응용통계연구 (2010)
23(5), 813–822
DOI: 10.5351/KJAS.2010.23.5.813

부분-수량화를 통한 시계열 자료 분석에서의 차원축소

박진아[1] · 황선영[2]

[1]숙명여자대학교 통계학과, [2]숙명여자대학교 통계학과

(2010년 7월 접수, 2010년 8월 채택)

요 약

차원 축소(dimension reduction) 기법은 주로 횡단면 자료 분석에서 널리 이용되어 왔으며 시계열 분석 분야에서의 적용은 상대적으로 미진한 실정이다. 본 논문에서는 부분-수량화를 통한 주성분분석 방법을 계절형 시계열에 적용시켜 시계열 자료의 차원 축소를 시도하고자 한다. 분석 방법론을 단계별로 제시하였으며 월별 실업률 자료 분석을 통해 설명하였다.

주요용어: 부분-수량화(partial quantification), 계절 시계열, 차원 축소(dimension reduction).

1. 서론

시계열 분석기법에서 특정한 모형가정이나 차원 또는 차수 등의 정보가 주어져 있지 않은 경우, 차원이나 차수가 증가하게 되면 모형에서 모수의 개수는 급격하게 증가하는 문제(the curse of dimensionality)가 발생한다. 이러한 현상은 특히 계절형 시계열에서 자주 나타날 수 있다. 이를 극복할 수 있는 시계열 분석에서의 차원 축소, 즉, DR(dimension reduction)에 대한 연구가 최근 들어 시작되고 있다. 시계열에서의 차원 축소를 다루기 위해, Xia와 Li (1999)와 Xia 등 (1999, 2002a)은 single-index 모형을 고려하였다. 또한, Xia 등 (2002b)은 시차에 대한 정보가 주어졌을 때, 시계열 자료에도 활용할 수 있는 회귀분석에서 차원축소방법을 제안하였고 최근에 Park 등 (2009)은 차원, 차수와 모형에 대한 사전 정의 없이 curse of dimensionality를 극복할 수 있는 시계열에서의 차원 축소방법인 sufficient dimension reduction을 제안하였다.

본 연구에서는 시계열 자료의 차원축소 방안으로 다변량 분석에서 주로 이용되는 주성분 분석 방법과 “부분-수량화(partial quantification)” 기법을 이용한 주성분 분석 방법을 시계열 자료에 적용해 보고자 한다. 부분-수량화란 선행 연구나 연구자의 사전 지식 등에 의해 주어진 자료에 선행적으로 제 1 인자를 전제하고 나머지 인자들을 찾는 수량화 절차를 의미한다. Suh와 Huh (1997)은 주성분 분석에서 부분-수량화에 대해 연구하였고, 서혜선 (1999)에서는 인자분석에 의한 부분-수량화 기법을 제안하였다. 황선영과 이연숙 (1999)에서는 정준판별분석에 부분-수량화 기법을 적용하였고, 황선영 등 (2001)은 부분-수량화를 통한 다차원 선호분석을 연구하였다. 시계열 자료 분석은 시간에 따라 관찰되는 자료들을 분석의 대상으로 하며, 관측 값들 사이에 자기상관(autocorrelation) 특성으로 인한 응용성 때문에 활발하게 연구되어 온 분야이다. 시계열 모형을 수립하는 경우, 전통적인 회귀분석이나 다변량 이론에서의 방법론들은 시계열 자료를 위한 선형/비선형 모형 설정에 유용하기는 하나, 관측치들 사이에 존재하는

본 연구는 숙명여자대학교 2009년도 교내연구비 지원에 의해 수행되었음.
[2]교신저자: (140-742) 서울시 용산구 효창원길 52, 숙명여자대학교 통계학과, 교수. E-mail: shwang@sm.ac.kr

종속성(dependence)이 시계열 자료의 가장 큰 특징임을 고려할 때, 기본적으로 독립자료인 횡단면자료 분석에 적합한 통계방법론인 회귀분석 및 다변량 이론의 직접적인 적용은 어느 정도 제한적이라 하겠다. 이 본문에서는 계절형 시계열 사례분석을 중심으로 시계열 자료의 부분-수량화를 통한 차원 축소 가능성에 대해 연구하고자 한다.

