

## 서울시 재정지출 예측모형에 관한 연구

The Econometric Forecasting Models  
for Seoul Metropolitan Government Expenditure

김 지 욱\*

## 목 차

- |          |            |
|----------|------------|
| . 서론     | . 계량예측모형설정 |
| . 자료특성분석 | . 결론       |

## ABSTRACT

Ji-Uk Kim

In this paper, we performed to develop the econometric forecasting models for government expenditure in Seoul city. An autoregressive integrated moving-average (ARIMA) process allows for integrated component in the Seoul time series after Unit-Root test. We found ARIMA(1,1,1) process as a best forecasting model for a single variable (Seoul metropolitan government expenditure). In dynamic models of a group of time series, the unrestricted vector autoregressions (UVAR's) cause overparameterization of the forecasting models. Therefore, we used a Bayesian procedure (BVAR) adjusting the general prior for estimating the VAR which largely improves forecasting performance. The responses of the two variables, the local tax and income per household, to shock in Seoul government expenditure did not come out in the near future.

## . 서론

지방재정규모가 증대되면서 재정을 계획적이고도 효율적인 운용을 위하여 각 지방자치단체는 중장기 지방재정계획을 수립하여 집행하고 있다.

지방재정의 활동은 안정적인 재원의 확보와 조달뿐만 아니라 확보된 재원을 얼마나 효율적으로 합리적으로 운영하는가에 달려 있다. 서울시 재정지출의 규모와 내용은 경제적 요인뿐만 아니라 정치·사회적 요인, 지리적·인구관련 요인 등에

\* 중앙대 정경대학 경제학과 조교수

의하여 영향을 받는다. 전국적인 국내총생산(GDP) 뿐만 아니라 서울지역내총생산, 지방세수입, 국민소득수준 향상에 따른 재정수요증대 요인 등 지역주민의 재정수요에 대한 다양한 욕구분출의 표현으로 나타난다. 기존의 연구에서는 중앙과 지방의 세원배분문제, 지방세 개편, 세입예측모형 등 재정확충 및 조달방안에 대한 많은 연구가 이루어졌으나 지방세출 측면에서는 연구가 부족하였다고 생각된다.

본 연구는 서울시의 재정지출 예측을 위한 최적모형을 개발하고자 하였다. 이러한 작업은 향후의 서울시 세수입 예측모형 개발과 더불어 세입-세출 예측에서 나타나는 서울시 재정활동의 과부족분(過不足分)을 어떻게 활용하며, 어떻게 조달하는 것이 바람직한 것인가에 대한 결과수준까지를 고려하는 사전 연구단계라고 할 수 있다.

재정지출 예측모형으로는 재정지출 자체의 움직임으로 과거의 관측치만을 이용하는 자기회귀이동평균(ARIMA)모형과 시계열변수 상호작용을 파악하여 미래의 방향을 예측하기 위해 선형적주관을 이용하는 베이지언 벡터자기회귀모형(BVAR) 등을 이용하였다.<sup>1)</sup>

재정지출과 경제변수들간의 관계를 설정하는데 있어서는, 먼저 지역경제의 활성화로 지역내총생산이 확대되고 이에 따른 요소소득의 재배분으로 가구당 소득이 증대되고 따라서 서울시의 세

수입이 증가한다. 세수입의 증가는 다시 재정지출의 증대로 이어지고 재정지출 확대는 지역경제 활성화로 이어진다고 보았다. 본 연구에서는 이러한 경제순환의 흐름을 전제로 지역내총생산, 가구당소득, 조세수입, 그리고 총재정지출 등을 주요 변수로 설정하였다.<sup>2)</sup>

## · 자료특성분석

### 1. 단위근 검정

먼저 시계열자료들을 계량분석에 이용하기 위해서는 시계열자료들이 안정적인 과정(stationary process)을 따르고 있는지를 확인하여야 한다. 왜냐하면 불안정적인 과정을 따르는 자료들을 회귀분석에 이용하면 변수간 상관관계가 없으면서도 높은 상관관계가 있는 것처럼 나타나는 허구적 회귀(spurious regression) 현상의 문제가 발생하기 때문이다. 따라서 변수들의 안정적 과정여부를 판단하는 단위근검정을 실시하기에 앞서 먼저 그림을 통한 대략적인 판단을 시도하고자 한다.

1970년부터 1998년까지의 서울시 1인당 총재정지출(PRTAC), 가구당소득(RINCOM), 지방세수입(RREVTX), 지역내총생산(RGRDP)의 수준변수자료를 이용하여, 모두 대수변환(log linear)시켜 서울시 소비자물가지수로 나눈 불변자료를 사용하

1) 변수간 공적분관계가 존재할 때에 오차벡터수정모형을 통한 예측모형이 제일 예측력이 뛰어난 것으로 알려져 있음. 예측모형 중 ARIMA모형의 제시는 서울시 재정지출 time series 그 자체의 움직임을 통하여 예측모형을 구축하는 것도 의미가 있다고 판단되었으며, 또한 일변량자기회귀모형, 일변량베이지언VAR모형, 단순베이지언VAR모형, 그리고 일반적 사전확률에 의한 베이지언VAR모형 등에 대한 예측력 비교에 연구목적이 있었음. ECM모형은 예측모형II 논문에서 다모형과 비교제시하고자 함.

