Assessing mortality inequality in the U.S.: What can be said about the future?

Han Li* and Rob J Hyndman†

February 14, 2021

Abstract

This paper investigates mortality inequality across U.S. states by modelling and forecasting mortality rates via a forecast reconciliation approach. Understanding the heterogeneity in state-level mortality experience is of fundamental importance, as it can assist decision making for policymakers, health authorities, as well as local communities who are seeking to reduce inequalities and disparities in life expectancy. A key challenge of multi-population mortality modeling is high dimensionality, and the resulting complex dependence structures across sub-populations. Moreover, when projecting future mortality rates, it is important to ensure that the state-level forecasts are coherent with the national-level forecasts. We address these issues by first obtaining independent state-level forecasts based on classical stochastic mortality models, and then incorporating the dependence structure in the forecast reconciliation process. Both traditional bottom-up reconciliation and the cutting-edge trace minimization reconciliation methods are considered. Based on the U.S. total mortality data for the period 1969–2017, we project the 10-year-ahead mortality rates at both national-level and state-level up to 2027. We find that the geographical inequality in the longevity levels is likely to continue in the future, and the mortality improvement rates will tend to slow down in the coming decades.

Keywords: Mortality modeling; Heterogeneity; Forecast reconciliation.

*Department of Actuarial Studies and Business Analytics, Macquarie University, Australia. Email address: han.li@mq.edu.au.

†Department of Econometrics and Business Statistics, Monash University, Australia. Email address: rob.hyndman@monash.edu.

1 Introduction

With rapid economic growth and medical advances, overall life expectancy in the U.S. has continuously improved in recent decades from 69.8 years in 1960 to 79.5 years in 2017. However, despite the significant improvement in national-level life expectancy, several recent studies have found large, long-lasting, and increasing geographical inequalities in mortality experience within the U.S. (see *e.g.* Kulkarni *et al.*, 2011; Wang *et al.*, 2013; Dwyer-Lindgren *et al.*, 2016). In 2014, the life expectancy gap between the most affluent and impoverished counties in the U.S. was more than 20 years, and this gap is expected to become even wider (Dwyer-Lindgren *et al.*, 2017).[1] Such inequality in mortality experience demands urgent action and presents new challenges for both the public sector (*e.g.* health policy design and planning) and the private sector (*e.g.* insurance pricing and longevity transfer).

The aim to "*Achieve health equity, eliminate disparities, and improve the health of all groups*" has been listed as an important mission in the "Healthy People 2020" framework proposed by the U.S. Department of Health and Human Services (US Department of Health and Human Services., 2010). Assessing the mortality inequalities at the state-level over time enables us to evaluate the progress toward this goal of more equitable health and longevity outcomes. Moreover, understanding and forecasting state-level mortality trend is essential for appropriate resource allocation, and efficient health and welfare policy design. Examining the converging or diverging trends in mortality rates across states is increasingly important, for policymakers, health authorities, researchers and local communities who are seeking to reduce the disparities and improve longevity.

While there is a vast literature on modeling and forecasting national-level mortality rates (see *e.g.* Rice and Feldman, 1983; Lee and Carter, 1992; Cairns *et al.*, 2006), relatively less attention has been paid to the disparities in mortality experience among individual U.S. states, and more importantly, whether these state-level mortality trends will converge or diverge in the future. To answer this question, we need to model and forecast mortality on not only the national level, but also the state level, in an integrated manner. However, joint modeling of state- and national-level mortality rates presents the challenge of high dimensionality. A multi-population model for state-age-gender-specific mortality rates can be extremely complex and overparameterized. Moreover, state-level mortality rates can be presented in a hierarchical setting which aggregates to the national level, but the coherence of mortality forecasts cannot be ensured if state-level and national-level mortality experience are projected separately. Fortunately, these issues can be addressed via a forecast reconciliation approach. Forecast reconciliation is an emerging state-of-the-art methodology which ensures the coherency of forecasts as well as improves the overall forecast quality, while joint modeling of each data series is not required for the reconciliation process (Athanasopoulos *et al.*, 2009; Wickramasuriya *et al.*, 2019). It has been used in the context of mortality modeling in recent years (see *e.g.* Shang and Hyndman, 2017; Shang and Haberman, 2017; Li *et al.*, 2019; Li and Tang, 2019), and has shown superior performance over alternative methods.

In this paper, we propose a forecast reconciliation approach to jointly projecting national- and state-level age-gender-specific mortality rates across 50 U.S. states and the District of Columbia. Our empirical study is based on the U.S. mortality data for ages 0–85+, during the years 1969–

[1] Also refer to the Guardian article which is available online at: https://www.theguardian.com/inequality/2017/may/08/life-expectancy-gap-rich-poor-us-regions-more-than-20-years

2017. We adopt the trace minimization (MinT) reconciliation method by Wickramasuriya *et al.* (2019) together with the sampling approach by Jeon *et al.* (2019) to reconcile both point and probabilistic mortality forecasts. A major advantage of MinT is that the dependence structure of data series in the hierarchy is taken into account during the reconciliation process, which is particularly desirable in our study as the mortality experience of different states are likely to be correlated in a spatial fashion. Therefore, we first obtain forecasts for each region as well as the total population by fitting a univariate Lee-Carter model (Lee and Carter, 1992), and then reconcile these forecasts in the second stage. Based on the proposed approach, we project 10-year-ahead U.S. mortality rates up to 2027, for each region-age-gender combination.

The contribution of our research to the literature is three-fold. First, we achieve higher overall forecast accuracy at all levels, for both males and females. To the best of our knowledge, our study is the first to apply the cutting-edge MinT reconciliation method on regional mortality forecasts. Earlier studies on a similar topic, such as Shang and Hyndman (2017) and Shang and Haberman (2017), adopt an ordinary least squares approach (Hyndman *et al.*, 2011) which did not consistently illustrate strong performances over other methods. Second, we improve upon the approach of Shang and Hyndman (2017) by adjusting the forecast of summing matrix to ensure that the underlying aggregation constraints are fully satisfied. Last and most importantly, we are among the first to provide reconciled probabilistic mortality forecasts as well as point mortality forecasts. This enables us to anticipate the level of heterogeneity in mortality experience across U.S. states in the future, which is crucial for decision-making and planning of various stakeholders such as government bodies, pension funds and insurance companies.

The rest of the paper is organized as follows. In Section 2, we describe the mortality data used and visualize the state-level mortality rates for different ages. Section 3 introduces the MinT forecast reconciliation approach, which is then applied to the U.S. mortality data described in Section 2. We present and discuss the results from our empirical study in Section 4. Conclusions are drawn in Section 5.

2 U.S. state-level mortality data

The empirical study of this research concerns age-gender-specific mortality rates in the U.S. at both state and national levels. We collect death and exposure information from three main data sources as follows:

- *The National Center for Health Statistics (NCHS)*:
  The NCHS records information of all individual deaths in the U.S. since 1959, including age, gender, marital status, cause of death, and geographic identifier. We collect state-level death data from the NCHS "Multiple Cause of Death Data, 1959–2017" vital statistics database for the period 1969–2004. Single-year ages 0–85+ are considered in this study. From 2005, the geographic identifier is omitted due to a restriction imposed on release of sub-national mortality data.[2]

- *The Centers for Disease Control and Prevention (CDC) WONDER online database*:
  The CDC WONDER online database provides a rich query system for public health data

[2]See more details at: http://www.nber.org/data/vital-statistics-mortality-data-multiple-cause-of-death.html

and vital statistics. We will collect summarized state-level death data for 2005–2017 via the online query "Underlying Cause of Death, 1999–2017 Request".

- *Survey of Epidemiology and End Results (SEER) U.S. State and County Population Data*: The Survey of Epidemiology and End Results (SEER) provide an excellent resource of data since 1969 on U.S. population at the level of the state/county and by age, race, sex, and more recently Hispanic origin (since 1990). We collect state-level annual population data for 86 single-year ages (0, 1, 2, ..., 84, 85+).

Mortality rates are estimated based on death data from the NCHS and CDC WONDER, and the SEER population data. We use the mid-year population data from SEER as population exposure. In the following, We define some notation which will be used throughout this paper. First, we define the state-age-gender specific mortality rate $m^j_{i,t}(x)$ as:

$$m^j_{i,t}(x) = \frac{D^j_{i,t}(x)}{E^j_{i,t}(x)}, \tag{1}$$

where $D^j_{i,t}(x)$ is the number of deaths for gender j, at age x and time t from region i. $E^j_{i,t}(x)$ is the corresponding population exposure.

Similarly, we define the age-gender-specific mortality rate as:

$$m^j_t(x) = \frac{D^j_t(x)}{E^j_t(x)}, \tag{2}$$

where $D^j_t(x)$ is the total number of deaths for gender j, at age x and time t. $E^j_t(x)$ is the corresponding population exposure.

