

# 결합추정기법을 이용한 분기 GDP 추정\*

분기 GDP를 추정하기 위한 다양한 계량모형이 연구·개발되어 활용되고 있다. 하지만 모형에 의한 개별 추정치는 서로 다른 이론적 배경에 기초하고 있어 동일 경제현상에 대해 서로 다른 추정치를 제공할 수 있다. 따라서 하나의 특정모형에만 전적으로 의존하는 것보다는 여러 모형에 의한 추정결과를 결합하여 보다 많은 정보를 활용하는 결합추정기법을 이용해 분기 GDP를 추정한다면 개별모형 추정치에 비해 정도 높은 추정치를 얻을 것으로 기대할 수 있다.

본고에서는 자기회귀분포시차모형, 인자모형, 베이지안 VAR모형 등 주요 계량모형을 이용하여 분기 GDP를 추정한 다음, 동 추정결과들을 결합하여 GDP를 추정하는 결합추정기법을 적용한 결과에 대하여 추정력의 개선여부를 평가해 보았다.

실증분석 결과 개별모형 추정치보다 결합추정치의 정도가 보다 개선된 것으로 나타났으며 특히 인자모형과 베이지안 VAR모형을 이용한 결합추정치가 가장 개선된 추정 결과를 나타내었다.

## I. 머리말

## III. 실증분석

## II. 분기 GDP 추정

## IV. 맺음말

1. 추정방법
  - 가. ADL모형
  - 나. 인자모형
  - 다. BVAR모형
  - 라. 결합추정기법
2. 추정력 검증

\* 본고는 경제통계국 국민소득팀 강창구 과장이 집필하였음. 본고의 내용은 집필자의 개인의견으로서 한국은행의 공식견해를 나타내는 것은 아님

## I. 머리말

경제주체가 의사결정을 할 때는 현재 경기를 정확히 판단하는 것이 무엇보다 중요하다. 한국은행에서는 경기판단에 이용되는 대표적 지표인 국내총생산(GDP)을 분기별로 해당분기 종료 후 28일 경에 속보치로, 70일 이내에 잠정치로 각각 공표하고 있다. 분기 GDP 잠정치 추계는 300여종에 이르는 방대한 기초자료를 수집하여 먼저 기초자료의 오류확인, 상품기준의 산출액, 중간투입 및 부가가치 추계, 경제현실과의 적정성 검토 등의 과정을 거쳐 최종 확정되기까지 많은 시간과 인력이 필요한 것이 사실이다.

따라서 경제지표의 입수와 동시에 실시간으로 GDP를 추정하기 위한 다양한 모형이 개발되고 있다. 민경삼·박진호·박순옥(2002)은 공급부문과 수요부문별 GDP를 추정하는 자기회귀분포모형(이하 ADL모형)을 기반으로 한 회귀모형을 구축하였다. Reijer(2005), Schneider와 Spitzer(2004)는 자국의 GDP를 추정하기 위해 인자모형을 이용한 추정모형을 개발하였다. 심상달·이항용(1992)은 내수(內需) 추정을 위해 베이지안 VAR모형(이하 BVAR모형)과 ADL모형을 구축하였다.

하지만 계량모형은 중장기적 추세를 벗어난 불규칙성을 정확히 추출해 내는 데 현실적으로 한계가 있을 수밖에 없다. 각 계량모형은 서로 다른 이론적 배경과 기본가정을 토대로 설정되어 있어, 변수들 사이에 존재하는 서로 다른 정보를 기초로 추정하며 이용

---

되는 변수들도 서로 다를 수가 있다. 어떤 계량모형을 이용하여 추정하는가에 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.

본고에서는 먼저 ADL모형, 인자모형, BVAR모형에 의해 분기 GDP를 추정하는 방법을 소개하고, 추정오차의 분산최소기법 및 회귀모형기법 등과 같은 결합추정기법을 간략히 설명하였다. 실증분석에서는 동 방법들을 이용하여 우리나라 분기 GDP를 추정해보았다.

## II. 분기 GDP 추정

### 1. 추정방법

여기에서는 분기 GDP를 추정하기 위한 다양한 계량방법 가운데 ADL모형, 인자모형, BVAR모형과 결합추정기법에 대해 간략히 정리하였다.

#### 가. ADL모형

ADL모형은 종속변수와 설명변수의 시차에 따른 움직임을 반영하고, 모형 설명력을 충분히 확보할 수 있는 한도 내에서 모수의 개수를 최소화 시키는 방법으로 기본 모형식은 다음과 같다.

$$A(L)y_t = \sum_{i=1}^p B_i(L)x_{it} + \varepsilon_t, \quad t = 1, \dots, T$$

여기서  $y_t$ 는 로그 차분한 분기 실질 GDP(또는 분기 실질 GDP 전년동기비),  $x_{it}$ 는 로그 차분한  $p$ 개 분가지표(또는 분가지표의 전년동기비)를 나타내며,  $A(L)$ 과  $B_i(L)$ 은  $y_t$ 와  $x_{it}$ 의 시차구조를 반영한 시차다항식,  $\varepsilon_t$ 는 백색잡음이다.

$\hat{B}_i(L)$ 을 오차항이 자기상관을 갖지 않는다는 조건에서 통상최소제곱법에 의해 추정한 시차다항식이라고 할 때, 1시차 후 예측

---

치(one-step ahead forecast)  $\widehat{y}_{t+1}$ 는 다음과 같이 산출된다.

$$\widehat{y}_{t+1} = E[y_{t+1}|y_t] = \sum_{i=1}^p \widehat{B}_i(L) x_{it+1}$$

ADL모형에서는 관련지표 및 시차변수를 초기모형식에 가능한 많이 포함시킨 후 통계적 검정을 거쳐 유의성이 떨어지는 지표 및 시차변수를 하나씩 제거하는 모수 축약방법(general to specific approach)을 주로 이용한다.

## 나. 인자모형

인자모형은 다수의 변수들 간에 존재하는 상호연관성을 미지의 잠재적인 공통인자(common factor)로 집약하여 해석하는 통계적 방법이다. 인자모형에서는 전체 변수들 사이에 공통인자가 존재한다고 가정하고, 변수들의 상관구조를 설명할 수 있는 소수의 잠재적 인자로 축약하여 회귀모형 등 추가적인 분석을 위한 설명변수로 이용할 수 있다. 따라서 GDP를 한 나라의 총체적 경제수준을 나타내는 지표라고 정의할 때 경제활동의 제(諸) 영역을 설명하는 다양한 경제지표에 공통적으로 내재되어 있는 인자를 추출하고 이를 이용해 GDP의 움직임을 포착하는 것이 가능하게 된다.

