분석기간	주기	예측범위
Witt et al. (1992)	라스베이거스 컨벤션 및 방문자 관리국에서 발행한 <i>Marketing bulletin</i> 과 1990년 <i>Ten Year Summary</i> 의 간행물	지수평활법	1973-1988	월	1989 (1년)
Athanasopoulos and Hyndman (2008)	국민방문객조사 (NVS) 방문객 숙박 수	회귀모형, 순수 시계열 모형-단일 오차 (또는 혁신) 상태 공간 모형, 외생변수 포함 혁신 상태공간모형	1998.1분기 - 2005.2분기	분기	2005-2014 (10년)
Athanasopoulos et al. (2009)	국민방문객조사 (NVS) 방문객 숙박 수	변형 하향식 접근법, 최적 결합 접근법, 지수평활법, 표본외 예측평가	1998.1분기 - 2006.4분기	분기	2007-2008 (1년)

자료: 선행연구를 바탕으로 연구자 정리

그러나 우리나라와 마찬가지로, 해외의 문헌에서도 국내관광(domestic tourism) 예측연구는 호주 및 라스베가스 등의 일부 사례를 대상으로 한정적으로 나타나고 있다. Witt et al. (1992)는 지수평활법과 나이브 모형(naive model)으로 라스베가스 방문자 예측을 수행하였다. 그 과정에서 국내관광 수요가 환율 변동 및 국제 정치 사건과 같은 외부 요인에 대해 국제관광 수요보다 덜 민감하기 때문에 변동성이

적고 따라서 국제관광 데이터에 적용되는 다양한 예측 방법이 국내관광 데이터에 적용될 때 유효하지 않을 수 있음을 주장하였다. Athanasopoulos and Hyndman (2008), Athanasopoulos et al. (2009)은 분기 데이터를 활용하여 호주 국내관광을 예측하였다. 이상의 내용은 <표 3-2>에 정리하였다. 그 외에도 호주관광연구소(TRA, Tourism Research Australia)에서 국민방문객조사(NVS, National Visitor Survey) 데이터를 활용하여 회귀모형(regression models), 혁신상태공간모형(innovations state space models with exogenous variables), 지수평활법(exponential smoothing), 연립방정식 모형(simultaneous equation model), 준이상수요체계 모형(Almost Ideal Demand System) 등 다양한 방법으로 호주 국내관광 수요 예측을 시도하고 있었으나 이는 별도의 절에서 소개하도록 한다.

2.2. 호주 국내관광 수요예측 사례

Athanasopoulos and Hyndman (2008)은 호주의 국내관광 수요예측 모형을 구축하고 수요예측을 수행하였다.³⁰⁾ Athanasopoulos and Hyndman는 전통적인 회귀모형(regression models), 혁신상태공간모형을 활용한 지수평활법(exponential smoothing via innovations state space), 외생변수를 포함하는 혁신상태공간모형(innovations state space models with exogenous variables)의 세 가지 모형을 제시하였다. 여기서는 Athanasopoulos and Hyndman (2008)의 주요 내용을 발췌하여 소개한다.

가. 회귀모형

Athanasopoulos and Hyndman (2008)은 다음과 같은 회귀모형을 제시하였다.

$$\ln VN_t^i = c + \delta \times t + \beta_1 D_{t-1} + \beta_2 P_{t-1} + \beta_3 Y_t + \beta_4 BALI_t + \beta_5 OLYMP_t + s_1 MAR_t + s_2 JUN_t + s_3 SEP_t + \varepsilon_t,$$

30) Athanasopoulos et al. (2009)은 Athanasopoulos and Hyndman (2008)을 발전시킨 후속연구로 볼 수 있으며, 연구 내용에서도 많은 부분이 중첩되므로 여기서는 생략하도록 한다.

여기서 t 는 기간(time period), $DEBT_t$ 는 1인당 실질 개인부채, D_{t-1} 는 $DEBT_t$ 의 성장률 시차변수, DPI_t 는 국내 휴일 여행 및 숙박(domestic holiday travel and accommodation)의 가격 지수(price index), P_{t-1} 은 DPI_t 성장률의 시차변수, GDP_t 는 1인당 실질 국내총생산, Y_t 는 GDP_t 성장률, $BALI_t$ 는 2002년 발리 폭탄 테러(2002 Bali bombings)의 효과를 포착하기 위한 변수로 2002년 4분기 이후로는 1 그 외에는 0의 값을 가지는 더미 변수, $OLYMP_t$ 는 2000년 시드니 올림픽의 효과를 포착하기 위한 변수로 2000년 4분기면 1 그 외에는 0의 값을 가지는 더미 변수, MAR_t , JUN_t , SEP_t 는 각각 분기 더미 변수, ε_t 는 무작위 오차항이다. i 는 여행의 목적에 따라 휴일(holiday), 친지방문(visiting friends and relatives), 업무(business), 기타(other)의 4개로 구분된다. 최종 모형의 (조정되지 않은) 결정계수는 종속 변수, 즉 여행 목적에 따라 상이하였으며 0.77~0.98 사이에 분포하고 있다.³¹⁾

나. 혁신상태공간모형을 활용한 지수평활법

지수평활법(exponential smoothing)을 개념적으로 단순하게 표현하면, 여러 가지 과거 값을 가중 평균하여 미래의 값을 예측하는 것이다. 더 과거의 값일수록 작은 가중치를 가지며, 가장 최근에 관측된 값이 가장 높은 가중치를 갖게 된다. Athanasopoulos and Hyndman (2008)은 Hyndman et al. (2002)에서 제시하는 바와 같이 일정한 제약이 가하여진 혁신상태공간모형을 사용한다.³²⁾ 이러한 모형은 최대가능도추정법(maximum likelihood estimation)으로 추정할 수 있다.³³⁾ 동 연구에서 고려한 모형은 <표 3-3>에서 요약·제시되어 있다.

31) 다만, 결정계수가 0.9를 넘는 모형은 종속변수가 휴일 여행(holiday travel)인 경우가 유일하다.

32) 혁신상태공간모형의 최적 예측치는 지수평활법으로 구한 수치와 동일함이 알려져 있다(Hyndman et al., 2005).

33) 해당 방법론은 R언어 패키지 “forecast”에 구현되어 있다(Hyndman, 2006).

〈표 3-3〉 모형 계산식

No-trend	Damped trend
$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = \ell_t + s_{t+h-m}$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = \Phi b_{t-1} + \beta \varepsilon_t$ $\hat{y}_{t+h t} = \ell_t + (1 + \Phi + \dots + \Phi^{h-1})b_t + s_{t+h-m}$

주: 표의 내용은 Athanasopoulos and Hyndman (2008, Table 4, p. 25)에서 발췌하였다. ℓ_t 는 t 시점에서 시계열의 수준을 나타내며, b_t 는 t 시점에서의 기울기, s_t 는 t 시점의 계절 성분(seasonal component)을 나타낸다. m 은 1년의 계절 수를 나타냄; $\hat{y}_{t+h|t}$ 는 t 시점까지의 모든 데이터를 기반으로 한 y_{t+h} 의 예측을 의미한다.

다. 외생변수를 포함하는 혁신상태공간모형

위에서 검토된 두 가지 모형에는 각각 장점과 단점이 있다. 우선 회귀 모형은 개인부채와 국내관광 간의 긍정적인 관계나 시드니 올림픽과 발리 폭탄 테러와 같은 중요한 국제적 이벤트의 영향을 파악할 수 있는 장점이 있다. 따라서 이러한 접근방법은 정책 입안자에게 중요한 시사점을 제안할 수 있다. 그러나 예측을 위해서는 회귀모형에 사용된 설명변수의 예측치도 필요하다는 단점도 있다.

반면, 지수평탄화와 같은 시계열 모형은 변수들 사이의 경제적 관계에 대한 시사점을 제공하지는 못하지만, 데이터의 동적 특성들을 포착하고 이를 활용하여 예측한다는 장점이 있다. Athanasopoulos and Hyndman (2008)은 (두 모형의 장점을 활용할 수 있도록) 두 모형을 결합하여, 다음과 같이 외생변수를 사용하는 혁신상태공간모형을 제시하였다.

$$\begin{aligned}
 y_t &= \ell_{t-1} + b_{t-1} + \mathbf{z}_t' \boldsymbol{\delta} + \varepsilon_t, \\
 \ell_t &= \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t, \\
 b_t &= \Phi b_{t-1} + \beta \varepsilon_t, \\
 \hat{y}_{n+h} &= \ell_n + (1 + \Phi + \dots + \Phi^{h-1})b_n + \hat{\mathbf{z}}_{n+h}' \boldsymbol{\delta},
 \end{aligned}$$

여기서 \mathbf{z}_t 는 외생변수들의 벡터이다. 앞에서 제시한 혁신상태공간모형과 여기서 제시하는 혁신상태공간모형의 차이는 평활식(smoothing equation)에서 계절 성분을 고려하느냐 외생변수를 고려하느냐이다.³⁴⁾ 따라서 모형에 포함되어야 하는 외생

34) 평활식이란 식 $y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ 을 말한다.

변수들을 식별하는 1단계 회귀를 먼저 실시한 다음, 최대가능도추정으로 외생변수가 포함된 완전한 모형을 추정하는 2단계 회귀를 수행한다.

라. 모형 평가

상기한 세 가지 모형에 대하여, 모형의 설명력을 제고하기 위해 여러 기준에 따라 표본내 평가(in-sample evaluation)를 수행하였다. 아울러, 표본내 평가, 즉 국내 관광 모형의 적합도가 예측력으로 귀결되는 것은 아님도 지적하였다. 따라서 예측 performance를 판단하기 위하여 표본외 평가(out-of-sample evaluation)도 실시하였다. 다만, 관측치의 수가 30개에 불과하기 때문에 26개의 분기를 추정에 사용하고 나머지를 모형 평가에 활용하였다.

제2절 호주 정부의 국내관광 수요예측 사례

1. 개요

관광산업은 호주에서 가장 빠르게 성장하고 있는 산업 중 하나로 2017-2018년 호주의 관광 국내총생산(GDP) 성장률은 호주 경제 전체 평균(2.8%)을 상회하는 5.0%를 기록하였다. 관광산업이 국가 핵심산업 가운데 하나로 지속 성장하기 위해서는 경제적, 환경적, 사회적 측면의 변화에 대응하기 위한 정부와 업계의 명확한 전략이 중요한데, 국제 경쟁 심화, 여행객 행태 변화, 신흥 시장에서의 새로운 기회 속에서 미래에 대한 계획 수립 및 투자 의사결정을 위한 고도의 정보력과 맞춤형 정보 확보의 필요성이 어느 때보다 강조되고 있다. 이에 따라, 호주 무역투자대표부(Austrade) 산하 호주관광연구소(Tourism Research Australia: TRA)는 국제관광 수요예측과 함께 향후 10년간의 호주 국내관광에 대한 수요예측을 실시하고 있다.³⁵⁾

〈표 3-4〉 호주 국민 국내관광 수요예측 개요

구분	주요 내용
목적	정책 입안자, 기획자, 투자자 등이 의사결정에 활용할 수 있는 기초자료를 제공
예측기간	10년
지역구분	호주 국내 전체, 각 주 및 테리토리(territory) ³⁶⁾ 별 예측
유형분류	숙박여행 및 당일여행으로 분류
예측항목	방문객수, 숙박일수, 관광지출

자료: 호주관광연구소(TRA) 홈페이지 내 국내관광 수요예측 내용을 바탕으로 연구자 재정리

35) 호주 국내관광 수요예측 사례는 호주관광연구소 홈페이지 내 국내관광 수요예측(<https://www.tra.gov.au/Domestic/domestic-tourism-forecasts>) 내용을 바탕으로 정리하였다.

36) 호주는 6개 주와 2개 테리토리(Australian Capital Territory, New South Wales, Northern Territory, Queensland, South Australia, Tasmania, Victoria)로 이루어져 있으나, 수요예측에서는 13개 지역으로 구분하였다.

2. 수요예측 데이터

호주 정부는 관광 수요예측을 위해 공항공사(Airport Coordination Australia), 통계청(Australian Bureau of Statistics), 중앙은행(Reserve Bank of Australia), 호주관광연구소(TRA) 등의 기초 통계자료를 활용하고 있다. 국민 국내관광 수요예측에 활용되는 시장정보는 글로벌 및 국내 경제 상황, 항공 수용력 및 항공운임, 국내 숙박시설 공급 및 객실료, 원천 시장에 영향을 미칠 수 있는 중요한 사건 등을 포함하고 있다.

〈표 3-5〉 호주 관광 수요예측 주요 데이터 출처

구분	주요 내용
호주공항공사	2019년 항공 스케줄
호주통계청	출입국통계 관광위성계정
인프라, 교통 및 지역경제국 (BITRE)	국내선 항공 현황 국제선 항공 현황
Consensus Economics Inc.	아시아-태평양 컨센서스 전망 (다양한 주제) 컨센서스 전망 (다양한 주제) 외환 컨센서스 전망 (다양한 주제)
재무부	예산 계획 및 전망
Deloitte Access Economics (DAE)	관광 및 호텔시장 전망
내무부	해외 입국 및 출국자료 (내부자료)
국제통화기금 (IMF)	세계 경제 전망
호주 준비 은행	-
호주관광연구소 (TRA)	외래관광객조사 국민여행조사- 2019년 3월 전국 분기별 결과 관광 투자 모니터 보고서, 2018 관광 전망 (다양한 주제)

자료: TRA (2019, p. 18).

한편, 호주관광연구소의 국가방문자조사(National Visitor Survey; NVS)는 국내관광 수요예측 핵심 데이터로서 매년 15세 이상 호주 국민 120,000명을 대상으로 전화 인터뷰 방식으로 조사하고 있다. 국가방문자조사는 국내 방문객의 여행 활동과 지출에 대한 정보를 제공하는데, 지출은 응답자의 전체 여정에 대해서만 집계되며, 개별 여행지에 대해서는 별도로 집계되지 않는다. 그러나 호주관광연구소는 방문객 활동이 특정 지역에 미치는 영향을 판단하기 위해 모형 기반의 접근법을 적용하여 방문객들의 지출을 호주 지역에 할당하고 있다.

〈표 3-6〉 호주 국가방문자조사(NVS) 개요

구분	주요 내용
개요	국가방문자조사는 1998년 호주 국민의 여행에 대한 공식적인 집계를 위해 시작되었으며, 호주의 가장 주요한 국내관광 현황을 측정하는 조사임
모집단	만 15세 이상 호주 국민, 표본크기는 연간 120,000명
조사방법	2019년부터는 100% 휴대전화 인터뷰 실시
조사항목	목적지, 목적, 교통수단, 여행 패키지 등

자료: 호주관광연구소(TRA) 홈페이지 내 국내관광 수요예측 내용을 바탕으로 연구자 재정리

국가방문자조사(NVS)의 변화 국내관광 수요예측 영향

15세 이상 호주 인구의 97%가 휴대전화를 소지함에 따라, 호주관광연구소(TRA)는 2019년부터 국가방문자조사(NVS) 샘플링을 100% 휴대전화 인터뷰로 전환하고 유선 샘플링을 중단하였다.

이러한 새로운 데이터 집합은 수요예측에 정확도를 개선할 수 있지만, 기존의 50% 휴대전화, 50% 유선으로 실시하던 조사에서의 전환 시점인 2018년 국가방문자조사(NVS) 데이터와 2019년도 데이터 간에 간극이 발생할 수밖에 없다.

이러한 문제 해결을 위해 호주관광연구소(TRA)는 국내관광 예측을 위한 새로운 방법론을 채택하였다. 이 방법론에 따라 새로운 국가방문자조사(NVS) 데이터 집합으로부터 모든 연령대에 대한 여행 성향(travel propensities)을 조사하여 도출하였다. 이를 호주 인구에 대한 인구 증가 예측치와 결합하여 국내 숙박여행 및 당일여행 증가율을 예측한다.

3. 수요예측 절차

3.1. 국민 국내관광 수요예측 절차

호주 국내관광의 관광객 활동과 지출에 대한 수요예측은 다음과 같은 방법으로 수행되고 있다.

- 계량 시계열 모형(Econometric Time Series Model)과 국내외 경제상황, 항공 수용력 및 항공요금, 국내 숙박 공급 및 숙박요금, 주요 이벤트 등에 바탕한 시장정보분석(Market Intelligence)을 함께 고려함
- 향후 가장 영향을 미칠만한 요소들에 대한 추가적인 통찰력을 얻고자 산업전망조사(industry sentiment survey) 실시
- 국내관광 수요예측(안)은 산업계 및 정부 전문가로 구성된 패널³⁷⁾들에게 검토 요청 및 피드백 수렴

3.2. 주 및 테리토리별 국내관광 수요예측 절차

호주 정부는 주 및 테리토리별 국내관광 수요예측의 수행을 위해 호주를 13개 도시 및 지역³⁸⁾으로 구분하여 수요예측치를 산출하고 있으며, 수요예측의 절차는 다음과 같다.

- 과거 방문객 숙박일수 데이터에 대한 최고치와 최저치를 조정 및 구조적 세분화를 위해 조정하여 각 지역별 예측치를 산출하고 미래 추세 예측치는 가장 최근 방문객 숙박일수 데이터를 활용하여 산출

37) 호주 정부는 예측의 정확도 제고를 위해 예측결과를 바탕으로 산업계 및 정부 전문가들로 구성된 패널(Tourism Forecasting Reference Panel)을 구성하여 국내관광 예측안에 대한 검토 및 조정기능을 수행하고 있다.

38) 13개 도시 및 지역은 다음과 같다. Sydney, regional New South Wales, Melbourne, regional Victoria, Brisbane/Gold Coast, regional Queensland, Perth, regional Western Australia, Adelaide, regional South Australia, Northern Territory, Tasmania, and the Australian Capital Territory.

- 계절성 추정치는 장기 성장 패턴에 기반하여 각 지역별로 산출
- 계절성 추정치 및 미래 추세 예측치를 결합하여 각 지역에 대한 예측치를 산출한 후 주 및 테리토리 예측으로 변환
- 주 및 테리토리별 예측치를 합산 시 국가 예측치와 일치하는지 최종 검토 진행

제3절 소결

본 장에서는 국민 국내관광(domestic tourism) 수요의 선행연구 및 호주 정부의 국내관광 수요예측 사례를 검토하였다. 우리나라에서 국가 전체의 국내관광 총량에 대해 수요 예측을 시도한 연구는 없었으며, 보고서에서 별도로 보고하지는 않았으나 대부분의 수요 예측 연구는 외국인의 방한관광(inbound tourism)을 다루고 있었다. 국내관광과 관련한 수요 예측 연구는 특정 지역에 대한 연구가 존재하였으나 방법론적으로는 거의 전적으로 시계열 분석에 의존하고 있으며 단기 예측에 치중하는 것으로 나타났다.

해외의 연구 역시 우리나라와 마찬가지로 대부분의 연구에서 인바운드 및 아웃바운드 수요 예측이 주된 관심사였으나, 예외적으로 호주에서는 국내관광 총량에 대하여 정부 및 민간 학술 차원의 연구가 꾸준히 축적되고 있는 것으로 판단된다. 특히, 호주에서는 국내관광 총량 예측을 수행하기 위하여 소득, 부채, 관광가격, 항공요금, 숙박요금과 같은 경제 변수들을 활용하여 중장기 예측을 수행하고 있어, 본 보고서가 수행하고자 하는 연구 방향에 잘 부합하는 실례를 보여주었다.

본 절에서 국내외의 선행연구를 검토한 결과 다음과 같은 시사점을 도출할 수 있었다. 첫째, 국민 국내관광 수요예측에 대한 연구는 전 세계적으로도 사례가 희소한 것으로 나타났다. 국내관광 수요예측 사례 자체가 국제적으로도 희소한 상황이기 때문에 문헌조사를 통해 예측 모형에 대한 지식을 습득하는데 명확한 한계도 존재한다. 특히, 우리나라에서는 국내관광 총량 예측에 대한 담론 자체가 전무한 상황이다. 국내관광 수요의 시범적 예측 등을 통해서 정부정책을 시의성 있게 지원할 수 있도록 선제적으로 연구를 수행해 나가야할 필요성이 있다.

둘째, 호주는 국내관광 수요예측의 예측력 제고를 위해 항공운임, 숙박요금과 같은 미시 가격정보를 풍부하게 수집하여 수요예측에 반영하고 있었다. 앞으로 우리

나라에서도 지속적으로 설명력 있는 설명변수를 발굴하는 동시에 문화체육관광부에서 수요예측에 필요한 정보에 접근할 수 있는 권한 부여가 검토되어야 한다.

셋째로 국내관광 수요예측은 외부 기관과의 협력을 통해서 수행하는 것이 바람직하다. 호주 정부는 계량 시계열 모형 외에도 공항공사, 통계청, 중앙은행, 호주관광연구소 등에서 생산하는 국내관광 관련 분석 결과를 국내관광 수요예측에 폭넓게 활용하고 있다. 또한 산업계 및 정부 전문가로 구성된 패널을 통해 예측모형을 발전시켜오고 있다. 우리나라도 교통과 같이 관광과 관련성이 높은 분야부터 정보를 상호 공유하고 발전시킬 수 있도록 협력 관계의 강화가 필요하다.

다만 본 보고서에서는 국내관광 수요의 시범적 예측을 중점적으로 논의하고 둘째, 셋째의 논의는 장기적으로 유관기관이 협의하여 실현시켜야 할 것이다. 다음 장에서는 상기의 내용을 바탕으로 국내관광 수요에 대한 시범적 예측을 수행한다.

국민 국내관광 수요의 모형화 및 예측연구

제4장

국내관광 수요에 대한
시범적 예측

제1절 시범적 예측의 의의

문헌조사 결과, 국내관광 예측에 대한 연구는 우리나라만이 아니라 전 세계적으로도 사례가 많지 않은 것으로 나타났다. 또한 일부 국내관광 예측 사례도 본 보고서의 관심사인 국내관광 총량이 아니라 특정 지역이나 특정 관광지점의 예측으로 한정되어 있다. 본 보고서와 가장 유사한 사례는 앞에서 소개한 호주 정부의 국내관광 예측이다. 따라서 우리도 장기적으로는 호주의 사례를 따라 독자적인 국내관광 예측 절차가 마련되어야 할 것으로 생각한다.

