The Korean Journal of Applied Statistics (2012) DOI: http://dx.doi.org/10.5351/KJAS.2012.25.4.553
25(4), 553–561

A Study of a Combining Model to Estimate Quarterly GDP

Changku Kang[1]

[1]Economic Statistics Department, The Bank of Korea

(Received May 15, 2012; Revised June 9, 2012; Accepted July 12, 2012)

Abstract

Various statistical models to Estimate GDP (measured as a nation's economic situation) have been developed. In this paper an autoregressive distributed lag model, factor model, and a Bayesian VAR model estimate quarterly GDP as a single model; the combined estimates were evaluated to compare a single model. Subsequently, we suggest that some combined models are better than a single model to estimate quarterly GDP.

Keywords: Combining estimates, quarterly GDP, factor model.

1. 서론

한 나라의 경제상황을 나타내는 대표적인 지표로는 국내총생산(GDP) 통계가 있다. GDP는 산업생산, 건설기성액, 전력판매량 등 다양한 기초자료를 이용하여 추계되고 있다. 이에 따라 GDP 추계에는 많은 시간과 인력이 필요한 것이 사실이다. 한편 경제분석 및 연구 등의 목적으로 보다 간편하게 GDP를 추정하기 위한 다양한 계량모형도 개발되고 있다. Min 등 (2002)은 공급부문과 수요부문별 GDP를 추정하는 자기회귀분포모형을 기반으로 한 회귀모형을 구축하였다. Reijer (2005), Schneider과 Spitzer (2004)는 자국의 GDP를 추정하기 위해 인자모형을 이용한 추정방법을 개발하였다. Shim과 Lee (1992)은 내수부문을 추정하기 위해 베이지안 VAR모형을 구축하였다.

하나의 계량모형은 중장기적 추세를 벗어난 불규칙성을 정확히 추출해 내는데 현실적으로 한계가 있을 수 밖에 없다. 이는 각 계량모형이 서로 다른 이론적 배경과 기본가정을 토대로 설정되어 있어 변수들 사이에 존재하는 서로 다른 정보를 기초로 추정하며 모형별로 이용되는 변수 선택도 서로 다를 수 있기 때문이다. 어떤 계량모형을 이용하여 추정하는 가에 따라 그 결과가 서로 상이할 뿐만 아니라, 특히 일부시점에서는 비현실적인 추정치가 산출될 수 있다. 따라서 하나의 개별모형에만 전적으로 의존하기 보다는 다른 모형에서 인식하는 고유정보를 부가적으로 활용함으로써 보다 나은 추정치를 유도해 낼 수 있다는 논리가 성립할 수가 있다. 이같은 점에 착안해 Bates와 Granger (1969)는 두 가지 이상의 모형을 결합하여 보다 정확한 예측값을 찾을 수 있는 결합모형을 제시하였다. 이후에도 Granger과 Ramanathan (1984)은 회귀분석을 이용한 결합모형을 개발하였으며 Diebold (1988)은 계열상관을 고려한 결합방법을 제시하는 등 개별모형을 결합하는 방법에 대한 연구가 지속되어 오고 있다. Collopy와

[1]Economist, Economic Statistics Department, The Bank of Korea, 110, 3-Ga, Namdaemun-Ro, Jung-Gu, Seoul 100-794, Korea. E-mail: koncap@bok.or.kr

Armstrong (1992)는 결합모형 관련 검증실험에서 실험참가자의 83%가 결합모형에 의한 예측치가 개별모형에 의한 예측치 보다 정확하다는 결과를 제시함으로써 결합모형을 실무적으로 이용할 수 있는 근거를 마련하였다.

본고에서는 분기 GDP를 추정하는 데에 결합모형이 개별모형에 의한 추정치보다 우월한지를 실증분석을 통해 알아보고자 한다. 개별모형은 회귀모형, 인자모형, 베이지안 VAR모형 등 일반적으로 널리 알려진 대표적인 추정모형을 선택하였으며 추정력 검정은 추정치와 실적치간 차이로부터 계산된 오차통계량을 통한 비교와 오차에 대한 통계적 검정 등을 실시하여 분석하였다.

