Probabilistic Forecasts for Hierarchical Time Series

Candidate:
Puwasala Gamakumara

Supervisors:
Rob J. Hyndman, George Athanasopoulos,
Anastasios Panagiotelis

Department of Econometrics and Business Statistics

Monash University

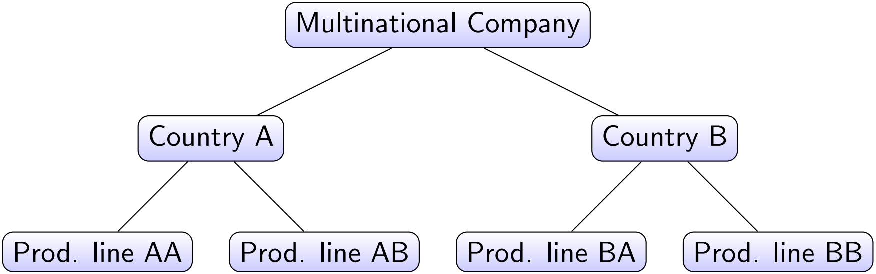
May 23, 2019

Overview

Project 1: Hierarchical Forecast Reconciliation in Geometric View

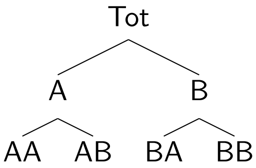
Project 1: *Motivation and Objective*

- **Example:**



- **Hierarchical time series:** A collection of multiple time series that has an inherent aggregation structure
- Forecasts should add up. We call these *coherent*
- **Objective:** Defining coherency and reconciliation of point forecasts in terms of geometric concepts.

Project 1: *Notations and Preliminaries*



$$\boldsymbol{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$\boldsymbol{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

$$n = 7$$

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

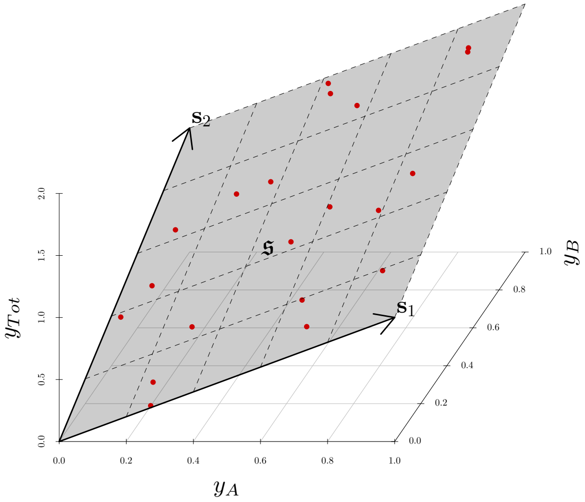
- Due to the aggregation nature of the hierarchy we have,

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

Coherent subspace

The m-dimensional linear subspace $\mathfrak{s} \subset \mathbb{R}^n$ that is spanned by the columns of $\boldsymbol{S}$, i.e. $\mathfrak{s} = \text{span}(\boldsymbol{S})$, is defined as the *coherent space*.

Project 1: *Coherent forecasts*



Project 1: *Point forecast reconciliation*

- Let $\hat{\boldsymbol{y}} \in \mathbb{R}^n$ be an incoherent forecast and $g(.)$ be a function $g : \mathbb{R}^n \to \mathbb{R}^m$.

Definition

Point forecasts $\tilde{\boldsymbol{y}}$ are reconciled with respect to $g(.)$ iff

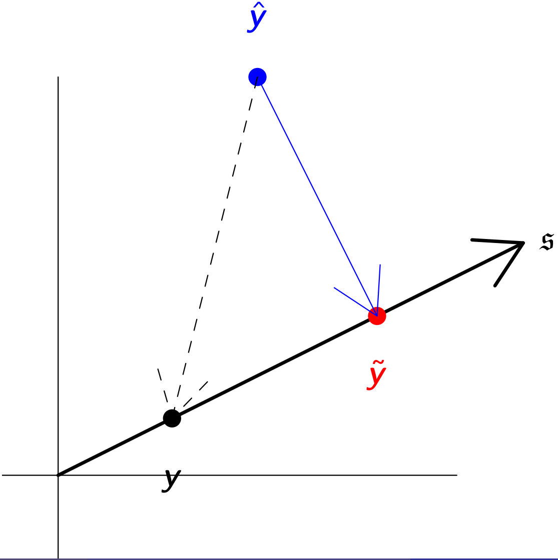
$$\tilde{\boldsymbol{y}} = \boldsymbol{S}g(\hat{\boldsymbol{y}})$$

- If $g(.)$ is linear,

$$\tilde{\boldsymbol{y}}_{T+h} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h}$$

Project 1: *Point forecast reconciliation - OLS*

$\boldsymbol{SG} = \boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}\boldsymbol{S}'$ (Hyndman et al., 2011)



Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, 2018)
- **Geometry**
  - 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 reconciled forecasts that are closer to the target.