그러나 현재 우리나라에서 실제적인 국내관광 총량 예측이 수행되지 않고 있기 때문에 국내관광 예측에 어떠한 어려움이 있는지 또한 장기적으로 어떠한 점을 보완하고 수정하여야 하는지에 대한 담론 자체가 존재하지 않는다.

이에 본 장에서는 국내관광 총량의 시범적 예측을 수행하고 그 결과를 제시하고자 한다. 우선 예측에서 사용되는 데이터에 대해 설명한다. 그리고 여러 모형의 추정(estimation) 결과를 제시하고, 추정 결과를 바탕으로 해당 모형이 왜 예측에는 적합하지 않은지도 언급할 것이다. 그리고 추정 결과 상당한 예측력이 있을 것으로 보이는 일부 모형을 대상으로 모형 검증 및 예측 결과를 제시하도록 한다. 다만, 본 보고서의 예측은 충분한 선행연구나 데이터의 축적이 이루어지지 않은 상태에서 시도되는 시범적인 분석인 만큼, 보고서의 결과를 바탕으로 지속적인 후속 연구가 필요할 것으로 생각한다.

제2절 국내관광 데이터 설명

본 연구에서는 선행연구 분석 및 전문가 의견을 바탕으로 ‘국민여행조사’ 국민여행 총량 데이터를 국내관광 수요예측을 위한 기초자료로 활용하고자 하였다. 국민여행조사의 대표 국내여행 총량인 국내여행 횟수를 활용하였었고, 국내관광 수요예측 모형 구축 및 검증 과정에 활용하고자 하였다.

〈표 4-1〉 분석 대상 기초자료

구분	자료명	주요변수	구분
1	국민여행조사	국내여행 횟수	숙박여행, 당일여행, 전체

가. 개요

국민여행조사는 우리 국민의 여행실태를 종합적으로 파악하여 국가의 관광정책 수립과 연구분석 등을 위한 기초자료를 제공한다. 관광을 통한 지역발전을 도모하고, 궁극적으로 국민의 복지 및 삶의 질 제고에 기여함을 목적으로 수행되고 있으며, 국민여행에 관해 포괄적인 내용들을 담고 있다.

국민여행조사는 통계법 제18조 규정에 의하여 승인된 국가승인통계(승인번호 제 314001호)로서 1976년 한국관광공사에서 「전국민여행동태조사」라는 명칭으로 최초 작성되었으며, 1993년 「국민여행실태조사」로 조사명칭이 변경되었고 2004년부터는 조사 및 작성주기가 기존 2년에서 1년으로 변경되어 오늘날까지 이어지고 있다.

〈표 4-2〉 국민여행조사 개요

구분	주요 내용
조사근거	통계법 제18조 규정에 의하여 승인된 일반통계(승인번호 제314001호)
조사주기	매월 조사 진행
조사대상	만 15세 이상의 국민
조사방법	구조화된 설문지를 바탕으로 1:1 가구방문 면접조사
조사항목	여행총량, 여행행태, 여행 평가, 여행 제약요인, 인구통계학적 특성 등으로 구성

자료: 문화체육관광부(2020). 「2019 국민여행조사」

국민여행조사는 여행경험률, 여행 횟수, 여행 일수, 여행지출액 등 우리나라 국민의 여행행태를 나타내는 주요 4가지 지표를 국민여행 총량으로 제시하고 있는데 세부내용은 〈표 4-3〉과 같다. 2019년 우리 국민의 국내여행 경험률은 92.4%로 나타났으며, 국내여행 횟수 총량은 3억 4,475만회로 조사되었다. 또한, 국내여행 일수 총량은 5억 8,571만일로 나타났고, 국내여행 지출액 총량은 44조 1,830억원으로 확인되었다.

〈표 4-3〉 2019년 국민여행조사 주요 내용

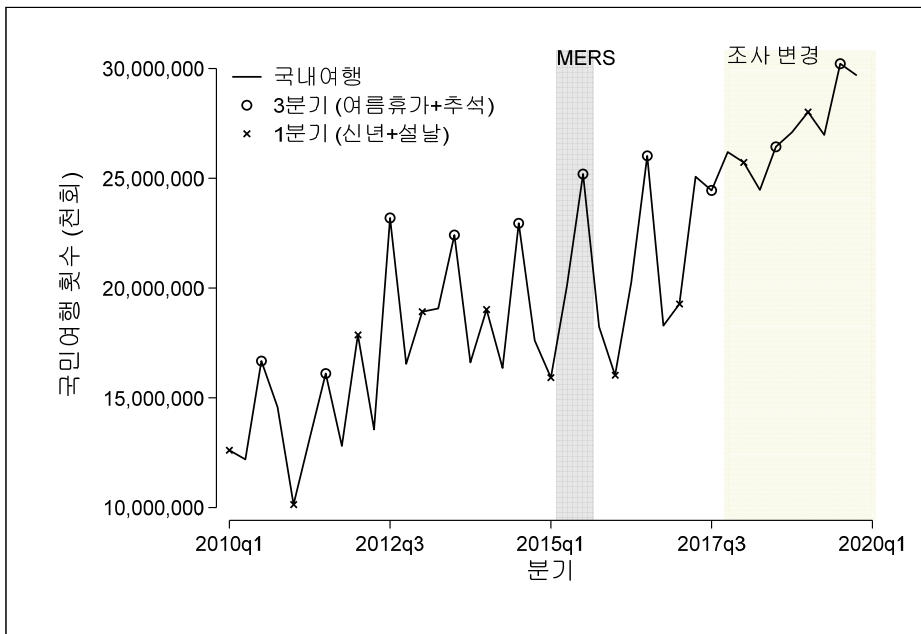
구분		내용	비고
여행경험률(%)		92.4	만 15세 이상의 전 국민이 2018년 한 해 동안 현 거주지역을 벗어나 다른 지역으로 여행을 다녀온 비중
여행 횟수 (천회)	전체	344,750	만 15세 이상의 전 국민이 2018년 한 해 동안 다녀온 국내 숙박 및 당일 여행 횟수의 총합계
	숙박	162,376	
	당일	182,374	
여행 일수 (천일)	정체	585,710	만 15세 이상의 전 국민이 2018년 한 해 동안 다녀온 국내 숙박 및 당일 여행 일수의 총합계
	숙박	403,336	
	당일	182,374	
여행 지출총액 (십억 원)	전체	44,183	만 15세 이상의 전 국민이 2018년 한 해 동안 국내 숙박 및 당일 여행을 위해 지출한 금액의 총합계
	숙박	30,041	
	당일	14,142	

자료: 문화체육관광부(2020). 「2019년 국민여행조사」

나. 주요 연혁 및 변동 사항

국민여행조사의 시작은 1976년 한국관광공사의 「전국민여행동태조사」로부터 시작되어 1993년 「국민여행실태조사」로 명칭이 변경되었다. 2004년에는 조사 모집단이 기존의 만 13세 이상 남녀에서 만 15세 이상 남녀로 변경되었으며 1년 주기 조사의 연속성이 확립되었다. 2009년부터는 작성기관이 한국문화관광연구원으로 이관되었는데 여행기록부 작성방식을 연간 수시조사로 변경하고 조사대상을 가구 및 만 15세 이상 가구원 전체로 확대하였다. 2018년에는 표본설계 변경을 통해 기존 연간 2,497가구 패넬, 6,309명을 조사하던 것을 매월 2,000명씩 표본을 추출하여 연간 24,000명을 조사하고 있다. 국민여행조사 연혁 및 변동사항은 <표 4-4>와 같다.

[그림 4-1] 국민여행조사 국민여행 횟수(2010.1분기 — 2019.4분기)



특히, 2018년에는 조사인원(표본 수) 외에도 표본틀(2010년 인구주택총조사 → 2015년 인구주택총조사), 조사 주기(반기별조사 → 월별조사), 조사 방법(여행기록부 자기기입 방식, 조사원 방문 면접조사) 등 여러 가지 변화가 있었다. [그림4-1]을 살펴보면 메르스(MERS) 발생 및 확산이 진행된 2015년 2분기에서 2015년 4분기

까지는 메르스 확산으로 인한 변화를 찾아보기 어려울 정도로 뚜렷한 계절성이 나타나고 있으나, 국민여행조사 조사 방식이 개편된 2018년 1분기 이후에는 조사방식 변화의 영향으로 국내여행 횟수에 있어서 계절성이 불분명한 것으로 확인된다. 따라서 국민여행조사 데이터의 경우 조사방식 변경 전(~2017년)의 조사 결과와 변경 후(2018년 이후)의 조사결과 간의 비교 및 자료의 활용에 있어서 분명한 한계가 존재 한다.

〈표 4-4〉 국민여행조사 연혁 및 변동사항 (2004~2008년)

구분	2004년	2005년	2006년	2007년	2008년
여행자수(명)	37,134,692	36,888,642	37,666,721	36,443,445	37,391,314
여행횟수(회)	227,537,202	257,790,190	284,575,435	302,860,566	245,669,452
여행총비용 (백만원)	14,208,206	14,511,123	16,081,318	15,757,038	15,711,047
여행이동총량(일)	358,942,843	388,836,797	416,982,061	477,372,260	408,026,189
1인평균여행횟수(회)	5.87	6.59	7.20	7.57	6.07
조사방법	횡단	횡단	횡단	횡단	횡단
표본수	12,600	12,600	12,600	26,494	12,758
방문횟수	2회	2회	2회	4회	2회
조사대상	가구원 중 1인	가구원 중 1인	가구원 중 1인	가구원 중 1인	가구원 중 1인
총량작성방법	요약표작성	요약표작성	요약표작성	요약표작성	요약표작성
여행종류 구분	국내여행	국내여행	국내여행	국내여행	국내여행
여행작성 범위	개인여행	개인여행	개인여행	개인여행	개인여행
여행공간 범위	구분없음	구분없음	구분없음	구분없음	구분없음
여행지출액 작성범위	1인여행 경비, 중분류 지출액 작성	1인여행 경비, 중분류 지출액 작성	1인여행 경비, 중분류 지출액 작성	1인여행 경비, 중분류 지출액 작성	1인여행 경비, 중분류 지출액 작성
여행행태 작성방법	매월첫번째	매월첫번째	매월첫번째	매월첫번째	매월첫번째
위탁업체	한국리서치	한국리서치	한국리서치	한국리서치	한국리서치

〈표 4-5〉 국민여행조사 연혁 및 변동사항 (2009~2013년)

구분	2009년	2010년	2011년	2012년	2013년
여행자수(명)	31,201,294	30,916,690	35,013,090	36,914,067	37,800,004
여행횟수(회)	219,585,834	168,147,987	156,594,347	213,468,294	231,034,535
여행총비용 (백만원)	16,364,538	16,859,774	20,205,430	23,891,080	23,234,315
여행이동총량(일)	375,340,664	339,607,551	286,947,961	365,282,249	389,220,312
1인평균여행횟수(회)	5.29	3.96	3.65	4.93	5.29
조사방법	패널	패널	패널	패널	패널
표본수	6,768	6,309	5,798	6,638	6,429
방문횟수	1회	2회	2회	2회	2회
조사대상	가구원모두	가구원모두	가구원모두	가구원모두	가구원모두
총량작성방법	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성
여행종류 구분	국내여행, 관광/기타로 구분	국내여행, 관광/기타로 구분	국내여행, 관광/기타로 구분	국내여행, 관광/기타로 구분	국내여행, 관광/기타로 구분
여행작성 범위	가구여행, 개인여행, 관광여행, 기타여행	가구여행, 개인여행, 관광여행, 기타여행	가구여행, 개인여행, 관광여행, 기타여행	가구여행, 개인여행, 관광여행, 기타여행	가구여행, 개인여행, 관광여행, 기타여행
여행공간 범위		거주하는 행정 구역을 벗어난 여행 광역시내 구간 이동은 제외	거주하는 행정 구역을 벗어난 여행, 시군구 포함	거주하는 행정 구역을 벗어난 여행, 시군구 포함	거주하는 행정 구역을 벗어난 여행, 시군구 포함
여행지출액 작성범위	세부 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함	세부 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함	중분류 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함	중분류 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함	중분류 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함
여행행태 작성방법	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성
위탁업체	한국갤럽	한국갤럽	한국갤럽	한국갤럽	한국갤럽

〈표 4-6〉 국민여행조사 연혁 및 변동사항 (2014~2018년)

구분	2014년	2015년	2016년	2017년	2018년
여행자수(명)	38,059,002	38,307,303	39,293,235	40,483,997	-
여행횟수(회)	227,815,219	238,297,311	241,749,955	284,966,001	311,153,000
여행총비용 (백만원)	24,938,983	25,395,649	25,748,487	29,455,943	43,133,000
여행이동총량(일)	398,308,414	406,818,700	412,378,155	479,673,688	556,890,000
1인평균여행횟수(회)	5.17	5.47	5.51	5.90	6.92
조사방법	패널	패널	패널	패널	표본
표본수	6,537	6,534	6,309	6,170	24,000
방문횟수	2회	2회	2회	2회	12회
조사대상	가구원모두	가구원모두	가구원모두	가구원모두	가구원모두
총량작성방법	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성
여행종류 구분	국내여행, 관광/기타로 구분	국내여행, 관광/기타로 구분	국내여행, 관광/기타로 구분	국내여행, 관광/기타로 구분	국내여행, 관광/기타로 구분
여행작성 범위	가구여행, 개인여행, 관광여행, 기타여행	가구여행, 개인여행, 관광여행, 기타여행	가구여행, 개인여행, 관광여행, 기타여행	가구여행, 개인여행, 관광여행, 기타여행	가구여행, 개인여행, 관광여행, 기타여행
여행공간 범위	거주하는 행정 구역을 벗어난 여행, 시군구 포함	거주하는 행정 구역을 벗어난 여행, 시군구 포함	거주하는 행정 구역을 벗어난 여행, 시군구 포함	거주하는 행정 구역을 벗어난 여행, 시군구 포함	거주하는 행정 구역을 벗어난 여행, 시군구 포함
여행지출액 작성범위	중분류 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함	중분류 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함	중분류 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함	중분류 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함	중분류 지출액 가계부 작성 본인이 직접 지출한 비용 및 타인이 본인의 경비를 부담· 지원한 비용도 포함
여행행태 작성방법	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성	매번여행 여행기록부 작성
위탁업체	한국리서치	한국리서치	한국리서치	한국리서치	케이스탯리서치

자료: 문화체육관광부(2019).

다. 통계 작성 절차

국민여행조사는 월별 단위 조사로써 매년 1월부터 12월까지 매월 조사를 진행하고 있다. 조사는 만 15세 이상의 국민을 대상으로 일대일 가구방문 면접조사³⁹⁾를 통해 진행되는데 구조화된 설문지를 Tablet PC에 탑재하여 자료를 수집한다. 자료 수집의 절차는 조사원 선발, 조사원 교육, 실사 관리, 자료 검증의 순서로 진행되는 데, 조사완료 후 조사원별로 완료된 조사의 30% 이상에 대해 전화검증을 실시하여 조사 결과의 신뢰성을 제고하고 있다.

라. 분석자료

본 연구에서는 국민 국내여행 수요예측 모형을 구축하기 위하여 국민여행조사 국내여행 총량을 활용하였다. 국내여행 총량은 국민의 여행행태를 나타내는 주요 항목 4가지를 수치로 표현한 것으로서 여행 경험률, 여행 횟수, 여행 일수, 여행 지출액 등을 포함하는 개념인데, 본 연구에서는 국내여행 총량의 주요 항목 중 여행 횟수를 국내관광 변수로 선정하였다. 여행 횟수는 ‘만 15세 이상의 전 국민이 2019년 한 해 동안 다녀온 국내 숙박 및 당일 여행 횟수의 총합계’로서 우리 국민의 국내관광 총량을 대표할 수 있는 변수이며, 관광동향에 관한 연차보고서 등 문화체육관광부의 간행물을 통해 국내여행 총량 지표로 활용되고 있어 국내관광 변수로서 대표성을 인정받고 있다. 시계열의 분석범위는 2010년 1월부터 2019년 12월 까지 총 10년간으로 한정하였는데, 국민여행조사 총량작성방법이 현재와 같이 ‘매년 여행기록부 작성 방식’으로 변경된 시점을 기준으로 하여 시계열의 연속성을 고려하였다. 다만, 2009년 국민여행조사는 관광여행에 한정하여 월간 국민여행 횟수를 제공하고 있으므로 본 연구의 분석 대상에서 제외하였다. 데이터의 단위는 월간자료로 분석범위 내 120개 시계열을 포함하고 있다.

〈표 4-7〉 국민여행조사 분석자료 개요

활용자료	분석범위	데이터 단위	시계열 수
국민여행조사 여행 횟수	2010.01. ~ 2019.12.	월간	120개

39) 1:1 가구방문 면접조사는 조사원이 조사대상 가구를 방문한 후 적격 대상자 1명을 선정하여 조사내용을 응답 받는 방식을 의미한다.

제3절 모형 설정 및 추정 결과

1. 모형 설정

문헌조사에서 살펴본 바와 같이, 국내관광(domestic tourism) 예측 연구가 활발하게 이루어지는 사례는 호주가 거의 유일하다. 따라서 본 보고서는 소수의 해외 사례와 국내 교통 분야의 예측에서 사용된 주요 설명 변수들을 고려하였다. 모형 설정은 동일한 설명력을 가지면 모수(parameter)가 작을수록 좋다는 모수 절약의 원칙(parsimony of parameters)에 따라서 설정하였다(Ledolter and Abraham, 1981; Box et al., 2015).

본 보고서는 변수에 자연 로그 값을 취하여 분석하도록 한다. 로그 값의 사용에 대해서는 여러 가지 장점과 이유가 존재한다(Wooldridge, 2019).⁴⁰⁾ 우선 추정된 계수를 탄력성(elasticity)으로 해석할 수 있다.⁴¹⁾ 로그 값을 취하면 변수의 변동성(variability)이 크게 감소하기 때문에, 회귀분석이 이상치(outlier)의 영향을 덜 받는다라는 장점도 있다. 또한 모형이 고전적인 동분산의 가정에 더 적합하게 된다.⁴²⁾ 비선형인 관계를 로그 변환에 의하여 선형으로 나타낼 수도 있으며, 이 경우 로그를 취한 모형도 비선형 회귀분석(nonlinear regression)으로 생각할 수도 있다

40) 로그의 사용은 상당히 일반화되었기 때문에 통계학이나 계량경제학 교과서에서 로그 값을 사용하는 이유에 대해서 별도의 절을 할애하여 설명하는 경우는 드문 편이다. Manning (1998)은 종속변수에 로그를 취하는 이유에 대하여 설명하고 있으며, 최근 많이 사용되는 교과서 중에서는 Wooldridge (2019)가 다양한 설명을 제시하고 있다. 국내관광과 관련한 연구는 아니지만, 최근에는 관광분야에서도 전효재·김하니(2020)와 같이 로그 변환의 유용성에 대해 언급하는 연구들도 존재한다.

41) 가장 간단한 형태의 로그-로그 회귀분석 $\ln y_t = \alpha + \beta \ln x_t + \epsilon_t$ 을 생각해보자. 라이프니츠 표기법으로 $\ln y_t$ 의 편미분을 나타내면 $\partial \ln y_t / \partial \ln x_t = \beta$ 와 같다. 연쇄 법칙(chain rule)에 의해 $\partial \ln y_t / \partial \ln x_t = (\partial y_t / y_t) / (\partial x_t / x_t)$ 이므로 추정된 β 를 y_t 의 x_t -연쇄탄력성으로 해석할 수 있다. Y 의 X -탄력성 정의는 $(\partial Y / Y) / (\partial X / X)$ 이며, $\partial \ln z / \partial z = 1/z$ 이다.

42) 동분산이 아닌 경우에 대해서는 일반화 최소제곱법(GLS), 강건한 표준오차(robust standard errors) 등의 해결책이 있기 때문에 동분산 가정이 분석에 필수적이지는 않다.

(Stock and Watson, 2020).

이처럼 통계학이나 계량경제학 기법을 사용하는 많은 정량연구에서 로그 변환한 수치를 사용하여 분석하고 있으며, <표 4-8>과 같이 관광 수요예측에서도 변수의 로그 값을 취하는 것이 일반적이다(Song et al., 2009).

<표 4-8> 주요 관광 수요예측 연구에서의 모형 함수 형태(functional form)

함수 형태		건수	비중
로그	로그-로그(double-log)	(69)	(59.0%)
	세미-로그(semi-log)	(20)	(17.1%)
	소 계	89	76.1%
그 외		28	23.9%

주: Song et al. (2009, pp. 14-26), 'Table 2.1'를 가공하였으며, 동일한 연구에서 다양한 함수 형태를 사용한 경우 중복하여 계산하였다.

또한 관광 분야의 데이터는 계절성이 매우 강하게 나타난다는 특성이 있기 때문에 모형에서 계절성을 어느 정도 고려할 필요성이 있다. 거시경제학 분야의 연구에서는 계절성이 조정된 시계열을 사용하거나 필터를 사용하여 이러한 계절성을 제거하는 것이 일반적이지만, 관광 분야에서는 계절성 또한 중요한 설명 요인의 하나이기 때문에 그러한 방법을 적용하는 것이 적절하지 않아 보인다. 또한 해외의 국내관광 예측에서도 더미를 사용하여 계절성을 고려하는 사례가 있음은 물론이다(Athanasopoulos and Hyndman, 2008).

하지만 더미 변수를 사용한 계절성의 반영은 모형에 매우 강한 가정을 가한다는 사실도 언급해둔다. 선형 회귀모형에서 이러한 더미의 사용은 특정 계절에는 종속 변수가 항상 특정한 만큼 증가함을 가정하는 것이다. 따라서 계절성의 적용은 계절성을 가지는 변수 자체를 사용하는 방식과 더미를 사용하는 방식 모두를 고려할 계획이다. 설날과 추석을 포함하는 공휴일 변수도 계절성을 설명할 수 있는 가능성이 있다. 주요 설명변수인 소득, 인구, 공휴일은 각각 별도의 소절에서 다룬다.

