Chapter 6: Behavioural models
Overview

So far we have focussed mainly on application scorecards. In this chapter we shall look at behavioural models. We shall cover the following topics:

- Behavioural models and data;
- Static behavioural models;
- Dynamic models of behaviour:
  - Survival models;
  - Markov transition models.
Why Behavioural Models?

Application scorecards only deal with application processing, deciding whether to accept or reject an application for credit. However, once a borrower is given credit, their behaviour needs to be monitored in terms of credit usage and repayments. There are several reasons for this:

- **Risk assessment.** It allows the lender to update their assessment of risk and warn of possible problems in the future.

- **Operational decisions.** The lender can choose to adjust lending parameters such as credit limit or interest rate, based on borrower behaviour.
• **Marketing and customer relations.** Offer new products to existing “good” customers, and respond to customer requests for new products or changes in credit limit and interest rate.

• **Risk management.** Estimation of default rate on loan portfolio and calculation of capital requirements.
Behavioural Models of Default or Delinquency

Rather like application scorecards, behavioural models are usually models of default, or possibly, delinquency.

- Values of predictor variables are taken across a performance period of the loan up to a pre-specified observation point.

- The outcome variable (usually, default / non-default) is taken at an outcome point some time after the observation point.

![Diagram showing the timeline of borrower behaviour, observation period, observation point, and outcome point](attachment:image.png)
What does the dynamic behaviour of credit accounts look like?

*Example 6.1*

Here is a case study of a credit card account. Notice how erratic credit card usage is.
Example 6.2

Here is another case study of a credit card account. Notice this time, card usage has some structure, but eventually the credit card holder defaults at 32 months.

![Graph showing credit card account activity over time.](image)

Values are fictional but based on a true account.
Traditionally, behavioural models have been built using the same kind of static models as are used in application scoring.

- For example, typically, logistic regression is used.

- The outcome of such a model is a behavioural score.

- Since values of predictor variables change over the performance period, aggregate values are used:
  - eg mean, maximum or last values of variables over time.
Typical Behavioural Variables

Some typical aggregate variables used in static behavioural models are:-

- The application variables;
- Generic credit score;
- Last current balance (ie at the end of the performance period);
- Mean balance;
- Last balance / starting balance ratio;
- Sum of credit advances;
- Mean monthly repayment amount;
- Total number of missed payments;
- Total number of months that credit limit is exceeded (for credit card).

The number of potential predictor variables can get quite large for behavioural models, so automated variable selection may be required to reduce their number.
Dynamic Models

In many ways, using static models based on aggregates on variables that naturally occur over time is missing a great opportunity, since these models will not fully represent the dynamic quality of the data.

There are several statistical models that allow us to naturally model behaviour over time.

Two modelling approaches which are finding favour in the industry are:

- Survival model;
- Markov transition model.

We will cover each of these models in this chapter.

Not only do dynamic models have the advantage that they can take account of changes in credit use over time, they can be used as the basis of profit estimation. This is covered in the next chapter.
Interest in using *survival analysis* for credit scoring is quite recent.

- Survival models allow us to model not just *if* a borrower will default, but *when*.

The advantages of using survival models are that:

1. They naturally model the loan default process and so incorporate situations when a case has not defaulted in the observation period;

2. Their use avoids the need to define a fixed period within which default is measured;

3. They provide a clear approach to assess the expected profitability of a borrower;
4. Survival estimates provide a forecast as a function of time from a single equation;

5. They allow the inclusion of behavioural and economic risk factors over time.

Survival analysis has been applied in:

- Behavioural scoring for consumer credit,
- Predicting default on personal loans, and
- The development of generic score cards for retail cards.
What is survival analysis?

- Survival analysis is used to study the time to **failure** of some population.
  - This is called the **survival time**.

- Survival analysis is able to facilitate the inclusion of observations that have not failed.
  - These are treated as **censored** data.
  - An observation time is given for a censored case indicating the last time it was observed.

- In the context of consumer credit, the population comprises individuals with credit in the form of loans or credit cards.
  - When a consumer **defaults** on a loan or credit card payment then this is a **failure** event.
  - Survival time is typically measured from the date the account was opened.
  - If a consumer never defaults during the observation period then they are **censored** at the observation point.
This graph demonstrates observations of four loan accounts with an observation period.

