An empirical application of a two-factor model of stochastic volatility

Alexandr Kuchynka
University of West Bohemia in Pilsen, Faculty of Economics
Institute of Information Theory and Automation of the ASCR

Abstract

This contribution focuses on the modelling of volatility of returns in Czech and US stock markets using a two-factor stochastic volatility model, i.e. the volatility process is modeled as a superposition of two autoregressive processes. As the volatility is not observable, the logarithm of the daily range is employed as the proxy. The estimation of parameters and volatility extraction are performed using the Kalman filter. We have obtained a meaningful decomposition of the volatility process into one highly persistent factor and another quickly mean-reverting factor. Moreover, we have shown that although the overall *level* of the volatility of returns is roughly the same in both markets, the US market exhibits substantially lower *volatility* of the volatility process.

1. Theory

1.1. Stochastic volatility models

The fact that volatility is not constant in most financial time series such as stock prices and exchange rates is widely recognised. From a practical point of view, estimating volatility is important in several fields within finance such as option pricing, portfolio optimization or risk management. There exist two prominent approaches to deal with time-dependent variances: ARCH/GARCH and stochastic volatility (SV) approaches. The GARCH model (Bollerslev, 1986) focuses on capturing the clustering of volatility in returns when the conditional variance at time t is modelled as a deterministic function of lagged values of conditional variances and squared returns. On the other hand, the stochastic volatility models understand the time-varying variance as a stochastic process which can be a continuous-time diffusion (Hull and White, 1987) or a more general Lévy process (Barndorff-Nielsen and Shepard, 2001). For econometric purposes, it is convenient to work with some discretized version of the model (pioneered by the seminal paper Taylor, 1982).

Now we introduce a simple example of a stochastic volatility model. Let the asset price $P(s)$ follow the geometric Brownian motion with local volatility $\sigma(s)$ and suppose that log-volatility evolves as a mean-reverting Ornstein-Uhlenbeck process with mean $\ln \bar{\sigma}$ and mean reversion parameter $\alpha > 0$:

$$dP(s) = \mu P(s)ds + \sigma(s)P(s)dW_1(s) \tag{1a}$$

$$d \ln \sigma(s) = \alpha\left(\ln \bar{\sigma} - \ln \sigma(s)\right)ds + \beta dW_2(s) \tag{1b}$$

where $W_1(s)$ and $W_2(s)$ are independent Brownian motions. This specification leads to a lognormal SV model which is widely used in volatility modelling. One reason for its popularity is given by the econometric tractability of its discrete-time counterpart, represented by a linear Gaussian first order autoregressive process.

1.2. Volatility proxies

It is important to note that the log-volatility is a latent variable and therefore is not directly observable. Instead, we are able to observe some of its *proxy* contaminated by a measurement error. Early attempts (for instance Ruiz, 1994, or Harvey et al., 1994) focused on squared returns as a volatility *proxy*, however the performance of Kalman filter turn out to be quite poor due to substantial skewness of squared returns. A novel and promising approach (Alizadeh et al., 2002) relies instead on the log range defined in the following way

$$R_t = \ln\left(\sup_{s\in[0,1]} p(s) - \inf_{s\in[0,1]} p(s) \right) \tag{2}$$

where $p(s)$ denotes the log price and supremum and infimum are taken over the daily interval which is normalized to unity purely for ease of notation.

It turns out that the log range exhibits several desirable properties: first, it is more efficient due to lower variance of the measurement errors relative to log absolute returns. Second, the log range is known to be robust to certain market microstructure effects as bid-ask bounce. Finally, contrary to log absolute returns, the distribution of the log range is nearly Gaussian. The asymptotic distribution of log range has been studied in Alizadeh et al. (2002); it is easily obtained in the series form (based on the result of Feller, 1951) but for practical purposes it can be well approximated by the normal distribution with mean $0.43 + h_t$ and variance 0.08 where $h_t \equiv \ln(\sigma_t)$ is the daily log-volatility. On the other hand, the distributional properties of the log absolute return are quite different. Recall that returns in discrete-time SV model are generated by

$$r_t = u_t \sigma_t \tag{3}$$

with $u_t \sim nid(0,1)$. Equation (3) implies

$$\ln\left(|r_r|\right) = -0.64 + h_t + \xi_t \tag{4}$$

where the measurement error $\xi_t \equiv \ln\left(|u_t|\right) - E\ln\left(|u_t|\right)$ has zero mean, variance $\pi^2/8 \doteq 1.23$ and is highly skewed.