Project 1: *Point forecast reconciliation - MinT*

$$\boldsymbol{SG} = \boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{W}_{T+h}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}_{T+h}^{-1}$$

Project 1: *Why projections?*

- Projections preserve unbiasedness
- What if the base forecasts are biased?
- Can we bias correct and proceed with projections?
  We investigate this further

Project 2: Probabilistic Forecasts for Hierarchical Time Series

Part 1: Definitions and A Parametric Approach

Project 2: *Motivation and Objectives*

- Extending "reconciliation" into a probabilistic framework

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure
- Existing literature
  - (Ben Taieb et al., 2017)
  - (Jeon, Panagiotelis, and Petropoulos, 2018)

- **Objectives:**
  1. Defining coherency and reconciliation for probabilistic forecasts
  2. Probabilistic forecast reconciliation in a parametric framework
  3. Probabilistic forecast reconciliation in a non-parametric framework

Project 2: *Coherent probabilistic forecasts*

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 $\boldsymbol{S}$

Project 2: *Reconciled Probabilistic Forecast*

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

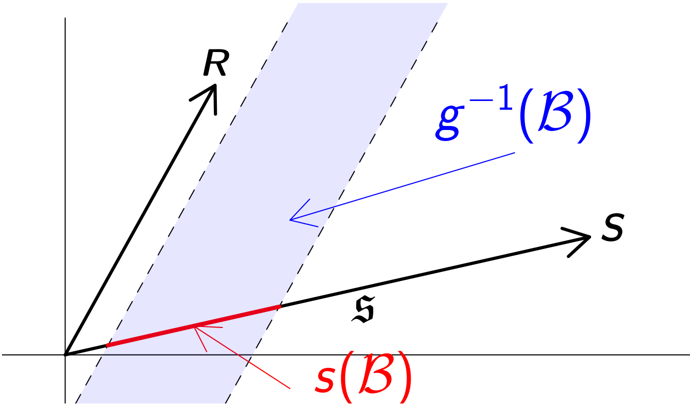
Definition

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.

Project 2: *Geometry*



Project 2: *Analytically*

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

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

Project 2: *For elliptical distributions*

Consider the 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.

Project 2: *Assuming Gaussian distribution*

- Let $\mathscr{N}(\hat{\boldsymbol{y}}_{T+h}, \boldsymbol{W}_{T+h})$ be an incoherent forecast distribution at time $T+h$ where $\hat{\boldsymbol{y}}_{T+h}$ is the incoherent mean and $\boldsymbol{W}_{T+h} = E_{\boldsymbol{y}_{T+h}}[(\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h})(\boldsymbol{y}_{T+h} - \hat{\boldsymbol{y}}_{T+h})^T|\mathcal{I}_T]$ is incoherent variance
- The reconciled Gaussian distribution is given by,

$$\mathscr{N}(\boldsymbol{SG}\hat{\boldsymbol{y}}_{T+h},\ \boldsymbol{SGW}_{T+h}\boldsymbol{G}'\boldsymbol{S}')$$

- $\boldsymbol{G} = \left(\boldsymbol{S}'\boldsymbol{W}_{T+h}^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}_{T+h}^{-1}$ minimizes the energy score in the limiting case A1
- Simulation study evident for improved predictive performance in reconciled Gaussian forecast distributions

Today's talk

- A non-parametric bootstrap approach for probabilistic forecasts reconciliation

- Hierarchical forecasts in macroeconomic variables - An application to Australian GDP forecasts

Part 2: A non-parametric bootstrap approach for probabilistic forecasts reconciliation

Probabilistic forecast reconciliation: Non-parametric approach

- Often parametric densities are unavailable but we can simulate a sample from the predictive distribution

- Suppose $\hat{\boldsymbol{y}}_{T+h}^{[1]}, ..., \hat{\boldsymbol{y}}_{T+h}^{[J]}$ is a sample from the incoherent predictive distribution

- Then setting $\tilde{\boldsymbol{y}}_{T+h}^{[j]} = \boldsymbol{SG}\hat{\boldsymbol{y}}_{T+h}^{[j]}$ produces a sample from the reconciled predictive distribution with respect to $\boldsymbol{G}$

Probabilistic forecast reconciliation: Non-parametric approach

1. Fit univariate models at each node using data up to time T

2. Let $\boldsymbol{\Gamma}_{(T\times n)} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_T)'$ be a matrix of in-sample residuals where $\boldsymbol{e}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$

3. Let $\boldsymbol{\Gamma}^b_{(H\times n)} = (\boldsymbol{e}^b_1, ..., \boldsymbol{e}^b_H)'$ be a block bootstrap sample of size h from $\boldsymbol{\Gamma}$ and repeat this for $b = 1, ..., B$.

4. Generate h-step ahead sample paths from the fitted models incorporating $\boldsymbol{\Gamma}^b$. Denote these by $\hat{\boldsymbol{y}}^b_{T+h}$, for $h = 1, ..., H$.

5. Repeat step 4 for $b = 1, ..., B$ times

6. Setting $\tilde{\boldsymbol{y}}^b_{T+h,j} = \boldsymbol{SG}\hat{\boldsymbol{y}}^b_{T+h,j}$ produces a sample from the reconciled distribution

Optimal reconciliation of future paths

- We propose to find an optimal $\boldsymbol{G}$ matrix by minimizing Energy score

$$\underset{\boldsymbol{G}_h}{\operatorname{argmin}} \quad \mathrm{E}_{\boldsymbol{Q}}[eS(\boldsymbol{S}\boldsymbol{G}_h \hat{\Upsilon}'_{T+h}, \boldsymbol{y}_{T+h})],$$

where,

$$\mathrm{eS}(\tilde{\Upsilon}_{T+h}, \boldsymbol{y}_{T+h}) = \mathrm{E}_{\tilde{\boldsymbol{F}}}\|\tilde{\boldsymbol{Y}}_{T+h} - \boldsymbol{y}_{T+h}\|^{\alpha} - \frac{1}{2}\mathrm{E}_{\tilde{\boldsymbol{F}}}\|\tilde{\boldsymbol{Y}}_{T+h} - \tilde{\boldsymbol{Y}}^{*}_{T+h}\|^{\alpha}, \quad \alpha \in (0,2]$$

