Distributions of stock price changes via normal hidden Markov models

Stanisław Galus

8 November 2010

Abstract

A method of modelling distributions of rates of return of stocks as stationary distributions of normal hidden Markov models is presented. The method is applied to fifty stocks quoted on the Stock Exchange in Warsaw and results are discussed.

1 Introduction

In the present paper an attempt is made to model distributions of rates of return of stocks quoted on the Stock Exchange in Warsaw by the means of normal hidden Markov models. A contemporary outline of modelling return distributions can be found in [5, chapter 7].

The next sections present methods and the set of data used in the paper followed by discussion of results.

2 Methods

Let $(X_t)_{t=0}^{\infty}$ be a homogeneous Markov chain over the state space $S = \{1, \dots, s\}$ with transition matrix $P = [p_{ij}]_{i,j \in S}$ and initial state distribution $\pi = [\pi_i]_{i \in S}$. Then, for each $i_0, \dots, i_T \in S$,

$$\Pr(X_0 = i_0, X_1 = i_1, \dots, X_T = i_T) = \pi_{i_0} p_{i_0 i_1} \dots p_{i_{T-1} i_T}.$$

For each state $i \in S$, let $f_i(y, \theta_i)$ be a corresponding probability density function. In each moment $t = 0, \dots, T$, a value y_t of a random variable Y_t is observed which comes from the density f_{i_t}. The likelihood of the sample $y_0, \dots, y_T$ is

$$\begin{aligned} L(p, P, \theta_1, \dots, \theta_s) &= \qquad (1) \\ &= \sum_{i_0, \dots, i_T = 1}^{s} \pi_{i_0} f_{i_0}(y_0, \theta_{i_0}) p_{i_0 i_1} f_{i_1}(y_1, \theta_{i_1}) \dots p_{i_{T-1} i_T} f_{i_T}(y_T, \theta_{i_T}) = \\ &= \sum_{i_0=1}^{s} \pi_{i_0} f_{i_0}(y_0, \theta_{i_0}) \sum_{i_1=1}^{s} p_{i_0 i_1} f_{i_1}(y_1, \theta_{i_1}) \dots \sum_{i_T=1}^{s} p_{i_{T-1} i_T} f_{i_T}(y_T, \theta_{i_T}). \end{aligned}$$

Throughout this paper, normal distributions of Y_t are considered, i. e. $f_i(y, \theta_i) = \phi((y - \mu_i)/\sigma_i)$, where ϕ denotes the standard normal probability density. In this case, the likelihood (1) can be maximized by the Baum-Welch estimation method [1].

If π, P, θ_1, ..., θ_s are known, the best sequence of states can be found, i. e. a sequence

$$i_0, \ldots, i_T \in S \tag{2}$$

which maximizes

$$\pi_{i_0} f_{i_0}(y_0, \theta_{i_0}) p_{i_0 i_1} f_{i_1}(y_1, \theta_{i_1}) \ldots p_{i_{T-1} i_T} f_{i_T}(y_T, \theta_{i_T}), \tag{3}$$

by the Viterbi algorithm [6, p. 264]. For the sequence (2), the χ^2 goodness-of-fit test [3, p. 453] can be used to establish whether or not the distribution of

$$\frac{y_0 - \mu_{i_0}}{\sigma_{i_0}}, \ldots, \frac{y_T - \mu_{i_T}}{\sigma_{i_T}} \tag{4}$$

differs from the standard normal distribution. If there is no reason to reject normality of (4) and sufficient conditions for the unique existence of stationary distribution [2, p. 130] of π are met, then

$$f(x) = \pi_1 f_1(x, \theta_1) + \ldots + \pi_s f_s(x, \theta_s) \tag{5}$$

can be considered to be a limiting distribution of y_t. Again, the χ^2 test can be used to establish whether distributions of y_t and (5) differ. In the sequel, the values of χ^2 statistic for these two tests are denoted by χ^2_1 and χ^2_2, respectively.

