
시계열 예측 기법에 대한 비교 분석*

- 상태공간모형을 이용한 지수평활법 및 시변자기회귀모형을 중심으로 -

계절성이 있는 월별 시계열자료의 예측에 일반적으로 이용되고 있는 방법으로는 계절ARIMA모형, Holt윈터스의 가법형 및 승법형 지수평활법을 들 수 있다. 한편, 상태공간모형 또는 동태적 선형모형은 이론적으로 베이지안 기법을 기초로 하고 있는 방법으로 모형설정방법에 따라 응용범위가 매우 넓어 시계열분석에 다양하게 사용되고 있다. 본 연구에서는 지수평활법을 상태공간모형으로 전환시켜 확률모형화한 지수평활법과 자기회귀모형에 시간 변동성을 고려한 시변자기회귀모형을 각각 살펴보고 그 예측력을 평가해 보았다.

본 연구에서는 우리나라 중분류 월별 산업생산지수 시계열을 계절ARIMA모형, 동태적 선형모형의 하나인 계절효과모형에 의해 예측해 보고 상태공간모형을 이용한 지수평활법과 시변자기회귀모형에 따른 예측결과와 비교분석해 보았다. 상태공간모형을 이용한 지수평활법과 시변자기회귀모형이 ARIMA모형에 비해 모든 시계열에서 우수한 결과를 보여주지는 못하지만 시계열이 가지고 있는 특성을 구성성분별 또는 시점별로 파악할 수 있다는 측면에서 단기시계열 예측에서의 활용가능성을 확인할 수 있었다.

* 본고는 경제통계국 국민소득팀 강창구 과장이 집필하였음. 본고의 내용은 집필자의 개인의견으로 한국은행의 공식견해를 나타내는 것이 아님.

<p>I. 서론</p> <p>II. 시계열의 예측 기법</p> <p>1. 개요</p> <p>2. HKSG모형 및 TVAR모형</p> <p>가. 상태공간모형을 이용한 지수평활법 (HKSG모형)</p> <p>나. 시변자기회귀모형(TVAR모형)</p>	<p>III. 실증 분석</p> <p>1. 분석 절차</p> <p>2. 분석 결과 및 예측력 비교</p> <p>IV. 결론</p>
---	--

I. 서론

시계열은 경제현상이나 자연현상에 대한 시간적 변화를 나타내는 역사적 계열(historical series)이므로 특정 시점의 시계열 관측치는 그 이전 자료들에 주로 의존하게 된다. 따라서 시계열분석을 통한 예측에서는 관측된 과거 자료들을 분석하여 규칙적인 패턴을 발견해서 이를 모형화하여 추정하고, 이 추정된 모형을 사용하여 미래시점의 관측치를 예측하게 된다. 시계열은 그 특성에 따라 다양한 형태로 나타날 수 있기 때문에 모든 시계열에 공통적으로 적합한 최적의 이론적, 객관적 모형은 존재할 수가 없다. 따라서 각 시계열별로 그 특성에 가장 적합한 예측모형을 각각 선정할 필요가 있다.

시계열 예측방법은 분해법, 지수평활법 등 전통적 시계열 분석 방법과 ARIMA모형으로 대표되는 모형기반(model-based) 접근법을 들 수 있다. 분해법은 경기변동을 예측하기 위해 시계열을 구성성분별로 따로 구분한 후 이를 이용하여 미래를 예측하거나 계절조정을 위해 주로 사용되고 있다.

지수평활법은 최근 자료에 더 큰 가중치를 주고 과거로 갈수록 가중치를 지수적으로 감소시킴으로써 시계열이 생성되는 특성에 변화가 있을 경우 이 변화를 쉽게 반영할 수 있을 뿐만 아니라 이상치(outlier)가 존재할 경우 로버스트(robust)하다는 장점을 가지고 있다. 반면 지수평활법은 필터링 효과를 결정해 주는 평활상수의 선택이 임의적일 뿐만 아니라 확률모형에 의한 우도함수의 계산 및 예측구간을 구할 수 없다는 단점을 가지고 있다.

ARIMA모형은 현 시점의 관측치를 과거 관측치들과 백색잡음의 선형결합형태로 표현하는 모형으로 단기예측력이 우수하여 시계열 예측에 일반적으로 사용되고 있다. 그러나 ARIMA모형은 모형 식별시 주관적인 판단이 요구되고 차분 또는 분산안정화를 위한 변수변환 등의 절차를 필요로 하고 있다.

따라서 본고에서는 먼저 개별 지수평활법의 이용상 한계를 보완하기 위해 단순지수평활법, 홀트의 이중지수평활법, 홀트-윈터스의 계절지수평활법 등 다양한 지수평활법을 상태공간모형을 이용하여 하나의 확률적 예측모형체계로 구축하고 각 지수평활모형 중에서 해당시계열에 최적 모형을 자동적으로 결정해 주는 Hyndman, Koehler, Snyder, Grose(2002)의 방법(이하 HKSG모형)을 소개하고 동 방식에 의한 단기예측력을 평가해 보았다.

또한 ARIMA모형의 일종인 AR모형을 확장하여 동태적 선형모형(Dynamic Linear Model: DLM)에 의해 시점별로 모수를 별도로 추정함으로써 시점별 변화를 반영한 시변자기회귀모형(Time-Varying Auto Regressive model: TVAR모형)[Prado,

Huerta, and West(2001)]을 소개하고 단기 예측력 입장에서 비교 분석하고자 한다.

II장에서는 시계열예측에 일반적으로 사용되고 있는 예측방법을 간략히 살펴본 후 HKSG모형과 TVAR모형에 대해 각각 소개하였다. III장에서는 위에서 제시한 각 시계열 예측기법의 비교분석을 위하여 통계청의 월별 산업생산지수를 이용한 실증분석 결과를 요약하였다.

II. 시계열의 예측 기법

이 장에서는 시계열의 여러 가지 예측 방법들을 살펴보기로 한다. 1절에서는 일변량 시계열 자료분석에서 많이 사용되는 ARIMA모형, 지수평활법 뿐만 아니라 GARCH 등과 같은 변동성 모형, 동태적 선형모형의 일종인 DLM-계절효과모형 등에 대하여 알아보았다. 2절에서는 상태공간모형으로의 전환을 통해 지수평활법을 확률모형화한 HKSG모형과 전통적인 AR모형에 시점별 변동을 반영한 TVAR모형을 소개하였다.