2) 사용된 모든 자료는 인구수로 나눈 1인당 시계열자료이며 소비자물가지수로 나눈 불변가격기준으로 사용하였음. 또한 재정지출예측시 GDP나 GNP자료도 중요한 자료로 판단되나 모형설정이 서울시 지역성장 가구당소득 지방세수입 재정지출의 구조로 되어 제외되었음. 향후 다모형설정시 연구자료로 사용가능함. 가구당소득은 서울시통계연보에 제시되어 있는 income per household 자료임.

였다. 시간이 지남에 따라 상향증가 추세에 있고 변수 모두 불안정적인 과정을 보이고 있는데 1차 차분만으로도 안정적 과정으로 변환될 수 있음을 보여주었다.

<표 1>은 1인당 총재정지출(PRTAC)과 1인당 일반회계지출(pgac)의 자연대수로 변환된 자료들의 수준변수와 1차 차분변수에 대한 단위근 존재 여부를 ADF검정 및 Phillips-Perron 방식의 검정 결과를 보이고 있다. 두 검정 모두 t-test를 포함하고 trend항은 제외되었으며 시차(lags)가 0, 1 및 3인 경우로 나누어 실시하였다.<sup>3)</sup>

결산기준으로 일반회계 및 특별회계를 합한 총 지출액과 일반회계 수준변수의 단위근 존재에 대한 귀무가설을 5% 유의수준에서 기각하지 못하여 단위근을 가지고 있는 것으로 판명되었다. 1차 차분변수에서는 안정적 시계열로 밝혀져 모든 자료들이 한개의 단위근을 가지고 있음을 나타내고 있어 1차 차분만으로 시계열자료의 안정성을 확보할 수 있다.

<표 1> 단위근 검정결과

		1인당 총지출		1인당 일반회계지출	
		PRTAC	dPRTAC	pgac	dpgac
ADF (t-test)	lags=0	1.52	-20.00*	2.52	-15.57*
	lags=1	1.33	-14.57*	1.46	-817.82*
	lags=3	1.07	-10.69*	2.72	-19.35*
P-P (t-test)	lags=0	1.52	-20.00*	2.52	-15.57*
	lags=1	1.51	-19.35*	2.47	-17.38*
	lags=3	1.51	-20.77*	2.73	-15.16*

- 주 : 1) ADF 및 P-P검정 모두 t-test 포함하고 trend항은 제외됨.  
 2) 변수앞의 d표시는 1차 차분변수를 표시함  
 3) 5%의 유의수준에서의 임계치는 n=25일 때 -3.00임.  
 4) \*표시는 5%의 유의수준에서 단위근존재에 대한 귀무가설이 기각됨.  
 5) PRTAC는 1인당 총재정지출액, pgac는 1인당 일반회계지출액.

<표 2>는 가구당소득(RINCOM), 지방세수입(RREVTX), 지역내총생산(RGRDP)의 자료의 단위근 검정 결과를 보이고 있다. 단위근 검정결과 모두 단위근이 존재하여 시계열이 불안정적임을 알 수 있다. 변수들의 1차 차분변수들에 대한 단위근 검정결과 모두 1% 유의수준에서도 단위근이 존재하지 않는 것으로 판명되었다.

<표 2> 변수별 단위근 검정결과

	가구당소득 (RINCOM)	지방세수입 (RREVTX)	지역내총생산 (RGRDP)
lags=0	0.255	-0.524	-0.219
lags=1	-2.192	-1.203	-2.169
lags=3	-1.965	-1.716	-1.657
	dRINCOM	dRREVTX	dRGRDP
lags=0	-14.341*	-19.119*	-15.279*
lags=1	-14.811*	-22.684*	-27.284*
lags=3	-15.519*	-52.097*	-25.775*

- 주 : 1) ADF 모두 t-test 포함하고 trend항은 제외됨  
 2) 변수앞의 d표시는 1차 차분변수를 표시함  
 3) 5%의 유의수준에서의 임계치는 n=25일 때 -3.00임  
 4) \*표시는 5%의 유의수준에서 단위근존재에 대한 귀무가설이 기각됨

## 2. 요한센(Johansen) 다변량공적분 검정

우리는 대부분의 변수들이 최소한 하나의 단위근을 가지고 있음을 보았다. 1차 차분을 통한 안정적 과정을 확보할 수 있으나 1차 차분은 자료의 특성을 손실할 가능성이 많다. 그러나 자료들간에 공적분관계(cointegration)에 있을 때는 수준변수들을 이용할 수 있다. Enger and Granger(1987)의 방식은 각각의 시계열자료가 단위근을 가지고 있다 할지라도 일정한 조건하에서 장기적 관계를

3) AIC 및 BIC 분석에서 대부분 lag=1, lag=3으로 나타났다.

유도하는 2단계공적분 방법을 제시한다<sup>4)</sup>. 그러나 그들의 방법은 공적분 벡터의 모수를 추정하는데 초일치성을 가정한 최소자승법을 이용하였는데 제한적인 표본에서는 심각한 편차를 야기할 수 있다. 또한 변수가 2개 이상인 다변량의 경우에는 공적분의 개수가 2개 이상 존재하는 가능성에도 불구하고 공적분의 개수를 파악할 수 없다. 뿐만 아니라 벡터자기회귀에 기초하지 않고 단일방정식을 추정하고 있어 독립변수와 종속변수간의 선택구별이 확실하지 않다.

