

계량경제모형간 국내 총화물물동량 예측정확도 비교 연구

정성환¹ · 강경우^{2*}

¹ 한양대학교 교통공학과, ² 한양대학교 교통물류공학과

A Comparative Study on the Forecasting Accuracy of Econometric Models : Domestic Total Freight Volume in South Korea

CHUNG, Sung Hwan¹ · KANG, Kyung Woo^{2*}

¹ Department of Transportation Engineering, Hanyang University, Gyeonggi, 426-791, Korea

² Department of Transportation and Logistics Engineering, Hanyang University, Gyeonggi, 426-791, Korea

Abstract

This study compares the forecasting accuracy of five econometric models on domestic total freight volume in South Korea. Applied five models are as follows: Ordinary Least Square model, Partial Adjustment model, Reduced Autoregressive Distributed Lag model, Vector Autoregressive model, Time Varying Parameter model. Estimating models and forecasting are carried out based on annual data of domestic freight volume and an index of industrial production during 1970~2011. 1-year, 3-year, and 5-year ahead forecasting performance of five models was compared using the recursive forecasting method. Additionally, two forecasting periods were set to compare forecasting accuracy according to the size of future volatility. As a result, the Time Varying Parameter model showed the best accuracy for forecasting periods having fluctuations, whereas the Vector Autoregressive model showed better performance for forecasting periods with gradual changes.

이 연구에서는 국내 총 화물물동량에 대한 5개 계량경제모형들의 예측정확도를 비교한다. 적용된 5개 모형은 통상최소사승모형, 부분조정모형, 축소된 자기회귀분포시차모형, 벡터자기회귀 모형, 시간변동계수모형이다. 모형의 추정과 예측은 1970-2011년 동안의 연간 국내 화물물동량 자료와 광공업생산지수를 이용하여 수행되었다. 5개 모형은 반복적인 예측방법을 이용하여 1년 후, 3년 후, 5년 후 예측성능이 비교되었다. 추가적으로 장래변동성의 크기에 따라 두 예측기간으로 나누어 예측정확도를 비교하였고, 결과적으로 시간변동계수모형은 변동을 갖는 예측기간에 대해서 가장 높은 정확도를, 반면에 벡터자기회귀 모형은 점진적인 변화를 갖는 예측기간에 대해서 다른 모형에 비해 우수한 성능을 보여주는 것으로 분석되었다.

Keywords

econometric model, forecasting accuracy comparison, freight volume, index of industrial production, recursive forecasting method

계량경제모형, 예측정확도 비교, 화물물동량, 광공업생산지수, 반복적 예측기법

* : Corresponding Author
kyungwoo@hanyang.ac.kr, Phone: +82-31-400-5153, Fax: +82-31-400-8147

Received 3 February 2014, Accepted 2 November 2014

서론

1. 연구의 배경 및 목적

장래 화물물동량, 수요의 예측은 경제활동과 화물교통 관계의 모형화를 통해 수행되고 있다. 화물물동량과 수요의 추정은 국가적, 지역간, 도시적 측면 세 범주로 나눌 수 있는데, 국가간 또는 국가적 관점에서 활용되는 모형으로는 거시적 계량경제모형, 산업연관모형 등이 있다. 지역간 모형에는 주로 4단계 수요추정모형이, 도시 화물모형에는 공급사슬망(supply chain) 특성을 보다 고려한 투어(tour)모형 등이 사용되고 있다. 수단별로는 항공과 해운의 경우, 장래화물수요 추정에 거시계량경제 모형이 주로 사용되며, 철도와 공로 수단의 화물수요추정에는 4단계 모형이 많이 활용된다.

각 범주별 장래 화물물동량과 수요의 추정결과는 국가적, 지역단위별 운영관리 및 장기계획 수립에 참고자료로 중요한 역할을 하기 때문에 이들의 예측정확성은 중요한 이슈가 되어 왔다. 본 연구에서는 국가적 관점에서 사회경제활동과 국내화물 물동량의 관계를 거시계량경제모형들을 통해 모형화하고, 이들 모형의 예측정확도를 비교하였다.

기존 문헌 및 이론적 모형 검토

1. 기존 문헌 고찰

개별 거시계량경제모형을 이용하여 여객 및 화물수요를 분석한 연구는 다음과 같다. Choi(2002)는 연간 시계열자료를 이용하여 국가단위의 화물발생 예측모형 개발을 시도하였다. 독립변수로 인구, 산업생산지수, 국내총생산, 화물자동차대수를 고려하였으나, 최종모형에서는 국내총생산(GDP)을 단일변수로 적용하였다.

Baek and Kim(2008)는 제주-내륙간 국내선 항공여객 수요모형추정에 계절변동을 고려한 단순시계열모형과 부분조정모형을 이용하고, 소득 및 운임에 대한 탄력도를 추정하였다.

Min et al.(2013)은 계절성 ARIMA 모형을 이용하여 인천공항발 유럽항공노선을 중심으로 항공화물 수요 예측모형을 추정하였다. 기존의 ARIMA모형과 계절성 ARIMA모형의 예측정확도를 비교한 결과, 유럽노선 화

물수송량에 대한 모형의 예측성능은 계절성 ARIMA모형이 더 높게 분석되었다.