1.1. 소득

소득은 대부분의 모형에서 핵심적인 변수로 간주된다. 소득과 소비의 관계는 경제학적으로 매우 탄탄한 근거를 가지고 있다. 한 개인의 선호가 동일할 때, 소득의 증가는 예산제약선을 이동시키고 최적 소비조합은 달라진다. 우리는 전체 소비에서 이동을 수반하는 특정한 형태의 소비를 묶어 ‘관광’이라 부르기 때문에 동일한 논리가 관광 수요에도 적용된다. 다만, 본 보고서에서 소득은 실질소득을 사용할 것이다. 개인의 선택은 명목소득이 아니라 실질소득에 기반할 뿐만 아니라, 우리나라의 중앙은행인 한국은행이 발표하는 경제성장률은 실질 국내총생산 성장률이기 때문이다. 따라서 우리는 한국은행 실질 국내총생산을 사용한 모형을 수립한다.⁴³⁾

1.2. 인구

인구 역시 종속변수가 총량인 경우 충분한 설명력을 가지는 변수이다. 다만, 그 중요성은 소득에 비해 떨어지며, 소득과 동시에 고려하게 되면 많은 경우에 소득 총량과의 다중공선성(multicollinearity) 문제가 발생한다(Song et al., 2009). 또한 인구는 늘지 않는데 국내관광은 빠르게 증가한다는 점도 중요하다. 따라서 인구를 직접 고려하지 않고, 소득 총량을 인구로 나눈 1인당 소득을 사용하는 방법도 널리 사용된다. 본 보고서에서는 두 가지 방법에 대한 추정결과를 모두 제시한다.

1.3. 공휴일

대부분의 해외연구에서는 공휴일을 중요한 설명 변수로 간주하지 않고 있다. 그러나 우리나라와 주요 국가의 공휴일 체계는 동일하지 않다. 간단한 예로 미국을 생각해보자. 미국의 경우 크리스마스(12월 25일)나 독립기념일(7월 4일)과 같은 공휴일은 우리나라와 동일하게 특정 일자를 지정하고 있으나, 메모리얼 데이(5월 마지막 월요일)와 같이 특정한 달의 특정한 월요일로 지정된 공휴일도 많다. 이 경우 주말에 연이어 쉴 수 있다는 사실을 인지할 수 있기 때문에 매년 여행이나 관광 계

43) 국내관광 변수로 사용하는 국민여행조사의 여행횟수는 사업 목적의 방문(business visits)을 포함하는 개념이다. 따라서 개인의 가처분 소득이 아니라 국내총생산을 사용하는 것이 타당하다고 보인다.

획에 큰 변동성이 발생하지 않는다.

한편 우리나라는 대부분의 공휴일은 특정 일자로 지정되었으며, 설이나 추석과 같이 태음태양력(lunisolar calendar)에 따라 임의로 결정되는 공휴일도 존재한다.⁴⁴⁾ 소위 “황금연휴”라는 말은 다른 해보다 추가적인 연휴를 더 누릴 수 있음이 반영된 용어이다. 따라서 공휴일의 임의성(randomness)으로 인하여 관광·여행의 추가적인 변동성(variability)이 발생할 수도 있다. 이 경우 공휴일은 소득이나 인구가 설명하지 못하는 국내관광의 변화를 설명할 수 있다. 또한 미래의 공휴일 정보는 정확하게 계산이 가능하므로, 공휴일 수는 예측에 사용하기에 매우 적합한 변수라고 판단한다.

지금까지 설명한 주요 설명변수의 기술통계량은 다음의 <표 4-9>에서 제시한다. 또한 변수의 출처나 정의 등에 대한 구체적인 내용은 별도의 부록으로 제시한다.

<표 4-9> 기술 통계량

변수 이름	관측치 수	평균	표준편차	최솟값	최댓값
로그 국내관광	32	17.796	0.242	17.230	18.180
로그 소득 (국내총생산)	32	15.193	0.070	15.064	15.307
로그 인구	28	3.936	0.008	3.923	3.947
로그 공휴일	32	3.372	0.044	3.258	3.526

2. 추정 결과

이하 소절에서는 (1) 소득 변수만을 설명 변수로 사용하는 모형, (2) 인구 변수만을 사용하는 모형, (3) 공휴일만을 사용하는 모형의 추정 결과를 제시하고, (4) 그 외에 추가적인 분석결과들을 제시한다. 예측 절차를 설명하면서 언급한 바와 같이, 높은 결정계수 값이 높은 예측력(predictability)을 보장하지 않기 때문에 모형 추정에서는 가용한 전체 데이터를 사용하지 않고 일부 데이터를 남겨둘 필요가 있다.⁴⁵⁾

44) 보다 정확하게는, 현행 음력 절기는 청나라 시헌력(時憲曆)에 따르고 있다.

45) 왜냐하면 모형의 추정 결과가 얼마나 예측력이 있는지를 단순히 표본내(in-sample)의 적합도만으로 판단할 수 없기 때문이다.

따라서 2015년 이전의 데이터만 사용하여 추정하고, 나머지 데이터는 모형 검증에 활용하며, 주요 결과에 대해서는 시계열도(time series plot)를 제공한다.⁴⁶⁾

2.1. 소득 모형

소득과 소비의 관계는 탄탄한 이론적 근거를 가지고 있다. 소득의 변화는 예산제 약선의 이동을 의미하고 최적 소비 역시 달라질 수 있다. 관광은 소비의 일부이기 때문에 소득과 상당한 관련성이 있다. 또한 현재는 관광총량이 빠르게 증가하여 왔으나, 앞으로도 지금과 같은 추세로 성장하지는 않을 가능성이 높다. 우리나라가 과거에는 매우 빠르게 성장하였으나 현재에는 과거 수준의 경제성장률을 달성하지 못하는 것과 유사한 문제이다.⁴⁷⁾ 이러한 성장을 둔화는 근사적으로 이차함수의 형태로 나타낼 수 있기 때문에, 로그 국내총생산 및 그 제곱항을 같이 고려하는 모형도 검토한다. 일반적인 계량경제 모형과 달리 관광에서 계절성은 매우 중요하므로 분기 터미를 추가한 모형도 검토한다. 소득 모형을 수식으로 나타내면 다음과 같다.

$$\ln(\text{tourist}_t) = \alpha + \beta_1 \ln GDP_t + \beta_2 (\ln GDP_t)^2 + \sum_{s=2}^4 \delta_s q_{t,s} + \epsilon_t,$$

여기서 $\ln(\text{tourist}_t)$ 는 t 분기의 로그 관광객 총량, $\ln GDP_t$ 는 t 분기의 로그 국내총생산 총량, $q_{t,s}$ 는 t 가 s 분기면 1 아니면 0의 값을 가지는 터미변수이다. 분기는 총 4분기이나 다중공선성 문제로 3개의 터미변수를 모형에 추가하는 것이 타당한 방식이다. 여기서 모든 $q_{t,s} = 0$ 또는 β_2 의 가정을 도입하면 다음 <표 4-10>의 모든 모형이 표현된다.

<표 4-10>의 1열은 α , β_1 을 제외한 모수가 모두 영인 모형, 2열은 (1)에 비영인

46) 시계열도는 모형의 추정결과가 얼마나 적합한지 직관적으로 이해할 수 있는 장점이 있다.

47) 즉, 우리나라 국내관광 시장(domestic tourism market) 규모가 앞으로 얼마만큼 빠르게 증가하느냐, 과거와 같은 수준으로 성장하느냐 아니면 다른 어떤 값으로 수렴하게 되느냐의 논쟁이 있을 수 있다. 이러한 의문에 대한 중요한 연구로 박준용 외(2011)를 들 수 있다. 박준용 외는 우리의 전력수요 소득탄력성이 선도국가 전력수요 소득탄력성을 따라간다는 선도추급(gap and catch-up) 모형을 제시한 바 있다. 다만 박준용 외(2011)의 방법론은 시간변동계수를 가지는 공적분회귀이기 때문에 본 연구에 바로 적용할 수 없다. 하지만 관광총량의 증가는 경제성장률처럼 둔화될 가능성이 매우 높고, 따라서 이차함수로 그 움직임을 근사할 수 있다. 장기적으로는 국내관광 수요예측도 시간에 따르는 탄력성 변화를 고려할 수 있는 시간가변 모형으로 발전하여야 할 것이다.

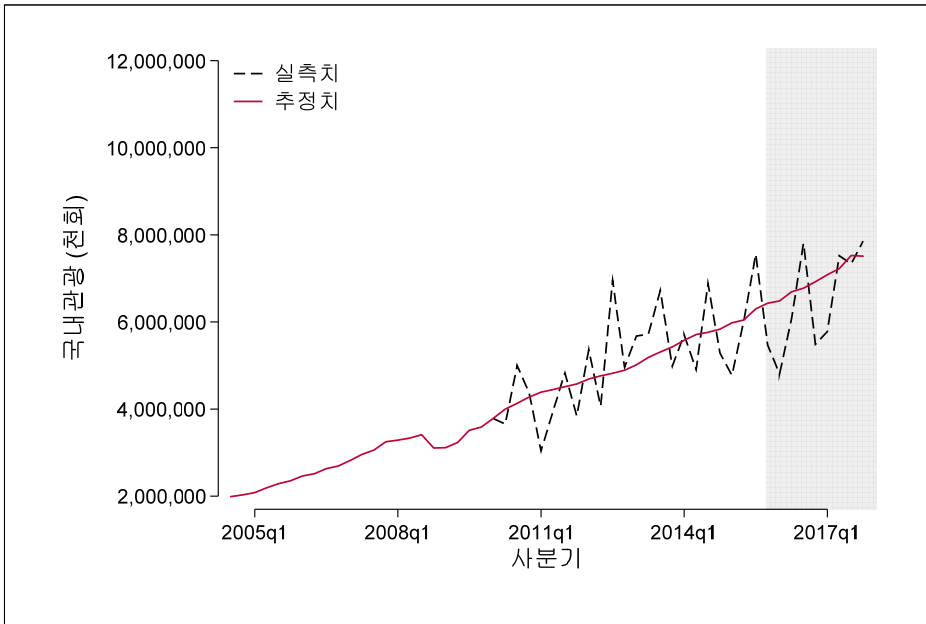
$q_{t,s}$ 를 추가한 모형, 3열은 (1)에 비영인 β_2 를 추가한 모형, 4열은 (1)에 비영인 $q_{t,s}$, β_2 를 추가한 모형이다. 다만 제곱항을 설명변수로 하는 3열, 4열의 모형에서 로그 국내총생산 및 그 제곱항이 모두 비유의하게 추정되었다. 또한 분기더미를 포함하는 경우 결정계수가 크게 상승하여, 분기더미가 모형의 설명력을 향상시킬 수 있는 것으로 나타났다. 다음으로 시계열도(time-series plot)를 사용하여 모형의 예측력을 검토해보도록 한다. 보다 엄밀한 예측력 비교는 표본외 검정을 통해서 수행되어야 하므로, 여기서는 예측력이 떨어지는 모형들을 걸러내기 위한 용도로 활용한다. [그림 4-2]는 오직 로그 국내총생산만을 설명변수로 하는 모형의 추정결과이며, [그림 4-3]은 로그 국내총생산에 분기 더미를 추가한 모형이다. [그림 4-2]에서는 계절성이 충분히 반영되지 않았다고 보이나, 분기 더미를 추가하면 주기적인 변동 추세를 보다 잘 반영함이 관찰된다. 음영 구간은 실제 추정에 사용되지 않은 ‘타당성 검사 구간’(validation periods)을 나타낸다.

〈표 4-10〉 로그 국내총생산을 설명변수로 하는 회귀모형 추정 결과

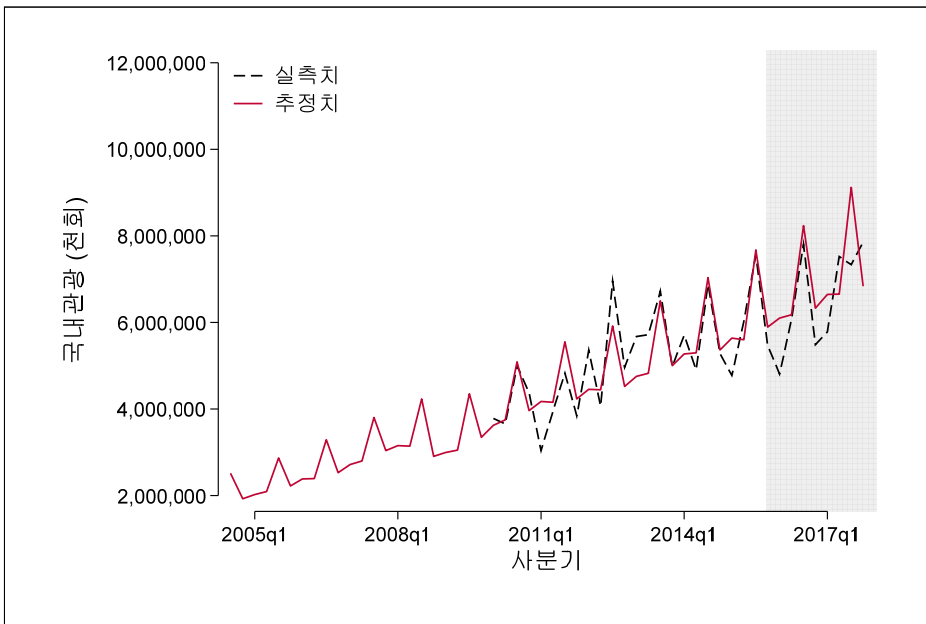
변수	(1)	(2)	(3)	(4)
로그 국내총생산	2.828*** (0.614)	2.744*** (0.410)	311.990 (345.422)	297.809 (253.067)
(로그 국내총생산) ²			-10.195 (11.399)	-9.730 (8.346)
상수항	-25.150** (9.311)	-23.930*** (6.225)	-2,368.893 (2,616.890)	-2,260.827 (1,918.276)
분기더미	No	Yes	No	Yes
관측치	24	24	24	24
결정계수	0.435	0.717	0.450	0.731

주: 괄호 안은 이분산성에 강건한 Eicker-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다.

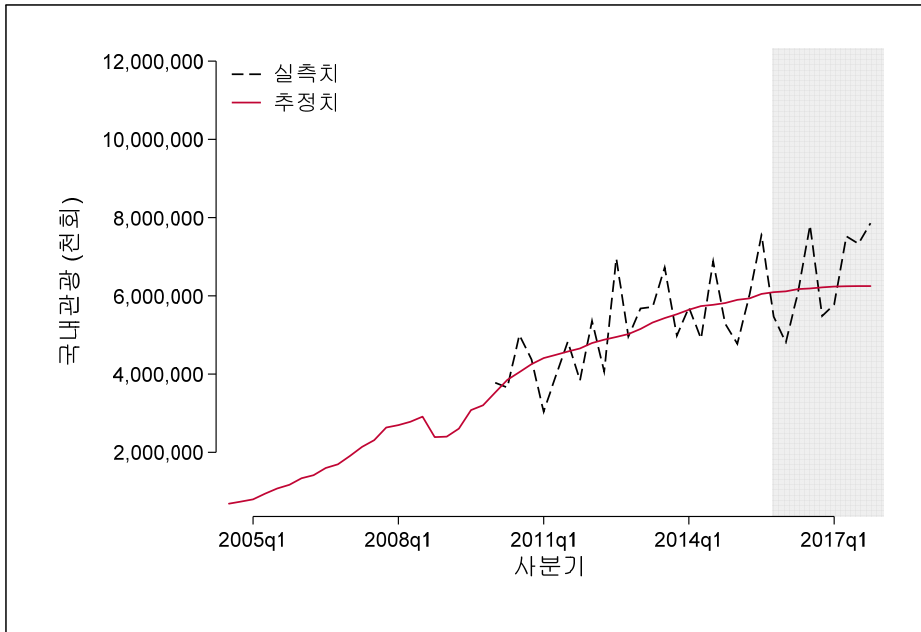
[그림 4-2] 국내관광과 <표 4-10> (1)열 모형 추정치 시계열도



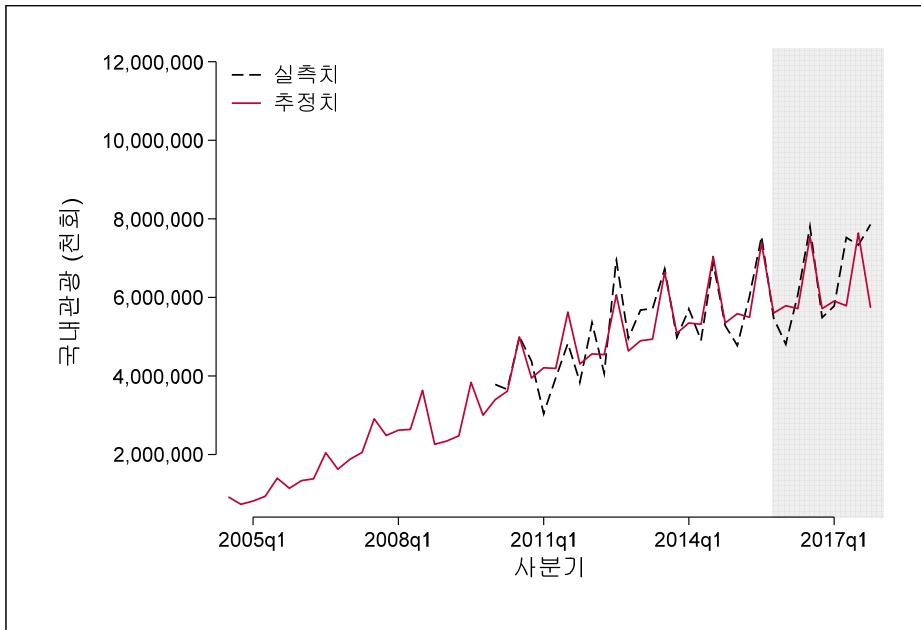
[그림 4-3] 국내관광과 <표 4-10> (2)열 모형 추정치 시계열도



[그림 4-4] 국내관광과 <표 4-10> (3)열 모형 추정치 시계열도



[그림 4-5] 국내관광과 <표 4-10>의 (4)열 추정치 시계열도



[그림 4-4], [그림 4-5]는 로그 국내총생산을 이차함수의 형태로 모형에 반영한 결과이다. 목적에 의거하면, [그림 4-4]는 추세의 측면에서는 국내관광에 대한 설명력이 있으나, 주기적 변동에 대해서는 설명력이 없어 보인다. 따라서 분기 더미를 활용한 계절성 반영이 국내관광 수요예측에 적합하다고 판단한다.

2.2. 인구 모형

가. 인구 총량 모형

인구 역시 관광 총량을 설명할 수 있는 중요한 설명변수의 하나이다. 다만 분석 기간에 대해서는 인구 증가율이 실질 국내총생산 증가율보다 높지 않은 점이나, 인구 총량이 아무리 많더라도 일정한 소득 수준 이상이 되지 못하면 관광이 늘어나지 않을 수 있다는 점에서는 소득에 비해서는 설명력이 떨어질 것이다.⁴⁸⁾ 모형을 수식으로 나타내면 아래와 같다.

$$\ln(\text{tourist}_t) = \alpha + \beta_1 \ln \text{pop}_t + \beta_2 (\ln \text{pop}_t)^2 + \sum_{s=2}^4 \delta_s q_{t,s} + \epsilon_t,$$

여기서 $\ln(\text{tourist}_t)$ 는 t 분기의 로그 관광객 총량, $\ln \text{pop}_t$ 는 t 분기의 로그 국내인구 총량, $q_{t,s}$ 는 앞서와 동일하게 t 가 s 분기면 1 아니면 0의 값을 가지는 더미변수이다.

〈표 4-11〉은 인구 모형의 추정결과를 제시한다. 로그 인구는 상당한 설명력이 있었으며, 특히 소득과는 달리 제곱항을 고려한 모형에서도 수준 변수와 제곱항이 모두 유의하였다. 그러나 앞서 설명한 것처럼 높은 결정계수와 통계적 유의성이 예측력을 보장하는 것이 아니다. 실제의 국내관광 총량과 모형에서 추정된 값의 비교는 다음의 [그림 4-6], [그림 4-7]과 같다. 소득모형에서 검토한 바와 유사하게, 여기서도 〈표 4-11〉에서 분기 더미를 고려하지 않은 모형의 시계열도는 계절성을 반영되지 않았다고 보여 제시하지 않는다.

[그림 4-6]은 일차함수, [그림 4-7]은 이차함수 모형이다 인구를 이차함수 형태

48) 소득이 미화 1불 인구가 50억명이라고 한다면, 소득 규모로는 50억불이지만 관광총량은 매우 낮고 인구에 비례하지도 않을 가능성이 높다.

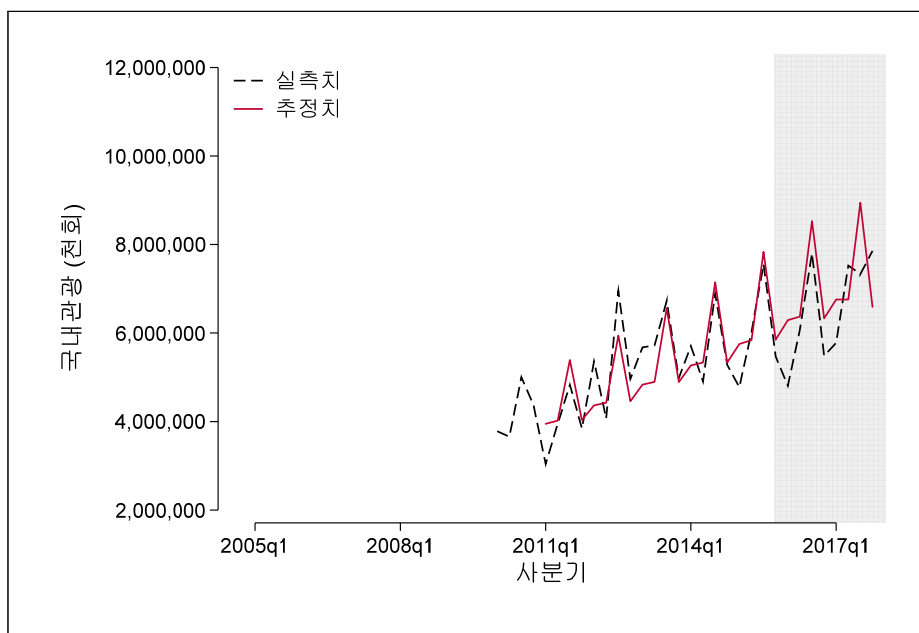
로 모형에 적용하는 경우, 국내관광이 둔화를 넘어서서 감소하는 것으로 추정되었으며, 이는 예측에 적합하다고 보기 어려우므로 표본외 검증에서는 <표 4-11>의 (2)열만을 검토하도록 한다.

