A statistical model for the temporal pattern of individual ATM withdrawals

Adam R. Brentnall\textsuperscript{a,1}, Martin J. Crowder\textsuperscript{b}, David J. Hand\textsuperscript{a,b}

\textbf{a} Institute for Mathematical Sciences, Imperial College London.

\textbf{b} Department of Mathematics, Imperial College London.

\section*{Abstract}

Models of consumer behaviour based purely upon empirical relationships in data can perform well in the short-term, but often degrade rapidly with changing circumstances. Superior longer term performance can sometimes be attained by developing models for the deeper processes underlying the consumer behaviour. In this paper, we develop a random-effects point-process model for ATM (Automated Teller Machine) withdrawals. Estimation, prediction and computational issues are discussed. The model may be used to predict behaviour for an individual, assess when state changes in individual behaviour have occurred, and as a description of behaviour for a portfolio of accounts.

\textbf{Keywords:} ATM, Consumer, Point-Process, Prediction, Random-Effects, Retail Financial Services

\section*{1 Introduction}

Financial institutions (banks, for short) collect large amounts of data on their customers, and aim to use these data to help make better decisions. Trends and patterns of behaviour contained in the data may be used to guide a bank’s customer-value management strategies in areas such as risk management, profit maximization, fraud detection, and marketing. For example, if a pattern of transactions known to be linked to a high probability of defaulting is detected in an individual’s account, then the bank can use this as a trigger to target that individual with appropriate management strategies. That is, some suitable action can be taken.

Such models for predicting outcomes are not new in retail banking; for example, scorecards for predicting likely defaulters have been used since the late 1960s. Statistical tools used in such models include logistic regression, linear discriminant analysis, and classification trees, and tools applied in fraud detection include neural networks and other more recently-developed methods (Thomas et al., 2005). Most of the data-analytic techniques that have been used seek out trends and patterns in the data without necessarily requiring an explanation for them. That is, they are purely empirical and data-driven, without any underlying theory or rationale. This has occasionally caused concern amongst the regulatory authorities when it has led to the inclusion of variables that have no apparent theoretical relationship with the response (with default probability, for example).

This data-driven approach may be effective in the short term, but it can also be vulnerable to the so-called plateau effect, in which a model that yields good predictions for data drawn from distributions very similar to those on which it
was based, degrades catastrophically as those distributions change: the predictive performance of the model falls off the edge of a plateau. In particular, if the circumstances under which the model is applied change, then a model that is based on an empirical fit to data collected under different circumstances may no longer yield good predictions. This is of especial relevance in the retail banking context, since the industry is a highly-changeable one, fluctuating in response to changing economic conditions, political decisions, changing competition, advances in technology, and the development of new kinds of financial products. Indeed, the very development of highly predictive models will itself influence the behaviour of the system. Given all this, all we can say for sure is that the future will be rather different from the past.

While one can never predict the future with perfect accuracy, dramatic degradation of the performance of the model can sometimes be eased by basing the model not purely on empirically observed relationships in the data but on deeper models of the underlying processes. This suggestion has parallels with that made by Robert Lucas in the 1970s in the context of certain econometric models. Lucas (Lucas Jnr., 1946) argued that, while econometric models based upon empirical evidence may be suitable for short-term forecasting, in the longer term their predictive power is of no use and can not be used as a basis of assessing the effect of interventions. A classic example of this is the breakdown in the 1970’s of the empirical law that high inflation corresponds to low unemployment (also known as the Phillips curve). In our context, the deepest level of model would be based on psychological studies of human behaviour, of the kind beginning to be explored in behavioural finance. At a level that is not quite so deep, stochastic models can be based on assumptions about the underlying data-generating process. It is this less-deep level that we explore in this paper. In particular, the aim of this paper is to explore such models in the context of withdrawals from Automated Teller Machines (ATMs).

Section 2 of the paper describes the data and Section 3 identifies four features of ATM withdrawal patterns. These form the basis for the point process model described in Section 4. The statistical form of this model, a number of statistical and computational issues in estimation and prediction, and some ways in which the model may be used for an individual account or a population of accounts are discussed. The final section demonstrates how the model may be used by applying estimation and prediction methods to the data set.

