

Cointegration Analysis with Mixed-Frequency Data of Quarterly GDP and Monthly Coincident Indicators

Byeongchan Seong¹

¹Department of Applied Statistics, Chung-Ang University

(Received November 12, 2012; Revised November 19, 2012; Accepted November 27, 2012)

Abstract

The article introduces a method to estimate a cointegrated vector autoregressive model, using mixed-frequency data, in terms of a state-space representation of the vector error correction(VECM) of the model. The method directly estimates the parameters of the model, in a state-space form of its VECM representation, using the available data in its mixed-frequency form. Then it allows one to compute in-sample smoothed estimates and out-of-sample forecasts at their high-frequency intervals using the estimated model. The method is applied to a mixed-frequency data set that consists of the quarterly real gross domestic product and three monthly coincident indicators. The result shows that the method produces accurate smoothed and forecasted estimates in comparison to a method based on single-frequency data.

Keywords: Missing data, state-space model, Kalman filter, expectation maximization algorithm, smoothing.

1. 서론

경제·경영 분야에서 관측되어지는 수많은 다변량 시계열 자료들은 혼합주기(mixed frequency) 또는 비완전 자료(incomplete data)의 형태로 흔히 나타난다. 예를 들어, 한국은행이나 통계청 데이터베이스의 시계열 변수들은 서로 다른 샘플링 주기를 가지고 있다. 즉, 각각의 변수들이 주별, 월별, 분기별 또는 연별로 관측되어 서로 다른 관측주기를 가지는 혼합주기형 자료 구조를 가진다. 대표적인 예로서, 국내 총생산(GDP)과 소비자물가지수(CPI)로 이루어진 이변량 시계열 자료를 고려한다면, GDP는 분기별로, CPI는 월별로 조사·관측되기 때문에 혼합주기형 이변량 시계열 자료로 정의할 수 있다.

혼합주기형 다변량 시계열 분석에 관련된 그동안의 연구는, Chow와 Lin (1971)과 Ghysels와 Valkanov (2006)에서처럼 고차 주기의 자료(high-frequency data)를 시간적으로 통합(temporal aggregation)하여 저차 주기(low-frequency)로 만들거나, Cuche와 Hess (2000)와 Liu와 Hall (2001)의 연구처럼 저차 주기의 자료를 보간(interpolation)하여 고차 주기로 만든 이후 단일주기형태(single-frequency)에 근거한 다변량 시계열 모형에 의존해 왔다. 그러나 이는 정보의 손실 및 예측력 저하를 일으키고 편익된 추정 결과를 산출할 수 있다 (Seong 등, 2012).

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (No. 2012R1A1A2008006).

¹Associate Professor, Department of Applied Statistics, Chung-Ang University, 221, Heukseok-dong, Dongjak-gu, Seoul 156-756, Korea. E-mail: bcseong@cau.ac.kr

비정상 다변량 시계열 분석에서 대표적 방법 중의 하나인 공적분(cointegration) 모형은, Engle와 Granger (1987) 이후, 계량경제학 분야에서 가장 심도있게 연구되어온 분야 중의 하나이다. 그럼에도 불구하고, 그 모형은 결측치가 없는 경우(complete data) 또는 단일주기형태와 같은 잘 정돈된 자료에 국한되어 주로 연구되어 왔다. VAR(vector autoregressive) 모형은 공적분 구조를 설명하기 위하여 오차수정모형(VECM)의 형태를 이용하여 결측치가 없는 경우에 널리 이용되어 왔다 (Johansen, 1996). 혼합주기형 자료를 위해서는 비구조적 VAR (Zadrozny, 1990) 및 구조적 VAR (Mariano와 Murasawa, 2003)가 동시에 이용되어 왔으며, Bai 등 (2009)은 VAR 모형과 Ghysels와 Valkanov의 MIDAS(mixed data sampling) 접근방법을 결측치가 있는 경우 또는 혼합주기자료를 위한 경우에 모형화 방법을 비교하였으나 공적분을 전혀 고려하지 않았다.

