AUSTRIAN JOURNAL OF STATISTICS
VOLUME 31 (2002), NUMBER 2&3, 211-220

Czech Stock Market Analysis

Jiří Trešl and Dagmar Blatná
University of Economics, Prague

Abstract: The analysis of relative returns of selected stocks at Prague Stock Exchange has been performed. As a rule, the kurtosis of the return distribution was greater than that of the standard normal distribution. Second, Box-Jenkins ARIMA models have been employed for return time series modelling. Instead of autoregressive terms of first order, higher terms are also necessary (particularly tenth order). Third, ARCH and GARCH models allowing for conditional heteroskedasticity provide closer approximation of return time series. In most cases, GARCH (1,1) is quite satisfactory.

Keywords: Stock Returns, Time Series Modelling, ARIMA Models, ARCH and GARCH Models.

1 Elementary Characteristics of Logarithmic Returns

Modelling of financial time series is very rapidly developing area at the boundary between statistics and econometrics. Plenty of new models and methodological approaches have appeared, mainly during the last ten years. The scope of this paper is to discuss the possibility of using **ARIMA** (**A**uto**R**egressive **I**ntegrated **M**oving **A**verage) and **GARCH** (**G**eneralized **A**uto**R**egressive **C**onditional **H**eteroskedastic) models for stock returns. Four types of input data are the following:

- 1004 closing daily prices of the following stocks (1996-1999)

CEZ	= CEZ	NHUT	= NOVA HUT
KB	= KOMERCNI BANKA	PRAZ	= PLZENSKY PRAZDROJ
CS	= CESKA SPORITELNA	SOUH	= SOKOLOVSKA UHELNA
IPB	= INVEST. A POST.BANKA	TEL	= CESKY TELECOM
SPLZ	= SKODA PLZEN		

- 1004 daily values of PX-50 index during 1996-1999
- 18575 daily values of Dow Jones Industrial Average (DJIA) during 1930-1999
- 18575 daily values of Dow Jones Agriculture and Transport (DJAT) during 1930-1999

As for the analysis, there are three possibilities in principle. First, we can introduce relative returns during k trading days

$$y_t(k) = \frac{x_t - x_{t-k}}{x_{t-k}} = \frac{x_t}{x_{t-k}} - 1 \qquad (1)$$

where x_t denotes daily prices. Second, chain price indexes

$$I_t(k) = \frac{x_t}{x_{t-k}} = 1 + y_t(k) \quad (2)$$

can be used. However, the most common measure are logarithmic returns

$$z_t(k) = ln\, I_t(k) = ln\, \frac{x_t}{x_{t-k}} = ln\, x_t - ln\, x_{t-k} = ln\,[1 + y_t(k)] \quad (3)$$

and we shall employ them in further analysis and call them simply returns.

In Table 1 and Table 2, results of skewness and kurtosis computation of returns distribution are given. As a rule, curtosis values are always greater than normal distribution ones; this seems to be a typical feature observed in financial time series (together with fat tails).

Table 1: Returns distribution: skewness

BCPP	1996	1997	1998	1999	1996-1997	1996-1998	1996-1999
CEZ	+0.05	+0.17	+0.12	+0.58	+0.19	+0.08	+0.22
KB	-0.91	+0.04	-1.04	-0.14	-0.27	-1.29	-0.88
CS	-0.33	-0.24	-1.58	-0.44	-0.24	-2.13	-1.53
IPB	+0.22	+0.12	-0.34	+0.52	+0.04	-0.32	-0.11
SPLZ	-0.45	-0.17	+0.23	+0.11	-0.33	-0.04	+0.03
NHUT	-0.24	-0.50	-0.17	-0.21	-0.32	-0.36	-0.30
PRAZ	-0.41	-0.25	-0.09	+0.27	-0.39	-0.28	+0.00
SOUH	-0.30	-0.55	+0.12	+0.44	-0.30	-0.14	+0.15
TEL	+0.03	-0.10	-0.22	+0.12	-0.09	-0.16	-0.07
PX 50	-0.79	-0.31	-0.41	+0.07	-0.52	-0.57	-0.33

DJ	30	40	50	60	70	80	90	30-40	30-50	30-60	30-70	30-80	30-90
DJ 1	+0.42	-1.29	-0.93	-0.05	+0.23	-4.62	-0.41	+0.36	+0.31	+0.31	+0.30	-0.54	-0.53
DJ 2	+0.26	-0.69	-0.41	+0.00	+0.17	-1.93	+0.10	+0.13	+0.09	+0.09	+0.09	-0.12	-0.11

Further, the chi-square goodness of fit test shows that normal distribution is not the best approximation for returns distribution (see Table 3). Instead of this, Laplace, logistic or Student distributions can be more appropriate. As for randomness (see Table 4), in all cases time series of returns exhibit less runs than expected; clearly, there is some tendency to creation of cycles and to persistent behaviour, as mentioned already in Peters (1994).

