Session 3: Results of the M6 Competition

The M6 Forecasting Competition: Bridging the Gap Between Forecasting and Investment Decisions *by Spyros Makridakis, Evangelos Spiliotis, Ross Hollyman, Fotios Petropoulos, Norman Swanson, Anil Gaba*

Select Volatility and Risk Models

Norman R. Swanson

Rutgers University

<u>The Future of Forecasting and the M6 Competition Conference</u> - *November 6-7, 2023*

Outline of the Talk

- **Introduction**
- **Brief Discussion of Select Volatility Models**

Introduction

- In the M6 paper we analyze investment decisions made by participants by using the ***investment risk model*** that Ross has developed.

- A key purpose of this model is to develop pseudo-optimal investment decisions that can be compared in a number of ways with competition ***investment submissions*** so that we can assess the optimality of the level of risk taken.

- Needless to say, models of volatility/risk play an important role in any investment risk model. For example, one approach is to construct implied volatilities from options prices.

- Some other key approaches to modelling and forecasting risk:

  - (i) Discrete parametric models based on low frequency historical time series data.

  - (ii) Continuous parametric models based on high frequency historical time series data.

  - (iii) Realized Volatility (RV): Simple nonparametric measures of volatility based on high frequency historical time series data.

  - (iv) RV Based Heterogenous Autoregressive (HAR) Type Models

Introduction

- Most approaches build on stylized facts concerning volatility.
  - *Variance evolves over time (e.g. heteroskedasticity).*
  - *Volatility spillover across/between assets and asset classes.*
  - *Volatility clustering over time.*
  - *Heavy tails and kurtosis.*
- Univariate models much easier to estimate than multivariate models.
- Multivariate models often use big-data and related machine learning, shrinkage and dimension reduction methods.
- In the sequel I briefly discuss key varieties of models (i)-(iv) above.

Volatility Models

- We first discuss the ubiquitous GARCH model.
  - These models are widely used in portfolio optimization.
  - They are very parsimonious, embed time variation, allow for volatility clustering and admit heavy tailedness, as kurtosis.

Volatility Models

(i) Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Models

(Engle (1982) and Bollerslev (1986))

Define the innovation as an error in a linear regression:

$$\varepsilon_t = X_t - Z_t'\beta$$

- and

$$\varepsilon_t | F_{t-1} \sim N(0, \sigma_t^2)$$

- A GARCH model specifies:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

$$p \geq 0, q > 0;\ \alpha_0 > 0,\ \alpha_i \geq 0,\ i = 1,\ldots,q;\ \beta_i \geq 0,\ i = 1,\ldots,p$$

- Normality may be relaxed. Higher moment assumptions needed for estimation and inference.
- Many extensions, such as IGARCH and EGARCH, and various asymmetric extensions. IGARCH has been found to be particularly useful, and assumes that:

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1$$

Volatility Models

- We now discuss the increasingly important class of ***continuous time models*** estimated using high frequency data. Discrete variants of these model are also used in practice.

Volatility Models

(ii) An Example of a Continuous Time Model – The Stochastic Volatility Model (CIR, Heston, etc.)

$$dX_t = (\alpha + \beta X_t)dt + \sigma dW_{1t},$$

$$d\sigma_t^2 = \kappa(\vartheta - \sigma_t^2)dt + \delta\sigma_t dW_{2t},$$

where $Cov(dW_{1t}, dW_{2t}) = \rho dt.$

- This model specifies the asset price and the instantaneous variance using a system of two stochastic differential equations. Here, the Wiener processes may be correlated.

- In the options pricing literature, many models are nested with the following model that allows for jumps:

$$dX_t = \mu_t dt + \sigma_t X_t dW_{1t} + dJ_{1t}$$

$$d\sigma_t^2 = \kappa(\alpha - \sigma_t^2)dt + \delta\sigma_t(\rho dW_{1t} + \sqrt{1-\rho^2}\, dW_{2t}) + dJ_{2t}$$

- This model is a generalization of the above model, where the drift term may be a function of the asset price, for example, in which case when we set the jump components equal to zero, then the first equation is the famous CIR model (when $\mu_t = \mu X_t$), and the system is the basic Heston model that is widely used.

Volatility Models

- Finally, we discuss the increasingly important class of nonparametric ***realized volatility (RV) estimators*** based on the use of high frequency data.

- These estimators are easy to construct, are consistent estimators of the true underlying instantaneous volatility associated with continuous time models and are often aggregated and used at a daily frequency and can be made robust to jumps (or not).

- These RV estimators are also used in an important class of models called HAR models.

