Forecast congruence, accuracy, and smoothing parameter shrinkage

Kandrika Pritularga

2024 Dijon ISF

3 July 2024

Marketing Analytics and Forecasting

Lancaster University Management School

Forecasting and decisions

Managers use forecasts to make decisions:

- Producing forecasts;
- Making decisions based on forecasts.

It is often difficult to merge these activities into a single activity.

The literature argues that accurate forecasts lead to desirable decisions.

However, this might not always be the case.

Forecasting and decisions

Pritularga and Kourentzes (2024) propose a new measure to connect forecasting with inventory decisions – **forecast congruence**

It measures the variance of forecasts for a specific date across forecast origins.

Forecasting and decisions

Pritularga and Kourentzes (2024) propose a new measure to connect forecasting with inventory decisions – **forecast congruence**

It measures the variance of forecasts for a specific date across forecast origins.

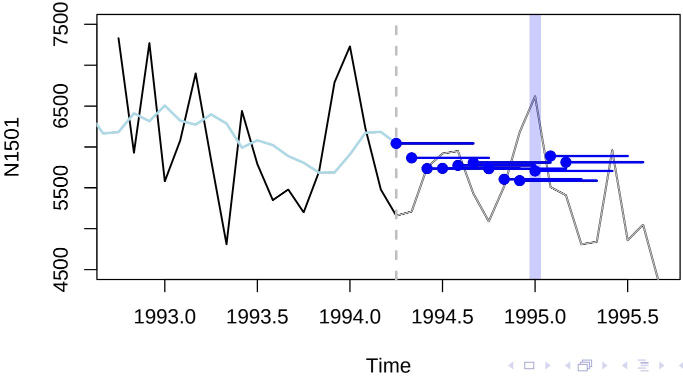
Let's forecast sales of double espresso tomorrow.

We produce many forecasts from many origins, e.g., from the last week to yesterday; from different step ahead forecasts.

The paper argues that we need to get congruent forecasts that do not harm accuracy.

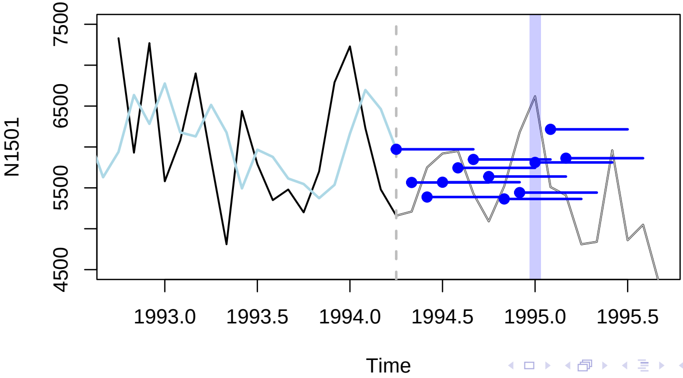
Forecasting and decisions

ETS(A,N,N) with $\alpha = 0.2$



Forecasting and decisions

ETS(A,N,N) with $\alpha = 0.5$



Forecasting and decisions

How can we achieve this?

Forecasting and decisions

How can we achieve this?

- Do not re-estimate the forecasting models across origins (ad-hoc approach);
- Multiple temporal aggregation approach (e.g., MAPA, THieF, (Kourentzes et al., 2014; Athanasopoulos et al., 2017));
- Exponential smoothing (ETS) with shrinkage estimators.
- Machine learning approach (Van Belle et al., 2023).

Changing the loss function to produce 'more stable' or congruent forecasts (Van Belle et al., 2023; Godahewa et al., 2023).

ETS and decisions

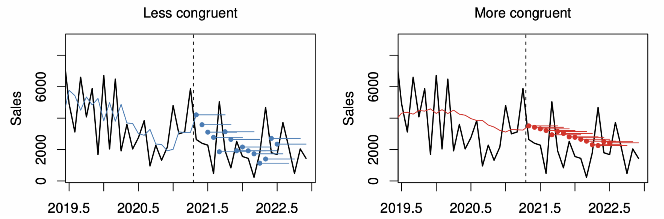
ETS shrinkage offers a way to control the value of the smoothing parameters.

ETS and decisions

ETS shrinkage offers a way to control the value of the smoothing parameters.

