Report on **Bachelor / Master Thesis**

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

Student:	**Bc. Eva Brabcová**
Advisor:	**PhDr. Jozef Baruník**
Title of the thesis:	**Volatility Modeling: Evidence from CEE Stock Markets**

***OVERALL ASSESSMENT** (provided in English, Czech, or Slovak):*

The thesis presents an application of a still relatively new yet powerful methodology based on the use of high-frequency data to the Central European (CE) equity index data. In particular, the author constructs realized volatility estimates and applies the HAR model of Corsi (2009), along with several of its extensions, to investigate the relative performance of the latter models in forecasting the realized volatility. The analysis is supplemented with the application of the theory of wavelet transforms with the aim to investigate the presence of (realized) volatility comovements between the CE stock markets.

General Description

Following a general introduction (Chapter 1), in Chapter 2 the author provides a detailed overview of the quadratic variation theory that lends theoretical justification to using a simple sum of squared high-frequency (or, intraday) returns as an observable proxy for the latent volatility. The corresponding nonparametric estimator, called realized volatility, is then used as the main variable in the Heterogenous Autoregressive model of the Realized Volatility (HAR-RV), described in the same chapter. Although not directly mentioned by the author, it remains implicit in the subsequent analysis why the HAR-RV model is used: relatively simple by construction, the HAR-RV model has been previously shown to successfully reproduce the main empirical features of financial data as well as to provide remarkably good out of sample forecasting performance relative to other models.

An important part of Chapter 2 is devoted to the description of various extensions of the HAR-RV model, including those that explicitly account for the volatility of realized volatility (HAR-RV-GARCH), and jumps (e.g., HAR-RV-J, HAR-RV-TCJ). The quadratic variation theory that explicitly accounts for the jump component in the underlying price process is also explained in this chapter, including the means to detect jumps in the first place.

Chapter 3 describes the set of high-frequency data employed by the author to construct the daily measures of realized volatility. In addition, the author examines the issues related to the effects of market microstructure on realized volatility estimator as well as the construction of the regression variables used in the HAR-RV(-X) models. Analysis of the high-frequency data leads to Chapter 4, where the HAR-RV(-X) models are estimated and the results discussed.

Finally, Chapter 5 extends the previous analysis by applying the theory of wavelet transforms to the realized volatility estimators. In particular, the author aims to identify the key periods and frequencies at which the pairs of stock market indices show high level of coherence.

Main Contribution

The main contribution of the thesis is a thorough analysis of both in-sample and out-of-sample forecasting performance of a set of HAR-RV(-X) models based on high-frequency data from three major Central European stock markets (Prague, Budapest and Warsaw). The analysis is all the more interesting as the HAR-RV model has the feature of considering volatilities realized over different time intervals and hence makes it possible to naturally relate the behavior of market agents with different reaction times to the news (→ different trading horizons).

Report on **Bachelor / Master Thesis**

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

Student:	**Bc. Eva Brabcová**
Advisor:	**PhDr. Jozef Baruník**
Title of the thesis:	**Volatility Modeling: Evidence from CEE Stock Markets**

Similarly, the wavelet analysis provides the first empirical findings from applying the wavelet transforms to the intraday Central European equity market data. An interesting finding is that there exists a relatively strong relationship between the realized volatilities of CE equity markets in the frequencies beyond 60 days.

General Comments

The thesis has a well-defined structure and, except for certain imprecisions in the methodology and the usual typographical/language errors (see below), is also well written. In addition, the author provides a very nice and detailed description of her findings; the observations that she makes are to the point, without any unnecessary comments being made. This, along with the ability of the author to have successfully applied her computational skills to perform the necessary/required calculations are also the main strengths of the thesis.

There are several problems that I find with the thesis, as summarized in the following paragraphs.

Chapter 1: The thesis lacks clear motivation, one that would go beyond the text presented in its first Chapter (Introduction). Although it is well understood why studying volatility is important, we learn only very little (if anything at all) of the advantages of using non-parametric volatility measures (realized volatility). Also, how does realized volatility go beyond other methodologies in measuring the actual volatility (other than by making use of intraday information, although even this fact is not entirely clear from the text in the first Chapter)? In particular, what are the advantages of applying the approach based on (daily) realized volatility to those based on daily squared returns to proxy for (daily) volatility? How exactly do GARCH-type models fare here? Is it only their inability to – only to cite the author – "reproduce main features of financial data and in their short memory" that disadvantages them? By the way, having mentioned the short memory, what about (fractionally) integrated ARFIMA and IGARCH models? Finally, why employ the HAR model in the first place? It is my opinion that many of these points should have been discussed in the Introduction, along with the references to the relevant literature.