이러한 Enger and Granger(1987)의 방식의 한계점을 극복하고 다변량 공적분벡터를 추정하고 검증하는데 매우 유용한 Johansen(1988)과 Johansen and Juselius(1990)이 제시한 최우추정 방법(maximum likelihood estimation)을 들 수 있다. Johansen의 다변량검정법을 간략하게 설명한다면 I(1)변수들이 k차의 벡터자기회귀과정(vector autoregressive process)을 따른 것으로 가정하면

다음 식으로 표현된다.

$$X_t = \Pi_1 X_{t-1} + \dots + \Pi_k X_{t-k} + \mu + \varepsilon_t, \\ (t=1, \dots, T)$$

위 식에서  $\varepsilon_1 \dots \varepsilon_T$  는 정규분포를 따르는 백색 잡음이다. 즉, 평균 0, 분산이 1인 iid(0, 1)인 정규분포로서 잔차간에 서로 독립적이며 모집단과 동일하게 분포하는 확률변수를 나타낸다. 대부분 경제시계열이 불안정적 과정을 나타내므로 1차 차분형태로 추정한다.

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \dots + \Gamma_k \Delta X_{t-k+1} + \Pi X_{t-k} + \mu + \varepsilon_t \\ \Gamma_i = - (I - \Pi_1 - \dots - \Pi_i), \\ \Pi = - (I - \Pi_1 - \Pi_k), \quad i = 1, \dots, k-1.$$

$\Pi X_{t-k}$  항의 계수행렬  $\Pi$ 를 통하여 계수간의 장기적 균형관계의 존재를 분석할 수 있으며 다음과 같은 3가지가 있다.

4)  $Y_t = \alpha + \beta X_t + \varepsilon_t$ 의 모형에서 구한 잔차항에 대하여 다음과 같은 ADF검정을 실시한다.

$\Delta \hat{\varepsilon}_t = \gamma \hat{\varepsilon}_{t-1} + \sum_{i=1}^p \rho_i \Delta \hat{\varepsilon}_{t-i} + v_t$ , 이 때 검정할 가설은  $H_0 : \gamma = 0$ 임.  $\gamma$ 의 t통계량 값이 임계치보다 크면 귀무가설을 기각하는데, 즉 단위근이 존재하지 않는다는 것이며 변수들간에는 공적분관계가 있다는 것이다. 시차를 3기로 설정하였으며 각 모형에 대한 공적분 검정결과가 <표>에 정리되어 있다. 가구당소득(RINCOM), 지방세수입(RREVTX), 지역내총생산(RGRDP)의 자료 중 제1모형은 전 독립변수를 포함한 모형이고 제2모형, 제3모형은 독립변수 중 각각 첫 번째, 두 번째 변수들을 제외한 모형이다. 제1모형에서는 CRDW통계량이 1.44로서 임계치 0.78을 상회하여 공적분이 존재하는 것으로 밝혀졌고 ADF검정법에서도 공적분관계가 존재하나 DF검정에서는 공적분관계가 나타나지 않았다. 제2모형, 제3모형에서도 동일한 결과가 도출되었으나 DF보다 ADF통계량이 상대적인 신뢰성을 감안할 때 공적분 결과를 채택.

<표> 공적분 검정 결과

주 : CRDW : cointegration regression Durbin-Watson statistic.

$CRDW = \sum (\hat{e}_t - \hat{e}_{t-1})^2 / \sum (\hat{e}_t)^2$ 에서  $\hat{e}_t$ 는 공적분분석에서 잔차항임.

1)  $\text{rank}(II) = n$

$II$ 가 완전계수(rank)이며 벡터과정  $X_t$ 가 안정적이다. 즉  $I(0)$ 임을 의미한다.

2)  $\text{rank}(II) = 0$

$II$ 가 공행렬(null matrix)이며 변수들 간에 장기적인 관계가 존재하지 않음을 나타낸다.

3)  $\text{rank}(II) = r, (1 < r < N)$

이 경우는  $II = \alpha\beta'$ 로 표현되는  $(N \times r)$ 행렬  $\alpha$ 와  $\beta$ 가 존재한다. 즉 장기적인 균형관계가 존재하며  $\beta'X_t$ 가  $I(0)$ 이다. 최대  $r$ 개의 공적분벡터가 존재한다는 가설에 대한 우도비검정통계량은 다음과 같이 표시되는 Trace 통계량과 같다.

$$\text{Trace} = -T \sum_{i=r+1}^n \ln(1 - \lambda_i)$$

여기에서  $T$ 와  $n$ 은 각각 표본의 크기와 변수의 계수를 나타낸다. 공적분의 계수(rank)를 결정하기 위하여  $n$ 개의 단위근이 존재한다( $\lambda_1 = \dots = \lambda_n = 0$ )는 가설부터 시작하여 차수를 하나씩 줄여나가면서 계속적으로 검정을 하게 된다<sup>5)</sup>. 즉  $n$ 개의 단위근 가설이 기각되면 다시  $n-1$ 개의 단위근이 존재한다는 가설을 검정하게 된다. 결국  $k$ 개의 단위근 가설이 채택되면  $r(=n-k)$ 개의 공적분계수가 존재한다. 또한 공적분 계수행렬  $\beta$

의 계수(rank)가  $r$ 인지  $(r+1)$ 인지에 대한 우도비검정은 다음의  $\lambda_{\max}$  통계량을 사용한다.

$$\lambda_{\max} = -T \cdot \ln(1 - \lambda_{r+1})$$

[공적분 검정결과]

자료 : 1970년부터 1998년까지 서울시의 지역내 총생산, 가구당소득, 지방세수입, 그리고 1인당 재정총지출의 자료이며 소비자물가지수로 불변가격으로 환산하였다.<sup>6)</sup>

