Abstract: This paper deals with a conditional volatility GARCH model and model based on realized volatility which is able to account for the main empirical features observed in data in financial markets. Inspired by well-known Heterogeneous Market Hypothesis and by the asymmetric behavior of volatility between long and short time horizons, we used an additive cascade of different volatility components generated by the actions of different types of market participants. This additive volatility cascade leads to a simple AR-type model in the realized volatility with the feature of considering volatilities realized over different time horizons. HAR-RV model successfully achieves the purpose of reproducing the main empirical features of volatility like long memory, fat tails, and self-similarity in a very simple and parsimoniously way. The aim of this paper is to compare estimates got by simple AR(1)-GARCH(1, 1) model and HAR-RV model using data from the Czech stock market represented by PX index. In our paper we work with daily, weekly and monthly returns of mentioned stock index. Preliminary results on the estimation and forecast of the HAR-RV model on PX stock index data show remarkably good in-sample forecasting performance which steadily and substantially outperforms those of standard models represented by AR(1)-GARCH(1, 1) model. There will be also very fruitful to compare results estimated by mentioned models in different time periods. We especially mention an impact of the global financial crisis on Czech stock market volatility. Therefore, in this paper we will investigate pre-crisis, crisis and post-crisis periods.

Key-Words: autoregressive conditional heteroskedasticity, forecast, HAR-RV, heterogeneous market hypothesis, model estimation, realized volatility, stock market.

1 Introduction
Volatility modelling was born 30 years ago, and is still and will remain one of the most active research topics of financial econometrics. The development of econometric models of volatility has gone along with their application in academia and progressive use in the financial industry.

Many recent investigations have considered volatility as an unobservable variable and therefore used a fully specified conditional mean and conditional variance model to estimate and analyse latent volatility. This approach led to all kinds of ARCH class and stochastic volatility models. One can see some weaknesses of that kind of models. The estimation procedure is often not trivial especially in stochastic volatility models and they are not able to replicate empirical features of financial data.

In order to simplify volatility estimation there would be used some observable proxy for the latent volatility. A standard proxy for ex-post volatility has been squared returns. This proxy has been named realized volatility (RV) by [3]. Our goal is to obtain a conditional volatility model based on realized volatility which is able to measure main empirical features and is easy to estimate, and compare in-sample forecasting performance with simple AR-GARCH model using data from Czech stock market.

2 Theoretical background
In this chapter there will be introduced Heterogeneous Market Hypothesis (HMH) which is key to further understanding the aim of our paper.

Moreover, some realized volatility measures will be defined as we need it for volatility model formulation.
2.1 Heterogeneous Market Hypothesis

Heterogeneous Market Hypothesis presented by [11] recognizes the presence of heterogeneity of the traders. This specific view on financial markets can be related with the Fractal Market Hypothesis of [13] and the Interacting Agent View of [10]. The idea of a presence of multiple components in the volatility process has been also suggested by [1] in the Hypothesis of Mixture of Distributions. This view on the multi-component structure stems from the heterogeneous nature of the information arrivals rather than from the heterogeneity of the agents.

According the Heterogeneous Market Hypothesis one can explain the empirical observation of a strong positive correlation between volatility and market presence. In heterogeneous markets, different actors are likely to settle for different prices and decide to execute their transactions in different market situations, hence they create volatility. The heterogeneity of the agents may be caused by various reasons: differences in degree of information, prior belief, temporal horizons, geographical location, institutional constraints, and risk profile and so on.

In this paper we concentrate on the heterogeneity which originates from the difference in investment time horizon. Financial market is usually composed by participants having a large spectrum of dealing prices and decide to execute their transactions in different market situations, hence they create volatility. The heterogeneity of the agents may be caused by various reasons: differences in degree of information, prior belief, temporal horizons, geographical location, institutional constraints, and risk profile and so on.