This then makes the states (level, trend, or seasons) tend to be less stochastic.

Then, the forecasts tend to be more congruent.



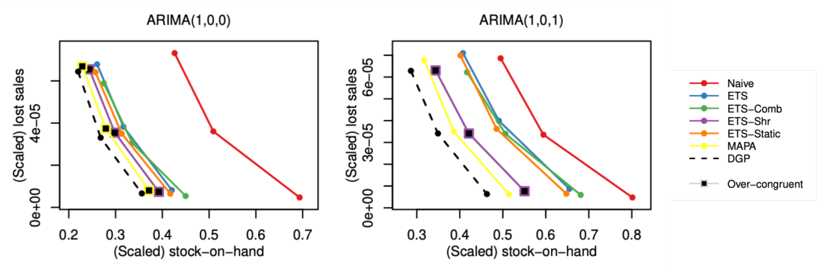
ETS and decisions

However, the previous study shows that ETS shrinkage tends to make over-congruent forecasts – which harms the accuracy.

ETS and decisions

However, the previous study shows that ETS shrinkage tends to make over-congruent forecasts – which harms the accuracy.

This leads to less desirable inventory decision metrics such as lost sales and service levels.



ETS and forecast congruence

This presentation will discuss the drawback of a mechanism in finding a hyperparameter in ETS with shrinkage estimators Pritularga et al. (2023).

ETS and forecast congruence

This presentation will discuss the drawback of a mechanism in finding a hyperparameter in ETS with shrinkage estimators Pritularga et al. (2023).

The previous study demonstrates that it is important to maintain accuracy

ETS and forecast congruence

This presentation will discuss the drawback of a mechanism in finding a hyperparameter in ETS with shrinkage estimators Pritularga et al. (2023).

The previous study demonstrates that it is important to maintain accuracy

...but also congruent enough so that the decision makers can benefit from the forecasts.

ETS and forecast congruence

This presentation will discuss the drawback of a mechanism in finding a hyperparameter in ETS with shrinkage estimators Pritularga et al. (2023).

The previous study demonstrates that it is important to maintain accuracy

...but also congruent enough so that the decision makers can benefit from the forecasts.

From here forecast congruence is our proxy for decisions

ETS and forecast congruence

Let's take a simple example of ETS(ANN)

$$y_t = l_{t-1} + \varepsilon_t$$
$$l_t = l_{t-1} + \alpha\varepsilon_t$$

where α is the smoothing parameter and it controls the stochasticity of the local mean. l_0 is the level initial value and it initialises the time series.

We need to estimate the parameters of α and l_0 with any estimators (MSE, MAE, MLE, etc.).

ETS and forecast congruence

The available shrinkage estimators implement 'LASSO' and 'RIDGE'-like estimators

$$\text{LASSO} = (1-\lambda)\text{MSE}_{t+1|t} + \lambda \sum_{j=1}^{K} |g_i| \tag{1}$$

$$\text{RIDGE} = (1-\lambda)\text{MSE}_{t+1|t} + \lambda \sum_{j=1}^{K} g_i^2 \tag{2}$$

where $g_i \in \boldsymbol{g}$, λ is a hyperparameter and λ is between 0 and 1.

ETS and forecast congruence

The available shrinkage estimators implement 'LASSO' and 'RIDGE'-like estimators

$$\text{LASSO} = (1-\lambda)\text{MSE}_{t+1|t} + \lambda \sum_{j=1}^{K} |g_i| \tag{1}$$

$$\text{RIDGE} = (1-\lambda)\text{MSE}_{t+1|t} + \lambda \sum_{j=1}^{K} g_i^2 \tag{2}$$

where $g_i \in \boldsymbol{g}$, λ is a hyperparameter and λ is between 0 and 1.

This means that we shrink the smoothing parameters (g_i) to zero and putting the importance of the initial values in ETS. Note: the dampening parameter is shrunk to 1.

ETS and forecast congruence

This shrinkage estimator gives us the flexibility to control the smoothing parameters,

ETS and forecast congruence

This shrinkage estimator gives us the flexibility to control the smoothing parameters,

... at the cost of having to find a hyperparameter.

ETS and forecast congruence

This shrinkage estimator gives us the flexibility to control the smoothing parameters,

... at the cost of having to find a hyperparameter.