However, a word of caution is needed here. In finite samples, the distribution of range estimators depends also on the number of observations per unit of time (day in this case). Therefore, we investigated the impact of discretization on mean and variance of the log range by a Monte Carlo simulation (results are reproduced in Table 1). The pattern is clear: reducing the number of observations during a trading day results in lower mean and higher variance. Nevertheless, the variance of the proxy seems to be quite close to the asymptotic one even for 50 trades per day.

N	5	10	50	100	200	500	1000
Mean	-0.115	0.097	0.300	0.340	0.366	0.401	0.415
Variance	0.233	0.152	0.104	0.097	0.092	0.086	0.084

Table 1. Mean and variance of log range for a Wiener process with zero drift and unit variance observed N times during a unit period. The Monte Carlo simulation was performed with 1 milion replications.

There exists a closely related estimator proposed by Parkinson (1980) given by $R_t^2/4\ln 2$ and several modifications thereof: Garman and Klass (1980) included the open and close prices in addition to the high and close prices, Rogers and Satchell (1991) suggested an estimator which allowed for a nonzero drift and also investigated the discretization bias (see also Christensen and Podolskij, 2007).

If higher efficiency is needed, intraday data should be used and appropriated volatility estimator constructed (realized volatility, see Andersen et al., 2001, or realized range, see Martens and van Dijk, 2007).

1.3. One and two-factor models

Stochastic volatility models can be conveniently written in the state space form. In the simplest case with one process governing the log-volatility the model reads as follows:

$$h_t = \overline{h} + \rho\left(h_{t-1} - \overline{h}\right) + \eta_t \tag{5a}$$
$$R_t = b + h_t + \varepsilon_t \tag{5b}$$

where h_t represents the latent log-volatility, $\rho \in (0,1)$ the autoregression parameter, $\overline{h}$ the mean of the log-volatility process, R_t the observed log-volatility proxy (log range), b the bias of the proxy, $\eta_t \sim nid(0, \sigma_\eta^2)$ and $\varepsilon_t \sim nid(0, \sigma_\varepsilon^2)$ are uncorrelated transition and measurement errors, respectively.

Equation (5a) is a transition equation which describes the dynamics of the latent variable, whereas the observation equation (5b) relates the latent variable and its (observable) proxy. The dynamics of the latent log-volatility can be enriched by including a second component. In this case, transition equations are modified in the following way:

3

$$h_t = \bar{h} + h_{1t} + h_{2t} \tag{6a}$$

$$h_{1t} = \rho_1 h_{1t-1} + \eta_{1t} \tag{6b}$$

$$h_{2t} = \rho_2 h_{2t-1} + \eta_{2t} \tag{6c}$$

where h_{1t}, h_{2t} are volatility factors, $\rho_1, \rho_2 \in (0,1)$ autoregression parameters and η_{1t}, η_{2t} mutually uncorrelated *n.i.d.* disturbances.

It is worth discussing in more details the benefits of employing two-factor models. Probably the most important reason is the ability thereof to capture several empirically observed patterns of the autocorrelation function, and the long memory-like behaviour in particular. The fact that the superposition of independent short memory processes (for instance, Ornstein-Uhlenbeck process in continuous time formulation or autoregressive process in discrete time) can mimic slow decay of the autocorrelation function or power laws empirically observed has been explored by several authors (LeBaron, 2001 or Barndorff-Nielsen, 2001, among others). The idea had appeared even in the context of GARCH models: Ding and Granger (1996) suggested a two-component GARCH model, one component describing the short-run effect whereas the persistent component specified as IGARCH process.

For the sake of illustration, suppose a composite process $x_t = y_{1t} + y_{2t}$ where both component processes are modeled as AR(1), i.e.

$$y_{it} = \gamma_i y_{it-1} + u_{it} \qquad i = 1,2 \tag{7}$$

with $\gamma_1 \neq \gamma_2$ and $\{u_{1t}\}, \{u_{2t}\}$ white noise sequences which are uncorrelated at all leads and lags. As the autocovariance function of a sum of independent components is equal to the sum of the autocovariances, it follows easily that $corr(x_t, x_{t-k}) = w_1 \gamma_1^k + w_2 \gamma_2^k$, $k = 1,2,...,$ and the weights are given by $w_i = \mathrm{var}(y_{it}) / (\mathrm{var}(y_{1t}) + \mathrm{var}(y_{2t}))$. An example of such an autocorrelation function for $w_1 = w_2 = 0.5, \gamma_1 = 0.99, \gamma_2 = 0.1$ is shown in Figure 1.