In this paper we concentrate on the heterogeneity which originates from the difference in investment time horizon. Financial market is usually composed by participants having a large spectrum of dealing prices and decide to execute their transactions in different market situations, hence they create volatility. The heterogeneity of the agents may be caused by various reasons: differences in degree of information, prior belief, temporal horizons, geographical location, institutional constraints, and risk profile and so on.

2.2 How to measure a realized volatility

According to [5] the standard definition for an equally spaced returns series of the realized volatility over a time interval of one trading day is:

\[ RV_{t}^{(d)} = \frac{1}{|\Delta|} \sum_{j=0}^{|\Delta|-2} r_{t-j|\Delta}, \]  

where \( \Delta = \frac{1}{M} \), and \( r_{t-j|\Delta} = p(t-j\Delta) - p(t-(j+1)\Delta) \) defines continuously compounded \( \Delta \)-frequency returns, that is, intraday returns sampled at time interval \( \Delta \). Under this assumption, the ex-post realized volatility is an unbiased volatility estimator. Definition of realized volatility involves two time parameters: the intraday return interval \( \Delta \) and the aggregation period \( 1d \). In the following we will also consider latent integrated volatility and realized volatility viewed over different time horizons longer than one day. These multi-period volatilities will simply be normalized sums of the one-period volatilities. For example, according [5] in our notation, a weekly realized volatility at time \( t \) will be given by

\[ RV_{t}^{(w)} = \frac{1}{(1\sigma)^{1/2}} \left( RV_{t-1d}^{d} + RV_{t-2d}^{d} + \ldots + RV_{t-1w}^{d} \right), \]  

where \( 1\sigma = 5d \) indicate a time interval of one week. In particular we will make use of weekly and monthly integrated and realized volatility. Indicating the aggregation period, the notation for weekly quantity will be respectively \( \sigma_{t}^{w} \) and \( RV_{t}^{w} \) while a monthly aggregation will be denoted as \( \sigma_{t}^{m} \) and \( RV_{t}^{m} \).

2.3 Model formulation

In the seminal papers of [4] and [6], GARCH or generalized autoregressive conditional heteroskedasticity models have become a standard tool in modeling the conditional variances of the returns from financial time series data. The popularity of these models lies in their compatibility with some major stylized facts for asset returns, the existence of efficient statistical methods for estimating model parameters, and the availability of useful volatility forecasts.

Standard GARCH class or stochastic volatility models are not able to model all the features typical for high frequency time series. Summarizing the main characteristics of the financial time series are: fat tails, long memory in the volatility, high level of skewness and kurtosis which decrease with temporal aggregation.

Normally, the AR(1)-GARCH(1, 1) model taking the form of discrete data and considering heteroskedasticity may be according to [4] defined as follows:

\[ r_{t} = \alpha_{0} + \alpha_{t} r_{t-1} + e_{t}, \]  

\[ e_{t} = \xi_{t} \sigma_{t}, \]  

\[ \xi_{t} \sim N(0,1), \]  

\[ \sigma_{t}^{2} = \beta_{0} + \beta_{1} e_{t-1}^{2} + \beta_{2} \sigma_{t-1}^{2}, \]  

where \( r_{t} \) is the return of stock index estimated here, which shows the volume of information inflow in the market in stock prices return. The GARCH heteroskedasticity is defined by equation (6).
\( \beta_0, \beta_1, \beta_2 \) are constants satisfying the conditions of \( \beta_0 > 0, \beta_1 + \beta_2 < 1 \).

For instance, in the GARCH(1, 1) model, there is a stringent trade-off between the possibility of having sharp changes in the short term volatility represented by high value of the parameters \( \beta_1 + \beta_2 \) and the ability to capture the long memory behaviour of volatility through high values of \( \beta_0 + \beta_2 \). Moreover, even with high value of \( \beta_1 + \beta_2 < 1 \), GARCH models are subject to exponential decline in the autocorrelation, which is at odds with the observed hyperbolic decline observed in the data. Hence the recent interest in long memory process.

