

## 소비자물가지수의 시계열모형 연구

이훈자<sup>1</sup>

<sup>1</sup>평택대학교 디지털응용정보학과

접수 2012년 4월 30일, 수정 2012년 5월 16일, 게재확정 2012년 5월 21일

### 요약

소비자물가지수는 국가의 중요한 경제 척도 중의 하나이다. 본 연구에서는 4개 도시, 서울, 부산, 대구, 광주지역의 소비자물가지수를 연구하였다. 자료는 모두 통계청에서 발췌하였고, 기간은 1998년-2011년 월별자료이며, 시계열분석 기법인 자기회귀오차모형으로 분석하였다. 소비자물가 분석을 위한 설명변수는 9가지 경제변수인 경기동행지수, 미국환율, 생산자물가지수, 원유수입단가, 원유수입물량, 국제경상수지, 수입물가지수, 실업율, 화폐통화량을 사용하였다. 분석 결과, 자기회귀오차모형으로 각 지역별 소비자물가지수를 46%-52% 정도 설명할 수 있다.

주요용어: 설명변수, 소비자물가, 자기회귀오차모형.

### 1. 서론

소비자물가지수는 국가의 경제동향, 경제정책, 화폐구매력 등을 보는 중요한 경제 지표이다. 지금까지 다양한 방법으로 소비자물가지수에 관해 연구되어 왔다. 국내연구의 경향을 살펴보면, 소비자물가에 대한 유가 및 환율충격의 비대칭성, 비선형성 분석 (김기호와 윤성훈, 2009), 소비자물가와 생산자물가 간의 구조적 격차 (김효율, 2008), 구조적 VAR모형을 이용한 환율전가 효과분석: 한국의 수입물가와 소비자물가를 중심으로 (주세우 등, 2010), 소비자물가와 생산자물가의 관계 및 소비자물가 변동요인에 관한 연구(이충열, 1996) 등으로 연구되었다.

소비자물가가 다른 경제지표에 의해 영향을 받는다는 사실이 국내적으로 알려져 왔다. 환율이 소비자물가지수에 영향을 미친다 (정동빈, 2007). 소비자물가에 유가와 환율은 매우 영향이 크며, 이 영향은 유가와 환율이 상승할 때가 하락할 때 보다 더 큰 것으로 나타났다 (김기호와 윤성훈, 2009). 소비자물가에 수입소비재가격이 매우 커다란 영향을 준 것으로 분석되었고 (이충열, 1996), 실업률도 소비자물가지수 인플레이션에 영향을 주는 것으로 나타났다 (임성식, 2011). 김영덕 (2000)은 소비자물가에 국제유가가 많은 영향을 주고, 산업생산지수와 통화량도 연관성이 있음을 보여준다. 그러나, 주세우 등 (2010)에서는 국제유가와 환율변화는 수입물가에 중요한 영향을 미치지만, 수입물가 자체는 소비자물가에 영향을 미치지 못하는 것으로 나타난 연구도 있는 것으로 나타났다.

지금까지 소비자물가지수에 관한 대부분의 선행연구는, 소비자물가에 관한 모형연구 보다는 소비자물가와 연관성이 있는 2-5 정도의 변수와의 연관성을 연구했다. 본 연구의 목적은 소비자물가에 영향을 주는 설명변수를 포함시킨 시계열 모형을 분석하는 것이다. 기대효과는 시계열 모형으로 분석된 소비자물가지수 모형에 따라 지역별로 미래의 값을 예측하는 것이다.