1.4. Model estimation and volatility extraction

We will shortly review the filtering algorithm known as Kalman filter (for a more detailed description see Hamilton, 1994). We assume that there exists a *(n x 1)* vector of observed variables $\mathbf{y}_t$ whose behaviour depends on a *(r x 1)* vector of unobserved (state) variables $\mathbf{z}_t$. The state space representation of the system is

$$\mathbf{z}_t = \mathbf{a} + \mathbf{F}\mathbf{z}_{t-1} + \mathbf{v}_t \tag{8a}$$

$$\mathbf{y}_t = \mathbf{b} + \mathbf{H}\mathbf{z}_t + \mathbf{w}_t \tag{8b}$$

where $\mathbf{a}, \mathbf{b}$ are *(r x 1)* and *(n x 1)* vectors, respectively, $\mathbf{F}$ and $\mathbf{H}$ are *(r x r)* and *(n x r)* matrices, respectively, $E(\mathbf{v}_t \mathbf{v}_\tau^T) = \mathbf{Q}$ for $t = \tau$ and $\mathbf{0}$ otherwise, and $E(\mathbf{w}_t \mathbf{w}_\tau^T) = \mathbf{R}$ for $t = \tau$ and $\mathbf{0}$ otherwise. The vectors $\mathbf{v}_t$ and $\mathbf{w}_t$ are uncorrelated for all t and τ. Moreover, it is convenient to

assume that the initial state $\mathbf{z}_1$ and the innovations $\{\mathbf{v}_t, \mathbf{w}_t\}$ are multivariate Gaussian. If the assumption of normality is dropped, the Kalman filter yields optimal (in the mean square sense) linear projections which in general differ from conditional expectations.

Step 1: Forecasting $\mathbf{y}_t$ given the information at time *t-1*

Given the initial value $\mathbf{z}_{1|0} = E(\mathbf{z}_1)$ with associated mean squared error (MSE) $\mathbf{P}_{1|0} = E\left\{ \left[\mathbf{z}_1 - \mathbf{z}_{1|0} \right]\left[\mathbf{z}_1 - \mathbf{z}_{1|0} \right]^T \right\}$ we can compute the forecast of $\mathbf{y}_t$ given the information observed up to time *t-1* as

$$\mathbf{y}_{t|t-1} = \mathbf{b} + \mathbf{H}\mathbf{z}_{t|t-1} \tag{9}$$

with the corresponding MSE

$$E\left[\left(\mathbf{y}_t - \mathbf{y}_{t|t-1} \right)\left(\mathbf{y}_t - \mathbf{y}_{t|t-1} \right)^T \right] = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R} \equiv \mathbf{V}_{t|t-1} \tag{10}$$

Step 2: Updating the inference about $\mathbf{z}_t$ given $\mathbf{y}_t$

Next we update the inference about the current value of the state variables $\mathbf{z}_t$ as a new observation of $\mathbf{y}_t$ becomes available:

$$\mathbf{z}_{t|t} = \mathbf{z}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}^T \mathbf{V}_{t|t-1}^{-1} \left(\mathbf{y}_t - \mathbf{y}_{t|t-1} \right) \tag{11}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}^T \mathbf{V}_{t|t-1}^{-1}\mathbf{H}\mathbf{P}_{t|t-1} \tag{12}$$

Step 3: Forecasting $\mathbf{z}_{t+1}$ given the information at time *t*

Subsequently, the transition equation (8b) is used to compute the forecast of $\mathbf{z}_{t+1}$ based on the information available at time *t* together with its MSE:

$$\mathbf{z}_{t+1|t} = \mathbf{a} + \mathbf{F}\mathbf{z}_{t|t} = \mathbf{a} + \mathbf{F}\mathbf{z}_{t|t-1} + \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}^T \mathbf{V}_{t|t-1}^{-1} \left(\mathbf{y}_t - \mathbf{y}_{t|t-1} \right) \tag{13}$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^T + \mathbf{Q} \tag{14}$$

If parameters of the model are unknown, their estimates can be obtained by maximizing the quasi-loglikelihood defined as (after omitting constant terms)

$$-\frac{1}{2}\sum_{t=1}^{T} \ln \det \left(\mathbf{V}_{t|t-1} \right) - \frac{1}{2}\sum_{t=1}^{T} \mathbf{u}_t^T \mathbf{V}_{t|t-1}^{-1}\mathbf{u}_t \tag{15}$$

where $\mathbf{u}_t \equiv \mathbf{y}_t - \mathbf{y}_{t|t-1}$ is a one step ahead prediction error.

