Hierarchical time-series decomposition for retail sales forecasting

Andrea Cognolato, Constantin Merbecks

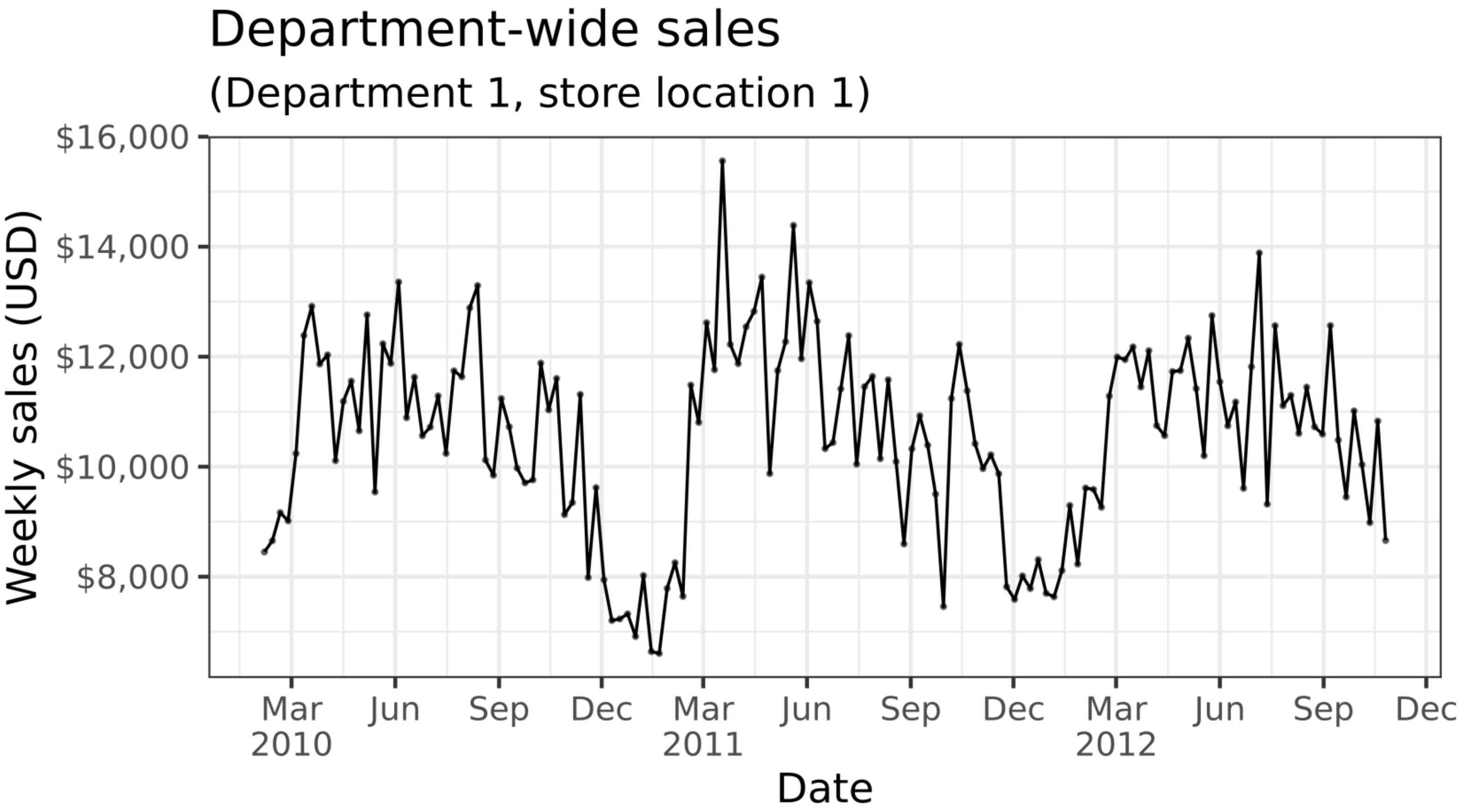
Bayesian Data Analysis, Aalto University

December 2021

Introduction and Motivation

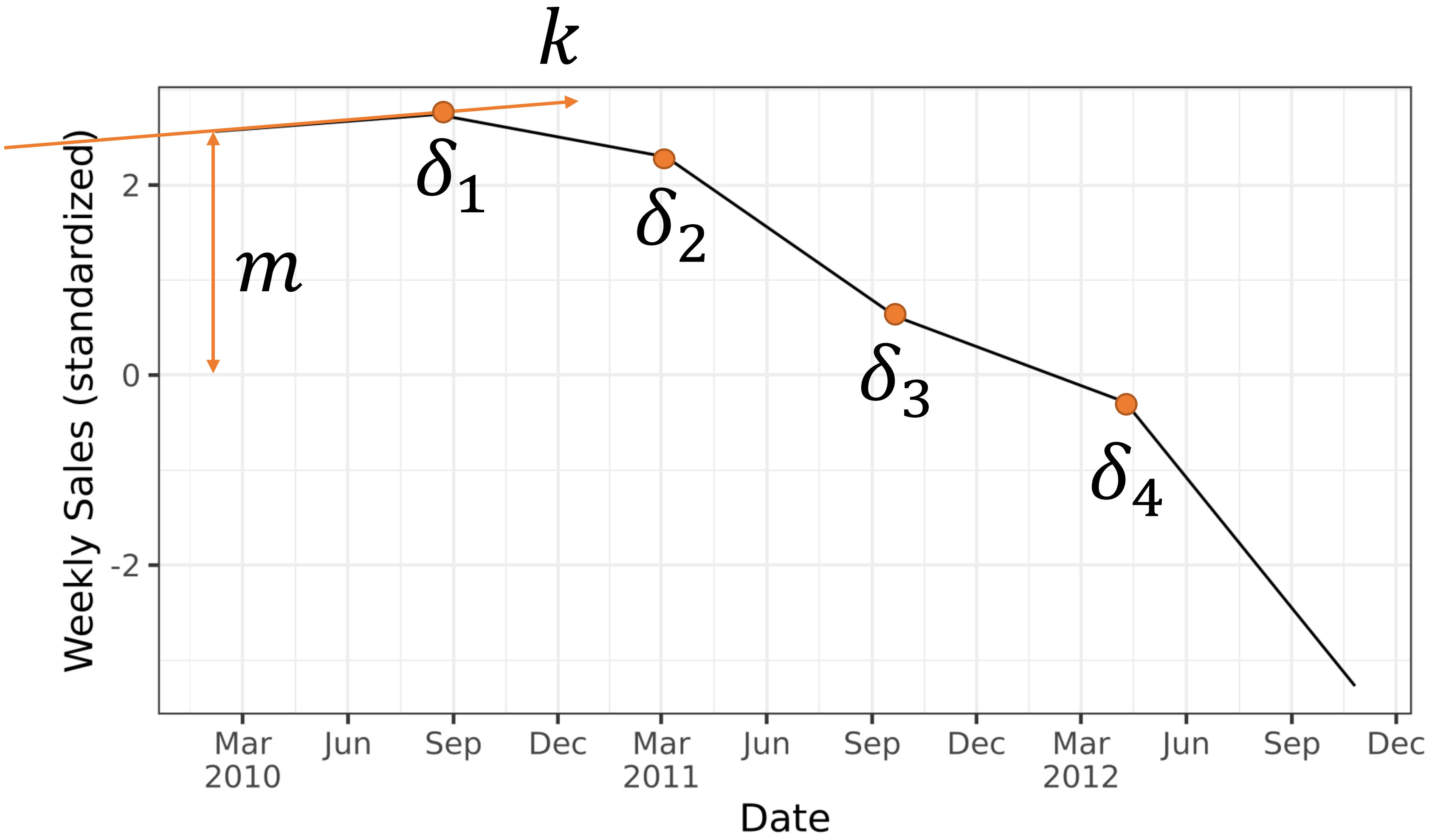
- Interest in time series analysis
- Discovery of Facebook Prophet
  - Built for time series modelling
  - Based on Bayesian statistics using separate model idea
- Goal
  - Rebuild the model and make it hierarchical
- Suitable, interesting data needed:
  - Walmart store sales time series
    - Multiple Stores and multiple departments

