

거시적 교통수요 분석모형 개발

목차

1. 연구의 개요	1
가. 연구의 필요성 및 목적	/ 1
나. 연구의 범위 및 방법론	/ 2
2. 기존 연구 고찰	4
가. 국내문헌 연구	/ 4
나. 국외문헌 연구	/ 5
3. 교통부문 자료설명 및 자료특성에 따른 연구방법론	9
가. 교통부문 자료설명	/ 9
나. 자료특성에 따른 연구방법론	/ 11
4. 횡단면 자료를 이용한 교통수요 분석	13
가. 중력방정식을 이용한 교통수요 분석	/ 13
나. Count Data Model	/ 21
5. 시계열 자료를 이용한 교통수요 분석	33
가. 교통수요 동태적 변화 모형	/ 33
나. 자기회귀 시차모형을 이용한 교통수요 분석	/ 45
다. 오차수정모형을 이용한 교통수요 분석	/ 59
라. 기타 시계열 분석 모형	/ 64
6. 결론 및 시사점	77
가. 국가교통수요조사 및 DB 구축사업과의 연계방안 모색	/ 77
나. 연구결과 요약 및 시사점	/ 78
참고문헌	79

1. 연구의 개요

가. 연구의 필요성 및 목적

- 거시적 교통수요 분석모형은 국토종합개발계획, 국가기간교통망계획을 비롯한 각종 교통계획 및 물류계획의 효과적인 수립, 시행, 평가를 위해 필요한 교통수요를 분석하는 모형임
 - 화물 혹은 여객부문 교통수요가 사회경제적 요인에 어떠한 영향을 받고 있는지 혹은 장래 어떠한 추세를 보일 것인지를 분석하는 것은 교통 및 물류계획에서 중요한 기초자료로 활용될 수 있음
- 거시적 교통수요 분석모형에서는 소득, 유류가격, 통행요금 등의 사회경제적 변수가 화물 혹은 여객부문 교통수요에 어떠한 영향을 미치는 지를 분석하고 장래 이러한 변수 변화에 기인한 외부 충격(shock)이 가해졌을 때 교통수요가 어떠한 양상을 나타낼 지를 분석하는 모형을 정립할 것임
- 또한 장래 교통수요가 어떠한 영향을 받을 것인지를 다양한 계량경제학적인 분석모형을 구축할 것임
- 교통공학 및 교통계획에서 교통수요분석에 이용되는 4단계 교통수요 분석모형이 정립되어 있으며 널리 이용되고 있음
 - 그러나 계량경제학을 응용한 거시적 교통수요 분석모형은 이와 달리 체계적으로 연구된 바가 없으며 연구자의 자의에 의해서 분석되곤 하였음
 - 즉 공통된 연구방법론을 따르는 것이 아니라 연구자의 필요에 따라 다양한 방법론을 취하고 있으며 연구결과의 해석 또한 다양하게 이루어지고 있는 실정
- 따라서 본 연구에서는 거시적 교통수요 분석모형의 일관된 틀을 제공하고자 할 것이며 이를 통해 연구자들이 준용할 수 있는 연구방법론과 정책결정자들을 위한 검증방법론을 제공할 것임
- 구체적으로 본 연구에서는 교통부문에서 이용되는 자료 특성에 대한 설명을 할 것이며 이를 적용할 수 있는 연구방법론을 구분할 것임
 - 또한 이러한 방법론에 따라 횡단면 및 시계열 분석방법론으로 크게 구분하여 중력방정식(gravity equation)을 적용하는 횡단면 자료 분석 부분과 적분된 자기회귀 이동평균(Autoregressive Integrated Moving Average, ARIMA) 모형, 자기회귀 시차(Autoregressive Distributed Lag, ADL) 모형, 오차수정모형(Error Correction Model, ECM)을 적용하는 시계열 분석 기법으로 구분하여

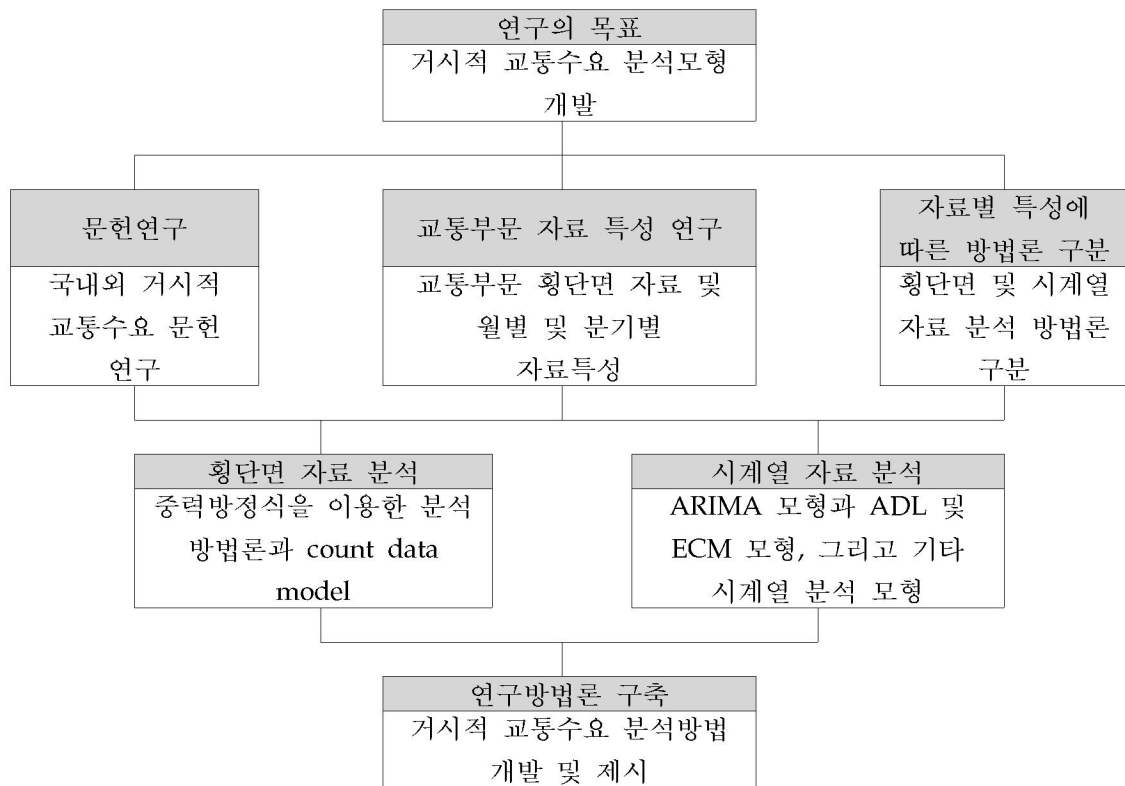
분석할 것임

- 이를 통해 자료별 특성에 따른 거시적 교통수요 분석 모형을 제공하여 연구자들에게 분석의 기초적인 틀을 제공할 것임

나. 연구의 범위 및 방법론

- 본 연구의 범위는 현시선호(Revealed Preference)에 입각한 교통실적 자료를 이용할 것임
 - 이용자들의 설문조사에 기반한 명시선호(Stated Preference)모형은 설문내용 및 설문에 응하는 이들의 판단에 크게 의존하는 모형으로 자의적인 해석이 나올 여지가 있음
 - 따라서 과거 실적자료에 입각한 교통수요분석 모형이 명시선호모형에 비해 교통이용자들의 행태를 좀 더 명확하게 분석할 수 있음
 - 물론 현재 및 장래에 다른 사회경제적 환경변화가 발생할 때는 과거 자료에 기반한 거시적 경제수요 모형이 한계가 있을 수 있음. 따라서 이러한 경우에는 전통적인 4단계 추정법을 이용해야 할 것임
 - 따라서 본 연구에서 수행하는 거시적 교통수요 분석모형은 전통적인 4단계 교통수요 분석모형과 서로 보완적인 관계를 이루면서 전개해야 할 것임
- 본 연구에서는 과거 교통수요 자료를 이용하지만 지나친 과거 자료 이용을 배제할 것임
 - 2000년대 이전의 교통실적 자료들은 최근의 교통이용 패턴과 판이한 양상을 나타내고 있으며 인구증가율 및 경제성장률도 최근과 상이한 수준을 나타내고 있음
 - 따라서 이러한 과거 교통실적 자료 이용을 통한 교통수요 분석은 장래 교통수요 예측에 심각한 오류를 야기할 수 있음
- 본 연구에서 이용되는 방법론은 계량경제학 방법론을 이용하여 분석할 것임
 - 최근 계량경제학은 다양한 주제 및 방법론을 이용하여 발전해 나가고 있으며 교통부문에서 적용할 수 있는 분야가 많은 편임
 - 따라서 일반적인 계량경제학 방법론 외에도 최근 방법론을 본 연구에서 소개하여 향후 연구자들이 이에 대해 좀 더 깊이 연구할 수 있는 지침을 제공할 것임

- 본 연구의 수행체계는 다음과 같음



<그림 1> 연구의 수행체계

2. 기존 연구 고찰

가. 국내문헌 연구

- 거시적 교통수요 분석 연구는 크게 2부분으로 구분할 수 있음
 - 횡단면 자료를 이용한 분석과 시계열 자료를 이용한 분석
- 횡단면 자료를 이용한 분석으로는 중력방정식(gravity equation)을 이용하여 통행요금 혹은 운임이 교통수요에 어떠한 영향을 미치는 지를 분석한 것임
- 이재민 외(2009)에서는 중력방정식을 이용하여 고속도로 통행요금에 대한 통행수요 탄력성을 분석하였음
 - 2008년 한국도로공사의 출발지 및 도착지 기준 통행실적을 바탕으로 중력방정식 모형을 적용하였으며 통행실적인 0인 경우를 감안하는 전통적인 토빗(Censored Tobit) 모형을 적용하였음
 - 고속도로 종별 통행요금에 대한 통행수요 탄력성을 추정하였는데 1종의 탄력성이 -0.15로 나타났으며 2종부터 6종까지의 탄력성이 약 -0.6~-0.5로 산정되었음
- 한국교통연구원(2011) 연구에서는 철도의 운임에 대한 탄력성을 중력방정식 모형으로 추정
 - 한국철도공사의 2011년 3월의 출도착지 기준 수송실적 자료를 이용하여 운임에 대한 통행수요 탄력성을 추정하였는데 -0.39로 산정되었음
- 시계열 자료를 이용한 분석으로 교통수요 동태적 변화를 계절성을 감안한 적분된 자기회귀 이동평균(Seasonal ARIMA, 이하 SARIMA) 모형을 이용하여 분석한 연구보고서와 논문이 있음
 - 한국교통연구원(2009)에서는 인천 지역 유료도로의 장래 교통수요를 ARIMA 모형을 이용하여 예측하였음
 - 한국교통연구원(2010)에서는 고속도로, 일반철도 및 수도권 전철 통행실적의 동태적 변화를 계절성을 감안한 ARIMA(SARIMA) 모형으로 분석하였음
 - 또한 이재민·권용재(2011)는 일반철도 통행실적의 동태적 변화를 SARIMA 모형으로 분석하였음
- 시계열 자료를 이용하여 사회경제적 변수 즉, 국내총생산, 유류가격, 통행요금 등이 교통수요에 어떠한 영향을 미치는 지를 분석한 보고서와 논문으로 다음을

들 수 있음

- 한국교통연구원(2004)은 연도별 한국도로공사 통행실적 자료를 이용하고 오차수정모형(ECM)을 이용하여 다양한 사회경제변수가 고속도로 통행수요에 어떠한 영향을 미치는 지를 분석하였음
- 이재민 외(2005)와 이재민·박수신(2005)은 연도별 자료 대신에 월별 한국도로공사 통행실적 자료를 이용하고 오차수정모형을 이용하여 사회경제변수로 인한 고속도로 통행수요 변화를 분석하였음
- 문진수·이재민(2007)은 연도별 철도화물 수송실적 자료와 오차수정모형을 이용하여 철도화물 수송수요 분석을 시도
- 한국개발원·한국교통연구원(2008) 연구에서도 2000년대 이후 한국도로공사의 월별 통행실적 자료와 오차수정모형(ECM)을 이용하여 국내총생산, 유류가격, 통행요금이 고속도로 통행실적에 미치는 영향을 분석하였음
- 한국교통연구원(2009)에서는 제주공항 항공수요 미래 예측을 위하여 2000년대 월별 제주공항 이용객수와 국내총생산 및 지역총생산(GRDP)과의 관계를 오차수정모형을 이용하여 분석하였음
- 이재민·한상용·이창운(2009)에서 유류가격 변화에 대한 고속도로 및 도시철도 수요변화를 자기회귀시차모형(Autoregressive Distributed Lag Model, ADL)을 이용하여 분석하였음
- 한국교통연구원(2010)에서는 고속도로, 일반철도 및 수도권 전철 통행실적과 국내총생산, 유류가격, 기타 사회경제적 변수 간에 어떠한 관계가 있는 지를 ADL 모형을 이용하여 분석하였음
- 한국교통연구원(2011) 연구에서는 고속도로 및 철도의 통행요금과 운임에 대한 탄력성을 오차수정모형으로 추정하였음

나. 국외문헌 연구¹⁾

- Box et al.(1994)은 월별 자료(1949-1960년)와 Seasonal ARIMA(SARIMA) 모형으로 장래 교통수요를 예측하는 방법을 제시하였음
- Box et al.(1994)은 특정 연도의 연속적인 월별자료(between observations for successive months in a particular year)간의 특성과 연속연도의 동일 시점

1) 본 내용은 한국교통연구원(2010, 2011)의 내용 중 일부를 발췌·요약한 것임

(월) 자료(between the observations for the same month in successive years)의 특성을 모두 고려하는 SARIMA 모형을 제시하였음

- Jong and Gunn (2001)과 Graham and Glaister(2002)는 EU를 대상으로 연구한 유가에 대한 교통수요 탄력성을 제시
 - European Commission에서 TRACE project(1998)를 수행하면서 다양한 유류가격 탄력성을 도출하였으며 주로 1985년 이후의 연구결과들임
 - Jong and Gunn (2001)과 Graham and Glaister(2002)는 이러한 연구결과를 정리하여 제시

<표 1> EU TRACE project의 도로교통수요에 대한 유류가격 탄력성

구분		유류가격에 대한 운행(car trips) 탄력성	유류가격에 대한 대-km 탄력성
단 기	통근(commuting)	-0.20	-0.12
	자택기반 비즈니스 (home-based business)	-0.06	-0.02
	이외의 비즈니스 (non home-based business)	-0.06	-0.02
	교육(education)	-0.22	-0.09
	기타(other)	-0.20	-0.20
	전체(total)	-0.16	-0.16
장 기	통근(commuting)	-0.14	-0.23
	자택기반 비즈니스 (home-based business)	-0.07	-0.20
	이외의 비즈니스 (non home-based business)	-0.17	-0.26
	교육(education)	-0.40	-0.41
	기타(other)	-0.15	-0.29
	전체(total)	-0.19	-0.26

주: 다양한 탄력성 수치들의 평균값을 제시한 것임

자료: 1. Jong, De and H. Gunn, "Recent Evidence on Car Cost and Time Elasticities of Travel Demand in Europe", *Journal of Transport Economics and Policy*, Vol. 35, 2001, pp. 137~160
 2. Graham, Daniel and Stephen Glaister, *Review of Income and Price Elasticities of Demand for Road Traffic*, Imperial College, 2002

- Graham and Glaister(2002)는 이들을 종합하여 유류가격에 대한 운행 및 대-km 탄력성을 다시 제시하였음
 - 대체적으로 유류가격이 증가하면 도로교통수요는 감소하는 것으로 나타남
 - 단기 및 장기 유류가격에 대한 차량운행 탄력성이 각각 -0.16과 -0.19
 - 단기 및 장기 유류가격에 대한 대-km 탄력성이 각각 -0.16과 -0.26

- Graham and Glaister(2002)에 의하면 유류가격에 대한 교통수요 탄력성 연구는 다양한 형태의 자료들을 이용하고 있음
 - 횡단면(cross sectional), 시계열(time series), 패널 및 풀링(cross sectional time series) 자료들을 이용하고 있음
 - 그러나 Graham and Glaister(2002)에 의하면 최근에는 횡단면자료는 잘 이용하지 않는다고 함
 - 추정모형으로 static model, ADL model, ECM model 등을 이용하고 있음
- Hanly et al.(2002)은 기존에 행해진 유류가격에 대한 교통수요 탄력성을 분석하여 정리하였음
 - Hanly et al.(2002)에 의하면 유류가격이 증가하면 도로교통수요는 감소하는 것으로 분석되었는데 dynamic model을 이용하면 단기보다 장기의 탄력성이 크게 산정되고 있음
 - 전체 대-km와 대당 대-km자료를 이용하여 탄력성을 제시하였고 또한 static model을 이용하였고 자료는 cross section data, cross section/time series data, time series data 등을 이용하였음

<표 2> 유류가격에 대한 도로교통수요 탄력성

구분	dynamic mode with time series data		static model		
	단기	장기	cross section data	cross section/time series data	time series data
대-km(합계)	-0.17~-0.05	-0.63~-0.10	-0.54~-0.21	-0.41~-0.13	-0.32~-0.32
대-km(대당)	-0.14~-0.06	-0.55~-0.11	-	-0.33~-0.33	-0.69~-0.69

자료: Hanly, Mark, Joyce Dargay, and Phil Goodwin, *Review of Income and Price Elasticity in the Demand for Road Traffic*, Economic and Social Research Council, 2002

- Hanly et al.(2002)은 기존에 행해진 소득에 대한 교통수요 탄력성을 분석하여 정리하였음
 - 일반적인 인식처럼 소득이 증가하면 교통수요가 증가하고 있음
 - 대체로 대당 자료는 전체 자료에 비해 탄력성이 낮게 나타나고 있음

<표 3> 소득에 대한 도로교통수요 탄력성

구분	dynamic mode with time series data		static model		
	단기	장기	cross section data	cross section/time series data	time series data
대-km(합계)	0.05~0.62	0.12~1.47	0.46~0.48	0.05~1.44	0.15~1.18
대-km(대당)	-0.02~0.005	0.00~0.41	0.06~0.08	-	0.03~0.03

자료: Hanly, Mark, Joyce Dargay, and Phil Goodwin, *Review of Income and Price Elasticity in the Demand for Road Traffic*, Economic and Social Research Council, 2002

- 도로부문의 통행요금 탄력성에 대한 다른 해외 연구사례를 살펴보면
 - 대체로 해외의 통행요금 탄력성이 교량 혹은 특정 지역 고속도로의 탄력성이 대부분을 차지하고 있음
 - 대체로 통행요금 탄력성이 -0.03에서 -0.3 혹은 -0.4 내외이며 예외적으로 -0.9 이상인 경우도 있음
- Wardman(2006)은 1990년부터 1998년까지 영국의 출도착지 기준 철도수송실적 자료를 변화된 중력방정식 모형과 가중최소자승법(Weighted Least Squares)을 적용하여 통행요금에 따른 통행수요 탄력성을 추정하였음
 - 추정결과에 따르면 영국의 통행요금에 대한 철도통행수요 탄력성은 -0.49~ -0.83으로 분석되었음

<표 4> 미국 및 다른 나라의 도로부문 통행요금 탄력성

지역 및 도로	탄력성
16 Toll Facilities	-0.03~-0.31
15 Toll Bridges	-0.15~-0.31
Golden Gate Bridge	-0.15
San Francisco Bay Bridge	-0.05
Numerous United State Facilities	-0.1~-0.35
Lee County(Cape Coral and Midpoint Bridges)	-0.02~-0.36
California(SR 91)	-0.9~-1.0
California(I-15)	-0.02~-0.42
Singapore(City Center)	-0.25
Toronto(407)	-0.3
Huston(Hardy Toll Road)	-0.4~-0.8

자료: Burris, M. W.(2003), "The Toll-Price Component of Travel Demand Elasticity", *International Journal of Transport Economics*, Vol. 30, No. 1, pp.45~59

3. 교통부문 자료설명 및 자료특성에 따른 연구방법론

가. 교통부문 자료설명

- 교통부문은 도로, 철도, 항공, 해운부문으로 구성되어 있음
- 개별 모드 수송실적 중 도로부문은 차량 통행대수 및 통행 대-km로 구성되어 있음. 이외에도 운송업 부문은 수송 여객인원 및 여객인-km로 구성되어 있음
 - 고속도로는 전체 총괄자료로 통행대수 혹은 통행 대-km자료가 구축되고 있으며 출발지 및 도착지(OD) 기준으로도 자료가 구축되고 있음²⁾
 - 운송업부문에서는 버스운송업에서 수송 인원 및 수송 인-km자료가 전체 총괄자료 혹은 출발지 및 도착지 기준으로 구축되고 있음
 - 고속도로 통행대수 및 통행 대-km 자료는 특정 시점 총괄자료는 한국교통연구원 KTDB(<http://www.ktdb.go.kr>) 및 한국도로공사(<http://www.ex.co.kr>)와 민자고속도로의 일부 홈페이지에서 제공하고 있음. 출도착지 기준 자료 구득은 한국도로공사 및 민자고속도로의 협조 혹은 한국교통연구원 KTDB에서 구득 가능
 - 또한 운송업부문의 수송 인원 및 수송 인-km 자료도 특정 시점 총괄자료는 한국교통연구원 KTDB에서 제공하고 있음. 그러나 보다 자세한 구간별 자료는 개별 운송회사의 협조가 필요
- 철도부문은 크게 일반철도, 수도권 전철 및 개별 도시철도로 구분할 수 있음
 - 철도부문 수송실적 자료는 총괄자료로서 수송인원 및 수송 인-km 자료가 축적되고 있으며 출발지 및 도착지 기준으로도 수송인원 및 수송 인-km 자료가 구득 가능
 - 이외에도 일반철도부문에서는 여객뿐만 아니라 화물수송실적도 톤 기준 및 톤-km 기준으로 총괄 및 출발지 및 도착지 기준으로 구축되고 있음
 - 철도부문 총괄실적 자료는 한국교통연구원 KTDB, 한국철도공사(<http://www.korail.com>) 및 해당 도시철도공사의 홈페이지에서 구득 가능. 구간별 자료는 한국철도공사 및 해당 도시철도공사의 협조 혹은 한국교통연구원 KTDB에서 구득 가능
- 항공 및 해운부문 역시 수송 인원 및 인-km 기준의 총괄 및 출도착지 기준 자

2) 전체 총괄자료라 함은 월별, 분기별 혹은 연도별로 전체 통행대수가 축적되는 경우를 의미하며 출발지 및 도착지 기준 자료는 어느 시점에 특정 지점과 다른 특정 지점간의 통행대수가 기록되는 경우를 의미함

- 료가 축적되고 있음
- 화물부문도 총괄 및 출도착지 기준으로 수송 톤 기준 및 톤-km 자료가 축적되고 있음
 - 항공부문 자료는 인천국제공항공사(<http://www.airport.kr>) 및 한국공항공사(<http://www.airport.co.kr>) 홈페이지와 항공정보포털시스템(<http://www.airportal.co.kr>)과 한국항공진흥협회(<http://www.airtransport.or.kr>)의 항공통계를 참조할 수 있음
- 교통부문 모드 자료는 크게 총괄자료와 출도착지 기준 자료로 구분 가능
- 총괄자료는 일별, 주별, 월별, 분기별 및 연도별로 구축이 되고 있는 자료로서 시계열 분석에 이용할 수 있음
 - 출도착지 기준 자료는 구간별 자료로서 횡단면 자료 분석에 이용될 수 있음
 - 출도착지 기준 자료가 월별, 분기별, 연도별로 누적된다면 패널자료(panel data)가 될 것이며 패널자료분석(panel data analysis)이 가능
- 교통부문 시계열 자료 이용에서 유의할 점은 지나친 과거 자료 이용을 배제해야 할 것임
- 2000년대 이전까지 교통수요는 크게 증가하는 양상을 보였지만 2000년대 이후에는 증가세가 완연히 꺾이고 있음
 - 한국도로공사의 고속도로 이용실적을 살펴보면 1990년대와 2000년대의 연평균 증가추세가 확연히 차이가 나고 있음
 - 철도부문에서도 일반구간과 달리 수도권 전철은 1990년대의 연평균 증가율이 2000년대 증가율보다 다소 높은 것을 알 수 있음
 - 1990년대는 교통시설 투자가 본격화된 시기이며 인구증가율 및 경제성장률이 크게 증가하는 시기였지만 2000년대 이후 교통수요 증가도 정체되는 시기이며 인구증가율 및 경제성장률이 정체되는 시기임. 향후에도 인구성장률 및 경제성장률의 증가세가 둔화될 것으로 예상됨
- 따라서 교통수요 이용에서 연도별 자료를 이용할 때 지나친 과거 자료를 이용해서는 안 될 것임
- 그러나 과거 자료의 배제는 관측치의 개수를 작게 해서 통계분석에 문제가 있을 수 있음. 즉 자유도(degree of freedom)가 30보다 작아서 중심극한정리(central limit theorem)를 적용하기 어려운 문제가 있음

- 따라서 시계열 분석에서는 과거 자료 배제와 자료의 개수문제를 해결하기 위해서 2000년대 이후 자료를 이용하고 분기별 혹은 월별 자료를 이용하여야 함

<표 5> 고속도로 및 철도 수송실적

단위: 천대, 천인, %

구분	고속도로	철도	
		일반구간	수도권 전철
1991	309,342	157,780	521,501
1992	392,647	161,912	554,452
1993	472,927	152,008	571,049
1994	563,412	136,987	592,017
1995	692,019	135,653	654,727
1996	839,085	133,401	686,141
1997	929,920	131,760	701,239
1998	824,895	120,102	708,948
1999	921,408	119,075	704,488
2000	1,054,241	115,914	721,354
2001	1,198,767	117,618	733,353
2002	1,398,778	109,935	741,781
2003	1,517,783	105,524	789,097
2004	1,540,982	111,214	810,009
2005	1,602,302	115,002	835,993
2006	1,646,399	114,331	854,814
2007	1,751,974	110,631	878,664
2008	1,726,748	113,098	905,879
2009	1,817,687	107,733	912,586
2010	2,008,306	112,093	948,848
1991-2000 연평균 증감률	14.60	-3.37	3.67
2001-2010 연평균 증감률	5.90	-0.53	2.90
1991-2010 연평균 증감률	10.35	-1.78	3.20

자료: 1. 한국도로공사 홈페이지(www.ex.co.kr)

2. 한국철도공사·코레일공항철도·한국철도시설공단, 『2010 철도통계연보』, 2011.

