THE SHANGHAI- HONG KONG STOCK CONNECT: AN APPLICATION OF THE SEMI-CGARCH AND SEMI-EGARCH

Christian Peitz¹*  
Yuanhua Feng²  
Bernard M. Gilroy³  
Nico Stoeckmann⁴  
¹²³⁴University of Paderborn, Paderborn, Germany.  
Email: christian.peitz@uni-paderborn.de

ABSTRACT

This paper examines the impact on volatility on the Shanghai Stock Exchange and the Hong Kong Stock Exchange before and after the connection on November 17, 2014. We test whether this event led to a structural break. For this purpose, the volatility series are shown and analysed in more detail using two new models. We test both semiparametric GARCH extensions, the Semi-EGARCH based on the EGARCH model (Nelson, 1991) and the Semi-CGARCH model, based on the CGARCH (Engle & Lee, 1999). Both univariate models work with a constrained local linear estimator for the scale function and a fully data-driven algorithm developed under weak moment conditions. The proposed method is applied to the two major Asian financial indices and to stocks from the banking sector. Furthermore, our focus is to improve the quantitative risk management analysis. When a parametric GARCH model provides satisfactory results for the calculation of e.g the Value at Risk (VaR), semiparametric can also be applied to improve the quality of measurement. This article compares the results of various parametric and semiparametric approaches regarding the VaR to show how the extensions increase the performance.

1. INTRODUCTION

The connection of the Shanghai Stock Exchange and the Hong Kong Stock Exchange was announced on April 10, 2014 and later established on November 17, 2014, giving domestic investors new access to a market previously dominated by institutional investors. Simultaneously it allows foreign investors to operate on an emerging market. Regardless of the arising possibilities, there are some limitations. Among others, investments from the mainland are exclusively available for institutional investors and wealthy private individuals due to a minimum investment size and only the best credit-rating shares are available for international investors through the channel of the Hong Kong Stock Exchange.

Its announcement had a significant impact on investors behaviour and expectations (see Huang and Lin (2017)), leaving the Hong Kong stock market with an overall higher volatility and response than the Shanghai counterpart (see Yan (2015)). Contrary to fundamental assumptions and announcements that partial liberalisation of markets will bring equal benefits to all sides by facilitating trade, reality shows unequal effects on both markets due to their
unique features and policies (see Chong and Kwok (2019-A) as well as Wang, Tsai, and Lin (2016)). Spill-over effects are found to be unidirectional from Hong Kong to Shanghai, before and after the connect (see Yang and Zhang (2015)). While the cross-correlation between Hong Kong and Shanghai gets stronger, the multifractality of Shanghai stock market shrinks after the Shanghai-Hong Kong stock connect (see Ruan, Zhang, Lv, and Lu (2018)). Long range correlation contributes to the multifractality of cross-correlation. The power, influence and market efficiency in the Chinese mainland therefore potentially grows due to an increase in mean and volatility spill-over effects from the mainland to Hong Kong (see Yang and Zhang (2015) as well as Huo and Ahmed (2018)). Chong and Kwok (2019-B) find a strong long term linkage between Shanghai and Hong Kong with fundamental changes after the connection.

ARCH and GARCH models serve as risk management and analytical tools since the 1980's. They ease the calculation of the VaR. These models and their extensions are highly popular due to their performance at estimating volatility in risk measurement (see e.g. Poon and Granger (2005)). For our extension, we rely on existing literature.

In this article, we will show the advantages of semiparametric models as opposed to the purely parametric models. The long-run risk component will be modelled by a smooth volatility trend in the return series and the conditional risk dynamics are analysed using two newly introduced extensions. We will show that this method can improve significantly the VaR measurement and that the proposed Semi-EGARCH and Semi-CGARCH are classified as the best variants. One way to test the performance of these models is to calculate the VaR and compare the results after a back-testing procedure. Furthermore, we analyse the long-term and short-term risk components in Chinese financial market.

2. SEMIPARAMETRIC MODELS

Semiparametric models are statistical models with parametric/non-parametric components, translating into a regression with a finite/infinite-dimensional component from an available data set. Parametric models often lack a fair representation of real-world characteristics. Non-parametric models may represent the real world more accurately. Semiparametric models may offer a compromise from both approaches, being adjustable while obtaining a complex representation of several features.

