MONASH University

Forecasting Hierarchical Time Series Through Trace Minimization

Shanika L. Wickramasuriya

with

George Athanasopoulos and Rob J. Hyndman

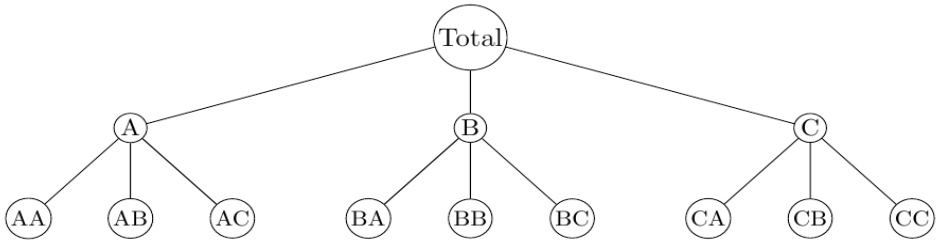
June 22, 2015

Department of Econometrics and Business Statistics

Outline

Hierarchical time series

- A ***hierarchical time series*** is a collection of time series linked in a hierarchical structure.



Example: Tourism demand: Australia, States, SLAs.

- **Challenge:** Independent forecasts do not add-up across the hierarchy.

Solutions:

- Top-down
- Bottom-up
- Middle-out
- Optimal combination (Hyndman et al., 2011)

Outline

Notations used in hierarchical time series

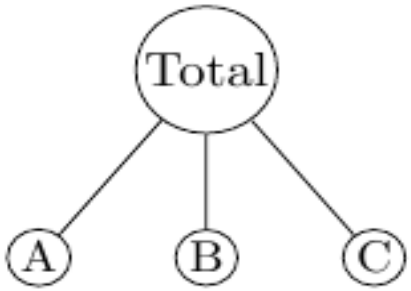
y_t : Aggregate of all time series at time t

$y_{X,t}$: Value of series X at time t

$\boldsymbol{b}_t$: Vector of all observations at bottom level at time t

m : Total no. of series in the hierarchy

m_K : no. of series in the bottom level



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

where,

$$\boldsymbol{y}_t = \begin{bmatrix} y_t & y_{A,t} & y_{B,t} & y_{C,t} \end{bmatrix}', \quad \boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{b}_t = \begin{bmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \end{bmatrix}.$$

Optimal combination

Regression model

$$\hat{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{\beta}_n(h) + \boldsymbol{\varepsilon}_h.$$

- $\hat{\boldsymbol{y}}_n(h)$: h-step ahead **base forecasts** stacked in the same order as $\boldsymbol{y}_t$
- $\boldsymbol{\beta}_n(h) = E[\boldsymbol{b}_n(h)|\boldsymbol{y}_1, ..., \boldsymbol{y}_n]$
- $\boldsymbol{\varepsilon}_h$, **reconciliation error**, with zero mean and covariance $\boldsymbol{\Sigma}_h$

If $\boldsymbol{\Sigma}_h$ is known, GLS estimator of $\boldsymbol{\beta}_n(h)$ leads

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\hat{\boldsymbol{\beta}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\hat{\boldsymbol{y}}_n(h).$$

$\tilde{\boldsymbol{y}}_n(h)$: **reconciled forecasts**

$\boldsymbol{\Sigma}_h^{\dagger}$: Moore-Penrose generalized inverse of $\boldsymbol{\Sigma}_h$

All existing methods of hierarchical forecasting:

$$\tilde{\boldsymbol{y}}_n(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h).$$

Pitfalls arise in the estimation of $\boldsymbol{\Sigma}_h$

Estimation is prohibited by the singularity of the matrices.

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}}_h &= \hat{\boldsymbol{y}}_n(h) - \tilde{\boldsymbol{y}}_n(h) \\ &= \hat{\boldsymbol{y}}_n(h) - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_n(h) \\ &= (\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})\hat{\boldsymbol{y}}_n(h),\end{aligned}$$

which gives

$$\begin{aligned}\text{Var}(\hat{\boldsymbol{\varepsilon}}_h) &= (\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})\text{Var}\left[\hat{\boldsymbol{y}}_n(h)\right](\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})' \\ &= (\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})\boldsymbol{\Sigma}_h(\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})'.\end{aligned}$$

- For any choice of $\boldsymbol{P}$, $(\boldsymbol{I}_m - \boldsymbol{S}\boldsymbol{P})$ is a rank deficient matrix.
- $\boldsymbol{\Sigma}_h$ is not identifiable.