인자모형에서는 상호 연관성을 가진  $p$ 개의 변수( $y_{it}$ )가 변수들 사이에 공통적으로 영향을 미치는  $m$ 개 공통인자( $F_{jt}$ )와 개별변수에만 영향을 미치는 특정인자( $\varepsilon_{it}$ : specific factor)로 구성된다고 가정한다.

$$y_{it} = \Lambda'_i F_{jt} + \varepsilon_{it}, \quad i = 1, \dots, p, \quad j = 1, \dots, m$$

여기서  $\Lambda_i$ 는 인자적재(factor loading)로서 변수  $y_{it}$ 에 대한 공통인자  $F_{jt}$ 의 중요도를 나타내는 가중치이다. 공통인자가  $M$ 개인 경우 인자적재를 가중치로 한 각 시점별 공통인자 추정치  $F_{jt}$ 는 다음과 같이 산출된다.

$$\hat{F}_{jt} = \hat{\Lambda}'_j (\hat{\Lambda}_j \hat{\Lambda}'_j + \hat{\psi}_j)^{-1} y_t, \quad j = 1, \dots, M \quad t = 1, \dots, T$$

여기서  $\hat{\Lambda}_j$ 는 인자적재 추정치,  $\hat{\psi}_j$ 는 특정인자의 분산추정치이다.

#### 다. BVAR모형

BVAR모형은 구조방정식과 시계열모형이 결합된 형태인 다음의 VAR모형에 베이시안 확률이론을 근거로 모수의 확률분포를 가정한 계량모형이다.

$$y_t = B_0 + B_1 y_{t-1} + \dots + B_K y_{t-K} + \omega_t$$

여기서  $y_t$ 는  $p$ 개 변수의 시계열 벡터,  $B_i, i = 1, \dots, K$ 는 시차  $i$ 의 계수행렬,  $\omega_t$ 는 백색잡음을 나타낸다. 모형에 포함되는 변수 및 시차구조의 설정에 따라 추정치가 민감하게 반응하고 과다모수(over-parameterization)로 인한 예측력 저하 등과 같이 VAR모형에서 빈번히 발생하는 문제점을 해결하고자 BVAR모형에서는 모수에 Litterman(1986)이 제시한 베이시안 사전확률(Bayesian Prior) 개념을 도입하였다. Litterman의 기본개념은 모수를 일정한 평균치를 갖는 확률변수로 간주하고 초모수(hyperparameter)를 이

용하여 사전평균(prior mean)에 대한 확률분포의 응집성(tightness)<sup>1)</sup>을 통제하는 데 있다(성병희, 2001). 즉, 개별 시차변수의 계수행렬  $B_i$ 에 대한 사전확률이 서로 독립인 정규분포를 따른다고 가정하고, 1시차변수의 계수가 평균이 1이고 다른 시차에서는 0인 분포를 따른다는 제약조건을 부여함으로써 종속변수의 대다수 변동이 자기시차계수에 의해 설명되는 현실적 전제를 반영하고 있다. 사전분포의 표준편차는 다음과 같이 3개의 초모수의 함수형태로 설정한다.  $i$ 번째 방정식,  $j$ 번째 변수,  $k$ 번째 시차에 대한 사전분포의 표준편차  $S(i, j, k)$ 는

$$S(i, j, k) = \lambda f(i, j) g(k) \frac{S_i}{S_j} \dots\dots\dots (1)$$

이며, 여기서  $S_i$ 는  $i$ 번째 변수의 표준편차이다. 초모수  $\lambda$ 는 전체적인 응집성,  $f(i, j)$ 는  $i$ 번째 방정식에서  $j$ 번째 변수의 응집성,  $g(k) = k^{-d}$ ,  $d = 1, 2$ 는 시차  $k$ 에 따른 응집성의 감소정도를 나타낸다.

BVAR모형에서 초모수는 예측력을 좌우하는 중요한 역할을 담당한다. 일반적으로 Theil의 U-통계량<sup>2)</sup> 값을 이용하여 예측력을 상호 평가함으로써 초모수를 선정하는 방식이 주로 이용된다. 시차의 길이는 BIC(Bayesian Information Criterion) 통계량을 이용하여 결정한다.

1) 표준편차와 대응되는 되는 개념으로 표준편차가 클 수록 응집성은 떨어지고 표준편차가 작을 수록 응집성은 커짐

2)

$$U = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}}{\sqrt{\frac{1}{N} \sum_{t=1}^N y_t^2} + \sqrt{\frac{1}{N} \sum_{t=1}^N \hat{y}_t^2}}, \quad y_t: \text{실적치}, \hat{y}_t: \text{예측치}, N: \text{예측기간}$$

## 라. 결합추정기법

동일한 경제현상에 적용되는 계량모형에 따라 산출되는 추정결과에는 차이가 발생할 수 있다. 이와 같은 차이는 적용되는 모형별 고유특성에 따라 자료로부터 확보되는 정보가 서로 다르기 때문이다. 따라서 단일모형에 의존하여 자료의 특정한 성질만을 유도해 내는 것보다는 다양한 계량모형을 적용해 보고 각 모형에서 제공하는 부수적인 정보를 종합한다면 보다 우월한 분석이 가능하다고 유추할 수 있을 것이다.

여기서는 분산최소기법(variance covariance method: Bate, Granger, 1969)과 회귀분석기법(regression based method: Granger and Ramanathan, 1984) 등 주요 결합기법을 소개하기로 한다(이종원 외, 1999).