- Monte-Carlo approximation to the above objective function is,

$$\underset{\boldsymbol{G}}{\operatorname{argmin}} \quad \sum_{i=1}^{N}\left\{\frac{1}{B}\sum_{j=1}^{B}||\boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}^{b}_{T+h,i,j} - \boldsymbol{y}_{T+h}|| - \frac{1}{2(B-1)}\sum_{j=1}^{B-1}||\boldsymbol{S}\boldsymbol{G}(\hat{\boldsymbol{y}}^{b}_{T+h,i,j} - \hat{\boldsymbol{y}}^{b}_{T+h,i,j+1})||\right\}$$

Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the $\boldsymbol{G}$ matrix

$$\boldsymbol{G} = \left(\boldsymbol{S}'\boldsymbol{W}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W} \tag{1}$$

- We propose four methods to optimise $\boldsymbol{G}$

Method 1: Optimising $\boldsymbol{W}$

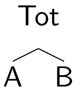
Method 2: Optimising cholesky decomposition of $\boldsymbol{W}$
$\boldsymbol{W} = \boldsymbol{R}'\boldsymbol{R}$ where $\boldsymbol{R}$ is an upper triangular matrix

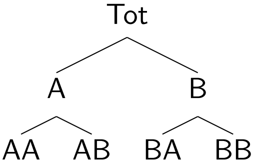
Method 3: Optimising cholesky of $\boldsymbol{W}$ - restricted for scaling
$\boldsymbol{W} = \boldsymbol{R}'\boldsymbol{R}$ s.t $\boldsymbol{i}'\boldsymbol{W}\boldsymbol{i} = 1$ where $\boldsymbol{i} = (1, 0, .., 0)'$

Method 4: Optimising $\boldsymbol{G}$ such that $\boldsymbol{G}\boldsymbol{S} = \boldsymbol{I}$

Monte-Carlo Simulation

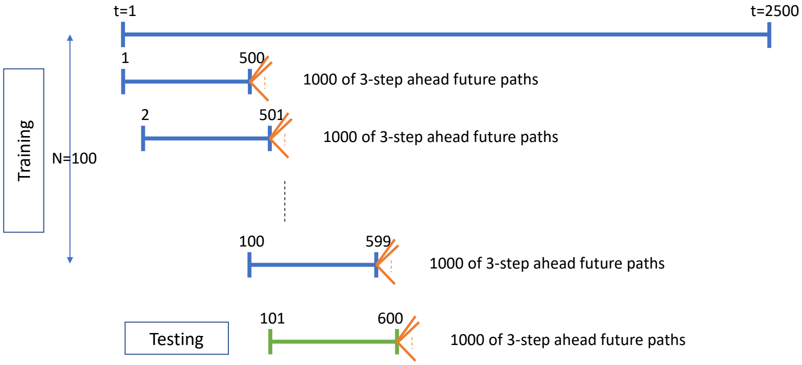
- Data generating process





- DGP was designed such that we have much noisier series in the bottom level

Monte-Carlo Simulation



Monte-Carlo Simulation *Cont.*

- Simulation setup:
  - First 2500 observations generated
  - Fit Univariate ARIMA models for a rolling window of 500 observations
  - $B = 1000$ of $h = 1, 2, 3$ steps-ahead bootstrap future paths generated
  - Training window is rolled one observation ahead and process was repeated until we get $N = 100$ of incoherent, $h = 1, 2, 3$ steps-ahead future paths
  - We find the optimal $\boldsymbol{G}_h$ for $h = 1, 2, 3$ that reconcile h-step-ahead future paths giving minimal average Energy score
  - This optimal $\boldsymbol{G}$ is then used to reconcile the incoherent future paths obtained for the test set
  - Process was repeated for 1000 times and average scores were calculated for the test set

Monte-Carlo Simulation *Cont.*

Optimisation method	Hierarchy 1				Hierarchy 2			
	$h=1$		$h=3$		$h=1$		$h=3$	
	ES	VS	ES	VS	ES	VS	ES	VS
Method 1 - Optimising $\boldsymbol{W}$	2.48	0.11	2.75	0.11	5.36	1.21	5.83	1.38
Method 2 - Optimising $\boldsymbol{R}$	2.48	0.11	2.75	0.11	5.37	1.21	5.83	1.37
Method 3 - Optimising $\boldsymbol{R}$(Restricted)	2.48	0.11	2.75	0.11	5.37	1.21	5.83	1.37
Method 4 - Optimising $\boldsymbol{G}$	2.48	0.11	2.75	0.11	5.38	1.21	5.83	1.38

- Parameterisation does not matter

Monte-Carlo Simulation *Cont.*

- Comparison with point forecast reconciliation methods

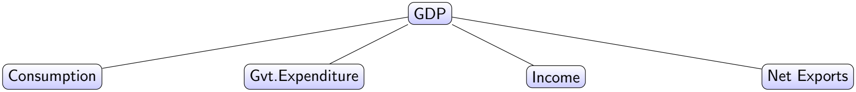
Reconciliation method	Hierarchy 1				Hierarchy 2			
	$h = 1$		$h = 3$		$h = 1$		$h = 3$	
	ES	VS	ES	VS	ES	VS	ES	VS
Optimal $\boldsymbol{G}$	2.48*	0.106	2.75*	0.106	5.36*	1.21*	5.83*	1.38*
MinT(Shrink)	2.47*	0.105	2.74*	0.105	5.33*	1.19*	5.77*	1.34*
WLS	2.46*	0.105	2.74*	0.105	5.43*	1.23	5.98*	1.40*
OLS	2.54*	0.105	2.80*	0.105	5.51*	1.23	5.98*	1.40*
Base	*2.67*	*0.105*	*2.94*	*0.105*	*5.71*	*1.28*	*6.27*	*1.49*