나. 자료특성에 따른 연구방법론

- 교통부문 자료는 크게 시계열 자료와 횡단면 자료로 구분할 수 있음
 - 월별, 분기별 및 연도별 자료로 구성된 시계열 자료와 구간별 자료로 구성된 횡단면 자료로 구성되어 있음
- 만일 시계열 자료를 이용하여 개별 변수들의 동태적 변이를 파악하고자 한다면 자기회귀 이동평균(ARMA) 혹은 적분된 자기회귀 이동평균(ARIMA) 모형을

이용하여야 함

- ARMA 및 ARIMA 모형은 개별 시계열변수들의 동태적 변이를 분석하고 이를 토대로 장래 예측치를 구하는 모형으로 볼 수 있음
- ARMA 모형과 ARIMA 모형의 차이점은 분석대상 시계열 변수가 안정적인지 혹은 불안정적인지에 달려 있음. 즉 개별 시계열 변수가 안정적(stationary)이라면 ARMA 모형을 이용하고 불안정적(non-stationary)이라면 ARIMA 모형을 이용할 수 있음. 이때 안정성 여부를 검증하기 위하여 단위근 검정(unit root test)을 수행해야 함
- 또한 개별 변수들이 분기별 혹은 월별변수이면 계절성(seasonality)을 보유할 수 있으므로 계절성을 감안한 ARIMA 모형 즉, Seasonal ARIMA(SARIMA) 모형을 적용할 수 있음
- 시계열 자료 분석에서 사회경제변수가 교통수요에 어떠한 영향을 미치는 지를 분석하는 것으로 자기회귀 시차모형(ADL)과 오차수정모형(ECM)을 적용할 수 있음
 - 대부분의 시계열 변수들이 불안정적일 가능성이 높는데 이에 대한 검증은 앞에서 지정한 단위근 검정을 수행
 - 만일 불안정적인 사회경제변수와 교통실적 변수들이 장기균형관계가 존재한다면 ADL 및 ECM 모형을 적용할 수 있음. 장기균형관계의 존재 여부는 공적분 검정(cointegration test)을 수행할 수 있음
 - 만일 불안정적인 시계열 변수들이 장기균형관계가 존재하지 않는다면 개별 변수들을 차분하여 OLS를 적용할 수 있음
- 횡단면 자료를 이용한 교통수요 분석에서는 개별 사회경제변수가 교통실적자료에 어떠한 영향을 미치는 지를 횡단면 자료를 이용하여 분석하는 경우를 의미
 - 이때 OD별 교통수요와 사회경제변수간의 관계를 보기 위해서는 중력방정식 모형(gravity equation model)을 구축하고 이를 OLS 혹은 토빗(Tobit) 추정방법으로 분석할 수 있음
 - OD별 통행수요 중 일부 구간에서 통행실적이 0인 경우가 존재하므로 이를 반영하기 위해서는 Tobit 추정방법을 이용하여야 함
 - 또한 OD별 인구 혹은 지역내총생산(GRDP) 자료와 구간별 통행요금(운임) 자료와 같은 구간별 자료 구축이 필요함

4. 횡단면 자료를 이용한 교통수요 분석

가. 중력방정식을 이용한 교통수요 분석

1) 모형 개요³⁾

- 횡단면 자료를 이용한 분석방법은 통행수단별로 출발지 및 목적지 수송(혹은 통행)실적 자료를 이용하여 분석할 수 있음
- 기존의 시계열 자료를 이용한 통행요금 탄력성 추정은 다음과 같은 점에서 모순이 있음
 - 한국도로공사는 2000년 이후 3번 정도 통행요금 체계를 개편하였음. 따라서 물가지수를 이용하여 실질화한 월별 통행요금 자료는 변화시점에만 전체 통행료 수준이 상승하고 그 이후에는 계속하여 감소하는 형태를 띠고 있음
 - 명목 통행료를 이용한다 할지라도 통행료 변화시점에만 전체 통행료 수준이 상승하고 그 이후에는 계속하여 일정한 형태를 유지하고 있음
 - 그러므로 시계열 자료를 이용한 통행요금 탄력성 분석은 엄밀한 의미에서 통행요금 탄력성 분석이 아니라 물가수준 고정에 따른 실질통행료 하락 혹은 명목통행료 고정의 효과분석에 지나지 않음
 - KDI·KOTI(2008)의 기존 연구는 시계열자료를 이용하여 앞에서 지적한 한계를 가지고 있음
- 따라서 본 연구에서는 수단별 출발지·목적지 통행실적 자료와 같은 횡단면자료를 이용한 통행요금 탄력성 분석을 수행할 수 있음
- 구체적으로 다음과 같은 중력방정식(gravity equation) 모형을 이용할 수 있음

$$\ln(1+y_{ij}^*)=\beta_0+\beta_1\ln po_i pd_j+\beta_2\ln p_{ij}+\beta_3 distance_{ij}+\beta_4 other\ variables_{ij}+u_{ij} \quad (\text{식 } 1)$$

$$u \sim Normal(0, \sigma^2)$$

y^* : 구간별 혹은 출도착지별 통행수요

po : 출발지 인구

pd : 도착지 인구

p : 출발지에서 도착지까지의 통행요금

$distance$: 출발지에서 도착지간의 거리

$other\ variables$: 다른 변수

3) 이 부분은 이제민 외(2009)와 한국교통연구원(2011)의 일부 내용을 참조하였음

u : 교란항

σ : 교란항의 표준오차

\ln : 자연로그

i, j : 출발지 및 도착지를 나타내는 하첨자

- 통행요금 탄력성을 산정하기 위하여 출발지와 도착지 사이의 통행요금을 설명변수로 이용하였으며 출발지와 도착지 사이의 관계를 나타내기 위하여 인구를 설명변수로 추가하였음
- 기타변수로는 출발지 혹은 도착지가 기종점을 나타내는 더미변수를 추가할 수 있는데 이는 인구로 표현될 수 없는 특성을 반영하고자 하였음
- 예를 들어 고속도로의 경우에 경기도 성남시 분당구 궁내동에 위치하는 서울 톨게이트는 분당구만의 인구로는 설명이 불가능한 특성을 보유하고 있으므로 이를 나타내는 더미변수를 추가할 수 있음
- 이외에도 출도착지 영업소가 신규 영업소인지를 나타내는 더미변수를 추가할 수 있는데 신규 영업소는 운전자들이 이용하기 어렵다는 점을 통제하기 위함임
- 그리고 본 연구에서는 일정한 통행수요 탄력성(constant travel elasticity)을 유도하기 위하여 로그-로그 모형(log-log model)을 설정하였음
- 본 연구에서 유의할 사항은 출발지 및 도착지간의 통행실적이 없는 경우가 발생할 수 있다는 것임
- 즉 통행실적이 0인 출발지와 도착지 쌍이 다수 존재할 수 있으며, 이는 로그-로그 모형으로 구성된 본 연구에서 missing data 문제가 발생할 수 있음
- 이러한 통행수요가 존재하지 않는 문제를 missing data로 볼 것인지 혹은 통행수요를 0으로 볼 것인지에 대해 다양한 의견이 있을 수 있음
- 만약 통행수요가 존재하지 않는다고 본다면 이는 표본편의문제(sample selection problem)로서 Heckman이 제안한 모형(Heckman sample selection model)으로 추정하여야 함
- 그러나 이를 통행수요가 0인 경우로 해석한다면 이는 표본편의문제가 아니라 코너해 문제(corner solution problem)로서 전통적인 토빗모형(Tobit model)으로 보아야 할 것임
- 본 연구에서는 출발지와 목적지 간에 통행수요가 없는 경우를 0으로 해석하여

토빗모형을 이용할 것임

- 그러나 본 모형이 로그-로그 모형임으로 종속변수인 통행수요가 0인 경우를 고려하기 위하여 모든 통행수요에 1을 더한 $\ln(1+y_{ij}^*)$ 를 종속변수로 고려할 것임
- 이러한 방법은 국제무역의 중력방정식 모형에서 주로 이용되는 기법으로 Eichengreen(1995), Eichengreen and Irwin(1997), 그리고 Felbermayr and Kohler(2006) 등을 참조할 수 있음
- 토빗모형에는 censored 토빗과 truncated 토빗이 있는데 본 연구에서는 통행수요가 존재하지 않는 경우에는 0으로 보는 것임으로 censored 토빗으로 보아야 함⁴⁾
- 본 연구에서는 0으로 censored 되는 부분과 censored 되지 않는 부분으로 구분하여 최우법(maximum likelihood estimation, MLE)으로 추정할 수 있음
- 비선형인 코너해 결과(coner solution outcome)와 선형인 부분이 결합되어 있는 형태로 최우법으로 추정할 수 있음

$$\ln(1+y_{ij}) = \max(0, \ln(1+y_{ij}^*)) \quad (\text{식 } 2)$$

- 본 연구의 모형은 비선형과 선형이 결합된 형태로서 로그-로그 모형이지만 추정치(estimate) 자체가 탄력성 계수가 될 수가 없음
- 탄력성 계수 산정을 위하여 다음과 같은 조건부 기대치를 고려할 수 있음

$$\begin{aligned} E(y|x) &= P(y=0|x) \times 0 + P(y>0|x) \times E(y|x, y>0) \\ &= P(y>0|x) \times E(y|x, y>0) \end{aligned} \quad (\text{식 } 3)$$

y: 종속변수, 본 연구에서는 $\ln(1+y_{ij})$

x: 설명변수, 본 연구에서는 (식 1)의 설명변수를 의미함.

P: 확률

- 이때 y가 0보다 클 확률과 0보다 클 때의 조건부 기대치를 산정하면

4) truncated 토빗은 0인 경우를 제외하고 하는 것임.

$$P(y > 0|x) = P(u > -x\beta|x) = P(u/\sigma > -x\beta/\sigma) \quad (\text{식 4})$$

$$= \Phi(x\beta/\sigma)$$

u : 평균이 0이며 분산이 σ^2 인 정규분포를 따르는 교란항

$$E(y|x, y > 0) = E(x\beta + u|x, y > 0) = x\beta + E(u|u > -x\beta) \quad (\text{식 5})$$

$$= x\beta + \sigma \left[\frac{\phi(x\beta/\sigma)}{\Phi(x\beta/\sigma)} \right] = x\beta + \sigma \lambda(x\beta/\sigma)$$

Φ : 표준정규분포의 누적확률분포함수(cumulative distribution function)

ϕ : 표준정규분포의 확률밀도함수(probability density function)

λ : Inverse Mills Ratio, $\phi(\cdot)/\Phi(\cdot)$

- 이를 기반으로 종속변수의 조건부 기대치를 산정하면

$$E(y|x) = P(y > 0|x) \cdot E(y|x, y > 0) \quad (\text{식 6})$$

$$= \Phi(x\beta/\sigma) \cdot [x\beta + \sigma \lambda(x\beta/\sigma)]$$

$$= x\beta \cdot \Phi(x\beta/\sigma) + \sigma \phi(x\beta/\sigma)$$

○ 조건부 기대치를 설명변수에 대해 미분하면 다음과 같은 탄력성 계수를 산정할 수 있음

$$\epsilon_{yp} = \Phi(x\beta/\sigma)\beta_2 \quad (\text{식 7})$$

ϵ_{yp} : 통행요금에 대한 통행수요 탄력성

2) 분석 사례

- 한국교통연구원(2011)에서 수행한 철도운임 변화에 따른 철도수송수요 탄력성 분석을 사례로 제시하고자 함
- 한국교통연구원(2011)에서는 한국철도공사에서 제공한 2011년 3월 기준 출발지 및 도착지 기준 구간별 수송실적 자료를 토대로 중력방정식을 이용하여 운임에 대한 수송수요 탄력성을 산정하였음

- 우선 한국철도공사의 운임체계를 살펴보면 기본적으로 km당 임률(총괄원가÷수송실적(인km))과 이용거리를 곱하여 산정하는 거리비례제를 기본으로 하고 있으며 KTX는 고속선 및 기존 선에 대한 복수 임률제를 적용하고 있음(한국교통연구원, 2010: 한국교통연구원, 2011)

<표 6> 한국철도공사 운임체계

구분	KTX	새마을호	무궁화호	통근열차
운임체계	복수임률제	거리비례제	거리비례제	거리비례제
	$P=a \times d \pm a$	$P=a \times d$	$P=a \times d$	$P=a \times d$

주: P는 운임, a는 임률, d는 거리, a는 기존선에 대한 체감률

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2011.

- 한국철도공사의 임율 및 최저운임을 살펴보면 KTX는 신선임율이 인km당 158.09원이며 기존선 임율은 좀 더 저렴한 100.35원/인km 수준
 - KTX 최저운임과 최저거리는 각각 8,100원과 52.5km 수준
 - 새마을의 임율은 93.29원/인km 수준이며 최저운임과 거리는 각각 4,700원과 50km 수준

<표 7> 한국철도공사 임율 및 최저운임 현황

구분	KTX	새마을	무궁화	통근열차	비고
신선임율 (기존선임율)	158.09원 (100.35원)	93.28원	62.83원	31.11원	1인 1km당
최저운임 (최저거리)	8,100원 (새마을 105%)	4,700원 (50km)	2,500원 (40km)	1,600원 (50km)	(좌석기준)

자료: 1. 한국철도공사 홈페이지(www.korail.com)

2. 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2011.

- 특실요금 현황을 살펴보면 KTX는 기준 운임의 40%를 추가로 받고 있으며 새마을 및 무궁화는 기준 운임의 15%를 추가적으로 받고 있음

<표 8> 한국 철도공사 특실요금 현황

구분	KTX	새마을	무궁화
특실요금	기준 운임의 40% 추가	기준 운임의 15% 추가	기준 운임의 15% 추가
최저요금	최저 4,600원 추가	최저 3,500원 추가	최저 1,600원 추가

자료: 1. 한국철도공사 홈페이지(www.korail.com)

2. 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2011.

- 이외에도 한국철도공사에서는 주말(금~일요일) 요금에 비해 주중(월~목요일) 요금을 할인하고 있는데 KTX의 경우에 7.0%, 새마을과 무궁화는 4.5%를 할인하고 있음

<표 9> 한국철도공사 요일별 운임 할인율 현황

구분	평균	KTX	새마을	무궁화	통근열차
할인율(월~목요일)	4.9%	7.0%	4.5%	4.5%	-

자료: 1. 한국철도공사 홈페이지(www.korail.com)

2. 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2011.

- 추정모형은 다음과 같음

$$\ln y_{ij} = \beta_0 + \beta_1 \ln po_i + \beta_2 \ln pd_j + \beta_3 \ln p_{ij} + \beta_4 se_{ij} \quad (\text{식 } 8)$$

$$+ \beta_6 ktx_{ij} + \beta_7 sam_{ij} + \beta_8 riro_{ij} + \beta_9 metro_{ij} + u_{ij}$$

se: 출발지 및 도착지가 기종점인 경우를 나타내는 더미변수

ktx: 출발지 및 도착지간 차량이 KTX 인지 아닌지를 나타내는 더미변수

sam: 출발지 및 도착지간 차량이 새마을 인지 아닌지를 나타내는 더미변수

riro: 출발지 및 도착지간 차량이 누리로 인지 아닌지를 나타내는 더미변수

metro: 출발지 및 도착지간 차량이 통근열차 인지 아닌지를 나타내는 더미변수

- 운임(*p*)변수가 거리비례 요금제를 따르고 있어서 거리변수를 따로이 추가하지 않았음
- 또한 기종점 영업소는 다른 영업소보다 더 많은 여객을 수송할 가능성이 높으므로 이를 나타내는 더미변수를 추가하였음
- 현재 한국철도공사에서 운영하는 차량은 KTX(KTX 신천 포함), 새마을, 무궁화, 누리로 및 통근열차임. 차량별로 수송실적에 차이가 있을 것으로 예상되어 이를 나타내는 더미변수를 추가하였음
 - 2011년 3월 기준 한국철도공사 자료에서는 무궁화 열차의 지역별 운행횟수가 가장 많이 나타나서 무궁화열차를 제외한 다른 차종을 더미변수로 추가한 것임
- 한국철도공사의 구간별 OD 여객수송실적 자료를 이용하여 중력방정식을 적용하여 추정
- 본 연구에 이용된 철도노선은 16개 노선으로 다음과 같음

- 경부선, 중앙선, 호남선, 전라선, 충북선, 장항선, 경의선, 경원선, 경춘선, 경북선, 영동선, 태백선, 동해남부선, 진해선, 대구선, 경전선
- 16개 노선의 차종별 OD 순서쌍 자료를 살펴보면 무궁화 열차의 지역별 OD 순서쌍 자료개수가 가장 많은 10,784개로서 전체의 약 79.40%를 점유하고 있음
- 새마을 차종의 OD 순서쌍 자료 개수가 1,694개로서 전체의 12.47%를 점유하고 있으며 KTX 차종의 OD 순서쌍 자료 개수가 570개로서 전체의 4.20%, 누리로 차량이 474개로서 전체의 3.49%, 마지막으로 통근차량이 60개로서 전체의 0.44%를 점유하고 있음
- 노선별로는 경부선 OD 순서쌍 자료 개수가 3,115개로서 전체의 22.93%를 점유하여 가장 높은 점유율을 보이고 있음
- 그 다음으로 호남선, 경전선, 동해남부선, 전라선, 영동선, 장항선, 중앙선, 태백선, 경북선, 충북선, 진해선, 대구선, 경원선, 경의선, 경춘선 순으로 나타나고 있음

<표 10> 노선별 차종별 OD 순서쌍 현황

노선명	기점 역명	종점 역명	KTX	새마을	무궁화	누리로	통근	소계
경부선	서울	부산	174	511	1,958	472	0	3,115
중앙선	청량리	경주	0	46	817	0	0	863
호남선	대전 조차장	목포	302	373	1,410	0	0	2,085
전라선	익산	여수	0	286	883	0	0	1,169
충북선	조치원	봉양	0	0	169	0	0	169
장항선	천안	익산	0	303	633	0	0	936
경의선	서울	도라산	0	0	0	0	6	6
경원선	용산	신탄리	0	0	0	0	54	54
경춘선	망우	춘천	0	0	0	2	0	2
경북선	김천	영주	0	0	489	0	0	489
영동선	영주	강릉	0	0	1,105	0	0	1,105
태백선	제천	백산	0	0	593	0	0	593
동해남부선	부산진	포항	0	90	1,155	0	0	1,245
진해선	창원	통해	0	85	0	0	0	85
대구선	가천	영천	0	0	69	0	0	69
경전선	삼랑진	광주송정	94	0	1,503	0	0	1,597
소계			570	1,694	10,784	474	60	13,582

주: 수치는 노선별 및 차종별 OD 순서쌍의 개수를 의미함

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2011.

<표 11> 철도 횡단면 분석 변수설명 자료

구분	평균	표준편차	극소	극대	자료개수
<i>y</i>	655.0925	5211.123	1	260,735	13,582
<i>po</i>	245,494.3	219,114.3	24,736	1,090,181	13,582
<i>pd</i>	244,792.6	217,913.0	24,736	1,090,181	13,582
<i>p</i>	9,479.789	7,743.245	800	49,100	13,582
<i>se</i>	0.3293	0.4700	0	1	13,582

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2011.

- 2011년 3월 현재 분석에 이용된 OD 순서쌍의 평균 여객인원은 약 655명 내외로 나타나고 있음
 - 2011년 3월 현재 부산발 서울행 KTX 여객 수송인원이 260,735명을 기록하여 가장 최대 수송실적을 기록하였음
- 이외에도 출발지 및 도착지의 평균인구는 약 24만명 수준
- 본 연구에 이용된 운임은 월~목요일 할인운임을 적용하였으며 평균 운임은 약 9,480원 수준
- 그리고 출발지 및 도착지 가운데 해당 노선의 기점 혹은 종점인 경우는 전체 순서쌍 OD의 약 32.93% 수준
- 2011년 3월 철도부문 횡단면자료를 이용한 중력방정식 추정결과에 의하면 출발지 및 도착지 인구 변수와 기종점 여부를 나타내는 더미변수, 그리고 KTX, 새마을, 누리로, 통근열차를 나타내는 더미변수가 여객 수송실적에 양(+)의 효과가 있음
 - 운임변수는 여객 수송실적에 음(-)의 효과가 있는 것으로 나타났음
- 출발지 및 도착지 인구가 많은 경우에 두 지역 간의 여객수송실적이 많다는 것은 중력방정식에서 예상된 바이며
 - 기종점이 포함된 구간은 주로 대도시와 연계된 지역이므로 수송실적인 많은 것임
 - 통행마찰변수(trip friction variable)인 운임은 두 지역 간의 통행을 저해하는 것으로 나타났음
 - 또한 무궁화 차종에 비해 KTX, 새마을, 누리로, 통근열차에서 점 더 많은 수송실적이 발생할 것으로 나타났음

- 운임에 대한 수송실적 탄력성은 약 -0.39로서 운임이 1% 증가하면 여객 수송실적이 약 0.39% 감소하는 것으로 나타났음
- *F*-statistic 통계량이 크게 산정되어 모형이 적절함을 보여 주고 있음
 - 누리로 여부를 나타내는 더미변수를 제외한 모든 변수들의 *t*-통계량이 통계적으로 유의하게 나타났음

<표 12> 철도 횡단면 분석 결과

구분	계수 추정치	강건한 표준오차
<i>lpo</i>	0.3288***	0.0217
<i>lpd</i>	0.3425***	0.0220
<i>lp</i>	-0.3906***	0.0252
<i>se</i>	0.9980***	0.0419
<i>ktx</i>	2.6524***	0.1042
<i>sam</i>	1.2651***	0.0521
<i>riro</i>	0.1113	0.1057
<i>metro</i>	1.3431***	0.3174
상수항	-1.5934***	0.4084
<i>F</i> -statistic	329.02***	-
R^2	0.1598	-
관측치 개수	13,582	-

주: ***, **, *는 각각 1%, 5%, 10% 수준에서 통계적으로 유의함을 나타내고 있음

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2011.

