

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

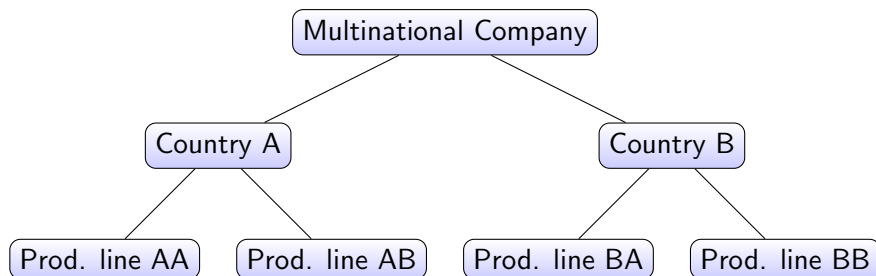
May 23, 2019

- 1 Project 1: Hierarchical Forecast Reconciliation in Geometric View
- 2 Project 2: Probabilistic Forecasts for Hierarchical Time Series
 - Part 1: Definitions and Parametric Approach
 - Part 2: A non-parametric bootstrap approach for probabilistic forecasts reconciliation
- 3 Hierarchical forecasts in macroeconomic variables - An application to Australian GDP forecasts
- 4 Summary and time plan for completion

Project 1: Hierarchical Forecast Reconciliation in Geometric View

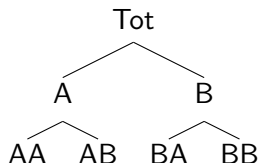
Project 1: Motivation

■ Example:



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

Project 1: Notations and Preliminaries



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

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

$$m = 4$$

$$n = 7$$

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

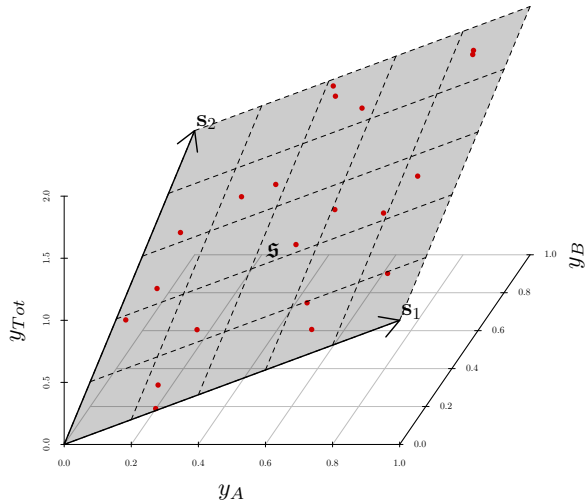
- Due to the aggregation nature of the hierarchy we have,

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

Coherent subspace

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

Project 1: *Coherent forecasts*



Project 1: *Point forecast reconciliation*

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

Definition

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

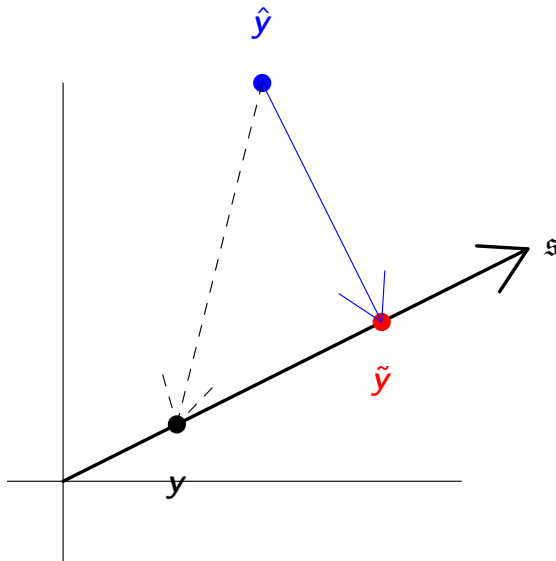
$$\tilde{\mathbf{y}} = \mathbf{S}g(\hat{\mathbf{y}})$$

- If $g(\cdot)$ is linear,

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

Project 1: *Point forecast reconciliation - OLS*

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



Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, [2018](#))

