

## 수산 관련 시계열 자료를 이용한 통계학적 분석에서의 자기상관에 대한 고찰

박영철<sup>+</sup> · 히야마 요시아끼  
일본 서해구수산연구소

### Autocorrelation in Statistical Analyses of Fisheries Time Series Data

Young Cheol PARK<sup>+</sup> and Yoshiaki HIYAMA  
Seikai National Fisheries Research Institute, #3-30, Kokubu,  
Nagasaki 850-0951, Japan

Autocorrelation in time series data can affect statistical inference in correlation or regression analyses. To improve a regression model from which the residuals are autocorrelated, Yule-Walker method, nonlinear least squares estimation, maximum likelihood method and 'prewhitening' method have been used to estimate the parameters in a regression equation. This study reviewed on the estimation methods of preventing spurious correlation in the presence of autocorrelation and applied the former three methods, Yule-Walker, nonlinear least squares and maximum likelihood method, to a 20-year real data set. Monte carlo simulation was used to compare the three parameter estimation methods. However, the simulation results showed that the mean squared error distributions from the three methods simulated do not differ significantly.

**Key words:** Autocorrelation, Yule-Walker method, Prewhitening, Monte carlo simulation, Fisheries time series data

#### 서 론

근년에 들어와 수산 자원 생물학에 있어서 가장 주목할 만한 연구들 중의 하나는 가변하는 어군의 생산률 및 가입 기작을 밝히려는 시도라 할 수 있다 (Pyper and Peterman, 1998). 어군의 가입과 사망과정은 여러가지 생물학적 또는 비생물학적 요인들에 의해 영향을 받는다. Kim et al. (1997)의 연구를 제외하면, 최근 우리나라에서 수행되어진 연구들에서 어군의 연변동에 영향을 끼치는 여러 물리적 또는 생물학적 요인들은 시계열상에 있어 추정된 잔차들과 독립적인 관계에 있는 것으로 가정되어져 왔으며 따라서 회귀 계수들은 시계열분석이 아닌 일반적인 회귀분석법을 사용하여 추정되어 왔다 (Kim et al., 1995; Kim et al., 1999; Park and Baik, 2001; Zhang and Lee, 2001). 그러나 이미 기존의 많은 연구들에서 지적되어 왔듯이 각각의 시계열 자료들은 시계열 상에서 서로 독립적이지 못하며 많은 경우에 같은 분포도 갖지 않는다 (Box et al., 1994; Hilborn and Walters, 1992). 이러한 관련성을 자기상관 (autocorrelation) 혹은 계열상관 (serial correlation)이라 하며 따라서 이들 요인 변수값들과 시계열 상에 있어 추정된 관계의 오차항은 서로 밀접한 관련을 가지게 된다.

회귀분석법은 특히 산란량-가입량 관계와 같이 함수적 관련식을 추정하기 위해 가장 많이 쓰이는 통계학적 방법이다. 그러나 수산학에서 추정되어지는 많은 함수적 관계들은 시계열 자료로부터 얻어지며 이 경우 비록 변량들에 대한 측정오차가 없다고 가정하더라도 시계열상의 변량들과 추정잔차들 사이에 존재하는 자기상관에 의해 계수 추정치들은 왜곡되게 된다. 또한 이로 인한

오차는 독립변수의 종류가 많을수록 커지게 된다 (Pyper and Peterman, 1998). 본 연구는 이러한 시계열분석에서 발생하는 자기상관으로 인한 오차에 대해 개관하고 그에 따른 보정 방법들에 대해 고찰하였다.

#### 자료 및 방법

##### 자기상관으로 인한 오차

Ricker의 산란량-자원량 관계를 예로 들어 회귀모형을 표현한다면, 즉

$$\ln(R_t/S_{t-1}) = a - bS_{t-1} \tag{1}$$

식 (1)은 다음의 회귀 모형으로 나타낼 수 있다.

$$y = \beta X + c$$

여기서  $R_t$ 는  $t$ 년의 가입량지수,  $S_{t-1}$ 는 그 전해의 산란량지수를 나타내고  $a$ 와  $b$ 는 상수이다. 또한  $y$ 는 종속변수의 시계열 벡터로서 가입율을,  $X$ 는 독립변수의 함수로 구성된 디자인행렬 (design matrix)로서 시계열 산란량지수의 조합을 나타내며  $\beta$ 와  $c$ 는 각각 추정된 계수들의 벡터와 오차항으로서, 다음과 같이 나타낼 수 있다.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{bmatrix}, y_t = \ln(R_t/S_{t-1}), \beta = \begin{bmatrix} a \\ -b \end{bmatrix}, X = \begin{bmatrix} 1 & S_0 \\ 1 & S_1 \\ \vdots & \vdots \\ 1 & S_{t-1} \end{bmatrix}, c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_t \end{bmatrix}$$

회귀분석법은 일반적으로 오차항에 대한 다음의 Gauss-Markov 조건들을 만족한다는 가정을 바탕으로 한다.

(i) 오차항  $c_t$ 는 서로 상관관계를 가지지 않는다.

<sup>+</sup>Corresponding author: youngyl96@yahoo.com

(ii)  $c_t$ 의 분산은 시점에 무관한 상수이다.