이러한 맥락에서 Seong 등 (2012)은 모든 다변량 시계열들이 최고차원의 주기(highest frequency)에서 생성된다고 가정하고 혼합주기형 공적분 분석을 위하여 VECM의 상태공간모형화(state space model) 및 추정 방법을 제안하였다. 이를 이용하여 EM(expectation maximization, Dempster 등, 1977) 및 칼만필터(Kalman filter)에 기반하여 최대가능도 추정법(maximum likelihood estimation)을 고려하고, 고차 주기 시점을 기준으로 저차 주기 시계열 자료들을 추정(또는 평활, smoothing)하였다. 그들은 고차 주기로 보간하거나 저차 주기로 통합하지 않고 공적분 모형을 수립함으로써 추정 및 예측의 효율성을 크게 개선하였다.

본 연구에서는 Seong 등의 방법을 이용하여 우리나라 분기별 GDP 및 3가지 월별 경기 동행지수들(coincident indicators)로 이루어진 4차원 혼합주기형 시계열에서 공적분 분석을 시도하였다. 특히, 글로벌 금융위기 이후 속보성 있는 거시 경제통계에 관심이 높은 만큼 본 연구를 통한 월별 GDP 추정 및 예측은 의미있는 연구로 사료된다.

본 논문의 구성은 다음과 같다. 2장에서 혼합주기형 공적분 모형을 간단하게 소개하고, 3장에서는 우리나라의 분기별 GDP와 월별 경기 동행지수로 이루어진 다변량 시계열 자료를 위한 공적분 모형을 추정하고 그 결과를 단일주기형 공적분 모형과 비교하였다.

2. 다변량 계절형 시계열모형

본 연구의 기본 모형은 공적분 계수가 h 인 $VAR(p)$ 이며, z_t 는 최고주기의 t 시점에서 n 차원의 다변량 시계열을 나타낸다. 일반적으로 공적분된 $VAR(p)$ 의 VECM은 다음과 같이 표현할 수 있다;

$$\Delta z_t = \alpha \beta' z_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta z_{t-j} + e_t. \quad (2.1)$$

여기서, α 와 β 는 각각 조정계수 및 공적분 벡터의 열로 이루어진 $n \times h$ 행렬을 나타내며, Γ_j 는 $n \times n$ 행렬, 그리고 e_t 는 IID $N(0, \Sigma)$ 로 분포되어 있다. 여기서, z_t 의 아래쪽 $d = n - h$ 개의 원소들은 공적분되어 있지 않다고 가정하며 이것은 일종의 정규화 $\beta = [I_h \ \beta_0]'$ 를 의미한다 (Ahn과 Reinsel, 1990). 단, I_h 는 $h \times h$ 항등행렬이며, β_0 는 $d \times h$ 이다. 특성방정식은 d 개가 단위근이며, 나머지 근들은 정상성을 만족하거나 단위원 바깥에 존재한다고 가정한다. Engle와 Granger (1987)의 용어로 표현하자면, z_t 는 (1, 1)의 차수로 공적분되어 있다. 본 연구에서는 편의상, $z_t = (z'_{1t}, z'_{2t})'$ 는 n_1 차원의 고차 주기의 시계열 변수 z_{1t} 가 위쪽에, n_2 차원의 저차 주기의 시계열 변수 z_{2t} 가 아래쪽에 배치되어 있다고 가정한다; 단, $n = n_1 + n_2$. 본 연구에서 z_{2t} 는 변화율 변수(flow variable)의 형태 또는 시간적 통합의 형태로 간주하며, 수준변수(stock variable)의 형태를 위해서는 유사한 방법으로 변형되어 적용될 수 있다.

