A forecast reconciliation approach to cause-of-death mortality modeling

Han Li, Hong Li, Yang Lu and **Anastasios Panagiotelis**

November 21, 2019

Forecasting death

- *"The first symptom of death is birth."*

Forecasting death

- "*The first symptom of death is birth.*"

–Stanislaw Jerzy Lec, Polish Poet

Forecasting death

- “*The first symptom of death is birth.*”

  –Stanislaw Jerzy Lec, Polish Poet

- “*Life is not fair. Life SUCKS and then you DIE.*”

Forecasting death

- "*The first symptom of death is birth.*"

–Stanislaw Jerzy Lec, Polish Poet

- "*Life is not fair. Life SUCKS and then you DIE.*"

–Vince McMahon, WWE President

Interesting questions

- Forecast mortality rate for both total and individual causes.
  - Input to bond pricing models.
  - Capital reserving implications.
- Cause Elimination
  - What happens if we cure cancer?

What are we forecasting?

Central mortality rate $m_t(x)$, reflects death probability for people aged x in the middle of calendar year t. It is estimated by:

What are we forecasting?

Central mortality rate $m_t(x)$, reflects death probability for people aged x in the middle of calendar year t. It is estimated by:

$$m_t(x) = \frac{d_t(x)}{E_t(x)}.$$

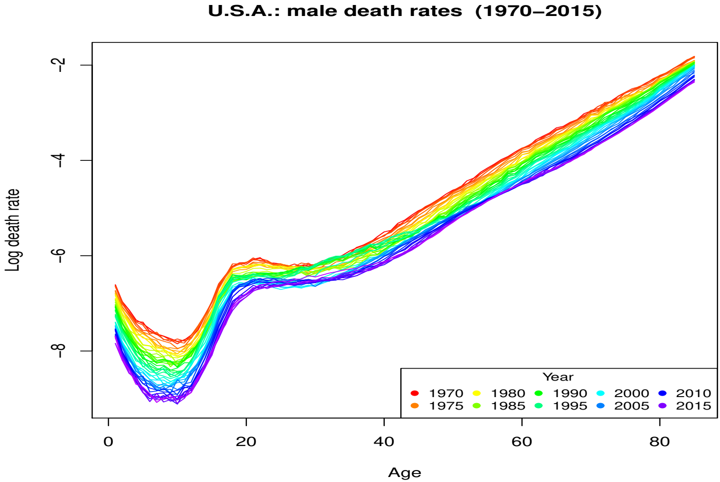
What are we forecasting?

Central mortality rate $m_t(x)$, reflects death probability for people aged x in the middle of calendar year t. It is estimated by:

$$m_t(x) = \frac{d_t(x)}{E_t(x)}.$$

Here, $d_t(x)$ is the number of deaths and $E_t(x)$ is the exposure for age x in year t.

What does this look like?



Causes of Death

- These are US data from the Human Mortality Database (HMD)and the National Center for Health Statistics (NCHS).

Causes of Death

- These are US data from the Human Mortality Database (HMD)and the National Center for Health Statistics (NCHS).
- Data on different causes of death are also available.

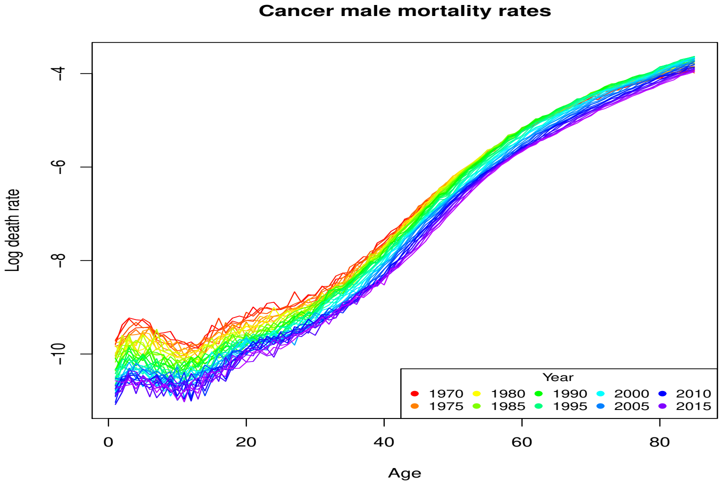
Causes of Death

- These are US data from the Human Mortality Database (HMD)and the National Center for Health Statistics (NCHS).
- Data on different causes of death are also available.
- We considered 8 causes: Cancer, Diabetes, External, Influenza, Mental, Nephritis, Vascular and Other.

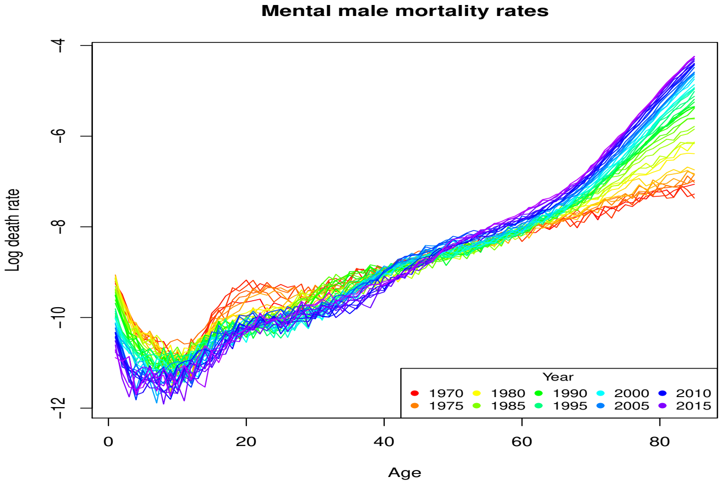
Causes of Death

- These are US data from the Human Mortality Database (HMD)and the National Center for Health Statistics (NCHS).
- Data on different causes of death are also available.
- We considered 8 causes: Cancer, Diabetes, External, Influenza, Mental, Nephritis, Vascular and Other.
- What do the data look like for different causes?

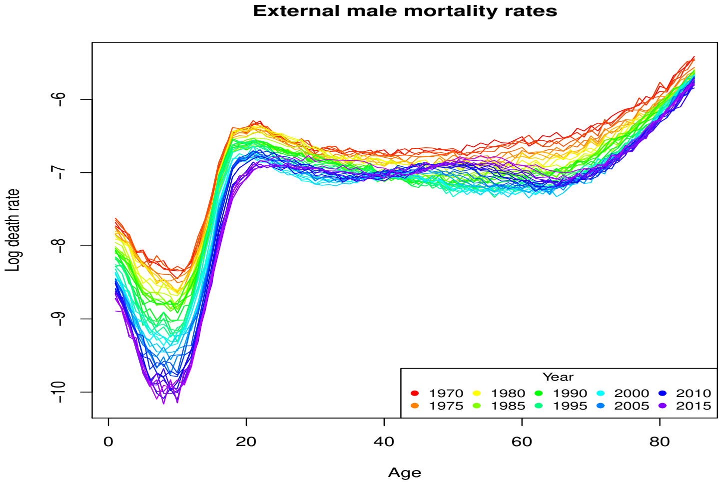
Cancer


Cancer male mortality rates
Log death rate
-4
-6
-8
-10
0
20
40
60
80
Age
Year
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015

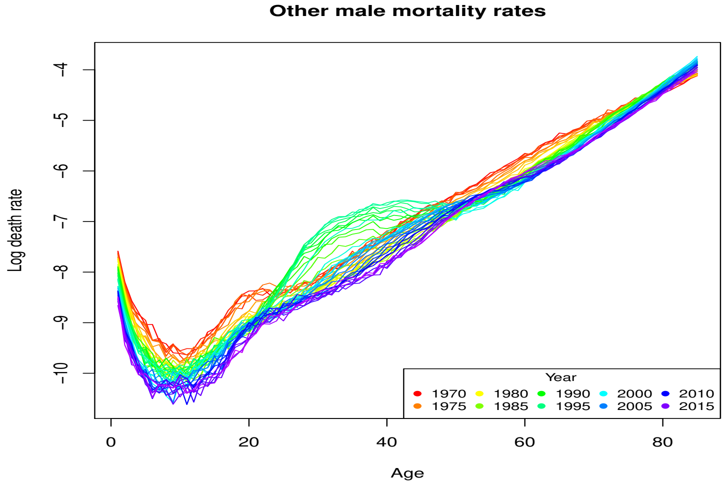
Mental


Mental male mortality rates
Log death rate
-4
-6
-8
-10
-12
0
20
40
60
80
Age
Year
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015

External


External male mortality rates
Log death rate
-6
-7
-8
-9
-10
0
20
40
60
80
Age
Year
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015

Other


Other male mortality rates
Log death rate
-4
-5
-6
-7
-8
-9
-10
0
20
40
60
80
Age
Year
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015

Hierarchy

- These are *leading* causes of death, so are mutually exclusive.

Hierarchy

- These are *leading* causes of death, so are mutually exclusive.
- The total number of deaths is the same as the sum of the deaths by each cause.

Hierarchy

- These are *leading* causes of death, so are mutually exclusive.
- The total number of deaths is the same as the sum of the deaths by each cause.
- All causes have the same exposed population (denominator).

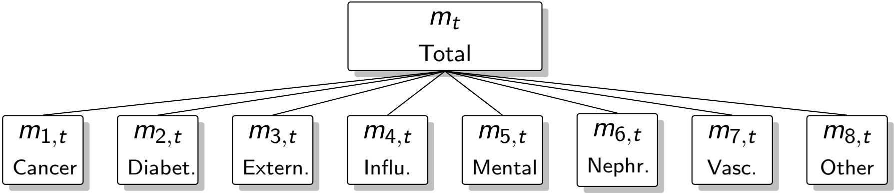
Hierarchy

- These are *leading* causes of death, so are mutually exclusive.
- The total number of deaths is the same as the sum of the deaths by each cause.
- All causes have the same exposed population (denominator).
- Therefore, for a given age, the total mortality rate must be the sum of the mortality rates by each cause.

Hierarchy

- These are *leading* causes of death, so are mutually exclusive.
- The total number of deaths is the same as the sum of the deaths by each cause.
- All causes have the same exposed population (denominator).
- Therefore, for a given age, the total mortality rate must be the sum of the mortality rates by each cause.
- These are a *hierarchical time series*.

Hierarchy


m_t
Total
$m_{1,t}$
Cancer
$m_{2,t}$
Diabet.
$m_{3,t}$
Extern.
$m_{4,t}$
Influ.
$m_{5,t}$
Mental
$m_{6,t}$
Nephr.
$m_{7,t}$
Vasc.
$m_{8,t}$
Other

Single level approaches

- **Top down** Forecast total mortality and allocate to causes using historical (or forecasted) proportions.

Single level approaches

- **Top down** Forecast total mortality and allocate to causes using historical (or forecasted) proportions.
  - Wastes information at bottom level.

Single level approaches

- **Top down** Forecast total mortality and allocate to causes using historical (or forecasted) proportions.
  - Wastes information at bottom level.
- **Bottom up:** Forecast cause specific mortality then add up.

Single level approaches

- **Top down** Forecast total mortality and allocate to causes using historical (or forecasted) proportions.
  - Wastes information at bottom level.
- **Bottom up:** Forecast cause specific mortality then add up.
  - Bottom level series are noisy.

Single level approaches

- **Top down** Forecast total mortality and allocate to causes using historical (or forecasted) proportions.
  - Wastes information at bottom level.
- **Bottom up:** Forecast cause specific mortality then add up.
  - Bottom level series are noisy.
- Why not forecast the total AND the cause specific mortality rates?

A new problem

- Excluding special cases, forecasts of cause-specific mortality not guaranteed to sum to forecast of total.

A new problem

- Excluding special cases, forecasts of cause-specific mortality not guaranteed to sum to forecast of total.
- Can lead to decisions that are not aligned.

A new problem

- Excluding special cases, forecasts of cause-specific mortality not guaranteed to sum to forecast of total.
- Can lead to decisions that are not aligned.
- We know that the target of the forecast *will* adhere to constraints

A new problem

- Excluding special cases, forecasts of cause-specific mortality not guaranteed to sum to forecast of total.
- Can lead to decisions that are not aligned.
- We know that the target of the forecast *will* adhere to constraints
- We should not waste this information.

Why not incorporate in modeling?

In mortality, a popular model is the Lee Carter (1992) model

Why not incorporate in modeling?

In mortality, a popular model is the Lee Carter (1992) model

$$\widehat{\log(m_t(x))} = a_x + b_x \kappa_t$$

Why not incorporate in modeling?

In mortality, a popular model is the Lee Carter (1992) model

$$\widehat{\log(m_t(x))} = a_x + b_x \kappa_t$$

- The intercept and coefficient a_x and b_x are age specific,
- The time trend κ_t follows a random walk with drift,
- Estimation using singular value decomposition.

Why not incorporate in modeling?

In mortality, a popular model is the Lee Carter (1992) model

$$\widehat{\log(m_t(x))} = a_x + b_x \kappa_t$$

- The intercept and coefficient a_x and b_x are age specific,
- The time trend κ_t follows a random walk with drift,
- Estimation using singular value decomposition.
- Hard to estimate a multivariate version of this,

Why not incorporate in modeling?

In mortality, a popular model is the Lee Carter (1992) model

$$\widehat{\log(m_t(x))} = a_x + b_x \kappa_t$$

- The intercept and coefficient a_x and b_x are age specific,
- The time trend κ_t follows a random walk with drift,
- Estimation using singular value decomposition.
- Hard to estimate a multivariate version of this, that also respects aggregation constraints.

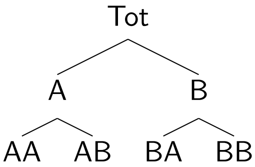
What can we do???

What can we do???

Forecast Reconciliation!!!

A simple hierarchy

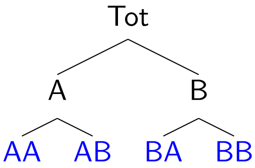
Consider a hierarchy given by



- Let n be the number of series, $\boldsymbol{y}_t$ be an n-vector of all series.

A simple hierarchy

Consider a hierarchy given by



- Let n be the number of series, $\boldsymbol{y}_t$ be an n-vector of all series.
- Let q be the number of bottom level series and $\boldsymbol{b}_t$ be an q-vector of the bottom level series.

The $\boldsymbol{S}$ matrix

Coherence holds when

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$$

The $n \times q$ matrix $\boldsymbol{S}$ defines the aggregation constraints, e.g.

$$\boldsymbol{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & \boldsymbol{I}_{4\times 4} & & \end{pmatrix}$$

As a regression model

- Cast the problem as a regression model with base forecasts $\hat{\boldsymbol{y}}_{T+h}$ as the "dependent variable" and $\boldsymbol{S}$ as the "design matrix".

$$\hat{\boldsymbol{y}}_{T+h} = \boldsymbol{S}\beta_{T+h} + \boldsymbol{e}_{T+h}$$

- Initial approach (Athanasopoulos et al, 2009; Hyndman et al, 2011) was to fit by OLS yielding reconciled forecasts:

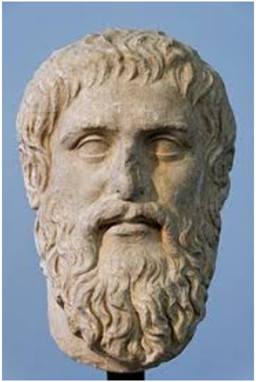
$$\tilde{\boldsymbol{y}}_{T+h} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_{T+h}$$

Generalisation

- Wherever we can use OLS we can use GLS

$$\tilde{\boldsymbol{y}}_{T+h} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\hat{\boldsymbol{y}}_{T+h}$$

- Diagonal $\boldsymbol{W}$ considered by Athanasopoulos et al (2017)
- MinT approach (Wickremasuriya et al, 2019) use a $\boldsymbol{W}$ that is an estimate of the *in-sample* forecast error covariance matrix.



ΑΓΕΩΜΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ

Those without knowledge of geometry may not enter.

Coherent Subspace

Definition

The **coherent subspace** is the q-dimensional linear subspace of $\mathbb{R}^n$ spanned by the columns of $\boldsymbol{S}$, i.e. $\mathfrak{s} = \text{sp}(\boldsymbol{S})$

Instead of using bottom-level series the top and $q-1$ bottom level could be used.
Although $\boldsymbol{S}$ would be different $\mathfrak{s}$ would be the same.

Coherent Point Forecast

Definition

A **coherent point forecast** is any forecast lying in the linear subspace $\mathfrak{s}$

Reconciled Point Forecast

Let $\hat{\boldsymbol{y}} \in \mathbb{R}^n$ be an incoherent forecast and $\psi(.)$ be a function $\psi : \mathbb{R}^n \rightarrow \mathfrak{s}$.

Definition

A **point forecast** $\tilde{\boldsymbol{y}}$ is reconciled with respect to *psi*(.) iff

$$\tilde{\boldsymbol{y}} = \psi(\hat{\boldsymbol{y}})$$

Projections

- While this definition is quite general, the special case where ψ is a projection has some interesting properties.

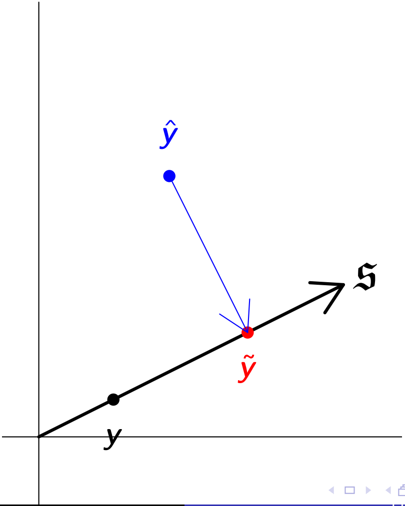
Projections

- While this definition is quite general, the special case where ψ is a projection has some interesting properties.
- For instance, if base forecasts are unbiased then reconciliation using a projection preserves unbiasedness.

Projections

- While this definition is quite general, the special case where ψ is a projection has some interesting properties.
- For instance, if base forecasts are unbiased then reconciliation using a projection preserves unbiasedness.
- Also reconciliation via a projection is guaranteed to reduce the distance between the forecast and the target.

Geometry



Why reconciliation works

- The realised observation always lies on $\mathfrak{s}$.

Why reconciliation works

- The realised observation always lies on $\mathfrak{s}$.
- Orthogonal projections always get us 'closer' to all points in $\mathfrak{s}$ including the actual realisation.

Why reconciliation works

- The realised observation always lies on $\mathfrak{s}$.
- Orthogonal projections always get us ‘closer’ to all points in $\mathfrak{s}$ including the actual realisation.
- Ergo reconciliation reduces the error and not just in expectation.

Why reconciliation works

- The realised observation always lies on $\mathfrak{s}$.
- Orthogonal projections always get us 'closer' to all points in $\mathfrak{s}$ including the actual realisation.
- Ergo reconciliation reduces the error and not just in expectation.
- What about the MinT approach?

Finding a direction

- Consider the covariance matrix of $\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h}$.

Finding a direction

- Consider the covariance matrix of $\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h}$.
- This can be estimated using in-sample forecast errors.

$$\boldsymbol{W} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)'$$

Finding a direction

- Consider the covariance matrix of $\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h}$.
- This can be estimated using in-sample forecast errors.

$$\boldsymbol{W} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)'$$

- This provides information about the likely direction of an error.

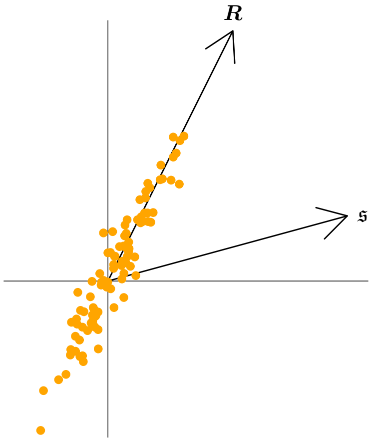
Finding a direction

- Consider the covariance matrix of $\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h}$.
- This can be estimated using in-sample forecast errors.

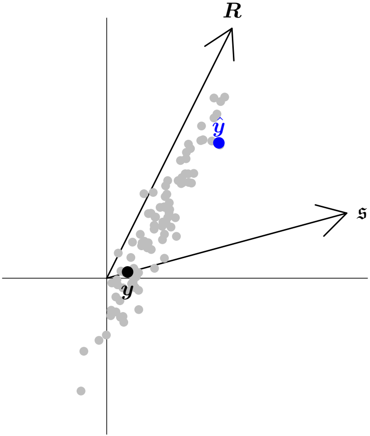
$$\boldsymbol{W} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)'$$

- This provides information about the likely direction of an error.
- Projecting along this direction is more likely to result in a reconciled forecast that is closer to the target.

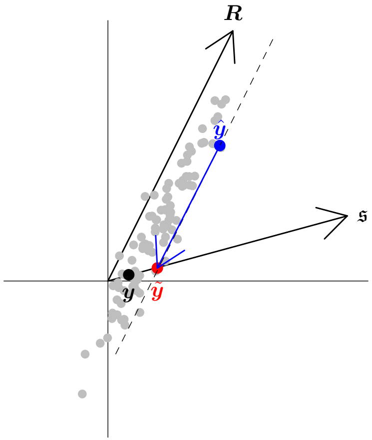
In-Sample errors



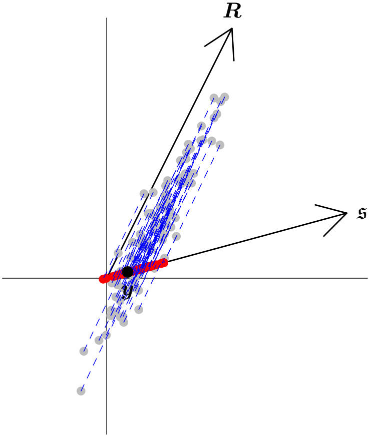
Base Forecast



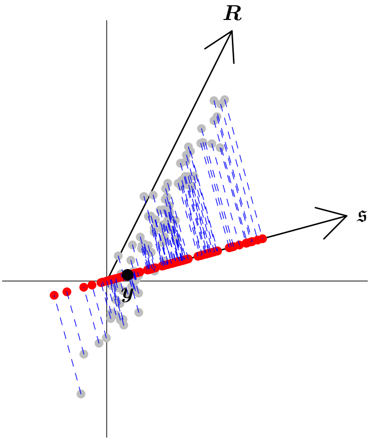
Geometry: Oblique Projection



Geometry: Oblique Projection



Geometry: Orthogonal Projection



Constructing Middle Level

- Forecast reconciliation can handle multi-level hierarchies.

Constructing Middle Level

- Forecast reconciliation can handle multi-level hierarchies.
- What happens if we *artificially* construct a middle level?

Constructing Middle Level

- Forecast reconciliation can handle multi-level hierarchies.
- What happens if we *artificially* construct a middle level?
- We did this by clustering causes of death.

Constructing Middle Level

- Forecast reconciliation can handle multi-level hierarchies.
- What happens if we *artificially* construct a middle level?
- We did this by clustering causes of death.
- This idea is novel in the forecast reconciliation literature.

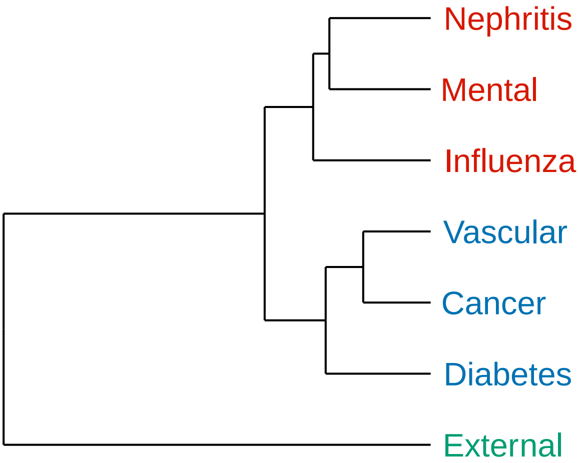
Cause clustering

- Consider a matrix $\boldsymbol{M}_j := \{m_{j,t}(x)\}_{x=1,\dots,85,t=1970,\dots,2015}$
- The distance between cause k and cause j is

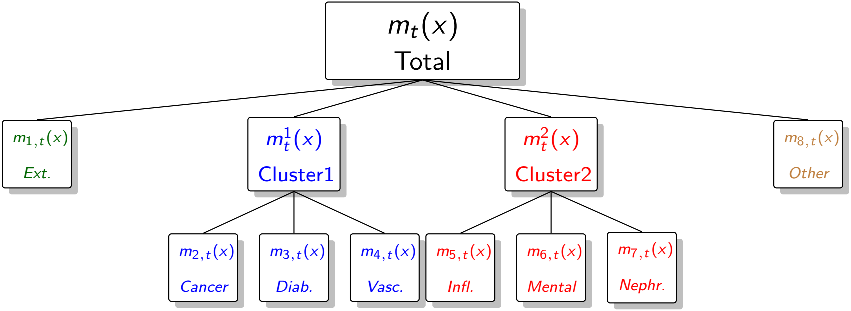
$$\delta_{j,k} = ||(1/s_j)\boldsymbol{M}_j - (1/s_k)\boldsymbol{M}_k||_F$$

- The s_j is a scale factor and $||.||_F$ is the Frobenius norm.
- Ward's method was used for clustering.
- Results were robust to different choices of scale, matrix norm and clustering method.

Dendrogam


Nephritis
Mental
Influenza
Vascular
Cancer
Diabetes
External

Three-level Hierarchy



Cause Elimination

- What should projections of total mortality look like if a cause (e.g. cancer) is eliminated?

Cause Elimination

- What should projections of total mortality look like if a cause (e.g. cancer) is eliminated?
- We would expect total mortality to be lower.

Cause Elimination

- What should projections of total mortality look like if a cause (e.g. cancer) is eliminated?
- We would expect total mortality to be lower.
- We would expect mortality from other causes to be higher.

Cause Elimination

- What should projections of total mortality look like if a cause (e.g. cancer) is eliminated?
- We would expect total mortality to be lower.
- We would expect mortality from other causes to be higher.
- With a subtle adjustment to forecast reconciliation this can be achieved.

Cause Elimination

Simply remove the column of $\boldsymbol{S}$ corresponding to cause. For 2-level hierarchy and first cause

$$\boldsymbol{S} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & & & & & & & \\ & & & \boldsymbol{I}_{8\times 8} & & & & \\ & & & & & & & \end{pmatrix}$$

Cause Elimination

Simply remove the column of $\boldsymbol{S}$ corresponding to cause. For 2-level hierarchy and first cause

$$\boldsymbol{S}_{-1} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & \boldsymbol{I}_{7\times 7} & & & \end{pmatrix}$$

Forecast Evaluation

- Use 1970-2000 as a training sample.
- Consider $h = 1$ through to $h = 15$ step ahead forecasts and compare to test sample 2001-2015.
- Evaluate forecasts using mean absolute percentage error (MAPE) averaged over 85 age groups.
- Consider 2-level and 3-level hierarchy.

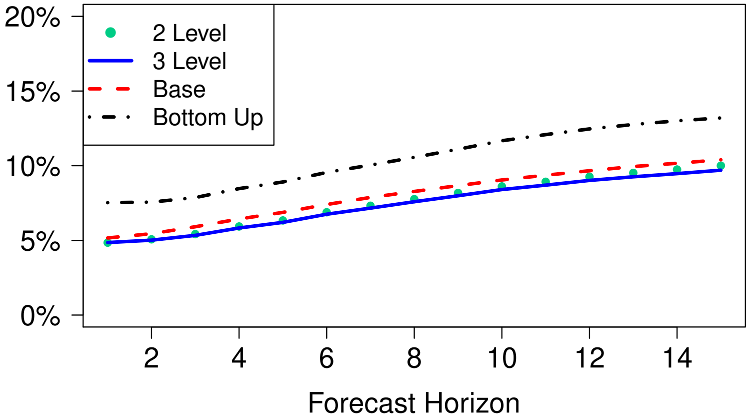
Results (h=15)

	2-level MinT	3-level MinT	Base	Bottom-up
Total	10.00%	9.70%	10.40%	13.20%
Cancer	17.98%	16.96%	20.06%	-
Diabetes	23.51%	21.78%	25.48%	-
External	19.11%	19.07%	20.62%	-
Influenza	37.17%	35.62%	38.36%	-
Mental	19.65%	20.02%	19.96%	-
Nephritis	59.60%	59.31%	61.31%	-
Vascular	16.37%	16.70%	17.37%	-
Other	26.64%	25.22%	34.46%	-

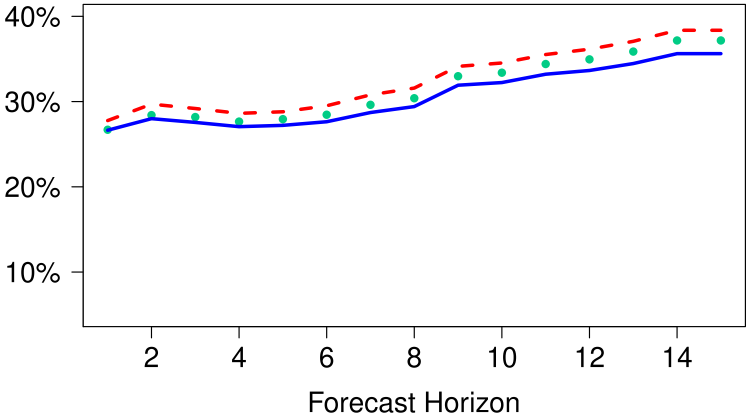
Summary

- The 2-level MinT always better than Base
- With one exception (mental), the 3-level MinT always better than Base.
- With two exceptions (mental and vascular), 3-level MinT better than 2-level MinT
- When reconciliation beats base it is always significant by Diebold Mariano (DM) test.
- Mixed results for DM tests between 2-level and 3-level.

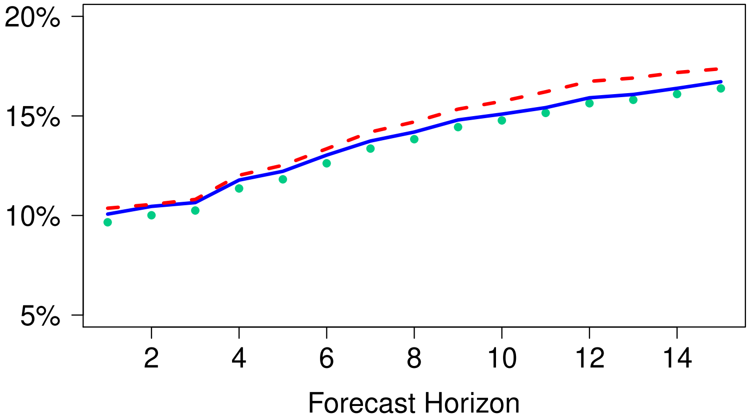
Total mortality MAPE



Influenza MAPE



Vascular MAPE



Cause Elimination

- We now use the modified $\boldsymbol{S}$ matrix to forecast mortality if cancer is eliminated.

Cause Elimination

- We now use the modified $\boldsymbol{S}$ matrix to forecast mortality if cancer is eliminated.
- Recall that what makes sense is

Cause Elimination

- We now use the modified $\boldsymbol{S}$ matrix to forecast mortality if cancer is eliminated.
- Recall that what makes sense is
  - Total mortality should be lower than the forecast of total mortality assuming cancer not eliminated. (benchmark).

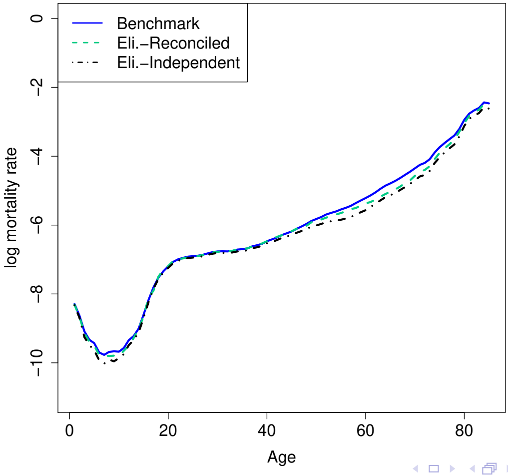
Cause Elimination

- We now use the modified $\boldsymbol{S}$ matrix to forecast mortality if cancer is eliminated.
- Recall that what makes sense is
  - Total mortality should be lower than the forecast of total mortality assuming cancer not eliminated. (benchmark).
  - Total mortality should be higher than the sum of forecasts of all non-cancer causes (independent).

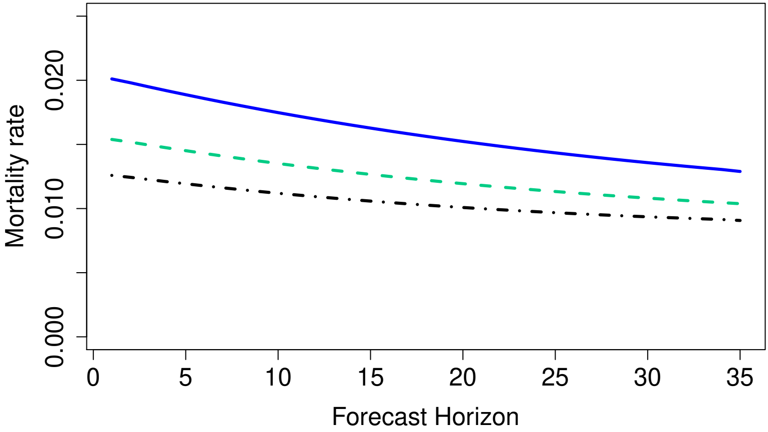
Cause Elimination

- We now use the modified $\boldsymbol{S}$ matrix to forecast mortality if cancer is eliminated.
- Recall that what makes sense is
  - Total mortality should be lower than the forecast of total mortality assuming cancer not eliminated. (benchmark).
  - Total mortality should be higher than the sum of forecasts of all non-cancer causes (independent).
- Results here are for the 2-level hierarchy.

Total Mortality 2050: Cancer eliminated



Total Mortality age 70: Cancer eliminated



Conclusions

- Forecast reconciliation works for mortality data.

Conclusions

- Forecast reconciliation works for mortality data.
- Using clustering to construct a middle level can improve reconciliation.

Conclusions

- Forecast reconciliation works for mortality data.
- Using clustering to construct a middle level can improve reconciliation.
- It can be modified for cause elimination.

Conclusions

- Forecast reconciliation works for mortality data.
- Using clustering to construct a middle level can improve reconciliation.
- It can be modified for cause elimination.
- Questions?