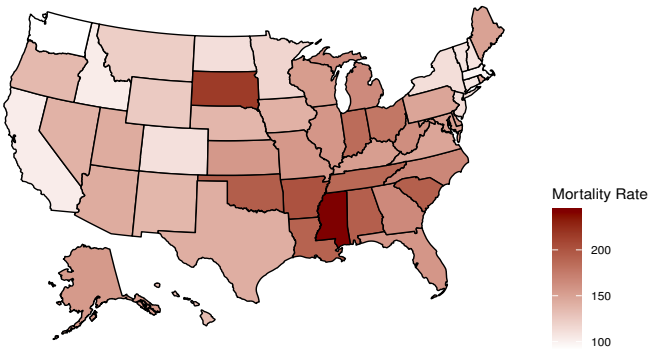
Finally, we have the age-specific mortality rate $m_t(x)$ which is calculated by:

$$m_t(x) = \frac{D_t(x)}{E_t(x)}, \tag{3}$$

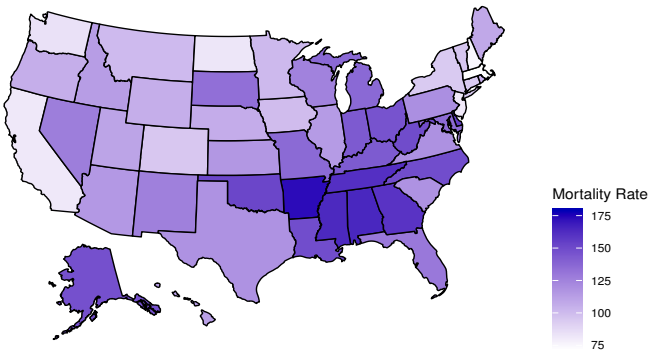
where $D_t(x)$ represents the total number of deaths at age x and $E_t(x)$ represents the corresponding population exposure.

Based on the collected data, we calculate the crude mortality rates of age group 0–85+ for 51 U.S. states[3] during 1969–2017. Figure 1 illustrates the geographical variations in mortality experience across the U.S. for four representative age brackets, namely, ages 0–4, ages 25–34, ages 55–64 and ages 85+ in 2017. There are very clear geographic patterns exhibited in these maps. On the left panel, we plot state-level male mortality rates. It can be observed that compared to the other three age groups, the "oldest-old" (ages 85+) do not show dramatic geographical variation in mortality rates. However, it should be noted that old-age mortality rates are much higher than the other three groups and thus, variations in old-age mortality experience would have a much bigger impact on the inequalities in life expectancy (Dwyer-Lindgren *et al.*, 2017). We plot state-level female mortality rates on the right panel, where a different color scale is used. Although female mortality rates are consistently lower than male mortality rates, it can be argued that the state-level male and female mortality experience illustrate similar geographical variation.

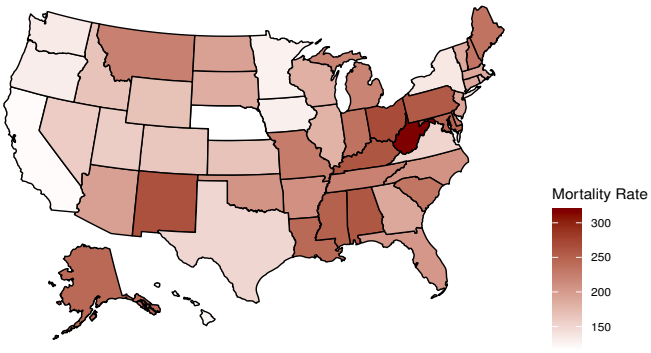
[3]For ease of exposition, in this paper we will refer to the District of Columbia as a "state".



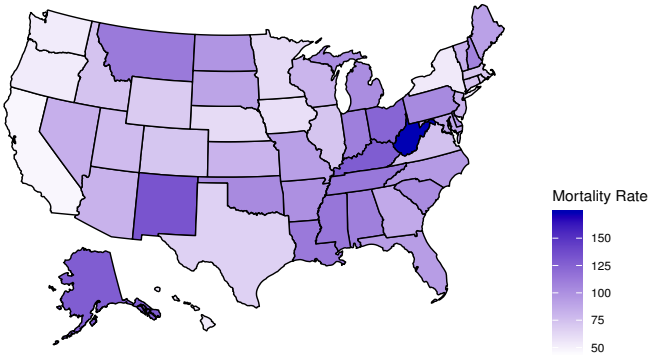
(a) Male ages 0–4



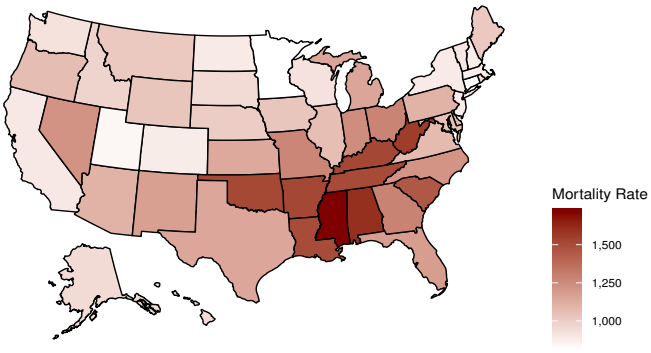
(b) Female ages 0–4



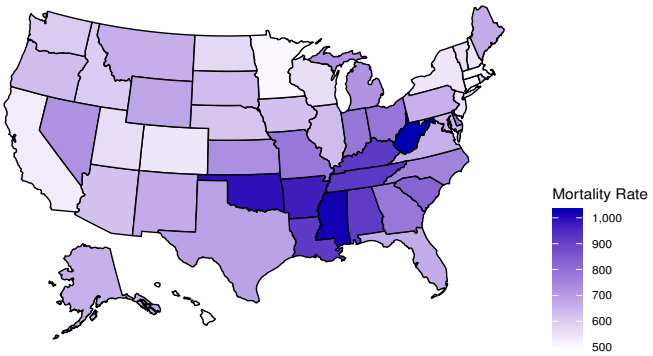
(c) Male ages 25–34



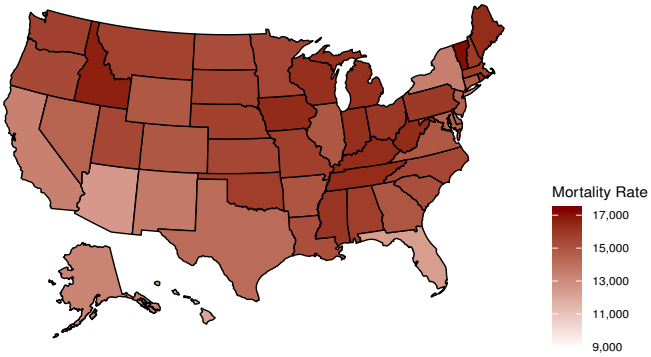
(d) Female ages 25–34



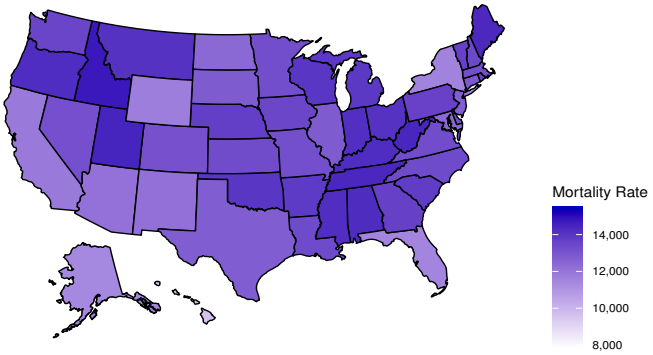
(e) Male ages 55–64



(f) Female ages 55–64



(g) Male ages 85+

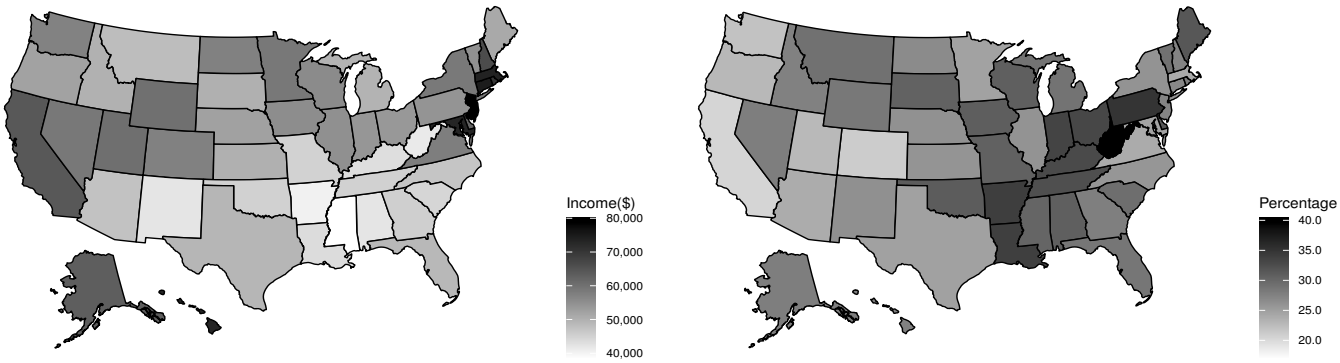


(h) Female ages 85+

Figure 1: U.S. state-level mortality rates (per 100,000) in 2017

Overall, we can conclude that Southern states have considerable survival disadvantages for almost all four groups, and for both genders. Another interesting finding is that there is a tendency for better mortality experience in those states with favorable weather or a large population of elderly. For example, Hawaii, Florida, Arizona, and Alaska stood out as states with low mortality rates for ages 85 and above. On the other hand, the worst old-age mortality experience are mostly in rural states, states with lower socio-economic profiles, and states with extremely cold winters. For example, high rates of oldest-old mortality were found in the Southeast and Midwest regions in the U.S. including the states of West Virginia, Tennessee, Kentucky, and Indiana. Our findings are consistent with those by Dwyer-Lindgren *et al.* (2016), Holman (2017), and Andreev *et al.* (2017).