나. Count Data Model

1) 개요

- count data model은 중력방정식을 이용한다는 점에서 앞절의 분석과 유사함
 - 앞절의 중력방정식을 추정할 때 Tobit 추정방법 혹은 최소자승법(OLS)를 이용할 수 있음
 - 최소자승법은 선형모형(linear model)을 가정하고서 중력방정식을 추정
 - Tobit 추정방법은 종속변수가 corner solution을 가진다는 가정 하에 모형을 추정하는 방법으로 최우법(MLE)을 이용하고 있음
- count data model은 종속변수를 셀 수 있다(가산가능성, countable)는 것에 포커스를 맞춘 것임. 특히 비음정수(non-negative integer)로 구성된 종속변수가 존재할 때 이용할 수 있는 추정모형
 - Poisson 추정방법과 Negative Binomial 추정방법이 존재하고 있음

- 또한 셀 수 있는 종속변수 중 0이 다수를 점유할 때 Zero Inflated Poisson(ZIP) 추정방법과 Zero Inflated Negative Binomial(ZINB) 추정방법이 존재하고 있음
- ZIP와 ZINB는 Poisson 및 Negative Binomial 추정방법과 마찬가지로 종속변수가 비음정수일 때 이용할 수 있지만 종속변수에 0이 다수가 존재하는 경우를 가정하는 추정방법으로 볼 수 있음
- 또한 Poisson 및 ZIP는 종속변수의 평균과 분산이 동일한 경우를 가정하는 것이며 Negative Binomial 및 ZINB는 종속변수의 평균과 분산이 상이한 경우를 가정하는 것임
 - 즉 Poisson 및 ZIP에서 $\text{Var}(y|x)=E(y|x)$ 로 가정하는 것임
 - Negative Binomial 및 ZINB에서는 $\text{Var}(y|x)=\sigma^2 E(y|x)$ 으로 좀 더 일반화된 모형으로 $\sigma^2 \neq 1$ 이 아닌 경우를 가정하는 것임. 이때 $\sigma^2 > 1$ 인 경우가 일반적인 경우로서 overdispersion이라 정의하며 $\sigma^2 < 1$ 인 경우는 희귀한 경우로서 underdispersion으로 정의함
- Poisson, Negative Binomial, ZIP 및 ZINB는 선형추정방법이 아니라 비선형(non-linear) 추정방법으로 볼 수 있음
 - 그러나 이들 추정모형은 일반적으로 exponential 함수 형태를 띠는데 자연로그를 취하면 모두 선형으로 변하기 때문에 일반화된 선형모형(generalized linear model, GLM)이라고도 함(Wooldridge, 1997, 2002)
- 종속변수가 count data일 때 최소자승법을 이용하면 다음과 같은 문제가 발생
 - 첫째 종속변수의 예측치(fitted value)가 0보다 작은 경우가 존재할 수 있음
 - 둘째 최소자승법을 이용하였을 때 count data로 구성된 종속변수의 특성으로 인한 이분산성(heteroscedasticity)을 해결할 수 없음. 이러한 이분산성은 최소자승법에서의 강건한 표준오차(robust standard error)로 해결할 수 없음
- 따라서 종속변수가 count data일 경우에 최소자승법을 이용하지 않고 count data 추정방법을 이용해야 함

<표 13> 추정모형 비교

구분	종속변수	추정방법	특징
OLS	• 연속적이 고 제한이 없음	• Least square(자승 법)으로 잔차항의 제곱을 최소화하는 과정에서 도출	• 추 정 치 (c o e f f i c i e n t estimates)와 한계적 효과 (marginal effects)가 동일
Tobit	• 코너해(coner solution)를 가지는 경우	• 로그우도함수를 극 대화하는 MLE(최우 법)으로 추정	• 추 정 치 (c o e f f i c i e n t estimates)와 한계적 효과 (marginal effects)가 상이
Poisson	• 셀 수 있는 비음정수	• 로그우도함수를 극 대화하는 MLE(최우 법) 혹은 QMLE(준 최우법) 으로 추정	• 종속변수의 평균과 분산 이 동일 • 추 정 치 (c o e f f i c i e n t estimates)와 한계적 효 과(marginal effects)가 상이
NB	• 셀 수 있는 비음정수	상동	• 종속변수의 평균과 분산 이 상이 • 추 정 치 (c o e f f i c i e n t estimates)와 한계적 효 과(marginal effects)가 상이
ZIP	• 셀 수 있는 비 음 정 수 이 지 만 0이 상당 히 많음	상동	• 종속변수의 평균과 분산 이 동일 • 추 정 치 (c o e f f i c i e n t estimates)와 한계적 효 과(marginal effects)가 상이
ZINB	• 셀 수 있는 비 음 정 수 이 지 만 0이 상당 히 많음	상동	• 종속변수의 평균과 분산 이 상이 • 추 정 치 (c o e f f i c i e n t estimates)와 한계적 효 과(marginal effects)가 상이

2) 이론적 고찰

☐ Poisson 추정모형

- Poisson 추정방법은 종속변수가 count data의 특성을 띠는 경우로서 조건부 확률밀도함수(conditional probability density function)가 다음과 같이 정의됨

$$f(y|x) = \frac{[\exp(-\mu(x))][\mu(x)]^y}{y!}, \text{ if } y=0,1,2,\dots \quad (\text{식 9})$$

$$= 0, \quad \text{otherwise}$$

- $\mu(x) \equiv E(y|x)$ 이며 $y!$ 은 y factorial을 의미함
- Poisson 추정방법에서 가장 중요한 특징은 분산과 평균이 동일하다는 점임. 즉 $Var(y|x) = E(y|x)$ 로 정의될 수 있음. 이는 Poisson 분포(distribution)의 특성을 따르는 것임
- 추정방법으로 만약 OLS를 이용할 경우에는 y 의 예측치(fitted value)가 음(-)의 값이 나타날 수 있으며 또한 OLS를 이용할 경우에 다양한 모형으로의 전환(transformation)이 어려운 점이 존재함
- 또한 비선형자승법(nonlinear least square, NLS)을 이용할 경우에는 $Var(y|x)$ 가 일정(constant 혹은 동분산, homoscedasticity)하다는 가정하에서만 효율적(efficient)임
 - 실제 $Var(y|x)$ 는 x (설명변수)에 의존적(dependent)인 경우가 많아서 NLS는 적절한 대안이 될 수 없음
 - 결국 $Var(y|x)$ 의 이분산성으로 인하여 OLS 및 NLS는 적절한 대안이 될 수 없음
- 결국 Poisson 추정방법은 조건부 로그우도함수(log likelihood function)를 정의하고 이를 극대화하여 추정치를 얻게 되는 최우법(MLE)을 이용하는 것임
- 그런데 Wooldridge(1997, 2002)는 count data model에서 (식 9) 혹은 $Var(y|x) = E(y|x)$ 와 같은 가정을 하지 않고 좀 더 일반화된 가정을 통하여 모형을 추정할 수 있다고 함
 - 표준적인 GLM 분석에서 x 에 대한 y 의 조건부 분포에 대하여 $Var(y|x) = \sigma^2 E(y|x)$ (단 $\sigma^2 > 0$)
 - 즉 x 에 대한 y 의 조건부 분포에 대하여 좀 더 유연한(weaker) 혹은 강건한(robust) 가정(assumption)을 하고서 로그우도함수를 추정하는 방법이 Quasi MLE(준최우법, QMLE)이며 y 의 조건부 분포가 어떠한 형태를 띠든 할 지라로 추정치의 일치성(consistency)을 보유했 것임(Poisson QMLE)

- 개별 관측치, i 에 대해 로그우도함수를 정의하면

$$l_i(\mu) = y_i \log(\mu) - \mu \quad (\text{식 10})$$

- 이때 $-\log(y_i!)$ 은 파라미터 μ 를 포함하지 않아서 제외하였음
- $E(l_i(\mu))$ 를 극대화하면 계수추정치를 얻을 수 있음. 그러나 이는 모집단(population)에 적용되는 경우이며 샘플(sample)에 적용되는 sample analog를 적용하면 $N^{-1} \sum_i^N l_i(\mu)$ 를 극대화하면 됨

- 이때 μ 에 대한 정의가 필요하며 일반적인 count data model에서는 exponent function을 적용함

$$\mu = m(x_i, \beta) = \exp(\beta_1 + \beta_2 x_2 + \dots + \beta_k x_k) = \exp(x\beta) \quad (\text{식 11})$$

- count data model에서 이와 같은 exponential function을 잘 이용하는데 이때 자연로그를 취하면 선형함수로 변화되기 때문에 일반화된 선형모형(GLM)이라 하며 linear exponential family(LEF)라고 함(Wooldridge, 1997, 2002)
- count data model에서는 일반적으로 추정치(coefficient estimates)가 한계적 효과(marginal effect)가 될 수 없음

$$\beta_j = \frac{\partial E(y|x)}{\partial x_j} \frac{1}{E(y|x)} = \frac{\partial \log E(y|x)}{\partial x_j} \quad (\text{식 12})$$

- (식 12)에서 보는 바와 같이 $100\beta_j$ 는 semi-elasticity of $E(y|x)$ with respect to x_j
- 만약 설명변수가 x_j 가 아니라 $\log(x_j)$ 라면 β_j 는 elasticity(탄력성) of $E(y|x)$ with respect to x_j 으로 볼 수 있음
- Poisson model은 보편적인 가정을 하여서 QMLE 추정방법을 이용하는 것임
 - 다만 일반적인 MLE 추정방법의 variance-covariance estimator를 이용하는 것이 아니라 QMLE의 variance-covariance estimator를 이용하는 것임
 - 이때 A 는 로그우도함수를 두 번 미분하여 음(-)을 부호를 취한 후에 기댓값을

취한 것으로 로그우도함수의 Hessian matrix의 negative Expected value로 볼 수 있음

- 또한 B 는 로그우도함수를 한 번 미분한 값을 제공한 후에 기댓값을 취한 것으로 로그우도함수의 score function의 product의 Expected value로 볼 수 있음
- Poisson QMLE의 추정계수는 다음과 같은 분포를 따름

$$\sqrt{N}(\hat{\beta} - \beta) \rightarrow \text{Normal}(0, A^{-1}BA^{-1}) \quad (\text{식 13})$$

- 이때 $\hat{\beta}$ 의 점근적 분산은 $A^{-1}BA^{-1}/N$ 으로 볼 수 있으며 QMLE를 이용하여 추정한 후에 추정계수와 잔차항을 이용하면 $\hat{\beta}$ 의 점근적 분산의 추정치 $\widehat{A^{-1}BA^{-1}}/N$ 을 구할 수 있음
- 이때 GLM 가정(assumption)이 충족된다면 즉 $\text{Var}(y|x) = \sigma^2 E(y|x)$ 이라면 $A^{-1}BA^{-1} = A^{-1}$ 가 되어 $\hat{\beta}$ 의 점근적 분산은 $\sigma^2 A^{-1}/N$ 이며 점근적 분산의 추정치는 $\hat{\sigma}^2 \widehat{A^{-1}}/N$ 이 됨
- 만일 $\text{Var}(y|x) = E(y|x)$ 이라면 $\sigma^2 = 1$ 이 되어(unit variance) $\hat{\beta}$ 의 점근적 분산과 점근적 분산의 추정치는 각각 A^{-1}/N 과 $\widehat{A^{-1}}/N$ 이 됨. 이는 일반적인 MLE 추정방법의 점근적 분산 및 점근적 분산의 추정치로 볼 수 있음
- 이러한 특성으로 인해 QMLE를 이용하는 것이 강건한 추정방법이 될 수 있음
 - 만일 x 에 대한 y 의 분포가 Poisson 분포를 따른다면 Poisson QMLE는 일반적인 MLE와 동일함
 - 또한 x 에 대한 y 의 분포가 완전한 형태의 Poisson 분포를 따르지 않는다고 할지라도 $\text{Var}(y|x) = \sigma^2 E(y|x)$ 만 성립한다면 Poisson QMLE는 효율적인(efficient) 추정방법이 될 수 있음

□ Negative Binomial 추정모형

- Poisson model은 매우 강한 가정(strong assumption)을 하고 있음
 - $\text{Var}(y|x) = E(y|x)$ 라는 equidispersion 가정을 하고 있음
- 만약 $\text{Var}(y|x) = \sigma^2 E(y|x)$ 이며 $\sigma^2 > 1$ 보다 크다면 이에 대한 대안으로 Negative Binomial(이하 NB) 추정방법을 이용할 수 있음. 만약 $\sigma^2 = 1 + \alpha$ 일 경우에

$Var(y|x) = \sigma^2 E(y|x) = (1 + \alpha)E(y|x)$ 로 정의될 수 있음

- 이러한 경우를 Wooldridge(2002)와 Cameron and Trivedi(1986)는 Negative Binomial model I(NB I)이라고 정의하였고 혹은 Constant Dispersion Model 이라고도 함
- Poisson model은 $\alpha=0$ 인 NB model의 특수한 경우(Washington et al., 2003)로 볼 수 있음
- NB model I의 경우에 만약 위와 같은 가정이 충족되지 않는다면 MLE 추정치가 일치성을 보유하지 못함
- 그러나 위와 같은 가정이 충족된다면 NB model의 MLE 추정치가 Poisson QMLE 추정치보다 효율적(efficient)임
- 그러나 Poisson QMLE 추정치는 $Var(y|x) = \sigma^2 E(y|x)$ 이 성립하는 한 조건부 평균이 정확하게 결정되는 것을 요구하는 어떠한 추정치보다 효율적이며 이러한 면에서 강건한 추정치로 볼 수 있음
- Wooldridge(2002)와 Cameron and Trivedi(1986)는 또 다른 형태의 NB model을 제시하였는데 이를 NB model II로 정의하였음. 혹은 Mean Dispersion Model이라고도 함
- 이때 y_i 는 Poisson 분포를 따르지만 x_i 뿐만 아니라 관찰되지 않는 이질적인 요소(unobserved heterogeneity), c_i 의 영향을 받고 있음.
- 즉 $y_i|x_i, c_i \sim Poisson(c_i m(x_i, \beta))$
- 이때 c_i 는 평균이 1이고 분산이 α 인 감마(Gamma) 분포(distribution)를 따르고 있음
- 따라서 x_i 에 대한 y_i 는 NB 분포를 따르고 있으며 조건부 평균과 조건부 분산은 다음과 같음

$$E(y_i|x_i) = m(x_i, \beta) = \mu_i \quad (\text{식 14})$$

$$\begin{aligned} Var(y_i|x_i) &= E[Var(y_i|x_i, c_i)|x_i] + Var[E(y_i|x_i, c_i)|x_i] \\ &= m(x_i, \beta)[1 + \alpha m(x_i, \beta)] = m(x_i, \beta) + \alpha[m(x_i, \beta)]^2 \\ &= E(y_i|x_i)[1 + \alpha E(y_i|x_i)] \end{aligned}$$

- y_i 의 확률밀도함수는 다음과 같이 정의됨

$$f(y_i|x) = \frac{\Gamma[(1/\alpha) + y_i]}{\Gamma(1/\alpha)y_i!} \left[\frac{1/\alpha}{(1/\alpha) + \mu_i} \right]^{1/\alpha} \left[\frac{\mu_i}{(1/\alpha) + \mu_i} \right]^{y_i}, \text{ if } y_i = 0, 1, 2, \dots \quad (\text{식 15})$$

$$= 0, \quad \text{otherwise}$$

- 이때 Γ 은 gamma function을 의미하며 $\mu_i = m(x_i, \beta) = \exp(x_i\beta)$ 를 의미함. 위 식에 자연로그를 취한 후에 로그우도함수를 극대화하는 것이 NB 추정방법으로 볼 수 있음
- Poisson QMLE와 마찬가지로 NB의 경우에도 QMLE를 적용할 수 있음

□ ZIP 및 ZINB 추정모형

- ZIP 및 ZINB model은 종속변수가 count data이면서 상당수의 관측치가 0인 경우를 설명하는 추정모형으로 볼 수 있음
 - 실제 교통수요모형에서 0이 관찰되는 경우는 매우 흔함
 - 예를 들어 개별 가구(household)가 도시철도를 이용하는 횟수를 종속변수로 할 때 어떤 주(week)에는 도시철도를 이용하지 않고 버스 혹은 승용차를 이용하여 0이 되는 경우가 존재
 - 또 다른 경우로서 어떤 주(week)에는 다른 도시에 여행을 간 관계로 도시철도를 이용하지 않은 경우가 존재
 - 이러한 경우에 normal count와 zero count로 구성된 2 state regimes이 존재한다고 함
- ZIP 및 ZINB model은 종속변수의 이러한 특성을 반영하는 모형
- ZIP에서는 종속변수의 확률밀도함수(probability density function)가 다음과 같이 정의됨

$$f(y|x) = (1 - p_i) \exp[-\mu(x)] [\mu(x)]^y / y!, \text{ if } y = 1, 2, \dots \quad (\text{식 16})$$

$$= p_i + (1 - p_i) \exp[-\mu(x)], \quad \text{if } y_i = 0$$

$$= 0, \quad \text{otherwise}$$

- 이때 p_i 는 y_i 가 0이 될 확률을 의미함

- ZINB model에서는 종속변수의 확률밀도함수(probability density function)가 다음과 같이 정의됨

$$f(y_i|x) = (1-p_i) \frac{\Gamma[(1/\alpha)+y_i]}{\Gamma(1/\alpha)y_i!} \left[\frac{1/\alpha}{(1/\alpha)+\mu_i} \right]^{1/\alpha} \left[\frac{\mu_i}{(1/\alpha)+\mu_i} \right]^{y_i}, \text{ if } y_i = 1, 2, 3, \dots \quad (\text{식 17})$$

$$= p_i + (1-p_i) \left[\frac{1/\alpha}{(1/\alpha)+\mu_i} \right]^{1/\alpha}, \quad \text{if } y_i = 0$$

$$= 0, \quad \text{otherwise}$$

- 마찬가지로 p_i 는 y_i 가 0이 될 확률을 의미함

□ Poisson 및 NB와 ZIP 및 ZINB 비교

- Vuong(1989)은 Poisson 및 NB와 ZIP 및 ZINB를 비교하는 검정방법을 제시
- 우선 다음과 같은 2개 모형의 확률밀도함수 비율의 로그함수를 정의

$$ml_i = \ln[f_1(y_i|x_i)/f_2(y_i|x_i)] \quad (\text{식 18})$$

- f_1 과 f_2 는 각각 모형1과 2의 확률밀도함수
- 이를 이용한 Vuong의 통계량은 다음과 같이 정의됨

$$V = \frac{\sqrt{N} \left[\frac{1}{N} \sum_{i=1}^N ml_i \right]}{\sqrt{\left(\frac{1}{N} \right) \sum_{i=1}^N (ml_i - \overline{ml})^2}} = \frac{\sqrt{N}(\overline{ml})}{S_{ml}} \quad (\text{식 19})$$

$$\overline{ml} = (1/N) \sum_{i=1}^N ml_i$$

- S_{ml} 는 ml 의 표준편차(standard deviation)이며 N 은 표본의 크기
- Vuong의 통계량은 점근적으로(asymptotically) 표준정규분포(standard normal distribution)를 따름
- 만약 Vuong의 통계량의 절대값이 1.96보다 작다면 특정한 모형 1과 2에 대한 우위가 존재하지 않으며 Vuong의 통계량이 1.96보다 크다면 모형 1이 모형 2

- 에 대해 우위가 존재하며 Vuong의 통계량이 -1.96보다 작다면 모형 2가 모형 1에 대해 우위가 존재(5% 유의수준 가정)
- 예를 들어 모형1이 ZINB이며 모형 2가 NB인 경우
 - Vuong의 통계량이 1.96보다 크다면 ZINB를 채택해야 하며
 - Vuong의 통계량이 -1.96보다 작다면 NB를 채택해야 함
 - Vuong의 통계량이 1.96과 1.96 사이에 존재하면 특별히 우위에 있는 모형이 존재하지 않는 것임
 - 또한 $Var(y_i|x_i) = E(y_i|x_i)[1 + \alpha E(y_i|x_i)]$ 라는 NB model의 조건부 분산의 특징을 이용하여 α (overdispersion parameter)에 대한 t 통계량에 대한 검정을 수행할 수 있음
 - α 에 대한 t 통계량의 절대값이 특정 임계치(예를 들어 1.96)보다 작다면 Poisson 혹은 ZIP를 채택해야 하며
 - α 에 대한 t 통계량의 절대값이 특정 임계치보다 크다면 NB 혹은 ZINB를 채택해야 함

<표 14> 모형선택 기준

구분		α 에 대한 t 통계량	
		$< 1.96 $	$> 1.96 $
Vuong 통계량 (ZINB(f_1)과 NB(f_2))	<-1.96	Poisson 혹은 ZIP	NB
	>1.96	ZIP	ZINB

주: NB는 Negative Binomial을 의미함

자료: Washington, S. P., Karlaftis, M. G., and Mannering F. L., *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC, 2003

3) 추정사례

- 앞절에서 이용한 2011년 3월 한국철도공사 OD 자료를 이용하여 Poisson 및 NB 모형을 적용할 수 있음
 - 2011년 3월의 OD 수송실적이 최소값이 0보다 크게 나타나서 ZIP 혹은 ZINB 모형을 적용할 필요는 없었음
- 추정모형은 다음과 같음

$$y_{ij} = \exp(\beta_0 + \beta_1 po_i + \beta_2 pd_j + \beta_3 p_{ij} + \beta_4 se_{ij} \quad (\text{식 } 20)$$

$$+ \beta_6 ktx_{ij} + \beta_7 sam_{ij} + \beta_8 riro_{ij} + \beta_9 metro_{ij}) \exp(u_{ij})$$

- 추정에 이용된 변수는 앞절 내용과 동일하지만 모든 변수에서 자연로그를 취하지 않고 수준변수를 이용하였다는 점에서 상이함
- Poisson 및 NB는 linear exponential family에 속하는 추정모형으로 개별 변수에 자연로그를 취하지 않았음

<표 15> count data regression 추정결과

구분	<i>Poisson</i>	<i>NB</i>
<i>po</i>	0.00000015 (0.00000016)	0.00000172*** (0.00000026)
<i>pd</i>	0.00000008 (0.00000016)	0.00000175*** (0.00000026)
<i>p</i>	0.00001950** (0.00000897)	-0.00004940*** (0.00000409)
<i>se</i>	1.0164*** (0.0869)	1.0658*** (0.0667)
<i>ktx</i>	2.1145*** (0.1153)	3.4143*** (0.2192)
<i>sam</i>	-0.1943** (0.0906)	0.3831*** (0.0726)
<i>riro</i>	0.1711 (0.1416)	-0.0087 (0.1279)
<i>metro</i>	1.4466*** (0.4052)	1.4304*** (0.3281)
상수항	5.3985*** (0.0952)	5.0853*** (0.1061)
α	- -	3.4438*** (0.0395)
관측치 개수	13,582	13,582
Wald 통계량	515.66***	639.04***
<i>log pseudolikelihood</i>	-16,636,066	-83,432.49

주: 1. ***, **, *는 각각 1%, 5%, 10% 수준에서 통계적으로 유의함을 나타내고 있음

2. 괄호안의 수치는 robust standard error를 의미함.