Volatility Models

(iii) Nonparametrically Estimated Realized Volatility (following Ait-Sahalia (various) and Duong and Swanson (2011), etc.)

- **Assume that the log price follows:** $X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s + \sum_{s \le t} \Delta X_s,$

where $X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s$ is the continuous semimartingale component of the process

The jump part of X_t in the time interval $[0, t]$ is defined to be $\sum_{s \le t} \Delta X_s$.

- When the jump is a Compound Poisson Process (CPP) - i.e. a finite activity jump process then

$$J_t = \sum_{s \le t} \Delta X_s = \sum_{i=1}^{N_t} Y_i,$$

where N_t is number of jumps in $[0, t]$, N_t follows a Poisson process,

and the Y_i's are i.i.d. and are the sizes of the jumps.

Volatility Models

(iii) Nonparametrically Estimated Realized Volatility (cont.)

- Here we can define and easily estimate volatility using high frequency data. This is done by estimating integrated and quadratic volatility.

$$IV_t = \int_0^t \sigma_s^2 ds = [\text{variation due to continuous component}]_t$$

$$QV_t = [X,X]_t = \int_0^t \sigma_s^2 ds + \sum_{t-1\leq s<t} \Delta X_s^2 = IV_t + [\text{variation due to jump component}]_t$$

Estimation is done using

$$RV_{t,n} = \sum_{i=1}^n (\Delta_i^n X)^2$$

For n large, RV converges almost surely to the quadratic variation of the process. Measuring volatility within a day we have:

$$RV_{t,n} = \sum_{i=1}^n (\Delta_i^n X)^2 \simeq \int_{t-1}^t \sigma_s^2 ds + [\text{jump component variation between day } t-1 \text{ and } t]_t$$

Jump robust estimators are often used in practical applications:

$$V_{r_1,r_2\ldots,r_j} = \sum_{i=j+1}^n |\Delta_i^n X|^{r_1} |\Delta_{i-1}^n X|^{r_2} \ldots . |\Delta_{i-j}^n X|^{r_j}.$$

$$BV_{t,n} = (\mu_1)^{-1} \sum_{i=2}^n V_{1,1} \simeq \int_{t-1}^t \sigma_s^2 ds$$

Volatility Models

(iv) RV Based Heterogeneous Autoregressive (HAR) Models (following Bollerslev et al. (2018))

- **The risk model uses a multivariant variant of the basic HAR model due to Corsi (2009).**

$$RV_{t+h}^{h} = \theta_0 + \theta_D RV_t + \theta_W RV_t^W + \theta_M RV_t^M + \varepsilon_{t+h}$$

- **A variant of this is based on the MIDAS approach of Ghysels, Santa-Clara, and Valkanov (2006) that uses smooth beta polynomials to avoid the stepwise changes inherent to the HAR model.**

$$RV_{t+h}^{h} = \theta_0 + \theta_1 [a(1)^{-1} a(L)] RV_t + \varepsilon_t$$

$$a_i = (\frac{i}{k})^{\theta_1 - 1}(1 - (\frac{i}{k})^{\theta_2 - 1}\Gamma(\theta_1 + \theta_2)\Gamma(\theta_1)^{-1}\Gamma(\theta_2)^{-1}, \quad i = 1, \ldots, k$$

- **This model is extended in Bollerslev et al. (2018) to specify models that use a mixture of "smooth" exponentially weighted moving averages of past RV values**

Volatility Models

(iv) RV Based Heterogeneous Autoregressive (HAR) Models – The HEXP Model

$$RV_{t+h}^{h} - RV_{t}^{LR} = \sum_{j=1,5,25,125} \theta_j (ExpRV_t^{j} - RV_t^{LR}) + \varepsilon_t$$

$$ExpRV_t^{CoM(\lambda)} = \sum_{i=1}^{500} \frac{e^{-i\lambda}}{e^{-\lambda}+e^{-2\lambda}+\ldots+e^{-500\lambda}}$$

where

$$CoM(\lambda) = e^{-\lambda}/(1 - e^{-\lambda})$$

- **This model is nested within the following omnibus model:**

$$RV_{t+1} - RV_t^{LR} = b_1(RV_t - RV_t^{LR}) + b_2(RV_{t-1} - RV_t^{LR}) + \ldots + \varepsilon_t$$

Here the long-run volatility factor equals the expanding sample mean of daily RV values from the beginning of the sample up until day *t.* This *centered* risk model replaces the intercept with the long-run volatility. Replacing the intercept in this way avoids forcing the intercept to be the same across different asset classes.

"Let's shrink Big Data into Small Data ...
... and hope it magically becomes Great Data."