Supplement materials of "Forecasting mortality rates with a coherent ensemble averaging approach"

A. Proof of Theorem 1

For the jth model ($j = 1, ..., J$), Gao and Shi (2021) show that a sufficient and necessary condition to achieve age coherence is that $|\hat{\mu}_{x,j} - \hat{\mu}_{x+k,j}| = O_p(1/h)$[1], where $x = 1, ..., N$, $k = 1, ..., N-1$, and $\hat{\mu}_{x,j}$ is the estimated long-run mean of mortality improvements of age x ($\triangle y_{x,t}$), using the model j. Hence, it can be inferred that

$$\hat{y}_{x,j,T+h} = O_p(1) + h\hat{\mu}_{x,j},$$

and age coherence is achieved as $|\hat{y}_{x,j,T+h} - \hat{y}_{x+k,j,T+h}| = O_p(1)$.

Using the MA approach, it is then easy to see

$$\hat{y}_{x,M,T+h} = O_p(1) + h\sum_{j=1}^{J_c} \hat{w}_{x,j}\hat{\mu}_{x,j} + h\sum_{j=1}^{J_n} \hat{w}_{x,j}\hat{\mu}_{x,j}. \tag{1}$$

Recall that $\hat{w}_{x,j}$ of age-incoherent models are estimates of (13) and subject to the coherent penalty λ_1. Since λ_1 is assumed to go large with T^2, $\hat{\boldsymbol{w}}'\boldsymbol{C}_1\hat{\boldsymbol{w}}$ will need to be $O_p(1/T^2)$, or the loss function will be "overwhelmed" by λ_1. Essentially, this suggests that $\hat{w}_{x,j} = O_p(1/T)$ for an age-incoherent model. Also, since $\hat{\mu}_{x,j}$ is not increasing with h, it is equivalent to the fact that $\hat{\mu}_{x,j}$ of all models have a lower or the same order of $O_p(1)$. Consequently, further with the assumption that h and T go to infinity at the same rate, we have that the highest order of the second summation in (1) is

$$h\sum_{j=1}^{J_n} \hat{w}_{x,j}\hat{\mu}_{x,j} = hO_p(1/h)O_p(1) = O_p(1).$$

Thus, (1) can be rewritten as

$$\hat{y}_{x,M,T+h} = O_p(1) + h\sum_{j=1}^{J_c} \hat{w}_{x,j}\hat{\mu}_{x,j}$$

[1] Note that if $\hat{\mu}_{x,j}$ is irrelevant to h (as in the STAR model), this condition is equivalent to $\hat{\mu}_{x,j} = \hat{\mu}_{x+k,j}$.

taking the difference between forecast logged mortality rates of two ages, we have that

$$\begin{aligned}
&|\hat{y}_{x,M,T+h} - \hat{y}_{x+k,M,T+h}| \\
=&O_p(1) + h|\sum_{j=1}^{J_c} \hat{w}_{x,j}\hat{\mu}_{x,j} - \sum_{j=1}^{J_c} \hat{w}_{x+k,j}\hat{\mu}_{x+k,j}| \\
=&O_p(1) + h|\sum_{j=1}^{J_c} \hat{w}_{x,j}(\hat{\mu}_{x,j} - \hat{\mu}_{x+k,j}) + \sum_{j=1}^{J_c-1} (\hat{w}_{x,j} - \hat{w}_{x+k,j})(\hat{\mu}_{x+k,j} - \hat{\mu}_{x+k,J_c})|
\end{aligned} \tag{2}$$

Further, since $\hat{w}_{x,j}$ is bounded in $[0, 1]$, it is easy to see that

$$h[\sum_{j=1}^{J_c} \hat{w}_{x,j}(\hat{\mu}_{x,j} - \hat{\mu}_{x+k,j})] = hO_p(1)O_p(1/h) = O_p(1).$$

This verifies that the first summation in (2) is $O_p(1)$.

For the second summation in (2), since all $\hat{\mu}_{x+k,j}$'s are asymptotically consistent,

$$\hat{\mu}_{x+k,j} - \hat{\mu}_{x+k,J_c} = \hat{\mu}_{x+k,j} - \mu_{x+k} - (\hat{\mu}_{x+k,J_c} - \mu_{x+k}) = o_p(1),$$

where μ_{x+k} is the true value of the mortality improvement at age $x + k$, it is then easy to see that

$$(\hat{w}_{x,j} - \hat{w}_{x+k,j})(\hat{\mu}_{x+k,j} - \hat{\mu}_{x+k,J}) = o_p(1)$$

for an age-coherent model.

Combine all the above results, we have that

$$|\hat{y}_{x,M,T+h} - \hat{y}_{x+k,M,T+h}| = O_p(1),$$

which completes the proof.