West et al.(2011)은 미국의 화물수요 추정에 적합한 설명변수를 찾고, 평가하는 연구를 수행하였다. 다양한 변수들이 고려되었으나, 화물수요에 직접적인 영향을 주는 요인들로는 인구통계학적 요인, 경제활동 요인, 유류가격 요인들로 나타났다. 특정 변수로는 국내총생산(GDP), 산업생산지수, 유류가격이 설명력이 좋은 변수로 분석되었다.

Meersman and Van de Voorde(2013)은 국가별로 국내총생산(GDP)을 이용하여 화물물동량을 추정하였다. 분석결과, 경제활동과 화물교통사이의 관계변화로 장기적인 화물수요예측 시 국내총생산(GDP)에 기초한 분석은 신뢰성을 얻기 어려우며, 미시경제적 행태를 고려한 비집합적 접근이 필요함을 설명하였다.

거시계량경제모형들의 예측정확도를 비교한 연구로 Song et al.(2003)는 덴마크로 유입되는 국가별 국제관광수요를 예측하는데 계량경제모형을 사용하고, 이들 모형의 정확도를 비교하였다. 적용모형으로 통상최소자승모형, Wickens-Breuch오차수정모형, Johansen 오차수정모형, 축소된 자기회귀 분포시차 모형, 시간변동계수 모형, ARIMA모형, Naïve no-change모형을 사용하였다. 결과로, 1년 후 예측에는 시간변동계수모형, 2년 후 예측에는 통상최소자승, 시간변동계수모형, 3년 그리고 4년 후 예측에는 통상최소자승모형이 가장 정확하다고 분석되었다.

Song(2007)은 일본 및 미국 방한수요를 예측하는 단변량 시계열모형과 계량경제모형들을 표본내 예측력과 표본후 예측력으로 나누어 분석하고, 표본후 예측의 중요성을 설명하였다.

Shen et al.(2009)는 계량경제모형을 영국화물수요 예측에 품목별로 적용하여 정확도를 비교하였다. 적용된 6가지 계량경제학적 모형, 1)통상최소자승모형, 2)부분조정모형, 3)축소된 자기회귀 분포시차모형, 4)벡터자기회귀 모형, 5)시간변동계수 모형, 6)구조적 시계열 모형(structural time series model)을 이용하여 Song et al.(2003)의 연구와 같은 방법론을 통해 정확도를 비교하였다. 단 이 연구에서는 시계열자료의 안정성에 대한 고려는 연구범위에 포함되지 않았다. 모형들의 정확도 순위는 화물품목별로 다르게 나타났으나, 총 화물수요(total freight demand, ton-km)에 대한 모형의 예측성능 비교 시 1년, 3년 후는 구조적 시계열모형, 5

년 후에는 부분조정모형과 축소된 자기회귀 분포시차모형이 가장 정확한 것으로 나타났다.

2. 이론적 모형

거시경제학뿐 아니라 여러 분야에서 통상최소자승(Ordinary Least Square)모형, 자기회귀(Autoregressive)모형, 이동평균(Moving Average)모형, ARMA(Autoregressive-moving-average)모형, 오차수정(Error-correction)모형, 상태공간(State Space)모형 등 다양한 모형들이 사용되고 있다. 이 연구에서는 교통 분야에서 많이 사용되는 1)통상최소자승(OLS)모형, 2)부분조정(Partial Adjustment)모형과, 3)축소된 자기회귀분포시차(Reduced Autoregressive Distributed Lag)모형, 벡터(vector)형태인 4)벡터자기회귀(Vector Autoregressive)모형, 5) 상태공간(State Space)형태의 대표적인 모형인 시간변동계수(Time Varying Parameter)모형을 국내화물 물동량자료를 이용하여 모형간 예측정확도를 비교하였다.

모형에 대한 검증 및 진단과정으로 잔차의 정규성에 대한 Jarque-Bera 검정(Jarque and Bera, 1980), 자기상관에 대한 Breusch-Godfrey의 라그랑주 승수검정(Breusch, 1978 and Godfrey, 1978), 이분산성에 대한 White 검정(White, 1980)을 수행하였고, 모형 표기(specification)에 대해서는 Ramsey의 RESET 검정(Ramsey, 1969)을 사용하였다.

1) 통상최소자승모형(OLS)

$$y_t = \alpha + \sum_{i=1}^I \beta_i x_{it} + \varepsilon_t \quad (1)$$

통상최소자승(Ordinary Least Square, 이하 OLS)모형은 시계열 회귀분석에서 가장 흔히 쓰이는 계량경제 모형으로 일반적인 형태는 식(1)과 같으며, y_t 는 종속변수, x_{it} 는 설명변수, ε_t 는 평균이 0이고 일정한 분산을 갖는 정규적, 독립적으로 분포된 확률적 오차를 나타낸다. I 는 설명변수의 개수를 의미하며, α 와 β 는 모형을 통해 추정할 계수를 가리킨다.

이 모형은 시계열자료의 안정성(stationary)을 가정으로 한다. 계열 자료가 안정적이지 않을 경우 더빈왓슨(Durbin-Watson)통계치가 적합도(R^2)보다 더 작은

허구적 회귀결과(spurious regression)가 나타나, 추정계수의 비효율(inefficient), 유효하지 않은 유의도 검정결과를 야기한다(Granger and Newbold, 1974). 계열자료의 안정성을 확인하는 방법으로는 단위근(unit-root) 검정을 많이 사용하고 있으며, 계열자료가 불안정한 경우 간단하게는 차분된 자료를 사용함으로써 시계열자료의 안정성을 확보할 수 있다.