1. 개요

Box and Jenkins의 ARIMA모형은 시계열을 일련의 확률과정 (stochastic process)으로 간주하고 평균과 분산이 시간에 무관하게 상수인 정상성을 만족하는 시계열에서만 모형의 식별이 가능하게 되어있다. 따라서 비정상시계열에 대해서는 차분과 분산안정화 등의 과정을 통한 정상시계열로의 전환 여부에 대한 판단을 필요로 하고 잠정모형의 식별단계, 특히 확률적 추세를 갖는 경우 자기상관함수 또는 편자기상관함수를 이용하는 데에 있어 주관성이 개입될 수 있다. 계절성과 추세를 가지고 있는 시계열은 다음과 같은 계절형 ARIMA(p, d, q)(P, D, Q)_s 모형으로 적합 시킬 수 있다.

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^Dy_t = \theta(B)\Theta(B^s)\omega_t$$

여기서 $\phi(B)$ 는 p차 AR후향연산자, $\theta(B)$ 는 q차 MA후향연산자, $\Phi(B^s)$ 는 P차 계절AR후향연산자, $\Theta(B^s)$ 는 Q차 계절MA후향연산자를 나타낸다. 이 때, s는 계절주기로 월자료에서는 12, 분기자료에서는 4를 갖는다. d와 D는 정상화에 필요한 차분 및 계절차분의 횟수를 나타내고 ω_t 는 평균 0, 분산 σ^2 을 갖는 백색잡음이다.

지수평활법은 시계열의 미래값을 예측하기 위해 가장 최근값에는 가장 큰 가중치를 주고 과거로 갈수록 가중치를 지수적으로 줄여 이동가중평균을 취함으로써 불규칙변동을 완화시키는 방법이다. 지수평활법은 확률모형에 의존하지 않는 대표적인 경험적 분

석방법 중 하나로 Brown(1959), Holt(1957), Winters(1960)등에 의해 다양한 방법이 제안되었다. 가장 간단한 방법으로서 단순지수평활법은 국지적으로는 동일한 평균수준을 갖지만 시간대별로 평균수준이 변화하는 자료에 적합하다. Holt는 시간대별로 추세가 변화하는 자료에 적절한 이중지수평활법을, Brown은 다항추세를 갖는 자료에 적합한 삼중지수평활법을 각각 제시하였다. 계절성을 가지고 있는 시계열에 대해서는 시계열의 변동폭이 시간의 흐름에 따라 일정한 경우에는 Winters의 가법형 계절지수평활법을, 시계열의 변동폭과 계절성이 추세에 따라 변화하는 경우에는 Winters의 승법형 계절지수평활법을 적용할 수 있다.

Engel(1982)은 일정시점에서 일어나는 충격이 일정기간 지속되는 변동성 클러스터링(volatility clustering)을 고려한 ARCH모형을 제안하였다. Bollerslev(1986)은 ARCH모형을 일반화한 GARCH모형으로 발전시켰으며, 이 후 IGARCH(Integrated GARCH), AR-GARCH 등 다양한 모형으로 확대 연구되고 있다. 황선영(2005)은 우리나라의 산업생산지수와 경기동행지수에 GARCH류 모형을 이용하여 경제시계열 예측에서의 활용 가능성을 제시하였다.

DLM-계절효과모형(Seasonality Effect model)은 동태적 선형모형을 이용하여 시계열을 추세순환성분, 계절성분으로 분해하여 분석하는 방법이다. 추세순환성분은 그 수준(μ_t)과 증가률(β_t)이 시간 t 에 따라 변화한다고 가정하고 이를 확률보행과정으로 근사

시켰다. 계절성(s_t)의 시점별 변화도 함께 고려한 DLM-계절효과 모형은 다음과 같다.

$$\begin{aligned}
 y_t &= \mu_t + s_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\
 \mu_t &= \mu_{t-1} + \beta_{t-1} + \xi_t, & \xi_t &\sim N(0, \sigma_\xi^2) \\
 \beta_t &= \beta_{t-1} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2) \\
 s_t &= -\sum_{j=1}^{m-1} s_{t-j} + \omega_t, & \omega_t &\sim N(0, \sigma_\omega^2)
 \end{aligned}$$

여기서 m 은 계절주기로 분기자료는 4, 월별자료는 12가 된다. 이 방법은 시계열의 추세순환성분과 계절성분의 변화를 파악할 수 있다는 장점이 있다.

2. HKSG모형 및 TVAR모형

가. 상태공간모형을 이용한 지수평활법(HKSG모형)

지수평활법은 미래값에 대한 정보를 과거와 현재 관측값의 가중평균 형태로 예측하는 기법으로 시계열의 특성에 따라 적용할 수 있는 다양한 기법들이 있다. Pegels(1969)와 Gardner(1985)는 지수평활법을 시계열이 가지고 있는 추세성분의 특성에 따라 상수형(N), 가법형(A), 승법형(M) 및 감폭추세형(D)으로 구분하고 또한 계절성의 변화양상에 따라 비계절성(N), 가법계절성(A), 승법계절성(M)의 3가지로 구분하여 총 12가지의 지수평활법을 포괄하는 체계를 구축하였다. <표 1>에서 NN은 단순지수평활법, AA는

가법형 Holt윈터스의 지수평활법, AM은 승법형 Holt윈터스의 지수평활법을 각각 나타낸다.

<표 1>

추세와 계절성에 따른 지수평활법

		계절성		
		비계절성(N)	가법계절성(A)	승법계절성(M)
추 세	상수(N)	NN	NA	NM
	가법(A)	AN	AA	AM
	승법(M)	MN	MA	MM
	감폭(D)	DN	DA	DM

12개 지수평활법은 다음과 같이 한 개의 방정식 체계로 표현될 수 있다.

$$\begin{aligned}
 l_t &= \alpha P_t + (1 - \alpha) Q_t, \\
 b_t &= \beta R_t + (\phi - \beta) b_{t-1}, \\
 s_t &= \gamma T_t + (1 - \gamma) s_{t-m}
 \end{aligned}
 \tag{1}$$

여기서, l_t 는 t 시점에서의 시계열의 수준값을 나타내며, b_t 와 s_t 는 t 시점에서의 추세의 기울기와 계절성분을 나타낸다. m 은 계절주기로서 분기자료의 경우에는 4, 월자료는 12의 값을 갖는다. $\alpha, \beta, \phi, \gamma$ 는 상수이다. 동 방정식 체계에 <표 2>에 주어진 모형별 P_t, Q_t, R_t, T_t 를 대입하면 12개의 지수평활법으로 유도할 수가 있다. <표 2>에서 F_{t+h} 는 t 시점에서의 h 시점 후 예측치를 나타낸다.

식 (1)은 다음과 같은 오차수정모형으로 표현할 수 있다.

$$\begin{aligned}
 l_t &= Q_t + \alpha(P_t - Q_t), \\
 b_t &= \phi b_{t-1} + \beta(R_t - b_{t-1}), \\
 s_t &= s_{t-m} + \gamma(T_t - s_{t-m})
 \end{aligned}
 \tag{2}$$

식 (2)에서 $\alpha=0$ 이면 고정상수, $\beta=0$ 이면 고정추세(drift), $\gamma=0$ 이면 고정 계절패턴을 반영한 방법임을 알 수 있다.

