

Institute of Economic Studies, Faculty of Social Sciences  
Charles University in Prague

# Volatility extraction using the Kalman filter

Alexandr Kuchynka

IES Working Paper: 10/2008



**Institute of Economic Studies,  
Faculty of Social Sciences,  
Charles University in Prague**

**[UK FSV – IES]**

**Opletalova 26  
CZ-110 00, Prague  
E-mail : [ies@fsv.cuni.cz](mailto:ies@fsv.cuni.cz)  
<http://ies.fsv.cuni.cz>**

**Institut ekonomických studií  
Fakulta sociálních věd  
Univerzita Karlova v Praze**

**Opletalova 26  
110 00 Praha 1**

**E-mail : [ies@fsv.cuni.cz](mailto:ies@fsv.cuni.cz)  
<http://ies.fsv.cuni.cz>**

**Disclaimer:** The IES Working Papers is an online paper series for works by the faculty and students of the Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic. The papers are peer reviewed, but they are *not* edited or formatted by the editors. The views expressed in documents served by this site do not reflect the views of the IES or any other Charles University Department. They are the sole property of the respective authors. Additional info at: [ies@fsv.cuni.cz](mailto:ies@fsv.cuni.cz)

**Copyright Notice:** Although all documents published by the IES are provided without charge, they are licensed for personal, academic or educational use. All rights are reserved by the authors.

**Citations:** All references to documents served by this site must be appropriately cited.

**Bibliographic information:**

**Kuchynka, A. (2008). “ Volatility extraction using the Kalman filter ” IES Working Paper 10/2008. IES FSV. Charles University.**

**This paper can be downloaded at: <http://ies.fsv.cuni.cz>**

# Volatility extraction using the Kalman filter

**Alexandr Kuchynka<sup>#</sup>**

<sup>#</sup> IES, Charles University Prague,  
Institute of Information Theory and Automation of the ASCR,  
Faculty of Economics, University of West Bohemia in Pilsen  
E-mail: alexk@kso.zcu.cz

June 2008

## **Abstract:**

**This paper focuses on the extraction of volatility of financial returns. The volatility process is modeled as a superposition of two autoregressive processes which represent the more persistent factor and the quickly mean-reverting factor. As the volatility is not observable, the logarithm of the daily high-low range is employed as its proxy. The estimation of parameters and volatility extraction are performed using a modified version of the Kalman filter which takes into account the finite sample distribution of the proxy.**

**Keywords:** volatility, stochastic volatility models, Kalman filter, volatility proxy

**JEL:** C22,G15.

## **Acknowledgements:**

**This research was supported by Czech Science Foundation GACR under Grant Nr. 402/07/0465 and Ministry of Education of the Czech Republic within the Center for Basic Research in Dynamic Economics and Econometrics (LC06075).**

## 1. Introduction

Modelling financial data comprises several issues, the most important being a proper specification of the time-varying variance of financial returns, so called volatility. As the variance of returns is used as a measure of risk, the importance of an adequate characterization of volatility is obvious in the areas of risk management or portfolio optimization. Another application arises in the context of option pricing. In the original Black-Scholes setting, the risk is quantified by a constant volatility parameter. If the true variance of the financial asset is time-varying rather than constant, the pricing formula has to be modified correspondingly.

There exist two prominent approaches to deal with volatility: GARCH and stochastic volatility (SV) approaches. The GARCH model (Engle, 1982, 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). Stochastic volatility models are typically formulated in the state space form and they are intimately related to the problem of signal extraction. It is well-known (see for instance Hamilton, 1994) that for linear system with Gaussian innovations the Kalman filter offers an optimal way (in the sense of minimizing mean squared errors) for sequentially updating a linear projection.

The paper is organized as follows: in the second chapter, we introduce the concept of the volatility proxy along the lines of de Vilder and Visser (2007) which represents an useful framework for assessing the quality of such proxies. Typical representants of volatility proxies are the high-low range and the absolute daily return because they are constructed from the easily available quaternion „high-low-open-close“ price. Since it turns out that logarithm of daily range (hereafter *log range*) has substantially lower variance than logarithm of absolute return and is nearly Gaussian, we will concentrate on this proxy. Although the asymptotic distribution of the log range has been studied in detail in Alizadeh et al. (2002), less attention has been paid to its finite sample counterpart. In particular, due to the discretization error, the mean and variance of the distribution vary with the number of points

in the grid over which minima and maxima are taken. We study the sensitivity of first four moments on the number of observations using a Monte Carlo simulation.

In the third chapter we suggest a modification of the standard Kalman filter algorithm which uses finite sample mean and variance of the log range as input. Finally in the fourth chapter, we apply our methodology to the data from the Prague Stock Exchange.

## 2. Measuring volatility using proxies

In our setup, we make use of the framework developed in de Vilder and Visser (2007) which is particularly useful when dealing with proxies for unobserved volatility. Recall that discrete time models typically exhibit a product structure

$$r_t = s_t Z_t \quad (1)$$

where the observed return for day  $t$   $r_t$  is described as a product of some positive scale factor  $s_t$  and *i.i.d.* innovation  $Z_t$ .

