<center>

Comments on:
"Forecast Evaluation of Explanatory Models of Financial Return Variability"
by Genaro Sucarrat

</center>

The aim of the paper is to develop a framework for forecast comparisons for volatility models of financial returns. In particular, in order to rank discrete time volatility models, the author considers frequently used loss-functions like the MSE or the MAE applied to forecasts of the conditional variability (approximated by the squared returns). The contribution of the paper is to compare by means of a Monte-Carlo study the finite-sample properties of the rankings obtained under the different loss functions. Based on those finite-sample properties a three-step ranking procedure is proposed, which is based on the MSE or MAE criterion (first step), comparative forecast accuracy tests like that of Diebold and Mariano (1995) (second step) and the R-squared of a Mincer regression (last step). In order to illustrate the proposed procedure it is applied to four volatility models which are fitted to exchange rate data.

I have the following comments:

1) The main problem I have with the paper is that it is unclear to me how general its main results are and whether they are really relevant for practical purposes. In particular, the MC study presented in Section 3 considers one particular DGP (a bivariate regression model for the returns with errors following a GARCH-type process) used to simulate artificial data and compares forecasts based on four models, the true one and 3 miss-specified models. The question is whether we get under alternative DGPs (like stochastic volatility or Markov-switching models) and/or the comparison of a different set of miss-specified model the same conclusions with respect to the finite sample properties of the ranking procedures obtained under different loss-functions.

Furthermore, in the simulation study the forecasts under different models are obtained by using the true parameter values for the correct model and 'calibrated pseudo-true' parameter values for the miss-specified models obtained by equating the unconditional variability $E(r_t^2)$ of the true model and that of the miss-specified models. However, in practical applications we need to estimate the parameter values. Hence, I think it is important to account for the effect of using a model fitted to the data rather than a model at the true or calibrated pseudo-true parameter values on the finite-sample properties of the ranking procedures.

2) It remains unclear to me why the "...volatility is not a variable in theory mechanism nor in the DGP... (see, p.3)". Many stochastic model used in empirical finance explicitly specify the volatility as the variable of key interest. Furthermore one can easily imagine a DGP where the volatility in terms of a conditional variance of the returns is explicitly specified. Finally, isn't it the case that the author itself uses in the simulation study a DGP (see, Equation 5) where the conditional variance is treated as a random variable?

<center>

1

</center>

3) If my interpretation is correct the discussion in Section 2.2. serves to motivate why one should use the squared returns as a measure for returns' conditional variability as the quantity to be forecasted and not volatility measures obtained from continuous-time models such as the realized volatilities.

One of the author's argument is that time is needed for an event to have an impact on the returns and, hence, explanatory variables are likely to account for a decreasing portion of return variability as the size of the time increments tends to zero. However, isn't it the case that the variability of the returns to be explained is also decreasing as the time increments tends to zero?

Furthermore, the authors suggestions not to use measures like the realized volatility for methods comparing volatility/variability forecast contradicts the typical suggestion in the literature (see, e.g. Andersen and Bollerslev, 1998 and Patton, 2006). Hence, I think the author should explain this contradiction more explicitly referring directly to that literature.

4) I have conceptional problems with the sample kurtosis of the standardized returns as a loss function (see Equation 12). The author argues that the sample kurtosis constitutes a goodness-of-fit measure in a time-varying volatility context, analogous to the standard error of a regression in constant-volatility context. However, in the present context, I think, the measure which would be analog to the usual standard error of regression is

$$\sum_{t=1}^{T}(r_t^2 - \hat{r}_{mt}^2)^2,$$

(or a normalized version of it) which is obviously different from the fourth sample moment entering the sample kurtosis

$$\sum_{t=1}^{T}(r_t - \hat{r}_{mt})^4 = \sum_{t=1}^{T}(r_t^2 - 2r_t r_{mt} + \hat{r}_{mt}^2)^2.$$

Hence it remains unclear to me to what extend the sample kurtosis of the standardized residuals could help to discriminate between models used to predict r_t^2. Additionally, the simulation results presented in Section 3 indicate that the performance of the sample kurtosis as a loss function is very bad.