<표 4-11> 로그 인구를 설명변수로 하는 회귀모형 추정 결과

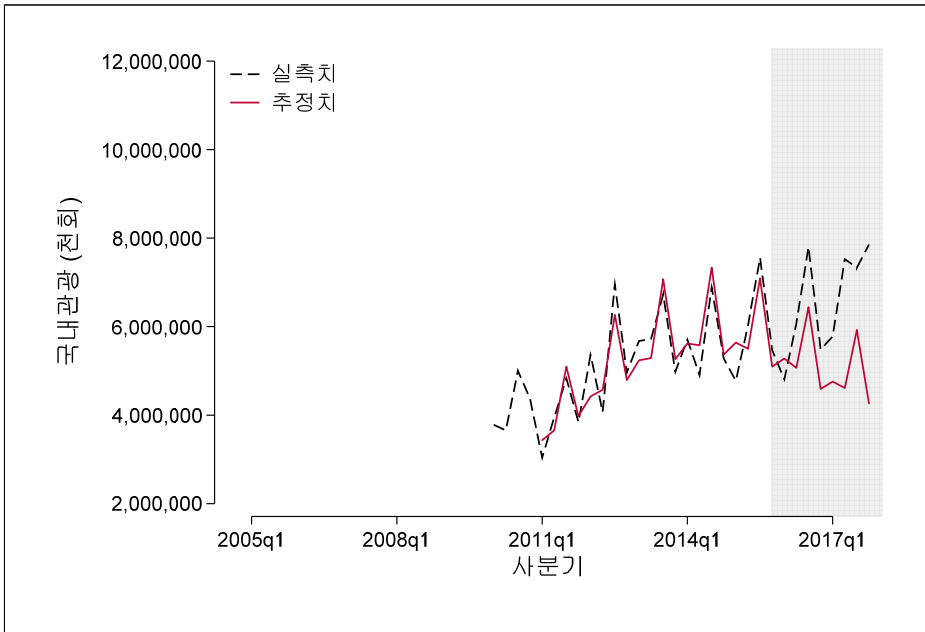
변수	(1)	(2)	(3)	(4)
로그 국내인구	24.172*** (7.617)	23.774*** (5.761)	20,199.159** (9,338.907)	19,645.798*** (6,090.956)
(로그 국내인구) ²			-2,565.169** (1,187.580)	-2,494.862*** (774.446)
상수항	-77.293** (29.964)	-75.776*** (22.678)	-39,746.150** (18,359.947)	-38,657.373*** (11,978.672)
분기더미	No	Yes	No	Yes
관측치	20	20	20	20
결정계수	0.391	0.711	0.506	0.820

주: 괄호 안은 이분산성에 강건한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다.

[그림 4-6] 국내관광과 <표 4-11>의 (2)열 추정치 시계열도



[그림 4-7] 국내관광과 <표 4-11>의 (4)열 추정치 시계열도



나. 일인당 모형

일인당 모형은 국내총생산 총량을 인구로 나누어 준 일인당 총생산의 로그 값을 활용하여 분석하는 모형이다. 일인당 총생산을 설명변수로 하는 모형의 장단점은 명확한 편이다. 우선 국내총생산과 인구의 정보를 모두 활용하고 있으며 국내총생산 총량이나 인구 총량보다 일인당 소득이 관광의 증가를 설명할 수 있는 보다 좋은 지표라고 생각되기 때문이다. 또한 국내총생산 총량과 인구 총량을 동시에 설명변수로 하는 경우 발생할 수 있는 다중공선성 문제에서 자유롭다는 장점도 있다. 다만 로그 일인당 총생산을 모형에서 설명변수로 하는 것은 모형에 추가적인 모수적 가정을 도입하는 것과 동일하다.

$$\ln(\text{tourist}_t) = \alpha + \beta_1 \ln y_t + \beta_2 (\ln y_t)^2 + \sum_{s=2}^4 \delta_s q_{t,s} + \epsilon_t,$$

여기서 $\ln(\text{tourist}_t)$ 는 t 분기의 로그 관광객 총량, $\ln y_t$ 는 t 분기의 로그 일인당 소득으로 $\ln(GDP_t/pop_t)$ 와 동일하다.⁴⁹⁾ $q_{t,s}$ 는 앞서서와 동일하게 t 가 s 분기면 1

아니면 0의 값을 가지는 더미변수이다. 회귀분석 결과는 <표 4-12>에서 제시한다.

<표 4-12> 로그 일인당 총생산을 설명변수로 하는 회귀모형 추정 결과

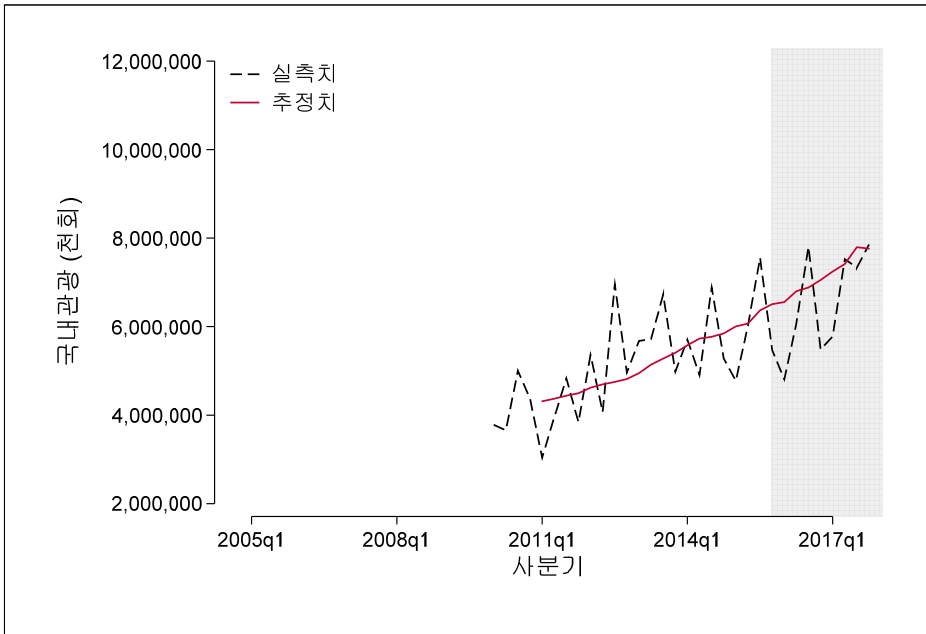
변수	(1)	(2)	(3)	(4)
로그 일인당소득	3.532*** (1.225)	3.423*** (0.924)	-16.998* (8.688)	-19.885*** (4.311)
(로그 일인당소득) ²			16.856** (6.951)	19.216*** (3.728)
상수항	-21.953 (13.793)	-20.782* (10.430)	-51.761*** (15.989)	-55.824*** (11.968)
분기더미	No	Yes	No	Yes
관측치	20	20	20	20
결정계수	0.344	0.661	0.489	0.840

주: 괄호 안은 이분산성에 강건한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다.

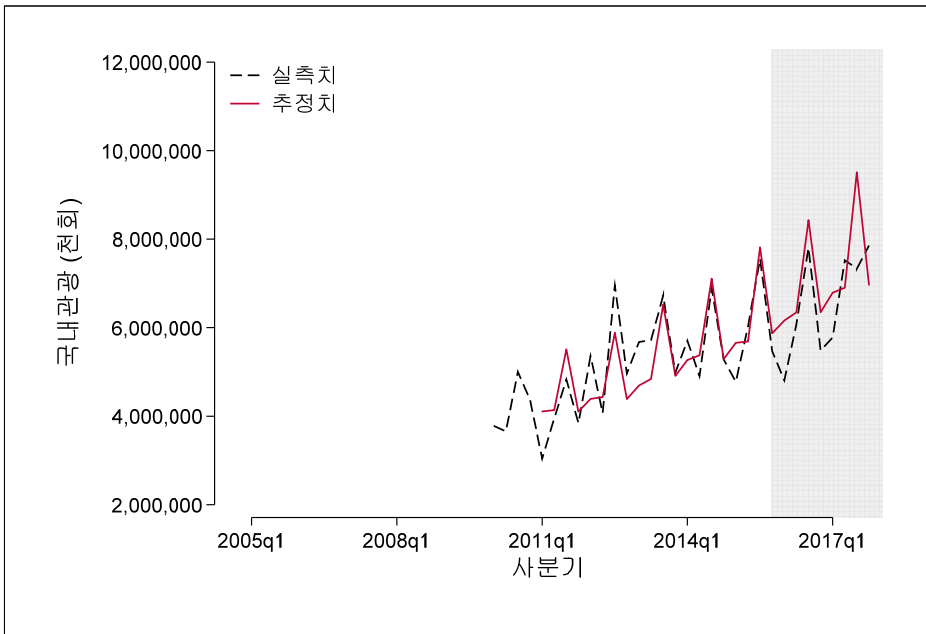
<표 4-12>의 결과는 로그 국내인구를 사용한 결과와 상당히 유사한 편이나, 이차함수의 형태가 반대라는 특징이 있다. 다만 [그림 4-8], [그림 4-9]에서 나타나는 것처럼 이차함수 모형은 예측력에 있어서는 충분한 설명력이 있다고 보기 어렵다. [그림 4-8]-[그림 4-11]는 각각 <표 4-12>에서 1-4열 추정 결과의 시계열도를 제시한다.

49) 또한, 로그 법칙에 의해 다음의 수식이 성립한다. $\ln y_t = \ln GDP_t - \ln pop_t$. 따라서 수식과 같이 로그 일인당 총생산을 설명변수로 하는 모형은 로그 국내총생산과 로그 국내인구의 계수 값은 동일하고 부호는 반대인 제약을 가하는 회귀모형과도 같다.

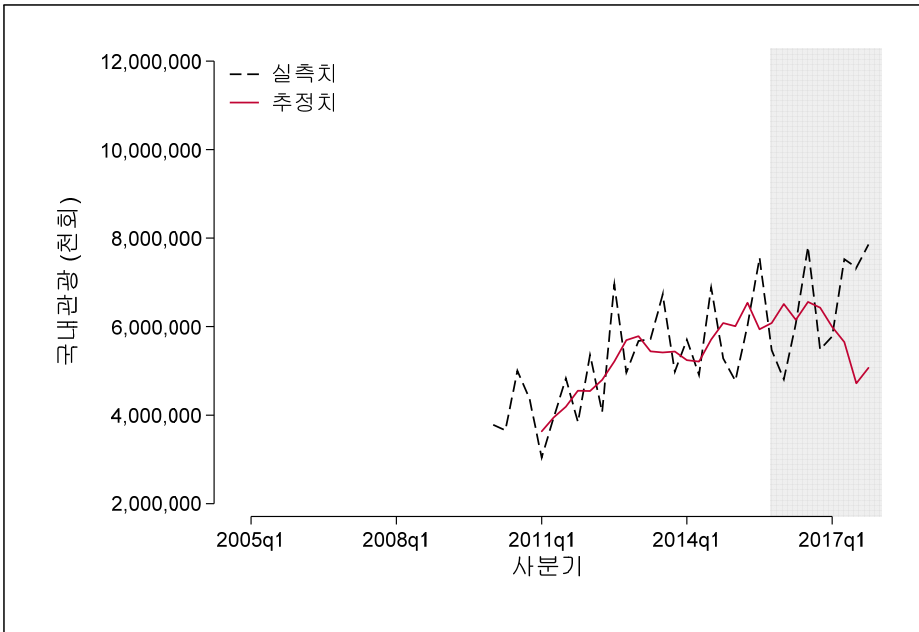
[그림 4-8] 국내관광과 <표 4-12>의 (1)열 추정치 시계열도



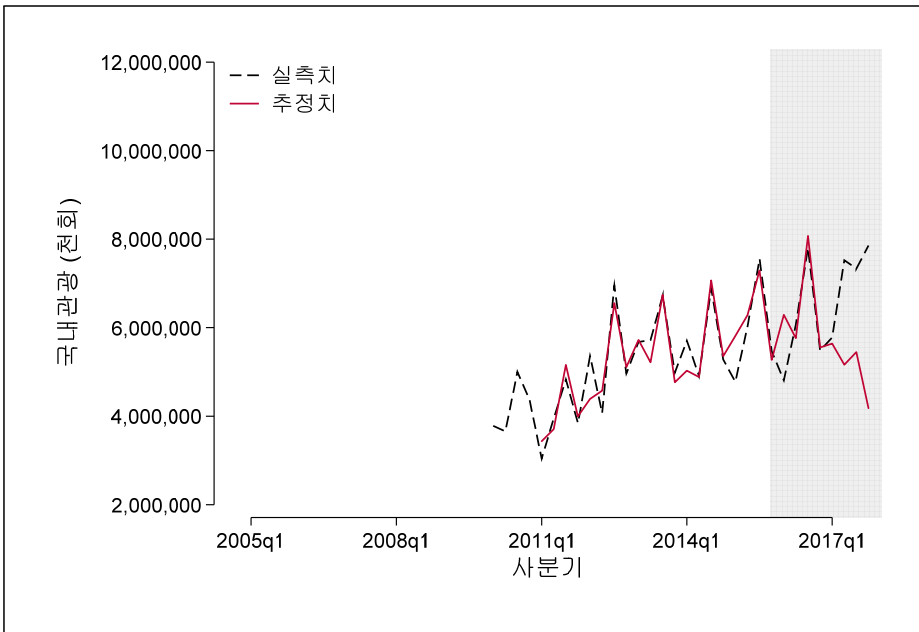
[그림 4-9] 국내관광과 <표 4-12>의 (2)열 추정치 시계열도



[그림 4-10] 국내관광과 <표 4-12>의 (3)열 추정치 시계열도



[그림 4-11] 국내관광과 <표 4-12>의 (4)열 추정치 시계열도



2.3. 공휴일 모형

관광하기 위해서는 소득과 함께 시간이 필요하다. 따라서 특정 시기에 공휴일의 유무는 관광에 영향을 줄 수 있는 요인이다. 공휴일 모형은 다음의 수식과 같다.

$$\ln(\text{tourist}_t) = \alpha + \beta_1 \ln \text{holiday}_t + \beta_2 (\ln \text{holiday}_t)^2 + \sum_{s=2}^4 \delta_s q_{t,s} + \epsilon_t,$$

여기서 $\ln(\text{tourist}_t)$ 는 t 분기의 로그 관광객 총량, $\ln \text{holiday}_t$ 는 t 분기의 로그 공휴일 총량, $q_{t,s}$ 는 t 가 s 분기면 1 아니면 0의 값을 가지는 더미변수이다. 공휴일 총량은 토요일, 일요일과 법정공휴일의 합계로 계산하였다.

〈표 4-13〉 로그 공휴일수를 설명변수로 하는 회귀모형 추정 결과

변수	(1)	(2)	(3)	(4)
로그 공휴일수	-0.007 (1.384)	-0.560 (1.112)	14.349 (22.685)	15.081 (18.963)
(로그 공휴일수) ²			-0.009 (0.015)	-0.010 (0.012)
상수항	17.757*** (4.643)	19.537*** (3.739)	-23.060 (63.984)	-24.953 (53.628)
분기더미	No	Yes	No	Yes
관측치	24	24	24	24
결정계수	0.000	0.325	0.012	0.339

주: 괄호 안은 이분산성에 강건한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다.

공휴일 모형의 추정결과는 〈표 4-13〉에서 제시되어 있다. 표의 모든 열에서 공휴일수는 통계적으로 유의하지 않았다.⁵⁰⁾ 따라서 주말과 법정공휴일을 모두 합산한 변수를 사용한 모형은 국내관광 예측에 적합하지 않다고 판단된다. 그 이유로는 여러 가지를 생각할 수 있다. 우선 토요일, 일요일은 관광보다 가족과의 여가를 즐기

50) 다만 이러한 결과가 국내관광에 달력효과(calendar effects)가 없음을 의미하지는 않음에 유의하여야 한다. 관광수요에 달력효과가 존재함을 시사하는 연구들이 있으며(박상근, 2017b; 김형중, 2019), 본 연구에서 제시한 공휴일 변수가 달력효과를 측정하기에 적절하지 않을 수 있기 때문이다. 또한 분기별 계절성 역시 큰 의미로 달력효과에 포함된다고 볼 수 있다. 국내관광 예측과 달력효과에 대해서는 추가적인 후속 연구가 필요하다.

는 사례가 많기 때문일 수 있다. 여기서 가족과의 여가란 이동을 수반하지 않고 거주하는 지역에서의 외식이나 문화활동 등을 지칭한다. 이러한 소비는 이동을 수반하지 않아 국내관광으로는 집계되지 않기 때문이다. 또한 우리나라 근로자들의 평균적인 휴가 기간은 ‘7말8초’라는 용어가 있을 만큼 특정 시기에 매우 집중되어 있으며, 따라서 공휴일수보다는 분기더미에 의해 그 영향이 잘 포착될 수 있다.⁵¹⁾ 실제로 지금까지 제시한 모든 모형에서 분기더미를 포함하는 경우에는 결정계수가 크게 상승함을 관찰할 수 있었다.

2.4. 교호항 모형

여기서는 앞에서 검토하지 않았던 교호항(interaction term)을 사용하여 추정한 결과를 보고한다. 검토하는 교호항은 분기더미와 로그 국내총생산, 로그 국내인구, 로그 일인당 소득과의 교호항이다. 이러한 모형은 설명변수가 미치는 영향이 분기별로 다를 수 있음을 의미하며, 따라서 시간가변계수모형(time-varying coefficient model)은 아니지만 계수의 동적 변화(dynamics)를 부분적으로 포착할 수 있다.⁵²⁾

$$\ln(\text{tourist}_t) = \alpha + \beta_1 \ln x_t + \sum_{s=2}^4 \delta_s q_{t,s} + \sum_{s=2}^4 \psi_s \ln x_t \times q_{t,s} + \epsilon_t,$$

여기서 $\ln(\text{tourist}_t)$ 는 t 분기의 로그 여행횟수 총량, $\ln x_t$ 는 t 분기의 로그 설명변수, $q_{t,s}$ 는 t 가 s 분기면 1 아니면 0의 값을 가지는 더미변수, $\ln x_t \times q_{t,s}$ 는 로그 설명변수와 분기 더미의 교호항이다.

〈표 4-14〉에서는 분기 더미와 설명변수의 교호항을 고려한 모형의 추정결과를 제시한다. 대부분의 추정결과에서 결정계수의 미미한 증가가 나타났으며 로그 국내총생산과 로그 국내인구를 같이 설명변수로 하는 모형에서는 다중공선성으로 인하여 로그 국내총생산의 계수가 음수로 추정되었다. 로그 국내인구와 로그 일인당 소득은 통계적으로 유의하지 않았다. 결정계수는 제3열의 모형이 가장 높았으며 그 다음으로는 1열, 2열, 4열의 순으로 결정계수가 높은 것으로 나타났다.

51) 7월말 8월초라는 단어 의미에 따라서 그 영향은 ‘3분기 더미’(7-9월)에 포착된다.

52) 분기마다 움직이는 계수의 움직임은 포착하지 못하고 분기 사이의 변화 차이를 포착하는 모형이 된다.

〈표 4-14〉 로그 일인당소득을 설명변수로 하는 회귀모형 추정 결과

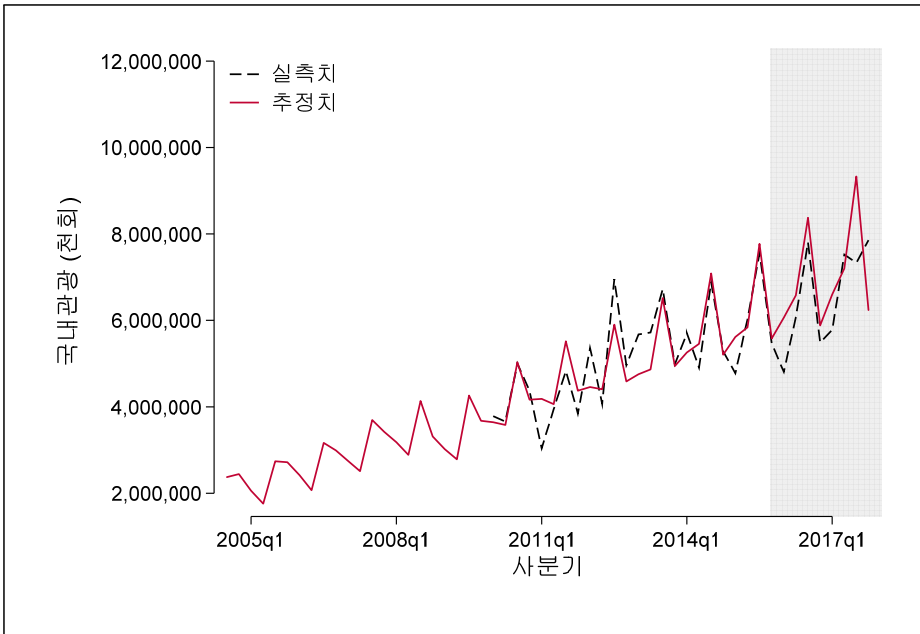
변수	(1)	(2)	(3)	(4)
로그 국내총생산	2.688** (1.262)		-27.939** (12.083)	
로그 국내인구		25.467 (22.069)	222.766** (82.697)	
로그 일인당소득				3.628 (3.696)
상수항	-23.083 (19.141)	-82.432 (86.791)	-434.247** (149.937)	-23.087 (41.589)
분기더미	Yes	Yes	Yes	Yes
분기더미×국내총생산	Yes	No	Yes	No
분기더미×국내인구	No	Yes	Yes	No
분기더미×일인당소득	No	No	No	Yes
관측치	24	20	20	20
결정계수	0.729	0.715	0.861	0.667

주: 괄호 안은 이분산성에 강건한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다. (분기더미×국내총생산), (분기더미×국내인구), (분기더미×일인당소득)은 각각 분기더미와 로그 국내총생산, 로그 국내인구, 로그 일인당소득의 교호항을 의미한다.