다변량 시계열의 혼합주기형 구조를 반영하기 위해서, 모형식 (2.1)을 상태공간모형으로 변환하는 것이

편리하다. 왜냐하면, 상태방정식(state equation) 및 관측방정식(measurement equation)의 틀 안에서 혼합주기의 형태가 쉽게 다루어질 수 있기 때문이다. 따라서, 모형식 (2.1)의 VECM을 상태방정식으로 변환하면 다음과 같은 형태를 얻게 된다;

$$x_t = Fx_{t-1} + Ge_t. \tag{2.2}$$

먼저, 상태벡터(state vector)는 $s \times 1$ 차원의 $x_t = (z'_t, \dots, z'_{t-r+1})'$ 이며, s 는 $s = nr$ 이고 $r = \max(p, v + 1)$, v 는 저차 주기 시계열의 시간통합에서 차수의 최대값을 나타낸다. 전이행렬(transition matrix) 및 입력행렬은 다음과 같이 만들 수 있다;

$$F = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_r \\ I_n & O_n & \cdots & O_n \\ O_n & I_n & \cdots & O_n \\ \vdots & \vdots & \ddots & \vdots \\ O_n & O_n & \cdots & O_n \end{bmatrix} : s \times s, \quad G = \begin{bmatrix} I_n \\ O_{(s-n) \times n} \end{bmatrix} : s \times n.$$

여기서, $\Phi_1 = I_n + \alpha\beta' + \Gamma_1$, $\Phi_j = \Gamma_j - \Gamma_{j-1}$ for $j = 2, \dots, p-1$, $\Phi_p = -\Gamma_{p-1}$, $\Phi_j = O_n$ if $j > p$ 이며, O_n 와 $O_{m \times n}$ 은 $n \times n$ 및 $m \times n$ 차원의 영행렬을 나타내고, 상태벡터의 초기값 x_0 은 $N(\delta, \Lambda)$ 로 분포되어 있다고 가정한다.

다음으로, 관측방정식을 만들기 위하여, $y_t^+ = (y'_{1t}, y'_{2t})'$ 를 정의한다. y_{1t} 는 고차 주기에서 관측된 z_{1t} 을 나타내고, 저차 주기를 위한 y_{2t} 는 관측되면 $y_{2t}^+ = z_{2t}$; 관측되지 않으면 $y_{2t}^+ = w_t$ 의 구조를 가지도록 한다. 여기서, w_t 는 Brockwell과 Davis (1991)에서처럼, z_t 와 독립인 IID $N(0, I_{n_2})$ 를 따르는 확률변수이다. 따라서, 다음과 같은 관측방정식을 만들 수 있다;

$$y_t^+ = H_t x_t + Q_t w_t. \tag{2.3}$$

$H_t = [H'_1, H'_{2t}]'$, $H_1 = [I_{n_1}, O_{n_1 \times (s-n_1)}]$, $H_{2t} = [O_{n_2 \times n_1}, C_0, O_{n_2 \times n_1}, C_1, \dots, O_{n_2 \times n_1}, C_{r-1}]$ 이고 $C_j = O_{n_2}$ for $j > v$; $Q_t = [O'_{n_1 \times n_2}, Q'_{2t}]'$; y_{2t} 는 관측되면 $Q_{2t} = O_{n_2}$, 관측되지 않으면 $Q_{2t} = I_{n_2}$. 여기서 C_j 들은 변화율 변수를 위한 시간적 통합의 형태를 표현하며, 행렬의 대각원소로서 0 또는 1을 가진다. 주목할 점은, H_1 은 상태벡터에서 고차 주기의 시계열 자료 z_{1t} 를 그대로 추출하는 역할을 하고, H_2 는 저차 주기의 시계열 자료를 시간적 통합의 형태 $\sum_{j=0}^v C_j z_{2t-j}$ 로 뽑아낸다는 것이다. 이러한 접근 방법은 Zadrozny (1990)을 참고하시오.

모형식 (2.1)의 VECM은 절편 및 추세와 관련된 여러가지 유용한 형태로 확장될 수 있다. 예를 들어, 상수항 및 결정적 선형 추세 $\gamma_1 + \gamma_2(t-1)$ 을 포함시킨다면 다음과 같이 표현될 수 있다;

$$\Delta z_t = \gamma_1 + \gamma_2(t-1) + \alpha\beta' z_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta z_{t-j} + e_t. \tag{2.4}$$

모형식 (2.4)를 위한 상태공간모형으로의 변환 및 다른 형태의 VECM을 위해서는 Seong 등 (2007)을 참고하시오.

