# Multi-step Ahead Prediction of Financial Returns

Dimitris N. Politis
Univ. of California, San Diego
(joint work with Jie Chen)

- The optimal one-step ahead predictor is
  - In the $L_1$ sense, conditional median.
  - In the $L_2$ sense, conditional mean.

- Under linearity, the $L_2$ optimal $h$-step ($h \geq 2$) ahead prediction can be obtained by iterating the one-step ahead predictor

- But: L1-estimators are more robust with respect to heavy-tailed distributions.

- Also: What happens under nonlinearity?

- How to do $L_1$ optimal $h$-step ($h \geq 2$) ahead predictor?

- Abadir, Luati, and Paruolo. 2018: The forecast density of a Garch(1,1).

- It is an important breakthrough but it hinges on knowing the error distribution that drives the GARCH(1,1) model.

Consider an ARCH($p$) process:

$$X_t = \sigma_t \epsilon_t, \;\; and \;\; \sigma_t^2 = \alpha + a_1 X_{t-1}^2 + ... + a_p X_{t-p}^2$$

where $\alpha \geq 0$, $a_j \geq 0$, and $\{\epsilon_t\} \sim i.i.d.\ N(0,1)$.

Focus on $h = 2$; then the squared return is

$$X_{n+2}^2 = \epsilon_{n+2}^2 \left(\alpha + a_1 X_{n+1}^2 + ... + a_p X_{n+2-p}^2\right)$$

Given $\{X_1, ..., X_n\}$, the $L_2$ optimal predictor of $X_{n+2}^2$ is $E(X_{n+2}^2 | X_1, ..., X_n) =$

$$= \alpha + a_1 E(X_{n+1}^2 | X_1, ..., X_n) + \sum_{i=2}^{p} a_i X_{n+2-i}^2$$

Here, we can directly plug in the one-step ahead predictors for the missing values.

$$(1): X_{n+1}^2 = \epsilon_{n+1}^2 \left(\alpha + a_1 X_n^2 + ... + a_p X_{n+1-p}^2\right)$$

Given $\{X_1, ..., X_n\}$, the $L_1$ optimal predictor of $X_{n+1}^2$ is $Median(X_{n+1}^2 | X_1, ..., X_n) =$

$$= Median(\epsilon^2) \left(\alpha + a_1 X_n^2 + ... + a_p X_{n+1-p}^2\right)$$

$$(2): X_{n+2}^2 = \epsilon_{n+2}^2 \left(\alpha + a_1 X_{n+1}^2 + ... + a_p X_{n+2-p}^2\right)$$

Given $\{X_1, ..., X_n\}$, the $L_1$ optimal predictor of $X_{n+2}^2$ is $Median(X_{n+2}^2 | X_1, ..., X_n) =$???

Our approach: $L_1$ case for $h = 2$

$$X_{n+1}^2 = \sigma_{n+1}^2 \epsilon_{n+1}^2 ; \sigma_{n+1}^2 = \alpha + a_1 X_n^2 + ... + a_p X_{n-p+1}^2$$

$$X_{n+2}^2 = \sigma_{n+2}^2 \epsilon_{n+2}^2 ; \sigma_{n+2}^2 = \alpha + a_1 X_{n+1}^2 + ... + a_p X_{n-p+2}^2$$

Plug-in $X_{n+1}^2 = \sigma_{n+1}^2 \epsilon_{n+1}^2$ to get:

$$X_{n+2}^2 = (\alpha + a_1 \sigma_{n+1}^2 \epsilon_{n+1}^2 + a_2 X_n^2 + ... + a_p X_{n-p+2}^2) \epsilon_{n+2}^2$$

RHS: only $\epsilon_{n+1}$ and $\epsilon_{n+2}$ are unknown given $\{X_1, \ldots, X_n\}$.

Under the assumption $\epsilon_t \sim$ i.i.d. $N(0, 1)$ it may be possible to derive formula for the median.