- Reconciliation methods perform better than Base forecasts

- MinT(Shrink) is at least as good as Optimal method. Thus going forward with MinT projection

Hierarchical forecasts for macroeconomic variables - An application to Australian GDP

Macroeconomic forecasting



- Common forecasting approaches involves univariate methods of multivariate methods such as VAR
- The era of big data led to the use of regularization and shrinkage methods - dynamic factor models, Lasso, Bayesian VARs
- The predictors in these methods commonly include the components of the variables of interest.
- This might fail to reflect the deterministic relationship between macroeconomic variables in the forecasts

Macroeconomic forecasting

- Both aligned decision making and forecast accuracy are key concerns for economic agents and policy makers
- Thus we propose to use hierarchical forecasting methods in macroeconomic forecasts.
- Related literature: Only one application on point forecasting for inflation (Capistrán, Constandse, and Ramos-Francia, 2010; Weiss, 2018)
- To the best of our knowledge we use hierarchical forecasting methods for point as well as probabilistic forecasts for the first time in macroeconomic literature

Australian GDP : *Data structures*

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3

- The Australian Bureau of Statistics (ABS) measures GDP using three main approaches - Production, Income and Expenditure

  The final GDP figure is obtained as an average of these three figures.

- We restrict our attention to nominal, seasonally unadjusted data

- Thus we concentrate on the Income and Expenditure approaches

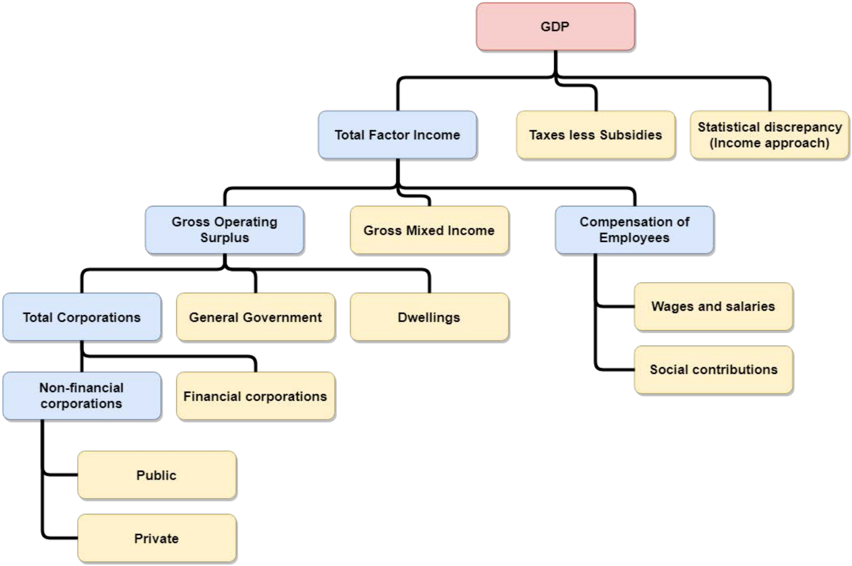
Australian GDP : *Data structures*

Income approach

$$\begin{aligned} GDP = & \textit{ Gross operating surplus} + \textit{Gross mixed income} \\ & + \textit{Compensation of employees} \\ & + \textit{Taxes less subsidies on production and imports} \\ & + \textit{Statistical discrepancy (I)}. \end{aligned}$$

- The hierarchy has two levels of aggregation below the top-level, with a total of $n = 16$ series and $m = 10$ bottom level series

