WEB APPENDIX

Predicting New Customers' Risk Type in the Credit Card Market

Yi Zhao, Ying Zhao, and Inseong Song

Web Appendix A: Priors

a. Prior on δ.

$\delta \sim \text{Beta}(a_{0\delta}, b_{0\delta})$,we set $a_{0\delta} = 0.01$ and $b_{0\delta} = 0.09$

b. Prior on Ω.

$\Omega \sim IW(v_{\Omega}, V_{\Omega})$. In our computations, we set $v_{\Omega} = 3.0$, $V_{\Omega} = \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 3.0 \end{pmatrix}$. The identification issue will be discussed in Appendix B.

c. Priors on ω, Σ where $\Sigma = \text{diag}(\lambda_1, \cdots, \lambda_k)$, and $\omega = (\omega_1, \cdots, \omega_k)'$.

$\lambda_j \sim \frac{v_j V_j}{x^2_{v_j}}$ and $\omega_j \mid \lambda_j \sim N(\overline{\omega}_j, \lambda_j A)$. We set $\overline{\omega}_j = 0$, $v_j = 3$, $V_j = 1$ and $A = I$.

d. Priors on a_1 and p_1 where $a_1 = (a_{11}, \cdots, a_{1k})$, $p_1 = (p_{11}, \cdots, p_{1k})$.

$p_{1j} \sim IG(v_{0p}, V_{0p})$ and $a_{1j} \mid p_{1j} \sim N(\overline{a}_1, \frac{p_{1j}}{\tau_a})$. We set $v_{0p_1} = 3$, $V_{0p_1} = 1$, $\overline{a}_{1j} = 0$ and $\tau_a = 1$.

e. Priors on $\overline{h}$ and V_h where $\overline{h} = (\overline{h}_1, \cdots, \overline{h}_k)$, $V_h = (V_{h1}, \cdots, V_{h2})$.

$V_{hj} \sim IG(v_{0h}, V_{0h})$ and $\overline{h}_j \mid V_{hj} \sim N(\overline{\overline{h}}, \frac{V_{hj}}{\tau_h})$. We set $v_{0h} = 3$, $V_{0h} = 1$, $\overline{\overline{h}} = 0$ and $\tau_h = 1$.

e. Priors on $\overline{\theta}$ and V_{θ} where $\overline{\theta} = (\overline{\theta}_1, \cdots, \overline{\theta}_k)$, $V_{\theta} = (V_{\theta 1}, \cdots, V_{\theta k})$

$V_{\theta j} \sim IG(v_{0\theta}, V_{0\theta})$ and $\overline{\theta} \mid V_{\theta j} \sim N(\overline{\overline{\theta}}, \frac{V_{\theta j}}{\tau_{\theta}})$. We set $v_{0\theta} = 3$, $V_{0\theta} = 1$, $\overline{\overline{\theta}} = 0$ and $\tau_{\theta} = 1$.

We tried different prior settings, and found little substantive differences among them.

Web Appendix B: Full Conditionals and Simulation Algorithm

For facilitating to apply MCMC Algorithm, we firstly define 'consumer oversight' variable D_{it} for data augmentation as follows:

If $y^*_{2it} > 0$ and $y_{2it} = 0$, then D_{it}=1; otherwise, D_{it}=0.

1. $D_{it} \mid \delta, y^*_{1it}, y_{2it}, \gamma_{it}, x_{1it}, x_{2it}, \Omega$

   If $y_{2it} > 0$ then $P(D_{it} = 0 \mid \delta, y^*_{1it}, y_{2it}, \gamma_{it}, x_{1it}, x_{2it}, \Omega) = 1$