The set of data consists of fifty series of rates of return of stocks quoted on the Stock Exchange in Warsaw till May 2010. The selection was made to select the stocks with the greatest possible number of quotes. The number of missing quotes did not exceed 30. The missing observations of prices were linearly interpolated, then rates of return were calculated. To avoid degenerated normal distributions in Baum-Welch estimation, for each series, all rates of return equal to zero were replaced by independent random numbers distributed normally with mean zero and standard deviation equal to one third of the smallest positive absolute value found in the series. The numbers of observations vary from 3062 to 4334.

For each series and each number of states s between 2 and 16, ten attempts were made to estimate π, P, $\theta_1, \ldots, \theta_s$ and that maximizing likelihood was taken as the result. For the estimated parameters both χ^2 tests mentioned above were carried.

3 Results

The results of estimation and testing are presented in table 1. It can be seen that the closer s to 16, the greater the number of stocks for which (4) seems to be normally distributed and (5) seems to be the same as distribution of y_t. Five of 750 estimation processes failed, all for one stock BANKBPH. For two stocks, ADVADIS and INGBSK, no satisfactory limiting distributions were found. It can be seen from the table that whenever (4) is normally distributed, the limiting distribution may be accepted as the distribution of y_t.

A typical histogram of rates of return and the corresponding estimated distribution (5) are depicted on figures 1 and 2, respectively. The histogram for ADVADIS is presented on figure 3.

4 Discussion and conclusion

A method of obtaining distributions of rates of return was presented. The method yields distributions which are apparently not rejected by the χ^2 goodness-of-fit test. However, several difficulties arise. At the moment, the true number of states should be identified, as for the majority of series more than one number of states give satisfactory results.

References

[1] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

[2] Patrick Billingsley. *Prawdopodobieństwo i miara*. Wydawnictwo Naukowe PWN, Warszawa, 2009.

[3] Marek Fisz. *Rachunek prawdopodobieństwa i statystyka matematyczna*. Państwowe Wydawnictwo Naukowe, Warszawa, 1969.

[4] David Freedman and Persi Diaconis. On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):453–476, 1981.

[5] T. C. Mills. *The econometric modelling of financial time series*. Cambridge University Press, Cambridge, 2008.

[6] Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Table 1: The results of estimation and testing.