모형식 (2.2)의 상태방정식과 모형식 (2.3)의 관측방정식으로 이루어진 상태공간모형은 EM 알고리즘 및 칼만필터를 통하여 추정될 수 있다. 즉, EM 알고리즘의 E-단계를 통하여 소위 혼합주기형 자료 하에서의 조건부 가능도함수가 유도될 수 있으며, M-단계를 통하여 모형식 (2.1)의 VECM 모수들이 추정될 수 있다. 또한, 칼만필터의 과정을 거쳐서 최고차 주기의 모든 시점들에서의 저차 시계열의 값들이 평활 또는 추정된다. 자세한 추정과정은 Seong 등 (2012)을 참고하시오. 또한, 혼합주기형 정상 다변량 시계열 분석을 위하여 유사한 접근 방법을 사용한 Shumway와 Stoffer (1982)을 참고하시오.

3. 한국의 거시경제자료 분석

본 연구에서는, 한국 거시경제의 중요 지표 중 하나인 경기동행지수와 GDP로 이루어진 다변량 시계열을 분석하기 위하여 2장에서 소개된 혼합주기형 공적분 모형을 이용한다. 경기동행지수는 월별로 통계청에서 수집되는 비농가취업자수(EMP), 도소매판매액지수(SLS), 산업생산지수(IIP)의 3가지를 이용하였다. 모든 시계열들은 계절조정된 것이며, 비농가취업자수는 천명 단위의 자료이고 두 가지 지수의 기준년도는 2005년(2005 = 100)이다. GDP는 분기별로 한국은행에서 집계되는 것으로서 실질 GDP이다. 분석에서 모형 수립을 위해서 1995년 1월부터 2011년 6월까지의 자료가 사용되었으며 2011년 7월부터 2012년 6월까지의 자료는 모형의 예측력을 비교하기 위해서 사용하였다. 자료의 분석은 Seong 등 (2012)의 MATLAB 코드를 이용하였다.

원자료는 다음과 같이 분산안정화를 위하여 로그 변환을 취하였으며 각 값들은 분석의 편의상 자료의 시작 시점의 값으로 빼서 100을 곱하였다 (Hamilton, 1994);

$$z_t = \begin{pmatrix} \text{emp}_t \\ \text{sls}_t \\ \text{iip}_t \\ \text{gdp}_t \end{pmatrix} = 100 \times \begin{pmatrix} \ln(\text{EMP}_t) - \ln(\text{EMP}_{1995:01}) \\ \ln(\text{SLS}_t) - \ln(\text{SLS}_{1995:01}) \\ \ln(\text{IIP}_t) - \ln(\text{IIP}_{1995:01}) \\ \ln(\text{GDP}_t) - \ln(\text{GDP}_{1995:03}) \end{pmatrix}, \quad t = 1, \dots, 198. \quad (3.1)$$

여기서, 시작 시점의 값으로 빼는 과정은 비정상 시계열을 정상화하는 차분과는 다르다. 분석의 목적은 다음의 3가지로 볼 수 있다; (1) '월별' GDP의 추정 및 예측, (2) 3가지 경기동행지수의 예측, (3) 4가지 변수간의 공적분 관계의 발견.

