Decreasing uncertainty in decision making using trace forecast likelihood

Ivan Svetunkov

Presentation at Higher School of Economics, Saint Petersburg

25th May 2016

Lancaster University Management School
Lancaster Centre for Forecasting

LCF

Motivation

Chatfield, 1995 argues that there are three major sources of uncertainty:

1. Structure of model;
2. Estimates of parameters;
3. White noise.

Motivation

(3) is irreducible,

(2) can be tackled when there is a known model or an appropriate estimation method,

(1) is the cause of the largest problems.

Conventional Estimation methods

The conventional estimation methods is based on MSE:

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T} e_{t+1|t}^2 \tag{1}$$

where $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1}$

MSE – “Mean Squared Error”.

If errors in a model are distributed normally, than using (1) is equivalent to maximising the following log-likelihood function:

$$\ell(\theta, \hat{\sigma}^2 | Y) = -\frac{T}{2} \left(\log(2 \pi e) + \log \hat{\sigma}^2 \right) \tag{2}$$

where $\hat{\sigma}^2$ is the estimated variance of residuals of the model, θ is a vector of parameters of the model.

This implies that we look at conditional distribution of one-step-ahead forecast error.

Advanced estimation methods

Sometimes the forecasting task is aligned to estimation:

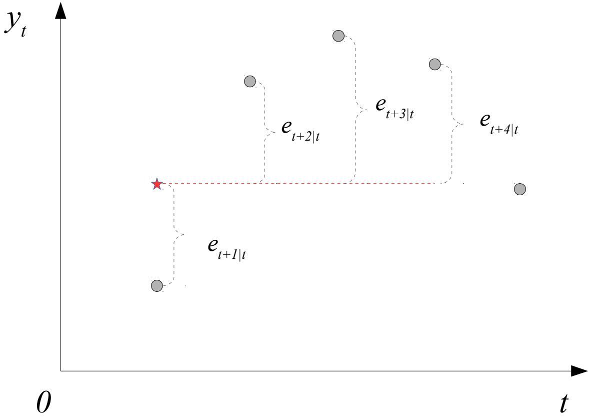
$$\text{MSE}_h = \frac{1}{T} \sum_{t=1}^{T} e_{t+h|t}^2 \tag{3}$$

or:

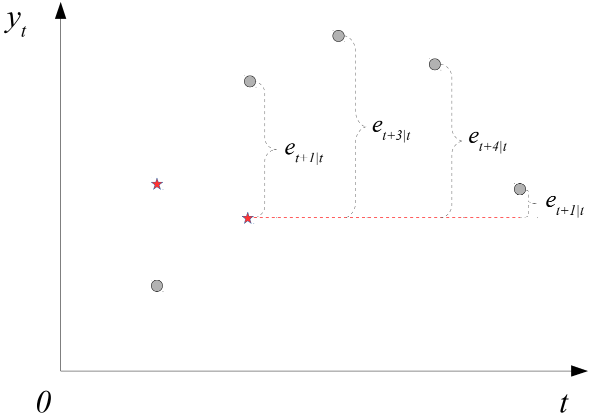
$$\text{MSTFE} = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{h} e_{t+j|t}^2 \tag{4}$$

MSTFE – "Mean Squared Trace Forecast Error".

These cost functions imply that we produce h-steps ahead forecasts from each observation:



These cost functions imply that we produce h-steps ahead forecasts from each observation:



Literature tells us...

MSE_h produces robust estimates of parameters.

(McElroy and Wildi, 2013, Tiao and Xu, 1993, Clements and Hendry, 2008)

The forecast accuracy increases.

(A. Weiss and Andersen, 1984, A. A. Weiss, 1991, Taylor, 2008, Xia and Tong, 2011)

MSTFE is consistent.

(A. A. Weiss, 1991)

BUT!

The efficiency of estimates of MSE_h is low.

(Tiao and Xu, 1993)

Marcellino, Stock and Watson, 2006 demonstrate on a set of 170 time series that the forecast accuracy using MSE_h is lower than using MSE.

