Financial Time Series – Forecasting and evaluation revisited

Andreas Petersson

TMS088/MSA410 – May 2020

Mathematical Sciences, Chalmers University of Technology & University of Gothenburg, Sweden

Parametric bootstrap/Monte Carlo simulation

Given $(X_1, X_2, \ldots, X_n)$, forecast X_{n+h} for some $h > 0$. Denote this by $X_n(h)$. Parametric bootstrap computes forecasts of $X_{n+1}, \ldots, X_{n+h}$ sequentially. For $i = 1, \ldots, h$ repeat:

1. Generate a random sample of the driving noise at time $n + i$ according to model.

Parametric bootstrap/Monte Carlo simulation

Given $(X_1, X_2, \ldots, X_n)$, forecast X_{n+h} for some $h > 0$. Denote this by $X_n(h)$. Parametric bootstrap computes forecasts of $X_{n+1}, \ldots, X_{n+h}$ sequentially. For $i = 1, \ldots, h$ repeat:

1. Generate a random sample of the driving noise at time $n + i$ according to model.
2. Compute $\tilde{X}_{n+i}$ using sample, model, data, and previous forecasts $X_n(1), \ldots, X_n(i-1)$.

Parametric bootstrap/Monte Carlo simulation

Given $(X_1, X_2, \ldots, X_n)$, forecast X_{n+h} for some $h > 0$. Denote this by $X_n(h)$. Parametric bootstrap computes forecasts of $X_{n+1}, \ldots, X_{n+h}$ sequentially. For $i = 1, \ldots, h$ repeat:

1. Generate a random sample of the driving noise at time $n + i$ according to model.
2. Compute $\tilde{X}_{n+i}$ using sample, model, data, and previous forecasts $X_n(1), \ldots, X_n(i-1)$.
3. Repeat the previous two steps K times to get K realizations $(\tilde{X}_{n+i}^{(k)}, k = 1, \ldots, K)$. Set $X_n(i) = K^{-1} \sum_{k=1}^{K} \tilde{X}_{n+h}^{(k)}$.

Parametric bootstrap/Monte Carlo simulation

Given $(X_1, X_2, \ldots, X_n)$, forecast X_{n+h} for some $h > 0$. Denote this by $X_n(h)$. Parametric bootstrap computes forecasts of $X_{n+1}, \ldots, X_{n+h}$ sequentially. For $i = 1, \ldots, h$ repeat:

1. Generate a random sample of the driving noise at time $n+i$ according to model.
2. Compute $\tilde{X}_{n+i}$ using sample, model, data, and previous forecasts $X_n(1), \ldots, X_n(i-1)$.
3. Repeat the previous two steps K times to get K realizations $(\tilde{X}_{n+i}^{(k)}, k = 1, \ldots, K)$. Set $X_n(i) = K^{-1} \sum_{k=1}^{K} \tilde{X}_{n+h}^{(k)}$.

If model is adequate, $X_n(h) \approx \mathbb{E}(X_{n+h} | X_n, X_{n-1}, \ldots, X_1)$.

Example: Markov switching AR(p) model

Let S be a Markov chain taking values in $\{1, 2\}$ with

$$P(S_t = 2 | S_{t-1} = 1) = w_1,$$
$$P(S_t = 1 | S_{t-1} = 2) = w_2$$

with $w_1, w_2 \in [0, 1]$. $X = (X_t, t \in \mathbb{Z})$ follows a *Markov switching autoregressive model* (MSA) with two states if

$$X_t = \begin{cases} c_1 + \sum_{i=1}^{p} \phi_{1i} X_{t-i} + Z_{1t} & \text{if } S_t = 1, \\ c_2 + \sum_{i=1}^{p} \phi_{2i} X_{t-i} + Z_{2t} & \text{if } S_t = 2. \end{cases}$$

Here $Z_1 = (Z_{1t}, t \in \mathbb{Z})$ and $Z_2 = (Z_{2t}, t \in \mathbb{Z})$ are $\mathrm{IID}(0, \sigma^2)$ noises for finite σ^2 and independent of each other.

Forecasting evaluation

- Subdivide $(X_1, X_2, \ldots, X_N)$, into: $(X_1, X_2, \ldots, X_n)$, (*training subsample* or *estimation subsample*) and $(X_{n+h}, X_{n+h+2}, \ldots, X_N)$ (*test subsample* or *forecasting subsample*)

Forecasting evaluation

- Subdivide $(X_1, X_2, \ldots, X_N)$, into: $(X_1, X_2, \ldots, X_n)$, (*training subsample* or *estimation subsample*) and $(X_{n+h}, X_{n+h+2}, \ldots, X_N)$ (*test subsample* or *forecasting subsample*)
- *Rolling forecasting procedure:* $(X_1, X_2, \ldots, X_n)$ is used to compute $X_n(h)$, $(X_1, X_2, X_3, \ldots, X_{n+1})$ is used to compute $X_{n+1}(h)$ and so on

