MKT6v99.0U1 Special Topics:
Bayesian Dynamic Models in Marketing

Summer 2013

Instructor: Dr. Norris Bruce, PhD
Meeting Time: Tuesday 10:30AM-2:30pm
Class Room: SOM 1.117
Office Hours: Tuesday 4:00pm-6:00pm, by appointment
Office: SOM 3.220
Phone: 972-883-6293
Email: norris.bruce@utdallas.edu
TA: Abhishek Rishabh

Course Description

This is an advanced PhD seminar that introduces you to Bayesian dynamic models. Its objective is to help you learn how to develop, estimate, and assess these dynamic models; and how to add to the marketing dynamics literature.

Recommended Textbooks

General Bayesian Statistics:
Bayesian Statistics and Marketing (2005). Peter E. Rossi, Greg M. Allenby, Rob McCulloch.

Dynamic Linear Models:
Petris, Petrone and Campagnoli (2009). Dynamic Linear Models with R. Springer.
West and Harrison (1999). Bayesian Forecasting and Dynamic Models. 2nd Edition, Springer.

Dynamic Non-Linear Models:
Branko Ristic, Sanjeev Arulampalam Neil Gordon (2004). Beyond the Kalman Filter: Particle Filters for Tracking Applications. Artech House Radar Library.

Other readings including journal articles and/or book chapters are referenced in the "Class Schedule, Topics and Readings" section below.

Course Requirements:

Course Project: For this project students are encouraged to work in pairs. The purpose of the project is to have you apply some of the techniques covered in the course to a substantive marketing problem. My hope is that some of your ideas will eventually be publishable. Please submit a two page report by June 25th containing detailed explanation of the potential problem you want to work on, the plans for collecting the data, and the modeling approach you hope to take. The project consists of two deliverables: 1) a copy of your code (properly documented and preferably written in Matlab); and 2) a 30-minutes presentation which should include a description of your problem, data, model, estimation and results.

Weekly Schedule, Topics and Readings

Readings with asterisks are recommended, not required.

Meeting 1, June 4: Course overview; introduction to Bayesian Dynamic Models
Richard J. Meinhold and Nozer D. Singpurwalla (1983). Understanding the Kalman Filter. The American Statistician, Vol. 37, No. 2, pp. 123-127.

Meeting 2, June 11: Bayesian Dynamic Linear Model (DLM)
West and Harrison (1999), "Bayesian Forecasting and Dynamic Models", 2nd Edition, Springer (Ch. 4).
*Carter, C., R. Kohn. 1994. On Gibbs sampling for state space models. Biometrika 81 541–553.

Frank M. Bass, Norris Bruce, Sumit Majumdar, and B. P. S. Murthi (2007). Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Advertising-Sales Relationship. Marketing Science, vol. 26, issue 2, pages 179-195.

Norris I. Bruce, Natasha Zhang Foutz, Ceren Kolsarici (2012). Dynamic Effectiveness of Advertising and Word of Mouth in Sequential Distribution of New Products. Journal of Marketing Research: Vol. 49, No. 4, pp. 469-486.

Meeting 3, June 18: Dynamic Factor Models
Lopes, H.F. and West, M. (2004). Bayesian Model assessment in factor analysis. Statistica Sinica, 14, 41-67.

Lopes, Hedibert F., Esther Salazar, and Dani Gamerman (2008), "Spatial Dynamic Factor Analysis," Bayesian Analysis, 3 (4), 759–92.

mcisam 1 Norris I. Bruce, Kay Peters, Prasad A. Naik (2012). Discovering How Advertising Grows Sales and Builds Brands. Journal of Marketing Research: Vol. 49, No. 6, pp. 793-806.

Meeting 4, June 25: Hierarchical DLM and its applications
Gamerman and Migon (1993). Dynamic Hierarchical Models. Journal of the Royal Statistical Society; Series B, 55, 629–642.

Neelamegham and Chintagunta (2004). Modeling and Forecasting the Sales of Technology Products. Quantitative Marketing and Economics, 2, 195-232.

Meeting 5, July 2: Dynamic GLMS and Conditional Linear Models
*Ferreira and Gamerman (2000). Dynamic Generalized Linear Models: in Generalized Linear Models: A Bayesian Perspective, CRC.

*Claudia Cargnoni, Peter Müller & Mike West (1997). Bayesian Forecasting of Multinomial Time Series through Conditionally Gaussian Dynamic Models. Journal of the American Statistical Association. Volume 92, Issue 438, pages 640-647.

Teixeira, Wedel and Pieters (2010). Moment-to-Moment Optimal Branding in TV-commercials: Preventing Avoidance by Pulsing. Marketing Science, Vol. 29, No. 5, pp. 783-804

(i) Rutz, Oliver J.; Sonnier, Garrett P. (2011). The Evolution of Internal Market Structure. Marketing Science. Mar/Apr2011, Vol. 30 Issue 2, p274.

Gamerman (1998). Markov Chain Monte Carlo for Dynamic Generalized Linear Models. Biometrika, 85, 215-227.

Meeting 6: July 9: No Class – *Marketing Science Conference, Istanbul Turkey*

Meeting 7, July 16: Latent Variable Approach to Measurement Error in DLMs

Bayesian instrumental variables pp183-187: *Bayesian Statistics and Marketing*, Rossi et al.

Naik, Prasad A.; Tsai, Chih-Ling (2000). Controlling Measurement Errors in Models of Advertising Competition. Journal of Marketing Research (JMR);Feb2000, Vol. 37 Issue 1, p113.

meisam Garrett P. Sonnier, Leigh McAlister, and Oliver J. Rutz (2011). A Dynamic Model of the Effect of Online Communications on Firm Sales. Marketing Science. Vol. 30, No. 4, 702-716.

Meeting 8, July 23: Nonlinear Filters
Understanding the Extended and Unscented Kalman Filters: in Beyond the Kalman Filter, Ristic et al (2004).

Jinhong Xie, X. Michael Song, Marvin Sirbu, Qiong Wang (1997). Kalman Filter Estimation of New Product Diffusion Models. Journal of Marketing Research, Vol. 34, No. 3 (Aug., 1997), pp. 378-393.

Prasad A. Naik, Ashutosh Prasad, Suresh P. Sethi (2208). Building Brand Awareness in Dynamic Oligopoly Markets. Management Science, Vol. 54, No. 1, pp. 129-138.

Meeting 9, July 30: The Ensemble Kalman Filter

Jonathan R. Strouda, Michael L. Steina, Barry M. Leshta, David J. Schwaba & Dmitry Beletskya (2010). An Ensemble Kalman Filter and Smoother for Satellite Data Assimilation. Journal of the American Statistical Association. Volume 105, Issue 491, 2010.

Chao Gao, Han Wang, Ensheng Weng, S. Lakshmivarahan, Yanfen Zhang, and Yiqi Luo (2011). Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics. Ecological Applications. Volume 21, Issue 5.

Meeting 10, Aug. 6: Project Presentations:
Research Question; Data Collection and Analysis; Model Development and Estimation; Results.

Baysian summer seminar

1

Understanding the Kalman Filter — American Statistician 1983

$Y_t, Y_{t-1}, \cdots, Y_1$ observable at $t, t-1, \cdots$

θ_t : State of nature unobservable

$Y_t = F_t\theta_t + v_t \sim N(0, V_t)$ obs-error, Regression AR(1)

↳ known quant

State of nature

↑

Dynamic, in system equation: $\theta_t = G_t\theta_{t-1} + w_t \sim N(0, W_t)$

↳ chng according to eng. or sci. fileds in time

assump) $v_t \perp w_t$ (indep)

θ_t: position of satelite at time t

Y_t: ① measurements of distance to the satelite ② angle of measure

v_t: measurement Error

G_t: position and speed change over time (physical laws of orbit)

w_t: deviation from laws (non uniformity of the earth gravitational field)

$\boxed{\text{param interest: } \theta_t}$

Familiarity / linear regression \ time series analysis

Y_t: number of defective observed

θ_1, θ_2: true defective index of the process, & drift

$Y_t = \theta_{1,t} + v_t$, $\theta_{1,t} = \theta_{2,t} + w_{1,t}$ $\theta_{2,t} = \theta_{2,t-1} + w_{2,t}$

$\theta_t = G\theta_{t-1} + u_t$ $\theta_t = \begin{bmatrix}\theta_{1,t}\\ \theta_{2,t}\end{bmatrix}$ $u_t = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}w_{1t}\\ w_{2t}\end{bmatrix}$ $G = \begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$

if $V_t = V$, $W_t = W \Rightarrow Y_t - Y_{t-1} \sim ARIMA(0,1,1)$ Box & Jenkins

Kalman filter ~ Recursive Procedure infer θ_t

$P\{St \text{ of Natural} | \text{Data}\} \propto P\{\text{Dat} | \text{Stat. Natre}\} \propto P\{\text{St. Natr}\}$

$P\{\theta_t | Y_t\} \propto P\{Y_t | \theta_t, Y_{t-1}\} \times P\{\theta_t | Y_{t-1}\}$ likelihood, prior dist