It is useful to consider a continuous time version of (1) with independent copies  $\{\Psi_t\}$  of some stochastic process  $\Psi$  in place of *i.i.d.* noise  $\{Z_t\}$ :

$$R_t(J) = s_t \Psi_t(J), \quad J \in [0,1] \quad (2)$$

Thus, the cumulated log return  $R_t$  is just a scaled version of some stochastic process  $\Psi_t$ , which can be a Wiener process (in the simplest setup) or a more complex process capturing intraday seasonality or jumps. Formally, we suppose that  $\Psi$  is a *càdlàg* (right continuous with left limits) process on  $[0,1]$ , left continuous at 1. For identification purposes we assume that  $\Psi$  is standardized, i.e.  $E\Psi(1) = 0$ ,  $\text{var}(\Psi(1)) = 1$ . The time  $J$  runs from the opening until the closing of the trading day, the overnight return from day  $t-1$  to  $t$  is given by  $s_t \Psi_t(0)$ . Daily close-to-close return then becomes

$$r_t = R_t(1) = s_t \Psi_t(1) \quad (3)$$

In order to compare this approach with the usual stochastic volatility models, suppose that  $\Psi$  is a diffusion obeying

$$d\Psi(J) = v(J) dB(J) \quad (4)$$

with instantaneous stochastic volatility  $v(J)$ . The cumulated log return process then follows

$$dR_t(J) = s_t v_t(J) dB(J) \equiv s_t(J) dB(J) \quad (5)$$

Thus, the local volatility  $s_t(J)$  can be decomposed into a daily scale factor  $s_t$  and an independent component  $v_t(J)$  capturing intraday effects.

The daily scale factor  $s_t$  is not directly observable and therefore has to be approximated by some random variable which serves as its *proxy*. A good proxy should exhibit large correlation with  $s_t$ . The notion of a proxy can be formalized as follows:

**Definition (de Vilder and Visser, 2007).** Let  $H$  be a measurable functional  $D \rightarrow [0, \infty)$  defined on a linear subspace  $D$  of  $D[0,1]$  (Skorohod space of càdlàg functions on  $[0,1]$  left continuous in 1) and  $\Psi \in D$  a.s. Assume

(i)  $H$  is positively homogeneous, i.e.

$$H(a\Psi) = aH(\Psi), \quad a \in [0, \infty), \Psi \in D$$

(ii)  $H(\Psi) > 0$  a.s.

Then

1.  $H$  is called a proxy functional.
2. the random variable  $\Pi_t$  is a proxy for the daily scale factor  $s_t$  :

$$\Pi_t = H(s_t \Psi_t) = s_t H(\Psi_t) \equiv s_t V_t,$$

$V_t$  is called a nuisance proxy.

An additive measurement equation is obtained by taking logs:

$$\log \Pi_t = \log s_t + \log V_t = \log s_t + E \log V_t + (\log V_t - E \log V_t) \quad (6)$$

This equation decomposes the log proxy into a sum of log-volatility, a constant bias and measurement errors

$$u_t \equiv \log V_t - E \log V_t$$

A quality of a proxy is determined by the variance of the measurement errors  $I^2 \equiv \text{var}(u_t)$ .

In the sequel, we will focus on approximating  $\log s_t$  and for simplicity we will use the word proxy in the meaning of  $\log \Pi_t$ . Typical examples of such proxies include log range

$$\log \left( \sup_{J \in [0,1]} R_t(J) - \inf_{J \in [0,1]} R_t(J) \right) \quad (7)$$

and log absolute return

$$\log |R_t(1) - R_t(0)|. \quad (8)$$

The asymptotic distribution of both proxies has been studied in Alizadeh et al. (2002). Based on the result of Feller (1951), they find out that if  $\Psi$  is a driftless standardized Wiener process,  $\log(\sup \Psi(J) - \inf \Psi(J))$  can be well approximated by the normal distribution with mean 0.43 and variance 0.084. On the other hand, the variance of  $\log |\Psi(1) - \Psi(0)|$  is much higher (equal to  $p^2/8 \approx 1.23$ ) and the proxy is left skewed and leptokurtic. For this reason, the use of the log range instead of the log absolute return is highly recommended.

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 the distribution of the log range and on the first four moments in particular by a Monte Carlo simulation (using one million replications). The results are reproduced in Table 1 and Figure 1. The pattern is clear: reducing the number of observations during a trading day results in lower mean and higher variance. From practical point of view, if number of transactions per day is in the order of hundred or even less (and this is the case for less liquid markets as the Prague Stock Exchange), the bias induced by the proxy can be substantial. More importantly, if the number of transactions differs from day to day the bias becomes time-varying. We abstract from microstructure effects: for instance, bid-ask spread will operate in the opposite direction inflating the log range on average.

$N$	5	10	50	100	500	1000
mean	-0.115	0.097	0.300	0.340	0.401	0.415
variance	0.233	0.152	0.104	0.097	0.086	0.084
skewness	-0.457	-0.124	0.077	0.105	0.139	0.150
kurtosis	3.509	2.893	2.762	2.757	2.762	2.761

Table 1. Mean, variance, skewness and kurtosis of log range for a Wiener process with zero drift and unit variance observed  $N$  times during a unit period.