2. 결합추정 기법

결합추정기법 가운데 Bates와 Granger (1969)가 제시한 분산최소기법은 개별추정치의 오차가 불편성(unbiasedness)을 가지며 분산이 안정적일 때 결합추정치의 평균제곱오차가 개별추정치의 평균제곱오차보다 작게 된다는 점에 근거하고 있다. y_t를 실적치, p_{1t}와 p_{2t}를 서로 다른 두 가지 계량모형에 의한 추정치라고 하면 2개 개별추정 모형식은 다음과 같다.

$$y_t = p_{1t} + \epsilon_{1t}, \qquad y_t = p_{2t} + \epsilon_{2t},$$

여기서 개별모형의 오차 ϵ_{1t}, ϵ_{2t}는 평균 0이고 분산 $\sigma_1^2 = E(\epsilon_{1t}^2)$, $\sigma_2^2 = E(\epsilon_{2t}^2)$, 공분산 σ_{12}를 갖는 추정오차이다. 결합추정치 c_t를 p_{1t}와 p_{2t}의 가중치 λ_1, λ_2를 이용한 선형결합 $c_t = \lambda_1 p_{1t} + \lambda_2 p_{2t}$라고 할 때 결합추정치 c_t의 오차는 다음과 같다.

$$y_t - c_t = y_t - \lambda_1(y_t - \epsilon_{1t}) - \lambda_2(y_t - \epsilon_{2t}).$$

불편성 유지를 위한 제약조건 $\lambda_1 + \lambda_2 = 1$을 만족하는 결합추정치 c_t의 오차분산은

$$\sigma_c^2 = \lambda_1^2\left(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}\right) - 2\lambda_1\left(\sigma_2^2 - \sigma_{12}\right) + \sigma_2^2$$

이다. 이 때 σ_c^2을 최소화하는 λ_1과 λ_2는 다음과 같이 유도된다.

$$\lambda_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}, \qquad \lambda_2 = 1 - \lambda_1 = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}.$$

Newbold와 Granger (1974)는 2개 추정치를 결합하는 Bates와 Granger (1969)의 방법을 일반화하여 m개 추정치를 결합하는 방법으로 확장하였다. 실적치 y_t에 대해 m개의 추정치, $p_{jt}, (j = 1, \dots, m)$가 존재한다고 가정할 때 j번째 추정치에 대한 오차는 $\epsilon_{jt} = y_t - p_{jt}$이다. 이 때 결합추정치는

$$c_t = \lambda_1 p_{1t} + \cdots + \lambda_m p_{mt}, \quad \lambda_1 + \cdots + \lambda_m = 1$$

이며, 오차분산을 최소화하는 가중치 벡터는

$$\hat{\lambda} = \frac{\Sigma^{-1} l}{l^T \Sigma^{-1} l}$$

가 된다. 여기서 Σ는 추정오차의 공분산행렬, l은 길이가 m이며 값이 모두 1인 벡터이다.

한편 Granger과 Ramanathan (1984)은 회귀모형 개념에 의해 실적치와 결합추정치의 오차가 최소화되는 가중치를 산출함으로써 개별추정치를 결합하는 방식을 제안하였다. $y^T = (y_1, \dots, y_n)$을 n개 관

측치 벡터, $p_j^T = (p_{j1}, \ldots, p_{jn})$, $j = 1, \ldots, m$는 j번째 모형에 의해 산출한 시점별 추정치 벡터라고 할 때 $P = (p_1, \ldots, p_m)$는 $n \times m$행렬이다. 결합추정치 $c^T = (c_1, \ldots, c_n)$는 개별추정치에 가중치 $\lambda^T = (\lambda_1, \ldots, \lambda_m)$을 각각 곱한 후 상수항 λ_0을 반영한 다음과 같은 회귀모형으로 나타낼 수 있다. 이 때 상수항을 추가하는 것은 개별추정치가 편향(biased) 추정치이더라도 결합추정치는 비편향(unbiased) 추정치가 될 수 있는 근거를 제공한다.

$$c = \lambda_0 l + P\lambda + \epsilon,$$

여기서 ϵ은 백색잡음, l은 길이가 n이며 값이 모두 1인 벡터이다. 가중치 λ는 통상 최소제곱법에 의해 다음과 같이 산출된다.

$$\hat{\lambda} = \hat{\alpha} - \hat{\lambda}_0 \left(P^T P\right)^{-1} P^T l,$$
$$\hat{\lambda}_0 = \frac{l^T y - l^T P\hat{\lambda}}{n} = \frac{l^T \hat{e}}{n - \hat{\theta}}, \tag{2.1}$$

여기서 $\hat{\alpha} = (P^T P)^{-1} P^T y$, $\hat{\theta} = l^T P (P^T P)^{-1} P^T l$, $\hat{e} = y - P\hat{\alpha}$는 추정오차벡터를 나타낸다. 따라서 결합추정치 $\hat{c}$은 다음과 같다.