2 Initial examination of ATM transaction patterns

The analysis described in this paper is based on one thousand accounts from a UK high-street bank, with behaviour recorded over a four-month period in 2005. These data are a purely random sample (not stratified) from all accounts with the bank that made at least one transaction during the time period. Accounts equate to card numbers, so there could be more than one cardholder for each account. We do not know the number of cardholders per account because these data were not supplied, but it is likely that there is only one cardholder for most accounts. For each ATM withdrawal the time (date, hour, minute, second) has been recorded. The withdrawals may be for cash or pay-as-you-go mobile-phone
top-ups; these cannot be separated in our data set. There were no serious data quality issues that we could identify and we believe the timings are accurate. Some exploratory analysis is presented in this section to describe the temporal process of ATM withdrawals.

An initial exploratory examination of the data found four main features of ATM withdrawal patterns: (1) variation in the overall rate, (2) a cyclical rate pattern within weeks, (3) a fairly constant long-term rate of withdrawal and (4) some accounts making consecutive withdrawals in quick succession. These findings are now described in more detail.

**Overall rate.**

There is a large variability in the distribution of the number of withdrawals made by individual accounts, shown as a histogram in Figure 1. The maximum number of withdrawals was 219, and 35 accounts did not withdraw any cash, but used ATMs for other reasons such as requesting a bank statement.

**Cyclical behaviour of withdrawals within week.**

Different accounts show different patterns of variation over the course of a week, and a number of different types of behaviour are apparent from the plots in Figure 2, which show the average weekly cumulative number of withdrawals of four selected accounts (solid lines); the straight line joining the end-points of the trace corresponds to a constant rate of withdrawal. The plots show differences in withdrawals in time of day (the seven steps) and day of week (departure from the line of constant rate). For example, the account in the top-left plot made relatively more withdrawals towards the start of the week (Monday), whereas the top-right account is skewed towards Fridays. For accounts with several withdrawals in the four month period, it is...
were made. Figure 3 shows the observed cumulative rate over the entire period, Kolmogorov-Smirnov tests rejected the hypothesis of independent, exponentially-distributed times between transactions. It is apparent that one reason for this is that the times of transaction often exhibit a cyclical pattern around the hour of day and day of week, making the rate non-constant. In general, although different patterns were observed in different accounts, the cyclical nature seems to be a common feature. So there is a need to look at the details of the event process, not just the average times between withdrawals. For example, if ‘x’ represents a withdrawal and ‘-’ no withdrawal in a time period then the pattern - x x - - x x - - x x - gives the same inter-event times as - x - - x x x - x -, but these correspond to very-different behaviours.

**Variability in number of weekly withdrawals within accounts.**

One might also expect cyclical patterns to occur between weeks. For example, Christmas always falls on the 25th December, and perhaps more visits to ATMs might be made in the run-up; bank holidays provide another example. However, the sort of variability that we observed in the data did not appear to be systematic across accounts. Most individuals appeared to have a fairly-constant rate through weeks, but some accounts had periods of very different rates where, for example, no withdrawals were made. Figure 3 shows the observed cumulative rate over the entire four month period for four selected accounts. Only one of these exceeds 95% confidence bars, based on the Kolmogorov-Smirnov statistic (Davidson, 2003), for the variation that might be expected from a stationary Poisson process. It might be argued that this account changes behaviour around half-way through the time period, with a reduced rate of withdrawal from this point on. Under a null Poisson model, we would expect
about 48 accounts from the 965 who made cash withdrawals to exceed the 95\% bars, but found 150. Many of these exhibited evidence of potential state-changes in behaviour. Longer-term changes of state might be linked to more permanent changes such as employment status or salary, whereas short-term variation between weeks and months might be explained by a change of state for the account holder, such as a holiday, or a business trip abroad. In general, the feature of between-week variation that we consider is likely to be linked to a behavioural-state change.