- Poisson 및 NB 추정모형을 적용하였을 때 운임변수의 추정부호가 상이하게 나타나고 있음
- Poisson 모형에서는 운임이 통행수요에 양(+)의 효과가 있는 것으로 나타났지만 NB 모형에서는 운임이 통행수요에 음(-)의 효과가 있는 것으로 나타났음

- Poisson 모형과 NB 모형의 선택을 위하여 α 추정치와 표준오차를 보면 각각 3.4438 및 0.0395로서 t 통계량이 매우 커서 NB 모형이 적절함을 나타내고 있음
- 따라서 NB 모형에서 운임변수가 음(-)의 부호를 가지는 것이 적절함을 알 수 있음
- 운임에 대한 철도 통행수요 탄력성을 계산하면 -0.4680으로 산정되었음
 - 즉 운임이 1% 상승하면 철도 통행수요는 약 0.47% 감소함
 - 이때 모든 변수들의 평균값을 이용한다는 가정을 하였을 때 탄력성 수치임

5. 시계열 자료를 이용한 교통수요 분석

가. 교통수요 동태적 변화 모형

1) 이론적 고찰⁵⁾

- 교통수요의 동태적 변화를 고찰하는 모형은 교통수요의 시계열 흐름에 대한 특성을 파악하여 장래 교통수요를 예측하고 교통정책 결정자들이 교통정책을 수립하는데 필수적인 자료를 제공하는데 있음
- 교통수요의 동태적 변화를 분석하는 모형으로 자기회귀 이동평균 (autoregressive moving average, ARMA)모형과 적분된 자기회귀 이동평균 (autoregressive distributed integrated moving average, ARIMA) 모형을 주목할 수 있음
 - ARMA(p, q)는 AR(p)모형과 MA(q)모형의 결합이며 ARIMA(p, d, q)는 ARMA(p, q)를 d 번 차분한 모형으로 볼 수 있음
- ARIMA(p, d, q)는 시계열변수가 불안정적일때 이를 안정적인 시계열변수로 변환시킨 모형으로 볼 수 있음
 - 즉 ARMA(p, q)에서 시계열변수를 d 번 차분하여 안정적인 시계열로 변환시킨 후에 ARMA(p, q)로 추정하는 모형임
- 교통실적 자료(ex: 고속도로 통행대수, 도시철도 통행실적 등)가 월별 혹은 분기별 시계열자료라면 이에 대한 안정성 검정(ex: 단위근검정 등)을 수행하여서 ARMA 혹은 ARIMA 모형으로 추정할 지를 결정하여야 함
- AR(p)는 p th order Autoregressive Model의 약자로서 “ p 차 자기회귀모형”으로 해석될 수 있음

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t \quad (\text{식 21})$$

$$\text{또는 } (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) y_t = u_t$$

$$\text{이는 다시 } \phi_p(L) y_t = u_t$$

- 이때 L 은 lag operator이며 u_t 는 기댓값이 0이고 분산이 σ^2 인 교란항
- 이때 y_t 가 안정적이기 위해서는 래그다항식 $\phi(L)$ 의 행렬식을 0으로 두었을때

5) 이 부분은 한국교통연구원(2010)의 일부 내용을 참조하였음

모든 근의 절대값이 1보다 커야 함(outside unit circle)

- MA(q)는 q th order Moving Average Model의 약자로서 “ q 차 이동평균모형”으로 해석될 수 있음

$$y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \quad (\text{식 } 22)$$

$$= (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) u_t$$

$$= \theta_q(L) u_t$$

- MA(q)의 가역성 조건(invertibility condition)은 래그다항식 $\theta_q(L)$ 의 행렬식을 0으로 두었을 때 모든 근의 절대값이 1보다 커야 함(outside unit circle)
 - 유의할 점은 가역성 조건(invertibility condition)은 해당 시계열 변수의 안정성에 영향을 미치지 못함. 즉 유한한 계수값⁶⁾을 가진 모든 이동평균(MA) 변수들은 안정적(stationary)임
 - 그 원인은 MA(q) process가 비가역적이라 할 지라도 동일한 자기상관계수를 보유하는 가역적인 MA(q) process가 존재하지 때문임
- 따라서 ARMA process의 안정성 조건은 MA process의 가역성 조건에 영향을 받는 것이 아니라 AR process의 안정성 조건에 영향을 받음(Greene, 1997; Hamilton, 1994)
- 자기회귀-이동평균모형(Autoregressive Moving-Average Model)은 AR(p)모형과 MA(q)모형의 결합으로 볼 수 있음
- 가장 간단한 ARMA(1,1)은 다음과 같이 정의됨

$$y_t = \phi_1 x_{t-1} + u_t + \theta_1 u_{t-1} \quad (\text{식 } 23)$$

- 이때 안정성(stationarity)을 위하여 $|\phi_1| < 1$ 이며 가역성(invertibility)을 위하여 $|\theta_1| < 1$ 을 가정하고 있음
- ARMA 모형은 조건부확률밀도함수(conditional probability density function)를 이용한 최우추정법(MLE) 혹은 비선형자승법(Nonlinear least squares)을 이용

6) 이를 달리 표현하면 MA process의 계수, θ_j 가 다음의 조건, $\sum_{j=0}^{\infty} \theta_j^2 < \infty$ 그리고 $\sum_{j=0}^{\infty} |\theta_j| < \infty$ 충족하여야 함을 의미함.

하여 추정할 수 있음

- 이때 적절한 p 와 q 를 선정하는 것이 중요함
 - 이를 위해 AIC(Akaike Information Criterion, AIC) 혹은 SC(Schwartz Criterion) 등을 이용할 수 있음
 - 혹은 Ljung-Box(1978)의 Q 통계량을 이용할 수 있음

$$Q = T(T+2) \sum_{j=1}^k (T-j)^{-1} \hat{\rho}_j^2 \sim \chi^2(k-p-q) \quad (\text{식 24})$$

- 이때 k 는 적절하게 큰 숫자이며 $\hat{\rho}_j$ 는 표본 자기상관함수(autocorrelation function)임
- 일반적인 ARIMA(p, d, q) 모형은 다음과 같이 정의됨

$$\phi_p(L)(1-L)^d y_t = \theta_q(L) u_t \quad (\text{식 25})$$

- 만약 $(1-L)^d$ 가 없다면 ARMA(p, q)와 동일함
 - 따라서 ARMA(p, q) 모형의 추정절차 및 p 와 q 차수 결정이 동일하게 적용될 수 있음
 - 다만 d 의 차수를 결정하기 위하여 변수(y_t)에 대한 단위근 검정을 수행해야 함
- 교통수요의 동태적 변화 분석 절차
 - 교통수요 실적자료에 대한 단위근 검정을 수행하여 ARMA(p, q)인지 ARIMA(p, d, q)를 결정할 것이며
 - AIC 혹은 SBC 또는 Ljung-Box(1978)의 Q 검정을 통하여 적절한 p 와 q 의 수치를 결정
 - 이후에 파라미터를 추정하고 장래 예측치를 구할 수 있음
- Seasonal ARIMA 모형은 Box et al.(1994)이 제시한 것으로 계절성을 띤 자료를 이용할 때 사용하는 추정방법임
 - 계절성(seasonality)이란 어떤 시계열변수(time series variables)의 값이 주기적(periodically)인 특성을 지니는 경우를 의미함
 - 예를 들어 항공수요가 여름휴가 및 성탄절 시즌에 피크를 이루며 국내총생산이 4/4분기에 가장 큰 값을 나타내는 경우임

- 월별 및 분기별 자료를 이용할 때 다음과 같은 방식을 이용할 수 있음
 - 원자료(raw data)와 ARIMA 모형을 이용하여 추정하고 장래 예측치를 구하는 경우와 계절조정(seasonal adjustment)을 한 자료와 ARIMA 모형을 이용하는 경우가 있음
- 첫 번째 방식은 계절성을 고려하지 않았기 때문에 올바른 추정방식으로 볼 수 없으며 두 번째 방법 역시 원자료를 이용하지 않고 계절조정을 수행한 자료를 이용하여 원자료가 내포하고 있는 의미를 왜곡할 수 있음
 - Judge et al.(1988)은 “the seasonality should be removed prior to a Box-Jenkins analysis by some sort of seasonal adjustment method because seasonal adjustment may distort the stochastic structure substantially.”라고 하여 계절 조정된 자료를 이용하지 말고 원자료를 이용하여 계절성을 고려한 ARIMA 모형(Seasonal ARIMA)을 이용할 것을 제시하였음
- ARIMA 모형은 특정 연도의 연속적인 월별자료(between observations for successive months in a particular year)간의 특성을 고려하는 모형이며 SARIMA 모형은 이외에도 연속연도의 동일 시점(월) 자료(between the observations for the same month in successive years)의 특성을 모두 고려하는 모형임⁷⁾
 - 첫 번째 부분은 ARIMA 모형으로 (식 13)과 같음. 첫 번째 부분은 2009년 12월 항공수요는 2009년 11월 및 10월 자료와 관계가 있음을 의미함
 - 계절성을 고려한 부분은 2009년 12월 항공수요는 2008년 12월 항공수요와 관련이 있음을 의미하며 다음과 같이 정의됨

$$\Phi_P(L)(1-L^s)^D y_t = \Theta_Q(L) u_t \quad (\text{식 26})$$

- 이때 s 는 만약 분기별 자료라면 4가 될 것이고 월별 자료라면 12가 될 것임. 또한 D 는 계절 단위근(Seasonal Unit Root)의 차수, P 는 계절 자기회귀(Seasonal Autoregressive)의 차수, Q 는 계절 이동평균(Seasonal Moving Average)의 차수임
- 최종적으로 첫 번째 부분과 두 번째 부분을 모두 함께 고려하면 다음과 같이 정의될 수 있음

7) 이하의 체계는 Box et al.(1994)를 따르고 있음

$$\phi_p(L)\Phi_P(L)(1-L)^d(1-L^s)^D y_t = \theta_q(L)\Theta_Q(L)u_t \quad (\text{식 27})$$

- 이는 SARIMA 모형으로 차수는 $(p, d, q) \times (P, D, Q)$ 로 정의될 수 있음
- 우선 SARIMA 모형을 이용하기 위해서는 원자료가 일반적인 단위근(unit root)과 계절 단위근을 보유하고 있는 지를 파악할 필요가 있음
 - 일반적인 단위근 검정은 Dickey and Fuller가 제시한 ADF(Augmented Dickey Fuller, ADF) test를 이용할 수 있고
 - 계절 단위근 검정은 Hasza and Fuller(1982), Dickey, Hasza, and Fuller(1984), Osborn, Chui, Smith, and Birchenhall(1988), Hylleberg et al.(1990), and Engle et al.(1993) 등이 제시하고 있음
- SARIMA 모형 혹은 ARIMA 모형의 추정절차는 크게 다음으로 구분할 수 있음
 - 모형구분(Identification), 추정(Identification), 진단(Diagnostic Testing)으로 구분할 수 있음
- 모형구분(Identification)에서는 SARIMA의 차수, 즉 $(p, d, q) \times (P, D, Q)$ 를 결정하는 것임
 - 이때 단위근(d) 및 계절 단위근(D)이 있는 지를 검정하기 위하여 단위근 검정을 수행하여야 함
 - 또한 자기회귀(Autoregressive) 및 계절 자기회귀(Seasonal Autoregressive) 부분의 차수(order)와 이동평균(Moving Average) 및 계절 이동평균(Seasonal Moving Average) 부분의 차수(order)를 결정하여야 함
 - 이를 위하여 원자료의 자기상관함수(autocorrelation function) 및 편자기상관함수(partial autocorrelation function)를 포함하는 correlogram을 관찰하여야 함
 - 일반적으로 원시계열 자료가 p 차 자기회귀모형(AR process)을 따른다면 자기상관계수가 지수적으로(exponentially) 혹은 기하학적으로(geometrically) 감소하는 추세를 보이며 p 차 이후의 편자기상관계수는 급격히 감소(cut off)하는 추세를 보임
 - 원시계열 자료가 q 차 이동평균모형(MA process)을 따른다면 편자기상관계수가 지수적으로(exponentially) 혹은 기하학적으로(geometrically) 감소하는 추세를 보이며 q 차 이후의 자기상관계수는 급격히 감소(cut off)하는 추세를 보임

- 만일 이에 해당되지 않는다면 이는 자기회귀 이동평균 모형(ARMA process)에 해당된다고 볼 수 있음
- 어떤 변수(y_t)의 자기상관함수(ρ_k)는 다음과 같이 정의됨

$$\rho_k = \text{Cov}(y_t y_{t+k}) / [\text{Var}(y_t) \text{Var}(y_{t+k})]^{1/2} \quad (\text{식 } 28)$$

$$k = 0, 1, 2, \dots$$

- 편자기상관함수(ϕ_k)는 개별 변수와 그의 시차변수 간의 회귀식의 회귀계수로서 다음과 같이 정의됨

$$\phi_k = \rho_1 \text{ for } k = 1 \quad (\text{식 } 29)$$

$$= \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_j} \text{ for } k > 1$$

$$\phi_{k,j} = \phi_{k-1,j} - \phi_k \phi_{k-1,k-j}$$

- $(p, d, q) \times (P, D, Q)$ 의 차수결정에서 무엇보다 중요한 것은 간결성(parsimonious)으로 가능한한 차수를 작게 하는 것이 모형의 추정에 유리함
- 추정(estimation)에서는 MA term이 있기 때문에 최소자승법(OLS)을 이용할 수 없음
 - 통계 package에 따라 조건부 최우법(conditional MLE) 혹은 비선형자승법(NLS)이 이용됨
 - 이때 다양한 차수에 따라 모형을 추정하여 AIC 및 SC를 최소화하는 모형을 선택할 수 있음. 또한 추정치들의 t-통계량도 고려하여 모형의 차수를 결정하여야 함
- 진단(Diagnostic Testing)에서는 진단과 추정에 의한 결과가 적정한 지를 판단하여야 함
 - 이를 위해 선택된 모형에 AR 혹은 MA term을 하나씩 추가하여 추정을 하여 그 결과를 비교할 수 있음

- 또 다른 방법으로 2단계 추정에서 얻은 잔차항(residuals)이 자기상관(autocorrelation)이 존재하는 지를 Ljung-Box(1978)의 Q 통계량을 이용하여 검정할 수 있음. 즉 만일 적절한 모형이 채택되었다면 잔차항에 자기상관이 존재하지 않음으로 자기상관이 존재하지 않는다는 귀무가설(null hypothesis)을 기각해서는 안될 것임(Q 통계량이 작거나 p-value가 0.05 혹은 0.1보다 커야 함)

2) 분석사례

□ 자료설명

- SARIMA 모형을 이용한 분석사례로 본 연구에서는 이재민·권용재(2011)의 일반철도부문에 적용한 것을 소개할 것임
 - 이재민·권용재(2011)의 연구에서는 일반철도 여객부문의 월별 시계열 자료(2000년 1월 ~ 2008년 12월)와 SARIMA 모형을 일반철도 여객부문의 동태적 변화추이를 분석하고 장래 교통수요를 예측하였음
- 한국철도공사·코레일공항철도·한국철도시설공단의 『철도통계연보』에서 2000년 1월부터 2008년 12월까지의 통행실적 중 수도권 전철을 제외한 자료를 이용
 - 2008년 12월 기준으로 지역 간 철도의 월별 수송인원과 수송인-km는 약 915만 명과 15억 159만 인-km 수준

<표 16> 일반 철도 통행실적(2008년 12월 기준)

단위: 인, 인-km

수송 인원	수송 인-km
9,150,414	1,501,594,900

자료: 한국철도공사·코레일공항철도·한국철도시설공단, 『2010 철도통계연보』, 2011

□ 모형 구분

- 일반 철도 수송인원 및 인-km의 자기상관 및 편자기상관계수값을 살펴보면 자기회귀(autoregressive) 혹은 이동평균(moving average)의 전형적인 특성을 나타내지 않고 있음
 - 특정 시차 이후에 자기상관계수나 편자기상관계수가 급격히 감소(cut off)하지

도 않으며 기하하적으로 감소하는 형태도 띄지 않고 있어서 자기회귀 및 이동 평균 과정을 따른다고 볼 수 없음

- 1계 차분 및 12계 계절차분의 correlogram에서도 특정 시차 이후에 자기상관계수나 편자기상관계수가 급격히 감소하지 않고 있으며, 지수적으로 감소하는 추세도 보이지 않음
- 그러므로 일반 철도 통행실적은 적분된(integrated) ARMA process를 따른다고 볼 수 있음

<표 17> 일반 철도 수송인원의 자기상관(AC) 및 편자기상관계수(PAC)

구분	원자료				1계 차분 및 12계 계절차분			
시차	AC	PAC	Q-Stat	P-value	AC	PAC	Q-Stat	P-value
1	0.095	0.095	0.997	0.318	-0.354	-0.354	12.293	0.000
2	0.111	0.103	2.389	0.303	-0.066	-0.219	12.726	0.002
3	0.314	0.301	13.567	0.004	0.022	-0.100	12.776	0.005
4	0.188	0.150	17.614	0.001	-0.014	-0.069	12.796	0.012
5	0.238	0.194	24.144	0.000	-0.081	-0.139	13.468	0.019
6	-0.068	-0.223	24.675	0.000	0.202	0.131	17.701	0.007
7	0.190	0.088	28.942	0.000	-0.203	-0.117	21.998	0.003
8	0.072	-0.098	29.551	0.000	0.063	-0.020	22.414	0.004
9	0.216	0.296	35.157	0.000	0.042	0.020	22.601	0.007
10	-0.022	-0.191	35.214	0.000	0.040	0.086	22.778	0.012

주: 1. 1계 차분 및 12계 차분은 금기변수에서 전기 및 12기 전변수를 모두 차분한 경우임
2. 자연로그(natural log)를 취하여 분석하였음

자료: 이재민·권용재, “계절성을 감안한 ARIMA 모델을 이용한 교통수요 동태적 변화 연구”, 『대한교통학회지』, 제29권 제5호, 2011, pp.139-155

<표 18> 일반 철도 수송인-km의 자기상관(AC) 및 편자기상관계수(PAC)

구분	원자료				1계 차분 및 12계 계절차분			
시차	AC	PAC	Q-Stat	P-value	AC	PAC	Q-Stat	P-value
1	0.313	0.313	10.862	0.001	-0.135	-0.135	1.798	0.180
2	0.005	-0.103	10.865	0.004	-0.377	-0.403	15.873	0.000
3	0.132	0.182	12.831	0.005	-0.005	-0.159	15.876	0.001
4	0.010	-0.109	12.843	0.012	0.058	-0.160	16.217	0.003
5	-0.017	0.042	12.876	0.025	0.021	-0.080	16.262	0.006
6	0.063	0.034	13.344	0.038	0.021	-0.030	16.308	0.012
7	0.010	-0.018	13.355	0.064	-0.099	-0.136	17.324	0.015
8	-0.034	-0.022	13.494	0.096	-0.025	-0.092	17.392	0.026
9	0.093	0.111	14.543	0.104	0.080	-0.045	18.082	0.034
10	-0.032	-0.124	14.667	0.145	0.164	0.155	20.990	0.021

주: 1. 1계 차분 및 12계 차분은 금기변수에서 전기 및 12기 전변수를 모두 차분한 경우임
2. 자연로그(natural log)를 취하여 분석하였음

자료: 이재민·권용재, “계절성을 감안한 ARIMA 모델을 이용한 교통수요 동태적 변화 연구”, 『대한교통학회지』, 제29권 제5호, 2011, pp.139-155

<표 19> 일반 철도 통행수요 단위근 검정 결과

구분		수송인원			수송인-km		
		모형1	모형2	모형3	모형1	모형2	모형3
원자료	통계량	-0.4417	-2.0456	-1.9467	-0.3070	-7.4116	-7.3968
	P-value	0.5207	0.2672	0.6222	0.5729	0.0000	0.0000
	시차	12	12	12	2	0	0
1계 차분	통계량	-3.7652	-3.7615	-3.8572	-12.5214	-12.4604	-12.3982
	P-value	0.0002	0.0046	0.0177	0.0000	0.0000	0.0000
	시차	11	11	11	1	1	1
1계 차분 및 12계 계절 차분	통계량	-9.8178	-9.7659	-9.7189	-11.0247	-10.9661	-10.9243
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	시차	1	1	1	1	1	1

주: 모형1은 선형추세가 없는 (no deterministic trend) 경우, 모형2는 상수항이 있는(intercept) 경우, 모형 3은 상수항과 시간추세(intercept & time trend)가 있는 경우를 의미함.

자료: 이재민·권용재, “계절성을 감안한 ARIMA 모형을 이용한 교통수요 동태적 변화 연구”, 『대한교통학회지』, 제29권 제5호, 2011, pp.139-155

- ADF test를 이용하여 일반 철도 통행실적을 수송 인원과 수송인-km로 구분하여 단위근 검정을 수행하였음⁸⁾
 - 수송인원은 모든 모형에서 p-value가 커서 단위근(unit root)이 있다는 귀무가설을 기각하지 못하였음
 - 수송인-km는 모든 모형에서 단위근이 존재하지 않는다는 가설을 기각하지 못함으로 단위근이 존재한다고 보아야 할 것임
 - 일반 철도 통행실적을 1계 차분 및 12계 계절차분을 하여 단위근 검정을 수행한 결과에 의하면 모두 p-value가 작아서 단위근이 존재하지 않는 것으로 나타났다
 - 이러한 분석에 입각할 때 일반 철도 통행수요 함수의 동태적 변화추정은 SARIMA 모형을 이용하여야 함

□ 모형 추정

- SARIMA $(p,d,q) \times (P,D,Q)$ 차수를 결정해야 하는데, d 와 D 는 ADF 검정에 의하여 1로 판명되었음

8) 단위근 검정에 대해서는 나. 자기회귀시차모형을 이용한 교통수요 분석의 (2) 시계열 분석 절차 설명에서 상세히 설명하고 있음

- p 와 q 및 P 와 Q 의 차수를 결정하기 위해서는 위에서 제시한 AIC와 SC 통계량을 이용할 수 있음

<표 20> SARIMA 모형 추정 결과

구분	일반철도 수송인원	일반철도 수송인-km
상수항	0.0002 (0.0008)	0.0005 (0.0011)
$AR(1)$	-0.4190*** (0.1242)	0.3833** (0.1500)
$SAR(12)$	- -	-0.3795 (0.2511)
$MA(1)$	-0.1960 (0.1189)	-0.8879*** (0.0748)
$SMA(12)$	-0.8521*** (0.0487)	- -
$Adj R^2$	0.5052	0.3947
F -statistic	32.6499	18.6050
관측치의 개수	94	82
AIC	-3.6531	-2.2036
SC	-3.5449	-2.0862
DW	1.9810	2.0047

주: 1. ***, **, *는 1%, 5%, 10% 유의수준에서 통계적으로 유의함을 의미함

2. 괄호안의 수치는 Newey-West 표준오차(standard error)임

3. AIC 및 SC는 Akaike Information Criteria와 Schwarz Criteria 통계량을 의미하며 DW는 Durbin Waston 통계량을 의미함

자료: 이재민·권용재, “계절성을 감안한 ARIMA 모형을 이용한 교통수요 동태적 변화 연구”, 『대한교통학회지』, 제29권 제5호, 2011, pp.139-155

- 또한 개별 추정치의 t -통계량이 의미 있는 수치가 되어야 한다. d 와 D 의 차수를 1로 보았을 때 p 와 q , 그리고 P 와 Q 의 가능한 차수, 즉 0과 1을 고려하면 모두 가능한 조합은 24로 16개의 모형을 고려할 수 있다. 이때 $(0,1,0) \times (0,1,0)^9$ 을 배제하면 모두 15가지 모형을 모두 고려해야 함
- 이러한 과정을 거쳐서 <표 20>과 같이 일반철도의 여객 수송인원과 인-km에 대한 SARIMA 모형을 추정하였음

□ 모형 진단

- 진단의 한 방법으로 추정에서 얻은 잔차항(residuals)이 자기상관(autocorrelation)이 존재하는지를 Q -통계량을 이용하여 검정
- 귀무가설은 잔차항 간에 자기상관이 존재하지 않는다는 것으로 Q -통계량이 작