$$\hat{c} = \hat{\lambda}_0 l + P\hat{\lambda}. \tag{2.2}$$

3. 추정력 평가방법

결합모형의 추정력을 평가하기 위해 우선 훈련자료(training data)를 이용하여 모형을 추정한다. 추정된 모형식에 시험자료(test data)를 대입하여 오차를 계산한다. 본고에서는 추정력 평가를 다음의 2가지 방법을 적용하였다. 첫번째는 GDP 추정치 $\hat{y_t}$와 실적지 y_t간 오차를 평균절대값오차(MAE)와 제곱근평균제곱오차(RMSE) 등 대표적인 오차통계량 값을 계산하여 비교하였다. 두 오차통계량은 대체로 비슷한 값을 나타내지만 MAE가 RMSE에 비해 특이치에 로버스트(robust) 하다는 특징이 있다.

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=t_1}^{t_N} |y_t - \hat{y_t}|, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=t_1}^{t_N} (y_t - \hat{y_t})^2}\ . \tag{3.1}$$

두 번째 평가방법은 개별 GDP 추정치와 결합 추정치의 상대적 정확성을 통계적으로 검정하기 위한 방법으로 Harvey 등 (1997)의 검정방법(이하 HLN검정)을 이용하였다. 서로 다른 두 가지 추정방법에 의한 추정치와 실적치간의 오차를 각각 e_{1t}, e_{2t} $(t = t_1, \ldots, t_N)$라고 할 때 HLN검정의 귀무가설

$$H_0 : g(e_{1t}) - g(e_{2t}) = 0, \quad g(x) = x^2$$

에 대한 검정 통계량 S는

$$S = \left(\frac{N-1}{N}\right)^{0.5} \cdot \frac{\overline{d}}{\sqrt{\widehat{V(\overline{d})}}}$$

이다. 여기서 $\overline{d} = (1/N)\sum_{t=t_1}^{t_N}(g(e_{1t}) - g(e_{2t}))$, $\widehat{V(\overline{d})} = (1/N)\widehat{\gamma_0}$이며 $\widehat{\gamma_0}$는 $g(e_{1t}) - g(e_{2t})$의 자기공분산이다. 검정통계량 S가 자유도 $N-1$인 t-분포의 임계치보다 작은 값을 갖는다면 모형 1의 추정력이 모형 2보다 우수한 것으로 판단한다.

4. 실증분석

개별모형은 GDP추정시 대표적으로 이용되고 있는 회귀모형과 베이지안 VAR모형 그리고 최근 개발된 여러변수의 정보로부터 공통성분을 추출하는 인자모형 등 3가지 모형을 선정하였다. GDP통계가 생산측면 지표를 중심으로 작성된다는 점에 착안하여 제조업생산지수, 서비스업생산지수(총지수), 전력판매량, 건설기성액 등의 생산량 변수들을 주요 후보변수로 선정하였다. 다만 인자모형은 다양한 변수들을 이용할 수 있다는 특성으로 인해 생산측면 변수 뿐만 아니라 소비, 투자, 수출입 등의 수요측면 변수까지도 추가하였다. 생산지수 등과 같은 변수는 가격요인이 제거된 불변지수를 이용하였으며 수출입 등은 경상금액을 해당물가지수를 이용하여 실질화하였다. 분석대상 시계열은 1992년 1/4분기∼2006년 4/4분기로 설정하였다. 모형추정은 1992년 1/4분기∼2001년 4/4분기의 훈련자료를 이용하였으며, 2002년 1/4분기∼2006년 4/4분기는 추정력 평가를 위한 시험자료로 이용하였다. 이때 서비스업생산지수 등과 같이 과거 시계열이 존재하지 않는 계열은 통계적방법을 이용하여 별도로 추정하였다. 시계열 추정은 다른 경제지표의 정보로부터 해당 시계열의 과거시점을 결측치로 간주하여 이를 추정하는 방법으로 Stock과 Watson (2002)의 EM 알고리즘 방법을 이용하였다.