The quasi-maximum likelihood (QML) estimation is known to yield consistent and asymptotically normal estimates of the unknown parameters, however, they are inefficient and their sampling properties can be quite poor if the normality approximation is inadequate.

2. Empirical application

2.1. Description of the dataset

ČEZ (CZ0005112300)	*Prague Stock Exchange (SPAD)*
Telefónica O2 C.R. (CZ0009093209)	*Prague Stock Exchange (SPAD)*
Erste Bank (AT0000652011)	*Prague Stock Exchange (SPAD)*
General Electric Co.	*New York Stock Exchange*
Microsoft Corp.	*NASDAQ*
Intel Corp.	*NASDAQ*

Table 2. Stocks included in the dataset.

We use daily high and low prices of six stocks (see Table 2) for the period from September 16, 2005 until November 13, 2007 (543 and 544 observations for Czech and US markets, respectively). The average number of transactions per day for ČEZ, Telefónica O2 C.R. and Erste Bank was 231, 102 and 86, respectively. In the case of Erste Bank there were only three transactions during the trading day in April 14, 2006 giving rise to the observed range very close to zero and a corresponding large negative outlier in the log range. Therefore, this observation was excluded from our analysis.

The data are depicted in Figures 2a and 2b together with their sample autocorrelation functions and QQ plots. The autocorrelation functions clearly show a certain degree of persistency in the volatility proxy. Empirical moments of the log range are reported in Table 3.

	Mean	*Std deviation*	*Skewness*	*Kurtosis*
ČEZ	*-3.7541*	*0.5796*	*0.1082*	*3.1717*
Telefónica O2 C.R.	*-4.2106*	*0.6215*	*0.2519*	*3.3516*
Erste Bank	*-4.2254*	*0.5840*	*-0.0179*	*2.9515*
General Electric	*-4.4138*	*0.3962*	*0.4316*	*3.1181*
Microsoft	*-4.2230*	*0.4194*	*0.3390*	*3.0718*
Intel	*-3.9355*	*0.3821*	*0.1501*	*2.7630*

Table 3. Unconditional moments of the observed log range

2.2. Estimation results

Now we proceed to estimate both stochastic volatility models by quasi-maximum likelihood method. Estimation results are shown in Tables 4 and 5 and smoothed extractions of volatility factors for the two-factor model are depicted in Figures 3a and 3b. In accordance with findings of Alizadeh et al. (2002), there exists a strong evidence that the log-volatility

process can be meaningfully decomposed into one highly persistent factor and another quickly mean-reverting factor (for a more detailed discussion concerning the inadequacy of one factor model see Alizadeh et al., 2002).

In order to obtain a better interpretation of our results, we computed estimated variances for both volatility factors and the total variance of the log-volatility process as their sum (due to zero cross-correlation) (see Table 6). The variance of both factors is roughly the same with notable exceptions of Erste Bank and Intel where the second (less persistent component) seems to be much more volatile. Comparing Czech and US markets, the most striking feature is substantially lower *volatility* of the log-volatility process for US stocks, even if the overall *level* of the volatility of returns (estimated by $\bar{h}$) is roughly the same.

	ρ	$\bar{h}$	var(η)
ČEZ	0.7507	-4.179	0.1087
	(0.0391)	(0.056)	(0.0141)
Telefónica O2 C.R.	0.6848	-4.6395	0.1593
	(0.041)	(0.0521)	(0.0168)
Erste Bank	0.5855	-4.6564	0.1681
	(0.0464)	(0.0439)	(0.017)
General Electric	0.8718	-4.8384	0.0147
	(0.0326)	(0.0426)	(0.0047)
Microsoft	0.7974	-4.6484	0.0292
	(0.0221)	(0.037)	(0.0023)
Intel	0.5036	-4.3656	0.0446
	(0.0301)	(0.0211)	(0.0072)

Table 4. Quasi-maximum likelihood estimates of the one-factor model (asymptotic standard errors appear in parentheses)