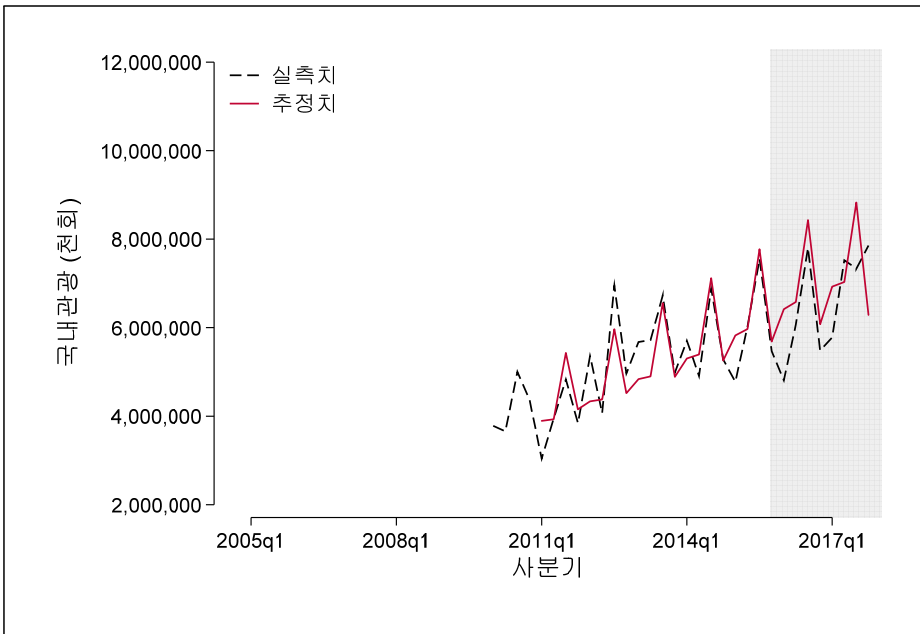
[그림 4-12]-[그림 4-15]는 각각 〈표 4-14〉의 1-4열에 상응하는 시계열도이다. 그림을 검토한 결과, 추가적인 교호항의 고려는 예측력에 긍정적인 영향이 있는 것으로 보이나, 그 영향이 크지는 않은 것으로 보인다. 다만 교호항의 포함으로 인한 부정적인 영향 역시 크지 않다고 보이기 때문에 교호항의 포함이 예측에 미치는 영향에 대해서는 표본의 검정에서 다시 논의하도록 한다. [그림 4-14]의 로그 국내총생산과 로그 국내인구를 동시에 고려하는 모형에서 추정된 시계열은, 예측력이 상당히 훼손된 결과가 제시되었다. 따라서 로그 국내총생산과 로그 국내인구를 동시에 고려하는 모형은 본 보고서에서는 배제한다.⁵³⁾

53) 잘못 추정된 계수의 부호를 바르게 추정하기 위한 방법으로 로그 국내총생산과 로그 국내인구의 교호항을 사용하여 로그 국내총생산의 순효과를 분리·추정하거나 도구변수를 사용하여 추정할 수 있다. 하지만 본 보고서의 목적은 바른 베타를 추정하는 것이 아니라 보다 정확한 미래 예측치를 구하는 것이기 때문에 여기서는 이러한 분석을 생략하도록 한다.

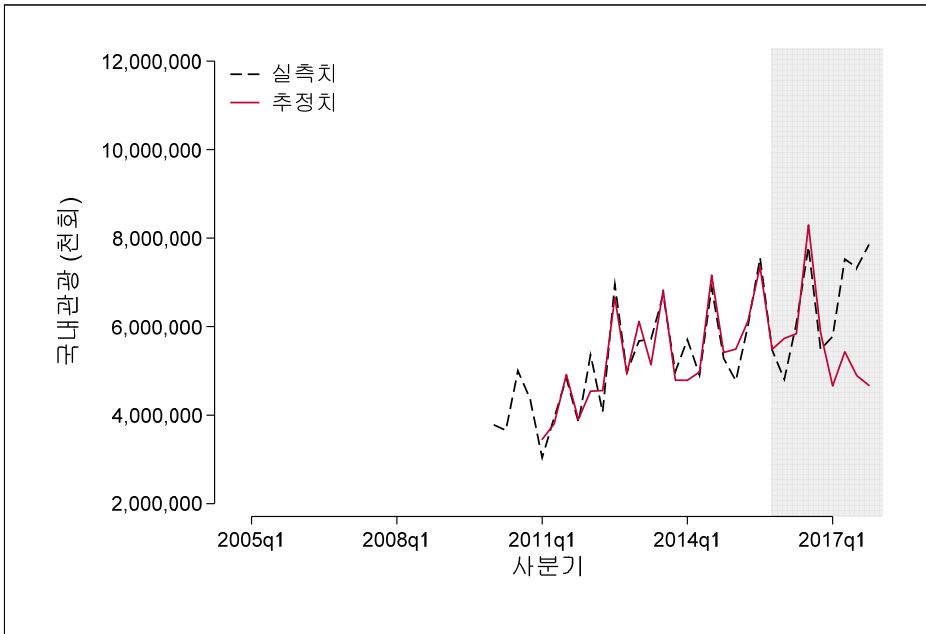
[그림 4-12] 국내관광과 <표 4-14>의 (1)열 추정치 시계열도



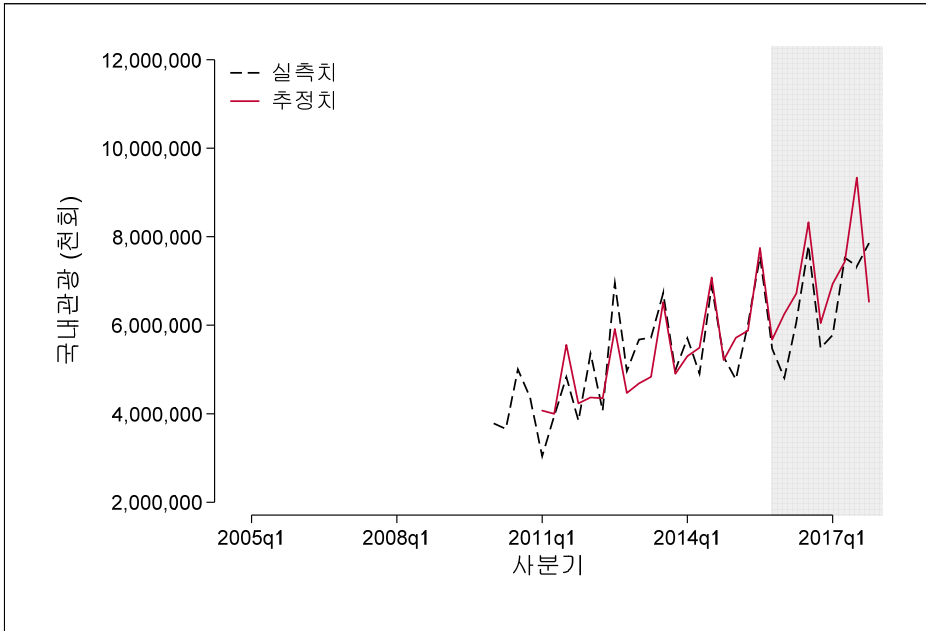
[그림 4-13] 국내관광과 <표 4-14>의 (2)열 추정치 시계열도



[그림 4-14] 국내관광과 <표 4-14>의 (3)열 추정치 시계열도



[그림 4-15] 국내관광과 <표 4-14>의 (4)열 추정치 시계열도



2.5. 습관 지속 모형

소비에는 어느 정도 관성적인 측면이 존재한다. 따라서 과거의 관광 소비가 현재의 소비에 영향을 주는 관계를 모형으로 만들 수도 있다. 관광 역시 과거의 관광 습관이 현재의 관광에 영향을 미치는 모형을 생각할 수 있다. 종속변수의 시차변수를 포함하면 과거의 관광이 현재의 관광에 영향을 주는 형태가 되며, 이러한 모형을 습관 지속 모형(habit persistence model)이라고 부른다. 관광 수요예측에서는 Witt (1980)가 최초로 사용하였다고 알려져 있다. 습관 지속 모형은 다음의 식처럼 나타낼 수 있다.

$$\ln(\text{tourist}_t) = \alpha + \rho \ln(\text{tourist}_{t-1}) + \beta \ln x_t + \sum_{s=2}^4 \delta_s q_{t,s} + \epsilon_t,$$

여기서 $\ln(\text{tourist}_t)$ 는 t 분기의 로그 여행횟수 총량, $\ln x_t$ 는 t 분기의 로그 설명변수, $q_{t,s}$ 는 t 가 s 분기면 1 아니면 0인 더미변수이다. 이전의 분석에서 인구 변수의 예측력은 미미하였기 때문에 국내총생산과 일인당소득만을 설명변수 x_t 로 고려하도록 한다. 모형의 추정결과는 <표 4-15>와 같다.

<표 4-15> 로그 국내총생산을 설명변수로 하는 습관 지속 모형 추정 결과 (제약: $\psi_s = 0$)

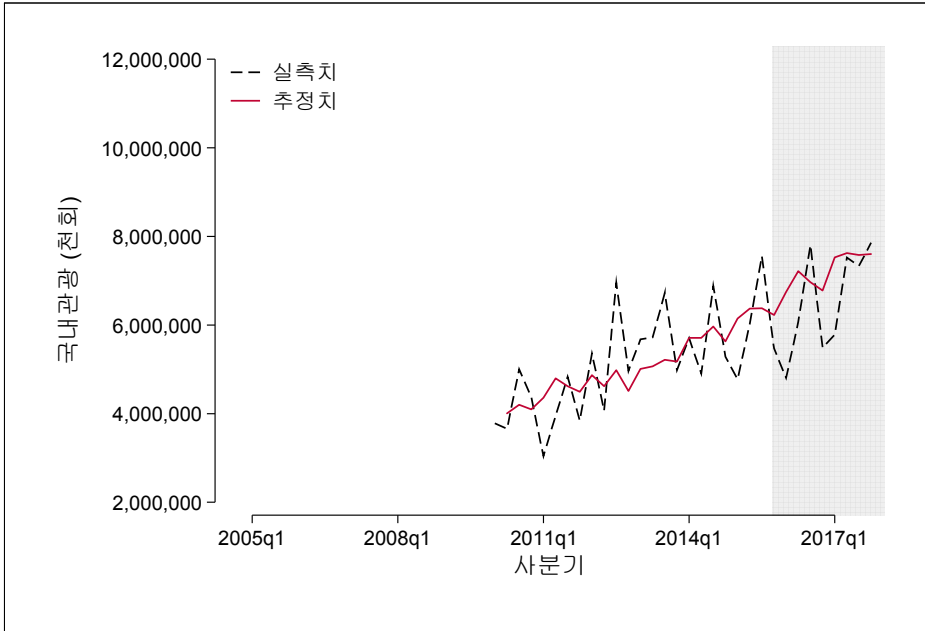
변수	(1)	(2)	(3)	(4)
로그 국내관광 (t-1)	-0.217 (0.192)	0.020 (0.226)	-0.210 (0.202)	0.043 (0.235)
로그 국내총생산 (t)	3.516*** (0.926)	2.768*** (0.953)		
로그 일인당소득 (t)			4.347*** (1.419)	3.276** (1.427)
상수항	-31.739** (12.279)	-24.658** (11.359)	-27.388* (14.492)	-19.881 (13.177)
분기더미	No	Yes	No	Yes
관측치	23	23	20	20
결정계수	0.419	0.698	0.371	0.661

주: 괄호 안은 이분산성에 강건한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다.

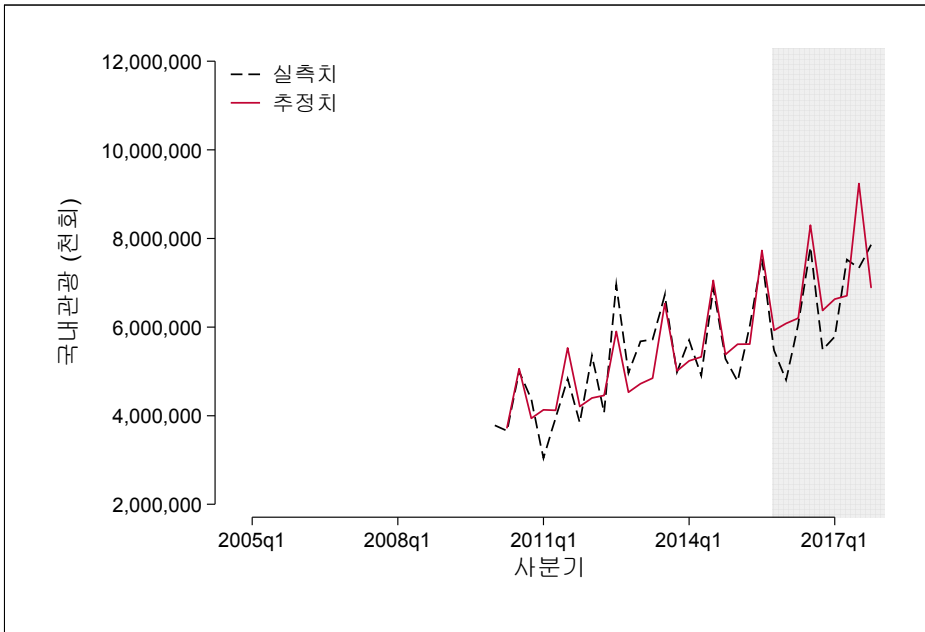
<표 4-15>의 추정결과는 시차변수의 포함이 예측력을 크게 향상시키지는 않음을 보여주고 있다. 결정계수가 예측력을 담보하는 것은 아니지만, 분기 더미 없이 시차

변수 $\ln(\text{tourist}_{t-1})$ 만을 포함하는 모형의 결정계수는 상대적으로 낮았다.

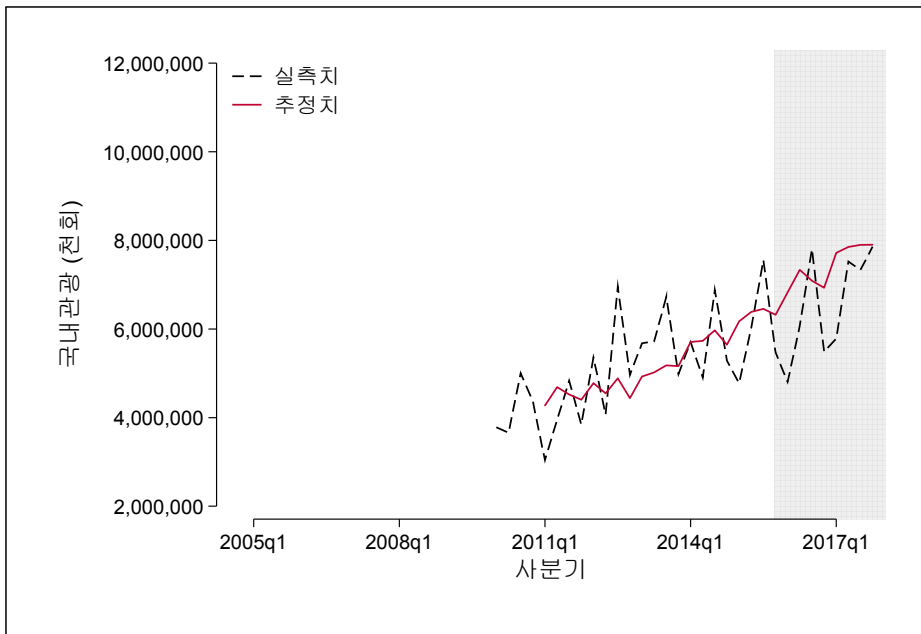
[그림 4-16] 국내관광과 <표 4-15>의 (1)열 추정치 시계열도



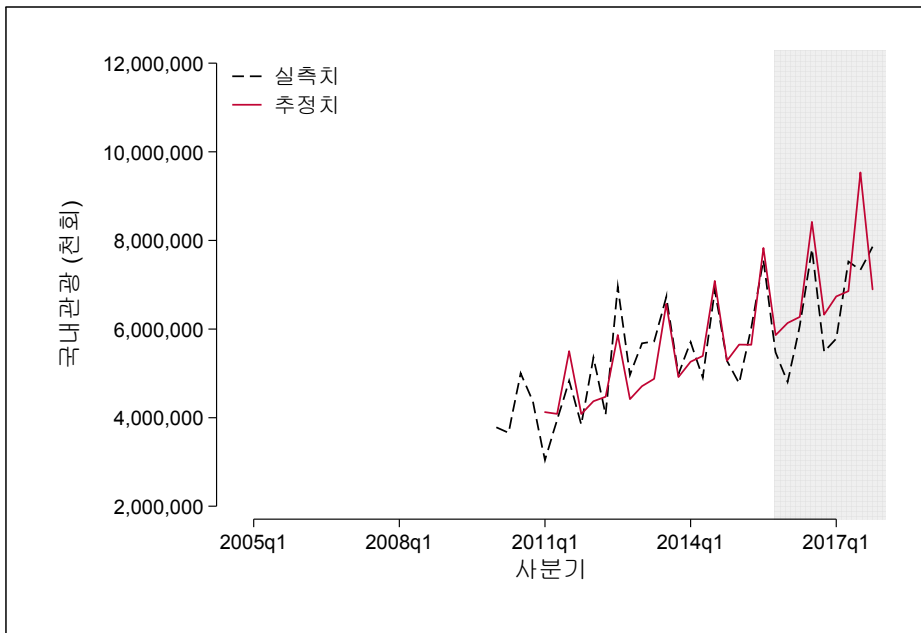
[그림 4-17] 국내관광과 <표 4-15>의 (2)열 추정치 시계열도



[그림 4-18] 국내관광과 <표 4-15>의 (3)열 추정치 시계열도



[그림 4-19] 국내관광과 <표 4-15>의 (4)열 추정치 시계열도



[그림 4-16]-[그림 4-19]은 각각 <표 4-15>의 1-4열에 상응하는 시계열도이다. 시차변수만을 포함하는 [그림 4-16]와 [그림 4-18]은 추세에 있어서는 국내관광과 유사하게 움직이고 있으나 계절성을 충분하게 추정하지는 못하고 있다. 반면 분기 더미를 포함하는 [그림 4-17]과 [그림 4-19]은 상대적으로 국내관광의 움직임을 적절히 포착하고 있다.

추가적으로 종속변수의 시차변수 및 설명변수와 분기 더미의 교호항을 고려하는 모형도 생각해 볼 수 있다. 모형은 다음의 식과 같다.

$$\ln(\text{tourist}_t) = \alpha + \rho \ln(\text{tourist}_{t-1}) + \sum_{s=2}^4 \xi_s \ln(\text{tourist}_{t-1}) \times q_{t,s} \\ + \beta \ln x_t + \sum_{s=2}^4 \delta_s q_{t,s} + \sum_{s=2}^4 \psi_s \ln x_t \times q_{t,s} + \epsilon_t,$$

여기서 $\ln(\text{tourist}_t)$ 는 t 분기의 로그 여행횟수 총량, $\ln x_t$ 는 t 분기의 로그 설명변수, $q_{t,s}$ 는 t 가 s 분기면 1 아니면 0인 더미변수, $\ln \text{tourist}_{t-1} \times q_{t,s}$ 는 로그 국내관광 시차변수와 분기 더미의 교호항, $\ln x_t \times q_{t,s}$ 는 로그 설명변수와 분기 더미의 교호항이다. 모형의 추정 결과는 <표 4-16>과 같다.

<표 4-16> 로그 국내총생산을 설명변수로 하는 습관 지속 모형 추정 결과

변수	(1)	(2)	(3)	(4)
로그 국내관광 (t-1)	0.025 (0.227)	0.039 (0.241)	-1.121 (1.485)	-1.060 (1.689)
로그 국내총생산 (t)	3.134 (3.260)		5.729 (5.292)	
로그 일인당소득 (t)		3.527 (4.001)		6.413 (6.944)
상수항	-30.296 (48.634)	-22.629 (44.110)	-49.420 (59.204)	-35.656 (54.279)
분기더미	Yes	Yes	Yes	Yes
분기더미×국내관광	No	No	Yes	Yes
분기더미×국내총생산	Yes	No	Yes	No
분기더미×일인당소득	No	Yes	No	Yes
관측치	23	20	23	20
결정계수	0.712	0.668	0.757	0.719

주: 괄호 안은 이분산성에 강건한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다.

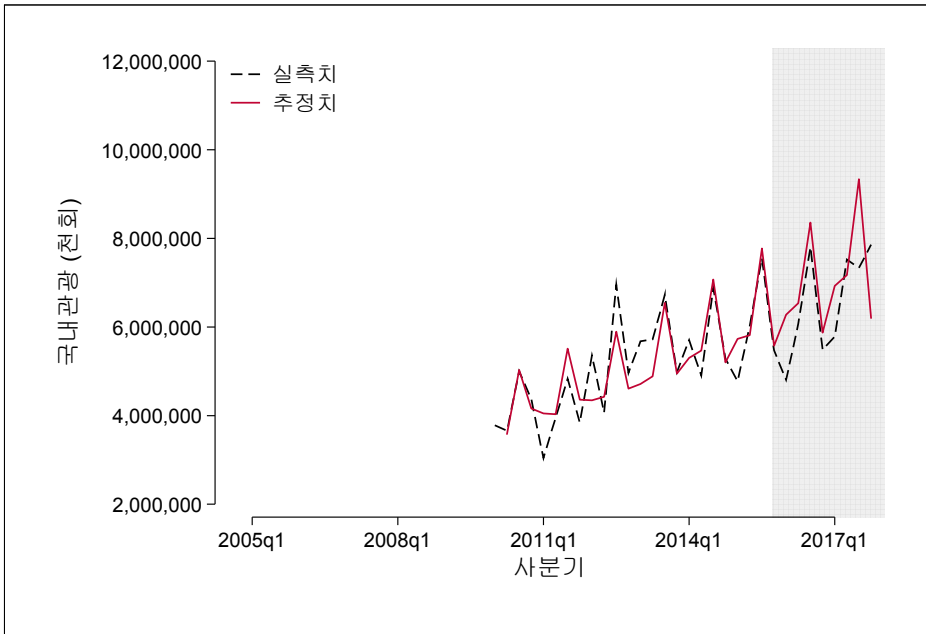
예측력과는 별개의 논의이기는 하지만, <표 4-16>에서 주요 변수들은 유의하지 않았다. 이는 한정된 표본에서 매우 많은 변수들을 추가한 영향일 개연성이 높다고 보인다. 교호항을 고려하지 않은 <표 4-15>에서는 국내총생산과 일인당소득이 유의하였기 때문이다. 또한 <표 4-15>과 <표 4-16>의 결과를 종합하면 종속변수의 시차변수가 가져다주는 예측력의 향상은 크지 않다고 보인다.