분석의 출발점으로서 단일주기형 공적분 모형을 수립하였다. 즉 경기동행지수들의 각 분기별 3번째(3월, 6월, 9월, 12월) 값들과 각 분기별 gdp_t 의 값만 모은, 자료의 길이가 $T = 65$ 인 단일주기형 자료를 이용하여 다음과 같은 VECM 모형을 수립하였다;

$$\begin{aligned} \Delta z_\tau &= \hat{\rho}_1 + \hat{\rho}_2(\tau - 1) + \hat{a}\hat{b}'z_{\tau-1} + u_\tau, \quad \tau = 1, \dots, 65, & (3.2) \\ \hat{\rho}_1 &= (0.989, 1.538, 1.533, 2.069)', & \hat{\rho}_2 &= (-0.024, -0.020, 0.008, -0.035)', \\ \hat{a} &= (-0.046, -0.038, 0.013, -0.052)', & \hat{b} &= (1, -1.197, -2.662, 4.508)', \\ \widehat{\text{var}}(u_\tau) &= \begin{pmatrix} 0.597 & 0.625 & 1.811 & 0.784 \\ & 5.294 & 5.359 & 1.510 \\ & & 20.061 & 5.500 \\ & & & 2.181 \end{pmatrix}. \end{aligned}$$

여기서, VAR(1)은 Schwartz의 BIC(Bayesian information criterion) 기준에 의하여 상수항 및 결정적 선형 추세를 가진 형태로 사용되었고, Johansen (1996)의 Trace 검정에 의하여 공적분 계수는 1로 추정되었다(p -value = 0.0891).

혼합주기형 공적분 모형의 수립은, AIC(Akaike's information criterion) 및 BIC 기준에 의하여 VAR(1)을 이용하였고, VECM 모형의 공적분 벡터의 초기값은 단일주기형 모형에서 추정된 $\hat{b} = (1, -1.197, -2.662, 4.508)'$ 의 마지막 원소의 값에 3배를 하여 $(1, -1.197, -2.662, 13.524)'$ 를 이용하였다. 이것은 GDP가 3개월의 합으로 계산되는 변화율 변수이므로 Pons와 Sansó (2005)의 변환을 사용한 결과이다. 추정된 모형식은 다음과 같다;

$$\Delta z_t = \hat{\gamma}_1 + \hat{\gamma}_2(t - 1) + \hat{\alpha}\hat{\beta}'z_{t-1} + e_t, \quad t = 1, \dots, 198, \quad (3.3)$$

Table 3.1. Monthly smoothed estimates of in-sample quarterly gdp_t , 2010:7 to 2011:6

Year: month	Observed	Temp. agg.	Skip-sampled	Year: month	Observed	Temp. agg.	Skip-sampled
2010:7		68.83	69.12	2011:1		70.50	71.17
2010:8		68.99	68.96	2011:2		70.85	70.70
2010:9	69.12	69.12	69.28	2011:3	71.05	71.05	71.27
2010:10		69.05	68.91	2011:4		71.09	71.30
2010:11		69.28	69.63	2011:5		71.50	71.93
2010:12	69.75	69.75	70.69	2011:6	71.89	71.89	72.44

Note: Reported “aggregated” and “skip-sampled” gdp_t in columns 3 and 4, respectively, reflect quarterly sums of monthly values ending in the indicated month and monthly values for that month multiplied by three (in order to be in quarterly form comparable to observed quarterly gdp_t in column 2).

$$\hat{\gamma}_1 = \begin{pmatrix} 0.2502 \\ 0.4240 \\ 0.5344 \\ 0.1697 \end{pmatrix}'_{\substack{(0.001) \\ (0.0043) \\ (<0.0001) \\ (0.005)}}, \quad \hat{\gamma}_2 = \begin{pmatrix} -0.0095 \\ -0.0111 \\ 0.0034 \\ -0.0020 \end{pmatrix}'_{\substack{(<0.0001) \\ (<0.0001) \\ (<0.0001) \\ (<0.0001)}},$$

$$\hat{\alpha} = \begin{pmatrix} -0.0545 \\ -0.0659 \\ 0.0188 \\ -0.0097 \end{pmatrix}'_{\substack{(<0.0001) \\ (0.0021) \\ (<0.0001) \\ (0.0002)}}, \quad \hat{\beta} = (1, -0.3281, -0.4820, 0.7482)',$$

$$\hat{\Sigma} = \begin{pmatrix} 0.0768 & 0.0585 & 0.2274 & 0.0371 \\ & 0.9672 & 0.7777 & 0.1023 \\ & & 5.6838 & 0.5719 \\ & & & 0.1055 \end{pmatrix}.$$