여기서 (ii)번의 가정이 만족되지 않을 때를 자료가 분산이질성(heteroscedasticity)이라고 하며 이 문제는 변수변환을 통하여 어느 정도 해결될 수 있다. 분산의 이질성 문제는 시계열자료에서 보다는 오히려 횡단면 자료(cross-section data)의 분석에서 주로 나타난다(최, 1992). 즉 어류들의 성장률을 분석할 경우 크기가 큰 어류의 성장률에 대한 분산이 크기가 작은 어류의 성장률에 비해 클 것이며 같은 크기라 하더라도 서식지역과 위도 및 산란시기 등에 따라 성장률의 분산이 다를 것이다. 반면 (i)번의 가정이 만족되지 않는 경우를 잔차항의 자기상관이라 하고 연도별 어획량과 같은 시계열자료의 분석시 발생하게 된다(Pindyck and Rubinfeld, 1991). 즉 잔차항의 자기상관은 자기상관을 가진 시계열 변량들에 의해 함수적 관계가 추정됨으로써 발생한다. 잔차항이 자기상관을 가지게 되면 (1) 먼저 추정된 계수들의 신뢰도 및 신뢰구간이 부정확해지고, (2) 추정계수들의 효율성이 현저히 떨어지게 된다. 즉, 잔차가 자기상관관계에 있다는 것은 잔차항이 추정된 관계식을 향상시킬 수 있는 정보를 내포하고 있음을 의미한다.

Hilborn and Walters (1992)는 부정확한 관계식의 추정을 피하기 위해 산란량-가입량 관계 추정 후 반드시 잔차들의 형태를 주의깊게 조사하여야 한다고 지적하였다. Walters (1985)와 Caputi (1988)는 산란량지수와 잔차항의 비독립적 관계에 의한 오차 발생에 대해 수식을 이용하여 설명하였으며 몬테카를로 방법을 이용하여 발생오차의 크기가 산란량-자원량관계의 변이에 비해 한다는 것을 보였다. 산란량-자원량 관계식의 적합도는 환경요인 변수들을 관계식에 첨가하게 되면 증가하지만 이로 인해 잔차항의 자기상관은 더욱 증가하게 되며 이러한 현상은 특히 관찰점들의 수가 적을 때 심화된다. 따라서 산란량-자원량 관계를 추정할 경우, 환경요인들을 독립변수에 추가할 때는 각별한 주의가 요망된다.

### 자기상관의 검증

잔차항 사이에 1차-자기상관(first-order autocorrelation)이 있는지를 검증하기 위해 흔히 쓰이는 방법이 Durbin-Watson 통계량 방법이다. 이 방법에서는 시간  $t$ 에서의 잔차와  $t-1$ 에서의 잔차를 비교하고 이들의 연속적인 비교를 통하여 상관성을 측정하기 위한 통계량을 개발하였다. 이 통계량은 다음 식에 의해 구해진다(Durbin and Watson, 1951).

$$d = \frac{\sum_{t=2}^n e_t - e_{t-1})^2}{\sum_{t=1}^n (e_t)^2}$$

여기서  $d$ 는 Durbin-Watson 통계량을,  $e_t$ 는 잔차 그리고  $t$ 는 시간을 나타낸다. 이 통계량의 범위는 0과 4사이에 존재하며 잔차들 사이의 관계가 양의 상관에 가까우면 0에 가깝고 음의 상관에 가까우면 4에 가깝게 나타난다. 이 통계량의 분포는 디자인행렬  $X$ 에 의존하기 때문에 Durbin and Watson (1951)은 분포가  $X$ 에 의존하지 않는  $d_U$ 와  $d_L$ 을 정의하고 다음과 같이 가설검정할 것을 제안하였다. 즉,

1)  $H_0$ : 잔차들 사이에 유의한 상관관계가 존재하지 않는다.

2) 만일  $0 < d < d_U$ 이면  $H_0$ 는 기각되고 잔차들 사이에 양의 자기상관이 존재한다.

3) 만일  $d_U < d < (4 - d_U)$ 이면  $H_0$ 를 채택한다.

4) 만일  $(4 - d_L) < d < 4$ 이면  $H_0$ 는 기각되고 잔차들 사이에 음의 자기상관이 존재한다.

5) 그 외의 경우에는 결론을 내리지 못한다.

유의수준 0.05, 0.025 및 0.01에서의  $d_U$ 와  $d_L$ 값은 Durbin and Watson (1951)에 의해 계산되어졌다(Appendix 1 참조).

### 자기상관의 보정

시계열자료의 자기상관을 보정하기 위한 방법으로는 여러 가지가 있으나 가장 널리 사용되는 것은 Appendix 2에 설명한 3가지 추정법과 사전백색화법(prewhitening; Box et al., 1994)을 들 수 있다. Appendix 2의 3가지 방법들은 1) 추정된 잔차를 재처리해 줌으로써 잔차에 포함된 상관 요인을 제거하는 율-워커법(Yule-Walker method), 2) 대수우도함수를 최대화시켜 최우추정량을 구하는 최우추정법(Maximum likelihood) 및 3) 대수우도함수의 지수부분을 최소화하는 비선형최소제곱법(Nonlinear least squares)이다(Appendix 2). 자기상관을 보정하기 위한 또 다른 방법은 입력시계열자료가 가진 자기상관현상을 분석 전에 제거해 주기 위해 시계열자료를 사전백색화하는 사전백색화법(Milicich et al., 1992)이며 차분을 이용하여 시계열자료가 가진 비정상성을 정상화 시켜주는 차분법(differencing; Thompson and Page, 1989)도 넓은 의미에서 사전백색화법의 일종으로 분류할 수 있다. 사전백색화란 입력 시계열변수들을 서로 독립이며 동일한 분포를 가지고 평균이 0인 확률변수화 시키는 것이다. 예를 들어 입력시계열자료  $X_t$ 가 AR(1) 모형을 따른다면 사전백색화된 시계열자료  $Z_t$ 는 다음과 같이 나타낼 수 있다.

$$Z_t = X_t - \phi_X X_{t-1}$$

여기서  $t$ 는 시점이며  $\phi_X$ 는 시계열  $X_t$ 의 자기상관계수이다. Thompson and Page (1989)는 복서대서양의 6개 대구 개체군과 3개의 볼락 개체군의 가입과정을 조사하면서 자기상관현상을 개선시키기 위해 다음과 같이 시계열자료의 차분을 이용하였다.