Table 2: Returns distribution: kurtosis

BCPP	1996	1997	1998	1999	1996-1997	1996-1998	1996-1999
CEZ	5.52	1.70	3.90	2.96	3.18	5.41	4.74
KB	9.47	0.94	5.87	7.48	2.88	11.35	10.10
CS	1.19	1.89	6.68	5.64	1.43	16.42	12.53
IPB	5.29	1.19	1.66	3.58	2.68	3.36	3.42
SPLZ	1.97	2.54	1.14	1.33	2.36	3.51	2.89
NHUT	0.20	1.31	3.22	0.83	0.66	3.26	2.38
PRAZ	1.40	4.41	4.63	4.35	2.51	5.68	5.60
SOUH	0.09	1.42	0.72	1.05	0.64	1.18	1.60
TEL	3.36	2.71	1.14	1.61	3.58	2.50	2.23
PX 50	7.09	2.36	2.04	2.63	3.96	4.08	3.66

DJ	30	40	50	60	70	80	90	30-40	30-50	30-60	30-70	30-80	30-90
DJ 1	5.50	11.62	7.36	5.53	1.66	105.6	5.19	10.01	12.82	14.66	13.57	30.28	29.16
DJ 2	5.13	4.98	4.40	3.26	1.96	30.52	2.71	7.74	9.58	11.67	11.73	13.57	12.91

Table 3: Returns distribution: Chi-square goodness-of-fit test

	NORMAL	LAPLACE	LOGISTIC	STUDENT		NORMAL	LAPLACE	LOGISTIC	STUDENT
	SL	SL	SL	SL		SL	SL	SL	SL
CEZ	0.00	0.15	0.00	0.00	PRA	0.00	0.00	0.00	0.03
KB	0.00	0.29	0.00	0.00	SOU	0.00	0.00	0.00	0.00
CS	0.00	0.20	0.00	0.00	TEL	0.00	0.14	0.00	0.00
IPB	0.00	0.00	0.00	0.00	PX50	0.00	0.09	0.51	0.39
SPLZ	0.00	0.00	0.00	0.00	DJ 1	0.00	0.00	0.00	0.00
NHUT	0.00	0.02	0.00	0.00	DJ 2	0.00	0.00	0.00	0.00

Table 4: Returns distribution: Runs test up and down for randomness

	R	**E(R)**	**R/E(R)**	**SL**		**R**	**E(R)**	**R/E(R)**	**S**
CEZ	623	669	0.93	0.00	PRAZ	497	669	0.74	0.
KB	621	669	0.93	0.00	SOUH	558	669	0.83	0.
CS	657	669	0.98	0.39	TEL	644	669	0.96	0.
IPB	533	669	0.80	0.00	PX50	615	669	0.92	0.
SPLZ	595	669	0.89	0.00	DJ 1	11538	12330	0.94	0.
NHUT	614	669	0.92	0.00	DJ 2	11549	12383	0.93	0.

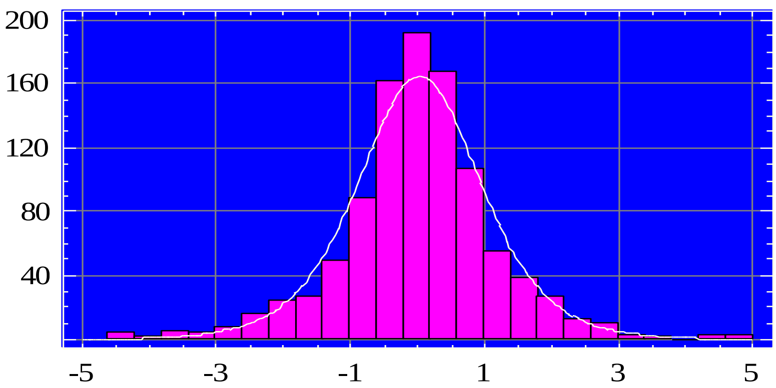

Figure 1: The approximation of PX 50 returns distribution by logistic distribution