Taking into consideration the above properties for high frequency time series, it is possible to formulate a multicomponent volatility model with an additive hierarchical structure which can leads to quite simple model of the realized volatility. The basic idea stems from mentioned Heterogeneous Market Hypothesis.

In our assumption we determine the market dynamics by the behaviour of the dealers. Hence the high frequency return process is determined by the highest frequency volatility component in the cascade with \( \sigma^d_t = \sigma^d \) the daily integrated volatility. Then the return process is

\[
 r_t = \sigma^d_t \varepsilon_t ,
\]

with \( \varepsilon_t \sim \text{NID}(0,1) \).

The model for the unobserved partial volatility processes \( \sigma^l(1) \) at each level of the time scale is assumed to be a function of the past realized volatility experienced at the same time scale and of the expectation of the next period values of the longer term partial volatilities.

According to [5] we proposed a very simple time representation of the cascade model of realized volatility:

\[
 RV^d_{t, id} = c + \beta^d RV^d_{t-1} + \beta^u RV^u_{t, id} + \beta^m RV^m_{t, id} + \omega_{t, id} ,
\]

Equation (8) has a simple autoregressive structure in the realized volatility. In general, denoting \( l \) and \( h \) respectively the lowest and highest frequency in the cascade (8) is an according \( AR(1 \frac{1}{h}) \) model reparametrized in a parsimoniously way by imposing economically meaningful restrictions. In other words (8) is an AR-type process but with the feature of considering volatilities realized over different interval sizes; it could then be titled as an Heterogeneous Autoregressive model for the Realized Volatility (HAR-RV). For more details, see [8] and [9].

In spite of its simplicity the proposed model is able to produce rich dynamics for the returns and the volatility which closely resemble the empirical ones. This dynamic is generated by the heterogeneous reaction of the different market components to a given price change which in turns affect the future size of price changes. This causes a complex process by which the markets react to its own price history with different reaction times. Thus market volatilities feed on themselves.

3 Empirical Analysis

In this chapter the data used for analysis will be described, their empirical features depicted, defined models estimated and forecasting performance evaluated.

3.1 The data

Empirical analysis is performed on the PX index in period from January 2004 till March 2012, it includes total of 2225 trading days. This period was chosen purposely, to investigate changes of the Czech equity market volatility during time with a special emphasis on the resolution of behaviour in the time before and during the global financial crisis in 2008-2009. For every trading day we have the information on its open and close price as well as the highest and the lowest, recorded at 5-minutes frequency.

The PX is a weighted index containing the most liquid titles with weights changing according to the market capitalization. At the present time, the actual number of the basic issues is variable.

The daily returns \( r_t \) at time \( t \) were for the purpose of AR-GARCH models defined in the logarithm of PX index \( p \), that is, \( r_t = \log( p_t - p_{t-1} ) \). Visual inspection of the plot of daily values and returns series of PX index proved very useful, for details see Fig. 1 and Fig. 2.

As it has been empirically confirmed, crises are not devoted to developed markets only. Emerging markets including Czech Republic aren’t excluded from this rule, and may face such instability sometime, for instance see [7], [12] or [15].

Following the spread of bad news about U.S financial crisis the Central European equity markets, Czech one included, have seen a more than 60 per cent decline in selected index, please see Fig. 1. This happened primarily due to the withdrawal by foreign portfolio investors between September and...
December 2008 and its psychological impact on national investors.

It can be seen from Fig. 2 that return fluctuates around mean value that is close to zero. Volatility is low for certain time periods and high for other periods. The movements are in the positive and negative territory and larger fluctuations tend to cluster together separated by periods of relative calm. The volatility of PX index was highest in 2008. Thus Fig. 2 show volatility clustering where large returns tend to be followed by small returns leading to continuous periods of volatility and stability. Volatility clustering implies a strong autocorrelation in squared return.