Data



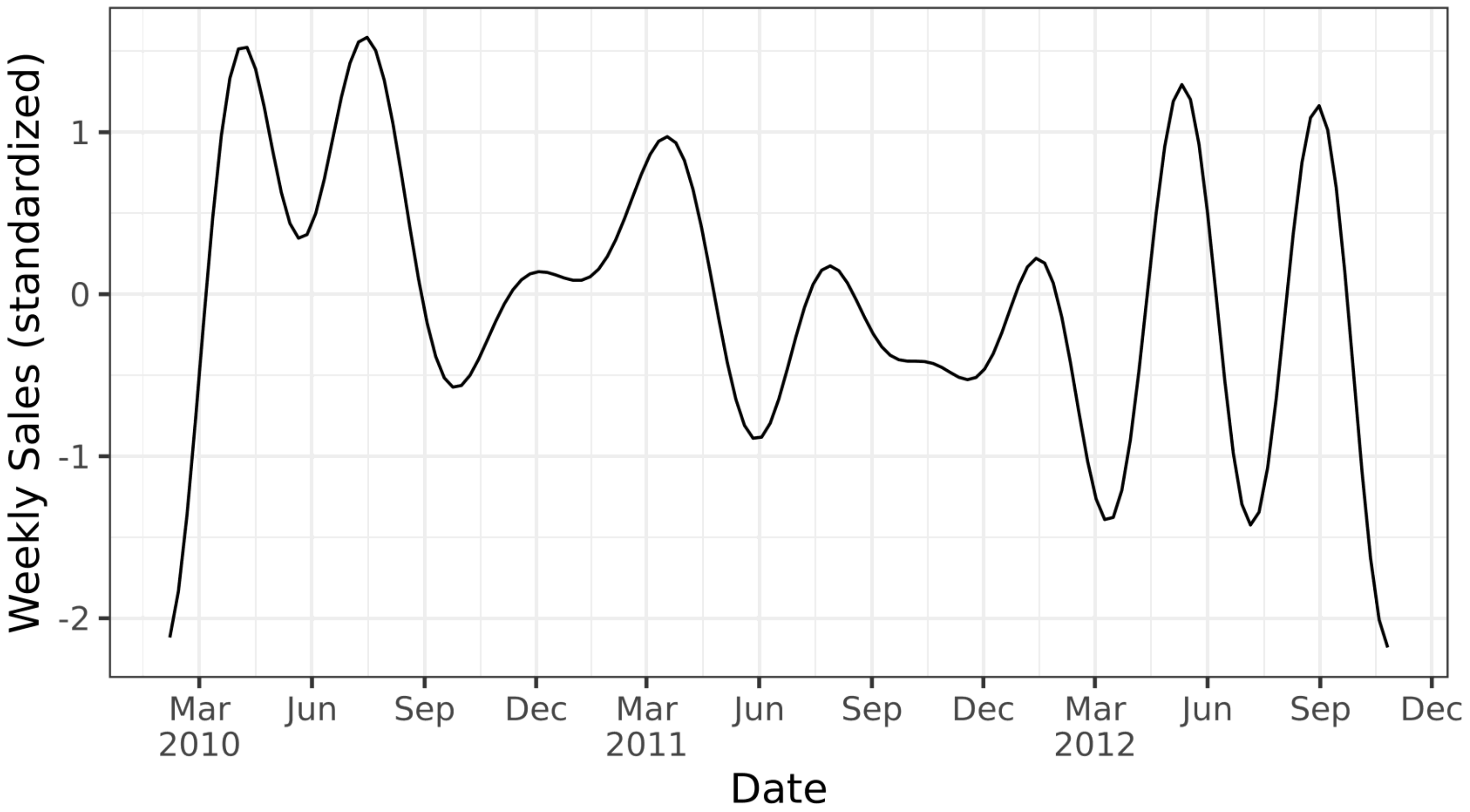
Department-wide sales
(Department 1, store location 1)
$16,000
$14,000
$12,000
$10,000
$8,000
Weekly sales (USD)
Mar
2010
Jun
Sep
Dec
Mar
2011
Jun
Sep
Dec
Mar
2012
Jun
Sep
Dec
Date