   If $y_{2it} = 0$ then

$$P(D_{it} = 0 \mid \delta, y^*_{1it}, y_{2it}, \gamma_{it}, x_{1it}, x_{2it}\Omega)$$
$$= \frac{P(y^*_{2it} \le 0 \mid y^*_{1it}, \gamma_{it}, x_{1it}, x_{2it}, \Omega)}{\delta \bullet P(y^*_{2it} > 0 \mid y^*_{1it}, \gamma_{it}, x_{1it}, x_{2it}, \Omega) + P(y^*_{2it} \le 0 \mid y^*_{1it}, \gamma_{it}, x_{1it}, x_{2it}, \Omega)}$$

   where $P(y_{2it} \le 0 \mid y^*_{1it}, \gamma_{it}, x_{1it}, x_{2it}, \Omega) = \Phi(\frac{M_{2|1}}{V_{2|1}})$,

$$M_{2|1} = x_{2it}\gamma_{2it} + \frac{\Omega_{12}}{\Omega_{11}}(y^*_{1it} - x_{1it}\gamma_{1it}) \text{ and } V_{2|1} = \Omega_{22} - \frac{\Omega^2_{12}}{\Omega_{11}}$$

2. δ| D_{it}

   As mentioned in paper, we consider consumer negligence only when $y^*_{2it} > 0$. We firstly define $\Theta = \{i, t;\, y^*_{2it} > 0\}$, then

   $\delta | D_{it} \sim \text{Beta}(a_{0\delta} + a_\delta, b_{0\delta} + b_\delta)$, where $a_\delta = \sum_{(i,t)\in\Theta} D_{it}$ and $b_\delta = \sum_{(i,t)\in\Theta}(1 - D_{it})$.

3. $y^*_{2it} \,|\, y_{2it}, y^*_{1it}, \Omega, \gamma_{it}, x_{1it}, x_{2it}$

   If $y_{2it} > 0$, then $y^*_{2it} = y_{2it}$

   If $y_{2it} = 0$, and D_{it}=0 then draw y^*_{2it} truncated by $y^*_{2it} \le 0$ from $N(M_{2|1}, V_{2|1})$

   If $y_{2it} = 0$, and D_{it}=1 then draw y^*_{2it} truncated by $y^*_{2it} > 0$ from $N(M_{2|1}, V_{2|1})$

   where $M_{2|1} = x_{2it}\gamma_{2it} + \frac{\Omega_{12}}{\Omega_{11}}(y^*_{1it} - x_{1it}\gamma_{1it})$ and $V_{2|1} = \Omega_{22} - \frac{\Omega^2_{12}}{\Omega_{11}}$

4. $y^*_{1it} \,|\, y_{1it}, y^*_{2it}, \Omega, \gamma_{it}, x_{1it}, x_{2it}$

   If $y_{1it} = 1$ then draw y^*_{1it} truncated by $y^*_{1it} > 0$ from $N(M_{1|2}, V_{1|2})$

If $y_{1it} = 0$ then draw y^*_{1it} truncated by $y^*_{1it} \le 0$ from $N(M_{1|2}, V_{1|2})$

where $M_{1|2} = x_{1it}\gamma_{1it} + \frac{\Omega_{12}}{\Omega_{22}}(y^*_{2it} - x_{2it}\gamma_{2it})$ and $V_{1|2} = \Omega_{11} - \frac{\Omega^2_{12}}{\Omega_{22}}$

5. $\gamma_{it} \mid y^*_{1it}, y^*_{2it}, \Omega, \Sigma, \theta_i, \omega, z_{it}, h_i, a_1, p_1$

   The conditional distribution for the state γ_{it} $(i = 1, \cdots, T)$ is also multivariate normal. We draw states recursively as explained in Appendix C.

6. $a_1, p_1 | \gamma_{i1}$

   The full conditional distributions for the mean, a_1, and the variance, p_1, of initial states are normal and inverted-gamma distributions, respectively. The conditional distribution of the jth element is given as follows.

$$p_{1j} \sim IG(v_{0p_1} + \frac{n}{2}, V_{0p_1} + \frac{1}{2}\sum_{i=1}^{n}(\gamma_{ij1} - \bar{\gamma}_{j1})^2 + \frac{\tau_a n(\bar{\gamma}_{j1} - \bar{a}_{1j})^2}{2(\tau_a + n)})$$

$$a_{1j} \mid p_{1j} \sim N(\frac{\bar{a}_{1j}\tau_a + n\bar{\gamma}_{j1}}{\tau_a + n}, \frac{p_{1j}}{\tau_a + n}),$$

where $\bar{\gamma}_{j1} = \frac{1}{n}\sum_{i=1}^{n}\gamma_{ij1}$.