A large body of research has discovered the association between life expectancy and socioeconomic factors (see *e.g.* Davids *et al.*, 2014; Chetty *et al.*, 2016; Dwyer-Lindgren *et al.*, 2017). Dwyer-Lindgren *et al.* (2017) found that socio-economic and ethnicity factors account for 60% of the county-level variation in life expectancy in the U.S. To see how mortality inequality is associated with socio-economics factors, we collect and present the following state-level socioeconomic data: 1) the median household income in 2017 from the Small Area Income and Poverty Estimates (SAIPE) Program, and 2) the proportion of population with high school diploma only (*i.e.* being the highest degree) during 2014–2018 from the United States Department of Agriculture, Economic Research Service (USDA ERS). These socio-economic data are plotted in Figure 2.



(a) Median Household Income

(b) Proportion of population with high school diploma only

Figure 2: U.S. state-level income and education. Panel (a): Median household income in 2017. Data source: Small Area Income and Poverty Estimates (SAIPE) Program. Panel (b): Proportion of population with high school diploma only during 2014–2018. Data source: United States Department of Agriculture, Economic Research Service (USDA ERS).

We can see that similar to mortality rates, there are very clear geographic patterns exhibited in income and education levels across the U.S. On the left panel, high median household income has mainly been found in California and North Eastern states. In particular, the "top" 3 states with the highest median household income are New Jersey, Connecticut, and Massachusetts. On the other hand, Southern states, stretching from Oklahoma to West Virginia, generally have low levels of household income. The "bottom" 3 states with the lowest median household income are West Virginia, Arkansas, and Mississippi. On the right panel, it can been seen that

Southern states have relatively larger proportions of people whose highest degree is high school diploma. The 3 states with the lowest percentage (and thus the highest level of education) are California, Colorado, and Washington. While the 3 states with the highest percentage (and thus the lowest level of education) are West Virginia, Pennsylvania, and Arkansas. Clearly, poorer states are more likely to have low levels of education. It is not surprising that Figure 2 shows a negative association between median household income and the proportion of population with high school diploma only.

Comparing Figure 2 with the mortality maps in Figure 1, we can see that high income levels are generally associated with low levels of mortality rates, and better educated states generally have more desirable mortality experience. These conclusions hold for all age groups included in our analysis, and are largely consistent with findings from the aforementioned existing studies. We have also computed the Kendall's tau coefficient which is a rank correlation coefficient (Kendall, 1938), for median household income (denoted as "Income") and mortality rates, as well as the proportion of population with high school diploma only (denoted as "Education") and mortality rates. These results are presented in Table 1. It is shown that for all age groups, there is a fairly strong negative association between income/education levels and mortality rates. These Kendall's tau coefficients confirm our observations based on the choropleth maps. Since mortality rates are largely influenced by socio-economic factors, policy actions targeting these factors can help reverse the trend of increasing disparities in life expectancy.

Table 1: Kendall's tau

Age group	**Income**		**Education**	
	Male	**Female**	**Male**	**Female**
0–4	-0.410	-0.426	0.359	0.345
25–34	-0.352	-0.427	0.466	0.491
55–64	-0.583	-0.540	0.367	0.400
85+	-0.231	-0.357	0.345	0.339

Understanding the level of heterogeneity in state-level mortality experience is of fundamental importance, as it can provide us with a more complete picture of mortality experience across the country, and help us to improve the mortality forecasts on the national level. Naturally, gender-specific state-level mortality rates can be presented in a hierarchical setting which aggregates to the national level, as illustrated in Figure 3.

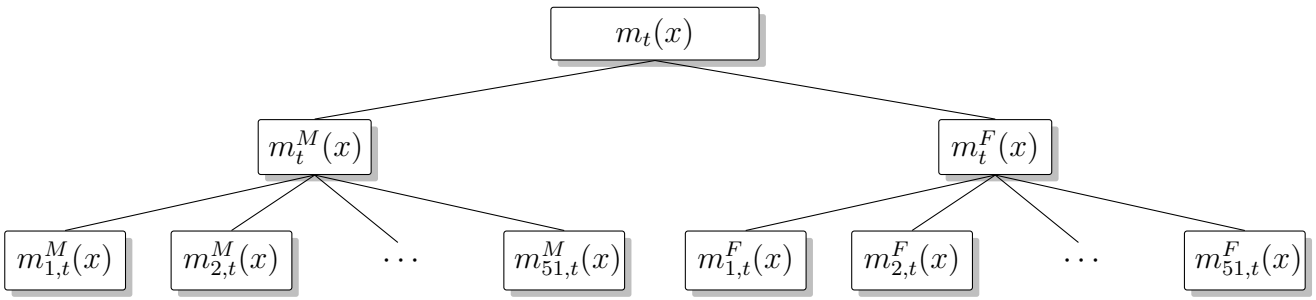
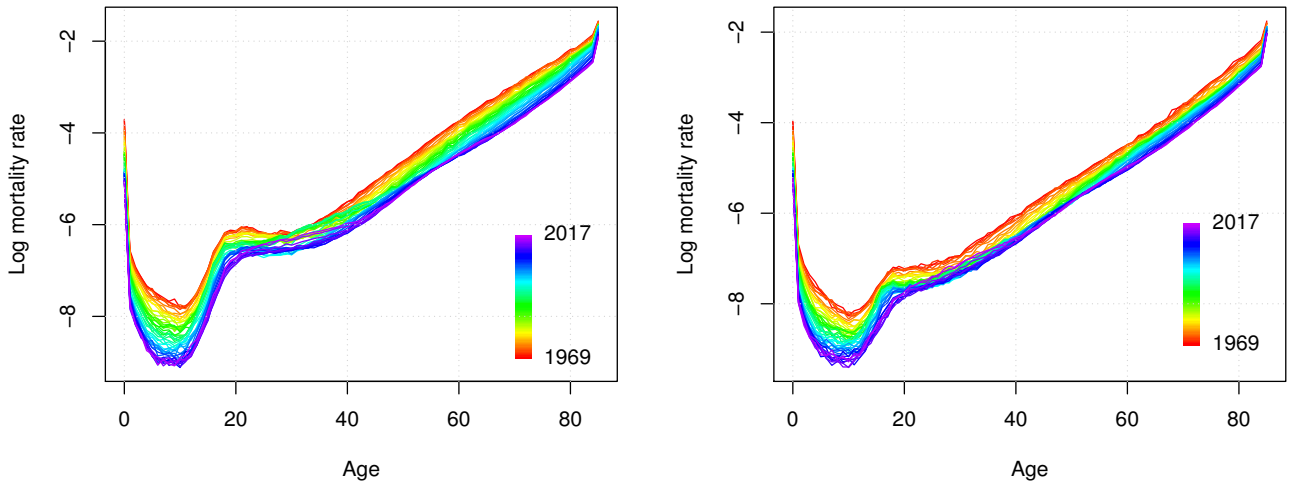


Figure 3: 3-level hierarchical tree for mortality rates

As shown in Figure 3, there are three levels in the hierarchy. For a given age x, at the top level, we have the total mortality rate $m_t(x)$. This can be divided into male mortality rate $m_t^M(x)$

and female mortality rate $m_t^F(x)$ at the middle level. At the bottom level, we further categorize gender-specific mortality rates by state, denoted by $m_{i,t}^M(x)$ and $m_{i,t}^F(x)$ for males and females, respectively. Therefore, we have $51 \times 2 = 102$ time series at the bottom level, 2 time series at the middle level and 1 time series at the top level. These sum up to 105 series in total.

To present the data at the middle level of the hierarchy, we plot the logarithm of gender-specific mortality rates for the U.S. during 1969-–2017 in Figure 4. Finally, for the top level of the hierarchy, we plot the total mortality rates for the U.S. in Figure 5.