Model: trend

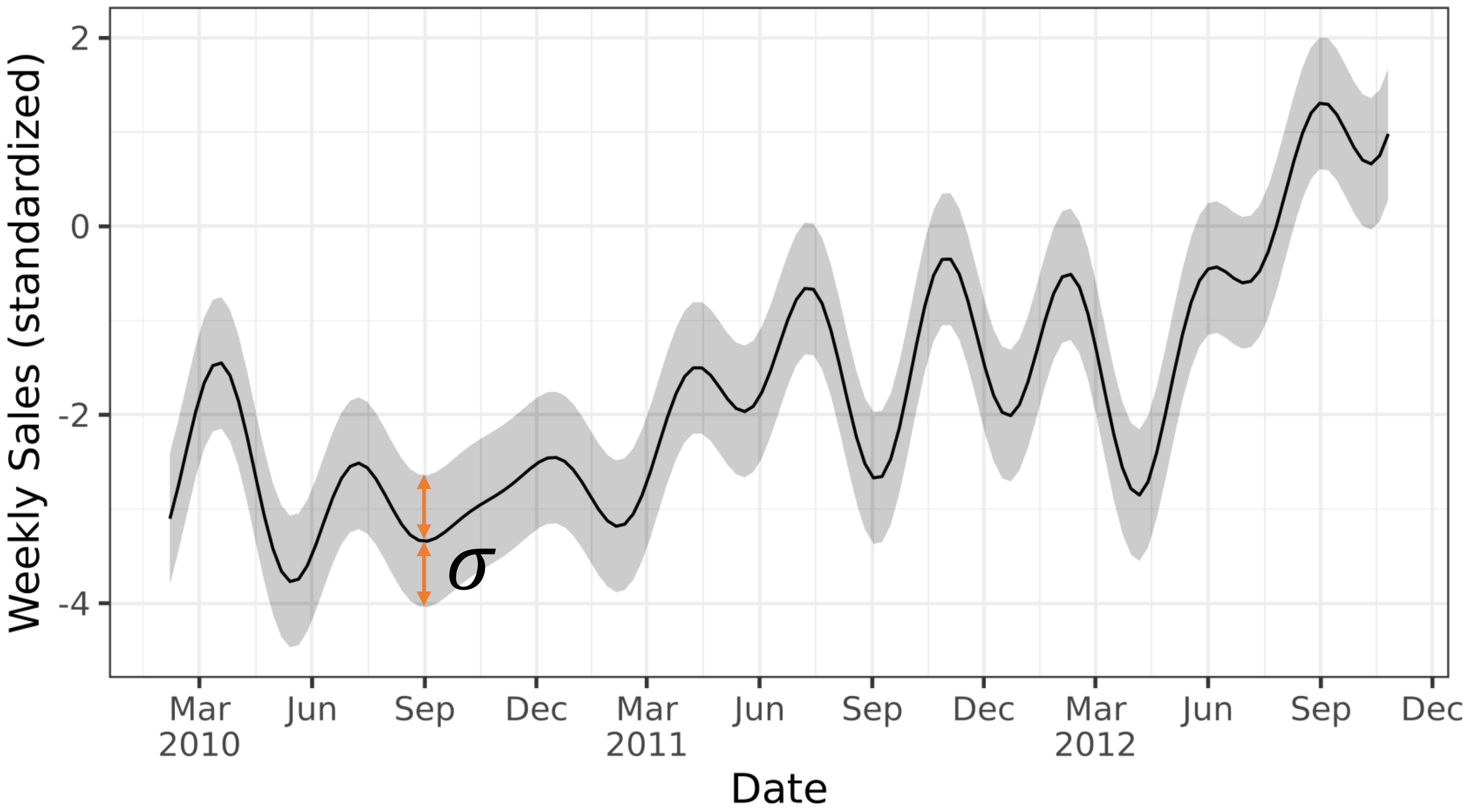


k
m
δ_1
δ_2
δ_3
δ_4
Weekly Sales (standardized)
2
0
-2
Mar 2010
Jun
Sep
Dec
Mar 2011
Jun
Sep
Dec
Mar 2012
Jun
Sep
Dec
Date