단, 괄호 안의 값은 각 추정값에 대한 표준오차이며 로그가능도 함수의 헤시안 행렬의 역행렬을 이용하여 계산되었다. 참고로 공적분 벡터의 추정값에서는 점근적 정규성(asymptotic normality)이 성립하지 않기 때문에 표준오차는 일반적으로 생략된다. 따라서, 모형식 (3.3)에 의한 3가지 경기동행지수와 GDP 간의 공적분 관계식은 다음과 같이 추정된다;

$$emp_t - 0.3281 \times sls_t - 0.4820 \times iip_t + 0.7482 \times gdp_t \approx 0. \tag{3.4}$$

Table 3.1은, 추정된 모형식 (3.3)을 사용하여 2010년 7월부터 2011년 6월까지 월별 gdp_t 를 저차 주기와 고차 주기에서 추정한 결과를 참값과 비교한 결과이다. 저차 주기에서의 추정값(smoothed estimates for aggregated gdp_t)이란 실제로 관측되는 GDP의 형태와 비교될 수 있는 3개월 국내총생산의 합을 나타내며, Table 3.1의 3열에 나타나 있다. 고차 주기에서의 추정값(smoothed estimates for skip-sampled gdp_t)은, Table 3.1의 4열 값으로서 각 분기가 끝나는 마지막 달의 gdp_t 의 추정값에 3을 곱한 값이다. 개별 월에 대한 gdp_t 의 추정값에 관심이 있다면 4열의 값을 3으로 나눠서 이용할 수 있다. 여기서, 주목할 점은 저차 주기에서의 gdp_t 의 값은 3월, 6월, 9월, 12월을 제외하고는 실제로 관측할 수 없으며, 특히 모든 월에서의 개별(또는 고차 주기에서의) gdp_t 의 값은 관측될 수 없다는 것이다.

Table 3.2와 Table 3.3은, 2011년 7월부터 2012년 6월까지 $z_{1t} = (emp_t, sls_t, iip_t)'$ 와 $z_{2t} = gdp_t$ 의 예측값을 추정된 모형식 (3.2)와 (3.3)을 통하여 각각 구하고 참값과 비교하여 예측오차를 계산한 것이다. Figure 3.1은 Table 3.3을 그래프화한 것이다. 이를 통하여, 예측오차 관점에서 혼합주기형 공적분 모형이 단일주기형 모형보다 더 효율적임을 알 수 있다. 상대적인 예측오차의 감소량($1 - |\text{혼합주기형 모형의 예측오차}| / |\text{단일주기형 모형의 예측오차}|$)은, gdp_t 와 emp_t 의 경우에 각각 50%와 85% 정도의 예측오차가 감소하였으며, sls_t 의 경우 2011년 9월(3분기)을 제외하고 예측오차가 약 17% 감소하였다. 단, iip_t 의 경우는 단일주기형 공적분 모형이 혼합주기형보다 예측오차가 약 1% 더 낮은 것으로 나타났다.

Table 3.2. Monthly out-of-sample forecast errors of emp_t , inc_t , and iip_t

Year:month	Single-frequency			Mixed-frequency		
	emp_t	inc_t	iip_t	emp_t	inc_t	iip_t
2011:7				-0.01	0.61	-1.11
2011:8				0.05	1.05	-2.43
2011:9	-0.84	0.07	-1.71	-0.36	0.32	-1.79
2011:10				-0.31	-0.64	-2.64
2011:11				-0.41	-1.69	-3.42
2011:12	-0.99	-2.62	-4.60	-0.13	-2.19	-4.74
2012:1				-0.18	-2.69	-2.10
2012:2				-0.09	-2.24	-1.96
2012:3	-1.19	-2.96	-5.35	-0.01	-2.42	-5.52
2012:4				0.04	-2.61	-5.16
2012:5				0.00	-3.42	-4.48
2012:6	-1.50	-4.05	-5.26	-0.07	-3.44	-5.45

Note: The table reports errors of out-of-sample (2011:7–2012:6) forecasts based on an in-sample (1995:1–2011:6) estimated model. For brevity, the table reports only forecast errors, not observed or forecasted values.