2 Box-Jenkins ARIMA Models

First, as the rough approximation, we attempted to model time series of returns by the use of Box-Jenkins (1970). The weakly stationary stochastic process has its autocorrelation function dependent only on the time argument difference h

$$\rho_x(h) = \frac{\operatorname{cov}\left(X_t, X_{t+h}\right)}{\sigma_t \sigma_{t+h}} = \frac{\gamma(h)}{\gamma(0)} \tag{4}$$

The sample counterpart of this function can be computed as follows

$$r(h) = \frac{c(h)}{c(0)} \qquad c(h) = \frac{1}{n}\sum_{t=1}^{n-h}\left(x_t - \bar{x}\right)\left(x_{t+h} - \bar{x}\right) \tag{5}$$

Autoregressive process of p-th order AR(p) is defined as (u_t is white noise)

$$\left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) X_t = \phi_p(B) X_t = u_t \tag{6}$$

where B denotes the backshift operator. Moving average process of q-th order MA(q) is written as

$$X_t = \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right) u_t = \theta_q(B)\, u_t \tag{7}$$

Combining both processes leads to the mixed ARMA(p,q) process

$$\phi_p(B)\, X_t = \theta_q(B)\, \varepsilon_t \tag{8}$$

The corresponding integrated ARIMA(p,d,q) process can be written as

$$\phi_p(B)(1 - B)^d X_t = \theta_q(B) u_t \tag{9}$$

The important special case ARIMA(0,1,0) is known as random walk

$$X_t = \left(1 + B + B^2 + \ldots\right) u_t = \sum_{i=1}^{t} u_i = u_1 + u_2 + \ldots + u_t \tag{10}$$

An example of simulated time series of this process is depicted in Figure 2.

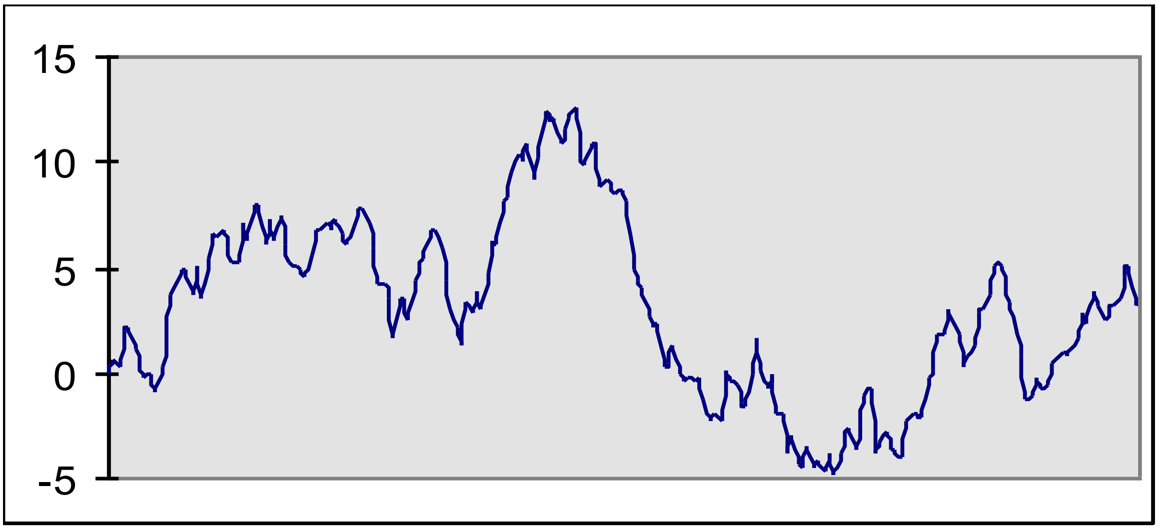

Figure 2: Simulation of random walk process $X_t = X_{t-1} + u_t$

Usually, Ljung-Box test is performed to decide, whether ARIMA model residuals a_t are white noise. The chi-square statistics used in the test for lack of fit are computed according to formulas

$$\chi_m^2 = n(n+2)\sum_{k=1}^{m} \frac{r_k^2}{n-k} \qquad\qquad r_k = \frac{\sum_{t=1}^{n-k} a_t\, a_{t+k}}{\sum_{t=1}^{n} a_t^2} \tag{11}$$

In Table 5, the results of ARIMA modelling are summarized. As a rule, autoregressive term of first order are present. However, higher autoregressive or moving-average terms, particularly of tenth order are still statistically significant. This tendency is even more distinctly expressed in return time series of both Dow-Jones indexes.