- **Geometry**

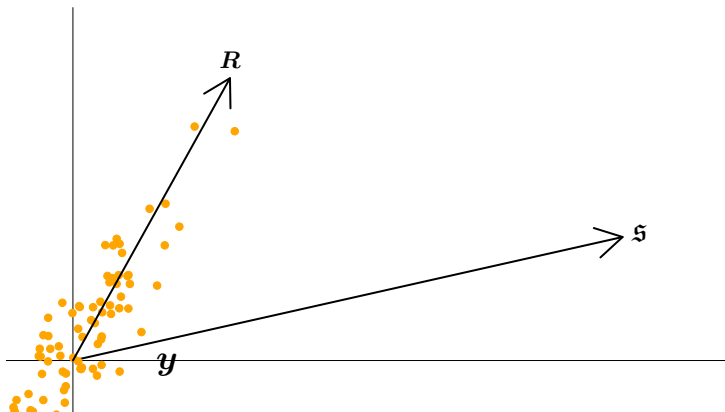
- Consider the covariance matrix of $\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h}$.
- This can be estimated using in-sample forecast errors.

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

- 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.

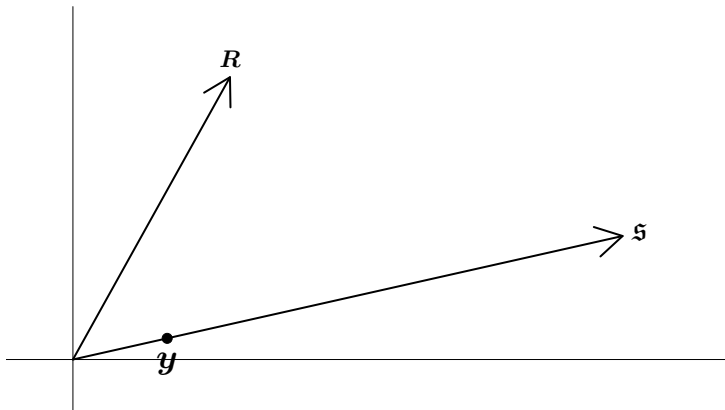
Project 1: *Point forecast reconciliation - MinT*

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$



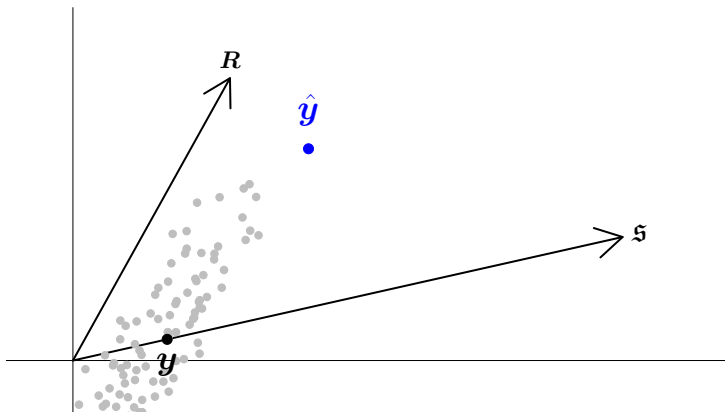
Project 1: *Point forecast reconciliation - MinT*

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$



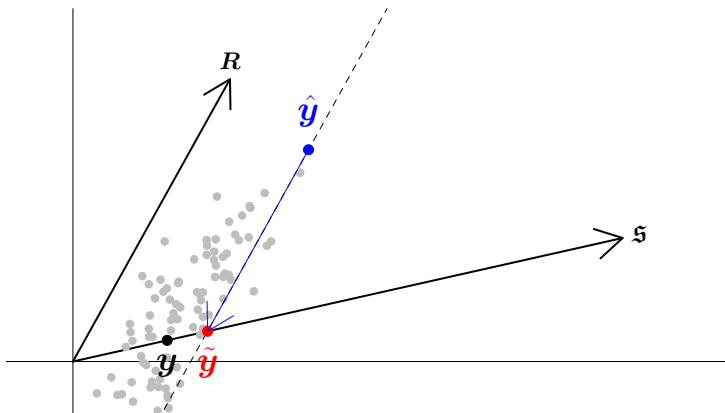
Project 1: *Point forecast reconciliation - MinT*