<표 2>

추세와 계절성에 따른 P_t, Q_t, R_t, T_t

		계절성		
		비계절성(N)	가법계절성(A)	승법계절성(M)
추세	상수(N)	$P_t = y_t$ $Q_t = l_{t-1}$ $R_t = \text{없음}$ $T_t = \text{없음}$ $F_{t+h} = l_t$	$P_t = y_t - s_{t-m}$ $Q_t = l_{t-1}$ $R_t = \text{없음}$ $T_t = y_t - Q_t$ $F_{t+h} = l_t + s_{t+h-m}$	$P_t = y_t / s_{t-m}$ $Q_t = l_{t-1}$ $R_t = \text{없음}$ $T_t = y_t / Q_t$ $F_{t+h} = l_t s_{t+h-m}$
	가법(A)	$P_t = y_t$ $Q_t = l_{t-1} + b_{t-1}$ $R_t = l_t - l_{t-1}$ $T_t = \text{없음}$ $F_{t+h} = l_t + hb_t$	$P_t = y_t - s_{t-m}$ $Q_t = l_{t-1} + b_{t-1}$ $R_t = l_t - l_{t-1}$ $T_t = y_t - Q_t$ $F_{t+h} = l_t + hb_t + s_{t+h-m}$	$P_t = y_t / s_{t-m}$ $Q_t = l_{t-1} + b_{t-1}$ $R_t = l_t - l_{t-1}$ $T_t = y_t / Q_t$ $F_{t+h} = (l_t + hb_t) s_{t+h-m}$
	승법(M)	$P_t = y_t$ $Q_t = l_{t-1} b_{t-1}$ $R_t = l_t / l_{t-1}$ $T_t = \text{없음}$ $F_{t+h} = l_t b_t^h$	$P_t = y_t - s_{t-m}$ $Q_t = l_{t-1} b_{t-1}$ $R_t = l_t / l_{t-1}$ $T_t = y_t - Q_t$ $F_{t+h} = l_t b_t^h + s_{t+h-m}$	$P_t = y_t / s_{t-m}$ $Q_t = l_{t-1} b_{t-1}$ $R_t = l_t / l_{t-1}$ $T_t = y_t / Q_t$ $F_{t+h} = l_t b_t^h s_{t+h-m}$
	감폭(D)	$P_t = y_t$ $Q_t = l_{t-1} + b_{t-1}$ $R_t = l_t - l_{t-1}$ $T_t = \text{없음}$ $F_{t+h} = l_t + (1 + \phi + \dots + \phi^{h-1}) b_t$	$P_t = y_t - s_{t-m}$ $Q_t = l_{t-1} + b_{t-1}$ $R_t = l_t - l_{t-1}$ $T_t = y_t - Q_t$ $F_{t+h} = l_t + (1 + \phi + \dots + \phi^{h-1}) b_t + s_{t+h-m}$	$P_t = y_t / s_{t-m}$ $Q_t = l_{t-1} + b_{t-1}$ $R_t = l_t - l_{t-1}$ $T_t = y_t / Q_t$ $F_{t+h} = [l_t + (1 + \phi + \dots + \phi^{h-1}) b_t] s_{t+h-m}$

* y_t 는 관측치를 나타냄

** ϕ 는 추세가 감폭(D)인 경우 $\beta < \phi < 1$ 이고 나머지 경우에는 $\phi = 1$ 로 설정

Hyndman, Koehler, Snyder, Grose(2002)는 위의 12개 지수평활법에 오차항을 가정한 상태공간모형으로 나타냄으로써 우도함수, AIC 통계량 및 예측구간의 계산이 가능한 확률모형으로 전환하였다. HKSG모형을 상태벡터 $\theta_t = (l_t, b_t, s_t, \dots, s_{t-(m-1)})$ 을 갖는 상태공간모형으로 나타내면 다음과 같다.

$$\begin{aligned} \text{[관측방정식]} : y_t &= h(\theta_{t-1}) + k(\theta_{t-1})\omega_t \\ \text{[상태방정식]} : \theta_t &= f(\theta_{t-1}) + g(\theta_{t-1})\omega_t \end{aligned} \quad (3)$$

여기서 오차항 ω_t 는 평균이 0이고 분산이 σ^2 인 백색잡음으로 가정한다. Ord(1997)은 동 방정식의 오차항 ω_t 를 각각 가법형과 승법형으로 구분하여 방정식체계를 24개 모형으로 확장하였다. 식 (3)에서 오차항 ω_t 가 가법형, 즉 $h(\theta_{t-1}) = m_t, k(\theta_{t-1}) = 1$ 을 따르고 가정한 관측방정식 $y_t = m_t + \omega_t$ 에 대해 함수 f, g 를 적절하게 설정하게 되면 <표 3>과 같은 12개의 지수평활법별 상태방정식으로 변환된다. 승법형 오차항을 가정하는 경우에는 $h(\theta_{t-1}) = m_t, k(\theta_{t-1}) = m_t$ 이 되고 <표 3>에서 ω_t 대신 $m_t\omega_t$ 로 바꾸어 대입한다.

상태공간모형에 의해 확률모형을 설정하게 됨에 따라 다음과 같은 우도함수를 통해 최우추정량을 산출할 수가 있다. 즉 식 (4)의 우도함수를 최대화 하는 모수($\alpha, \beta, \phi, \gamma$)와 상태벡터의 초기값 $\theta_0 = (l_0, b_0, s_0, \dots, s_{-m+1})$ 을 추정한다.

$$L^*(\alpha, \beta, \gamma, \phi, \theta_0) = n \log \left(\sum_{t=1}^n e_t^2 / k^2(\theta_{t-1}) \right) + 2 \sum_{t=1}^n \log |k(\theta_{t-1})| \quad (4)$$

<표 3>

12개의 지수평활법에 대한 상태방정식(가법형 오차 가정)