본 연구에서는, 지출목적별 소비자물가지수 가운데 총지수 (2010=100)를 사용했다. 총지수에는 식품, 가정용품, 의류 및 신발, 음료, 주류 및 담배, 주택·수도·전기, 보건, 교통, 음식, 교육, 서비스 등이

<sup>1</sup> (450-701) 경기도 평택시 용이동 111 평택대학교 디지털응용정보학과, 부교수. E-mail: esther@ptu.ac.kr

포함되어 있다. 1988년 1월부터 2011년 12월까지의 14년간 월별 자료를 시계열 모형에 적합하고자 한다. 소비자물가의 분석을 위해서는 소비자물가와 연관된 자료를 포함시켜야 한다. 본 논문에서는 소비자물가와 연관이 있고, 통계청에서 구할 수 있는 9종류의 자료를 설명변수로 포함한 다변량 분석인 ARE (Autoregressive Error; 자기회귀오차) 모형으로 분석하였다. 지역별로 차이를 보기위해 서울, 부산, 대구, 광주 4지역의 소비자물가 지수를 비교하였다.

## 2. 소비자물가 자료 및 설명변수 자료

소비자물가 자료는 통계청 KOSIS 자료를 사용했다. 기간은 1988년부터 2011년까지의 14년간 월별 자료를 사용했다. 통상적으로 자료분석 할 때, 70-80% 자료는 모델링기간으로 최근 20-30% 자료는 검증기간으로 사용한다. 그러나 본 연구에서는 14년간의 월별 168자료가 비교적 적은 관계로 모든 자료를 모델링 분석에 사용했다. 그리고 상반기, 하반기가 차이가 있을 것으로 생각해, 상반기 1월-6월, 하반기 7월-12월로 나누어 분석하였다.

설명변수로 사용한 자료는 소비자물가와 연관이 있으면서, 통계청에서 쉽게 구할 수 있는 변수를 사용하였다. 설명변수도 소비자물가와 같이 1998년-2011년 월별자료를 사용하였다. 9종류의 설명변수는 이충열 (1996), 김영덕(2000), 김기호와 윤성훈 (2009), 주세우 등 (2010), 임성식 (2011)을 참고로 하여 경기동행지수, 미국환율, 생산자물가, 원유수입단가, 원유수입물량, 국제경상수지, 수입물가지수, 실업률, 화폐통화량을 사용하였다. 국내총생산 (GDP)도 소비자물가와 연관성이 있는 것으로 알려졌으나, 분기별 자료이기 때문에 사용하지 못했고, 산업생산지수도 소비자물가와 연관성이 있으나, 2000년 이후부터 자료가 있기 때문에 사용하지 못했다. 또한, 실업률은 각 도시별 자료가 1998년부터 있기 때문에 본 연구에서는 지역별로 사용하지 못하고 전국 실업율을 이용하였다. 이분산 모형으로 심주용과 이장택 (2010), 황창하와 신사임 (2010)을 참고하였고, 설명변수 선택방법을 위해 Hwang (2010)을 참고하였고 시계열모형을 위해 이훈자 (2010)를 참고하였다.

## 3. 소비자물가의 ARE 모형과 적합성

반응변수에 영향을 주는 설명변수가 있을 때는 설명변수를 분석에 이용하는 것이 효율적이다. ARE (Autoregressive Error; 자기회귀오차) 모형은 설명변수를 분석에 사용 할 수 있는 모형으로 시계열자료를 회귀모형에 적합 시킬 때 적합한 모형이다. 반응변수인 소비자물가에 영향을 주는 수입물가지수나, 환율 같은 설명변수를 사용하여 자료를 분석한다. ARE 모형은 전이함수 모형의 특수한 경우로, 오차항 ( $\epsilon_t$ )가 독립이 아니라 서로 상관관계를 갖게 되며, 특히 AR (Autoregressive; 자기회귀) 형태를 갖는다. 일반적인  $k$ 차 ARE 모형은 다음과 같다.