[그림 4-20]-[그림 4-22]은 각각 <표 4-16>의 1-4열에 상응하는 시계열도이다. <표 4-15>의 결과를 그린 [그림 4-16]-[그림 4-19]과 비교하면, 시차변수와의 교호항 포함이 예측력을 상당히 향상시키는 것으로 보기 어렵다. 로그 국내총생산과 분기 더미의 교호항을 시차변수와 같이 고려하는 모형도 상당한 수준의 예측력 향상이 존재한다고 판단하기는 어려워 보인다.⁵⁴⁾ 지금까지 분석한 결과로 판단하면 현재 주어진 데이터의 수준이 충분하지 못한 상황에서 많은 더미변수의 사용이나 설명변수의 추가가 오히려 예측력 저하로 이어지는 과적합(overfitting)이 나타난다고 보인다. 과적합 문제의 주요 해결책이 데이터 증강(data augmentation) 및 모형내 변수의 감소임을 감안하면 현재의 모형에 추가적으로 다양한 가정이나 변수의 추가는 바람직하지 않다. 문제의 해결을 위해서는 국내관광 데이터를 보완하고 시계열을 연장하려는 시도가 필요하며, 동시에 지속적인 국내관광 예측 경험이 누적되어야 한다고 판단한다.

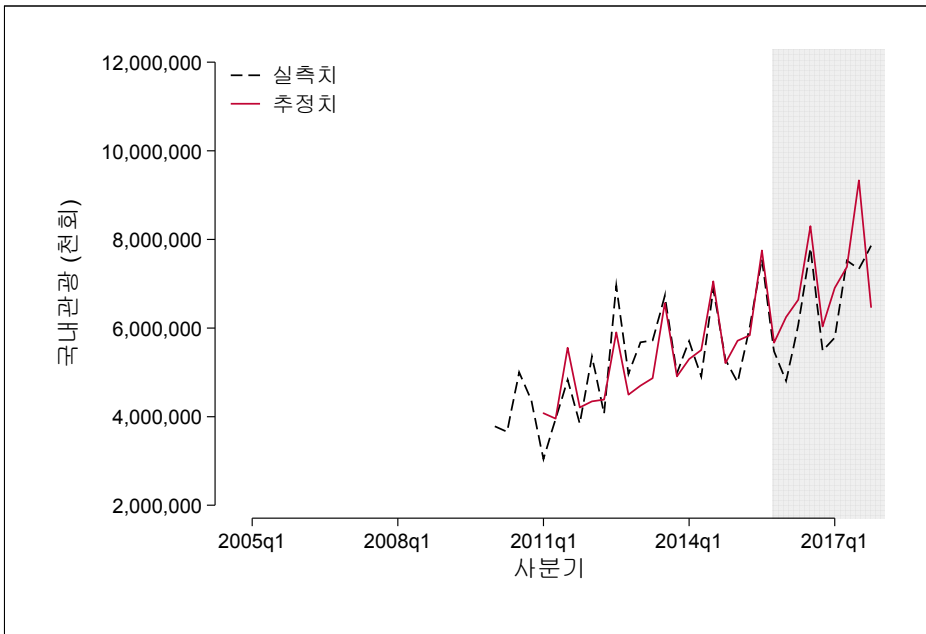
전기의 시차 변수만이 아니라 보다 과거의 시차변수들, 즉 $t-1$, $t-2$, ..., $t-s$ 의 시차 변수도 고려할 수 있으나, 분석 결과에서 나타나는 것처럼 예측에 사용할 수 있는 국내관광 관측치가 20여개에 불과할 정도로 매우 제약되어 있으며, 최적의 시차를 선택하여야 하는 문제도 추가적으로 발생한다. 이러한 모형은 주어진 데이터로는 추정이 불가능하다고 판단하여 시도하지 않았다. 하지만 시계열 데이터가 축적되고 과거의 시계열이 보정됨을 전제로 하여, 추가적인 시차를 고려하는 모형도 분석될 필요가 있을 것이다. 우리나라의 국내관광에서 나타나는 중요한 특징의 하나가 계절성인 만큼, 국내관광 역시 전 분기의 국내관광보다 더 과거의 국내관광과 강한 관련성이 존재할 수 있기 때문이다.

54) 다만, 이러한 결과는 월별 데이터나 연도별 데이터를 활용하는 경우 달라질 수 있다. 하지만 월별 데이터의 경우 분기별로 생산되는 설명변수 데이터와의 결합이 불가능하기 때문에 우리가 예측에 사용할 수 있는 모형은 순수한 시계열 모형으로 수렴되며, 연도별 데이터를 사용하는 경우는 시계열이 지금보다 크게 짧아진다는 문제점이 존재한다.

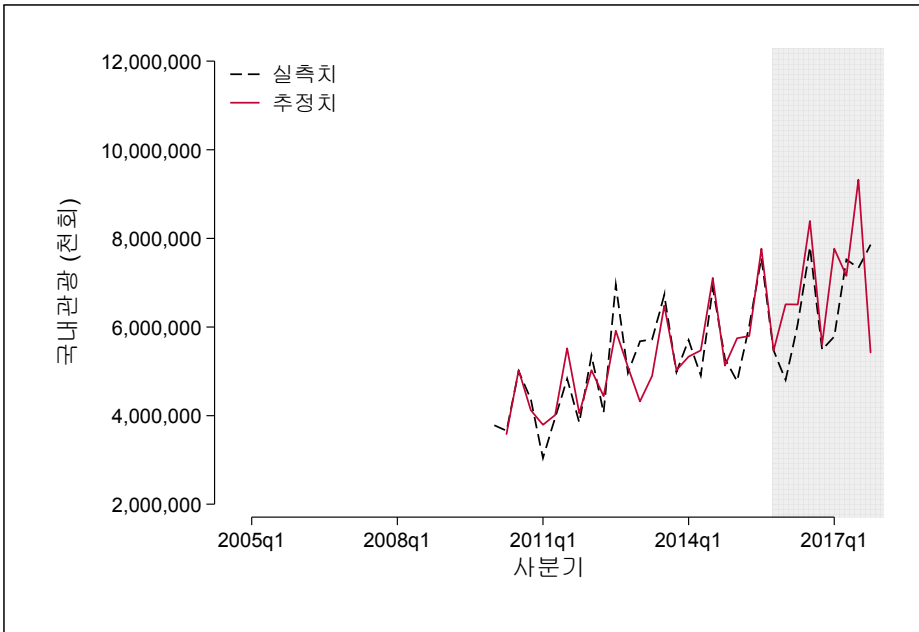
[그림 4-20] 국내관광과 <표 4-16>의 (1)열 추정치 시계열도



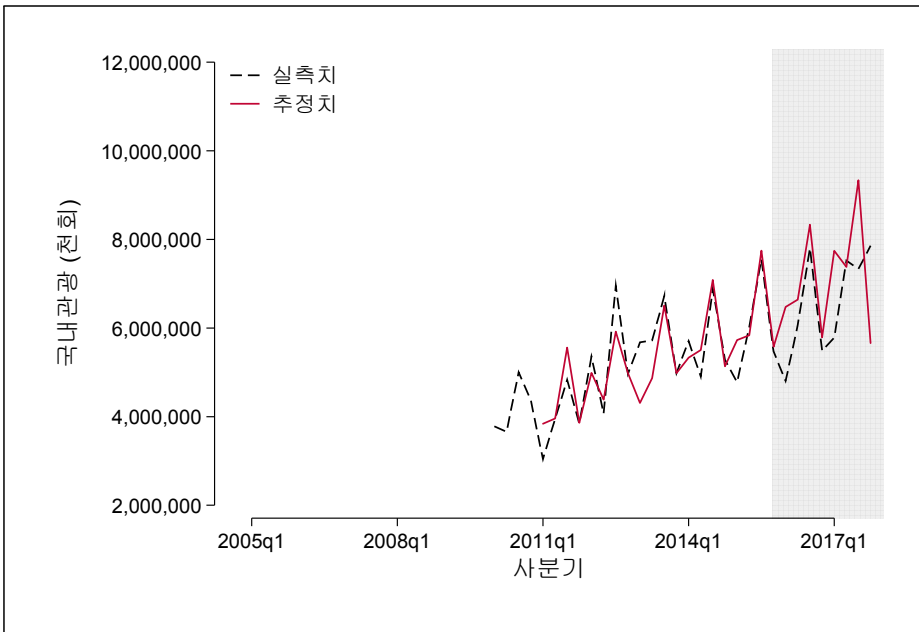
[그림 4-21] 국내관광과 <표 4-16>의 (2)열 추정치 시계열도



[그림 4-22] 국내관광과 <표 4-16>의 (3)열 추정치 시계열도



[그림 4-23] 국내관광과 <표 4-16>의 (4)열 추정치 시계열도



제4절 모형 검증 및 예측

1. 표본외 검정

추정 결과를 검토하여 예측력이 충분하지 않다고 판단되는 모형들은 모형 검증에서 제외하였다. 표본외 검정(out-of-sample test)을 위해 최종적으로 선택된 모형은 다음의 식과 같다.⁵⁵⁾

$$(1) \ln(\text{tourist}_t) = \alpha + \beta_1 \ln GDP_t + \sum_{s=2}^4 \delta_s q_{t,s} + \epsilon_t,$$

$$(2) \ln(\text{tourist}_t) = \alpha + \beta_1 \ln GDP_t + \beta_2 (\ln GDP_t)^2 + \sum_{s=2}^4 \delta_s q_{t,s} + \epsilon_t,$$

$$(3) \ln(\text{tourist}_t) = \alpha + \beta_1 \ln GDP_t + \sum_{s=2}^4 \gamma_s q_{t,s} \cdot \ln GDP_t + \sum_{s=2}^4 \delta_s q_{t,s} + \epsilon_t,$$

$$(4) \ln(\text{tourist}_t) = \alpha + \beta_1 \ln GDP_t + \beta_2 (\ln GDP_t)^2 + \sum_{s=2}^4 \delta_s q_{t,s} \\ + \sum_{s=2}^4 \gamma_s q_{t,s} \cdot \ln GDP_t + \sum_{s=2}^4 \psi_s q_{t,s} \cdot (\ln GDP_t)^2 + \epsilon_t,$$

여기서 $\ln(\text{tourist}_t)$ 는 t 분기 로그 관광객, $\ln GDP_t$ 는 t 분기 로그 국내총생산 총량, $q_{t,s}$ 는 $t = s$ 이면 1, $t \neq s$ 이면 0의 값을 가지는 더미변수이다.

제시된 모형은 단순한 형태이기는 하나, 이전 절의 추정결과에서는 국내관광 모형에 다양한 변수를 추가하는 것보다 분기더미와 같은 추가적인 모수적 가정이 예측력 제고에 도움이 되었다. 그리고 실무적으로도 다양한 변수를 고려하는 예측모형은 그만큼의 예측치가 필요하기 때문에 단일한 설명변수를 사용하는 모형에 비해서 실용성은 떨어진다.

55) 항공교통 분야에서 사용되는 수요예측 모형 역시 여기서 제시하는 모형과 상당히 유사하다.

사전적인 검토는 목적에 의거하였으므로, 여기서는 보다 수리적인 방법에 의하여 표본외 검정을 수행하며, 예측하고자 하는 미래는 1년 후를 가정하였으므로 기본적인 표본외 검정 역시 그에 상응하게 실시한다. 검정의 척도는 평균절대백분율오차(mean absolute percentage error; 이하 MAPE)와 정규화 제곱근평균제곱오차(normalized root mean squared error; 이하 NRMSE)를 사용한다.

평균절대백분율오차와 정규화 제곱근평균제곱오차

y_t 가 실제 값이고 \hat{y}_t 가 모형에서 예측한 값이라고 하자. 평균절대백분율오차 공식은 다음과 같다.

$$MAPE = \sqrt{\frac{\sum_{t=\tau_s}^{\tau_e} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{\tau_s - \tau_e + 1}}.$$

제곱근평균제곱오차의 공식은 다음과 같이 쓸 수 있다.

$$RMSE = \sqrt{\frac{\sum_{t=\tau_s}^{\tau_e} (y_t - \hat{y}_t)^2}{\tau_s - \tau_e + 1}}.$$

제곱근평균제곱오차는 y_t 의 단위에 따라서 값이 크게 달라지는 문제가 있고, 이러한 문제는 정규화(normalization)로 해결할 수 있다. 정규화 제곱근평균제곱오차는 다음과 같다.

$$NRMSE = \frac{RMSE}{\max(y_t) - \min(y_t)}$$

1.1. 표본외 검정결과 1

데이터에서 시계열 단절이 일어나지 않은 기간은 2010년 1분기부터 2017 4분기까지이다. 따라서 2010년 1분기부터 2016년 4분기까지를 추정에 사용하고 나머지 1년을 사용하여 표본외 검정을 수행한다. 식 (1)-(4)에서 제시한 모형의 추정결과 및 표본외 검정결과는 다음의 <표 4-17>과 같다. MAPE와 NRMSE의 모형 평가가 다르기는 하지만 전반적으로는 가장 단순한 식 (1) 모형과 식 (3) 모형이 예측 오차가 작은 것으로 나타났다. 다만 식 (3) 모형이 MAPE 순위가 1등, NRMSE 순위가 2등으로 나타나 가장 적합한 모형으로 판단하였다. 제곱항을 고려한 모형은 제곱항이 없는 모형들과 비교하면 예측 오차가 상대적으로 큰 것으로 판명되었다.

〈표 4-17〉 모형 (1)~(4) 추정결과 및 표본외 검정결과 (분석에 사용한 기간: 2010q1~2016q4)

변수	(1)	(2)	(3)	(4)
로그 국내총생산	2.342*** (0.341)	318.339* (178.836)	1.904* (0.990)	373.411* (191.886)
(로그 일인당소득) ²		-10.410* (5.891)		-12.252* (6.341)
상수항	-17.869*** (5.182)	-2,415.763* (1,357.282)	-11.228 (15.044)	-2,827.340* (1,451.532)
분기더미	Yes	Yes	Yes	Yes
분기더미×국내총생산	No	No	Yes	Yes
분기더미×국내총생산 ²	No	No	No	Yes
관측치	28	28	28	28
결정계수	0.725	0.750	0.744	0.773
MAPE	0.144	0.143	0.137	0.175
NRMSE	0.552	0.699	0.614	0.734

주: 괄호 안은 이분산성에 강한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다. (분기더미×국내총생산), (분기더미×국내총생산²), (분기더미×일인당소득)은 각각 분기더미와 로그 국내총생산, 로그 국내총생산 제곱항, 로그 일인당소득의 교호항을 의미한다. NRMSE는 정규화 제곱근평균제곱 오차를 의미한다.

1.2. 표본외 검정결과 2

여기서는 시계열 단절 문제를 무시하고 조사가 크게 개편된 2018년, 2019년 데이터를 활용하여 표본외 검정을 수행한다. 〈표 4-18〉은 2010년 1분기부터 2017년 4분기까지를 추정에 사용하고 2018년 국내관광을 예측한 결과이다. 즉, 국민여행조사 개편 이전의 가용한 데이터 전체를 가지고 추정하고 2018년 조사 결과와 비교하는 표이다.

〈표 4-19〉은 2010년 1분기부터 2018년 4분기까지를 추정에 사용하고 2019년 국내관광을 예측한 결과이다. 즉, 국민여행조사 개편 이전의 가용한 데이터 전체에 개편된 2018년 데이터를 추가하여 모형을 추정하고 2019년 조사 결과와 비교하는 표이다.

〈표 4-18〉는 이전의 〈표 4-17〉과 동일하게 식 (3) 모형의 예측오차가 가장 작은 것으로 나타났다. 식 (3) 모형의 예측력은 MAPE로 평가하면 1등, NRMSE로 평가하면 2등으로 나타났다. 〈표 4-19〉에서도 역시 식 (3) 모형의 예측오차가 가장 작게 나타났다. 그러나 평가 순위에는 변동이 있었는데, 식 (3)의 예측력을 MAPE로 평가하면 2등이었으며, NRMSE로 평가하면 1등으로 나타났다. MAPE 순위에서 2등은 식 (4) 모형이었으며, NRMSE 순위에서 2등은 식 (2) 모형이었다.

〈표 4-18〉 모형 (1)~(4) 추정결과 및 표본외 검정결과 (분석에 사용한 기간: 2010q1~2017q4)

변수	(1)	(2)	(3)	(4)
로그 국내총생산	2.442*** (0.329)	114.6 (157.4)	1.889** (0.760)	147.0 (141.6)
(로그 일인당소득) ²		-3.693 (5.184)		-4.781 (4.670)
상수항	-19.40*** (5.013)	-871.7 (1,195)	-10.99 (11.57)	-1,112 (1,073)
분기더미	Yes	Yes	Yes	Yes
분기더미×국내총생산	No	No	Yes	Yes
분기더미×국내총생산 ²	No	No	No	Yes
관측치	32	32	32	32
결정계수	0.740	0.744	0.764	0.771
MAPE	0.128	0.126	0.080	0.090
NRMSE	0.439	0.508	0.477	0.572

주: 괄호 안은 이분산성에 강건한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다. (분기더미×국내총생산), (분기더미×국내총생산²), (분기더미×일인당소득)은 각각 분기더미와 로그 국내총생산, 로그 국내총생산 제곱항, 로그 일인당소득의 교호항을 의미한다. NRMSE는 정규화 제곱근평균제곱 오차를 의미한다.

〈표 4-19〉 모형 (1)~(4) 추정결과 및 표본외 검정결과 (분석에 사용한 기간: 2010q1~2018q4)

변수	(1)	(2)	(3)	(4)
로그 국내총생산	2.553*** (0.267)	26.64 (95.32)	2.386*** (0.736)	29.03 (99.62)
(로그 일인당소득) ²		-0.792 (3.136)		-0.877 (3.286)
상수항	-21.06*** (4.074)	-204.2 (724.3)	-18.52 (11.20)	-220.9 (755.0)
분기더미	Yes	Yes	Yes	Yes
분기더미×국내총생산	No	No	Yes	Yes
분기더미×국내총생산 ²	No	No	No	Yes
관측치	36	36	36	36
결정계수	0.775	0.775	0.794	0.795
MAPE	0.126	0.126	0.089	0.088
NRMSE	0.431	0.458	0.409	0.442

주: 괄호 안은 이분산성에 강건한 Eicker-Huber-White 표준 오차이다. ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다. (분기더미×국내총생산), (분기더미×국내총생산²), (분기더미×일인당소득)은 각각 분기더미와 로그 국내총생산, 로그 국내총생산 제곱항, 로그 일인당소득의 교호항을 의미한다. NRMSE는 정규화 제곱근평균제곱 오차를 의미한다.

2. 예측 결과 비교

〈표 4-20〉은 본 연구에서 표본외 검정을 통해 선정한 예측 결과와 국민여행조사 국내여행 추정치 및 과거 연구원에서 생산한 예측결과를 비교하였다. 다만 〈표 4-20〉의 1열은 추정 구간을 달리한 식 (1)의 예측 결과를 추가적으로 제시하였다.

〈표 4-20〉 예측 결과 비교

(단위 : 천회, 천일)

연도	(1) 〈표 4-10〉 2열 예측	(2) 〈표 4-17〉 3열 예측	(3) 〈표 4-18〉 3열 예측	(4) 〈표 4-19〉 3열 예측	(5) 국민여행 조사 (여행횟수)	(6) 국민여행 조사 (여행일수)	(8) 2000년 국내관광 예측	(9) 2011년 국내관광 예측
2002	-	-	-	-	-	-	404,648	-
2005	-	-	-	-	257,790	388,837	507,436	-
2006	-	-	-	-	284,575	416,982	-	-
2007	-	-	-	-	302,861	477,372	-	-
2008	-	-	-	-	245,669	408,027	536,876	-
2009	-	-	-	-	219,586	375,341	-	-
2010	-	-	-	-	168,148	339,608	-	-
2011	-	-	-	-	156,594	286,947	605,968	473,833
2012	-	-	-	-	213,468	365,282	-	492,617
2013	-	-	-	-	231,035	389,220	-	512,220
2014	-	-	-	-	227,100	397,847	-	532,683
2015	-	-	-	-	238,297	406,819	-	554,055
2016	269,055	-	-	-	241,750	412,378	-	576,347
2017	293,721	276,445	-	-	284,966	479,674	-	599,637
2018	317,406	295,451	298,964	-	311,153	556,890	-	623,960
2019	335,502	310,117	314,607	319,311	344,750	585,710	-	649,367

주: 2000년, 2011년 예측은 내부자료 수치이다. 각 예측 값은 해당 연도의 실질 국내총생산에 대입하여 계산하였다. 실제 예측에서는 한국은행이나 한국개발연구원의 경제성장률 전망을 활용할 수 있다.

〈표 4-20〉의 제8열과 제9열의 예측 결과는 과거 정책적 요청에 따라 연구원 내부에서 국내관광 수요 예측을 수행한 결과이다.⁵⁶⁾ 제1열부터 제4열까지의 수치는 본 연구에서 표본외 검정 결과, 예측력이 높은 모형들을 사용하여 예측한 결과이다. 예측치는 실제 값과 상당한 유사성이 있었다. 다만 2018년, 2019년 국민여행조사 결과를 비교하여 보면, 예측 결과가 모두 작게 추정되었음을 확인할 수 있다.

56) 예측이 수행된 이후로 많은 시간이 경과되어 구체적인 방법론이나 데이터는 보존되어 있지 않다.

2018-2019년 조사결과에서 국민여행이 증가한 원인에 대해서 여러 가지 설이 존재하지만, 2018-2019년 조사는 이전 조사와 많은 면에서 큰 차이가 발생하였기 때문에 시계열 단절이 일어났다는 가설은 매우 설득력이 높다. 본 보고서의 결과는 2018년의 국내여행 수치가 증가는, 실제로 국내여행이 증가한 것이 아니라 시계열 단절로 인한 변동임을 시사하고 있다. 따라서 국민여행조사에 시계열 보정이 필요하다 판단되지만 현재의 결과가 과대추정인지 과거의 결과가 과소추정인지에 대해서는 추가적인 검토가 필요하다.

또한 제8열, 제9열에 제시된 연구원의 과거 예측 결과와 비교하면 오차가 상당히 작은 수준임을 알 수 있다. 따라서 본 연구에서 제시한 모형은 과거 사용되어오던 시계열 모형에 비하여 예측력의 측면에서는 상당한 개선이 있었다고 보인다. 다만 과거에도 국민여행조사 기관이 변경되면서 시계열 단절이 발생하였고, 글로벌 금융 위기로 인한 경제충격이 있었으며, 과거 예측결과는 10년 이후의 미래를 예측하여 구조적으로 오차가 많을 수 없다는 점은 감안하여야 할 것이다.