		계절성		
		비계절성(N)	가법계절성(A)	승법계절성(M)
추 세	상수 (N)	$m_t = l_{t-1}$ $l_t = l_{t-1} + \alpha\omega_t$	$m_t = l_{t-1} + s_{t-m}$ $l_t = l_{t-1} + \alpha\omega_t$ $s_t = s_{t-m} + \gamma\omega_t$	$m_t = l_{t-1}s_{t-m}$ $l_t = l_{t-1} + \alpha\omega_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\omega_t/l_{t-1}$
	가법 (A)	$m_t = l_{t-1} + b_{t-1}$ $l_t = l_{t-1} + b_{t-1} + \alpha\omega_t$ $b_t = b_{t-1} + \alpha\beta\omega_t$	$m_t = l_{t-1} + b_{t-1} + s_{t-m}$ $l_t = l_{t-1} + b_{t-1} + \alpha\omega_t$ $b_t = b_{t-1} + \alpha\beta\omega_t$ $s_t = s_{t-m} + \gamma\omega_t$	$m_t = (l_{t-1} + b_{t-1})s_{t-m}$ $l_t = l_{t-1} + b_{t-1} + \alpha\omega_t/s_{t-m}$ $b_t = b_{t-1} + \alpha\beta\omega_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\omega_t/(l_{t-1} + b_{t-1})$
	승법 (M)	$m_t = l_{t-1}b_{t-1}$ $l_t = l_{t-1}b_{t-1} + \alpha\omega_t$ $b_t = b_{t-1} + \alpha\beta\omega_t/l_{t-1}$	$m_t = l_{t-1}b_{t-1} + s_{t-m}$ $l_t = l_{t-1}b_{t-1} + \alpha\omega_t$ $b_t = b_{t-1} + \alpha\beta\omega_t/l_{t-1}$ $s_t = s_{t-m} + \gamma\omega_t$	$m_t = (l_{t-1}b_{t-1})s_{t-m}$ $l_t = l_{t-1}b_{t-1} + \alpha\omega_t/s_{t-m}$ $b_t = b_{t-1} + \alpha\beta\omega_t/(s_{t-m}l_{t-1})$ $s_t = s_{t-m} + \gamma\omega_t/(l_{t-1}b_{t-1})$
	감폭 (D)	$m_t = l_{t-1} + b_{t-1}$ $l_t = l_{t-1} + b_{t-1} + \alpha\omega_t$ $b_t = \phi b_{t-1} + \alpha\beta\omega_t$	$m_t = l_{t-1} + b_{t-1}$ $l_t = l_{t-1} + b_{t-1} + \alpha\omega_t$ $b_t = \phi b_{t-1} + \alpha\beta\omega_t$ $s_t = s_{t-m} + \gamma\omega_t$	$m_t = (l_{t-1} + b_{t-1})s_{t-m}$ $l_t = l_{t-1} + b_{t-1} + \alpha\omega_t/s_{t-m}$ $b_t = \phi b_{t-1} + \alpha\beta\omega_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\omega_t/(l_{t-1} + b_{t-1})$

여기서 $e_t = k(\theta_{t-1})\omega_t$ 이다. 24개 지수평활모형 중 최적예측모형을 선정하기 위해서는 다음과 같은 AIC통계량을 이용한다.

$$AIC = L^*(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\phi}, \hat{\theta}_0) + 2p$$

여기서 p 는 각 모형별 모수($\alpha, \beta, \phi, \gamma$)의 개수이다.

나. 시변자기회귀(TVAR) 모형

TVAR모형은 AR모형의 자기회귀계수가 시점별로 동일하다는 가정 대신, 시간이 변하면서 다른 모수값을 갖는 동태적모형을 가

정하였다. 즉, 현재 관측치가 과거 자료의 선형결합이라는 AR모형에 자기회귀계수 및 분산에 시변(time-varying)이라는 동태적성질을 추가함으로써 단기예측력 향상을 도모하고자 하였다.

TVAR모형을 방정식으로 나타내면 다음과 같다.

$$y_t = \phi_{t,1}y_{t-1} + \phi_{t,2}y_{t-2} + \dots + \phi_{t,p}y_{t-p} + \omega_t$$

여기서 ω_t 는 평균이 0이고 시변 분산 σ_t^2 을 갖는다. 시변계수벡터 $\phi_t = (\phi_{t,1}, \phi_{t,2}, \dots, \phi_{t,p})'$ 는 다음과 같은 확률보행과정을 따른다고 가정한다.

$$\phi_t = \phi_{t-1} + \xi_t, \quad \xi_t \sim N(0, W_t)$$

이 때 시변계수벡터 ϕ_t 는 시점에 따른 ϕ_t 의 변동폭을 조절하는 분산 W_t 에 의해 조정된다. 시변 분산 σ_t^2 도 다음의 승법 확률보행과정을 따르며 할인 상수 δ 에 의해 조정된다. 할인상수의 값은 0과 1사이의 값을 가지며 일반적으로 0.9~0.99의 값을 사용한다. [West and Harrison(1997)]

$$\sigma_t^2 = \sigma_{t-1}^2 \cdot \left(\frac{\delta}{\eta_t} \right)$$

여기서 η_t 는 베타분포를 따르는 확률변수이다.

한편, West(1997)는 동태적 선형모형을 이용하여 시계열이 잠재된(latent) 몇 개의 성분으로 이루어져 있음을 증명하였다. 즉 시계열을 추세순환성분과 계절성분 등 체계적 변동과 불규칙 변동으로 나누는 일반적인 분해법과는 달리 시계열을 다수의 확률과정

성분으로 분해가 가능함을 보였다. TVAR모형도 아래와 같이 동태적 선형모형으로 전환할 수 있다.

$$[\text{관측방정식}] : y_t = F'\theta_t + \omega_t$$

$$[\text{상태방정식}] : \theta_t = G_t\theta_{t-1} + \epsilon_t$$

상태벡터(θ), 오차항(ω_t, ϵ_t) 및 관측벡터(F), 상태행렬(G_t)은 아래와 같다.

$$F = (1, 0, \dots, 0)', \quad \theta_t = (y_t, y_{t-1}, \dots, y_{t-p+1}), \quad \omega_t = 0, \quad \epsilon_t = F'\omega_t,$$

$$G_t = G(\phi_t) = \begin{pmatrix} \phi_{t,1} & \phi_{t,2} & \dots & \phi_{t,p-1} & \phi_{t,p} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

Prado, Huerta, West(2001)는 동태적 선형모형화한 TVAR모형이 아래와 같이 분해가 됨을 보였다.

$$y_t = \sum_{j=1}^C z_{t,j} + \sum_{j=1}^R a_{t,j}$$

이 때, $z_{t,j}$ 는 ARMA(2, 1)을 따르며 $a_{t,j}$ 는 AR(1)을 따르는 확률과정이고 $z_{t,j}$ 와 $a_{t,j}$ 의 개수는 상태행렬 G_t 의 특성값의 복소수, 실수의 개수에 의해 각각 얻어진다. 따라서 시계열이 다수의 확률과정의 결합으로 표현될 수 있음을 알 수 있다.