$\lambda_{\max}$  통계량분석결과  $h_0: r = 1$ 에서  $\lambda_{\max}$  통계량이 20.43으로서 5% 유의수준에서 기각되고  $h_1: r = 2$ 에서는  $\lambda_{\max}$  통계량이 7.67로서 5% 유의수준에서 기각시킬 수 없다. 따라서 지역내총생산, 가구당소득, 지방세수입, 그리고 1인당 재정총지출 등의 네가지 변수 사이에는 적어도 2개의 공적분계수가 존재하는 것으로 밝혀졌다. Trace 통계량분석도  $h_0: r = 1$ 에서 Trace통계량이 32.83으로서 5% 유의수준에서 기각되고  $h_1: r = 2$ 에서는 Trace통계량이 12.40으로서 5%유의수준에서 기각시킬 수 없어 적어도 2개의 공적분계수가 존재하는 것으로 밝혀졌다(<표 3>). 그러므로 지역내총생산, 가구당소득, 지방세수입, 그리고 1인당 재정총지출 등의 네 변수 사이에는 장기적인 균형관계가 존재함을 알 수 있다. 위 모형에 의해 공적분계수를 추정한 결과가 다음 <표 4>에 나타나 있다.

5)  $\lambda_i$ 는  $X_{t-k}$ 와  $\Delta X_t$ 간의 평방정준상관(squared canonical correlation)을 의미하고  $\Delta X_t$ 와  $X_{t-k}$ 를  $\Delta X_t$ 의 시차와 상수에 대하여 다음과 같이 회귀분석을 실시함.  $\Delta X_t = \sum_{i=1}^k A_{0i} \Delta X_{t-i} + \hat{\mu} + R_{0t}$ ,  $\Delta X_{t-k} = \sum_{i=1}^k A_{1i} \Delta X_{t-i} + \hat{\mu} + R_{kt}$ , 두 회귀식에서 추정된 잔차항  $R_{0t}$ 와  $R_{kt}$ 를 이용하여  $S_{kk}$ 에서  $S_{k0} S_{00}^{-1} S_{0k}$ 의 고유근  $\lambda$ 를 찾는 것임. 즉 다음의 방정식을 풀.

$$|\lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k}| = 0, S_{ii} = T^{-1} \sum_{t=1}^T R_{it} R_{it}', i, l = 0, k.$$

6) 서울통계연보(각 연도), 서울특별시

&lt;표 3&gt; 공적분 검정결과

$H_0: r$	n-r	eigenvalue	$\lambda_{max}$	Trace	$\lambda_{max,90}$	Trace,90
0	4	0.9196	40.34	73.17	17.15	43.84
1	3	0.7211	20.43	32.83	13.39	26.70
2	2	0.3807	7.67	12.40	10.60	13.31
3	1	0.2559	4.73	4.73	2.71	2.71

&lt;표 4&gt; 공적분 계수벡터의 추정

$\alpha$	$\beta$	$\Pi = \alpha\beta'$
-1.547 0.025	1.000 0.307	-1.539 -0.669 0.468 1.152
1.048 0.087	0.419 -0.435	1.075 0.401 -0.408 -0.676
2.593 0.738	-0.316 -0.832	2.820 0.764 -1.433 -1.149
1.163 -0.156	-0.728 1.000	1.115 0.555 -0.237 -1.003

## · 계량예측모형설정

### 1. ARIMA 모형

박스-젠킨스(Box-Jenkins) 모형이라 불리는 ARIMA모형(Autoregressive Integrated Moving Average Model)이 시계열자료의 정밀한 미래 예측을 위하여 최근에 많이 사용되고 있다. 분석대상이 되는 변수의 과거 관측치만으로 미래예측을 실시하며 경제이론에 근거하지 않고서도 전통적인 회귀분석보다 높은 정확성을 보여준다. ARIMA모형은 확률적 과정에 의하여 정의될 수 있는데 대표적으로 자기회귀확률과정(AR process)과 이동평균확률과정(MA process)으로 표현될 수 있다. 대부분의 시계열자료가 불안정적 과정을 보

일 때 1차 또는 2차 차분만으로 안정적 과정으로 전환될 수 있음을 앞절에서 보았다. ARIMA(p, d, q)모형은 d번 차분한 시계열을 p차의 AR과정과 q차의 MA과정의 결합함수로 표현된다.

ARIMA(p, d, q)모형을 후방전위연산자(barkward shift operator,  $L$ )을 이용하여 표현하면 다음과 같다. 여기서  $\phi(L)$ 은 AR연산자,  $\theta(L)$ 은 MA연산자를 의미한다.

$$\varphi(L)X_t = \theta(L)\varepsilon_t$$

$$\varphi(L) = \Delta^d \phi(L), \quad \Delta^d = (1 - L)^d$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 \cdots - \phi_p L^p$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 \cdots - \theta_q L^q$$

추정결과 모델전체가 바람직한 상태인지를 확인하는 적합도검정은 RMSE방식과 MAPE방식이 사용된다<sup>7)</sup>. 또한 확률적 충격의 독립성을 검정하기 위해서는 Bartlett검정법 및 Box-Pierce검정법등이 사용된다.<sup>8)</sup> 본 실증분석에서는 결산기준으로 일반회계 및 특별회계를 합한 1인당 총지출액(PRTAC)의 대수변수값으로 1970년대부터 1998년까지 연간자료를 이용하였다. 자료에 대한 단위근 검정은 앞 절에서 수행하였다. <표 5>에서는 확률과정의 단순성(parsimony)을 고려하여 각 ARIMA모델 추정계수의 유의성 검정을 실시하였다.