The current implementation relies on the one-step ahead MSE in the validation set:

$$\mathcal{L}(\lambda, \hat{\boldsymbol{g}}) = \min_{\lambda} \frac{1}{L} \sum_{l=1}^{L} \mathsf{MSE}(\hat{y}_{t+1} | \mathcal{I}_{t,l}, \hat{\boldsymbol{g}})$$

where L is the number of forecast origins and $\mathcal{I}_{t,l}$ is the information at time t and origin l.

ETS and forecast congruence

This shrinkage estimator gives us the flexibility to control the smoothing parameters,

... at the cost of having to find a hyperparameter.

The current implementation relies on the one-step ahead MSE in the validation set:

$$\mathcal{L}(\lambda, \hat{\boldsymbol{g}}) = \min_{\lambda} \frac{1}{L} \sum_{l=1}^{L} \mathsf{MSE}(\hat{y}_{t+1} | \mathcal{I}_{t,l}, \hat{\boldsymbol{g}})$$

where L is the number of forecast origins and $\mathcal{I}_{t,l}$ is the information at time t and origin l. What does it mean?

ETS and forecast congruence

This means that λ is optimised for a very short forecast horizon.

ETS and forecast congruence

This means that λ is optimised for a very short forecast horizon.

This does not align well with the nature of forecast congruence – **multistep forecast errors**.

$$
\begin{array}{cc}
\begin{pmatrix}
e_{1+l|l} & e_{1+l|l-1} & \cdots & e_{1+l|1+l-(l-1)} & e_{1+l|1} \\
e_{2+l|1+l} & e_{2+l|l} & \cdots & e_{2+l|2+l-(l-1)} & e_{1+l|2} \\
\vdots & \ddots & & & \vdots \\
 & & e_{t|t-h} & & \\
\vdots & & & \ddots & \vdots \\
e_{n-1|n-2} & e_{n-1|n-3} & \cdots & e_{n-1|n-1-(l-1)} & e_{n-1|n-1-l} \\
e_{n|n-1} & e_{n|n-2} & \cdots & e_{n|n-(l-1)} & e_{n|n-l}
\end{pmatrix}
&
\begin{array}{l}
\rightarrow \tau^2_{l+1} \\
\rightarrow \tau^2_{l+2} \\
\\
\quad \vdots \\
\\
\rightarrow \tau^2_{n-1} \\
\rightarrow \tau^2_{n}
\end{array}
\\
\begin{array}{ccccc}
\downarrow & \downarrow & & \downarrow & \downarrow \\
\sigma^2_1 & \sigma^2_2 & \cdots & \sigma^2_{l-1} & \sigma^2_l
\end{array}
&
\end{array}
$$

Research problem

Research problem

How do we estimate the hyperparameter of ETS shrinkage estimators so that the forecasts are accurate and congruent enough?

Research problem

Research problem

How do we estimate the hyperparameter of ETS shrinkage estimators so that the forecasts are accurate and congruent enough?

Steps in answering this question:

1. exploring the relationship between accuracy (MSE) and congruence (τ), conditional to different circumstances;
2. using the exploration findings to propose a new loss function for finding the hyperparameter.

Forecast congruence

Let's define forecast congruence. We are interested in collecting a set of forecasts from different origins and h for a specific t.

Forecast congruence

Let's define forecast congruence. We are interested in collecting a set of forecasts from different origins and h for a specific t.

$$\tau = \frac{1}{n}\sum_{t=1}^{n}\left(\sqrt{\mathrm{var}_h(\hat{y}_{t|t-h})}\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n}\left(\sqrt{\mathrm{var}_h(y_t - e_{t|t-h})}\right)$$

$$\tau = \frac{1}{n}\sum_{t=1}^{n}\left(\sqrt{\mathrm{var}_h(e_{t|t-h})}\right)$$

τ measures the variance of $e_{t|t-h}$ over h

Accuracy & congruence: analytical approach

Random walk

$$\tau_t = \sqrt{\mathsf{E}\left[\sum_{i=1}^{h}(h-i+1)\varepsilon^2_{t-i+1}\right] - \mathsf{E}\left[\sum_{i=1}^{h}(h-1+1)\varepsilon_{t-i+1}\right]^2},$$