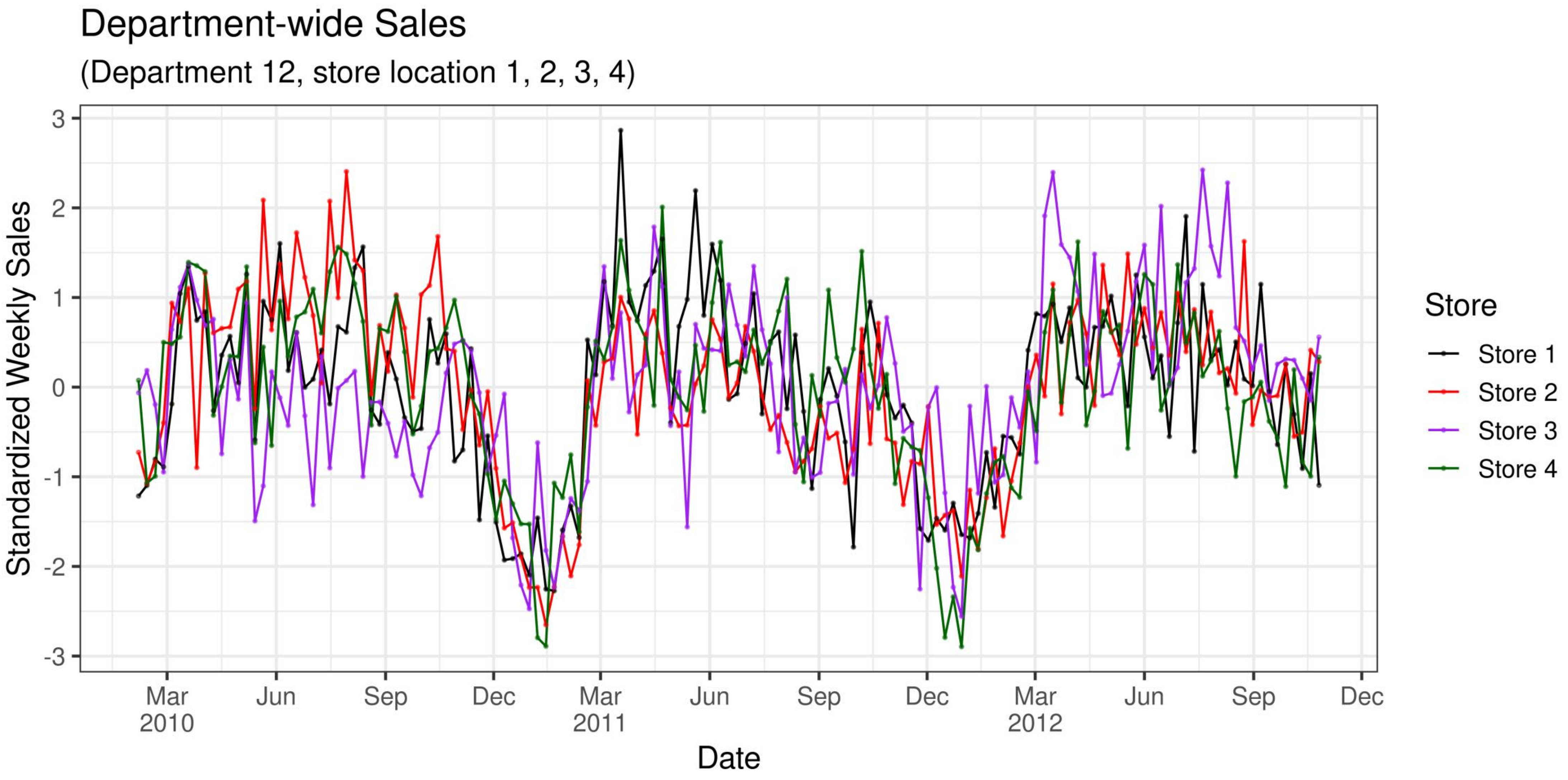
Model: seasonality



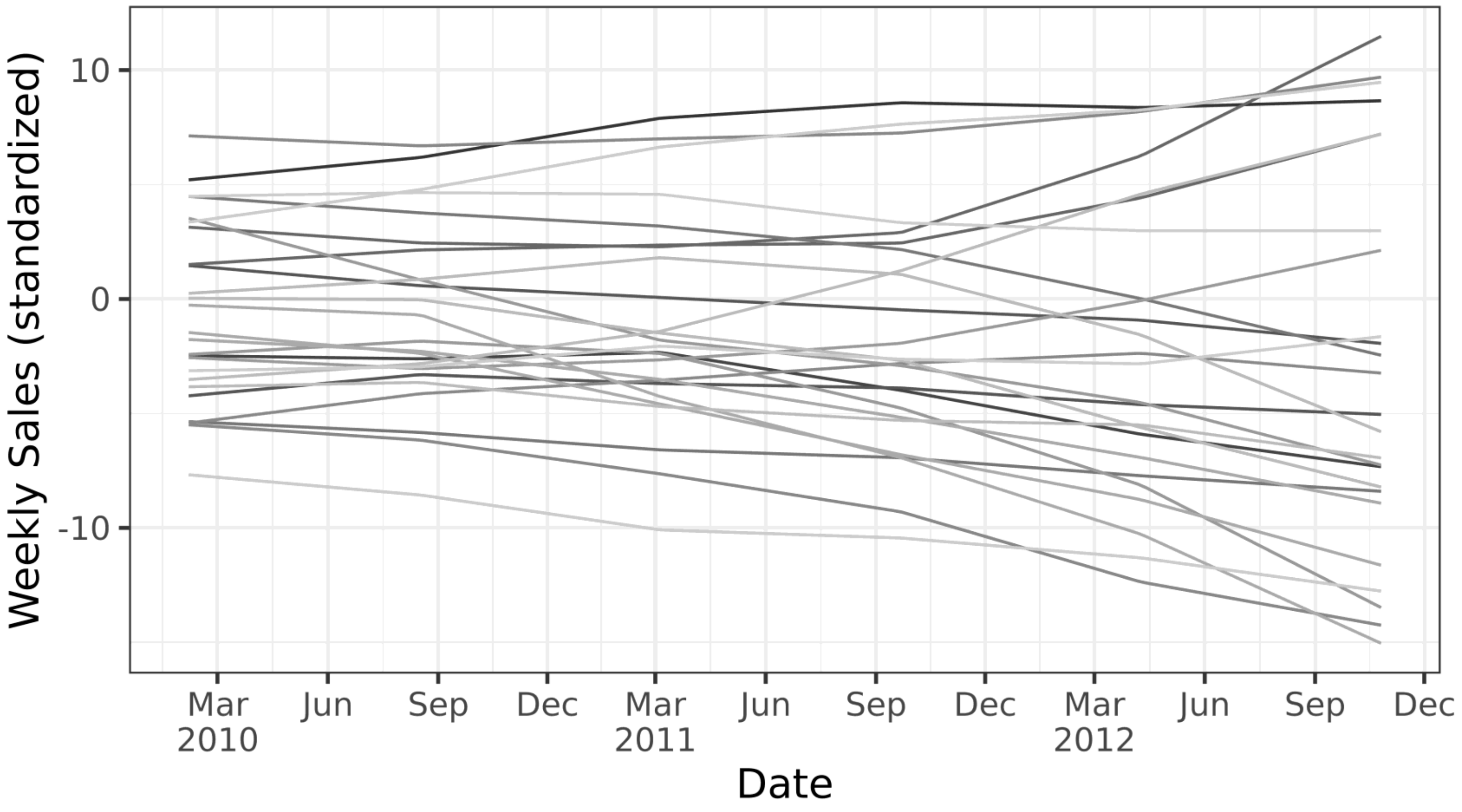
Model: trend, seasonality, residuals



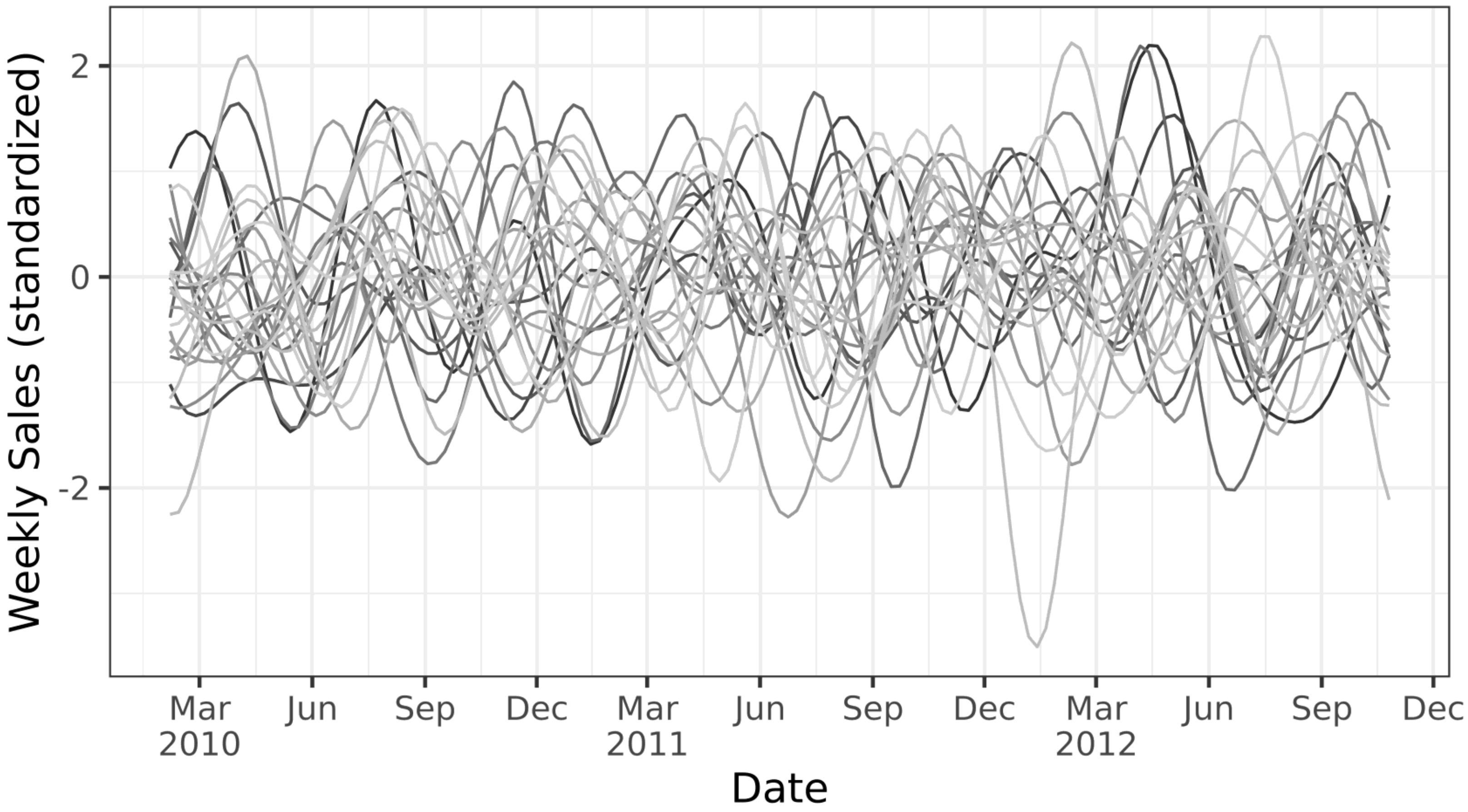
Model: hierarchical structure