7) RMSE는 오차자승의 평균에 평방근을 씌운 값으로 상대적으로 적은 값을 가질수록 실제자료에 대한 적합도가 높다. 물론  $\hat{\varepsilon}_t$ 는 알지 못하므로 그 추정량인  $\hat{\varepsilon}_t$ 를 사용한다. MAPE(Mean absolute percent error)는 잔차항을 원시계열로 나눈 후 다시 관측치로 나눈 평균비율을 구한 뒤 100을 곱하여 주어 %값으로 환산하여 표시된다.  $RMSE = \sqrt{\frac{\sum \hat{\varepsilon}_t^2}{T}} = \sigma$ ,  $MAPE = 100 \times \frac{1}{T} \sum \left| \frac{\hat{\varepsilon}_t}{Y_t} \right|$

8) 자세한 내용은 이종원(1994) 참조

&lt;표 5&gt; ARIMA(p, d, q)모형 추정계수 유의성검정

모형	(0, 1, 1)	(1, 1, 0)	(1, 1, 1)
constant ( $t_{ik}^1$ )	0.19 (6.30)	0.18 (5.6)	-0.13 (-0.01)
$\Phi_1$ ( $t_{ik}^1$ )	-	0.12 (0.58)	0.99 (3.44)
$\theta_1$ ( $t_{ik}^1$ )	0.07 (0.33)	-	-0.92 (-2.28)
모형	(1, 1, 2)	(2, 1, 1)	(2, 1, 2)
constant ( $t_{ik}^1$ )	0.17 (4.28)	0.18 (3.35)	0.18 (6.21)
$\Phi_1$ ( $t_{ik}^1$ )	-0.04 (-0.08)	0.09 (0.19)	0.04 (0.26)
$\Phi_2$ ( $t_{ik}^1$ )	-	0.37 (1.79)	-0.34 (-2.79)
$\theta_1$ ( $t_{ik}^1$ )	0.13 (0.25)	0.04 (0.07)	0.08 (0.25)
$\theta_2$ ( $t_{ik}^1$ )	0.41 (2.02)	-	1.78 (5.58)

추정계수들의 안정성 및 가역성조건을 충족시키기 위해서는 다음의 조건을 만족시켜야 한다.

$$|\hat{\phi}_i| < 1, \quad |\hat{\theta}_i| < 1,$$

$$\sum \phi_i < 1, \quad \sum \theta_i < 1$$

$$\phi_2 - \phi_1 < 1, \quad \theta_2 - \theta_1 < 1, \quad \text{단 } i = 1, 2.$$

ARIMA(2,1,2)모형만이 위의 조건을 만족시키지 못하고 다른 ARIMA(p,d,q)모형들은 위의 조건을 만족시키고 있다. <표 6>의 모형 적합도검정에서 높은  $R^2$ 와 낮은 MAPE(Mean Absolute percent error)는 실제자료에 대한 적합도가 높음을 의미한다. 잔차항의 독립성검정에서 Ljung-Box 검정을 적용하면 자유도 6 또는 7일 경우 임계치가 12.59 또는 14.07을 기준으로  $\chi^2$ 값 모두가 잔차항의 독

립성이 보장됨을 보인다. 또한 확률충격의 독립성 여부를 검정하기 위하여 Bartlett검정법을 사용하면 자기상관함수의 t값이  $k=1, 2, 3$ 에서 1.25보다 작고  $k \geq 4$ 에서 1.6보다 작다면 확률충격의 독립성이 보장된다. 검정결과  $k=2$ 일때 ARIMA(0,1,1)모형과 ARIMA(1,1,0)모형은 위의 조건을 충족시키지 못하고 있다.

이제 남은 3가지 모형 ARIMA(1,1,1)모형, ARIMA(1,1,2)모형 및 ARIMA(2,1,1)모형 중에서 <표 5>의 추정계수의 t값을 살펴볼 때 ARIMA(1,1,1)모형만이 추정계수가 통계적으로 유의함을 보여 ARIMA(1,1,1)모형이 최선의 모형으로 판명되었다.

&lt;표 6&gt; ARIMA모형 적합도 검정결과

	(0,1,1)	(1,1,0)	(1,1,1)	(1,1,2)	(2,1,1)
MAPE	0.72	0.70	0.65	0.65	0.57
$R^2$	0.99	0.99	0.99	0.99	0.99
Ljung-Box Q	3.52	3.32	2.53	0.51	1.67
Bartlett( $t_{ik}^1$ ) $\rho_1$	0.12	-0.12	-0.15	0.13	0.01
$\rho_2$	1.75	1.64	1.22	0.06	0.24
$\rho_3$	0.53	0.26	-0.15	0.56	0.78
$\rho_4$	0.17	0.01	-0.45	0.11	-0.64

## 2. 베이지언 벡터자기회귀모형

VAR(Vector Autoregression)모형이란 시계열 변수들간에 상호작용을 파악하여 미래의 방향을 예측하기 위한 목적으로 개발한 다변량시계열모형이다. 선형적인 경제이론을 배제한 상태에서 변수들간의 관계에서 나타나는 특징을 분석한다.”