$Y_{t-1} = (Y_{t-1}, Y_{t-2}, \cdots, Y_1)$ matrix form

$(\theta_{t-1} | Y_{t-1}) \sim N(\hat{\theta}_{t-1}, \Sigma_{t-1})$ posterior

Look time t

prior to t Y_t ⟵ ⟶ after time t Y_t

Rec Proc start at $t=0$, $\hat{\theta}_0, \Sigma_0$

Stg 1 $(\theta_t | Y_{t-1}) \sim N(G_t\hat{\theta}_{t-1}, R_t = G_t\Sigma_{t-1}G_t' + W_t)$

Stg 2 $(\theta_t | Y_t) = ?$ on observing $\theta_t \Rightarrow$ nd: $L(\theta_t = Y_t) = ? \equiv P(Y_t | \theta_t, Y_{t-1})$?

err prediction: $e_t = Y_t - \hat{Y}_t = Y_t - F_tG_t\hat{\theta}_{t-1}$

$P(\theta_t | Y_t, Y_{t-1}) = P(\theta_t | e_t, Y_{t-1}) \propto \underbrace{P(e_t | Y_{t-1}, \theta_t)}_{\text{likelihood}} \times P(\theta_t | Y_{t-1})$

↳ ① $Y_t = F_t\theta_t + v_t \Rightarrow e_t = F_t(\theta_t - G_t\hat{\theta}_{t-1})$

$\Rightarrow$ ② $v_t \sim N(0, V_t) \Rightarrow (e_t | \theta, Y_{t-1}) \sim N(F_t(\theta_t - G_t\hat{\theta}_{t-1}), V_t)$

$\Rightarrow$ use bays theorem $P(\theta_t | Y_t, Y_{t-1}) = \dfrac{P(e_t | \theta_t, Y_{t-1}) \times P(\theta_t | Y_{t-1})}{\int_{\theta_t} P(e_t, \theta_t | Y_{t-1})\, d\theta_t}$

$\begin{pmatrix}X_1\\ X_2\end{pmatrix} \sim N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right)$

$(X_1 | X_2 = x_2) \sim N(\underbrace{\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)}_{\text{Regression func}}, \underbrace{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}_{})$ Coefficient of least sq. reg x_1 on x_2

$X_1 \Leftrightarrow e_t$ $X_2 \Leftrightarrow \theta_t$

$(\theta_t | Y_{t-1}) \sim N(G_t\hat{\theta}_{t-1}, R_t)$ $\mu_2 \Leftrightarrow G_t\hat{\theta}_{t-1}$ $\Sigma_{22} \Leftrightarrow R_t$

$(e_t | \theta_t, Y_{t-1}) \sim N(F_t(\theta_t - G_t\hat{\theta}_{t-1}), V_t)$

$\mu_1 + \Sigma_{12}R_t^{-1}(\theta_t - G_t\hat{\theta}_{t-1}) \Leftrightarrow F_t(\theta_t - G_t\hat{\theta}_{t-1})$

$\Rightarrow (e_t | \theta_t, Y_{t-1}) \sim N(F_t(\theta_t - G_t\hat{\theta}_{t-1}), V_t)$, $\mu_1 \Leftrightarrow 0$, $\Sigma_{12} \Leftrightarrow F_tR_t$

$\Sigma_{11} - \Sigma_{12}\Sigma_{22}\Sigma_{21} = \Sigma_{11} - F_tR_tF_t' \Leftrightarrow V_t$, $\Sigma_{11} \Leftrightarrow V_t + F_tR_tF_t'$

$\Rightarrow \left[\begin{pmatrix}\theta_t\\ e_t\end{pmatrix} \middle| Y_{t-1}\right] \sim N\left[\begin{pmatrix}G_t\hat{\theta}_{t-1}\\ 0\end{pmatrix}, \begin{pmatrix}R_t & R_tF_t'\\ F_tR_t & V_t + F_tR_tF_t'\end{pmatrix}\right]$

$\Rightarrow (\theta_t | e_t, Y_{t-1}) \sim N[G_t\hat{\theta}_{t-1} + R_tF_t'(V_t + F_tR_tF_t')^{-1}e_t,$
$R_t - R_tF_t'(V_t + F_tR_tF_t')^{-1}F_tR_t] \Rightarrow$ posterior

Prior: $\theta_t \sim (G_t\hat{\theta}_{t-1}, R_t = G_t\Sigma_{t-1}G_t' + W_t)$

$e_t = Y_t - F_tG_t\hat{\theta}_{t-1}$

Posterior: $\hat{\theta}_t = G_t\hat{\theta}_{t-1} + R_tF_t'(V_t + F_tR_tF_t')^{-1}e_t$

$\Sigma_t = R_t - R_tF_t'(V_t + F_tR_tF_t')^{-1}F_tR_t$

evolution of series of regression functions of (θ_t, e_t)

≡ Learning process

- GPS devices tracking mobile phones, signal processing, tracking
- inverting Cov matrix (huge matr) is costly $\boxed{\text{Bottle neck}}$
- Satelite picture ⇒ ensemble Common Filter
- Awareness as measure of enthusiasm, loyalty
- sequential monte carlo $\boxed{\text{Stat: Construct we dont observe}}$
- $\boxed{\text{Assumptions}}$ { ① normality ② linearity
- Name: $\boxed{\text{Linear Space model}}$ ↑ θ: state parameter
- Sales function of advertising effectiveness / loyalty, ↑

F_t: Vector of known constant (linear or nonlinear)

G_t: known matrix related to vec. state

↳ Could be advertising (kevin keller brand Equity model)

- strong data over time
- Attribution: Last touch point attributed to sales (all learning done beforehand)
- State evolve for indiv ⇒ heterogeneity ⇒ identify segments
- Filter step: create Cond. dist. ⇒ create posterior move to
- next side of bar ≡ assume that as data
- heterogeneity is problem in max Likelihood
- Gibbs Sampler → very flexible
- MAD: mean absolute deviation
- Goodness of fit: how good you track data

~~How to estimate~~

Bayesian Seminar

- Retrospective analysis $\Leftrightarrow$ go back & revise the prior (Smoothing)
- general goal: (1) make inference (2) make Forecast
- model well defined by $\{F_t, G_t, V_t, W_t\}$
  - vector (reg), matrix, Variance v_t, variance w_t

$$\begin{cases} Y_t = F_t \Phi_t + v_t \\ \Phi_t = G_t \Phi_{t-1} + w_t \end{cases}$$

Prior: $\Phi_0 | D_0 \sim N(m_0, C_0)$; D_0: info set

$\Phi_{t-1} | D_{t-1} \sim N(m_{t-1}, C_{t-1})$ Posterior on $t-1$

Prior on t: $\Phi_t | D_{t-1} \sim N(a_t, R_t)$

$E(\Phi_t | D_{t-1}) = a_t = G_t m_{t-1}$

$V(\Phi_t | D_{t-1}) = R_t = G_t C_{t-1} G_t' + W_t$

Predictive dist: $Y_t | D_{t-1} \sim N(f_t, Q_t)$

$E(Y_t | D_{t-1}) = f_t = F_t a_t$

$V(Y_t | D_{t-1}) = Q_t = F_t R_t F_t + v_t$

Posterior on t: $\Phi_t, Y_t | D_{t-1} \sim N\left(\begin{pmatrix} a_t \\ f_t \end{pmatrix}, \begin{pmatrix} R_t & R_t F_t \\ (R_t F_t)' & Q_t \end{pmatrix}\right)$

$Cov(\Phi_t, Y_t) = Cov(\Phi_t, F_t \Phi_t + v_t) = Var(\Phi_t | D_{t-1}) F_t = R_t F_t$

MVN theory: $\Phi_t | Y_{t-1} \sim N(m_t, C_t)$

$m_t = a_t + A_t(Y_t - f_t)$ $\quad C_t = R_t(I - F_t A_t)$ $\quad A_t = R_t F_t Q_t^{-1}$

$MSE = e^2$ $\quad MAD(t) = abs(e)$

play with: $w >> v$ $\quad$ $v >> w \rightarrow$ problematic

The Dynamic Linear Model (DLM)

- given time θ_t, past, present future independent (Conditional independence)
  - $\{m_t, C_t\}$ Contains all relevant info about Fut. $\equiv$ suffic. for $\{Y_{t+1}, \theta_{t+1}, \dots, Y_{t+k}, \theta_{t+k}\}$
- Any Lin Comb. of indep. normal DLMs is a normal DLM
- $\theta_t' = (\theta_{t1}', \theta_{t2}')$ two series of Compont evolve independently
- $P(Y_t, \theta_t | D_{t-1}) = P(Y_t | \theta_t, D_{t-1}) \cdot P(\theta_t | D_{t-1})$ - inf from any relevant src
- Forecast of Component Consistent with total forecast
- Mgmnt by Except. { ① relevant src of info external to syst. ② Feedback
- DLM: characterized by quadruple $\{F, G, V, W\}_t = \{F_t, G_t, V_t, W_t\}$ $[n \times r]\ [n \times n]\ [r \times r]\ [n \times n]$

$\Rightarrow (Y_t | \theta_t) \sim N[F_t' \theta_t, V_t]$ $\rightarrow$ observation equation

$(\theta_t | \theta_{t-1}) \sim N[G_t \theta_{t-1}, W_t]$ $\rightarrow$ evolution, state, syst. eq.

