Measuring the Effectiveness of Marketing Activities and Baseline Sales from POS Data using Bayesian State Space Models

Tomohiro Ando

Graduate School of Business Administration, Keio University, Japan

Presented at, Cherry Bud Workshop 2007, March 16, 2007

Business environment in Japan

The uncertainty is increasing !!

- Globalization of the economy
- Development of the information technology
- Change in customer's sense of value
- Deregulation

Main aim

- To develop the methodology for predicting daily and baseline sales of individual items by considering specific factors

- The prediction is very useful for
  - inventory management
  - marketing strategy planning
  - manpower planning
  - production planning

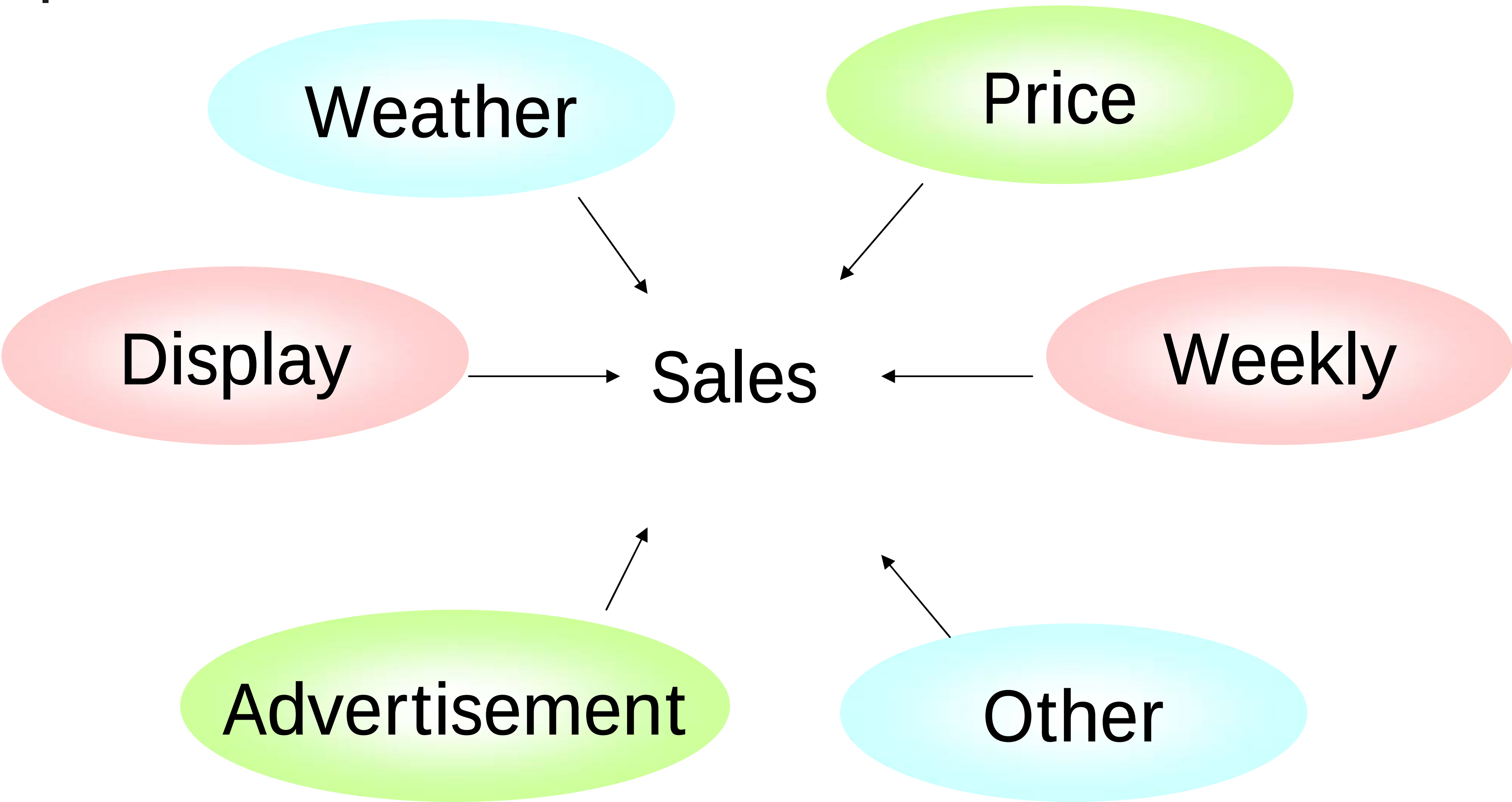
Agenda

- Data description
- Model description
- Bayesian MCMC Estimation
- Model evaluation
- Application results
- Conclusion

