Probabilistic Forecast Reconciliation

Puwasala Gamakumara, Anastasios Panagiotelis, George Athanasopoulos and Rob Hyndman

November 26, 2018

Motivating Examples

- Multiple time series $\rightarrow$ some series are aggregates of others.

Motivating Examples

- Multiple time series $\rightarrow$ some series are aggregates of others.
- Gross Domestic Product

Motivating Examples

- Multiple time series $\rightarrow$ some series are aggregates of others.
- Gross Domestic Product
  - An aggregate of consumption, investment, government spending and trade balance
  - Further breakdown, e.g. consumption of food, rent, etc.

Motivating Examples

- Multiple time series $\rightarrow$ some series are aggregates of others.
- Gross Domestic Product
  - An aggregate of consumption, investment, government spending and trade balance
  - Further breakdown, e.g. consumption of food, rent, etc.
- Wind power
  - Forecasts required at a daily and hourly resolution.
  - Daily series is aggregate of 24 hourly series.

Motivating Examples

- Multiple time series $\rightarrow$ some series are aggregates of others.
- Gross Domestic Product
  - An aggregate of consumption, investment, government spending and trade balance
  - Further breakdown, e.g. consumption of food, rent, etc.
- Wind power
  - Forecasts required at a daily and hourly resolution.
  - Daily series is aggregate of 24 hourly series.
- Potentially need forecasts of all time series.

Incoherent Forecasts

- Potential approaches

Incoherent Forecasts

- Potential approaches
  - Multivariate models

Incoherent Forecasts

- Potential approaches
  - Multivariate models
  - Univariate models

Incoherent Forecasts

- Potential approaches
  - Multivariate models
  - Univariate models
  - Judgemental forecasts

Incoherent Forecasts

- Potential approaches
  - Multivariate models
  - Univariate models
  - Judgemental forecasts
- Forecasts do not respect aggregation structure

Incoherent Forecasts

- Potential approaches
  - Multivariate models
  - Univariate models
  - Judgemental forecasts
- Forecasts do not respect aggregation structure (**Incoherent**)

Incoherent Forecasts

- Potential approaches
  - Multivariate models
  - Univariate models
  - Judgemental forecasts
- Forecasts do not respect aggregation structure (**Incoherent**)
- Outcome *does* respect aggregation structure

Incoherent Forecasts

- Potential approaches
  - Multivariate models
  - Univariate models
  - Judgemental forecasts
- Forecasts do not respect aggregation structure (**Incoherent**)
- Outcome *does* respect aggregation structure (**Coherent**)

Incoherent Forecasts

- Potential approaches
  - Multivariate models
  - Univariate models
  - Judgemental forecasts
- Forecasts do not respect aggregation structure (**Incoherent**)
- Outcome *does* respect aggregation structure (**Coherent**)
- Motivation is aggregation but can be generalised to any linear constraints.

Reconciliation

- Begin with a vector of *base* forecasts that are incoherent.

Reconciliation

- Begin with a vector of *base* forecasts that are incoherent.
- Adjust these **ex post** to make them coherent.

Reconciliation

- Begin with a vector of *base* forecasts that are incoherent.
- Adjust these **ex post** to make them coherent.
- There are good solutions for point forecasting that:

Reconciliation

- Begin with a vector of *base* forecasts that are incoherent.
- Adjust these **ex post** to make them coherent.
- There are good solutions for point forecasting that:
  - Guarantee coherent forecasts.
  - Improve forecast accuracy overall.

Reconciliation

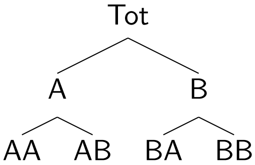
- Begin with a vector of *base* forecasts that are incoherent.
- Adjust these **ex post** to make them coherent.
- There are good solutions for point forecasting that:
  - Guarantee coherent forecasts.
  - Improve forecast accuracy overall.
- Generalisation to probabilistic forecasts is our contribution.

Reconciliation

- Begin with a vector of *base* forecasts that are incoherent.
- Adjust these **ex post** to make them coherent.
- There are good solutions for point forecasting that:
  - Guarantee coherent forecasts.
  - Improve forecast accuracy overall.
- Generalisation to probabilistic forecasts is our contribution.
- Getting there necessitates a rethink of the existing point forecasting literature.