2) 부분조정 모형(PA)

$$y_t = \delta\alpha + \sum_{i=1}^I \delta\beta_i x_{it} + (1-\delta)y_{t-1} + \delta\varepsilon_t \quad (2)$$

부분조정(Partial Adjustment, 이하 PA)모형은 모형 내 설명변수로 종속변수의 1차 시차변수를 포함하는 자기회귀형태를 갖고 있다. y_t 는 종속변수, x_{it} 는 설명변수, ε_t 는 확률적 오차를 나타내며, I 는 설명변수의 개수를 의미한다. 조정계수(adjustment parameter)로 알려진 δ 는 $0 < \delta \leq 1$ 범위 내 값이며, 이 값이 1에 가까울수록 조정속도는 빨라진다. 만약 δ 이 0이라면 $y_t = y_{t-1}$ 이므로 변화가 없다는 것을 의미하고, δ 이 1이라면 x_{it} 의 변화에 따른 y_t 의 변화가 t기에 모두 이루어졌다는 것을 의미한다.

3) 축소된 자기회귀분포시차모형(ReADLM)

$$y_t = \alpha + \sum_{i=1}^I \sum_{j=0}^J \beta_{ij} x_{i,t-j} + \sum_{j=1}^J \phi_j y_{t-j} + \varepsilon_t \quad (3)$$

식(3)은 일반적인 자기회귀 분포시차 모형(General Autoregressive Distributed Lag Model)의 형태이며, 분포시차를 가진다는 것은 설명변수에 따른 종속변수의 변화가 바로 일어나지 않고 시간에 걸쳐 발생함을 의미한다. α, β, ϕ 는 추정할 계수이며, 시차 길이 J 는 실험적으로 결정되나, 분석 시 연간자료를 사용하는 경우 $J=1$ 로 널리 사용하고 있다(Thomas, 1997). 본 연구에서도 시차길이를 1로 사용하였는데, 이는 다른 시차길이 이를 적용하여 모형을 추정했으나 축소과정에서 변수들이 모두 제외되었기 때문이다.

모형의 축소과정은 회귀결과로부터 가장 유의하지 않은 설명변수 또는 가변수를 제거 후, 모형을 재추정하고 또 다시 가장 유의하지 않은 변수를 제거하는 과정을 반복한다. 이 과정은 남은 변수들의 유의수준이 5%이내

며, 옳은 부호를 가질 때까지 계속 수행된다(Song et al., 2003).

4) 벡터자기회귀 모형(VAR)

$$\begin{cases} y_t = \alpha_{00} + \sum_{j=1}^J \beta_{0j} y_{t-j} + \sum_{i=1}^I \sum_{j=1}^J \delta_{0ij} x_{i,t-j} + \varepsilon_{0t} \\ x_{1t} = \alpha_{10} + \sum_{j=1}^J \beta_{1j} y_{t-j} + \sum_{i=1}^I \sum_{j=1}^J \delta_{1ij} x_{i,t-j} + \varepsilon_{1t} \\ \vdots \\ x_{It} = \alpha_{I0} + \sum_{j=1}^J \beta_{Ij} y_{t-j} + \sum_{i=1}^I \sum_{j=1}^J \delta_{Iij} x_{i,t-j} + \varepsilon_{It} \end{cases} \quad (4)$$

벡터자기회귀(Vector Autoregressive, 이하 VAR) 모형은 구조적 형태(structural form)와 식(4)에 나타난 축소된 형태(reduced form)로 표현될 수 있다. 현상 설명에는 구조적 형태의 모형이 용이하나 추정이 어려워, 축소형태의 추정 이후 제약을 통한 변형하여 사용하고 있다. 본 연구에서는 예측에 초점을 맞추고 있어, 축소된 형태의 벡터자기회귀모형을 사용하였다. 위 식에서 ε_{it} 는 독립적, 정규분포를 갖는 확률오차항이나, 축소된 형태에서는 이들 오차항간의 공분산은 0이 아니다. 벡터자기회귀모형에서는 모든 변수를 내생변수로 다루어지며, 따라서 다른 모형과 달리 장래예측 시 설명변수의 장래 값을 요구하지 않는다.

벡터자기회귀모형에서 시차길이(J)의 결정은 시차길이 조정에 따라 추정된 모형의 Akaike Information Criterion (AIC), Schwarz Information Criterion (SC), Likelihood Ratio (LR) 비교를 통해 수행되며, AIC나 SC값이 가장 작거나 LR이 가장 클 때의 시차길이를 선택한다.

5) 시간변동계수 모형(TVP)

$$y_t = \beta_{0t} + \sum_{i=1}^I \beta_{it} x_{it} + \varepsilon_t \quad (5)$$

$$\beta_{it} = \beta_{it-1} + \mu_t \quad i=0,1,\dots,I \quad (6)$$

시간변동계수(Time Varying Parameter, 이하 TVP) 모형은 상수항과 설명변수의 계수들이 시간에 따라 변화하며, 모형 내에서 이들은 잠재(unobserved, latent)변수로 다루어진다. 모형은 상태공간(State Space)의 구조적 형태로 나타나며, 식(5)는 신호방정식(signal equation), 식(6)은 상태방정식(state equation)이라

고 불린다. 상태공간모형에서 계수들은 Kalman filter 알고리즘에 의해 추정되는데, 이 알고리즘은 잠재변수의 최적 추정치를 계산하는데 예측과 업데이트 두 단계를 반복적으로 수행하는 기법이다.