Quick-succession withdrawals.
In the four-month period, 678 withdrawals were made within two minutes of a previous one, which compares to an expected 9 or 10 such withdrawals under a null Poisson model. Around 36\% of accounts had at least one withdrawal within ten minutes of a previous one, some 4\% of all withdrawals occurred within 10 minutes of a previous withdrawal and, for approximately 1 in 5 accounts, 10\% of their withdrawals took place within ten minutes of another. Quick-succession withdrawals may be partly explained by the practice of some banks of limiting the amount that can be withdrawn at one time to less than the total amount that may be withdrawn in a twenty-four hour period. Alternatively, since our data are a mixture of cash withdrawals and mobile top-ups, a partial explanation may be that the same ATMs are used for cash withdrawals and mobile top-ups within quick succession.

3 Point process models for individuals

We take the primary feature for ATM withdrawals to be the rate at which they occur. A Non-Homogeneous Poisson Process (NHPP) is often used when the rate of events is assumed to be independent of the history $H$ of previous event times. For a NHPP, structure in the rate $\lambda(t)$ may be modelled through links to a vector of explanatory variables $\mathbf{x}(t)$, commonly via a log-linear form $\log \lambda(t) = \mathbf{x}(t)^T \beta$ (Cox and Lewis, 1966). A NHPP may be extended by allowing the rate $\lambda(t|H_t)$ to be conditional on the entire history $H_t$ up to time $t$. For example, the occurrence of a withdrawal might increase the probability that there will be another withdrawal soon afterwards. This type of point process is called a self-exciting point process (Cox and Isham, 1980). In this section we present some way of incorporating some features of the ATM withdrawals into a rate function $\lambda(t|H_t)$.

The first feature we consider is cyclical within-week variation. We have examined three ways to model this:

1. **Piecewise-constant rate.**
   In a log-linear NHPP model we might take $x_1, \ldots, x_7$ as indicators for the day of week and $x_8$ as an indicator for daylight. This approach has the disadvantage of requiring judgement on where to draw the lines between day and night. If further coefficients are used then more complex piecewise constant rates might be modelled.

2. **Fourier series.**
   We take the components of $\mathbf{x}(t)$ to be periodic in a log-linear model.
Figure 3: Cumulative ATM withdrawal rate for four accounts. Variation between weeks is compared relative to a stationary Poisson process. The 95% and 99% point-wise confidence bars based on a Kolmogorov-Smirnov statistic are shown as dashed lines.
This removes dependency of the model on selected fixed points and allows more forms of behaviour than a piecewise-constant rate. We experimented with various different versions of Fourier series in which two fundamental frequencies were used, one to accommodate variation during the day, the other to accommodate variation over the week. However, we found that, in practice, the Fourier series could be quite parameter-hungry; the model we used required a total of 16 parameters before the shape of the rate functions appeared realistic for a number of individuals.

3. **Cyclic weighting.**

The rate may be scaled using weighting functions so that events are more likely to occur at certain times. Suitable weighting functions include the probability density functions of the Beta distribution and Kumaraswamy’s double-bounded distribution. Both require two parameters to be estimated. Two such weight functions may be used, one to model within-day variation and another for within-week variation. If Beta or Kumaraswamy functions are used then five parameters need to be estimated: one for an overall rate plus two for each weight function. Choosing such a shape restricts the cyclical pattern around the time in week or day to be unimodal. However, it is advisable to avoid models with many parameters when the data are highly variable and sparse, as here. The approach is attractive because the model requires relatively few parameters, does not require any subjective judgement on cut-off points, is easy to interpret and allows for a wide range of patterns of behaviour including the null Poisson.

We also consider how quick-succession withdrawals might be modelled; a quick-succession withdrawal is one that takes place very soon after a previous one. For an individual who systematically makes such withdrawals the rate of withdrawals should be increased just after a withdrawal and then return to normal over time. This may be modelled by a boost followed by an exponential decay: for \( t > t_r \), with \( t_r \) as the time of the most recent event, \( \lambda(t) \) is multiplied by \( \exp\{\beta_1 \exp[\beta_2 (t_r - t)]\} \) where \( \beta_1 \) and \( \beta_2 > 0 \) are parameters. In this way the process is defined to be dependent on its history through the time of the previous event. With \( \beta_1 > 0 \), \( \lambda(t) \) is increased in the time period following a withdrawal. However, \( \beta_1 < 0 \) is also allowed which permits the rate to be reduced following a withdrawal. Such behaviour might be referred to as self-inhibiting rather than self-exciting.