그렇다고 하더라도 과거의 예측 결과는 거의 항상 양의 편의를 가지는 것으로 보이는데, 문제의 원인은 시계열 모형을 사용하였기 때문에 과거의 증가추세만이 반영된 결과일 수 있다. 즉, 우리나라가 급격히 성장하던 시기에는 과거의 추세를 반영하는 시계열 모형으로 정책 수립에 적합한 예측 결과를 도출할 수도 있었으나, 지금은 경제가 상대적으로 저성장하면서 인구 감소도 예상되는 만큼 과거의 추세로 빠르게 성장함을 가정하는 예측 모형은 적절하지 않을 수 있다. 앞으로는 예측에서 과대 추정이 발생하지 않도록 '수요 포화'나 계수의 '시간 변동'과 같은 개념을 도입한 예측 모형을 연구할 필요성이 있다.

또한 현재 세계를 휩쓸고 있는 코로나바이러스감염증-19로 인하여 우리나라에서도 국민의 건강을 위한 많은 조치들이 실시되고 있다. 또한 소강상태로 접어들었던 많은 국가에서 코로나바이러스가 재유행하고 있으며, 불행하게도 우리나라 역시 예외가 아니다. 따라서 과거의 데이터를 사용하여 2020-2021년에 대해 어떠한 예측 결과를 내놓더라도 사실과는 상당한 괴리가 존재할 것으로 예상된다. 이는 예측 모형의 한계라기보다 과거의 데이터로 예측할 수 없는 매우 특수한 형태의 충격으로 이해하여야 할 것이다. 따라서 본 보고서는 코로나바이러스의 영향이 상당하리라 예상되는 2020-2021년에 대해서는 별도의 예측치를 제시하지 않는다.

국민 국내관광 수요의 모형화 및 예측연구

제5장

요약 및 결론

제1절 연구결과 요약

본 보고서에서는 시의적절하고 정확한 국내관광 수요예측으로 정책적 요구 충족에 도움을 주기 위한 첫걸음으로, 지난 기간 적극적으로 수행되어오지 못하였던 국내관광 수요의 모형화 및 예측연구를 새롭게 추진하였다. 과업 내용은 크게 다음과 같다. 첫째, 국내관광 수요 예측과 관련한 연구에 대하여 문헌조사를 수행하였다. 둘째, 수요예측 절차 및 그에 따라서 모형을 개발하는 과정을 제시하였다. 셋째, 교차 검증을 통해 예측력이 우수한 것으로 평가된 모형으로 예측치를 제시하였다.

문헌조사는 일부 국내 문헌의 경우 주요 관광 지역이나 지점을 예측하는 연구도 포함하였는데, 이는 국내외를 막론하고 국내관광 수요의 예측에 대한 연구가 매우 적었기 때문이다. 다수의 국내외 연구(김상원·이덕순, 2013; 현수영 외, 2015; 김상원·박미선, 2016; 김상원, 2018; Witt, 1992; Athanasopoulos and Hyndman, 2008; Athanasopoulos et al., 2014)를 검토한 결과 국내의 연구들은 대부분 시차변수 만을 사용하는 시계열 모형을 사용하는 반면에, 외국의 연구들은 단순 시계열 모형보다는 소득, 관광 가격, 항공 요금, 숙박 요금을 설명변수로 하는 회귀 모형을 활용하는 것으로 나타났다. 다만 국내 연구도 최근에는 인과 모형을 같이 검토하고 있었다. 특히, 호주에서는 국가적으로 국내외 관광 수요에 대한 정교한 예측을 수행하고 있었으며, 따라서 호주의 연구자들도 국내관광 수요예측에 대하여 여러 가지 연구를 발표하고 있었다.

문헌조사를 검토하고 전문가 자문을 바탕으로 하여, 본 보고서도 회귀 모형을 사용한 예측을 시도하였다. 종속변수로는 국내관광 횟수를 사용하였고 설명변수로는 국내총생산, 인구, 공휴일수를 검토하였으며 분기별 더미 변수로 국내관광의 계절성을 반영하였다. 모형 설정은 모수 절약의 원칙(parsimony of parameters)에 따라 제한된 수의 변수를 사용하였다. 다양한 회귀 모형을 검토한 결과 표본내(in-sample)에서의 설명력이 충분하다고 보이는 경우가 예측력으로 이어지지는 않

으며, 국내총생산은 많은 회귀 분석에서 상당한 설명력과 예측력을 보여주었다.

따라서 최종적으로는 국내총생산을 사용한 여러 가지 예측 모형의 예측력을 검토하였다. 예측력의 검토는 표본외 데이터에 대해서 평균절대백분율오차(MAPE)와 정규화 제곱근평균제곱오차(NRMSE)를 계산하여 여러 예측 모형을 비교하였다. 표본외 검정결과 국내관광 수요예측은 분기별 더미와 로그 국내총생산의 교호항을 고려하는 다음 회귀모형의 예측력이 가장 높은 것으로 나타났다. 또한 예측 결과에서 2018년 이후의 국내관광 증가는 실제로 국내관광이 증가한 것으로 해석하기 보다는 조사방식의 변경에 의한 시계열 단절일 가능성이 높다고 추측된다.

제2절 제언

본 보고서는 우리나라에서 그동안 수행되지 않았던 국내관광 수요예측을 수행하였으며, 그 과정에서 발견한 정책제언 및 국내관광 수요예측의 발전 방향에 대하여 언급하고자 한다.⁵⁷⁾

□ 정책제언

1) 장기 시계열 확보

연구 결과, 시의성 있고 정확한 국내관광 수요예측을 수행하여 정부의 정책수립을 지원하는데 있어서 가장 크고 시급한 처리가 필요한 걸림돌은 데이터의 부족이라고 판단된다. 국민여행조사가 개편되고 표본이 수도 크게 증가하고 있으나, 현재의 긍정적인 변화가 과거의 데이터를 향상시키는 것은 아니다. 예측에는 과거의 정보도 필요하기 때문에 과거의 시계열을 보정하거나 국내관광을 나타낼 수 있는 동시에 과거의 수치를 제시할 수도 있는 지표의 개발이 필요하다.

2) 정보 수집기능 강화

호주의 사례에서 나타나는 바와 같이, 호주 정부의 수요예측에서는 교통 가격, 숙박 가격과 같은 다양한 정보들이 활용되고 있다. 그러나 우리나라에서는 지금까지 이러한 정보를 수집할 수 있는 근거가 미약하기 때문에, 필요성에도 불구하고 수집되지 못하고 있다. 문화체육관광부가 관광과 관련한 정보를 제공받거나 수집할 수 있는 법적인 근거를 마련하고 해당 업무를 수행할 수 있는 별도의 조직과 인력을 충원하는 것이 바람직하다.

57) 문제점·한계 및 그에 대한 제언은 해당 장절에서도 언급하기도 하였으나, 여기서는 이를 종합적으로 언급한다는데 의의가 있다.

또한 국내관광과 관련이 있는 외부 기관과의 협력관계 구축도 정보 수집에 큰 도움이 될 것이다. 특히 관광의 정의상 이동을 필수적으로 수반함에도 불구하고, 아직까지 교통과의 협력관계 구축은 충분히 이루어지지 못하고 있다. 국내관광과 관련하여 다양한 연구를 공동으로 수행하고 정보를 공유하는 방안을 고민할 필요가 있다. 관련 협회와의 협력도 필수적이다. 호주의 사례를 참고로 하여 협회는 가격정보를 문화체육관광부에 제공하고, 정부와 데이터를 제공하는 협회가 공동으로 국내관광 수요예측 결과를 공유하는 위원회를 구성하는 방안도 검토할 필요성이 있다.

3) 데이터 댐 구축

데이터의 품질을 관리하고 필요한 적기에 공급하는 데이터 댐을 구축할 필요성이 있다. 빅데이터의 인기로 인하여 외부에서는 데이터 댐의 역할로 주로 방대한 데이터의 저장이나 데이터 사이의 연계 및 통합과 관련한 이야기들을 언급하고 있다. 하지만 관광분야의 데이터 댐 구축은 데이터의 연계 이상을 고려하여 설계하여야 한다. 다목적 댐의 목적은 단순히 수자원을 저장하기만 하는 것이 아니다. 저장되어 있는 수자원의 질도 관리하여야 하며, 적절한 시기에 방류하고 전기를 생산하는 것도 중요하다. 데이터 댐에서 잘 관리된 데이터를 기반으로 하는 수요예측이나 관련 연구는 정부의 정책 의사 결정에도 도움이 될 것이다.

현재는 여러 사이트에 산재되어 있는 정보를 개인이 취합하여 데이터를 구축하고 있으나, 관광 데이터 댐은 관광 데이터뿐만 아니라 연구에 필요한 데이터를 댐에서 한꺼번에 제공하는 형태로 구축하는 방안을 검토하여야 한다. 또한 이러한 형태의 데이터 제공이야말로 데이터 댐의 취지에도 부합하는 형태일 것으로 판단한다. 여러 정보를 취합하여 완성된 형태로 데이터를 제공하는 것은, 정부의 정책적 수요예측 요구에 대한 즉시적인 대응력을 크게 향상시킬 것으로 기대한다. 또한 장기적으로는 자체적으로 데이터 댐에서 데이터를 분석하고 예측이 가능하도록 데이터 댐을 발전시켜 나아가야 할 것이다.

데이터 댐에서 저장된 데이터를 제공하는 방식은 세계은행(World Bank)이나 국제통화기금(International Monetary Fund)을 참고로 하여, 이용자가 데이터 댐의 정보를 사용하기에 편리한 형태로 제공하는 방안도 같이 검토할 필요가 있다.

□ 모형의 발전방향

국내관광 수요예측은 장기적으로 시간가변계수모형(time-varying coefficient model), 상태공간모형(state-space model)과 같이 동적 움직임을 나타낼 수 있는 모형과 결합한 형태나 통계적 학습(statistical learning)을 통해 예측을 수행하는 방향으로 발전하는 것이 바람직하다고 보인다. 가격변수나 부채변수 같은 추가적인 경제적 변수의 확보도 모형의 설명력과 예측력을 향상시킬 수 있을 것이다.⁵⁸⁾

다만 앞에서 언급한 바와 같이 여러 고급 시계열 기법을 적용하기에는 국내관광 시계열의 길이가 충분하지 않다고 생각되며, 이러한 기법의 적용을 위해서는 선제적으로 장기 시계열 확보가 필요하다고 보인다. 본 보고서에서 사용한 고전적인 점 추정(point estimation)과 베이저안 추정(Bayesian estimation)은 많은 면에서 상이하기 때문에 같이 검토하지 못하였다. 이러한 고급 기법들은 앞으로 심도 있는 연구를 진행할 필요가 있다.

또한 데이터는 매년 축적되기 때문에, 데이터를 추가할 때마다 예측 모형도 매년 달라진다. 따라서 한 번 만들면 끝나는 것이 아니라 장기간 예측 결과를 실제 값과 비교하고 개선하는 과정이 요구되며, 이러한 과정을 이해하고 지원하여 주는 관행이 정착될 필요성이 있다.

58) 다만, 예측과 별개로 미시 자료 확보 또는 미시 자료 가공이 필요하기 때문에 본 연구에서 시도되지는 못하였다.

참고문헌

- 권지호·김도완·지정구·김건·노경서 (2019). 우리나라의 잠재성장률 추정. 조사통계월보, 72(8), 16-32.
- 관계부처합동 (2019). 관광혁신 추진성과 및 과제. 국가관광전략회의 발표자료.
- 기획재정부 (2020). 서비스산업혁신 관계부처 TF 첫 회의 개최. 보도자료.
- 김도완·이상협 (2019). 산업별 노동생산성 변동요인 분석. 조사통계월보, 73(3), 16-38.
- 김상원·이덕순 (2013). 익산시 관광객 수요예측: 윈터스지수평활법을 이용하여. 동북아관광연구, 9(4), 151-169.
- 김상원·박미선 (2016). 수원시 관광객 수요예측 모형의 정확도 비교. 동북아관광연구, 12(4), 121-142.
- 김상원 (2018). 제주특별자치도 관광수요 예측모형의 정확도 비교. 동북아관광연구, 14(2), 219-237.
- 김형중 (2019). 대기오염이 관광수요에 미치는 영향. 한국문화관광연구원.
- 김형중 (2020). 관광 시장 규모 및 기여도 추산. 내부자료.
- 문혜정·문수정 (2018). 2017년 잠정 GDP 통계 계절조정계열 작성 결과.
- 문화체육관광부 (2020). 2019 국민여행조사.
- 문화체육관광부·한국문화관광연구원 (2020). 2019년 주요관광지점 입장객통계.
- 박상곤 (2017a). 국내관광 수요예측 모형 개발. 한국문화관광연구원.
- 박상곤 (2017b). 연휴가 관광수요에 미치는 영향. 한국문화관광연구원.
- 박준용·김인무·김창식·이성로 (2011). 선도추급과정을 이용한 새로운 예측기법: 장기전력수요예측에의 응용. 경제학연구 59(3), 113-147.
- 이강욱 (1999). 한국 관광계량 모형(KTRI-99) 구축. 한국문화관광연구원.
- 이강욱 (2017). 국제관광수요 영향요인 분석. 한국문화관광연구원.
- 이강욱·이성태·黑順宏志·손용훈 (2009). 한·일 관광시장의 영향요인 분석 및 중·장기 전망. 한국문화관광연구원.
- 이강욱·최승묵·권태일 (2010). 환율변화에 따른 관광수요 변화 분석. KCTI 문화예술포럼 관광 동향분석 제5호.

- 이수일 (2019). 관광산업의 도약, 무엇을 어떻게 할 것인가? 선진국형 서비스산업 발전방향: 서비스산업 업그레이드로 삶의 질 높이고, 일자리도 창출. 발표자료.
- 이연택 (2003). 관광정책론. 일신사.
- 이연택 (2015). 관광학. 백산출판사.
- 이종연·이현정·박미수·김진경·성기택·최홍빈·유승훈·최효연·변태근·정세리 (2015). 문화관광 부분 사업의 예비타당성조사 표준지침 연구. 한국개발연구원 공공투자 관리센터.
- 이충기 (2011). 관광응용경제학. 대왕사.
- 장병권 (2003). 국민관광론. 기문사.
- 전효재 (2003). 국제관광 수요예측. 한국문화관광정책연구원.
- 전효재·김하니 (2020). 거시경제 변수와 서비스업생산지수가 관광수입에 미치는 영향에 관한 연구. 관광레저연구, 32(2), 21-37.
- 조아라·김형중 (2019). 국민의 해외여행 동향 분석. 한국문화관광연구원.
- 최승묵(2017). 광역지자체 국내 관광여행 성장 특성 분석 국내 관광여행 총비용 및 이동총량 자료를 중심으로. 관광연구논총, 29(2), 3-23.
- 통계청 (n.d.). 국가통계포털. <http://kosis.kr>
- 한국관광학회 (2019). 문화관광론. 백산출판사.
- 한국은행 (n.d.). 경제통계시스템. <https://ecos.bok.or.kr>
- 현수영·손윤환·유상욱 (2015). 제주도 관광객 수요 변화와 예측에 대한 연구. 통계연구, 19, 1-24.
- Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. International Journal of Forecasting, 25(1), 146-166.
- Athanasopoulos, G., Deng, M., Li, G., & Song, H. (2014). Modelling substitution between domestic and outbound tourism in Australia: A system-of-equations approach. Tourism Management, 45, 159-170.
- Athanasopoulos, G., & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. Tourism Management, 29(1), 19-31.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis, Forecasting and Control (5th ed.). John Wiley and Sons.
- Engle, R. F., & Granger, W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. Econometrica, 55(2), 251-276.

- Hyndman, R. J. (2006). Forecast: Forecasting time series. R package version 1.0. <http://www.robhyndman.info/Rlibrary/forecast>
- Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2005). Prediction intervals for exponential smoothing state space models. *Journal of Forecasting*, 24(1), 17-37. <https://doi.org/10.1002/for.938>
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18, 439-454.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. New York: Springer.
- Ledolter, J., & Abraham, B. (1981). Parsimony and its importance in time series forecasting. *Technometrics* 23(4): 411-414.
- Manning, W. G. (1998), The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of health economics*, 17(3), 283-295. [https://doi.org/10.1016/S0167-6296\(98\)00025-3](https://doi.org/10.1016/S0167-6296(98)00025-3)
- Song, H., Li, G., Witt, S. F., & Fei, B. (2010). Tourism demand modelling and forecasting: How should demand be measured? *Tourism Economics*, 16(1), 63-81. <https://doi.org/10.5367/000000010790872213>
- Song, H., Witt, S. F., Li, G. (2009). *The Advanced Econometrics of Tourism Demand*. New York: Routledge.
- Stock, J. H., & Watson, M. W. (2020). *Introduction to Econometrics* (4th ed.). Pearson.
- Tourism Research Australia (2019). *Tourism Forecasts*.
- United Nations (2010). *International Recommendations on Tourism Statistics (IRTS)*.
- Witt, S. F. (1980). An abstract mode-abstract (destination) node model of foreign holiday demand. *Applied Economics*, 12(2), 163-180.
- Witt, S. F., Newbould, G. D., & Watkins, A. J. (1992). Forecasting domestic tourism demand: application to Las Vegas arrivals data. *Journal of Travel Research*, 31(1), 36-41.
- Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach* (7th ed.). Cengage.

ABSTRACT

Modeling and Forecasting Korean Domestic Tourism Demand

This policy report deals with modeling and forecasting Korean domestic tourism demand. As South Korea's economy grows, domestic tourism receipts are nearly doubling that of inbound tourism. Although domestic tourism contributes considerably to the economy, awareness of the domestic tourism market has been mostly indifferent. Consequently, in South Korea, there are only a few domestic tourism forecasting studies in terms of national aggregates, whereas numerous studies have been conducted on forecasting tourists for regional tourism attractions. Nonetheless, establishing appropriate tourism policy responses by the relevant authority requires accurate statistics and forecasts. Thus, there is a need to explore suitable forecasting methods for domestic tourism. Herein, we attempt to fulfill this need. In this report, we model and forecast Korean domestic tourism demand by adapting a linear regression framework. The limitations of both forecasts and the current statistics on domestic tourists are also addressed. The report's organization is as follows: Chapter I provides the background and aim of the report. Chapter II introduces the concept of domestic tourism and the forecasting procedure. In Chapter III, we conduct a literature review for forecasting domestic tourism demand. Chapter IV explains the dataset, provides the regression specifications, and presents both model validations and resultant forecasts. Chapter V concludes.

Keywords

Cross Validation, Domestic Tourism, Forecasting, Out of Sample Test, Time Series

국민 국내관광 수요의 모형화 및 예측연구



부록

【부록1】 수요예측 모형 설명변수 개요

본 보고서에서는 선행연구(박상곤, 2017a) 분석 및 호주 정부 국내관광 수요예측 사례, 전문가 의견 등을 바탕으로 ‘소득’, ‘인구’ ‘공휴일 수’ 등을 국내관광 수요예측을 위한 설명변수로 선정하였다. 소득의 대리변수는 국내총생산(GDP)으로 하였으며, 인구는 행정안전부의 주민등록인구 현황 자료를 활용하였다. 또한 공휴일, 대체공휴일, 토요일 등의 합을 공휴일 총량으로 변수화 하여 국민의 여가시간이 국내관광 수요에 미치는 영향을 확인하고자 하였다.

〈분석 대상 기초자료〉

구분	변수명	주요변수	활용방안
1	소득	국내총생산 (비계절조정, 시장가격)	국민 국내관광 수요예측 설명변수
2	인구	주민등록인구현황	국민 국내관광 수요예측 설명변수
3	공휴일 수	공휴일, 대체공휴일, 토요일 등의 합계	국민 국내관광 수요예측 설명변수

가. 국내총생산

1) 개요

국민소득통계는 한국은행법 제86조, 통계청 승인번호 제30109호에 근거하여 생산되는 지정통계이다. 국민소득통계는 일국의 경제력 또는 그 국민의 생활수준을 가늠하는 기준점이 되는 동시에 경제 상황에 대한 각종 분석지표를 작성하여 경제 정책을 수립하고 평가하는데 기초자료로 유용하게 활용되고 있다. 국내총생산(GDP)은 국민소득 3면등가원칙에 따라 생산접근방법으로 생산된 생산국민소득 지표로서 일정한 기간 동안 우리나라 국내에서 생산된 모든 부가가치의 합계를 의미한다. 국내총생산(GDP)은 분기 및 연간 주기로 작성되며 농림어업, 광업, 제조업 등 15개 대분류로 구성된다. 2019년 우리나라의 국내총생산(시장가격, GDP)은 1분기 473조 1,379억원, 2분기 479조 9,356억원, 3분기 481조 4,105억원, 4분기 484조 5,559억원 등으로 연간 기준1,919조 399억원으로 나타났다.

〈국내총생산(GDP) 개요〉

구분	주요 내용
조사근거	한국은행법 제86조, 통계청 승인(30109)
조사주기	분기 및 연간
조사대상	전국의 가계, 기업, 정부 등 경제활동을 수행하는 모든 경제주체
조사방법	통계작성기관으로부터 기초통계를 입수하거나 조사대상 경제주체를 대상으로 직접적인 설문 조사
적용분류	15개 대분류: 농림어업, 광업, 제조업, 전기가스 및 수도사업, 건설업, 도소매 및 음식숙박업, 운수 및 보관업, 금융 및 보험업, 부동산 및 임대업, 정보통신업, 사업서비스, 공공행정 및 국방, 교육서비스업, 보건 및 사회복지서비스업, 문화 및 기타서비스업
조사(작성) 기준시점	분기 속보치: 해당분기 종료 후 1일~20일경 분기 잠정치: 해당분기 종료 후 25일~65일경 연간 잠정치: 익년 4월~5월 20일경 연간 확정치: 익년 4월~5월 20일경

자료: 한국은행 (n.d.)을 바탕으로 재정리

2) 주요 연혁 및 변동 사항

국민소득통계는 1957년 한국은행이 공식 편제기관으로 지정됨에 따라 국민소득추계를 전담하고 있으며, 1958년부터 UN이 권고한 편제기준(1953 SNA)에 따라 국민소득통계를 추계하였다. 1995년에는 경제성장률의 중심지표를 국민총생산(GNP)에서 국내총생산(GDP)으로 변경하여 국민소득통계의 대표 지표로 활용하고 있다.