식(5)와 식(6)에서 I는 설명변수의 개수를 의미하고 ε_t 와 μ_t 는 평균이 0이고 일정한 분산을 가지며 정규적, 독립적으로 분포된 확률적 오차를 나타낸다.

분석 자료 및 변수설정

화물물동량과 연관된 사회경제지표들은 여러 가지가 있는데, 많은 연구에서 화물물동량의 설명변수로 국내총생산(Gross Domestic Production, 이하 GDP)과 광공업생산지수(Index of Industrial Production, 이하 IIP)를 많이 적용하고 있다. 화물수요에 영향을 미치는 사회경제요인들을 조사한 West et al.(2011)의 연구에서도 광공업 생산지수, 국내총생산, 연료가격은 화물물동량의 좋은 설명변수로 분석되었다.

GDP는 다른 분야별 사회경제지표들뿐 아니라 광공업생산지수에 기초한 가공통계기기에 GDP와 광공업생산지수(IIP)는 높은 상관관계를 갖고 있다. 본 연구에서는 경제생산성의 대리(proxy)변수로 광공업생산지수를 유일한 설명변수로 선택하였다. 연료가격요소로는 특정 변수의 추가 없이, 오일쇼크를 반영하기 위한 더미변수를 설정하였다. 그 외 다른 변수들은 VAR모형 추정 시 자기회귀변수가 지나치게 많아질 수 있어, 분석편의를 위해 고려되지 않았다.

분석에는 1970-2011년의 연도별 국내화물 물동량의 톤단위 자료와 광공업생산지수 자료를 이용하였다. 연도별 국내화물물동량 자료는 국가교통데이터베이스에서 수집하였다(KTDB, 2014). 통계청에서 얻은 광공업 생산지수는 1970-2011년의 시계열 자료가 연속적이지 않아, 조정과정을 통한 2005년 기준으로 시계열을 연결하였다(KOSTAT, 2014; KOSIS 2014).

Figure 1은 국내화물물동량의 연도별 추이를 나타낸 것으로, 그림에 나타난 지수형 추세선(점선)은 1990년 중반까지의 추세와 이후 추세가 다름을 보여주고 있다. 국내 총 화물물동량은 1997년까지는 점차 증가하다가 경제활동의 축소와 함께 다소 변동을 보여주었다.

수단별로는 2011년 기준 전체 국내화물 물동량의 약 81%가 공로(화물트럭)를 이용하였으며, 이는 1970년 전체대비 59%에서 지속적으로 증가한 것으로 나타났다.

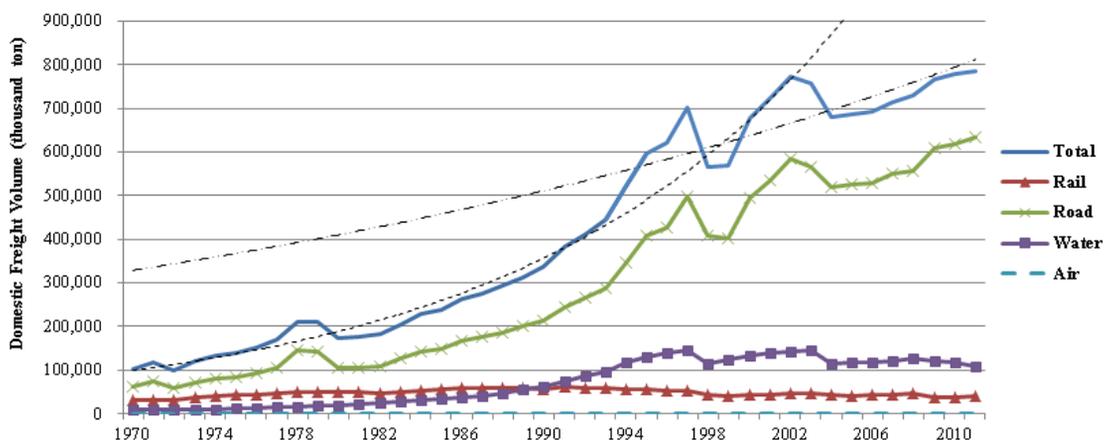


Figure 1. 1970-2011 Domestic freight volume in South Korea

해운과 철도 수송은 전체 물동량 대비 비중이 점차 낮아졌다.

모형 내 종속변수와 설명변수의 관계로는 계수가 탄력도로 표현되는 로그-로그 형태를 사용하였고, 수식의 일반 형태는 식(7)과 같다.

$$\ln TON_t = f(\ln IIP_t, Dummy\ variables) \quad (7)$$

$\ln TON_t$ 는 t년도의 총 국내화물 물동량의 자연로그 값을 가리키며, IIP는 광공업생산지수를 가리킨다. 추정된 모형에는 광공업생산지수 외 일회성 사회경제현상(one-off event)을 설명하기 위하여 더미변수를 모형내에 포함하였다. 총 국내화물 물동량에 영향을 주었을 일회성 사회경제현상으로 유류가격과 관련 있는 1차(1974년) 및 2차 오일쇼크(1980-81년), 국가경제에 큰 영향을 준 아시아 금융위기(1998-99년), 공로의 높은 비중을 고려한 2003년 및 2008년 화물연대 파업을 고려하였다.