Since the volatility was highest in 2008 when the values of PX index reached the minimum values in investigated period we divided the basic period 2004-2012 into three testing period. First period was defined from 2004 to the end of July 2007, the second one started at the beginning of June 2007 and finished by the end of March 2009; the last period was defined from March 2009 to the end of March 2012.

Our goal is to analyse and compare the estimates of HAR-RV and AR(1)-GARCH (1, 1) models of equity volatility in particular periods and evaluate in-sample forecasting performance of those models.

### 3.2 Empirical features of PX returns

Our results for the empirical study of PX index are in harmony with those already found for other stock indexes, see [14]. Table 1 shows selected descriptive statistics of raw and RV-standardised daily PX returns in particular periods.

<table>
<thead>
<tr>
<th></th>
<th>Pre-crisis</th>
<th>Crisis</th>
<th>Post-crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw</td>
<td>RV</td>
<td>Raw</td>
</tr>
<tr>
<td>Mean</td>
<td>0,001</td>
<td>0,019</td>
<td>-0,00</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>0,011</td>
<td>1,076</td>
<td>0,025</td>
</tr>
<tr>
<td>Kurtos.</td>
<td>8,604</td>
<td>2,967</td>
<td>6,529</td>
</tr>
<tr>
<td>J-B test</td>
<td>1254</td>
<td>0,243</td>
<td>1492</td>
</tr>
<tr>
<td>Probab.</td>
<td>0,000</td>
<td>0,887</td>
<td>0,000</td>
</tr>
</tbody>
</table>

Table 1: Comparison of raw and RV-standardized PX returns distributions

Summarizing the main empirical features of the PX index data in general are:

1. Long range dependence: although the autocorrelation of the returns is not significant at all measures, on the other hand the autocorrelation of the square and absolute values of returns displays very strong persistence even at very long lags. This persistence means the autocorrelation of realized volatilities where the long memory of the process becomes even more clear and visible. The autocorrelation of PX index realized volatility remains significant for at least 2 months.

2. Distribution properties of returns and realized volatility: the unconditional distributions of realized variances possess high level of kurtosis which decrease with temporal aggregation but remain far from normal even at monthly scale. Realized volatility is instead much closer to normal distributions.

3. Fat tails: the kurtosis value of the returns is much higher than is usual for normal distribution at daily frequency and clearly tends to decrease as the return length increases. Thus returns have leptokurtic distribution with shapes depending on the time scale and presenting a quite slow convergence of the Central Limit Theorem to the normal distribution.

### 3.3 Estimation results

The purpose of this section is to empirically analyse the performance of models proposed in the Section 2. As a first step we estimated AR(1)-GARCH(1, 1)
model for all the three investigated periods according (3)-(6). The results are shown in Table 2. Statistically significant parameters at 5% significance level are marked in bold.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Pre-crisis</th>
<th>Crisis</th>
<th>Post-crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>0.0016</td>
<td>5.49</td>
<td>-0.001</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0264</td>
<td>1.72</td>
<td>0.0140</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.0001</td>
<td>2.64</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.1076</td>
<td>4.79</td>
<td>0.1856</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.8341</td>
<td>21.74</td>
<td>0.8007</td>
</tr>
</tbody>
</table>

Table 2: AR(1)-GARCH(1, 1) model for PX index

In the following step, we considered HAR-RV model in the terms of (8) and then easily estimate its parameters $\beta(\cdot)$ by applying simple linear regression. Standard OLS regression is consistent and normally distributed.

Since the uses of for instance intraday measures of realized volatility poses problems either of measurement accuracy and strong intraday seasonality, we choose to estimate the variance equation (8) at daily frequency. Table 3 reports the results of the estimation of the HAR-RV model for nearly nine years of PX daily realized volatilities.

It seems that realized volatilities aggregated over different time horizons are reasonable proxy for volatilities generated by the corresponding market components.