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$



Project 1: *Point forecast reconciliation - MinT*

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$



Project 1: *Why projections?*

- Projections preserve the unbiasedness
- What if the incoherent 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 Parametric Approach

Project 2: *Motivation and Objectives*

- Extending the “reconciliation” method into 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 the parametric framework
 - 3 Probabilistic forecast reconciliation in the 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 \mathbf{S}

Project 2: *Reconciled Probabilistic Forecast*

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

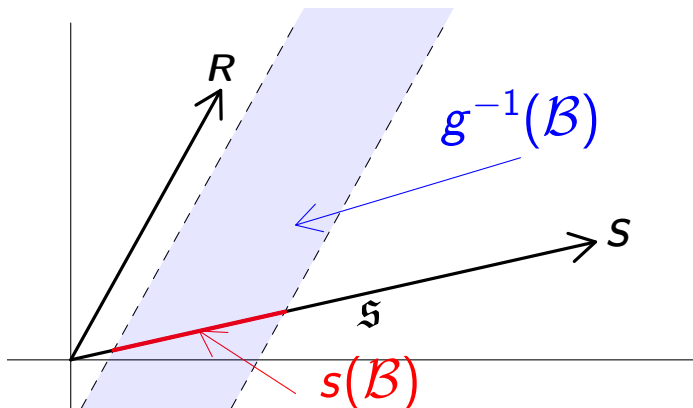
Definition

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 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.

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

Project 2: *For elliptical distributions*

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

Theorem

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

$$\mathcal{N}(\mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}'\mathbf{S}')$$

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

- 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 $\hat{\mathbf{y}}_{T+h}^{[1]}, \dots, \hat{\mathbf{y}}_{T+h}^{[J]}$ is a sample from the incoherent predictive distribution
- Then setting $\tilde{\mathbf{y}}_{T+h}^{[j]} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}^{[j]}$ produces a sample from the reconciled predictive distribution with respect to \mathbf{G}

Probabilistic forecast reconciliation: Non-parametric approach

- 1 Fit univariate models at each node using data up to time T
- 2 Let $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$ be a matrix of in-sample residuals where $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$
- 3 Let $\mathbf{\Gamma}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$ be a block bootstrap sample of size h from $\mathbf{\Gamma}$ and repeat this for $b = 1, \dots, B$.
- 4 Generate h -step ahead sample paths from the fitted models incorporating $\mathbf{\Gamma}^b$. Denote these by $\hat{\mathbf{y}}_{T+h}^b$, for $h = 1, \dots, H$.
- 5 Repeat step 4 for $b = 1, \dots, B$ times
- 6 Setting $\tilde{\mathbf{y}}_{T+h,j}^b = \mathbf{SG} \hat{\mathbf{y}}_{T+h,j}^b$ produces a sample from the reconciled distribution

Optimal reconciliation of future paths

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

$$\operatorname{argmin}_{\mathbf{G}_h} \mathbb{E}_Q[\mathbf{eS}(\mathbf{S}\mathbf{G}_h \hat{\mathbf{Y}}'_{T+h}, \mathbf{y}_{T+h})],$$

where,

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

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

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

Optimal reconciliation of future paths *Cont.*

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

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}\mathbf{S})^{-1} \mathbf{S}'\mathbf{W} \quad (1)$$