분산최소기법은 개별추정치 오차가 불편성을 가지며 분산이 안정적일 때 결합추정치의 평균제곱오차가 개별추정치의 평균제곱오차보다 작게 된다는 점에 근거하고 있다.  $y_t$ 가 분기 GDP(전년 동기비)의 실적치,  $p_{1t}$ 와  $p_{2t}$ 가 서로 다른 두 가지 계량모형에 의해 추정한 분기 GDP(전년동기비)라고 하자.

$$y_t = p_{1t} + \epsilon_{1t}$$

$$y_t = p_{2t} + \epsilon_{2t}$$

여기서  $\epsilon_{1t}$ ,  $\epsilon_{2t}$ 는 평균 0이고 분산  $\sigma_1^2 = E(\epsilon_{1t}^2)$ ,  $\sigma_2^2 = E(\epsilon_{2t}^2)$ , 공분산  $\sigma_{12}$ 를 갖는 추정오차이다. 결합추정치  $c_t$ 를  $p_{1t}$ 와  $p_{2t}$ 의 가중치  $\lambda_1$



과  $\lambda_2$ 를 이용한 선형결합  $c_t = \lambda_1 p_{1t} + \lambda_2 p_{2t}$  라고 할 때 결합추정치  $c_t$ 의 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  $\lambda_1 + \lambda_2 = 1$ 를 만족하는 결합추정치  $c_t$ 의 오차분산은

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

이다. 이 때  $\sigma_c^2$ 를 최소화하는  $\lambda_1$ 과  $\lambda_2$ 는 다음과 같이 유도된다.

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

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

Newbold and Granger(1974)는 2개 추정치를 결합하는 Bates의 (1969) 방법을 일반화하여  $m$ 개의 추정방식으로 다음과 같이 확장하였다. 실적치  $y_t$ 에 대해  $m$ 개의 추정치  $p_{jt}$ , ( $j=1, \dots, m$ )가 존재한다고 가정할 때  $j$ 번째 추정치에 대한 오차는  $\epsilon_{jt} = y_t - p_{jt}$ 이다. 결합추정치는

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

이며, 오차분산을 최소화하는 가중치 벡터는

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

가 된다. 여기서  $\Sigma$ 는 추정오차의 공분산행렬,  $l$ 은 길이가  $m$ 이며

값이 모두 1인 벡터이다.

Granger와 Ramanathan(1984)은 회귀모형의 개념에 의해 실적치와 결합추정치 오차가 최소화되는 가중치를 산출하는 방식을 제안하였다.  $y^T = (y_1, \dots, y_n)$ 을 분기 GDP의  $n$ 개 관측치 벡터,  $p_j^T = (p_{j1}, \dots, p_{jn})$ ,  $j = 1, \dots, m$ 는  $j$ 번째 모형에 의해 산출한 시점별 추정치 벡터라고 할 때  $P = (p_1, \dots, p_m)$ 는  $n \times m$  행렬이다. 결합추정치  $c = (c_1, \dots, c_n)$ 는 개별추정치에 가중치  $\lambda = (\lambda_1, \dots, \lambda_m)$ 을 각각 곱한 후 상수항  $\lambda_0$ 을 반영한<sup>3)</sup> 다음과 같은 회귀모형으로 나타낼 수 있다.

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

여기서  $\epsilon$ 는 백색잡음,  $l$ 은 길이가  $m$ 이며 값이 모두 1인 벡터이다. 가중치  $\lambda$ 는 통상 최소제곱법에 의해 다음과 같이 산출된다.

$$\hat{\lambda} = \hat{\alpha} - \hat{\lambda}_0 (P^T P)^{-1} P^T l$$

$$\hat{\lambda}_0 = (l^T y - l^T P \hat{\lambda}) / n = l^T \hat{e} / (n - \hat{\theta})$$

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 추정 오차벡터를 나타낸다. 따라서 결합추정치  $\hat{c}$ 은

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

에 의해 산출된다.

3) 상수항이 있는 결합추정치는 평균제곱오차를 최소화하며 개별 추정치가 편향(biased) 추정치 이더라도 결합추정치는 비편향(unbiased) 추정치임 (Granger and Ramanathan, 1984)

## 2. 추정력 검정

본고에서는 ADL모형, 인자모형, BVAR모형 등 개별 GDP 추정치와 결합기법에 의한 추정치의 상대적 정확성을 검정하기 위해 Harvey외(1997)의 검정 방법(이하 HLN검정)을 이용하였다.

서로 다른 두 가지 추정방법에 의한 추정치와 실적치간의 오차를 각각  $e_{1t}$ ,  $e_{2t}$  ( $t=1, \dots, n$ )라고 할 때 HLN검정의 귀무가설

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

에 대한 검정 통계량  $S$ 는

$$S = \bar{d} \cdot \hat{V}(\bar{d}) \cdot \left( \frac{n-1}{n} \right)^{-\frac{1}{2}}$$

이다. 여기서  $\bar{d} = \frac{1}{n} \sum_{t=1}^n (g(e_{1t}) - g(e_{2t}))$ ,  $\hat{V}(\bar{d}) = \frac{1}{n} \hat{\gamma}_0$  이며,  $\hat{\gamma}_0$ 는  $\bar{d}$ 의 자기공분산이다. 검정통계량  $S$ 가 자유도  $n-1$ 인 t-분포의 임계치보다 작은 값을 갖는다면 모형 2가 모형 1에 비해 통계적으로 추정력이 우수한 것으로 판단한다.

계량분석에서는 시점별로 추정치의 정확성뿐만 아니라 최근 시점에서의 상승 하강 국면을 신속히 파악할 수 있는 방향성 또한 확보될 필요가 있다. 추정치와 실적치간 증감 방향의 일치여부를 검정하기 위해 Pesaran과 Timmermann(1992)은 다음의 검정방법(이하 PT검정)을 제안하였다. 검정통계량  $S_n^2$ 은

$$S_n^2 = \frac{(\hat{p} - p^*)^2}{V(\hat{p}) - V(p^*)}$$

이며,  $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$ 이며  $X_i$ 는 추정치의 증감방향이 실적치와 같은 경우 1, 다를 때는 0의 값을 가지는 이항변수이다.  $p^*$ 는 실적치와 추정치가 서로 독립이라는 가정하에서  $p$ 의 기대값이며,  $V(\hat{p})$ ,  $V(p^*)$ 는 각각  $\hat{p}$ ,  $p^*$ 의 분산을 나타낸다. 검정통계량  $S_n^2$ 가 자유도  $n$ 인 카이제곱분포에서의 임계치보다 큰 값을 갖는 경우에 추정치의 증감방향(+, -)이 통계적 관점에서 실적치와 일치한 것으로 판단한다.

### Ⅲ. 실증분석

여기에서는 결합기법에 의한 추정력 개선효과를 평가해 보기 위해 II장에서 제시한 ADL모형, 인자모형 및 BVAR모형 등 3가지 계량모형과 2가지 결합기법을 이용해 우리나라 분기 실질 GDP를 각각 추정해 보았다.