$Y_t = F_t' \theta_t + v_t \sim N[0, V_t]$ $\quad \theta_t = G_t \theta_{t-1} + w_t \sim N[0, W_t]$ — evolution Error Var; Trans. Fed State matrix; observational Error

$\mu_t = F_t' \theta_t$: mean response / level

$\hookrightarrow$ design matrix of known values (independent variables)

Def: ① $\{F, G\}_t$: Constant $\Rightarrow$ time series DLM TSDLM

② TSDLM: obs. & evol. Var = cte $\forall t$: Const. DLM. $\{F, G, V, W\}$ $\leftarrow$ classical time series

general DLM: $\{F_t, G_t, V_t, W_t\}$

$(Y_t | \theta_t) \sim N[F_t' \theta_t, V_t]$ $\quad (\theta_t | \theta_{t-1}) \sim N[G_t \theta_{t-1}, W_t]$

obs. eq.: $Y_t = F_t' \theta_t + v_t$ $\quad v_t \sim N[0, V_t]$

syst. eq.: $\theta_t = G_t \theta_{t-1} + w_t$ $\quad w_t \sim N[0, W_t]$

Initial info: $(\theta_0 | D_0) \sim N[m_0, C_0]$

Covar and k-step marginal dist:

(a) State dist: $(\theta_{t+k} | D_t) \sim N[a_t(k), R_t(k)]$

(b) Forecast dist: $(Y_{t+k} | D_t) \sim N[f_t(k), Q_t(k)]$

(c) State Covar: $C[\theta_{t+k}, \theta_{t+j} | D_t] = C_t(k, j)$

(d) Obsn. Covar: $C[Y_{t+k}, Y_{t+j} | D_t] = F_{t+k}' C_t(k,j) F_{t+j}$

(e) Other Covar: $C[\theta_{t+k}, Y_{t+j} | D_t] = C_t(k, j) F_{t+j}$

$C[Y_{t+k}, \theta_{t+k} | D_t] = F_{t+k}' C_t(k, j)$

$f_t(k) = F_{t+k}' a_t(k)$ $\quad Q_t(k) = F_{t+k}' R_t(k) F_{t+k} + V_{t+k}$

$a_t(k) = G_{t+k} a_t(k-1)$ $\quad R_t(k) = G_{t+k} R_t(k-1) G_{t+k}' + W_{t+k}$

$C_t(k, j) = G_{t+k} C_t(k-1, j)$ $\quad k = j+1, \dots$

$a_t(0) = m_t$ $\quad R_t(0) = C_t$ $\quad C_t(j, j) = R_t(j)$

$V_t = V \ \forall t$ $\quad \Phi = 1/V$ observation percision

Conjugate: DLM definition:

obs. equation: $Y_t = F_t' \theta_t + v_t$ $\quad v_t \sim N[0, V]$

Sys. eq.: $\theta_t = G_t \theta_{t-1} + w_t$ $\quad w_t \sim N[0, V W_t^*]$

Initial info: $(\theta_0 | D_0, \Phi) \sim N[m_0, V C_0^*]$

$(\Phi | D_0) \sim G[\frac{n_0}{2}, \frac{n_0 S_0}{2}]$

initial quant: m_0, C_0^*, n_0, S_0, specif: $\{F_t, G_t, W_t^*\}$

$E(\Phi | D_0) = 1/S_0 \rightarrow$ prior point est. of obsv. variance V

Student t: θ: $[n \times 1]$ multivar T dis (h-deg fredm) mode m, postiv def. scale mtrx: C

$p(\theta) \propto \{h + (\theta - m)' C^{-1} (\theta - m)\}^{-\frac{(n+h)}{2}}$

$\theta \sim T_h[m, C]$ $\quad E(\theta) = m$ $\quad h > 1$ $\quad V(\theta) = \frac{Ch}{h-2}$ $\quad h > 2$

DLM dist result

① Cond on V: $(\theta_{t-1} | D_{t-1}, V) \sim N[m_{t-1}, V C_{t-1}^*]$

$(\theta_t | D_{t-1}, V) \sim N[a_t, V R_t^*]$

$(Y_t | D_{t-1}, V) \sim N[f_t, V Q_t^*]$

$(\theta_t | D_t, V) \sim N[m_t, V C_t^*]$

$a_t = G_t m_{t-1}$ $\quad R_t^* = G_t C_{t-1}^* G_t' + W_t^*$

$f_t = F_t' a_t$ $\quad Q_t^* = 1 + F_t' R_t^* F_t$

$e_t = Y_t - f_t$ $\quad A_t = R_t^* F_t / Q_t^*$

$m_t = a_t + A_t e_t$ $\quad C_t^* = R_t^* - A_t A_t' Q_t^*$

② for precision: $\Phi = V^{-1}$

$(\Phi | D_{t-1}) \sim G[\frac{n_{t-1}}{2}, \frac{n_{t-1} S_{t-1}}{2}]$

$(\Phi | D_t) \sim G[\frac{n_t}{2}, \frac{n_t S_t}{2}]$

③ Unconditional on V:

$(\theta_{t-1} | D_{t-1}) \sim T_{n_{t-1}}[m_{t-1}, C_{t-1}]$

$(\theta_t | D_{t-1}) \sim T_{n_{t-1}}[a_t, R_t]$

Bayesian Seminar ⑤

$(Y_t \mid D_{t-1}) \sim T_{n_{t-1}}[f_t, Q_t]$ $\quad R_t = S_{t-1} R_t^*$

$(\theta_t \mid D_t) \sim T_{n_t}[m_t, C_t]$ $\quad Q_t = S_{t-1} Q_t^*$

$C_t = S_t C_t^*$

④ operational def. of updating eq:

$Q_t = F_t' R_t F_t + S_{t-1}$ $\quad A_t = R_t F_t / Q_t$

$n_t = n_{t-1} + 1$ $\quad S_t = S_{t-1} + \frac{S_{t-1}}{n_t}\left(\frac{e_t^2}{Q_t} - 1\right)$

$m_t = a_t + A_t e_t$ $\quad C_t = \frac{S_t}{S_{t-1}}(R_t - A_t A_t' Q_t)$

- Conditionally independent model components
- Forecast as prob. dist. - incorp expert info
- sequential model def. & anal. $\Rightarrow$ as info arrives mdl param revised

step ahead forecast: $p(Y_t \mid \Phi_t, D_{t-1})$ $\quad$ prior $(\Phi_t \mid D_{t-1})$

posterior: $p(\Phi_t \mid y_t, D_{t-1})$

- simple reg: (sales as func. of prom. elast. change in time

$y_t = \alpha_t + \beta_t X_t + v_t$ $\quad v_t \sim N(0, V)$

$\alpha_t = \alpha_{t-1} + \delta\alpha_t$ $\quad$ - info set: $\boxed{D_t = D_{t-1} \cup y_t}$

$\beta_t = \beta_{t-1} + \delta\beta_t$ $\rightarrow$ random changes

- when $w \gg v$ responding properly could happen, but
$v \gg w \equiv$ data noisy you cant

- θ_t: latent measure (goal: find true measure)
- Dynamic bayesian $\equiv$ dynamic w.r.t parameter
- monitoring and assessment through bays factors

posterior at t-1: $(\Phi_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1})$

$y_t = \Phi_t + v_t$ $\quad$ Prior: $(\Phi_t \mid D_{t-1}) \sim N(m_{t-1}, R_t)$ $\quad R_t = C_{t-1} + W$

$\Phi_t = \Phi_{t-1} + \omega_t$ $\quad$ 1-step ahead: $(y_t \mid D_{t-1}) \sim N(f_t, Q_t)$, $f_t = m_{t-1}$, $Q_t = R_t + V$

fore cast error, e_t $\quad$ posterior at t: $(\Phi_t \mid D_t) \sim N(m_t, C_t)$

intuition: $m_t = m_{t-1} + A_t(y_t - f_t)$, $C_t = (R_t - R_t^2/Q_t) = R_t V / Q_t$

rate of update of mean $\Leftarrow$ $A_t = R_t / Q_t$ $\Rightarrow$ bigger when Q_t lower
bound by one $\equiv$ data more informative

Variance evolue: $C_t = R_t\left(1 - \frac{R_t}{R_t + V_t}\right)$ learning as info comes

- in sample measure: MAD: mean absolute deviation; MSE: mean square errors

$[n \times r]$ $[n \times n]$ $[r \times r]$ $[n \times n]$

- $\{F_t, G_t, V_t, W_t\}$: $\boxed{\text{well fully defines}}$ DLM $\quad$ $Y_t \mid \Phi_t \sim N[F_t' \Phi_t, V_t]$
$\Phi_t \mid \Phi_{t-1} \sim N[G_t \Phi_t, W_t]$

intuition - next side of bar $\equiv$ param is now known $\equiv$ data $\equiv$ no variation