$$\nabla X_t = X_t - X_{t-1}$$

여기서  $\nabla X_t$ 는  $X_t$ 의 일차차분이다. 이와 같이 차분을 이용하면 자료가 가진 자기상관현상을 개선시킬 수 있는 경우가 많다. 관찰점의 수가 크지 않을 때 이 방법들의 효율성은 잘 알려져 있지 않으나 일반적으로 사전백색화법은 관찰점의 수가 50개 이상에서 사용할 것을 권고하고 있다(Brockwell and Davis, 1995). 또한 사전백색화법의 사용이 추정계수들에 대한 Type-II 오차를 증가시킬 수도 있다(Pyper and Peterman, 1998).

시계열자료에서 나타나는 변이는 크게 2가지로 나눌 수 있다. 하나는 단기변동으로 고주파 변이(high-frequency variability)라고 하며 변동의 주기가 짧은 변이를 말한다. 반면 저주파 변이(low-frequency variability)는 변동의 주기가 비교적 긴 현상으로 흔히 변동주기가 10년 또는 20년 이상 되는 변이를 일컫는다. 일

반적으로 사전백색화법이나 차분법을 사용할 때 전제되는 가정은 저주파변이가 (1)고주파변이에 비해 미미하고 중요하지 않으며 (2)고주파변이와 동일한 원인에 의해 야기된다는 것이다. Pypers and Peterman (1998)은 저주파 변이를 가진 시계열 자료의 분석에 있어서는 사전백색화법이 효과적인 분석법이 아니라는 것을 보였다. 저주파 변이의 주기성은 비교적 관찰되기가 어려우나 최근 들어 물리적 또는 생물학적 환경요인들의 장기변동에 있어 일정한 규칙성이 드러나고 있어 (Beamish et al., 1994) 시계열자료의 변이를 주의 깊게 파악할 필요가 있다.

#### 실제자료에의 적용

본 연구에서 사용한 실제자료는 동지나해 일본측 쌍끌이 창포뚜기 (*Loligo edulis*) 어업의 20년간의 어획자료이며 각 변수들의 측정오차는 없는 것으로 가정하였다. 동지나해 일본측 쌍끌이 조업에 있어 창포뚜기는 주요 상업종이며 1980년대 이후 계속해서 어획량이 증가하고 있는 종이다. 창포뚜기 개체군의 가입 및 산란량 지수는 VPA (Virtual Population Analysis; Pope, 1972)를 이용하여 구하였으며 중간관계지수와 (Park et al., 2002) 함께 산란량-자원량 (SR: spawning stock-recruitment) 관계를 구하는데 이용하였다. SR 관계는 환경요인을 고려한 Ricker 모델을 사용하였다. 즉

$$\ln(R/S) = \alpha - \beta S + \gamma_1 E_1 + \dots + \gamma_n E_n + e$$

여기서 R은 가입량지수, S는 산란량지수, E는 중간관계지수, e는 잔차  $\alpha, \beta, \gamma$ 는 각각 상수인 SR계수이다. 이때 잔차는 평균이 0이고 분산이 일정하며 정규분포를 따른다. 중간관계지수로는 툼블때통이 (*Saurida wanieso*)와 갯장어 (*Muraenesox cinereus*)의 trawl당 평균 어획량의 역수를 각각 사용하였다.

각 방법의 오차를 비교하기 위해 몬테카를로법 (Monte Carlo simulation)을 이용하였다. 사전백색화법의 효율성을 높이려면 최저 50년간의 장기간 자료가 필요하나 실제 50년 이상의 자료를 얻기는 매우 어렵다. 그러므로 관찰점의 수가 50개 이하인 많은 수산자료분석의 경우, 사전백색화법은 유용한 방법이 되지 못한다. 따라서 본 연구에서는 SR계수  $\alpha, \beta, \gamma_1, \gamma_2$ 를 추정하기 위해 1) Yule-Walker법, 2) 최우추정법, 3) 비선형최소제곱법을 각각 이용하였다.

각 방법의 오차를 비교하기 위해 각 방법 당 1000회의 몬테카를로자료를 발생시켜 ( $\sigma=0.3$ ) 평균 잔차제곱 (MSE: mean squared error)과 그 분포를 비교하였다. 몬테카를로 자료의 발생은 MS-Excel 컴퓨터 패키지의 rand 함수를 이용하였다. 각 방법들의 추정 매개변수들과 평균잔차제곱 추정치의 비교에는 분산분석을 위한 F-검정법이 이용되었다.

#### 결과 및 고찰

VPA를 이용하여 추정된 가입량 및 산란량지수는 Fig. 1과 같다. 가입과 산란량지수 모두 1987년에서 1989년 사이에 가장 높은 값을 나타냈으나 90년대에 들어오면서 급격히 감소함을 보여주었다.