	ρ_1	ρ_2	$\bar{h}$	var(η_1)	var(η_2)
ČEZ	0.9427	0.3992	-4.1688	0.0156	0.0941
	(0.0234)	(0.0944)	(0.0917)	(0.0066)	(0.0143)
Telefónica O2 C.R.	0.9403	0.4938	-4.6317	0.013	0.1427
	(0.0385)	(0.0955)	(0.0919)	(0.0117)	(0.021)
Erste Bank	0.9395	0.4235	-4.6493	0.0081	0.1542
	(0.0369)	(0.0838)	(0.0699)	(0.0065)	(0.0185)
General Electric	0.9551	0.1602	-4.8295	0.0039	0.0299
	(0.0228)	(0.0666)	(0.0579)	(0.002)	(0.008)
Microsoft	0.946	0.1086	-4.6366	0.0056	0.0404
	(0.0225)	(0.1292)	(0.0604)	(0.0024)	(0.0093)
Intel	0.9799	0.3444	-4.3664	0.0004	0.0457
	(0.0171)	(0.1141)	(0.0387)	(0.0004)	(0.0083)

Table 5. Quasi-maximum likelihood estimates of the two-factor model (asymptotic standard errors appear in parentheses)

	First factor	*Second factor*	*Total variance*
ČEZ	*0.1401*	*0.1119*	*0.2520*
Telefónica O2 C.R.	*0.1122*	*0.1887*	*0.3009*
Erste Bank	*0.0690*	*0.1879*	*0.2569*
General Electric	*0.0444*	*0.0307*	*0.0751*
Microsoft	*0.0533*	*0.0409*	*0.0942*
Intel	*0.0101*	*0.0518*	*0.0619*

Table 6.Variances of individual factors and the total variance of the estimated log-volatility.

3. Conclusion

We have analysed the volatility of returns of six representative Czech and US stocks using a simple but quite flexible two-factor model which specifies the log-volatility as a superposition of two independent autoregressive processes with different persistence rates. This approach offers a much richer dynamics than the usual autoregressive models allowing to model the temporal dependence in a parsimonious way. Moreover, the availability of a nearly Gaussian proxy contaminated with a low amount of the measurement error allows to estimate parameters of the model and extract the latent log-volatility in a computationally efficient way. The main empirical finding of this paper is that Czech and US stock markets seem to differ considerably in the variability of the log-volatility process rather than in the overall level thereof. Given the importance of the proper volatility modelling in the risk management area it would be interesting to extend our analysis to a multivariate setting. However, this issue is left for further research.

References

ALIZADEH, S. - BRANDT, M. - DIEBOLD, F. (2002): *Range-based estimation of stochastic volatility models.* Journal of Finance 57, pp. 1047–1091.

ANDERSEN, T.G. - BOLLERSLEV, T. - DIEBOLD, F.X. - LABYS, P. (2001): *The distribution of realized exchange rate volatility.* Journal of the American Statistical Association 96, pp. 42–55.

BARNDORFF-NIELSEN, O. (2001): *Superposition of Ornstein-Uhlenbeck type processes.* Theory Prob. Its Appl. 45, pp. 175-194.

BARNDORFF-NIELSEN, O. E. – SHEPARD, N. (2001): *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economic..* Journal of the Royal Statistical Society, Series B 63, pp. 167-241.

BOLLERSLEV, T. (1986): *Generalized autoregressive conditional heteroskedasticity.* Journal of Econometrics 31, pp. 307–327.

CHRISTENSEN, K. - PODOLSKIJ, M. (2007): *Realized range-based estimation of integrated variance,* Journal of Econometrics, 141, no 2, pp. 323-349.

DING, Z. – GRANGER, C.W.J. (1996): *Modeling volatility persistence of speculative returns: a new approach.* Journal of Econometrics, 73, no 1, pp. 185-215.

ENGLE, R.F. (1982): *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.* Econometrica 50, pp. 987–1007.

FELLER, W. (1951): *The asymptotic distribution of the range of sums of independent random variables.* Annals of Mathematical Statistics, 22, pp. 427-432.

GARMAN, M. – KLASS, M.J. (1980): *On the estimation of security price volatilities from historical data.* Journal of Business, 53, pp. 67-78.

HAMILTON, J. (1994): *Time Series Analysis.* Princeton, Princeton University Press.

HARVEY, A.C. – RUIZ, E. – SHEPARD, N. (1994): *Multivariate stochastic variance models.* Rev. Economic Studies, 61, pp. 247-264.

HULL, J. – WHITE, A. (1987): *The pricing of options on assets with stochastic volatilities.* Journal of Finance 42, pp. 281 - 300.

