

기하브라우니안모션 모형을 이용한 주가시계열 분석*

김진경¹⁾

요약

기하브라우니안모션(geometric Brownian motion) 모형과 자기상관(autoregressive) 모형을 이용하여 최근 우리나라의 주가(지수)시계열을 분석하고, 이 두 모형을 예측의 관점에서 비교하였다. 고려한 7개의 주가(지수)시계열 모두에서 예측을 시행할 때 이용하는 자료의 개수가 작을수록 기하브라우니안모션 모형이 상대적으로 더 나은 예측치를 주는 것으로 나타났다.

1. 서론

주가시계열은 한나라 경제의 중요한 지표로서 그에 대한 적절한 분석법의 개발은 오랫동안 연구대상이 되어 왔다. 주가시계열에 대한 연구들은 매우 방대하며 각 나라에 따라 또한 각 시기에 따라 여러 모형이 제안되었다. (참고: Shapiro(1991)) 그 중 가장 단순하면서도 광범위하게 받아들여지는 모형으로는 자기상관(AR)모형이 있다. 이 논문에서는 이 AR 모형과 확률과정모형 중 기하브라우니안모션(Geometric Brownian Motion(GBM), Ross(1993)) 모형을 이용하여 최근 우리나라의 주가시계열을 분석하고 이 두 모형을 비교하고자 한다. 기하브라우니안모션 모형은 주가시계열을 설명하는데 적합하다고 알려져 있는 것으로, 파생금융상품의 가치결정에 주로 쓰이는 모형이다. 옵션의 가치를 결정하는 모형으로 유명한 Black-Scholes 모형도 기초자산이 기하브라우니안모션 모형을 따른다는 가정하에 얻어진 것이다. (참고: Hull(1993)) 분석대상이 되는 주가시계열은 종합주가지수(1989.3.2 - 1996.5.22)와 다양한 특성을 나타내면서 주식시장에서 몇 개의 전형을 나타내는 개별의 주식가격 시계열; 태광산업, 한국이동통신, 성장기업, 삼성전자, 한국전력, 포항제철이다. (그림 1.1 - 1.7 참조) 이 각각의 시계열에 AR 모형과 GBM 모형을 이용하여 값들을 예측하고 두 모형 중 어느 모형이 주가시계열을 예측하는데 더 적합한가를 판단하고 각각의 시계열에 어떠한 공통점과 다른점이 있는지 살펴보고자 한다.

또한 주말(토요일 - 월요일)의 주가변동이 주중의 일일 주가변동보다 크리라고 예상된다. 이와같은 현상을 주말효과(weekend effect, Roll(1984))라 부를 수 있는데, 일일자료에서 이와같은 주말효과가 뚜렷하게 나타나는 경우 일일자료에 대한 분석에는 어려움이 있다. 이런 경우에 주말 관측치를 누락된 것으로 판단하여 분석을 시행하거나 일일자료에 대한 분석 대신 주말효과가 없는 주단위자료에 대한 분석이 더 바람직할 수도 있다. 다음 절에서는 종합주가지수에 대하여 주말효과가 있는지 조사하였다.

* 본 연구는 인하대학교 1996-97년도 교내연구비에 의해 수행되었다.

1) (402-751) 인천광역시 남구 용현동 253번지, 인하대학교 이과대학 통계학과, 교수



그림 1.1: 종합주가지수 (1989.3.2.-1996.5.22)

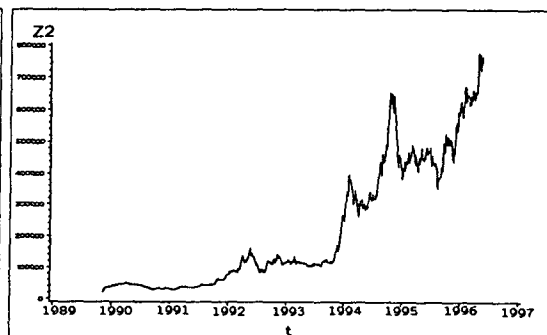
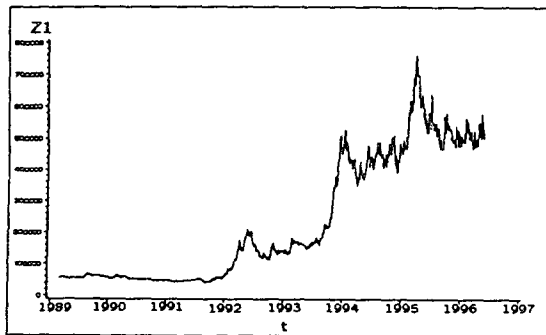


그림 1.2: 태광전자주가 (89.3.2-96.5.22) 그림 1.3: 한국이동통신주가 (89.11.7-96.5.22)

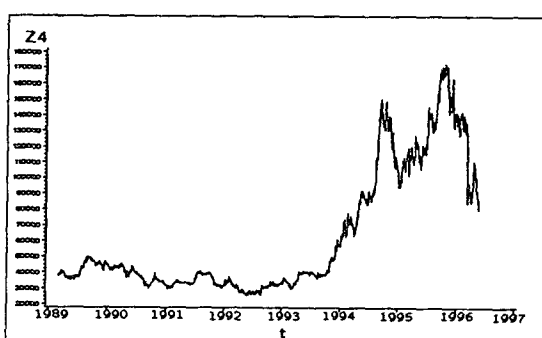
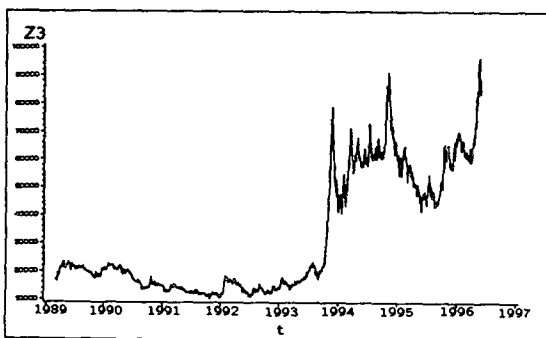


그림 1.4: 성장기업주가 (89.3.2-96.5.22) 그림 1.5: 삼성전자주가 (89.3.2-96.5.22)