2. 주성분 회귀분석

다중회귀분석에서 다수의 설명변수들을 회귀모형에 포함시킬 경우, 설명변수들 사이에 강한 선형관계가 존재하는 다중공선성의 문제가 발생한다. 주성분회귀분석은 설명변수들 간에 이런 다중공선성 문제를 해결하기 위해 설명변수들을 독립성이 만족하는 회귀변수인 주성분으로 변환하여 회귀분석을 시도하는 것이다. 성웅현 (1997)을 바탕으로 정리하면 다음과 같다.

다중회귀모형을 행렬로 표현하면

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{2.1}$$

이다. 여기서 $\mathbf{Y}$는 $n \times 1$ 확률벡터, $\mathbf{X}$는 $n \times (p+1)$ 자료행렬이다. 변수들 사이에 강한 선형관계가 존재할 때에는 일반적으로 표준화변수를 사용하게 된다. 식 (2.1)의 표준화 회귀모형은 다음과 같다.

$$\mathbf{Y}^* = \mathbf{Z}\beta^* + \varepsilon, \tag{2.2}$$

여기서 $\mathbf{Z}^t\mathbf{Z}$는 변수벡터 $\mathbf{X}$들의 $p \times p$ 상관행렬인 $\mathbf{R}$을 의미하므로 다음과 같이 $\mathbf{Z}^t\mathbf{Z}$를 스펙트럼 분해할 수 있다.

$$\mathbf{Z}^t\mathbf{Z} = \mathbf{R} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^t,$$

여기서 $\mathbf{\Lambda}$은 행렬 $\mathbf{R}$의 고유치를 대각원소로 하는 대각행렬이고, $\mathbf{C}$은 행렬 $\mathbf{R}$의 고유벡터들을 열로 갖는 직교행렬이다. 이때 구해진 고유치와 고유벡터를 이용하여 상대적으로 작은 고유치에 대응하는 주성분을 제외시키고 추정치를 구하는 과정을 실행한다. 고유벡터의 직교성을 이용하여 식 (2.2)를 다시 표현하면

$$\begin{aligned} \mathbf{Y}^* &= \mathbf{Z}\mathbf{C}\mathbf{C}^t\beta^* + \varepsilon \\ &= \mathbf{V}\alpha + \varepsilon \end{aligned} \tag{2.3}$$

이 된다. 여기서 $\mathbf{V} = \mathbf{Z}\mathbf{C}$는 p개의 주성분으로 이루어진 주성분벡터이고, $\alpha = \mathbf{C}^t\beta^*$이므로 표준화변수의 회귀계수벡터는 $\beta^* = \mathbf{C}\alpha$로 표시된다. 이 때, 식 (2.3)은 주성분들을 독립변수로 하고 표준화변수 $\mathbf{Y}^*$를 종속변수로 하는 주성분회귀모형이다. 따라서 독립변수인 주성분은 표본상관행렬의 고유벡터를 원소로 하는 표준화변수의 선형결합이고, 식 (2.3)의 주성분회귀모형에서 회귀계수벡터 α를 최소자승법에 의해 추정하면

$$\hat{\alpha} = (\mathbf{V}^t\mathbf{V})^{-1}\mathbf{V}^t\mathbf{Y}^* = (\mathbf{C}^t\mathbf{Z}^t\mathbf{Z}\mathbf{C})^{-1}\mathbf{V}^t\mathbf{Y}^* = \mathbf{\Lambda}^{-1}\mathbf{V}^t\mathbf{Y}^* \tag{2.4}$$

이 되고 추정된 주성분 회귀함수는 $\hat{\mathbf{Y}}^* = \mathbf{V}\hat{\alpha}$가 된다.

3. 부분 수량화(Partial Quantification)

본 연구에서 분석의 대상이 되는 $\mathbf{X}$은 $n \times p$ 자료행렬이고, 행렬 $\mathbf{Z}$은 표준화 행렬로 가정한다.

$$\mathbf{Z} = \begin{bmatrix} \vec{z}_1^t \\ \vdots \\ \vec{z}_n^t \end{bmatrix} = (\mathbf{z}_1,\ \mathbf{z}_2, \ldots, \mathbf{z}_p)$$

단위길이를 갖는 $\vec{v}_1 = (v_1, \ldots, v_p)^t$가 주어진 경우, 사전적(a priori) 의미의 첫 번째 주성분(first principal component)은 다음과 같이 정의된다.