4.1. 회귀모형에 의한 GDP추정

회귀모형의 설명변수는 우선 분기 GDP와 각 분기지표간의 개별 회귀분석을 통해 설명변수로서의 유용성을 평가하여 단계별 변수선택과정(stepwise selection procedure)을 적용하여 유의한 변수를 골라내었다. 그 결과 건설기성액을 제외한 4개 지표가 GDP에 대해 유의한 설명력을 가지고 있었으며 최종적으로 추정된 모형식은 식 (4.1)과 같다.

$$\begin{aligned}\mathrm{GDP}^R_t &= \underset{(3.02)}{0.283\,\mathrm{GDP}^R_{t-1}} + \underset{(6.11)}{0.234\,\mathrm{IPI}_t} - \underset{(-2.24)}{0.102\,\mathrm{IPI}_{t-1}} + \underset{(5.06)}{0.183\,\mathrm{ELE}_t}\\ &\quad + \underset{(2.84)}{0.256\,\mathrm{EMP}_t} + \underset{(3.86)}{0.302\,\mathrm{SP}_t} - \underset{(-2.54)}{0.172\,\mathrm{SP}_{t-1}} + \mathrm{DUM},\end{aligned} \tag{4.1}$$

여기서 GDP^R은 회귀모형에 의한 분기 실질 GDP, IPI는 제조업생산지수, ELE는 전력판매량, EMP는 비농가취업자수, SP는 서비스업생산지수의 전년동기비를 각각 나타내며 DUM은 외환위기를 나타내는 이상치 조정항이다. ( )안은 각 회귀계수들에 대한 t-통계량 값이다. 회귀모형에서 결정계수 R^2는 0.973으로 매우 높은 모형적합도를 가지며 DW-통계량도 2.064로 오차의 독립성가정을 만족하고 있다.

4.2. 인자모형에 의한 GDP추정

인자모형을 이용하여 분기 GDP를 추정하기 위해서 다수의 경제지표들 간에 존재하는 공통적인 특성인 미지의 공통인자를 추출하였다. 이를 위해 p개 변수 y_{it}가 변수들 사이에 공통적으로 영향을 미치는 m개 공통인자 F_{jt}와 개별변수에 영향을 미치는 특정인자 ϵ_{it}로 구성된다고 가정한다.

$$y_{it} = \Lambda_i^T F_{jt} + \epsilon_{it}, \quad i = 1, \ldots, p,\ j = 1, \ldots, m,\ t = 1, \ldots, n,$$

여기서 Λ_i는 인자적재로서 변수 y_{it}에 대한 공통인자 F_{jt}의 중요도를 나타내는 가중치이다. 각 시점별 공통인자 추정치 $\hat{F}_{jt}$는 다음과 같다.

$$\hat{F}_{jt} = \hat{\Lambda}_j^T \left(\hat{\Lambda}_j \hat{\Lambda}_j^T + \hat{\psi}_j^{-1}\right) y_t, \quad j = 1, \ldots, M,\ t = 1, \ldots, n,$$

Table 4.1. Estimation result of factor loading and common factor

Variable	Coefficient of Correlation with GDP	Factor Loading		Common Factor	
		FAC1	FAC2	FAC1	FAC2
Manufacturing Production Index	0.78	0.38	0.92	−0.681	1.364
(Food, Beverages, Tobacco)	0.75	0.69	0.40	0.007	−0.003
(Textiles, Leather)	0.75	0.50	0.74	0.005	−0.002
(Wood, Paper, Printing)	0.88	0.82	0.37	0.016	−0.007
(Coke, Hard-coal, chemicals)	0.77	0.65	0.32	0.005	−0.002
(Non-metallic mineral)	0.83	0.76	0.34	0.010	−0.004
(Metal Products)	0.91	0.65	0.56	0.008	−0.003
(Machinery, Electrical equipment)	0.65	0.22	0.93	−0.009	0.004
(Transport Equipment)	0.76	0.53	0.62	0.004	−0.002
(Furniture, Other manufacturing)	0.54	0.36	0.64	0.001	0.000
Value of Construction Completed	0.50	0.60	−0.07	0.005	−0.002
Power Sold	0.86	0.77	0.31	0.010	−0.004
Index of Services	0.86	0.77	0.31	0.010	−0.004
(Wholesale and Retail Trade)	0.95	0.89	0.45	0.697	−0.288
(Accommodation and Food Service)	0.92	0.89	0.41	0.081	−0.033
(Transportation)	0.93	0.67	0.63	0.013	−0.006
(Real estate, Renting, Leasing)	0.73	0.82	0.26	0.013	−0.006
(Business Support)	0.93	0.85	0.47	0.065	−0.027
(Education)	0.82	0.80	0.38	0.015	−0.006
(Arts, Sports, Recreation)	0.88	0.85	0.42	0.030	−0.013
(Repair and Other personal services)	0.87	0.90	0.32	0.048	−0.020
Retail Sales	0.92	0.88	0.38	0.046	−0.019
Equipment Investment Index	0.89	0.60	0.75	0.020	−0.008
BOP Export(FOB)	0.22	−0.12	0.53	−0.002	0.001
BOP Import(FOB)	0.91	0.61	0.67	0.009	−0.004
Producer's Inventory	0.35	0.40	−0.22	0.003	−0.001