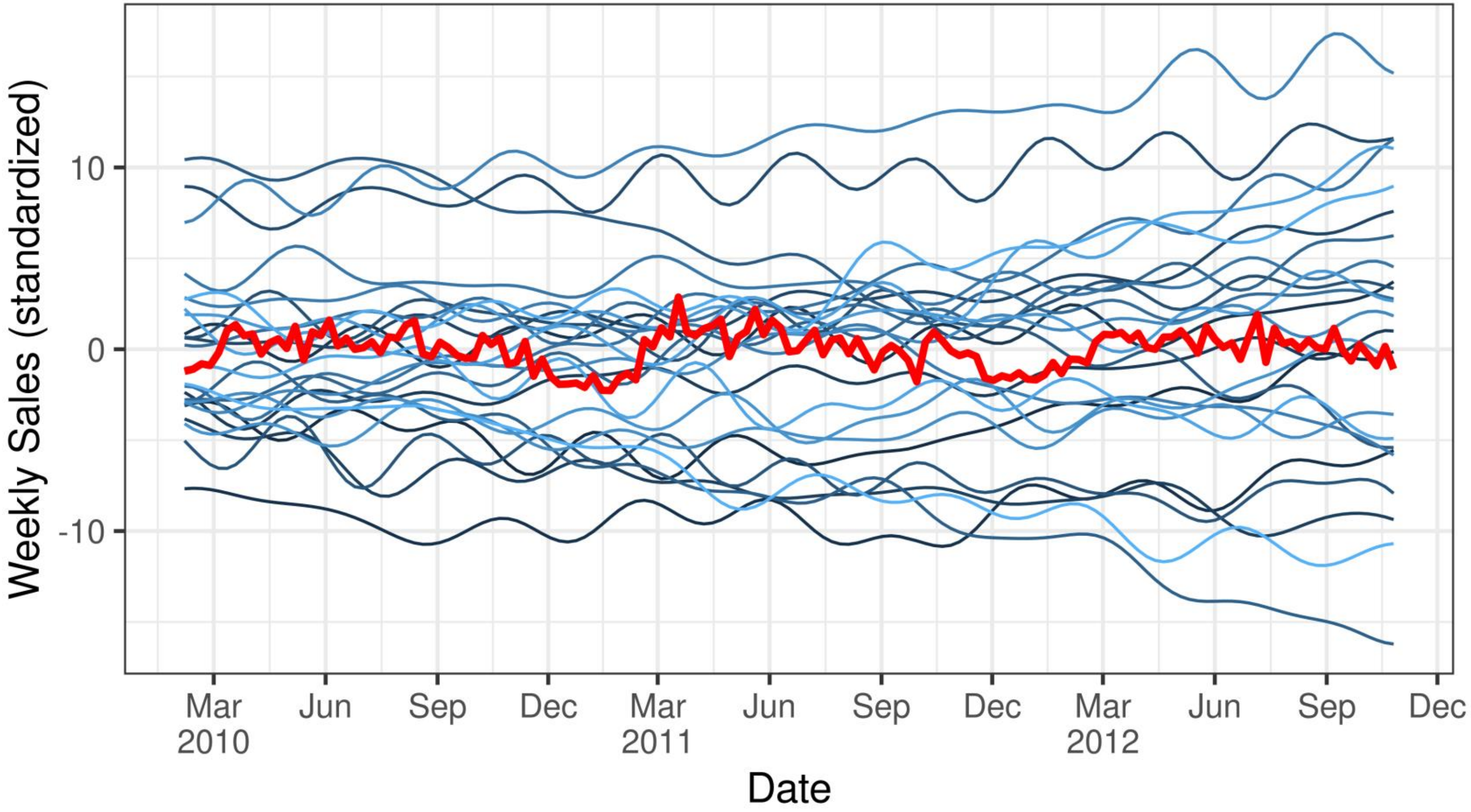
Priors: trend samples



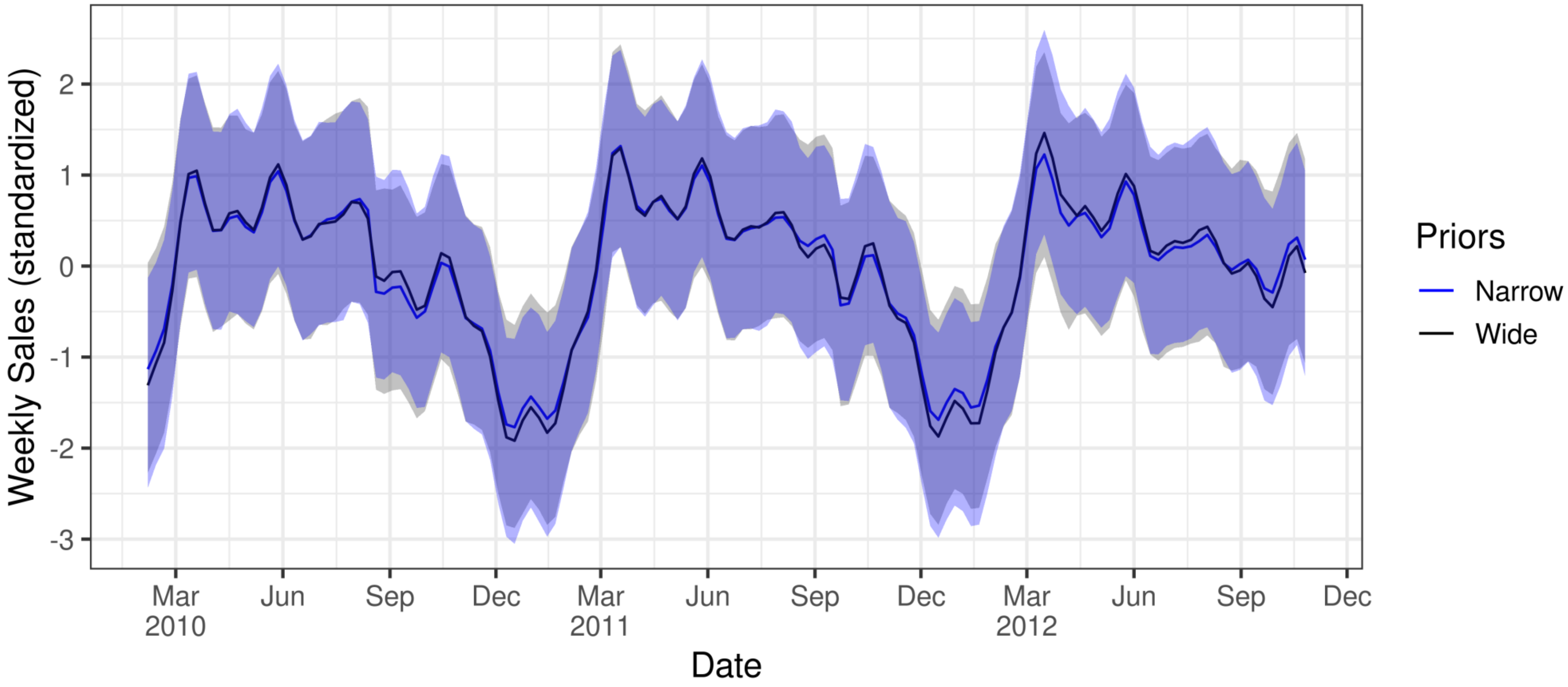
Priors: seasonality samples



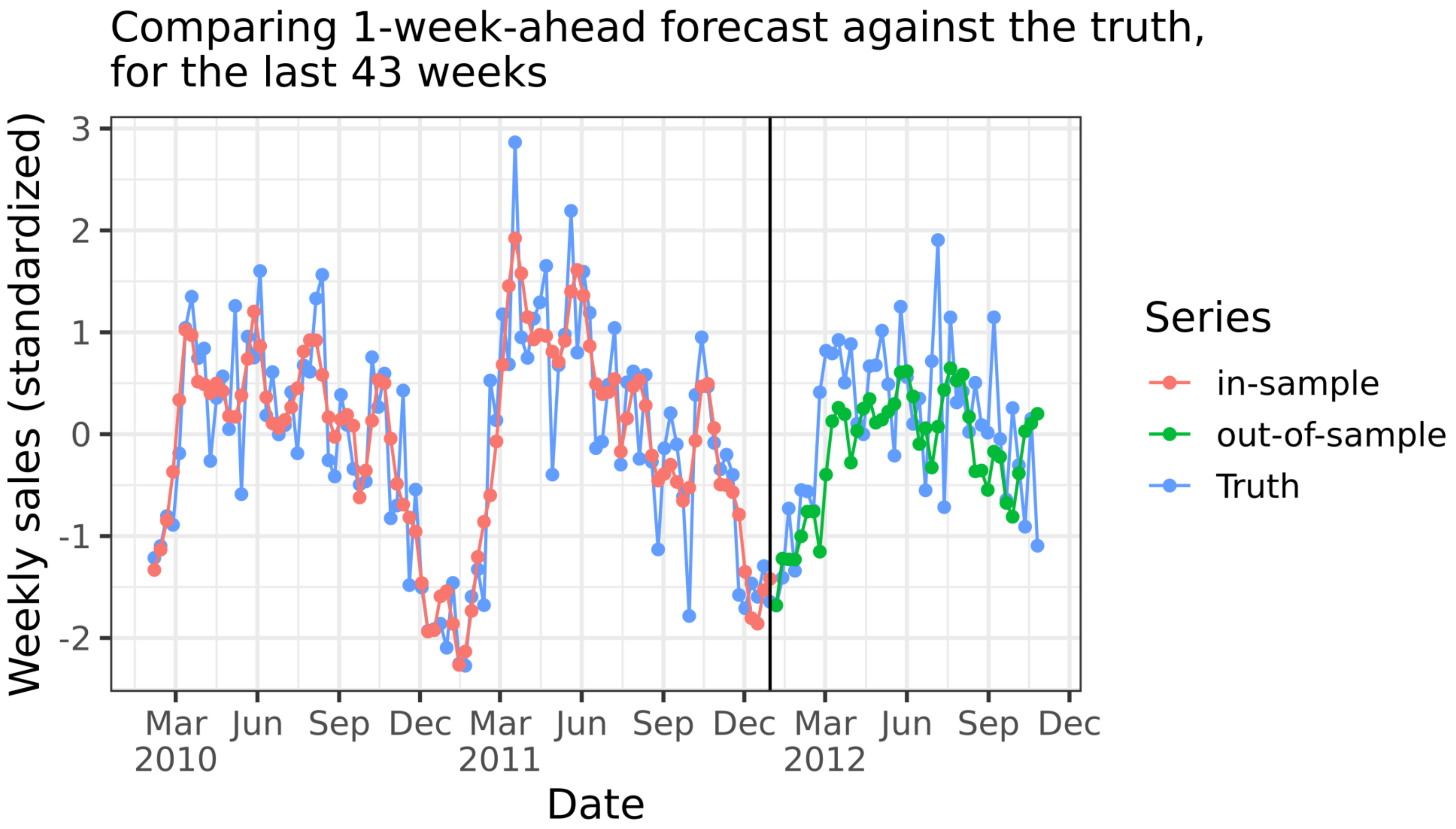
Priors: seasonality + trend samples vs. actuals