모형추정 및 예측성능 비교

1. 모형의 추정

모형추정과 예측에는 통계프로그램인 EViews 7.1을 사용하였다. 본 연구에서는 모형 추정 시, 시계열자료에 대한 안정성 여부를 확인하기 위하여 국내 총 화물물동량과 광공업생산지수 자료에 대해 ADF(Augmented Dicky-Fuller)단위근 검정과 공적분(cointegration)

검정을 먼저 수행하였다. 검정결과 두 자료는 1차 차분 후에 안정화 되는 것으로 나타났고, 공적분 관계에 있지 않았다. 따라서 시계열의 불안정성으로 인한 허귀적 회귀를 피하기 위해 OLS모형과 VAR모형은 1차 차분된 시계열 자료를 사용하였다. 차분된 자료의 사용은 Δ (delta)기호로 표시하였다. 시간변동계수모형(TVP)의 경우, 초기에 상수항과 설명변수의 계수 모두를 시간변동계수로 설정하였으나, 추정 시 오차의 분산이 유일하지 않은 관계로 상수항은 고정계수로 설정하고 설명변수의 계수만을 시간변동계수로 적용하였다.

1970-2011년 기간의 자료를 이용하여 총 국내화물 물동량과 광공업생산지수간의 관계를 추정한 모형별 초기추정결과는 Table 1에 나타나있다. 모형별 적합도는 OLS모형과 VAR모형이 ReADLM과 PA모형에 비해 낮게 나타났는데, 이는 앞서 설명한 것처럼 차분된 자료를 사용하지 않았을 시 적합도는 높게 나타날 수 있지만, 허귀적 회귀결과가 얻을 수 있으므로 본 연구에서 차분된 자료를 사용했기 때문이다.

설명변수로 사용된 광공업생산지수의 경우, VAR모형을 제외하고는, 모두 추정된 계수가 유의수준 5% 이내에서 유의하게 나타났다. 이를 바탕으로 광공업생산지수에 국내총화물물동량의 탄력도는 모형별 다소 차이가 있으나, 0.493(OLS모형), 0.512(PA모형), 0.598(ReADLM), 0.456(TVP모형)로 분석되었다.

초기모형의 잔차 검정 결과, 유의수준 5%내에서 귀무가설(잔차에 대한 가정이 만족)을 기각하는 것으로 나타난 OLS모형이 정규성 검정(NORM)을 통과하지 못했다. Ramsey의 모형표기검정(RESET)은 유의수준

Table 1. Initial estimation results of five models (1970-2011)

Model	OLS	PA	ReADLM	VAR	TVP	
Dependent Independent	$\Delta \ln \text{TON}_t$	$\ln \text{TON}_t$	$\ln \text{TON}_t$	$\Delta \ln \text{TON}_t$	$\Delta \ln \text{IIP}_t$	$\ln \text{TON}_t$
$\ln \text{TON}_{t-1}$		0.799*** (0.086)	0.863*** (0.080)			
$\Delta \ln \text{TON}_{t-1}$				-0.251 (0.165)	-0.553*** (0.147)	
$\Delta \ln \text{TON}_{t-2}$				-0.023 (0.182)	0.165 (0.162)	
$\ln \text{IIP}_t$		0.103** (0.049)	0.548*** (0.185)			0.456*** (0.019)
$\Delta \ln \text{IIP}_t$	0.493*** (0.163)					
$\ln \text{IIP}_{t-1}$			-0.466** (0.188)			
$\Delta \ln \text{IIP}_{t-1}$				0.066 (0.212)	0.365* (0.190)	
$\Delta \ln \text{IIP}_{t-2}$				0.087 (0.218)	0.136 (0.195)	
Constant	0.013 (0.022)	2.281** (0.937)	1.482 (0.888)	0.073** (0.027)	0.077*** (0.024)	11.300*** (0.107)
DUM74	-0.032 (0.078)	0.023 (0.081)		0.040 (0.090)	0.173** (0.081)	-0.002 (0.269)
DUM80	-0.134** (0.054)	-0.179*** (0.057)	-0.134** (0.055)	-0.207*** (0.062)	-0.137** (0.055)	-0.125*** (0.043)
DUM98	-0.156*** (0.054)	-0.117* (0.061)	-0.125** (0.054)	-0.193*** (0.060)	-0.045 (0.053)	-0.130 (0.122)
DUM03	-0.058 (0.075)	-0.030 (0.082)		-0.078 (0.079)	-0.025 (0.070)	0.063 (0.067)
DUM08	-0.009 (0.076)	-0.042 (0.080)		-0.055 (0.076)	-0.061 (0.068)	-0.017 (0.266)
Adj. R-square	0.369	0.987	0.989	0.276	0.284	
S.E.	0.073	0.077	0.070	0.074	0.066	
LMSC	0.089	5.381	3.343		8.332*	
HETRO	1.613	2.998	3.262		27.238	
NORM	30.334***	0.273	2.736		2.425	
RESET	1.207	0.194	3.197*			

- $\ln \text{TON}_t$ means the natural logarithm of total freight volume in year t, and $\Delta \ln \text{TON}_t = \ln \text{TON}_t - \ln \text{TON}_{t-1}$

- Values in parentheses are standard error.