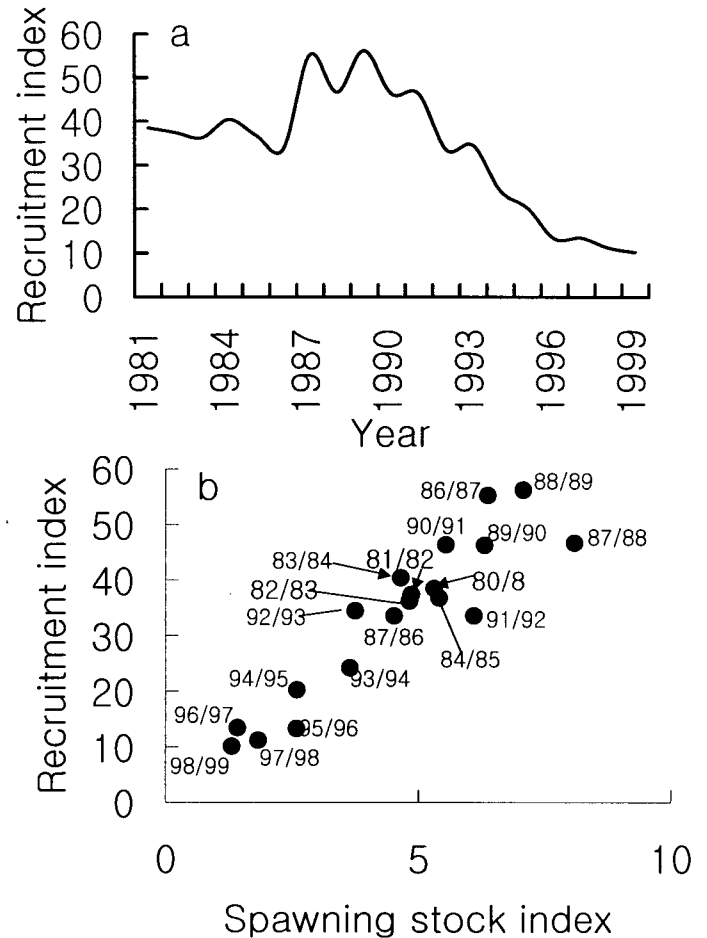


Fig. 1. Annual variation in recruitment index (a) and spawning stock-recruitment relationship (b) for *Loligo edulis* in the East China Sea.

산란 및 가입지수에 대한 90년대 감소의 가장 중요한 원인은 일본측 쌍끌이 어획노력량의 감소 때문으로 생각된다. 즉, 80년대 동지나해 일본측 쌍끌이 어업의 총어선톤수는 50만톤 내외를 유지하였으나 1990년부터 급격히 감소하여 1999년에는 4천9백톤에 불과하였다 (Park et al., 2002). Fig. 1a의 가입량 변동은 80년대 후반의 증가와 90년대의 감소로써 요약될 수 있으며 이것은 산란량 변동에서도 마찬가지로 나타난다. 이러한 산란과 가입의 유사한 경향성은 가입량 (R)이 가입의 결과로 일어나는 산란량 (S)과 상관 관계를 가지기 때문에 나타나는 것으로 시계열분석에서 발생하는 또 다른 오차의 원인이 된다 (Hilborn and Walters, 1992). 그러나 이로 인한 오차는 환경요인들을 분석에 포함시키므로 현저히 줄일 수 있다 (Caputi, 1988).

문헌에 나타난 예를 들어보면, Zhang and Lee (2001)는 전갱이 (*Trachurus japonicus*)의 자원 평가 연구에서 가입량을 추정하였는데 이 또한 1983년 이전의 상대적으로 낮은 가입과 1983년 이후의 증가로 요약된다 (see Zhang and Lee, 2001, Figs. 3 and 8). 이는 1965년과 1964년을 제외하면 생체량 변동에서도 유사하게 나타난다. 따라서 환경요인을 고려한 산란량-가입량 관계 (즉

$R_t$ 와  $S_{t-1}$ 의 관계)를 추정하는 것이 가입량 ( $R_t$ )과 그 가입의 결과로 나타나는 산란량 ( $S_t$ ) 사이의 상관 관계 때문에 발생하는 오차를 줄일 수 있을 것이다.

가입량-산란량관계에 포함시킬 환경요인을 선정할 경우에도 세심한 주의가 필요하다. 환경요인은 보통 종속변수 즉 가입량과의 상관도가 높은 항목을 선정하게 되는데 그 선택에 있어 환경요인들 상호간의 상관성 또한 고려하여야 한다. 환경요인들 사이에 유의한 상관관계가 있을 때를 다중공선성 (multicollinearity)이라 하며 이 경우 계수 추정값들의 분산이 커지기 때문에 추정된 회귀식의 신뢰도를 떨어뜨릴 수 있다. Zhang and Lee (2001)의 연구에서는 서로 상관관계를 가진 쿠로시오 난류의 수송량 (volume transport)과 동물플랑크톤의 생체량을 선택하였지만 stepwise방법을 사용하였기 때문에 모형의 유의한 신뢰도를 확인할 수 있었다 (see Zhang and Lee, 2001, Table 6). 다중공선성에 대한 대책으로는 기존모형에 독립변수를 추가하거나 제거함으로써 모형의 신뢰도와 적합도를 확인하는 stepwise법 이외에도 주성분분석을 통해 독립변수를 재편하는 방법 등이 있으나 본 연구의 주제가 아니므로 통계학 문헌을 참조하기 바란다 (e.g. Draper and Smith, 1998).

각 추정방법들로부터 얻어진 산란량-가입량관계 계수들과 평균잔차제곱, 결정계수값 그리고 Durbin-Watson 통계량은 Table 1에 나타내었다. 일반적인 최소제곱추정법에서 얻은 Durbin-Watson 통계량은 3.14로서 잔차들 사이에 어떤 상관관계도 없다는 귀무가설을 유의수준 0.05와 0.025에서 각각 기각하며 유의수준 0.01에서는 결론을 내리지 못한다. 따라서 이 방법에 의해 추정된 산란량-가입량관계 계수추정치들은 잔차 사이의 자기상관에 의해 왜곡되어진 값들이라고 생각할 수 있다. 반면 Yule-Walker법, 최우추정법, 그리고 비선형최소제곱법을 사용한 경우의 Durbin-Watson 통계량은 2.3~2.49로 유의 수준 0.01에서 귀무가설을 채택하게 되며 결정계수값 또한 일반적인 최소제곱법에 비해 향상된 결과를 나타내었다. 그러나 Yule-Walker 법을 이용했을 경우는 유의수준 0.05와 0.025에서 각각 잔차항 사이에 상관관계가 존재하는가에 대한 결론을 내리지 못하였다. 각 방법을 통해 추정된 계수들의 분포범위는 Fig. 2와 같다. 즉 각각의 그림들은 위의 3가지 추정법에 의해 추정된 자원량 산란량 계수들의 5% 하한값, 중앙값