AR(1)

$$\tau_t = \sqrt{\mathsf{E}\left[\sum_{i=1}^{h}(h-i+1)\alpha^{2i}\varepsilon^2_{t-i+1}\right] - \mathsf{E}\left[\sum_{i=1}^{h}(h-i+1)\alpha^{i}\varepsilon_{t-i+1}\right]^2},$$

Accuracy & congruence: analytical approach

Random walk

$$\tau_t = \sqrt{\mathsf{E}\left[\sum_{i=1}^{h} (h-i+1)\varepsilon_{t-i+1}^2\right] - \mathsf{E}\left[\sum_{i=1}^{h} (h-1+1)\varepsilon_{t-i+1}\right]^2},$$

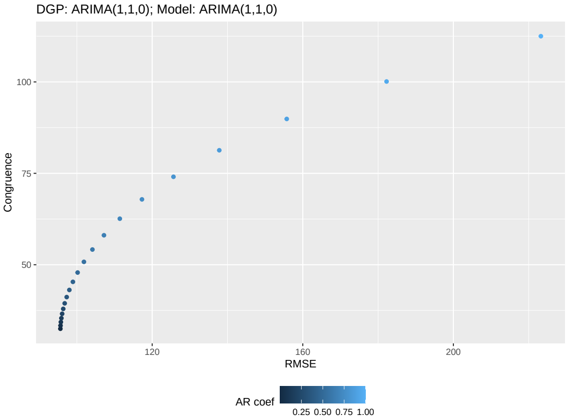
AR(1)

$$\tau_t = \sqrt{\mathsf{E}\left[\sum_{i=1}^{h}(h-i+1)\alpha^{2i}\varepsilon_{t-i+1}^2\right] - \mathsf{E}\left[\sum_{i=1}^{h}(h-i+1)\alpha^{i}\varepsilon_{t-i+1}\right]^2},$$

The positive relationship between τ and σ is not only mediated by the horizon but also the model parameters.

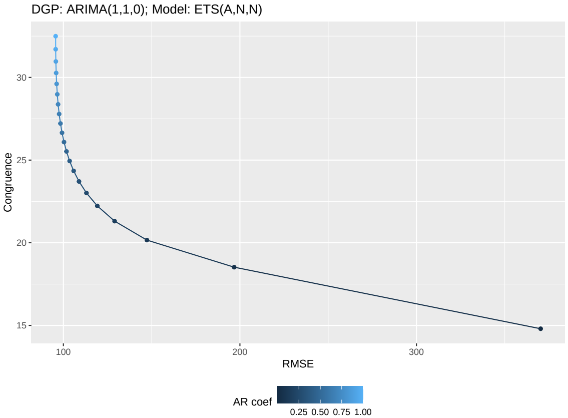
Accuracy & congruence: simulation approach

Correct models, 1000 repetitions



Accuracy & congruence: simulation approach

Incorrect models, 1000 repetitions



Multistep estimators

We need a metric that takes into account the 'multi-step' nature – **multistep estimator**

Multistep estimators

We need a metric that takes into account the 'multi-step' nature – **multistep estimator**

Literature has studied the efficacy of multistep loss function in forecasting (Svetunkov et al., 2024, and the reference therein).

Minimising h-step ahead forecast variance eventually shrinks the coefficients of SSOE-based models.

Multistep estimators

We need a metric that takes into account the ‘multi-step’ nature – **multistep estimator**

Literature has studied the efficacy of multistep loss function in forecasting (Svetunkov et al., 2024, and the reference therein).

Minimising h-step ahead forecast variance eventually shrinks the coefficients of SSOE-based models.

GMSE_h is the geometric mean of $\log \text{MSE}_j$ for $j \in h$ - with a slight modification.

$$\mathsf{GMSE_h} = \frac{1}{h}\sum_{j=1}^{h}\log\frac{1}{T-h}\sum_{t=1}^{T-h}e^2_{t+j|t} = \frac{1}{h}\sum_{j=1}^{h}\log \mathsf{MSE_j}$$

A composite loss: multistep accuracy + congruence

Let's balance between the multistep loss function and congruence as both have the same 'nature'.

A composite loss: multistep accuracy + congruence

Let's balance between the multistep loss function and congruence as both have the same 'nature'.