### 3. Kalman filter algorithm

It should be stressed that computing a proxy for volatility is not the same thing as extracting the latent volatility. In order to accomplish this task, we should specify the dynamics of the volatility process itself. We suppose that the log scale factor  $h_t \equiv \log s_t$  can be expressed as a sum of zero mean independent components  $h_1, h_2$  and the overall volatility level  $\bar{h}$ . The dynamics of each factor is governed by a first order autoregression:

$$h_{1t} = r_1 h_{1t-1} + \mathbf{h}_{1t} \quad (9a)$$

$$h_{2t} = r_2 h_{2t-1} + \mathbf{h}_{2t} \quad (9b)$$

with  $0 < r_1, r_2 < 1$  and  $\mathbf{h}_{1t}, \mathbf{h}_{2t}$  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.

$$y_{it} = g_i y_{i,t-1} + u_{it} \quad i = 1, 2 \quad (10)$$

with  $g_1 > g_2$  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  $corr(x_t, x_{t-k}) = w_1 g_1^k + w_2 g_2^k$ ,  $k = 1, 2, \dots$ , and the weights are given by  $w_i = \text{var}(y_{it}) / (\text{var}(y_{1t}) + \text{var}(y_{2t}))$ . Figure 2 shows that a choice of the autoregression coefficients in (10) has a profound impact on the rate of decay of the autocorrelation function of the composite process.

Transition equations (9a) and (9b) together with measurement equation (6) represent linear state space system whose general form reads

$$\mathbf{z}_{t+1} = \mathbf{F}\mathbf{z}_t + \mathbf{v}_{t+1} \quad (11a)$$

$$\mathbf{y}_t = \mathbf{b}(\mathbf{x}_t) + \mathbf{H}\mathbf{z}_t + \mathbf{w}_t \quad (11b)$$

where  $\mathbf{y}_t$  is a  $(n \times 1)$  vector of observed variables whose behaviour depends on a  $(r \times 1)$  vector of unobserved (state) variables  $\mathbf{z}_t$ ,  $\mathbf{b}(\mathbf{x}_t)$  is a  $(n \times 1)$  vector-valued function of exogenous variables collected in the vector  $\mathbf{x}_t$ ,  $\mathbf{F}$  and  $\mathbf{H}$  are  $(r \times r)$  and  $(n \times r)$  matrices. In other words, (11a) represents the transition equation whereas (11b) is the measurement equation.

If the conditional normality can be (at least approximatively) assumed, the extraction of the unobserved process is quite straightforward and relies on the well-known Kalman filter. The following exposition of the filter goes along the lines in Hamilton (1994, pp. 399). We assume that conditional on  $\mathbf{x}_t$  and  $\mathbf{F}_{t-1} \equiv (\mathbf{y}_{t-1}^T, \mathbf{y}_{t-2}^T, \dots, \mathbf{y}_1^T, \mathbf{x}_{t-1}^T, \mathbf{x}_{t-2}^T, \dots, \mathbf{x}_1^T)^T$  (i.e. data observed through date  $t-1$ ) the vector  $(\mathbf{v}_{t+1}^T, \mathbf{w}_t^T)^T$  is normally distributed with zero mean and covariance matrix given by



$$\begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}(\mathbf{x}_t) \end{bmatrix} \quad (12)$$

The exogenous vector  $\mathbf{x}_t$  has been introduced in order to deal with small sample effects which have influence on the bias and the variance of the proxy.

Further, suppose that  $\mathbf{z}_t | \mathbf{x}_t, F_{t-1} \sim N(\mathbf{z}_{t|t-1}, \mathbf{P}_{t|t-1})$ . Then the distribution of the vector  $(\mathbf{z}_t^T, \mathbf{y}_t^T)^T$  conditional on  $\mathbf{x}_t$  and  $F_{t-1}$  is normal with mean vector

$$(\mathbf{z}_{t|t-1}^T, \mathbf{b}^T(\mathbf{x}_t) + \mathbf{z}_{t|t-1}^T \mathbf{H}^T)^T \quad (13)$$

and covariance matrix

$$\begin{bmatrix} \mathbf{P}_{t|t-1} & \mathbf{P}_{t|t-1} \mathbf{H}^T \\ \mathbf{H} \mathbf{P}_{t|t-1} & \mathbf{V}_t \end{bmatrix} \quad (14)$$

with  $\mathbf{V}_t \equiv \mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{R}(\mathbf{x}_t)$ .