Outline

Motivation

- Goal of a set of forecasts is to produce estimates close to the actual values yet to be observed.
- Statistical point of view: obtain a set of minimum variance unbiased estimates of future values.
- Current study proposes a method of forecasting hierarchical time series through minimizing the sum of **variances of the reconciled forecast errors** under the property of **unbiasedness**.

Theoretical properties

If base forecasts are unbiased:

$$E[\hat{\boldsymbol{y}}_n(h)|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n] = E[\boldsymbol{y}_{n+h}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_n],$$

then reconciled forecasts are unbiased if and only if

$$\boldsymbol{SPS} = \boldsymbol{S}.$$

Lemma (Forecast error variance)

For any given $\boldsymbol{P}$ which results in unbiased reconciled forecasts,

$$\text{Var}[\boldsymbol{y}_{n+h} - \tilde{\boldsymbol{y}}_n(h)] = \boldsymbol{SPW}_h\boldsymbol{P}'\boldsymbol{S}'.$$

- $\boldsymbol{y}_{n+h} - \tilde{\boldsymbol{y}}_n(h)$: forecast errors of reconciled forecasts
- $\boldsymbol{W}_h = \text{Var}[\boldsymbol{y}_{n+h} - \hat{\boldsymbol{y}}_n(h)]$

Hierarchical forecasting using MinT

Theorem

The solution for minimizing the objective function,

$$\min_{\boldsymbol{P}} \quad \text{tr}\left[\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}'\right]$$

$$\text{s.t.} \quad \boldsymbol{S}\boldsymbol{P}\boldsymbol{S} = \boldsymbol{S},$$

where $\boldsymbol{W}_h$ is the positive definite covariance matrix of the h-step ahead base forecast errors, is uniquely attained at,

$$\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}.$$

Alternative representation of MinT

Let

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{C}_{m^* \times m_K} \\ \boldsymbol{I}_{m_K} \end{bmatrix},$$

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{0}_{m_K \times m^*} & \boldsymbol{I}_{m_K} \end{bmatrix},$$

$$\text{and} \qquad \boldsymbol{U}' = \begin{bmatrix} \boldsymbol{I}_{m^*} & -\boldsymbol{C}_{m^* \times m_K} \end{bmatrix},$$

where $m^* = m - m_K$.

$\boldsymbol{P}$ can be re-expressed,

$$\boldsymbol{P} = \boldsymbol{J} - \boldsymbol{J}\boldsymbol{W}_h\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}'.$$

- Convenient, as requires inversion of a $m^* \times m^*$ matrix ($m^* < m_K$).
- Coincides with the work of Stone (1976) and Byron (1978) in the area of balancing national income accounts.

Remarks

Special structures on $\boldsymbol{W}_h$ leads:

- When $\boldsymbol{W}_h = k_h \boldsymbol{I}$, $\forall h$, **OLS** and **MinT** coincide.
- When $\boldsymbol{W}_h = k_h \text{diag}(\hat{\boldsymbol{W}}_1), \forall h$, **approximated WLS** (Hyndman et al., 2014) and **MinT** coincide.

  $\text{diag}(\hat{\boldsymbol{W}}_1)$: diagonal matrix with elements given by variance of the in-sample one-step-ahead base forecast errors

Estimating $\boldsymbol{W}_h$

$$\boldsymbol{P} = \boldsymbol{J} - \boldsymbol{J}\boldsymbol{W}_h\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}'.$$

For simplicity, it is assumed that,

$$\boldsymbol{W}_h = k_h\boldsymbol{W}_1, \qquad \forall h,$$

- $\boldsymbol{W}_1, \boldsymbol{W}_h$: covariance matrix of 1-step & h-step ahead base forecast errors

Sample covariance matrix

$$\hat{\boldsymbol{W}}_1 = \frac{1}{n-1}\sum_{t=1}^{n-1}\hat{\boldsymbol{e}}_{t+1|t}\hat{\boldsymbol{e}}'_{t+1|t},$$