- trajectory of param = e.g. adv. goodwill

intuition - latent sample: utility, goodwill (no singularity problem)

- No data: moment prior = posterior moment

R_t: inflated (slight inflation in error)

intuition: keep have missing data $\equiv$ increased uncertainty
missing data tells us something about it = uncertainty

- out of sample forecast: break 100 sample into two
80 for in sample 20 for out of sample

- k-step ahead

- smoothing by going back: Retrospective approach ⑥

- $m_t = m_{t-1} + A_t(y_t - f_t)$

$A_t = R_t / Q_t = \frac{C_{t-1} + 1}{C_{t-1} + 1 + \frac{V}{W}}$ $\Rightarrow$ $\frac{dA_t}{d(V/W)} < 0$

V/W influences how posterior mean adapts to data

$MAD = \frac{\sum_{t=1}^T |e_t|}{T}$ $\quad MSE = \frac{\sum_{t=1}^T e_t^2}{T}$ $\quad e_t = y_t - f_t$

two phases: ① use prior $(\Phi_t \mid D_{t-1}) \Rightarrow$ calc posterior $(\Phi_t \mid D_t)$
② take posterior $\Phi_{t-1} \mid D_{t-1} \Rightarrow$ calc prior next $(\Phi_t \mid D_t)$ (main equation)

$\boxed{\text{phase 2}}$:
posterior: $\Phi_{t-1} \mid D_{t-1} \sim N(m_{t-1}, C_{t-1})$

Calc prior: $\Phi_t \mid D_{t-1} \sim N(a_t, R_t)$

$E(\Phi_t \mid D_{t-1}) = a_t = G_t m_{t-1}$

$Var(\Phi_t \mid D_{t-1}) = R_t = G_t C_{t-1} G_t' + W_t$

① pred. dist: $Y_t \mid D_{t-1} \sim N(f_t, Q_t)$

$E(Y_t \mid D_{t-1}) = f_t = F_t a_t$ $\quad Var(Y_t \mid D_{t-1}) = Q_t = F_t' R_t F_t + W_t$

② posterior cont: $\Phi_t, Y_t \mid D_{t-1} \sim N\left[\begin{Bmatrix} a_t \\ f_t \end{Bmatrix}, \begin{Bmatrix} R_t & R_t F_t \\ (R_t F_t)' & Q_t \end{Bmatrix}\right]$

$Cov(\Phi_t, Y_t) = Cov(\Phi_t, F_t \Phi_t + v_t) = Var(\Phi_t \mid D_{t-1}) F_t = R_t F_t$

③ k-step ahead forecast: $f_t(k) = E(Y_{t+k} \mid D_t) = F_{t+k} a_t(k)$
$Q_t(k) = F_{t+k}' R_t(k) F_{t+k} + V_{t+k}$

recursive calc: $a_t(k) = G_{t+k} a_t(k-1)$

$R_t(k) = G_{t+k} R_t(k-1) G_{t+k}' + W_{t+k}$

$a_t(0) = m_t$ $\quad R_t(0) = C_t$

④ $\boxed{\text{Retrospective analysis}}$:

$\Phi_t, \Phi_{t-1} \mid D_{t-1} \sim N\left[\begin{Bmatrix} a_t \\ m_{t-1} \end{Bmatrix}, \begin{Bmatrix} R_t & G_t C_{t-1} \\ (G_t C_{t-1})' & C_{t-1} \end{Bmatrix}\right]$

push Φ_t to the other side:

$\Phi_{t-1} \mid \Phi_t, D_{t-1} \sim N[m_t + B_t(\Phi_t - a_t), C_{t-1} - B_t' R_t B_t]$

where $B_t = G_t C_{t-1} R_t^{-1}$ $\Rightarrow$ update idea about past

$E(\Phi_{t-1} \mid D_t) = E(E(\Phi_{t-1} \mid \Phi_{t-1}, D_{t-1}) \mid D_{t-1}) = E(m_{t-1} + B_t(\Phi_t - a_t)) \mid$
$= m_{t-1} + B_t(m_t - a_t)$

$Var(\Phi_{t-1} \mid D_t) = Var(E(\Phi_{t-1} \mid \Phi_{t-1}, D_{t-1}) \mid D_t) + E(Var(\Phi_{t-1} \mid \Phi_t, D_{t-1})$
$\mid D_t) = C_{t-1} - B_t' R_t B_t + B_t' C_t B_t = C_{t-1} - B_t'(R_t - C_t) B_t$

⑤ Φ_t, y_t conditionally independent given Φ_{t+1}, since it know Φ_{t-1} don't need y_{t+1} $\Rightarrow$ sequence of conditions (once we know Φ_t, Φ_{t-1} does not require the data)

⑥ Forward filtering backward sampling (FFBS):

① go forward and get: $(\Phi_t \mid D_t) \sim N(m_T, C_t)$
determine dist. fully

② go backward: $p(\Phi_t \mid \hat{\Phi}_{t+1}, D_t)$

⑦ MCMC baysian: not just point estimate, but dist.

⑧ Conditioned independant:

$\Phi^T = \{\Phi_0, \Phi_1, \Phi_2, \ldots, \Phi_T\}$

$P(\Phi^T \mid D_T) = P(\Phi_T \mid D_T) P(\Phi_{T-1} \mid \Phi_T, D_{T-1}) \cdots P(\Phi_1 \mid \Phi_2, D_1) \cdot P(\Phi_0 \mid \Phi_1, D_0)$

Bayesian seminar

9 simulation based computation:
objective: Joint posterior sample: $p(\Phi^T | D_T)$

10 each cycle: FFBS
1. first do FF: get m_t, C_t
2. Do bw sampling: access dist
draw last & come back

11 m_0, C_0: initial guess ⇒ revised in bw

12 if not have variance ↳ { 1. Frequentist: max likelihood; 2. Bayesian: simple reg. Gibbs sampler }

13 independent ⇒ estimate each variance separately

14 Elements of F and G: { 1. weirdest 2. forgetting 3. Ban et al. }

15 MCMC in DLM:
Repeat $p(\Omega, \theta | D_T)$ iterative resample
1. apply FFBS to draw θ^* from $p(\theta | \Omega^*, D_T)$
2. Draw new value Ω^* from $p(\Omega | \theta^*, D_T)$

16 standard Gibbs - Conjugate prior for ∀ compnt of Ω^*
if not ⇒ metropolis hast.
other Monte Carlo

17 $y_t = \alpha_t + \beta_t' X_t + \nu_t$ $\nu_t \sim N(0, V)$ ≡ Rand. walk Param evol.
$\alpha_t = \alpha_{t-1} + w_{1t}$, $\beta_t = \beta_{t-1} + w_{2t}$ $(w_{1t}, w_{2t}) \sim N(0, W)$
$\psi_t = (\alpha_t, \beta_t)$: State vector $\psi_0 \sim N(m_0, C_0)$

$p(\psi^T | y^T, V, W)$ FFBS
↓
$p(W, V | \psi_t, y^T)$ Gibbs Sampling

18 $F_t = [1 \ \ X_t]$ G: identity matrix (transition matrix)

19 Pritt: Fix parameter effect

Reversible Jump MCMC Factor loadings

1. number of factors k as parameter
2. k-factor model: $y_t = \beta f_t + \varepsilon_t$ $f_t \sim iid\ N(0, I_k)$
β: [m×k] $\varepsilon \sim iid\ N(0, \Sigma)$, $diag(\sigma_1^2, \dots, \sigma_m^2)$

$\Omega = V(y_t | \Omega) = V(y_t | \beta, \Sigma)$ $\Omega = \beta\beta' + \Sigma$

$v(y_{it} | f) = \sigma_i^2$ $Cov(y_{it}, y_{jt} | f) = 0$ $Var(y_{it}) = \Sigma_{l=1}^{k} \beta_{il}^2 + \sigma_i^2$

$Cov(y_{it}, y_{jt}) = \Sigma_{l=1}^{k} \beta_{il}\beta_{jl}$

$y = F\beta' + E \to [T \times m]$
[T×m] [T×k]

$p(y | F, \beta, \Sigma) \propto |\Sigma|^{-T/2} etr(-.5\, \Sigma^{-1} E'E)$
↳ $exp(tr(A))$

$\Rightarrow p(y | \beta, \Sigma) \propto |\Omega|^{-T/2} etr(-.5\, \Omega^{-1} y'y)$

model invariance → $\beta^* = \beta P'$, $f_t^* = P \cdot f_t$ P → any orth. k×k matrix
block lower triang. matrix (positive semi def.)
identification & useful interpretation
loading matrix elements: $r = mk - \frac{k(k-1)}{2}$ Free param

β rank r: $r < k \Rightarrow \exists\ Q: \beta Q = 0,\ Q'Q = I$
orthogonal matrix M: $\Omega = \beta\beta' + \Sigma = (\beta + MQ')(\beta + MQ')' + (\Sigma - MM')$ lack of identification, symmetry & multimodal