Volatility clustering is a generally acknowledged challenge throughout the analysis of financial return data. Semiparametric ARCH and GARCH models show high potential for application in economic contexts. Semiparametric GARCH models have been proposed to minimize the risk of a potential efficiency loss as well as assuming the wrong parametric family of conditional distributions (see Ji and Lucas (2019)). Semiparametric models capture a substantial proportion of the potential by the estimators in economic data and generally tend to show significant potential at improving the quality of risk measurements. We point out that these models are not only better in terms of analysis, but a general improvement of the GARCH class.


The subject of this section is the introduction of the semiparametric extension of the EGARCH model and the CGARCH model. Both model specifications are not yet well known in the literature.
2.1. The Semi-EGARCH and the Semi-CGARCH

The class of semiparametric models and especially the Semi-GARCH is a well-known improvement of parametric models. The advantages of the Semi-EGARCH and Semi-CGARCH will be demonstrated empirically. The numerical results are also compared with those of the known parametric models and the Semi-GARCH model. Following Feng and Sun (2013) who developed a Semi-APARCH model, the general semiparametric equation is to be considered:

\[ r_t = s(\tau)\sigma_t \varepsilon_t \]  

(1)

where \( s(\tau) \geq 0 \) is the scale function of the long-term trend, \( \tau \) is the rescaled time and \( \varepsilon_t \) is an i.i.d. random variable with an average of zero and a variance of one (see Feng (2013)). For the explicit derivation of \( s(\tau) \), we refer to Feng (2013) as well as Beran and Feng (2002). After adjustment for the trend, returns can be estimated using a parametric model. Due to its characteristics and its popularity we consider a further logical adjustment, and thus another proposed semiparametric extension in this regard, the Semi-EGARCH model. The original parametric model according to Nelson (1991) is defined by:

\[ \log \sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i z_{t-i} + \gamma (|z_{t-i}| - E|z_{t-i}|) + \sum_{j=1}^q \beta_j \log \sigma_{t-j}^2 \]  

(2)

With Equation 2 and the general semiparametric Equation 1, we can define the Semi-EGARCH. The semiparametric extension of the CGARCH model should also be considered in this context. The general Equation 1 can be used by integrating the CGARCH model introduced by Engle and Lee (1999):

\[ \sigma_t^2 = q_t + \sum_{i=1}^p \alpha_i (z_{t-i} - q_{t-i}) + \sum_{j=1}^q \beta_j (\sigma_{t-j}^2 - q_{t-j}) \]  

(3)

and

\[ q_t = \omega + p q_{t-1} + \phi (\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \]  

(4)

must be adjusted. For further details, refer to the explanations in Engle and Lee (1999). It can be stated here once again that Equation 1 does not depend on the parametric model and the first, non-parametric, part of Equation 1 can be estimated independently of the second part of the equation. Following the definition of these two new models, the efficiency will be reviewed empirically.

2.2. Calculation of Value-At-Risk Based on Semiparametric Models

We estimate the risk measure VaR with a parametric model and a semiparametric model, as well as to compare these estimates with each other, in addition to a further comparison of a parametric model with a semiparametric model.

Analyses with various parametric GARCH extensions for calculating the VaR have been established and often tested in the literature for many years.

If a parametric GARCH model provides satisfactory results for the calculation of the VaR as mentioned before, a VaR calculation based on semiparametric models is also to be considered. The models GARCH, EGARCH and CGARCH should be used. These models are therefore to be extended semiparametrically. The Semi-GARCH model (see Yuanhua (2004)) and the Semi-APARCH model (see Feng and Sun (2013)) are used as a benchmark in this context. For this purpose, the local variance has to be considered again. In the semiparametric case, the VaR is calculated as follows:
\[ VaR^T_{\alpha} = \mu_{t+1} + s(\tau) \sigma_t q_\alpha(Z) \] (5)