7. $\omega, \Sigma | \gamma_{it}, h_i, \theta_i, z_{it}$

   Given γ_{it}, h_i, z_{it}, the conditional distribution of each element in ω is normal. And the conditional distribution of λ_j is inverted Chi-square. Let us define

$$\delta_j \equiv \begin{pmatrix} \hat{\gamma}_{1j} \\ \vdots \\ \hat{\gamma}_{nj} \end{pmatrix}, \quad D_j = \begin{pmatrix} \tilde{z}_1 \\ \vdots \\ \tilde{z}_n \end{pmatrix}, \ \hat{n} \equiv \sum_{i=1}^{n} T_i - n,$$

where $\hat{\gamma}_{ij} = \begin{pmatrix} \gamma_{ij2} - \theta_{ij}\gamma_{ij1} - h_{ij} \\ \vdots \\ \gamma_{ij,T_i} - \theta_{ij}\gamma_{ij,T_i-1} - h_{ij} \end{pmatrix}$, $\tilde{z}_i = \begin{pmatrix} z_{i2} \\ \vdots \\ z_{iT_i} \end{pmatrix}$, and T_i is the number of observations for consumer i and n is the total number of consumers in the data set. Based on the standard results for a multiple regression model, we have

$$\lambda_j \sim \frac{vV}{x_v^2} \quad \text{with } v = v_j + \hat{n}, V = \frac{v_j V_j + \hat{n}s^2}{v}$$

$$\omega_j \big| \lambda_j \sim N\left(\tilde{\omega}_j, \lambda_j \left(D_j' D_j + A\right)^{-1}\right)$$

where $\tilde{\omega}_j = (D_j' D_j + A)^{-1}(D_j' \delta_j + A\bar{\omega}_j)$ and

$\hat{n}s^2 = (\delta_j - D_j\tilde{\omega}_j)'(\delta_j - D_j\tilde{\omega}_j) + (\tilde{\omega}_j - \bar{\omega}_j)' A(\tilde{\omega}_j - \bar{\omega}_j)$

8. $h_i, \theta_i \mid \bar{h}, V_h, \bar{\theta}, V_\theta, \gamma_{it}, \omega, z_i, \Sigma$

The conditional distribution of h_i and θ_i is normal. For each element j, define

$\underset{\sim}{x}_{ijt} = (1/\sqrt{\lambda_j} \quad \gamma_{ij,t-1}/\sqrt{\lambda_j})$, and $\underset{\sim}{y}_{ijt} = (\gamma_{ijt} - \omega_j z_{it})/\sqrt{\lambda_j}$, for $t = 2, \cdots, T_i$. And also

define $\beta_{ij} = (h_{ij}, \theta_{ij})'$ and $V_{\beta j} = \begin{pmatrix} V_{hj} & 0 \\ 0 & V_{\theta j} \end{pmatrix}$.

Based on standard results for linear regression model, we have

$\beta_{ij} \sim N(\bar{b}_j, (\underset{\sim}{x}_{ij}' \underset{\sim}{x}_{ij} + V_{\beta j}^{-1})^{-1})$ where $\underset{\sim}{x}_{ij}$ is a $(T_i-1)\times 2$ matrix of the stacked $\underset{\sim}{x}_{ijt}$,

$\bar{b}_j = (\underset{\sim}{x}_{ij}' \underset{\sim}{x}_{ij} + V_{\beta j}^{-1})^{-1}(\underset{\sim}{x}_{ij}' \underset{\sim}{y}_{ij} + V_{\beta j}^{-1}\bar{\beta}_j)$ and $\bar{\beta}_j = (\bar{h}_j \quad \bar{\theta}_j)'$.