- We propose four methods to optimise \mathbf{G}

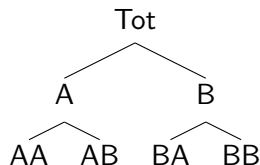
Method 1: Optimising \mathbf{W}

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

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

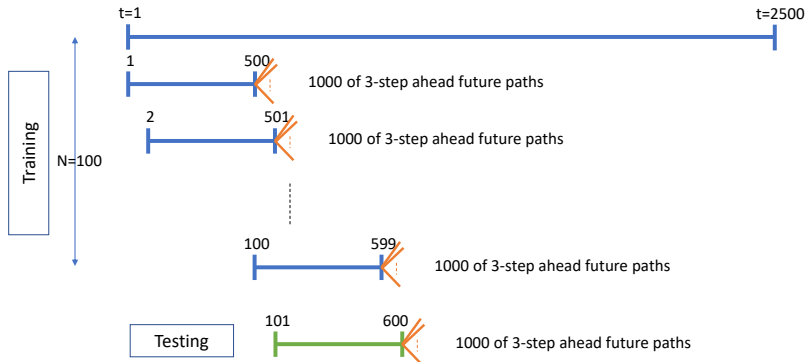
Method 4: Optimising \mathbf{G} such that $\mathbf{GS} = \mathbf{I}$

- Data generating process



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

Monte-Carlo Simulation



■ 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 \mathbf{G}_h for $h = 1, 2, 3$ that reconcile h -step-ahead future paths giving minimal average Energy score
- This optimal \mathbf{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 W	2.48	0.11	2.75	0.11	5.36	1.21	5.83	1.38
Method 2 - Optimising R	2.48	0.11	2.75	0.11	5.37	1.21	5.83	1.37
Method 3 - Optimising R (Restricted)	2.48	0.11	2.75	0.11	5.37	1.21	5.83	1.37
Method 4 - Optimising G	2.48	0.11	2.75	0.11	5.38	1.21	5.83	1.38

- Parameterisation does not matter

■ Comparison with point forecast reconciliation methods

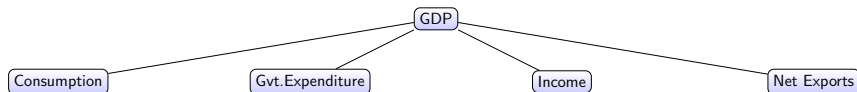
Reconciliation method	Hierarchy 1				Hierarchy 2			
	$h = 1$		$h = 3$		$h = 1$		$h = 3$	
	ES	VS	ES	VS	ES	VS	ES	VS
Optimal \mathbf{G}	2.48*	0.106	2.75*	0.106	5.36*	1.21*	5.83*	1.38*
MinT(Shrink)	2.47*	0.105	2.74*	0.105	5.33*	1.19*	5.77*	1.34*
WLS	2.46*	0.105	2.74*	0.105	5.43*	1.23	5.98*	1.40*
OLS	2.54*	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 in macroeconomic variables - An application to Australian GDP forecasts

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

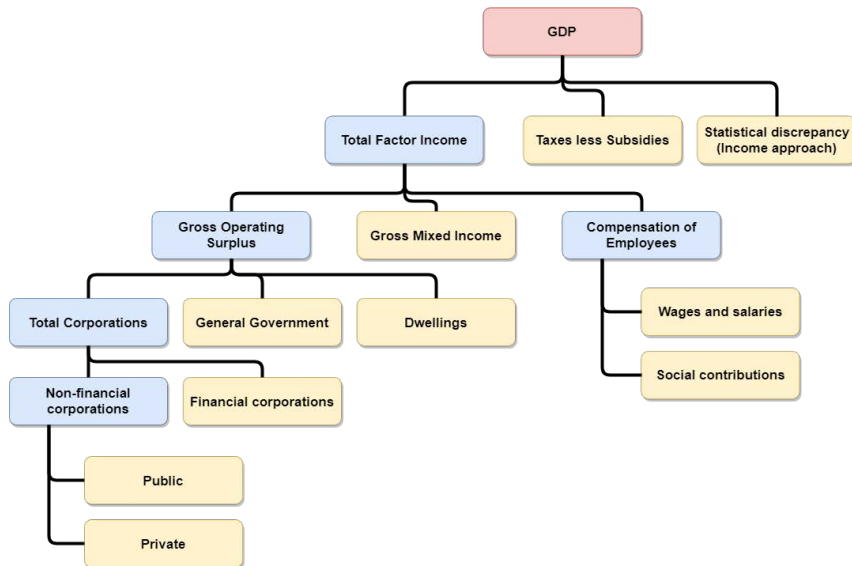
- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3
- The Australian Bureau of Statistics (ABS) measures GDP using three main approaches - Production, Income and Expenditure
The final GDP figure is obtained as an average of these three figures.
- We restrict our attention to nominal, seasonally unadjusted data
- Thus we concentrate on the Income and Expenditure approaches