9) $(0,1,0) \times (0,1,0)$ 은 AR, MA, SAR, SMA 항이 모두 존재하지 않은 경우로서 추정의 의미가 없으므로 고려할 필요가 없음

게 나타나서 귀무가설을 기각하지 않아야 모형선택이 적절함

<표 21> SARIMA 모형 추정 이후 잔차항의 Q 검정 결과

구분 시차	수송인원		수송인-km	
	Q-통계량	P-value	Q-통계량	P-value
1	0.0006	-	0.0186	-
2	4.8276	-	0.4106	-
3	5.4163	-	0.8759	-
4	5.4504	0.020	1.0749	0.300
5	5.9215	0.052	1.1488	0.563
6	6.4749	0.091	1.5705	0.666
7	7.5242	0.111	3.9543	0.412
8	7.5492	0.183	4.0986	0.535
9	7.5523	0.273	4.6201	0.593
10	7.5523	0.374	5.1147	0.646
11	8.0629	0.427	5.1333	0.743
12	8.0993	0.524	7.8682	0.547
13	8.2078	0.609	8.2352	0.606
14	8.3832	0.679	8.726	0.647
15	9.0911	0.695	8.9668	0.706

자료: 이재민·권용재, “계절성을 감안한 ARIMA 모형을 이용한 교통수요 동태적 변화 연구”, 『대한교통학회지』, 제29권 제5호, 2011, pp.139-155

- 수송인원의 경우에 시차 6까지는 Q-통계량이 다소 크게 산정되어 잔차항 간에 자기상관이 존재하는 것으로 나타나지만 그 이후에는 자기상관이 존재하지 않는 것으로 나타났음
- 또한 수송인-km의 경우도 잔차항 간의 자기상관이 존재하지 않는다는 것을 Q-통계량이 보여 주고 있음
- 따라서 수송인-km의 경우, 모형 선택이 적절하였다고 볼 수 있으며 수송인원의 경우에도 다소 미흡하지만 모형 선택이 적절하다고 볼 수 있음

□ 장래 통행수요 예측

- 수송인원 및 수송인-km 기준 자료를 이용하여 동태적 변화과정을 SARIMA 모형으로 추정하였는데 이를 기반으로 장래 교통수요를 예측할 수 있음
- 이의 기본 원리는 다음과 같음

$$\hat{y}_T(k) = E(y_{T+k} | y_T, y_{T-1}, \dots) \quad (\text{식 } 30)$$

- $\hat{y}_T(k)$ 는 예측시점이 T (본 연구에서는 2008년 12월)이고 k 시점 이후의 y_{T+k} 에 대한 예측치를 의미. 과거 및 현재 정보, y_T, y_{T-1} 등을 이용하여 미래를 예측하는 것이며 이는 과거 및 현재 정보를 이용한 조건부 기대치로 해석
- SARIMA모형에 의거하여 일반 철도의 장래 교통수요를 예측하면 수송인원 및 수송인-km 기준으로 2010년에는 1억 3,212만 명과 189억 3,699만인-km 수준으로 예측되었음
 - 이는 2008년의 1억 1,310만 명과 186억 7,136만 인-km와 비교하여 각각 16.82%와 1.42% 증가한 수치
 - 2015년과 2020년의 수송인원은 각각 1억 5,409만 명과 1억 9,338만 명으로 산정되었으며, 이는 2008년의 1.36배와 1.71배에 해당하는 수치
 - 2015년과 2020년의 수송인-km는 각각 232억 4,543만 인-km와 333억 1,731만인-km로 산정되었으며 이는 2008년의 1.25배와 1.78배에 해당하는 수치
- 이재민·권용재(2011)연구에서는 약 10 여년치(정확하게 9년치)의 실적자료를 이용하였기 때문에 10여년 정도의 장래자료를 예측하였음
 - 약 10 여년치 실적자료를 이용하여 추정하고 지나치게 먼 미래까지 예측한다면 다소 문제가 있을 수 있음
 - 현재의 제약된 정보를 이용하여 먼 미래를 예측하는 것은 왜곡된 예측치를 얻을 수 있기 때문. 따라서 SARIMA 모형에 의한 예측은 실제 모형에 이용된 수준(연도)의 예측을 하여야 할 것임

<표 22> 통행수요 예측

구분	수송인원	수송인-km
2010	132,119,229	18,936,991,011
2011	135,451,675	19,486,412,204
2012	139,275,618	20,176,459,611
2013	143,627,690	21,020,850,980
2014	148,550,330	22,036,764,958
2015	154,092,476	23,245,431,745
2016	160,310,369	24,672,867,037
2017	167,268,498	26,350,802,532
2018	175,040,711	28,317,851,498
2019	183,711,502	30,620,972,463
2020	193,377,522	33,317,306,415
2008년 실적 대비 2020년 예측치 비중	1.71	1.78

자료: 이재민·권용재, “계절성을 감안한 ARIMA 모형을 이용한 교통수요 동태적 변화 연구”, 『대한교통학회지』, 제29권 제5호, 2011, pp.139-155

나. 자기회귀 시차모형을 이용한 교통수요 분석

1) 모형의 개요¹⁰⁾

- 다양한 사회경제적 변수가 교통수요에 어떠한 영향을 미치는 지를 분석할 필요성이 다양한 교통분야 연구에서 발생하고 있음
 - 구체적으로 국내총생산, 유류가격, 고속도로 연장, 자동차 대수, 철도연장 등이 고속도로, 수도권 전철, 일반 철도 등 통행수요에 어떠한 영향을 미치는 지는 주요한 관심사임
- 이러한 변수를 고려하는 모형으로 자기회귀시차모형(Autoregressive Distributed Lag Model, ADL 모형)을 고려할 수 있음
- 자기회귀시차모형은 Hendry et al. (1984)이 제안한 것으로 $ADL(m,n;p)$ 는 p 개의 설명변수, 종속변수 및 설명변수의 시차가 각각 m 및 n 개로 구성되어 있음을 의미함
- 가장 기본적인 $ADL(1,1;1)$ 은 다음과 같은 구조를 띠고 있음(단 $|\alpha_1| < 1$)

$$\ln y_t = \beta_0 + \alpha_1 \ln y_{t-1} + \beta_1 \ln x_t + \alpha_2 \ln x_{t-1} + u_t \quad (\text{식 31})$$

y : 교통수요 변수

x : 설명변수로 유류가격 혹은 국내총생산 등

\ln : 자연로그

u : 교란항

t : 시점을 나타내는 하첨자

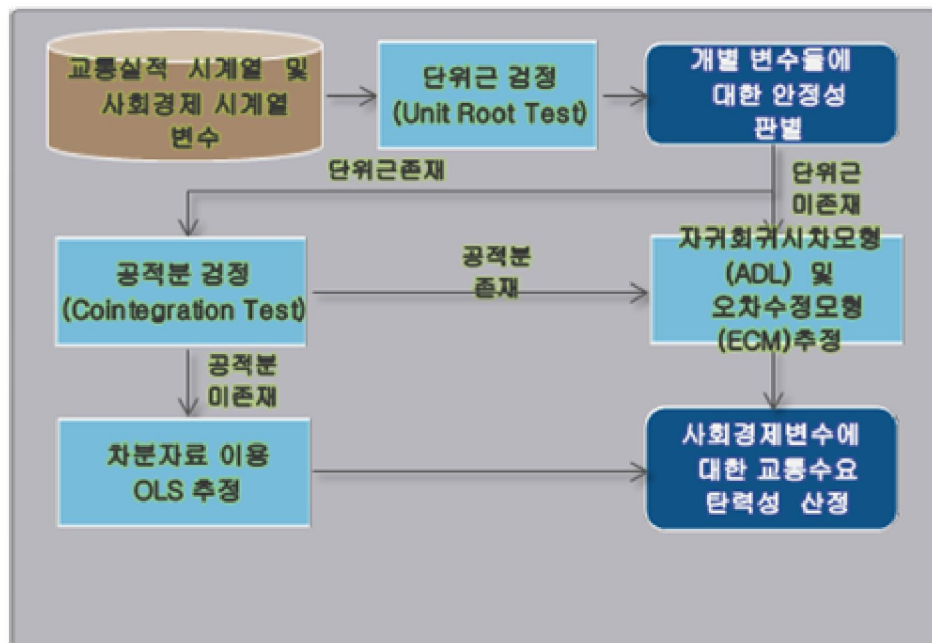
- 이때 설명변수에 대한 종속변수(교통수요)에 대한 장기탄력성은 다음과 같이 정의됨

$$\text{장기수요 탄력성} = \frac{\beta_1 + \alpha_2}{1 - \alpha_1} \quad (\text{식 32})$$

- 만일 설명변수가 유류가격이라면 유류가격에 대한 교통수요의 장기탄력성이 될 것이며 국내총생산이라면 소득에 대한 교통수요의 장기탄력성이 될 것임

10) 이 부분은 한국교통연구원(2010)의 일부 내용을 참조하였음

- 또한 고속도로 차로환산 연장, 수도권 전철 연장, 지역간 철도 연장, 자동차 등록대수 증가량에 대한 장기탄력성을 산정할 수 있음
- 설명변수와 종속변수가 모두 안정적인 시계열변수(stationary series variables)라면 ADL 모형을 이용할 수 있음
 - 설명변수와 종속변수가 불안정적인 시계열변수라 할지라도 이들 간에 공적분 관계가 존재한다면 역시 ADL 모형을 이용할 수 있음
- ADL 모형은 ECM 모형에 비해 추정이 다소 용이하다는 점에서 장점이 있음. 하지만 설명변수가 외생적(exogenous)이라는 가정이 필요함
 - 하지만 설명변수가 1개($p=1$)인 ADL(1,1;1)은 ECM 모형과 동일한 결과를 나타낼 수 있음(Banerjee et al.(1993))
- 오차수정모형(ECM)에 비하여 자기회귀시차모형(ADL)이 분석이 용이
 - 향후 전문가 그룹들이 실제 분석을 수행할 경우 ECM에 비해 ADL이 좀 더 용이한 측면이 있음
 - 오차수정모형은 추정해야 할 계수들이 매우 많고 결정계수가 작게 나타날 가능성이 높으며 이로 인해 예측치와 실제치의 차이가 큰 편임
- 일반적으로 알려져 있는 시계열 분석기법인 단위근 검정(unit root test), 그랜저 인과관계 검정(Granger Test of Causality), 공적분 검정(Cointegration Test), 그리고 ADL 모형을 이용한 추정방법을 따를 수 있음
- 단위근 검정에서 모든 변수들이 안정적이라면 ADL 모형을 이용할 수 있음
 - 그러나 안정적이지 않다면 공적분 검정을 수행하여야 하며 장기균형관계가 존재한다면 ADL 모형을 이용하고 존재하지 않는다면 차분하여 OLS로 추정할 수 있음
 - 또한 설명변수들이 적절하게 선택되었는지를 살펴보기 위하여 Granger 인과관계 검정을 시도할 수 있음
- 이러한 시계열 분석 절차를 나타내면 <그림 2>와 같음



<그림 2> 시계열 분석 절차

2) 시계열 분석 절차 설명

□ 단위근 검정

- 단위근 검정(Unit Root Test)이란 개별 변수들이 불안정적인(Non-Stationary) 시계열인지 혹은 안정적인(Stationary) 시계열인지를 판별하는 과정
- 어떤 시계열 변수가 안정적(stationary)이라는 것은 다음 조건을 충족시켜야 함¹¹⁾
 - 변수의 평균($E(Z_t)$)과 분산($Var(Z_t)$)이 일정해야 하며 변수의 자기공분산(auto covariance, $Cov(Z_t, Z_{t+k})$)이 시차(래그, k)의 영향은 받지만 시간(t)의 영향은 받지 않아야 함
 - 이러한 안정성 조건을 충족시키면 두 개 관측치의 시차(k)가 커질수록 자기상관함수(auto-correlation Function, ACF)¹²⁾가 0으로 수렴하게 됨. 따라서 안정적인 시계열변수들은 어떠한 외부충격이 발생할 때 평균으로 복귀하려는 현상을 보유하게 됨
- 단위근을 보유하고 있는 시계열변수들은 확률적인 추세를 보유하고 있어서 불

11) 이러한 안정성을 Weak Stationary 혹은 Covariance Stationary라고 한다. 이외에도 보다 강한 안정성으로 Strongly Stationary 혹은 Strictly Stationary도 있다.

12) 자기상관함수는 다음과 같이 정의할 수 있음.

$$ACF_k = Cov(Z_t, Z_{t+k}) / [Var(Z_t) Var(Z_{t+k})]^{1/2}$$

안정적(non-stationary)인 특성을 띠게 됨

- 이때 불안정적인 시계열을 안정적인 시계열로 변환시키기 위해서는 불안정적인 시계열을 차분(difference)하여야 함
- 불안정적인 시계열이 안정적인 시계열이 될 때까지 계속하여 차분하는 과정을 거쳐야 하며 안정적인 시계열인지를 ADF(Augmented Dickey Fuller, ADF) test를 통하여 검정하여야 함
- 이때 처음부터 안정적인 변수들을 $Z_t \sim I(0)$ 라고 정의하고 d 번의 차분과정을 거쳐서 불안정적인 시계열로 변하였을 때 그 변수들을 $Z_t \sim I(d)$ 라고 표기함
- 단위근 검정방법으로 DF(Dickey Fuller) test, ADF(Augmented DF) test, PP(Phillips Perron) test가 있는데, 일반적으로 ADF 검정방법을 많이 이용함
- ADF test에서 귀무가설(H_0)은 “단위근이 존재한다”는 것임. 이때 Z 가 선형추세(deterministic linear trend)를 보유하고 있는지를 알 수 없기 때문에 다음의 3가지 경우를 가정하고서 단위근 검정을 함
- 즉 선형추세가 없다고 가정(no deterministic trend)하는 경우, 상수항이 있다고 가정(intercept)하는 경우, 상수항과 시간추세(intercept and time trend)가 있다고 가정하는 경우로 구분하여 단위근 검정을 수행
- 선형추세가 없다고 가정(no deterministic trend)하는 경우(모형 1)에서 귀무가설과 대립가설은 각각 “단위근이 존재한다”($\rho = 0$)와 “단위근이 존재하지 않는다”($\rho < 0$)임

$$\Delta Z_t = \rho Z_{t-1} + \sum_{j=1}^p \theta_j \Delta Z_{t-j} + u_t \quad (\text{식 33})$$

∇ : 종속변수 혹은 설명변수

Δ : 차분기호

u : 교란항

- 상수항(intercept)이 있다고 가정하는 경우(모형 2)의 귀무가설, “단위근이 존재한다”($\rho = 0$, $\alpha = 0$)이고 대립가설은 “단위근이 존재하지 않는다”($\rho < 0$, $\alpha \neq 0$)임. 그러나 귀무가설, “단위근이 존재한다”($\rho = 0$)를 검정하는 것이 일반적임

$$\Delta Z_t = \alpha + \rho Z_{t-1} + \sum_{j=1}^p \theta_j \Delta Z_{t-j} + u_t \quad (\text{식 34})$$

- 상수항과 시간추세(intercept and time trend)가 있다고 가정하는 경우(모형 3)의 귀무가설은 “단위근이 존재한다”($\rho = 0$, $\alpha = \gamma = 0$)이고 대립가설은 “단위근이 존재하지 않는다” ($\rho < 0$, $\alpha \neq 0$, $\gamma \neq 0$)임. 그러나 귀무가설, “단위근이 존재한다”($\rho = 0$)를 검정하는 것이 일반적임

$$\Delta Z_t = \alpha + \gamma trend + \rho Z_{t-1} + \sum_{j=1}^p \theta_j \Delta Z_{t-j} + u_t \quad (\text{식 35})$$

trend: 시간을 나타내는 변수

- 과연 이 세 가지 경우 중에 어떠한 것이 적절한 지를 검정하여야 함
 - 이종원(1994)과 Dickey · Fuller(1981)는 이 세가지중 어느 것이 적절하냐를 검정하는 방법으로 *F*-test를 추천하고 있음
 - 그러나 일반적으로 이용되는 *F*-test의 임계치(critical value)를 이용하는 것이 아니라 변화된 임계치를 이용하여야 함
 - 또한 몇 개의 시차(Lag term)를 두어야 하느냐가 중요한 문제가 될 수 있음. 일반적으로 AIC(Akaike Information Criteria) 혹은 SC(Schwartz Criteria)를 이용하여 시차를 선택할 수 있음

□ Granger 인과관계 검정

- 회귀분석에서 어느 변수가 원인변수(설명변수)이고 어느 변수가 결과변수(종속변수)인지를 확인하기 위하여 Granger는 다음과 같은 인과관계 검정 방법을 제시
 - 어떠한 변수 y_1 이 다른 변수 y_2 와 y_1 자신의 과거값에 의해서 영향을 받고 있다면 y_2 가 y_1 변수를 “Granger cause한다”라고 함
 - 만일 y_2 변수의 과거치가 y_1 변수에 영향을 미치지 못한다면 “Granger cause하지 않는다”고 함
- 다음의 경우에는 “Granger cause 하지 않는다”로 함

$$MSE(y_{1t+1}|y_{1t}, y_{1t-1}, \dots) = MSE(y_{1t+1}|y_{1t}, y_{1t-1}, \dots, y_{2t}, y_{2t-1}, \dots) \quad (\text{식 } 36)$$

- 이때 고려하는 모형은 다음과 같음

$$y_{1t+1} = \sum_{j=1}^p \phi_{11}(j)y_{1t+1-j} + \sum_{j=1}^q \phi_{12}(j)y_{2t+1-j} + u_t \quad (\text{식 } 37)$$

- 이때 귀무가설(null hypothesis)은 y_2 가 y_1 을 “Granger cause 하지 않는다”임.
즉 $\phi_{12}(1) = \phi_{12}(2) = \dots = \phi_{12}(q) = 0$ 를 F -test를 이용하여 검정할 수 있음

□ 공적분 검정

- 모든 변수들이 불안정적인 시계열 변수의 특성을 나타내고 있으므로 ADL 모형을 이용하기 위해서는 변수들간에 장기균형관계(long-run relationship)가 존재하는지를 검정하여야 함
 - 이를 검정하는 방법으로 공적분 검정(cointegration test)을 이용할 수 있음
- 공적분 검정(Cointegration Test)이란 불안정적인 시계열 변수들간에 장기균형관계가 존재하는지를 분석하는 방법
 - 따라서 공적분 검정은 모든 변수가 $I(1)$ 혹은 $I(2)$ 변수의 특성을 지니고 있을 때 수행하는 검정방법임
 - 공적분관계가 존재할 때 안정적인 시계열로 변환한 차분변수들끼리 회귀분석을 시도하는 것도 오류가 있을 수 있음. 즉 차분변수들끼리 회귀분석을 시도하는 것은 장기균형관계를 누락하는 오류를 범하는 것임
- 개별 변수들간에 공적분관계가 있는지를 검정하기 위하여 Johansen Cointegration Test를 이용할 수 있음
 - 만일 공적분 관계가 존재한다면 ADL 모형을 이용하여 방정식을 추정하여야 하며 공적분 관계가 존재하지 않는다면 개별 변수들을 차분하여 추정할 수 있음
- 공적분 검정은 Johansen Cointegration Test¹³⁾를 이용할 수 있는데 기본적인 아이디어는 다음과 같음. P 차 자기상관벡터(P th Order Vector Autoregressive), Z_t 를 정의하면¹⁴⁾

13) Johansen Cointegration Test는 Johansen(1991)과 Johansen and Juselius(1994)의 논문을 참조할 것.

14) 이하의 논의는 김명자·장국현(2002)과 이재민·박수신(2005)에서 요약발췌하였음.

$$Z_t = A_1 Z_{t-1} + \dots + Z_p X_{t-p} + u_t \quad (\text{식 } 38)$$

- Z_t 는 k 차원 불안정적인 $I(1)$ 벡터이고 A_i 는 회귀계수벡터이며, u_t 는 교란항 벡터이다. Z_t 를 다시 정의하면,

$$\Delta Z_t = \Pi Z_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Z_{t-i} + v_t \quad (\text{식 } 39)$$

- 이때 $\Pi = \sum_{i=1}^p (A_i - I)$, $\Gamma_i = - \sum_{j=i+1}^p A_j$
- 만약 Π 의 위수(rank)가 k 이면 Z_t 의 모든 변수들은 안정적 과정이 되고 k 개의 공적분 관계가 존재하며, Π 의 위수가 0이면 공적분 관계가 존재하지 않음
- Trace Test라는 우도비검정(Likelihood Ratio Test)을 이용하여 공적분 관계를 추정할 수 있음
- Trace Test의 통계량을 구하기 위하여 ΔZ_t 와 Z_{t-1} 을 $\Delta Z_{t-1}, \Delta Z_{t-2}, \dots, \Delta Z_{t-p+1}$ 으로 회귀분석을 하여 구한 잔차항, $\widehat{\Delta Z_t}$ 와 $\widehat{Z_{t-1}}$ 으로 다음과 같은 표준상관관계(Canonical Correlation)를 구함

$$S_{00} = \frac{1}{T} \sum_{t=1}^T \widehat{\Delta Z_t} \widehat{\Delta Z_t}' \quad (\text{식 } 40)$$

$$S_{11} = \frac{1}{T} \sum_{t=1}^T \widehat{Z_t} \widehat{Z_t}'$$

$$S_{01} = \frac{1}{T} \sum_{t=1}^T \widehat{\Delta Z_t} \widehat{Z_t}', \quad S_{01} = S_{10}'$$

- 이러한 표준상관관계(Canonical Correlations)를 이용하여 다음과 같은 식을 만족하는 특성근(Eigenvalues)을 구할 수 있음

$$|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0 \quad (\text{식 } 41)$$

- 여기에서 구한 특성근()을 이용하여 다음과 같은 Trace Test의 검정통계량을 구함. 우도비 검정의 귀무가설(H_0)은 “공적분 벡터의 수가 r 개이다”이고 대립가설(H_1)은 “ r 개 보다 크다”임

$$LR(r|k) = -T \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i) \quad (\text{식 42})$$

- 이때 T 는 관측치의 수를 나타내며 $LR(r|k)$ 는 점근적(Asymptotically)으로 비표준적인 분포를 따름 (김명직 · 장국현, 2002)

3) 추정사례

□ 개요

- ADL 모형을 적용하여 추정한 사례로 한국교통연구원(2010)을 참조할 수 있음
 - 한국교통연구원(2010)에서는 월별 고속도로 통행수요 함수가 유류가격, GDP, 차로환산 고속도로 연장, 자동차 등록대수의 로그차분, 유효철도연장에 어떠한 영향을 받고 있는 지를 분석하였음
- 고속도로 통행수요에 대한 추정모형은 다음과 같음

$$\begin{aligned} \ln Y_t = & \beta_0 + \alpha_1 \ln Y_{t-1} + \beta_1 \ln P_t + \alpha_2 \ln P_{t-1} + \beta_2 \ln GDP_t + \alpha_3 \ln GDP_{t-1} \\ & + \beta_3 \ln EX_t + \alpha_4 \ln EX_{t-1} + \beta_4 \ln NC_t + \alpha_5 \ln NC_{t-1} + \beta_5 \ln EXR_t \\ & + \alpha_6 \ln EXR_{t-1} + \beta_6 trend + u_t \end{aligned} \quad (\text{식 43})$$

Y : 통행수요 변수

P : 실질 가중평균 유류가격

GDP : 실질 GDP

EX : 차로환산 고속도로 연장

NC : 자동차 등록대수

EXR : 유효철도연장

$trend$: time trend 변수

t : 월을 나타내는 시간하첨자

u : 교란항

\ln : 자연로그

□ 자료설명

- 우리나라의 도로교통부문 유류가격을 나타내기 위하여 가중평균 유류가격을 이용하였음

- 또한 2005년 소비자물가지수를 이용하여 실질가격으로 변환할 수 있으며 최종적으로 월별가중평균 유류가격은 다음과 같이 정의됨

$$P_{y,m} = \sum_{i=1}^3 w_{y,m}^i OP_{y,m}^i \quad (\text{식 44})$$

$P_{y,m}$: 월별 가중평균 유류가격

$OP_{y,m}^i$: 휘발유, 경유, LPG의 명목가격

$w_{y,m}^i$: 도로교통부문의 유류 이용량 대비 휘발유, 경유, 혹은 LPG 이용량 비율

i : 휘발유, 경유, LPG

y : 연도(2000, 2001, ..., 2009)

m : 월 (1, 2, ..., 12)