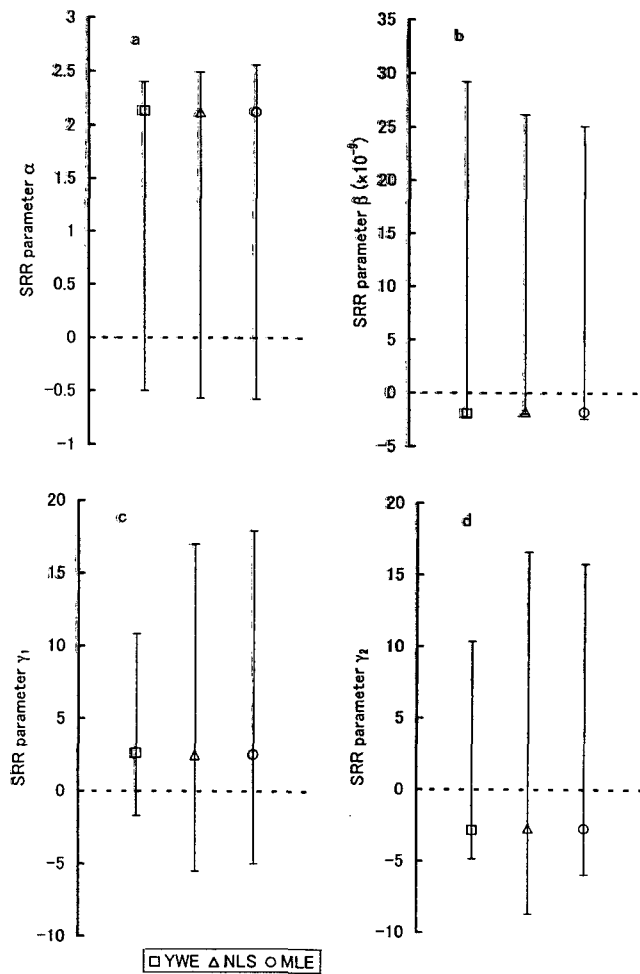


Fig. 2. Distributions of SRR (spawning stock-recruitment relationship) parameter estimates,  $\alpha$ (a),  $\beta$ (b),  $\gamma_1$ (c) and  $\gamma_2$ (d), for *Loligo edulis* in the East China Sea. YWE = Yule-Walker method; NLS = nonlinear least square method; MLE = maximum likelihood method.

(median) 그리고 5% 상한값들을 나타내었다. Ricker의 산란량-가입량관계 계수  $\alpha$ 와  $\beta$ 의 분포는 추정방법들 사이에 유의한 차이가 없었다. 반면 중간관계계수  $\gamma_1$ 와  $\gamma_2$ 의 추정치 분포범위는 Yule-Walker법을 이용했을 때가 비선형최소제곱법이나 최우추정법을 이용했을 때보다 유의하게 작았다 ( $p < 0.01$ ; Fig. 2).

평균잔차제곱분포의 경우, 비선형최소제곱법으로 추정된 평균잔차제곱이 가장 작았으나 추정치의 변이는 최우추정법을 이용했을 때가 가장 작았다. 그러나 각 추정방법사이에 유의한 차이는 발견되지 않았다. 즉 본 연구재료인 동지나해 창끝뚜기 개체군의 경우는 위에서 사용한 3가지 추정법들 가운데 상대적으로 유의하게 효과적인 방법을 가릴 수는 없었다 (Fig. 3).

수산생물학의 시계열분석에 있어서 자기상관에 의한 오차발생과 그 보정에 관한 주제는 오랫동안 여러 학자들에 의해 다루어져 왔다 (Walters, 1985; 1990; Caputi, 1988; Milicich et al., 1992; Pypers and Peterman, 1998). 특히 환경요인을 고려한 산란량과 가

Table 1. Parameter estimates of spawning stock-recruitment relationships for *Loligo edulis* in the East China Sea.

	OLS	YWE	NLS	MLE
$\alpha$	2.249	2.132	2.121	2.124
$\beta$	$-3.881 \times 10^{-9}$	$-1.926 \times 10^{-9}$	$-1.744 \times 10^{-9}$	$-1.786 \times 10^{-9}$
$\gamma_1$	3.299	2.636	2.501	2.537
$\gamma_2$	-3.710	-2.841	-2.7023	-2.735
MSE	$1.943 \times 10^{-2}$	$1.178 \times 10^{-2}$	$1.143 \times 10^{-2}$	$1.146 \times 10^{-2}$
$R^2$	0.44	0.68	0.69	0.69
d	3.14	2.49	2.30	2.31

OLS=ordinary least square method; YWE=Yule-Walker method; NLS=nonlinear least square method; MLE=maximum likelihood method; MSE=mean squared error; d=Dubin-Watson statistic.