Ⅲ. 실증 분석

1. 분석 절차

한국은행에서는 2005년 3/4분기부터 분기GDP속보치를 분기종료 후 약 25일내에 발표하고 있다. 그러나 분기GDP추계에 있어 활용도가 높은 월별 산업생산지수는 익월말경 발표됨에 따라 분기GDP속보치 추계시점에서 산업생산지수 3개월째 자료를 이용하는 것이 현실적으로 불가능하다. 본장에서는 ARIMA모형, DLM-계절효과모형, TVAR모형 및 HKSG모형 등 II장에서 제시한 각종 시계열모형을 이용해 월별 산업생산지수의 단기예측력을 비교해 봄으로써 분기GDP속보치 추계에서의 활용가능성을 평가해 보았다.

여기서는 23개 중분류 업종별 산업생산지수 중에서 6개 업종(음식료제조(D15), 봉제의복 및 모피제품 제조(D18), 출판, 인쇄 및 기록매체 복제(D22), 컴퓨터 및 사무용 기기 제조(D30), 전자부품, 영상, 음향 및 통신장비 제조(D32), 기타 운송장비 제조(D35))을 선정하였다. 업종별 산업생산지수 원계열(1975년 1월~2006년 5월)로부터 공휴일수, 요일효과 및 이상치의 영향을 제거한 사전조정계열¹⁾을 대상으로 실증분석을 실시함으로써 상기 4가지 시계열분석 기법의 상대적 비교가 가능토록 하였다. 모의실험은 먼저 1975년 1월부터 2004년 5월까지의 사전조정계열에 대해 각 시계열

1) BOK-X-12-ARIMA의 RegARIMA를 이용하여 설, 추석등 명절효과, 요일효과 및 공휴일효과의 영향을 제거한 사전조정계열을 작성

예측기법에 의해 2004년 6월 생산지수(사전조정)를 예측한 후 동 예측치에 사전조정지수를 곱하여²⁾ 원계열로 환산한 후 2004년 6월 실적치와의 차이, 즉 1시점 후 예측오차(1-period ahead forecast error)를 계산하였다. 동 방식을 2004년 6월~2006년 5월까지 24개 시점에서 반복 적용함으로써 2년간 평균적으로 발생하는 오차의 크기를 평가해 보았다.

예측력 비교를 위한 통계량으로는 RMSPE(Root Mean Square Percentage Error)³⁾와 특정시점에서 극단적으로 크게 발생하는 이상(異常)오차를 제외한 로버스트한 예측력 비교를 위해 중앙값을 이용한 RMdSPE(Root Median Square Percentage Error)⁴⁾도 함께 이용하였다.

2. 분석 결과 및 예측력 비교

실증분석 대상으로 선정한 6개의 산업생산지수 중 음식료제조업(D15)의 시도표는 시계열이 성장추세와 계절성을 가지고 있으며 시간에 따라 분산이 확대되고 있음을 보여준다. 따라서 이분산성을 모형에서 직접 반영해 줄 수 있는 TVAR모형을 제외한, 기타 모형에서는 분산 안정화를 위해 사전조정계열에 로그변환을 실시하였다.

2) 가법모형일 경우에는 예측치에 사전조정지수를 더하여 산출

$$3) \text{RMSPE} = 100 \sqrt{\sum_{t=1}^n \left(\frac{y_t - \hat{y}_t}{y_t} \right)^2}, \quad y_t: \text{실적치}, \hat{y}_t: \text{예측치}$$

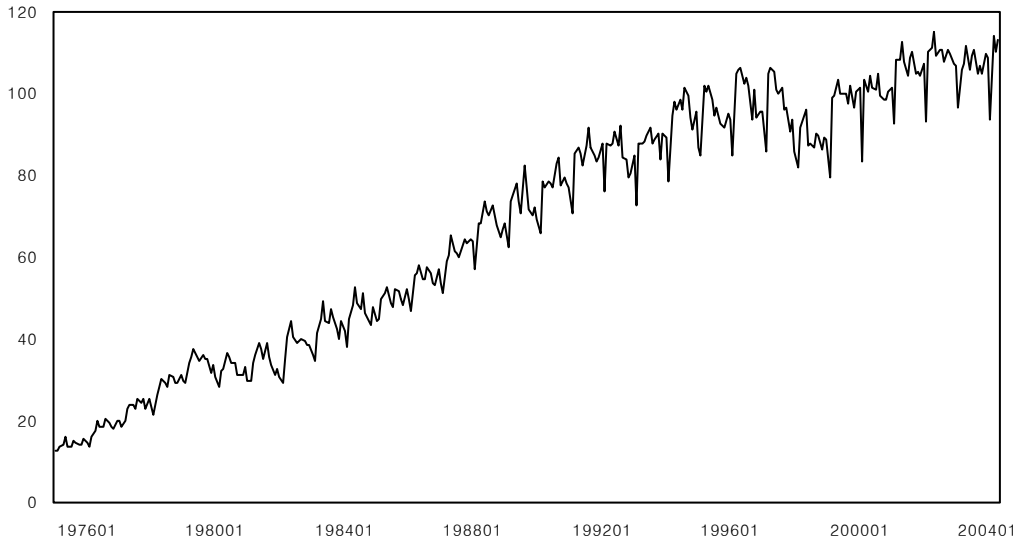
$$4) \text{RMdSPE} = 100 \sqrt{\text{Median}_{t=1, \dots, n} \left(\frac{y_t - \hat{y}_t}{y_t} \right)^2}, \quad y_t: \text{실적치}, \hat{y}_t: \text{예측치}$$

<그림 1>

음식료제조업(D15) 생산지수 시도표(사전조정계열)

(1975년 1월~2004년 5월)

(2000=100)



2004년 6월을 기준으로 각 예측기법을 적용⁵⁾해 산출된 모형 추정식은 다음과 같다. 먼저 ARIMA모형에 의한 최적선택모형⁶⁾은 ARIMA(1 1 1)(0 1 1)₁₂로 추정모형식은 다음과 같다.

$$y_t = 1.075y_{t-1} - 0.075y_{t-2} + y_{t-12} - 1.075y_{t-13} + 0.075y_{t-14} \\ + \widehat{\omega}_t + 0.451\widehat{\omega}_{t-1} + 0.741\widehat{\omega}_{t-12} + 0.334\widehat{\omega}_{t-13}$$

여기서 $\widehat{\omega}_{t-j}$, $j=0,1,12,13$ 는 평균 0, 분산 σ^2 는 0.035²인 백색잡음이다.