여기서 공통인자는 M개라고 가정하였으며 $\hat{\Lambda}_j$는 인자적재 추정치이고 $\hat{\psi}_j$는 특정인자의 분산추정치이다.

인자모형에서는 GDP와 연관성 있는 경제지표를 가능한 많이 모형에 포함시킬 필요가 있다. 여기서는 분기 GDP 추계시 주요 지표로 이용되는 산업생산지수(제조업 및 9개 하위부문 지수), 서비스업생산지수(총지수 및 8개 하위부문 지수), 건설기성액, 전력판매량, 소비재판매액지수, 설비투자추계지수, 상품수출입, 생산자재고지수 등 총 26개 지표를 선정하였다. 회귀모형 경우와 마찬가지로 서비스업생산지수, 건설기성액, 설비투자추계지수, 소비재판매액지수 등 최초 공표시점이 늦어 발생하는 결측치는 EM 알고리즘을 이용한 추정치를 이용하였다.

공통인자는 2개를 선택하여 변수들의 누적설명력이 약 78%가 되도록 하였으며 베리멕스 회전을 통해 인자구조를 단순화하였다. 공통인자에 대응하는 각 변수들의 상대적 영향력을 나타내는 인자적재는 최대우도법(Maximum Likelihood Method)을 적용하였으며 추정결과는 Table 4.1과 같다. 공통인자 1은 서비스업을 중심으로 소비, 투자, 수입의 전반적인 움직임을, 공통인자 2는 제조업생산 및 수출의 전반적인 움직임을 나타내는 경기지표로의 특징을 보인다고 할 수 있다. 산출된 2개의 공통인자를

설명변수로 이용한 분기 GDP추정식은 식 (4.2)와 같다.

$$\begin{aligned} \mathrm{GDP}_t^F &= 0.910\,\mathrm{GDP}_{t-1}^F + 0.825\,\mathrm{FAC1}_t - 0.677\,\mathrm{FAC1}_{t-1} \\ &\qquad (9.82) \qquad\qquad (16.30) \qquad\quad (-7.53) \\ &\quad + 0.626\,\mathrm{FAC2}_t - 0.559\,\mathrm{FAC2}_{t-1} + \mathrm{DUM}, \\ &\qquad (13.90) \qquad\quad (-8.48) \end{aligned} \tag{4.2}$$

여기서 GDP^F는 평균 0, 분산 1이 되도록 표준화한 분기 실질 GDP 전년동기비이며 FAC1과 FAC2는 각각 공통인자 1과 공통인자 2를 나타내며 DUM은 이상치 조정항이다. ( )안은 계수들에 대해 t-통계량을 나타낸다. 인자모형에서의 결정계수 R^2은 0.972로 매우 높은 모형적합도를 가지며 DW-통계량 값도 2.185로 오차의 독립성가정을 만족하고 있다.

4.3. 베이지안VAR모형에 의한 GDP추정

베이지안VAR모형은 구조방정식과 시계열모형이 결합된 형태인 다음의 VAR모형에 베이지안 확률이론을 근거로 모수의 확률분포를 가정한 계량모형이다.

$$y_t = \mu_t + B_1 y_{t-1} + \cdots + B_K y_{t-K} + \epsilon_t,$$

여기서 $y_t^T = (y_{1t}, \ldots, y_{pt})$, $\mu_t^T = (\mu_{1t}, \ldots, \mu_{pt})$이며 B_i는 시차가 i인 계수행렬, $\epsilon_t^T = (\epsilon_{1t}, \ldots, \epsilon_{pt})$이다. VAR모형은 이용되는 변수의 갯수와 동일한 갯수의 방정식으로 이루어진다는 특성을 가지고 있다. 첫번째 방정식의 종속변수y_{1t}를 분기 GDP 전년동기비로 놓고 나머지 설명변수는 제조업생산지수(IPI), 전력판매량(ELE), 비농가취업자수(EMP), 서비스업생산지수(SP) 등 4개 주요 지표를 이용하였다. 총 5개 방정식으로 이루어진 VAR모형에서 시차항계수의 사전분포가 0이고 시차가 길어질 수록 표준편차가 작아진다는 베이지안 개념을 추가하였다. 보다 구체적으로 계수행렬에 대한 사전분포(prior distribution)는 서로 독립인 정규분포를 따른다고 가정하고 1시차변수의 계수가 평균 1, 다른 시차에서는 0인 분포를 따른다고 가정하였다. 이는 종속변수의 대다수 변동이 자기시차계수에 의해 설명되는 현실적 전제를 반영하고 있다. 사전분포의 표준편차를 다음의 3개 초모수(hyperparameter)의 함수형태로 설정하였다.