Income approach

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

- 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*

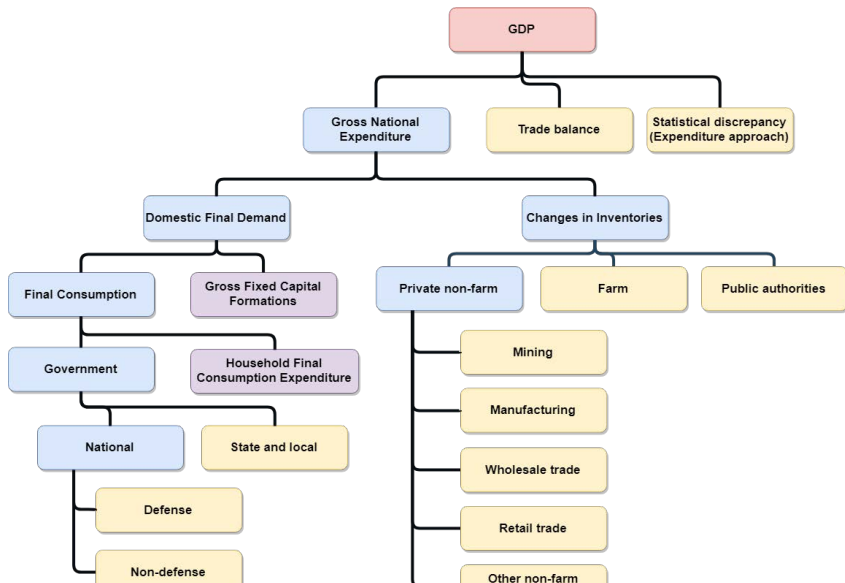


Expenditure approach

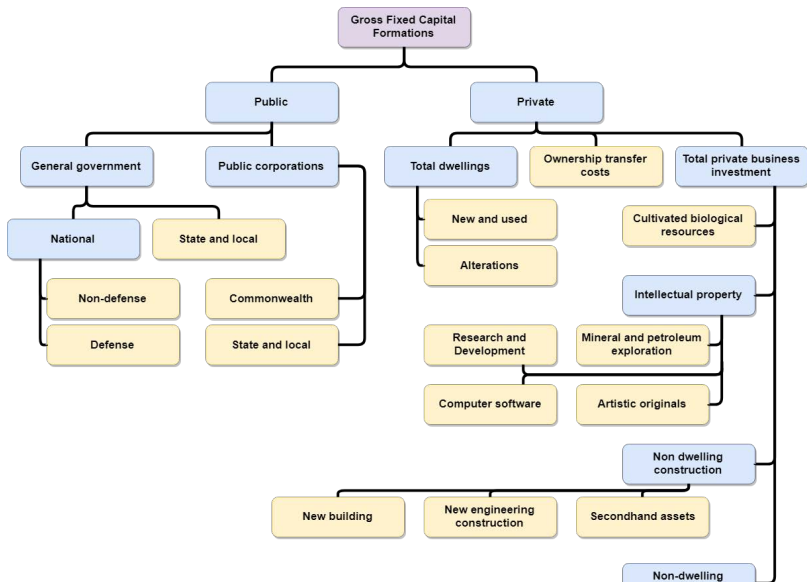
$$\begin{aligned} \text{GDP} = & \text{Final consumption expenditure} + \text{Gross fixed capital formation} \\ & + \text{Changes in inventories} + \text{Trade balance} \\ & + \text{Statistical discrepancy (E)}. \end{aligned}$$