A simple hierarchy

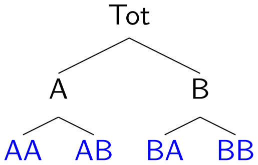
Consider a hierarchy given by



- 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 m be the number of bottom level series and $\boldsymbol{b}_t$ be an m-vector of the bottom level series.

The $\boldsymbol{S}$ matrix

Coherence holds when

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

The $n \times m$ 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:

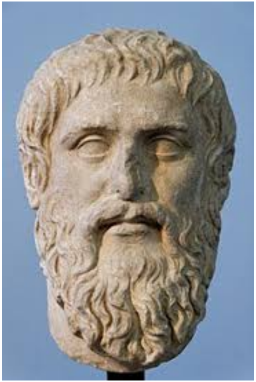
$$\tilde{\boldsymbol{y}}_{T+h} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_{T+h}$$

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, 2018) use a $\boldsymbol{W}$ that is an estimate of the *in-sample* forecast error covariance matrix.



ΑΓΕΩΜΤΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ

Those without knowledge of geometry may not enter.

Coherent Subspace

Definition

The **coherent subspace** is the m-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 a different combination of m **basis series** could be used (e.g. top and $m-1$ bottom). 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 $g(.)$ be a function $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

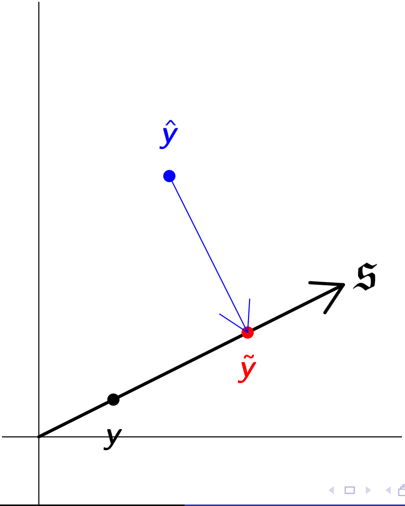
Definition

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

$$\tilde{\boldsymbol{y}} = \boldsymbol{S}g(\hat{\boldsymbol{y}})$$

when $g(.)$ is linear it is easier to write $\tilde{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}$

Geometry



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.

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

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.