Stock	Number of states														
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
04PRO	○	○	○	○	○	○	○	○	●	○	●	◑	●	◑	●
06MAGNA	○	○	○	○	○	◑	●	○	○	○	●	●	●	◑	●
08OCTAVA	○	○	◑	○	◑	○	◑	◑	◑	●	◑	◑	◑	◑	◑
ADVADIS	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
ALMA	○	○	○	○	○	○	○	○	○	○	●	○	○	○	●
AMICA	○	○	○	●	○	○	○	○	○	○	○	●	○	●	◑
APATOR	○	◑	○	◑	◑	◑	◑	◑	●	◑	◑	◑	●	◑	●
ATLANTIS	○	○	○	○	●	○	◑	●	○	●	●	◑	◑	◑	●
BANKBPH		○	○			○			○	●	◑	◑	○	◑	◑
BBICAPNFI	○	○	○	○	○	○	○	◑	◑	○	◑	●	◑	◑	●
BBIDEVNFI	○	○	○	○	○	○	○	○	○	◑	○	●	○	◑	○
BBIZENNFI	○	○	○	○	○	○	◑	◑	◑	●	●	◑	◑	●	●
BORYSZEW	○	○	○	○	○	○	◑	●	◑	●	◑	◑	●	◑	●
BRE	○	○	○	○	○	○	○	○	◑	●	●	○	●	●	●
BUDIMEX	○	○	○	○	○	○	○	○	●	○	●	●	◑	◑	◑
DEBICA	○	○	○	○	○	○	○	◑	◑	○	◑	◑	●	◑	●
ECHO	○	○	○	○	○	○	◑	◑	●	●	●	◑	●	●	●
FORTE	○	○	○	◑	◑	◑	◑	●	●	◑	◑	●	◑	●	◑
HANDLOWY	○	○	○	○	○	○	◑	◑	●	○	◑	●	◑	◑	◑
IGROUP	○	○	○	○	○	○	◑	◑	◑	◑	◑	◑	◑	◑	◑
IMPEXMET	○	○	○	○	○	○	◑	●	◑	◑	●	◑	◑	◑	◑
INGBSK	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
JUPITER	○	○	○	○	●	◑	○	○	○	●	◑	◑	◑	◑	●
JUTRZENKA	○	○	○	○	○	○	○	○	◑	◑	○	●	◑	○	◑
KABLE	○	○	○	○	○	◑	○	○	●	●	◑	◑	◑	○	◑
KETY	○	○	○	○	○	○	○	○	○	◑	○	◑	◑	●	◑
KGHM	○	○	○	○	○	○	●	○	◑	○	○	●	◑	●	●
KREDYTB	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○
KREZUS	○	○	○	○	○	○	○	○	●	○	●	◑	◑	◑	●
KRUSZWICA	○	○	○	○	○	○	○	○	●	●	●	●	●	●	◑
LENTEX	○	○	○	○	○	○	○	○	○	○	●	●	○	◑	◑
MIDAS	○	○	○	○	○	○	○	●	○	○	●	○	◑	○	●
MIESZKO	○	○	○	○	○	◑	○	○	○	◑	○	●	◑	◑	◑
MILLENNIUM	○	○	○	○	○	◑	◑	○	●	●	◑	◑	●	◑	●
MOSTALEXP	○	○	○	○	○	○	○	○	○	○	◑	●	◑	◑	●
MOSTALZAB	○	○	○	○	○	◑	◑	●	◑	◑	◑	●	●	◑	◑
NFIEMF	○	○	◑	◑	○	○	◑	◑	●	◑	●	◑	●	●	◑
ODLEWNIE	○	○	○	○	○	●	◑	◑	●	◑	◑	◑	◑	◑	◑
ORBIS	○	○	○	○	○	○	○	◑	○	○	○	●	◑	◑	◑
PGF	○	●	○	○	○	○	○	◑	○	○	◑	○	○	●	○
PROCHNIK	○	○	○	○	○	○	○	◑	◑	◑	◑	◑	○	◑	◑
RAFAKO	○	○	○	○	○	○	○	○	○	○	○	○	●	○	◑
RELPOL	○	○	○	○	○	○	◑	●	◑	●	◑	◑	◑	●	◑
RUBICON	○	○	○	○	○	○	○	◑	●	●	◑	◑	○	○	●
STALEXP	○	○	○	○	○	○	◑	◑	◑	◑	◑	◑	◑	◑	◑
SWARZEDZ	○	○	○	○	○	◑	○	◑	○	●	◑	◑	●	◑	○
SWIECIE	○	○	○	○	○	○	○	○	●	◑	◑	●	●	◑	◑
SYGNITY	○	○	○	○	○	○	○	○	○	●	◑	○	○	◑	○
VISTULA	○	○	○	○	○	○	◑	○	◑	◑	◑	◑	◑	◑	●
WILBO	○	○	○	○	○	○	○	◑	○	○	●	○	◑	◑	●

○ $\Pr(\chi^2 \geq \chi^2_1) < 0.1$ and $\Pr(\chi^2 \geq \chi^2_2) < 0.1$.

◑ $\Pr(\chi^2 \geq \chi^2_1) < 0.1$ but $\Pr(\chi^2 \geq \chi^2_2) \geq 0.1$.

● $\Pr(\chi^2 \geq \chi^2_1) \geq 0.1$ and $\Pr(\chi^2 \geq \chi^2_2) \geq 0.1$.

Empty cells denote unsuccessful estimation attempts.