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp} + \epsilon_t, t = 1, 2, \dots, n,$$

$$\epsilon_t = e_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2} - \dots - \phi_k \epsilon_{t-k}, \text{ 이고 } e_t \sim i.i.d.N(0, \sigma^2) \quad (3.1)$$

식 (3.1)에서 오차항 ( $\epsilon_t$ )는 서로 상관관계가 있고 AR( $k$ ) 형태를 갖는다. 또한 식 (3.1)에서  $Y_t$ 는 반응변수인 소비자물가를 나타내며,  $X_{tj}$  ( $j = 1, 2, \dots, p$ )는  $Y_t$ 에 영향을 주는 미국 환율, 수입물가지수 등의 설명변수들을 나타낸다. 위의 ARE 모형과 회귀모형의 차이점은 ARE 모형은 식 (3.1)에서 오차항이 서로 독립이 아니라 시차에 따른 자기 상관을 갖고 있으며 특히 AR( $k$ ) 모형의 형태를 갖는다는 점이다. 따라서 식 (3.1)을 ARE 모형이라 부르며 시계열자료에 회귀모형을 적합 시킬 경우 많이 사용된다.

ARE 모형들의 적합성을 측정하는데 있어 결정계수 ( $R^2$ )와 잔차 분석의 RMSE를 사용하였다. 결정계수는 소비자물가  $Y_t$ 가 설명변수에 의해 어느 정도 설명되는지를 나타내는 지표로 값이 1에 가까울수록 설명력이 높은 좋은 모형이다. 예측값과 실제값의 분산정도를 나타내는 RMSE (Root Mean Square Error)의 공식은 다음과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2}, \quad (3.2)$$

여기서,  $o_i$ 와  $p_i$ 는 각각  $i$ 일의 관측 소비자물가와 추정 소비자물가를 의미하며,  $n$ 은 통계량에 사용된 일수이다. RMSE는 값이 작을수록 좋은 모형이다.

#### 4. 소비자물가 ARE 모형

4개 도시, 서울, 부산, 대구, 광주지역의 소비자물가 순차도표 (그림 4.1)를 통해 살펴보았다. 4개 도시 도표가 거의 비슷한 형태로 나타났다. 1988년부터 계속 꾸준히 증가하는 추세이고, 주목할 점은 IMF 기간 초기인 1997년 11월부터 급등하는 형태이다. 서울지역 경우, 1997년 11월에는 소비자물가지수가 66.177, 12월에는 67.621로 1998년 1월에는 69.245, 2월에는 70.207로 급등했고, 그 후는 서서히 증가하다가 1998년 9월에서 1999년 1월까지는 약하게 감소하다가 1999년 2월부터는 다시 점차적으로 증가하였다. 그러나 미국 글로벌 금융위기로 촉발된 세계 금융위기 영향으로, 우리나라 금융 위기가 한창이던, 2008년 5월-11월 그래프는 4지역 조금 다르게 나타났다. 이 기간 동안 서울지역 소비자물가는 크게 변동이 없는 반면, 부산지역은 약간의 변동이 있고, 대구와 광주지역은 비교적 변동이 크게 나타났다. 예를 들면, 대구지역 소비자물가지수가 2008년 5월에는 95.07, 6월에는 95.85, 7월에는 96.63, 8월에는 96.54, 9월에는 96.20, 10월에는 96.02로 계속 급등하다가, 11월에 95.59로 조금 낮아지며, 평년의 형태로 돌아왔다.

##### 4.1. 소비자물가 자료에 관한 ARE 모형

소비자물가는 그림 4.1에서 보듯이 추세가 있어 차분한 값을 분석에 이용하였다. ARE 모형의 변수 선택은 9종류의 설명변수들 간의 다중공선성 (multi-collinearity)을 고려하면서 설명변수들 간의 다양한 모형을 분석한 후 RMSE 값이 작은 모형을 선택하였다. ARE 모형의 설명변수 차수 선택은 경제변수임을 고려해, 당월자료 ( $t$ ), 전월자료 ( $t-1$ ), 2개월 전 자료 ( $t-2$ )까지 사용하였다.

ARE 모형의 모수 추정에는 YW (Yule-Walker), ULS (Unconditional Least Square), MLE (Maximum Likelihood Estimate) 등의 방법이 있지만, 본 연구에서는 소비자물가 모형과 같이 과거시점의 종속변수 (lagged dependent variable)인  $Y_{t-12}$  등이 설명변수로 사용되는 경우에 적합한 MLE 방법을 사용하였다.