9) 본 절의 주요내용을 본문에서 제시하고 분석과정에 관한 내용을 부록처리 가능하나, 분석모형을 단계별 설명하는데 있어 분리하여 부록으로 제시하기에는 어려움이 있어 그대로 기술함.

또한 VAR모형은 전통적인 거시계량모형이 안고 있는 '0의 제약'이라는 경직성 및 자의적인 내생 변수와 외생변수와의 구분 등의 문제점을 극복하고 보다 정확한 예측을 얻기 위한 대안으로 개발되었다. VAR모형은  $n$ 개의 선형회귀방정식으로 구성되는데 각 방정식은 각 변수들의 현재 관측치를 종속변수로 하고 자신과 타 변수들의 과거 관측치를 설명변수로 설정한다.

$$X_t = A(L)X_t + e_t = \sum_{k=1}^L A_k X_{t-k} + e_t$$

( $X_t$ 는  $n \times 1$  벡터, 시차  $L$ )

그런데 VAR모형에서도 다음과 같은 문제점이 있다. VAR모형은 모형의 추정이나 충격함수, 예측오차의 분산분해 등 자료의 분석에서 모형체계에 포함되는 변수의 종류, 변수의 순서 등의 설정에 따라 분석결과가 예민하게 반응한다. 시차분포에 사전제약이 없는 비제약VAR모형은 한정된 자료를 가지고 계수를 추정하므로 과다모수로 인해 자유도가 급격히 떨어져 계수추정치가 불안정해지고 이에 따라 예측자체의 정확도를 크게 떨어뜨리는 결과를 초래한다. 대부분 경제시계열이 짧으므로 시차를 길게 할수록 문제가 있게 된다.

따라서 추정기간의 표본크기가 작아 예측력을 저하시키는 과다계수문제(overparameterization)가 발생할 가능성이 높을 것으로 우려됨에 따라 Todd, Litterman 등이 제시한 선험적 주관(베이지안 사전확률, Bayesian prior)을 이용하여 추정 계수에 사전적인 제약을 부과하는 베이지안VAR모형을 사용하였다. 베이지안VAR모형은 통계적으로 시차가 길어질수록 시차항계수의 사전분포는 0이고 표준편차도 작은 정규분포에 따른다는

개념에 기초하고 있다. 시차가 길어질수록 계수의 표준편차에 대하여 0에 가까운 사전확률을 부여함으로써 추정상의 문제를 해결한다. VAR에서 각 방정식내의 각 결정항에 대한 사전확률은 비정보적이다. 시차의 사전분포는 독립된 정규분포에 따른다. 모든 계수의 사전분포에 대한 평균은 0. 단, 각 식에서 종속변수에 대한 첫 번째 시차계수의 평균은 특별히 다른 값을 주지 않는 한 1로 가정하여 시계열 변수들이 랜덤워크 가설(random walk hypothesis)을 따른다.

위의 세가지 가정을 기초로 사전확률을 설정하기 위해 다음과 같은 두가지 정보를 필요로 한다. 첫째는 각 식에서 종속변수의 첫 번째 시차변수의 계수에 대한 사전분포의 평균이며 둘째는  $i$ 식에서 변수  $j$ 의  $l$ 번째 시차에 대한 사전분포의 표준편차  $S(i, j, l)$ 이다. 보통 첫 번째 시차계수의 사전평균을 1로 정하므로 표준적인 사전확률은 다음의 표준편차를 정해주는 문제이다.

$$S(i, j, l) = \frac{[\lambda g(lf(i, j))] * S_i}{S_j}$$

$$i, j = 1, 2, \dots, n,$$

$$l = 1, 2, \dots, p, \quad f(i, l) = g(l) = 1.0$$

$S_i$ 는  $i$ 식에서 일변량자기회귀식의 표준오차이며 상이한 변수크기를 조정할 목적으로  $S_i/S_j$ 를 곱한다. 이때  $\lambda g(lf(i, j))$ 은 각 계수에 부여되는 사전확률에 대한 가중치를 의미한다. 베이지안VAR모형은 이 값을 결정하는 문제로 요약된다.  $\lambda$ 는 전체적인 가중치(overall tightness)로서 종속변수의 첫 번째 시차항계수에 대한 표준편차를 의미한다. 일반적으로 0.1과 0.2사이의 값을 택하



나  $\lambda$ 값이 크면 사전확률이 느슨해져 베이저언 VAR모형은 비베이저언VAR모형으로 바뀌게 된다.  $g(l)$ 은 첫 번째 시차와 비교한  $l$ 번째 시차의 상대적 중요도로서 시차의 길이에 따른 시차 형태를 정한다. VAR모형에서는 사전확률을 정하지 않고 시차를 짧게 설정하는 것보다는 시차를 다소 길게 설정한 후 긴 시차일수록 작은 값의 사전확률을 정하는 것이 보다 나은 방법으로 알려져 있다. 함수  $g(l)$ 의 형태는 크게 두 가지로 구분된다.

hamonic인 경우  $g(l) = l^{-d}$ , 여기서  $d$ 값을 크게 정하면 시차가 길어질수록 계수값이 0에 더욱 근접한다. geometric인 경우  $g(l) = d^{l-1}$ , 여기서  $d$ 값을 작게 정하면 시차가 길어질수록 계수값이 0에 더욱 근접한다. 일반적으로  $d$ 가 1 또는 2인 hamonic 시차감소유형을 이용한다.  $f(i,j)$ 는  $i$ 식에서 종속변수  $i$ 와 비교한 변수  $j$ 의 상대적 중요도로 다음과 같이 표현된다.