Data description

- Daily unit sales for 10 foods
  - Cold beverages
  - Coffee
  - Tea beverages
  - Cup noodle
  - Rice balls
  - Sandwiches
  - Packed lunches
  - Jelly
  - Yogurt

Data description



Main aim

- To develop the methodology for predicting daily and baseline sales of individual items by considering specific factors

- The prediction is very useful for
  - inventory management
  - marketing strategy planning
  - manpower planning
  - production planning

Agenda

- Model description
- Bayesian MCMC Estimation
- Model evaluation
- Application results
- Conclusion

Model description

- A general state space model (Kitagawa, 1987) enables us to conduct flexible time series modeling.
- Observation model

$$y_t \sim f(y_t \mid F_t, h_t, \ldots, h_1)$$

- System model

$$h_t \sim f(h_t \mid F_t, h_{t-1}, \ldots, h_1)$$

Model description

- Observation model

Daily unit of sales
= Baseline sales
+ Sales affected by environmental effect
+ Sales obtained from marketing promotion

Model description

- Observation model (Poisson model)

Daily unit sales

$$y_t \sim Poisson(\lambda_t \mid \mathbf{x}_t)$$

$$\mathbf{x}_t = (w_t, r_t, p_t, d_t, a_t)'$$

$$\lambda_t = h_t + \beta_1 w_t + \beta_2 r_t + \beta_3 p_t + \beta_4 d_t + \beta_5 a_t$$

Baseline unit sales Weather effect Price promotion effect

Weekly effect Display promotion effect Advertisement effect

Model description

- System model (q-th order trend model)

$$\Delta^q h_t = \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2)$$

- Example

$$(q=1): h_t = h_{t-1} + \varepsilon_t$$

$$(q=2): h_t = 2h_{t-1} - h_{t-2} + \varepsilon_t$$

Model description

- Each of the 10 foods follow the Poisson model.

$$y_{jt} \sim Poisson(\lambda_{jt} \mid \mathbf{x}_{jt}), \quad j = 1, \ldots, 10$$

- We consider that the daily unit sales are mutually dependent on each other.

$$\begin{cases} \Delta^q h_{jt} = \varepsilon_{jt}, \ \ \varepsilon_{jt} \sim N(0, \sigma_j^{\ 2}), \quad j = 1, \ldots, 10 \\ Cov(\varepsilon_{j,t}, \varepsilon_{k,t}) = \sigma_{jk} \end{cases}$$

Model description

- Summarizing the above specification leads to the following observation equation and the system equation:

Observation equation

$$y_{jt} \sim f(y_{jt} \mid h_{jt}; \mathbf{x}_{jt}, \boldsymbol{\beta}_j)$$

System equation

$$\mathbf{h}_t \sim f(\mathbf{h}_t \mid \mathbf{h}_{t-1}, \ldots, \mathbf{h}_{t-q}, \Sigma), \ \Sigma = (\sigma_{ij})$$

Agenda

- Model description
- Bayesian MCMC Estimation
- Model evaluation
- Application results
- Conclusion

Bayesian estimation via Markov chain Monte Carlo (1/2)

- A set of n independent observations
- Likelihood function

$$
\begin{aligned}
L(D_n|\boldsymbol{\theta}, X_n) &= \prod_{t=1}^{n} f(\boldsymbol{y}_t|F_{t-1}, \boldsymbol{x}_t, \boldsymbol{\theta}) \\
&= \prod_{t=1}^{n} \left[\prod_{j=1}^{p} f(y_{jt}|F_{t-1}, \boldsymbol{x}_{jt}, \boldsymbol{\theta}) \right] \\
&= \prod_{t=1}^{n} \left[\prod_{j=1}^{p} \int \underbrace{f(y_{jt}|h_{jt}; \boldsymbol{x}_{jt}, \beta_j)}_{\text{Observation model}} \underbrace{f(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \ldots, \boldsymbol{h}_{t-q}; \Sigma)}_{\text{System model}} d\boldsymbol{h}_t \right]
\end{aligned}
$$