where \( \mu_{t+1} \) is close to zero and can therefore be neglected, \( s(\tau) \) is the local variance estimated by the observed values within the period and is approximately the same for the same \( t \) over a short period. The conditional standard deviation is also referred to \( \sigma_t \). Furthermore, \( q_\alpha(Z) \) depends on the innovations and is assumed to be normally distributed. Therefore \( q_\alpha(Z) \) is equal to \( \varphi^{-1}(\alpha) \). In the semiparametric model fitting, the deterministic trend is again estimated in a first step, which is then eliminated. Afterwards, the parametric model is adapted and a retransformation is performed. An important aspect of the trend function estimation is once again the choice of bandwidth. In the calculation of semiparametric models, the bandwidth is realized by the data-driven IPI algorithm (see Feng (2013)). The local linear estimator of the scale function should be applied again to absolute returns and not to squared returns, which is a common method (see Feng (2013)). This aspect ensures that the data-driven algorithm converges under weak moment conditions and that the estimation works better at the limits than estimating with squared returns. The fundamental approach of the VaR estimation is the same for all models and is only mentioned briefly in this paper (see McNeil, Frey, and Embrechts (2015)).

Based on the approaches of the Semi-GARCH model and the Semi-APARCH model, in the context of the VaR estimate, logical extensions are the Semi-EGARCH model and the Semi-CGARCH model.

3. APPLICATION

In order to examine the impact of the Shanghai- Hong Kong stock connect on volatility via semiparametric extensions, a closer look at the time series of the two major Asian indices, the Shanghai Composite (SSE) and the HangSeng (HSI) is indispensable, especially around the stock connect on November 17, 2014.

Figure 1 on the left hand side shows the price series as well as the normal log returns. Figure 2 on the right hand side shows the scale function as well as the standardized returns. The standardized returns multiplied by the scale function represents the returns from the Figure 1.
The red line in Figure 2 shows the smooth scale function which reflects the long term risk. For the sake of completeness, Figure 3 shows the same procedure for the HSI.

![Figure 2: HSI Price, log-returns, scale function and standardized returns.](image)

Considering these illustrations, it is noticeable that Figure 3 is one-peaked while Figure 2 is two-peaked. This is the first significant difference between these two indices in the long term.

The only huge long-term risk phase while looking at the HSI is the global financial crisis from 2007 onwards. When looking at the SSE, there is another massive increase in the scale function around 2014. As an intermediate result, the risk of the SSE is higher than the risk of the HSI measured by our models. The measured volatility of both series in Figure 4 and their comparison are used for further investigations.

3.1. Test of Structural Breaks in Volatility

The following Figure 4 shows the volatility according to a normal GARCH model, to compare the two time-series. The respective illustrations with the models were created for the t-distribution which is the most suitable for this data set.
The red vertical line shows the exact date of the merger of the two markets on November 17, 2014. After that, a massive increase in volatility and a possible structural break can be obtained, especially in the time-series of the SSE. In particular, the SSE, whose volatility rises to an average of 142% in the year after the merger compared to the year before. The volatility of the HSI in the year after the merger was 37% higher than the average volatility in the previous year. Therefore, also the risk for both markets is rising sharply. We consider the test of possible structural breaks in volatility series estimated by GARCH models at some known date, in this case November 17, 2014.

Two sub-volatility series in one year before that date (Series B, $y_{t-1}, \ldots, y_{t}$) and in one year thereafter (Series A, $x_{t}$) are chosen to carry out the following test. Denote the two means of those series by $\mu_{x}$ and $\mu_{y}$.

Here, we assume that the structural break should be a jump. Hence, the null- and alternative hypotheses are:

$$H_{0}: \mu_{x} \leq \mu_{y} \text{ against } H_{1}: \mu_{x} > \mu_{y}$$

The approach for a two-side test is similar. The proposed test approach is developed under following assumptions:

- $A_{1}$: $x_{t}, t = 1, \ldots, n_{x}$ is stationary with absolutely summable acf (autocovariances) $y_{x}(k), k = \ldots, 1, 0, 1, \ldots$ and $y_{t}, t = 1, \ldots, n_{y}$ is stationary with absolutely summable acf $y_{y}(k), k = \ldots, 1, 0, 1, \ldots$

- $A_{2}$: $x_{t}$ and $y_{t}$ are independent of each other and both of $n_{x}$ and $n_{y}$ are large enough.