(a) Male (b) Female

Figure 4: Rainbow plots for U.S. gender-specific mortality rates

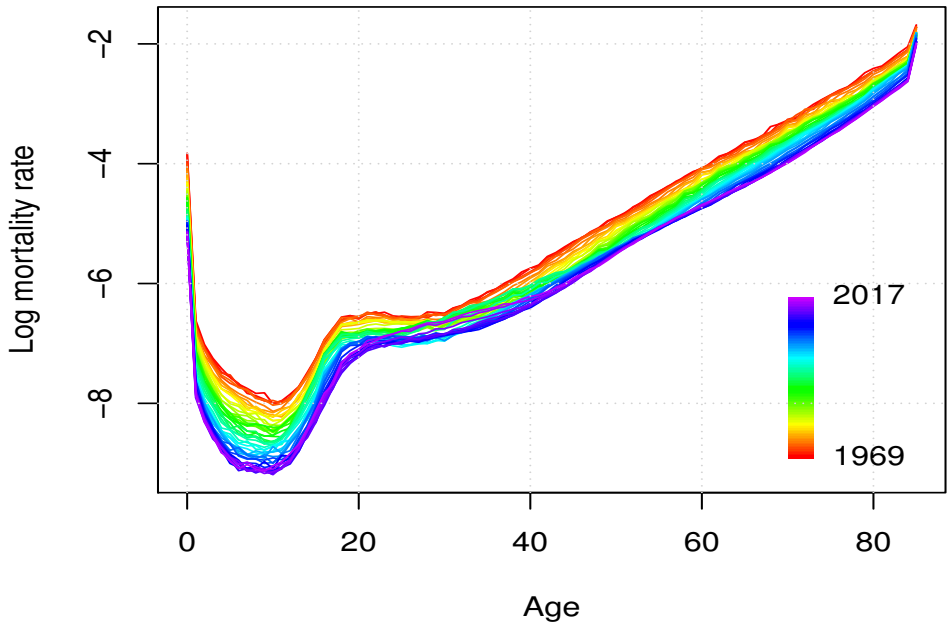


Figure 5: Rainbow plots for U.S. total mortality rates

3 A hierarchical reconciliation approach

Forecast reconciliation refers to the process of adjusting hierarchical time series forecasts such that the underlying aggregation constraints are met. There has been a fast growing literature on forecast reconciliation during the last few decades (see *e.g.* Dangerfield and Morris, 1992; Kahn, 1998; Zellner and Tobias, 2000; Athanasopoulos *et al.*, 2009; Hyndman *et al.*, 2011; Wickramasuriya *et al.*, 2019). There are three main advantages of forecast reconciliation. First, the reconciliation process ensures forecast coherency which helps people to make aligned decisions. In reality, it is almost impossible for independently projected forecasts at different levels to add up in a manner that is consistent with the underlying hierarchical structure. Therefore, reconciliation becomes a useful tool to reduce the discrepancy resulting from conflicting forecasts. Second, forecast reconciliation can circumvent the need for heavily parameterised and possibly misspecified joint models. This is a desirable property of the method as it is less subject to problems associated with high dimensionality. Third, by incorporating information at all levels into the forecasting process, certain reconciliation methods can improve the overall forecast accuracy of hierarchical time series (see, *e.g.* Capistrán *et al.*, 2010; Borges *et al.*, 2013; Syntetos *et al.*, 2016).

Traditional forecast reconciliation methods include the "bottom-up" and the "top-down" methods[4]. The common limitation of the two methods is that neither method utilizes all available information in the reconciliation process. In the case of the "bottom-up" method, only bottom-level time series is considered in the forecasting process. Therefore, another disadvantage of the method is that bottom-level data are the most disaggregated and thus are the noisiest time series in the hierarchy. Any forecast errors from the bottom level will be amplified into higher levels. Moreover, by simply adding up bottom-level forecasts, the "bottom-up" method does not necessarily take into account the dependence structure across disaggregated series.

To address these issues, an optimal combination method (commonly referred to as the "OLS" method) was introduced to reconcile hierarchical time series (Hyndman *et al.*, 2011). This method combines forecasts at all levels to achieve higher overall forecast accuracy. It has been applied to a wide range of disciplines and has shown strong performance over traditional reconciliation methods as well as unreconciled forecasts. It is worth noting that the application of the OLS forecast reconciliation approach in mortality modeling is not a new phenomenon, several earlier attempts have been made in reconciling Japanese infant mortality rates (Shang and Hyndman, 2017; Shang and Haberman, 2017).

Wickramasuriya *et al.* (2019) improved on the OLS reconciliation method and proposed a trace minimization (MinT) reconciliation approach. The variance-covariance matrix used in the OLS approach is generally unknown or unidentifiable, and the MinT approach provides a practical solution to this problem. Wickramasuriya *et al.* (2019) formally justified the use of variances and covariances of the in-sample forecast errors in their approach. It is worth noting that the MinT approach has also been adopted in actuarial research recently. Li *et al.* (2019) reconciled cause-of-death mortality forecasts via the MinT approach and achieved better forecast accuracy for both cause-specific mortality rates and total mortality rates. Li and Tang (2019) applied the MinT approach to analyze the Swiss Re Kortis bond index, and illustrate the strong performance of the approach over the bottom up method and the OLS method.

[4]For more information on the two methods, see Hyndman and Athanasopoulos (2018), Chapter 10.

In this paper, we adopt the MinT approach to reconcile state- and national-level mortality forecasts described and shown in Section 2.

3.1 Notation

As shown in Figure 3, we can present mortality rates in a hierarchical setting. However, it should be noted that, unlike traditional applications of forecast reconciliation, the aggregation rules of mortality rates are not as straightforward. For the hierarchical structure in Figure 3, at any given t and x, we have the following aggregation constraints:

$$\sum_{i=1}^{51} \frac{E_{i,t}^{M}(x)}{E_t^{M}(x)} \times m_{i,t}^{M}(x) = m_t^{M}(x). \tag{4}$$

$$\sum_{i=1}^{51} \frac{E_{i,t}^{F}(x)}{E_t^{F}(x)} \times m_{i,t}^{F}(x) = m_t^{F}(x). \tag{5}$$

$$m_t^{M}(x) \times \frac{E_t^{M}(x)}{E_t(x)} + m_t^{F}(x) \times \frac{E_t^{F}(x)}{E_t(x)} = m_t(x). \tag{6}$$

To address these aggregation constraints, we follow a similar approach introduced in an earlier work by Shang and Hyndman (2017). Before introducing the MinT reconciliation method, we first express the aforementioned aggregation constraints in a matrix form and formally define the following notation:

- Define $\boldsymbol{y}_t(x) = (m_t(x),\, m_t^{M}(x),\, m_t^{F}(x),\, m_{1,t}^{M}(x),\, \ldots,\, m_{51,t}^{M}(x),\, m_{1,t}^{F}(x),\, \ldots,\, m_{51,t}^{F}(x))'$ as a vector that contains observations at all levels in the hierarchy;
- Define $\boldsymbol{b}_t(x) = (m_{1,t}^{M}(x),\, \ldots,\, m_{51,t}^{M}(x),\, m_{1,t}^{F}(x),\, \ldots,\, m_{51,t}^{F}(x))'$ as a vector that contains observations at the bottom level only.

The two vectors can then be linked by the equation

$$\boldsymbol{y}_t(x) = \boldsymbol{S}_t(x)\, \boldsymbol{b}_t(x), \tag{7}$$

where $\boldsymbol{S}_t(x)$ is a "summing matrix" of dimension 105×102, which aggregates state-age-gender-specific mortality rates to obtain mortality rates at higher levels. It is given by

$$\boldsymbol{S}_t(x) = \begin{pmatrix} \frac{E_{1,t}^{M}(x)}{E_t(x)} & \frac{E_{2,t}^{M}(x)}{E_t(x)} & \cdots & \frac{E_{51,t}^{M}(x)}{E_t(x)} & \frac{E_{1,t}^{F}(x)}{E_t(x)} & \frac{E_{2,t}^{F}(x)}{E_t(x)} & \cdots & \frac{E_{51,t}^{F}(x)}{E_t(x)} \\ \frac{E_{1,t}^{M}(x)}{E_t^{M}(x)} & \frac{E_{2,t}^{M}(x)}{E_t^{M}(x)} & \cdots & \frac{E_{51,t}^{M}(x)}{E_t^{M}(x)} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \frac{E_{1,t}^{F}(x)}{E_t^{F}(x)} & \frac{E_{2,t}^{F}(x)}{E_t^{F}(x)} & \cdots & \frac{E_{51,t}^{F}(x)}{E_t^{F}(x)} \\ & & & & \boldsymbol{I}_{102} & & & \end{pmatrix},$$

where $\boldsymbol{I}_{102}$ denotes an identity matrix of dimension 102×102. Therefore, the aggregation constraints (4)–(6) are reflected in the first three rows of the $\boldsymbol{S}_t(x)$ matrix.

Although Equation (7) holds for all observed values, it is unlikely to hold for independently obtained forecasts in the hierarchy. Let $\hat{\boldsymbol{y}}_{T+h}(x)$ denote the unreconciled h-step-ahead forecasts at all levels, and let $\tilde{\boldsymbol{y}}_{T+h}(x)$ denote the correspondingly reconciled h-step-ahead forecasts which satisfy all aggregation constraints. Any linear reconciliation method, according to Wickramasuriya *et al.* (2019), can be expressed as

$$\tilde{\boldsymbol{y}}_{T+h}(x) = \boldsymbol{S}_{T+h}(x)\boldsymbol{P}\hat{\boldsymbol{y}}_{T+h}(x), \tag{8}$$

where in our example $\boldsymbol{P}$ is a 102×105 matrix. The selection of $\boldsymbol{P}$ is not unique and is a key step in forecast reconciliation.

3.2 Bottom-up reconciliation

As mentioned earlier, the bottom-up reconciliation method aggregates bottom-level forecasts to produce higher-level forecasts in the hierarchy, and thus ensures the coherence in forecasts at all levels (Dangerfield and Morris, 1992; Zellner and Tobias, 2000). In the case of bottom-up approach, the $\boldsymbol{P}$ matrix is chosen as

$$\boldsymbol{P} = (\boldsymbol{0}_{102\times 3}, \boldsymbol{I}_{102}), \tag{9}$$

where $\boldsymbol{0}_{102\times 3}$ is a zero matrix of dimension 102×3.