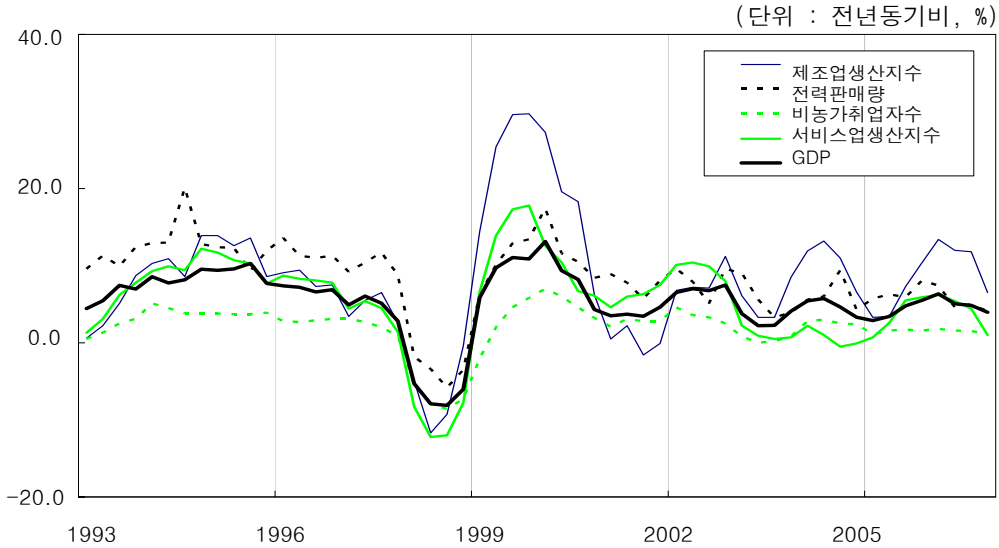
#### 가. ADL모형

ADL모형에서는 생산측면 분기 GDP 추계시 사용되는 주요지표를 중심으로 제조업생산지수, 서비스업생산지수(총지수), 전력판매량, 건설기성액 및 비농가 취업자수 등 5개 지표를 후보변수로 선정하였다. 분석대상 시계열은 1992년 1/4분기~2006년 4/4분기로 설정하였으며 1999년 1월부터 공표된 서비스업생산지수 과거계열의 결측치는 통계적 기법<sup>4)</sup>에 의해 연장하여 이용하였다.

4) Stock, Waton(2002)의 EM알고리즘을 이용하였으며, 동 방법은 실질 GDP를 포함한 우리나라 경제의 제 부문을 설명하는 26개 주요지표의 정보를 이용하여 서비스업생산지수의 결측치를 추정

<그림 1>

GDP 및 주요지표의 분기 성장률 추이 비교



분기 GDP와 각 분기 지표간의 개별 회귀분석을 통해 설명변수로서의 유용성을 평가해 본 결과, 건설기성액을 제외한 4개 지표가 GDP에 대해 유의한 설명력을 가지고 있으며, <그림 1>에서도 GDP 추이에 밀접한 관계가 있음을 확인할 수 있다. 특히 제조업생산지수가 평균제곱합(=21.6), 평방근평균제곱오차(=0.89) 기준으로 볼 때 GDP를 가장 잘 설명하고 있는 것으로 나타났다(<표 1> 참조). 따라서 가장 높은 설명력을 가진 제조업생산지수에 3개 변수(건설기성액 제외)들이 각각 부가적으로 설명하는 영향력을 평가해 보기 위해 encompassing 회귀분석<sup>5)</sup>을 실시해본 결과, 3개 변수 모두 GDP 추정에 유의한 부가정보를 제공하였다.

5)  $\hat{y}_{t,IP}$  을 제조업생산지수에 의한 GDP 추정치,  $\hat{y}_{t,j}$  을 기타 지표에 의한 GDP추정치라고 할 때, encompassing 회귀모형  $y_t = \lambda \hat{y}_{t,IP} + (1-\lambda) \hat{y}_{t,j}$  은 모수  $\lambda$ 가 1보다 작을 때 부가적인 정보를 제공하는 유의성 있는 변수로 판별

<표 1>

**GDP에 대한 분기지표의 영향력 평가**

	평균제곱합 (SSR)	평방근 평균 제곱오차 (RMSE)	F통계량 ( $H_0 : B(L) = 0$ )	(p값)	encompassing regression $\lambda$
제조업생산지수	21.6	0.89	28.5	(0.00)** <sup>1)</sup>	-
서비스업생산지수	37.8	1.18	14.3	(0.00)**	0.41
취업자수(비농림)	27.0	1.02	14.8	(0.00)**	0.40
전력판매량	45.5	1.27	13.4	(0.00)**	0.78
건설기성액(실질)	57.8	1.46	1.85	(0.17)	-

주: 1) \*\*는 유의수준 1%에서 통계적으로 유의함을 의미

GDP 및 4개 지표의 시차변수에 대해 단계별 변수선택과정<sup>6)</sup> (stepwise selection procedure)을 적용하여 산출된 최종 ADL 추정 모형은 다음과 같다.

$$\widehat{GDP}_t = 0.283 GDP_{t-1} + 0.234 IPI_t - 0.102 IPI_{t-1} + 0.183 ELE_t$$

(3.02)                      (6.11)                      (-2.24)                      (5.06)

$$+ 0.256 EMP_t + 0.302 SP_t - 0.172 SP_{t-1} + DUM$$

(2.84)                      (3.86)                      (-2.54)

여기서 GDP는 분기 실질 GDP, IPI는 분기 제조업생산지수, ELE는 전력판매량, EMP는 취업자수(비농가), SP는 서비스업생산지수의 전년동기비를 나타내며 DUM은 이상치 조정항이다. ( )내 숫자는 각 회귀계수들에 대한 t-통계량이다.