Each account exhibits different censoring and failure characteristics.
A common means to analyze survival data is through the **hazard function** which gives the instantaneous chance of failure at time $t$:

$$h(t) = \lim_{\delta \to 0} \left( \frac{P(t \leq T < t + \delta \mid T \geq t)}{\delta} \right)$$

where $T$ is a random variable associated with survival time.

In consumer credit, several studies demonstrate the classic shape for default hazard as:

- Highest risk of default is within the first few months,
- then the risk tails off over the lifetime of the loan or credit card.
Example 6.3

Hazard rates for Default on a Store Card.

95% confidence intervals on the estimate are also shown (Andreeva, Ansell, Crook 2007).
The survival probability at time $t$ can be given in terms of the hazard function:

$$S(t) \triangleq P(T \geq t)$$

This is the probability of survival from time 0 to some time $t$.

For credit data, this gives the probability of default (PD) as:

$$P_D = 1 - S(t)$$
The survival probability is related to the hazard function, since

\[ h(t) = \lim_{\delta \to 0} \left( \frac{P(t \leq T < t + \delta | T \geq t)}{\delta} \right) = \lim_{\delta \to 0} \left( \frac{P(t \leq T < t + \delta)}{\delta} \right) / P(T \geq t) = \frac{f(t)}{S(t)} \]

where \( f \) is the probability density function of \( t \).

Since \( S(t) = 1 - F(t) \), where \( F \) is the cumulative distribution function on \( t \),

\[ f(t) = -\frac{dS(t)}{dt} \]

Therefore, integrating over \( t \), substituting \( s = S(t) \),

\[ \int_0^t h(u)du = \int_{S(0)}^{S(t)} \frac{1}{s} ds = [-\log s]_{S(0)}^{S(t)} = -\log S(t) \]

since \( S(0) = 1 \). Therefore,

\[ S(t) = \exp \left( -\int_0^t h(u)du \right) \]
Cox Proportional Hazards Model

There are several alternative survival models to estimate the hazard function.

We will look at perhaps the most popular in the credit scoring literature:

The Cox Proportional Hazards (PH) model.

- Named after Sir David Cox (Professor of Statistics at Imperial College London from 1966 to 1988).

The Cox PH model allows us to model survival in terms of the borrower characteristics. In particular, the hazard function changes with the values of predictor variables.
Suppose we have a vector of predictor variables \( x(t) \).

Then the Cox PH model is a semi-parametric model which estimates the hazard function exponentially on a linear combination of the predictor variables:

\[
h(t, x(t), \beta) = h_0(t) \exp(\beta \cdot x(t))
\]

- The vector of coefficients \( \beta \) needs to be estimated.

- The function \( h_0(t) \) is a non-parametric baseline hazard rate which is true for all observations.
  - It is similar to an intercept in a regression model, except that it changes over time.

- The model is called semi-parametric because it is composed of a non-parametric part (the baseline hazard) and a parametric part.
• Notice that the predictor variables are indexed by time. This means they can change over time.

  o They are called **time varying covariates** (TVCs).
  o It is the availability of TVCs that enable us to include dynamic behavioural data.

• This model is a **proportional hazards** (PH) model since the hazard ratio between two observations is constant over time when TVCs are not included:

\[
\text{Hazard ratio } h_r(t, x, x', \beta) = \frac{h(t,x,\beta)}{h(t,x',\beta)} = \exp \left( \beta \cdot (x - x') \right)
\]

  o However, this principle is no longer true when TVCs are included and then “PH” is a misnomer!
Partial likelihood function

The Cox PH model is estimated using maximum likelihood estimation (MLE) based on a training data set.

Suppose we have \( n \) observations for \( i=1 \) to \( n \):
- observation times \( t_i \)
- indicator variables \( c_i \) where
  - \( c_i=0 \) for a censored observation and
  - \( c_i=1 \) for a failure event (default);
    - if \( c_i=1 \) then \( t_i \) is the survival time,
- predictor variable values \( x_i(t) \).