Priors: sensitivity analysis



Posterior predictive checking



Model comparison

Expected Log Posterior Density (OOS)

		Store			
Model		1	2	3	4
Separate		-57	-45	-67	-50
Hierarchical		**-48**	**-42**	**-53**	**-43**

Mean Average Error (OOS)

		Store			
Model		1	2	3	4
Separate		0.7	0.5.	0.9	0.6
Hierarchical		**0.5**	**0.5**	**0.7**	**0.5**

- Computed using 1-week-ahead leave-future-out cross validation
- 100 week in-sample training
- 43 week out-of-sample testing

Discussion

- Issues:
  - Runtime and model size
  - Seasonality decomposition not perfectly matching
  - Lower effective sample size for the hierarchical model

- Future Improvements:
  - External Variables: Weather, gas prices, holidays
  - Extending the data sets to more stores

Summary and take-home message

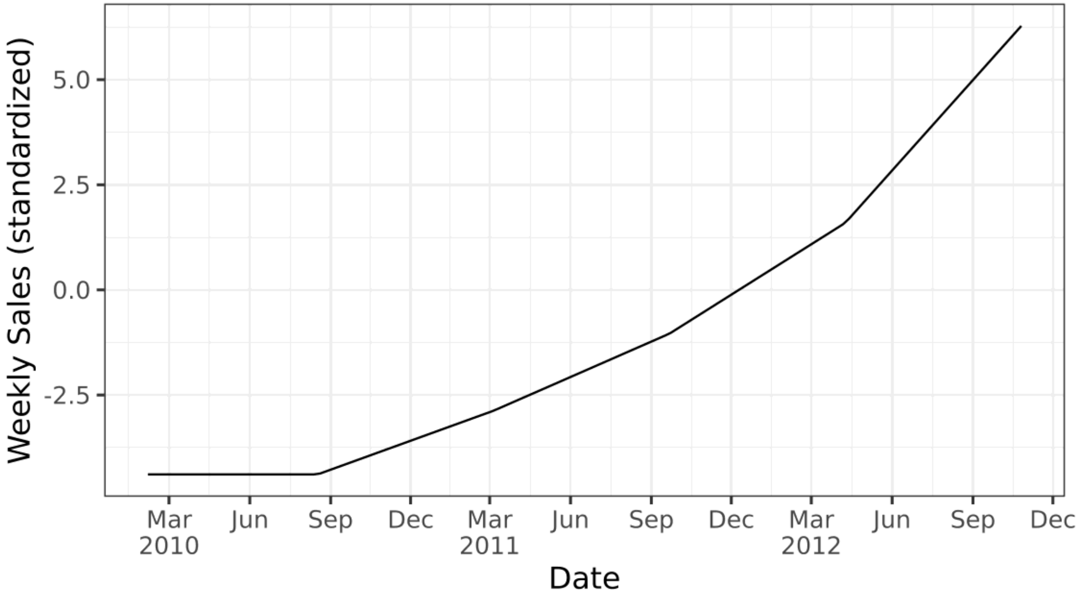
- Additive time series model captures sales dynamics
- Hierarchical model able to generalize and share information across stores

- Contact info:

  - andrea.cognolato@aalto.fi
  - constantin.merbecks@aalto.fi

Appendix

Model: trend component formulation



$$\text{trend}_t = \text{growth}_t + \text{offset}_t$$

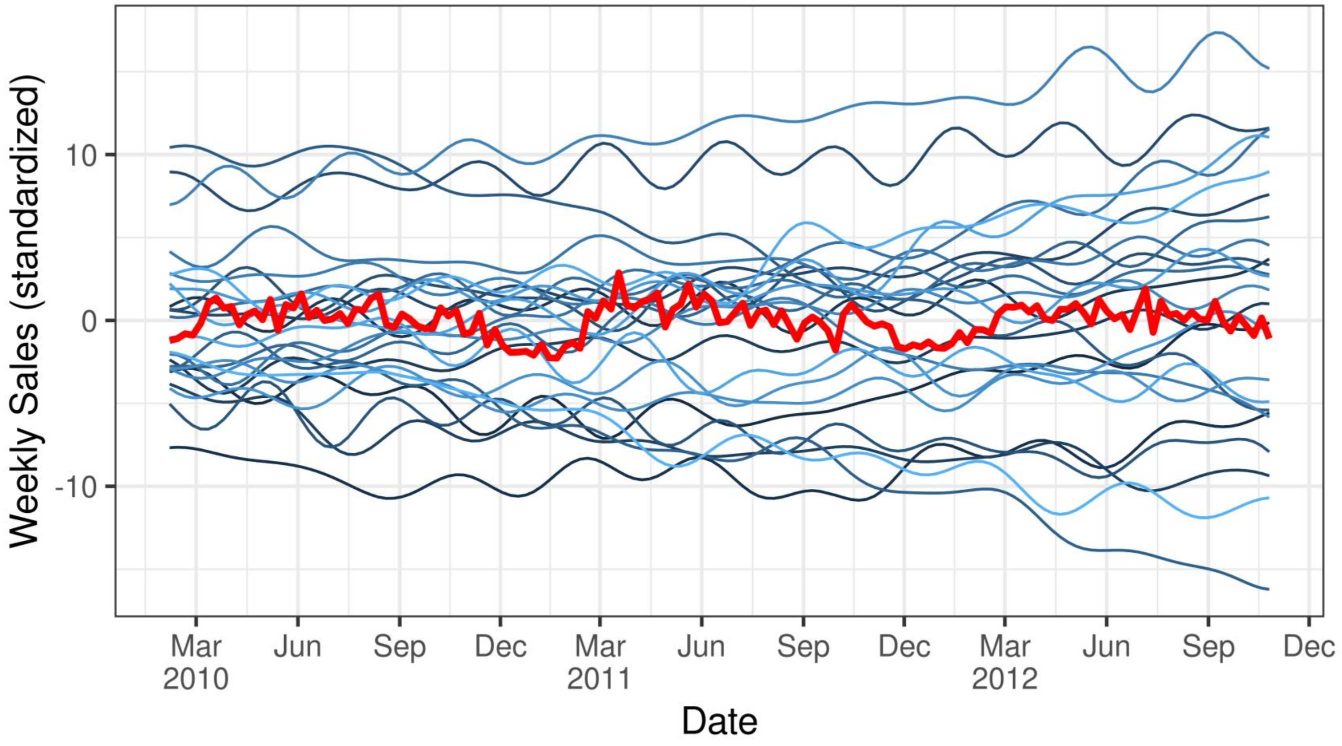
$$\text{growth}_t = (k + A\delta)t$$

$$\text{offset}_t = m - As\delta$$

$$(A)_{t's'} = t' > s'$$

Priors – Separate Model

$$\sigma_{\text{obs}} \sim N(0, 0.5)$$
$$k_j \sim N(0, 5)$$
$$m_j \sim N(0, 5)$$
$$\delta_{lj} \sim N(0, 5)$$
$$\beta_{kj} \sim N(0, 5/20)$$