$$f(i,j) = \begin{cases} 1.0, & \text{if } i=j \\ w, & \text{otherwise} \end{cases}$$

$w$  값의 결정이 중요한데 일반적으로 대칭적, 일반적, 주변핵심 사전확률 등 세가지 방법이 있다. 대칭적(symetric) 사전확률은 각 식에서 종속변수에 대한 타 변수들의 중요도를 모두 동일하게 취급하는 방법으로 주로 5개 이하의 변수들로 구성되는 소규모 VAR체계에서 많이 이용된다. 일반적(general) 사전확률은 변수의 상대적 중요도(relative tightness)에 따라  $w$ 를 모두 다르게 하는 방법이다. 주변핵심 사전확률은 예측에 대한 기여도 및 변수의 상대적 중요도 측면에서 우월

하다고 판단되는 변수를 핵심으로 하고 그렇지 못한 변수를 주변변수로 하여 핵심변수의 시차항 계수에 대한 표준편차가 주변변수에 비해 커지도록 사전확률을 정하는 방법이다.

베이저언VAR모형을 구축하기 위해서는 대상변수의 선정, 시차길이의 선택, 베이저언사전확률탐색 등을 하여야 한다. 본 분석에 사용된 자료는 1970년부터 1998년까지 29년간의 연간자료, 대상변수의 선정, 변수의 형태 : 1995년을 기준(=100)으로 한 소비자 물가지수를 조정한 불변가격으로 전환하여 자연대수를 취함. 총재정지출의 경우 인구수로 나눈 1인당 총재정지출을 사용,

시차구조 : 심즈(Sims)의 우도비검정에 의한 오차항의 회귀잔차가 백색오차가 되는 최소 시차 결정법이 있으나 예측오차를 최소화하기 위한 시차길이를 모형예측력을 최소화하는 시차구조 3년을 설정, 결정항 : 상수항과 추세항, 감소유형 :  $d$ 가 1인 hamonic을 선택하여 각 방정식에서 종속변수 자산의  $l$ 번째 시차계수의 표준편차\* $1/l$ 을 표준편차로 갖도록 한다.

모형별 계수값은 일변량AR, 일변량BVAR, 단순BVAR모형과 일반적 사전확률에 의한 베이저언VAR모형 등의 사후예측력을 비교하였다.

- 일변량자기회귀모형 :  $\lambda$ 값=2.0,  $w$  값=0.001
- 일변량베이저언VAR모형 :  $\lambda$ 값=0.1,  $w$  값=0.001
- 단순베이저언VAR모형 :  $\lambda$ 값=0.1,  $w$  값=0.5
- 일반적 사전확률에 의한 베이저언VAR모형  
 $\{(1.00 \ 0.50 \ 0.50 \ 0.50), (0.1 \ 1.50 \ 0.10 \ 0.10),$   
 $(0.5 \ 0.5 \ 1.00 \ 0.50), (0.50 \ 0.50 \ 0.50 \ 1.00)\}$

1차시차변수계수의 표준편차( $\lambda$ )와 타변수간의

상대적 표준편차의 비율( $w$ )을 각각 0.1, 0.2, 2.0까지 변경시키면서  $\lambda$ 값과  $w$ 값별로 칼만

<표 7> 모형별 예측력 비교

		일반량 AR	일반량 BVAR	단순 BVAR	BVAR
RINCOM	1	0.84	0.88	0.81	0.80
	2	0.78	0.76	0.64	0.64
	3	0.69	0.66	0.51	0.51
	4	0.73	0.63	0.47	0.47
	5	0.67	0.57	0.43	0.43
	6	0.74	0.58	0.44	0.44
	7	0.77	0.52	0.44	0.43
	8	0.71	0.46	0.39	0.38
	9	0.84	0.56	0.40	0.40
	10	0.92	0.40	0.20	0.19
RREVTX	1	0.79	0.87	0.77	0.85
	2	0.83	0.85	0.72	0.83
	3	0.86	0.93	0.71	0.89
	4	1.10	1.16	0.85	1.13
	5	1.34	1.38	0.99	1.34
	6	1.52	1.54	1.06	1.50
	7	1.58	1.54	1.07	1.51
	8	1.36	1.33	0.91	1.31
	9	1.28	1.23	0.72	1.21
	10	1.48	1.09	0.62	1.07
RGRDP	1	0.78	0.79	0.80	0.79
	2	0.64	0.64	0.64	0.64
	3	0.52	0.51	0.51	0.51
	4	0.44	0.44	0.44	0.44
	5	0.44	0.43	0.43	0.43
	6	0.39	0.38	0.37	0.37
	7	0.42	0.42	0.42	0.42
	8	0.38	0.36	0.38	0.38
	9	0.22	0.21	0.20	0.20
	10	0.15	0.14	0.12	0.10
PRTAC	1	1.02	0.96	0.96	0.96
	2	0.90	0.84	0.83	0.83
	3	0.91	0.82	0.82	0.82
	4	0.90	0.78	0.79	0.79
	5	0.89	0.77	0.78	0.78
	6	0.94	0.80	0.80	0.80
	7	0.85	0.77	0.81	0.80
	8	0.85	0.69	0.77	0.76
	9	0.94	0.75	0.85	0.84
	10	1.09	0.82	0.97	0.97

주 : RINCOM(가구당소득), PREVTX(지방세수입), RGRDP(지역내총생산), PRTAC(총재정지출)

(Kalman)의 여과법을 이용하여 1999년부터 2007년까지 9년간 사후예측을 하였다. 여기서 사후예측을 위한 예측단계별 예측력검정통계량은 관측자료의 크기나 단위와 무관한 타일(Theil)의 불균등계수  $U$ 를 이용하였다.

