A hierarchical approach to forecasting Scandinavian unemployment

Nikolaos Kourentzes[a,b]

Rickard Sandberg[b]

[a]Lancaster University; [b]Stock School of Economics

International Symposium on Forecasting 2017 - Cairns

Marketing Analytics
and Forecasting

LUMS

STOCKHOLM SCHOOL OF ECONOMICS

Forecasting
Society

Agenda

1. Motivation

2. The case study

3. How to solve it

4. Results

5. Findings

CMAF

The problem

- Time series that exhibit relationships are typically modelled as a VAR system.

- VAR (and more complete forms) are very cumbersome models to specify and estimate.

- Typically we end up using low dimensionality and order models so as to be able to build them and use them:

  - get insights from them apart from forecasts!

  - or use BVAR and other "arcane" specifications – eventually the dimensionality is still small.

- But in reality we may have to model high dimensional systems (many time series) or require higher order VAR specifications.

- We will deal with one such case: modelling Scandinavian unemployment.

  - They have high workforce mobility due to geographical closeness, legislation, cultural and linguistic proximity.

 - The economies of these countries are very interconnected.

The data

- **Sixteen** unemployment time series across the following dimensions:

  - Age {15-24; 25 and above}

  - Country {Denmark; Finland; Norway; Sweden}

  - Gender {Female; Male}

- Monthly data with 312 observations (Jan 1989 – Dec 2014).

- From these we can construct multiple hierarchies, resulting in 29 unique aggregate series (16 + 29 = 45 series in total).

	Top Level	Level 1	Level 2	Level 3
Hierarchy 1	Total	Country	Gender	Age
Hierarchy 2	Total	Country	Age	Gender
Hierarchy 3	Total	Gender	Country	Age
Hierarchy 4	Total	Gender	Age	Country
Hierarchy 5	Total	Age	Country	Gender
Hierarchy 6	Total	Age	Gender	Country

CMAF

The data

Illustration of Hierarchy 1: Total → Country → Gender → Age
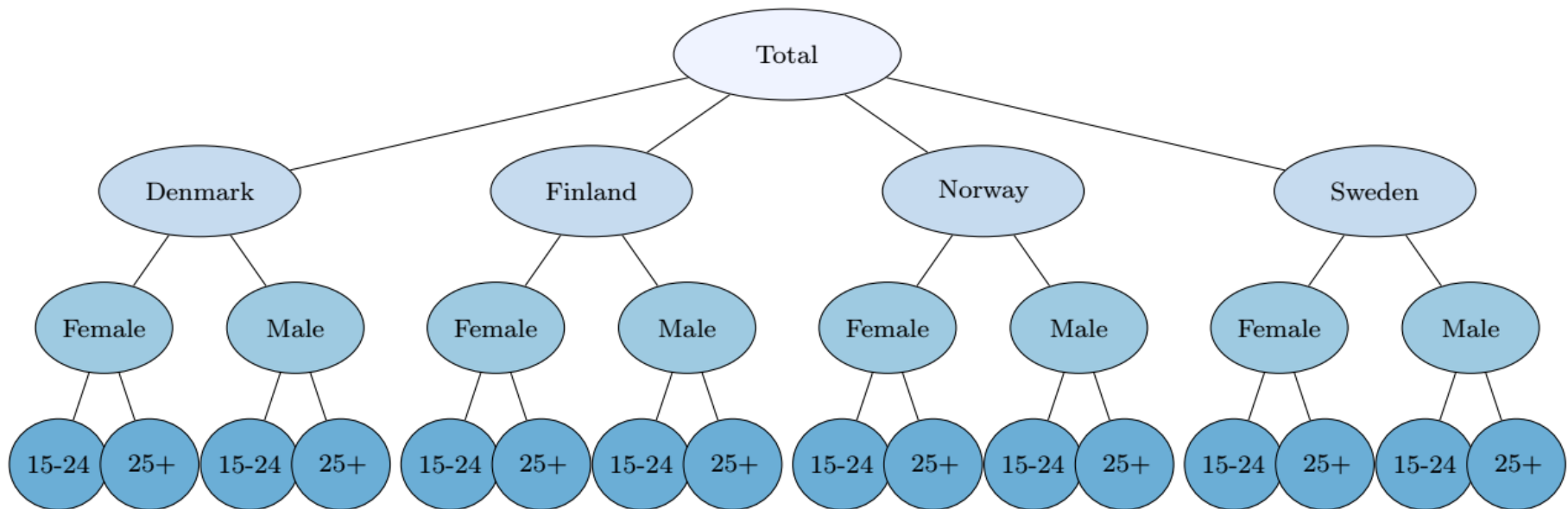
The data