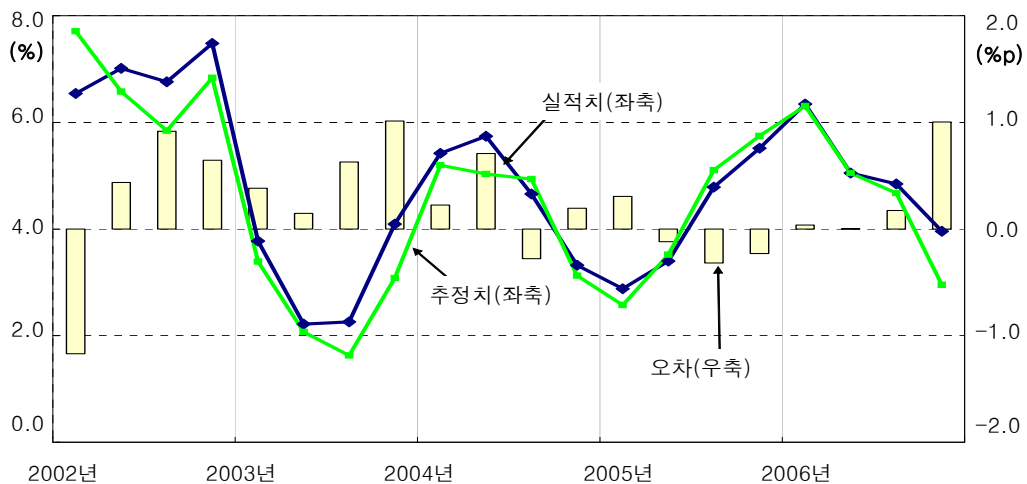
6) 종속변수에 영향을 미치는 다수의 변수중에서 어떤 변수들을 독립변수로 모형 속에 포함해야 하는가를 결정하기 위해 사용되는 변수선택방법으로 지나치게 많은 변수를 모형에 포함 시킴에 따라 발생할 수 있는 다중공선성문제를 해결하고, 결정계수 및 MSE 등 통계적 기준에서 최적모형을 설정하기 위해 사용

ADL모형에서의 결정계수는 0.973으로 매우 높은 모형적합도를 가지며 DW-통계량도 2.064로 오차의 독립성가정을 만족하고 있다. 또한 시차변수들의 부호(+, -)에서도 모형에서의 적절성을 확인할 수가 있다.

2002년 1/4분기부터 2006년 4/4분기까지 총 20개 분기를 대상으로 역사적 모의실험을 실행해 본 실적치와 추정치의 추이는 <그림 2>와 같다. 실적치와 추정치간 차이의 절대평균을 나타내는 MAPE 통계량은 0.4%p, 오차의 상대평균을 나타내는 RMSPE 통계량은 0.5%p이었다. 그러나 2002년 1분기(-1.2%p), 3분기(0.9%p), 2003년 4분기(1.0%p), 2006년 4분기(1.0%p) 등 일부 시점에서 지나치게 큰 오차가 발생하고 있어 안정적 추정에는 한계가 있는 것으로 보인다.

<그림 2>

ADL모형에 의한 분기 GDP 추정



## 나. 인자모형

다수의 변수들 간에 존재하는 공통적인 특성인 미지의 공통인자

를 추출한 후 이를 이용하는 인자모형의 특성을 충분히 활용하기 위해서는 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. 단 서비스업생산지수, 건설기성액, 설비투자추계지수, 소비재판매액지수 등 최초 공표시점이 늦어 발생하는 결측치<sup>7)</sup>는 EM알고리즘을 이용하여 추정하였다.

변수들의 누적설명력이 약 78%가 되도록 공통인자의 개수는 2개로 결정하였으며 베리맥스 회전을 통해 인자구조를 단순화하였다. 공통인자에 대응하는 각 변수들의 상대적 영향력을 나타내는 인자적재  $\Lambda$ 는 최대우도법(Maximum Likelihood Method)을 적용하였으며 추정결과는 <표 2>와 같다. 공통인자 1은 서비스업을 중심으로 소비, 투자, 수입의 전반적인 움직임, 공통인자 2는 제조업생산 및 수출의 전반적인 움직임을 나타내는 경기지표로의 특징을 보인다고 할 수 있다.

앞에서 산출된 2개의 공통인자를 설명변수로 이용한 회귀추정식은 다음과 같다.

7) 서비스업생산지수는 1999년 1월, 건설기성액은 1997년 7월, 설비투자추계지수는 1995년 2월, 소비재판매액지수는 1995년 1월부터 통계청에서 공표



<표 2>

인자적재 및 공통인자 추정 결과

변 수	GDP와의 상관계수	인자적재 추정치		공통인자 계수	
		공통인자1	공통인자2	공통인자1	공통인자2
제조업 생산지수	0.78	0.38	0.92	-0.681	1.364
(음식료품및담배)	0.75	0.69	0.40	0.007	-0.003
(섬유 및 가죽제품)	0.75	0.50	0.74	0.005	-0.002
(목재, 종이, 출판 및 인쇄)	0.88	0.82	0.37	0.016	-0.007
(석유, 석탄 및 화학제품)	0.77	0.65	0.32	0.005	-0.002
(비금속광물제품)	0.83	0.76	0.34	0.010	-0.004
(금속제품)	0.91	0.65	0.56	0.008	-0.003
(기계 및 전기전자기기)	0.65	0.22	0.93	-0.009	0.004
(운수장비)	0.76	0.53	0.62	0.004	-0.002
(가구 및 기타 제품 제조업)	0.54	0.36	0.64	0.001	0.000
건설기성액	0.50	0.60	-0.07	0.005	-0.002
전력판매량	0.86	0.77	0.31	0.010	-0.004
서비스업생산지수(총지수)	0.94	0.89	0.45	0.335	-0.138
(도소매업)	0.95	0.89	0.45	0.697	-0.288
(숙박 및 음식점업)	0.92	0.89	0.41	0.081	-0.033
(운수업)	0.93	0.67	0.63	0.013	-0.006
(부동산 및 임대업)	0.73	0.82	0.26	0.013	-0.006
(사업서비스업)	0.93	0.85	0.47	0.065	-0.027
(교육 서비스업)	0.82	0.80	0.38	0.015	-0.006
(오락, 문화및운동관련서비스업)	0.88	0.85	0.42	0.030	-0.013
(기타공공, 수리및개인서비스업)	0.87	0.90	0.32	0.048	-0.020
소비재액판매액지수	0.92	0.88	0.38	0.046	-0.019
설비투자추계지수	0.89	0.60	0.75	0.020	-0.008
BOP수출액(FOB)	0.22	-0.12	0.53	-0.002	0.001
BOP수입액(FOB)	0.91	0.61	0.67	0.009	-0.004
생산자재고지수	0.35	0.40	-0.22	0.003	-0.001

$$\widehat{GDP}_t = 0.910 GDP_{t-1} + 0.825 FAC1_t - 0.677 FAC1_{t-1} + 0.626 FAC2_t - 0.559 FAC2_{t-1}$$