4 ATM withdrawal model

In this section we build on our exploratory analysis to construct a likelihood function for the ATM withdrawal times. Methods for estimation and prediction are outlined and we suggest some ways in which the model may be interpreted and used to help banks make better decisions.

4.1 Intensity function

The following intensity function has been designed to allow the majority of individuals to make withdrawals at a fairly constant rate week-on-week, but also to allow the rate of withdrawals to vary within weeks and within days. For
a generic individual the model is defined using a rate function $\lambda(t|H_t)$ that is
c conditional on the history $H_t$ of previous events up to time $t$; in fact, $H_t$ is
t here taken to specify only the most recent event-time up to time $t$, say $t_r$, and
the unit of time is one day. Also, to reflect the fact that the parameters are to
be regarded as random effects, varying over individuals, we denote them by $u$
rather than $\beta$.

Let $b(s; a_1, a_2) = B(a_1, a_2)^{-1} s^{a_1-1} (1 - s)^{a_2-1}$ where $B(\ldots)$ is the Beta function,
and $w_1(t; u_2, u_3) = \int_{I_t} b(s; u_2, u_3) ds$, where $I_t$ denotes the interval $\{ [d(t) - 1]/7, d(t)/7 \}$ and $d(t) = 1 + t \mod 7$ (the day of week). Let $w_2(t; u_4, u_5) = \begin{cases} b(a(t); u_4, u_5) & \text{where } a(t) = t \mod 1 (\text{the time in day}). \end{cases}$
For $t > t_r$, we define

$$\lambda(t|H_t) = w_1(t; u_2, u_3) \times w_2(t; u_4, u_5) \times \exp[u_1 + u_6 \exp[u_7(t_r - t)]] \quad (1)$$

where $u_1$ and $u_6$ are unrestricted while the other parameters are greater than
zero. The intensity function may be interpreted in the following way.

- Each individual has a propensity to make withdrawals linked to the cal-
endar. A constant rate, $\exp(u_1)$, is scaled during the week to take into
account the hour of the day, through $w_2(t; u_4, u_5)$, and the day-of-week
proportion $w_1(t; u_2, u_3)$.

- Each individual also has a propensity to make withdrawals linked to their
history $H_t$, through the previous withdrawal time. Withdrawal rate may
be boosted (or reduced) by a multiplying factor $\exp(u_6)$, decaying by
$\exp[u_7(t_r - t)]$ following a withdrawal at time $t_r$. This allows for a form
of quick-succession or delayed-succession behaviour.

- If $u_6$ is greater (less) than zero then $\exp(u_1)$ may be interpreted as a lower
(upper) bound on the cumulative rate over a week.

- If $u_4$ is greater (less) than $u_2$ then relatively more (less) withdrawals will
occur at the start of the week. If both parameters are equal then the day-
of-week distribution will be symmetric around the middle of the week,
and if they are both unity then there is no day-of-week effect. A similar
interpretation may be made for the time-of-day parameters.

4.2 Likelihood function

In the following we use the general notation $p(D|u)$ to denote a likelihood
function, where $p$ denotes a probability density or mass function, $D$ the data
set and $u$ the parameter set. The likelihood for individual $i$, based on data
$D_i(t)$, comprising withdrawal times $t_{ij}$ up to time $t$, is (Cox and Lewis, 1966):

$$p(D_i(t)|u_i) = e^{-\Lambda_i(t|u_i)} \prod_{j=1}^{N_i(t)} \lambda_i(t_{ij}|H_{it_{ij}}, u_i), \quad (2)$$

where $N_i(t)$ is the number of events in period $(0,t]$, $\Lambda_i(t|u_i) = \int_0^t \lambda_i(s|H_{is}, u_i) ds$
and $u_i = (u_1, u_2, \ldots)$ is the individual random-effect parameter vector for
individual $i$.  

8
Let the distribution of the random effect parameters $u_i$ over individuals be $p(u; \theta)$, where $\theta$ is the underlying parameter vector. Then the likelihood function for individual $i$ is:

$$p(D_i; \theta) = \int p(D_i | u_i) p(u_i; \theta) \, du_i$$

(3)

Assuming that the $n$ individual records are independent, the likelihood for the complete data $D = \{D_1, \ldots, D_n\}$ is:

$$p(D; \theta) = \prod_{i=1}^{n} p(D_i; \theta).$$

(4)

### 4.3 Estimation of random effects distributions

The likelihood for the random-effects model in equation (4) depends on the seven-dimensional integral in equation (3). It is not possible to integrate this analytically and so computational methods need to be employed. We considered several approaches.

A first approach that might be taken is to approximate equation (3) using Monte-Carlo integration and to explore the likelihood in (4) using a general search algorithm. This suffers from requiring the Monte-Carlo integral to be sufficiently accurate that the search algorithm does not become confused by sampling error. In order to achieve this a large number of Monte-Carlo samples need to be generated at each iteration, to reasonably approximate the seven-dimensional integral. This procedure has been suggested by Geyer and Thompson (1992) and is called simulated maximum likelihood by McCulloch (1997) in the context of generalized linear mixed models.

The computational problem faced in estimating generalized linear mixed models is identical to ours: an integral with dimension equal to the number of random-effects parameters needs to be evaluated. A number of authors have proposed versions of the Expectation-Maximisation (EM) algorithm in this context. McCulloch (1997) developed a Monte Carlo EM algorithm, where calculation of the expectation step is based on samples generated using a Metropolis algorithm. Booth and Hobert (1999) built on the algorithm, using Monte-Carlo simulation instead on the Metropolis algorithm and developed rules to increase the simulation effort, based on a sandwich estimator of the Monte Carlo error. Further developments were proposed in Caffo et al. (2005), who recovered the normal EM algorithm ascent property with high probability. However, in all these algorithms there is no guarantee of convergence to the MLE and the number of Monte-Carlo simulations required for our problem make their use computationally prohibitive.

Because of the difficulties described above, instead of making parametric assumptions for the joint probability distribution for $u_i$ over individuals, we propose to use the empirical distribution function (EDF) of the estimated random effects from fits to many individual accounts. In consequence, $p(u; \theta)$ is replaced by $\hat{\pi}(u)$, say, which is a discrete probability mass function with atoms $1/n$ at each $\hat{u}_i$. In addition to avoiding specification of $p(u; \theta)$ there is the computational advantage that integrals involving $p(u; \theta)$, such as (3), become finite summations. For generalized linear mixed models this approach is unattractive.
because some approximation would be needed to estimate fixed effects, but for our point process model there are only random effects.

The following argument supports the use of such an empirical distribution. We assume that the unknown random-effects parameters \( \{u_1, \ldots, u_n\} \) form a random sample from some underlying density \( p(u; \theta) \). These random effects parameters give rise to the observations \( \{D_1, \ldots, D_n\} \) from which maximum likelihood estimates \( \{\hat{u}_1, \ldots, \hat{u}_n\} \) are obtained. As the individual sample sizes of the \( D_i \), become large \( \{\hat{u}_1, \ldots, \hat{u}_n\} \rightarrow \{u_1, \ldots, u_n\} \) in probability for fixed \( n \). Thus, the EDF of \( \{\hat{u}_1, \ldots, \hat{u}_n\} \) will provide a reasonable approximation to that of \( \{u_1, \ldots, u_n\} \), which is the standard estimator for the distribution function corresponding to \( p(u; \theta) \).

### 4.4 Prediction

Using the point-process model of ATM withdrawals it is possible to make predictions for individual account behaviour. Two aspects of behaviour that might be predicted for a temporal point process are the number of events and their times. Other model-based statistics may also be used to help identify changes in behaviour, perhaps linked to a change in state.

The posterior density of the random effects for each individual \( i \) may be estimated using the empirical distribution of random effects \( \hat{\pi}(u) \) in the following way:

\[
\hat{p}(u|D_i) = \frac{p(D_i|u)\hat{\pi}(u)}{\sum_u p(D_i|u)\hat{\pi}(u)}
\]

(5)

where \( \sum_u \) denotes summation over the observed \( u \)-values. Note that \( \hat{p}(u|D_i) \) is thus a discrete distribution with non-zero probability masses at the observed \( u \)-values. In general, a point estimate of \( u \) based on the expected value of the posterior distribution will tend to shrink towards the overall mean. Point estimates of the parameters may be of interest in, for example, understanding the weekly profile shape of withdrawal activity.