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

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

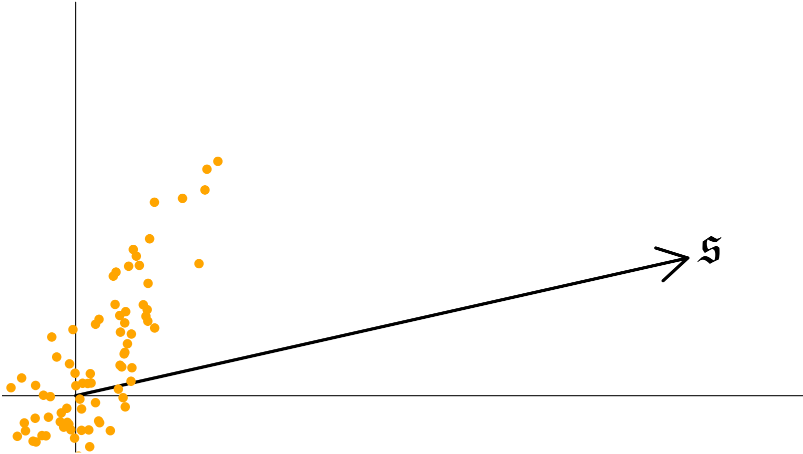
Finding a direction

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

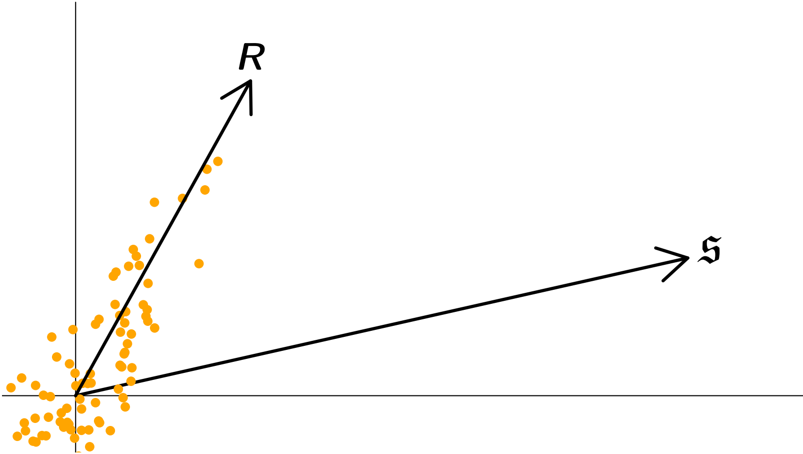
$$\boldsymbol{W} = \sum_{t=1}^{T} (\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)'$$

- 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.