- $\hat{\boldsymbol{e}}_{t+1|t}$: 1-step ahead in-sample base forecast errors

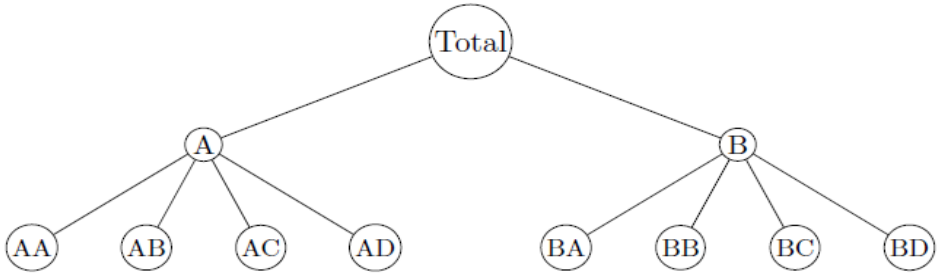
Shrinkage estimator

$$\hat{\boldsymbol{W}}^*_{1,D} = \lambda_D\hat{\boldsymbol{W}}_{1,D} + (1-\lambda_D)\hat{\boldsymbol{W}}_1,$$

- $\hat{\boldsymbol{W}}_{1,D}$: $\text{diag}(\hat{\boldsymbol{W}}_1)$
- λ_D : shrinkage intensity parameter

Outline

Simulation - Design 1



Bottom level series are generated from ARIMA(p, d, q) process.

- $p \in \{1, 2\}$ with equal prob. and $d, q \in \{0, 1, 2\}$ with equal prob.
- Parameters are chosen randomly from uniform distribution satisfying stationarity & invertibility conditions.

Parameter space of AR and MA components

AR component			MA component		
p	Coef	Parameter space	q	Coef	Parameter space
1	ϕ_1	$[0.5, 0.85]$	1	θ_1	$[0.1, 0.3]$
2	ϕ_1	$[\phi_2 - 1, 1 - \phi_2]$	2	θ_1	$[-(1+\theta_2), 1+\theta_2]$
	ϕ_2	$[0.5, 0.85]$		θ_2	$[0.1, 0.3]$

Simulation - Design 1

- ARIMA processes have contemporaneous error correlation

$$\begin{bmatrix} 5 & 3 & 2 & 1 & 0 & 0 & 0 & 0 \\ 3 & 5 & 2 & 1 & 0 & 0 & 0 & 0 \\ 2 & 2 & 6 & 3 & 0 & 0 & 0 & 0 \\ 1 & 1 & 3 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 & 5 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 6 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 & 3 & 6 \end{bmatrix}.$$

- 100 obs. generated (training=90, testing=10).
- Summed appropriately to get data for higher levels.
- Using training set, ARIMA model is fitted by minimizing AICc (using auto.arima function from forecast package in R).
- 1-10 steps ahead base forecasts are obtained.
- Process is repeated 1000 times.