- *** Significant at 1% level; ** Significant at 5% level; * Significant at 10% level.

- LMSC: Lagrange multiplier test for Serial correlation, HETRO: White's heteroscedasticity test, NORM: Jarque-Bera normality test, RESET : Ramsey's misspecification test

5%이내에서는 모두 통과한 것으로 나타났다.

고려된 일회성 사회경제현상 중 1차 오일쇼크, 2003년과 2008년의 화물연대 파업에 대한 더미변수의 계수는 모든 모형에서 유의하지 않은 것으로 분석되었다. 1차 오일쇼크의 경우, 국내경제에 대한 영향이 크지 않았기 때문에 여겨진다. 화물연대 파업의 경우 단기간 발생한 사건이라 연도별 시계열 자료에는 영향이 미비하며,

파업의 충격이 다른 수단으로 흡수되어 총 물동량 자료에 영향이 나타나지 않았을 수 있기 때문으로 생각된다. 따라서 유의하지 않은 세 가지 더미변수를 제외하고, 나머지 2개 더미변수, 2차 오일쇼크 반영을 위한 DUM80(1980~81=1, 그 외 0),과 아시아 금융위기를 반영하기 위한 DUM98(1998~99=1, 그 외 0)를 최종모형추정에 사용하여 예측성능 비교를 진행하였다.

2. 모형간 예측성능 비교

본 연구에서는 장래변동성에 따른 모형간 예측성능 차이를 알아보기 위해, Figure 1을 바탕으로 점진적 추세를 갖는 기간(1986-1992년)과 추세를 중심으로 변동이 있는 기간(2002-2011년)을 예측기간으로 설정하였다. 또한, 각 설정된 기간에서는 1년 후, 3년 후, 5년 후에 대한 모형들의 예측정확도가 비교되었다. 각 기간별 모형의 추정에는 광공업생산지수와 더미변수인 DUM80, DUM98을 이용하였다.

VAR모형에서는 종속변수인 총 화물 물동량과 설명변수인 광공업생산지수 모두를 내생변수로 다루기 때문에 다른 모형과는 달리, 장래 화물물동량의 추정을 위해 장래 광공업생산지수 값을 요구하지 않는 특성이 있다. 즉, 예측정확도는 설정된 예측기간 내 화물물동량 추정치와 실측치의 비교로 분석되는데, OLS, PA, ReADLM, TVP모형은 예측기간 내 화물물동량 추정을 위해 설명변수인 예측기간 내 광공업생산지수 값을 요구하지만, VAR모형은 필요로 하지 않는다. 이러한 VAR모형의 특수성을 남겨둔 채 국내 총 화물물동량에 대한 모형간 예측정확도 비교를 수행하였다.

각 기간 별 각 모형의 예측치 추정에는 반복적 예측기법(recursive forecasting method)가 사용되었다. 예를 들어, 2002-2011년 기간의 예측성능을 비교 시 반복적 예측기법의 절차는 다음과 같다.

먼저, 1970-2001년 기간의 자료를 이용해 모형을 추정하고 2002-2011년을 화물물동량을 예측하였다. 그 후, 1970-2002년 자료를 통해 모형을 다시 추정하고 2003-2011년 기간의 화물물동량을 예측하였다. 반복적 과정은 1970-2010년 자료를 이용하여 2011년 화물물동량을 추정하기까지 총 10번이 수행되었다. 이를 통해, 반복적 예측기법의 결과로 2002-2011년 기간에서는

10개의 1년 후 예측치, 8개의 3년 후 예측치, 6개의 5년 후 예측치를 얻었고, 마찬가지로 방법으로 1986-92년 기간에서는 7개의 1년 후 예측치, 5개의 3년 후 예측치, 3개의 5년 후 예측치를 얻었다.

예측치와 실측치의 오차계산을 통한 모형들의 예측성능비교방법으로는 절대오차평균비율법(Mean Absolute Percentage Error, MAPE)과 루트오차제곱평균 비율법(Root Mean Square Percentage Error, RMSPE)를 사용하였다.

$$MAPE = \frac{\sum_{t=1}^n |y_{est} - y_{obs}| / y_{obs}}{n} \times 100 \quad (8)$$

$$RMSPE = \sqrt{\frac{\sum_{t=1}^n ((y_{est} - y_{obs}) / y_{obs})^2}{n}} \times 100 \quad (9)$$

Table 2와 Table 3은 1986-1992년, 2002-2011년 두 예측기간별 5개 모형의 예측성능을 측정한 결과이다. 이들 결과는 사회적 통념과 같이 단기 예측이 장기 예측보다 더 정확함을 보여주고 있다.

점진적으로 증가하는 추세를 가지는 1986-1992년 기간의 예측 시, VAR모형이 1년 후, 3년 후, 5년 후 예측에서 5개 모형 중 가장 우수한 예측성능을 보여주었다. OLS모형의 경우 1년 후 예측에서, PA모형은 3년 후 예측, ReADLM은 5년 후 예측에서 두 번째로 우수한 예측정확도를 보여주었다. 반면에 TVP모형은 5개 모형 중 예측정확도가 가장 낮게 나타났다.