Table 3.3. Monthly out-of-sample forecasts of quarterly gdp_t

Year:month	Observed	Single-frequency		Mixed-frequency		
		Forecast	Error	Forecast Low Freq.	Error	Forecast High Freq.
2011:7				72.37		72.75
2011:8				72.75		73.07
2011:9	72.72	73.42	-0.70	73.07	-0.34	73.38
2011:10				73.38		73.69
2011:11				73.69		74.00
2011:12	73.05	74.83	-1.78	74.00	-0.96	74.31
2012:1				74.31		74.62
2012:2				74.62		74.93
2012:3	73.90	76.15	-2.25	74.93	-1.03	75.24
2012:4				75.24		75.55
2012:5				75.55		75.85
2012:6	74.18	77.40	-3.22	75.85	-1.68	76.16

Note: "Out-of-sample" means the same as in Table 3.2.

마지막으로, Figure 2는 추정된 월별 gdp_t 의 전반적인 추세 및 합리성을 살펴보기 위하여 1995년 1월부터 2011년 6월까지의 저차 주기에서 gdp_t 의 추정값과 통계청에서 발표하는 경기종합지수 중 경기동행지수(2005 = 100)를 그린 것이다. 여기서, 경기동행지수는 현재 사용된 자료와 비교를 위하여 로그변환 후 1995년 1월의 값을 차감한 것이다. 두 시계열은 비슷한 증가 추세와 순환적 움직임(cyclic pattern)을 가지고 있음을 알 수 있다. 즉, 추정된 gdp_t 값은 경기동행지수의 움직임을 적절하게 반영하고 있다.