- $A_{3}$: It is assumed that $\sum_{k=0}^{\infty}|k + 1|^4|y_{x}(k)| < \infty$ and $\sum_{k=0}^{\infty}|k + 1|^4|y_{y}(k)| < \infty$

$A_{1}$ is a typical regularity assumption in the current context. $A_{2}$ is made for simplicity and is usually practically relevant. $A_{3}$ is much stronger than the absolute summability and is required by the application of the nonparametric data-driven lag-window estimator of the factor in the variance of the sample mean of a stationary time series (see Bühlmann (1996) as well as Feng, Gries, and Fritz (2019)).
Let $\bar{x}$ and $\bar{y}$ denote the two sample means. Under A1 we have: $\text{var}(\bar{x}) \approx c_{f,x}/n_x$ and $\text{var}(\bar{y}) \approx c_{f,y}/n_y$,

where $c_{f,x} = \sum_{k=-\infty}^{\infty} Y_x(k)$ and $c_{f,y} = \sum_{k=-\infty}^{\infty} Y_y(k)$. Under the additional assumption A2 we have

$$\text{var}(\bar{x} - \bar{y}) \approx \frac{c_{f,x}}{n_x} + \frac{c_{f,y}}{n_y} \quad (6)$$

The result in Equation 6 provides a simple way to test, whether $d = \bar{x} - \bar{y}$ is significantly positive or not, provided that $c_{f,x}$ and $c_{f,y}$ are estimated properly.

Under A3, $c_{f,x}$ and $c_{f,y}$ can be estimated using the nonparametric data-driven lag-window estimator proposed by Bühlmann (1996). Feng, et al. (2019) adjusted his idea slightly and developed an R function for the practical implementation of this approach, which will be applied in the following to test the possible structural break in the volatility of HSI and SSE caused by the Shanghai–Hong Kong stock connect on November 17, 2014.

For the HSI we have

$$\bar{y} = 0.009172475, \bar{x} = 0.01254373 \text{ and } d = 0.00337126$$

The estimated standard error of $d$ is $se(d) = 0.000904134$ with $t = 3.73$ and a $p$-value of $9.622864 \times 10^{-05}$. We see the structural break is very highly significant. For the SSE we have

$$\bar{y} = 0.00978263, \bar{x} = 0.02377473 \text{ and } d = 0.0139921$$

The estimated standard error $se(d) = 0.002021008$ with $t = 6.92$ and a $p$-value of $2.20568 \times 10^{-12}$. Here, the structural break is extremely highly significant. The effects of the Shanghai–Hong Kong stock connect are therefore very large in terms of quantitative risk.

Note:

- The suggested test can also be carried out by means of fitting two ARMA models to the two sub-series. But this this parametric method may be with misspecification and is unstable.
- This idea can be extended to cases, where the date of possible structural break is assumed to be unknown. Now, the proposal should be applied in a moving way and the detected structural break corresponds to the time point, where the $t$ statistic is maximized.
- It should be mentioned that the volatility of both markets has calmed down within about 1.5 years after the Shanghai–Hong Kong stock connect in 2014 and are currently at a very low level as Figure 5 shows.
3.2. **Comparison of Different GARCH Models to Measure Risk**

The current aim is to make more accurate statements about the risk within the indices. Therefore, we calculate various risk measures on the basis of normal GARCH models and the semiparametric extended models, based on the procedure of section 2.2. On the one hand, it enables us to make statements about the market risk and, on the other hand, to make statements about which model is most suitable for assessing the risk.

Three parametric models and three models after the semiparametric extension were calculated. All under the orders \((1,1)\) which was the best following the BIC for all models. Furthermore, all models were adapted for \(t\) - distribution. Based on this, the VaR is calculated according to \[\text{Equation 5}\] with the limit \(\alpha = 5\%\).