Results of simulation design 1

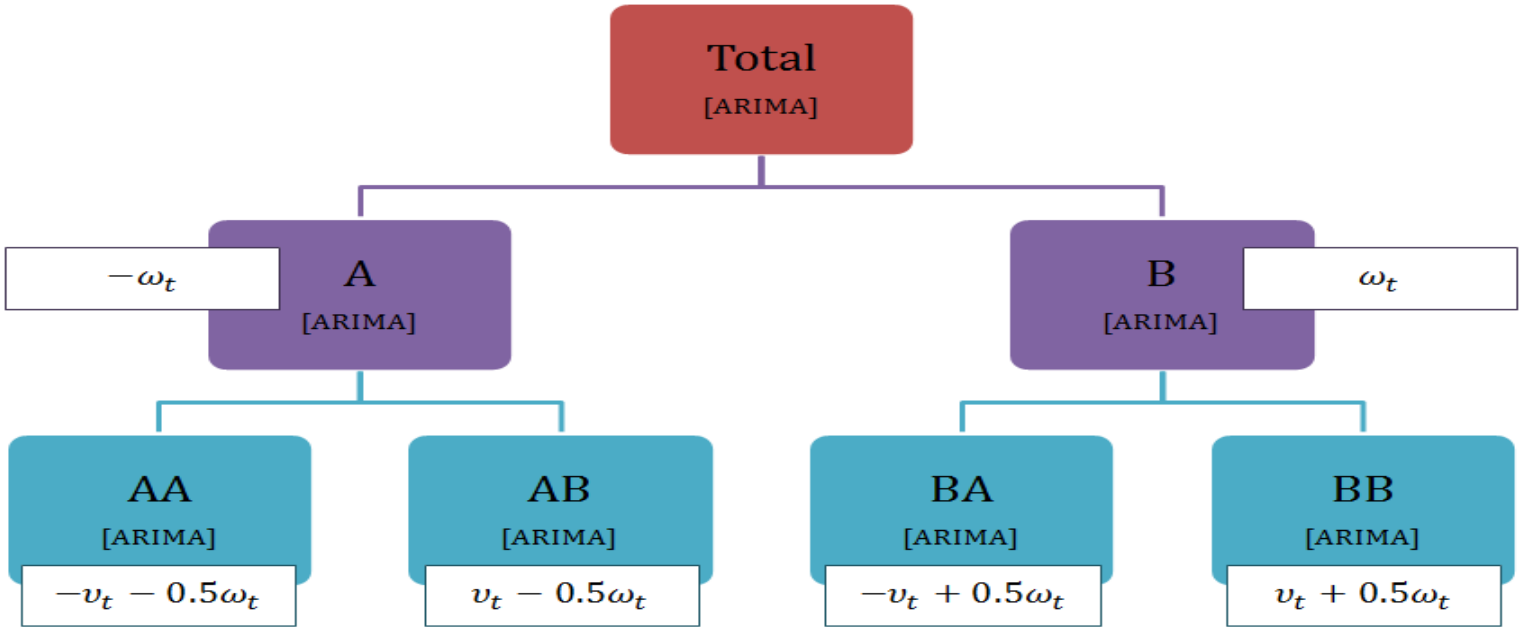
Out-of-sample forecast performance of design 1

	Mean Absolute Error (MAE)					
Series	Bottom-up	OLS	WLS	MinT(Sample)	MinT(Shrink)	Base
Total	96.288	91.121	78.601	77.818	**77.419**	94.867
A	59.403	64.551	52.681	52.491	**52.097**	65.913
B	63.366	62.323	51.610	51.473	**51.094**	59.851
AA	19.400	23.596	18.028	18.189	**17.905**	19.400
AB	16.102	20.577	15.506	15.894	**15.490**	16.102
AC	17.157	21.737	16.869	17.039	**16.702**	17.157
AD	22.750	26.240	18.973	18.946	**18.695**	22.750
BA	**16.458**	21.899	16.596	16.828	16.546	**16.458**
BB	17.109	21.642	**15.624**	15.936	15.656	17.109
BC	17.886	23.206	17.737	17.555	**17.281**	17.886
BD	28.291	29.250	18.635	18.764	**18.457**	28.291
Average	34.019	36.922	29.169	29.176	**28.849**	34.162

p-values for paired t-test between MinT(Shrink) and other methods for each level

	p-value				
Level	Bottom-up	OLS	WLS	MinT(Sample)	Base
0	**0.0403**	**0.0000**	**0.0127**	0.1190	**0.0000**
1	**0.0345**	**0.0000**	**0.0094**	**0.0034**	**0.0000**
2	**0.0478**	**0.0004**	**0.0006**	**0.0000**	**0.0478**

Simulation - Design 2



- $\tau_{AA,t}, \tau_{AB,t}, \tau_{BA,t}, \tau_{BB,t} \sim \mathcal{N}(0, \sigma_0^2)$, $\nu_t \sim \mathcal{N}(0, \sigma_1^2)$, $\omega_t \sim \mathcal{N}(0, \sigma_2^2)$ are independent of time and independent of each other.

- To ensure that aggregate data are much smoother than disaggregated data:

$$2\sigma_0^2 \leq \sigma_2^2 \leq \frac{4}{3}(\sigma_1^2 - \sigma_0^2).$$

Set $\sigma_0^2 = 1, \sigma_1^2 = 10$ and $\sigma_2^2 = 6$.

