

# Discussion of Generalized Density Forecast Combinations

Anne Opschoor

VU University Amsterdam

[a.opschoor@vu.nl](mailto:a.opschoor@vu.nl)

8th ECB Workshop on Forecasting Techniques  
Frankfurt, June 13-14, 2014

- ★ This paper introduces a new methodology to combine  $N$  density forecasts, based on Sieve estimation (Chen and Shen, 1998)

- ★ This paper introduces a new methodology to combine  $N$  density forecasts, based on Sieve estimation (Chen and Shen, 1998)
- ★ The combination weights depends on the variable that is tried to be forecasted (here the S&P 500 asset return)

- ★ This paper introduces a new methodology to combine  $N$  density forecasts, based on Sieve estimation (Chen and Shen, 1998)
- ★ The combination weights depends on the variable that is tried to be forecasted (here the S&P 500 asset return)
- ★ Use of piece-wise linear functions: Density is divided in certain parts by thresholds, each part gets a different weight.

- ★ This paper introduces a new methodology to combine  $N$  density forecasts, based on Sieve estimation (Chen and Shen, 1998)
- ★ The combination weights depends on the variable that is tried to be forecasted (here the S&P 500 asset return)
- ★ Use of piece-wise linear functions: Density is divided in certain parts by thresholds, each part gets a different weight.
- ★ The log-score is used as a loss function, which is minimized.

- ★ Nice idea the weights differ per region of the density

- ★ Nice idea the weights differ per region of the density
- ★ Evidence both in simulation study as well in empirical application

- ★ Nice idea the weights differ per region of the density
- ★ Evidence both in simulation study as well in empirical application
- ★ In addition, the theoretical part establishes the appealing asymptotic features of the Sieve estimates



- ★ Non-parametric estimator to estimate an unknown high-dimensional function as *more data becomes available*

## Comments: Sieve estimation (I) and $T$

- ★ Non-parametric estimator to estimate an unknown high-dimensional function as *more data becomes available*
- ★ Simulation/empirical application shows it works for large  $T$ ; what about macro-economic examples when  $T$  is small? (quarterly inflation, output etc.)

# Comments: Sieve estimation (I) and $T$

- ★ Non-parametric estimator to estimate an unknown high-dimensional function as *more data becomes available*
- ★ Simulation/empirical application shows it works for large  $T$ ; what about macro-economic examples when  $T$  is small? (quarterly inflation, output etc.)
- ★ Moreover, other studies (e.g. Del Negro et al., 2013; Billio et al., 2014) show that optimal weights of density forecasts are time-varying → hence static weights imply the use of a moving window of reasonable length (no large  $T$ ). Curious to see if Sieve estimation still works properly.

- ★ The paper shows results of combining two (simulation) or four (application) density forecasts; what about  $N > 5$ ? Does estimation still work (also given the number of thresholds) ? Probably show this also in simulation design?

- ★ The paper shows results of combining two (simulation) or four (application) density forecasts; what about  $N > 5$ ? Does estimation still work (also given the number of thresholds)? Probably show this also in simulation design?
- ★ Uniqueness of parameter estimates not proven: Authors propose to modify (log-score) loss function as follows:

$$L_T = \sum_{t=1}^T l(p_t(y_t); y_t) + T^\gamma \sum_{i=1}^N \sum_{s=1}^{\infty} |\nu_s|^\delta \quad \delta > 0, 0 < \gamma < 1$$

Show by simulation setting? Do we have to care about it?

- ★ DGP 1: Rejection probabilities low, even if  $p = 2$  and  $r_1 = 0$  (their correct values). What happens if you estimate these values? Implication for test in application?

- ★ DGP 1: Rejection probabilities low, even if  $p = 2$  and  $r_1 = 0$  (their correct values). What happens if you estimate these values? Implication for test in application?
- ★ DGP 2: DGP = Standard Normal, combination of again two Normals with different means. Curious to see  $N > 2$  (as noted before). Why a Normal distribution (from practical perspectives) ?

# Comments: Application (I): log-score and $p$

- ★ Table 8/9: Significant drop in performance (log-score) if one adds one model. Desirable? (compare with linear method)

Component Densities				
(1:GARCH, 2:EGARCH, 3:SV, 4:TGARCH)				
	In-sample		Out-of-sample	
	2,3,4	1,2,3,4	2,3,4	1,2,3,4
G	2.762	2.572	0.735	-0.296
L	-1.183	-1.169	-0.558	-0.554
$\hat{p}$	7	1		



## Comments: Application (I): log-score and $p$

- ★ Table 8/9: Significant drop in performance (log-score) if one adds one model. Desirable? (compare with linear method)

Component Densities				
(1:GARCH, 2:EGARCH, 3:SV, 4:TGARCH)				
	In-sample		Out-of-sample	
	2,3,4	1,2,3,4	2,3,4	1,2,3,4
G	2.762	2.572	0.735	-0.296
L	-1.183	-1.169	-0.558	-0.554
$\hat{p}$	7	1		

- ★ Probably related with the estimation of  $p$ : a "non-linear trade-off between the complexity of the generalized combination and estimation error" (also shown in simulation setting).

# Comments: Application (I): log-score and $p$

- ★ Table 8/9: Significant drop in performance (log-score) if one adds one model. Desirable? (compare with linear method)

Component Densities				
(1:GARCH, 2:EGARCH, 3:SV, 4:TGARCH)				
	In-sample		Out-of-sample	
	2,3,4	1,2,3,4	2,3,4	1,2,3,4
G	2.762	2.572	0.735	-0.296
L	-1.183	-1.169	-0.558	-0.554
$\hat{p}$	7	1		

- ★ Probably related with the estimation of  $p$ : a "non-linear trade-off between the complexity of the generalized combination and estimation error" (also shown in simulation setting).
- ★ In simulation, often  $p = 3$  and  $p = 4$  is chosen (according to Figure 1), while the true one is  $p = 2$ . How to set  $p_{max}$  ?

- ★ I truly believe the generalized method improves upon the linear method, but I have difficulties with understanding where the gain comes from? (What are the estimated thresholds?)

- ★ I truly believe the generalized method improves upon the linear method, but I have difficulties with understanding where the gain comes from? (What are the estimated thresholds?)
- ★ E.g.: Combining GARCH with TGARCH, the (in-sample) log-scores of the Generalized vs. Linear method are 2.762 and -1.169 respectively. The 8 weights corresponding with the TGARCH are (0.40, 1,1,0.84,1,1,1,0.5). Hence the huge difference arises from the tails?

- ★ I truly believe the generalized method improves upon the linear method, but I have difficulties with understanding where the gain comes from? (What are the estimated thresholds?)
- ★ E.g.: Combining GARCH with TGARCH, the (in-sample) log-scores of the Generalized vs. Linear method are 2.762 and -1.169 respectively. The 8 weights corresponding with the TGARCH are (0.40, 1,1,0.84,1,1,1,0.5). Hence the huge difference arises from the tails?
- ★ In general: what about interpreting this weights?