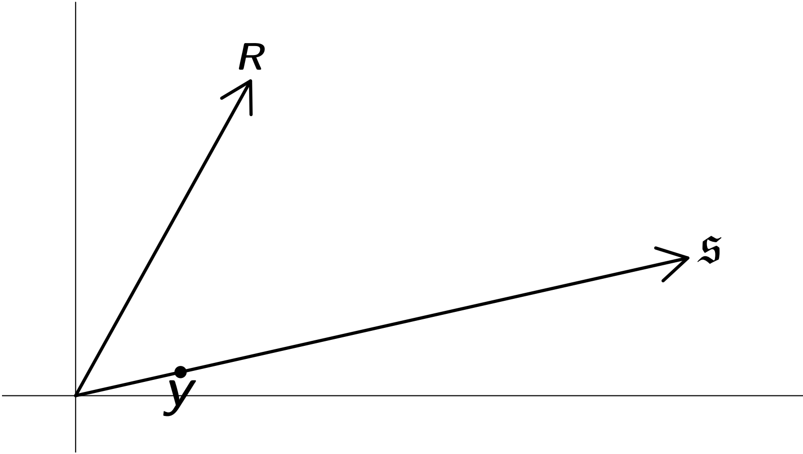
In-Sample errors



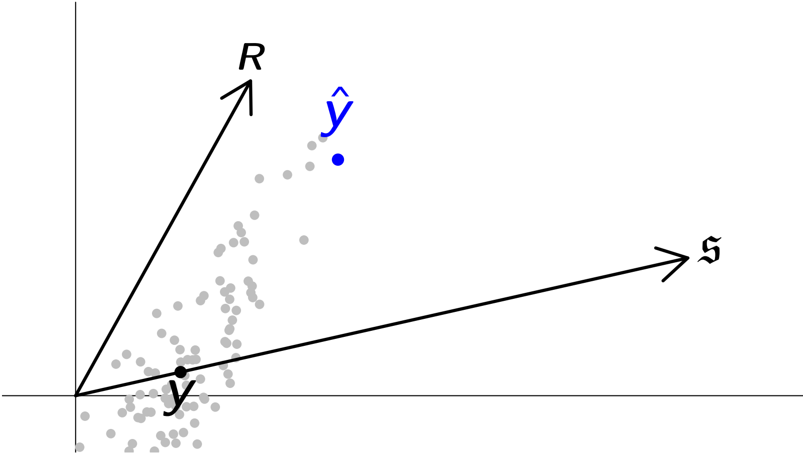
Most likely direction



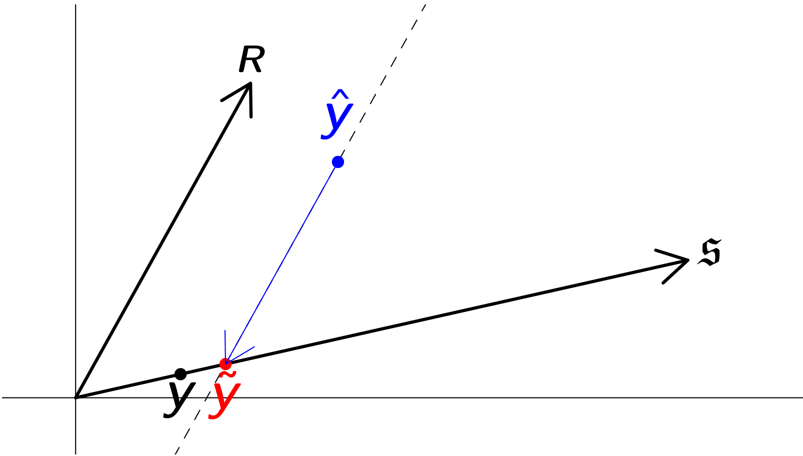
Geometry: Oblique Projection



Geometry: Oblique Projection



Geometry: Oblique Projection



Is this overkill?

- Is this geometric interpretation really necessary?

Is this overkill?

- Is this geometric interpretation really necessary?
- When we generalise to probabilistic forecasts the regression interpretation does not really fit.

Is this overkill?

- Is this geometric interpretation really necessary?
- When we generalise to probabilistic forecasts the regression interpretation does not really fit.
- Geometric ideas can easily be generalised.

Coherent Probabilistic Forecast

Let $(\mathbb{R}^m, \mathcal{F}_{\mathbb{R}^m}, \nu)$ and $(\mathfrak{s}, \mathcal{F}_{\mathfrak{s}}, \mu)$ be probability triples on m-dimensional space and the coherent subspace respectively.

Definition

The probability measure μ is coherent if

$$\nu(\mathcal{B}) = \mu(s(\mathcal{B})) \quad \forall \mathcal{B} \in \mathcal{F}_{\mathbb{R}_m}$$

where $s(\mathcal{B})$ is the image of $\mathcal{B}$ under premultiplication by $\boldsymbol{S}$

Reconciled Probabilistic Forecast

Let $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a function. Then

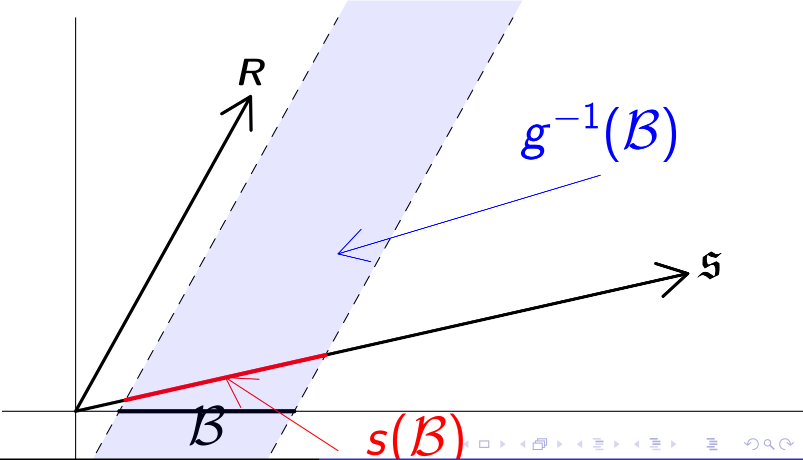
Definition

The probability triple $(\mathfrak{s}, \mathcal{F}_{\mathfrak{s}}, \tilde{\nu})$ reconciles the probability triple $(\mathbb{R}^n, \mathcal{F}_{\mathbb{R}^n}, \hat{\nu})$ with with respect to g iff

$$\tilde{\nu}(s(\mathcal{B})) = \nu(\mathcal{B}) = \hat{\nu}(g^{-1}(\mathcal{B})) \quad \forall \mathcal{B} \in \mathcal{F}_{\mathbb{R}_m}$$

where g^{-1} is the pre-image of g.

Geometry



Analytically

If we have an unreconciled density the reconciled density can be obtained by linear transformations and marginalisation.

$$
\begin{aligned}
\Pr(\tilde{\boldsymbol{b}} \in \mathcal{B}) &= \Pr(\hat{\boldsymbol{y}} \in g^{-1}(\mathcal{B})) \\
&= \int\limits_{g^{-1}(\mathcal{B})} f(\hat{\boldsymbol{y}}) d\hat{\boldsymbol{y}} \\
&= \int\limits_{\mathcal{B}} \int f(\boldsymbol{S}\tilde{\boldsymbol{b}} + \boldsymbol{R}\tilde{\boldsymbol{a}}) \left| \begin{pmatrix} \boldsymbol{S} & \boldsymbol{R} \end{pmatrix} \right| d\tilde{\boldsymbol{a}} d\tilde{\boldsymbol{b}}
\end{aligned}
$$

Elliptical distributions

Consider case where the base and true predictive distributions are elliptical.