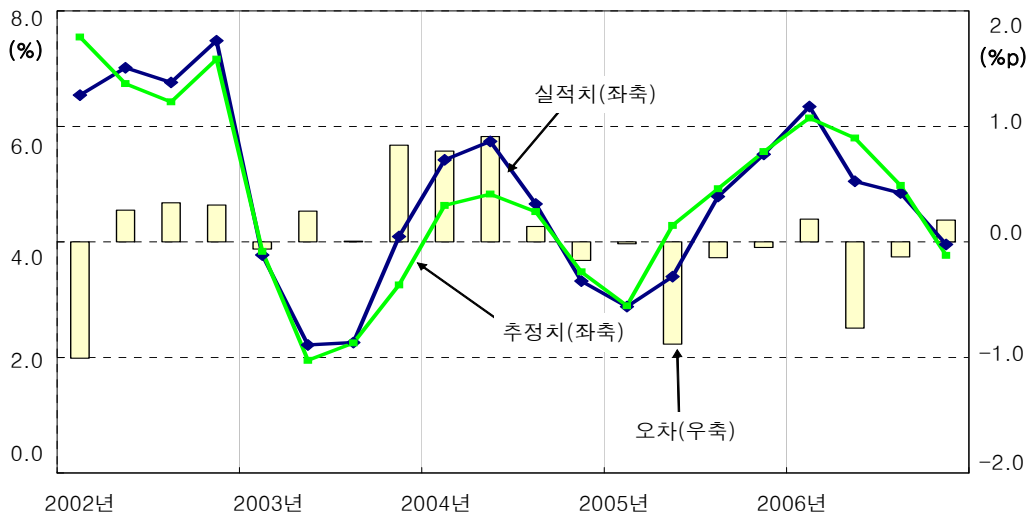
(9.82)                      (16.30)                      (-7.53)  
(13.90)                      (-8.48)

여기서 GDP는 표준화한<sup>8)</sup> 분기 실질 GDP 전년동기비, FAC1과 FAC2는 공통인자 1과 공통인자 2를 나타내며 ( )내는 회귀계수들에 대해 t-통계량을 나타낸다.

인자모형에서의 결정계수는 0.972로 매우 높은 모형적합도를 가지며 DW-통계량 값도 2.185로 오차의 독립성가정을 만족하고 있다. 또한 회귀계수의 크기 및 방향(+, -)에서도 의미 있는 결과를 산출하고 있어 모형의 타당성을 확인할 수 있다.

<그림 3>

인자모형에 의한 분기 GDP 추정



8) 평균 0, 분산 1이 되도록 변수변환

역사적 모의실험에 의한 오차는 MAPE 통계량 기준으로는 약 0.4%p, RMSPE 통계량 기준으로는 약 0.5%p로 나타났다. 대다수 시점에서 추정오차가 0.3%p 이내로 추정력이 비교적 우수한 것으로 나타나고 있으나, 2002년 1분기, 2003년 4분기, 2004년 1, 2분기, 2005년 2분기 등 일부 시점에서 지나치게 큰 오차가 발생하고 있어 추정력의 안정성은 미흡한 것으로 보인다.

#### 다. BVAR모형

BVAR모형을 이용해 분기 GDP를 추정하기 위해 제조업생산지수(IPI), 전력판매량(ELE), 비농가취업자수(EMP), 서비스업생산지수(SP) 등 4개 주요 지표를 이용하였다. 식(1)의 초모수  $\lambda, d, f(i, j)$ 는 다음과 같은 모의실험 과정을 통해 산출되었다. 1992년 1/4분기~1999년 4/4분기를 시계열을 대상으로 BVAR모형을 적합한 후 2000년 1/4분기~2006년 4/4분기까지 자료를 추가해 가면서 표본외 1시차 후 예측을 실시하여 Theil의 U-통계량 기준으로 예측오차를 가장 작게 하는 초모수 값을 선정하였다. 전반적인 응집성을 나타내는 초모수  $\lambda=0.2$ , 시차변수에 대한 감소정도를 나타내는 초모수  $d=1$ , 변수간 상대적 응집성을 나타내는 초모수  $f(i, j)$ 는 다음과 같다.

i \ j	GDP	IPI	ELE	EMP	SP
G D P	1.00	0.50	0.01	0.01	0.50
I P I	0.50	1.00	0.01	0.01	0.50
E L E	0.01	0.01	2.00	0.01	0.01
E M P	0.01	0.01	0.01	2.00	0.01
S P	0.50	0.50	0.01	0.01	1.00

동 모형에서는 BIC통계량 값이 최소가 되는 6분기까지의 시차 변수를 모형에 포함하였다. BVAR모형에 의한 GDP 추정결과는 <표 3>과 같다. GDP의 1시차계수가 1에 근사하고 다른 시차에서는 0에 근사하여 사전분포의 가정을 만족하고 있으며, DW-통계량 값도 1.954로 오차의 독립성가정을 만족하고 있다.