5) ARIMA모형은 BOK-X-12-ARIMA, TVAR모형은 West에 의해 작성된 Matlab 코드(<http://www.isds.duke.edu/~mw/tvar.html>), DLM-계절모형과 HKSG모형은 R 프로그램(<http://cran.r-project.org/>)을 이용하였음. R 프로그램은 통계분석 패키지로 프리웨어임

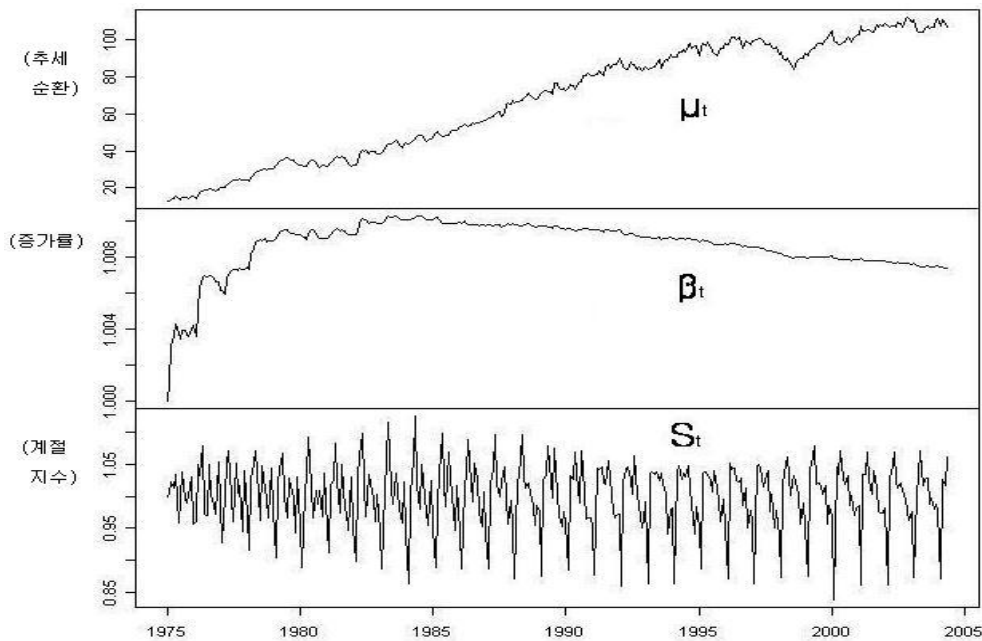
6) 최근 3개년 최소예측오차 기준으로 최적모형을 선택

DLM-계절효과모형을 적용하여 산출된 관측방정식과 상태방정식의 추정모형식은 다음과 같다.

$$\begin{aligned}
 y_t &= \mu_t + s_t + \epsilon_t, & \epsilon_t &\sim N(0, 0.007^2) \\
 \mu_t &= \mu_{t-1} + \beta_{t-1} + \xi_t, & \xi_t &\sim N(0, 0.013^2) \\
 \beta_t &= \beta_{t-1} + \eta_t, & \eta_t &\sim N(0, 0) \\
 s_t &= -\sum_{j=1}^{11} s_{t-j} + \omega_t, & \omega_t &\sim N(0, 0.028^2)
 \end{aligned}$$

<그림 2>

DLM-계절효과모형의 각 성분 추정결과(D15)



<그림 2>로부터 평균수준이 시간에 따라 상승하고 있으며, 각 시점별 증가율을 나타내는 β_t 는 급격한 상승추세를 유지한 후 1990년부터 다소 감소하는 모습을 볼 수 있다. 계절성분은 1985년까지

는 계절진폭이 커지다가 1990년대 이후 계절적 패턴이 안정화 되고 있음을 알 수 있다.

HKSG모형은 AIC통계량에 의해 상수형 추세와 가법형 계절성 및 가법형 오차항을 가진 NA모형을 선정하였다. 모수 추정값은 $\hat{\alpha}=0.714$, $\hat{\gamma}=0.164$ 이며, 상태벡터 초기치는 $\hat{\theta}_0 = (\hat{l}_0, \hat{s}_0, \hat{s}_{-1}, \dots, \hat{s}_{-11}) = (2.65, -0.082, -0.14, -0.02, 0.03, 0.12, 0.04, 0.01, 0.08, 0.03, -0.01, -0.06, 0.01)$ 이다. 이를 상태공간모형으로 나타내면 다음과 같다.

$$\begin{aligned} y_t &= m_t + \omega_t, \\ \mu_t &= l_{t-1} + s_{t-12}, \\ l_t &= l_{t-1} + 0.714\omega_t, \\ s_t &= s_{t-12} + 0.164\omega_t \end{aligned}$$

TVAR모형에서는 자기회귀차수별로 우도함수를 각각 계산하여 그 중에서 최대우도를 제공하는 TVAR(25)모형을 선정하였다.

TVAR모형에 의한 추정식은 $y_t = \sum_{i=1}^{25} \hat{\phi}_{t,i} y_{t-i}$ 이며, t가 2004년 5월

일 경우의 $\hat{\phi}_{t,i}$ 값은 <표 4>에 나타나 있다. <그림 3>은 25개 시

변자기회귀계수 중 상대적으로 영향력이 큰 계수($\hat{\phi}_{t,1}$, $\hat{\phi}_{t,2}$, $\hat{\phi}_{t,3}$, $\hat{\phi}_{t,12}$, $\hat{\phi}_{t,14}$, $\hat{\phi}_{t,15}$, $\hat{\phi}_{t,24}$, $\hat{\phi}_{t,25}$)들의 시간에 따른 움직임을 나타낸다.

최근 관측값들의 상대적 영향력을 나타내는 $\hat{\phi}_{t,1}$, $\hat{\phi}_{t,2}$, $\hat{\phi}_{t,3}$ 를 살펴 보면 1990년대 이후로는 t월 관측값에 대한 전월 관측값의 영향력이 다소 감소한 반면 2~3개월전 관측값의 영향력이 다소 커지는

모습을 보이고 있다. 1년 주기의 계절패턴을 보여주는 $\widehat{\phi}_{t,12}$, $\widehat{\phi}_{t,24}$ 는 1990년대 이후 계절성의 영향이 더 커지고 안정화됨을 보여준다.

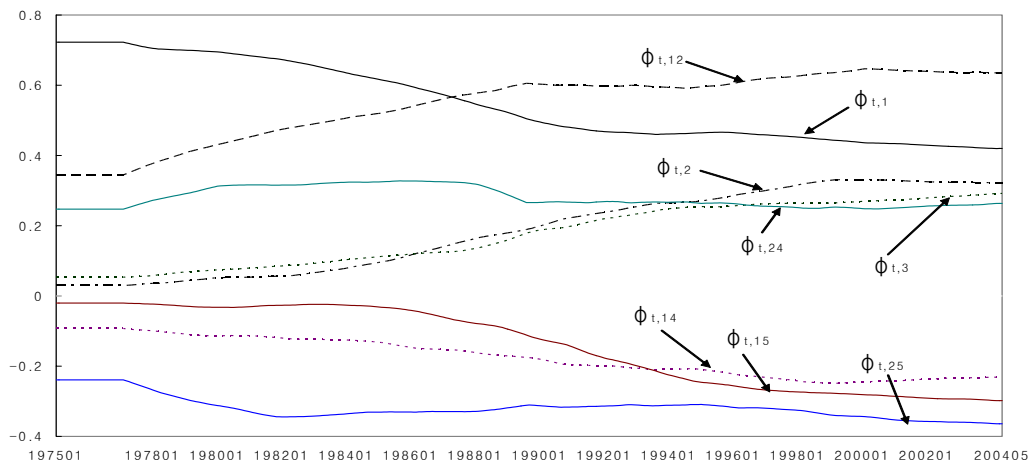
<표 4>

시변자기회귀계수 추정치 (D15)