Illustration of Hierarchy 2
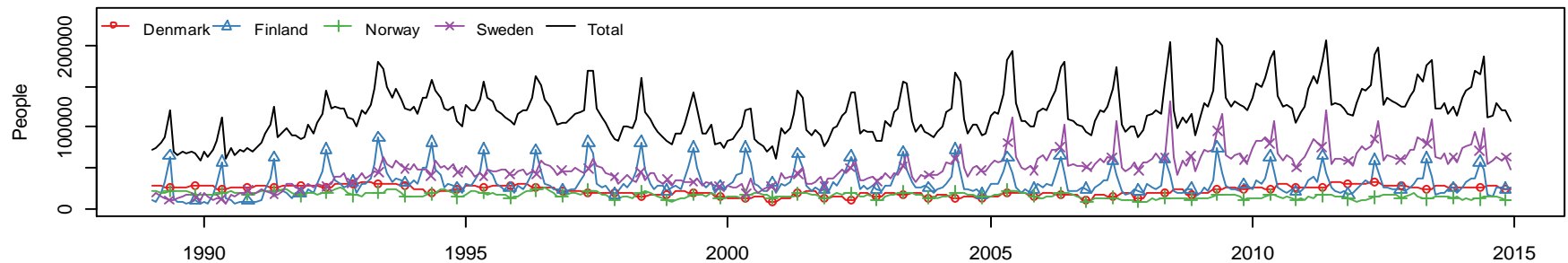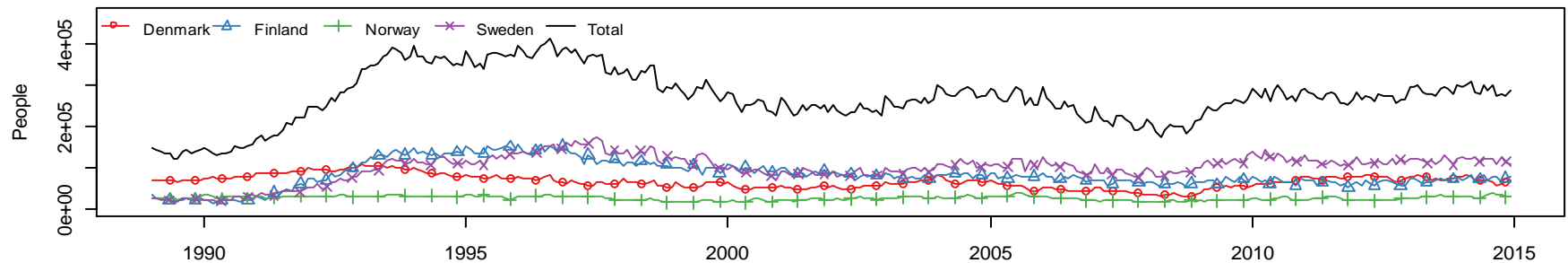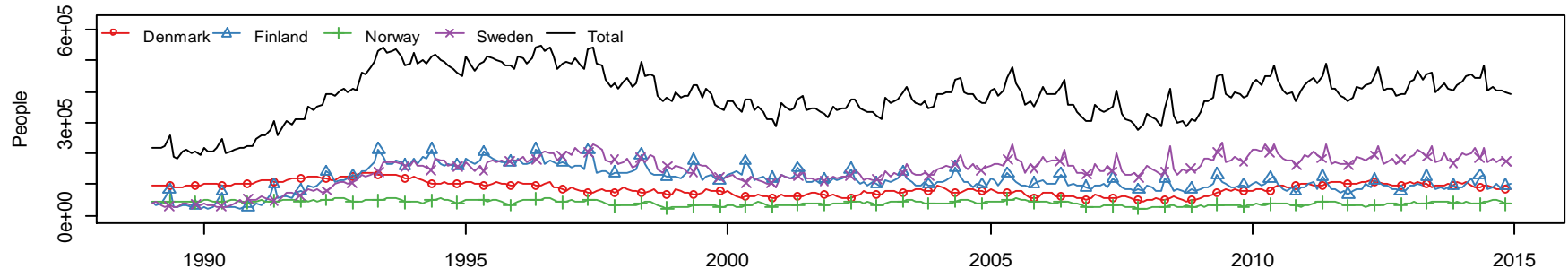
The idea