4. 결론

본 연구에서는 혼합주기형태로 나타나는 비정상 다변량 시계열 자료 분석을 위한 공적분 모형을 이용하

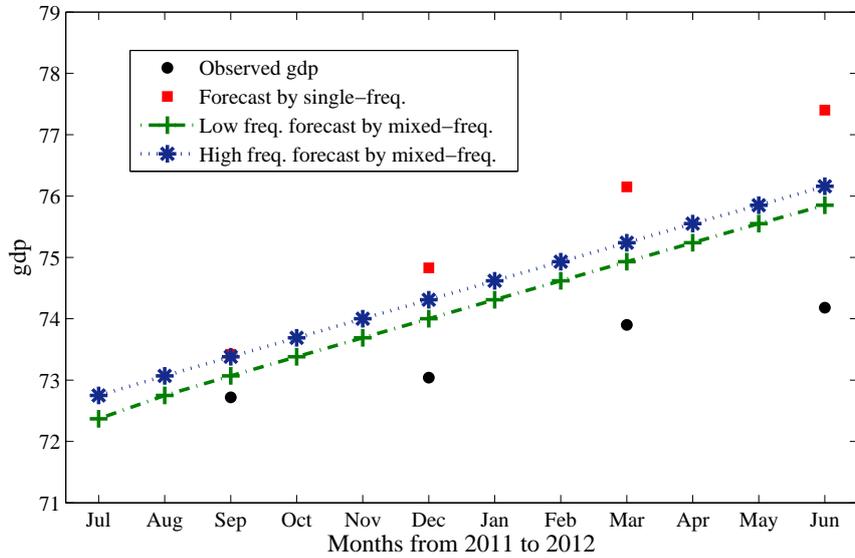


Figure 3.1. Comparison of monthly out-of-sample forecasts of quarterly gdp_t

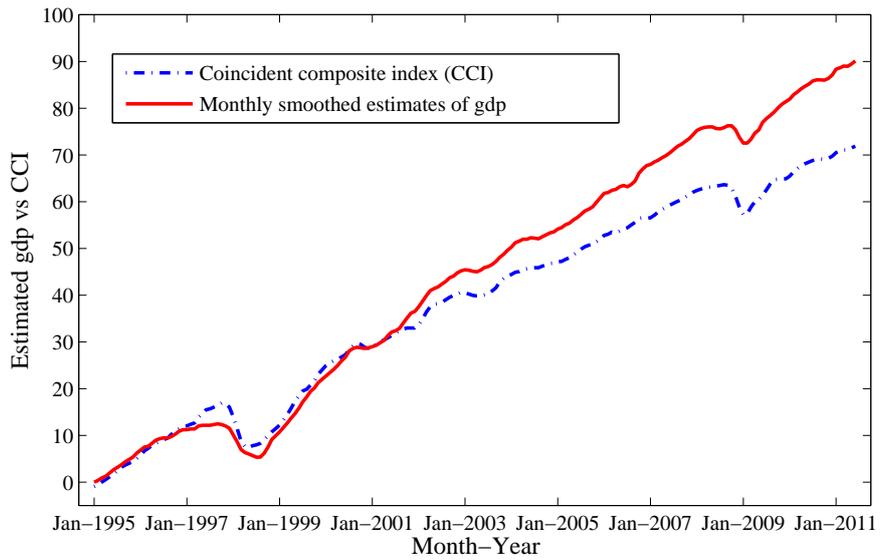


Figure 3.2. Time series plot of the coincident composite index and the monthly smoothed estimates of gdp_t

여, 우리나라의 분기별 GDP 및 월별 경기 동행지수를 분석하고 그 분석 결과가 단일주기형 공적분 모형보다 우수함을 보였다. 그동안 혼합주기형 다변량 시계열의 분석을 위해서 단일주기형 모형에 제한되었던 한계를 극복하고, 월별 주기에 토대한 공적분 관계의 추정 및 예측을 가능하게 한다는 점에 큰 의의를 둘 수 있다. 특히 단일주기형 공적분 모형으로는 추정하기 어려운 월별 GDP 값을 추정하고 예측할 수 있다는 점이 더욱 그러하다고 하겠다.

References

- Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association*, **85**, 813–823.
- Bai, J., Ghysels, E. and Wright, J. (2009). State space models and MIDAS regressions, Working paper, Department of Economics, Johns Hopkins University, Baltimore, MD.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag, New York.
- Chow, G. C. and Lin, A. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series, *Review of Economics and Statistics*, **53**, 372–375.
- Cuče, N. A. and Hess, M. K. (2000). Estimating monthly GDP in a general Kalman filter framework: Evidence from Switzerland, *Economic and Financial Modelling*, **7**, 153–194.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing, *Econometrica*, **55**, 251–276.
- Ghysels, E. and Valkanov, R. (2006). Linear time series processes with mixed data sampling and MIDAS regression models, Working Paper, University of North Carolina, Chapel Hill, NC.
- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.
- Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed., Oxford University Press, Oxford.
- Liu, H. and Hall, S. G. (2001). Creating high-frequency national accounts with state-space modeling: A Monte Carlo experiment, *Journal of Forecasting*, **20**, 441–449.
- Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series, *Journal of Applied Econometrics*, **18**, 427–443.
- Pons, G. and Sansó, A. (2005). Estimation of cointegrating vectors with time Series measured at different periodicity, *Econometric Theory*, **21**, 735–756.
- Seong, B., Ahn, S. K. and Zdrozny, P. A. (2007). Cointegration analysis with mixed-frequency data, CESifo Working Paper 1939, CESifo, Munich.
- Seong, B., Ahn, S. K. and Zdrozny, P. A. (2012). Estimation of vector error correction models with mixed-frequency data, *Journal of Time Series Analysis*, doi: 10.1111/jtsa.12001.
- Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis*, **3**, 253–264.
- Zdrozny, P. A. (1990). Estimating a multivariate ARMA model with mixed-frequency data: an application to forecasting U.S. GNP at monthly intervals, *Center for Economic Studies Discussion Paper 90-5*, Bureau of the Census, Washington, DC.