For updating the inference about the current value of the state variables  $\mathbf{z}_t$  as a new observation of  $\mathbf{y}_t$  becomes available we use the fact that

$$\mathbf{z}_t | \mathbf{y}_t, \mathbf{x}_t, F_{t-1} \equiv \mathbf{z}_t | F_t \sim N(\mathbf{z}_{t|t}, \mathbf{P}_{t|t}) \quad (15)$$

where

$$\mathbf{z}_{t|t} = \mathbf{z}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{H}^T \mathbf{V}_t^{-1} (\mathbf{y}_t - \mathbf{b}(\mathbf{x}_t) - \mathbf{H} \mathbf{z}_{t|t-1}) \quad (16)$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{H}^T \mathbf{V}_t^{-1} \mathbf{H} \mathbf{P}_{t|t-1} \quad (17)$$

The matrix  $\mathbf{K}_t \equiv \mathbf{P}_{t|t-1} \mathbf{H}^T \mathbf{V}_t^{-1}$  is often called the (Kalman) gain matrix or weight matrix and expresses the weights of innovations  $\mathbf{y}_t - \mathbf{b}(\mathbf{x}_t) - \mathbf{H} \mathbf{z}_{t|t-1}$  in producing the filtered estimate of the state variable  $\mathbf{z}_{t|t}$ . Roughly speaking, if the variance of the measurement noise is high, the weights attributed to the recent observation are relatively low and vice versa. In the standard Kalman filter case (with constant coefficients) under suitable conditions the sequence of matrices  $\{\mathbf{K}_t\}$  converges to a fixed, steady-state matrix (Proposition 13.1, Hamilton, 1994). If the covariance matrix of the measurement noise  $\mathbf{R}$  is time-dependent rather than constant then  $\{\mathbf{K}_t\}$  will fluctuate as well.

Subsequently, the conditional distribution of the state vector at time  $t+1$  given the observations known through date  $t$ , i.e.  $\mathbf{z}_{t+1} | F_t \sim N(\mathbf{z}_{t+1|t}, \mathbf{P}_{t+1|t})$  where

$$\mathbf{z}_{t+1|t} = \mathbf{F}\mathbf{z}_{t|t} \quad (18)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^T + \mathbf{Q} \quad (19)$$

The sample log likelihood function reads (after omitting constant terms)

$$-\frac{1}{2} \sum_{t=1}^T \log |\mathbf{V}_t| - \frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{b}(\mathbf{x}_t) - \mathbf{H}\mathbf{z}_{t|t-1})^T \mathbf{V}_t^{-1} (\mathbf{y}_t - \mathbf{b}(\mathbf{x}_t) - \mathbf{H}\mathbf{z}_{t|t-1}) \quad (20)$$

Similarly to the standard case with constant coefficients, estimates of the parameters contained in the matrices  $\mathbf{F}, \mathbf{H}, \mathbf{Q}$  can be obtained by maximizing (20) numerically.

#### 4. Empirical application

We apply Kalman filter algorithm fed both with asymptotic and finite sample values to the data from Prague Stock Exchange. We use daily high and low prices of ČEZ and Telefónica O2 C.R. stocks traded in the SPAD system. The log range proxy has been constructed from the daily highs  $H_t$  and lows  $L_t$  according to the formula  $\log(\log H_t - \log L_t)$ . The time span covers the period running from January 2, 2006 until December 28, 2007 giving rise to 501 observations. The average number of transactions per day for ČEZ and Telefónica O2 C.R. in the sample period was 231 and 101, respectively. All the calculations were carried out using MATLAB 7.1.

Sample autocorrelation functions of the data and QQ plots are depicted in Figure 3. The autocorrelation functions show a higher degree of persistency than it would correspond to a simple autoregressive process. Empirical moments of the log range are reported in Table 2. In comparison with their theoretical values, excess kurtosis is present in both series and moreover, Telefónica O2 C.R. exhibits a more pronounced skewness. Concerning the dependence between both series, the sample correlation coefficient amounts to 0.4017 and the positive correlation is clearly visible from the scatterplot (Figure 4).

	<i>mean</i>	<i>std deviation</i>	<i>skewness</i>	<i>kurtosis</i>
<i>ČEZ</i>	-3.7498	0.5691	0.0918	3.1730
<i>Telefónica O2 C.R.</i>	-4.1652	0.5870	0.3775	3.4424

Table 2. Unconditional moments of the observed log range

Now we proceed to estimating the two factor model in the state space form given by equations (6), (9a) and (9b). Considering the fact that the distribution of the log range is very similar to the normal distribution, we approximate the true density of the measurement errors  $u_t$  (see Equation (6)) by the Gaussian density. Therefore, the quasi maximum likelihood estimates of the unknown parameter vector  $(r_1, r_2, \bar{h}, \text{var}(h_1), \text{var}(h_2))^T$  has been obtained by numerically maximizing the likelihood function whose general form is given in Equation (20)

with respect to these parameters. As stated above, we employ two versions of the Kalman filter:

(i) Kalman filter with constant coefficients in the measurement equation corresponding to the asymptotic values, that is  $\mathbf{b}(\mathbf{x}_t) = 0.43$  and  $\mathbf{R}(\mathbf{x}_t) = 0.084$  in the notation of the Chapter 3,

(ii) Kalman filter with time-varying coefficients; the finite sample mean and variance have been computed by means of Monte Carlo simulation (cf. Table 1).

Estimation results are shown in Table 3 and observed log range together with filtered extractions of the daily log volatility (i.e. sum of both factors plus the constant term  $\bar{h}$ ) are depicted in Figure 5. In accordance with findings of Alizadeh et al. (2002), there exists a strong evidence that the volatility process can be meaningfully decomposed into one highly persistent factor and another quickly mean-reverting factor.