- VAR model imposes explicit connections through lagged inputs, e.g.:

$$Y_{1t} = \alpha_{10} + \alpha_{11} Y_{1t-1} + \alpha_{12} Y_{2t-1} + \varepsilon_{1t}$$
$$Y_{2t} = \alpha_{20} + \alpha_{21} Y_{2t-1} + \alpha_{22} Y_{1t-1} + \varepsilon_{2t}$$

- An alternative would be to consider that any hierarchy implies an aggregation consistency constraint. Any lower level time series must aggregate to the higher level, implying the following:

$$Y_t = \beta_1 Y_{1t} + \beta_2 Y_{2t} + \tilde{\varepsilon}_t$$

  - The aggregation consistency constraint allows to connect time series implicitly. This gives us a different path to model systems of time series.

- Naturally, these two can be combined to enforced both implicit and explicit connections:

$$Y_t = \beta_1 \left(\alpha_{10} + \alpha_{11} Y_{1t-1} + \alpha_{12} Y_{2t-1} + \varepsilon_{1t} \right) +$$
$$\beta_2 \left(\alpha_{20} + \alpha_{21} Y_{2t-1} + \alpha_{22} Y_{1t-1} + \varepsilon_{2t} \right) + \tilde{\varepsilon}_t$$
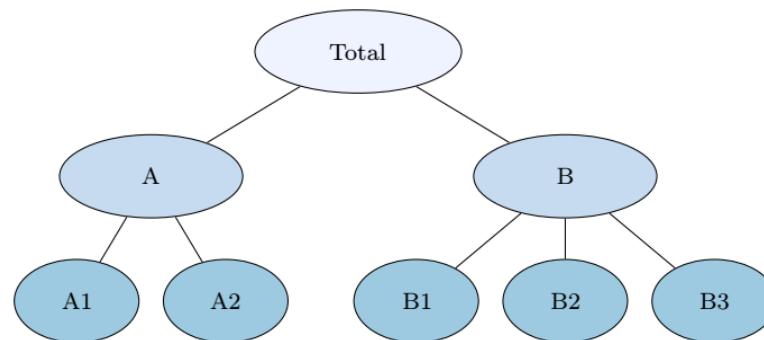
CMAF

Hierarchical forecasting

- Traditionally Bottom-Up and Top-Down (Fliedner, 2001)
    - For our case these are useless, as we cannot enforce aggregate consistency across multiple hierarchies (we have 6 possible aggregation pathways).

- Optimal combination (Hyndman et al., 2011, Athanasopoulos et al., 2009).