Forecasting evaluation

- Subdivide $(X_1, X_2, \ldots, X_N)$, into: $(X_1, X_2, \ldots, X_n)$, (*training subsample* or *estimation subsample*) and $(X_{n+h}, X_{n+h+2}, \ldots, X_N)$ (*test subsample* or *forecasting subsample*)
- *Rolling forecasting procedure:* $(X_1, X_2, \ldots, X_n)$ is used to compute $X_n(h)$, $(X_1, X_2, X_3, \ldots, X_{n+1})$ is used to compute $X_{n+1}(h)$ and so on
- $m = N - n - h + 1$ the size of the test subsample

Forecasting evaluation

Method (Directional measure)

Contingency table that summarizes “hits” and “misses” of predicting ups and downs up of X_{n+h} in the test subsample:

Actual \ Predicted	Up	Down	
Up	m_{11}	m_{12}	m_{10}
Down	m_{21}	m_{22}	m_{20}
	m_{01}	m_{02}	m

Calculate row sums and column sums. Larger values in m_{11} and m_{22} indicate better forecasts.

Method (Directional measure)

Contingency table that summarizes "hits" and "misses" of predicting ups and downs up of X_{n+h} in the test subsample:

Actual \ Predicted	Up	Down	
Up	m_{11}	m_{12}	m_{10}
Down	m_{21}	m_{22}	m_{20}
	m_{01}	m_{02}	m

Calculate row sums and column sums. Larger values in m_{11} and m_{22} indicate better forecasts. Large values of the test statistic

$$\chi^2 := \sum_{i,j=1}^{2} \frac{(m_{ij} - m_{i0}m_{0j}/m)^2}{m_{i0}m_{0j}/m}$$

signifies that the model outperforms the chance of random choice. Under mild assumptions, $\chi^2 \sim \chi^2_1$.

Forecasting evaluation

Method (Magnitude measure)

Three statistics for forecasting performance:

- the *mean squared error*

$$\mathsf{MSE}(h) := m^{-1} \sum_{j=0}^{m-1} (X_{n+h+j} - X_{n+j}(h))^2$$

Forecasting evaluation

Method (Magnitude measure)

Three statistics for forecasting performance:

- the *mean squared error*

$$\mathsf{MSE}(h) := m^{-1} \sum_{j=0}^{m-1} (X_{n+h+j} - X_{n+j}(h))^2$$

- the *mean absolute deviation*

$$\mathsf{MAD}(h) := m^{-1} \sum_{j=0}^{m-1} |X_{n+h+j} - X_{n+j}(h))|$$

Forecasting evaluation

Method (Magnitude measure)

Three statistics for forecasting performance:

- the *mean squared error*

$$\mathsf{MSE}(h) := m^{-1} \sum_{j=0}^{m-1} (X_{n+h+j} - X_{n+j}(h))^2$$

- the *mean absolute deviation*

$$\mathsf{MAD}(h) := m^{-1} \sum_{j=0}^{m-1} |X_{n+h+j} - X_{n+j}(h))|$$

- the *mean absolute percentage error*

$$\mathsf{MAPE}(h) := m^{-1} \sum_{j=0}^{m-1} \left| \frac{X_{n+j}(h)}{X_{n+h+j}} - 1 \right|$$

Forecasting evaluation

- If X is a continuous random variable with cumulative distribution function F_X, then the distribution of the random variable $Y = F_X(X) \sim \mathcal{U}([0,1])$

Forecasting evaluation

- If X is a continuous random variable with cumulative distribution function F_X, then the distribution of the random variable $Y = F_X(X) \sim \mathcal{U}([0,1])$
- A *distributional measure* (goodness of fit): for X_{n+h+j}, $j = 0, \ldots, m-1$, in the test set, compute *empirical CDF* $\hat{F}$

$$\hat{F}_j(y) = \frac{1}{K} \sum_{k=1}^{K} I(\tilde{X}^{(k)}_{n+h+j} \leq y)$$

out of the parametric bootstrap sample.

Forecasting evaluation

- If X is a continuous random variable with cumulative distribution function F_X, then the distribution of the random variable $Y = F_X(X) \sim \mathcal{U}([0,1])$
- A *distributional measure* (goodness of fit): for X_{n+h+j}, $j = 0, \ldots, m-1$, in the test set, compute *empirical CDF* $\hat{F}$

$$\hat{F}_j(y) = \frac{1}{K} \sum_{k=1}^{K} I(\tilde{X}^{(k)}_{n+h+j} \leq y)$$

out of the parametric bootstrap sample. Use the test set to compute

$$u_{n+j}(h) := \hat{F}_j(X_{n+h+j})$$

for all $j = 0, \ldots, m-1$

Forecasting evaluation

For sufficiently large m, the Kolmogorov–Smirnov statistic

$$D = \sup_{x \in [0,1]} \left| \frac{1}{m} \sum_{j=0}^{m-1} I(u_{n+j}(h) \leq x) - x \right|$$

can be used to test the sample with respect to the uniform distribution. The (asymptotic) distribution for this statistic is complex but if the model is adequate the statistic D should be small. This fact can be used to choose between several models.