The VaR quantifies the loss which, with a certain probability, is not exceeded within a fixed time horizon. **Figure 6** shows the VaR frontier after a GARCH calculation together with the returns of the SSE and the red marked exceedances. **Figure 7** shows the same after the semiparametric adaption (in this case a simple Semi-GARCH model).

Since the VaR is calculated purely parametrically, the prediction accuracy of both approaches can be compared. The term "negative returns" refers to a reflection of the yields in the figures. In this case, the VaR limits represent the value of the negative returns with a probability of \(95\%\) in the period of one day.
Figure 7. Value at Risk after a semi GARCH calculation (SSE).

Only on closer inspection, it is noticeable that the VaR frontier with these models adapts better to returns. This point is going to be investigated later with numerical results.

After the calculation and adjustment of the VaR, its quality must be checked as mentioned above. There are a number of ways to do this, with a simple back-test to assess the quality of the estimates. Consequently, the forecasting ability is going to be compared by recording the number of observations in which the loss, e.g. the negative returns, exceed the VaR. Theoretically, the VaR should be exceeded in exactly \( \alpha \% \) of all cases. The number of trading days is a decisive criterion for calculating the theoretical or expected overshoots. Back-testing in VaR is a technique used to compare the predicted losses from the calculated VaR with the actual losses realized. We want to show that the semiparametric models work more efficiently. Therefore, we want to measure the points over or out of the VaR frontier for every model.

We can only make accurate statements if we take a closer look at the exceedances, so the number of points that are above the limit. In the following, we will focus on these points. The ugarch-roll function in R is quite useful for testing the suitability of models in a back-test application, especially for the rolling forecast procedure.

Table 1 and Table 2 show the numerical results, respectively the deviations (the VaR for the parametric models at the \( \alpha = 5\% \) level), from 2001 to August 2019.

<table>
<thead>
<tr>
<th></th>
<th>GARCH</th>
<th>EGARCH</th>
<th>CGARCH</th>
<th>Semi-GARCH</th>
<th>Semi-EGARCH</th>
<th>Semi-CGARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected Exceed</td>
<td>183</td>
<td>183</td>
<td>183</td>
<td>183</td>
<td>183</td>
<td>183</td>
</tr>
<tr>
<td>Actual Exceed</td>
<td>241</td>
<td>229</td>
<td>239</td>
<td>241</td>
<td>182</td>
<td>201</td>
</tr>
<tr>
<td>Difference</td>
<td>-58</td>
<td>-46</td>
<td>-56</td>
<td>-21</td>
<td>1</td>
<td>-19</td>
</tr>
</tbody>
</table>

For the back-test length of the example we have an expected exceed of 183 (which are 5% of about 3700 observations, the back-test length). With the first parametric model, the standard GARCH, we achieve an actual VaR exceed of 241, so a relatively high upward deviation. For the best semiparametric model, the Semi-EGARCH model we have of course the same expected exceed of 183 and an actual VaR exceed from 182, much closer to the
expected value and a nearly perfect result in the back-testing procedure. The results of the back-test are much more accurate according to the semiparametric model.

Table 2. HSI: Number and expected exceedances of the used data.

<table>
<thead>
<tr>
<th></th>
<th>GARCH</th>
<th>EGARCH</th>
<th>CGARCH</th>
<th>Semi-GARCH</th>
<th>Semi-EGARCH</th>
<th>Semi-CGARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected Exceed</td>
<td>181</td>
<td>181</td>
<td>181</td>
<td>181</td>
<td>181</td>
<td>181</td>
</tr>
<tr>
<td>Actual Exceed</td>
<td>196</td>
<td>228</td>
<td>198</td>
<td>191</td>
<td>200</td>
<td>193</td>
</tr>
<tr>
<td>Difference</td>
<td>-15</td>
<td>-47</td>
<td>-17</td>
<td>-10</td>
<td>-19</td>
<td>-12</td>
</tr>
</tbody>
</table>