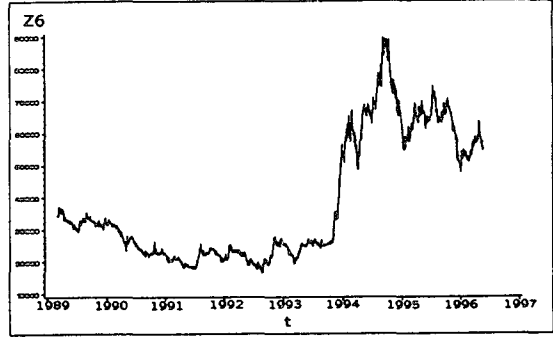
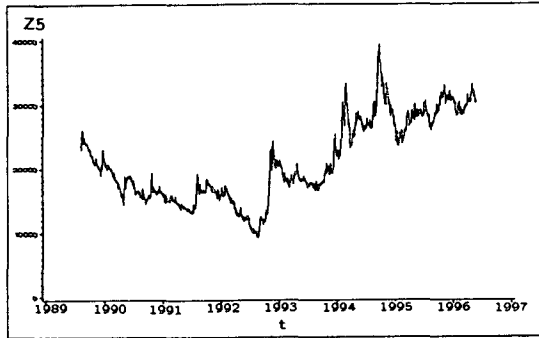


그림 1.6: 한국전력주가 (89.8.10-96.5.22) 그림 1.7: 포항제철주가 (89.3.2-96.5.22)

2. 주말효과

일일 종합주가지수(Z_t)가 일반적인 AR(1) 모형을 따른다고 하자. 즉,

$$Z_t = \alpha_0 + \alpha_1 Z_{t-1} + U_t, \quad U_t \sim iid(0, \sigma^2), \quad 0 < \alpha_1 \leq 1$$

이 관계는 매일 매일의 주가지수의 변동을 설명하는 것으로 이러한 관계가 주말에도 이어진다면 주말에 일어나는 변동($Z_t - Z_{t-2}$)은 주중에 일어나는 변동($Z_t - Z_{t-1}$)에 비해 그 값의 변화가 크리라 예상된다. 이론적으로 $\alpha_1 = 1$ 인 경우 주말의 변동에 대한 표준편차는 주중의 변동의 표준편차의 $\sqrt{2}$ 배가 된다. (참고: 김진경(1995)) 이와 같은 주말효과를 실제 종합주가지수 시계열자료로부터 확인하기 위해서 다음 두가지의 수치를 각년도 상·하반기에 대해서 주중과 주말 두 경우에 계산하였다. 연속되는 날, 또는 토요일과 다음 월요일 사이에 일어나는 종합주가지수의 변동을 d 라 할 때

평균($|d|$): 주가지세열의 변동의 절대치의 평균

표준편차(d): 주가지세열의 변동의 표준편차

를 구해 보니 그 결과가 아래의 표 2.1과 같았다. 이 표를 보면 평균($|d|$)와 표준편차(d) 모두에 있어서 15번 중 12번의 경우에 주말의 수치가 주중보다 크게 나타남을 알 수 있다. 이로부터 어느 정도 주말효과가 존재함을 알 수 있다. 그러나 주말변동의 표준편차(d)가 주중변동의 표준편차의 1.181배로 $\sqrt{2} = 1.414$ 배 만큼 크지 않으므로 AR(1)의 특성이 일요일에도 완전하게 유지된다고 보기는 어렵다. 이는 우리나라에 있는 상·하반기 제도로 인하여 변동의 폭이 어느정도 제한을 받기 때문인 것으로 보인다. 이와같은 제한은 개별 주가의 경우에 더욱 확실하여 종합주가지수의 경우보다 주말효과가 크지 않으리라고 예상된다. 실제로 한국이동통신 주가에 대하여 주말과 주중 변동의 표준편차를 계산한 결과, 주말

표 2.1: 주말효과 (주중/주말)

		평균(d)	표준편차(d)
1989	상	6.37 / 8.97	8.36 / 10.55
1989	하	6.42 / 5.98	9.02 / 7.27
1990	상	7.76 / 10.26	10.48 / 11.80
1990	하	8.34 / 9.41	11.40 / 12.86
1991	상	5.32 / 5.67	8.03 / 7.80
1991	하	6.67 / 8.79	8.90 / 11.45
1992	상	6.43 / 4.97	8.26 / 6.36
1992	하	7.42 / 11.84	9.48 / 13.64
1993	상	5.30 / 7.91	6.82 / 9.76
1993	하	6.45 / 9.32	8.68 / 11.03
1994	상	7.32 / 8.51	9.03 / 10.86
1994	하	6.37 / 9.62	7.83 / 10.43
1995	상	6.73 / 10.35	8.69 / 13.65
1995	하	7.14 / 8.11	8.97 / 10.02
1996	상	6.53 / 7.49	8.31 / 8.76

의 표준편차가 주중의 표준편차의 0.976배로 오히려 작게 나타났다. 따라서 이 논문에서는 주말효과가 어느정도 존재함은 인정하되 이를 묵인할 수준이라 판단하고 일일자료에 대한 분석을 시도하기로 한다.

3. AR 모형과 기하브라우니안모션(GBM) 모형

이 절에서는 주가 시계열에 AR 모형과 GBM 모형을 적합시켜 두 모형을 비교하고자 한다. 앞에 주어져 있는 그림을 보면 종합주가지수를 제외한 6개 개별 주가시계열의 경우 진동의 폭이 시간에 따라 변화하는 것을 볼 수 있다. 즉, 1993-4년을 전후하여 진동의 폭이 증가하는 것을 볼 수 있다. 이와 같은 경우에는 시계열에 변환을 행한 후 분석을 시도하는 것이 바람직하므로 여기서는 log변환을 행하였다. 그림 3.2를 보면 태광전자주가 시계열의 경우 log변환으로 진동의 폭이 어느정도 일정하게 유지되는 것을 볼 수 있다. 다른 개별 시계열의 경우에도 여기에 제시하지는 않으나 log변환으로 태광전자 경우와 같은 경향이 나타나는 것을 알 수 있다. 종합주가지수의 경우에는 그림 3.1에서와 같이 별 변화를 감지할 수 없으나 분석의 편리를 위하여 변환을 시행하였다. 실제로 AR(1)의 경우에 변환전, 후의 R^2 값을 비교해보면 대부분의 경우에 큰 변화는 없으나 전반적으로 약간 증가하는 것으로 나타났다.