Results of simulation design 2

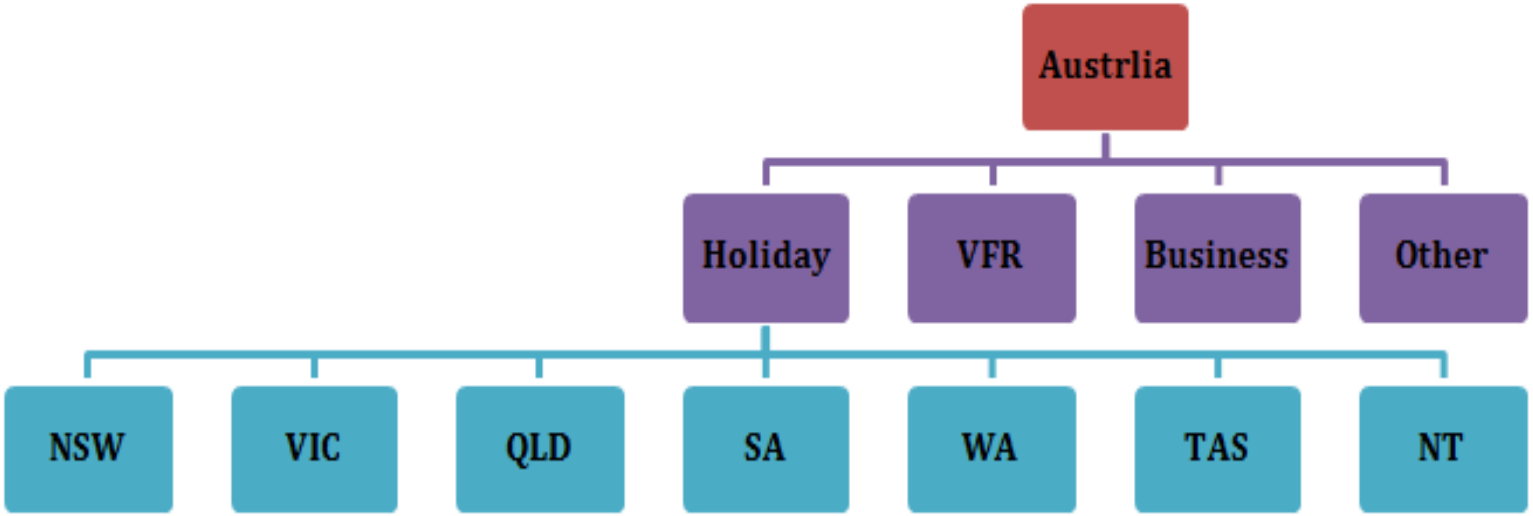
Out-of-sample forecast performance of design 2

	Mean Absolute Error (MAE)					
Series	Bottom-up	OLS	WLS	MinT(Sample)	MinT(Shrink)	Base
Total	26.986	22.759	21.477	21.477	**21.442**	23.484
A	17.640	16.041	15.222	15.396	**15.220**	16.599
B	17.386	16.211	15.167	15.058	**14.962**	16.779
AA	11.947	12.192	11.336	11.432	**11.131**	11.947
AB	11.585	11.709	11.046	11.117	**10.785**	11.585
BA	10.726	11.587	10.548	10.668	**10.441**	10.726
BB	11.924	12.102	10.886	10.924	**10.693**	11.924
Average	15.456	14.657	13.669	13.725	**13.525**	14.721

p-values for paired t-test between MinT(Shrink) and other methods for each level

	p-value				
Level	Bottom-up	OLS	WLS	MinT(Sample)	Base
0	**0.0000**	**0.0401**	0.7271	0.5197	0.0586
1	**0.0000**	**0.0056**	0.1665	**0.0005**	**0.0001**
2	**0.0001**	**0.0000**	**0.0033**	**0.0000**	**0.0001**

Application to Australian domestic tourism



Structure of the hierarchy

Level	Number of series	Total series per level
Australia	1	1
Purpose of Travel	4	4
Geographical region	7	28

Source: National visitor survey managed by Tourism Research Australia

- Observations from January, 2003 to December, 2013.

Procedure

1. Data are divided into two: 100 obs for training and 32 obs for testing.
2. Using training data, ARIMA and ETS models are fitted by minimizing the AICc using the forecast package.
3. 1-6 step ahead base forecasts are obtained for each series in the hierarchy.

This was continued using a rolling origin forecasting approach.

- For each forecast horizon reconciled forecasts are obtained for bottom-up, top-down, OLS and approximated WLS and MinT (MinT(Sample) and MinT(Shrink)).