3 Prior: $\beta_{ij} \sim N(0, C_0)$ when $i \neq j$ ↳ rather large
$\beta_{ii} \sim N(0, C_0) \cdot 1(\beta_{ii} > 0)$ upper diag. $i = 1, \dots, k$ ⇒ positive diag
σ_i^2: idiosyncratic variance $\sim IG$, indep $(\nu/2, \nu s^2/2)$
prior deg frdm ← prior ↵
(low to diffuse prior)

$\otimes\ p(\sigma_i^2) \propto \frac{1}{\sigma_i^2}$ Heywood problem (tend to zero) ⇒ singularity

4 Full Cond posterior:
$f_t \sim N((I_k + \beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}y_t, (I_k + \beta'\Sigma^{-1}\beta)^{-1})$
$\sigma_i^2 \sim IG((\nu + T)/2, (\nu s^2 + d_i)/2)$ $d_i = (y_i - F\beta_i')'(y_i - F\beta_i')$

5 Full Conditional: $\beta_i \sim N(m_i, C_i) 1(\beta_{ii} > 0)$ $i = 1 - k$
$m_i = C_i(C_0^{-1}\mu_0 1_i + \sigma_i^{-2} F_i' y_i)$
$C_i^{-1} = C_0^{-1} I_i + \sigma_i^{-2} F_i' F_i$
$\beta_i \sim N(m_i, C_i)$ $i = k+1, \dots, m$
$m_i = C_i(C_0^{-1}\mu_0 1_k + \sigma_i^{-2} F' y_i)$
$C_i^{-1} = C_0^{-1} I_k + \sigma_i^{-2} F'F$

6 dependance on k show $\beta \to \beta_k$, $F \to F_k$ $\theta_k = (\beta_k, \Sigma)$

7 RJMCMC uses Metropolis Hasting $(k, \theta_k) \longrightarrow (k', \theta_{k'})$
reversible jump to maintain ~~jump~~ detail balance of chain diff. dim. ↵

8 for $\forall k \in K$ $q_k(\beta_k) = N(b_k, b B_k)$ $i = 1, \dots, m$
$q_k(\sigma_i^2) = IG(a, a \nu_{ki}^2)$ a, b scale param to specify

$\Rightarrow q_k(\theta_k) \equiv q_k(\beta_k, \Sigma) = q_k(\beta_k) \prod_{i=1}^{m} q_k(\sigma_i^2)$

9 accept reject ratio:
$\alpha = \min\left\{1, \frac{p(y | k', \theta_{k'}) \cdot p(\theta_{k'} | k') p(k')}{p(y | k, \theta_k) \cdot p(\theta_k | k) \cdot p(k)} \frac{q_k(\theta_k) J(k' \to k)}{q_{k'}(\theta_{k'}) J(k \to k')}\right\}$
likelihood ↓ prior density

Spatial Dynamic Factor analysis 2008

1. temporal dependance → latent factors (trend, seasonality)
spatial dependance → Factor loadings (dtrm, stoch)
(reduce complexity)

2. nonseperable & nonstationary space-time models (standard dynamic factor model):
$y_t = \mu_t^{y} + \beta f_t + \varepsilon_t$ $\varepsilon_t \sim N(0, \Sigma)$
mean level of space time ← $f_t = \Gamma f_{t-1} + w_t$ $w_t \sim N(0, \Lambda)$
$y_t = (y_{1t}, \dots, y_{Nt})'$ N-dim. vect. of obs. location $s_1, \dots, s_N$ time $t = 1, \dots, T$
f_t: m dim. vect. of common factor $m \le N$
$\Sigma = diag(\sigma_1^2, \dots, \sigma_N^2)$ $\Lambda = diag(\lambda_1, \dots, \lambda_m)$
dynamic evolution of factors $\Gamma = diag(\gamma_1, \dots, \gamma_m)$

3. Conditionally independent, distance-based Gaussian process or Gaussian random field (GRF)
$\beta_{(j)} \sim GRF(\mu_j^{\beta^*}, \tau_j^2 \rho_{\phi_j}(\cdot)) \equiv N(\mu_j^{\beta^*}, \tau_j^2 R_{\phi_j})$
$r_{lk} = \rho_{\phi_j}(|s_l - s_k|)$ $l, k = 1, \dots, N$ $j = 1 \dots m$

~~Spatial Dynamic~~ Bayesian seminar

(4) Bessel function

(5) mean level of the ~~time~~ space-time process options μ_t^{y*}

i) Constant mean level model: $\mu_t^{y*} = \mu^y \quad \forall t$

ii) regression model: $\mu_t^{y*} = X_t^y \mu^y$

$$X_t^y = (1_N, X_{1t}^y, \dots, X_{qt}^y) \quad q \text{ time varying cov}$$

iii) dynamic coeff model

$$\mu_t^{y*} = X_t^y \mu_t^y \qquad \mu_t^y \sim N(\mu_{t-1}^y, W)$$

(6) $Cov(y_{it}, y_{it+h}) = \Sigma_{k=1}^m (\lambda_k \gamma_k^h)(1-\gamma_k^2)^{-1}(\Gamma_k^2 \rho(u,\phi_k) + \mu_{ik}^\beta \mu_{jk}^\beta)$

(7) weekly seasonal: $\beta = (\beta_{(1)}, 0, \dots, \beta_{(h)}, 0)$ $\quad T = diag(T_1, \dots, T_h)$

$$T_2 = \begin{pmatrix} \cos(2\pi l/p) & \sin(2\pi l/p) \\ -\sin(2\pi l/p) & \cos(2\pi l/p) \end{pmatrix} \quad l = 1, \dots, h = p/2$$

(8) $\mu_t^{y*} = 0, \quad \mu_j \beta^* = x_j^\beta \mu_j^\beta \quad \Rightarrow \quad y = F\beta' + \varepsilon \quad y = (y_1, \dots, y_T)'$ $[T \times N]$

$F = (f_1, \dots, f_T)'$ $[T \times m]$

likelihood: $p(y \mid \theta, F, \beta, m) = (2\pi)^{-TN/2} |\Sigma|^{-T/2} \, etr\{-\frac{1}{2}\Sigma^{-1}(y-F\beta')' \ast (y - F\beta')\}$

$\theta = (\sigma, \lambda, \gamma, \mu, \tau, \phi) \qquad \sigma = (\sigma_1^2, \dots, \sigma_N^2)'$

$\lambda = (\lambda_1, \dots, \lambda_m)' \qquad \gamma = (\gamma_1, \dots, \gamma_m)' \qquad \mu = (\mu_1^\beta, \dots, \mu_m^\beta)$

$\tau = (\tau_1^2, \dots, \tau_m^2)' \qquad \phi = (\phi_1, \dots, \phi_m)' \qquad etr(x) = exp(trace(x))$

(9) Conjugate prior

(10) $\boxed{\text{prior}}$ $f_0 \sim N(m_0, C_0) \qquad \sigma_i^2 \sim IG(n_\sigma/2, n_\sigma S_\sigma/2) \quad i=1-N$

$\lambda_j \sim IG(n_\lambda/2, n_\lambda S_\lambda/2)$

(11) γ specification: (i) $\gamma_j \sim N tr_{(-1,1)}(m_\gamma, S_\gamma)$ — truncate normal $(-1,1)$

Prior

(ii) $\gamma_j \sim \alpha N tr_{(-1,1)}(m_\gamma, S_\gamma) + (1-\alpha)\delta_1(\gamma_j)$

$\alpha \in (0,1]$, $\alpha, m_\gamma, S_\gamma$ hyperparam.