The baseline hazard complicates the likelihood function.

Therefore the likelihood function is decomposed into two components:

1. The probability that a failure event occurs at a time \( t \);
2. The probability that it was a specific observation \( i \) that failed at time \( t \), given that a failure occurred.
It turns out that using just the second component is sufficient to get estimates of $\beta$.

This is called \textit{partial likelihood estimation}.

The practical effect is that partial likelihood estimates have higher standard errors than using MLE.

The probability that an observation fails at some time $t$, amongst all other observations is therefore given by

$$\frac{h(t, x_i(t), \beta)}{\sum_{j \in R(t)} h(t, x_j(t), \beta)} = \frac{\exp(\beta \cdot x_i(t))}{\sum_{j \in R(t)} \exp(\beta \cdot x_j(t))}$$

where $R(t)$ is called the \textit{risk set} and includes all observations that are uncensored and have not failed by time $t$.

Specifically, $R(t) = \{j : t_{(j)} \geq t\}$ where $t_{(j)}$ are ordered survival times.
Partial likelihood function

This gives the partial likelihood function for the Cox PH model:

\[
l_p(\beta) = \prod_{i=1}^{n} \left( \frac{\exp(\beta \cdot x_i(t_i))}{\sum_{j \in R(t_i)} \exp(\beta \cdot x_j(t_i))} \right)^{c_i}
\]

Maximizing this with respect to \( \beta \) gives an estimate of \( \beta \).

Typically the Cox PH model is used as an explanatory model, in which case an estimate \( \hat{\beta} \) of \( \beta \) is sufficient.
Forecasting survival probability

For retail finance, we are primarily interested in forecasting the survival function for an individual, $\hat{S}_i(t)$, since this is related to the PD, $1 - \hat{S}_i(t)$.

For forecasting, the baseline hazard will also need to be estimated. A nonparametric MLE is used to do that, based on the initial estimate of $\beta$.

In the survival setting, for forecasting, an estimate of the survival curve is required for each observation.

Since, in general, for the Cox PH model,

$$S(t) = \exp \left( - \int_0^t h(u) \, du \right) = \exp \left( - \int_0^t h_0(u) \exp(\beta \cdot x(u)) \, du \right),$$

it follows for each observation an estimate is given by

$$\hat{S}_i(t) = \exp \left( - \int_0^t \hat{h}_0(u) \exp(\hat{\beta} \cdot x_i(u)) \, du \right).$$

Notice that an estimate $\hat{h}_0$ of $h_0$ is needed.
Estimating the baseline hazard

There are different ways to estimate $\hat{h}_0$, but one approach that has been suggested is to estimate the **cumulative baseline hazard**

$$\hat{H}_0(t) = \int_0^t \hat{h}_0(u) \, du \approx \sum_{t(i) \leq t} \frac{c_i}{\sum_{j \in R(t(i))} \exp \left( \beta \cdot x_j(t(i)) \right)}$$

Then

$$\hat{h}_0(u) \approx \frac{\hat{H}_0(u + \Delta u) - \hat{H}_0(u)}{\Delta u}$$

for some appropriately small $\Delta u$.

Since the formula for $\hat{s}_i(t)$ includes an integral, in practice, the estimation of survival probability requires a numerical integration method, if TVCs are included.
However, if no TVCs are included in the model, so $x_i(t) = x_i(0)$ for all $t > 0,$

$$\hat{S}_i(t) = \exp \left(-\hat{H}_0(t)\exp \left(\hat{\beta} \cdot x_i(0)\right)\right)$$

which does not require numerical integration.

Indeed, many statistical packages, such as R and SAS, have standard functions to estimate the baseline hazard and survival probability when no TVCs are in the model, but they do not work if TVCs are included.
It is straightforward to include behavioural variables directly as TVCs.

- However, for the models to be useful for forecasting, it is necessary that they are entered with a lag in relation to outcome.

- That is, if survival time is $t$, then behavioural data from time $t-k$ is included, for some lag time $k$.

- This means we can forecast outcome for some time $k$ ahead.
Example 6.4

A behavioural model with default as failure event for a credit card data set ($n \approx 400,000$).