Assuming $\epsilon_t \sim$ i.i.d. $N(0,1)$, we can compute

$$F_{X^2_{n+2}}(x) = \frac{\gamma(\frac{1}{2}, \frac{x}{2A})}{B\sqrt{\pi}}, \quad x \geq 0$$

$$Conditional\ Median \text{ of } X^2_{n+2} \approx \tfrac{\pi}{8} AB^2 + \pi^4 B^4$$

with $A = \alpha + a_2 X_n^2 + a_3 X_{n-1}^2 + ... + a_p X_{n-p+2}^2$,

$$B = a_1 \sigma_{n+1}^2 = a_1(\alpha + a_1 X_n^2 + ... + a_p X_{n-p+1}^2),$$

$$\gamma(s,x) = \int_0^x t^{s-1} \mathrm{e}^{-t} \mathrm{d}t, \quad s > 0$$

What happens if errors are NOT normal and/or $h > 2$??

## Bootstrap Algorithm to predict $g(X_{n+h})$

1. Derive $g(X_{n+h})$ as a function of $\{\epsilon_{n+1}, \ldots, \epsilon_{n+h}\}$ based on model assumptions (ARCH/GARCH/NoVaS).

2. Using Monte Carlo N times, each time generate $\{\epsilon^*_{n+1}, \ldots, \epsilon^*_{n+h}\}$ from $F_\epsilon$ (or $\widehat{F}_\epsilon$) and plug in the function obtained in step 1 to compute $\{g(X^{(1)}_{n+h}), \ldots, g(X^{(N)}_{n+h})\}$.

3. Calculate the predictor $g(\widehat{X_{n+h}})$ of $g(X_{n+h})$ by taking the sample median (if $L_1$ optimal) or mean (if $L_2$ optimal) of the replicates $\{g(X^{(1)}_{n+h}), \ldots, g(X^{(N)}_{n+h})\}$.

Recall the NoVaS transformation of Politis (2007) mapping $\underline{X}$ to $\underline{W}$

$$W_t = \frac{X_t}{\sqrt{a + a_0 X_t^2 + \sum_{i=1}^p a_i X_{t-i}^2}}$$

Choose $a$ and $a_0, a_1, \ldots, a_p$ to render $W_t$ i.i.d.

Simple NovaS: $a_0 = a_1 = \ldots = a_p$; choose $p$.

Exponential Novas: $a_i = ce^{-di}$; choose $d$.

NoVaS transformation can be inverted:

$$X_t^2 = \frac{W_t^2}{1 - a_0 W_t^2} \left( a + \sum_{i=1}^{p} a_i X_{t-i}^2 \right)$$

The above can be used as a “model” equation for prediction purposes.

**Simulation:** 500 data-sets are generated separately by 7 different GARCH(1,1) models. Compare the performance for point prediction and 95% prediction intervals.

Focus on Model 1: Standard GARCH(1,1) with finite 4th moment and Gaussian errors: $X_t = \sigma_t \epsilon_t,\ \sigma_t^2 = .00001 + .73\sigma_{t-1}^2 + .10X_{t-1}^2$, $\{\epsilon_t\} \sim i.i.d.\ N(0,1)$.

Simulation settings:

- Each dataset is of size $n = 100$
- up to 5-step ahead predictions
- Monte Carlo $N$=5000, Bootstrap $B$= 500
- Model-fitting:
  - GARCH(1,1)
  - Generalized Simple NoVaS (GS-NoVaS)
  - Generalized Exponential NoVaS (GE-NoVaS)

| $h=$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| GARCH | 4.84E-05 | 4.69E-05 | 5.18E-05 | 5.45E-05 | 5.99E- |
| $GS_{NoVaS}$ | 4.83E-05 | 4.67E-05 | 5.17E-05 | 5.44E-05 | 5.94E- |
| $GE_{NoVaS}$ | 4.84E-05 | 4.69E-05 | 5.25E-05 | 5.43E-05 | 6.00E- |