그림 3.1: log변환된 종합주가지수의 초기값대비

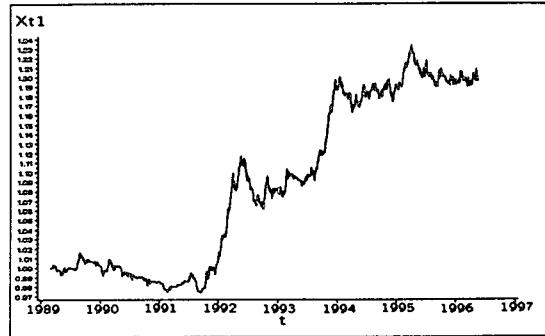


그림 3.2: log변환된 태광전자주가의 초기값대비

주어진 시간 t 에서의 log 변환된 주가(지수)를 Z_t 라 하고, X_t 를 초기값대비, 즉 $X_t = Z_t/Z_0$ 라 하자. 이 X_t 를 예측하기 위한 AR 모형은

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + U_t, \quad U_t \sim NI(0, \sigma_u^2)$$

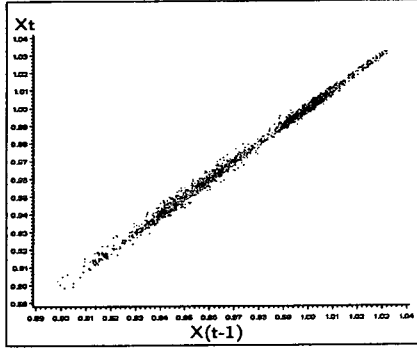
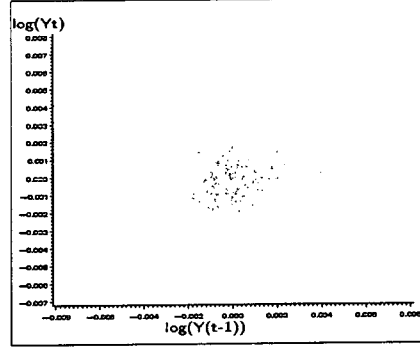
이다. 종합주가지수의 경우 AR 모형을 적합시킨 결과 AR(1)과 AR(2)가 R^2 값에서 별 차이가 없고, 부분자기상관함수(Partial Autocorrelation Function, ϕ_k)의 값도 다음 표와 같이 $k = 1$ 인 경우 0.998이고 $k \geq 2$ 인 경우 거의 0에 가까운 값을 갖는 것을 알 수 있다. 따라서 AR 모형으로 AR(1)을 사용하기로 한다. (참고: Wei(1994)) 종합주가지수 이외의 주가 시계열의 경우에도 여기 제시하지는 않으나 종합주가지수의 경우와 유사하다.

k	1	2	3	4	5	6	7	8	9	10
ϕ_k	0.998	-0.072	0.036	-0.044	0.002	-0.016	0.006	0.017	0.026	0.001

따라서 여기서는 AR(1) 모형을 이용하여 X_t 를 다음과 같이 예측할 수 있는데,

$$\hat{X}_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_{t-1} \tag{3.1}$$

이때 $\hat{\alpha}_0$ 와 $\hat{\alpha}_1$ 은 최소제곱추정량으로 구하기로 한다. 그림 3.3은 종합주가지수 경우의 X_t 대 X_{t-1} 의 그림을 그린 것으로 거의 직선에 가깝다는 것을 볼 수 있다. 다른 주가 시계열의 경우도 거의 직선에 가까우며 다만 그 폭이 약간 변동하는 것을 볼 수 있다.

그림 3.3: X_t 대 X_{t-1} 그림 3.4: $\log Y_t$ 대 $\log Y_{t-1}$

GBM 모형을 이용하여 X_t 를 예측하는 과정을 간단히 설명하면 다음과 같다. 우선 Y_t 를 전일값대비, 즉, $Y_t = Z_t/Z_{t-1} = X_t/X_{t-1}$ 라 하면,

$$\begin{aligned} X_1 &= Y_1 \times X_0 \\ X_2 &= Y_2 \times X_1 = Y_2 \times Y_1 \times X_0 \\ &\dots \\ X_t &= Y_t \times Y_{t-1} \times \dots \times Y_1 \times X_0 \end{aligned}$$

이다. 양편에 로그를 취하면

$$\log X_t = \sum_{i=1}^t \log Y_i + \log X_0$$

이 된다. 이때 $X_0 = 1$ 이므로 $\log X_0 = 0$ 이다. 만약 $\log Y_i$ 들이 서로 독립이고 평균이 μ 이고 분산이 σ^2 인 같은 분포를 따른다면 주어진 시간 t 가 클 때 $\log X_t$ 는 근사적으로 평균이 μt 이고 분산이 $\sigma^2 t$ 인 정규분포를 따르고, $Cov(\log X_s, \log X_t) = \sigma^2 \min(s, t)$ 이 된다. 따라서 $\log Y_i$ 들이 위의 가정들(독립성과 동일한 분포)을 만족한다면 $\log X_t$ 가 브라운이안모션을 따른다고 할 수 있으므로 X_t 는 기하브라운이안 모션을 따른다고 할 수 있다. 그림 3.4는 $\log Y_{t-1}$ 대 $\log Y_t$ 의 그림을 그린 것으로 $\log Y_t$ 가 서로 상관관계가 없는 것으로 보인다. 종합 주가지수의 $\log Y_t$ 를 나타내는 그림 3.5를 보면 $\log Y_t$ 의 평균과 분산이 어느 정도는 변동하는 것처럼 보인다. 실제로 그림 3.6에서 보는 바와 같이 개별주가의 경우 분산 변동의 경향은 더욱 심하게 나타난다. 한편 위에서 잠깐 언급하였듯이 개별주가에 AR(1) 모형을 이용하는 경우에도 오차의 그림을 그려보면 분산이 변동하는 것을 볼 수 있다. 따라서 두 모형 모두에게 있어서 분산의 변동은 영향을 주게 되어있고 그 영향이 어느 모형에게 일방적으로 유리하게 작용하게 되는지는 확실하지 않으며 이에 대한 연구도 필요하리라고 생각한다. 그러나 여기서는 일단 분산의 변동에 대한 고려는 하지 않기로 한다.