ARE모형의 적합성 검증으로는 소비자물가 모형과 같이 과거시점의 종속변수가 설명변수로 사용되는 경우에 적합한 Durbin-h 검정 (조신섭과 이정형, 1997)을 사용해 검정하였고 또한 잔차의 ACF (Auto-Correlation Function)와 PACF (Partial Auto-Correlation Function)의 결과를 검정하였다 (조신섭과 이정형, 1997).

4개 도시, 서울, 부산, 대구, 광주지역에 관한 소비자물가에 관한 ARE모형을 분석한 결과를 표 4.1에 나타내었다. 각 도시에서 자료전체에 관한 분석과 상반기 (1월-6월), 하반기 (7월-12월)로 나누어 각각 분석하였다. 원래 분기별로 모형을 분석하여 차이를 보고 싶었지만, 자료수가 적어 상반기와 하반기로만 나누어 비교하였다. 사용된 모든 설명변수들은 유의수준 10%에서 유의한 변수들만 선택했다.

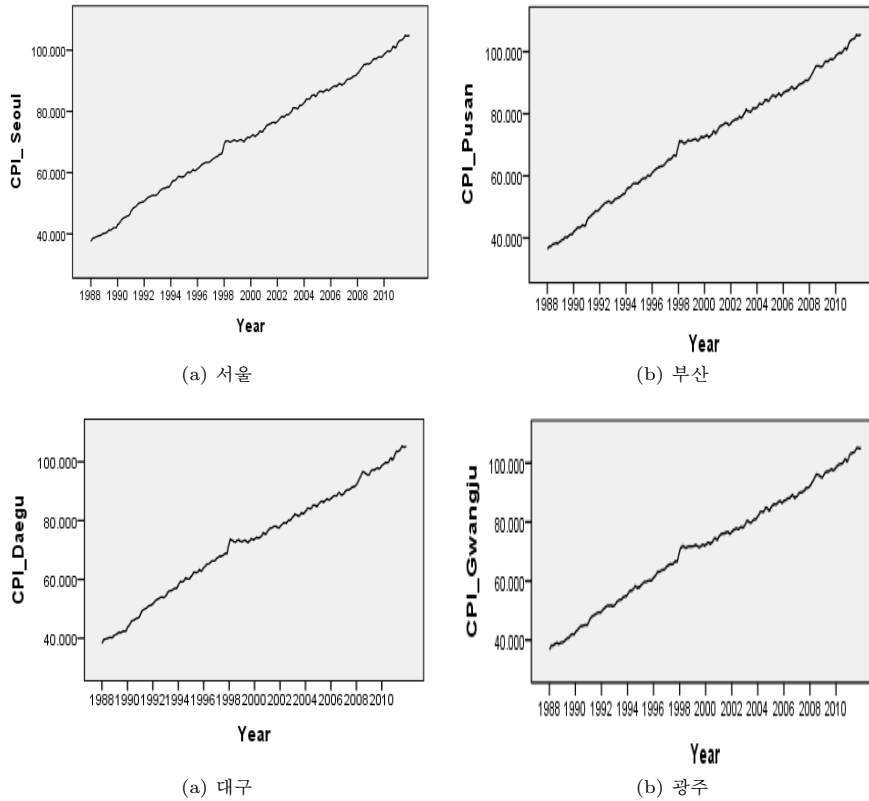


그림 4.1 지역별 소비자물가의 순차도표

먼저, 전체자료를 살펴보면, 지역별로 조금씩 다르게 나타났지만, 서울, 부산, 대구, 광주 4지역 모두 공통적으로 나타나는 소비자물가 설명변수는 1년전-4년전 소비자물가 ( $Y_{t-12}$ ,  $Y_{t-24}$ ,  $Y_{t-36}$ ,  $Y_{t-48}$ ) 와 미국 환율 ( $D_t$ ), 수입물가지수 ( $I_t$ ), 실업률 ( $U_t$ ) 로 분석되어, 이 변수들이 영향을 많이 주는 것으로 나타났다. 미국 환율과 수입물가지수의 계수는 양수로 나타나 환율과 수입물가지수가 올라갈수록 소비자물가도 높아짐을 알 수 있다. 그러나 실업률 계수는 음수로 나타나 소비자물가와 음의 관계가 있는 것으로 나타났다. 