$$S(i,j,k) = \lambda \cdot g(k) f(i,j) \frac{S_i}{S_j}, \quad i,j = 1,\ldots,n,\ k = 1,\ldots,K,$$

여기서 S_i는 i번째 방정식에서 일변량자귀회귀식의 표준오차이며 상이한 변수들의 크기를 조정할 목적으로 S_i/S_j를 이용한다. 이 때 $\lambda \cdot g(k)f(i,j)$은 각 계수에 부여되는 사전확률에 대한 가중치를 의미한다. 여기서 λ는 전체적인 가중치(overall tightness)를 나타내는 상수이다. 첫번째 시차와 비교한 k번째 시차의 상대적 중요도를 나타내는 함수는 $g(k) = k^{-1}$로 설정하였다. $f(i,j)$는 i번째 방정식에서 종속변수와 비교한 j번째 변수의 상대적 중요도를 나타내며 $i = j$인 경우에만 1의 값을 가지고 나머지는 상수 w값을 갖는다. 이 때 상수 λ와 w는 별도의 모의실험을 통하여 결정하였다. 두 상수 값을 각각 0.1부터 2.0까지 0.1씩 변경시키면서 1992년 1/4분기∼ 1999년 4/4분기 시계열을 대상으로 VAR모형을 적합한 후 2000년 1/4분기∼2006년 4/4분기까지 자료를 추가해 가면서 표본외 1시차 후 예측을 실시하여 Theil의 U-통계량 기준으로 예측오차를 가장 작게 하는 값으로 선정하였다. 또한 시차길이는 BIC통계량 값이 최소가 되는 6분기로 선택하였다. 베이지안VAR모형에 의한 GDP추정결과는 Table 4.2와 같다. GDP의 1시차계수가 1에 근사하고 다른 시차에서는 0에 근사하여 사전분포의 가정을 만족하고 있으며, DW-통계량 값도 1.954로 오차의 독립성가정을 만족하고 있다.

Table 4.2. Estimation result of Bayesian VAR model coefficients

	lag 1	lag 2	lag 3	lag 4	lag 5	lag 6
GDP^B	0.951	−0.109	−0.122	0.004	0.046	0.028
	(7.03)	(−0.76)	(−0.85)	(0.03)	(0.32)	(0.20)
IPI	0.001	−0.011	−0.006	0.014	0.024	−0.003
	(0.02)	(−0.23)	(−0.12)	(0.28)	(0.50)	(−0.07)
ELE	−0.002	−0.024	−0.001	−0.001	−0.010	0.022
	(−0.04)	(−0.38)	(−0.01)	(−0.01)	(−0.18)	(0.39)
EMP	0.023	−0.080	−0.082	−0.167	0.080	−0.007
	(0.15)	(−0.51)	(−0.52)	(−1.06)	(0.51)	(−0.05)
SP	0.108	−0.005	−0.007	−0.011	0.001	−0.039
	(1.47)	(−0.06)	(−0.10)	(−0.14)	(0.01)	(−0.53)
Constant	1.183					
	(1.34)					

Note: Numbers in parentheses are t statistics.