- 국내총생산 자료는 한국은행(www.bok.or.kr)에서 발표하는 분기별 실질 GDP를 이용할 수 있음. 그러나 본 연구에서는 월별 GDP를 이용하여야 함으로 이에 대한 조정이 필요
- 분기별 GDP 자료를 월별 GDP 자료로 변환하기 위하여 한국은행에서 발표하는 산업생산지수를 이용하여 월별 GDP를 다음과 같이 산정하였음

$$\text{월별 GDP} = \text{분기별 GDP} \times \frac{\text{당월 산업생산지수}}{\text{당분기 3개월 산업생산지수의 합}} \quad (\text{식 45})$$

- 월별 가중평균 유류가격은 2005년 기준 소비자 물가지수를 이용하여 실질화하였으며 월별 GDP 산정에서도 실질 분기별 GDP를 이용하여 2005년 기준으로 실질화하였음
- 2009년 12월 현재 명목 및 실질 가중평균 유류가격은 리터당 각각 1,385.16원과 1,217.18원으로 산정되었음
- 2009년 12월 현재 명목 GDP 및 실질 GDP는 각각 93조 6,800억원과 85조 4,460억원으로 산정되었으며 당월 현재 자동차 대수는 1,733만여대이며 전달에 비해 62,020대가 증가하였음
- 또한 차로환산 고속도로 연장은 약 4,012.7km이며 유효철도연장은 약 4,692.5km에 달하고 있음

<표 23> 설명변수 자료(2009년 12월 기준)

월별 명목 가중평균 유류가격 (원/리터)	월별 실질 가중평균 유류가격 (원/리터)	월별 실질 GDP(십억원)	월별 자동차 등록 대수 증가량	월별 차로환산 고속도로 연장(km)	유효철도 연장 (km)
1,385.16	1,217.18	85,446	62,020	4,012.7	4,692.5

주: 실질변수는 2005년을 기준으로 하였음

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

□ 단위근, Granger 인과관계 및 공적분 검정

○ 설명변수를 ADF test를 이용하여 단위근 검정을 수행한 결과

- 실질 가중평균 유류가격, 실질 GDP, 차로환산 고속도로 연장, 유효철도연장, 수도권 전철 연장은 $I(1)$ 변수로 판명됨. 그러나 자동차 대수는 $I(2)$ 변수로 판명되었음

<표 24> 단위근 검정 결과(고속도로 통행수요)

구분		Level			1계 차분		
		통계량	P-value	시차	통계량	P-value	시차
실질가중평균 유류가격($\ln P$)	모형 1	0.7974	0.8835	0	-9.9733	0.0000	0
	모형 2	-1.0971	0.7154	0	-9.9836	0.0000	0
	모형 3	-3.1615	0.0974	2	-9.9411	0.0000	0
실질 GDP($\ln GDP$)	모형 1	2.9249	0.9991	3	-11.8007	0.0000	1
	모형 2	-1.0664	0.7272	2	-9.7875	0.0000	2
	모형 3	-5.6574	0.0000	0	-9.7766	0.0000	2
차로환산 고속도로 연장($\ln EX$)	모형 1	2.7131	0.9984	1	-7.4785	0.0000	0
	모형 2	-1.3587	0.6002	1	-8.1601	0.0000	0
	모형 3	-2.4472	0.3537	1	-8.1623	0.0000	0
자동차 대수($\ln NC$)	모형 1	1.4432	0.9625	3	-1.1709	0.2196	2
	모형 2	-2.4566	0.1289	3	-1.8429	0.3582	2
	모형 3	-3.0268	0.1295	3	-2.3085	0.4257	2
자동차 대수 증가량($d\ln NC$)	모형 1	-	-	-	-12.7624	0.0000	1
	모형 2	-	-	-	-12.7181	0.0000	1
	모형 3	-	-	-	-12.6876	0.0000	1
유효철도연장 ($\ln EXR$)	모형 1	1.7635	0.9810	0	-9.8192	0.0000	0
	모형 2	-0.7084	0.8397	0	-10.0203	0.0000	0
	모형 3	-1.4552	0.8393	0	-9.9785	0.0000	0

주: 자동차 대수 증가는 자동차 대수를 1계 차분한 것임

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

- 다른 변수와 달리 자동차 대수는 플로우 변수(flow variable)가 아닌 스톡 변수(stock variable)로서 계속하여 증가하는 추세를 따르고 있음

- 이러한 변수의 특성으로 말미암아 자동차 대수는 $I(2)$ 변수의 특성을 나타내고 있음
- 따라서 ADL 모형에서는 자동차 대수가 아니라 자동차 대수 증가량 변수를 이용할 것임
- Granger 인과관계 검정에서 국내총생산과 차로환산 고속도로 연장은 고속도로 통행수요를 Granger cause하는 것으로 나타났음
 - 자동차 등록대수 증가율은 고속도로 이용대수를 부분적으로 Granger cause하는 것으로 보이지만 이용 대-km를 Granger cause 하지 않은 것으로 나타났음
 - 유류가격 변수와 유효철도 연장은 이용대수 및 이용 대-km를 Granger cause 하지 않은 것으로 나타났음
- 그랜저 인과관계 분석을 기준으로 국내총생산과 차로환산 고속도로 연장만이 고속도로 통행수요에 영향을 미치는 것으로 보임
 - 그러나 본 연구의 목적과 이외에 다른 상황을 고려할 때 유류가격, 자동차 등록대수, 유효철도연장을 분석에서 배제하는 것은 바람직하지 않음. 따라서 언급된 변수들을 모두 회귀분석모형의 설명변수로 고려할 것임

<표 25> Granger 인과관계 검정 결과(고속도로 통행수요)

구분	시차	F-statistic	P-value	구분	시차	F-statistic	P-value
이용대수				이용 대-km			
$\ln P$	1	1.9514	0.1651	$\ln P$	1	2.5284	0.1145
	2	1.1732	0.3131		2	1.6201	0.2024
	3	0.7911	0.5014		3	0.9389	0.4245
$\ln GDP$	1	10.8921	0.0013	$\ln GDP$	1	10.4047	0.0016
	2	5.1775	0.0071		2	4.2104	0.0172
	3	4.9514	0.0029		3	3.2972	0.0232
$\ln EX$	1	18.8209	0.0000	$\ln EX$	1	22.4906	0.0000
	2	10.6405	0.0001		2	6.5386	0.0021
	3	11.2525	0.0000		3	4.3970	0.0058
$d\ln NC$	1	3.3001	0.0719	$d\ln NC$	1	1.5659	0.2133
	2	1.8549	0.1612		2	0.0838	0.9197
	3	2.2555	0.0860		3	0.2665	0.8494
$\ln EXR$	1	3.8051	0.0535	$\ln EXR$	1	3.9689	0.0487
	2	1.6009	0.2063		2	1.4388	0.2415
	3	1.2203	0.3058		3	1.2184	0.3065

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

<표 26> 공적분 검정 결과(고속도로 통행수요)

구분		Trace Test			Max-Eigenvalue Test		
전체 이용대수	공적분 관계의 개수	통계량	임계치	p-value	통계량	임계치	p-value
	0	133.6486	117.7082	0.0034	66.1505	44.4972	0.0001
	1	67.4980	88.8038	0.6050	24.4891	38.3310	0.7075
	2	43.0090	63.8761	0.7357	17.5506	32.1183	0.8292
전체 이용 대-km	공적분 관계의 개수	통계량	임계치	p-value	통계량	임계치	p-value
	0	140.0980	117.7082	0.0009	72.1826	44.4972	0.0000
	1	67.9154	88.8038	0.5887	27.5177	38.3310	0.4892
	2	40.3977	63.8761	0.8355	16.3768	32.1183	0.8940

주: 임계치는 5% 유의수준에서의 임계치를 의미함

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

- 종속변수(대기준 및 대-km 기준 고속도로 통행실적)와 설명변수와의 공적분 관계를 살펴보면 장기균형관계가 존재하는 것으로 판명되었음
- 대기준 이용대수, 유류가격, GDP, 차로환산 고속도로 연장, 자동차 등록대수의 로그차분, 유효철도연장 간에 장기균형관계가 존재하는 것으로 판명되었음
- 대-km 기준 통행수요, 유류가격, GDP, 차로환산 고속도로 연장, 자동차 등록대수의 로그차분, 유효철도연장 간에도 장기균형관계가 존재하는 것으로 판명되었음
- “공적분 관계가 존재하지 않는다”라는 귀무가설을 기각하고 있지만 “공적분 관계가 1개이다”라는 귀무가설을 기각하지 못하고 있음
- 따라서 ADL 모형을 이용할 수 있음

□ ADL 모형 추정

- ADL 모형을 이용하여 고속도로 통행수요 함수를 추정하였음
 - 설명변수로는 유류가격, GDP, 차로환산 고속도로 연장, 자동차 등록대수의 로그차분, 유효철도연장 등을 이용하였음
 - 모두 자연로그를 취한 변수를 적용하였지만 자동차 등록대수의 경우에 로그차분을 이용하였음
 - 종속변수와 설명변수 모두 계절조정(seasonal adjustment)을 하였음
- 이때 유효철도연장은 부호가 예측된 것과 달리 나타나서 이를 제외하고 분석하

였음

- 실제 고속도로 이용자들은 철도를 고속도로의 완전한 대체재로 여기지 않는 것으로 보임
- 본 연구에서는 자동차 등록대수 증가율을 추가하는 경우와 추가하는 않는 경우로 구분하였음
- 이는 자동차 등록대수가 $I(2)$ 변수로 나타났기 때문임
- 또한 *trend* 변수를 추가하는 경우와 추가하지 않는 경우로 구분하여 분석하였음

<표 27> 고속도로 이용대수 추정(ADL 모형)

구분	모형 1	모형 2	모형 3	모형 4
$\ln Y_{t-1}$	0.7896*** (0.0543)	0.7632*** (0.0600)	0.8267*** (0.0524)	0.8016*** (0.0578)
$\ln P_t$	-0.0906 (0.0600)	-0.0992 (0.0617)	-0.0890 (0.0597)	-0.0964 (0.0603)
$\ln P_{t-1}$	0.0142 (0.0497)	0.0239 (0.0514)	0.0243 (0.0516)	0.0314 (0.0520)
$\ln GDP_t$	0.2396*** (0.0806)	0.2708*** (0.0857)	0.2091*** (0.0733)	0.2385*** (0.0778)
$\ln GDP_{t-1}$	-0.1044 (0.0815)	-0.0719 (0.0830)	-0.0912 (0.0825)	-0.0608 (0.0853)
$\ln EX_t$	0.4681*** (0.1402)	0.4981*** (0.1362)	0.5205*** (0.1393)	0.5470*** (0.1367)
$\ln EX_{t-1}$	-0.3003* (0.1612)	-0.2857* (0.1657)	-0.3781** (0.1603)	-0.3647** (0.1646)
$d\ln NC_t$	- -	- -	2.8054** (1.3755)	2.7144* (1.3809)
$d\ln NC_{t-1}$	- -	- -	-1.9947 (1.4508)	-2.0966 (1.4943)
<i>trend</i>	- -	-0.0003 (0.0002)	- -	-0.0003 (0.0002)
상수항	0.0728 (0.3399)	-0.6835 (0.7058)	-0.0370 (0.3545)	-0.7192 (0.6982)
관측치 개수.	119	119	118	118
R^2	0.9889	0.9890	0.9891	0.9892
F-statistic	1,412.441***	1,234.46***	1,087.935***	976.7998***

주: 1. ***, **, *는 1%, 5%, 10%에서 통계적으로 유의함을 나타냄

2. 괄호안의 수치는 Newey-West standard error임

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

- 대기준 고속도로 통행수요함수에서 유류가격에 대한 고속도로 장기 통행수요 탄력성은 -0.32~-0.37로 나타났음¹⁵⁾

- 또한 국내총생산에 대한 고속도로 장기 통행수요 탄력성은 0.64~0.90로 나타남
- 차로환산 고속도로 연장에 대한 고속도로 장기 통행수요 탄력성은 0.80~0.92로 산정되었음
- 자동차 등록대수 증가율에 대한 고속도로 장기 통행수요 탄력성은 3.11~4.68로 산정됨

<표 28> 고속도로 이용 대-km 추정(ADL 모형)

구분	모형 1	모형 2	모형 3	모형 4
$\ln Y_{t-1}$	0.5884*** (0.1185)	0.4864*** (0.1424)	0.6215*** (0.1071)	0.5253*** (0.1290)
$\ln P_t$	-0.2423** (0.0975)	-0.2718*** (0.1014)	-0.2470*** (0.0887)	-0.2740*** (0.0915)
$\ln P_{t-1}$	0.0518 (0.0959)	0.0861 (0.1032)	0.0900 (0.0872)	0.1189 (0.0927)
$\ln GDP_t$	0.1499 (0.1556)	0.2906*** (0.1561)	0.0908 (0.1385)	0.2259 (0.1387)
$\ln GDP_{t-1}$	0.0313 (0.1503)	0.1591 (0.1603)	0.0663 (0.1522)	0.1854 (0.1622)
$\ln EX_t$	0.5941*** (0.1650)	0.7386*** (0.1611)	0.7011*** (0.1678)	0.8316*** (0.1685)
$\ln EX_{t-1}$	-0.1932 (0.2286)	-0.1326 (0.2388)	-0.3041 (0.2108)	-0.2449 (0.2194)
$d \ln NC_t$	- -	- -	6.5308** (3.1341)	6.1688* (3.1450)
$d \ln NC_{t-1}$	- -	- -	-2.7465 (2.7081)	-2.7792 (2.6637)
$trend$	- -	-0.0015** (0.0007)	- -	-0.0014** (0.0006)
상수항	2.2892** (1.0603)	-0.7608 (1.4966)	1.8461* (1.0171)	-0.9942 (1.4297)
관측치 개수.	119	119	118	118
R^2	0.9597	0.9618	0.9603	0.9622
F-statistic	377.4481***	345.7356***	290.314***	272.4139***

주: 1. ***, **, *는 1%, 5%, 10%에서 통계적으로 유의함을 나타냄

2. 괄호안의 수치는 Newey-West standard error임

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

15) 월별 자료를 이용하고 있어서 장기 탄력성이란 용어는 정확하기 않음(Hughes et al., 2006). 그러나 편의상 장기 탄력성이란 용어를 본 과업에서 계속 이용할 것임.

- 대-km 기준 고속도로 통행수요함수에서 유류가격에 대한 고속도로 장기 통행수요 탄력성은 -0.33~-0.46으로 나타났음
- 또한 국내총생산에 대한 고속도로 장기 통행수요 탄력성은 0.41~0.88으로 나타남
- 차로환산 고속도로 연장에 대한 고속도로 장기 통행수요 탄력성은 0.97~1.24로 산정되었음
- 자동차 등록대수 증가율에 대한 고속도로 장기 통행수요 탄력성은 7.14~10.00으로 산정됨

다. 오차수정모형을 이용한 교통수요 분석

1) 모형의 개요

- ADL 모형과 별도로 불안정적인 변수들의 장기균형관계를 분석하는 방법으로 오차수정모형(Error Correction Model, ECM)을 고려할 수 있음
 - 오차수정모형(ECM)은 앞에서 분석한 자기회귀시차모형(Autoregressive Distributed Lag Model, ADL)과 밀접한 관계가 있음¹⁶⁾
 - 또한 설명변수가 하나이며 종속변수와 설명변수의 시차가 1개인 ADL(1,1;1)모형은 공적분 관계가 하나인 단순한 오차수정모형(ECM)으로 전환될 수 있음 (Maddala & Kim(2004), Banerjee et al.(2003))
- 오차수정모형(ECM)의 추정절차는 다음과 같음
 - 첫째 개별 변수들이 안정적인지를 단위근 검정을 통하여 수행
 - 둘째 불안정적이라면 개별 변수들이 장기균형관계가 존재하는지를 공적분 검정을 통하여 검증
 - 셋째 오차수정모형을 추정
- 추정하여야 할 기본적인 ECM 모형은 다음과 같음

$$\Delta \ln Y_t = \alpha_0 + \alpha_1 \Delta \ln Y_{t-1} + \alpha_2 \Delta \ln X_{t-1} + \gamma_1 (\ln Y_{t-1} + \beta_0 + \beta_1 \ln X_{t-1}) + \Delta u_{t-1} \quad (\text{식 46})$$

16) 일부 논문 및 보고서에서 시계열변수들의 안정성 혹은 공적분 관계를 검정하지 않고서 자기회귀시차모형을 이용하는 경우가 있는데 이는 유사회귀모형의 결과일 수 있음.

- 추정모형에서 설명변수로 삽입된 것을 살펴보면¹⁷⁾ 차분시차 종속변수 ($\Delta \ln Y_{t-1}$), 차분시차 설명변수($\Delta \ln X_{t-1}$), 그리고 오차수정항으로 이루어져 있음
 - 차분시차 종속변수의 계수값은 전기의 종속변수가 금기의 종속변수에 영향을 미치는 정도를 파악하며 차분시차 설명변수는 단기적으로 전기의 설명변수가 금기의 설명변수에 미치는 영향을 의미하고 있음
 - 오차수정모형의 계수는 균형상태로부터의 이탈이 금기의 종속변수에 미치는 효과를 의미
 - 오차수정모형안의 계수($-\beta_1$)들은 장기적인 균형관계를 나타내는 것으로 설명변수가 종속변수에 미치는 장기적인 관계를 의미
- 오차수정모형의 추정절차는 자기회귀 시차모형의 추정절차와 동일함
 - 1단계에서 개별 변수들이 단위근이 존재하는 지를 검정하며
 - 2단계에서 개별 변수들이 단위근이 존재한다면 장기균형관계가 존재하는 지를 검정
 - 3단계에서 오차수정모형을 추정
- 오차수정모형과 자기회귀 시차모형의 차이점은 다음과 같음
 - 오차수정모형은 벡터자기회귀모형(Vector Autoregressive model, VAR model)에서 오차수정항이 삽입된 것으로 모든 변수들이 내생적(endogenous)이라고 가정하고 있음
 - 이러한 점에서 자기회귀 시차모형보다 덜 제한적(less restrictive)이라 볼 수 있음

2) 추정사례

□ 개요

- ECM 모형을 적용하여 추정한 사례로 한국교통연구원(2011)을 참조할 수 있음
 - 한국교통연구원(2011)에서는 월별 철도 통행수요 함수가 운임, 유류가격, 국내 총생산, 열차 정차횟수 및 열차 정차가능역수에 어떠한 영향을 받고 있는 지를 분석하였음

17) 벡터오차수정모형의 설명은 이종원(1994)을 참조하였음.

- 구체적으로 2005년 4월부터 2011년 6월까지 월별 자료를 이용하였으며 KTX와 새마을 노선을 분석하였는데 본 연구에서는 KTX 사례를 소개할 것임
- 철도 통행수요에 대한 추정모형은 다음과 같음

$$\ln Y_t = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln OP_t + \beta_3 \ln RGDP_t + \beta_4 \ln NOA_t + \beta_5 \ln NOS_t + u_t \quad (\text{식 47})$$

y : 월별 철도 여객인원

p : 월별 운임

op : 월별 리터당 가중평균 유류가격

$RGDP$: 월별 국내총생산

NOA : 월별 정차횟수

NOS : 월별 정차가능역수

u : 교란항

- 월별 정차횟수는 예를 들어 KTX 열차가 특정 연월에 정차한 횟수를 의미하며 월별 정차가능역수는 KTX 열차가 정차가 가능한 특정 역의 개수를 의미함
 - 월별 정차횟수와 정차가능역수는 해당 차종의 여객수송인원에 큰 영향을 미칠 것이라고 가정하여 모형의 설명변수로 설정하였음

□ 자료설명

- 월별 가중평균 유류가격과 월별 국내총생산 자료 구축은 고속도로 부문에서 설명하였음
 - 월별 가중평균 유류가격은 명목기준으로 약 1,661.45원이며 월별 국내총생산은 실질기준으로 약 88,884.9 십억 원 수준
- KTX 수송인원은 2011년 6월 현재 각각 4,008,508명 수준이며 KTX의 1인당 운임은 명목기준으로 각각 29,668.89원 수준
 - 1인당 운임은 (운임 수입÷여객인원)으로 정의하였음
 - 이때 설명변수의 내생성 문제가 있을 수 있지만 현재 오차수정모형을 이용하므로 이를 어느 정도 해결할 수 있음
- KTX의 정차횟수 및 정차가능역수는 2011년 6월 현재 각각 38,526번과 61개 수준

<표 29> 철도 시계열 분석 변수설명 자료(2011년 6월 현재)

구분	명목	실질
KTX 수송인원(명)	4,008,508	-
KTX 운임(원)	29,668.89	24,601.07
KTX 정차횟수	38,526	-
KTX 정차가능역수	61	-
가중평균 유류가격(원/리터)	1,661.45	1,377.65
국내총생산(십억원)	-	88,884.9

주: 1. 실질은 2005년을 기준으로 하였음

2. 한국은행에서 2011년 2분기 명목 국내총생산을 발표하지 않아서 이를 제시하지 않았음

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

□ 단위근 및 공적분 검정

- ADF(Augmented Dickey Fuller) 단위근 검정 결과를 살펴보면
 - KTX 수송인원, KTX 운임, KTX 정차횟수, KTX 정차가능역수, 실질국내총생산, 실질 가중평균 유류가격 모두 단위근이 있는 것으로 분석되었음
- 모든 변수들을 1계 차분을 하여 ADF 단위근 검정을 시도하였을 때, 모두 단위근이 없는 것으로 분석되었음
 - 따라서 모든 변수들을 $I(1)$ 변수로 볼 수 있음
- KTX 수송인원에 대한 운임 탄력성을 산정하기 위하여 단위근 검정을 수행한 변수들을 대상으로 공적분 검정을 수행하였음
- (식 47)의 주요 변수들에 대한 공적분 검정을 수행한 결과에 의하면 KTX 수송인원, KTX 운임, 가중평균 유류가격, 국내총생산 등이 의미 있는 공적분 결과와 장기균형관계를 보이고 있음
 - 이들 변수간의 공적분관계를 살펴보면 모형 1~3까지는 공적분 관계가 존재하지 않지만 모형 4와 5는 공적분 관계가 존재하는 것으로 나타났음
- 이런 경우 매우 애매한 검정 결과이지만 본 연구에서는 공적분 관계가 존재한다고 판단하였음

<표 30> 단위근 검정 결과(KTX 통행수요)

구분		수준변수			1계 차분		
		통계량	p-value	시차	통계량	p-value	시차
KTX 수송인원	모형1	1.6178	0.9733	2	-9.9136	0.0000	1
	모형2	-0.9739	0.7584	2	-10.1593	0.0001	1
	모형3	-1.9946	0.5942	2	-10.0963	0.0000	1
KTX 운임	모형1	-0.7335	0.3955	0	-10.8189	0.0000	0
	모형2	-2.2677	0.1851	0	-10.8089	0.0001	0
	모형3	-3.5599	0.0404	0	-10.7339	0.0000	0
KTX 정차횟수	모형1	0.4122	0.7996	2	-9.2760	0.0000	1
	모형2	-5.7057	0.0000	0	-9.2357	0.0000	1
	모형3	-7.1128	0.0000	0	-9.1656	0.0000	1
KTX 정차가능 역수	모형1	1.6285	0.9740	0	-8.3777	0.0000	0
	모형2	0.4788	0.9849	0	-8.6124	0.0000	0
	모형3	-1.2094	0.9010	0	-8.8669	0.0000	0
실질국내 총생산	모형1	1.3683	0.9560	2	-11.5898	0.0000	1
	모형2	-1.3956	0.5799	2	-11.7442	0.0001	1
	모형3	-3.3996	0.0594	2	-11.6610	0.0001	1
실질 가중평균 유통가격	모형1	0.5453	0.8316	1	-4.5338	0.0000	0
	모형2	-2.8872	0.0517	1	-4.5491	0.0004	0
	모형3	-3.2524	0.0826	1	-4.5177	0.0028	0