I propose a composite loss function:

$$\mathcal{L} = \min\left((1-\lambda)\mathsf{GMSE}_h + \lambda\mathsf{congruence}\right)$$

A composite loss: multistep accuracy + congruence

Let's balance between the multistep loss function and congruence as both have the same 'nature'.

I propose a composite loss function:

$$\mathcal{L} = \min\left((1-\lambda)\mathsf{GMSE}_h + \lambda \mathsf{congruence}\right)$$

How do we find the 'optimal' λ? Note: $0 \leq \lambda < 1$

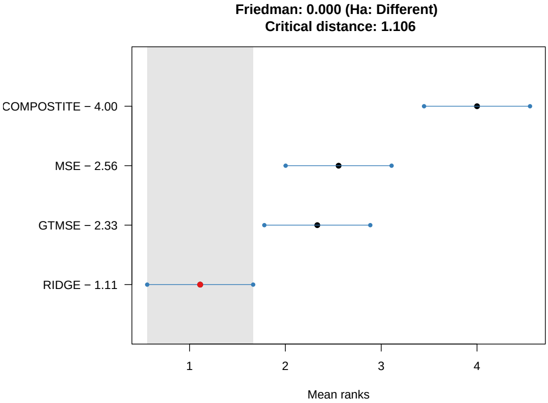
1. Split the dataset into (a) training and (b) validation set.
2. Estimate ETS parameters with a composite loss function for each λ.
3. Measure the composite loss in the validation set.
4. Find λ that results in the lowest composite loss.

Empirical analysis

- Dataset: M3 dataset for monthly only.
- Split the training set into an in-sample and a validation set (6 origins).
- The model structure is set to be incorrect: ETS(M,A,M)
- Forecast horizon: 18.
- MSE as the benchmark.
- Evaluated the point forecast accuracy for 1-h step ahead in the test set using RMSSE and scaled AME.
- Include τ that is calculated in the validation set.

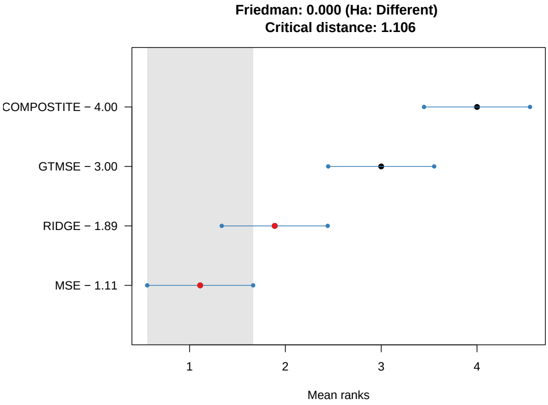
Empirical analysis

Accuracy – MSE (across hi



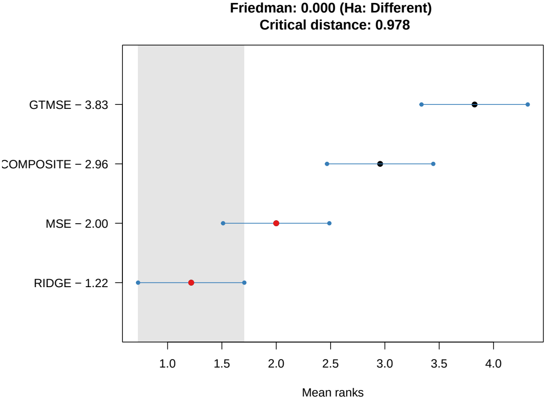
Empirical analysis

Bias – AME



Empirical analysis

Congruence – τ



Discussion & conclusion

RIDGE and MSE perform well for accuracy, bias, and congruence when the model structure is incorrectly estimated.

Discussion & conclusion

RIDGE and MSE perform well for accuracy, bias, and congruence when the model structure is incorrectly estimated.

This experiment is motivated by the fact that we often have to estimate an incorrect model structure.

The experiment design could have been improved by providing the case when the model structure is estimated via an information criterion.

Discussion & conclusion

RIDGE and MSE perform well for accuracy, bias, and congruence when the model structure is incorrectly estimated.

This experiment is motivated by the fact that we often have to estimate an incorrect model structure.

The experiment design could have been improved by providing the case when the model structure is estimated via an information criterion.

The data-driven approach to find λ is not suitable in this particular situation.