$$s_i = \vec{z}_1^t \vec{v}_1, \quad i = 1, \ldots, n.$$

주어진 다변량자료에서 제 1주성분을 제외하고 분석을 하기 위해 $\vec{z}_i$을 다음과 같이 두 부분으로 분해한다.

$$\vec{z}_i = s_i \vec{z}_i + (\vec{z}_i - s_i \vec{z}_i), \quad i = 1, \ldots, n.$$

따라서

$$\vec{z}_i - s_i \vec{v}_1, \quad i = 1,\ 2, \ldots, n$$

에 대한 일반적인 주성분분석을 다시 실시할 수 있다. 이와 같은 주성분분석의 부분-수량화 과정을 Suh와 Huh (1997)를 바탕으로 정리하면 다음과 같다.

(단계 1) 단위벡터 $\vec{v}_1$이 제 1주성분계수(first principal coefficient)로 주어진 경우 다음 수량화를 위해 자료행렬 $\mathbf{Z}$은 다음과 같이 대체한다.

$$\mathbf{Z}_{[1]} = \mathbf{Z}(\mathbf{I}_p - \mathbf{P}_{\vec{v}_1}), \tag{3.1}$$

여기서, $\mathbf{P}_{\vec{v}_1} = \vec{v}_1(\vec{v}_1^t \vec{v}_1)^{-1} \vec{v}_1^t$, 즉 $\vec{v}$로의 사영행렬(projection matrix)이다.

(단계 2) $p \times 1$ 단위벡터 $\vec{v}_2$는 $\min ||\mathbf{Z}_{[1]} - \mathbf{Z}_{[1]}\mathbf{P}_{\vec{v}_2}||^2$ 또는 $\max ||\mathbf{Z}_{[1]}\mathbf{P}_{\vec{v}_2}||^2$를 통해 얻을 수 있고, 이전 단계에서 넘어온 행렬 $\mathbf{Z}_{[1]}$을 다음과 같이 대체한다.

$$\mathbf{Z}_{[2]} = \mathbf{Z}_{[1]}(\mathbf{I}_p - \mathbf{P}_{\vec{v}_2}). \tag{3.2}$$

(단계 3) 위와 같은 과정을 단계 p까지 실행하여 주성분계수벡터 $\vec{v}_3, \ldots, \vec{v}_p$를 얻는다.

부분수량화 과정의 (단계 1)에서 식 (3.1)로부터 $\mathbf{Z} = \mathbf{Z}\vec{v}_1\vec{v}_1^t + \mathbf{Z}_{[1]}$으로 표현되며, (단계 2)의 식 (3.2)로부터 $\mathbf{Z}_{[1]} = \mathbf{Z}_{[1]}\vec{v}_2\vec{v}_2^t + \mathbf{Z}_{[2]}$를 얻을 수 있는데, $\vec{v}_1\vec{v}_2 = 0$이므로

$$\mathbf{Z} = \mathbf{Z}\vec{v}_1\vec{v}_1^t + \mathbf{Z}\vec{v}_2\vec{v}_2^t + \mathbf{Z}_{[2]}$$

로 정리된다. 따라서 $\vec{v}_3, \ldots, \vec{v}_p$를 고려하여 전개하면

$$\mathbf{Z} = \mathbf{Z}\vec{v}_1\vec{v}_1^t + \mathbf{Z}\vec{v}_2\vec{v}_2^t + \cdots + \mathbf{Z}\vec{v}_{p-1}\vec{v}_{p-1}^t + \mathbf{Z}\vec{v}_p\vec{v}_p^t + \mathbf{Z}_{[p]}$$

가 된다. 이때 $\mathbf{Z}$는 단계 p까지의 과정을 통한 p개 성분에 의해 모두 표현되므로 $\mathbf{Z}_{[p]}$는 $\mathbf{0}$행렬이다. 그러므로 다음과 같이 분해 할 수 있다.

$$\mathbf{Z} = \sum_{j=1}^{p} \mathbf{Z}\vec{v}_j\vec{v}_j^t.$$

이때, λ_j와 단위벡터 $\mathbf{u}_j$를 각각 다음과 같이 정의하자.