Problems:

- The results are ambiguous;
- Estimates of parameters are inefficient;
- Estimates of parameters could be unstable;
- Nobody has ever explained why MSE$_h$ and MSTFE work / don't work;

- There is no likelihood function for both MSE$_h$ and MSTFE;
- Model selection using MSE$_h$ and MSTFE is really tricky;

LCF

Why multiple steps methods work / don't work

It can be shown that MSE is proportional to variance of one-step-ahead error:

$$\text{MSE} = \sigma_1^2 + V(\hat{y}_{t+1}) + E(\hat{y}_{t+1})^2 \quad (5)$$

MSE_h is then proportional to variance of h-step-ahead error.

MSTFE is in fact the sum of MSE_h .

And variance of h-step-ahead error consists of two parts (Hyndman, Koehler, Ord and Snyder, 2008):

1. variance of one-step-ahead error,
2. parameters of model.

Example ETS(ANN): $\sigma_h^2 = \sigma_1^2 \left(1 + \sum_{j=1}^{h-1} \alpha^2\right)$.

This means that minimising MSE_h (or MSTFE) leads to:

1. decrease of variance of one-step-ahead error,
2. shrinkage of values of smoothing parameters towards zero,

This is the root of the problem and the main advantage of MSE_h and MSTFE.

If model is wrong, shrinkage allows to get rid of redundant parameters.

If model is correct, the parameters “overshrink”.

The shrinkage effect becomes stronger when h increases.
This allows to decrease the uncertainty, caused by model structure.

But it may increase the uncertainty caused by estimates of parameters.

Solution – Trace Forecast Likelihood (TFL)

Let's derive likelihood for multistep cost function.
We need to study multivariate distribution of errors:

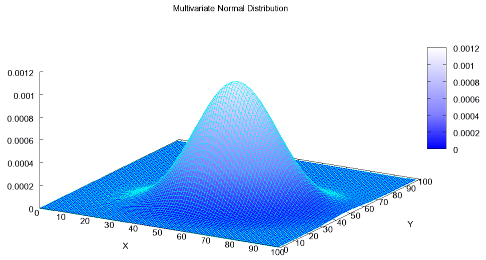


Figure: Multivariate normal distribution.

Based on multivariate normal distribution, we have (skipping derivations):

$$\ell(\theta, \hat{\Sigma}|Y) = -\frac{T}{2}\left(h \log(2\pi e) + \log|\hat{\Sigma}|\right) \tag{6}$$

Looks similar to:

$$\ell(\theta, \hat{\sigma_1}^2|Y) = -\frac{T}{2}\left(\log(2\pi e) + \log \hat{\sigma_1}^2\right) \tag{7}$$

Σ is covariance matrix that has the structure:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,h} \\ \sigma_{1,2} & \sigma_2^2 & \dots & \sigma_{2,h} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,h} & \sigma_{2,h} & \dots & \sigma_h^2 \end{pmatrix}, \tag{8}$$

Note that $MSE_h \propto \sigma_h^2$, which makes it a special case of Σ.

And MSTFE is just the sum of diagonals of Σ.

What does $|\hat{\Sigma}|$ mean?

Example with $h = 2$:

$$|\Sigma| = \begin{vmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{vmatrix} = \sigma_1^2 \sigma_2^2 - \sigma_{1,2}^2 \quad (9)$$

Minimising determinant of $|\hat{\Sigma}|$ will:

1. decrease variances,
2. increase covariances.

They can be rewritten as interaction between variances and parameters.

For example:

$$|\Sigma| = \sigma_1^2 \cdot \sigma_1^2 (1 + \alpha^2) - (\sigma_1^2 \alpha)^2 \quad (10)$$

In a more general case:

$$|\boldsymbol{\Sigma}| = \prod_{j=1}^{h} \sigma_j^2 |\mathbf{R}|, \qquad (11)$$

where $R = \begin{pmatrix} 1 & r_{1,2} & \dots & r_{1,h} \\ r_{1,2} & 1 & \dots & r_{2,h} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1,h} & r_{2,h} & \dots & 1 \end{pmatrix}$ is correlation matrix and $r_{i,j} = f(\theta)$.

This means that in general shrinkage effect is weakened.

Likelihood ensures that parameters are cool!

