

## 확률적 변동성 모형과 자기회귀이분산 모형의 비교분석

이용훈<sup>1)</sup> 김삼용<sup>2)</sup> 황선영<sup>3)</sup>

### 요약

시간의 경과에 따라 관측된 시계열 자료를 통해 데이터 분석을 하고 적당한 모형을 생성함으로써 미래 시점을 예측하는 방법들은 그 동안 많은 방법들이 제시되었고 연구되고 있다. 그 중 최근 들어 과거의 데이터를 바탕으로 관측된 각 시점에서의 분산을 서로 다른 분산(조건부 이분산성)을 따른다고 가정하고, 이를 분석하는 모형(ARCH, GARCH, Stochastic Volatility(SV))들이 옵션 가격분석이나 환율 변화 등 경제 시계열 자료의 예측 모형을 위하여 활발히 연구되고 있다. 본 논문에서는 한국의 KOSPI 데이터(1995년 1월 3일부터 2001년 12월 28일, 총 1906일)를 바탕으로 (조건부) 우도함수 모수 추정 방법을 이용한 GARCH(1,1) 모형과, MCMC 방법을 이용하여 모수를 추정한 SV 모형을 적용시켜 보고 각 모형들의 예측 정확도를 비교하여 보았다.

주요용어: GARCH, KOSPI, MCMC, Volatility

### 1. 서론

전통적으로 시계열 자료 분석 방법은 AR, MA, 및 ARMA 모형 등 선형성을 가정한 모형들이 많이 활용되었다. 그러나 여러 경제 시계열 자료에서 나타나는 도약 현상(jump phenomena)이나 한계 순환(limit cycles)과 같은 현상은 선형성 모형으로 잘 설명이 되지 않는다. 최근에 들어 Engle(1982)이 오차의 분산이 자기회귀적으로 변하는 조건부 이분산 자기회귀모형인 ARCH(autoregressive conditional heteroscedastic)모형을 제시하였으며, 1986년에 Bollerslev가 오차의 분산에 이전 데이터들의 값과 분산을 고려하여 ARCH 모형을 일반화 시킨 GARCH(generalized ARCH) 모형을 제시 하였다.

반면에 확률적 변동성(stochastic volatility, 이하, SV) 모형은 변동성을 잠재변수로 이용하여 각 시점에서의 오차의 분산을 구하는 모형으로서, 이는 1973년에 Black과 Scholes가 옵션가격 분석에 처음 제시 하였으며, 1987년에 Hull과 White는 Black과 Scholes가 제안한 모형을 좀 더 일반화하여 옵션가격 분석에 이용하였다. 본 논문에서는 한국의 KOSPI(한국 종합주가지수)데이터를 GARCH(1,1) 모형과 Hull과 White가 제안한 SV 모형에 적용시켜 보고, 오차의 분산에 대한 예측 정도를 비교하여 보았다. 여기서 GARCH 모형 예측 방법

1) (150-070) 서울시 동작구 흑석동 221 중앙대학교 통계학과 박사과정

E-mail : withus@pharmacodesign.com

2) (150-070) 서울시 동작구 흑석동 221 중앙대학교 통계학과 조교수

E-mail : sahm@cau.ac.kr

3) (140-742) 서울시 용산구 청파동 2가 53-12 숙명여자대학교 통계학과 교수

E-mail : shwang@sookmyung.ac.kr

은 현재 가장 보편적으로 사용되는 최우추정법(이하, MLE) 방법을 사용하였고, SV 모형에서는 근사적 추정 방법인 유사우도(quasi-likelihood) 모수 추정 방법에 비해 정확한 모수의 추정치를 구할 수 있는 MCMC(Markov chain Monte Carlo) 방법을 이용하여 모수를 추정하였다(cf. Kim et al. (1998)).

## 2. 확률적 변동성 (SV) 모형

SV 모형이란 1973년에 Black과 Scholes가 옵션가격 분석에 처음 제안한 이래로 여러 학자들에 의해 여러 가지 변환된 모형들과, 연관된 타당성, 모수 추정 방법에 대해 많은 연구가 되고 있다. 1987년에 Hull과 White는 Black과 Scholes가 제안한 모형을 좀 더 일반화하여 옵션가격 분석에 이용 했는데, 이 모형이 SV 모형의 표준적인 형태로 사용되고 있다.

$$\begin{aligned} y_t &= \beta e^{h_t/2} \epsilon_t \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t \\ h_1 &\sim N\left(\mu, \frac{\sigma^2}{1 - \phi^2}\right) \end{aligned} \quad (2.1)$$