변동성이 존재하는 2002-2011년 기간의 예측 시, 5개 모형 중 1년 후, 3년 후, 5년 후 모두 TVP모형이 가장 정확한 것으로 분석되었다. TVP모형에 이어서 OLS 모형의 예측성능이 우수했다. 시차변수를 상대적으로 많

Table 2. Forecasting performance of the models over the period 1986-1992

Horizon	Measure	OLS	PA	ReADLM	VAR	TVP
1-year ahead	MAPE	4.447%(2)	5.399%(3)	5.508%(4)	3.634%(1)	7.337%(5)
	RMSPE	5.204%(2)	6.586%(4)	6.349%(3)	4.395%(1)	8.122%(5)
3-year ahead	MAPE	11.849%(4)	10.103%(2)	11.170%(3)	4.447%(1)	13.402%(5)
	RMSPE	13.723%(4)	12.029%(2)	12.535%(3)	5.424%(1)	14.770%(5)
5-year ahead	MAPE	18.453%(4)	15.556%(3)	13.159%(2)	4.120%(1)	19.151%(5)
	RMSPE	19.868%(4)	16.130%(3)	14.821%(2)	4.882%(1)	19.562%(5)

Figures in parenthesis are rankings.

Table 3. Forecasting performance of the models over the period 2002-2011

Horizon	Measure	OLS	PA	ReADLM	VAR	TVP
1-year ahead	MAPE	5.319%(2)	6.577%(3)	8.286%(4)	9.726%(5)	4.880%(1)
	RMSPE	7.586%(2)	8.217%(3)	10.302%(4)	12.078%(5)	6.803%(1)
3-year ahead	MAPE	15.927%(2)	20.484%(3)	26.669%(4)	28.633%(5)	10.850%(1)
	RMSPE	20.456%(2)	22.643%(3)	31.303%(4)	31.763%(5)	14.399%(1)
5-year ahead	MAPE	26.216%(2)	31.318%(3)	46.327%(4)	50.278%(5)	14.551%(1)
	RMSPE	31.171%(2)	33.306%(3)	52.031%(4)	52.883%(5)	18.618%(1)

Figures in parenthesis are rankings.

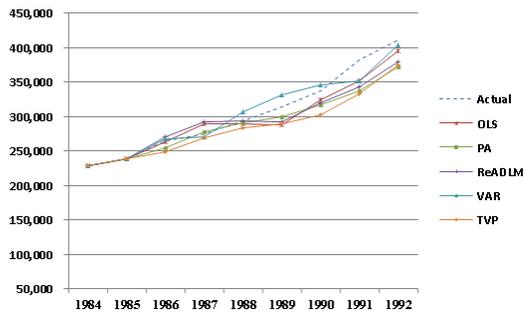


Figure 2. 1-year ahead forecasts vs. actual (1986-1992)

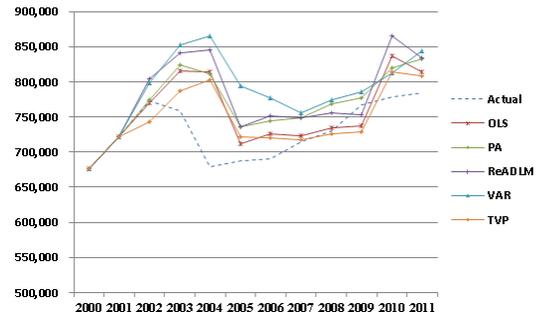


Figure 3. 1-year ahead forecasts vs. actual (2002-2011)

이 포함한 ReADLM과 VAR모형은 변동이 존재하는 기간에서 상대적으로 낮은 예측성능을 보여주었다.

전체적으로 점진적 추세에 대한 예측에는 자기회귀변수를 포함한 VAR, ReADLM, PA모형이 상대적 우수함을, 변동에 존재하는 기간에 대한 예측에는 교통류 분석에도 많이 활용되는 TVP모형이 우수하게 나타났다. 많이 사용되는 OLS모형 역시 두 기간 모두에서 중간 정도의 예측성능을 보여주었다. Figure 2와 Figure 3은 각 모형별 1년 후 예측치와 실측치를 비교한 그림이다.

결론

본 연구에서는 국내 총 화물물동량에 대해 5가지 계량경제모형을 이용하여 예측정확도를 비교하였다. 적용된 5가지 모형은 통상최소자승모형(OLS), 부분조정모형(PA), 축소된 자기회귀분포시차모형(ReADLM), 벡터자기회귀모형(VAR), 시간변동계수모형(TVP)이다. 모형의 추정과 예측에는 1970-2011년 동안의 연간 국내 총화물물동량, 광공업생산지수 자료를 사용하였다. 모형 추정 결과, 모형별 다소 차이가 있으나, 광공업생산지수에 대한 국내총화물물동량의 탄력도는 0.456-0.598 사이로 분석되었다. 광공업생산지수 외에는 일회성 사회경

제현상인 2차오일쇼크(1980-81)와 아시아금융위기(1998-99)를 고려한 더미변수가 모형추정 시 포함되었다. 예측성능 비교에는 화물물동량의 변동성에 따라 1986-92년, 2002-11년 두 기간으로 나누어 수행되었다.