**MADs of $L_1$ predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and** $\{\epsilon_t\} \sim i.i.d.\ N(0,1)$.

| $h=$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| GARCH | 7.28E-09 | 7.47E-09 | 6.50E-09 | 7.64E-09 | 8.65E- |
| $GS_{NoVaS}$ | 7.12E-09 | 7.40E-09 | 6.41E-09 | 7.90E-09 | 8.43E- |
| $GE_{NoVaS}$ | 6.99E-09 | 7.44E-09 | 6.43E-09 | 7.68E-09 | 8.53E- |

**MSEs of $L_2$ predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and** $\{\epsilon_t\} \sim i.i.d.\ N(0,1)$.

| | L2 | | | L1 | | |
|---|---|---|---|---|---|---|
| | GARCH(1,1) | | | GARCH(1,1) | | |
| $h$ | CVR | LEN | ST.ERR | CVR | LEN | ST.ERR |
| 1 | 0.948 | 4.39E-03 | 2.11E-03 | 0.92 | 3.76E-03 | 3.18E-03 |
| 2 | 0.940 | 4.53E-03 | 2.23E-03 | 0.936 | 5.49E-03 | 5.21E-03 |
| 3 | 0.950 | 4.47E-03 | 2.74E-03 | 0.938 | 5.99E-03 | 5.52E-03 |
| 4 | 0.952 | 4.02E-03 | 1.89E-03 | 0.922 | 7.16E-03 | 6.31E-03 |
| 5 | 0.934 | 3.77E-03 | 2.74E-03 | 0.92 | 4.57E-03 | 4.21E-03 |

**Interval predictions for Data generated from GARCH(1,1) with** $\omega = .00001, \alpha = .73, \theta = .10$ **and** $\epsilon \sim i.i.d\ N(0,1)$.

| | L2 | | | L1 | | |
|---|---|---|---|---|---|---|
| | GS-NoVaS | | | GS-NoVaS | | |
| $h$ | CVR | LEN | ST.ERR | CVR | LEN | ST.ERR |
| 1 | 0.949 | 4.37E-03 | 2.53E-03 | 0.946 | 4.26E-03 | 2.37E-03 |
| 2 | 0.948 | 4.63E-02 | 2.78E-03 | 0.95 | 4.26E-03 | 2.33E-03 |
| 3 | 0.938 | 4.17E-03 | 2.59E-03 | 0.95 | 4.22E-03 | 2.42E-03 |
| 4 | 0.945 | 3.76E-03 | 2.71E-03 | 0.948 | 4.20E-03 | 1.91E-03 |
| 5 | 0.950 | 4.58E-03 | 2.49E-03 | 0.948 | 4.19E-03 | 2.32E-03 |

**Interval predictions for Data generated from GARCH(1,1) with** $\omega = .00001, \alpha = .73, \theta = .10$ **and** $\epsilon \sim i.i.d\ N(0,1)$.

| | L2 | | | L1 | | |
|---|---|---|---|---|---|---|
| | GE-NoVaS | | | GE-NoVaS | | |
| $h$ | CVR | LEN | ST.ERR | CVR | LEN | ST.ERR |
| 1 | 0.946 | 4.92E-03 | 2.57E-03 | 0.96 | 5.37E-03 | 2.00E-03 |
| 2 | 0.946 | 4.68E-03 | 2.38E-03 | 0.948 | 5.13E-03 | 3.34E-03 |
| 3 | 0.958 | 4.39E-03 | 2.35E-03 | 0.952 | 4.03E-03 | 2.05E-03 |
| 4 | 0.954 | 4.30E-03 | 2.03E-03 | 0.95 | 4.80E-03 | 2.03E-03 |
| 5 | 0.948 | 4.15E-03 | 2.93E-03 | 0.944 | 4.42E-03 | 2.78E-03 |

**Interval predictions for Data generated from GARCH(1,1) with** $\omega = .00001, \alpha = .73, \theta = .10$ **and** $\epsilon \sim i.i.d\ N(0,1)$.