여기서  $y_t$  는 관측값의 평균을 보정한 값이고, 대수-변동성(log-volatility)이라 불리는  $h_t$  는 ( $|\phi| < 1$ )인 정상 시계열을 따른다고 가정한다. 또  $\epsilon_t$ 와  $\eta_t$ 는 서로 독립이고 각각이 모든 시점에서 서로 독립이며 표준정규분포를 따른다. 위 모형에서 모수  $\beta$ 와  $e^{\mu/2}$ 는 모형 전체에서의 곱셈요소(scale factor)로서의 역할을 하며, 모형화 되지 않는 본래의 변동 부분이라고 생각할 수 있으며, 모수  $\phi$ 는 변동성에서의 지속적인 관계의 척도, 즉 오차의 분산에서 예측되는 부분의 정도를 의미하며,  $\sigma_\eta$ 는 대수-변동성의 표준편차이다.

위 모형은 관측되지 않는 잠재(latent) 변수인 대수-변동성을 바탕으로 관측값 변동의 패턴을 분석함으로써, 이전 시점들에서의 관측값과 변동의 값을 고려하여 현 시점에서의 값의 변동량을 예측하는 GARCH모형과는 차이가 있다. (2.1) 모형에서 나타나는 모수인  $\beta, \mu, \phi, \sigma_\eta$  들을 예측하는 방법으로는 여러 가지 방법들이 제안되었는데, 본 논문에서는 정확한 모수의 추정값을 구하기 위한 방법으로 잘 알려진 MCMC 기법을 이용하도록 한다. MCMC 및 연관된 Gibbs sampling에 대해서는 Casella and George(1992), Gelfand and Smith(1990), Geman and Geman(1984) 그리고 Tanner(1993)를 참고하기 바란다.

MCMC의 가장 기본적인 개념은 여러 변수들이 서로 연관관계를 가지는 다변량 변수일 때 주어진 다른 변수들을 조건으로 변수 값들을 하나씩 생성해 기존의 연관관계를 유지하는 새로운 데이터를 생성하고 이를 바탕으로 마코프 연쇄와 이전 단계에서 만들어진 데이터  $V$ 을 통해 다음 단계의 데이터  $V$ 을 생성하는 것이다.

예를 들어  $Z = (z_1, z_2, \dots, z_j)$ ;  $j$ 개의 변수를 갖는 다변량 분포  $Z$ 가 있을 때, 각각  $t, t+1$  단계에서,  $Z^{(t)}, Z^{(t+1)}$  는 다음과 같이 표현 될 수 있다.

$$Z^{(t)} = (z_1^{(t)}, z_2^{(t)}, \dots, z_j^{(t)}), \quad Z^{(t+1)} = (z_1^{(t+1)}, z_2^{(t+1)}, \dots, z_j^{(t+1)})$$

여기서 주어진  $Z^{(t)}$ 로 부터 바로 다음 단계의  $Z^{(t+1)}$ 를 생성한다.

$$\begin{aligned} z_1^{(t+1)} &= P(Z_1|z_2^{(t)}, \dots, z_j^{(t)}) \\ z_2^{(t+1)} &= P(Z_2|z_1^{(t+1)}, z_3^{(t)}, \dots, z_j^{(t)}) \\ z_j^{(t+1)} &= P(Z_j|z_1^{(t+1)}, z_2^{(t+1)}, \dots, z_{j-1}^{(t+1)}) \end{aligned}$$

이제  $Z^{(t+1)}$ 을 바탕으로 바로 다음 단계인  $Z^{(t+2)}$ 를 생성한다. 이와 같은 과정은 베이저안 접근 방법에서의 사후확률을 바탕으로 새로운 모수 값을 생성하는 사후시뮬레이션방법과 동일하다.

SV 모형을 이용하여 KOSPI 데이터에서 분석함에 있어서 Kim et al.(1998)의 방법론을 적용시켜 보았으며 다음의 6단계를 거쳐 수행하였다.

1단계. 잠재 벡터  $h$ 와 모수 벡터  $\theta$ 를 초기화 한다.

2단계.  $h_t, h_t | h_{\{t\}^c}, y, \theta, t = 1, \dots, n$ 에서  $h_t$ 를 추출한다.

(주어진 모수 벡터, 관측치 벡터  $y$ , 그리고 잠재 벡터를 바탕으로 새로운 잠재 벡터의 원소를 하나씩 새로 생성한다.)

3단계.  $\theta_\eta^2 | y, h, \phi, \mu, \beta$ 를 추출한다.

(주어진 벡터  $y$ , 잠재 벡터  $h, \phi, \mu, \sigma_\eta^2$  하에서 잠재 벡터의 오차를 생성한다.)