<표 3>

**BVAR모형의 추정계수**

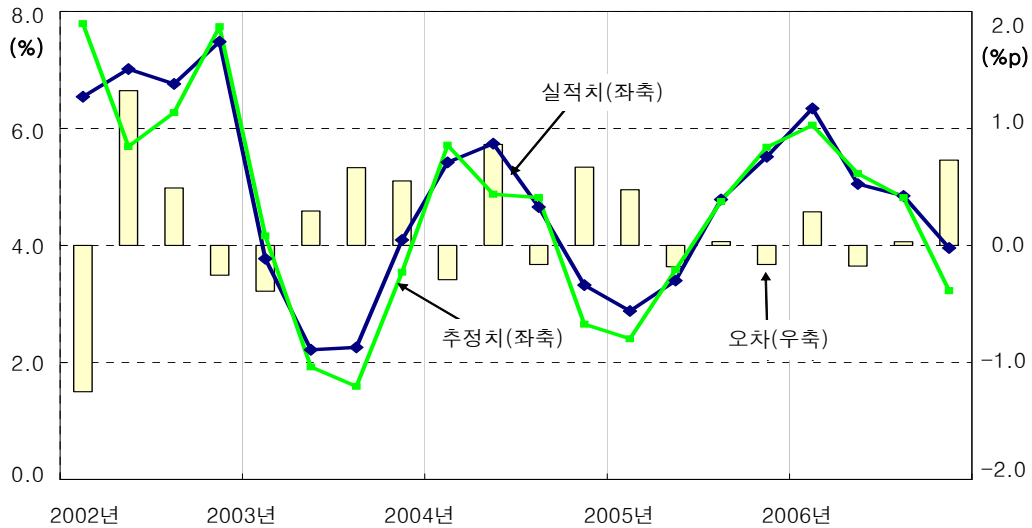
지 표 \ 시 차	1	2	3	4	5	6
G D P	0.951 (7.03)	-0.109 (-0.76)	-0.122 (-0.85)	0.004 (0.03)	0.046 (0.32)	0.028 (0.20)
I P I	0.001 (0.02)	-0.011 (-0.23)	-0.006 (-0.12)	0.014 (0.28)	0.024 (0.50)	-0.003 (-0.07)
E L E	-0.002 (-0.04)	-0.021 (-0.38)	-0.001 (-0.01)	-0.001 (-0.01)	-0.010 (-0.18)	0.022 (0.39)
E M P	0.023 (0.15)	-0.080 (-0.51)	-0.082 (-0.52)	-0.167 (-1.06)	0.080 (0.51)	-0.007 (-0.05)
S P	0.108 (1.47)	-0.005 (-0.06)	-0.007 (-0.10)	-0.011 (-0.14)	0.001 (0.01)	-0.039 (-0.53)
상 수 항	1.1833 (1.34)	DW-통계량 : 1.954				

주: ()내는 t-통계량 값

역사적 모의실험을 실행해본 결과 MAPE 통계량 기준으로는 약 0.4%p, RMSPE 통계량 기준으로는 약 0.5%p의 오차가 발생하고 있다. 대다수 시점에서의 추정오차는 0.4%p 이내이나 2002년 1분기, 2분기에는 지나치게 큰 오차가 발생하여 추정력의 안정성이 미흡한 것으로 보인다.

<그림 4>

### BVAR모형에 의한 분기 GDP 추정



## 라. 결합추정 기법

ADL모형, 인자모형, BVAR모형 등 개별모형에 의한 추정치는 일부시점에서 지나치게 큰 오차가 발생할 뿐만 아니라 안정적 결과를 도출하는데도 한계를 보이고 있다. 따라서 분산최소기법 및 회귀모형기법에 의해 산출 가능한 8개 조합에 의해 결합 추정치를 작성해 보았다(<표 4> 참조). 대부분 결합추정식에서 인자모형의 가중치가 가장 커 인자모형이 상대적으로 보다 많은 정보를 제공하는 것으로 보인다.

계량모형에 의한 GDP 추정치는 실적치와의 절대적 차이에 의한 정확성 뿐만 아니라 경기의 정저점 국면전환 등도 신속히 판단할 필요가 있다. 따라서 여기에서는 ADL모형, 인자모형, BVAR모형에 의한 개별 추정치에 대한 결합추정치의 상대적 비교를 위해

RMSPE 통계량을 산출해 보고, HLN검정을 실행하였다. 또한 경기 상승·하강 등 방향성에 대한 정보의 적절성 평가를 위해 TP 검정도 함께 실행하였다.

<표 4>

**결합기법에 의한 추정식**

결합기법	결합모형	결합추정식
분산 최소 기법	A D L + 인 자	0.4822 (ADL) + 0.5178 (인자)
	인 자 + B V A R	0.5872 (인자) + 0.4128 (BVAR)
	A D L + B V A R	0.5526 (ADL) + 0.4474 (BVAR)
	ADL+인자+BVAR	0.1220 (ADL) + 0.6086(인자) + 0.2694 (BVAR)
회귀 모형 기법	A D L + 인 자	0.5454 + 0.5030 (ADL) + 0.4127 (인자)
	인 자 + B V A R	0.5251 + 0.4869 (인자) + 0.4249 (BVAR)
	A D L + B V A R	0.7766 + 0.5136 (ADL) + 0.3633 (BVAR)
	ADL+인자+BVAR	0.5771 + 0.2813 (ADL) + 0.3349 (인자) + 0.2919 (BVAR)

먼저 개별추정치의 정확성을 판단하기 위한 기준으로 ARIMA모형에 의해 분기 GDP를 추정해보고 이를 벤치마크로 설정하였다. <표 5>는 ADL모형, 인자모형, BVAR모형 모두 RMSPE 기준으로 ARIMA모형에 비해 평균 오차의 크기가 크게 축소될 뿐만 아니라, HLN검정에 의해서도 통계적으로 유의한 차이를 보이고 있다.

RMSPE 기준으로는 인자모형의 평균오차가 가장 낮지만 ADL 모형 및 BVAR모형과 큰 차이를 보이지 않았다. 이와 같은 결과는 ARIMA모형에 의한 추정치를 벤치마크로 설정하고 ADL모형, 인자모형, BVAR모형에 의한 부가적 정보제공의 유무성 판단을

위한 HLN검정에서도 큰 차이를 보이지 못하고 있다.

<표 5>

**추정방법별 추정력 비교**

추 정 방 법	RMSPE	HLN검정 <sup>1)</sup>	TP검정
A R I M A	1.01	-	9.27
인 자	0.48	-2.14	14.09
A D L	0.54	-2.01	14.09
B V A R	0.55	-2.05	14.09

주: 1) 5% 유의수준에 따른 임계치 -1.7보다 작은 경우 ARIMA모형보다 추정력이 우수한 것으로 판단  
 2) 5% 유의수준에 따른 임계치 5.0보다 큰 경우 결합추정치와 증감방향이 실적치와 대부분 일치하는 것으로 판단

결합추정치의 유용성을 평가해 보기 위해 RMSPE 통계량을 계산하여, ADL모형, 인자모형 및 BVAR모형에 의한 추정치를 벤치마크로 설정한 후 다양한 결합에 의한 결합추정치의 정확성을 평가해 보았으며 그 결과는 <표 6>과 같다. RMSPE 기준으로 볼 때 분산최소기법에 의한 추정치는 개별추정치에 비해 개선의 효과가 미미하지만 회귀분석 방식에서는 정확도가 상대적으로 높아지고 있다. 이와 같은 결과는 HLN검정에 의한 통계적 유의성 검정에서도 확인 할 수 있다.