B. Additional tables

Table 1: Summary of the tuning parameter selection

Model	Parameter	Purpose	Selection procedure
STAR	λ_{α}	A larger (smaller) value increases (reduces) the smoothness between α_x and α_{x+1}	The selection takes place at the estimation stage. A grid search is performed to select the three smoothness penalties simultaneously. At each step, model coefficients are fitted after all tuning parameters are specified. Forecasts are then produced as in (6). An RMSFE is then produced following (12). The final selection of tuning parameters is that minimizes the RMSFE over all grids.
	λ_{β_1}	A larger (smaller) value increases (reduces) the smoothness between $\beta_{x,x-1}$ and $\beta_{x+1,x}$	
	λ_{β_2}	A larger (smaller) value increases (reduces) the smoothness between $\beta_{x,x-2}$ and $\beta_{x+1,x-1}$	
LC-H	d_1^l	A larger (smaller) value indicates slower (faster) decay at the first age	The selection takes place at the forecasting stage. A grid search is performed to select both parameters simultaneously. Once a pair is specified, mortality rates are forecast as in (9). RMSFE is then produced following (12). The final selection of tuning parameters is that minimizes the RMSFE over all grids.
	b^w	A larger (smaller) value indicates decline in r_x^l starts from an older (younger) age	
LC-G	r_1^l b^w	Same as in LC-H	Same as in LC-H
SVAR-H	λ_x	A larger (smaller) value indicates fewer (more) none-zero AR coefficients	The selection takes place at the estimation stage. A grid search is performed. At each step, model coefficients are fitted after all tuning parameters are specified. Forecasts are then produced as in (10). An RMSFE is then produced following (12). The final selection of tuning parameters is that minimizes the RMSFE over all grids.
	d_1^s b^w	Same as in LC-H	Same as in LC-H
SVAR-G	λ_x r_1^s b^w	Same as in SVAR-H	Same as in SVAR-H

Table 2: Summary of robustness checks

	Mean	Std. Dev.	Median	Q_1	Q_3
Panel A: Sixteen forecasting steps					
STAR	0.2172	0.0626	0.2283	0.1814	0.2649
LC-H	0.2277	0.0656	0.2216	0.1748	**0.2417**
LC-G	0.2319	0.0634	0.2284	0.1846	0.2446
SVAR-H	0.2311	0.0686	0.2227	0.1890	0.2873
SVAR-G	0.2314	0.0685	0.2226	0.1902	0.2876
LC	0.2519	0.0611	0.2309	0.2177	0.2761
SVAR	0.2297	0.0737	0.2160	0.1875	0.2820
APC	0.3256	0.0630	0.3205	0.2741	0.3735
HU	0.2427	**0.0607**	0.2219	0.2094	0.2770
RH	0.2675	0.0893	0.2760	0.2194	0.3192
MA	**0.2039**	0.0611	**0.2011**	**0.1571**	0.2470
Panel B: Starting year of 1960					
STAR	0.1844	0.0645	**0.1794**	0.1413	0.2451
LC-H	0.2041	0.0679	0.2041	0.1385	0.2594
LC-G	0.2055	0.0676	0.2063	0.1430	0.2612
SVAR-H	0.2054	0.0828	0.2070	0.1377	0.2805
SVAR-G	0.2058	0.0824	0.2068	0.1390	0.2807
LC	0.2305	0.0672	0.2289	0.1690	0.2909
SVAR	0.2064	0.0842	0.1994	0.1351	0.2838
APC	0.2638	**0.0523**	0.2820	0.2242	0.3010
HU	0.2024	0.0680	0.2010	0.1512	0.2537
RH	0.2170	0.0649	0.2205	0.1780	0.2617
MA	**0.1837**	0.0712	0.1938	**0.1242**	**0.2413**
Panel C: Ages 0–89					
STAR	0.1862	0.0688	0.1821	0.1380	**0.2457**
LC-H	0.2059	0.0746	0.1859	0.1527	0.2500
LC-G	0.2106	0.0719	0.1942	0.1562	0.2535
SVAR-H	0.2067	0.0862	0.1876	0.1492	0.2835
SVAR-G	0.2069	0.0863	0.1877	0.1495	0.2837
LC	0.2469	**0.0654**	0.2452	0.1894	0.2989
SVAR	0.2102	0.0929	0.1910	0.1351	0.2913
APC	0.3274	0.0657	0.3114	0.2753	0.3805
HU	0.2085	0.0768	0.1775	0.1573	0.2653
RH	0.2570	0.0838	0.2717	0.1957	0.3128
MA	**0.1834**	0.0757	**0.1695**	**0.1245**	0.2471

Note: bold numbers are the smallest quantity for each statistic across the eleven models.

References

Gao, G., Shi, Y., 2021. Age-coherent extensions of the Lee–Carter model. Scandinavian Actuarial Journal , 1–19.