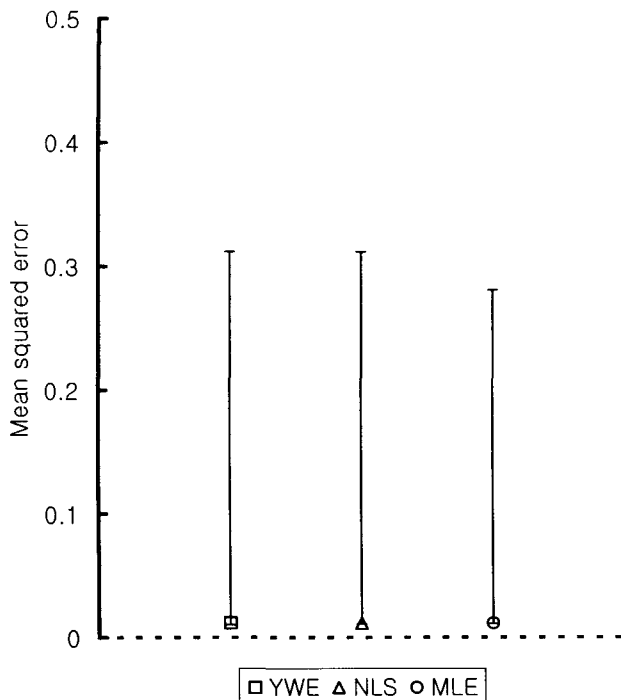


Fig. 3. Mean squared error distributions of estimated spawning stock-recruitment relationships for *Loligo edulis* in the East China Sea.

입량 분석에 있어 오차항의 자기상관은 일반적인 현상이며 이를 보정해 주는 과정이 적합한 관계추정의 필수요건으로 지적되어 왔다 (Hall, et al., 1988; Kope and Botsford, 1990; Walters, 1990). 실제로 Penn and Caputi (1986)의 연구를 제외하면 많은 환경요인을 고려한 산란량-가입량관계 연구들이 입력자료의 측정오차와 가입자료의 자기상관으로 인한 추정관계의 왜곡으로 인해 받아들여지지 못했다 (Penn et al., 1995). Hall et al. (1988)은 태평양 청어 (*Clupea harengus pallasii*)의 산란량-가입량 관계연구에서 일반적인 회귀분석에 의해 추정된 Ricker 계수  $\alpha$ 와  $\beta$ 가 각각 10%와 18% 과대 추정되었음을 보였다. 본 연구에서는 일반적인 회귀분석법을 사용했을 때 자기상관을 보정해주는 추정방법들을 사용했을 때 보다  $\alpha$ 는 약 5.5~6% 과대 추정되었으며  $\beta$ 는 약 100% 이상 과소 추정되어진 것으로 나타났다. 자기상관으로 인한 오차의 정도는 입력시계열자료가 가진 분포에 따라 달라질 것이다. 그러나 입력시계열자료가 가진 분포가 자기상관으로 인한 오차를 생성시키는지의 여부를 떠나 회귀 분석으로 계수들을 추정한 후 오차항을 검사하는 절차는 시계열 분석에서 필수적인 과정이다 (Caputi, 1993). Zhang and Lee (2001)는 우리나라 전갱이 가입과 산란량 자료를 Beverton and Holt model과 Ricker model에 적합시켰는데 그 결과를 보면 가입이 높은 값을 가지는 80년대 후반과 1993~1994년, 그리고 가입이 낮은 값을 가지면서 산란지수는 비교적 높은 1965~1966년에 비교적 큰 오차를 나타내었다 (see Zhang and Lee, 2001, Figs. 3). 이 결과만으로는 오차항이 어떤 경향성을 가진다고 말하기는 어려우나 오차의 산포도나 Dubin-

Watson 통계량을 구해볼 필요는 있다고 생각된다.

최근 우리나라에서 산란과 가입에 대한 연구들이 활발히 진행되고 있다. 또한 중요한 자원군에 대해 TAC (total allowable catch)를 산정하기 시작하면서 산란과 가입과의 관계는 더욱 중요한 연구주제가 되었다. 그러나 가입과 환경과의 관계, 산란과 가입과의 관계 심지어는 어황예측연구에 있어서도 오차항의 자기상관에 대해서는 아직까지 고찰된 바가 없었다. 따라서 본연구가 앞으로 수행되어질 많은 상관관계추정연구들 특히 산란과 가입연구들에 있어 요인들간의 상호관계들을 향상시키는데 중요한 참고가 될 것으로 사료된다.

## 요 약

시계열자료가 가진 자기상관은 추정된 상관관계를 왜곡시키는 요인들 중의 하나로 작용한다. 회귀모형의 잔차항에 자기상관이 있는 지를 검정하기 위해 Durbin-Watson 통계량이 흔히 쓰인다. 잔차항에 자기상관을 가진 회귀모형의 효율성을 향상시키기 위해 Yule-Walker 법, 비선형최소제곱법, 최우추정법 및 사전백색화법이 사용되어 왔다. 본 연구는 자기상관으로 인한 상관관계의 왜곡을 방지하기 위한 이들 방법들에 대해 고찰하였다. 사전백색화법을 제외한 앞의 3가지 방법을 20년간의 실제 시계열 자료에 적용하였으며 몬테카를로법을 이용하여 각 방법의 오차변이를 조사하였다. 각 방법의 평균잔차제곱분포의 경우, 최우추정법으로 추정된 평균잔차제곱이 가장 작았으며 분포 범위도 가장 작았으나 각 추정방법 사이에 유의한 차이가 발견되지는 않았다.