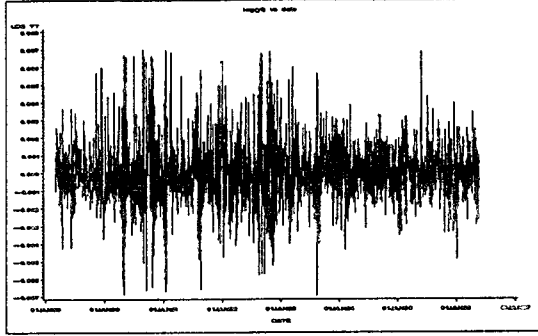


그림 3.5: $\log Y_t$ 대 t 의 그래프
(종합주가지수)

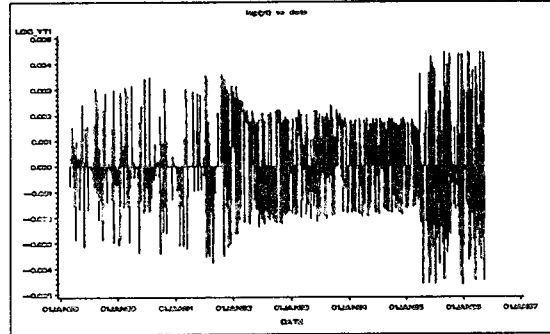


그림 3.6: $\log Y_t$ 대 t 의 그래프
(태광전자주가)

시계열 X_t 가 GBM 모델을 따른다고 할 때, 시간 s 까지의 정보가 주어져 있는 조건 하에서 X_t 의 기대값의 식은 Ross(1993), Karatzas와 Shreve(1991)나 Oksendal(1991) 등에 자세하게 설명되어 있으므로 여기서 그 결과만을 이용하면 다음과 같다.

$$E[X_t | X_s; 0 \leq s \leq t] = X_s e^{(t-s)(\mu + \sigma^2/2)}$$

따라서 X_{t-1} 까지의 정보가 주어져 있을 때 X_t 의 예측값은 다음과 같이 구할 수 있다.

$$\hat{X}_t = X_{t-1} \exp(\hat{\mu} + \hat{\sigma}^2/2) \tag{3.2}$$

이때 μ 와 σ^2 는 $\log Y_t$ 의 기대값과 분산이므로 t 이전의 관측치들의 표본평균과 표본분산으로부터 추정할 수 있다.

표 3.1: AR 모형과 GBM 모형에서의 모수의 추정치

	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\sigma}_u$
Z (종합주가지수)	5.754E-7	0.00189	0.00199	0.997966	0.00182
Z_1 (태광전자)	0.0000844	0.00140	0.000134	0.999962	0.00158
Z_2 (한국이동통신)	0.000158	0.00152	0.000438	0.999786	0.00180
Z_3 (성창기업)	0.0000720	0.00228	-0.000140	1.000088	0.00237
Z_4 (삼성전자)	0.0000318	0.00170	0.000560	0.999488	0.00178
Z_5 (한국전력)	0.0000139	0.00187	0.00140	0.998599	0.00184
Z_6 (포항제철)	0.0000207	0.00164	0.000433	0.999589	0.00164

표 3.1은 종합주가지수와 6개의 개별 주가에 대하여 AR 모형을 적합시켰을 때의 $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\sigma}_u$ 값과 GBM 모형을 적합시켰을 때의 $\hat{\mu}$, $\hat{\sigma}$ 값을 구한 것이다. AR 모형을 적합시켰을 때 모든 경우에서 $\hat{\alpha}_0$ 은 0, $\hat{\alpha}_1$ 은 1에 아주 가깝게 나타났으며, 이는 주가(지수)를 간단하게 기술할 때 자주 이용하는 랜덤워크 모형이 크게 틀리지 않음을 보여준다. 실제로 Dickey와 Fuller(1979)가 제시한 단위근검정을 위의 7개의 시계열에 대하여 시행한 결과 가장 작은 p -값이 0.4885로 랜덤워크 모형을 기각할 수 없다. 또한 모든 주가(지수)의 경우에 $\hat{\mu}$ 이 양수로 나타났으며 이는 분석의 시점인 1989년부터 점차 증가하는 경향이 존재한다는 것을 나타낸다. 그러나 분석의 시점이 바뀔 경우 이는 변화할 수 있다.

4. AR(1) 모형과 GBM 모형의 예측결과 비교

주가 시계열에 대하여 AR(1) 모형과 GBM 모형의 예측능력을 비교하기 위해서 우선 1991.1.3부터 1996.5.22까지의 X_t 값을 두 모형을 이용하여 예측하였다. 실제의 관측값과 예측치의 차이, 즉, 오차로부터 오차의 제곱합(Sum of Squared Errors; SSE), 절대오차의 합(Sum of Absolute Errors; SAE), 절대백분율오차의 합(Sum of absolute Proportional Errors; SAPE)을 계산하였다. 예측에 이용된 관측치의 개수(n)는 두 모형의 예측력의 추이를 잘 보여주는 몇개의 경우, 즉 예측하고자 하는 날 이전의 관측치 30개, 60개, 90개, 300개를 이용하는 경우를 고려하였다. 각각의 경우에 세개의 수치를 계산하고, 이들을 AR(1) 모형을 이용한 경우에는 SSE1, SAE1, SAPE1이라 하고 GBM 모형을 이용한 경우에는 SSE2, SAE2, SAPE2라 하고 그 값들의 비를 표 4.1에 정리하였다. 즉, 각 칸의 수치는 AR(1) 경우의 값을 GBM 경우의 값으로 나누어 준 것이다. 즉, 이 값이 1보다 큰 경우는 GBM 모형이,

표 4.1: AR(1) 모형과 GBM 모형의 예측력 비교

	n	X_t	X_{t1}	X_{t2}	X_{t3}	X_{t4}	X_{t5}	X_{t6}
$\frac{SSE1}{SSE2}$	30	1.1022	1.0661	1.0966	1.0748	1.1069	1.0837	1.0992
	60	1.0561	1.0321	1.0325	1.0198	1.0614	1.0576	1.0624
	90	1.0321	0.9935	1.0148	1.0218	1.0417	1.0358	1.0337
	300	1.0043	0.9939	1.0032	1.0119	1.0035	1.0092	1.0036
$\frac{SAE1}{SAE2}$	30	1.0466	1.0211	1.0160	1.0278	1.0455	1.0415	1.0371
	60	1.0262	1.0089	1.0080	1.0052	1.0258	1.0286	1.0313
	90	1.0236	0.9971	0.9833	0.9990	1.0123	1.0236	1.0105
	300	1.0071	0.9815	0.9811	0.9959	0.9987	1.0079	1.0042
$\frac{SAPE1}{SAPE2}$	30	1.0468	1.0217	1.0158	1.0286	1.0460	1.0419	1.0374
	60	1.0264	1.0089	1.0082	1.0055	1.0259	1.0288	1.0315
	90	1.0238	0.9977	0.9845	0.9994	1.0126	1.0239	1.0109
	300	1.0072	0.9811	0.9818	0.9967	0.9987	1.0084	1.0044

표 4.2: AR(1) 모형과 GBM 모형의 기간별 예측력 비교 (SAPE1/SAPE2)

	n	X_t	X_{t1}	X_{t2}	X_{t3}	X_{t4}	X_{t5}	X_{t6}
기간 I	90	1.040	1.001	1.009	1.023	1.013	1.051	1.029
	150	1.020	0.981	1.001	1.017	1.008	1.035	1.023
	300	1.013	0.961	1.003	1.011	1.002	1.029	1.015
기간 II	15	1.077	1.046	1.019	1.029	1.070	1.058	1.039
	30	1.032	0.995	1.020	1.005	1.026	1.029	1.021
	60	1.016	0.996	0.971	0.982	1.015	1.026	1.013
	90	1.003	0.989	0.932	0.975	1.015	1.013	0.999
	150	1.000+	0.980	0.938	0.978	0.999	1.009	0.994
	300	1.004	0.980	0.946	0.975	0.992	0.997	0.995
기간 III	60	1.017	1.007	1.021	1.015	1.021	1.028	1.017
	90	1.024	1.001	0.994	1.008	1.011	1.020	1.007
	150	1.006	1.005	0.993	1.008	1.016	1.003	0.999
	300	1.005	1.000+	0.996	1.002	0.998	0.996	0.997

1보다 작은 경우에는 AR(1) 모형이 예측에 더 좋다는 것을 의미한다.