LEBARON, B. (2001): *Stochastic volatility as a simple generator of apparent financial power laws and long memory.* Quantitative Finance, Vol 1, No 6, pp. 621-31.

MARTENS, M. – VAN DIJK, D. (2007): *Measuring volatility with the realized range.* Journal of Econometrics, 138, pp. 181-207.

PARKINSON, M. (1980): *The extreme value method for estimating the variance of the rate of return.* Journal of Business, 53(1), pp. 61-65.

ROGERS, L. C. G. - SATCHELL, S. E. (1991): *Estimating variances from high, low, and closing prices*, Annals of Applied Probability 1(4), pp. 504–512.

RUIZ, E. (1994): *Quasi-maximum likelihood estimation of stochastic volatility models.* Journal of Econometrics, 63, pp. 289 – 306.

TAYLOR, S. J. (1982): *Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices 1961-79.* In O. D. Anderson (ed.), Time Series Analysis: Theory and Practice, 1, pp. 203-226. Amsterdam: North-Holland.

Figures
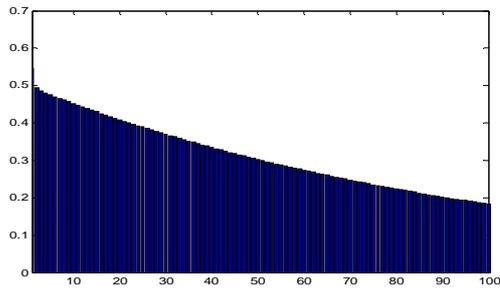
Figure 1. Example of an autocorrelation function for the two-component model
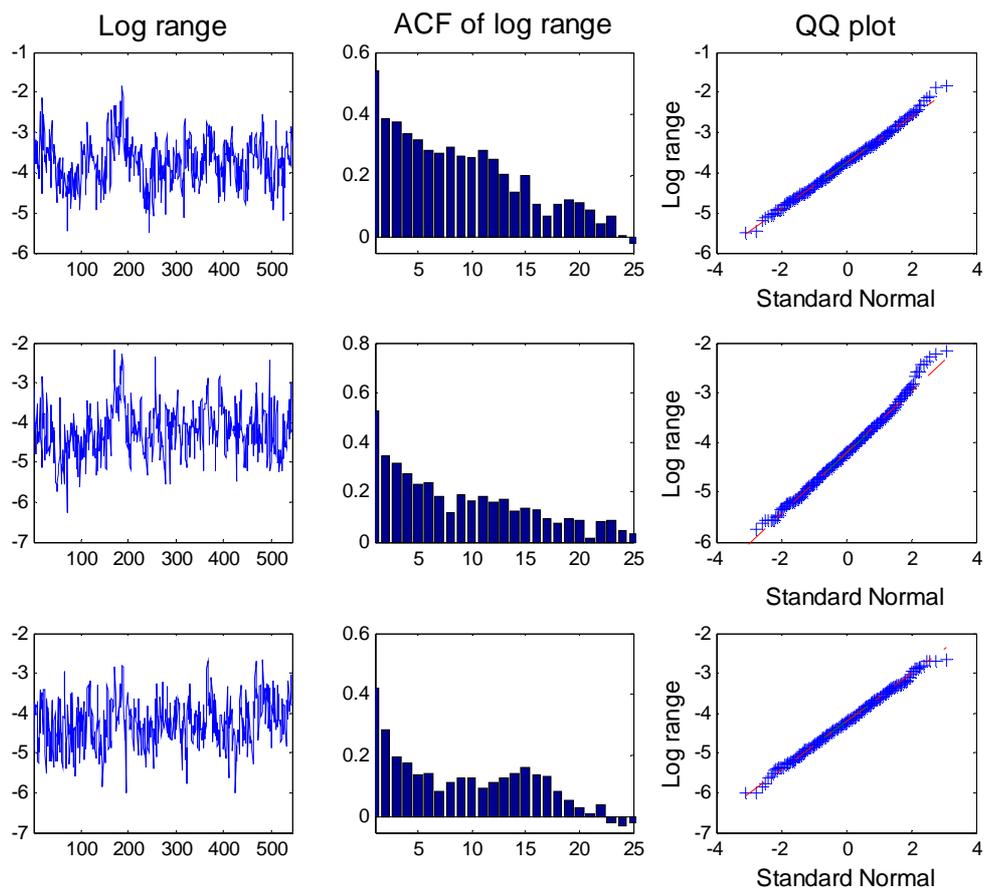
Figure 2a. Log range, its sample autocorrelation function and QQ plots for ČEZ, Telefónica O2 C.R. and Erste Bank (from top to bottom).
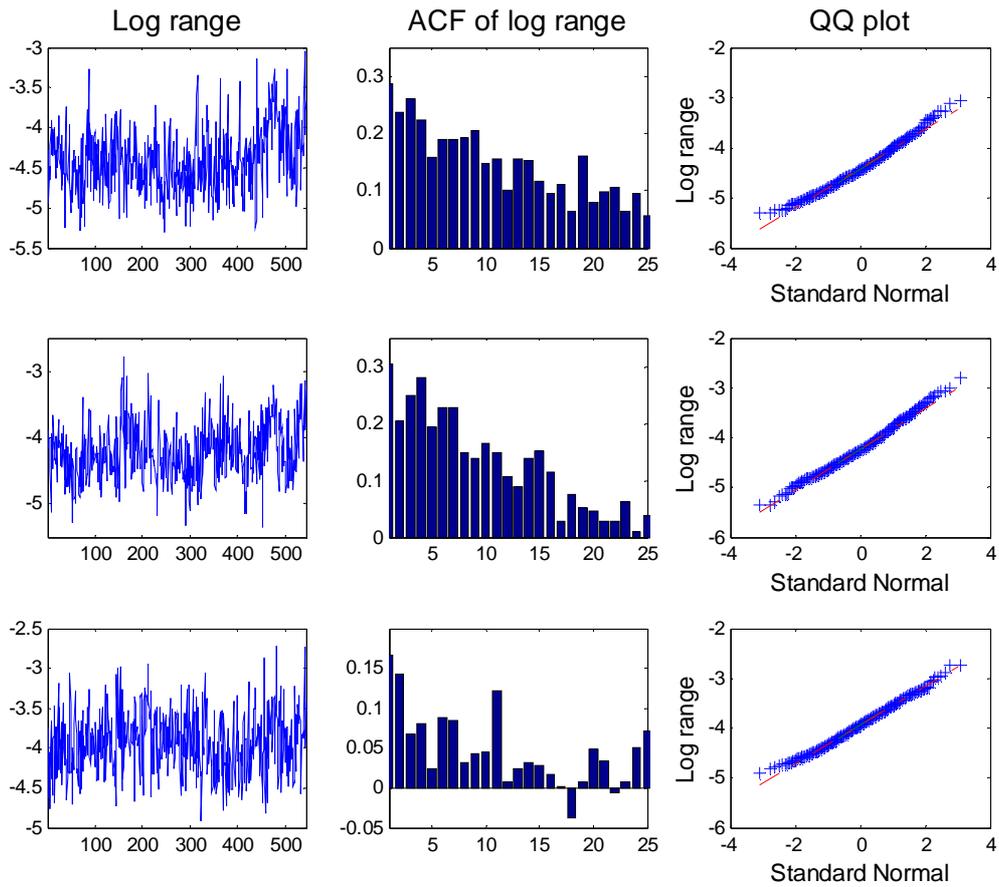
Figure 2b. Log range, its sample autocorrelation function and QQ plots for General Electric, Microsoft and Intel (from top to bottom).
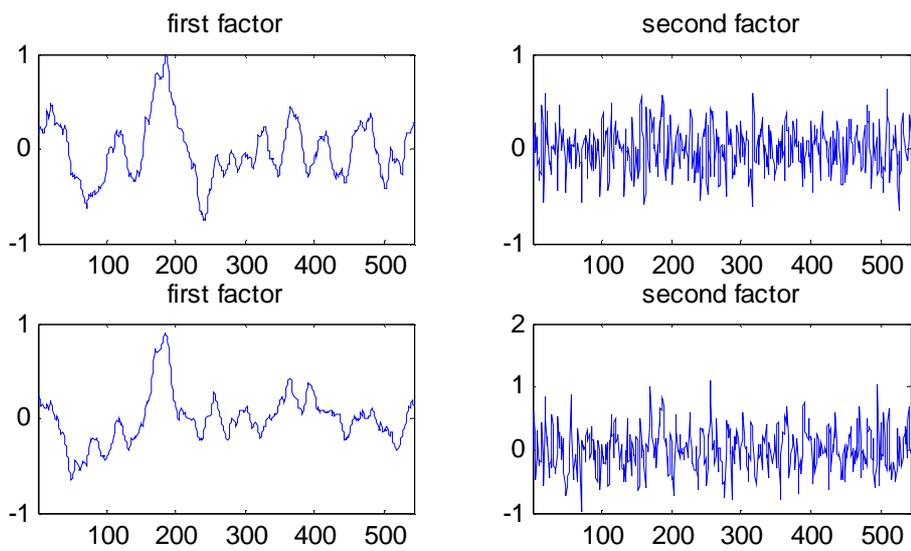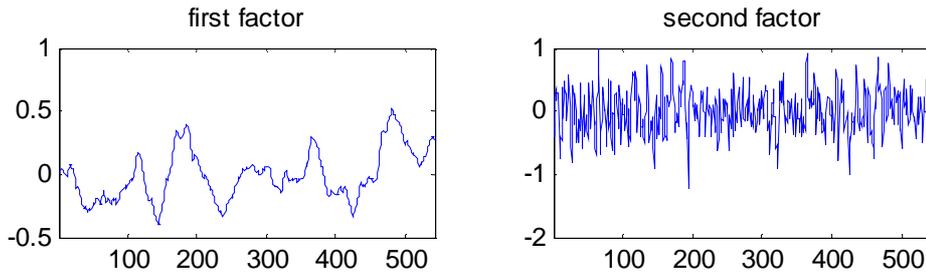
Figure 3a. Smoothed extractions of volatility factors for the two-factor model for ČEZ, Telefónica O2 C.R. and Erste Bank (from top to bottom).
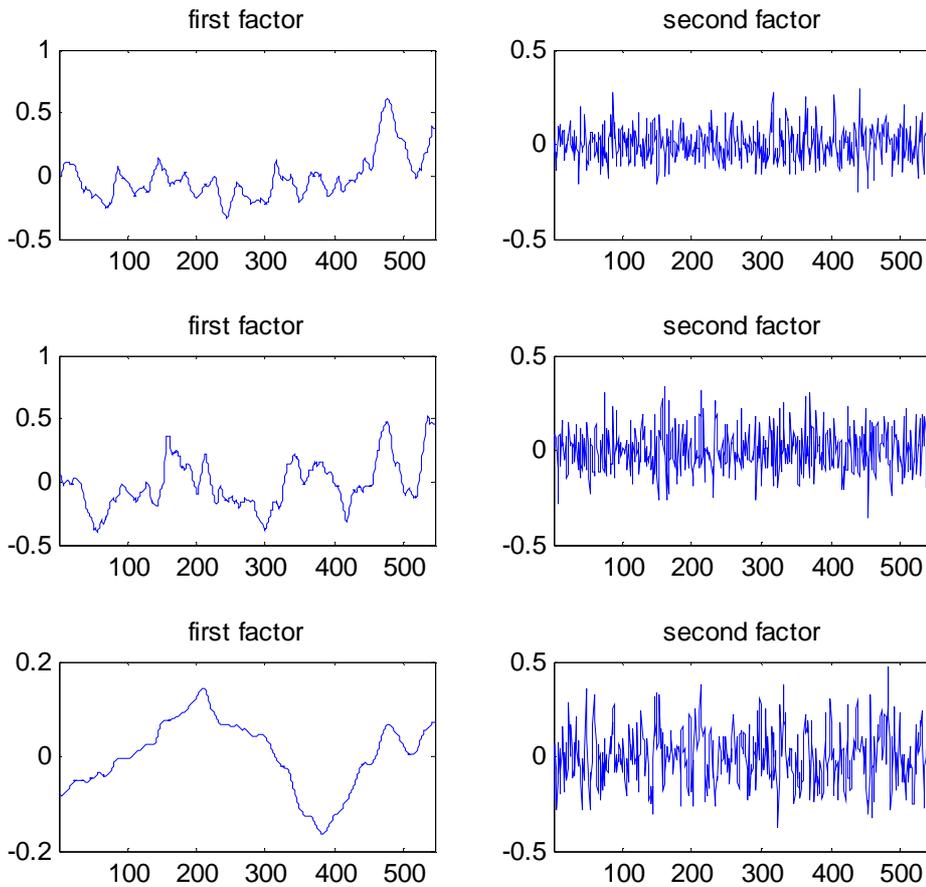
Figure 3b. Smoothed extractions of volatility factors for the two-factor model for General Electric, Microsoft and Intel (from top to bottom).