## 참 고 문 헌

- Beamish, R.J., C.E.M. Neville, B.L. Thomson, P.J. Harrison and M. St. John. 1994. A relationship between Fraser River discharge and interannual production of Pacific salmon (*Oncorhynchus* spp.) and Pacific herring (*Clupea pallasii*) in the strait of Georgia. *Can. J. Fish. Aquat. Sci.*, 51, 2843~2855.
- Box, G.E., G.M. Jenkins and G.C. Reinsel. 1994. *Time series analysis: Forecasting and control*. 3rd ed. Prentice-Hall, Englewood Cliffs, N.J., 592 pp.
- Brockwell, P.J. and R.A. Davis. 1995. *Time series: Theory and methods*. Springer-Verlag, New York, 577 pp.
- Caputi, N. 1988. Factors affecting the time series bias in stock-recruitment relationships and the interaction between time series and measurement error bias. *Can. J. Fish. Aquat. Sci.*, 45, 178~184.
- Caputi, N. 1993. Aspects of Spawner-recruit relationships with particular reference to crustacean stocks; A review. *Aust. J. Mar. Freshw. Res.*, 44, 589~607.
- Draper, N.R. and H. Smith. 1998. *Applied regression analysis*. 3rd ed. John Wiley & Sons, New York, 706 pp.
- Durbin, J. and G.S. Watson. 1951. Testing for serial correlation in least squares regression II. *Biometrika*, 38, 159~178.
- Hall, D.L., R. Hilborn, M. Stocker and C.J. Walters. 1988. Alternative harvest strategies for Pacific herring (*Clupea harengus pallasii*). *Can. J. Fish. Aquat. Sci.*, 45, 888~897.

- Hilborn, R. and C.J. Walters, 1992. Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. Chapman and Hall, New York, 570 pp.
- Kim, J.Y., Y.S. Kang and H.D. Jeong. 1999. Long-term variation in population biomass of Mackerel, *Scomber japonicus* and environmental factors in Korean waters. J. Korean Soc. Fish. Res., 2, 92~100 (in Korean).
- Kim, J.Y., Y.M. Kim and J.I. Kim. 1995. Variation of mackerel recruitment in the Korean Waters. Bull. Nat'l. Fish. Res. Dev. Agency, 49, 17~23 (in Korean).
- Kim, S., S. Jung and C.I. Zhang. 1997. The effect of seasonal anomalies of seawater temperature and salinity on the fluctuation in yields of small yellow croaker, *Pseudosciaena polyactis*, in the Yellow Sea. Fish. Oceanogr., 6, 1~9.
- Kope, R.G. and L.W. Botsford. 1990. Determination of factors affecting recruitment of chinook salmon, *Oncorhynchus tshawytscha*, in central California. Fish. Bull. U.S., 88, 257~269.
- Milicich, M.J., M.G. Meekan and P.J. Doherty. 1992. Larval supply: A good predictor of recruitment of three species of reef fish (Pomacentridae). Mar. Ecol. Prog. Ser., 86, 153~166.
- Park, J.H. and C.I. Baik. 2001. Prediction modeling of fishing conditions of common squid, *Todarodes pacificus*, in the East Sea, by multiple regression analysis. Bull. Nat'l. Fish. Res. Dev. Inst. Korea, 59, 9~19 (in Korean).
- Park, Y.C., M. Yoda and Y. Hiyama. 2002. Stock assessment of swordtip squid, *Loligo edulis*, in the East China Sea and the southwest sea of Japan. Fisheries Sci., (in press).
- Penn, J.W. and N. Caputi. 1986. Spawning stock-recruitment relationships for the tiger prawn (*Penaeus esculentus*) fishery in Exmouth Gulf, Western Australia. Aust. J. Mar. Freshwater Res., 37, 491~505.
- Penn, J.W., N. Caputi and N.G. Hall. 1995. Stock-recruitment relationships for the tiger prawn (*Penaeus esculentus*) stocks in Western Australia. ICES mar. Sci. Symp., 199, 320~333.
- Pindyck, R.S. and D.L. Rubinfeld. 1991. Economic models and economic forecasts, 3rd ed. McGraw-Hill, New York, 700 pp.
- Pope, J.G. 1972. An investigation of the accuracy of virtual population analysis. International Commission for the Northwest Atlantic Fisheries Research Bulletin, 9, 65~74.
- Pyper, B.J. and R.M. Peterman. 1998. Comparison of methods to account for autocorrelation in correlation analyses of fish data. Can. J. Fish. Aquat. Sci., 55, 2127~2140.
- Thompson, K.R. and F.H. Page. 1989. Detecting synchrony of recruitment using short, autocorrelated time series. Can. J. Fish. Aquat. Sci., 46, 1831~1838.
- Walters, C.J. 1985. Bias in the estimation of functional relationships from time series data. Can. J. Fish. Aquat. Sci., 42, 147~149.
- Walters, C.J. 1990. A partial bias correction factor for stock recruitment parameters in the presence of autocorrelated environmental effects. Can. J. Fish. Aquat. Sci., 47, 516~519.
- Zhang, C.I. and J.B. Lee. 2001. Stock assessment and management implications of horse mackerel (*Trachurus japonicus*) in Korean waters, based on the relationship between recruitment and the ocean environment. Progress in Oceanography, 49, 513~537.
- 최병선. 1992. 단변량 시계열분석1. 세경사, 736 pp.

## Appendix 1. Durbin-Watson 통계량에 대한 유의성 범위

Table A-1. Significance points of  $d_L$  and  $d_U$ : 5%

N	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Table A-2. Significance points of  $d_L$  and  $d_U$ : 2.5%

N	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
20	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.70	0.70	1.87
25	1.18	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
30	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	0.98	1.73
35	1.30	1.42	1.25	1.48	1.19	1.55	1.13	1.63	1.07	1.70
40	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
45	1.39	1.48	1.34	1.53	1.30	1.58	1.25	1.63	1.21	1.69
50	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
55	1.45	1.52	1.41	1.56	1.37	1.60	1.33	1.64	1.30	1.69
60	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
65	1.49	1.55	1.46	1.59	1.43	1.62	1.40	1.66	1.36	1.69
70	1.51	1.57	1.48	1.60	1.45	1.63	1.42	1.66	1.39	1.70
80	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
90	1.57	1.61	1.55	1.64	1.53	1.66	1.50	1.69	1.48	1.71
100	1.59	1.63	1.65	1.65	1.55	1.67	1.53	1.70	1.51	1.72

