Making Regional Forecasts Add Up

Tim van Erven\textsuperscript{1,2}

Joint work with: Jairo Cugliari\textsuperscript{2}
Regional Electricity Consumption

We want to forecast:

1. Electricity consumption in K regions
2. The total consumption of those regions

(A “region” could be any group of customers.
- E.g. customers with the same contract.)
Measuring Performance

- **Real consumptions**
  - Regions: \( y = (y_1, \ldots, y_K) \)
  - Total: \( y_\star = y_1 + \ldots + y_K \)

- **Predictions**
  - Regions: \( \hat{y} = (\hat{y}_1, \ldots, \hat{y}_K) \)
  - Total: \( \hat{y}_\star \)

- **Weighted squared loss**

\[
\ell(y, (\hat{y}, \hat{y}_\star)) = \sum_{k=1}^{K} a_k (y_k - \hat{y}_k)^2 + a_\star (y_\star - \hat{y}_\star)^2
\]
Measuring Performance

- Real consumptions
  - Regions: $y = (y_1, \ldots, y_K)$
  - Total: $y_* = y_1 + \ldots + y_K$
- Predictions
  - Regions: $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_K)$
  - Total: $\hat{y}_*$
- Weighted squared loss

$$
\ell(y, (\hat{y}, \hat{y}_*)) = \sum_{k=1}^{K} a_k (y_k - \hat{y}_k)^2 + a_*(y_* - \hat{y}_*)^2
$$

Weights represent electricity network configurations
For example:

$$
a_k = 1 \text{ for all } k
$$

$$
a_* = K
$$
The Operational Constraint

Prediction for the total
= sum of predictions for the regions

\[ \hat{y}_* = \hat{y}_1 + \ldots + \hat{y}_K \]

Imposed, for example, in the Global Energy Forecasting Competition 2012 on Kaggle.com
The Forecasters' Rebellion

- Constraint: $\hat{y}_* = \hat{y}_1 + \ldots + \hat{y}_K$
  - Maybe the total is easier to predict than the regions...
  - What if we have a better predictor for the total consumption?

We don't want this constraint!
A Peace Treaty Allowing a Separation of Concerns

- Forecasters produce *ideal* predictions
  \[ \bar{y} = (\bar{y}_1, \ldots, \bar{y}_K, \bar{y}_*) \]
- Map to predictions that satisfy the constraint
  - Regions: \( \hat{y} = (\hat{y}_1, \ldots, \hat{y}_K) \)
  - Total: \( \hat{y}_* = \hat{y}_1 + \ldots + \hat{y}_K \)
Related Work

- Let $z = \bar{y}_* - \sum_k \bar{y}_k$ measure how much we violate the constraint
- HTS [Hyndman et al, 2011]: $\hat{y}_k = \bar{y}_k + \frac{1}{K + 1} z$

- Disadvantages:
  - Designed under probabilistic assumptions about distribution of predictions and consumptions
  - Does not take into account weights $a_k$ of the regions and of the total $a_*$. 
Game-theoretically Optimal Predictions (GTOP)

• Difference between ideal and real loss:

\[ \ell(y, \hat{y}) - \ell(y, \bar{y}) \quad (1) \]

where \( \hat{y} = (\hat{y}_1, \ldots, \hat{y}_K, \sum_k \hat{y}_k) \) satisfies the constraint

• Idea: model as a zero-sum game
  - We first choose our predictions \( \hat{y} \)
  - Then an opponent chooses \( y \) to make (1) as large as possible
Game-theoretically Optimal Predictions (GTOP)

- Difference between ideal and real loss:
  \[ \ell(y, \hat{y}) - \ell(y, \bar{y}) \quad (1) \]
  where \( \hat{y} = (\hat{y}_1, \ldots, \hat{y}_K, \sum_k \hat{y}_k) \) satisfies the constraint

- Idea: model as a zero-sum game
  - We first choose our predictions \( \hat{y} \)
  - Then an opponent chooses \( y \) to make (1) as large as possible

- No probabilistic assumptions!
Game-theoretically Optimal Predictions (GTOP)

- The optimal move chooses $\hat{y}$ to achieve
  \[
  \min_{\hat{y}} \max_y \{ \ell(y, \hat{y}) - \ell(y, \bar{y}) \}
  \]

- Assume confidence bands:
  \[
  y_k \in [\bar{y}_k - B_k, \bar{y}_k + B_k]
  \]
Game-theoretically Optimal Predictions (GTOP)