Priors – Hierarchical Model

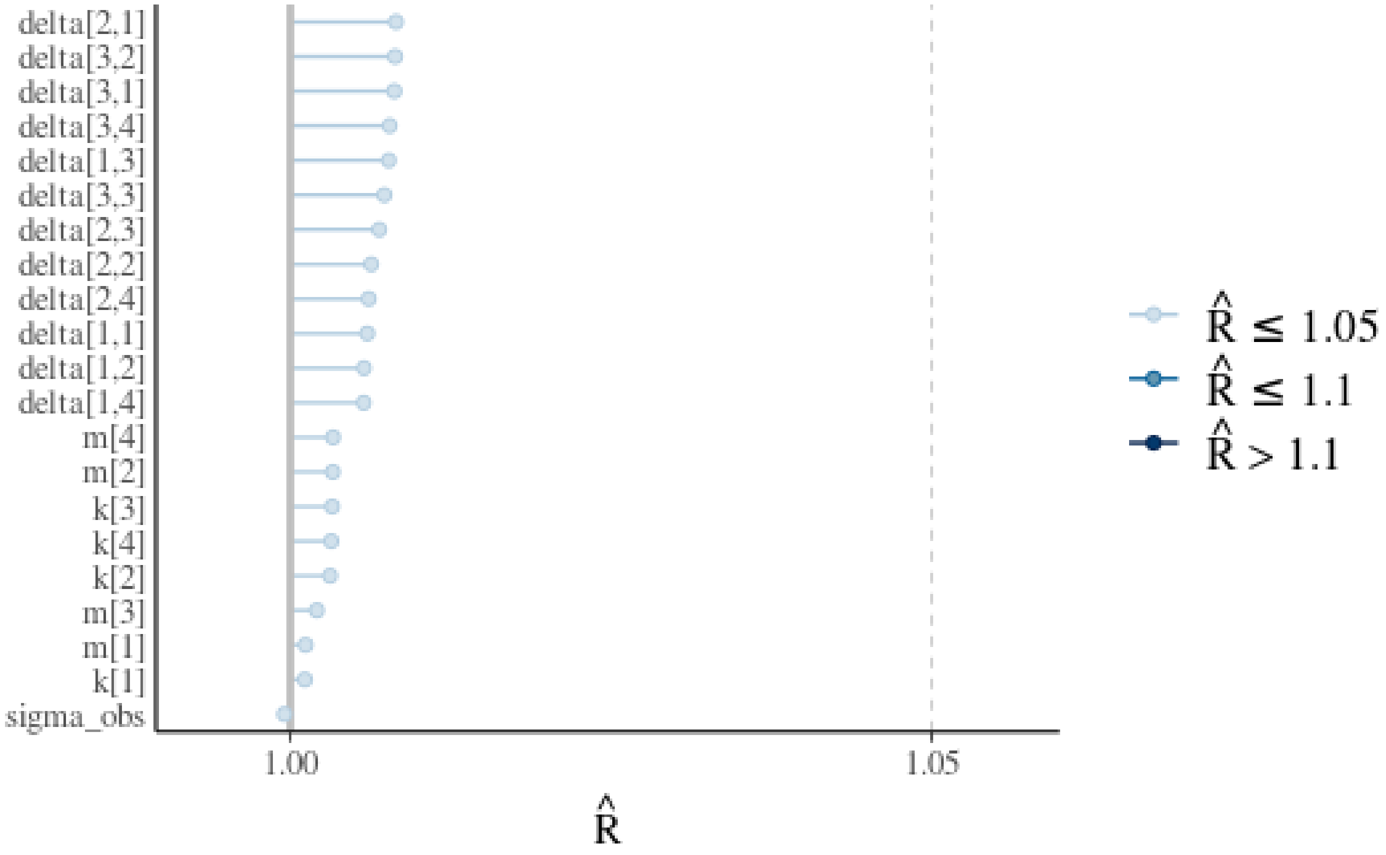
Priors

$$\sigma_{\text{obs}} \sim N(0, 0.5)$$
$$k_j \sim N(\mu_k, \sigma_k)$$
$$m_j \sim N(\mu_m, \sigma_m)$$
$$\delta_{lj} \sim N(\mu_\delta, \sigma_\delta)$$
$$\beta_{kj} \sim N(\mu_\beta, \sigma_\beta)$$

Hyperpriors

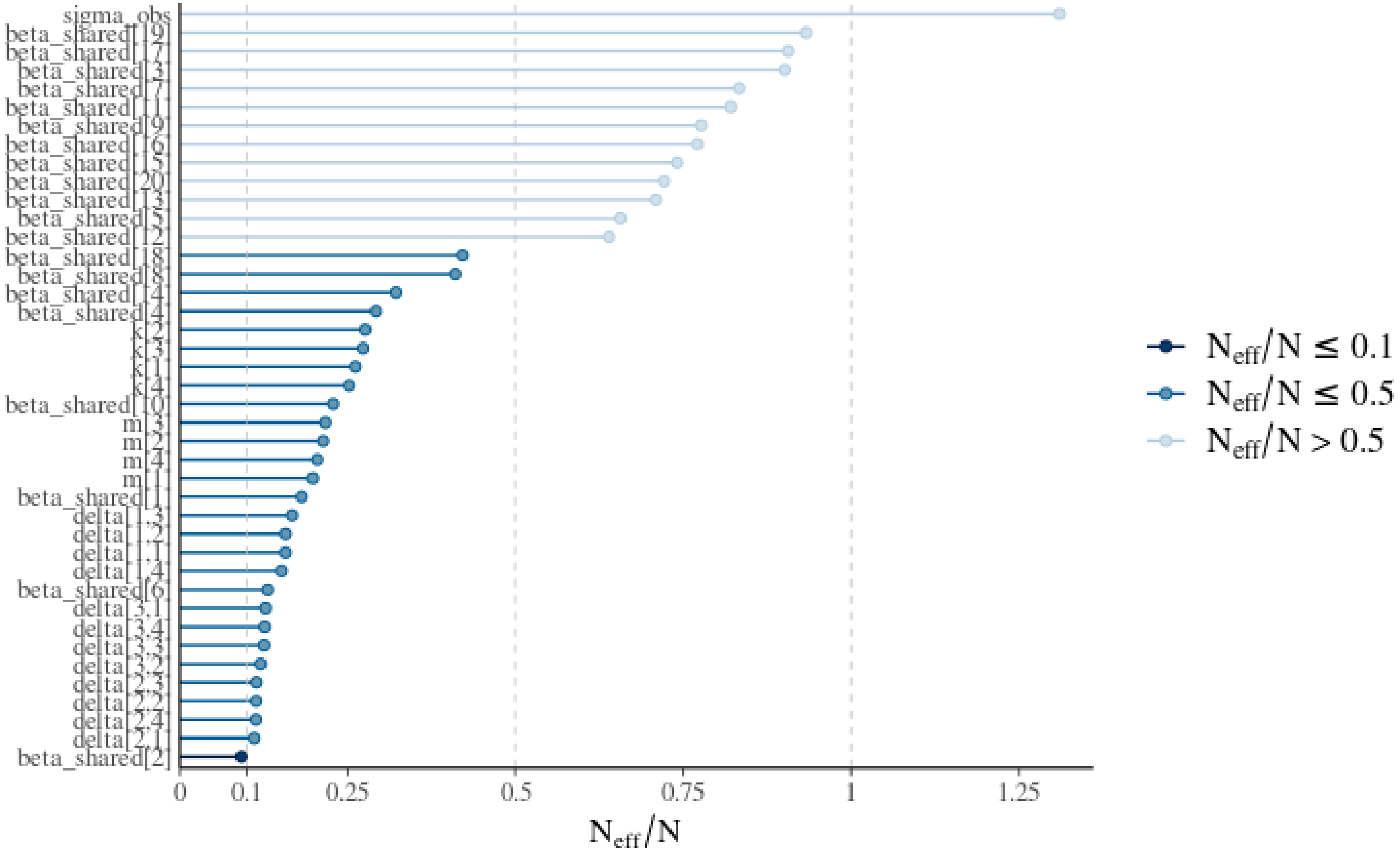
$$\mu_k \sim N(0, 5)$$
$$\mu_m \sim N(0, 5)$$
$$\mu_\delta \sim N(0, 5)$$
$$\mu_\beta \sim N(0, 5)$$
$$\sigma_k \sim \text{Inv-}\chi^2(1)$$
$$\sigma_m \sim \text{Inv-}\chi^2(1)$$
$$\sigma_\delta \sim \text{Inv-}\chi^2(1)$$
$$\sigma_\beta \sim \text{Inv-}\chi^2(1)$$