Bayesian estimation via Markov chain Monte Carlo (2/2)

- Markov chain Monte Carlo considers the state vector as model parameters

$$\mathbf{h} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n)$$

- A posterior distribution

$$\pi(\boldsymbol{\theta}, \boldsymbol{h} | D_n, X_n) \propto \pi(\boldsymbol{\theta}) \times \prod_{t=1}^{n} \prod_{t=1}^{n} f(y_{jt} | h_{jt}; \boldsymbol{x}_{jt}, \beta_j) f(\boldsymbol{h}_t | \boldsymbol{h}_{t-1}, \ldots, \boldsymbol{h}_{t-q}; \Sigma).$$

Agenda

- Model description
- Bayesian MCMC Estimation
- Model evaluation
- Application results
- Conclusion

Maximisation of the posterior mean of the expected likelihood

- Ando (2006) considered the maximization of the posterior mean of the expected log-likelihood (PMELL):

$$\eta = E_{Z_n}\left[E_{\boldsymbol{\theta}|D_n,X_n}[\log L(Z_n|\boldsymbol{\theta}, X_n)]\right]$$
$$= \int \left[\int \log L(Z_n|\boldsymbol{\theta}, X_n)\pi(\boldsymbol{\theta}|D_n, X_n)d\boldsymbol{\theta}\right] dG(Z_n|X_n),$$

- The best model is chosen by maximising η .

Ando (2006)

- The asymptotic bias is

$$
\begin{aligned}
n\hat{b} \approx\ & E_{\boldsymbol{\theta}|D_n,X_n}\left[\log\{L(D_n|\boldsymbol{\theta},X_n)\pi(\boldsymbol{\theta})\}\right] + \dim\{\boldsymbol{\theta}\}/2 \\
& - \log\{L(D_n|\hat{\boldsymbol{\theta}}_n,X_n)\pi(\hat{\boldsymbol{\theta}}_n)\} + \mathrm{tr}\left\{J_n^{-1}(\hat{\boldsymbol{\theta}}_n)I_n(\hat{\boldsymbol{\theta}}_n)\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
I_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial\eta_n(\boldsymbol{y}_t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\eta_n(\boldsymbol{y}_t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\}, \\
J_n(\boldsymbol{\theta}) &= -\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2\eta_n(\boldsymbol{y}_t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\},
\end{aligned}
$$

with $\eta_n(\boldsymbol{y}_t,\boldsymbol{\theta}) = \log f(\boldsymbol{y}_t|F_{t-1},\boldsymbol{x}_t,\boldsymbol{\theta}) + \log\pi(\boldsymbol{\theta})/n.$

Ando (2006)

- The prior is dominated by the likelihood as the sample size becomes large.

- The specified parametric models contain the true model or are close to the true model

$$\Rightarrow \quad n \times \hat{b} = \dim(\boldsymbol{\theta})$$

Bayesian predictive information criterion

- Bayesian predictive information criterion (BPIC Ando (2006)):

$$\mathbf{BPIC} = -2E_{\boldsymbol{\theta}|D_n, X_n}\left[\log L(D_n \mid \boldsymbol{\theta}, X_n)\right] + 2\dim\{\boldsymbol{\theta}\}$$

- We choose the predictive distribution that minimises the BPIC score.

Agenda

- Model description
- Bayesian MCMC Estimation
- Model evaluation
- Application results
- Conclusion

Application results

- The posterior mean of the log-likelihood and BPIC scores for each of q-th order trend models.

q	1	2	3	4
$\hat{\eta}$	-1543.081	-1507.530	-1492.693	-1483.921
BPIC	3296.162	3245.060	3235.386	3237.842
CPU time	16.75	17.54	18.28	19.09

- The CPU time (seconds) to generate 100 iterations is also reported.