## Extension: multi-step ahead prediction of nonlinear autoregressions

For some fixed $p \geq 0$, consider a model equation

$$Y_t = f(Y_{t-1}, \ldots, Y_{t-p}; \epsilon_t) \text{ for } t \in Z \quad (1)$$

where $f$ is some (typically unknown) function, $\{\epsilon_t, t \in Z\}$ are iid with cdf $F$, and $\epsilon_t$ is independent of $\{Y_s, s < t\}$. Denote $X_{t-1} = (Y_{t-1}, \ldots, Y_{t-p})'$.

A first aim is to construct an optimal predictor of $Y_{n+1}$ given observed data $Y_1, \ldots, Y_n$ with $n > p$. Eq. (1) implies: $Y_{n+1} = f(X_n; \epsilon_{n+1})$.

- Hence, the $L_2$–optimal predictor of $Y_{n+1}$ is $E(Y_{n+1}|Y_1, \ldots, Y_n) = \int f(X_n; \epsilon) \, dF(\epsilon)$.

- Similarly, the $L_1$–optimal predictor of $Y_{n+1}$ is $M(Y_{n+1}|Y_1,\ldots,Y_n)$ where $M$ denotes Median. In view of (1), we have $M(Y_{n+1}|Y_1,\ldots,Y_n) = M(Y_{n+1}|X_n)$ which is the Median of $f(X_n;\epsilon)$ when $X_n$ is kept fixed and $\epsilon$ takes values according to $F$.

Both $E(Y_{n+1}|Y_1,\ldots,Y_n)$ and $M(Y_{n+1}|Y_1,\ldots,Y_n)$ can easily be computed by Monte Carlo: (i) Let $N$ be a large integer (say 1,000) and draw $N$ values of $\epsilon$ from $F$; (ii) compute the $N$ values of the quantity $f(X_n;\epsilon)$, and (iii) take the sample mean or median of these $N$ values.

If $f$ and $F$ are unknown, consistent estimates (say $\widehat{f}$ and $\widehat{F}$) can used in the above procedures for either $L_1$ or $L_2$–optimal predictor.

The aim now is to construct an optimal $h$-step ahead predictor, i.e., predict $Y_{n+h}$ given observed data $Y_1, \ldots, Y_n$ with $n > p$.

Let $h = 2$; then

- the $L_2$–optimal predictor of $Y_{n+2}$ is $E(Y_{n+2}|Y_1, \ldots, Y_n)$.

- Similarly, the $L_1$–optimal predictor of $Y_{n+2}$ is $M(Y_{n+2}|Y_1, \ldots, Y_n)$ .

Neither conditional expectation or median can be computed in closed-form.

Eq. (1) implies: $Y_{n+1} = f(X_n; \epsilon_{n+1})$ and

$$Y_{n+2} = f(X_{n+1}; \epsilon_{n+2})$$

$$= f(Y_{n+1}, Y_n, \ldots, Y_{n+1-p}; \epsilon_{n+2})$$

$$= f\left(f(X_n; \epsilon_{n+1}), Y_n, \ldots, Y_{n+1-p}; \epsilon_{n+2}\right). \tag{2}$$

Hence, the conditional expectation (or the conditional median) can be computed by Monte Carlo simulation, namely: (i) draw $2N$ values of $\epsilon$ from $F$; (ii) use half as replicates of $\epsilon_{n+1}$, and the other half as replicates of $\epsilon_{n+2}$; (iii) plug them in expression (2) to yield $N$ replicates of values from the conditional distribution of $Y_{n+2}$ given $Y_1, \ldots, Y_n$; and (iii) take the sample mean (or sample median) of the $N$ replicates of $Y_{n+2}$.