베이지언 사전확률을 선택하기 위해 <표 7>은 각 모형에 대한 내생변수들의 타일  $U$ 값을 나타내고 있다. 각각 내생변수에 대해 일반량자귀회귀모형, 일반량베이지언VAR모형, 단순베이지언VAR모형, 그리고 일반적 사전확률에 의한 베이지언VAR모형 등 4가지 유형에 의한 사후예측결과가 나타나 있다. 일반적 사전확률에 의한 베이지언VAR모형의 사후 예측력이 일반량자귀회귀모형, 일반량베이지언VAR모형보다는 현저히 타일  $U$ 값이 개선되고 있다. 그러나 단순 BVAR모형과 비교시 가구당소득, 총재정지출, 지역내총생산 등에서는 사후 예측력이 우수하나 지방세수입에서는 타일  $U$ 값의 개선이 뒤떨어지고 있다. 따라서 단순BVAR모형을 사용하여도 큰 차이는 없을 것으로 보인다.<sup>10)</sup>

일반적 사전확률에 의한 베이지언VAR모형에 대한 사후적 예측의 정확도를 분석하기 위하여 RMS백분율오차(root mean square percent error)를 살펴보았다.<sup>11)</sup> 4년간의 사후적 예측에서 <표 8>의 결과와 같이 RMS백분율오차(RMSP)가 0.5663%로 나타나 예측력이 우수한 것으로 판정되었다.

10) 충격반응함수에 관한 논의에서는 직교화과정에서의 변수들의 배열순서에 의해서도 견고성이 밝혀졌으며 충격반응에 대한 추가적인 설명체제가 가능하나 본 연구의 목적이 추정모형의 탐색에 있으므로 최종논문에서는 삭제하였음. 추정모형II, ECM모델에서 언급하고자 함.

11)  $RMSP = \sqrt{\frac{1}{T} \sum_{t=1}^T \left( \frac{G_t^p - G_t}{G_t} \right)^2}$   $G_t^p$ 는 예측치이며  $G_t$ 는 실제관측치임.

&lt;표 8&gt; BVAR모형의 사후 예측결과

예측기간(년)	실제관측치	예측치	RMS
1995	13441	13411	0.5663
1996	13515	13465	
1997	13558	13519	
1998	13432	13567	

주 : 실제관측치 및 예측치자료는 일인당 총재정지출을 서울시 소비자물가지수로 나눈 불변자료로 단위조정을 거친 자연대수값임

## · 결론

본 연구는 서울시의 재정지출 예측을 위한 최적모형을 개발하고자 하였다. 이러한 작업은 연결되는 서울시 세수입 예측모형 개발과 더불어 예측에서 나타나는 서울시 재정활동의 세입세출 부족분을 어떻게 조달하는 것이 바람직한 것인가 하는 결과수준까지를 고려하는 사전 연구단계라고 할 수 있다.

서울시 재정지출 자료뿐만 아니라 주요 변수로 선정된 서울시의 가구당소득, 지방세수입, 지역내 총생산 등의 자료들이 단위근검정을 통하여 모두 불안정적인 시계열과정을 따르는 것으로 판명되었다. 또한 동 변수들간 공적분검정에서는 공적분계수가 적어도 두 개 존재하는 등 공적분관계에 있는 것으로 밝혀져 차분에 따르는 자료의 손실을 막을 수 있었다.

서울시 실증분석 첫 번째 모형에서는 ARIMA (1, 1, 1)모형이 가장 적합한 추정모형으로 판정되었으며 추정계수에 사전적인 제약을 부과하는 베이즈안VAR모형에서 일반적 사전확률에 의한 모형이 사후 예측력이 높은 것으로 나타났다. RMS 백분율오차를 이용한 사후적 예측 정확도 분석에서도 0.5663%로 나타나 예측력이 우수한 것으로

판정되었다.

ARIMA모형과 BVAR모형을 이용하여 경상가격 결산기준으로 서울시 재정지출의 사전적 단기적인 예측을 수행하면 ARIMA모형은 2001년 10조4,466억원, 2002년 11조339억원으로 예측되었으며, BVAR모형의 경우에는 2001년 11조81억원, 2002년 11조6,505억원으로 추정되었다.

향후 연구과제로서는 일반행정부, 사회개발비, 경제개발비 등 일반회계 중 기능별 분류에 따른 세부적 분석도 필요하다고 본다. 또한 중앙 및 지방정부간 업무조정과 그에 따른 재원부담의 조정도 이루어지고 있는 바 이에 따른 모형의 조정도 필요하다고 본다.

## 참고문헌

- 서울특별시, 『서울통계연보』 각 연도  
 이종원, 『계량경제학』 박영사, 1994  
 Engle, R., and C. Granger, "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica* 55(1987), p.251-276  
 Johansen, S., "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* 12(1988), p.231-254  
 \_\_\_\_\_, and K. Juselius, "The full information maximum likelihood procedure for inference on cointegration-with applications to the demand for money", *Oxford Bulletin of Economics and Statistics*, 52(1990), p.169-210  
 Stock, S. and M. Watson, "Testing for common trends," *Journal of the American Statistical Association* 83(1988), p.1097-1107