$$y_t = Sb_t$$

| All series | Disaggregate series |

| Summing matrix |
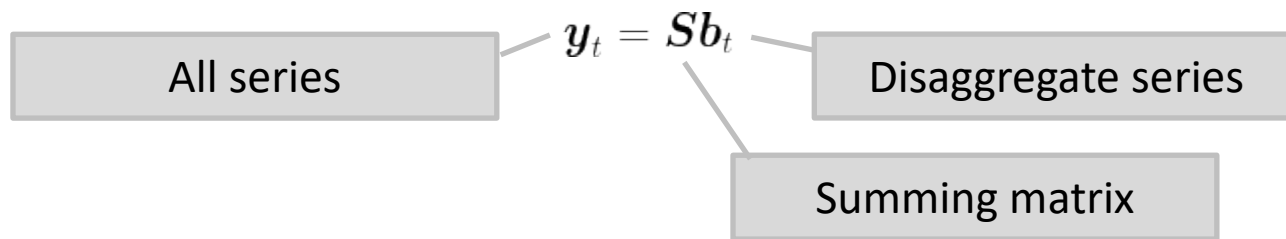
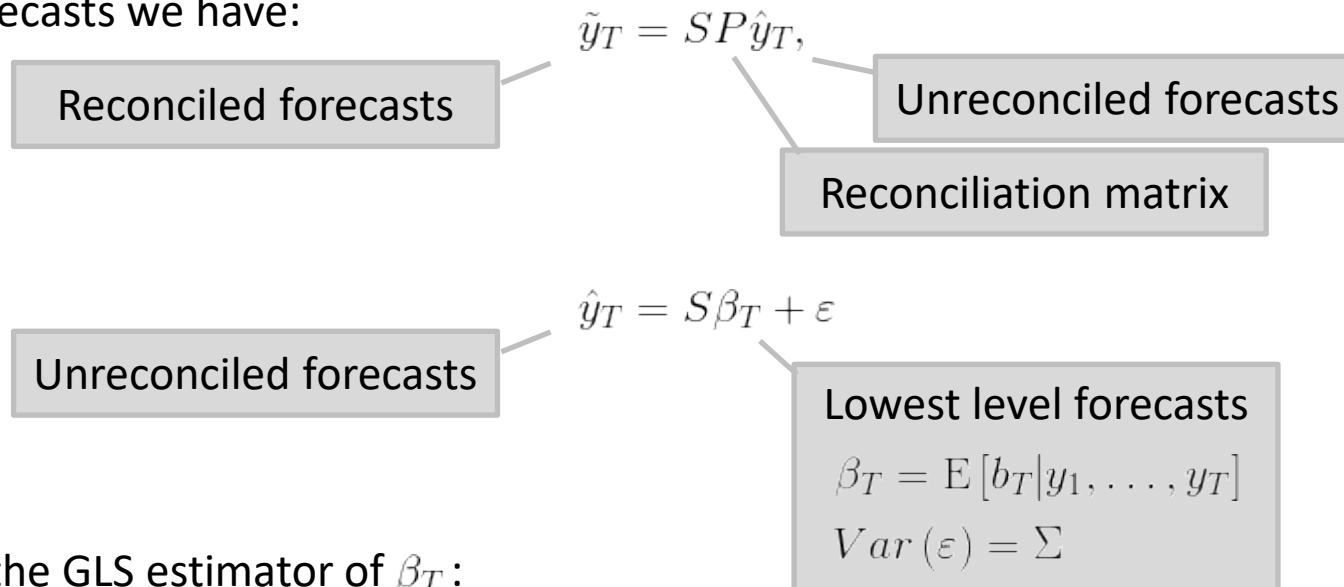To understand the summing matrix let us use an example:

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ & & I_n & & \end{bmatrix}$$

Hierarchical forecasting

For forecasts we have:

$$\tilde{y}_T = SP\hat{y}_T,$$

Reconciled forecasts

Unreconciled forecasts

Reconciliation matrix

$$\hat{y}_T = S\beta_T + \varepsilon$$

Unreconciled forecasts

Lowest level forecasts

$$\beta_T = \mathrm{E}\left[b_T | y_1, \ldots, y_T\right]$$
$$Var\left(\varepsilon\right) = \Sigma$$

Using the GLS estimator of β_T :

$$\tilde{y}_T = S \left(S'\Sigma^{\dagger}S\right)^{-1} S'\Sigma^{\dagger}\hat{y}_T$$

Wickramasuriya et al. (2005) showed that Σ is not identifiable and instead show that W can be used, which is the covariance matric of the forecast errors.

- W can be effectively estimated using the shrinkage estimator by Schafer and Strimmer (2005), but also can be empirical or constrained to zero off-diagonals.

CMAF

The experiment

- Withhold last 120 observations; forecast 12 steps ahead using rolling origin evaluation.

- Models are re-fit (re-specified) at each forecast origin.

- Use AvgRelMAE (Davydenko and Fildes, 2013)

- Four types of forecats:
  - forecast each series independently (benchmark)
  - forecast lowest level using VAR and aggregate
  - use hierarchical forecasting
 - combination of hierarchical forecasting and VAR

$$\text{MAE} = \frac{1}{h} \sum_{i=1}^{h} |y_i - \hat{y}_i|,$$

$$\text{AvgRelMAE} = \sqrt[o]{\prod_{j=1}^{o} \frac{\text{MAE}_{A,j}}{\text{MAE}_{B,j}}}.$$

- Four types of "individual forecasts"
  - AR(p) [SARIMA(p,d,0)(0,D,0); matches VAR; benchmark (Edlund and Karlsson, 1993)]
  - SARIMA(p,d,q)(0,D,0)
  - SARIMA(p,d,q)(P,D,Q)
  - SARIMA(p,d,q)(P,D,Q) + Temporal hierarchies (Athanasopoulos et al., 2017)

Percentage of times of on non-zero coefficient across 109 forecast origins.