(12) $\mu_j^\beta, \phi_j, \tau_j^2$:

(i) vague but proper priors

(ii) reference type: $\mu_j^\beta \sim N(m_\mu, S_\mu)$, $\phi_j \sim IG(2, b)$

$\tau_j \sim IG(n_\tau/2, n_\tau S_\tau/2)$ hyperparam: $m_\mu, S_\mu, n_\tau S_\tau$

(13) $\pi_{IG}(\tau_j^2, \phi_j) = \pi_{IG}(\tau_j^2)\pi_{IG}(\phi_j) \propto \tau_j^{-(n_\tau+2)} e^{-n_\tau S_\tau/\tau_j^2} \phi_j^{-3} e^{-b/\phi_j}$

(14) $\pi_R(\mu_j^\beta, \tau_j^2, \phi_j) = \pi_R(\mu_j^\beta \mid \tau_j^2, \phi_j)\, \pi_R(\tau_j^2, \phi_j)$

$\pi_R(\mu_j^\beta \mid \tau_j^2, \phi_j) = 1$

$\pi_R(\tau_j^2, \phi_j) = \pi_R(\tau_j^2)\pi_R(\phi_j) \propto \tau_j^{-2} \{tr(W_{\phi_j}^2) - \frac{1}{N-p_j}[tr(W_{\phi_j})]^2\}^{1/2}$

$W_{\phi_j} = ((\partial/\partial\phi_j) R\phi_j) R\phi_j^{-1} (I_N - X_j^\beta (X_j^{\beta'} R\phi_j^{-1} X_j^\beta)^{-1} X_j^{\beta'} R\phi_j^{-1})$

(15) $\boxed{\text{fore cast}}$ $p(y_{T+h} \mid y) = \int p(y_{T+h} \mid f_{T+h}, \beta, \theta) \cdot p(f_{T+h} \mid f_T, \beta, \theta)$

$\ast\, p(f_T, \beta, \theta \mid y)\, df_{T+h}\, df_T\, d\beta\, d\theta$

$(y_{T+h} \mid f_{T+h}, \beta, \theta) \sim N(\beta f_{T+h}, \Sigma)$

$(f_{T+h} \mid f_T, \beta, \theta) \sim N(\mu_h, V_h) \qquad \mu_h = T^h f_T$

$V_h = \Sigma_{k=1}^h T^{k-1} \Lambda (T^{k-1})'$

(16) $\boxed{\text{Interpolation}}$

$S_n = \{S_{N+1}, \dots, S_{N+N_n}\}$

$p(\beta^n \mid y^o) = \int p(\beta^n \mid \beta^o, \theta)\, p(\beta^o, \theta \mid y^o)\, d\beta^o\, d\theta$

$p(\beta^n \mid \beta^o, \theta) = \pi_{j=1}^m p(\beta_{(j)}^n \mid \beta_{(j)}^o, \mu_j^\beta, \tau_j^2, \phi_j)$

$$\begin{pmatrix} \beta_{(j)}^o \\ \beta_{(j)}^n \end{pmatrix} \sim N\left[\begin{pmatrix} x_j^{\beta^o} \\ x_j^{\beta^n} \end{pmatrix}\mu_j^\beta, \tau_j^2 \begin{pmatrix} R\phi_j^o & R\phi_j^{o,n} \\ R\phi_j^{n,o} & R\phi_j^n \end{pmatrix}\right]$$

$\beta_{(j)}^n \mid \beta_{(j)}^o, \theta \sim N(x_j^{\beta^n}\mu_j^\beta + R\phi_j^{n,o} R\phi_j^{o\,-1}(\beta_{(j)}^o - x_j^{\beta^o}\mu_j^\beta),$

$\tau_j^2(R\phi_j^n - R\phi_j^{n,o} R\phi_j^{o\,-1} R\phi_j^{o,n}))$

$\hat{p}(\beta^n \mid y^o) = L^{-1}\Sigma_{l=1}^L p(\beta^n \mid \beta^{o(l)}, \theta^{(l)})$

$\hat{E}(y^n \mid y^o) = L^{-1}\Sigma_{l=1}^L \beta^{n(l)} f^{(l)}$

(17) Conditional on m, posterior dist. of (F, β, θ)

$p(F, \beta, \theta \mid y) \propto \pi_{t=1}^T p(y_t \mid f_t, \beta, \sigma)\, p(f_0 \mid m_0, C_0) \prod_{t=1}^T p(f_t \mid f_{t-1}, \beta, \theta)$

$\times \prod_{j=1}^m p(\beta_{(j)} \mid \mu_j^\beta, \tau_j^2, \phi_j)\, p(\gamma_j)\, p(\lambda_j) \cdot p(\mu_j^\beta)\, p(\tau_j^2, \phi_j)$

$\times \pi_{i=1}^N p(\sigma_i^2)$

(18) $\Psi_m = (F_m, \beta_m, \theta_m)$ $\quad \boxed{\text{proposal dist.}}$

$q_m(\Psi_m) = \pi_{j=1}^m f_N(f_{(j)} \mid M_{f(j)}, a V_{f(j)})\, f_N(\beta_{(j)} \mid M_{\beta(j)}, b V_{\beta(j)})$

$f_N(\gamma_j \mid M_{\gamma_j}, c V_{\gamma_j}) \times \prod_{j=1}^m f_{IG}(\lambda_j \mid d, d M_{\lambda_j})\, f_N(\mu_j \mid M_{\mu_j^\beta}, e V_{\mu_j^\beta})$

$f_{IG}(\phi_j \mid f, f M_{\phi_j}) \times \pi_{j=1}^m f_{IG}(\tau_j^2 \mid g, g M_{\tau_j})\, f_{IG}(\sigma_j^2 \mid h, h M_{\sigma_j^2})$

a, b, c, d, e, f, g: tuning param $\quad M_x, V_x$ sample mean, var

$a, b, c, e < 1$, $\quad 1.5 < d, f, g$

$p(y, m, \Psi_m) = p(y \mid m, \Psi_m)\, p(\Psi_m \mid m)\, pr(m)$

(19) metropolis hastings

accept $(m', \Psi_{m'})$ with probability:

$$\alpha = \min\left\{1, \frac{p(y, m', \Psi_{m'})}{p(y, m, \Psi_m)} \times \frac{q_m(\Psi_m)\, q(m', m)}{q_{m'}(\Psi_{m'})\, q(m, m')}\right\}$$

Bayesian Seminar   Marketing papers   ①

① Wearout Effects of different advertising themes, (Bam, Bruce 2007) – adv. sales relationship

- theme: price vs. prod ad.   - wearout vs. forgetting
- reallocation of resources across themes → improvmnt dmnd

Rsrch Q: ① How diffrnt wearout effect for diffrnt theme?
② How researcher can assess mgntd of wearout effect?
③ How nature of interaction b/w themes?
④ How alloc resrc for adv bdgt?

- Gibbs sampling and DLM
- Dynamic path of response coeff over time

Literature related
① Response model
② wearin and wearout
③ Forgetting
④ Differential effects of wearout across themes (Exprmnt)
⑤ Need for time-varying coeff.
⑥ need for interaction

- demand ~ advertising expenditure

$$\underbrace{\frac{dG(t)}{dt}}_{\text{rate of change in Goodwill}} = \underbrace{q}_{\text{effectiveness of spending}} \underbrace{A(t)}_{\text{adv expend.}} - \underbrace{\delta}_{\text{rate of decay (forgetting)}} G(t)$$

copy wearout param, repitition

$$\frac{dq}{dt} = -a(A)q + (1-I(A))\delta(1-q) \qquad a(A) = c + wA(t)$$

$$\frac{dG}{dt} = \sum_{i=1}^{m} q_i \big(\underbrace{g(A_i)}_{\text{func adv. expndtr}} + \underbrace{\lambda_i \sum_{j=1, j\neq i}^{m} h(A_i, A_j)}_{\text{adv expnd. each theme}}\big) - \delta \underbrace{G_t}_{\text{Goodwill}}$$

m → # themes; interaction

$g(A_i) = \ln(1+A)$ reason: diminishing return

$h(A_i, A_j) = \ln(1+A_i)\ln(1+A_j)$

change in effectiveness $\quad \frac{dq_i}{dt} = -a(A_i)q_i + \delta\{(1-I(A_i))\}\,{}_{*}(1-q_i) \quad i = 1, 2, \dots, m$

$a(A_i) = c_i + w_i A_{it}$

DLM good method to handle non-stationarity

Compete random coeff: problem of no est. point in time

demand ← $y_t = G_t + \beta X_t + \varepsilon_t$ where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$
↳ price, number of lines, and competitive adv

w: Instruments: (1) retail price index (2) # household (3) Con Sntmnt (4) household spending

$P_t = P_t(w|\alpha) + \eta_t$

$$y_t = \underbrace{[1\ 0 \cdots 0]}_{F_t} \begin{bmatrix} G_t \\ q_{1t} \\ \vdots \\ q_{mt} \end{bmatrix} + \beta' X_t + \varepsilon_t$$

$$\underbrace{\begin{bmatrix} G_t \\ q_{1t} \\ \vdots \\ q_{mt} \end{bmatrix}}_{\Phi_t} = \underbrace{\begin{bmatrix} (1-\delta) & \tilde{g}(A_{1t}) & \cdots & \tilde{g}(A_{mt}) \\ 0 & (1-a(A_{1t})) & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & (1-a(A_{mt}))-\delta(1-I(A_{mt})) \end{bmatrix}}_{H_t} \underbrace{\begin{bmatrix} G_{t-1} \\ q_{1t-1} \\ \vdots \\ q_{mt-1} \end{bmatrix}}_{\Phi_{t-1}}$$

$$+ \underbrace{\begin{bmatrix} 0 \\ \delta(1-I(A_{1t})) \\ \vdots \\ \delta(1-I(A_{mt})) \end{bmatrix}}_{u_t} + \underbrace{\begin{bmatrix} w_{0t} \\ w_{1t} \\ \vdots \\ w_{mt} \end{bmatrix}}_{w_t}$$

$\tilde{g}(A_{it}) = g(A_i) + \lambda_i \sum_{i\neq j}^{m} h(A_{it}, A_{jt}) \quad i = 1, \dots, m$

$y_t = F_t \Phi_t + \beta X_t + \varepsilon_t \qquad \Phi_t = H_t \Phi_{t-1} + u_t + w_t \qquad \varepsilon_t \sim N[0, \sigma_\varepsilon^2] \quad w_t \sim N(0, W]$

DLM spec: $\{F_t, H_t, \sigma_\varepsilon^2, W\}$

②

Solve large scal nonliner:

$$\max_{A_{11} \cdots A_{1m} \cdots A_{T1} \cdots A_{TM}} \sum_{t=1}^{T} E(y_t | D_{t-1})$$

s.t. $\sum_{i=1}^{m} A_{it} \le b_t \quad A_{it} \ge 0 \quad t = 1, \dots, T$ linear constraint

$Var(y_t | D_{t-1}) \le \sigma_t^2 \quad t = 1, \dots, T$

② Dynamic effectiveness of advertising and word of mouth in sequential dist. of new products
Bruce, Foutz, ~~Kol~~ SARICI 2012

- sequential stages release (adv. dist. stges) – spill over
- WOM more powerful at later stages
- reallocate advertising   – windowing (sequential distribution)
- Extends product life cycle
- forgetting vs. carryover   – wearout vs. wear-in
- spillover advertisement across stages
- WOM effectiveness interaction with advertising effectiveness

Q: ① how Ad effe. & WOM effect fluc across dist stgs
② how differ and interact?
③ how vary across products?
④ more efficient to alloc. more adv?