그 외에 서울을 제외한 부산, 대구, 광주지역엔 화폐 통화량 ( $M_t$ ) 이 설명변수로 포함되었고, 서울지역엔, 생산자물가 ( $P_t$ )가 설명변수로 포함된 것으로 분석되었다. 화폐 통화량의 계수는 음수로, 생산자물가 계수는 양수로 나타났다. 결정계수  $R^2$ 는 0.46에서 0.52로 나타났다. 예를 들면, 서울 소비자물가는 1년전-4년전 소비자물가 ( $Y_{t-12}$ ,  $Y_{t-24}$ ,  $Y_{t-36}$ ,  $Y_{t-48}$ ), 생산자물가 ( $P_t$ ), 환율 ( $D_t$ ), 수입물가지수 ( $I_t$ ), 실업률 ( $U_t$ )에 의해 52% 설명된다. 적합도 척도로 실제값과 예측값의 차이를 나타내는 RMSE 값은 0.25~0.32로 나타났다.

지역별로 상반기와 하반기로 나누어 분석한 결과를 살펴보면, 서울지역 하반기모형에서는 전체자료에서는 설명변수로 포함되지 않았던 전월 원유수입단가 ( $O_{t-1}$ )가 포함되어 있고 부산 상반기모형과 대구 상반기모형에서도 당월 원유수입단가 ( $O_t$ )가 포함되어 있어 원유수입단가도 소비자물가에 영향을 주는 것으로 나타났다. 광주 하반기모형에서는 경기동행지수 ( $B_t$ )가 포함되어 있다. 각 지역 상반기와 하반기 모형들의 결정계수  $R^2$ 는 0.36~0.55로 대부분 전체모형 보다 작게 나타났고, 대구

하반기 모형은 0.55로 조금 높게 나타났다. RMSE는 0.26~0.35로 각 지역 전체자료 모형과 거의 비슷하게 나타났다.

본 연구에서 사용한 설명변수는 통계청에서 쉽게 구할 수 있는 자료이므로, 표 4.1을 이용하면, 소비자물가를 예측하는데 도움이 될 것으로 생각된다.