주: 1. 자연로그를 취한 값에 대해 단위근 검정을 수행하였음

2. KTX 및 새마을 운임은 2005년 실질 기준 변수를 이용하여 단위근 검정을 수행하였음

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

<표 31> 공적분 검정 결과(KTX 통행수요)

구분	모형 1	모형 2	모형 3	모형 4	모형 5
Trace	0	0	0	1	1
Max-Eigenvalue	0	0	0	1	1

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

□ 오차수정모형 추정

- KTX 운임 변화에 대한 KTX 수송인원 변화를 추정하는 기본 모형은 다음과 같음

$$\Delta \ln Y_t = \alpha_1 \Delta \ln Y_{t-1} + \alpha_1 \Delta \ln P_{t-1} + \alpha_2 \Delta \ln OP_{t-1} + \alpha_3 \Delta \ln RGDP_{t-1} \quad (\text{식 48})$$

$$+ \gamma (\ln Y_{t-1} + \beta_1 \ln P_{t-1} + \beta_2 \ln OP_{t-1} + \beta_3 \ln RGDP_{t-1}) + \Delta u_t$$

- 오차수정모형 추정결과에 의하면 KTX 운임은 KTX 수송인원에 음(-)의 영향을 미치고 있음을 알 수 있음

- 즉 KTX 운임이 인상되면 KTX 수송인원이 감소하는 것으로 나타났음
- 또한 가중평균 유류가격과 국내총생산은 KTX 수송인원에 양(+)의 효과가 있는 것으로 분석되었음
 - 경기가 활성화되거나 소득이 증가하였을 때 KTX 이용수요가 증가하는 것으로 보이며 유류가격이 인상되면 승용차의 대체재인 KTX 이용이 증가하는 것으로 볼 수 있음
- 그러나 장기균형 관계식에서 KTX 운임과 가중평균 유류가격이 통계적으로 유의하지 않게 나타나고 있음
 - 한국교통연구원(2010)에서는 KTX 운임과 가중평균 유류가격의 t -통계량이 각각 1.2955, -1.2536로서 한계적으로 유의하다(marginally significant)고 보았음
- 장기균형 관계식에서 KTX 운임의 계수추정치는 약 0.2454로 산정되었고 가중평균 유류가격의 계수추정치는 약 -0.3391, 국내총생산의 계수추정치는 약 -1.3338로 산정되었음
 - 개별 변수들의 부호는 일반적인 예상과 부합되지만 국내총생산의 계수값이 다소 크게 산정되었음
- 장기균형관계식에서 KTX 운임의 계수추정치가 장기탄력성이 되는데 KTX 운임에 대한 KTX 수송수요의 장기탄력성은 약 0.2454로 나타났음
 - KTX 운임이 1% 증가하면 KTX 수송인원이 약 0.25% 감소하는 것으로 볼 수 있음

<표 32> 오차수정모형 추정 결과(KTX)

구분	$\Delta \ln Y_{t-1}$	$\Delta \ln P_{t-1}$	$\Delta \ln OP_{t-1}$	$\Delta \ln RGDP_{t-1}$	γ
추정치	-0.3071**	-0.3398	-0.0682	-0.1056	-0.3094***
표준오차	0.1132	0.4285	0.3324	0.2098	0.1055
구분	$\ln P_{t-1}$	$\ln OP_{t-1}$	$\ln RGDP_{t-1}$	R^2	관측치 개수
추정치	0.2454	-0.3391	-1.3338***	0.3313	
표준오차	0.1894	0.2705	0.2515		

주: ***, **, *는 각각 1%, 5%, 10% 수준에서 통계적으로 유의함을 나타내고 있음

자료: 한국교통연구원, 『교통정책지원 및 분석시스템 개발』, 한국건설교통기술평가원, 2010

라. 기타 시계열 분석 모형

1) Time Varying Coefficient or Random Coefficient Model

- 지금까지 모형은 계수값들이 시간에 따라 변하지 않는다는 가정을 하고 있었음
 - 즉 다음과 같은 모형을 가정하고 있었음

$$y_t = x_t\beta + u_t \quad (\text{식 49})$$

y, x : 종속변수 및 상수항(constant term)이 포함된 설명변수

β : 계수(parameter)

u : 교란항

t : 시간을 의미하는 하첨자

- 그러나 실제로는 다양한 원인에 의해 계수값들이 시간에 따라 변할 수 있음
 - 정책변화(policy changes) 및 경제의 상태 변화(evolution of the economy)에 의해 계수값들이 변화할 수 있음
 - 따라서 다음과 같이 정의할 수 있음(Chow, 1984)

$$y_t = x_t\beta_t + u_t \quad (\text{식 50})$$

- 계수값들이 시간(t)에 의존적이며(time dependent) 시간의 함수로 볼 수 있음
 - 이러한 모형을 Time Varying Coefficient model 혹은 Random Coefficient Model이라고 함
- 만약 계수값들이 시간에 따라 조금씩 변화한다면 특정 기간의 계수값들은 인접한 기간의 계수값들에 영향을 받을 것이며 유사할 수 있음
 - 혹은 금기의 계수값들이 전기의 계수값들의 함수로 볼 수 있음

$$\beta_t = M\beta_{t-1} + e_t \quad (\text{식 51})$$

M : 설명변수(x)와 차수가 동일한 정방행렬(square matrix)

e : 교란항

- 이 때 e_t 의 분산이 0이고 M 이 단위행렬(identity matrix)이라면 (식 50)은 (식 49)와 같은 일반적인 회귀분석모형으로 돌아감
- 그러나 e_t 의 분산이 0이 아니고 M 이 단위행렬(identity matrix)이 아니라면

(식 50)은 유효할 것임

- 따라서 계수값들은 특정 시점 혹은 기간의 정보(information)에 따라 변화할 수 있는 것이 Time Varying Coefficient Model 혹은 Random Coefficient Model의 특징임
- 이때 특정 시점 혹은 기간의 계수값들의 시점(혹은 기간)에 대한 조건부 기대치는 다음과 같이 정의할 수 있음

$$E(\beta_t | I_s) = \beta_{t|s} \quad (\text{식 52})$$

I : 특정 시점(기간)의 정보(information)

E : 기대치

- 이때 $t=s$ 이면 filtering, $t>s$ 이면 prediction, 그리고 $t<s$ 이면 smoothing
- 이러한 계수값들의 시간 의존성을 고려하면 좀 더 다른 방법으로의 추정이 필요함을 알 수 있음

2) Chow test or Bai-Perron test

□ Chow test

- 가장 기본적인 Time Varying Coefficient Model로서는 추정모형의 자료들을 구분하여 추정하는 방법을 들 수 있음
- 예를 들어 1997년 하반기와 1998년 하반기가 포함된 자료를 이용하여 모형을 추정할 때 이 기간은 외환위기 기간으로 다른 기간과 다른 특성을 보유했을 것임
- 따라서 이 기간 전후를 구분하여 추정하는 방법을 들 수 있음
- 예를 들어 1970년부터 2010년까지의 연도별 자료를 이용할 때 1970년부터 1997년까지를 기간 I로 하고 1998년부터 2010년까지를 기간 II로 구분하여 따로이 추정
- 즉 (식 50)을 기간 I과 기간 II로 구분하여 추정하면 기간 I의 계수추정치($\hat{\beta}_I$)와 기간 II($\hat{\beta}_{II}$)의 계수추정치를 얻을 수 있음
- 이러한 추정 이후에 전체 기간을 추정하고 기간별 추정과 비교할 필요가 있음
- 비교의 방법으로 Chow test를 들 수 있음
- Chow test는 두 기간간의 추정계수값이 동일하다는 귀무가설(null hypothesis)

과 동일하지 않다는 대립가설(alternative hypothesis)을 비교하는 검정방법

- Chow test 통계량은 다음과 같음

$$\frac{(SSE_R - SSE_U)/k}{SSE_U/(T-2k)} \quad (\text{식 53})$$

SSE_R : 두 기간의 추정계수값이 동일하다는 경우의 sum of squared residuals

SSE_U : 두 기간의 추정계수값이 동일하지 않다는 경우의 sum of squared residuals
 로서 두 개 기간의 회귀분석 수행 후의 sum of squared residuals의 합으로
 정의할 수 있음

T : 전체 자료의 개수

k : 설명변수의 개수

- Chow test 통계량은 자유도가 k 와 $T-2k$ 인 F 분포를 따르고 있음

○ Chow test를 적용한 연구보고서로서 이재민 외(2009)를 들 수 있는데 서울지역
 지하철 통행수요가 유류가격, 국내총생산 등과 같은 사회경제변수에 어떠한 영
 향을 받는 지를 분석하였음

- 이재민 외(2009)에서는 2000년 1월부터 2008년 12월까지의 우리나라의 교통부
 문 유류(휘발유, 경유, LPG)의 실질 가중평균 가격을 산정하면, 2004년 6월까지
 는 리터당 1,000원 미만이었지만 그 이후부터는 리터당 1,000원을 넘어섰다고
 하였음¹⁸⁾

- 따라서 2000년 1월부터 2004년 6월까지를 기간 I로 구분하고 2004년 7월부터
 2008년 12월까지를 기간 II로 구분하여 추정하고 그 결과를 모든 기간에 대해
 추정한 것과 비교하였음

○ 추정모형은 다음과 같으며 기본적으로 자기회귀 시차모형(ADL)을 적용하였음

$$LNNPSA_t = \beta_0 + \gamma_1 LNNPSA_{t-1} + \beta_1 LNRPOR_t \quad (\text{식 54})$$

$$+ \gamma_2 LNRPOR_{t-1} + \beta_2 LNRGDP_t + \gamma_3 LNRGDP_{t-1} + u_t$$

18) 휘발유, 경유 및 LPG의 교통부문 이용량을 기준으로 가중평균하였으며 2005년 소비자물가지수를 이용하여
 실질화하였음

$LNNPSA$: 월별 서울 지하철 통행실적(인원 기준)¹⁹⁾

$LNRPOR$: 월별 가중평균 유류가격

$LNRGDP$: 월별 국내총생산

u : 교란항

t : 특정 연도(year)와 월(month)을 나타내는 하첨자

<표 33> 지하철 통행수요 추정 결과(수송인원)

구분	전체 기간 (2000.1-2008.12)	기간 I (2000.1-2004.6)	기간 II (2004.7-2008.12)
$LNNPSA_{t-1}$	0.0937 (0.1390)	-0.0048 (0.1853)	0.0248 (0.1827)
$LNRPOR_t$	0.0371 (0.1154)	-0.1824 (0.1602)	0.1231 (0.1174)
$LNRPOR_{t-1}$	0.1904 (0.1254)	0.2704* (0.1569)	0.0945 (0.1188)
$LNRGDP_t$	0.0991 (0.1496)	-0.3133 (0.3443)	0.1217 (0.1834)
$LNRGDP_{t-1}$	-0.1768 (0.1640)	0.1717 (0.3343)	-0.3284* (0.1895)
상수항	15.9611*** (2.4871)	19.4257*** (3.8354)	18.7541*** (3.7423)
관측치 개수.	107	53	54
R^2	0.4264	0.1468	0.1850
유류가격 장기탄력성	0.2510	0.0876	0.2231
F 통계량	6.7816 (0.0001)	1.5835 (0.2059)	2.9831 (0.0404)

주: 1. 괄호안의 수치는 Newey-West robust standard errors임

2. ***, **, *는 각각 1%, 5%, 10% 수준에서 통계적으로 유의함을 나타내고 있음

3. F 통계량은 $LNNPSA_{t-1}$, $LNRPOR_t$, 및 $LNRPOR_{t-1}$ 에 대한 joint significant test의 통계량을 의미하며
괄호안의 수치는 p-values임

자료: 이재민 외, 『Oil Price and Travel Demand』, 한국교통연구원, 2009.

- 지하철 통행수요를 전체 기간과 두 기간으로 구분하여 추정하였으며 이때 두 기간의 통행수요가 동일한가를 살펴보기 위하여 Chow test를 수행한 결과에 의하면 F 통계량이 약 3.0688로 산정되어 귀무가설 즉, 두 기간의 통행수요가 동일하다는 가설을 기각하고 있었음
- 즉 기간 I과 기간 II의 수송인원 기준 통행수요함수가 상이하며 유류가격 탄력성도 상이하게 나타나는 것을 알 수 있음

19) 모든 변수는 자연로그를 취하였음

- 즉 유류가격이 낮을 때는 탄력성이 낮게 나타났으며 유류가격이 높을 때에는 탄력성이 크게 나타나고 있음을 알 수 있음

□ Bai-Perron test²⁰⁾

- Chow test는 계수값들이 기간에 따라 상이한지 혹은 동일한지를 검정하는 방법으로 볼 수 있음
 - 그러나 Chow test에서는 기간을 어떻게 구분해야 하는 지에 대한 정보는 제공하지 못함
 - 이재민 외(2009)의 연구에서는 2004년 7월을 기점으로 실질 가중평균 유류가격이 1,000원을 상회하고 있다는 사실에 주목하여 이를 근거로 기간을 구분하였음. 그러나 이러한 기간 구분은 다소 자의적으로 볼 수 있음
- 좀 더 논리적인 방법에 의해 기간을 구분하는 방법으로 Bai-Perron test(1998, 2003)를 이용할 수 있음
 - Bai-Perron test는 어떤 기간 동안에 구조적인 변화가 존재하는 지를 검정하는 것이며 만일 구조적인 변화가 존재한다면 어떻게 기간 구분을 해야 하는지에 대한 정보를 제공함
 - 예로 들어 새로운 운송수단의 등장이나 도로망의 확장은 기존의 교통흐름에 급격한 변화를 줄 수 있으며 이러한 변화는 과거 데이터에서는 예측될 수 없는 것임. 따라서 이러한 불연속적인 사건을 내포하는 시계열 자료를 분석할 때는 교통 시계열에 구조적인 변화가 있었는지를 확인하여야 함
- 다음과 같은 k 개의 다중회귀분석을 가정할 때

$$y_t = \beta_j + u_t, (t = T_{j-1} + 1, \dots, T_j, j = 1, \dots, k) \quad (\text{식 55})$$

- y_t 는 분석하고자 하는 시계열 자료이고 β_j 는 j 번째 국면(regime)에서의 시계열의 평균값이며 u_t 는 t 기의 잔차항
- 총 k 개의 구조변화 시점들(structural break points)이 있고, 따라서 $k+1$ 개의 국면이 존재. (T_1, T_2, \dots, T_k) 는 구조변화 시점들의 집합²¹⁾
- Bai-Perron 구조변화검정의 목적은 구조변화가 존재하는지 여부와 존재한다면

20) 이후의 내용은 이재민·권용재(2011)의 내용 중 일부를 발췌·요약하였음

21) m 이 최대 구조변화 시점의 수라면 $k = 1, \dots, m$ 이다.

구조변화 시점을 내생적으로 추정하는 것임. 우선 (식 56)과 같이 각 k 개의 분할에 대하여 모든 경우의 수를 고려하여 잔차제곱의 합을 계산

$$S_T(T_1, T_2, \dots, T_k) = \sum_{j=1}^{k+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - \beta_j)^2, \quad (k=1, \dots, m) \quad (\text{식 56})$$

- 이 중 최소의 잔차제곱의 합을 내는 구조변화 시점들의 집합이, 주어진 구조변화시점의 수에서 최적의 구조변화 시점들의 집합이 됨

$$(\hat{T}_1, \hat{T}_2, \dots, \hat{T}_k) = \argmin S_T(T_1, T_2, \dots, T_k), \quad (k=1, \dots, m) \quad (\text{식 57})$$

- 위에서 추정된 $(\hat{T}_1, \hat{T}_2, \dots, \hat{T}_k)$ 에 대응하는 개별 국면 하에서의 시계열 평균값인 $\hat{\beta}(\{\hat{T}_1, \hat{T}_2, \dots, \hat{T}_k\})$ 을 얻을 수 있으며 이러한 과정을 통해 총 m 개의 구조변화점과 그에 대응하는 시계열의 평균값을 구하게 됨
- 다음으로 구조변화가 통계적으로 유의한 수준으로 존재하느냐에 대한 검정을 해야 함. 이를 위해 $\sup F$ 검정통계량과 비가중 이중극대값 검정통계량(unweighted double maximum statistic)을 사용. $\sup F$ 검정통계량은 k 개의 구조변화점이 있다는 가정 하에 ‘시계열에 k 개의 구조변화가 존재한다.’는 대립가설 하에서 ‘시계열에 구조변화가 없다’는 귀무가설을 검정하는 F -통계량임

$$H_0: \delta_1 = \delta_2 = \dots = \delta_{k+1} \quad (\text{식 58})$$

$$\sup F_T(\lambda_1, \dots, \lambda_k; 1) = \sup \frac{1}{T} \left(\frac{T - (k+1)}{k} \right) ((R\hat{\delta})' (R\hat{V}(\hat{\delta})R')^{-1} (R\hat{\delta})) \quad (\text{식 59})$$

- $\sup F$ 검정통계량과 달리 구조변화의 수를 명시적으로 지정하지 않고 구조변화가 존재하는 지 여부를 검정하는 방법도 있는데 이는 비가중 이중극대값 통계량을 통해 가능
- 비가중 이중극대값 통계량은 여러 k 값 하에서의 $\sup F$ 검정통계량 중 최대값이며 통계량의 유의성은 별도의 통계표를 이용하여 검정
- 시계열에 구조변화가 존재한다면 몇 번의 구조변화가 있는지를 통계적으로 측정해야 하며 이를 위해 베이즈 정보기준(Bayesian Information Criteria)과 Liu et al. (1997)이 제안한 수정된 슈와르츠 기준(modified Schwarz Criteria)을 이용하

여 최적의 구조변화의 횟수를 추정

- 이재민·권용재(2011)는 일반철도 수송실적을 SARIMA 모형을 이용하여 추정하면서 동 기간(2000년 1월-2008년 12월) 중에 어떠한 구조적 변화가 존재하는지를 Bai-Perron test를 이용하여 검정하였음

<표 34> 수송인원과 수송인-km 시계열에 대한 Bai & Perron 구조변화 검정 결과

구분		수송인원	수송인-km
SupF test	0 vs 1	1.2071 (7.0400)	1.4882 (7.0400)
	0 vs 2	2.7148 (6.2800)	5.3852 (6.2800)
	0 vs 3	2.2981 (5.2100)	4.3330 (5.2100)
	0 vs 4	1.6403 (4.4100)	3.2807 (4.4100)
	0 vs 5	1.4238 (3.4700)	1.4255 (3.4700)
UDmax test		2.7148 (7.4600)	5.3852 (7.4600)
BIC	0 breaks	11.4780	11.5832
	1 break	11.5560	11.6604
	2 breaks	11.6342	11.7258
	3 breaks	11.7096	11.7977
	4 breaks	11.7933	11.8812
	5 breaks	11.8779	11.9786
	선택된 구조변화의 수	0	0
LWZ	0 breaks	11.4874	11.5925
	1 break	11.6392	11.7435
	2 breaks	11.7915	11.8831
	3 breaks	11.9414	12.0295
	4 breaks	12.1000	12.1880
	5 breaks	12.2600	12.3607
	선택된 구조변화의 수	0	0

주: 1. UDmax test는 비가중 이중극대값 통계량을 이용한 구조변화 검정임.

2. SupF test와 UDmax test의 괄호 안의 숫자는 10% 유의수준의 임계치(critical value)임.

3. BIC: 베이즈 정보기준(Bayesian Information Criteria), LWZ: Liu et al.(1997)가 제안한 수정된 슈와르츠 기준

자료: 이재민·권용재, “계절성을 감안한 ARIMA 모형을 이용한 교통수요 동태적 변화 연구”, 『대한교통학회지』, 제29권 제5호, 2011, pp.139~155

- 우선 supF 검정통계량을 보면 두 시계열 모두 한 번에서 다섯 번의 구조변화가 있었다는 대립가설들에 대해 구조변화가 없었다는 귀무가설이 10% 유의수준에서 기각되지 않았음

- 귀무가설에서 구조변화의 수가 명시되지 않는 비가중 이중극대값 통계량을 살펴 보면 두 시계열에서 10% 유의수준에서 귀무가설이 기각되지 않았음
- 따라서 이는 유의한 수준의 구조변화가 없었다는 통계적인 증거임
- 베이즈 정보기준과 수정된 슈와르츠 정보기준으로 구조변화의 수를 추정
 - 베이즈 정보기준의 경우 수송인원과 수송인-km 모두 구조변화의 수가 0일 때 베이즈 정보기준 통계량이 가장 작았으며 수정된 슈와르츠 정보기준 역시 베이즈 정보기준과 동일하게 구조변화의 수가 0일 때 통계량이 가장 작았음²²⁾
 - 두 정보기준 모두 산출된 통계량이 작은 모형일수록 더 적합한 모형으로 판정되므로 결국 두 교통시계열에서 동 기간 동안(2000년 1월-2008년 12월) 유의한 수준의 구조변화가 없었다고 판단하였음

3) Rolling and Recursive Estimation

□ Rolling Estimation

- time varying coefficient model에서 개별 계수값들이 시간에 따라 변화하고 인접 기간의 계수값과 밀접한 관련을 가진다면 rolling estimation을 방법을 이용할 수 있음
- rolling estimation이란 어느 기간(예를 들어 2000년 1월부터 2009년 12월까지)의 샘플을 이용할 때 특정 개수의 샘플로 구분하여 추정하는 방법을 의미함
 - 특정 시점, t 에서 특정 기간을 $[t-w/2, \dots, t+w/2]$ 로 정하여 회귀분석을 수행하는 것임. 이때 w 는 window width라고 하며 자료의 개수와 동일
 - 예를 들어 2000년 1월부터 2003년 12월까지 회귀분석을 수행하고 그 다음에는 2000년 2월부터 2004년 1월까지, 2000년 3월부터 2004년 2월까지, 2000년 4월부터 2004년 3월까지, 시점을 달리하여 구간별 회귀분석을 수행할 수 있음
- rolling estimation은 계수값(true coefficients)과 샘플링 에러(sampling error)의 조합으로 볼 수 있음
- rolling estimation을 서울지역 도시철도 통행수요 함수의 경우에 적용해 볼 수 있음
 - 2000년 1월부터 2008년 12월까지 서울지역 도시철도 통행수요가 유류가격, 국

22) 베이즈 정보기준 통계량은 수송인원이 11.4780, 수송인-km이 11.5832였고 수정된 슈와르츠 기준 통계량은 수송인원이 11.4874, 수송인-km가 11.5925였다.