Model description

- Observation model

$$y_t \sim Poisson(\lambda_t \mid \mathbf{x}_t)$$

$$\mathbf{x}_t = (w_t, r_t, p_t, d_t, a_t)'$$

$$\lambda_t = h_t + \beta_1 w_t + \beta_2 r_t + \beta_3 p_t + \beta_4 d_t + \beta_5 a_t$$

Baseline unit sales (h_t)

Weekly effect ($\beta_1 w_t$)

Weather effect ($\beta_2 r_t$)

Display promotion effect ($\beta_3 p_t$)

Price promotion effect ($\beta_4 d_t$)

Advertisement effect ($\beta_5 a_t$)

		PMs	95% CIs				PMs	95% CIs	
CB	β_{11}	0.66	[0.14	1.18]	SA	β_{61}	0.52	[0.06	0.97]
	β_{12}	7.52	[6.52	8.59]		β_{62}	0.22	[0.02	0.60]
	β_{13}	41.2	[40.2	42.3]		β_{63}	1.53	[0.17	2.56]
	β_{14}	6.38	[5.71	7.16]		β_{64}	2.36	[1.85	2.71]
	β_{15}	1.37	[0.09	2.47]		β_{65}	0.64	[0.04	1.59]
CF	β_{21}	0.58	[0.15	1.06]	P1	β_{71}	0.52	[0.07	1.28]
	β_{22}	13.2	[12.4	13.8]		β_{72}	0.99	[0.13	1.51]
	β_{23}	19.9	[19.3	20.6]		β_{73}	3.06	[1.04	4.21]
	β_{24}	4.82	[3.97	6.19]		β_{74}	17.0	[16.2	17.8]
	β_{25}	1.81	[1.10	2.68]		β_{75}	1.64	[0.98	2.40]
TB	β_{31}	0.63	[0.05	1.67]	P2	β_{81}	0.29	[0.04	0.64]
	β_{32}	9.67	[8.60	10.5]		β_{82}	0.71	[0.03	1.33]
	β_{33}	1.03	[0.21	1.59]		β_{83}	104.57	[103.6	105.4]
	β_{34}	3.15	[2.44	3.70]		β_{84}	0.45	[0.03	1.45]
	β_{35}	11.2	[10.4	12.2]		β_{85}	0.38	[0.04	0.84]
CN	β_{41}	1.01	[0.06	1.97]	JE	β_{91}	5.65	[4.47	6.89]
	β_{42}	4.61	[3.95	5.24]		β_{92}	5.65	[4.06	6.56]
	β_{43}	1.66	[1.01	2.17]		β_{93}	0.55	[0.08	1.15]
	β_{44}	9.47	[8.50	11.0]		β_{94}	49.3	[48.4	51.3]
	β_{45}	2.68	[1.27	4.12]		β_{95}	36.3	[35.0	37.8]
RB	β_{51}	1.70	[0.30	2.68]	YO	β_{101}	1.52	[0.78	2.13]
	β_{52}	16.5	[15.4	17.3]		β_{102}	14.6	[14.1	15.1]
	β_{53}	3.23	[2.60	3.88]		β_{103}	0.36	[0.04	0.84]
	β_{54}	13.1	[12.5	13.7]		β_{104}	13.5	[11.7	14.9]
	β_{55}	0.95	[0.50	1.69]		β_{105}	26.2	[25.7	26.9]