$$\lambda_j = \vec{v}_j^t \mathbf{Z}^t \mathbf{Z} \vec{v}_j, \qquad \mathbf{u}_j = \frac{\mathbf{Z}\vec{v}_j}{\sqrt{\lambda_j}}, \quad j = 1, \ldots, p.$$

그러면, 다음과 같은 표현식이 가능하며

$$\mathbf{Z} = \sum_{j=1}^{p} \sqrt{\lambda_j}\mathbf{u}_j\vec{v}_j^t = \mathbf{U}\mathbf{D}_{\sqrt{\lambda}}\mathbf{V}^t, \tag{3.3}$$

여기서 $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$, $\mathbf{V} = (\vec{v}_1, \ldots, \vec{v}_p)$, $\mathbf{D}_{\sqrt{\lambda}} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p})$이다. 식 (3.3)의 $\mathbf{U}$와 $\mathbf{V}$는 다음을 만족한다.

$$\mathbf{V}^t\mathbf{V} = \mathbf{V}\mathbf{V}^t = \mathbf{I}_p, \qquad \mathbf{U}^t\mathbf{U} = \begin{pmatrix} 1 & \mathbf{m}^t \\ \mathbf{m}^t & \mathbf{I}_{p-1} \end{pmatrix},$$

여기서, $\mathbf{m} = (m_2, \ldots, m_p)^t$, $m_j = \mathbf{u}_1^t\mathbf{u}_j$ $(j = 2, \ldots, p)$을 의미한다. 따라서 식 (3.3)을 행렬 $\mathbf{Z}$의 유사-비정칙치분해라하고 λ_j를 유사-고유값이라한다. 유사-비정칙분해에 의하여 $\mathbf{Z}$는 다음과 같이 표현할 수 있다.

$$\mathbf{Z} = \mathbf{G}\mathbf{H}^t, \tag{3.4}$$

여기서, $\mathbf{G} = \mathbf{U}\mathbf{D}_{\sqrt{\lambda}}$, $\mathbf{H} = \mathbf{V}$이다. 즉 행렬 $\mathbf{Z}$의 x_{ij}는

$$x_{ij} = \vec{g}_i^t\vec{h}_j, \quad i = 1, \ldots, n,\ j = 1, \ldots, p.$$

로 표현되고, $\vec{g}_i^t$는 행렬 $\mathbf{G}$의 i번째 행벡터이며 $\vec{h}_j$는 행렬 $\mathbf{H}$의 j번째 열벡터이다. 벡터 $\vec{g}_i^t$는 '행효과'로, $\vec{h}_j$는 '열효과'로 고려된다. 이를 r차원으로 축소하여 근사적으로 표현하면 다음과 같다.

$$\mathbf{Z} \approx \mathbf{G}_{(r)}\mathbf{H}_{(r)}^t, \tag{3.5}$$

여기서 $\mathbf{G}_{(r)}$와 $\mathbf{H}_{(r)}$은 각각 행렬 $\mathbf{G}$와 $\mathbf{H}$의 처음 r개의 열을 갖는다. 이와 같이 r차원으로 축소된 자료 행렬의 근사도(goodness of fit)는

$$\frac{||\mathbf{G}_{(r)}||^2}{||\mathbf{G}||^2} = \frac{\left|\left|\mathbf{U}_{(r)}\mathbf{D}_{\sqrt{\lambda_{(r)}}}\right|\right|^2}{||\mathbf{U}\mathbf{D}_{\sqrt{\lambda}}||^2} = \frac{\sum_{j=1}^{r}\lambda_j}{\sum_{j=1}^{p}\lambda_j} \tag{3.6}$$

로 정의한다 (*cf.* Suh와 Huh, 1997).