$\widehat{\phi}_{t,1}$	$\widehat{\phi}_{t,2}$	$\widehat{\phi}_{t,3}$	$\widehat{\phi}_{t,4}$	$\widehat{\phi}_{t,5}$
0.420	0.322	0.292	-0.095	0.016
$\widehat{\phi}_{t,6}$	$\widehat{\phi}_{t,7}$	$\widehat{\phi}_{t,8}$	$\widehat{\phi}_{t,9}$	$\widehat{\phi}_{t,10}$
0.042	0.085	-0.061	-0.091	-0.120
$\widehat{\phi}_{t,11}$	$\widehat{\phi}_{t,12}$	$\widehat{\phi}_{t,13}$	$\widehat{\phi}_{t,14}$	$\widehat{\phi}_{t,15}$
-0.052	0.635	-0.021	-0.231	-0.298
$\widehat{\phi}_{t,16}$	$\widehat{\phi}_{t,17}$	$\widehat{\phi}_{t,18}$	$\widehat{\phi}_{t,19}$	$\widehat{\phi}_{t,20}$
0.082	0.017	-0.103	-0.061	0.123
$\widehat{\phi}_{t,21}$	$\widehat{\phi}_{t,22}$	$\widehat{\phi}_{t,23}$	$\widehat{\phi}_{t,24}$	$\widehat{\phi}_{t,25}$
0.028	0.099	0.078	0.263	-0.364

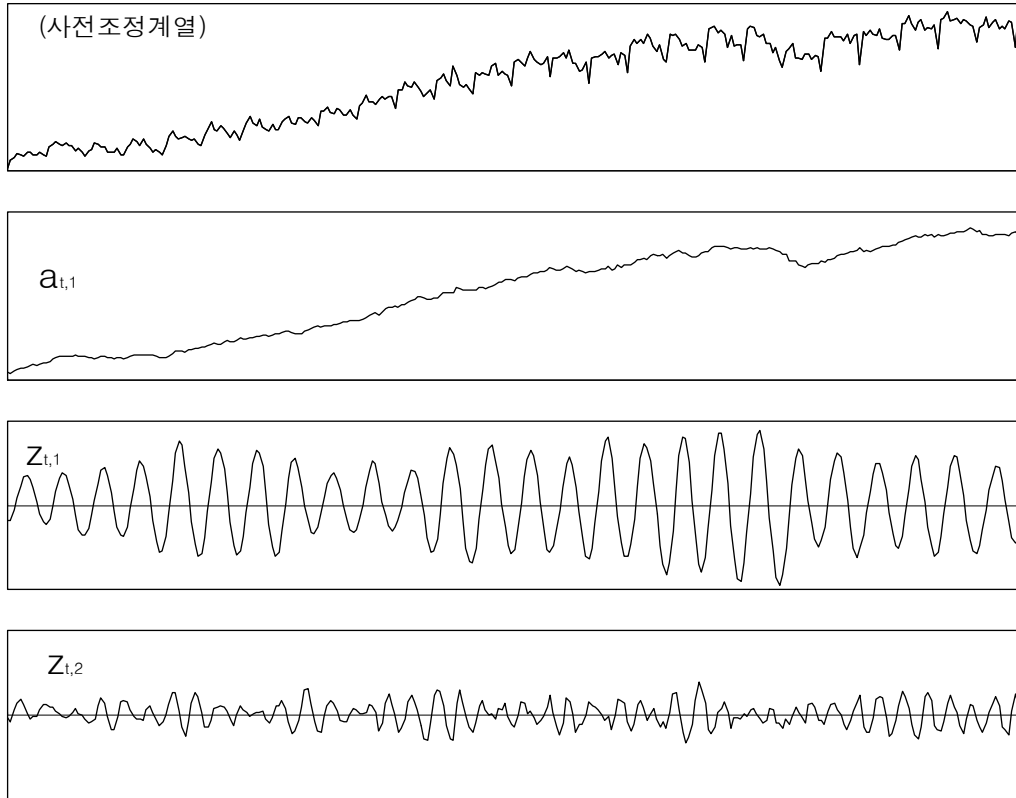
<그림 3>

시변자기회귀계수 추정치 추이 (D15)



<그림 4>

사전조정계열 및 TVAR모형의 잠재 성분 (D15)

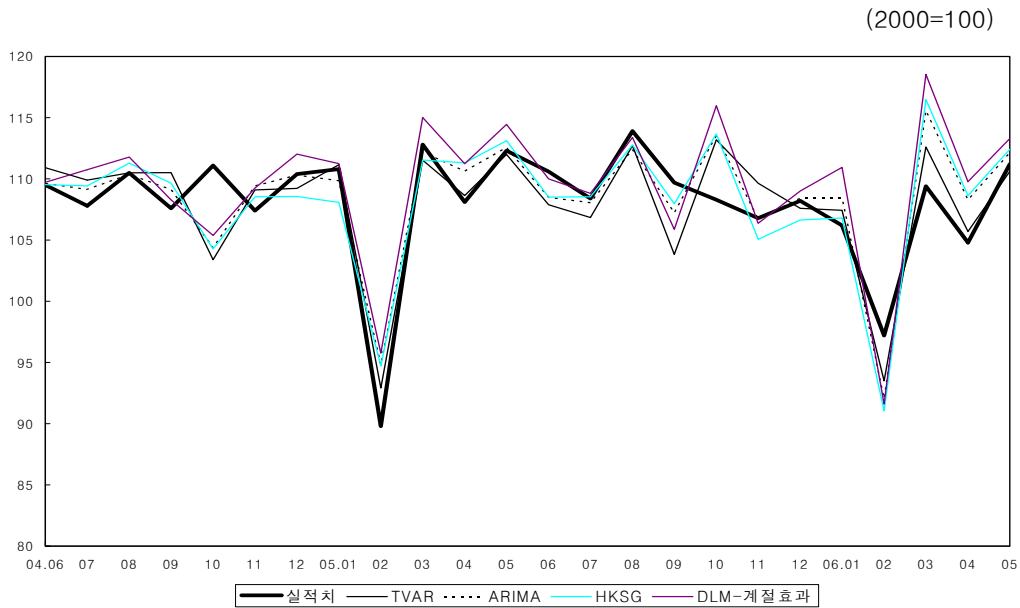


음식료제조업(D15)의 사전조정계열에 TVAR모형을 적용해보면 시계열이 장기적 추세를 나타내는 $a_{t,1}$ 성분과 다양한 주기에서 발생하는 순환패턴 등 25개 성분으로 구성됨을 알 수 있다. <그림 4>에서 $z_{t,1}$ 은 1년 주기로 변화하는 순환변동을 나타내고, $z_{t,2}$ 는 보다 단주기에서 발생하는 순환 특성을 나타내고 있다. 따라서 TVAR모형의 25개 잠재성분이 시계열의 장기추세 및 장단주기에서 발생하는 다양한 순환변동을 나타냄을 의미한다.