4단계.  $\phi | h, \mu, \beta, \sigma_\eta^2$ 를 추출한다. (잠재 벡터  $h, \mu, \beta, \sigma_\eta^2$  하에서  $\phi$ 를 생성한다.)

5단계.  $\mu | h, \phi, \sigma_\eta^2$ 를 추출한다. (잠재 벡터  $h, \phi, \sigma_\eta^2$  하에서  $\mu$ 를 생성한다.)

6단계. 2단계로 돌아간다.

이번 연구에서는 먼저 KOSPI 데이터를 평균이 0인 정상 시계열 자료로 만들기 위하여 다음과 같은 변환을 행하였다.

$$y_t = 100 \times \{(\log r_t - \log r_{t-1}) - \frac{1}{n} \sum_{i=1}^n (\log r_i - \log r_{i-1})\} \quad (2.2)$$

각 모수의 초기치 값으로는  $\phi, \mu, \sigma_\eta^2$ 에 각각 0.95, 1, 0.02를 할당하였으며, 잠재 벡터  $h$ 는 모든 원소에 0을 할당하였다. MCMC과정을 수행하기 과정 중 2번째 단계의 과정이 가장 복잡하고 어려운 과정으로, 관측되지 않은 잠재 벡터를 생성하는 과정은 채택-기각 절차(accept-reject procedure)를 이용하여 주어진 관측 자료의 값 하에서 특정 확률 값을 기준으로 새로이 생성된 잠재 벡터의 원소의 값을 받아 들이고 그 확률 값보다 작은 경우에는 다시 새로운 값을 생성하는 과정을 거쳤다. 본 연구에서는 채택 확률로 0.995를 사용하였다. 그리고  $h$ 의 각 원소의 생성은 식(2.1)에 의해 정규분포를 이용하였다. 또한  $\sigma_\eta^2$ 는 inverse gamma(IG) 분포를 이용하였고 따라서 새로운  $\sigma_\eta^2$ 는 IG-분포에서 생성되었다. 본 연구에서는 prior로 IG(2.5, 0.025)를 사용하였다.  $\phi$ 는 normal 분포에서 생성하였으며 현 단계에서 생성된  $\phi^*$ 의 값이 정상 영역에 있지 않을 때에는 이전 단계에서 만들어진 값으로 현 단계의  $\phi$ 를 대체하였다.  $\mu$ 는 정규 분포에서 생성하였으며,  $\beta$ 는  $e^{\mu/2}$ 로 대체하였다(cf. Chib과 Greenberg(1996)).

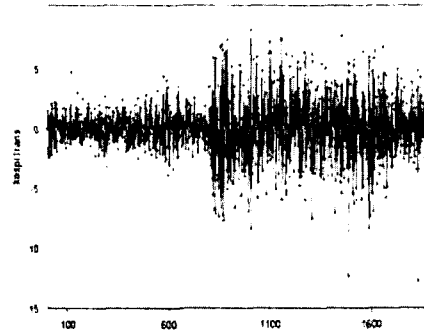
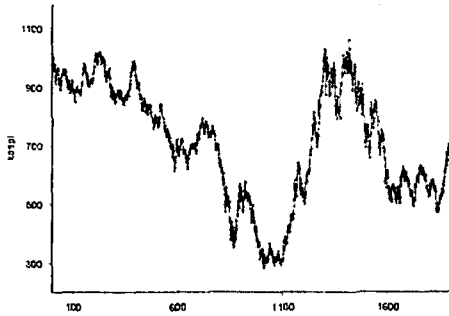


그림 2.1: 6년간의 KOSPI 데이터      그림 2.2: 정상화된 KOSPI 데이터(평균수준=0)

전체 과정으로는 최초 1000번의 단계를 통하여 잠재 벡터를 생성하였으며, 15000번의 단계 중 후반부 10000번의 각 모수의 생성 값을 구하여 모수 추정값의 평균과 분산을 계산해 보았다. 다음 표와 그림은 MCMC에 의한 SV모형의 모수추정 결과와 이를 이용한 3배 표준편차( $3\sigma$ ) 한계에서의 신뢰(예측)구간이다.