Convergence Diagnostics – R-hat



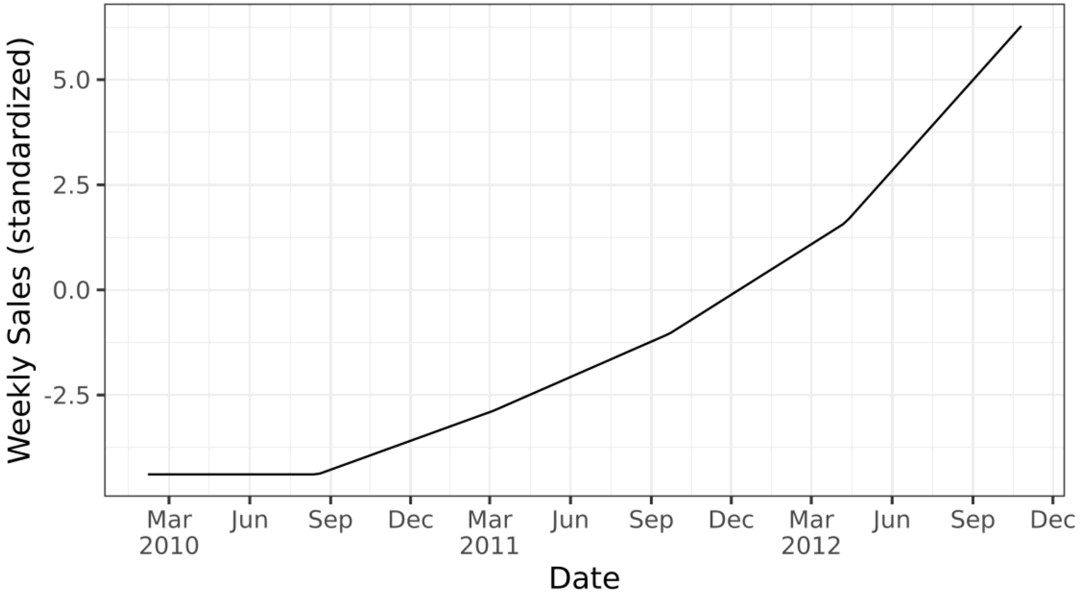
Hierarchical Model

Convergence Diagnostics – N_{eff}



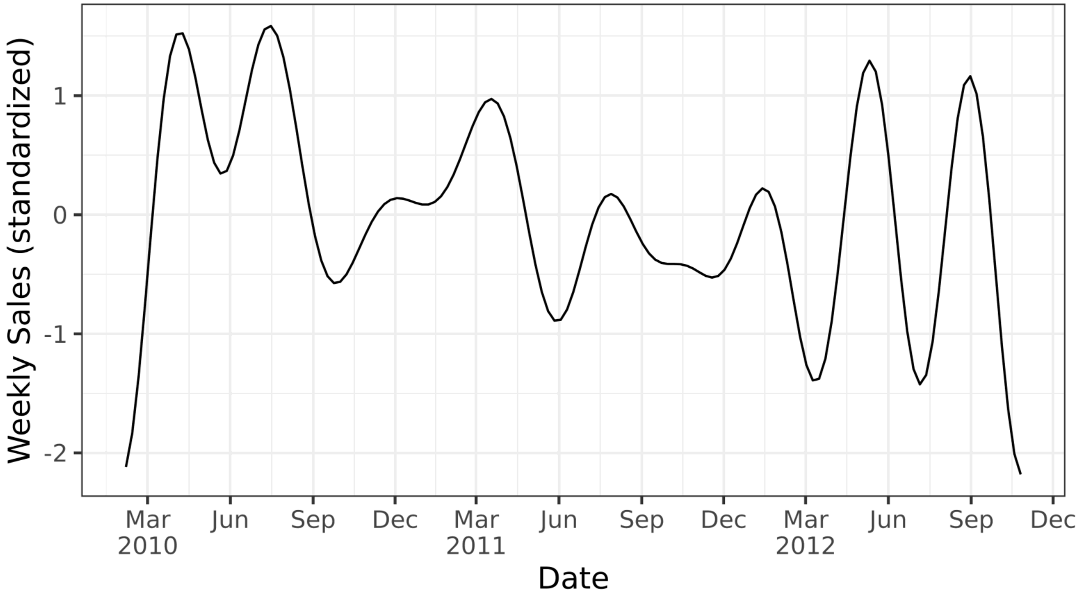
Hierarchical Model

Model: trend component formulation



$$\mathrm{trend}_t = \mathrm{growth}_t + \mathrm{offset}_t$$
$$\mathrm{growth}_t = (k + A\delta)t$$
$$\mathrm{offset}_t = m - As\delta$$
$$(A)_{t's'} = t' > s'$$

Model: seasonality



$$\text{seasonality}_t = \sum_{k=1}^{K} \beta_{k,1} cos(2\pi k \frac{t}{T}) + \beta_{k,2} sin(2\pi k \frac{t}{T})$$

```
transformed parameters {
  vector[T] growth = k + A * delta;
  vector[T] offset = m + A * (-s .* delta);
  vector[T] trend = (growth .* t) + offset;

  vector[T] seasonality = X * beta;

  vector[T] prediction = trend + seasonality;
}

model {
  k ~ normal(0, 5);
  m ~ normal(0, 5);

  delta ~ normal(0, 5);

  beta ~ normal(0, 5);

  sigma ~ normal(0, 0.5);

  y ~ normal(prediction, sigma);
}
```