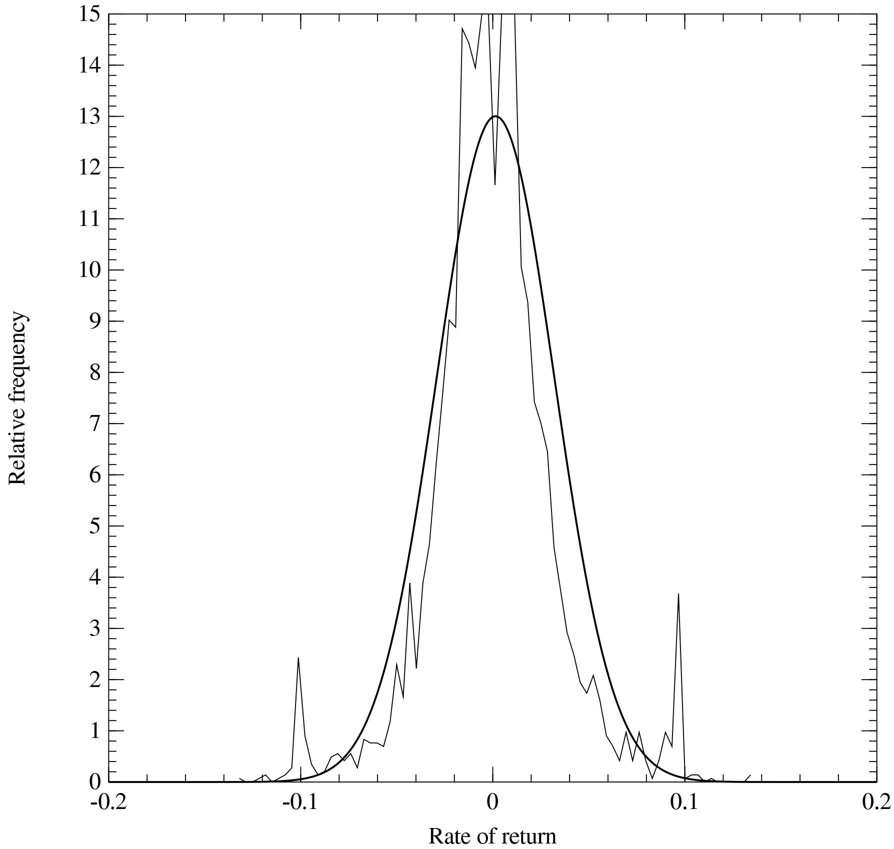

Figure 1: Histogram of rates of return of BRE based on 4222 observations and 79 cells. The cell width has been chosen as twice the interquantile range of the data divided by the cube root of the sample size [4, p. 455]. Segments of the broken line which lie above the ordinate 15 are not shown as the maximum ordinate exceeds 40. The normal curve fitted by the maximum likelihood method is drawn with a thicker line.