표1. MCMC를 이용한 SV 모형 모수 추정치 결과

Parameter	N	Mean	Std Dev	Minimum	Maximum
$\phi$	10000	0.9686577	0.0075967	0.923361	0.9995085
$\sigma_{\eta}$	10000	1.0686523	0.085967	0.943463	1.1457832
$\beta$	10000	1.3646331	1.0515767	-2.2992959	4.2224444

### 3. 자기회귀이분산(GARCH) 모형 및 비교분석

GARCH 모형은 오차의 분산이 자기회귀적으로 변하는 조건부 이분산 자기회귀모형인 ARCH 모형의 일반화된 모형으로써 Bollerslev(1986)에 의해 제시되었다. GARCH 모형은 장기기억(long memory) 특성을 가지고 있으며 본 논문에서는 지수적으로 감소하는 장기기억 모형 중 대표적인 GARCH(1,1) 모형을 고려하도록 한다.

$$y_t = \epsilon_t, \quad t = 1, 2, \dots, n$$

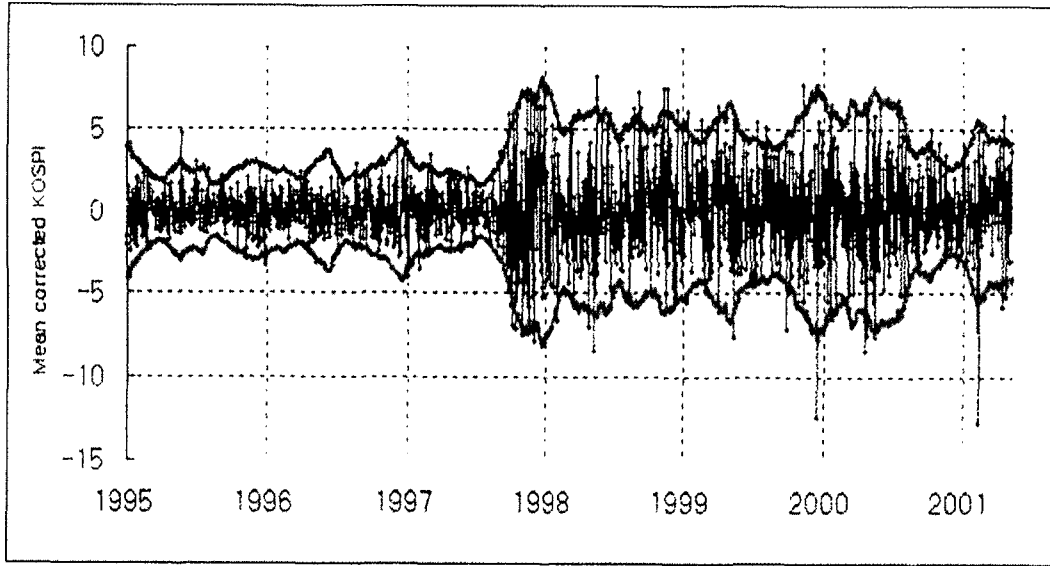


그림 2.3: SV모형을 이용한 KOSPI 자료의  $3\sigma$  한계에서의 신뢰구간과 관측값

$$\epsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

GARCH 모형의 모수는 일반적으로 MLE 방법을 통하여 추정한다.  $\epsilon_t$ 가 조건부적으로 정규분포를 따르므로 로그우도함수를 다음과 같이 표현할 수 있으며 이를 최대화 시키는 과정을 통해 MLE를 구할 수 있다.

$$l(\epsilon_y; y_t) = \sum_{i=1}^n \left( -\ln(2\pi) - \ln(\sigma_t) - \frac{\epsilon_t^2}{2\sigma_t^2} \right)$$

[표2]는 GARCH(1,1) 모형에서 MLE 추정치에 대한 결과이며 연관된 신뢰구간( $3\sigma$ 한계)은 [그림3.1]에 수록하였다. 단,  $\sigma^2 = \alpha_0 / (1 - \alpha_1 - \alpha_2)$ 이다.

표2. GRACH(1.1)모형에서의 모수의 추정치

Parameter	Estimate	p-value
$\alpha_0$	0.01	0.0078
$\alpha_1$	0.07	0.0001
$\beta_1$	0.92	0.0001