Frequency of a term being included in VAR across origins
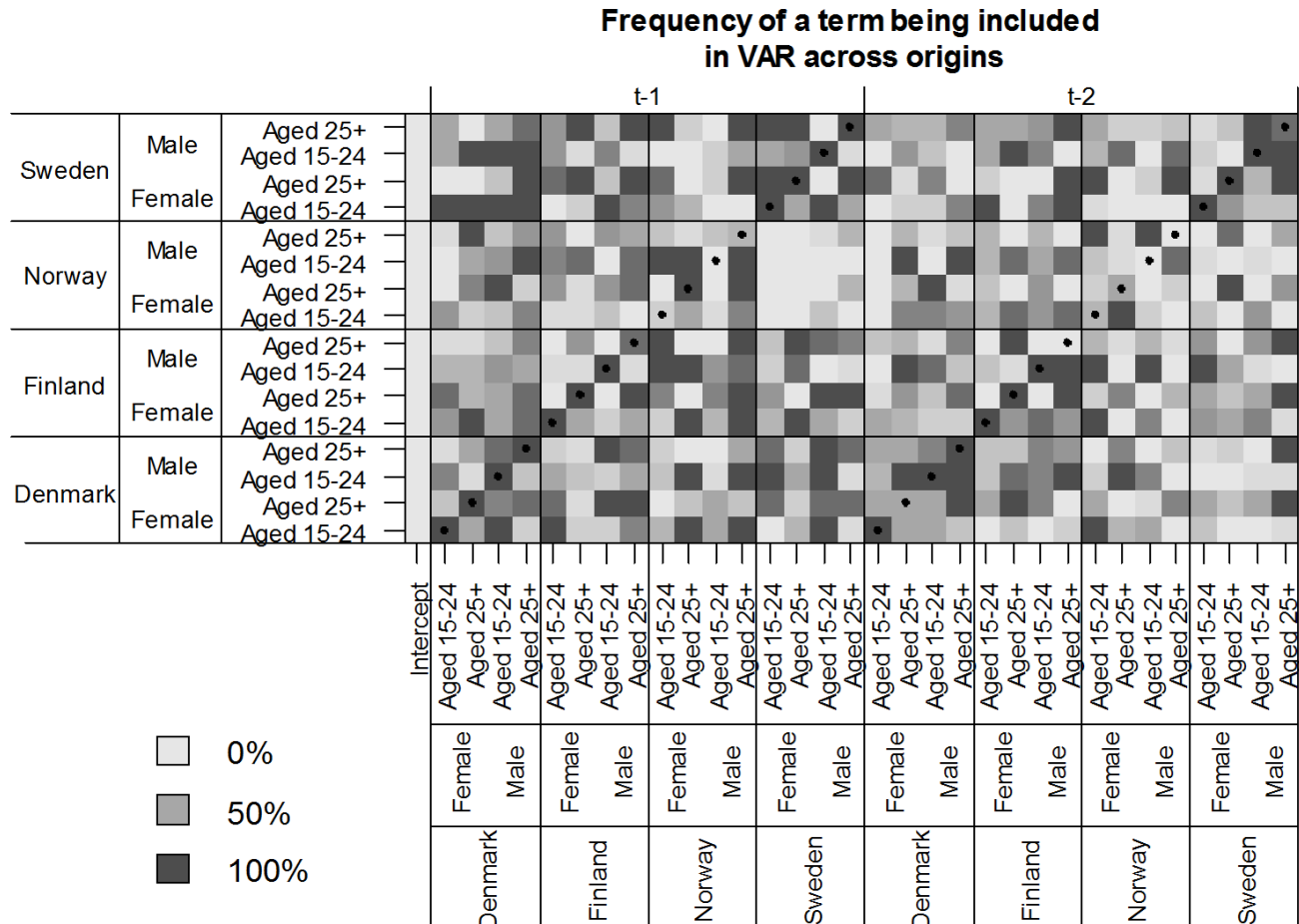
Results