- 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

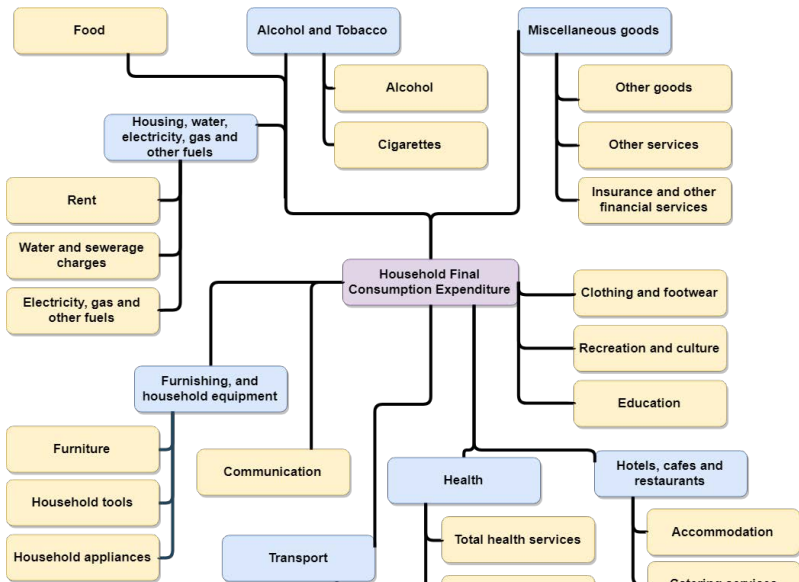
Australian GDP : *Data structures - Expenditure approach*



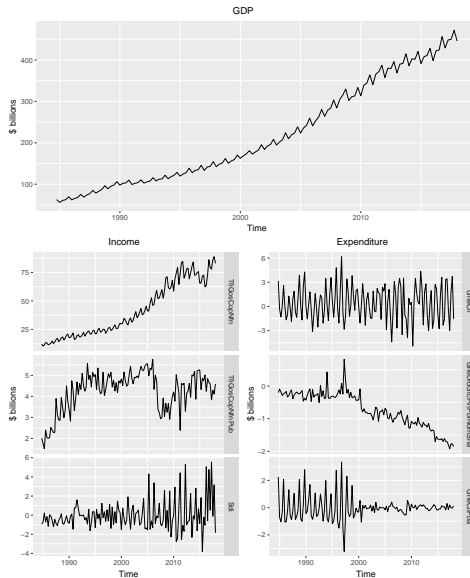
Australian GDP : *Data structures - Expenditure approach*



Australian GDP : *Data structures - Expenditure approach*



Australian GDP : *Time plots for different levels*



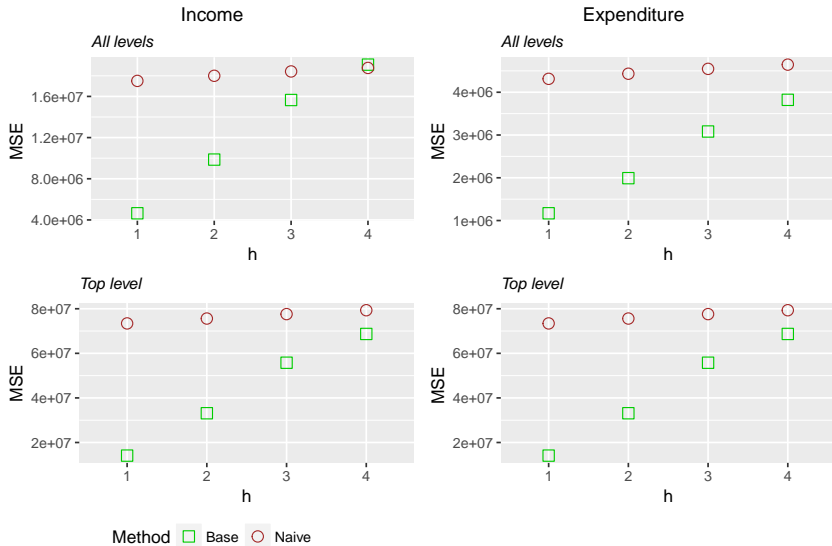
■ Analysis set up:

- 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,

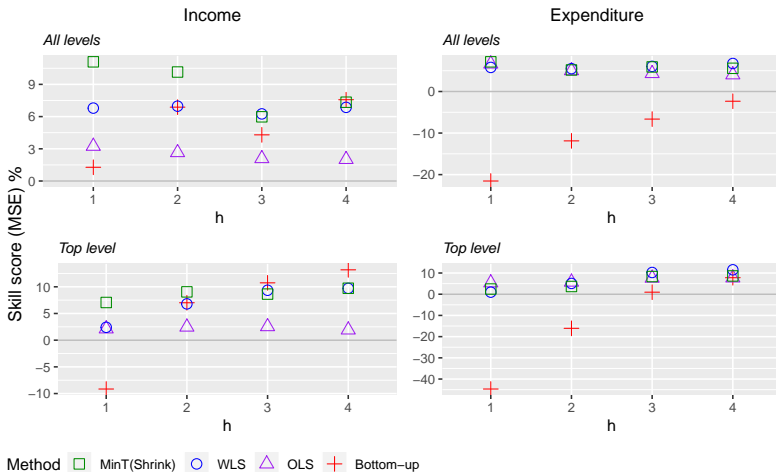
$$\tilde{\mathbf{y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}$$

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

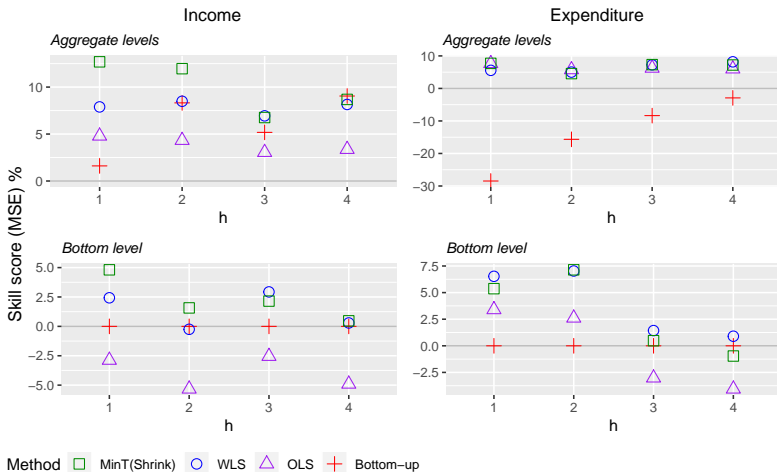
$$\hat{\mathbf{W}}_{T+1}^{shr} = \tau \text{Diag}(\hat{\mathbf{W}}_{T+1}^{sam}) + (1 - \tau)\hat{\mathbf{W}}_{T+1}^{sam}$$

$$\hat{\mathbf{W}}_{T+1}^{wls} = \text{Diag}(\hat{\mathbf{W}}_{T+1}^{shr})$$

Reconciled Point forecasts - Results



Reconciled Point forecasts - Results



- Gaussian approach :

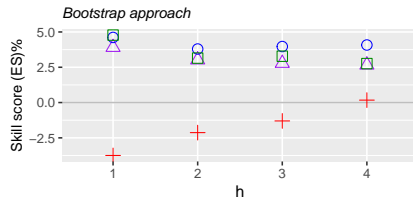
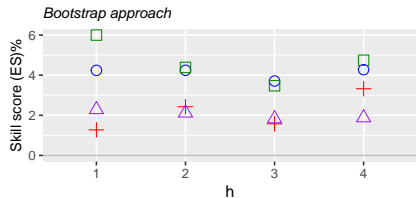
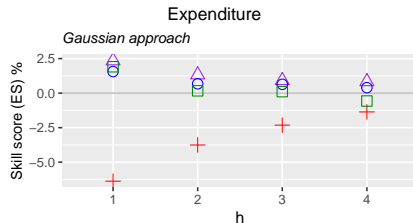
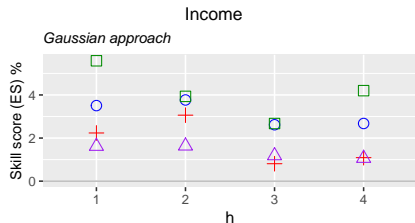
$$\mathcal{N}(\mathbf{SG}\hat{\mathbf{y}}_{T+h}, \mathbf{SGW}_{T+h}\mathbf{G}'\mathbf{S}')$$