Application results

- Observation model

$$y_t \sim Poisson(\lambda_t \mid \mathbf{x}_t)$$

$$\mathbf{x}_t = (w_t, r_t, p_t, d_t, a_t)'$$

$$\lambda_t = h_t + 1.7w_t + 16.5r_t + 3.2p_t + 13.1d_t + 0.95a_t$$

Baseline unit sales

Weather effect

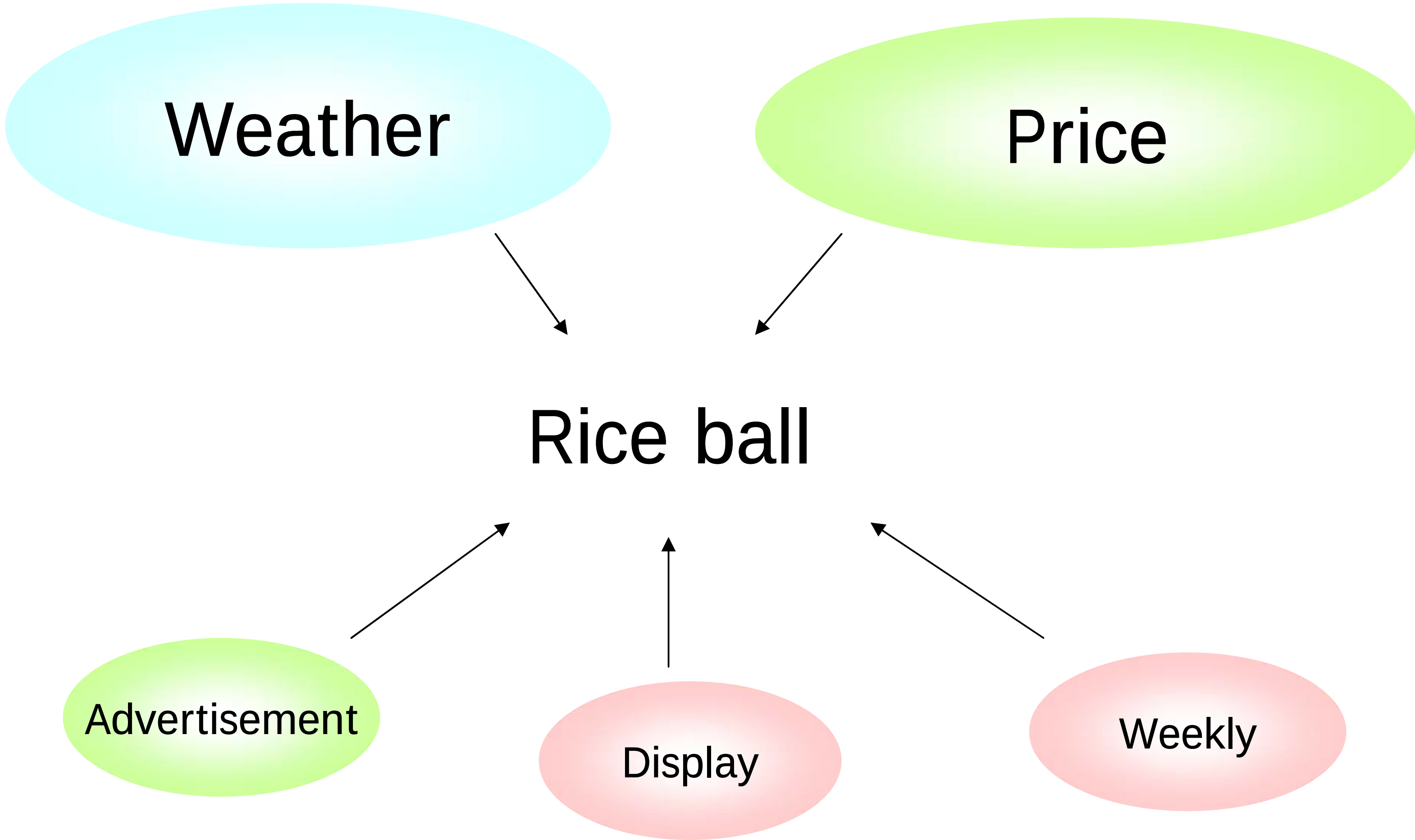
Price promotion effect

Weekly effect

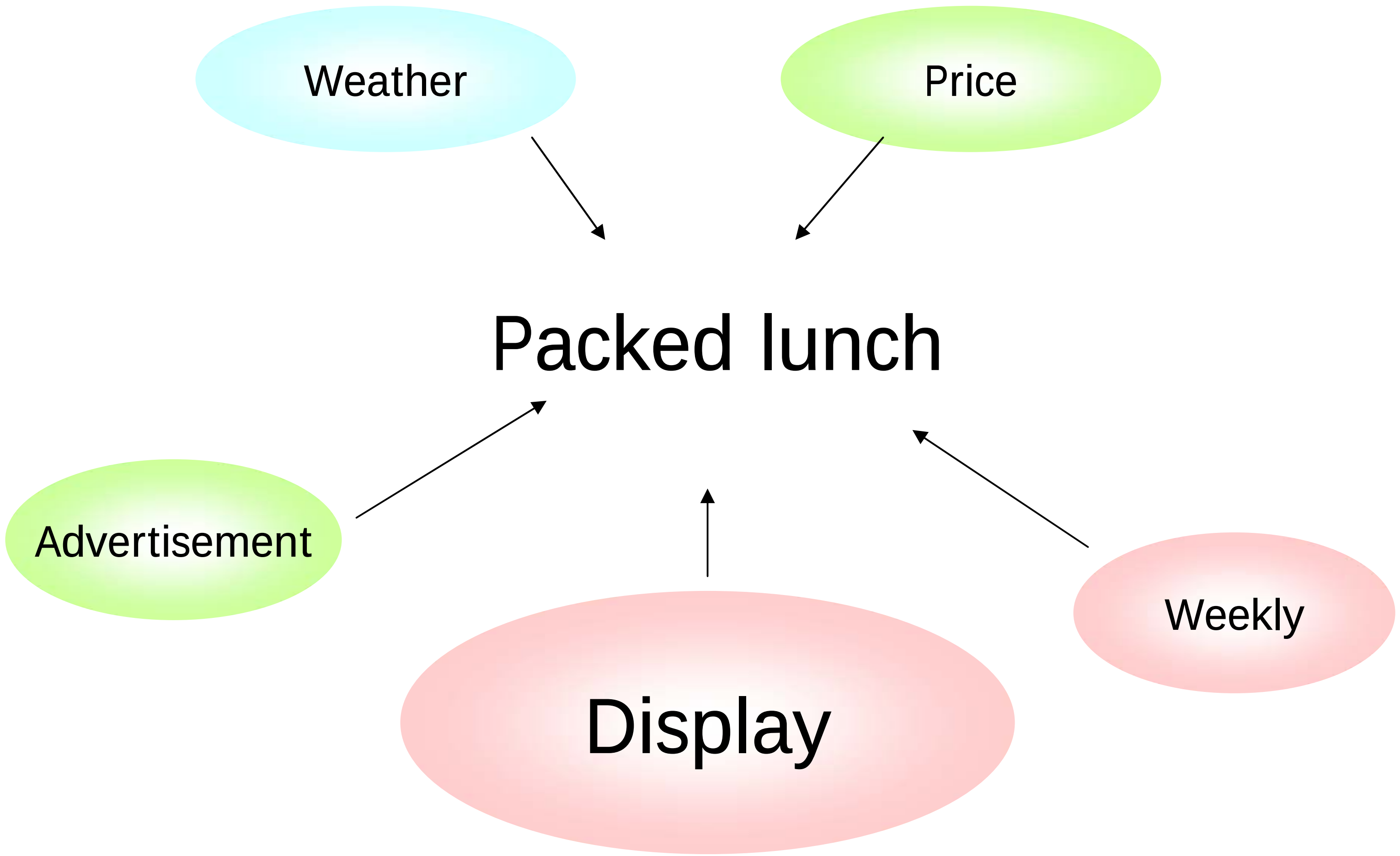
Display promotion effect

Advertisement effect

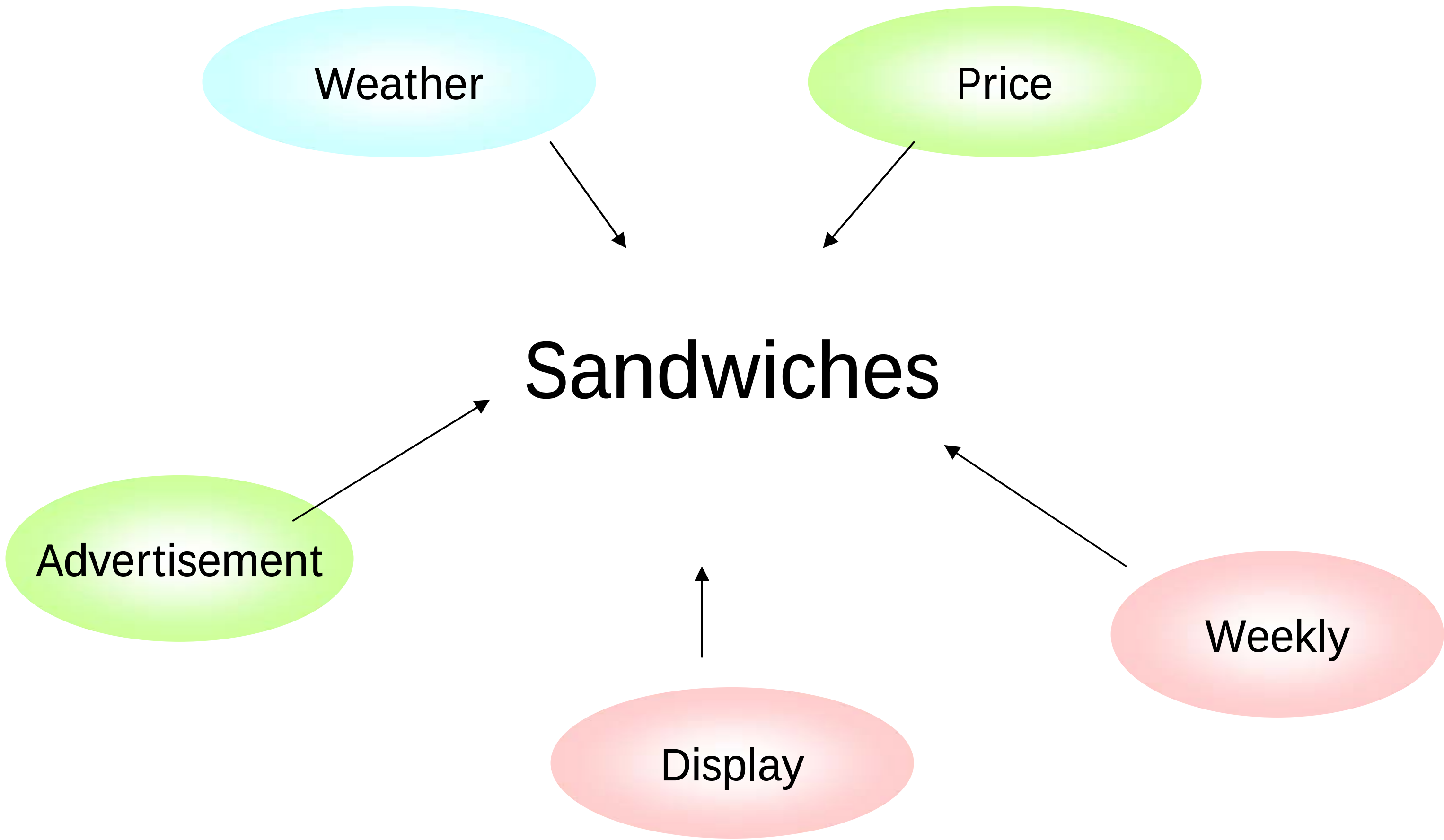
Application results



Application results



Application results



Model description

- Each of the 10 foods follow the Poisson model.

$$y_{jt} \sim Poisson(\lambda_{jt} \mid \mathbf{x}_{jt}), \quad j = 1, \ldots, 10$$

- We consider that the daily unit sales are mutually dependent on each other.