Coefficient estimates for a model with *fixed* application variables (AV) and *time varying* monthly behavioural variables (BV).

- Indicator variables are denoted by a plus sign (+).
- Statistical significance levels are denoted by asterisks:
  - ** is less than 0.001 and
  - * is less than 0.01 level.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selected AVs:</strong></td>
<td></td>
</tr>
<tr>
<td>Time customer with bank (years)</td>
<td>-0.00250**</td>
</tr>
<tr>
<td>Income (log)</td>
<td>-0.146**</td>
</tr>
<tr>
<td>Number of cards</td>
<td>-0.0610**</td>
</tr>
<tr>
<td>Time at current address</td>
<td>-0.00129</td>
</tr>
<tr>
<td>Covariate</td>
<td>Estimate</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Employment + :</td>
<td></td>
</tr>
<tr>
<td>Self-employed</td>
<td>+0.303**</td>
</tr>
<tr>
<td>Homemaker</td>
<td>+0.072</td>
</tr>
<tr>
<td>Retired</td>
<td>+0.111</td>
</tr>
<tr>
<td>Student</td>
<td>-0.035</td>
</tr>
<tr>
<td>Unemployed</td>
<td>+0.231</td>
</tr>
<tr>
<td>Part time</td>
<td>-0.365**</td>
</tr>
<tr>
<td>Other</td>
<td>-0.037</td>
</tr>
<tr>
<td>Excluded category: Employed</td>
<td></td>
</tr>
<tr>
<td>Age + : 18 to 24</td>
<td>+0.074</td>
</tr>
<tr>
<td>25 to 29</td>
<td>-0.058</td>
</tr>
<tr>
<td>30 to 33</td>
<td>+0.010</td>
</tr>
<tr>
<td>34 to 37</td>
<td>+0.100**</td>
</tr>
<tr>
<td>38 to 41</td>
<td>+0.046</td>
</tr>
<tr>
<td>48 to 55</td>
<td>-0.108**</td>
</tr>
<tr>
<td>56 and over</td>
<td>-0.243**</td>
</tr>
<tr>
<td>Excluded category: 42 to 47</td>
<td></td>
</tr>
<tr>
<td>Generic credit score</td>
<td>-0.00322**</td>
</tr>
</tbody>
</table>
### Covariate Estimate

<table>
<thead>
<tr>
<th>Behavioural variables, lag 12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payment status + :</td>
</tr>
<tr>
<td>Fully paid</td>
</tr>
<tr>
<td>Greater than minimum paid</td>
</tr>
<tr>
<td>Minimum paid</td>
</tr>
<tr>
<td>Less than minimum paid</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
</tbody>
</table>

*Excluded category*: No payment

<table>
<thead>
<tr>
<th>Current balance (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(log squared)</td>
</tr>
<tr>
<td>is zero +</td>
</tr>
<tr>
<td>is negative +</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Credit limit (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Payment amount (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>is zero +</td>
</tr>
<tr>
<td>is unknown +</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of months past due</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Past due amount (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>is zero +</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Covariate</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>Number of transactions</td>
</tr>
<tr>
<td>Transaction sales amount (log)</td>
</tr>
<tr>
<td>&quot; is zero +</td>
</tr>
<tr>
<td>APR on purchases</td>
</tr>
<tr>
<td>&quot; is zero +</td>
</tr>
<tr>
<td>Behavioural data is missing +</td>
</tr>
</tbody>
</table>
Example 6.4 continued

Forecasting using survival models with BVs, using a Deviance model fit measure.

Forecasts of time to default improve with the inclusion of BVs and shorter lag time.