표 4.1의 값을 보면 수치가 1에서 크게 벗어난 것은 아니나, 1보다 큰 경우가, 즉, GBM 모형이 AR(1) 모형보다 예측을 더 잘한 경우가 월등히 많은 것을 알 수 있다. 특히 적은 수의 관측치를 이용할수록 GBM모형에서의 수치가 AR 모형에서의 수치보다 작아지는 경향이 확실히 나타난다. 이와같은 경향은 약간의 차이는 있으나 세 개의 수치 모두에서 나타난다.

또한 두 모형에 대한 차이점을 알아내기 위해서 전체 시계열에 대한 것 뿐 아니라 특성을 달리하는 몇 개의 시점에서 부분적으로 예측을 시도하였다. 즉, 전체적으로 완만한 경향을 보이는 1991.1.3부터의 500개(기간 I : 1991.1.3 - 1992.9.15), 완만하던 시계열이 갑자기 급격히 증가하기 시작하는 시기를 포함하는 1993.7.1부터의 500개(기간 II : 1993.7.1 - 1995.3.10)와 전체적으로 증가나 감소가 크게 일어나는 뒷부분의 500개(기간 III : 1994.9.8 - 1996.5.22)의 시계열값을 두 모형으로부터 예측하고 이를 비교하여 표 4.2에 정리하였다. 표 4.1에서는 비교의 기준으로 SSE, SAE, SAPE를 계산하였는데 SAE와 SAPE는 그 경향이 일치하고 SSE보다는 더 안정적이므로 여기서는 SAPE 값의 비율만을 제시하기로 한다.

표 4.2를 보면 일단 전체적으로 하나의 경향이 나타나는데 그것은 이용하는 관측치의 개수, n 이 작을수록 GBM 모형의 예측력이 상대적으로 좋아진다는 것을 볼 수 있다. 그러나 시점에 따라서 그 속도가 빠르기도 하고 늦기도 한 것을 알 수 있다. 즉, 표 4.2에서 보면 완만하게 변하는 기간 I에서나 진동의 폭이 전반적으로 큰 기간 III에서는 GBM 모형이 AR(1) 모형보다 좋아지는 경향이 빠르게 나타난다. 한편 완만하게 진행되던 시계열이 급

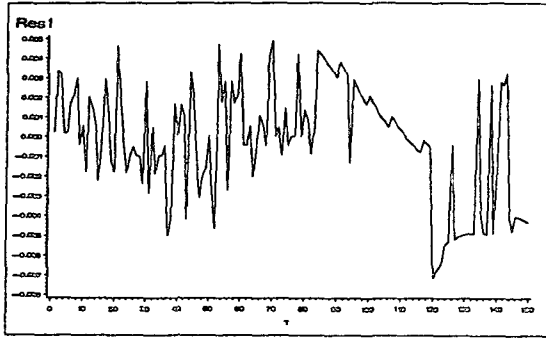


그림 4.1: AR(1), $n=300$ 인 경우의 오차 (기간IV)

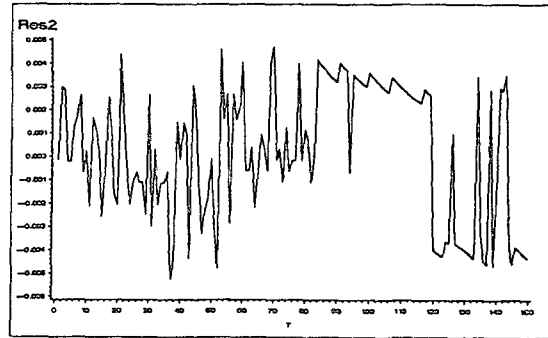


그림 4.2: GBM, $n=300$ 인 경우의 오차 (기간IV)

격히 증가하는 즉, 시계열의 경향이 급격히 변화하는 시점을 포함하는 기간 II에서는 n 이 감소하면서 GBM 모형이 AR(1) 모형에 비해 상대적으로 좋아지는 경향이 완만하게 일어난다. 이는 급격한 변화가 일어나기 시작할 때 상대적으로 AR(1) 모형이 빠르게 대응하는 것으로 판단할 수 있다.

이와 같은 특성을 좀더 자세히 살펴보기 위해서 다른 시계열에 비해 급격한 변화가 조금 더 극적으로 일어나는 성장기업의 주가시계열을 대상으로 급격한 변화의 시점을 포함하는 짧은 기간인 기간 IV(1993.7.1-1993.12.26)에 두 모형에서의 오차가 어떻게 변화하는지를 그려보았다. 이 기간은 완만하게 변화하는 시계열의 값을 이용하여, 급격하게 증가한 후 또 급격히 감소하는 시계열의 값을 예측하여야 하는 기간이다. 그림 4.1과 그림 4.2에서

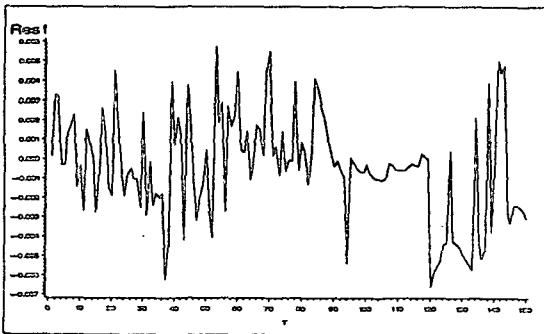


그림 4.3: AR(1), $n=30$ 인 경우의 오차 (기간IV)