Australian GDP : *Data structures - Income approach*



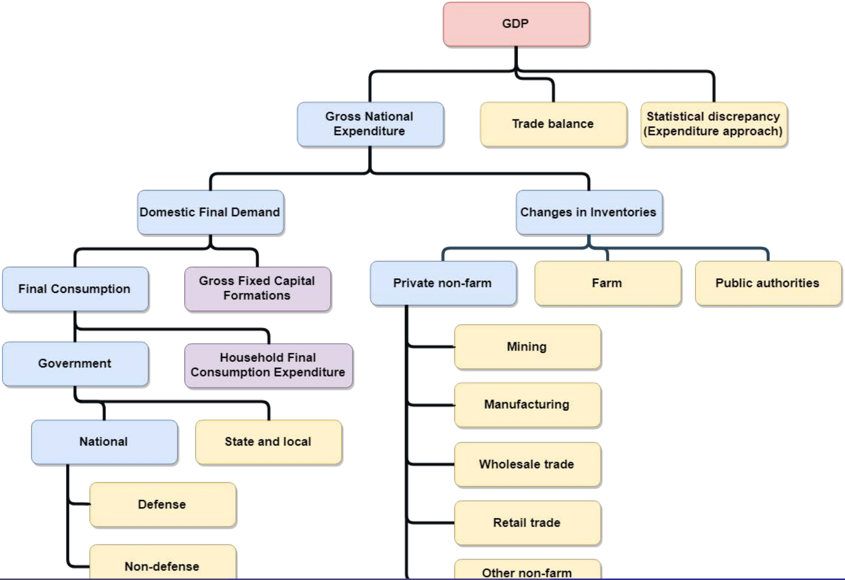
Australian GDP : *Data structures*

Expenditure approach

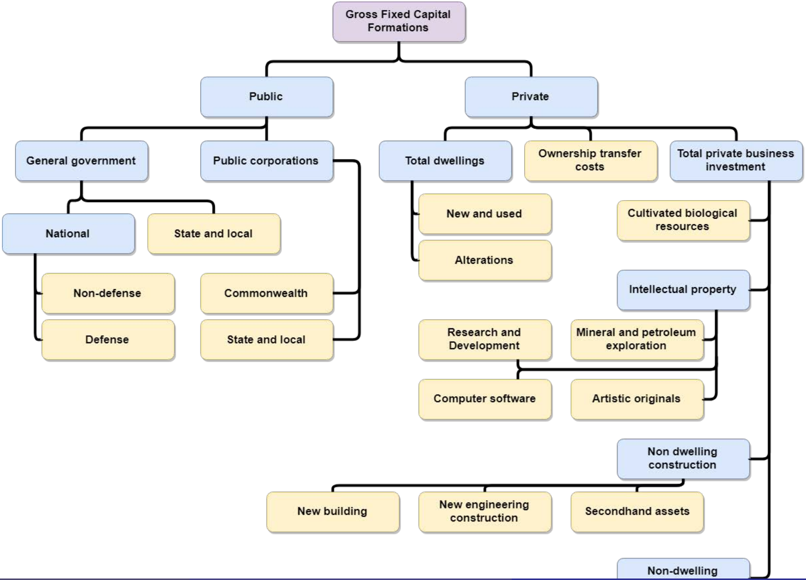
$$\begin{aligned} \mathit{GDP} = {} & \mathit{Final\ consumption\ expenditure} + \mathit{Gross\ fixed\ capital\ formation} \\ & + \mathit{Changes\ in\ inventories} + \mathit{Trade\ balance} \\ & + \mathit{Statistical\ discrepancy\ (E)}. \end{aligned}$$

- The hierarchy has three levels of aggregation below the top-level, with a total of $n = 80$ series and $m = 53$ series at the bottom level

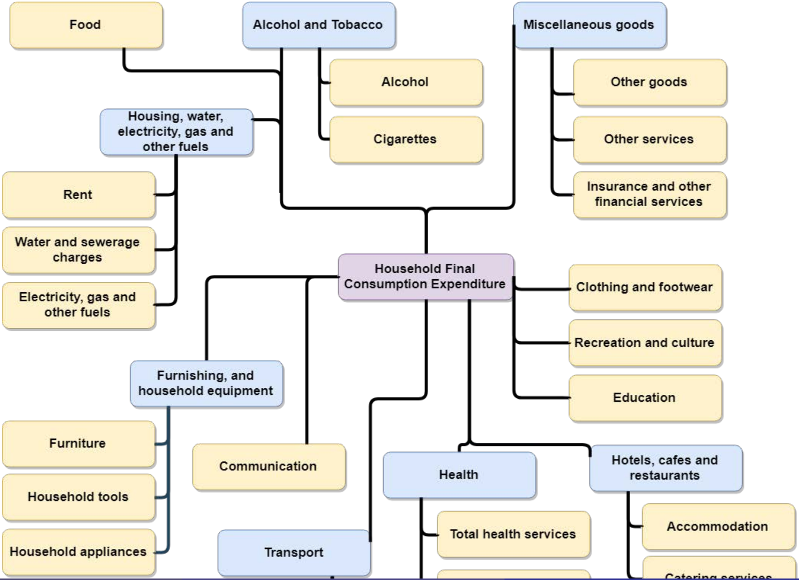
Australian GDP : *Data structures - Expenditure approach*



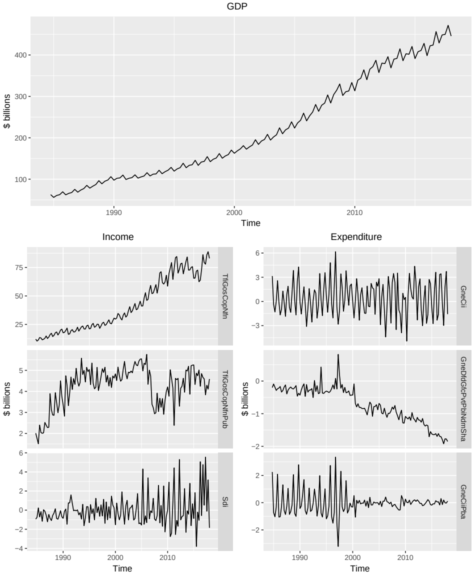
Australian GDP : *Data structures - Expenditure approach*