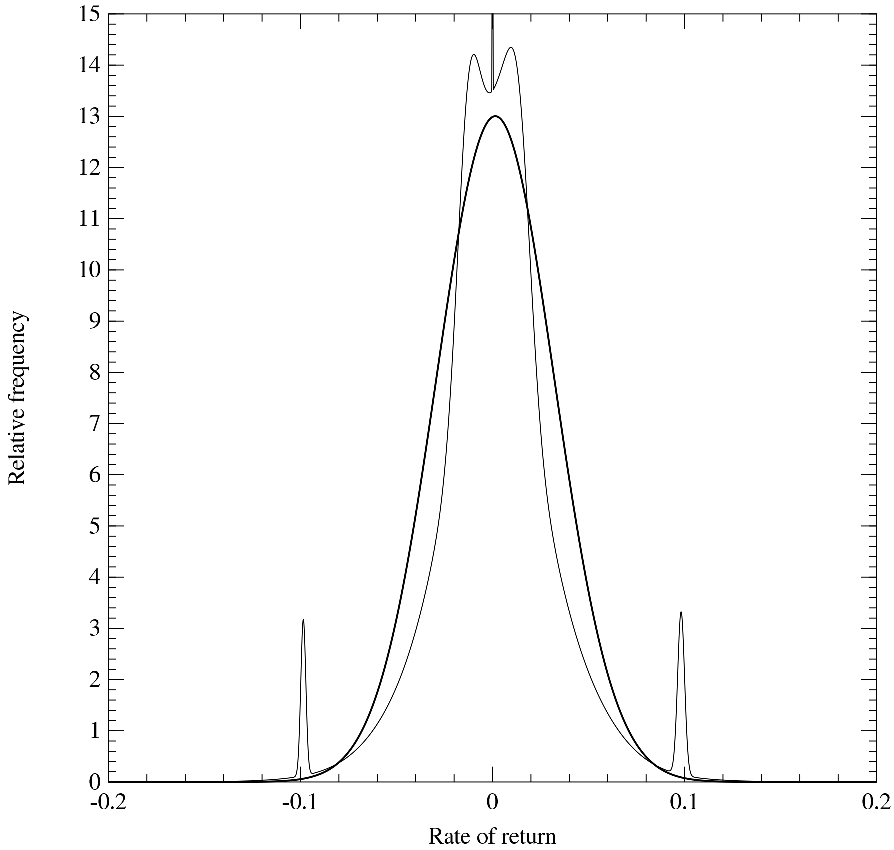

Figure 2: Limiting distribution (5) with 11 states fitted for BRE.

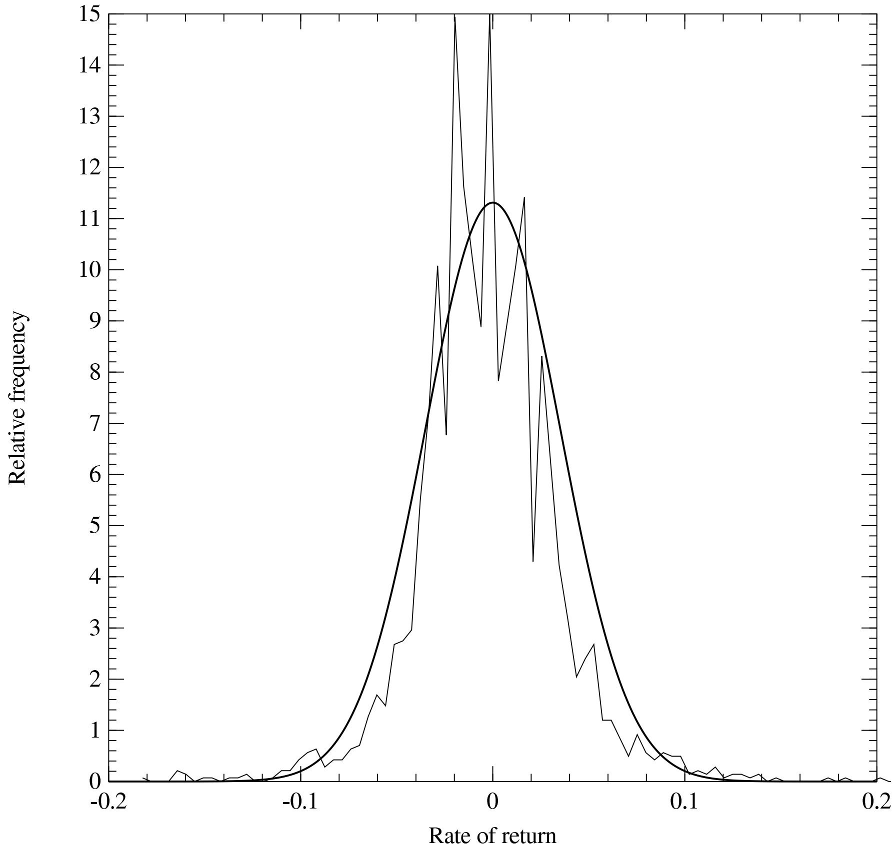

Figure 3: Histogram of rates of return of ADVADIS based on 3140 observations and 151 cells. Cf. notes under figure 1.