- Compare across the stages

related literature
① Sequential Distribution
② Dynamic Ad Effectiveness
③ Dynamic WOM Effectiveness
④ Aggregate Demand and Goodwill stock

$$y_{it}^{(j)} = \underbrace{G_{it}^{(j)}}_{\text{Goodwill}} + \beta_i^{(j)} X_{it}^{(j)} + \varepsilon_{it}^{(j)} \quad \text{(N-A model)} \quad \text{where } \varepsilon_{it}^{(j)} \sim N(0, \sigma_i^{2(j)})$$

→ Forgetting rate

$$G_{it}^{(j)} = (1-\delta_i^{(j)}) G_{it}^{(j)} + q_{1it}^{(j)} g(\underbrace{A_{it}^{(j)}}_{\text{studio spending}} - 1) + q_{2it-1}^{(j)} \underbrace{R_{it}^{(j)}}_{\text{online rating}} + \underbrace{w_{1it}^{(j)}}_{\text{lagged goodwill}}, \quad w_{1it} \sim N(0, w_i^{(j)})$$

$g(A) = \ln(1+A)$: diminishing return of advertising

Copy wearout: time ⇒ (0)   rate of forgetting
repitition wearout ⇒ (w)

$$q_{1it}^{(j)} = (1 - O_{1i}^{(j)} - w_{1i}^{(j)} A_{it}^{(j)}) q_{1it}^{(j)} + \delta_i^{(j)} [1 - I(A_{it}^{(j)})](1 - q_{1it}^{(j)}) + \varphi_{1i}^{(j)} q_{2it-1}^{(j)} I(Q_{it}^{(j)}) + w_{2it}^{(j)}$$

→ word of mouth effectiveness; → interaction; ↳ ad effectiveness

$$I(A_{it}^{(j)}) = \begin{cases} 1 & \text{if } A_{it}^{(j)} \neq 0 \\ 0 & \text{if } A_{it}^{(j)} = 0 \end{cases}, \quad I(G_{it}^{(j)}) = \begin{cases} 1 & \text{if } G_{it}^{(j)} \neq 0 \\ 0 & \text{if } G_{it}^{(j)} = 0 \end{cases}$$

$w_{2it}^{(j)} \sim N(0, w_{2i}^{(j)})$

WOM evolution:

$$q_{2it}^{(j)} = (1 - O_{2i}^{(j)}) q_{2it-1}^{(j)} + w_{2i}^{(j)} Q_{it}^{(j)} (1 - q_{2it-1}^{(j)}) + \varphi_{2i}^{(j)} q_{1it}^{(j)} I(A_{it}^{(j)}) + w_{3it}^{(j)}, \quad \text{where } w_{3it}^{(j)} \sim N(0, w_{3i}^{(j)})$$

$$\{\underbrace{\beta_i^{(j)}}_{\text{demand param}}, \underbrace{\delta_i^{(j)}}_{\text{forgetting rate}}, \underbrace{\bar{\omega}_i}_{\text{wearout wearin rate}}, \underbrace{\varphi_{1i}^{(j)}, \varphi_{2i}^{(j)}}_{\text{ad-WOM interdp. parameters}}\} = \Omega^{(j)} Z_i + \xi_i^{(j)} \sim N(0, M^{(j)})$$

→ set of film charact. e.g genre, sequel

$\{O_{1i}^{(j)}, w_{1i}^{(j)}, O_{2i}^{(j)}, w_{2i}^{(j)}\}$

obs. eq: $$y_{it}^{(j)} = \underbrace{[1\ 0\ 0]}_{F} \begin{bmatrix} G_{it}^{(j)} \\ q_{1it}^{(j)} \\ q_{2it}^{(j)} \end{bmatrix} + \beta_i^{(j)} X_{it}^{(j)} + \varepsilon_{it}^{(j)}$$

St. eq: $$\underbrace{\begin{bmatrix} G_{it}^{(j)} \\ q_{1it}^{(j)} \\ q_{2it}^{(j)} \end{bmatrix}}_{\theta_{it}^{(j)}} = \underbrace{\begin{bmatrix} (1-\delta_i^{(j)}) & g(A_{it}^{(j)}) & R_{it}^{(j)} \\ 0 & (1-O_{1i}^{(j)} - \omega_{1i}^{(j)} A_{it}^{(j)}) & \varphi_{1i}^{(j)} I(G_{it}^{(j)}) \\ 0 & \varphi_{2i}^{(j)} I(A_{it}^{(j)}) & (1-O_{2i}^{(j)} - \omega_{2i}^{(j)} G_{it}^{(j)}) \end{bmatrix}}_{H_{it}^{(j)}}$$

$$\underbrace{\begin{bmatrix} G_{it-1}^{(j)} \\ q_{1it-1}^{(j)} \\ q_{2it-1}^{(j)} \end{bmatrix}}_{\theta_{it-1}^{(j)}} + \underbrace{\begin{bmatrix} 0 \\ \delta_i^{(j)}[1 - I(A_{it}^{(j)})] \\ \omega_{2i}^{(j)} Q_{it}^{(j)} \end{bmatrix}}_{u_{it}^{(j)}} + \underbrace{\begin{bmatrix} \omega_{1it}^{(j)} \\ \omega_{2it}^{(j)} \\ \omega_{3it}^{(j)} \end{bmatrix}}_{\omega_{it}^{(j)}}$$

$$\Rightarrow y_{it}^{(j)} = F\theta_{it}^{(j)} + \beta_i^{(j)} X_{it}^{(j)} + \varepsilon_{it}^{(j)} \qquad \varepsilon_{it}^{(j)} \sim N(0, \sigma_i^{2(j)})$$

$$\theta_{it}^{(j)} = H_{it}^{(j)} \theta_{it}^{(j)} + u_{it}^{(j)} + \omega_{it}^{(j)} \qquad \omega_{it}^{(j)} \sim N(0, W_i^{(j)})$$

① model comparisons

② optimization $\max_{A_{it}^{(j)}} \sum_t E(y_{it}^{(j)})$. s.t. $\sum_t A_{it}^{(j)} \le b_i^{(j)}$ and $A_{it} \ge 0$

---

① Themes of advertising:

① Reconnect ad. ③ Emotional ad. ⑤ Price ad
② Reassurance ad. ④ product ad

② optimization: optimize one step ahead Expectation

---

1. Discovering how advertising grows sales and builds Brands. Bruce, Peters, Naik 2012

① think-feel-do hierarchy, intermediate effect of adv. induce sales (importance of intermediary role) Factor of mindset matrix
② Cognition, affect, and experience: unobservable
③ batteries of mindset matrices to assess how adv. builds brands
④ Cont: augment dynamic adv. sales response model by integrating dyn. evol., Purchase reinf of intr. effects,
⑤ Filtering out msrmnt noise (Extract factor loading)
⑥ Brand operating hierarchy:
adv $\to$ experience $\to$ cognition $\to$ affect $\leftrightarrow$ sales
⑦ purchase reinforcement Exists for all
⑧ adv $\to$ simult affect { ① Sales growth ② Brand building
⑨ Riccati equation, reversible jump markov chain monte carlo
⑩ mk $\to$ sales agencies: mind-set metrics.
⑪ Cognition (C), affect (A), Experience (E) { samping? creative ad?
⑫ advantage of factor loading: ① Avg losses info $\Rightarrow$ inaccuracy
② Factorization reduces measurement noise (more reliable than any single matrix alone)
③ Factor loading corr b/w metric & factor
⑬ Corr b/w interm. Factors
⑭ media exposure $\to$ purchase intention

- State space model: $\boxed{\text{obs eq.}}$ Composition of 3 intermediate factors (reflected mindset matrix)

$\boxed{\text{transition eq}}$ dynamic evolution of sales & 3 intermediate factors, altntv hierarchies & purchase reinforcement effect

$\boxed{\text{drift vector}}$ how advertising triggers one or more intermediate factors & brand sales

Dynamic advertising

$$\begin{bmatrix} C_t \\ A_t \\ E_t \\ S_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_{44} \end{bmatrix} \begin{bmatrix} C_{t-1} \\ A_{t-1} \\ E_{t-1} \\ S_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \beta_4 g(u_t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \omega_{4t} \end{bmatrix}$$

$\omega_t \sim N(0, W) \to$ specific error

Classical hierarchical model $E \to C \to A$

$$\begin{bmatrix} C_t \\ A_t \\ E_t \\ S_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & \gamma_{13} & 0 \\ \gamma_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \gamma_{42} & 0 & 0 \end{bmatrix} \begin{bmatrix} C_{t-1} \\ A_{t-1} \\ E_{t-1} \\ S_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \beta_3 g(u_t) \\ 0 \end{bmatrix} + \begin{bmatrix} \omega_{1t} \\ \omega_{2t} \\ \omega_{3t} \\ \omega_{4t} \end{bmatrix}$$

advertising affects experience $\omega_t \sim N(0, W)$

$(\omega_1, \omega_2, \omega_3) \to$ specific error.

Vakratsas-Ambler model adv. ignites all 3 fact. simult then all factors affect sales, Brand purchase reinf.

Experience (adv $\to$ (E, C, A) $\to$ S ; S $\to$ E)

$$\begin{bmatrix} C_t \\ A_t \\ E_t \\ S_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_{34} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & 0 \end{bmatrix} \begin{bmatrix} C_{t-1} \\ A_{t-1} \\ E_{t-1} \\ S_{t-1} \end{bmatrix} + \begin{bmatrix} \beta_1 g(u_t) \\ \beta_2 g(u_t) \\ \beta_3 g(u_t) \\ 0 \end{bmatrix} + \begin{bmatrix} \omega_{1t} \\ \omega_{2t} \\ \omega_{3t} \\ \omega_{4t} \end{bmatrix}$$

ignore dynamics

lagged integrated $E \to C \to A$

$$\begin{bmatrix} C_t \\ A_t \\ E_t \\ S_t \end{bmatrix} = \begin{bmatrix} \gamma_{11} & 0 & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & 0 & \gamma_{24} \\ 0 & 0 & \gamma_{33} & \gamma_{34} \\ 0 & \gamma_{42} & 0 & \gamma_{44} \end{bmatrix} \begin{bmatrix} C_{t-1} \\ A_{t-1} \\ E_{t-1} \\ S_{t-1} \end{bmatrix} + \begin{bmatrix} \beta_1 g(u_t) \\ \beta_2 g(u_t) \\ \beta_3 g(u_t) \\ \beta_4 g(u_t) \end{bmatrix} + \begin{bmatrix} \omega_{1t} \\ \omega_{2t} \\ \omega_{3t} \\ \omega_{4t} \end{bmatrix}$$

diag: carry overs tangible & intangible effects of adv

Transformation

$$\underbrace{\begin{bmatrix} 1 & 0 & -\gamma_{13} & 0 \\ -\gamma_{21} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -\gamma_{42} & 0 & 1 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} C_t \\ A_t \\ E_t \\ S_t \end{bmatrix}}_{f_t} = \underbrace{\begin{bmatrix} \gamma_{11} & 0 & 0 & \gamma_{14} \\ 0 & \gamma_{22} & 0 & \gamma_{24} \\ 0 & 0 & \gamma_{33} & \gamma_{34} \\ 0 & 0 & 0 & \gamma_{44} \end{bmatrix}}_{\varphi} \underbrace{\begin{bmatrix} C_{t-1} \\ A_{t-1} \\ E_{t-1} \\ S_{t-1} \end{bmatrix}}_{f_{t-1}}$$

$$+ \underbrace{\begin{bmatrix} \beta_1 g(u_t) \\ \beta_2 g(u_t) \\ \beta_3 g(u_t) \\ \beta_4 g(u_t) \end{bmatrix}}_{\beta g(u_t)} + \underbrace{\begin{bmatrix} \omega_{1t} \\ \omega_{2t} \\ \omega_{3t} \\ \omega_{4t} \end{bmatrix}}_{\omega_t}$$

⑥ - nests

$\boxed{\text{general model}}$ $f_t = \Gamma f_{t-1} + d_t + \omega_t$ all permutation

$\boxed{\text{Intermediate Factor Composition}}$ $\{x_{it}\}_{i=1}^{n}$ mindset metrics

- measurement noise

$$\underbrace{\begin{bmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{bmatrix}}_{\text{mindset metric}} = \underbrace{\begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1J} \\ & \lambda_{ij} & \vdots \\ \lambda_{n1} & \cdots & \lambda_{nJ} \end{bmatrix}}_{\text{loading coeff}} \overset{\text{factors}}{\begin{bmatrix} f_{1t} \\ \vdots \\ f_{Jt} \end{bmatrix}} + \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}$$

need to set some to zero, 1 for identification

Simple structure

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{nt} \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \lambda_{n1} & \lambda_{n2} & \lambda_{n3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C_t \\ A_t \\ E_t \\ S_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \\ \varepsilon_{n+1,t} \end{bmatrix}$$

purchase reinforcement

$x_t = \Lambda f_t + \varepsilon_t$

Bayesian seminar Marketing papers 5 6

Estimation techniques
- (1) max likelihood
- (2) Expectation maximization algorithm
- (3) Bayesian techniques

estimation
- (1) Kalman filter smoother recursion
- (2) MCMC
- (3) stable factor correl. by solving discrete algebric Riccati eq. (non orthog. deg)
- (4) RJ-MCMC: determin # of factors to retain

$f^T = \{f_0, f_1, \dots, f_T\} \Rightarrow$ state param.: (c, A, E, S) x^T: observation

$\theta = \{\Lambda, \Gamma, \beta\}$ $\{\Lambda, \Gamma, \beta, Q, W\}$

factor loading intermediate dynamics ad effectiveness inver wishart

direct gibs sampling $\Rightarrow$ joint posterior: $p(f^T, W, Q, \theta, x^T)$

iteration resampling from $p(f^T \mid x^T, W, Q, \theta)$ & $p(W, Q, \theta \mid x^T, f^T)$ Gibbs sampling

(15) repeated measurements $\Rightarrow$ relax orthogonality assumption $\Phi_t = V(f_t)$

(16) time invariant: $\Phi_\infty = \lim_{t \to \infty} \Phi_t$

$$\Gamma \Phi_\infty \Gamma' - \Phi_\infty + \Gamma \Phi_\infty \Lambda' (\Lambda \Phi_\infty \Lambda' + Q)^{-1} \Lambda \Phi_\infty \Gamma' + W = 0$$

discrete algebric riccati equation

(17) factor retention: RJ-MCMC # param: uncertainty

(18) joint posterior: $p(\theta^{(k)} \mid x^T) \propto \prod_{t=1}^{T} p(x_t \mid f_t, \beta Q)\, p(f_0 \mid m_0, C_0)$

$$\times\, p(\Gamma)\, p(Q) \times \prod_{t=1}^{T} p(f_t \mid f_{t-1}, \Lambda, \beta, W)\, p(\Lambda)\, p(\beta)\, p(W)$$

$$q(\theta^{(k)}) = \prod_{t=1}^{T} f_N(f_{(k)t}, M_{f(k)t} \mid a V_{f(k)t}) \cdot f_N(\beta_k, M_{\beta(k)} \mid b V_{\beta(k)}) \cdot f_N(\gamma_k, M_{\gamma(k)} \mid c V_{\gamma(k)}) \cdot f_N(\lambda_k, M_{\lambda(k)} \mid d V_{\lambda(k)})\, f_{IG}(\omega_k^2, e, e M_{\omega(k)}) \cdot f_{IG}(\sigma_k^2, g, g M_{\sigma(k)})$$

normal

inverse gamma

a_1 — a_k: tunning param.

M_2: sample mean

V_2: sample variance

(19) $p(x^T, k, \theta^{(k)}) = p(x^T \mid \theta^{(k)}, k)\, p(\theta^{(k)} \mid k) \cdot pr(k)$ $pr(k) = 1/k$

MCMC initial values: $(k, \theta^{(k)})$

draw candidate model: $k' \sim pr(k' \mid k)$

Conditional on model k' draws $\theta^{(k')}$ from $q(\theta^{(k')})$

accept with prob: $\alpha = \min\left\{1, \dfrac{p(x \mid k', \theta^{k'}) \cdot p(\theta^{k'} \mid k')\, p(k')\, q(\theta^{k})}{p(x \mid k, \theta^{(k)})\, p(\theta^{(k)} \mid k)\, p(k) \cdot q(\theta^{(k')})}\right\}$

(20) test on DGP.

(21) unit root: Dickey Fuller

(22) advertising: budget approved in advance

(23) Instrument: two lagged

(24) emotional advertising prevails