Results of ARIMA

Out-of-sample forecast performance from ARIMA

	Forecast horizon (h)						
MAE	1	2	3	4	5	6	Average
			Top level				
Base	337.2	372.5	380.8	386.8	423.9	419.6	386.8
Bottom-up	308.8	313.2	292.0	**252.0**	283.1	301.1	291.7
Top-down FP	337.2	372.5	380.8	386.8	423.9	419.6	386.8
OLS	318.6	336.6	325.3	332.1	366.5	357.1	339.4
WLS	301.1	287.7	**278.4**	**252.0**	268.8	**280.8**	278.1
MinT(Sample)	310.3	299.5	298.3	281.8	281.3	307.7	296.5
MinT(Shrink)	**298.6**	**282.5**	282.2	252.1	**266.7**	285.1	**277.9**
			Level 1				
Base	121.0	114.7	116.1	114.7	114.3	113.2	115.7
Bottom-up	123.4	125.4	120.4	115.1	120.2	121.2	121.0
Top-down FP	128.7	134.8	130.1	131.4	140.1	140.5	134.3
OLS	125.9	127.4	123.6	120.1	126.2	130.9	125.7
WLS	118.2	117.0	**112.9**	109.2	114.1	**112.3**	114.0
MinT(Sample)	119.5	114.6	115.6	110.2	**110.6**	116.2	114.5
MinT(Shrink)	**117.4**	**114.0**	113.2	**108.2**	111.0	113.6	**112.9**
			Level 2				
Base	29.0	29.3	28.9	28.7	28.9	29.4	29.0
Bottom-up	29.0	29.3	28.9	28.7	28.9	29.4	29.0
Top-down FP	29.5	29.9	29.8	30.0	30.7	31.0	30.1
OLS	30.2	30.6	30.6	29.9	30.7	31.7	30.6
WLS	28.5	28.5	28.2	28.2	28.4	28.8	28.4
MinT(Sample)	28.8	28.4	28.7	27.9	**27.8**	**28.4**	28.3
MinT(Shrink)	**28.3**	**28.1**	**28.0**	**27.7**	**27.8**	**28.4**	**28.1**

Results of ETS

Out-of-sample forecast performance from ETS

	Forecast horizon (h)						
MAE	1	2	3	4	5	6	Average
			Top level				
Base	266.7	253.4	**244.5**	**205.2**	**214.7**	220.0	234.1
Bottom-up	**254.1**	254.7	258.7	214.9	224.3	240.6	241.2
Top-down FP	266.7	253.4	**244.5**	**205.2**	**214.7**	220.0	234.1
OLS	263.6	249.7	245.8	205.9	215.1	**219.3**	**233.2**
WLS	254.6	248.5	250.3	207.1	216.6	229.1	234.4
MinT(Sample)	276.4	250.6	270.3	233.4	225.9	227.1	247.3
MinT(Shrink)	259.3	**243.9**	248.8	208.6	215.9	224.0	233.4
			Level 1				
Base	105.2	105.6	107.4	101.8	102.4	104.6	104.5
Bottom-up	107.1	105.3	106.3	99.2	102.0	102.1	103.7
Top-down FP	**104.1**	105.6	**104.5**	99.8	102.2	102.8	103.2
OLS	104.3	105.2	105.4	99.7	101.5	102.2	103.0
WLS	105.4	104.4	105.7	**99.0**	101.2	101.9	102.9
MinT(Sample)	110.4	106.2	109.9	101.6	103.4	102.8	105.7
MinT(Shrink)	105.2	**103.7**	105.6	99.2	**100.6**	**101.0**	**102.5**
			Level 2				
Base	25.5	25.4	25.6	25.1	25.8	25.6	25.5
Bottom-up	25.5	25.4	25.6	25.1	25.8	25.6	25.5
Top-down FP	25.5	25.4	**25.5**	**25.0**	25.8	25.6	25.5
OLS	25.7	25.6	25.7	25.3	25.9	25.7	25.6
WLS	**25.4**	25.3	**25.5**	**25.0**	25.6	25.5	25.4
MinT(Sample)	25.9	25.4	25.9	25.2	25.6	25.6	25.6
MinT(Shrink)	**25.4**	**25.2**	**25.5**	**25.0**	**25.5**	**25.4**	**25.3**

Conclusions

- MinT uses information about the correlation structure of the hierarchy.
- Simulation and empirical results illustrated the better performance of MinT over others.
- However, needs a good estimate of the covariance matrix to maintain the gains made.

THANK YOU!!

Email: shanika.wickramasuriya@monash.edu