It can be easily verified that

$$\tilde{\boldsymbol{y}}_{T+h}(x) = \boldsymbol{S}_{T+h}(x)\hat{\boldsymbol{b}}_{T+h}(x), \tag{10}$$

where $\hat{\boldsymbol{b}}_{T+h}(x)$ contains h-step-ahead unreconciled forecasts of the bottom-level data series. Therefore, the reconciled age-gender-specific mortality forecasts are obtained as

$$\tilde{m}^{M}_{T+h}(x) = \sum_{i=1}^{51} \frac{E^{M}_{i,T+h}(x)}{E^{M}_{T+h}(x)} \times \hat{m}^{M}_{i,T+h}(x). \tag{11}$$

$$\tilde{m}^{F}_{T+h}(x) = \sum_{i=1}^{51} \frac{E^{F}_{i,T+h}(x)}{E^{F}_{T+h}(x)} \times \hat{m}^{F}_{i,T+h}(x). \tag{12}$$

Similarly, the reconciled age-specific mortality forecasts at the top level are obtained as

$$\tilde{m}_{T+h}(x) = \tilde{m}^{M}_{T+h}(x) \times \frac{E^{M}_{T+h}(x)}{E_{T+h}(x)} + \tilde{m}^{F}_{T+h}(x) \times \frac{E^{F}_{T+h}(x)}{E_{T+h}(x)}. \tag{13}$$

3.3 MinT reconciliation

Hyndman *et al.* (2011) provided a method to estimate $\boldsymbol{P}$ using all available information such that the reconciled forecasts are unbiased given that the unreconciled forecasts are also unbiased. Wickramasuriya *et al.* (2019) extended the OLS method by proposing an alternative estimator of $\boldsymbol{P}$. Wickramasuriya *et al.* (2019) selected $\boldsymbol{P}$ to be the matrix that minimizes the trace of the covariance matrix of the in-sample reconciled forecast errors. Therefore, the method is referred to as the trace minimization method.

Based on the MinT approach, the reconciliation matrix $\boldsymbol{P}$ is given by

$$\boldsymbol{P} = (\boldsymbol{S}'_{T+h}(x)\boldsymbol{W}_h^{-1}\boldsymbol{S}_{T+h}(x))^{-1}\boldsymbol{S}'_{T+h}(x)\boldsymbol{W}_h^{-1}, \tag{14}$$

where $\boldsymbol{W}_h$ represents the covariance matrix of the h-step-ahead in-sample unreconciled forecast errors, which is VAR$[\boldsymbol{y}_{t+h}(x) - \tilde{\boldsymbol{y}}_{t+h}(x)|\boldsymbol{y}_1(x), \boldsymbol{y}_2(x), \ldots, \boldsymbol{y}_t(x)]$.[5]

Wickramasuriya *et al.* (2019) also showed that the reconciled forecasts will be at least as accurate as the unreconciled forecasts. Moreover, the reconciliation process takes into account the dependence structure across different levels of the hierarchy, which is particularly beneficial when dependence is ignored in the process of obtaining unreconciled forecasts for each individual time series. In our case, the bottom level contains 102 mortality time series, which makes it extremely challenging to apply a joint modeling approach to producing forecasts. Therefore, the MinT reconciliation method becomes a particularly effective tool when we obtain unreconciled forecasts based on independent models.

3.4 Reconciling probabilistic forecasts

Besides mean forecasts, in this paper we are also interested in the probabilistic forecasts of mortality rates at all levels of the hierarchy. In other words, we want to produce coherent prediction intervals for state- and national- level mortality rates.

As far as we know, there have been very few studies on reconciling probabilistic forecasts in the literature (see *e.g.* Ben Taieb *et al.*, 2017; Jeon *et al.*, 2019; Panagiotelis *et al.*, 2020b). In this paper, we adopt the "ranked sample" approach introduced by Jeon *et al.* (2019), as it has shown strong performance when there is a positive correlation between the underlying series. In our empirical study, we also expect a positive dependence structure across mortality experience of different states and genders.

To illustrate probabilistic forecast reconciliation, we further define the following terms:

- Define $\boldsymbol{f}(\boldsymbol{y}_{T+h}(x)|\boldsymbol{y}_1(x), \boldsymbol{y}_2(x), \ldots, \boldsymbol{y}_T(x))$ as the predictive probabilistic distribution of h-step-ahead forecasts at age x;
- Let $\hat{\boldsymbol{y}}^k_{T+h}$ denote the k^{th} sample of unreconciled forecasts generated from the predictive probabilistic distribution $\boldsymbol{f}(\boldsymbol{y}_{T+h}(x)|\boldsymbol{y}_1(x), \boldsymbol{y}_2(x), \ldots, \boldsymbol{y}_T(x))$;
- A sample of size N generated from $\boldsymbol{f}(\boldsymbol{y}_{T+h}(x)|\boldsymbol{y}_1(x), \boldsymbol{y}_2(x), \ldots, \boldsymbol{y}_T(x))$ is denoted by $\hat{\boldsymbol{Y}}_{T+h}(x)$, where $\hat{\boldsymbol{Y}}_{T+h}(x) = (\hat{\boldsymbol{y}}^1_{T+h}(x), \hat{\boldsymbol{y}}^2_{T+h}(x), \ldots, \hat{\boldsymbol{y}}^N_{T+h}(x))$.

Similar to point forecast reconciliation, we can express the reconciliation process as

$$\tilde{\boldsymbol{Y}}_{T+h}(x) = \boldsymbol{S}_{T+h}(x)\boldsymbol{P}\hat{\boldsymbol{Y}}_{T+h}(x), \tag{15}$$

where $\boldsymbol{S}_{T+h}(x)$ and $\boldsymbol{P}$ are the same as defined in Section 3.1, and $\tilde{\boldsymbol{Y}}_{T+h}(x)$ represents the reconciled forecasts for N sample paths.

Following the "ranked sample" approach by Jeon *et al.* (2019), we first arrange the generated sample in ascending order for each data series from the predictive distribution. We then construct the unreconciled forecasts $\hat{\boldsymbol{Y}}_{T+h}(x)$ based on the ranked samples from all series. Finally, MinT reconciliation approach can be applied to estimate matrix $\boldsymbol{P}$, and thus reconciled probabilistic forecasts can be obtained.

[5] In this paper, we have used the "shrinkage" estimator of $\boldsymbol{W}_h$ as described in Wickramasuriya *et al.* (2019), Section 2.4.

4 Empirical study

In this section, we present the empirical results based on the U.S. mortality experience. First, we briefly review the Lee-Carter model (Lee and Carter, 1992) and the ARIMA time series model which are used to produce unreconciled mortality forecasts and the h-step-ahead summing matrix, respectively.[6] Second, we choose the mortality data during 2008–2017 as our holdout sample and compare the out-of-sample forecast accuracy of the MinT reconciliation method with the bottom-up method and the independent method (unreconciled forecasts). Lastly, we project a 10-year-ahead mortality rates up to 2027 at state- and national-level and visualize these results.

4.1 The Lee-Carter model and the ARIMA model

Stochastic mortality modelling was popularised after the introduction of the Lee Carter model (Lee and Carter, 1992). It was originally introduced to model the U.S. mortality experience and has shown robust performance in long-term forecasting. The model is specified as follows

$$\log m_t(x) = a(x) + b(x)\kappa_t + \varepsilon_t(x), \tag{16}$$

where $a(x)$ and $b(x)$ are age effects, κ_t is the time effect and $\epsilon_t(x)$ is the error term. To project future mortality rates, κ_t is normally modelled as a random walk with drift process.

As multi-population mortality modeling is not required for the reconciliation approach, we first project mortality rates at all levels in the hierarchy based on the univariate Lee-Carter model. Another important input of the reconciliation process is the h-step-ahead forecasts of the summing matrix. Recall that $\boldsymbol{S}_{T+h}(x)$ is given by

$$\boldsymbol{S}_{T+h}(x) = \begin{pmatrix} \frac{E^M_{1,T+h}(x)}{E_{T+h}(x)} & \frac{E^M_{2,T+h}(x)}{E_{T+h}(x)} & \cdots & \frac{E^M_{51,T+h}(x)}{E_{T+h}(x)} & \frac{E^F_{1,T+h}(x)}{E_{T+h}(x)} & \frac{E^F_{2,T+h}(x)}{E_{T+h}(x)} & \cdots & \frac{E^F_{51,T+h}(x)}{E_{T+h}(x)} \\ \frac{E^M_{1,T+h}(x)}{E^M_{T+h}(x)} & \frac{E^M_{2,T+h}(x)}{E^M_{T+h}(x)} & \cdots & \frac{E^M_{51,T+h}(x)}{E^M_{T+h}(x)} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \frac{E^F_{1,T+h}(x)}{E^F_{T+h}(x)} & \frac{E^F_{2,T+h}(x)}{E^F_{T+h}(x)} & \cdots & \frac{E^F_{51,T+h}(x)}{E^F_{T+h}(x)} \\ & & & & & & & \\ & & & & \boldsymbol{I}_{102} & & & \end{pmatrix}.$$

To project the exposure ratios in the first three rows of the matrix, we adopt a similar approach as in Shang and Hyndman (2017). We forecast all ratios using independent ARIMA models, where Akaike Information Criteria is used to select the optimal ARIMA model for each ratio series (Akaike, 1974). In contrast to Shang and Hyndman (2017), we adjust the $\boldsymbol{S}$ such that all rows of the matrix add up to one. This is to make sure that the aggregation constraints in the hierarchy are fully satisfied. We adjust the first three rows of the matrix by dividing the forecast ratios by the row sum. The MinT reconciliation method and the bottom-up method are then applied for each age group as described in Section 3.