Australian GDP : *Data structures - Expenditure approach*



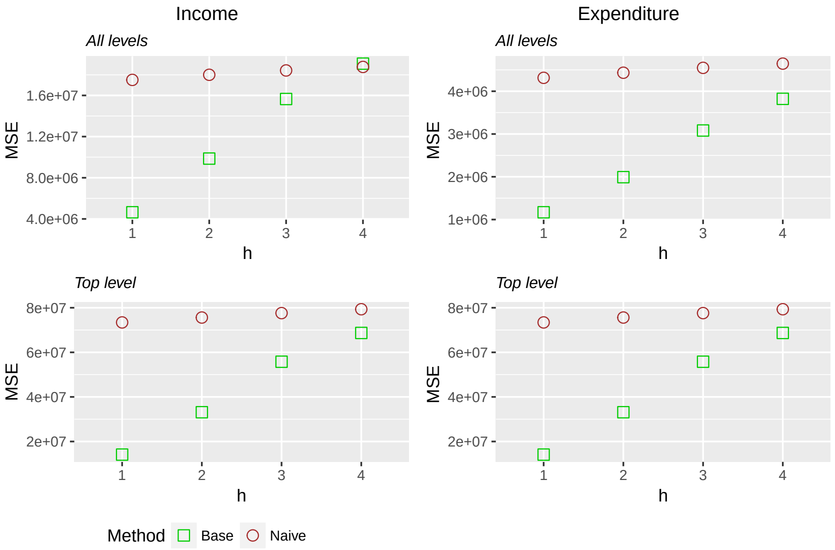
Australian GDP : *Time plots for different levels*



Methodology

- **Analysis set up:**
  - First training sample is set from 1984:Q4 to 1994:Q3 and forecasts produces for four quarters ahead (1994:Q4 to 1995:Q3)
  - Then the training window is expanded by one quarter at a time
  - This leads to 94 1-step-ahead, 93 2-steps-ahead, 92 3-steps-ahead and 91 4-steps-ahead forecasts available for evaluation.
- **Base forecasting models:**
  - Univariate ARIMA and ETS models were fitted for each training set
  - Four step ahead forecasts were generated using the fitted models

Point forecasts: *Base vs Naive*



Point forecasts: *Reconciliation*

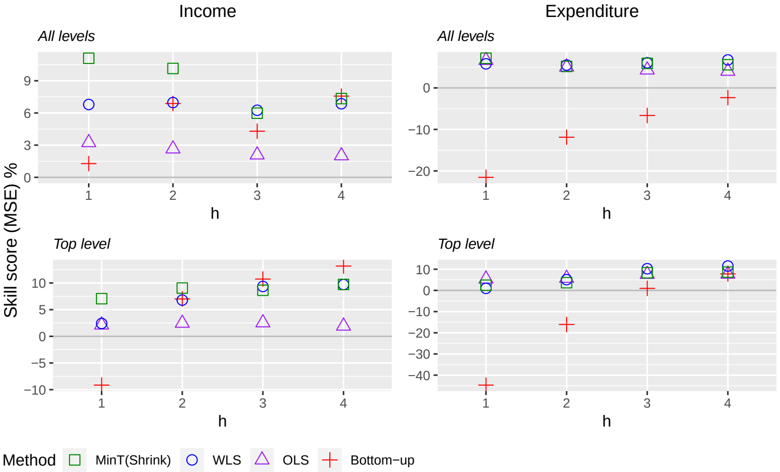
Reconciled forecasts are given by,

$$\tilde{\boldsymbol{y}}_{T+h} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h}$$

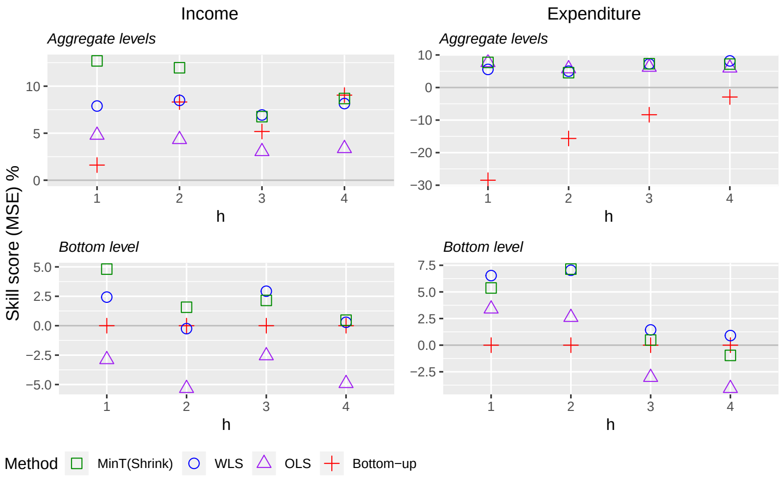
Method	$\boldsymbol{G}$
BU	$(\boldsymbol{0}_{m\times n-m} \ \boldsymbol{I}_{m\times m})$
OLS	$(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$
WLS	$\left(\boldsymbol{S}'\hat{\boldsymbol{W}}^{wls}_{T+1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\hat{\boldsymbol{W}}^{wls}_{T+1}$
MinT(Shrink)	$\left(\boldsymbol{S}'\hat{\boldsymbol{W}}^{shr}_{T+1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\hat{\boldsymbol{W}}^{shr}_{T+1}$

$$\begin{aligned}\hat{\boldsymbol{W}}^{shr}_{T+1} &= \tau\mathrm{Diag}(\hat{\boldsymbol{W}}^{sam}_{T+1}) + (1-\tau)\hat{\boldsymbol{W}}^{sam}_{T+1} \\ \hat{\boldsymbol{W}}^{wls}_{T+1} &= \mathrm{Diag}(\hat{\boldsymbol{W}}^{shr}_{T+1})\end{aligned}$$

Reconciled Point forecasts - *Results*



Reconciled Point forecasts - *Results*



Probabilistic forecasts

- Gaussian approach :