표 4.1 지역별 소비자물가의 ARE 모형

지역	기간	ARE 모형	R <sup>2</sup>	RMSE
전체	전체	$Y_t = -0.27 - 0.84Y_{t-12} - 0.67Y_{t-24} - 0.34Y_{t-36} - 0.20Y_{t-48} - 0.02P_t + 0.001D_t + 0.01I_t - 0.06U_t + \epsilon_t,$	0.52	0.25
		$\epsilon_t = 0.15\epsilon_{t-1} + e_t$	0.52	0.25
			0.52	0.25
서울	상반기	$Y_t = -0.68Y_{t-12} - 0.27Y_{t-24} + 0.001D_t - 0.05U_t + \epsilon_t,$	0.41	0.26
		$\epsilon_t = 0.04\epsilon_{t-1} - 0.71\epsilon_{t-6} + e_t$	0.41	0.26
	하반기	$Y_t = -0.59 - 0.72Y_{t-12} - 0.43Y_{t-24} - 0.35Y_{t-36} - 0.31Y_{t-48} + 0.001D_t + 0.003O_{t-1} + 0.007I_t - 0.04U_t + \epsilon_t,$	0.43	0.26
		$\epsilon_t = -1.09\epsilon_{t-6} - 0.30\epsilon_{t-12} + e_t$	0.43	0.26
			0.43	0.26
			0.43	0.26
부산	전체	$Y_t = -0.91 - 0.78Y_{t-12} - 0.59Y_{t-24} - 0.34Y_{t-36} - 0.21Y_{t-48} + 0.001D_t + 0.009I_t - 0.10U_t - 0.001M_t + \epsilon_t,$	0.49	0.31
		$\epsilon_t = -0.17\epsilon_{t-3} + e_t$	0.49	0.31
			0.49	0.31
	상반기	$Y_t = -0.58Y_{t-12} - 0.22Y_{t-24} + 0.001D_t + 0.003O_t - 0.06U_t + \epsilon_t,$	0.37	0.32
		$\epsilon_t = 0.68\epsilon_{t-6} + e_t$	0.37	0.32
			0.37	0.32
하반기	$Y_t = 0.56 - 0.09Y_{t-12} - 0.55Y_{t-24} - 0.22Y_{t-36} - 0.40Y_{t-48} + 0.001D_t - 0.004A_t + \epsilon_t,$	0.36	0.31	
	$\epsilon_t = -0.95\epsilon_{t-6} - 0.77\epsilon_{t-12} - 0.68\epsilon_{t-18} + e_t$	0.36	0.31	
		0.36	0.31	
대구	전체	$Y_t = -0.93 - 0.85Y_{t-12} - 0.63Y_{t-24} - 0.35Y_{t-36} - 0.20Y_{t-48} + 0.001D_t + 0.01I_t - 0.11U_t - 0.0001M_t + \epsilon_t,$	0.51	0.30
		$\epsilon_t \equiv 0.15\epsilon_{t-1} - 0.20\epsilon_{t-3} + e_t$	0.51	0.30
			0.51	0.30
	상반기	$Y_t = -0.57Y_{t-12} - 0.21Y_{t-24} + 0.001D_t + 0.003O_t - 0.06U_t + \epsilon_t,$	0.41	0.30
		$\epsilon_t = -0.10\epsilon_{t-3} - 0.73\epsilon_{t-6} + e_t$	0.41	0.30
			0.41	0.30
하반기	$Y_t = -0.94 - 0.82Y_{t-12} - 0.58Y_{t-24} - 0.34Y_{t-36} - 0.23Y_{t-48} + 0.001D_t + 0.01I_t - 0.05U_t + \epsilon_t,$	0.55	0.30	
	$\epsilon_t = -1.29\epsilon_{t-6} - 0.50\epsilon_{t-12} + e_t$	0.55	0.30	
		0.55	0.30	
광주	전체	$Y_t = -0.80 - 0.75Y_{t-12} - 0.58Y_{t-24} - 0.34Y_{t-36} - 0.17Y_{t-48} + 0.001D_t + 0.01I_t - 0.11U_t - 0.001M_t + \epsilon_t,$	0.46	0.32
		$\epsilon_t = -0.15\epsilon_{t-3} - 0.17\epsilon_{t-15} - 0.16\epsilon_{t-18} + e_t$	0.46	0.32
			0.46	0.32
	상반기	$Y_t = -0.69 - 0.54Y_{t-12} - 0.25Y_{t-24} + 0.001D_t + 0.07I_t - 0.07U_t + \epsilon_t,$	0.36	0.31
		$\epsilon_t = -0.66\epsilon_{t-6} + e_t$	0.36	0.31
			0.36	0.31
하반기	$Y_t = -3.19 - 0.53Y_{t-12} - 0.25Y_{t-24} - 0.16Y_{t-36} - 0.15Y_{t-48} + 0.002D_t + 0.02B_t + 0.03I_t - 0.12U_t + \epsilon_t,$	0.36	0.35	
	$\epsilon_t = -0.87\epsilon_{t-6} - 0.14\epsilon_{t-12} + e_t$	0.36	0.35	
		0.36	0.35	

#### 4.2. ARE 모형의 실제값과 추정치의 산점도

소비자물가 실제값과 ARE 모형 추정치의 산점도를 서울과 광주지역 전체모형을 그림 4.2에 나타내었다. 그림 4.2는 X축은 소비자물가의 실제값이고, Y축은 ARE 모형 추정치를 나타낸다. 그러므로 원점을 지나는 대각선 주변에 값들이 많을수록 잘 적합 된다고 볼 수 있다. 서울지역과 광주지역 둘 다 비교적 적합이 잘 되었으나, 상대적으로 광주지역이 대각선에서 벗어나는 값들이 좀 더 많이 나타났다. 두 지역 모두 소비지물가지수가 높은 경우에는 약간 과소추정 되는 경향이 있는 것으로 나타났다.