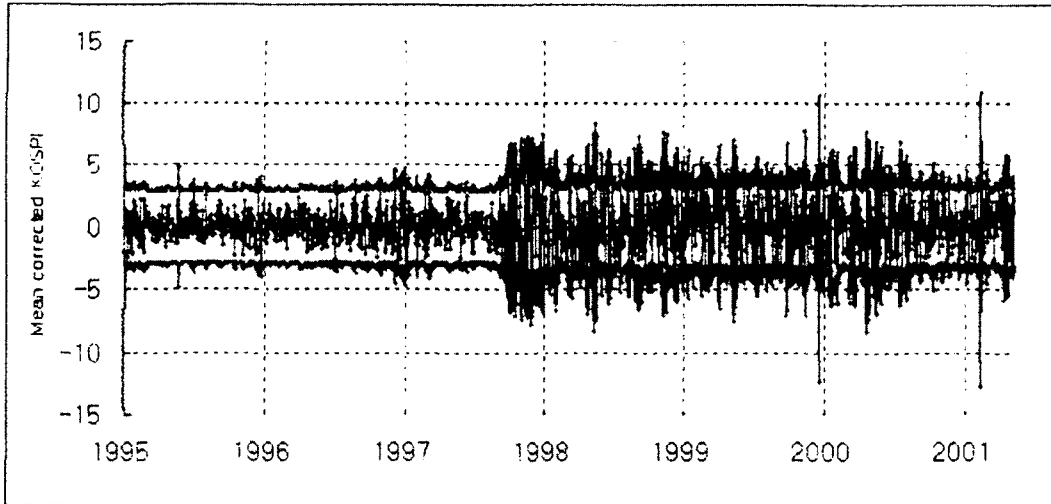


그림 3.1: GRACH(1,1)모형을 이용한  $3\sigma$  한계에서의 신뢰구간과 관측값

끝으로, 두 모형의 예측정도를 비교하기 위해 그림 2.3, 3.1로부터 각 관측치가 3배의 표준편차 밖으로 나갈 기준확률(0.0027) 하에서 GARCH모형을 적합 시켜서 나온 예측치와 SV모형 하에서의 예측치가 이 3배의 표준편차 밖으로 나갈 확률을 [표3]에 수록하였다. 이 표에서 GARCH 모형 하에서의 확률은 0.10 인 반면 SV모형 하에서의 확률은 0.046으로서 SV 모형이 약 2배 이상 우수함을 나타낸다고 할 수 있다. 또한, [표3]을 보면 오차 절대값의 기대값이 수록되어 있는데 SV모형이 0.04, GARCH(1,1)모형이 0.10로서 SV가 우수하다고 볼 수 있다. 결론적으로 SV 모형이 GARCH(1,1)보다 더 우수한 분석 결과를 제공한다고 할 수 있다. [표4]의 결과도 비슷한 양상을 보이고 있는데 GARCH 모형 하에서의 확률은 0.22인 반면 SV모형 하에서의 확률은 0.17로서 SV 모형의 우수성을 나타낸다고 할 수 있다.

표3. SV 모형과 GARCH 모형의 비교 ( $3\sigma$ )

모형	$P( X  > 3\sigma)$ $X \sim N(0, \sigma^2)$	Number	$P( y_t  > 3\sigma)$
SV모형	0.0027	89	0.046719
GARCH(1,1)모형		193	0.101312

표4. SV 모형과 GARCH 모형의 비교 ( $3\sigma$ )

모형	$P( X  > 1.96)$ $X \sim N(0, 1)$	Number	$P( y_t  > 3\sigma)$
SV모형	0.05	327	0.171654
GARCH(1,1)모형		414	0.217354

#### 4. 감사의 글

본 논문을 심사해 주신 두분의 심사위원님께 깊은 감사를 드립니다.

#### 참고문헌

- [1] Bollerslev, T.(1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 69, 543-547.
- [2] Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46, 167-174.
- [3] Chib, S. and Greenberg, E. (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, 12, 409-431.
- [4] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.
- [5] Geman, D. and Geman, S. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian reconstruction of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- [6] Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility; Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65, 361-393.
- [7] Tanner, M. A. (1993). *Tools for Statistical Inference, Methods for the Exploration of Posterior Distribution and Likelihood Function*. Springer-Verlag, New York.

[ 2003년 4월 접수, 2003년 4월 채택 ]

## Stochastic Volatility Model vs. GARCH Model : A Comparative Study

Y.H. Lee<sup>1)</sup> S. Kim<sup>2)</sup> S.Y. Hwang<sup>3)</sup>

### ABSTRACT

The volatility in the financial data is usually measured by conditional variance. Two main streams for gauging conditional variance are stochastic volatility (SV) model and autoregressive type approach (GARCH). This article is conducting comparative study between SV and GARCH through the Korean Stock Prices Index (KOSPI) data. It is seen that SV model is slightly better than GARCH(1,1) in analyzing KOSPI data.

*Keywords:* GARCH; KOSPI; MCMC; volatility

---

1) Doctoral student, Dept. of Statistics, Chung-Ang Univ.

E-mail : withus@pharmacodesign.com

2) Assistant Professor, Dept. of Statistics, Chung-Ang Univ.

E-mail : sahm@cau.ac.kr

3) Professor, Dept. of Statistics, Sookmyung Women's Univ.

E-mail : shwang@sookmyung.ac.kr