A debt behaviour model

Wenjun Zhang, John Holt

Figure 1: This diagram depicts the underlying causal structure of the model. See the text for the definitions of D,Y,B,T,S.

The model concerns the following random variables:

- A discrete Markov process \( B_t \) which records the *behavioural state* of the debtor during the time period \( t \) - measured in months. The state is measured in the middle of each month.

- A discrete-valued process \( T_t \) which records the *strongest debt management intervention* that was applied to the debtor during the time period \( t \).
• \( R \) an entity-specific variable, \( R \) gives the final result of the debtor’s most immediate previous debt case - NA, paid in full, liquidation/bankruptcy, full write-off, partial write-off.

• \( X_t \) is the economic state at time period \( t \). This measure is obtained through clustering a pertinent collection of economic variables: change in CPI, change in unemployment, change in the average weekly wage, etc. The underlying variables for \( X_t \) are varying quarterly, so \( X_t \) will be constant in blocks of three months.

• \( S_t \) is a latent discrete Markov process which categorizes debtors in a time period into the \textit{behavioural scheme} that governs the generation of \( B_t \). The model supposes that \( T_{t-1} \) influences \( S_t \), and hence influences \( B_t \) indirectly.

• \( D_t \) is a positive real-valued variable, given by

\[
D_t = \frac{\text{Debt amount at time } t, \text{ including penalties and interest}}{\text{Largest amount of debt owed up to time } t, \text{ excluding penalties and interest}}
\]

• \( Y_t \) is a categorization of \( D_t \) into \( \{0, 1\} \) - this is governed by a parameter \( \alpha \) that needs to be inferred. the notion is that as a debtor gets closer to being paid in full, its probability of making a large lump-sum payment to clear its debt may change.

We introduce a set of parameters as follows:

• \( \alpha \): defined by \( Y_t := 0 \) if and only if \( D_t \leq \alpha \).

• \( Q_S \): a list of transition matrices, one for each combination of values of \( R, X_t, T_{t-1} \).

• \( \pi_S \): a list of initial probabilities, one for each combination of values of \( R, X_t \).

• \( Q_B \): a list of transition matrices, one for each combination of values of \( Y_{t-1} \) and \( S_t \).

• \( \pi_B \): a list of initial probabilities, one for each value of \( S_1 \).
Figure 2: This diagram depicts the underlying causal structure of the model, including the parameters. Refer to the text for definitions of the parameters $\pi_B, Q_B, \pi_S, Q_S, \alpha$.

Figure 2 depicts the causal structure of the variables and the parameters - we have now expressed each of the variables as a vector of length as long as the number of observation periods.

Every debt case begins at a time period $u$ and ends at a time period $l$. If the debt case is indexed by $i$, the the beginning is $u_i$ and the end is $l_i$. There will be observations of $T_t, B_t, D_t, X_t$ from $u_i$ through to $l_i$.

The log-likelihood of observing a single debt case is maximized when we maximize:

$$l_0 = \sum_{t=u+1}^{l_i} \left( \ln(Q_B^{Y_i-1:S_i}(B_{t-1}, B_t)) + \ln(Q_S^{X_t,R:T_t-1}(S_{t-1}, S_t)) + \ln(\pi_B(B_u)) + \ln(\pi_S^{X_u,R}(S_u)) \right)$$

We apply the EM algorithm to $l_0$, taking the expected value of $l_0$ conditional on $\{B_t, X_t, D_t, T_t, R\}$ and the $k$-th iteration of the parameters $\{\alpha, Q_B, Q_S, \pi_B, \pi_S\}$.
For this we define the responsibilities for each debt case, \( i \), and time \( t \), \( t = u_i, \ldots, l_i \):

\[
\gamma_{i,t}(s) := p(S_t = s | T_{u_i}^{l_i-1}, X_{u_i}^{l_i}, B_{u_i}^{l_i}, R_i, D_{u_i}^{l_i-1})
\]

for \( t \geq u_i \); and for \( t > u_i \),

\[
\Gamma_{i,t}(p,q) := p(S_t = q, S_{t-1} = p | T_{u_i}^{l_i-1}, B_{u_i}^{l_i}, R_i, D_{u_i}^{l_i-1})
\]

It is clear that \( \gamma_{i,t}(s) = \sum_p \Gamma_{i,t}(p,s) \), or if \( t = u_i \), \( \gamma_{i,u_i}(s) = \sum_q \Gamma_{i,u_i+1}(s,q) \) - hence we need only compute \( \Gamma_{i,t} \).