5가지 모형의 예측성능 비교결과로, 예측기간이 큰 변동성 없이, 점진적 추세를 갖는 1986-92년 기간에는 벡터자기회귀모형(VAR), 부분조정모형(PA), 축소된 자기회귀분포시차모형(ReADLM)의 성능이 우수하게 나타났다, 예측기간 내 변동성이 존재하는 2002-11년 기간의 경우, 시간변동계수모형(TVP)이 상대적으로 가장 우수한 예측정확도를 보여주었다. 통상최소자승(OLS) 모형은 두 기간 모두에서 중간 정도의 결과를 보여주었다. 두 예측기간에서 모형들의 예측성능이 상반된 결과를 보이는 이유는 시차변수들이 가지는 관성력에 의한 것으로 여겨진다.

본 연구의 결과는 앞선 연구들의 결과와 다소 일치하지 않는 부분도 있었으나, 연구결과를 통해 변동성이 큰 계열자료에는 시간변동계수모형(TVP)이, 점진적 추세를 갖는 계열자료에는 벡터자기회귀모형(VAR)을 사용하는 것이 적절할 것임을 보여주었다. 이러한 정보는 교통 분야뿐 아니라 계량경제모형을 활용하는 다른 분야에서도 모형 선택 시 참고자료로 쓰일 수 있을 것이다.

향후 연구과제로, 본 연구에서는 연구편의를 위해 설 명변수로 광공업생산지수만을 사용하였으나, 추가적인 설명변수를 고려한 연구가 필요할 것으로 생각된다. 또한, 점점 비집합화(disaggregate)된 자료를 통한 분석이 시 도되고 요구되고 있는 만큼, 수단별, 화물품목별 연구가 진행된다면 활용가치가 더욱 높을 것으로 생각된다.

REFERENCES

- Baek S. H., Kim S. S. (2008), Estimation of Air Travel Demand Models and Elasticities for Jeju-mainland Domestic Routes, *J. Korean Soc. Transp.*, 26(1), Korean Society of Transportation, 51-63.
- Breusch T. (1978), Testing for Autocorrelation in Dynamic Linear Models, *Australian Economic Papers*, 17(31), 334-355.
- Choi C. H. (2002), Development of Commodity Generation Models Using GDP, *KSCE Journal of Civil Engineering*, 22(3D), Korean Society of Civil Engineers, 351-362.
- Dargay J. M., Hanly M. (2002), The Demand for Local Bus Services in England, *Journal of Transport Economics and Policy*, 36(1), 73-91.
- Godfrey L. G. (1978), Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Contain Lagged Dependent Variables, *Econometrica: Journal of the Econometric Society*, 46(6), 1303-1310.
- Granger C. W. J., Newbold P. (1974), Spurious Regressions in Econometrics, *Journal of Econometrics*, 2(2), 111-120.
- Jarque C. M., Bera A. K. (1980), Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, *Economics Letters*, 6(3), 255-259.
- Korea Transport DataBase (KTDB) (2014), Domestic/International Freight Transport Records, Retrieved From <http://www.ktdb.go.kr/en/web/guest/291>.
- KOrean Statistical Information Service (KOSIS) (2014), By Metropolitan Area/By Industry in IIP, Retrieved From <http://kosis.kr/eng/>.
- Meersman H., Van de Voorde E. (2013), The Relationship Between Economic Activity and Freight Transport, *Freight Transport Modelling*, Bingley, Emerald, 17-43.
- Min K. C., Jun Y. I., Ha H. K. (2013), Forecasting the Air Cargo Demand With Seasonal ARIMA Model: Focusing on ICN to EU Route, *J. Korean Soc. Transp.*, 31(3), Korean Society of Transportation, 3-18.
- Ramsey J. B. (1969), Test for Specification Errors in Classical Linear Least Squares Regression Analysis, *Journal of the Royal Statistical Society, Series B (Methodological)*, 31(2), 350-371.
- Shen S., Fowkes T., Whiteing T., Johnson D. (2009), Econometric Modeling and Forecasting of Freight Transport Demand in Great Britain, *Proceedings of European Transport Conference*.
- Song H., Witt S. F., Jensen T. C. (2003), Tourism Forecasting: Accuracy of Alternative Econometric Models, *International Journal of Forecasting*, 19(1), 123-141.
- Song K. S. (2007), A Comparison of Accuracy Between Univariate and Econometric Models, *Journal of Tourism Sciences*, 31(4), The Tourism Sciences Society of Korea, 33-50.
- Statistics Korea (KOSTAT) (2014), Korea Statistical Yearbook, Retrieved From <http://kostat.go.kr/portal/korea/index.action>.
- Thomas R. L. (1997), *Modern Econometrics: An Introduction*, Harlow, Addison-Wesley.
- West R., Rubin D., Villa J. C. (2011), Identification and Evaluation of Freight Demand Factors (No.NCFRP Project 11).
- White H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica: Journal of the Econometric Society*, 48(4), 817-838.
- 알림 : 본 논문은 동아시아교통학회 2013 학술발표회 (2013.09.10)에서 발표된 내용을 수정·보완하여 작성된 것입니다.
- ✉ 주 작성자 : 정성환
 ✉ 교신저자 : 강경우
 ✉ 논문투고일 : 2014. 2. 3
 ✉ 논문심사일 : 2014. 3. 25 (1차)
 2014. 11. 2 (2차)
 ✉ 심사판정일 : 2014. 11. 2
 ✉ 반론접수기한 : 2015. 6. 30
 ✉ 3인 익명 심사필
 ✉ 1인 abstract 교정필