9. $\bar{h}, V_h \big| h_i$

The conditional distribution of $\bar{h}$ is normal and that of V_{hj} is inverted gamma. For each element, j,

$$V_{hj} \sim IG\left(v_{0h} + \frac{n}{2}, V_{0h} + \frac{1}{2}\sum_{i=1}^{n}(h_{ij} - m_{h_j})^2 + \frac{\tau_h n(m_{h_j} - \bar{\bar{h}}_j)^2}{2(\tau_h + n)}\right)$$

$$\bar{h}_j \mid V_{hj} \sim N\left(\frac{\bar{\bar{h}}_j \tau_h + m_{h_j} n}{\tau_h + n}, \frac{V_{hj}}{\tau_h + n}\right)$$

where $m_{h_j} = \frac{1}{n}\sum_{i=1}^{n} h_{ij}$.

10. $\bar{\theta}, V_\theta \mid \theta_i$

The conditional distribution of $\bar{\theta}$ is normal and that of $V_{\theta j}$ is inverted gamma. For each element, j,

$$V_{\theta j} \sim IG(v_{0\theta} + \frac{n}{2}, V_{0\theta} + \frac{1}{2}\sum_{i=1}^{n}(\theta_{ij} - m_{\theta_j})^2 + \frac{\tau_\theta n(m_{\theta_j} - \overline{\overline{\theta}}_j)^2}{2(\tau_\theta + n)})$$

$$\overline{\theta}_j \mid V_{\theta j} \sim N(\frac{\overline{\overline{\theta}}_j \tau_\theta + m_{\theta_j} n}{\tau_\theta + n}, \frac{V_{\theta j}}{\tau_\theta + n})$$

where $m_{\theta_j} = \frac{1}{n}\sum_{i=1}^{n}\theta_{ij}$.

11. $\Omega \big| y^*_{1it}, y^*_{2it}, x_{1it}, x_{2it}, \gamma_{1it}, \gamma_{2it}$

First, we ignore the identification issue. The conditional distribution of Ω is inverted Wishart,

$$\Omega^{-1} \big| y^*_{1it}, y^*_{2it}, x_{1it}, x_{2it}, \gamma_{1it}, \gamma_{2it} \sim W(N + v_\Omega, (V_\Omega + S)^{-1}),$$

where $N = \sum_{i=1}^{n} T_i$, $S = \sum_{i,t} e_{it} e_{it}'$, $e_{it} = \begin{pmatrix} y^*_{1it} - x_{1it}\gamma_{1it} \\ y^*_{2it} - x_{2it}\gamma_{2it} \end{pmatrix}$.

We draw from the inverted Wishart distribution without imposing any restriction on it. Next, we solve the identification problem by adopting the method suggested by Edwards and Allenby (2003). Basically, we transform the draws from the unrestricted distributions into an identified parameter space. Consider a draw of the unrestricted distribution, Ω:

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

Define Δ and Λ as follows:

$$\Delta = \text{diag}\left(1/\sqrt{\Omega_{11}}, 1 \right), \text{ and } \Lambda = \text{diag}\left(1/\sqrt{\Omega_{11}}, ..., 1/\sqrt{\Omega_{11}}, 1, ..., 1 \right),$$

where Λ is a diagonal matrix of dimension k_1+k_2 with the first k_1 elements being all $1/\sqrt{\Omega_{11}}$ and the next k_2 elements being all 1. i.e.,

$$\Lambda = \begin{pmatrix} 1/\sqrt{\Omega_{11}} & & & & & \\ & \ddots & & & & \\ & & 1/\sqrt{\Omega_{11}} & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}.$$

In each step of the iteration, we transform the unrestricted draws as follows:

$$\tilde{\gamma}_{it} = \Lambda\gamma_{it}$$

$$\tilde{\Omega} = \Delta\Omega\Delta$$

$$\tilde{\omega} = \Lambda\omega$$

$$\tilde{h}_i = \Lambda h_i$$

$$\tilde{\bar{h}} = \Lambda\bar{h}$$

$$\tilde{V}_h = \Lambda V_h \Lambda$$

$$\tilde{\Sigma} = \Lambda\Sigma\Lambda$$

We use the transformed draws for inferences, i.e., $\tilde{\gamma}_{it}$, $\tilde{\Omega}$, $\tilde{\omega}$, $\tilde{h}_i$, $\tilde{\bar{h}}$, $\tilde{\sigma}_h^2$, $\tilde{\Sigma}$, and θ.