This is done using the Forward-Backward algorithm:

\section{Calculating \( \Gamma_{i,t} \)}

This calculation is standard, but we present it for completeness.

Define the following four sets of probabilities:

\begin{itemize}
  \item \( \pi_t(s) = p(S_t = s | T_u^t, X_u^t, R, D_u^{t-1}, B_u^t) \)
  \item \( \pi'_t(s) = p(S_t = s | T_u^t, X_u^t, R, D_u^{t-1}, B_u^t), t \geq u \).
  \item \( F_t(p,q) = p(S_{t-1} = p, S_t = q | T_u^{t-1}, X_u^{t-1}, R, D_u^{t-1}, B_u^t), t > u \)
  \item \( \Gamma_t(p,q) = p(S_{t-1} = p, S_t = q | T_u^{t-1}, X_u^{t-1}, R, D_u^{t-1}, B_u^t), t > u. \)
\end{itemize}

Then

\[
F_t(p,q) \propto Q_B^{q,Y_{t-1}}(B_{t-1}, B_t)Q_S^{T_{t-1},X_t,R}(p,q)\pi'_t\]

\[
= (Q_B^{q,0}(B_{t-1}, B_t)I_{[0,\alpha]}(D_{t-1}) + Q_B^{q,1}(B_{t-1}, B_t)I_{(\alpha,\infty)}(D_{t-1}))Q_S^{T_{t-1},X_t,R}(p,q)
\]

and

\[
\pi'_t(q) = \sum_p F_t(p,q)
\]

with \( \pi'_u(s) \propto \pi_B^*(B_u)\pi_S^{X_u,R}(s) \). The normalizing constants can be found by noting that \( \sum_{p,q} F_t(p,q) = 1 \) and \( \sum_s \pi'_u(s) = 1. \)

Having obtained \( F_t(p,q) \) (the forward matrices) we can calculate the backward matrices \( \Gamma_t \) as follows: 4
Set $\Gamma_t = F_t$.

For $t < l$, \[ \Gamma_t(p, q) = p(S_t = q | S_{t-1} = p, T_{u_t}^{t-1}, X_{u_t}^{t-1}, R, D_{u_t}^{t-1}, B_{u_t}^t) p(S_t = q | T_{u_t}^{t-1}, X_{u_t}^{t-1}, R, D_{u_t}^{t-1}, B_{u_t}^t) \]

\[ = F_t(p, q) \frac{\pi_t(q)}{\pi_t'(q)} \]

2 Update equations for the M-step

The formulas that follow are the result of straightforward calculations.

\[ Q^{s, y}_{B}(b, c) = \frac{\sum_i \sum_{t=u_i+1}^{l_i} \delta(B_{i,t} - c) \delta(B_{i,t-1} - b) \delta(Y_{i,t-1} - y) \gamma_{i,t}(s)}{\sum_i \sum_{t=u_i+1}^{l_i} \delta(B_{i,t-1} - b) \delta(Y_{i,t-1} - y) \gamma_{i,t}(s)} \]

\[ \pi^{y}_{B}(b) = \frac{\sum_i \delta(B_{i,u_i} - b) \gamma_{i,u_i}(s)}{\sum_i \gamma_{i,u_i}(s)} \]

\[ Q^{T, R, X}_{S}(p, q) = \frac{\sum_i \sum_{t=u_i}^{l_i-1} \delta(T_{i,t} - T) \delta(R_i - R) \delta(X_t - X) \gamma_{i,t}(p) \gamma_{i,t+1}(q)}{\sum_i \sum_{t=u_i}^{l_i-1} \delta(T_{i,t} - T) \delta(X_t - X) \delta(R_i - R) \gamma_{i,t}(p)} \]

\[ \pi^{R, X}_{S}(s) = \frac{\sum_i \delta(R_i - R) \delta(X_{u_i} - X) \gamma_{i,u_i}(s)}{\sum_i \delta(R_i - R) \delta(X_{u_i} - X)} \]

Note that $Q_B$ depends on an unknown value of $\alpha$. The approach will be to fit $Q_B$ for a range of values of $\alpha$, and to choose the $\alpha$ that gives the maximum value to:

\[ l_1 = \sum_i \sum_{t=u_i+1}^{l_i} \sum_s \ln(Q^{s,0}_{B}(B_{i,t-1}, B_{i,t}) I_{[0,\alpha]}(D_{i,t-1}) + Q^{s,1}_{B}(B_{i,t-1}, B_{i,t}) I_{(\alpha,\infty)}(D_{i,t-1})) \gamma_{i,t}(s) \]