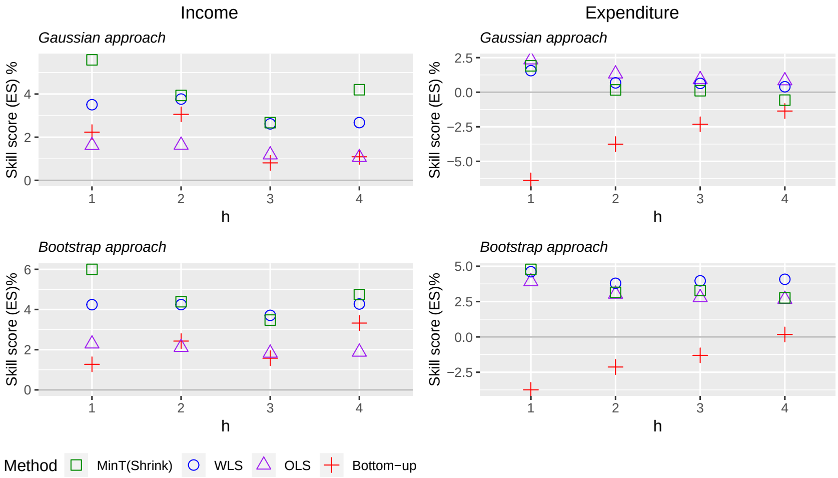
$$\mathcal{N}(\boldsymbol{SG}\hat{\boldsymbol{y}}_{T+h},\ \boldsymbol{SGW}_{T+h}\boldsymbol{G}'\boldsymbol{S}')$$

- Non-parametric Bootstrap approach :

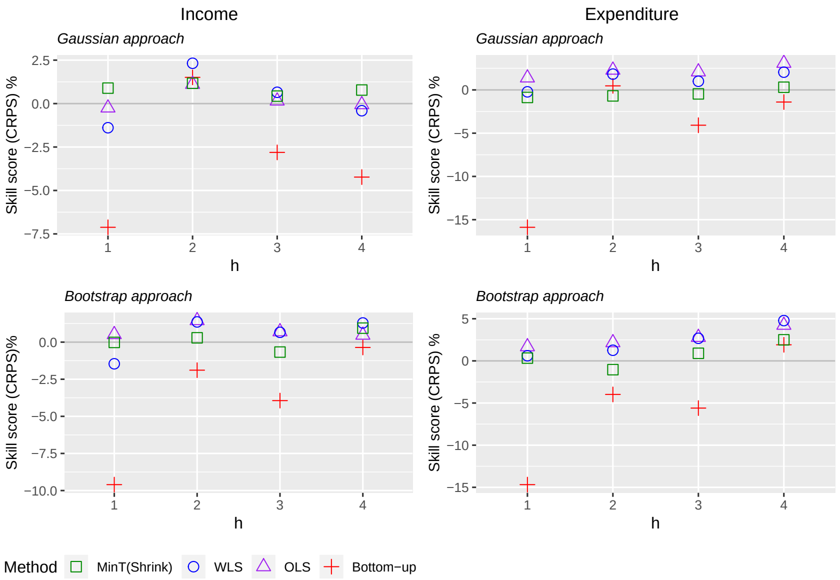
$$\tilde{\Upsilon}_{T+h} = \boldsymbol{SG}\hat{\Upsilon}'_{T+h}$$

where, $\hat{\Upsilon}_{T+h} = (\hat{\boldsymbol{y}}^1_{T+h}, ..., \hat{\boldsymbol{y}}^B_{T+h})'$

Reconciled Probabilistic Forecasts



Reconciled Probabilistic Forecasts



Summary and time plan for completion

Summary

- We define point and probabilistic forecast reconciliation in geometric terms
- We propose a parametric approach for probabilistic forecast reconciliation
- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts

- Simulation study evident that the optimal reconciliation with respect to energy score is equivalent to reconciling each sample path via MinT approach

- We apply hierarchical forecast reconciliation methods to forecast Australian GDP in point as well as probabilistic framework

Time plan for completion

	Thesis Chapter	Task description	Time duration	Progress
1 and 2.	Introduction and Background Review	Writing the chapter.	September/2019 - October/2019	40% complete
3.	Hierarchical forecast reconciliation in Geometric view	Bias correction and application	May/2019 - July/2019	75% Completed
4.	Probabilistic forecast reconciliation for hierarchical time series	Completing the paper.	June/2019 - August/2019	90% Completed
5.	Application	Forecasting Australian GDP		100% Completed

References I

Ben Taieb, S., R. Huser, R. J. Hyndman, and M. G. Genton (2017). Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Transactions on Smart Grid* **7**(5), 2448–2455.

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

Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102**(477), 359–378.

Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test* **17**(2), 211–235.

References II

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

Jeon, J., A. Panagiotelis, and F. Petropoulos (2018). "Reconciliation of probabilistic forecasts with an application to wind power".

Scheuerer, M. and T. M. Hamill (2015). Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities *. *Monthly Weather Review* **143**(4), 1321–1334.

Weiss, C. (2018). "Essays in Hierarchical Time Series Forecasting and Forecast Combination". PhD thesis. University of Cambridge.

Wickramasuriya, S. L., G. Athanasopoulos, and R. J. Hyndman (2018). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association* **1459**, 1–45.

Thank You!!