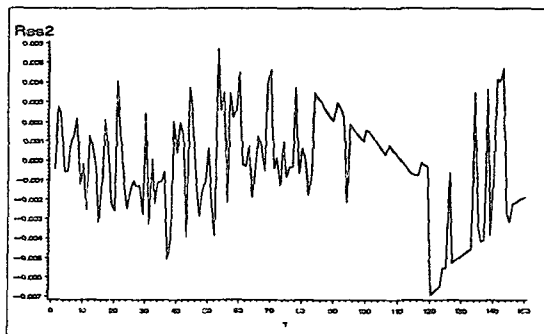


그림 4.4: GBM, $n=30$ 인 경우의 오차 (기간IV)

표 4.3: 등락에 대한 예측 적중력의 비교

n	모형	X_t	X_{t1}	X_{t2}	X_{t3}	X_{t4}	X_{t5}	X_{t6}
300	AR(1)	46.4	42.0	51.6	47.2	47.0	43.4	47.8
	GBM	51.4	43.5	51.6	47.6	46.8	41.6	44.0
150	AR(1)	50.8	44.0	51.4	49.2	47.0	41.2	45.6
	GBM	50.6	43.8	49.8	47.2	46.6	41.0	45.4
90	AR(1)	52.0	42.8	50.4	46.4	45.4	44.2	44.6
	GBM	50.0	41.8	49.8	48.4	46.8	44.2	43.6
60	AR(1)	51.8	43.2	49.6	45.4	45.2	43.6	42.6
	GBM	49.0	42.2	51.6	44.6	47.0	44.6	43.6
30	AR(1)	49.4	42.0	54.0	45.4	47.4	41.6	43.8
	GBM	52.8	43.2	54.6	46.0	47.8	45.4	47.8
15	AR(1)	49.4	41.0	54.0	48.0	41.2	43.6	44.2
	GBM	51.8	44.2	54.6	52.2	48.4	46.0	46.8

t 가 80-120 부분을 비교해보면 관측치를 300개 이용하는 경우에 AR(1) 모형은 비교적 빠르게 증가하는 추세를 예측에 반영하여 예측치도 빠르게 증가시켜서 양수의 오차를 빠르게 0에 가깝게 만드는 반면에, GBM 모형의 경우는 증가 추세를 아주 완만하게 반영하느라 오랫동안 관측치보다 작은 예측치를 주게되어 큰 양수의 오차를 갖게되는 것을 볼 수 있다. 한편 그림 4.3과 4.4에서 위와 같은 기간을 비교하여 보면 관측치를 30개만 이용하는 경우에 AR(1) 모형의 경우에는 너무 급격히 증가의 추세를 반영하느라 곧이은 하락하는 시점에서 실제 관측치보다 훨씬 크게 예측하게되어 절대치가 큰 음수의 오차를 갖게되고 ($t=94$), 그러나 GBM의 경우에는 그림 4.2의 경우보다는 빠르게 증가의 추세를 반영하여 양수의 오차를 빠르게 0에 가깝게 한 반면 AR(1) 모형에 비해 상대적으로 완만하게 대응하여 하락의 시점($t=94$)에서도 절대치가 큰 음수의 오차를 갖지 않는 것을 볼 수 있다.

이와 같은 경향은 두 모형으로부터 설명이 가능하리라고 본다. AR(1) 모형의 경우에 X_t 의 값은 식 3.1을 이용하여 X_{t-1} 에 대한 회귀식으로부터 구한다고 하였다. 회귀모수의 추정치는 기존의 (X_{t-1}, X_t)로부터 구하게 되는데, X 값들이 급격히 증가하기 시작하는 경우에 새로 들어오는 (X_{t-1}, X_t)값은 기존의 관측치들과 멀리 떨어져 있게되고, 따라서 영향력이 큰 관측치가 된다. 즉, 기존 n 개의 자료보다 큰 비중으로 회귀모수에 영향을 주게되어 증가의 추세가 회귀모수에 빠르게 반영된다. 한편 GBM 모형의 경우에 X_t 값을 예측하는데 이용되는 식 3.2의 $\hat{\mu}$ 은 이용하는 n 개의 관측치 값을 고르게 반영한다. 즉, 관측치들이 급격히 증가하여 새로 들어오는 값들이 큰값을 갖는다 하더라도 그 관측치에 대한 비중은 $1/n$ 로 n 이 클 경우에는 증가하는 추세가 예측치에 잘 반영되지 않게 된다.

이와 같이 두 모형 모두에게 있어서 n 이 작을수록 실제값의 변화에 민감하게 반응하게

되는데 상대적으로 AR(1) 모형이 GBM 모형보다 변화에 더 민감하게 반응하는 것을 볼 수 있다. 일반적으로 시계열의 경향을 나타내는 AR(1) 모형의 α_0, α_1 의 값이나 GBM 모형의 μ 값이 변화하는 경우에 n 이 큰 경우에는 상대적으로 GBM 모형이 변화에 더디게 반응함으로써 예측력이 떨어지게 되고, n 이 작아지면서 AR(1)이 너무 민감하게 변화에 반응하는 반면 상대적으로 GBM 모형은 완만하게 반응함으로써 예측력이 나아지는 것이다. 이와같은 경향은 변화의 폭이 급격할수록 n 이 더 작아야만 GBM 모형이 상대적으로 나아지게 된다.

또한 여기서 한가지 고려한 것은 주가의 등락에 대한 두 모형의 예측력의 비교이다. 표 4.3은 기간 II에서 등락을 맞게 예측한 비율(%)을 이용한 관측치의 개수와 모형에 따라 정리한 것이다.

표 4.3을 보면 우선 두 모형 모두에 있어서 등락의 예측력이 별로 좋지 않음을 알 수 있다. 즉, 훨씬 많은 경우에 적중률이 50%에 못미치는 것을 알 수 있다. 특히 개별 주가의 경우에는 등락의 적중률이 더 떨어지는 것을 볼 수 있다. 또한 표 4.3에서도 앞의 표 4.1, 4.2에서와 같이 n 이 작을 때 상대적으로 GBM 모형이 AR(1) 모형에 비해 나은 적중률을 보인다는 것이다. 그러나 이 경우는 앞서와는 달리 n 이 감소하면서 GBM 모형이 일관되게 나아지는 것이 아니라 결과가 오락가락하는 경향을 보이다가 n 이 30보다 작은 경우에는 비교적 확실하게 GBM 모형이 나아지는 것이다. 또한 n 이 커짐에 따라 적중률이 증가하는 경향이 나타나지 않는다는 것이다. 일반적으로 SSE, SAE, SAPE를 계산하는 경우 두 모형 모두에 있어서 n 이 클수록 그값이 작아지는 경향이 나타난다. 이는 많은 관측치를 이용하면 좀더 정밀한 예측이 가능하다는 것을 의미한다. 그러나 등락만을 예측하고자 하는 경우에는 관측치의 개수가 크게 영향을 미치지 않는다는 것을 알 수 있다.