특히 회귀분석기법 중에서도 BVAR모형과 인자모형을 결합한 추정치가 최소 RMSPE를 가지며 TP검정에 의한 방향성에서도 만족할 만한 결과를 보이고 있어 가장 우월한 것으로 나타났다.

<표 6>

결합기법별 추정력 비교

결합기법	벤치마크	결 합 모 형	RMSPE	HLN검정 <sup>1)</sup>	TP검정 <sup>2)</sup>
분 산 최 소 기 법	인자	인 자 + A D L	0.46	-0.33	14.09
		인 자 + B V A R	0.45	-0.20	14.09
		인자+ADL+BVAR	0.44	-0.80	14.09
	ADL	A D L + 인 자	0.46	-1.31	14.09
		A D L + B V A R	0.48	-0.71	14.09
		ADL+인자+BVAR	0.44	-1.37	14.09
	BVAR	B V A R + 인 자	0.45	-0.77	14.09
		B V A R + A D L	0.48	-0.16	14.09
		BVAR+인자+ADL	0.44	-0.63	14.09
회 귀 분 석 기 법	인자	인 자 + A D L	0.42	-1.12	17.22
		인 자 + B V A R	0.33	-1.10	10.42
		인자+ADL+BVAR	0.39	-1.24	10.42
	ADL	A D L + 인 자	0.42	-2.51**	17.22
		A D L + B V A R	0.39	-2.80**	10.42
		ADL+인자+BVAR	0.39	-2.06**	10.42
	BVAR	B V A R + 인 자	0.33	-2.90**	10.42
		B V A R + A D L	0.39	-2.92**	10.42
		BVAR+인자+ADL	0.39	-1.37	10.42

주: 1) 5% 유의수준에 따른 임계치 -1.7보다 작은 경우(\*\*) 벤치마크보다  
결합추정치 추정력이 우수한 것으로 판단

2) 5% 유의수준에 따른 임계치 5.0보다 큰 경우 결합추정치의 증감방향  
실적치와 대부분 일치하는 것으로 판단



---

## V. 맺음말

ADL모형, 인자모형 및 BVAR모형에 의해 각각 분기 실질 GDP를 추정해 본 결과 추정치의 정확도에 통계적으로 유의한 차이를 보이지 않음을 확인하였다. 개별모형의 결합을 통한 개선효과를 확인해 보기 위해 분산최소기법 및 회귀분석기법 등 결합추정기법을 ADL모형, 인자모형, BVAR모형을 대상으로 실시해 보았다. 분산최소기법에서는 개별모형에 비해 결합추정치에 추정력이 큰 차이를 보이지 않았으나, 회귀분석기법에 의한 결합추정치는 평균적인 오차의 크기뿐만 아니라 통계적 검정을 통한 유의성도 확보하는 개선효과를 보이는 것으로 나타났다. 또한 TP검정을 통한 추정치의 방향성에서도 실적치와 대부분 일치하고 있어 결합기법에 의한 분기 GDP추정이 보다 유용함을 확인할 수가 있었다.

본고에서는 3개 계량모형을 대상으로 결합추정치를 산출하여 평가하는 데 그쳤으나, 새로운 계량모형 등 보다 다양한 분석을 통한 결합기법을 적용해 봄으로써 우리나라 분기 GDP추정에 활용 가능성이 높은 조합을 산출하기 위한 지속적 연구가 필요할 것으로 보인다.

## < 참고문헌 >

민경삼, 박진호, 박순옥, “월별 지표를 이용한 분기예측모형”, 『통계분석연구』, 제7권 제2호, 통계청, 2002, pp 97-126.

성병희, “Bayesian VAR모형을 이용한 경제전망”, 『경제분석』, 제7권 제2호, 한국은행, 2001, pp 59-90

심상달, 이항용, “월별 경기지표를 이용한 내수의 단기예측”, 『KDI 분기별 경제전망』, 한국개발연구원, 1992, pp 63-75

이종원, 이상돈, “결합방식을 이용한 예측모형의 효율성 분석”, 『응용경제』, 제1권, 제1호, 한국응용경제학회, 1999, pp 3-34

Bates, J. M. and Granger, C. W. J., "The combination of forecasts", *Operational Research Quarterly*, 1969, pp 451-468.

Doan, T. A., Litterman, R. B. and Sims, C. A., "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Reviews*, 3, 1984, pp 1-100.

---

Granger, C. W. J. and Ramanathan, R., "Improved Methods of Combining Forecasts", *Journal of Forecasting*, **3**, 1984, pp 197-204.

Harvey, D. I., Leybourne, S. J. and Newbold, P., "Testing the equality of prediction mean square errors", *International Journal of Forecasting*, **13**, 1997, pp 273-281.

Litterman, R. B., "Forecasting with Bayesian Vector Autoregressions: Five years of experience", *Journal of Business and Economic Statistics*, **4**, 1986, pp 25-38.

Newbold, P. and Granger, C. W. J., "Experience with forecasting univariate time series and the combination of forecasts", *Journal of the Royal Statistical Society*, 1974, pp 131-164.

Reijer, A. H. J., "Forecasting Dutch GDP using large scale factor models", working paper, 2005, pp 1-27.

Roberson, J. C. and Tallman, E. W., "Vector Autoregression and Reality", *Economic Review* 1, Federal Reserve Bank of Atlanta, 1999.

Pesaran, M. H. and Timmerman, A. G., "A simple Non parametric test of predictive performance", *Journal of Business and Economic Statistics*, 10, 1992, pp 461-465.

Schneider, M. and Spitzer, M., "Forecasting Austrian GDP using the generalized dynamic factor model", working paper, 2004, pp 1-36.

Sedillot, F. and Pain N., "Indicator Models of Real GDP growth in selected OECD countries", *Economics Department Working Papers*, 364, 2003.

Sims, C. A. and Zha, T. A., "Bayesian Methods for Dynamic Multivariate Models", *International Economic Review*, 39, 1998, pp 949-968.

Stock, J. and Waston, M., "Forecasting Using Principal Components From a Large number of Predictors", *Journal of American Statistical Association* 97, 2002, pp1167-1179.