- Non-parametric Bootstrap approach :

$$\tilde{\mathbf{Y}}_{T+h} = \mathbf{SG}\hat{\mathbf{Y}}'_{T+h}$$

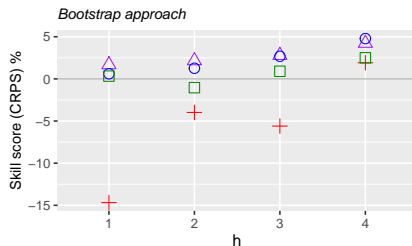
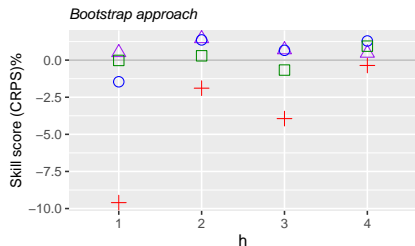
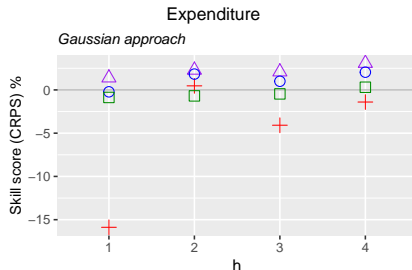
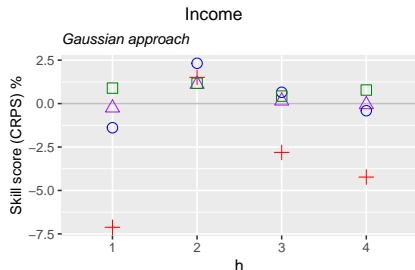
where, $\hat{\mathbf{Y}}_{T+h} = (\hat{\mathbf{y}}_{T+h}^1, \dots, \hat{\mathbf{y}}_{T+h}^B)'$

Reconciled Probabilistic Forecasts



Method ■ MinT(Shrink) ○ WLS △ OLS + Bottom-up

Reconciled Probabilistic Forecasts



Method ■ MinT(Shrink) ○ WLS △ OLS + Bottom-up

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

- 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* **145**9, 1–45.

Thank You!!

Project 2: Probabilistic forecasts evaluation

A1

Energy score (Gneiting et al., 2008)

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

Log score (Gneiting and Raftery, 2007)

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

Variogram score (Scheuerer and Hamill, 2015)

$$\text{VS}(\check{\mathbf{F}}, \mathbf{y}_{T+h}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left(|y_{T+h,i} - y_{T+h,j}|^p - \mathbb{E}_{\check{\mathbf{F}}} |\check{Y}_{T+h,i} - \check{Y}_{T+h,j}|^p \right)^2$$

CRPS (Gneiting and Raftery, 2007)

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

$\check{\mathbf{Y}}_{T+h}$ and $\check{\mathbf{Y}}_{T+h}^*$: Independent random vectors from the coherent forecast distribution $\check{\mathbf{F}}$.

\mathbf{y}_{T+h} : Vector of realizations.

$\check{Y}_{T+h,i}$ and $\check{Y}_{T+h,j}$: i th and j th components of the vector $\check{\mathbf{Y}}_{T+h}$

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

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

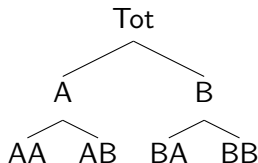
where $\hat{\Sigma}_{T+1}^D$ is the diagonal matrix comprising diagonal entries of $\hat{\Sigma}_{T+1}$ and

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

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

■ y_t are then generated as follows

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

- To get less noisier series at aggregate levels, we choose Σ , σ_u^2 and σ_v^2 such that,

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

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

Sample version of the scoring rules

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

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