- The optimal move chooses \( \hat{y} \) to achieve

\[
\min_{\hat{y}} \max_y \{ \ell(y, \hat{y}) - \ell(y, \bar{y}) \}
\]

- Assume confidence bands:

\[
y_k \in [\bar{y}_k - B_k, \bar{y}_k + B_k]
\]

Example: If \( B_1 = \ldots = B_K = B \) and \( a_1 = \ldots = a_K = a_* \)

\[
\hat{y}_k = \bar{y}_k + \left[ \frac{1}{K+1} \bar{z} \right] B
\]

where \( z = \bar{y}_* - \sum_k \bar{y}_k \)

\[
[x]_B = \min \{ B, \max \{-B, x\} \}
\]
Non-uniform Weights: L2-projection

- If confidence bands $B_k$ are sufficiently large:

\[ \hat{y}_k = \bar{y}_k + \frac{1/a_k}{1/a_* + \sum_{k'} 1/a_{k'}} z \]

- This is the L2-projection
  - of $\bar{y}$ unto the hyperplane of predictions satisfying summation constraint,
  - with axes rescaled to take into account the region weights $a_k$, $a_*$

- In simulations we see that GTOP exactly predicts like this already for very small $B_k$. 
General Computation

- In general no closed-form solution for GTOP, but can rewrite as LASSO optimization problem.
- Size of problem depends on number of regions $K$
- Standard software to quickly compute LASSO solutions can deal with very large problems; $K$ is typically much smaller
Experiments with Simulated Data

- **K = 2 regions:**
  \[
  y_1 = 1 + 5x + \sigma \xi + \tau \zeta_1 \\
  y_2 = 1 + 5x - \sigma \xi + \tau \zeta_2
  \]

- Noise r.v. \(\xi, \zeta_1, \zeta_2\) are uniform on \([-1, 1]\)
- Parameters \(\sigma, \tau\) control amount of noise

- **Train set:** \(x \in \left\{ \frac{1}{100}, \frac{2}{100}, \ldots, 1 \right\}\)
- **Test set:** \(x \in \left\{ 1 + \frac{1}{100}, 1 + \frac{2}{100}, \ldots, 2 \right\}\)
Ideal Predictions

- For the regions \((\bar{y}_1, \bar{y}_2)\):
  - Fit linear function \(y = \beta_0 + \beta_1 x\) to the data
  - Use LASSO to estimate \(\beta_0, \beta_1\) per region
- For the total \((\bar{y}_*)\), \(\bar{y}_1 + \bar{y}_2\) already very good predictor. How do we do better???
Ideal Predictions

- For the regions \((\bar{y}_1, \bar{y}_2)\):
  - Fit linear function \(y = \beta_0 + \beta_1 x\) to the data
  - Use LASSO to estimate \(\beta_0, \beta_1\) per region

- For the total \((\bar{y}_*), \bar{y}_1 + \bar{y}_2\) already very good predictor. How do we do better???
  - 1. Fit \(y = \beta_0 + \beta_1 x + \beta_2 \bar{y}_1 + \beta_3 \bar{y}_2\) with LASSO
  - 2. Regularize by
    \[|\beta_0| + |\beta_1| + |\beta_2 - 1| + |\beta_3 - 1|\]
  - Behaves like \(\bar{y}_1 + \bar{y}_2\) unless data say otherwise
Results

- GTOP calibration
  - $B_k$ are set to maximum absolute value of residuals on train set
- Loss HTS – loss GTOP summed over test set
Summary

- We want to forecast:
  - Electricity consumption in K regions
  - The total consumption of those regions
- Unpleasant operational constraint:
  - prediction for the total = sum of regional predictions
- Approach:
  - Ignore the constraint to get ideal predictions
  - Use GTOP to adjust ideal predictions to satisfy the constraint
Experiment with EDF data

- The data
  - K = 17 groups of customers
  - Half-hourly energy consumption records
  - Train set: 1 Jan 2004 to 31 Dec 2007
  - Test set: 1 Dec 2008 to 31 Dec 2009

- The model (presented yesterday by Jairo)
  - Non-parametric functional model
  - Based on matching similar contexts in previous observations
Preliminary Results

- GTOP calibration
  - $B_k$ are set heuristically as $0.01 \times y_k$

- Preliminary results
  - Ideal loss of $\bar{y}$ vs GTOP loss
  - Desired outcome: GTOP should not be much worse than $\bar{y}$
  - GTOP actually reduces the mean loss by 2.5% compared to $\bar{y}$!