Table 4.3. Comparison of single and combined model's estimation accuracy

		Combined				Single		
		GDP^{RF}	GDP^{RB}	GDP^{FB}	GDP^{RFB}	GDP^R	GDP^F	GDP^B
MAE	VM	0.363	0.392	0.361	0.347	0.425	0.354	0.442
	RB	0.350	0.344	0.321	0.329			
RMSE	VM	0.460	0.508	0.462	0.444	0.536	0.477	0.555
	RB	0.423	0.422	0.416	0.407			

Note: VM means Variance Minimization method and RB means Regression based method

4.4. 개별모형 및 결합모형의 추정력 비교

여기서는 3개 개별모형을 결합하여 만든 결합모형과 개별모형의 추정력을 비교하였다. 분산최소기법 및 회귀모형기법을 이용하여 2개 및 3개 개별모형 추정치를 결합하여 결합모형 추정치로 산출하였다. 개별모형 및 결합모형에 대해 시험자료인 2002년 1/4분기부터 2006년 4/4분기까지 총 20개 분기를 대상으로 역사적 모의실험을 실행하여 추정치를 계산하였다. 먼저 실질 GDP 전년동기비의 실적치와 추정치 간 차이의 절대값 평균오차(MAE)와 제곱근평균제곱오차(RMSE) 등 각 모형별 오차통계량을 계산하여 비교하였다. 회귀모형, 인자모형, 베이지안VAR모형 추정치를 각각 GDP^R, GDP^F, GDP^B로 표시하고 결합모형은 위첨자로 구분하여 표시하였다. 예를들어 회귀모형과 인자모형의 결합모형은 GDP^{RF}로 3개 모형을 모두 결합한 모형은 GDP^{RFB}로 표시하였다.

Table 4.3에서 모형별 오차통계량을 계산하여 비교한 결과 대부분의 결합모형 추정오차가 개별모형에 비해 작게 나타났다. 다만 인자모형 GDP^F의 오차가 회귀모형과 베이지안VAR모형을 분산최소기법에 의해 결합한 GDP^{RF}의 오차에 비해 오히려 작게 나타나는 경우도 있었다. 결합기법별로는 회귀모형기법(RB)이 분산최소기법(VM)에 비해 오차가 모두 작게 나타났다. MAE 기준으로 가장 우수한 모형은 각각 인자모형과 베이지안VAR모형의 회귀모형기법 결합인 GDP^{FB}이었다. 또한 RMSE 기준으로는 3개 모형을 회귀모형기법에 의해 결합한 GDP^{RFB}가 가장 우수한 것으로 나타났다.

이번에는 HLN검정을 통해 결합모형의 추정력이 통계적으로 우월한지를 평가하였다. Table 4.4에서 분산최소기법에 의한 결합모형은 총 9번의 검정가운데 5번이 개별모형에 비해 추정력이 우수한 것으로 나타났다. 하지만 인자모형의 경우에는 어떤 결합모형도 인자모형보다 통계적으로 유의하게 추정력이 우

Table 4.4. Comparison of Combined Model using HLN Test

	Combined Model	RMSE	HLN Test: Comparable Model	HLN Test: Test Statistics
VM	$0.482\,\mathrm{GDP}^R + 0.518\,\mathrm{GDP}^F$	0.460	GDP^R	-1.659^*
			GDP^F	-0.333
	$0.553\,\mathrm{GDP}^R + 0.447\,\mathrm{GDP}^B$	0.508	GDP^R	-0.730
			GDP^B	-3.803^{**}
	$0.587\,\mathrm{GDP}^F + 0.413\,\mathrm{GDP}^B$	0.462	GDP^F	-0.199
			GDP^B	-3.731^{**}
	$0.122\,\mathrm{GDP}^R + 0.609\,\mathrm{GDP}^F + 0.269\,\mathrm{GDP}^B$	0.444	GDP^R	-1.469^*
			GDP^F	-0.803
			GDP^B	-3.731^{**}
RB	$0.545 + 0.503\,\mathrm{GDP}^R + 0.413\,\mathrm{GDP}^F$	0.423	GDP^R	-1.793^{**}
			GDP^F	-1.116
	$0.777 + 0.514\,\mathrm{GDP}^R + 0.363\,\mathrm{GDP}^B$	0.422	GDP^R	-1.451^*
			GDP^B	-4.658^{**}
	$0.525 + 0.487\,\mathrm{GDP}^F + 0.425\,\mathrm{GDP}^B$	0.416	GDP^F	-1.100
			GDP^B	-3.519^{**}
	$0.577 + 0.281\,\mathrm{GDP}^R + 0.335\,\mathrm{GDP}^F + 0.292\,\mathrm{GDP}^B$	0.407	GDP^R	-1.828^{**}
			GDP^F	-1.241
			GDP^B	-4.489^{**}

Note: *, ** means that Combined Model is superior to comparable model with significant level 10%, 5%, respectively