And model selection can easily be done using IC:

$$AIC = 2kh - 2\ell(\theta, \hat{\Sigma}|Y) \tag{12}$$

Examples

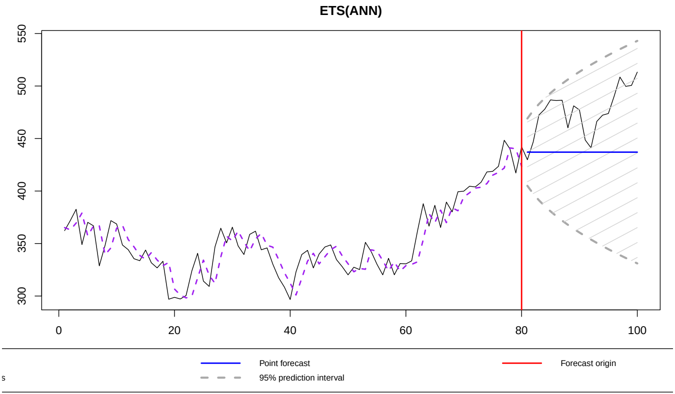


Figure: Model selection and estimation using MSE.

Examples

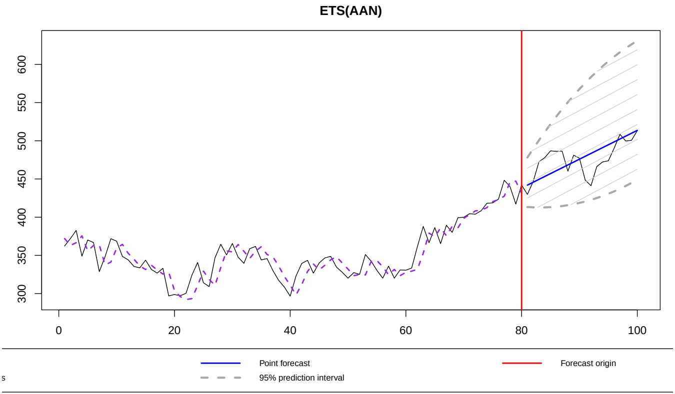


Figure: Model selection and estimation with TFL.

Conclusions

- Multiple steps ahead objective functions imply shrinkage of parameters;
- Parameters of univariate models shrink;
- This gives robustness to models and help decrease potential model structure uncertainty;
- Parameters may overshrink when estimated using MSE_h and MSTFE;

Conclusions

- Trace Forecast Likelihood (TFL) do not overshrink the parameters;
- TFL gives consistent, efficient and unbiased estimates of parameters;
- Using TFL increases long-term forecast accuracy and decreases uncertainty;
- Model selection with TFL is easy.

Finale

Thank you for your attention!

Ivan Svetunkov
i.svetunkov@lancaster.ac.uk

Lancaster University Management School | Lancaster Centre for Forecasting

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society*, *158*, 419–466.

Clements, M. P. & Hendry, D. F. (2008). *Multi-step estimation for forecasting*.

Hyndman, R. J., Koehler, A., Ord, K. & Snyder, R. (2008). *Forecasting with exponential smoothing*. Springer Series in Statistics. Springer Berlin Heidelberg. doi:10.1007/978-3-540-71918-2

Marcellino, M., Stock, J. H. & Watson, M. W. (2006, November). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1-2), 499–526. doi:10.1016/j.jeconom.2005.07.020

McElroy, T. & Wildi, M. (2013, July). Multi-step-ahead estimation of time series models. *International Journal of Forecasting*, *29*(3), 378–394. doi:10.1016/j.ijforecast.2012.08.003

Taylor, J. W. (2008). An evaluation of methods for very short-term load forecasting using minute-by-minute british data. *International Journal of Forecasting*, *24*(4), 645–658. doi:10.1016/j.ijforecast.2008.07.007

Tiao, G. & Xu, D. (1993). Robustness of maximum likelihood estimates for multi-step predictions: the exponential smoothing case. *Biometrika*, *80*(3), 623–641. Retrieved from http://biomet.oxfordjournals.org/content/80/3/623.short

Weiss, A. A. (1991, April). Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics*, *48*, 135–149. doi:10.1016/0304-4076(91)90035-C

Weiss, A. & Andersen, A. P. (1984, September). Estimating time series models using the relevant forecast evaluation criterion. *Journal of the Royal Statistical Society. Series A (General)*, *147*(3), 484. doi:10.2307/2981579

LCF

Xia, Y. & Tong, H. (2011). Feature matching in time series modeling. *Statistical Science*, *26*(1), 21–46. doi:10.1214/10-STS345