Theorem

There exists a matrix $\boldsymbol{G}$ such that the true predictive distribution can be recovered by linear reconciliation.

This follows from the closure property of elliptical distributions under affine transformations and marginalisation.

Multivariate Scores

- Scoring rules can be used to evaluate probabilistic forecasts

Multivariate Scores

- Scoring rules can be used to evaluate probabilistic forecasts
  - Log Score
  - Energy Score
  - Variogram score

Multivariate Scores

- Scoring rules can be used to evaluate probabilistic forecasts
  - Log Score
  - Energy Score
  - Variogram score
- We may want to compare
  - Coherent v Incoherent
  - Coherent v Coherent

Coherent v Incoherent

When using log score

Theorem

Let $f(\boldsymbol{y})$ be the true predictive density (on $\mathfrak{s}$) and LS be the (negatively-oriented) log score. Then there exists an unreconciled density $\hat{f}(\boldsymbol{y})$ on $\mathbb{R}^n$ such that

$$E_{\boldsymbol{y}}\left[LS(\hat{f},\boldsymbol{y})\right] < E_{\boldsymbol{y}}\left[LS(f,\boldsymbol{y})\right]$$

The log score is not proper **in this context**.

Simulations

- We have run lots of simulations.

Simulations

- We have run lots of simulations.
- The main takeaway messages are:

Simulations

- We have run lots of simulations.
- The main takeaway messages are:
  - Reconciliation is better than no reconciliation.

Simulations

- We have run lots of simulations.
- The main takeaway messages are:
  - Reconciliation is better than no reconciliation.
  - Bottom up does not do well.

Simulations

- We have run lots of simulations.
- The main takeaway messages are:
  - Reconciliation is better than no reconciliation.
  - Bottom up does not do well.
  - OLS (an orthogonal projection) does reasonably well.

Simulations

- We have run lots of simulations.
- The main takeaway messages are:
  - Reconciliation is better than no reconciliation.
  - Bottom up does not do well.
  - OLS (an orthogonal projection) does reasonably well.
  - MinT (an oblique projection) does best.

Looking ahead

- The optimal feasible reconciliation method remains an open question even for elliptical distributions.

Looking ahead

- The optimal feasible reconciliation method remains an open question even for elliptical distributions.
  - It is likely to depend on the specific score used.

Looking ahead

- The optimal feasible reconciliation method remains an open question even for elliptical distributions.
  - It is likely to depend on the specific score used.
- How should probabilistic reconciliation work for non-elliptical distributions.

Looking ahead

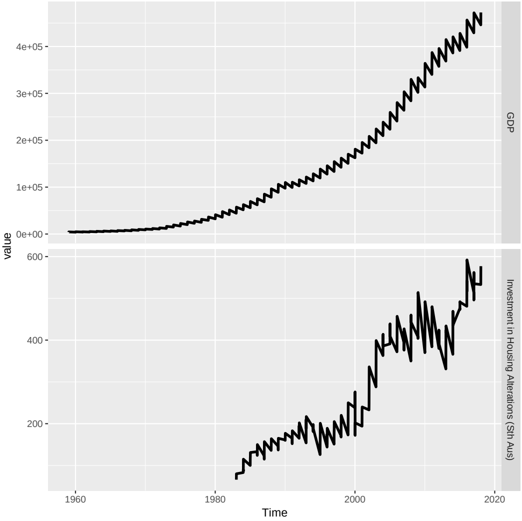
- The optimal feasible reconciliation method remains an open question even for elliptical distributions.
  - It is likely to depend on the specific score used.
- How should probabilistic reconciliation work for non-elliptical distributions.
- Are non-linear reconciliation methods worthwhile?