Of course, a shorter lag implies a shorter period to forecast ahead.
Exercise 6.1

Interpret the association of each of these behavioural variables with the default hazard rate in the model given in Example 17.1:

- Payment status
- Current balance
- Credit limit
- Number of months past due
- Number of transactions
- Transaction sales amount
Exercise 6.2

a) Let $h(t)$ be the hazard function at time $t$. Show that the survival probability is given by $S(t) = \exp\left(-\int_0^t h(u)du\right)$.

b) A hazard function for default is given by

$$h(t) = \begin{cases} 0, & \text{for } t < 3 \\ re^{-kt}, & \text{for } t \geq 3 \end{cases}$$

for some $r > 0$ and $k > 0$. Suppose we want to ensure probability of default at time $t$ is less than a given value $p_t$ and $k$ is fixed. Then, what is the inequality constraint on $r$?
c) Interpret the following Cox Proportional Hazards model of time to default:

   i. Which are the statistically significant variables at a 1% level?
   ii. What effect does each variable have on default hazard risk?

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Range of values</th>
<th>Coefficient Estimate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment status at application</td>
<td>1 (yes) or 0 (no)</td>
<td>-0.50</td>
<td>0.001</td>
</tr>
<tr>
<td>Generic credit score</td>
<td>0 to 999</td>
<td>-0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>Current balance (log), lag 6 months</td>
<td>0 to 6</td>
<td>+0.20</td>
<td>0.121</td>
</tr>
<tr>
<td>Payment missed, lag 6 months</td>
<td>1 (yes) or 0 (no)</td>
<td>+1.20</td>
<td>0.001</td>
</tr>
</tbody>
</table>
References for Survival models


There are actually many other good text books on Survival modelling.
We now move on to a new dynamic model structure...

Markov transition models (or *Markov chains*) are a dynamic approach to modelling processes with changes of state.

They are valuable in credit scoring since they allow us to model changes in the state of an account over time. For instance,

- Modelling the number of account periods of delinquency.
- Changes in behavioural score.

Markov transition models are especially useful for modelling revolving credit with highly variable credit usage.

- For instance, for tracking credit card use.
Some definitions:

- Let $X_0, X_1, X_2, \ldots$ be a sequence of random variables taking values from \{1,2, \ldots, K\} for some fixed $K$.

- The sequence is a **finite-valued first-order Markov chain** if
  \[
  P(X_{n+1} = j \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = i) = P(X_{n+1} = j \mid X_n = i)
  \]
  for all $n$ and $i,j$ such that $1 \leq i \leq K$ and $1 \leq j \leq K$.

- We denote $p_n(i,j) \triangleq P(X_n = j \mid X_{n-1} = i)$ and call this the **transition probability**.
  - The transition probability represents the probability of moving from one state $i$ to another state $j$.

- The **transition matrix** $P_n$ is defined as a $K \times K$ matrix such that $P_n[i,j] \triangleq p_n(i,j)$. 

"First-order Markov transition model"
• If we make a prior assumption that a transition from state $i$ to state $j$ in the $n$th period then we fix $p_n(i, j) = 0$ and call this a **structural zero**.

• A Markov chain is **stationary** if $P_n = P$ for all $n$ for some transition matrix $P$. That is, the transition probabilities are the same over all periods.

Notice that

$$P(X_n = j \mid X_{n-2} = i) = \sum_{k=1}^{K} P(X_n = j, X_{n-1} = k \mid X_{n-2} = i)$$

$$= \sum_{k=1}^{K} P(X_n = j \mid X_{n-1} = k, X_{n-2} = i)P(X_{n-1} = k \mid X_{n-2} = i)$$

$$= \sum_{k=1}^{K} P(X_n = j \mid X_{n-1} = k)P(X_{n-1} = k \mid X_{n-2} = i)$$

by the law of total probability and first-order Markov chain assumption.
Therefore,

\[ P(X_n = j \mid X_{n-2} = i) = \sum_{k=1}^{K} p_{n-1}(i,k)p_n(k,j) \]

\[ = (P_{n-1}P_n)[i,j] \]

We can easily extend this result to get

\[ P(X_n = j \mid X_0 = i) = (P_1 P_2 \ldots P_n)[i,j] \]

Let \( \pi_n \) be the distribution of \( X_n \) so that \( \pi_n(i) \equiv P(X_n = i) \).

Then, since \( P(X_n = j) = \sum_{i=1}^{K} P(X_n = j \mid X_0 = i)P(X_0 = i) \),

\[ \pi_n = \pi_0(P_1 P_2 \ldots P_n) \]
Example 6.5

Consider a two state stationary Markov chain for behavioural score change (state 1=high score, 2=low score) with transition matrix

\[ P = \begin{pmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{pmatrix} \]

Suppose we start with an individual having high score. What are the distributions after one and two periods?

Solution
\[ \pi_0 = (1 \ 0) \]

Therefore, after one period:
\[ \pi_1 = \pi_0 P = (1 \ 0) \begin{pmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{pmatrix} = (0.95 \ 0.05) \]

And, after two periods,
\[ \pi_2 = \pi_0 P^2 = (0.95 \ 0.05) \begin{pmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{pmatrix} = (0.9075 \ 0.0925) \]
Use maximum likelihood estimation (MLE) for each $p(i,j)$.

Given a sequence of $n$ realizations $x_0, x_1, \ldots, x_n$, the probability of this realization is given as

$$P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \left( \prod_{t=1}^{n} P(X_t = x_t | X_0 = x_0, \ldots, X_{t-1} = x_{t-1}) \right) P(X_0 = x_0)$$

$$= P(X_0 = x_0) \prod_{t=1}^{n} P(X_t = x_t | X_{t-1} = x_{t-1})$$

$$= P(X_0 = x_0) \prod_{t=1}^{n} p(x_{t-1}, x_t)$$
Therefore, the log-likelihood function is

\[ L(p) = \log P(X_0 = x_0) + \sum_{t=1}^{n} \log p(x_{t-1}, x_t) = \log P(X_0 = x_0) + \sum_{i=1}^{K} \sum_{j=1}^{K} n_{ij} \log p(i, j) \]

where \( n_{ij} \triangleq |\{t: x_{t-1} = i, x_t = j\}|. \)

However, the likelihood function is constrained by \( \sum_{j=1}^{K} p(i, j) = 1. \)

Therefore, choose some \( r \) such that \( 1 \leq r \leq K \) and substitute

\[ p(i, r) = 1 - \sum_{j \in \{1, \ldots, K\} - \{r\}} p(i, j) \]

to get

\[ L(p) = \log P(X_0 = x_0) + \sum_{i=1}^{K} \left[ \left( \sum_{j \in \{1, \ldots, K\} - \{r\}} n_{ij} \log p(i, j) \right) + n_{ir} \log \left( 1 - \sum_{j \in \{1, \ldots, K\} - \{r\}} p(i, j) \right) \right] \]
Then find the derivative with respect to each $p(i,j)$ where $j \neq r$ and set to zero to find the maxima:

$$\frac{\partial L(p)}{\partial p(i,j)} = \frac{n_{ij}}{p(i,j)} + \frac{n_{ir}}{p(i,r)} \times -1 = 0$$

Therefore,

$$\hat{p}(i,j) = n_{ij} \frac{\hat{p}(i,r)}{n_{ir}} \propto n_{ij}$$

But, the choice of $r$ is arbitrary so for consistency the result must hold generally for all $j$.
In particular, the MLE is

$$\hat{p}(i,j) = \frac{n_{ij}}{\sum_{l=1}^{K} n_{il}}$$

Notice that this result easily generalizes to the case when we have multiple sequences of realizations (eg more than one borrower), so long as we assume independence between each sequence.
Example 6.6

Consider three states (1=high score, 2=low score, 3=default) for a stationary process. Transition probabilities are given as:

- from a high score to a low score is 0.05;
- from a low score to a high score is 0.1;
- from a low score to default is 0.02.

It is impossible to move from high score to default. Also, it is impossible to move out of default.

1. What is the transition matrix?
2. How many structural zeroes are there in the matrix?

Solution

\[
P = \begin{pmatrix}
0.95 & 0.05 & 0 \\
0.1 & 0.88 & 0.02 \\
0 & 0 & 1
\end{pmatrix}
\]

There are 3 structural zeroes.
An obvious omission from the Markov chain formulation is the lack of predictor variables.

There are two ways to include borrower details in the model:

1. Include behavioural variables within the state space.

2. Segment the population on static variables and build segmented Markov transition models.
Both methods suffer from similar problem:

1. Increasing the state space means more transition probabilities need to
   be estimated and this will mean reduced estimation efficiency.

2. Segmentation will mean several distinct Markov chains, each based on a
   reduced training sample.

3. Neither method allows for continuous data, unless it is discretized, and
   there is a limit to the number of categorical variables that can be used in
   states or separate models.
Example 6.7

Suppose we want to include credit usage, in terms of monthly spend in a model for behavioural score (Low or High).

- First discretize credit usage into levels:
  - eg three levels: monthly spend < £200, ≥£200 and < £1000, ≥£1000,
- Then, form 6 states, instead of 2:

<table>
<thead>
<tr>
<th>Behavioural score</th>
<th>Monthly spend</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>&lt; £200</td>
<td>1</td>
</tr>
<tr>
<td>Low</td>
<td>≥£200 and &lt; £1000</td>
<td>2</td>
</tr>
<tr>
<td>Low</td>
<td>≥£1000</td>
<td>3</td>
</tr>
<tr>
<td>High</td>
<td>&lt; £200</td>
<td>4</td>
</tr>
<tr>
<td>High</td>
<td>≥£200 and &lt; £1000</td>
<td>5</td>
</tr>
<tr>
<td>High</td>
<td>≥£1000</td>
<td>6</td>
</tr>
</tbody>
</table>
Example 6.8

Research suggests two broad categories of credit card usage: the movers and stayers.

- **Movers** are those whose credit card usage is erratic; having periods of heavy credit card usage then quiet periods.

- **Stayers**, by contrast, tend to be steady, and stay in the same state over long periods.

We could build a static behavioural model to broadly categorize borrowers into one of the two categories.

Then separate Markov transition models could be built separately for the two segments.
Exercise 6.3
Three credit card account states are defined as
   1 = Good; account being paid off fully;
   2 = Minimum repayments in a month;
   3 = Bad; minimum repayment is not made.

A mover account profile is then given as a sequence of states:
   1,1,1,1,2,2,3,1,2,1,1,1,2,2,2,3.

A stayer account profile is then given as a sequence of states:
   1,1,1,1,1,1,2,2,2,2,3,3,2,2,2,1.

1. Use maximum likelihood estimation to compute probability transition matrices for both accounts for a first-order Markov transition model.
2. If each account is in state 2 in time \( t \), what is the probability that it will move to state 3 in time \( t + 1 \) or \( t + 2 \)?
A roll-rate model is a type of Markov transition model but the focus is on the number of accounts or value of loans that rolls over from one level of delinquency to another over several months.

- Consider $K$ states where 0 corresponds to no delinquency, states >0 correspond to increasing levels of delinquency and $K$ corresponds to loan default with write-off.
- Let $A$ be a vector of initial number of accounts or value of loans.
- Let $P$ be a $K \times K$ transition matrix.

Then the vector of values in each state at month $t$ is given by $AP^t$. 
Example 6.9

Let $K = 3$.

Let $A = (50000, 10000, 5000, 1000)$, in GB£.

Let $P = \begin{pmatrix} 0.98 & 0.02 & 0 & 0 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0.1 & 0.05 & 0.55 & 0.3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

Let first month ($t = 0$) be January 2013.

Then roll-rate table (projection) for six months is computed as:-

<table>
<thead>
<tr>
<th>Month</th>
<th>Computation</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Jan 13</td>
<td>$A$</td>
<td>50000</td>
</tr>
<tr>
<td>Feb 13</td>
<td>$AP$</td>
<td>52500</td>
</tr>
<tr>
<td>Mar 13</td>
<td>$AP^2$</td>
<td>53600</td>
</tr>
<tr>
<td>Apr 13</td>
<td>$AP^3$</td>
<td>54033</td>
</tr>
<tr>
<td>May 13</td>
<td>$AP^4$</td>
<td>54121</td>
</tr>
<tr>
<td>Jun 13</td>
<td>$AP^5$</td>
<td>54020</td>
</tr>
</tbody>
</table>
We covered the following topics on behavioural models:

- Behavioural models and data;
- Static behavioural models;
- Dynamic models of behaviour:
  - Survival models;
  - Markov transition models.