In summary, providing the algorithm with finite sample mean and variance of the measurement errors has two effects: the most important one is the reduction of the extraction bias (note that using asymptotic values in the Kalman algorithm would result in an *underestimation* of the overall volatility level). Second, the weights contained in the Kalman gain vector are no longer constant: they are smaller when the number of transactions is low (due to higher variance of the measurement error). However, this effect is quite weak and the weights stabilize around 0.21 and 0.34 for the first and second factor, respectively (see Figures 6a, 6b).

	<i>ČEZ</i>		<i>Telefónica O2 C.R.</i>	
	<i>asymptotic</i>	<i>finite sample</i>	<i>asymptotic</i>	<i>finite sample</i>
$r_1$	0.9358	0.9361	0.8968	0.9094
$r_2$	0.425	0.4545	0.4266	0.514
$\bar{h}$	-4.171	-4.1139	-4.5917	-4.4924
$\text{var}(\mathbf{h}_1)$	0.0168	0.0159	0.0254	0.0191
$\text{var}(\mathbf{h}_2)$	0.0848	0.073	0.1045	0.085

Table 3. *Quasi-maximum likelihood estimates of the two-factor model*

In order to obtain a better interpretation of the estimated coefficients, 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 4). Each factor explains about one half of the variance of the log volatility process.

	<i>ČEZ</i>		<i>Telefónica O2 C.R.</i>	
	<i>asymptotic</i>	<i>finite sample</i>	<i>asymptotic</i>	<i>finite sample</i>
<i>first factor</i>	0.1352	0.1285	0.1298	0.1104
<i>second factor</i>	0.1035	0.092	0.1277	0.1155
<i>total variance</i>	0.2387	0.2205	0.2575	0.2259

*Table 4. Decomposition of the total variance of the estimated log-volatility into variances of individual factors*

Finally, we performed a Monte Carlo experiment to investigate the bias reduction obtained when finite sample values are used. We generated series of the length of 501 observations with the log scale process with parameters equal to the rightmost column of Table 3 (corresponding to Telefónica O2 C.R.). These volatility factors were used to scale the daily Wiener process. Further, the number of observations in a trading day was set to the true number of transactions for the given day. Then, the log range was computed and used as an input for both asymptotic and finite sample versions of the Kalman filter (with true parameters fed into the filter). Finally, the true log scale factor values and filtered estimates are compared by computing the mean extraction error and the root mean squared (RMS) extraction error. The simulation was performed with 10000 replications and average values of mean and RMS extraction errors are reported in Table 5. The result is that using the finite sample version of the filter practically eliminates the bias leading to a lower average value of the RMS extraction error.

<i>mean extraction error</i>		<i>root mean squared extraction error</i>	
<i>asymptotic</i>	<i>finite sample</i>	<i>asymptotic</i>	<i>finite sample</i>
-0.083	0.0003	0.2544	0.2389

*Table 5. Mean extraction error and root mean squared extraction error (average for 10000 replications)*

## 5. Conclusion

We have shown that the use of standard Kalman filter fed with asymptotic values in the measurement equation advocated for instance in Alizadeh et al. (2002) can be inappropriate for less liquid shares like those one traded on the Prague Stock Exchange. Therefore, we suggest a modified procedure which takes into account the small sample distribution of the volatility proxy. This modification leads to more precise extractions of the log scale factor due to elimination of the finite sample bias. However, we have relied on the (possibly unrealistic) assumption of the driftless Wiener process during a trading day. It would be interesting to study the behaviour of the log range if this assumption were relaxed. Nevertheless, these issues are 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. (2006): *Realized range-based estimation of integrated variance*, Journal of Econometrics (forthcoming).
- DE VILDER, R.G. – VISSER, M.P. (2007): *Volatility Proxies for Discrete Time Models*, MPRA Paper 4917, University Library of Munich, Germany.
- 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

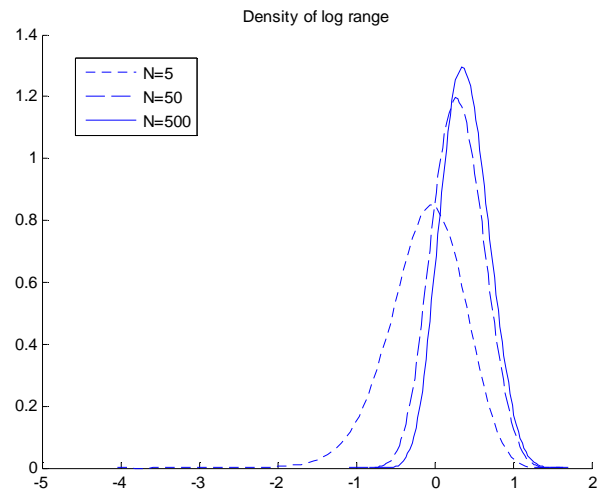


Figure 1. Density of the log range with 5, 50 and 500 observations during a unit interval.