5. 시뮬레이션

앞 절에서 고려한 7개의 주가시계열의 경우 그 값을 예측할 때 n 이 작을수록 GBM 모형이 AR(1) 모형보다 상대적으로 더 좋다는 것을 알았다. 이와 같은 시계열의 경향, 즉 AR(1)의 α_0, α_1 이나 GBM 모형의 μ 의 값이 변하는 경우에 n 이 예측에 어떻게 영향을 미치는가를 살펴보았다. 이절에서는 시뮬레이션을 통해서 몇 개의 AR(1)을 따르는 시계열의 경우와 GBM을 따르는 시계열의 경우에 두 모형을 이용한 예측력이 어떠한 경향을 보이는지 살펴보고자 한다.

우선 몇개의 μ 와 σ 에 대하여 GBM 모형을 따르는 길이가 1000인 시계열을 생성하였다. 즉, 평균이 μ 이고 분산이 σ^2 인 정규확률변수 U_1, U_2, \dots 를 생성한 후, X_k 을 다음과 같이 계산하였다.

$$X_k = \exp[U_1 + \dots + U_k] \quad k = 1, \dots, 1000$$

따라서 X_k 는 GBM모형을 따르는 시계열의 이산시점 k 에서의 관측치라 할 수 있다. 이렇게 생성된 시계열에서 뒷부분의 500개 자료값에 대하여 AR(1)과 GBM을 이용하여 값을 예측한 후, 실제 생성된 자료값 500개와 각각 대응된 예측값 500개의 차이, 즉, 예측오차를 계산하였다. 이들로부터 AR(1) 모형을 이용하여 SAPE1을, GBM 모형을 이용하여 SAPE2를 계산하여 비교하였다. 이를 500번 반복하여 그 중 GBM모형이 나은 예측을 한 경우, 즉, 더

표 5.1: GBM 시계열에 대한 AR(1)과 GBM 모형의 비교

$\mu \backslash n$	SAPE1/SAPE2 > 1의 비율					
	$\sigma = 0.001$			$\sigma = 0.005$		
	60	150	300	60	150	300
-0.01	99.8	100	100	99.8	100	100
-0.001	99.6	96.6	88.8	99.4	96.6	90.6
-0.0001	97.0	93.4	86.8	93.4	86.4	74.4
-0.00001	93.8	86.4	74.2	94.6	85.4	72.4
0	94.6	85.4	72.6	95.0	85.4	72.6
0.00001	95.0	85.4	74.2	94.8	86.2	73.2
0.0001	97.6	93.8	87.0	94.6	83.6	75.4
0.001	99.2	95.8	88.6	99.8	97.8	91.8
0.01	99.0	92.6	89.4	99.0	92.4	89.4

$\mu \backslash n$	$\sigma = 0.001$						$\sigma = 0.005$					
	P1			P2			P1			P2		
	60	150	300	60	150	300	60	150	300	60	150	300
-0.01	100	100	100	100	100	100	97.7	97.7	97.7	97.7	97.7	97.7
-0.001	84.1	84.1	84.1	84.1	84.1	84.1	52.4	55.1	56.9	56.9	57.8	57.9
-0.0001	50.3	50.7	51.4	52.3	53.1	53.7	50.0	50.0	49.9	50.1	50.1	50.1
-0.00001	50.0	49.8	49.9	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	49.9
0	50.0	49.9	49.9	50.0	50.0	49.9	50.0	49.8	49.9	50.0	50.1	49.8
0.00001	50.0	49.9	50.0	50.1	50.1	50.0	50.0	49.8	50.0	50.0	50.0	49.9
0.0001	52.2	50.7	51.5	52.4	53.2	53.8	50.0	49.9	50.1	50.2	50.2	50.1
0.001	84.2	84.2	84.2	84.2	84.2	84.2	52.5	55.3	57.4	57.1	57.9	57.9
0.01	100	100	100	100	100	100	97.7	97.7	97.7	97.7	97.7	97.7

작은 SAPE를 갖는 경우를 세어 %로 계산하고, 또한 두 모형의 등락에 대한 적중률(AR(1)의 경우를 P1, GBM의 경우를 P2라 하자.)을 정리한 것이 표 5.1이다. 예측에 이용된 관측치의 개수가 60개, 150개, 300개인 경우를 고려하였는데 이 경우들로부터 예측력의 변화추이를 어느정도 알아볼 수 있었기 때문이다. 고려된 μ 와 σ 의 값은 위의 주가(지수)시계열자료가 GBM 모형을 따른다고 했을 때 갖게되리라고 추정하는 값에 가까운 몇 개의 값으로 하였다.

표 5.1에서 우선 SAPE1/SAPE2가 1보다 큰 비율을 보면 여기서 고려한 모든 경우에 50%보다 크다는 것을 알 수 있다. 즉, GBM모형이 AR(1)모형보다 예측력이 더 좋다는 것

표 5.2: AR(1) 시계열에 대한 AR(1)과 GBM 모형의 비교

$\alpha_0 \backslash n$	SAPE1/SAPE2 > 1의 비율					
	$\sigma_u = 0.001$			$\sigma_u = 0.002$		
	60	150	300	60	150	300
0.0001	97.6	94.0	86.4	93.0	87.4	81.0
0.0005	99.8	95.0	83.0	99.4	97.2	86.2
0.0010	99.2	91.8	49.8	99.6	93.8	76.8
0.0015	99.2	80.0	4.4	99.2	91.6	50.6
0.0020	99.2	50.6	0.2	99.2	87.2	19.0

$\alpha_0 \backslash n$	$\sigma_u = 0.001$						$\sigma_u = 0.002$					
	P1			P2			P1			P2		
	60	150	300	60	150	300	60	150	300	60	150	300
0.0001	52.4	50.7	51.4	52.4	53.2	53.8	50.1	50.0	50.1	50.7	50.9	51.2
0.0005	69.2	69.1	69.2	69.2	69.2	69.2	54.6	58.3	59.7	59.4	59.9	59.9
0.0010	84.2	84.2	84.2	84.2	84.2	84.2	68.1	69.1	69.2	69.2	69.2	69.2
0.0015	93.3	93.3	93.3	93.3	93.3	93.3	77.3	77.3	77.3	77.3	77.3	77.3
0.0020	97.7	97.7	97.7	97.7	97.7	97.7	84.2	84.2	84.2	84.2	84.2	84.2