$$\begin{cases} \Delta^q h_{jt} = \varepsilon_{jt}, \quad \varepsilon_{jt} \sim N(0, \sigma_j^{\ 2}), \quad j = 1, \ldots, 10 \\ Cov(\varepsilon_{j,t}, \varepsilon_{k,t}) = \sigma_{jk} \end{cases}$$

Application results

- The posterior means of correlation

	CB	CF	TB	CN	RB	SA	P1	P2	JE	YO
CB:	1.00	0.54	0.35	0.15	0.52	0.18	0.17	0.24	0.35	0.46
CF:	0.54	1.00	0.37	0.29	0.40	0.21	0.36	0.12	0.49	0.44
TB:	0.35	0.37	1.00	0.55	0.74	0.13	0.34	0.34	0.28	0.36
CN:	0.15	0.29	0.55	1.00	0.47	0.23	0.28	0.17	0.41	0.29
RB:	0.52	0.40	0.74	0.47	1.00	0.07	0.31	0.35	0.59	0.61
SA:	0.18	0.21	0.13	0.23	0.07	1.00	0.43	0.34	0.22	0.23
P1:	0.17	0.36	0.34	0.28	0.31	0.43	1.00	-0.24	0.41	0.61
P2:	0.24	0.12	0.34	0.17	0.35	0.34	-0.24	1.00	0.20	0.13
JE:	0.35	0.49	0.28	0.41	0.59	0.22	0.41	0.20	1.00	0.83
YO:	0.46	0.44	0.36	0.29	0.61	0.23	0.61	0.13	0.83	1.00

CB: cold beverages, CF: coffee, TB: tea beverages, CN: cup noodle, RB: rice balls, SA: sandwiches, P1: packed lunches (type1), P2: packed lunches (type2), JE: jelly and YO: yogurt.

Agenda

- Model description
- Bayesian MCMC Estimation
- Model evaluation
- Application results
- Conclusion

Conclusion

- The problem of identifying unobserved baseline sales, marketing promotion effects and other specific effects using POS data is considered.
- General state space model
- Bayesian MCMC estimation
- Bayesian predictive information criterion is utilized.
- The unobserved baseline sales, marketing promotion effects and other specific effects are estimated simultaneously.

Thank you for your attention