Thank You!

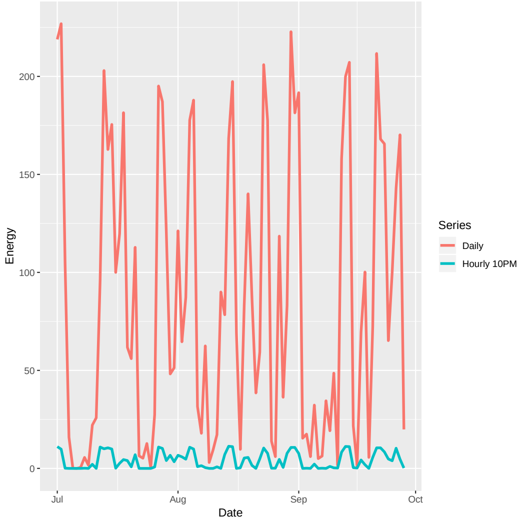
Thank You!

Questions?

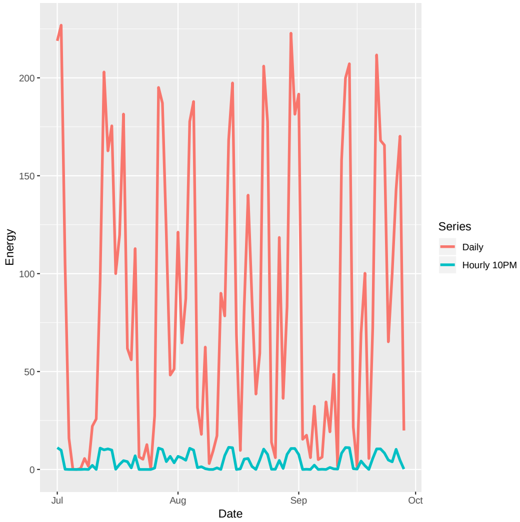
GDP


4e+05
3e+05
2e+05
1e+05
0e+00
600
400
200
value
GDP
Investment in Housing Alterations (Sth Aus)
1960
1980
2000
2020
Time

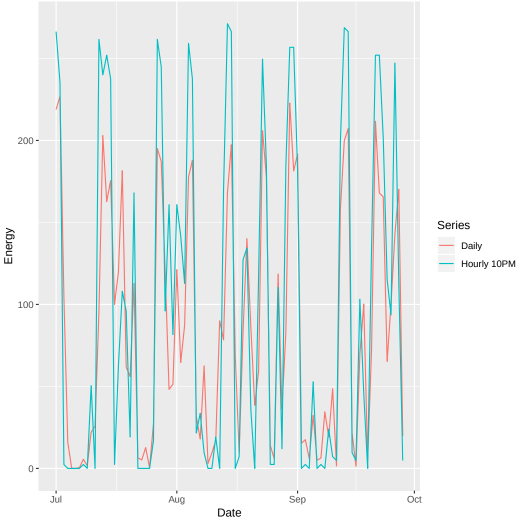
Aeolos Wind Farm



Aeolos Wind Farm



Aeolos Wind Farm



What we do NOT do

- Use disaggregate series as predictors (although we can).
- All information is contained in the most disaggregate series.
- In principle using the correct multivariate model for the most disaggregate series and aggregating them should work.
- Disaggregate series are:
    - Very noisy
    - High-dimensional
    - Prone to model misspecification

Macro Example

Improvement in RMSE of GDP forecasts over seasonal random walk using ARIMA models both without and with reconciliation.

Method	h=1	h=2
Base	56.1496	33.8093
BU	54.1873	36.1718
OLS	56.6164	34.6302
WLS	56.6819	36.0987
MinT Sam.	56.2204	31.6136
MinT Shr.	57.7249	36.8764

Simulation Results

Hierarchy from earlier bottom series are ARIMA models. Training sample of 500, one-step ahead forecasts, 1000 replications.

Forecasting	Energy score	Variogram score	Log score
MinT(Sample)	**10.01**	**8.41**	**11.29**
OLS	10.53	8.86	11.54
Bottom-up	12.35	9.22	12.05
Incoherent	11.12	9.53	