Project 2: *Probabilistic forecasts evaluation*

A1

Energy score (Gneiting et al., 2008)

$$\text{eS}(\breve{\boldsymbol{Y}}_{T+h}, \boldsymbol{y}_{T+h}) = \text{E}_{\breve{\boldsymbol{F}}}\|\breve{\boldsymbol{Y}}_{T+h} - \boldsymbol{y}_{T+h}\|^{\alpha} - \frac{1}{2}\text{E}_{\breve{\boldsymbol{F}}}\|\breve{\boldsymbol{Y}}_{T+h} - \breve{\boldsymbol{Y}}^{*}_{T+h}\|^{\alpha}, \quad \alpha \in (0,2]$$

Log score (Gneiting and Raftery, 2007)

$$\text{LS}(\breve{\boldsymbol{F}}, \boldsymbol{y}_{T+h}) = -\log \breve{\boldsymbol{f}}(\boldsymbol{y}_{T+h})$$

Variogram score (Scheuerer and Hamill, 2015)

$$\text{VS}(\breve{\boldsymbol{F}}, \boldsymbol{y}_{T+h}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\Big(|y_{T+h,i} - y_{T+h,j}|^{p} - \text{E}_{\breve{\boldsymbol{F}}}|\breve{Y}_{T+h,i} - \breve{Y}_{T+h,j}|^{p}\Big)^{2}$$

CRPS (Gneiting and Raftery, 2007)

$$\text{CRPS}(\breve{F}_i, y_{T+h,i}) = \text{E}_{\breve{F}_i}|\breve{Y}_{T+h,i} - y_{T+h,i}| - \frac{1}{2}\text{E}_{\breve{F}_i}|\breve{Y}_{T+h,i} - \breve{Y}^{*}_{T+h,i}|$$

$\breve{\boldsymbol{Y}}_{T+h}$ and $\breve{\boldsymbol{Y}}^{*}_{T+h}$	:	Independent random vectors from the coherent forecast distribution $\breve{\boldsymbol{F}}$.
$\boldsymbol{y}_{T+h}$	:	Vector of realizations.
$\breve{Y}_{T+h,i}$ and $\breve{Y}_{T+h,j}$	:	ith and jth components of the vector $\breve{\boldsymbol{Y}}_{T+h}$

Appendix

- A2 Shrinkage estimator for 1-step ahead base forecast errors

$$\hat{\boldsymbol{\Sigma}}^{shr}_{T+1} = \tau \hat{\boldsymbol{\Sigma}}^{D}_{T+1} + (1-\tau)\hat{\boldsymbol{\Sigma}}_{T+1},$$

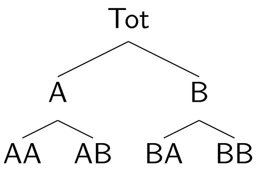
where $\hat{\boldsymbol{\Sigma}}^{D}_{T+1}$ is the diagonal matrix comprising diagonal entries of $\hat{\boldsymbol{\Sigma}}_{T+1}$ and

$$\tau = \frac{\sum_{i \neq j} \hat{Var}(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}^2_{ij}}$$

is a shrinkage parameter. $\hat{r}_{ij}$ is the *ij*-th element of sample correlation matrix. In this estimation, the off-diagonal elements of 1-step ahead sample covariance matrix will be shrunk to zero depending on the sparsity.

Monte-Carlo simulation

- **Data generating process** A3



- $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim ARIMA(p, d, q)$
- $p \in \{1, 2\}$ and $d \in \{0, 1\}$
- $\{\epsilon_{AA,t}, \epsilon_{AB,t}, \epsilon_{BA,t}, \epsilon_{BB,t}\} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$
- Parameters for *AR* and *MA* components were randomly and uniformly generated from $[0.3, 0.5]$ and $[0.3, 0.7]$ respectively

- $\boldsymbol{y}_t$ are then generated as follows

Bottom level	**Aggregate level 1**	**Total**
$y_{AA,t} = w_{AA,t} + u_t - 0.5v_t$	$y_{A,t} = w_{AA,t} + w_{AB,t} - v_t$	$y_{Tot,t} = w_{AA,t} + w_{AB,t} + w_{BA,t} + w_{BB,t}$
$y_{AB,t} = w_{AB,t} - u_t - 0.5v_t$	$y_{B,t} = w_{BA,t} + w_{BB,t} + v_t$	
$y_{BA,t} = w_{BA,t} + u_t + 0.5v_t$		
$y_{BB,t} = w_{BB,t} - u_t + 0.5v_t$		

Monte-Carlo simulation *Cont..*

- To get less noisier series at aggregate levels, we choose $\boldsymbol{\Sigma}, \sigma_u^2$ and σ_v^2 such that,

$$\mathrm{Var}(\epsilon_{AA,t} + \epsilon_{AB,t} + \epsilon_{BA,t} + \epsilon_{BB,t}) \leq \mathrm{Var}(\epsilon_{AA,t} + \epsilon_{AB,t} - v_t) \leq \mathrm{Var}(\epsilon_{AA,t} + u_t - 0.5v_t),$$

- Thus we choose, $\boldsymbol{\Sigma} = \begin{pmatrix} 5.0 & 3.1 & 0.6 & 0.4 \\ 3.1 & 4.0 & 0.9 & 1.4 \\ 0.6 & 0.9 & 2.0 & 1.8 \\ 0.4 & 1.4 & 1.8 & 3.0 \end{pmatrix}$, $\sigma_u^2 = 19$ and $\sigma_u^2 = 18$.

Sample version of the scoring rules

- For a possible finite sample of size B from the multivariate forecast density $\breve{\boldsymbol{F}}$, the variogram score is defined as,

$$\text{VS}(\breve{\boldsymbol{F}}, \boldsymbol{y}_{T+h}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(|y_{T+h,i} - y_{T+h,j}|^p - \frac{1}{B}\sum_{k=1}^{B}|\breve{Y}^k_{T+h,i} - \breve{Y}^k_{T+h,j}|^p\right)^2$$