Web Appendix C: Simulating States

In simplifying our presentation of the simulation procedure, we define some additional parameters and variables. In this section, all expressions are for one consumer, so we drop the subscript i for expositional simplicity.

Define $y_t \equiv \begin{pmatrix} y_{1t}^* \\ y_{2t}^* \end{pmatrix}$, $x_t \equiv \begin{pmatrix} x_{1t} & 0 \\ 0 & x_{2t} \end{pmatrix}$, $\gamma_t \equiv \begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \end{pmatrix}$, $\varepsilon_t \equiv \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$, and $m_t \equiv \omega z_t + h_i$. Our model can be rewritten as:

$$y_t = x_t \gamma_t + \varepsilon_t \qquad \gamma_t = \theta \gamma_{t-1} + m_t + \xi_t,$$

where $\varepsilon_t \sim \text{iid}\, N(0, \Sigma_\varepsilon)$, $\xi_t \sim \text{iid}\, N(0, \Sigma_\xi)$, t=1,…T.

Now, our model is a standard state space model. There are many studies on the procedure of drawing from the conditional distribution of a state vector given observations $y_t, t = 1, \cdots, T$. We use the method by Durbin and Koopman (2001b) as it is simple and computationally efficient. The basic idea is as follows. Define $\hat{a} \equiv E(\gamma_1, \cdots, \gamma_T | y_1, \cdots y_T)$ and $W \equiv \text{var}(\gamma_1, \cdots, \gamma_T | y_1, \cdots, y_T)$. In a multivariate normal distribution, the conditional variance-covariance matrix of a vector conditional on the second vector does not depend on the value of the second vector. It depends only on the variance-covariance matrix of the second vector. That is, W does not depend upon the value of y. That is, the conditional variance covariance matrix of the state W is the same as long as the sets of observations come from the same distribution. If we can calculate the conditional mean of the states, the process of drawing $\gamma_1, \cdots, \gamma_T$ is straightforward as their variance does not depend upon the value of y.

1. Draw a random vector (ε, ζ) from the normal distribution given above and use it to generate γ^+ and y^+ iteratively conditional on x and m using the initial state draw from $N(a_1, p_1)$.
2. Compute $\hat{a} = E(\gamma | y)$, $\hat{a}^+ = E(\gamma^+ | y^+)$ using the filtering and smoothing algorithm discussed below.
3. The random draw of the state is given by $\tilde{\gamma} = \gamma^+ - \hat{a}^+ + \hat{a}$.

Now we discuss how to calculate the conditional mean of the state, i.e., $\hat{a} = E(\gamma|y)$ and $\hat{a}^+ = E(\gamma^+|y^+)$. According to Durbin and Koopman (2001a), we can derive this conditional mean as follows based on the usual Kalmam filtering algorithm.

Define $a_t \equiv E(\gamma_t|y_1,\cdots y_{t-1})$ and $p_t \equiv var(\gamma_t|y_1,\cdots,y_t)$. By applying Kalman filtering, we get

$$v_t = y_t - x_t a_t \ , F_t = x_t P_t x_t' + \Sigma_\varepsilon, \ K_t = \theta P_t x_t' F_t^{-1} \ , L_t = \theta - K_t x_t$$

$$a_{t+1} = \theta a_t + K_t v_t + m_{t+1}, \ P_{t+1} = \theta P_t L_t' + \Sigma_\xi .$$

We also need to prepare another variable by backwards recursion. Compute the values c_t recursively as follows:

$$c_T = 0, \ c_{t-1} = x_t' F_t^{-1} v_t + L_t' c_t, \ \text{t=T-1,...,1.}$$

Finally, we can get the conditional mean of the states:

$$\hat{a}_1 = a_1 + P_1 c_0, \ \hat{a}_{t+1} = \theta \hat{a}_t + \Sigma_\xi c_t$$

The proofs for the above procedure are provided in Anderson and Moore (1979) and Durbin and Koopman (2001a).

References are available from the authors on request.