Table 5: ARIMA models selected for returns

KB	**MA (1 2 x x 5 x x 8 9 10)**
CS	**AR (1 x x x x x 7 x x 10)**
CEZ	**MA (1 2 x x x x x x x 10)**
IPB	**AR (1 2 x x 5 x x x x x 11) + MA (3 6)**
SPLZ	**AR (1 x x x x x x 8 x x)**
NHUT	**MA (1 x x x x x x x x x)**
PRAZ	**AR (1 x x x x x x x x x)**
SOUH	**AR (1 x x x x x x x x x)**
TEL	**MA (1 x x x 5 x 7 x x 10)**
PX 50	**AR (1 x x x 5 x x x x 10) + MA (2)**
DJIA	**AR (1 2 6 10 11 19 26 27 30 32 42)**
DJAT	**MA (1 2 3 5 11 13 14 16 17 20 23 24 27 38 45)**

3 ARCH and GARCH Models

First, let us introduce the product process

$$\varepsilon_t = X_t - \mu = \sigma_t U_t \qquad E(U_t) = 0 \qquad \mathrm{var}(U_t) = 1 \tag{12}$$

where σ_t is conditional standard deviation of X_t process. Further, let us suppose some type of functional dependence

$$\sigma_t = f(X_{t-1}, X_{t-2}, \ldots) \tag{13}$$

According to Engle (1995), an ARCH (1) process is defined as

$$\varepsilon_t = U_t \sqrt{\alpha_0 + \alpha_1 \left(X_{t-1} - \mu\right)^2} \tag{14}$$

or in alternative expression

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \nu_t \qquad \nu_t = \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2 \left(U_t^2 - 1\right) \tag{15}$$

Thus, an ARCH (1) process is an AR (1) process for ε_t^2 with heteroskedastic errors. Of course, we can generalise further and obtain an ARCH(q) process

$$\varepsilon_t = U_t \sqrt{\alpha_0 + \sum_{i=0}^{q} \alpha_i \varepsilon_{t-1}^2} \qquad \alpha_0 \rangle 0 \qquad \alpha_i \ge 0 \tag{16}$$

More substantial generalisation is accomplished in GARCH(1,1) model

$$\varepsilon_t = U_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2} \tag{17}$$

or, in alternative expression

$$\varepsilon_t^2 = \alpha_0 + \left(\alpha_1 + \beta_1\right) \varepsilon_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1} \tag{18}$$

Thus, GARCH (1,1) process is ARMA (1,1) process for ε_t^2 with heteroskedastic errors. Like before, we can introduce the GARCH(p,q) process

$$\varepsilon_t = U_t \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}} \tag{19}$$

or, in an alternative expression

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{m} \left(\alpha_i + \beta_i\right) \varepsilon_{t-i}^2 - \sum_{i=1}^{q} \beta_i \nu_{t-i} + \nu_t \tag{20}$$