4. 부분-수량화를 통한 시계열자료의 차원축소 : 사례분석

본 절에서는 시계열 자료에 차원축소기법을 적용하기 위해 국내 실업률 자료를 고려하고자 한다. 자료는 1999년 1월부터 2010년 3월까지의 월별시계열로 원자료의 시도표는 그림 4.1과 같다. 전통적인 시계열 분석방법인 ARIMA(p, d, q) 모형을 적용할 때, 특히 월별 실업률 자료의 경우 계절성이 존재하기 때문에 Seasonal ARIMA(p, d, q) 모형을 이용한 분석이 일반적이다. 이와 같은 모형을 사용할 경우, 시차가 늘어남에 따라 증가하는 모수의 추정에 부담이 증가한다. 본 연구에서는 주성분 분석 방법과 부분-수량화(partial quantification) 기법을 이용한 주성분 분석 방법을 이용해 월별 시계열 자료의 차원을 축소하고, 이를 통해 얻어진 주성분점수를 이용하여 시계열회귀분석을 시행하고자 한다. 첫 번째로는 안정화되지 않은 월별 실업률 x_t에 대해 분석하고, 두 번째로는 안정화된 $\nabla \ln x_t$에 대해 동일한 분석을 시행하도록 한다. 월별 실업률 시계열 데이터를 $\mathbf{X}_t$라 할 때, 주성분분석에 사용할 자료 행렬 $\mathbf{X}$는 다음과 같다. $\mathbf{X} = (\mathbf{X}_{t-1},\ \mathbf{X}_{t-2}, \ldots, \mathbf{X}_{t-12})$, 여기서 자료 행렬 $\mathbf{X}$는 열별로 표준화된 행렬이다.

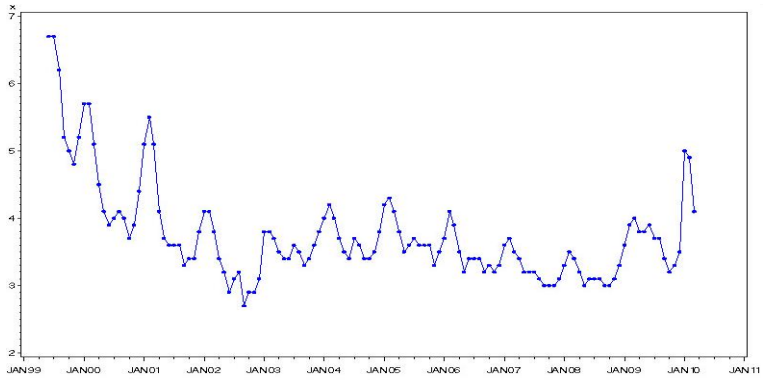

그림 4.1. 월별실업율(단위%, 1999년 1월–2010년 3월)

표 4.1. 주성분 분석에 의한 주성분 점수벡터

Variable	주성분점수벡터	
	제 1축	제 2축
$\mathbf{X}_{t-1}$	0.1425	0.4598
$\mathbf{X}_{t-2}$	0.2305	0.4537
$\mathbf{X}_{t-3}$	0.2873	0.3538
$\mathbf{X}_{t-4}$	0.3143	0.224
$\mathbf{X}_{t-5}$	0.3231	0.0983
$\mathbf{X}_{t-6}$	0.3268	0.0176
$\mathbf{X}_{t-7}$	0.3288	−0.0341
$\mathbf{X}_{t-8}$	0.3296	−0.1026
$\mathbf{X}_{t-9}$	0.3199	−0.2033
$\mathbf{X}_{t-10}$	0.2961	−0.3279
$\mathbf{X}_{t-11}$	0.2625	−0.3865
$\mathbf{X}_{t-12}$	0.2417	−0.2953
고유값	7.1574	2.2804
근사도(%)	59.6%	19.0%

4.1. 원자료 x_t에 대한 차원축소

일반적인 월별시계열 자료에 많이 적용하는 AR(12) 모형을 적합 시킨 후 변수선택을 이용해서 최종 선택된 모형은 다음과 같다.

$$\mathbf{X}_t - \overline{\mathbf{X}} = \nu_t, \tag{4.1}$$
$$\nu_t = 1.1053\nu_{t-1} - 0.3437\nu_{t-2} + 0.1364\nu_{t-6} - 0.1427\nu_{t-8} + 0.1362\nu_{t-10}.$$

모형에 대한 $\sqrt{\text{MSE}} = 0.36883$이고 $R^2 = 0.8682$이다.