수하다고 결론내릴 수는 없었다. 한편 회귀모형기법에 의한 결합모형은 총 9번의 검정가운데 6번이 개별모형에 비해 우수한 것으로 나타나 분산최소기법과 비슷한 결과를 나타내었다. 회귀모형기법의 검정통계량이 분산최소기법 경우와 비교해 상대적으로 임계치에 보다 근접하여 두 개의 결합기법 가운데 더 나은 결과를 보일 수 있다는 가능성은 확인할 수 있었다. 이는 회귀모형기법 결합이 상수항을 반영함으로써 편의(bias)를 제거하는 특징이 반영된 것으로 추측할 수 있다. 회귀모형기법에 의한 결합모형의 추정력 검정에서도 인자모형 보다 우수한 모형은 나타나지 않았는데 이는 인자모형에 포함된 변수가 생산측면 뿐만 아니라 수요측면 변수까지도 모두 포함한 것이 모형의 추정력을 일정부분 높인 것으로 예상할 수 있다. 하지만 회귀모형 및 베이지안VAR모형을 개별모형으로 선정하였을 경우에는 개별모형보다는 결합모형의 추정력이 통계적으로 유의하게 개선되었음을 확인할 수 있었다.

5. 맺음말

본 연구에서는 분기 GDP추정을 위해 회귀모형, 인자모형, 베이지안VAR모형 등을 이용하여 산출한 개별추정치와 이를 결합한 결합추정치간 추정력을 비교해 보았다. 개별모형에서는 인자모형의 추정력이 가장 우수한 것으로 결론 지을 수 있는데 이는 인자모형이 많은 정보를 축약한 공통인자를 이용한다는 점이 GDP통계가 방대한 자료를 이용하여 추계된다는 점과 유사한 특징을 지닌 것으로 볼 수 있기 때문이다. 한편 결합모형과 개별모형의 추정력 평가에서는 개별모형에 따라 상반된 결과가 나왔다. 회귀모형, 베이지안VAR모형의 경우는 개별모형보다는 2개 또는 3개 모형을 결합한 모형의 추정력이 통계적으로도 개선되는 것으로 나왔다. 이는 동일한 변수를 이용하는 서로 다른 모형의 경우에 결합모형이 유용하다는 것을 보여준다. 하지만 인자모형과 같이 이용되는 설명변수의 개수가 크게 달라지는 경우에는

결합모형의 추정력 개선 여부는 확신하기 어렵다는 것을 간접적으로 나타내고 있다.

각종 경제지표를 설명변수로 하고 이를 이용하여 보다 정확한 GDP추정을 위해 결합기법이 유용한지에 대해 분석하는 것은 GDP를 신속하고 간편하게 가늠할 수 있다는 장점이 있다. 또한 최근 시점의 경제지표가 입수 되기 전에 GDP를 예측하는 예측치의 결합에 대한 연구도 향후 경기판단에 유용할 것으로 보인다. 본 연구에서는 3개 모형을 이용하였으나 새로 개발되는 추정모형을 포함하여 결합하는 개별모형의 수를 더 늘렸을 때의 결합모형 추정력을 평가하는 등 우리나라 분기 GDP추정에 활용가능성이 높은 조합을 산출하기 위한 지속적인 연구가 필요할 것으로 보인다.

References

Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts, *Operational Research Quarterly*, 451–468.

Collopy, F. and Armstrong, J. S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations, *Management Science*, **38**, 1394–1414.

Diebold, X. (1988). Serial correlation and the combination of forecasts, *Journal of Business Economic Statistics*, **6**, 105–111.

Granger, C. W. J. and Ramanathan, R. (1984). Improved methods of combining forecasts, *Journal of Forecasting*, **3**, 197–204.

Harvey, D. I., Leybourne, S. J. and Newbold, P. (1997). Testing the equality of prediction mean square errors, *International Journal of Forecasting*, **13**, 281–291.

Min, K. S., Park, J. H. and Park, S. O. (2002). Quaterly Projection Model by using monthly indicators, *Journal of The Korean Official Statistics*, **7**, 97–126.

Newbold, P. and Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts, *Journal of the Royal Statistical Society*, **137**, 131–164.

Reijer, A. H. J. (2005). *Forecasting Dutch GDP using large scale factor models*, Working paper, 1–27.

Schneider, M. and Spitzer, M. (2004). *Forecasting Austrian GDP using the generalized dynamic factor model*, Working paper, 1–36.

Shim, S. and Lee, H. (1992). Short-term Forecasting Domestic Demand using the monthly economic indicators, *KDI Quarterly Economic Outlook*, 63–75.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors, *Journal of American Statistical Association*, **97**, 1167–1179.