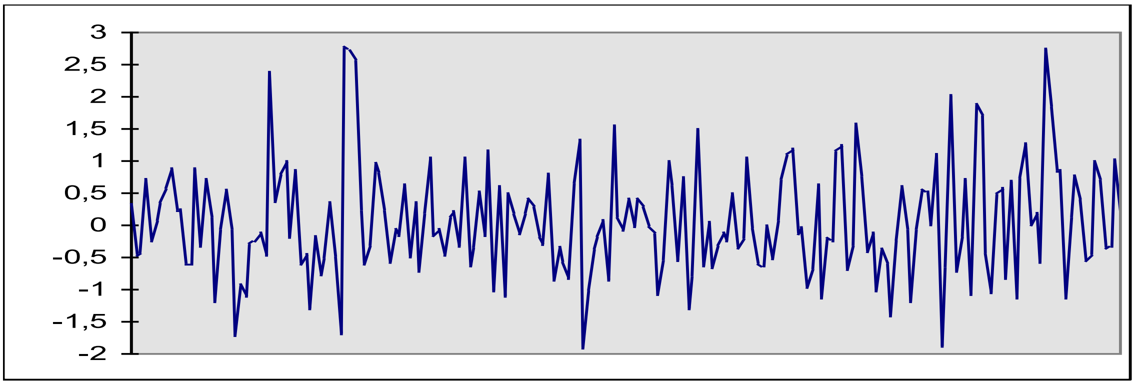

Figure 3: ARCH (1) process: $X_t = U_t \sqrt{0.5 + 0.5 X_{t-1}^2}$

Table 6: AR+GARCH models selected for returns

KB	**AR (1 8 10) + GARCH (p=1 q=1)**
CS	**AR (1 10) + GARCH (p=1 q=1)**
CEZ	**AR (1 10) + GARCH (p=1 q=1)**
IPB	**AR (1 3 10) + GARCH (p=0 q=1 2 3)**
SPLZ	**AR (1) + GARCH (p=1 q=1)**
NHUT	**AR (1) + GARCH (p=1 q=1)**
PRAZ	**AR (1) + GARCH (p=0 q=1 3)**
SOUH	**AR (1 3) + GARCH (p=0 q=1 3 5)**
TEL	**AR (1 10) + GARCH (p=1 q=1)**
PX 50	**AR (1 10) + GARCH (p=2 q=1)**
DJIA	**AR (1 2 6 10) + GARCH (p=2 q=1)**
DJAT	**AR (1 2 3 5 11 16 17 20) + GARCH (p=1 q=1)**

4 Rescaled Range Analysis

Rescaled Range Analysis proved to be powerful and as a general tool for detecting of generating process in time series. Extensive applications in finance time series modelling were given in Peters (1994). The general approach is the following:

- given a time series $x_1, x_2, \dots, x_N$ registered for $t = 1,2,\dots,N$
- divide N values of this time series into m contiguous intervals with length n, $N = mn$
- compute averages for each interval

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \qquad j = 1,2,\dots,m \tag{21}$$

- compute the time series of accumulated departures from the mean value for each interval

$$z_{kj} = \sum_{i=1}^{k}\left(x_{ij} - \bar{x}_j\right) \qquad k = 1,2,\dots,n \tag{22}$$

- compute the range of accumulated departures for each interval

$$R_j = \max\left(z_{kj}\right) - \min\left(z_{kj}\right) \geq 0 \tag{23}$$

- compute the standard deviation for each interval

$$s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2} \qquad (24)$$

- compute the standardized range R_j / s_j for each interval
- the average rescaled range value R / S for an interval of length n

$$\left(R / S\right)_n = \frac{1}{m}\sum_{j=1}^{m}\left(R_j / S_j\right) \qquad (25)$$

According to Hurst, the general type of R / S dependence on time can be written as

$$\left(R / S\right)_n = C n^H \qquad (26)$$

where C is a constant and H Hurst exponent. In practice, H can be estimated using linear regression

$$\log\left[\left(R / S\right)_n\right] = \log C + H \log n \qquad (27)$$

The interpretation of Hurst exponent is the following:

- If $H=0.50$, then a time series is generated by IID (Independent Identically Distributed) process. R / S analysis is nonparametric procedure and it does not require assumptions about the shape of underlying distribution.
- If $0.50 \langle H \langle 1.00$, than a time series is generated by some kind of persistent process characterized by long memory efffects. There is no characteristic time scale (typical feature of fractal time series).
- If $0 \langle H \langle 0.50$, than a time series is generated by some kind of antipersistent process that reverses itself more frequently than a random process.

Table 7: Results of R / S analysis

	H returns	**H AR(1) residuals**
KB	**0.656**	**0.624**
SPLZ	**0.635**	**0.584**
CS	**0.606**	**0.595**
NHUT	**0.598**	**0.580**
SOUH	**0.595**	**0.566**
CEZ	**0.594**	**0.577**
IPB	**0.593**	**0.567**
TEL	**0.524**	**0.516**
PRAZ	**0.512**	**0.494**
PX 50	**0.637**	**0.610**

In Table 7, the Hurst coefficient is given both for returns and their AR(1) residuals. Practically in all cases, the values are higher than 0.5 indicating a departure from random walk behaviour. Again, some tendency of cycles creation is manifested here.

This study was prepared with support from the Grant Agency of the Czech Republic under contract number 402 / 00 / 0459.

References

E. Peters. Fractal Market Analysis: Applying Chaos Theory to Investment and Economics. Wiley, New York, 1994.

G. Box and G. Jenkins. Time Series Analysis, Forecasting and Control. Holden Day, San Francisco, 1970.

R. Engle (editor). ARCH (Selected Readings). Oxford Univ. Press, Oxford 1995.

Authors' addresses:

Jiří Trešl, Dagmar Blatná

University of Economics
4 W. Churchill Sq.
13067 Prague 3
Czech Republic
E-mail: TRESL@VSE.CZ