Discussion & conclusion

Congruence offers a theoretical understanding of the ‘good’ forecasts, where forecasts agree with each other.

Forecasting models/ methods that have the smoothing feature on them:

- MAPA;
- ETS shrinkage;
- Forecast combination.

Discussion & conclusion

Congruence offers a theoretical understanding of the ‘good’ forecasts, where forecasts agree with each other.

Forecasting models/ methods that have the smoothing feature on them:

- MAPA;
- ETS shrinkage;
- Forecast combination.

However, more questions than answers...

Discussion & conclusion

The congruence measures the balance between the column-wise and the row-wise of the forecast error matrix.

- Can we create a ‘seemingly-like ’covariance matrix of the multistep errors that contains accuracy and congruence?
  - a multistep error matrix has a dimension of $n \times h$, where n is the number of the origins and h is the forecast horizon
  - the covariance matrix of the multistep error will have a dimension of $n \times n$
- But then τ can have a matrix of $h \times h$ – does τ have another version of the covariance matrix?

Is λ too restrictive? Should we have a different way of formulating the hyperparameter loss function?

Why not forecast stability?

The concept of stability has been well known in time series analysis – ARIMA concept!

The stability feature is attached to a statistical forecasting model, which explains the long-term behaviour of a model.

Why not forecast stability?

The concept of stability has been well known in time series analysis – ARIMA concept!

The stability feature is attached to a statistical forecasting model, which explains the long-term behaviour of a model.

... which is often related to a stationary ARMA.

Why not forecast stability?

The concept of stability has been well known in time series analysis – ARIMA concept!

The stability feature is attached to a statistical forecasting model, which explains the long-term behaviour of a model.

... which is often related to a stationary ARMA.

Nonetheless, we consider a few alternatives until we find a 'free' word.

Why not forecast stability?

The concept of stability has been well known in time series analysis – ARIMA concept!

The stability feature is attached to a statistical forecasting model, which explains the long-term behaviour of a model.

... which is often related to a stationary ARMA.

Nonetheless, we consider a few alternatives until we find a ‘free’ word.

Congruence here means ‘agreement’ between forecasts, due to updating information (re-estimate each model on each origin).

Thank you for your attention!

Kandrika Pritularga

k.pritularga@lancaster.ac.uk

Linkedin: Kandrika Pritularga

Marketing Analytics and Forecasting | Lancaster University Management School

References I

Athanasopoulos, G., Hyndman, R.J., Kourentzes, N., Petropoulos, F., 2017. Forecasting with temporal hierarchies. European journal of operational research 262, 60–74.

Godahewa, R., Bergmeir, C., Baz, Z.E., Zhu, C., Song, Z., García, S., Benavides, D., 2023. On forecast stability. URL: https://arxiv.org/abs/2310.17332, arXiv:2310.17332.

Kourentzes, N., Petropoulos, F., Trapero, J.R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. International journal of forecasting 30, 291–302.

References II

Pritularga, K., Kourentzes, N., 2024. Forecast congruence: a quantity to bridge between forecasts and inventory decisions. SSRN , 1–40.

Pritularga, K.F., Svetunkov, I., Kourentzes, N., 2023. Shrinkage estimator for exponential smoothing models. International journal of forecasting 39, 1351–1365.

Svetunkov, I., Kourentzes, N., Killick, R., 2024. Multi-step estimators and shrinkage effect in time series models. Computational statistics 39, 1203–1239.

Van Belle, J., Crevits, R., Verbeke, W., 2023. Improving forecast stability using deep learning. International journal of forecasting 39, 1333–1350.

Forecast congruence

$e_{t|t-h}$ contains the innovation at time t (ε_t) and the 'other' errors $\eta_{t|t-h}$ due to incorrect model structures and parameters.

$$\begin{aligned}
\tau &= \frac{1}{n}\sum_{t=1}^{n}\left(\sqrt{\mathsf{var}_h(\hat{y}_{t|t-h})}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\sqrt{\mathsf{var}_h(y_t - e_{t|t-h})}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\sqrt{\mathsf{var}_h(\eta_{t|t-h} + \varepsilon_t)}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\sqrt{\mathsf{var}_h(\eta_{t|t-h})}\right)
\end{aligned}$$

τ eventually measures the variance of misspesification errors