〈국민소득통계 국내총생산(GDP) 연혁〉

구분	주요 내용
1957년	• 한국은행이 공식 편제기관으로 지정되면서 국민소득추계를 전담
1958년	• 1953~1957년까지 각 년의 '산업별 국민총생산(GNP)'과 '국민총생산에 대한 지출'을 1953 SNA에 따라 추계
1995년	• 경제성장률의 중심지표를 국민총생산(GNP)에서 국내총생산(GDP)으로 변경
1999년	• 국민총소득(GNI) 도입 등 1993 SNA를 일부 반영하였으며 IMF의 특별통계 공표기준(SDDS)에 따라 통계공표일정을 사전에 지정. 또한 분기 명목 국내총생산 및 계절변동조정 통계를 새로이 공표하기 시작
2005년	• 국내총생산 분기 속보치 발표
2006년	• 분기 국내총생산의 주지표를 원계열 전년동기비에서 계절조정계열 전기비로 변경
2009년	• 국제기준(1993 SNA)에 맞춰 실질 국내총생산(GDP) 추계방법을 기존의 고정가중법에서 연쇄가중법으로 변경 • 지출국민소득(GDE) 추계방법을 민간소비, 재고증감 등 지출 항목을 직접 추계하는 지출 접근법으로 변경 * 기존에는 생산국민소득을 기반으로 하는 상품흐름법 활용

자료: 한국은행 (n.d.)을 바탕으로 재정리

다) 통계 작성 절차

한국은행은 분기별로 실질 국내총생산 통계를 작성하고 있다. 일부 부문의 경우 해당 분기 2개월 실적 자료와 분기 말월 결측치의 추정자료를 사용하여 잠정치 편제방법과 유사한 방법으로 작성하고 있다.⁵⁹⁾ 분기 실질 국내총생산(GDP)은 경제활동별로 부가가치를 추계하는데, 산업별 또는 재화별 국내총산출물의 가치, 즉 총산출액(gross output)을 먼저 추계한 뒤 생산에 투입된 원재료와 기타비용, 즉 중간투입액(intermediate input)을 총산출액에서 공제하여 국내산출물의 부가가치(value added)를 산출한다.⁶⁰⁾

$$\text{부가가치} = \text{총산출} - \text{중간투입} = [\text{생산수량} \times \text{가격}] \times \text{부가가치율}$$

또한, 추정의 정확도 제고를 위해 달력효과가 있다고 판단되는 계열에 대해서는 공휴일수 효과, 명절 효과 등 회귀변수의 유의성을 검정한다(달력효과(calendar effect) 조정). 한편, 계절변동조정 국내총생산 자료는 미 상무부의 X-13-ARIMA를 우리나라 실정에 맞도록 조정한 방법을 이용하여 작성하는데, 원계열에 명확한 계절성이 존재하는 경우 계절조정을 실시한다. 계절조정계열에는 잔여계절성이 남아 있지 않아야 하며, 계절조정계열은 최대한 안정성이 확보되도록 작성하는 것을 원칙으로 한다.

4) 분석자료

본 연구에서는 국민 국내관광 수요예측의 설명 변수로서 소득 데이터를 활용하였다. 소득 데이터는 한국은행 경제통계시스템에서 제공하고 있는 두 종류의 국내총생산(계절조정 및 원계열)을 각각 사용하였으며, 국내관광 변수의 분석범위를 고려하여 2004년 7월부터 2019년 12월 까지 총 15년 6개월을 분석 대상으로 하였다. 데이터의 단위는 분기자료로 전체 분석범위 내 62개 시계열을 포함하고 있다.

〈국내총생산(GDP) 분석자료 개요〉

활용자료	분석범위	데이터 단위	시계열 수
국내총생산(계절조정, 실질, 분기) 국내총생산(원계열, 실질, 분기)	2004.07. ~ 2019.12.	분기	62개

59) 분기 말월 결측치의 추정은 기업체 등에 대한 모니터링과 이동평균법, 추세분석, 회귀분석 등으로 실시하고 있다.

60) 부가가치(value added)란 경제주체들이 노동, 자본, 토지 및 경영 등의 생산요소를 투입하여 생산활동을 수행한 결과로 발생한 새로이 늘어난 가치로서 급여, 이자, 임료 및 이윤 등의 합계와 같다.

나. 주민등록인구현황

1) 개요

주민등록인구현황은 주민등록법에 의한 주민등록인구 및 세대현황에 대하여 전국단위의 기관별(시도, 시군구, 읍면동), 연령별 현황통계를 집계 및 안내하는 통계로서 통계청 국가승인통계 승인번호 제110026호(2008.01.08.) 일반통계 / 보고통계로 분류되고 있다. 전국단위 주민등록인구 및 세대현황을 기반으로 교육, 교통, 복지, 선거, 조세, 지역개발 등 각 분야에 적시성 있는 통계를 제공하고 있으며, 국가 정책에 필요한 기초자료로 활용되고 있다.

〈주민등록인구현황 개요〉

구분	주요 내용
조사근거	• 통계법 제18조 규정에 의하여 승인된 일반통계(승인번호 제110026호)
작성주기	• 월별 작성
공표범위	• 전국
공표주기	• 월별 공표(작성기준 월 익월 1일)
공표항목	• 주민등록인구현황, 주민등록기준 읍면동별 출생자수, 사망자수, 세대원수별 세대수, 평균연령, 인구 증감을 등

자료: 통계청 (n.d.)을 바탕으로 재정리

2019년 기준 전국 주민등록인구 수는 전년대비 23,802명(0.1%) 증가한 5,184만 9,861명으로 경기도1,323만 9,666명, 서울특별시 972만 9,107명, 부산광역시 341만 3,841명, 경상남도 336만 2,553명, 인천광역시 2,957,026명 등의 순서로 나타났다. 우리나라 인구는 지난 10년간 133만 4,195명 증가하였는데, 연평균 0.3% 수준의 완만한 증가세를 나타내고 있다.

〈전국 주민등록인구 현황(2010-2019년)〉

(단위: 천명)

2010년	2011년	2012년	2013년	2014년	2015년	2016년	2017년	2018년	2019년
50,515	50,734	50,948	51,141	51,327	51,529	51,696	51,778	51,826	51,849

자료: 통계청 (n.d.)의 행정구역(시군구)별, 성별 인구수를 바탕으로 재정리

주: 주민등록인구는 절사 후 표기하였다.

2) 통계 작성 절차

주민등록인구현황은 주민등록인구 및 세대현황에 대하여 전국단위의 기관별(시·도, 시군구, 읍면동), 연령별 현황통계를 행정안전부의 주민등록정보시스템으로 자동 집계하고 있으며,⁶¹⁾ 변동자료에 대한 실시간 반영과 전산시스템으로 자동 집계를 통해 전국단위의 일관성 있는 통계를 제공하고 있다.⁶²⁾

3) 분석자료

본 연구에서는 국민 국내관광 수요예측의 설명 변수로서 인구 데이터를 활용하였다. 인구 데이터는 통계청 국가통계포털에서 제공하고 있는 주민등록인구현황 데이터를 적용하였는데, 분석범위는 2011년 1월부터 2019년 12월 까지 총 9년으로, 주민등록인구현황 월간데이터의 최대 구득기간을 고려하였다. 데이터의 단위는 월간자료로 전체 분석범위 내 108개 시계열을 포함하고 있다.

〈주민등록인구현황 분석자료 개요〉

활용자료	분석범위	데이터 단위	시계열 수
주민등록인구현황	2011.01. ~ 2019.12.	월	108개

61) 자동 집계 흐름: 시·군·구 주민등록시스템(자치행정과, 정보통신담당관실) → 행정안전부 주민등록전산정보센터(주민과).

62) 현재 시·도별로 개별 공표하는 주민등록인구통계는 작성시기와 발표시기 등이 시도별로 서로 상이하여 전국단위의 일관성 있는 통계확보가 어려운 실정임

다. 공휴일 수

1) 개요

우리나라에는 국민 전체를 대상으로 하는 공휴일에 관한 법률이 부재하며, 대통령령인 ‘관공서의 공휴일에 관한 규정’을 통해 정부기관이나 지방정부 등 관공서의 공휴일을 규정하고 있다. 동 규정 제2조는 관공서 공휴일을 각 호에 따라 규정하고 있으며, 제3조는 대체공휴일에 관한 사항을 규정하고 있는데 그 내용은 다음과 같다.

「관공서 공휴일에 관한 규정」

제2조(공휴일) 관공서의 공휴일은 다음 각 호와 같다. 다만, 재외공관의 공휴일은 우리나라의 국경일 중 공휴일과 주재국의 공휴일로 한다.

1. 일요일
2. 국경일 중 3·1절, 광복절, 개천절 및 한글날
3. 1월 1일
4. 설날 전날, 설날, 설날 다음날 (음력 12월 말일, 1월 1일, 2일)
5. 석제 <2005. 6. 30.>
6. 부처님오신날 (음력 4월 8일)
7. 5월 5일 (어린이날)
8. 6월 6일 (현충일)
9. 추석 전날, 추석, 추석 다음날 (음력 8월 14일, 15일, 16일)
10. 12월 25일 (기독탄신일)
- 10의2. 「공직선거법」 제34조에 따른 임기만료에 의한 선거의 선거일
11. 기타 정부에서 수시 지정하는 날

제3조(대체공휴일) ① 제2조제4호 또는 제9호에 따른 공휴일이 다른 공휴일과 겹칠 경우 제2조제4호 또는 제9호에 따른 공휴일 다음의 첫 번째 비공휴일을 공휴일로 한다.

② 제2조제7호에 따른 공휴일이 토요일이나 다른 공휴일과 겹칠 경우 제2조제7호에 따른 공휴일 다음의 첫 번째 비공휴일을 공휴일로 한다.

2) 주요 연혁 및 변동 사항

우리나라는 1949년 6월 4일 첫 공휴일이 제정된 이래 대체휴일제도의 실시(1959년 3월 27일)와 폐지(1960년 12월 30일), 관공서 주5일 근무제 시행에 따른 식목일, 제헌절 등 공휴일 제외(2005년 6월 30일), 설·추석연휴와 어린이날에 대한 대체휴일제도 실시(2013년 11월 5일) 등 사회 환경의 변화에 따라 공휴일의 확대와 축소를 반복해 왔다. 그러나 2013년 관공서의 공휴일에 관한 규정 일부개정예 따라 설날·추석연휴에 대한 대체휴일제도가 시행되고, 내수활성화 및 국민여가시간 확대를 목적으로 임시공휴일 지정을 활성화 하는 등 우리나라의 공휴일 수는 점진적으로 증가하는 추세이다.

3) 분석자료

본 연구에서는 국민 국내관광 수요예측의 설명 변수로서 공휴일 데이터를 활용하였다. 공휴일 데이터는 관공서의 공휴일에 관한 규정에서 명시하고 있는 공휴일 및 대체공휴일을 대상으로 하였으며, 주5일 근무제 정착에 따라 토요일을 함께 포함하여 산정하였다. 분석기간은 2004년 7월부터 2019년 12월까지 총 15년 6개월로 하였는데 국내관광 변수의 분석범위를 고려하였다. 데이터의 단위는 월간자료로 전체 분석범위 내 186개 시계열을 포함하고 있다.

〈공휴일 수 분석자료 개요〉

활용자료	분석범위	데이터 단위	시계열 수
공휴일 수	2004.07. ~ 2019.12.	월	186개

【부록2】 국내관광 대리변수 탐색

연구진의 지식 한계 내에서는, 현재 국내관광을 정기적으로 측정하는 조사는 본 연구원의 “국민여행조사”가 유일하다. 그러나 앞에서 고찰한 바와 같이 국민여행조사의 시계열은 단절되어 있으며, 단절을 고려하지 않고 과거 시계열과 연결하여 사용하는 경우에도 분기별 시계열이 40개를 넘지 않는 수준이다. 또한 ‘국민여행조사’는 국내관광 전수를 조사하는 것이 아니라 임의추출을 통해 국내관광을 추정하는 조사이다. 따라서 조사방식의 변경은 마치 본 연구의 예측 결과와 마찬가지로 상이한 결과로 귀결될 수 있다. 따라서 시계열 단절을 회피하면서 충분한 길이의 시계열을 가지는 국내관광 대리변수를 탐색할 필요성이 제기된다.

대리변수(proxy variable)란 직접적으로 관측이나 획득이 어려운 변수를 대신하여 사용할 수 있는 변수로 원래의 변수와 상당한 관련성이 있는 변수를 의미한다. 만약 국내관광과 상당한 관련성이 있으면서 시계열은 국민여행조사의 국내관광 시계열보다 긴 데이터가 존재한다면, 대리변수를 활용한 국내관광 수요예측이 가능할 것이다.

본 보고서에서는 다음의 두 가지 방법으로 국내관광과 관련성이 높은 대리변수 확보에 성공하였음을 밝힌다. 우선 한 가지 방법으로는 결정계수가 높은 일부 관광 지점을 선정하여 다시 국내관광을 종속변수로 하는 다중회귀분석을 수행하는 경우 결정계수가 높은 사례를 얻을 수 있었다.⁶³⁾ 다른 방법으로는 특정 지역의 관광지점을 합산하여 단순회귀분석(simple linear regression)을 수행하는 경우 높은 결정계수를 얻는 사례가 있었다.⁶⁴⁾ 다음은 대리변수의 작성을 위해 어떤 데이터를 어떻게 가공하였는지 기술한다.

본 보고서에서는 대리변수의 탐색을 위하여 관광지식정보시스템에 수록된 4,222개의 관광지점 입장객 통계를 국민여행조사의 국내관광 총량, 당일관광 총량, 숙박

63) 이러한 분석은 시험적인 것으로 다중회귀분석의 설명변수로 어떠한 지점을 선정할지 또한 어떠한 지점들을 선정하면 추가적인 설명력을 확보할 수 없는지에 대해서 추가적인 연구가 필요하다.

64) 서울의 관광지점 입장객을 모두 합산하는 경우, 당일관광을 종속변수로 합산변수를 설명변수로 하는 회귀분석의 결정계수는 0.999에 달한다.

관광 총량에 대해서 모두 회귀분석하였다. 회귀분석에서 계산되는 결정계수가 바로 두 변수의 움직임이 얼마나 유사한지에 대한 척도이기 때문이다. 본 보고서는 임의로 0.7의 결정계수를 기준으로 하여 그 이상의 값을 보고한 관광지점을 별도로 분류하였다.⁶⁵⁾ 그 결과는 부록으로 제시한다.

또한 선택된 관광지점을 세부적으로 들여다보면 특이한 패턴을 관찰할 수 있었는데, 선택된 대부분의 관광 지점은 당일관광을 종속변수로 하는 회귀분석에서 높은 결정계수를 보였다. 상대적으로 국내관광 총량 및 숙박관광 총량을 회귀분석하여 선정된 경우의 수는 작았다. 그러한 원인으로는 몇 가지 가능성을 생각할 수 있다. 우선 ‘주요관광지점 입장객통계’의 주요관광지점이 균등하게 선정되지 아니하고 당일관광에 관련된 관광지점으로 선정되었을 수 있다. 다른 하나로는 ‘국민여행조사’가 당일관광은 충실하게 조사되고 있으나 숙박관광이 충분하게 조사되지 못하였을 수 있다.⁶⁶⁾ 마지막으로 선정된 관광지점이 실제로 당일관광객이 많을 가능성 있다. 그러나 하나의 가능성이 다른 가능성을 완전히 배제하거나 충분한 설명을 제시하지는 못하기 때문에 이를 규명하기 위해서는 추가적인 연구가 필요하다.

그러나 국민여행조사와 매우 높은 관련성을 보여주는 국내관광 대리변수의 탐색에는 성공하였으나, 그러한 대리변수의 시계열은 기존의 국민여행조사와 큰 차이가 없었다. 따라서 새롭게 찾아낸 대리변수를 사용한 예측 결과는 본 보고서에서 제시하는 결과와 큰 차이점이 없으리라 예상되므로 별도로 시도하지 않았다. 본 보고서의 목적은 국내관광 수요예측의 모형화와 예측이지, 수요예측에 적합한 국내관광 데이터의 작성이나 확보가 아니기 때문이다.⁶⁷⁾ 그러나 본 보고서의 결과를 바탕으로 국내관광의 장기적인 변화 추이를 분석할 수 있는 국내관광의 장기 시계열 작성을 시도한다면 좋은 후속연구가 될 것이라고 생각하며, 추후 국내관광 데이터 확보를 위해서 국내관광 시계열의 보정과 함께 가장 먼저 시도되어야 할 과제이다.

65) 국내관광 총량의 경우 결정계수가 높은 회귀분석 결과가 부족하여 0.65를 기준으로 관광지점을 선택하였다.

66) 특이하게도 당일관광 회귀분석에서 높은 결정계수를 보인 일부 관광지역은 숙박시설이었다. 물론 이러한 관계가 가성적일 수도 있으며 이에 대해서는 추가적인 연구가 필요하다.

67) 물론 두 문제가 동일하지는 않으나 상당한 관련성이 있는 것은 사실이다. 하지만 좋은 데이터의 확보가 좋은 예측모형이나 예측력을 담보하지는 않음도 사실이다.

〈주요관광지점별 결정계수〉

지역	유형	관광지명	결정계수
서울(4개)	당일	경복궁	0.766
	당일	창경궁	0.787
	당일	창덕궁	0.728
	당일	덕수궁	0.789
	당일	국립중앙박물관	0.715
	당일	서대문역사박물관	0.709
	당일	서대문형무소 역사관	0.778
부산(1개)	당일	선릉, 정릉	0.702
	전체	감천문화마을	0.604
대구(1개)	당일	대구국제사격장	0.839
인천(2개)	당일	수도국산 달동네 박물관	0.784
	당일	인천대공원	0.735
광주(1개)	당일	자산유원지	0.721
대전(2개)	당일	계룡산수통골	0.868
	당일	드림랜드	0.760
울산(3개)	당일	롯데호텔 울산	0.715
	당일	올림피아 관광호텔	0.741
	당일	울산대곡박물관	0.733
세종(2개)	당일	IMGNational:	0.709
	전체	함강공원오토캠핑장	0.645
경기(5개)	전체	수원박물관	0.609
	숙박	유엔군초전기념비	0.706
	당일	유엔군초전기념비	0.842
	당일	안양베네스트CC	0.705
	당일	용인자연휴양림	0.711
강원(16개)	당일	금강산랜드	0.706
	당일	아쿠아랜드	0.785
	전체	용설호수문화마을	0.640
	전체	조병활문학관	0.624
	당일	포천아트벨리	0.786
	당일	허브아일랜드	0.710
	당일	고대산	0.787
	숙박	백운산자연휴양림	0.708
	전체	치악산국립공원	0.634
	당일	숲사랑홍보관	0.841
	전체	아바이마을	0.687
	당일	척산온천휴양림	0.793
	당일	한화리조트	0.716
	당일	코레스코 치악산	0.792
	숙박	국립두타산자연휴양림	0.747
	당일	오색그린야드호텔온천	0.759
충북(9개)	전체	상당산성자연휴양림	0.616
	당일	동양파크호텔 수안보	0.765
	당일	밀란호텔 수안보	0.706
	당일	부림장 수안보	0.724
	당일	신흥장 수안보	0.731
	당일	원저궁전모텔 수안보	0.721
	당일	크라온호텔 수안보	0.759

지역	유형	관광지명	결정계수
	전체	충북알프스자연휴양림	0.639
	전체	황정산자연휴양림	0.624
충남(2개)	당일	공주한옥마을	0.856
	당일	삼교호관광지	0.736
전북(5개)	전체	선유도	0.652
	당일	남원 항공우주천문대	0.821
	당일	사선대관광지	0.759
	당일	옥정호	0.719
	숙박	전북종합사격장	0.702
전남(15개)	전체	김대중 노벨상기념관	0.615
	당일	해안골프코스	0.705
	전체	엑스포해양공원EDG	0.656
	전체	여수 하멜전시관	0.676
	전체	순천드라마촬영장	0.648
	당일	순천자연휴양림	0.723
	당일	국립고흥청소년우주체험센터	0.778
	당일	나로도	0.756
	당일	팔령산	0.782
	당일	보성CC	0.705
	당일	도곡온천	0.720
	당일	무당산CC	0.813
	전체	백수노을전시관	0.604
	당일	완도타워	0.724
	전체	장산도	0.659
경북(4개)	당일	경주월드	0.702
	당일	석정온천	0.740
	당일	경보화석박물관	0.723
	당일	부경온천	0.715
경남(18개)	당일	꿈의궁전모텔	0.735
	당일	마금산온천관광지	0.799
	당일	EML	0.738
	당일	마금산온천관광지	0.764
	당일	창원자동차경기장	0.736
	당일	진해해양공원	0.742
	당일	용원CC	0.756
	전체	계도어촌체험마을	0.601
	전체	청천량해전공원	0.692
	당일	동부산CC	0.826
	당일	AOneCC	0.808
	당일	통도 PineEastCC	0.781
	숙박	산신봉	0.700
	당일	삼신봉	0.801
	전체	지산생태과학박물관	0.658
	전체	지산국립공원 내원	0.725
	전체	지산국립공원 여평	0.629
	당일	황금원송이마을	0.728
제주(3개)	당일	비자림	0.723
	당일	신영영화박물관	0.717
	당일	여미지식물원	0.730

집필내역

연구책임

김형중 관광정책연구실 부연구위원: 제1장, 제4장, 제5장, 연구총괄
강현수 관광정책연구실 연구원: 제2장, 제3장, 제5장

국민 국내관광 수요의 모형화 및 예측연구

발행인 김 대 관

발행처 한국문화관광연구원

서울시 강서구 금남화로 154

전화 02-2669-9800 팩스 02-2669-9880

<http://www.kcti.re.kr>

인쇄일 2020년 10월 19일

발행일 2020년 10월 19일

인쇄인 (사)한국장애인이워크협회 일자리사업장

ISBN 978-89-6035-826-3 93300

DOI <https://doi.org/10.16937/kcti.rep.2020.e17>

www.kcti.re.kr

Modeling and Forecasting Korean Domestic Tourism Demand



아래의 DOI 또는 QR코드를 통해
이 보고서를 무료로 다운로드할 수 있습니다.
<https://doi.org/10.16937/kcti.rep.2020.e17>

Hyoungjong Kim Hyunsoo Kang