월별 실업률 자료행렬 $\mathbf{X}$에 대한 주성분 분석 결과 2차원으로 축소된 자료 행렬의 근사도는 약 78.6%이며 이 중 제 1축이 59.6%, 제 2축이 19.0%를 설명하고 있다. 표 4.1은 주성분 점수벡터와 고유값과 근사도에 대한 결과이다. 이 주성분 점수를 이용하여 새로운 변수 V_{1t}와 V_{2t}를 생성하여, 월별 실업률 데이터 $\mathbf{X}_t$와 시계열 회귀분석을 시행하였다. 적합된 모형(P)은 다음과 같다.

$$P: \ \mathbf{X}_t = 3.7941 + 0.2593V_{1t} + 0.0621V_{2t}. \tag{4.2}$$

표 **4.2.** 부분수량화에 의한 주성분 점수벡터 (제 1주성분: 12개월 단순평균으로 지정)

Variable	주성분점수벡터	
	제 1축	제 2축
$\mathbf{X}_{t-1}$	0.2886751	0.4775714
$\mathbf{X}_{t-2}$	0.2886751	0.4434186
$\mathbf{X}_{t-3}$	0.2886751	0.3242138
$\mathbf{X}_{t-4}$	0.2886751	0.186178
$\mathbf{X}_{t-5}$	0.2886751	0.0567778
$\mathbf{X}_{t-6}$	0.2886751	−0.02342
$\mathbf{X}_{t-7}$	0.2886751	−0.072595
$\mathbf{X}_{t-8}$	0.2886751	−0.137338
$\mathbf{X}_{t-9}$	0.2886751	−0.23101
$\mathbf{X}_{t-10}$	0.2886751	−0.344245
$\mathbf{X}_{t-11}$	0.2886751	−0.38907
$\mathbf{X}_{t-12}$	0.2886751	−0.29048
유사고유값	820.74	270.46
근사도(%)	57.9%	19.2%

이제, 3장에서 소개한 부분-수량화기법을 적용하도록 하자. 제 1주성분 점수벡터를 과거 시차들의 단순평균으로 사전 정의(predetermined)하였을 경우, (단계 1)∼(단계 3)를 이용한 부분-수량화 나머지 주성분 분석 결과는 표 4.2와 같다. 이차원으로 축소된 근사도는 약 77%이며, 이중 제 1축이 57.9%, 제 2축이 19.2%를 설명하고 있다. 단순평균인 제 1축을 사전 정의하였을 경우에 유도된 제 2축은 상반기와 하반기의 대비(contrast)로 해석할 수 있다. 부분 수량화 기법으로 결과를 이용하여 적합한 모형(Q)은 다음과 같다.

$$Q:\ \mathbf{X}_t = 3.8013 + 0.2676V_{1t} + 0.0381V_{2t} \tag{4.3}$$

두 모형 Q와 P의 적합도인 $\sqrt{\text{MSE}}$와 R^2는 다음과 같다.

차원축소 모형	P	Q
$\sqrt{\text{MSE}}$	0.0657	0.06927
R^2	0.9814	0.9794

부분수량화 기법을 적용한 주성분 회귀모형(Q)의 경우 설명력은 약 97.9%로 주성분 회귀모형(P)의 98.14% 보다는 약간 낮게 나타나지만 모형 Q의 해석상 용이한 장점이 약간의 설명력 저하를 충분히 보상하고 있음을 볼 수 있다. 또한 차원 축소 모형인 Q와 P의 적합도는 일반적인 AR(12) 모형(식 (4.1))의 적합도인 $\sqrt{\text{MSE}} = 0.3688$와 $R^2 = 0.8682$ 보다 상당히 우수함을 알 수 있다.

4.2. 로그 차분한 자료($\nabla \ln x_t$)의 차원 축소 분석

두 번째로 안정화된 로그차분($\nabla \ln x_t$) 시계열에 대해 위와 같은 방법으로 분석을 시행해 보았다. 안정화된 $\nabla \ln x_t$에 대한 시도표는 그림 4.2에 제시하였다. 우선 일반적인 AR(12) 모형을 적합 시켜 얻은 최종 모형은 다음과 같다.

$$\nabla \ln x_t = \mathbf{W}_t; \quad \mathbf{W}_t - \overline{\mathbf{W}} = \nu_t, \tag{4.4}$$
$$\nu_t = 0.198242\nu_{t-1} - 0.200044\nu_{t-2} - 0.20541\nu_{t-4} + 0.408108\nu_{t-12}.$$