Simple OLS regression is a direct estimate of the market participants weights, that is, a readily evaluation of the contribution of each market component to the overall market activity. Moreover, if a moving window regression is performed, a time series evolution of such weights is easily achieved as well.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Pre-crisis</th>
<th>Crisis</th>
<th>Post-crisis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.0185</td>
<td>3.76</td>
<td>0.0477</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.3272</td>
<td>11.65</td>
<td>0.2461</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.2176</td>
<td>5.52</td>
<td>0.0656</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.1422</td>
<td>3.81</td>
<td>0.1105</td>
</tr>
</tbody>
</table>

Table 3: HAR-RV model for PX index

As a next step of our analysis, logically a comparison of estimated models will be provided. According results achieved in Table 2 and Table 3 it can be said in case of HAR-RV model of volatility estimated parameter are statistically significant in all investigated periods whereas the parameter estimates for AR(1)-GARCH models are not always statistically significant.

However, a real quality of the estimate can be verified on the basis of the forecasting performance.

### 3.4 Forecasting performance

The in-sample forecast of the models defined in the Section 2 is shown in Table 4. These forecasts are obtained by first estimate the parameters of the models on the full sample and then comparing with volatility measured as a standard deviation of series.

For comparison purposes these two models are added: the standard AR(1)-GARCH(1, 1) and HAR-RV model of the realized volatility. In Table 4 the forecasting performance are evaluated on the basis of Root Mean Square Errors (RMSE), Mean Absolut Error (MAE) and Theil Inequality coefficient.

<table>
<thead>
<tr>
<th>Pre-crisis</th>
<th>RMSE</th>
<th>MAE</th>
<th>Theil Inequality coeff. x 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>3.297</td>
<td>2.628</td>
<td>12.659</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>2.822</td>
<td>2.068</td>
<td>10.768</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crisis</th>
<th>RMSE</th>
<th>MAE</th>
<th>Theil Inequality coeff. x 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>4.764</td>
<td>3.649</td>
<td>15.064</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>4.212</td>
<td>3.176</td>
<td>14.271</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post-crisis</th>
<th>RMSE</th>
<th>MAE</th>
<th>Theil Inequality coeff. x 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>3.569</td>
<td>2.982</td>
<td>13.412</td>
</tr>
<tr>
<td>HAR-RV</td>
<td>3.043</td>
<td>2.254</td>
<td>11.422</td>
</tr>
</tbody>
</table>

Table 4: Comparison of the in-sample performances of the 1 day ahead forecast of AR(1)-GARCH(1, 1) and HAR-RV models for PX index

Based on the results shown in Table 4, it seems that in-sample performance the HAR-RV model steadily outperform the AR(1)-GARCH model at all three investigated periods considered. It is noteworthy noticing that the performance of the HAR-RV model has appeared at daily horizon.

The reason for that is that AR(1)-GARCH model has a memory which is too short compared to the forecasting horizon or they adjust too late to the movements of the realized volatility. What is surprising, the ability of the HAR-RV model to achieve good results with only a few parameters.
even in crisis period. Moreover, our results are not in harmony with other authors; see [2] for instance.

4 Conclusion
The volatility models inspired by the Heterogeneous Market Hypothesis leads to a simple autoregressive type model of the realized volatility which has the feature of considering volatilities realized over different interval sizes. In our paper we used Heterogeneous Autoregressive model of the Realized Volatility. The HAR-RV model seems to successfully achieves the purpose of modelling, the long memory behaviour of volatility in a very simple and parsimoniously way. In spite of the simplicity of its structure and estimation, the HAR-RV model shows remarkably good in-sample forecasting performance which in general beats AR-GARCH model.

By projecting a dynamic process on its own past values aggregated over different time horizons, the HAR - RV model is a general and flexible approach to fit the autocorrelation function of any persistent process in a very simple and tractable way.

In our paper we have briefly surveyed the nature, construction, and properties of HAR-RV class models for realized volatility estimation and forecasting performance using data from the Czech stock market in the period 2004-2012.

Acknowledgements
The support provided by VSB-TU Ostrava under the SGS project SP2012/172 is kindly announced.

References: