1 Temporal hierarchies

2 Probabilistic Hierarchical Forecasting

3 Probabilistic Gaussian Hierarchical Forecasting

4 Probabilistic Nonparametric Hierarchical Forecasting

5 Conclusions
Temporal hierarchies

Basic idea:

- Forecast series at each available frequency.
- Optimally reconcile forecasts within the same year.
Basic idea:

- Forecast series at each available frequency.
- Optimally reconcile forecasts within the same year.
Monthly series

- **Annual**
  - **Semi-Annual\textsubscript{1}**
    - **Q\textsubscript{1}**
      - **M\textsubscript{1}**
      - **M\textsubscript{2}**
      - **M\textsubscript{3}**
      - **M\textsubscript{4}**
      - **M\textsubscript{5}**
      - **M\textsubscript{6}**
    - **Q\textsubscript{2}**
      - **M\textsubscript{7}**
      - **M\textsubscript{8}**
      - **M\textsubscript{9}**
      - **M\textsubscript{10}**
      - **M\textsubscript{11}**
      - **M\textsubscript{12}**
  - **Semi-Annual\textsubscript{2}**
    - **Q\textsubscript{3}**
      - **M\textsubscript{7}**
      - **M\textsubscript{8}**
      - **M\textsubscript{9}**
      - **M\textsubscript{10}**
      - **M\textsubscript{11}**
      - **M\textsubscript{12}**
    - **Q\textsubscript{4}**
      - **M\textsubscript{7}**
      - **M\textsubscript{8}**
      - **M\textsubscript{9}**
      - **M\textsubscript{10}**
      - **M\textsubscript{11}**
      - **M\textsubscript{12}**

- **k = 2, 4, 12 nodes**
- **k = 3, 6, 12 nodes**
- **Why not k = 2, 3, 4, 6, 12 nodes?**
Monthly series

**Probabilistic Hierarchical Forecasting**

**Temporal hierarchies**

- $k = 2, 4, 12$ nodes
- $k = 3, 6, 12$ nodes
- Why not $k = 2, 3, 4, 6, 12$ nodes?
Monthly series

Annual

FourM₁
  BiM₁  BiM₂
  M₁  M₂  M₃  M₄

FourM₂
  BiM₃  BiM₄
  M₅  M₆  M₇  M₈

FourM₃
  BiM₅  BiM₆
  M₉  M₁₀  M₁₁  M₁₂

\( k = 2, 4, 12 \) nodes

\( k = 3, 6, 12 \) nodes

Why not \( k = 2, 3, 4, 6, 12 \) nodes?
Monthly data

\[
\begin{pmatrix}
A \\
SemiA_1 \\
SemiA_2 \\
FourM_1 \\
FourM_2 \\
FourM_3 \\
Q_1 \\
\vdots \\
Q_4 \\
BiM_1 \\
\vdots \\
BiM_6 \\
M_1 \\
\vdots \\
M_{12}
\end{pmatrix}
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
M_1 \\
M_2 \\
M_3 \\
M_4 \\
M_5 \\
M_6 \\
M_7 \\
M_8 \\
M_9 \\
M_{10} \\
M_{11} \\
M_{12}
\end{pmatrix}
\]

where \(S\) is the matrix of semi-asymmetric coefficients, and \(B_t\) is the vector of月初月份数 coefficients.
In general

For a time series $y_1, \ldots, y_T$, observed at frequency $m$, we generate aggregate series

$$y_j^{[k]} = \sum_{t=1+(j-1)k}^{jk} y_t, \quad \text{for } j = 1, \ldots, \left\lfloor \frac{T}{k} \right\rfloor$$

- $k \in F(m) = \{\text{factors of } m\}$.
- A single unique hierarchy is only possible when there are no coprime pairs in $F(m)$.
- $M_k = m/k$ is seasonal period of aggregated series.
In general

For a time series $y_1, \ldots, y_T$, observed at frequency $m$, we generate aggregate series

$$y_j^{[k]} = \sum_{t=1+(j-1)k}^{jk} y_t, \quad \text{for } j = 1, \ldots, \lfloor T/k \rfloor$$

- $k \in F(m) = \{\text{factors of } m\}$.
- A single unique hierarchy is only possible when there are no coprime pairs in $F(m)$.
- $M_k = m/k$ is seasonal period of aggregated series.
In general

For a time series $y_1, \ldots, y_T$, observed at frequency $m$, we generate aggregate series

$$y_j^{[k]} = \sum_{t=1+(j-1)k}^{jk} y_t, \quad \text{for } j = 1, \ldots, \lfloor T/k \rfloor$$

- $k \in F(m) = \{\text{factors of } m\}$.
- A single unique hierarchy is only possible when there are no coprime pairs in $F(m)$.
- $M_k = m/k$ is seasonal period of aggregated series.
UK Accidents and Emergency Demand

Temporal hierarchies

- Annual (k=52)
  - Forecast
  - 5100 5300 5500

- Semi-annual (k=26)
  - Forecast
  - 2500 2600 2700 2800 2900

- Quarterly (k=13)
  - Forecast

- Monthly (k=4)
  - Forecast
  - 360 380 400 420 440 460

- Bi-weekly (k=2)
  - Forecast
  - 180 200 210 220 230

- Weekly (k=1)
  - Forecast
  - 90 100 110

---

Base: - - - -
Reconciled: - - - -
# UK Accidents and Emergency Demand

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Type 1 Departments — Major A&amp;E</td>
</tr>
<tr>
<td>2</td>
<td>Type 2 Departments — Single Specialty</td>
</tr>
<tr>
<td>3</td>
<td>Type 3 Departments — Other A&amp;E/Minor Injury</td>
</tr>
<tr>
<td>4</td>
<td>Total Attendances</td>
</tr>
<tr>
<td>5</td>
<td>Type 1 Departments — Major A&amp;E &gt; 4 hrs</td>
</tr>
<tr>
<td>6</td>
<td>Type 2 Departments — Single Specialty &gt; 4 hrs</td>
</tr>
<tr>
<td>7</td>
<td>Type 3 Departments — Other A&amp;E/Minor Injury &gt; 4 hrs</td>
</tr>
<tr>
<td>8</td>
<td>Total Attendances &gt; 4 hrs</td>
</tr>
<tr>
<td>9</td>
<td>Emergency Admissions via Type 1 A&amp;E</td>
</tr>
<tr>
<td>10</td>
<td>Total Emergency Admissions via A&amp;E</td>
</tr>
<tr>
<td>11</td>
<td>Other Emergency Admissions (i.e., not via A&amp;E)</td>
</tr>
<tr>
<td>12</td>
<td>Total Emergency Admissions</td>
</tr>
<tr>
<td>13</td>
<td>Number of patients spending &gt; 4 hrs from decision to admission</td>
</tr>
</tbody>
</table>
**Minimum training set:** all data except the last year

- Base forecasts using `auto.arima()`.
- Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
UK Accidents and Emergency Demand

- **Minimum training set**: all data except the last year
- **Base forecasts using `auto.arima()`**.
- Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
Minimum training set: all data except the last year

Base forecasts using `auto.arima()`.

Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
UK Accidents and Emergency Demand

- **Minimum training set**: all data except the last year
- Base forecasts using `auto.arima()`.
- Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
UK Accidents and Emergency Demand

- **Minimum training set**: all data except the last year
- Base forecasts using `auto.arima()`.
- Mean Absolute Scaled Errors for 1, 4 and 13 weeks ahead using a rolling origin.

<table>
<thead>
<tr>
<th>Aggr. Level</th>
<th>$h$</th>
<th>Base</th>
<th>Reconciled</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>1</td>
<td>1.6</td>
<td>1.3</td>
<td>$-17.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>4</td>
<td>1.9</td>
<td>1.5</td>
<td>$-18.6%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>13</td>
<td>2.3</td>
<td>1.9</td>
<td>$-16.2%$</td>
</tr>
<tr>
<td>Weekly</td>
<td>1–52</td>
<td>2.0</td>
<td>1.9</td>
<td>$-5.0%$</td>
</tr>
<tr>
<td>Annual</td>
<td>1</td>
<td>3.4</td>
<td>1.9</td>
<td>$-42.9%$</td>
</tr>
</tbody>
</table>
thief package for R

**Temporal Hierarchical Forecasting**

Install from CRAN

```r
install.packages("thief")
```

Usage

```r
library(thief)
thief(y)
```
thief package for R

Temporal Hierarchical Forecasting

Install from CRAN
install.packages("thief")

Usage
library(thief)
thief(y)
Outline

1. Temporal hierarchies
2. Probabilistic Hierarchical Forecasting
3. Probabilistic Gaussian Hierarchical Forecasting
4. Probabilistic Nonparametric Hierarchical Forecasting
5. Conclusions
Coherent density forecasts

Definition: Coherence
Suppose $y_t \in \mathbb{R}^n$. $y_t$ is coherent if $y_t$ lies in an $m$-dimensional subspace of $\mathbb{R}^n$ spanned by the columns of the summing matrix $S$.

Definition: Coherent density forecasts
Any density $p(y_{t+h})$ is coherent if $p(y_{t+h}) = 0$ for all $y_{t+h}$ in the null space of $S$.

- Corollary: The probability distribution at each node is a convolution of the child distributions.
- Coherent point forecasts: $\tilde{y}_{T+h|T} = SP\hat{y}_{T+h}$.
- Coherent variance forecasts: $\tilde{\Sigma}_{T+h|T} = SP\hat{\Sigma}_{T+h|T}P'S'$.
Coherent density forecasts

**Definition: Coherence**
Suppose $y_t \in \mathbb{R}^n$. $y_t$ is coherent if $y_t$ lies in an $m$-dimensional subspace of $\mathbb{R}^n$ spanned by the columns of the summing matrix $S$.

**Definition: Coherent density forecasts**
Any density $p(y_{t+h})$ is coherent if $p(y_{t+h}) = 0$ for all $y_{t+h}$ in the null space of $S$.

- Corollary: The probability distribution at each node is a convolution of the child distributions.
- Coherent point forecasts: $\tilde{y}_{T+h|T} = SP\hat{y}_{T+h}$.
- Coherent variance forecasts: $\tilde{\Sigma}_{T+h|T} = SP\hat{\Sigma}_{T+h|T}P'S'$. 

Probabilistic Hierarchical Forecasting
Coherent density forecasts

**Definition: Coherence**
Suppose $\mathbf{y}_t \in \mathbb{R}^n$. $\mathbf{y}_t$ is coherent if $\mathbf{y}_t$ lies in an $m$-dimensional subspace of $\mathbb{R}^n$ spanned by the columns of the summing matrix $\mathbf{S}$.

**Definition: Coherent density forecasts**
Any density $p(\mathbf{y}_{t+h})$ is coherent if $p(\mathbf{y}_{t+h}) = 0$ for all $\mathbf{y}_{t+h}$ in the null space of $\mathbf{S}$.

- **Corollary**: The probability distribution at each node is a convolution of the child distributions.
- **Coherent point forecasts**: $\tilde{\mathbf{y}}_{T+h|T} = \mathbf{S}\hat{\mathbf{y}}_{T+h}$.
- **Coherent variance forecasts**: $\tilde{\Sigma}_{T+h|T} = \mathbf{S}\hat{\Sigma}_{T+h}\mathbf{T}\mathbf{P}'\mathbf{S}'$. 
Coherent Gaussian forecasts

\[ y_{T+h|T} \sim N(\tilde{y}_{T+h|T}, \tilde{\Sigma}_{T+h|T}) \]

Let \( L \) be the Energy Score (a proper scoring rule):

\[
L(\tilde{F}_{T+h|T}, y_{T+h}) = E\|\tilde{Y}_{T+h} - y_{T+h}\|^\alpha - \frac{1}{2}E\|\tilde{Y}_{T+h} - \tilde{Y}'_{T+h}\|^\alpha
\]

for \( \alpha \in (0, 2] \), where \( \tilde{Y}_{T+h} \) and \( \tilde{Y}'_{T+h} \) are independent rvs from \( \tilde{F}_{T+h|T} = N(\tilde{y}_{T+h|T}, \tilde{\Sigma}_{T+h|T}) \).

- There is no closed form expression for \( L(\tilde{F}_{T+h|T}, y_{T+h}) \) for \( \alpha \in (0, 2) \) under the Gaussian predictive distribution.
- When \( \alpha = 2 \), \( L(\tilde{F}_{T+h|T}, y_{T+h}) = E\|\tilde{Y}_{T+h|T} - y_{T+h}\|^2 \)
- This is equivalent to MinT solution.
Monte-Carlo simulation

Hierarchy 1: Case A

Probabilistic Hierarchical Forecasting
Probabilistic Gaussian Hierarchical Forecasting
Monte-Carlo simulation

Hierarchy 1: Case B

Probabilistic Hierarchical Forecasting
Probabilistic Gaussian Hierarchical Forecasting
Monte-Carlo simulation

Hierarchy 2

A  B  C  .....  T
Monte-Carlo simulation

- Bottom level series generated from univariate ARMA(1,1) processes.
- Contemporaneous errors randomly generated from multivariate Gaussian distribution with mean zero and correlation structures described before.
- Parameters for AR and MA components from uniform distribution, satisfying stationarity and invertibility conditions.

<table>
<thead>
<tr>
<th>Hierarchy</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchy 1: Case A</td>
<td>[0.4, 0.7]</td>
</tr>
<tr>
<td>Hierarchy 1: Case B</td>
<td>[0.4, 0.7]</td>
</tr>
<tr>
<td>Hierarchy 2</td>
<td>[0.3, 0.7]</td>
</tr>
</tbody>
</table>
Monte-Carlo simulation

- 501 observations generated for each series.
- Univariate ARIMA models fitted for first 500 observations and 1-step ahead base forecasts generated.
- Predictive means and variances obtained using different reconciliation methods.
- Process replicated 1000 times from same DGP.
Monte-Carlo simulation

- 501 observations generated for each series.
- Univariate ARIMA models fitted for first 500 observations and 1-step ahead base forecasts generated.
- Predictive means and variances obtained using different reconciliation methods.
- Process replicated 1000 times from same DGP.

<table>
<thead>
<tr>
<th>Reconciliation method</th>
<th>Average Energy Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hierarchy 1A</td>
</tr>
<tr>
<td>Base</td>
<td>9.26</td>
</tr>
<tr>
<td>Bottom up</td>
<td>9.19**</td>
</tr>
<tr>
<td>OLS</td>
<td>9.23**</td>
</tr>
<tr>
<td>MinT(Sample)</td>
<td>9.20*</td>
</tr>
<tr>
<td>MinT(Shrink)</td>
<td>9.19**</td>
</tr>
</tbody>
</table>
Monte-Carlo simulation

- 501 observations generated for each series.
- Univariate ARIMA models fitted for first 500 observations and 1-step ahead base forecasts generated.
- Predictive means and variances obtained using different reconciliation methods.
- Process replicated 1000 times from same DGP.

<table>
<thead>
<tr>
<th>Reconciliation method</th>
<th>Diebold-Mariano test: best pairwise method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hierarchy 1A</td>
</tr>
<tr>
<td>BU vs OLS</td>
<td></td>
</tr>
<tr>
<td>BU vs MinT(Sample)</td>
<td></td>
</tr>
<tr>
<td>BU vs MinT(Shrink)</td>
<td></td>
</tr>
<tr>
<td>OLS vs MinT(Sample)</td>
<td></td>
</tr>
<tr>
<td>OLS vs MinT(Shrink)</td>
<td></td>
</tr>
<tr>
<td>MinT(Shrink) vs MinT(Sample)</td>
<td></td>
</tr>
</tbody>
</table>
1 Temporal hierarchies

2 Probabilistic Hierarchical Forecasting

3 Probabilistic Gaussian Hierarchical Forecasting

4 Probabilistic Nonparametric Hierarchical Forecasting

5 Conclusions
Coherent nonparametric forecasts

1. Fit univariate models at each node using data up to time $T$.
2. Let $R = (e_1, \ldots, e_T)'$ be a matrix of residuals where $e_t = y_t - \hat{y}_t$.
3. Let $E^b = (e_{i+1}, \ldots, e_{i+h})'$ be a block bootstrap sample of size $h$ from $R$.
4. Generate $h$-step ahead sample paths from the fitted models incorporating $E^b$. Denote by $y^b_{T+h}$.
5. Project sample paths to coherent space: $\tilde{y}^b_{T+h} = SPy^b_{T+h}$ where $\tilde{y}^b_{T+h}$ denote coherent $h$-step ahead sample paths.
6. Repeat step 3–5 $J$ times.
Monte-Carlo simulation

- 501 observations generated for each series.
- Univariate ARIMA models fitted for first 500 observations and 1-step ahead base forecasts generated.
- 5000 1-step future paths constructed for 500 replications from same DGP.
Monte-Carlo simulation

- 501 observations generated for each series.
- Univariate ARIMA models fitted for first 500 observations and 1-step ahead base forecasts generated.
- 5000 1-step future paths constructed for 500 replications from same DGP.

<table>
<thead>
<tr>
<th>Reconciliation method</th>
<th>Average Energy Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hierarchy 1A</td>
</tr>
<tr>
<td>Base</td>
<td>14.54</td>
</tr>
<tr>
<td>Bottom up</td>
<td>13.87**</td>
</tr>
<tr>
<td>OLS</td>
<td>14.17**</td>
</tr>
<tr>
<td>MinT(Sample)</td>
<td>15.12</td>
</tr>
<tr>
<td>MinT(Shrink)</td>
<td>14.15**</td>
</tr>
</tbody>
</table>
Monte-Carlo simulation

- 501 observations generated for each series.
- Univariate ARIMA models fitted for first 500 observations and 1-step ahead base forecasts generated.
- 5000 1-step future paths constructed for 500 replications from same DGP.

### Diebold-Mariano test: best pairwise method

<table>
<thead>
<tr>
<th>Reconciliation method</th>
<th>Hierarchy 1A</th>
<th>Hierarchy 1B</th>
<th>Hierarchy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BU vs OLS</td>
<td>BU</td>
<td>BU</td>
<td>MinT(Sample)</td>
</tr>
<tr>
<td>BU vs MinT(Sample)</td>
<td>BU</td>
<td>BU</td>
<td>MinT(Shrink)</td>
</tr>
<tr>
<td>BU vs MinT(Shrink)</td>
<td>BU</td>
<td>BU</td>
<td>MinT(Shrink)</td>
</tr>
<tr>
<td>OLS vs MinT(Sample)</td>
<td>OLS</td>
<td>OLS</td>
<td>MinT(Shrink)</td>
</tr>
<tr>
<td>OLS vs MinT(Shrink)</td>
<td></td>
<td></td>
<td>MinT(Shrink)</td>
</tr>
<tr>
<td>MinT(Shrink) vs MinT(Sample)</td>
<td>MinT(Shrink)</td>
<td>MinT(Shrink)</td>
<td>MinT(Shrink)</td>
</tr>
</tbody>
</table>
Copula-based distributions of sums

Sklar’s theorem

For any continuous distribution $F$ with marginals $F_1, \ldots, F_d$, there exists a unique “copula” function $C : [0, 1]^d \rightarrow [0, 1]$ such that

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

Empirical copula

If $x_k^i \sim F_i$ and $u_k = (u_k^1, \ldots, u_k^d) \sim C$, then

$$\hat{F}_i(x) = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\{x_k^i \leq x\}$$

and empirical copula is

$$C(u) = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\left\{ \frac{rk(u_k^1)}{K} \leq u_1, \ldots, \frac{rk(u_k^d)}{K} \leq u_d \right\}$$

$$\hat{F}(x_1, \ldots, x_d) = \hat{C}(\hat{F}_1(x_1), \text{dots}, \hat{F}_d(x_d))$$
We can efficiently compute $\hat{F}$ using permutations.

We can compute copulas recursively in the tree structure, rather than find the joint distribution or the entire hierarchy.
We can efficiently compute $\hat{F}$ using permutations.

We can compute copulas recursively in the tree structure, rather than find the joint distribution or the entire hierarchy.
Coherent nonparametric forecasts

1. Forecast at every node using whatever method you choose to get marginal forecast distributions for each node.
2. Apply MinT to reconcile the means of the forecast distributions.
3. Simulate from the forecast distributions at each bottom level node.
4. Compute empirical copulas for each parent+children group to obtain coherent forecast distributions at the next level up.
5. Repeat working up the tree.
1578 households from Great Britain.


Training data: to 30 April 2010.

Forecasting 48 periods ahead (one day).

Geographical hierarchy with five levels.
Application: Smart Meter Data

- 3 groups at level 2
- 11 groups at level 3
- 40 groups at level 4.
- 1578 households at bottom level.
Application: Smart Meter Data
One week of demand at different levels of aggregation.
Forecast individual series using Taylor’s double-seasonal Holt-Winters’ method.

Kernel density estimation by 48 half-hours and for 3 different day types (weekday, Saturday, Sunday) for density forecasts.

KDE use decay parameter to “forget” the past.

Decay and bandwidth chosen to minimize CRPS
Application: Smart Meter Data

Top aggregated series

Individual smart meter

Time of day

Consumption (kWh)

00:00 03:30 07:30 11:30 15:30 19:30 23:30

CRPS skill relative to base forecasts.

Probabilistic Hierarchical Forecasting

Probabilistic Nonparametric Hierarchical Forecasting
MinT (Shink) not only optimally reconciles point forecasts, it is also optimal for probabilistic Gaussian forecasts.

MinT (Shrink) can also be used to generate coherent future sample paths.

Combining MinT (Shrink) with empirical copulas allows for efficient nonparametric coherent probabilistic forecasting.


Plus contributions from: Anastasios Panagiotelis, George Athanasopoulos, Puwasala Gamakumara.