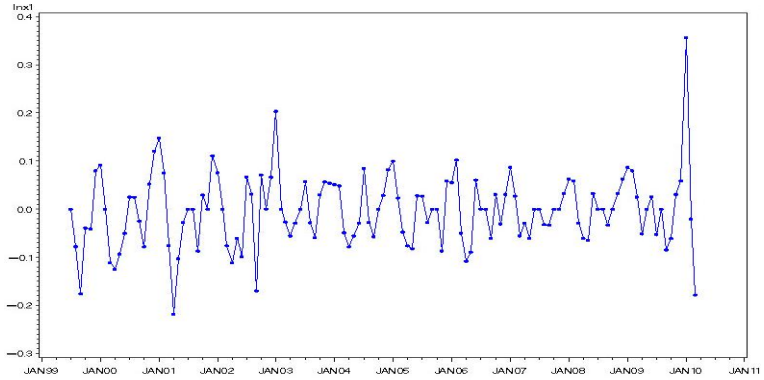

그림 4.2. 로그 차분($\nabla \ln x_t$)한 월별 실업율

표 4.3. 주성분 분석에 의한 주성분 점수벡터

Variable	주성분점수벡터	
	제 1축	제 2축
$\mathbf{W}_{t-1}$	0.2405	0.4947
$\mathbf{W}_{t-2}$	0.2963	0.3114
$\mathbf{W}_{t-3}$	0.3228	0.0214
$\mathbf{W}_{t-4}$	0.3263	−0.2135
$\mathbf{W}_{t-5}$	0.2887	−0.2779
$\mathbf{W}_{t-6}$	0.2773	0.0687
$\mathbf{W}_{t-7}$	0.2940	0.3099
$\mathbf{W}_{t-8}$	0.3147	0.1887
$\mathbf{W}_{t-9}$	0.3266	−0.0868
$\mathbf{W}_{t-10}$	0.2946	−0.3883
$\mathbf{W}_{t-11}$	0.2450	−0.4789
$\mathbf{W}_{t-12}$	0.2109	0.1054
고유값	6.7534	1.2650
근사도(%)	56.3%	10.5%

적합된 모형의 $\sqrt{\text{MSE}} = 0.76326$이고 $R^2 = 0.4311$로 우리가 일반적으로 시차선택을 한 후 사용하는 모형의 경우 설명력이 떨어지는 것을 알 수 있다.

로그 차분한 월별 시계열 자료행렬인 $\mathbf{W}_t$의 주성분 분석결과 2차원으로 축소된 자료 행렬의 근사도는 약 66.8%이며 이 중 제 1축이 56.3%, 제 2축이 10.5%를 설명하고 있다. 표 4.3은 연관된 주성분점수벡터와 고유값과 근사도에 대한 결과이다. 주성분 분석 결과 제 1주성분은 시차 전반에 걸쳐 비슷한 가중치가 주어져 과거 자료의 평균적 지표라 해석할 수 있고, 제 2주성분은 1, 3분기에서는 양의 값을 갖고 2, 4분기에서는 음의 값을 가지므로 분기에 따른 대비라 해석할 수 있다. 이 주성분 점수를 이용하여 새로운 변수 V_{1t}와 V_{2t}를 생성, 로그 차분한 월별 실업률 데이터 $\mathbf{W}_t$와 시계열 회귀분석을 시행하였다. 적합된 모형(P)은 다음과 같다.

$$P: \ \mathbf{W}_t = -0.0041 + 0.2787V_{1t} + 0.0005V_{2t}. \tag{4.5}$$

제 1주성분 점수벡터를 과거 시차들의 전반적인 단순평균으로 사전 정의하였을 경우, 부분-수량화 기법을 이용한 주성분 분석 결과는 표 4.4와 같다.

표 **4.4.** 부분수량화에 의한 주성분 점수벡터

Variable	주성분점수벡터	
	제 1축	제 2축
$\mathbf{W}_{t-1}$	0.2886751	0.5000835
$\mathbf{W}_{t-2}$	0.2886751	0.2997935
$\mathbf{W}_{t-3}$	0.2886751	0.0034284
$\mathbf{W}_{t-4}$	0.2886751	−0.226055
$\mathbf{W}_{t-5}$	0.2886751	−0.282016
$\mathbf{W}_{t-6}$	0.2886751	0.0697414
$\mathbf{W}_{t-7}$	0.2886751	0.3017874
$\mathbf{W}_{t-8}$	0.2886751	0.1707676
$\mathbf{W}_{t-9}$	0.2886751	−0.103902
$\mathbf{W}_{t-10}$	0.2886751	−0.396322
$\mathbf{W}_{t-11}$	0.2886751	−0.468269
$\mathbf{W}_{t-12}$	0.2886751	0.1309622
유사고유값	772.96	146.95
근사도(%)	55.5%	10.2%