Reconciliation Method	Base Forecast			
	SARIMA (p,d,0)(0,D,0)	SARIMA (p,d,q)(0,D,0)	SARIMA (p,d,q)(P,D,Q)	THieF SARIMA
Bottom Level - Disaggregate Series				
ARIMA	1.000	0.966	0.854	0.795
ARIMA.MSE	0.958	0.926	**0.850**	**0.791**
ARIMA.Empir	0.949	1.076	0.902	0.832
ARIMA.Shrink	0.890	**0.870**	0.854	0.798
VAR.BU	0.978	0.978	0.978	0.978
VAR.Shrink	**0.884**	0.877	0.871	0.831
All Levels				
ARIMA	1.000	0.969	0.805	0.751
ARIMA.MSE	0.916	0.881	**0.783**	**0.737**
ARIMA.Empir	0.901	1.033	0.834	0.770
ARIMA.Shrink	0.840	0.813	0.785	0.744
VAR.BU	0.911	0.911	0.911	0.911
VAR.Shrink	**0.825**	**0.812**	0.789	0.761

☐ < VAR.BU

☐ < VAR.Shrink

CMAF

Findings

- The idea works! Indeed we can effectively model VAR style connections with hierarchical forecasting.

- The combined approach (VAR + Hierarchies) works even better for relatively simply models [AR and SARIMA(p,d,q)(0,D,0)]

- However as in hierarchical forecasting we build models for each series independently, it is trivial to specify more complex models than with VAR and eventually substantially outperform it (19% gains over VAR).

- The best forecasts here are by no means the best models for forecasting unemployment, they still rely solely on unemployment information. However, both VAR and individual forecasts can be enhanced. Yet, our VAR is already 16 x 32, it is not trivial to specify bigger ones!

- Potential for future work on hybrid VAR + Hierarchical.

- Prediction intervals are either empirical or bootstrapped in hierarchical forecasting. This needs to be resolved.

- (Presentation not finished yet!)

Modelling supply chains

- Realistic supply chains are messy to forecast as a system due to their complexity; currently most work done with simulations, and oversimplistic (Trapero et al., 2012).

Upstream

Consumer
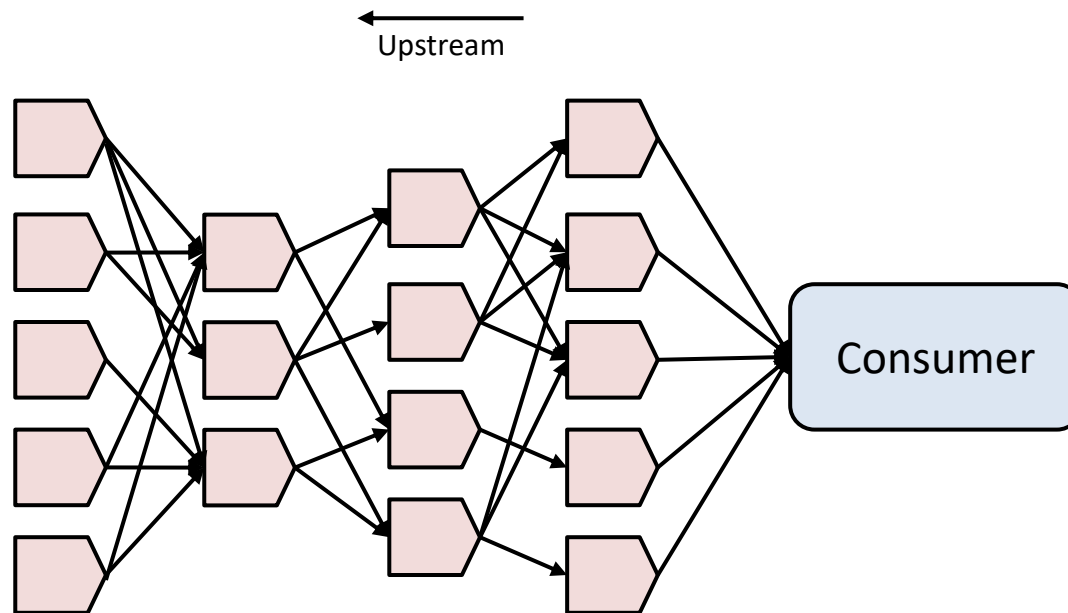
- Not fully connected.
- Multiple layers and multiple actors in each layer.
- Different demand patterns at each level and decision frequency.

Fits to the framework!

Thank you for your attention! Questions?

Working paper available!

Nikolaos Kourentzes (@nkourentz)

email: nikolaos@kourentzes.com

blog: http://nikolaos.kourentzes.com

Marketing Analytics and Forecasting | Lancaster University Management School