ARE 모형은 소비자물가지수에 영향을 주는 설명변수를 수식으로 표현하는 장점이 있으므로, 모형에 소비자물가에 영향을 주는 경제 설명변수를 좀 더 포함시켜 개발하면, 실제값에 좀 더 가까이 접근하는 좋은 결과를 가질 수 있을 것으로 기대한다.



## 참고문헌

- 김기호, 윤성훈 (2009). 소비자물가에 대한 유가 및 환율충격의 비대칭성, 비선형성 분석. <국제경제연구>, **15**, 131-152.
- 김영덕 (2000). 국제유가와 소비자물가의 변동. <자원·환경경제연구>, **9**, 373-391.
- 김효을 (2008). 소비자물가와 생산자물가 간의 구조적 격차. <한국무역학회 무역학자 전국대회 학술대회 발표 논문집>, 363-383.
- 심주용, 이장택 (2010). 비선형 평균 일반화 이분산 자기회귀모형의 추정. <한국데이터정보과학회>, **21**, 831-839.
- 이충열 (1998). 소비자물가와 생산자물가의 관계 및 소비자물가 변동요인에 관한 연구. <KIF 금융논문집>, 101-121.
- 이훈자 (2010). 경기도 수원시 미세먼지 농도의 시계열모형 연구, <한국데이터정보과학회>, **21**, 1117-1124.
- 임성식 (2011). 시계열모형에 의한 인플레이션 예측력 비교. <서경대학교 산업기술연구소 논문집>, **26**, 19-26.
- 정동빈 (2007). 실물경기변동에 관한 연구. <한국국제회계학회 추계 공동국제학술발표논문집>, 1-5.
- 조신섭, 이정형 (1997). <SAS/ETS를 이용한 경제시계열 분석>, 자유아카데미, 서울.
- 주세우, 이민환, 황규선 (2010). 구조적 VAR 모형을 이용한 환율전가 효과분석: 한국의 수입물가와 소비자물가를 중심으로. <경제연구>, **28**, 85-108.
- 황창하, 신사임 (2010). 커널기계 기법을 이용한 일반화 이분산자기회귀모형 추정. <한국데이터정보과학회>, **21**, 419-425.
- Hwang, H. (2010). Variable selection for multiclassification by LS-SVM, *Journal of the Korean Data & Information Science Society*, **21**, 959-965.

## Analysis of time series models for consumer price index

Hoonja Lee<sup>1</sup>

<sup>1</sup>Department of Information Statistics, Pyeongtaek University

Received 30 April 2012, revised 16 May 2012, accepted 21 May 2012

### Abstract

The consumer price index (CPI) data is one of the important economic measurement of the country. In this article, the Autoregressive Error (ARE) model has been considered for analyzing the monthly CPI data at Seoul, Pusan, Daegu, and Gwangju Cities in Korea. In the ARE model, nine economic variables are used as the explanatory variables for the CPI data set. The nine explanatory variables are CCI (coincident composite index), won-dollar rate, producer price index, oil import price, oil import volume, international current account, import price index, unemployment rate, and amount of currency. The result showed that the monthly ARE models explained about 46-52% for describing the CPI.

*Keywords:* Autoregressive error model, consumer price index, explanatory variable.

---

<sup>1</sup> Associate professor, Department of Information Statistics, Pyeongtaek University, Pyeongtaek, 450-701, Korea. E-mail: esther@ptu.ac.kr