내총생산 등과 같은 사회경제변수에 어떠한 영향을 받는 지를 분석한 이재민 외(2009)의 연구자료를 이용하여 rolling estimation을 적용할 수 있음

- 추정모형은 time constant 회귀분석모형과 rolling estimation을 비교하였음. 단순한 회귀분석모형은 다음과 같음

$$LNNPSA_t = \beta_0 + \beta_1 LNRPOR_t + \beta_2 LNRGDP_t + u_t \quad (\text{식 } 60)$$

- 즉 서울지역 도시철도 통행수요가 유류가격 및 국내총생산에 어떠한 영향을 받는 지를 Time Constant 회귀분석모형으로 추정할 수 있음
- 또한 이와 별도로 2000년 1월부터 2008년 12월까지의 통행실적을 32개씩 구분하여 추정할 수 있음. 현재 설명변수가 2개이므로 중심극한정리(central limit theorem)가 적용되기 위해서는 자유도(degree of freedom)가 최소한 30개가 되어야 하므로 32개씩 구분하여 회귀분석을 수행하였음
- 즉 2000년 1월부터 2002년 8월까지, 2000년 2월부터 2002년 9월까지, ..., 2006년 5월부터 2008년 12월까지 계속하여 회귀분석을 수행하는 것임
- Time Constant 회귀분석모형의 추정결과에 의하면 국내총생산은 서울지역 도시철도 통행수요에 큰 영향을 미치지 못하는 것으로 나타났음
- 그러나 유류가격은 서울지역 도시철도 통행수요에 양(+)의 효과가 있으며 5% 수준에서 통계적으로 유의함을 알 수 있음
- 즉 유류가격이 1% 상승하면 서울지역 도시철도 통행수요는 인원기준으로 약 0.18% 증가하는 것을 알 수 있음

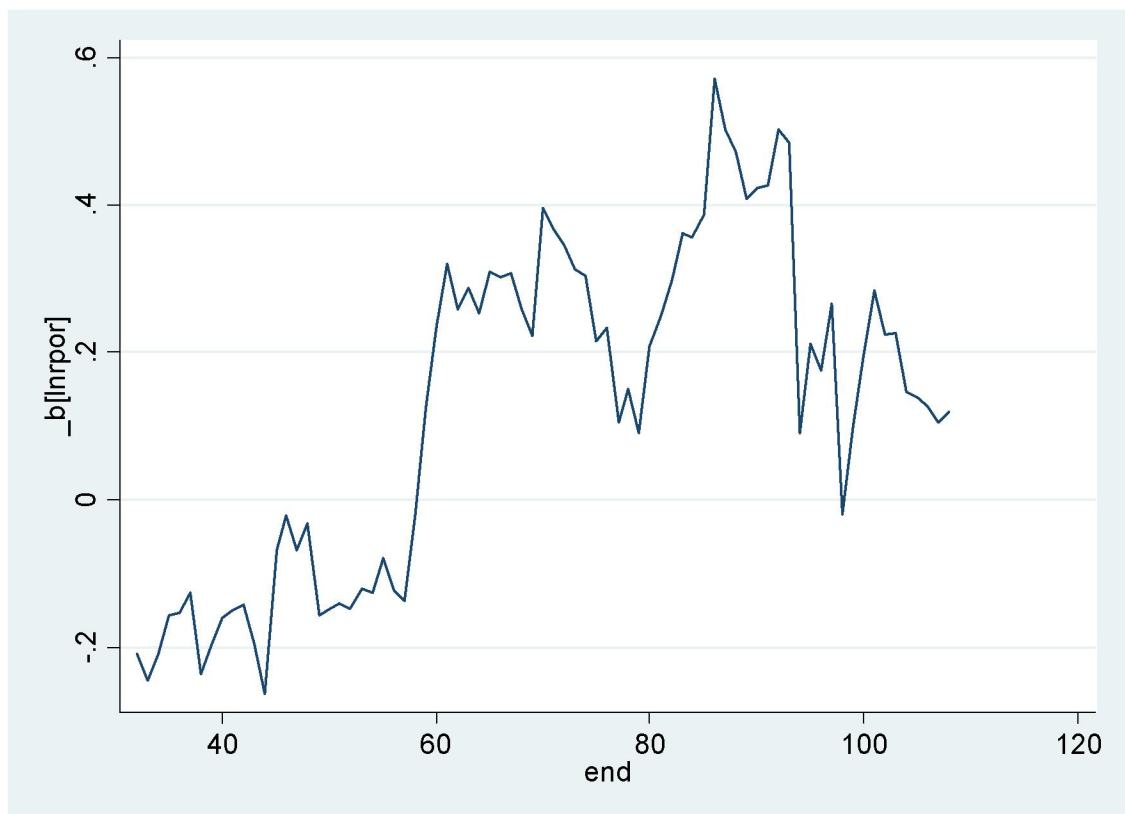
<표 35> Time Constant 회귀분석 추정결과

구분	추정계수	표준오차
$\ln rpor_t$	0.1834**	0.0840
$\ln rgdp_t$	-0.0042	0.1092
상수항	17.1711***	0.7326
관측치 개수	108	-
R^2	0.17	-
F-통계량	12.67***	-

주: 1. 표준오차는 robust standard errors임

2. ***, **, *는 각각 1%, 5%, 10% 수준에서 통계적으로 유의함을 나타내고 있음

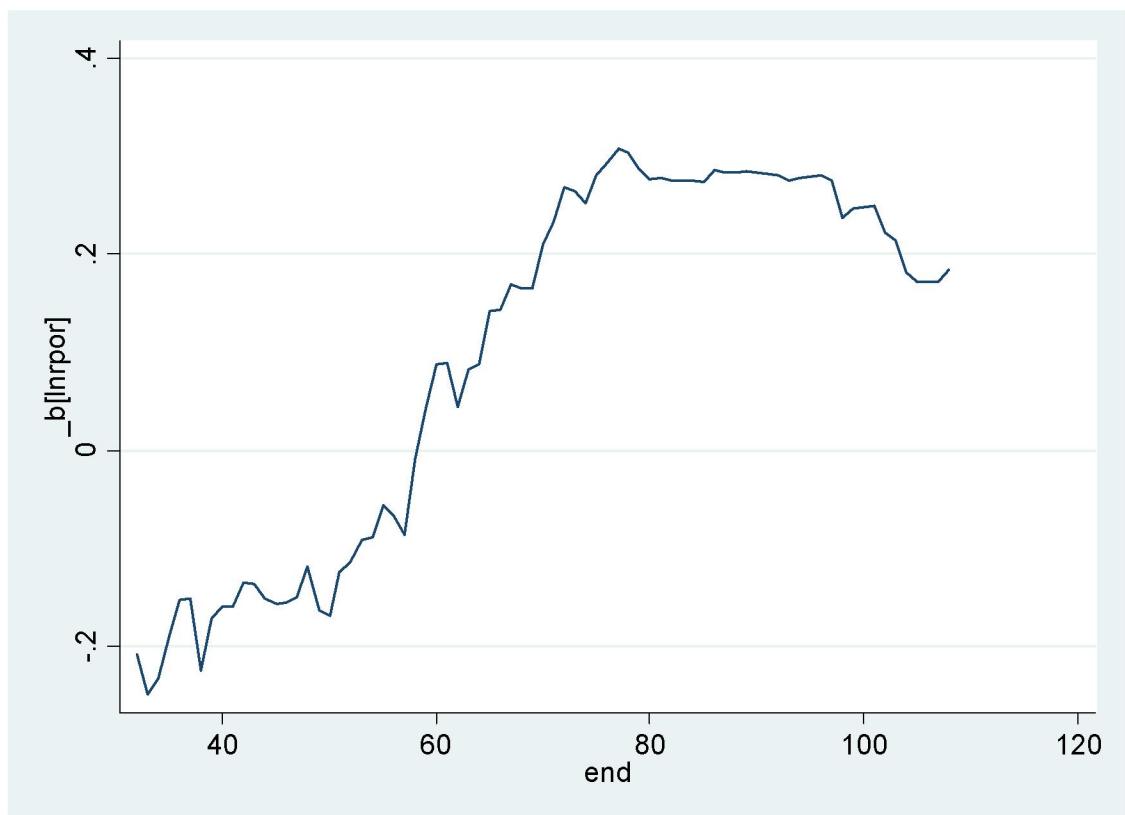
- Time Varying Coefficient Model의 일종인 rolling estimation 추정결과는 <그림 3>에서 볼 수 있는데 유류가격 계수추정치를 보이고 있음
 - 분석초기에는 유류가격 계수추정치가 음(-)의 부호를 나타내고 있지만 종점(end date)이 2004년 12월 이후부터는 양(+)의 부호를 나타내고 있음
 - 2004년 12월 이후의 유류가격 탄력성이 최대 0.6 수준이며 평균적으로 약 0.2 수준임을 알 수 있으며 이는 Time Constant 회귀분석모형에서의 유류가격 탄력성 0.18과 유사한 수준
- 2004년 12월 이전에는 유류가격이 상승하더라도 서울시민들은 도시철도 통행을 증가시키지 않고 감소시키는 경향을 나타내었음
 - 그러나 유류가격이 높은 시기(2004년 12월 이후)에는 유류가격이 상승하면 도시철도 통행을 증가시키는 방향으로 서울시민들이 대응하고 있음
 - 그러나 유류가격이 높은 시기(2004년 12월 이후)에도 유류가격 탄력성의 fluctuation이 심하여 나타나고 있음
- rolling estimation model에서의 계수추정치는 시간에 따른 변화를 보여주고 있어서 유용한 정보를 제공하고 있음



<그림 3> rolling estimation 추정결과(유류가격 계수추정치)

□ Recursive Estimation

- recursive estimation(혹은 sequential estimation)은 rolling estimation을 확장한 개념임
 - 즉 첫 번째는 $[1, w]$, 두 번째는 $[1, w+1]$, 마지막에는 $[1, T]$ 의 자료를 이용하는 분석방법임
 - 앞의 자료를 이용할 때 초기에 2000년 1월부터 2002년 8월까지의 자료를 이용하여 회귀분석을 수행하고 두 번째는 2000년 1월부터 2002년 9월까지의 자료를 이용하며 세 번째는 2000년 1월부터 2002년 10월까지의 자료를 이용하여 마지막에는 2000년 1월부터 2008년 12월까지의 자료를 이용하여 회귀분석을 수행하는 것을 의미함



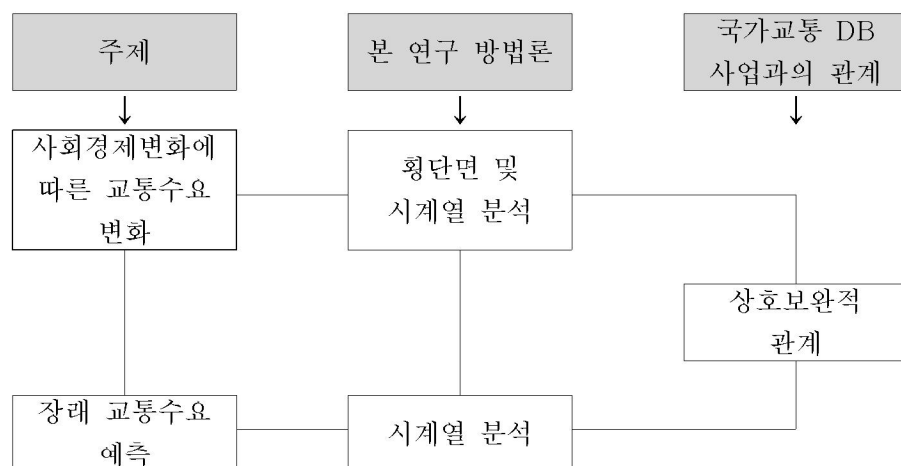
<그림 4> recursive estimation 추정 결과(유류가격 계수추정치)

- recursive estimation을 수행한 후에 유류가격 탄력성을 살펴보면
 - 초기에 유류가격 계수추정치가 음(-)의 부호를 나타내고 있지만 종점(end date)이 2004년 12월 이후부터는 양(+)의 부호를 나타내고 있음
 - 종점(end data)이 2005년 12월경에 유류가격 탄력성이 최대치(약 0.3)에 도달하지만 그 이후 감소하여 안정적인 추세를 보이고 있음
- 종점(end data)이 2004년 12월 이후부터는 유류가격이 증가하면 서울지역 도시철도 통행수요가 증가하는 형태를 보이고 있음
 - 또한 2004년 12월 이전에는 유류가격이 증가하더라도 서울지역 도시철도 통행수요가 증가하지 않고 감소하는 추세를 보이고 있음
- 2000년대 후반의 유류가격이 높은 시기에는 서울시민들이 유류변화에 대응하여 도시철도 통행을 증가시키는 행태를 보이고 있음

6. 결론

가. 국가교통수요조사 및 DB 구축사업과의 연계방안 모색

- 한국교통연구원 국가교통DB센터에서는 국토해양부의 국가교통수요조사와 DB 구축사업을 대행하여 수행하고 있음. 국가교통 수요 및 DB 사업에서는 전국을 대상으로 조사사업을 수행하여 여객 및 화물O/D를 구축하고 장래의 여객 및 화물교통 발생량을 예측하여 교통투자사업에 대한 평가기준을 제시하고 있음
- 그런데 이러한 국가교통 수요 및 DB 사업에서는 주로 4단계 교통수요 분석모형을 이용하고 있음
- 본 연구에서 구축된 교통수요분석 방법론은 기존 4단계 교통수요 분석모형을 보완하는 형태가 될 수 있음
- 구체적으로 시계열 및 횡단면분석에서 수행된 사회경제변수에 따른 탄력성 분석은 4단계 모형에서 도출되는 결과와 비교분석할 수 있음
- 또한 시계열분석에서 구축되는 장래 교통수요는 국가교통 수요 및 DB 사업에서 구축된 장래 교통수요와 비교분석을 위한 기초자료로 이용될 수 있음
- 따라서 본 연구에서 수행된 거시적 교통수요 분석모형은 국가교통 수요 및 DB 사업에서 이용되는 4단계 교통수요 분석모형의 보완적 방법론으로 볼 수 있음



<그림 5> 본 연구방법론과 국가교통 수요 및 DB 구축사업과의 관계

- 실제 본 연구에서 수행된 통행요금(운임) 변화에 따른 교통수요 변화량 분석은 전국 OD 분석을 통한 통행요금 변화에 따른 교통수요 변화량 분석과 비교할

수 있음

- 예를 들어 2011년 11월에 수행된 고속도로 통행요금 변화에 따른 전국 고속도로 통행량 변화를 본 연구방법론에 따라 분석할 수 있으며 또한 2011년 12월에 수행된 전국 철도 운임변화에 따른 철도이용객 변화를 본 연구방법론에 따라 분석할 수 있음
- 이러한 분석결과와 국가교통수요 및 DB 사업을 통하여 분석된 일반화된 비용 혹은 요금변화에 따른 OD 통행량 변화를 비교분석할 수 있음
- 또한 시계열 분석에서 수행된 철도 혹은 고속도로 통행량 예측은 국가교통 수요 및 DB 사업에서 구축된 장래 교통수요 변화와 비교분석할 수 있음
- 본 연구에서 수행된 고속도로 및 철도의 통행수요 예측치는 국가교통 수요 및 DB 사업의 장래 교통수요 예측과 비교분석하여 장래 교통수요 예측의 정확도를 높이는데 참고자료로 이용할 수 있음

나. 연구결과 요약 및 시사점

- 본 연구에서는 거시적 교통수요 분석모형을 고찰하였음. 구체적으로 자료특성에 따라서 횡단면 자료를 이용한 분석과 시계열 자료를 이용한 분석방법으로 구분하여 고찰하였음
- 구체적으로 횡단면 자료를 이용한 분석에서는 중력방정식을 이용한 최소자승법 및 Tobit 추정방법을 이용하는 모형을 제시하였으며 count data model인 Poisson, Negative Binomial, Zero Inflated Poisson 및 Zero Inflated Negative Binomial 모형을 제시하였음
- 또한 시계열 자료를 이용한 분석에서는 ARIMA 및 SAIMA를 이용한 교통수요 동태적 변화모형과 자기회귀 시차모형 및 오차수정모형을 제시하였음. 또한 Time Varying Coefficient model로 Rolling Estimation 및 Recursive Estimation model을 제시하였음
- 개별 모형의 이론적 개요와 이를 이용한 추정사례를 다른 논문 및 연구보고서의 부분을 인용하거나, 새로이 본 연구를 통하여 분석한 사례를 제시하였음
- 본 연구는 다음과 같은 점에서 기본 연구보고서와 차별성이 있음
- 자료별 특성에 따라 분석모형을 달리 제시하였음
- 기존 연구에서 이용된 모형과 사례를 다수 제시하였음

- 이외에도 국내에서 잘 이용되지 않았던 분석모형을 제시하였으며 새로운 분석한 사례를 제시하였음
- 전체적으로 본 연구보고서는 기존 교통부문 논문과 연구보고서에서 이용된 경제학 방법론을 이용한 교통수요 분석모형을 총망라한 것이라 볼 수 있음
- 이러한 점에서 향후 경제학적 방법론을 이용한 교통수요 분석모형의 기본적인 지침서가 될 것이라 예상할 수 있음
- 한편 본 보고서는 많은 유형과 자료별 특성에 따른 분석방법론을 제시하였지만 패널자료(panel data)를 이용한 분석방법론을 제시하지 않은 점에서 다소 아쉬움이 있음
- 또한 본 연구보고서에서 언급되지 않은 다수의 모형이 있으리라 짐작됨
- 그러나 본 연구보고서는 향후 더 나은 연구환경에서 좀 더 세련된 계량경제학 방법론을 이용한 교통수요 분석모형을 구축할 수 있는 계기가 될 것으로 기대

참고문헌

1. 김명직 · 장국현, 『금융시계열분석』, 경문사, 2002
2. 문진수 · 이재민, 『철도화물증대를 위한 제도개선 방안』, 한국교통연구원, 2007
3. 이재민 · 권용재, “계절성을 감안한 ARIMA 모형을 이용한 교통수요 동태적 변화 연구”, 『대한교통학회지』, 제29권 제5호, 2011, pp.139~155
4. 이재민 · 박수신, “시계열 분석을 통한 고속도로 통행수요함수의 추정”, 『대한교통학회지』, 제23권 제2호, 2005, pp. 83~92.
5. 이재민 · 박수신 · 박선주 『통행요금 원가산정을 위한 수요예측 연구』, 한국교통연구원, 2005
6. 이재민 · 한상용 · 이창운, 『Oil Price and Travel Demand』, 한국교통연구원, 2009
7. 이재민 · 한상용 · 유정복, 『고속도로 차종별 통행요금 조정에 관한 연구』, 한국교통연구원, 2009
8. 이종원, 『계량경제학』, 박영사, 1994
9. 한국개발연구원 · 한국교통연구원, 『고속도로 통행요금 제도 및 법령 개선을 위한 교통수요 분석』, 한국도로공사, 2008
10. 한국교통연구원, 『교통정책 지원 및 분석시스템 개발-1차 연도』, 한국건설교통기술평가원, 2010

11. 한국교통연구원, 『교통정책 지원 및 분석시스템 개발-2차 연도』, 한국
건설교통기술평가원, 2011
12. 한국교통연구원, 『제주공항 마스터플랜 수립연구 용역』, 한국공항공사,
2009
13. 한국철도공사·코레일공항철도·한국철도시설공단, 『2010 철도통계연
보』, 2011.
14. Bai, J. & Perron, P., "Estimating and Testing Linear Models with
Multiple Structural Changes", *Econometrica*, Vol. 66, 1998, pp. 47~78
15. Bai, J. & Perron, P., "Computation and Analysis of Multiple Structural
Change Models", *Journal of Applied Econometrics*, vol. 18, 2003, pp. 1~
22
16. Benerjee, Anindya, Dolado, Juan J., Galbraith, John W., and Hendry,
David F., *Co-Integration, Error Correction, and the Econometric Analysis of
Non-Stationary Data*, Oxford University Press, Oxford, 1993
17. Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel, *Time
Series Analysis: Forecasting and Control*, 3rd edition, Prentice Hall,
1994.
18. Burris, M. W., "The Toll-Price Component of Travel Demand
Elasticity", *International Journal of Transport Economics*, Vol. 30, No. 1,
2003, pp.45~59
19. Cameron. A. C. and Trivedi, P. K., "Econometric Models Based on
Count data: Comparisons and Applications of Some Estimation and
Tests", *Journal of Applied Econometrics*, vol. 1, Issue 1, pp. 29~53
20. Chow, G. C., "Random and Changing Coefficient Models", in *Handbook
of Econometrics II* edited by Griliches, Z. and Intriligator, M. D., pp.
1213~1245, North-Holland, 1984
21. Dickey. D. A. and Fuller, W. A., "Likelihood Ratio Statistics for
Autoregressive Time Series with a Unit Root," *Econometrica*, vol. 49,
1981, pp. 1057~1072
22. Dickey, D. A., Hasza, D. P., and Fuller, W. A., "Testing for a Unit
Root in Seasonal Time-Series", *Journal of the American Statistical
Association*, Vol. 79, No. 386, 1984, pp. 355-367
23. Eichengreen, B., "Trade Blocs, Currency Blocs, and the Reorientation of
World Trade in the 1930s", *Journal of International Economics*, Vol. 38,

Issue 1 - 2, 1995, pp. 1~24

24. Eichengreen, B., and Irwin, D. A., "The Role of History in Bilateral Trade Flows", *The Regionalization of the World Economy* edited by Frankel, J. A., University of Chicago Press, 1997
25. Engle, R. F., Granger, C. W. J., Hylleberg, S., and Lee. H. S., "Seasonal Co-Integration: The Japanese Consumption Function", *Journal of Econometrics*, Vol. 55, Issue 1-2, 1993, pp. 275-298
26. Felbermayr, G. J. and Kohler, W., "Exploring the Intensive and Extensive Margins of World Trade", *Review of World Economics*, Vol. 142, No. 4, 2006, pp. 642~674
27. Graham, Daniel and Stephen Glaister, *Review of Income and Price Elasticities of Demand for Road Traffic*, Imperial College, 2002
28. Greene, william H., *Econometric Analysis*, 3rd edition, Prentice Hall, 1997
29. Hanly, Mark, Joyce Dargay, and Phil Goodwin, *Review of Income and Price Elasticity in the Demand for Road Traffic*, Economic and Social Research Council, 2002
30. Hamilton, James D., *Time Series Analysis*, Princeton University Press, 1994
31. Hasza, D. P. and Fuller, W. A., "Testing for Nonstationary Parameter Specifications in Seasonal Time-Series Models", *Annals of Statistics*, Vol. 10, No. 4, 1982, pp. 1209-1216
32. Hendry, D. F., Pagan, A. R., and Sargan, J. D., "Dynamic Specification", in *Handbook of Econometrics II* edited by Griliches, Z. and M. D. Intrilligator, North-Holland, Amsterdam, 1984, pp. 1023~1100
33. Hylleberg, S., Engle, R. F., Granger, C. W. J., and Yoo, B. S., "Seasonal Integration and Co-Integration", *Journal of Econometrics*, Vol. 44, Issue 1-2, 1990, pp. 215-228
34. Johansen, Soren, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, Vol. 59, No. 6, 1991, pp. 1551~1580
35. Johansen, S. and Juselius, K., "Identification of the Long-Run and the Short-Run Structure: An Application to the ISLM Model", *Journal of*

- Econometrics*, Vol. 63, No. 1, 1994, p.7~36
36. Jong, De and H. Gunn, "Recent Evidence on Car Cost and Time Elasticities of Travel Demand in Europe", *Journal of Transport Economics and Policy*, Vol. 35, pp. 2001, 137~160
 37. Judge George J., Hill, R. Carter, Griffiths, William E., Lutkepohl, Helmut, and Lee, Tsoung-Chao, *Introduction to the Theory and Practice of Econometrics*, 2nd edition, Wiley, 1988
 38. Liu, J., Wu, S., and Zidek, J. V., "On Segmented Multivariate Regression", *Statistica Sinica*, Vol. 7, 1997, pp. 497~525
 39. Ljung, G. M., and Box, G. E. P., "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, Vol. 65, 1978, pp. 297~303
 40. Maddala, G. S. and Kim, I., *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, 2004
 41. Osborn, D. R., Chui, A. P., Smith, J. P., and Birchenhall, C. R., "Seasonality and the Order of Integration for Consumption", *Oxford bulletin of Economics and Statistics*, Vol. 50, Issue 4, 1988, pp. 361-378
 42. Wardman, M., "Demand for Fair Travel and the Effects of External Factors", *Transportation Research Part E*, Vol. 42, Issue 3, 2006, pp. 129~148
 43. Washington, S. P., Karlaftis, M. G., and Mannering F. L., *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC, 2003
 44. Wooldridge, J. M., "Quasi-Likelihood Methods for Count Data", in *Handbook of Applied Econometrics Volume II: Microeconometrics* edited by Pesaran, H. M. and Schmidt, P., 1997, pp. 352~406
 45. Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data*, MIT press, 2002
 46. Vuong, Q., "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", *Econometrica*, Vol. 57, No. 2, 1989, pp. 307~333