이 표에서 보면 2차원으로 축소된 자료 행렬의 근사도는 약 66%이며, 이중 제 1축이 55.5%, 제 2축이 10.5%를 설명하고 있다. 제 1축을 사전 정의하였을 경우에 제 2축을 분기별 대비로 해석할 수 있다. 부분 수량화 기법으로 결과를 이용하여 적합한 모형(Q)은 다음과 같다.

$$Q: \ \mathbf{W}_t = -0.0041 + 0.2811V_{1t} - 0.0076V_{2t}. \tag{4.6}$$

두 모형의 Q(식 (4.6))와 P(식 (4.5))의 $\sqrt{\text{MSE}}$와 R^2는 다음과 같다.

차원축소 모형	P	Q
$\sqrt{\text{MSE}}$	0.00391	0.00446
R^2	0.9973	0.9965

첫 번째 분석(4.1 절)과 마찬가지로 부분-수량화 기법을 적용한 주성분 회귀모형의 경우 설명력은 약 99.65%로 주성분 회귀모형의 99.73% 보다는 약간 낮게 나타난다. 그러나 기존의 분석기법들이 일단 자료와 분석방법이 정해지면 사전에 분석자의 주관이나 지식이 개입되는 과정 없이 기계적으로 계산되어진다는 단점을 가지고 있다면, 부분-수량화 절차에 의한 차원 축소에서는 연구자의 주관이 개입되어 이전의 방법에 비해 약간의 정보손실은 야기할 수 있으나, 자료 분석과 사전지식의 조화를 얻을 수 있다는 점에서 다양한 분야의 실증분석에 유용하게 사용될 수 있을 것으로 생각된다.

참고문헌

서혜선 (1999). 인자분석에 의한 부분수량화, <응용통계연구>, **12**, 165–179.

성웅현 (1997). <응용 다변량분석 -이론, 방법론, SAS 활용->, 탐진.

황선영, 이연숙 (1999). 정준판별분석의 조건부계량화, <응용통계연구>, **12**, 517–525.

황선영, 정수진, 김영원 (2001). 다차원선호분석의 최적척도화 및 부분 수량화, <응용통계연구>, **14**, 305–320.

Park, J. H., Sriram, T. N. and Yin, X. (2009). Dimension reduction in time series, *Statistica Sinica*, **20**, 747–770.

Suh, H. S. and Huh, M. H. (1997). Partial quantification in principal component analysis, *The Korean Communications in Statistics*, **4**, 637–644.

Xia, Y. and Li, W. K. (1999). On the estimation and testing of functional coefficient linear models, *Statistica Sinica*, **9**, 735–757.

Xia, Y., Tong, H. and Li, W. K. (1999). On extended partially linear single-index models, *Biometrika*, **86**, 831–842.

Xia, Y., Tong, H. and Li, W. K. (2002a). Single-index volatility models and estimation, *Statistica Sinica*, **12**, 785–799.

Xia, Y., Tong, H., Li, W. K. and Zhu, L. X. (2002b). An adaptive estimation of dimension reduction methods, *Journal of the Royal Statistical Society: Series B*, **64**, 363–410.

Dimension Reduction in Time Series via Partially Quantified Principal Component

J.A. Park[1] · S.Y. Hwang[2]

[1]Department of Statistics, Sookmyung Women's University
[2]Department of Statistics, Sookmyung Women's University

(Received July 2010; accepted August 2010)

Abstract

We investigate a possible achievement in dimension reduction of time series via partially quantified principal component. Partial quantification technique allows us in modeling to accommodate artificial variable(s) of practical importance which is defined subjectively by the data analyst. Suggested procedures are described and in turn illustrated in detail by analyzing monthly unemployment rates in Korea.

Keywords: Partial quantification, seasonal time series, dimension reduction.

This work was supported by a grant (2009) from Sookmyung Women's University.
[2]Corresponding author: Professor, Department of Statistics, Sookmyung Women's University, 52 Hyochangwon-gil, Yongsan-gu, Seoul 140-742, Korea. E-mail: shwang@sm.ac.kr