[6] We have also adopted the 6-factor mortality model proposed by Hyndman and Ullah (2007) to produce independent (*i.e.* unreconciled) forecasts, and find the results in-line with conclusions and findings based on the Lee-Carter model. These results are available upon request.

4.2 Point forecast accuracy

As mentioned earlier in the section, we choose mortality data during 2008–2017 as holdout sample to evaluate the out-of-sample forecast performance of proposed methods. Therefore, the Lee-Carter model and the ARIMA model are fitted using the data from 1969 to 2007. To evaluate the point forecast accuracy, we choose the mean absolute percentage error (MAPE) as the error measure which has been frequently used in the field of mortality modelling (see *e.g.* Lee and Miller, 2001; O'Hare and Li, 2012; Shang and Haberman, 2017). The MAPE for h-step-ahead forecasts across all ages is defined as follows

$$\text{MAPE} = \frac{1}{86 \times h} \sum_{x=0}^{85+} \sum_{u=1}^{h} \left| \frac{\hat{m}_{2007+u}(x) - m_{2007+u}(x)}{m_{2007+u}(x)} \right| \times 100\%.$$

In Table 2, we report the values of MAPE based on the independent method (*i.e.* unreconciled forecasts), the bottom-up method and the MinT method. All levels in the hierarchy are considered which include total mortality rates, gender-specific mortality rates, and state-gender-specific mortality rates. It should be noted that as the bottom-up method generates high-level forecasts by adding up bottom-level forecast, at the bottom-level independent method and bottom-up method provide the same MAPE values. We can see that, overall, MinT reconciliation method gives most accurate point forecasts among the three methods across all forecast horizons. At the top level, the MinT reconciliation is the best performing method while the bottom-up reconciliation is the worst performing method. At the middle level, on average MinT still has the strongest forecast performance.

It is worth noting that although the MinT approach gives the best overall forecasts and better forecasts for male mortality experience across the 51 states, the independent method and bottom-up method give better forecasts for state-level female mortality experience. A motivation for forecast reconciliation is to improve forecast accuracy for the hierarchy overall. In the MinT approach, optimality is understood to be minimisation of the trace of the variance covariance matrix of forecast errors for the entire hierarchy. There is now a substantial body of theoretical (Panagiotelis *et al.*, 2020a) and empirical (see *e.g.* Zhang and Dong, 2018; Fildes *et al.*, 2019; Yagli *et al.*, 2019) literature that supports the properties of MinT. However, despite an improvement in forecast accuracy overall, it is common to find some series in a hierarchy, for which base forecasts are more accurate than the reconciled forecasts obtained by MinT. Therefore, for females, although the reconciled forecasts provide higher overall accuracy, for some individual series in the hierarchy, the unreconciled forecasts have smaller errors. A deeper theoretical understanding of why reconciliation leads to better forecast accuracy for some series and worse forecast accuracy for other series remains an open question, and is beyond the scope of this paper.

Hyndman *et al.* (2011) have shown that if the base forecasts (*i.e.* independent forecasts in this paper) are unbiased, the OLS (and thus MinT) reconciled forecasts will also be unbiased. To check if the independent forecasts produced by the Lee-Carter model are biased, we compute the mean percentage error (MPE) of the Lee-Carter mortality forecasts, which is defined as follows

$$\text{MPE} = \frac{1}{86 \times h} \sum_{x=0}^{85+} \sum_{u=1}^{h} \left(\frac{\hat{m}_{2007+u}(x) - m_{2007+u}(x)}{m_{2007+u}(x)} \right) \times 100\%.$$

We present the MPE results together with the Mean Absolute Percentage Error (MAPE) results based on independent forecasts in Table 4 of the Appendix. Since the values of MPE are

much smaller than MAPE for all forecasting horizons, it implies that the unreconciled forecasts based on the Lee-Carter model are unbiased (Li *et al.*, 2015, show unbiasedness in a similar way).

Table 2: MAPE (×100) of out-of-sample forecasts for period 2008–2017, all ages

Method	h	**Total**	**Male**	**Female**	**Male-state**	**Female-state**
Independent	1	0.0816	0.0816	0.1000	0.2016	**0.2226**
	2	0.0870	0.0864	0.1060	0.2062	**0.2291**
	3	0.0922	0.0912	0.1114	0.2114	**0.2353**
	4	0.0980	**0.0967**	0.1171	0.2171	**0.2417**
	5	0.1044	**0.1029**	0.1228	0.2233	**0.2481**
	6	0.1110	**0.1096**	0.1284	0.2298	**0.2545**
	7	0.1177	**0.1164**	0.1341	0.2365	**0.2610**
	8	0.1243	**0.1232**	0.1399	0.2433	**0.2675**
	9	0.1309	**0.1301**	0.1458	0.2503	**0.2741**
	10	0.1375	**0.1372**	0.1518	0.2574	**0.2807**
Bottom-up	1	0.0884	0.0929	0.0996	0.2016	**0.2226**
	2	0.0949	0.0977	0.1063	0.2062	**0.2291**
	3	0.1013	0.1032	0.1128	0.2114	**0.2353**
	4	0.1084	0.1098	0.1199	0.2171	**0.2417**
	5	0.1155	0.1167	0.1272	0.2233	**0.2481**
	6	0.1225	0.1237	0.1345	0.2298	**0.2545**
	7	0.1295	0.1307	0.1417	0.2365	**0.2610**
	8	0.1363	0.1376	0.1489	0.2433	**0.2675**
	9	0.1431	0.1446	0.1558	0.2503	**0.2741**
	10	0.1498	0.1517	0.1628	0.2574	**0.2807**
MinT	1	**0.0790**	**0.0814**	**0.0914**	**0.1977**	0.2259
	2	**0.0847**	**0.0862**	**0.0970**	**0.2019**	0.2325
	3	**0.0904**	**0.0910**	**0.1021**	**0.2066**	0.2388
	4	**0.0967**	0.0975	**0.1078**	**0.2118**	0.2452
	5	**0.1033**	0.1041	**0.1137**	**0.2174**	0.2517
	6	**0.1100**	0.1108	**0.1198**	**0.2234**	0.2580
	7	**0.1165**	0.1175	**0.1258**	**0.2296**	0.2644
	8	**0.1229**	0.1242	**0.1320**	**0.2359**	0.2708
	9	**0.1292**	0.1309	**0.1383**	**0.2424**	0.2772
	10	**0.1357**	0.1376	**0.1445**	**0.2489**	0.2837

4.3 Interval forecast accuracy

To evaluate the interval forecast accuracy of the proposed methods, we compute the Winkler score for prediction intervals (Winkler, 1972). Let $[l, u]$ be the 100(1-α)% prediction interval and the actual observed value is m, the Winkler score is calculated as

$$(u-l)+\frac{2}{\alpha}(l-m)\mathbb{1}(m<l)+\frac{2}{\alpha}(m-u)\mathbb{1}(m>u), \tag{17}$$

where $\mathbb{1}$ denotes the indicator function. In this paper, we set α to be 0.05 thus evaluating the accuracy of 95% prediction intervals.

In Table 3, we report the Winkler score based on all three methods. It is shown that the MinT reconciliation method substantially improves the interval forecast accuracy at the top level. The Winkler score based on the MinT method is around 40% smaller compared to the independent forecasts. At the middle level where gender-specific mortality forecasts are considered, the MinT method still outperforms the other two methods in most cases. Similar to the results on point forecasts, for state-level mortality forecasts, the MinT method gives better results for males, and the independent method and the bottom-up method gives better results for females. However, these differences in the Winkler scores are rather marginal. Again, we do not further comment on this result and leave it for future research.