Table A-3. Significance points of  $d_L$  and  $d_U$ : 1%

N	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

**Appendix 2. Yule-Walker 법, 최우추정법, 비선형최소제곱법**

식 (2)에서 오차항  $c$ 가  $m$ 차 자기회귀 (AR(m)) 모형을 갖는다면 다음식을 만족한다.

$$c_t = \Phi_1 c_{t-1} + \Phi_2 c_{t-2} + \dots + \Phi_m c_{t-m} + v_t$$

$$(c = c_1, c_2, \dots, c_N)$$

여기서  $\Phi$ 는 자기회귀계수이고  $N$ 은 관찰점의 수이며  $\{v_t\}$ 는 서로 독립이며 평균이 0이고 분산이  $\sigma_v^2$ 인 확률과정이다.

**Yule-Walker 법**

만약 자기회귀계수의 벡터를  $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_m)'$  그리고 오차벡터의 변수행렬을  $c = (c_1, c_2, \dots, c_N) = \Omega \lambda$  라 두면 오차항의 공분산행렬은 다음과 같다.

$$E(cc') = \Omega = \sigma^2 V$$

이때의  $V$ 는 상관계수 행렬이라 한다. 만약 자기회귀계수들의 벡터를 안다면  $V$ 는 자기회귀계수들에 의해 계산되어질 수 있다. 따라서 주어진  $V$ 로부터 효율적인 회귀계수 추정값  $\beta$ 은 일반화최소제곱법 (Generalized Least Squares; see <http://www.sinica.edu.tw/as/ssrc/ckuan/pdf/et01/ch4.pdf>)을 이용하여 구해질 수 있다.

상관계수행렬  $V$ 를 모를 경우에도  $\beta$ 는 자기상관함수를 만족하는 Yule-Walker 방정식을 풀어서 구한  $\Phi$  추정치와 일반화최소제곱법을 이용하여 구할 수 있다. 이때 자기상관함수는 보통의 최소제곱법을 사용해서 종속변수를 독립변수들에 회귀시켜 얻은 잔차  $e$ 로부터 구해진다. 즉 자기상관함수  $r$ 은

$$r_j = \frac{\sum_{i=1}^{N-j} e_i e_{i+j}}{\sum_{i=1}^N e_i^2} \quad (\text{단 } j=0, 1, 2, \dots, m)$$

오차항이 AR(m) 모형인 자기회귀계수들은 다음과 같은 Yule-Walker 방정식의 해로서 추정할 수 있다 (for more details, see 최, 1992, pp. 380~381).

$$\begin{bmatrix} r_0 & r_1 & r_2 & \dots & r_{m-1} \\ r_1 & r_0 & r_1 & \dots & r_{m-2} \\ r_2 & r_1 & r_0 & \dots & r_{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{m-1} & r_{m-2} & r_{m-3} & \dots & r_0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_m \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_m \end{bmatrix}$$

그 다음  $V$ 는 추정된 자기회귀계수들로부터 계산되며 마지막으로  $\beta$ 는  $V$ 를 사용하여 일반화최소제곱법에 다음과 같이 구해진다.

$$\beta = (X'V^{-1}X)^{-1}X'V^{-1}y$$

Yule-Walker 추정법은 반복적으로 사용할 수 있는데 이 경우, 일반화최소제곱 추정법으로부터 얻은 잔차를 사용하여 새로운 자기상관함수를 계산하고 이 결과는 다시 새로운 자기회귀계수들과  $V$ 값을 추정하는데 이용되며 일반화최소제곱 추정치는 새로운  $V$ 값에 의해 다시 계산되게 된다. 이러한 과정은 새로 계산되는 자기회귀계수 추정치가 일정한 값에 수렴할 때까지 반복적으로 수행하게 된다.

**최우추정법**

시계열자료  $y_1, y_2, \dots, y_N$ 에 대해 우도함수는  $N$ 관찰에 대한 결합확률밀도함수로 나타내어진다. 우도함수  $L$ 은 계수들의 함수가 되며  $N$ 관찰에 대한 결합확률밀도를 최대화시킴으로 추정될 수 있다. 일반적으로 관찰에 대한 확률과정은 정규분포를 따른다고 가정되며 (즉,  $\int p dy = 1$ ) 우도함수보다 대수를 취한 대수우도함수 (log likelihood function)를 사용하는 것이 편리하다.

대수우도함수는 다음과 같이 나타낼 수 있다.

$$\ln(L) = -N \ln(2\pi) / 2 - N \ln(\sigma^2) - \ln(|V|) / 2 - S / (2\sigma^2)$$

여기서  $|V|$ 는 상관계수행렬  $V$ 의 결정계수 (즉, 자기회귀계수  $\Phi$ 의 함수)이며  $S$ 는 비조건부 오차제곱합이다. 오차항의 분산에 대해 대수우도함수를 최대화시킨 다음 상수항을 제외하면  $-N \ln(|S|V|^{1/N}) / 2$ 가 되며 최우추정값은  $S$ 를 최소화함으로 결정된다. 이때  $S$ 는 다음과 같이 나타낼 수 있다.

$$S = |L|^{1/N} e' e |L|^{1/N} \quad (\text{단, } e = L'c)$$

**비선형최소제곱법**

오차항이 AR(m) 모형인 경우 비조건부 제곱합  $S$ 는 다음과 같이 나타낼 수 있다.

$$S = c' V c$$

비선형최소제곱값은 계수  $\beta$ 와  $\Phi$ 에 대해  $S$ 를 최소화함으로써 구해진다.