A similar result can be seen, even if the models perform differently, because of the structure of the respective time series and whether it matches the selected model. The EGARCH model seems to fit badly to the structure of the time series, but the semiparametric model is much better and has less than half of the deviations. So also this example shows that the semiparametric models are generally more suitable than the parametric models. In these models, we measure less minor deviations from the theoretical VaR limits. Furthermore, it has to be investigated which of the adapted semiparametric models is the most suitable. The requirements for a good econometric model are robustness, safety and especially variability. For this reason, we consider additional examples. It makes sense to examine the Chinese banks, which are among the largest in the world. As an additional factor, risk management is most important for banks. So we are adding three more data sets to our example to test the models, especially the differences between the parametric and the semiparametric models. When considering these examples, we concentrate only on the measured exceeds. Table 3 shows the sum of deviations of all models. The results from Table 1 and Table 2 will only be displayed in a shortened form. A non-robust model and a comparable static model would result in a high sum of deviations at this point.

Table 3. Sum of deviations.

<table>
<thead>
<tr>
<th></th>
<th>GARCH</th>
<th>EGARCH</th>
<th>CGARCH</th>
<th>Semi-GARCH</th>
<th>Semi-EGARCH</th>
<th>Semi-CGARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICBC</td>
<td>18</td>
<td>22</td>
<td>-14</td>
<td>-12</td>
<td>-6</td>
<td>-12</td>
</tr>
<tr>
<td>ABC</td>
<td>23</td>
<td>14</td>
<td>21</td>
<td>-1</td>
<td>-3</td>
<td>0</td>
</tr>
<tr>
<td>BOC</td>
<td>30</td>
<td>23</td>
<td>5</td>
<td>1</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>SSE</td>
<td>60</td>
<td>48</td>
<td>58</td>
<td>21</td>
<td>-3</td>
<td>-15</td>
</tr>
<tr>
<td>HSI</td>
<td>-14</td>
<td>-48</td>
<td>-16</td>
<td>-10</td>
<td>-3</td>
<td>-12</td>
</tr>
<tr>
<td>Sum (Absolute)</td>
<td>145</td>
<td>155</td>
<td>114</td>
<td>43</td>
<td>33</td>
<td>41</td>
</tr>
</tbody>
</table>

Note: In the analyses in this section, the slightly deviating observation figures among the individual companies and among the individual indices were disregarded. In the analyses, in which the results of the companies are combined with those of the indices, corresponding weighted values were used for the indices.

If we combine the results to a sum of differences and order them, we can see that all semiparametric models beat the parametric ones and both new models beat the normal Semi-GARCH. In order to clarify the sequence once again, the models were arranged in ascending order according to the sum of deviations. The smallest sum of deviations should show the model that is most robust and variable with chosen time series. Table 4 shows the summarized results.

This table clearly shows that semiparametric models are better suited for VaR calculation than parametric models according to the results of the eight used data sets. The sequence should be noted, especially the point that all semiparametric models are preferable to parametric models. The Semi-EGARCH model in particular is to be classified as the best model after the analyses in this work, which is also illustrated in Table 4.
4. CONCLUSION

The analyses of the volatility have shown the massive influence of the merger of November 17, 2014 on the two markets SSE and HSI. A structural break in the volatility of the indices was recognized from this point on. The risk in the markets rose sharply as a result.

Furthermore, the risk was estimated using parametric and semiparametric models. The semiparametric extended models, in particular Semi-EGARCH and Semi-CGARCH, provide significantly better results after the back-testing procedure than the parametric models. Consequently, the models are better suited for risk forecasting. However, here are the clear advantages of semiparametric adjustments. Model adaptation on trend-adjusted data generally works better. In particular, the Semi-EGARCH model is validated to be a suitable model when considering all five data sets.

The procedure of the semiparametric extension is of course transferable to other related model categories, such as the semiparametric ACD models for daily average trade durations and daily trade durations. It is also applicable to semiparametric multivariate models (MGARCH) such as the Semi-DCC or the FIGARCH (Baillie, Bollerslev, & Mikkelsen, 1996).

**Funding:** This study received no specific financial support.

**Competing Interests:** The authors declare that they have no competing interests.

**Acknowledgement:** All authors contributed equally to the conception and design of the study.

**REFERENCES**


*Views and opinions expressed in this article are the views and opinions of the author(s), Asian Economic and Financial Review shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.*