Table 3: Winkler score of out-of-sample forecasts for period 2008–2017, all ages

Method	h	**Total**	**Male**	**Female**	**Male-state**	**Female-state**
Independent	1	0.9605	1.7259	0.8946	1.8984	**0.8530**
	2	1.1301	1.9584	1.0422	2.0897	**0.9464**
	3	1.3172	2.1789	1.2251	2.2861	**1.0522**
	4	1.5014	2.4088	1.4083	2.5076	**1.1734**
	5	1.6829	2.6436	1.5899	2.7386	**1.3084**
	6	1.8576	2.8611	1.7620	2.9676	**1.4483**
	7	2.0411	3.0842	1.9310	3.2022	**1.5951**
	8	2.2117	3.2853	2.0926	3.4294	**1.7434**
	9	2.3807	3.4814	2.2476	3.6594	**1.8928**
	10	2.5492	3.6831	2.3979	3.8958	**2.0426**
Bottom-up	1	0.5601	1.1339	0.4924	1.8984	**0.8530**
	2	0.6948	1.3004	0.5415	2.0897	**0.9464**
	3	0.8390	1.4721	0.6156	2.2861	**1.0522**
	4	0.9791	1.6397	0.7137	2.5076	**1.1734**
	5	1.1275	1.8160	0.8298	2.7386	**1.3084**
	6	1.2741	1.9882	0.9497	2.9676	**1.4483**
	7	1.4212	2.1698	1.0741	3.2022	**1.5951**
	8	1.5673	2.3380	**1.1975**	3.4294	**1.7434**
	9	1.7124	2.5065	**1.3198**	3.6594	**1.8928**
	10	1.8535	2.6718	**1.4387**	3.8958	**2.0426**
MinT	1	**0.5300**	**1.0732**	**0.4540**	**1.8596**	1.1385
	2	**0.6558**	**1.2381**	**0.5092**	**2.0510**	1.2315
	3	**0.7996**	**1.4122**	**0.5837**	**2.2462**	1.3372
	4	**0.9418**	**1.5836**	**0.6872**	**2.4662**	1.4589
	5	**1.0928**	**1.7579**	**0.8107**	**2.6980**	1.5947
	6	**1.2430**	**1.9302**	**0.9381**	**2.9305**	1.7347
	7	**1.3944**	**2.1140**	**1.0701**	**3.1716**	1.8828
	8	**1.5465**	**2.2848**	1.2020	**3.4068**	2.0333
	9	**1.6996**	**2.4578**	1.3339	**3.6456**	2.1862
	10	**1.8504**	**2.6302**	1.4628	**3.8919**	2.3410

4.4 Projecting mortality rates to 2027

In this subsection we apply the MinT reconciliation method to project the state- and national-level mortality rates for both genders for the period 2018–2027. Currie and Schwandt (2016) found that mortality inequality has decreased among children and adolescents, while increasing for older adults, in the U.S. over the period 1990–2010. Dwyer-Lindgren *et al.* (2017) also found that the trajectory of mortality inequalities differs by age, where older adults experience the greatest geographical inequalities. It is also stated that the increase in life expectancy gap over the last 3 decades have been largely driven by the growing mortality inequality in older ages (Dwyer-Lindgren *et al.*, 2017). Therefore, in this section we focus on mortality forecasts for the "oldest-old" (ages 85+), and present the mortality forecasts of males and females for all states in Figures 6 and 7, respectively.

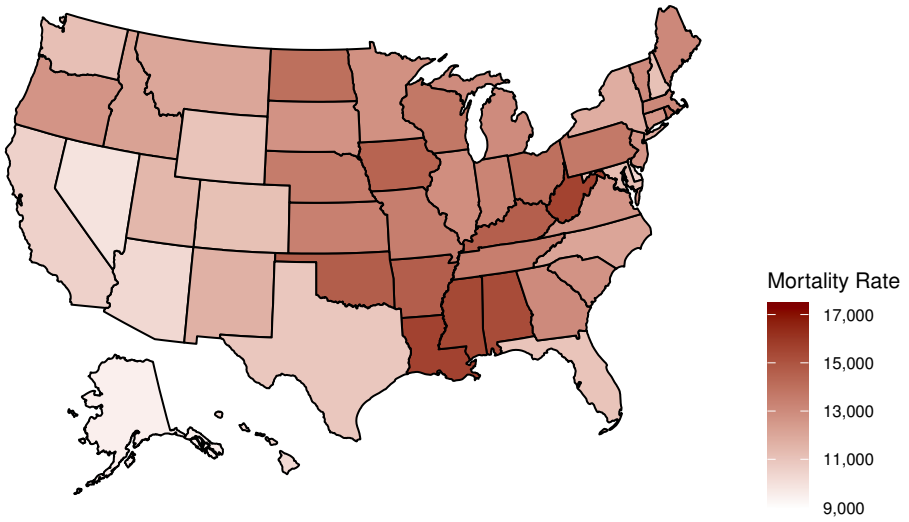


Figure 6: U.S. state-level male mortality forecasts at age 85+ in 2027

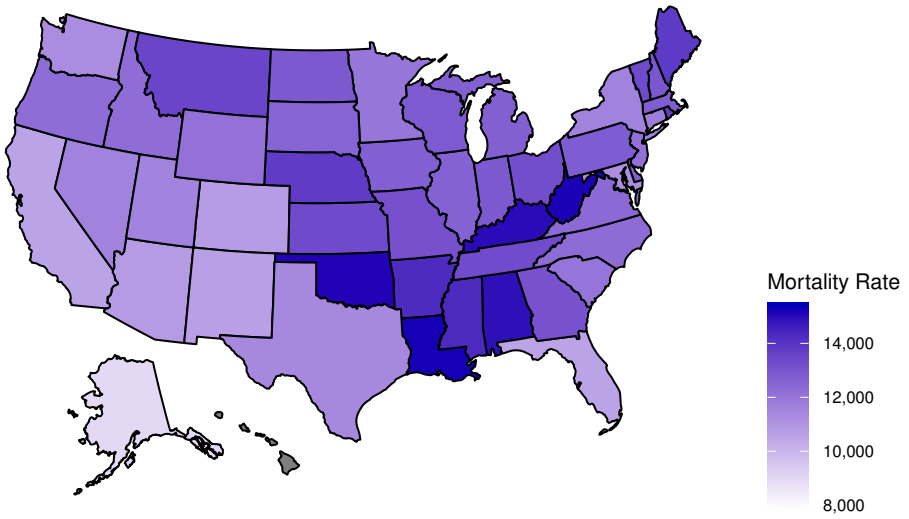


Figure 7: U.S. state-level female mortality forecasts at age 85+ in 2027

Figures 6 and 7 show that despite the overall reduction in mortality rates for ages 85+, the old-age survival disadvantage across Southern states will continue to exist. In fact, based on the reconciled forecasts, some Southern states are not expected to experience any mortality improvement at all over the next decade. For 85+ males, the "bottom" 3 states with the highest mortality predictions are Louisiana, West Virginia, and Mississippi. On the other hand, for 85+ females, the "bottom" 3 states with the highest mortality predictions are Louisiana, West Virginia, and Oklahoma. Moreover, high mortality rate forecasts are also observed in some "rust belt states" such as Pennsylvania, Ohio, and Indiana. Compared to the plots (g) and (h) in Figure 1, we also observe improvements in mortality in several Pacific states and mountain states, especially for male mortality experience.

In Figure 8, we plot the national-level forecasts for males, females, and total, over the period 2018–2027. We plot the reconciled point forecasts as well as the reconciled interval forecasts at the 95% confidence level. It can be seen that males are predicted to experience a more rapid mortality improvement compared to females. However, we also observe a slight slow-down in the decreasing trend of mortality rates, especially in the case of female mortality rates. In fact, Figure 8 shows that the slow-down may have already happened since the early 2010s. In addition, based on our forecasts, we find that the future old-age mortality improvement is more likely to be result from the improvement in male mortality rates, which is broadly consistent with the finding of narrowing gap between female and male life expectancy by Wang *et al.* (2013). Finally, it should be noted that the reconciled forecasts produced in this research are based on the scenario where no significant change or reform is made in federal/state government policies or health care systems.

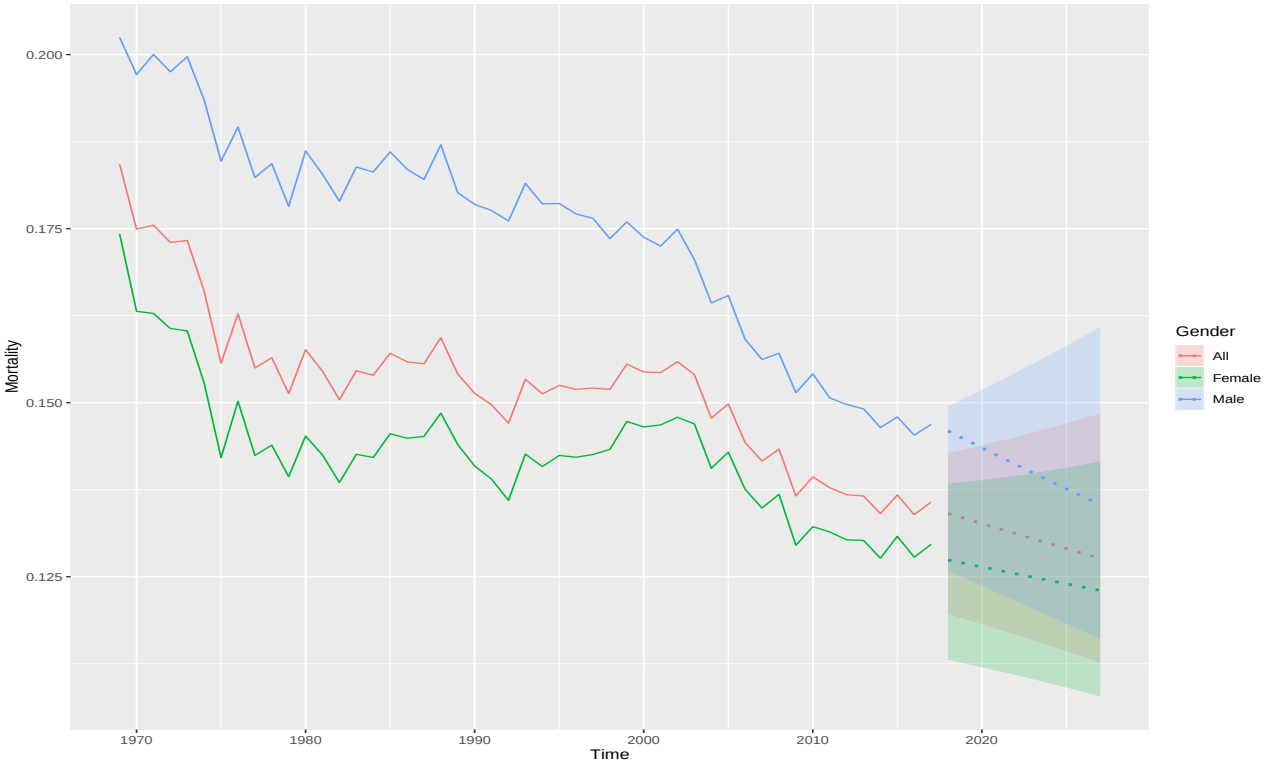


Figure 8: U.S. mortality forecasts at age 85+ for the period 2018–2027

5 Conclusion

The geographical heterogeneity in mortality experience remains an important issue to be addressed and investigated. Examining mortality inequality across geographical regions allows for tracking longevity and health disparities over time, which in turn enables assessing factors related to these disparities. In this paper, we propose a forecast reconciliation approach to project U.S. state- and national-level mortality rates in a hierarchical setting. Coherent point and interval mortality forecasts are produced using the MinT reconciliation method by Wickramasuriya *et al.* (2019) and Jeon *et al.* (2019).