을 알 수 있다. 또한 n 이 작을수록 GBM 모형이 AR(1) 모형보다 예측하는데 적합한 것을 알 수 있다. 여기에 제시하지는 않았으나 n 이 30, 15인 경우에는 더욱 더 확실하게 GBM 모형이 예측력이 뛰어나다는 것을 볼 수 있다. 이와 같은 경향은 두 개의 σ 값에 있어서 일치한다. 여기서 한가지 눈여겨 볼만한 것은 고려한 구간에서 μ 가 0으로부터 멀어질 때 GBM 모형이 더 나은 예측력을 갖게되는 경향이 어느정도 나타난다는 것이다. (μ 가 0.01이상인 경우에는 이 경향이 약간 불안정해지는 것을 볼 수 있었다.) 한편 등락의 적중률을 보면 μ 의 절대값이 커질수록 적중률이 높아지는 경향이 뚜렷하게 나타나는 것을 볼 수 있다. 이는 증가나 하락의 경향이 뚜렷할 때 증가하거나 하락하리라 예측하기는 쉬운 것이기 때문이다. 이와같은 경향은 모든 n 의 경우와 σ 에서 일치한다. 한편 n 의 값에 따른 적중률의 변화는 거의 없으며, 앞절에서 나타난 n 이 작을 때 상대적으로 GBM의 적중률이 높아지는 경향은 어느 정도는 감지되나 뚜렷하지는 않다.

GBM 모형에서와 마찬가지로 AR(1) 모형을 이용하여 $\alpha_1 = 1$ 이고 α_0 와 오차의 분산 σ_u^2 의 값을 변화시켜가면서 각각에 대하여 길이가 1000인 시계열을 생성한 후, 두 모형에 대하여 예측력을 비교하였다. 앞에서 고려한 주가시계열이 AR(1) 모형을 따른다고 할 때 α_1 이 1에 가까웠으므로 α_1 은 1로 고정시키고, α_0 와 σ_u 는 주가시계열이 AR(1) 모형을 따른다고 할 때 갖게되리라 예상되는 몇 개의 값으로 정하였다. 이 때에도 n 은 60개, 150개, 300개인 경우를 고려하였고, 표 5.2의 각란의 수치는 표 5.1에서와 같은 방법으로 얻어졌다.

이 표에서 우선 SAPE1/SAPE2가 1보다 큰 비율을 보면 위의 표 5.1에서와 같이 n 이 작을수록 GBM 모형이 예측력에서 더 나은 결과를 주는 것을 알 수 있다. 또한 전반적으로 볼 때 절편의 값(α_0)이 커갈수록 AR(1) 모형이 더 나은 결과를 준다고 할 수 있다. 이와 같은 경향은 α_0 값이 0.0001로 아주 작을 때는 조금 불안정하게 나타나지만 여기서 고려한 것보다 조금 더 큰 경우에는 아주 급격히 나타난다. 이런 경향을 두 모형으로부터 설명해 보면 다음과 같다. 즉,

$$GBM : E[X_{k+1}|X_i; 0 \leq i \leq k] = X_k e^{(\mu + \sigma^2/2)}$$

$$AR(1) : E[X_{k+1}|X_k] = \alpha_0 + \alpha_1 X_k$$

에서 AR(1) 모형을 따르는 시계열에서 $\alpha_1 = 1$ 이고 α_0 가 큰 경우 X_t 는 t 에 대해서 선형으로 증가하는 형태를 취한다. 이 증가의 경향을 GBM 모형으로 설명할 경우 μ 가 0보다 큰 값으로 추정되어 X_t 값이 비선형으로 증가하는 것으로 예측하게 되는 것이다. 따라서 α_0 가 클수록 나쁜 예측력을 보이는 것은 당연한 결과라 하겠다. 한편 오차항의 분산(σ_u^2)이 큰 경우에는 상대적으로 α_0 값이 더 클 때 AR(1) 모형이 더 나은 것으로 나타나는데, 이는 α_0 의 영향이 σ_u 에 대한 상대적인 크기에 의해서 결정되는 것을 의미한다. 한편 등락에 대한 예측은 α_0 값이 클수록 그 적중률이 높아지는 것을 볼 수 있다. 이는 GBM 모형에서와 똑같은 논리로 설명할 수 있다. 즉, α_0 가 클수록 증가의 경향이 확실하게 되고 따라서 증가하리라고 예측하는 것을 쉬운 일이기 때문이다. 이 경우에도 n 에 따른 적중률의 변화는 거의 없으며 n 에 따라 GBM 모형이 상대적으로 좋아지는 경향도 어느 정도는 나타나지만 뚜렷하지는 않다.

이 절의 결과와 제 4절의 결과를 비교해보면 우선 공통적으로 나타나는 특징은 n 이 작을수록 GBM 모형이 상대적으로 AR(1) 모형보다 예측에 적합하다는 것이다. 제 4절에서는 μ 나 α_0 , α_1 이 급격히 변화하는 경우에 대하여 설명하였는데, 이 절의 결과는 그 값들이 고정되어 있는 경우에도 n 에 대한 특성이 유지된다는 것을 보여준다. 이에 대한 설명은 또 다른 연구의 과제가 될 수 있을 것으로 생각한다.

참고문헌

- [1] 김진경 (1995). <환율 시계열에 적용된 조건부 비등분산 자동회귀 모형에 관한 재검토>. 인하대학교 기초과학연구소 논문집, 16, 89-104.
- [2] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root, *Journal of the American Statistical Association*, 74, 427-431.
- [3] Hull, J. C. (1993). *Options, Futures, and other Derivative Securities*, 2ed. Prentice Hall, New Jersey.
- [4] Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

- [5] Oksendal, B. (1991). *Stochastic Differential Equations*, Springer, New York.
- [6] Roll, R. (1984). Orange Juice and Weather, *The American Economic Review*, 74, 861-880.
- [7] Ross, S. M. (1993). *Introduction to Probability Models*, Academic Press.
- [8] Shapiro, A. C. (1991). *Modern Corporate Finance*, Macmillan, New York.
- [9] Wei, W. W. S. (1994). *Time Series Analysis*, Addison-Wesley.

[1997년 11월 접수, 1998년 6월 최종수정]

The Analysis of the Stock Price Time Series using the Geometric Brownian Motion Model*

Kim, Jeankyung¹⁾

ABSTRACT

In this study, I employed the autoregressive model and the geometric Brownian motion model to analyze the recent stock prices of Korea. For all 7 series of stock prices(or index) the geometric Brownian motion model gives better predicted values compared with the autoregressive model when we use smaller number of observations.

* This research was supported by the Inha University Research Grant 1996-1997.

1) Dept. of Statistics, Inha University, 253 Yonghyun-dong Nam-gu Incheon, 402-751, KOREA.