위의 네 가지 예측방법에 의해 총 24회의 모의실험을 실시하여 2004년 6월~2006년 5월까지 1시점 후 예측치의 추이는 <그림 5>에 나타나 있다.

<그림 5>

모의실험을 통한 예측치 (D15)



위와 같은 방법으로 나머지 5개의 산업생산지수에 대하여 모의 실험을 실시하였다. <표 5>와 <표 6>은 모의실험에 의해 각 시 계열모형의 예측력을 비교하기 위해 사용된 RMSPE와 RMdSPE 통계량을 나타낸다. RMSPE 기준으로는 업종 D15, D18, D30에서 TVAR모형이 ARIMA모형보다 우수하였고 나머지 업종에서는 ARIMA모형이 TVAR모형보다 우수하였다. 한편, RMdSPE 기준으로는 업종 D22를 제외한 나머지 업종에서 TVAR모형과 HKSG 모형이 ARIMA모형보다 예측력이 우수한 것으로 나타났다.

<표 5>

예측오차(RMSPE) 비교

코드 번호	BOK-X-12- ARIMA	TVAR모형	HKSG모형	DLM-계절 효과 모형
D15	2.83	2.66	3.08	3.62
D18	7.93	7.17	7.48	6.94
D22	5.07	5.35	5.11	5.30
D30	6.73	6.20	6.83	7.00
D32	5.46	5.97	5.85	5.70
D35	8.71	10.43	8.49	9.07

<표 6>

예측오차(RMdSPE) 비교

코드 번호	BOK-X-12- ARIMA	TVAR모형	HKSG모형	DLM-계절 효과 모형
D15	1.37	1.36	1.60	1.90
D18	6.09	5.07	5.91	5.59
D22	1.96	2.69	2.52	2.57
D30	3.97	5.44	3.88	4.43
D32	4.13	3.76	3.53	5.29
D35	4.65	6.39	4.08	5.53

IV. 결론

시계열자료의 단기 예측방법으로는 다양한 형식의 지수평활법과 Box-Jenkins의 ARIMA모형이 일반적으로 사용되고 있다. 지수평활법은 이용상의 편리성 측면에서, ARIMA모형은 단기예측력의 우수성 측면에서 선호되어 왔다. 반면 지수평활법은 주관적인 판단에 의존하여 결정되는 평활상수에 따라 영향을 받을 뿐만 아니라 시계열의 확률적 움직임을 포착하지 못하는 단점이 있다. ARIMA모형은 확률과정을 전제로 정상성의 조건을 필요로 하며 모형식별과정에서 주관성이 개입될 수가 있다.

본 연구에서는 상태공간모형의 원리를 이용하여 24개 지수평활법을 체계화하고 그 중에서 최적모형을 선정해 주는 HKSG모형과 동태적 선형모형에 의해 시계열을 다수의 구성성분으로 분해하고 각 성분의 움직임을 시점별로 파악하는 시변자기회귀모형시계열을 고려해 보았다.

6개 업종에 대한 월별산업생산지수를 이용하여 각 예측기법별로 모의실험을 실시해 본 결과 모든 시계열에서 예측력이 뛰어난 기법은 채택되지 않았다. 이는 시계열이 가지고 있는 고유의 추세와 계절성의 특성 등 체계적 움직임뿐만 아니라 파업, 기후변동 등 불규칙적으로 발생하는 요인이 다르고 이를 포착하는 예측기법별 차이에 기인하는 것으로 판단된다.

그러나 상태공간모형에 의한 지수평활모형은 기존의 지수평활법에 비해 시계열의 확률적 움직임을 포착하고 AIC통계량 등에

의해 최적모형을 선정해 준다는 점에서 주관적으로 선정되는 단일
지수평활방식보다 예측력 개선을 확보할 수 있다는 장점이 있다.
또한 시변자기회귀모형(TVAR모형)에서는 매 시점별로 모수를 별
도 추정함으로써 전체 시계열에서 개별 자기회귀계수에 단일 모수
를 가정한 자기회귀모형에 비해 시계열의 움직임을 보다 정교히
파악할 수 있다는 측면에서 장점을 가진 것을 알 수가 있었다.

<참 고 문 헌>

황선영, “월별 시계열 자료 분석기법들의 비교 분석 및 변동성 모형을 이용한 단기예측력 개선 방안”, 『계간 국민계정』, 2005년 제4호, 한국은행, pp33-68.

Bollerslev, T., "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics*, **31**, 1986, 307-327.

Brown, R. G., *Statistical forecasting for inventory control*, 1959, New York: McGraw-Hill.

Bureau of the Census, *X-12-ARIMA Reference manual (version 0.2.10)*, 2002.

Engel, R. F., "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, **50**, 1982, 987-1007.

Gardner, E. S., "Exponential smoothing: the state of the art", *Journal of Forecasting*, **4**, 1985, pp1-28.

Holt, C. C., "Forecasting seasonal and trends by exponentially weighted averages" *O.N.R Memorandum*, **52**, 1957.

Hyndman, R. J., Koehler, A. B., Snyder, R. D. and Grose, S., "A State space framework for automatic forecasting using exponential smoothing methods", *International Journal of Forecasting*, **18**, 2002, pp439-454.

Ord, J. K., Koehler, A. B. and Snyder, R. D., "Estimation and prediction for a class of dynamic nonlinear statistical models", *Journal of the American Statistical Association*, **92**, 1997, pp1621-1629.

Pegels, C. C., "Exponential forecasting: some new variation", *Management Science*, **12**, 1969, pp311-315.

Prado, R., Huerta, G. and West, M., "Bayesian time-varying autoregressions: Theory, methods and applications", *Resenhas*, **4**, 2001, pp405-422.

West, M., "Time series decomposition", *Biometrika*, **84**, 1997, pp489-494.

West, M. and Harrison, J., *Bayesian forecasting and dynamic linear models*, New York: Springer-Verlag, 1997.

Winters, P. R., "Forecasting sales by exponentially weighted moving averages", *Management Science*, 1960, **6**, 324-342.