Chapter 2: There is very little to say about the methodological part of the thesis. At times, the author takes entire passages from the referenced literature, including - for obvious reasons - the texts of definitions, propositions, et cet. Still, the text contains many mistakes. Among others, in Proposition 2.1 (p. 5), for $x < 0$, we have $\mathrm{sgn}(x) \equiv -1$ (and not $\mathrm{sgn}(x) \equiv 0$). In part ii) of Proposition 2.2 (p. 6), we have "decomposition in Proposition 2.1" (and not "in definition 2.1"). On p. 7, it would be better to call Proposition 2.3 a Theorem as in the original study (similarly, on p. 8, we would then have a corresponding Corrolary and not Proposition 2.4). To name a few more, in Proposition 2.5 (p. 9), the integrals on the right hand side of (2.11) should have μ's and not σ's. Proposition 2.6 (p. 10) should reference Proposition 2.5 and not Proposition 5, and in formula 2.35 (p. 16), the difference between the realized variation and the standardized realized bipower variation converges (and is not equal to) to the sum of the (squared) jumps in the logarithmic price process, et cetera.

Chapter 3: First, the description of high-frequency data used in the analysis is insufficient. What is the source of the data? Furthermore, the text lacks any descriptive statistics for the raw (intraday) data. Given the descriptive statistics, one might be interested to investigate the raw (intraday) returns further: e.g., we might want to learn about the number of zero returns given the choice of a sampling frequency. This would be also interesting from the point of view of testing for jumps - it is easy to see from the definition of the bi-power variation (p. 16, formula 2.33, in the thesis) that zero returns introduce a downward bias into the estimator; as a result the test for jumps based on the difference between RVt and BVt has a potential to become heavily oversized if too few changes take place during a given sampling interval.

Second, the author performs no data cleaning; instead, she reasons that this is not necessary, "as the index price represents the whole market and it is comprised of the most liquid stocks". Clearly, this explanation is not sufficient. What about observations plagued by recording errors, such as misplaced decimal points, et cet. Such observations could seriously affect further results, especially if the source - which remains unknown - is not known to be reliable.

Third, how exactly did the author select the 5-minute frequency for the construction of the realized volatility estimator? The author employs the *scaling analysis* of Corsi et al. (2001) but does not provide any details about the analysis; also, what is the definition of annualized volatilities employed in the scaling analysis? Dismissing the choice of a sampling frequency at 5 minutes in one short paragraph by stating that the same choice was suggested in the previous literature is inappropriate; it is also the case that the previous studies (at least the once reference in the text) have dealt with entirely different and much more liquid markets. If anything, it would be interesting to see whether/how the results would be affected by the choice of a different sampling frequency and/or of a market microstructure robust estimator.

Fourth, how – in the end – did the author construct the realized volatility? Was it "constructed using equation (2.48)" as stated on p. 24? Was it constructed using 2.18? Related to this point are the "different computations of realized volatility" (p. 24); I do not understand the purpose of showing these computations, let alone the fact that the *different* approaches are mutually incomparable (e.g.: both delta and M in Corsi et al. (2001) and Corsi et al. (2008) are defined differently, and follow the different analysis in the corresponding studies).

Final comment concerns the testing for jumps. As already stated, the tests that the author employs are based on the difference between RV_t and BV_t (or similar) that can become can become oversized if too many zero returns are present. How does the index data look in this case? Still, even if the zero returns are not a concern, it would be interesting to employ an alternative approach to detecting for jumps (see, e.g., Mancini (2004) who employs a well-know result established by Lévy regarding the rate of decay of the increments of Brownian motion as the sampling interval tends to zero).

Chapters 4 and 5: I have no comments for these chapters.

References:

Corsi F. (2009). A Simple Approximate Long Memory Model of Realized Volatility, J. of Financial Econometrics, 7, pp. 174-196.

Mancini, C. (2004). Disentangling diffusion from jumps, J. of Financial Economics, 74, pp. 487–528.

In the case of successful defense, I recommend to award the thesis with the grade “vyborne” (excellent, 1).

Report on **Bachelor / Master Thesis**

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

Student:	**Bc. Eva Brabcová**
Advisor:	**PhDr. Jozef Baruník**
Title of the thesis:	**Volatility Modeling: Evidence from CEE Stock Markets**

SUMMARY OF POINTS AWARDED *(for details, see below):*

CATEGORY		**POINTS**
Literature	*(max. 20 points)*	20
Methods	*(max. 30 points)*	25
Contribution	*(max. 30 points)*	25
Manuscript Form	*(max. 20 points)*	20
TOTAL POINTS	*(max. 100 points)*	**90**
GRADE	**(1 – 2 – 3 – 4)**	**1**

NAME OF THE REFEREE: Vít Bubák

DATE OF EVALUATION: 18. 6. 2009

Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

LITERATURE REVIEW: *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

Strong	*Average*	*Weak*
20	*10*	*0*

METHODS: *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

Strong	*Average*	*Weak*
30	*15*	*0*

CONTRIBUTION: *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

Strong	*Average*	*Weak*
30	*15*	*0*

MANUSCRIPT FORM: *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

Strong	*Average*	*Weak*
20	*10*	*0*

Overall grading:

TOTAL POINTS	**GRADE**		
81 – 100	**1**	= excellent	= výborně
61 – 80	**2**	= good	= velmi dobře
41 – 60	**3**	= satisfactory	= dobře
0 – 40	**4**	= fail	= nedoporučuji k obhajobě