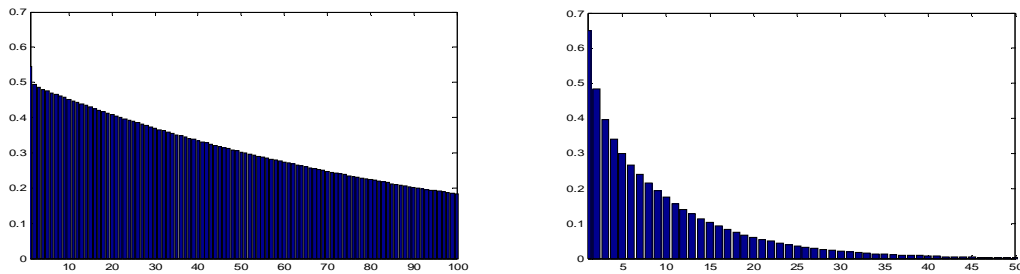
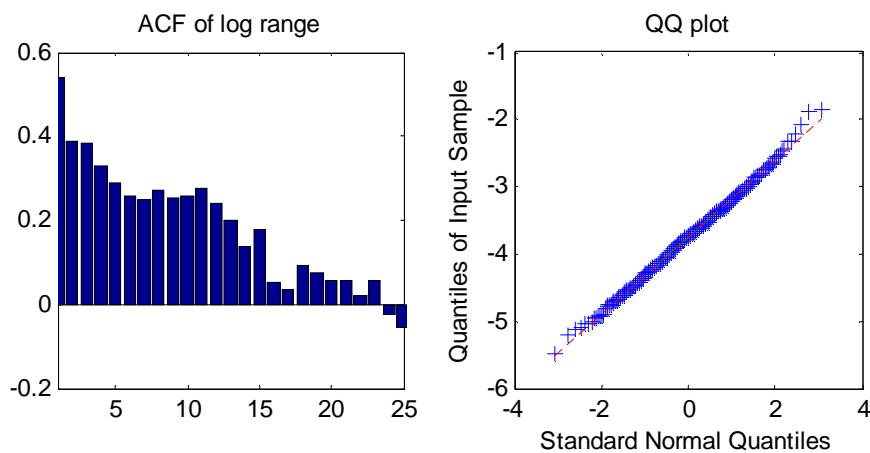


Figure 2. Two examples of the autocorrelation function for the two-component model for equally weighted components (i.e.  $w_1 = w_2 = 0.5$ ) with  $g_1 = 0.99, g_2 = 0.1$  (left) and  $g_1 = 0.9, g_2 = 0.4$  (right).



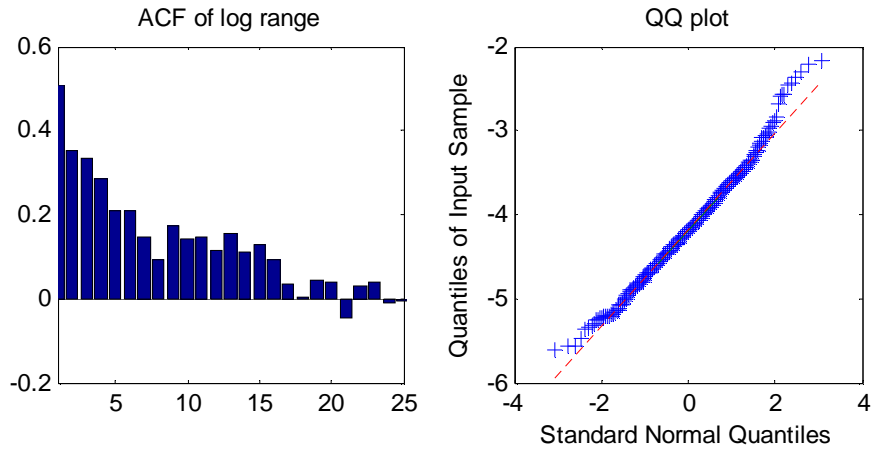


Figure 3. Sample autocorrelation function and QQ plots for ČEZ (top) and Telefonica O2 C.R. (bottom).

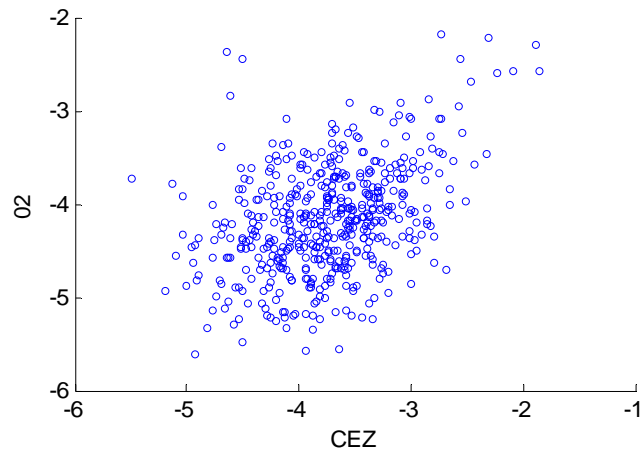
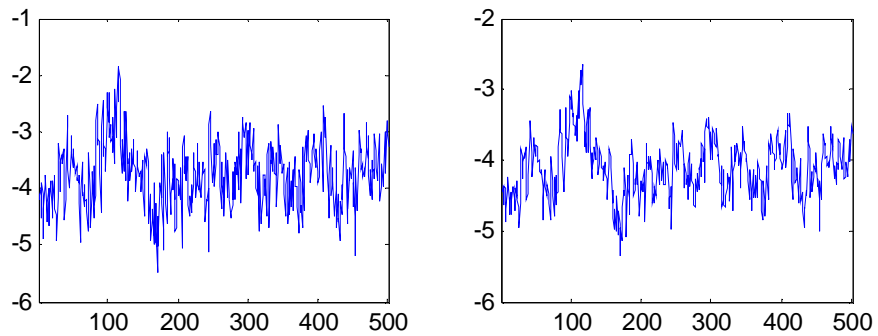


Figure 4. Scatterplot for observed log range (x-axis: ČEZ, y-axis: Telefónica O2 C.R.).





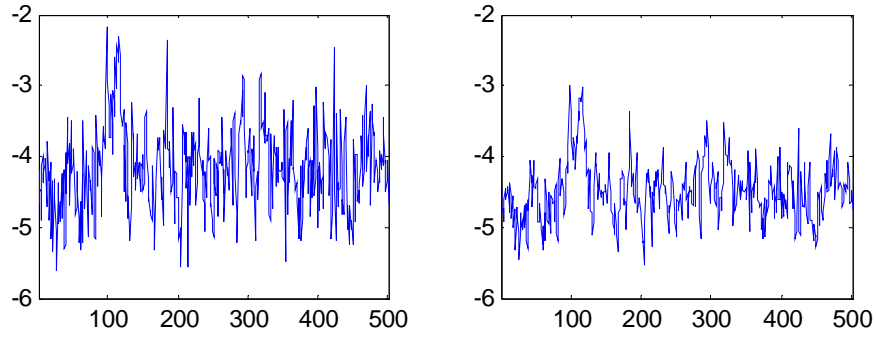


Figure 5. Observed log range and corresponding filtered extractions of the log volatility for ČEZ (top) and Telefonica O2 C.R. (bottom).

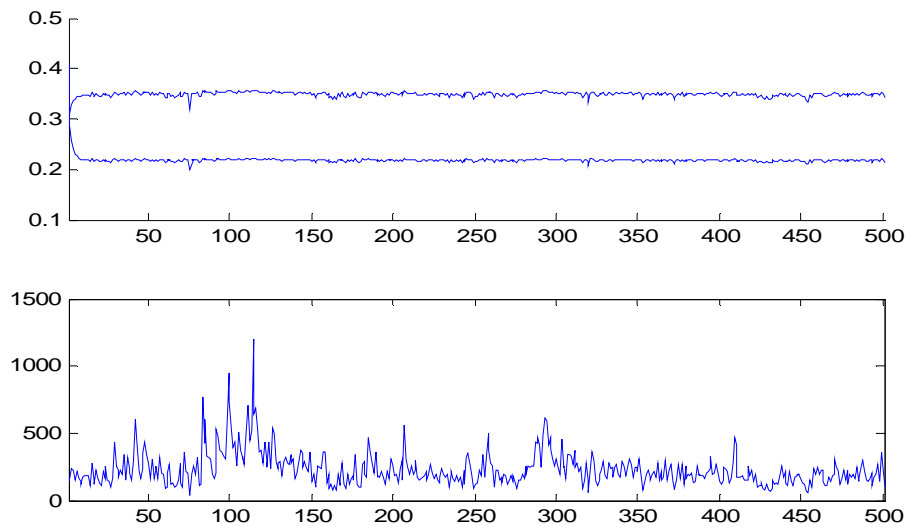
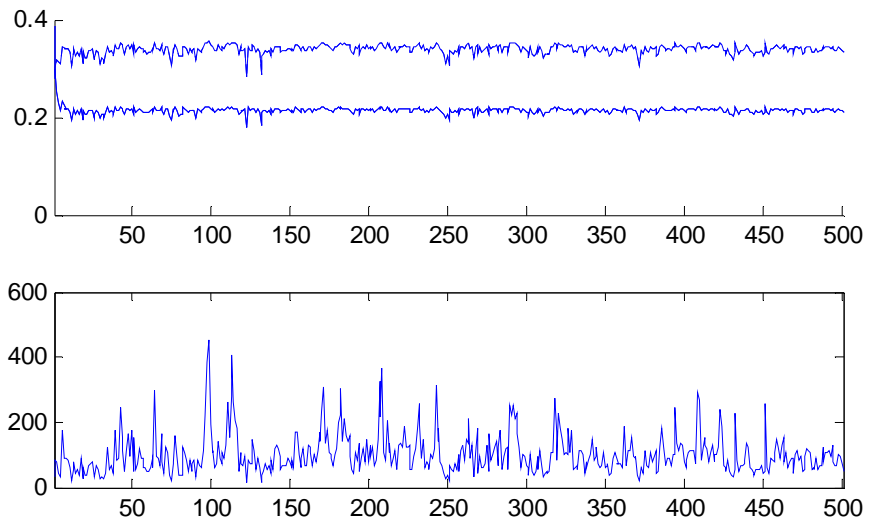


Figure 6a. Kalman gain coefficients for both factors (top) and number of transactions (bottom) for ČEZ.



*Figure 6b. Kalman gain coefficients for both factors (top) and number of transactions (bottom) for Telefónica O2 C.R.*

# IES Working Paper Series

2007

1. Roman Horváth : *Estimating Time-Varying Policy Neutral Rate in Real Time*
2. Filip Žikeš : *Dependence Structure and Portfolio Diversification on Central European Stock Markets*
3. Martin Gregor : *The Pros and Cons of Banking Socialism*
4. František Turnovec : *Dochází k reálné diferenciaci ekonomických vysokoškolských vzdělávacích institucí na výzkumně zaměřené a výukově zaměřené?*
5. Jan Ámos Víšek : *The Instrumental Weighted Variables. Part I. Consistency*
6. Jan Ámos Víšek : *The Instrumental Weighted Variables. Part II.  $\sqrt{n}$  - consistency*
7. Jan Ámos Víšek : *The Instrumental Weighted Variables. Part III. Asymptotic Representation*
8. Adam Geršl : *Foreign Banks, Foreign Lending and Cross-Border Contagion: Evidence from the BIS Data*
9. Miloslav Vošvrda, Jan Kodera : *Goodwin's Predator-Prey Model with Endogenous Technological Progress*
10. Michal Bauer, Julie Chytilová : *Does Education Matter in Patience Formation? Evidence from Ugandan Villages*
11. Petr Jakubík : *Credit Risk in the Czech Economy*
12. Kamila Fialová : *Minimální mzda: vývoj a ekonomické souvislosti v České republice*
13. Martina Mysíková : *Trh práce žen: Gender pay gap a jeho determinanty*
14. Ondřej Schneider : *The EU Budget Dispute – A Blessing in Disguise?*
15. Jan Zápál : *Cyclical Bias in Government Spending: Evidence from New EU Member Countries*
16. Alexis Derviz : *Modeling Electronic FX Brokerage as a Fast Order-Driven Market under Heterogeneous Private Values and Information*
17. Martin Gregor : *Rozpočtová pravidla a rozpočtový proces: teorie, empirie a realita České republiky*
18. Radka Štiková : *Modely politického cyklu a jejich testování na podmínkách ČR*
19. Martin Gregor, Lenka Gregorová : *Inefficient centralization of imperfect complements*
20. Karel Janda : *Instituce státní úvěrové podpory v České republice*
21. Martin Gregor : *Markets vs. Politics, Correcting Erroneous Beliefs Differently*
22. Ian Babetskii, Fabrizio Coricelli, Roman Horváth : *Measuring and Explaining Inflation Persistence: Disaggregate Evidence on the Czech Republic*
23. Milan Matejašák, Petr Teplý : *Regulation of Bank Capital and Behavior of Banks: Assessing the US and the EU-15 Region Banks in the 2000-2005 Period*
24. Julie Chytilová, Michal Mejstřík : *European Social Models and Growth: Where are the Eastern European countries heading?*
25. Mattias Hamberg, Jiri Novak : *On the importance of clean accounting measures for the tests of stock market efficiency*
26. Magdalena Morgese Borys, Roman Horváth : *The Effects of Monetary Policy in the Czech Republic: An Empirical Study*
27. Kamila Fialová, Ondřej Schneider : *Labour Market Institutions and Their Contribution to Labour Market Performance in the New EU Member Countries*

28. Petr Švarc, Natálie Švarcová : *The Impact of Social and Tax Policies on Families with Children: Comparative Study of the Czech Republic, Hungary, Poland and Slovakia*
29. Petr Jakubík : *Exekuce, bankroty a jejich makroekonomické determinanty*
30. Ibrahim L. Awad : *Towards Measurement of Political Pressure on Central Banks in the Emerging Market Economies: The Case of the Central Bank of Egypt*
31. Tomáš Havránek : *Nabídka pobídek pro zahraniční investory: Soutěž o FDI v rámci oligopolu*

2008

1. Irena Jindrichovska, Pavel Körner : *Determinants of corporate financing decisions: a survey evidence from Czech firms*
2. Petr Jakubík, Jaroslav Heřmánek : *Stress testing of the Czech banking sector*
3. Adam Geršl : *Performance and financing of the corporate sector: the role of foreign direct investment*
4. Jiří Witzany : *Valuation of Convexity Related Derivatives*
5. Tomáš Richter : *Použití (mikro)ekonomické metodologie při tvorbě a interpretaci soukromého práva*
6. František Turnovec : *Duality of Power in the European Parliament*
7. Natalie Svarciva, Petr Svarc : *Technology adoption and herding behavior in complex social networks*
8. Tomáš Havránek, Zuzana Iršová : *Intra-Industry Spillovers from Inward FDI: A Meta-Regression Analysis*
9. Libor Dušek, Juraj Kopecsni : *Policy Risk in Action: Pension Reforms and Social Security Wealth in Hungary, Czech Republic, and Slovakia*
10. Alexandr Kuchynka : *Volatility extraction using the Kalman filter*

All papers can be downloaded at: <http://ies.fsv.cuni.cz>