Based on the empirical results, we show that the MinT reconciliation method provides both superior point forecasts and interval forecasts, compared to the independent method and the bottom-up method. Projecting the 10-year-ahead mortality rates to 2027, we anticipate that less favorable old-age mortality experience in the Southern states is likely to continue in the next decade. We also find that the future mortality improvement rate at old ages will tend to slow down, in particular for females. Results of this research can provide insight into resource allocation, as well as health and welfare policy design. Future research could be built upon this paper, exploring what factors might be driving the geographical variations in the life expectancy, and more importantly how to reduce these inequalities and achieve more equitable health outcomes.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**(6), 716–723.

Andreev, K., Gu, D., and Dupre, M. (2017). *Regional Mortality in the United States at Ages 80 and Older: An Analysis of Direct Estimates, 1959–2011*. Living to 100 Symposium, Society of Actuaries.

Athanasopoulos, G., Ahmed, R. A., and Hyndman, R. J. (2009). Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, **25**(1), 146–166.

Ben Taieb, S., Taylor, J. W., and Hyndman, R. J. (2017). Coherent probabilistic forecasts for hierarchical time series. In *International Conference on Machine Learning*, pages 3348–3357.

Borges, C. E., Penya, Y. K., and Fernandez, I. (2013). Evaluating combined load forecasting in large power systems and smart grids. *IEEE Transactions on Industrial Informatics*, **9**(3), 1570–1577.

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, **73**(4), 687–718.

Capistrán, C., Constandse, C., and Ramos-Francia, M. (2010). Multi-horizon inflation forecasts using disaggregated data. *Economic Modelling*, **27**(3), 666–677.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *JAMA*, **315**(16), 1750–1766.

Currie, J. and Schwandt, H. (2016). Inequality in mortality decreased among the young while increasing for older adults, 1990–2010. *Science*, **352**(6286), 708–712.

Dangerfield, B. J. and Morris, J. S. (1992). Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting*, **8**(2), 233–241.

Davids, B.-O., Hutchins, S. S., Jones, C. P., and Hood, J. R. (2014). Disparities in life expectancy across US counties linked to county social factors, 2009 community health status indicators (CHSI). *Journal of Racial and Ethnic Health Disparities*, **1**(1), 2–11.

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Kutz, M. J., Huynh, C., Barber, R. M., Shackelford, K. A., Mackenbach, J. P., Van Lenthe, F. J., *et al.* (2016). Us county-level trends in mortality rates for major causes of death, 1980-2014. *Jama*, **316**(22), 2385–2401.

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Mackenbach, J. P., van Lenthe, F. J., Mokdad, A. H., and Murray, C. J. (2017). Inequalities in life expectancy among US counties, 1980 to 2014: temporal trends and key drivers. *JAMA internal medicine*, **177**(7), 1003–1011.

Fildes, R., Ma, S., and Kolassa, S. (2019). Retail forecasting: Research and Practice. *International Journal of Forecasting*.

Holman, R. (2017). *US Population Mortality Rate Study - Variation by Age Group, Cause of Death and Region from 2000–2015*. Society of Actuaries, United States.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, **51**(10), 4942–4956.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, **55**(9), 2579–2589.

Jeon, J., Panagiotelis, A., and Petropoulos, F. (2019). Probabilistic forecast reconciliation with applications to wind power and electric load. *European Journal of Operational Research*.

Kahn, K. B. (1998). Revisiting top-down versus bottom-up forecasting. *The Journal of Business Forecasting*, **17**(2), 14.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**(1/2), 81–93.

Kulkarni, S. C., Levin-Rector, A., Ezzati, M., and Murray, C. J. (2011). Falling behind: life expectancy in us counties from 2000 to 2007 in an international context. *Population health metrics*, **9**(1), 16.

Lee, R. and Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, **38**(4), 537–549.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**(419), 659–671.

Li, H. and Tang, Q. (2019). Analyzing mortality bond indexes via hierarchical forecast reconciliation. *ASTIN Bulletin: The Journal of the IAA*, **49**(3), 823–846.

Li, H., O'Hare, C., and Zhang, X. (2015). A semiparametric panel approach to mortality modeling. *Insurance: Mathematics and Economics*, **61**, 264–270.

Li, H., Li, H., Lu, Y., and Panagiotelis, A. (2019). A forecast reconciliation approach to cause-of-death mortality modeling. *Insurance: Mathematics and Economics*, **86**, 122–133.

O'Hare, C. and Li, Y. (2012). Explaining young mortality. *Insurance: Mathematics and Economics*, **50**(1), 12–25.

Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., and Hyndman, R. J. (2020a). Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*.

Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., and Hyndman, R. (2020b). Probabilistic forecast reconciliation: Properties, evaluation and score optimisation. Monash econometrics and business statistics working paper series 26/20.

Rice, D. P. and Feldman, J. J. (1983). Living longer in the united states: Demographic changes and health needs of the elderly. *The Milbank Memorial Fund Quarterly. Health and Society*, pages 362–396.

Shang, H. L. and Haberman, S. (2017). Grouped multivariate and functional time series forecasting: An application to annuity pricing. *Insurance: Mathematics and Economics*, **75**, 166–179.

Shang, H. L. and Hyndman, R. J. (2017). Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, **26**(2), 330–343.

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., and Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, **252**(1), 1–26.

US Department of Health and Human Services. (2010). Healthy People 2020 Framework. Available at: https://www.healthypeople.gov/sites/default/files/HP2020Framework.pdf.

Wang, H., Schumacher, A. E., Levitz, C. E., Mokdad, A. H., and Murray, C. J. (2013). Left behind: widening disparities for males and females in us county life expectancy, 1985–2010. *Population health metrics*, **11**(1), 8.

Wickramasuriya, S. L., Athanasopoulos, G., and Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, **114**(526), 804–819.

Winkler, R. L. (1972). A decision-theoretic approach to interval estimation. *Journal of the American Statistical Association*, **67**(337), 187–191.

Yagli, G. M., Yang, D., and Srinivasan, D. (2019). Reconciling solar forecasts: Sequential reconciliation. *Solar Energy*, **179**, 391–397.

Zellner, A. and Tobias, J. (2000). A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting*, **19**(5), 457–465.

Zhang, Y. and Dong, J. (2018). Least squares-based optimal reconciliation method for hierarchical forecasts of wind power generation. *IEEE Transactions on Power Systems*.

Appendix

Table 4: MAPE and MPE based on independent forecasts

h	**Total**		**Male**		**Female**		**Male-state**		**Female-state**	
	MAPE	**MPE**	**MAPE**	**MPE**	**MAPE**	**MPE**	**MAPE**	**MPE**	**MAPE**	**MPE**
1	0.0816	0.0090	0.0816	0.0075	0.1000	0.0246	0.2016	0.0381	0.2226	0.0531
2	0.0870	0.0073	0.0864	0.0056	0.1060	0.0279	0.2062	0.0386	0.2291	0.0531
3	0.0922	0.0056	0.0912	0.0036	0.1114	0.0311	0.2114	0.0391	0.2353	0.0534
4	0.0980	0.0039	0.0967	0.0018	0.1171	0.0343	0.2171	0.0396	0.2417	0.0539
5	0.1044	0.0023	0.1029	0.0003	0.1228	0.0376	0.2233	0.0404	0.2481	0.0547
6	0.1110	0.0007	0.1096	-0.0017	0.1284	0.0409	0.2298	0.0412	0.2545	0.0558
7	0.1177	-0.0007	0.1164	-0.0033	0.1341	0.0443	0.2365	0.0421	0.2610	0.0571
8	0.1243	-0.0020	0.1232	-0.0047	0.1399	0.0478	0.2433	0.0430	0.2675	0.0585
9	0.1309	-0.0032	0.1301	-0.0061	0.1458	0.0515	0.2503	0.0440	0.2741	0.0602
10	0.1375	-0.0042	0.1372	-0.0074	0.1518	0.0553	0.2574	0.0450	0.2807	0.0621