The posterior distribution may also be used to predict some future aspect, say \( C \), of the behaviour for \( t > T \):

\[
p(C|D) = \sum_u p(C|D, u)\hat{p}(u|D)
\]

(6)

Unfortunately it is difficult to write down explicit expressions for all \( p(C|D, u) \) related to numbers and times of events for self-exciting processes such as ours. However, predictions about the next one or two events are possible. Examples of future occurrences which may be of interest include:

- The probability density of the time \( t \) of next withdrawal for an individual account, which is

\[
p(t|t_r, u) = \lambda(t|t_r, u)e^{-\Lambda(t|u)+\Lambda(t_r|u)},
\]

where \( t_r \) is the most recent withdrawal time.

- The probability of zero withdrawals can be computed using

\[
P[N(t_r, t) = 0|t_r, u] = e^{-\Lambda(t|u)+\Lambda(t_r|u)}.
\]
The probability of one withdrawal

\[ P\{N(t_r, t] = 1 | t_r, u\} = \int_{t_r}^{t} \lambda(s | t_r, u) e^{-\Lambda(t_r | u) + \Lambda(t | u)} ds. \]

More generally, the probability of \( k \) withdrawals over the period \((t_r, t] \) is found by integrating over withdrawal times \( t_r < t_1 < t_2 < \ldots < t_k \leq t \) as

\[ P\{N(t_r, t] = k | t_r, u\} = \int_{t_r}^{t} \cdots \int_{t_{k-1}}^{t} \lambda(t_1 | t_r, u) \cdots \lambda(t_k | t_{k-1}, u) e^{-\Lambda(t_r | u) + \Lambda(t | u)} dt_1 \cdots dt_k. \]

The last of these expressions rapidly becomes computationally prohibitive as \( k \) increases.

Another approach to monitoring observed against predicted behaviour is to use the cumulative rate function. Suppose we have observed data \( D_1 \) on individual \( i \) over time period 1 and then new data \( D_2 = \{t_1, \ldots, t_{n_i}\} \) arrives. Then the observed cumulative rate function in the new time period may be compared to the estimated function \( \Lambda(t | D_2, \hat{u}) \), where \( \hat{u} \) is based on \( D_1 \).

One way to do this is to transform the observed times to be a realisation from a stationary Poisson Process through \( \{\Lambda(t | D_2, \hat{u}), \ldots, \Lambda(t_n | D_2, \hat{u})\} \). This can be tested using goodness-of-fit methods such as the Kolmogorov-Smirnov statistic. A disadvantage of transforming the times is that the tests will not be very powerful if few observations are recorded, and also may not detect change in overall rate. An alternative way to monitor behaviour is to take the difference between the observed and predicted cumulative rate functions. If the model is correct then these should form a Martingale sequence. Plots of standardised Martingale residuals may be used as a way to visually assess change in patterns of behaviour.

### 4.5 Dynamic updating

In practice, the model will be updated through time. When new data \( D_2 \) arrives the predictive density of an individual’s random effect parameters can be updated using:

\[
p(u|D^1, D^2) = \frac{p(D^2|u, D^1)p(u|D^1)}{p(D^2|D^1)} = \frac{p(D^2|u, D^1)p(u|D^1)}{p(D^2|D^1)} = \frac{p(D^2, D^1|u)p(u|D^1)}{p(D^1|u)p(D^2|D^1)} \tag{7}
\]

where \( p(D^2|D^1) = \int p(D^2|u, D^1)p(u|D^1)du \). A problem with updating in this way is that the possible values of \( u \) from the empirical distribution \( \hat{\pi}(u) \) cannot change. To overcome this problem we propose an alternative approach where the empirical distribution \( \hat{\pi}(u) \) is re-calculated at each update step by re-computing individual maximum likelihood estimates.
4.6 Why use the model?

On an account level, the random-effects model encapsulates behaviour into seven interpretable parameters. These may provide important insight into the behaviour of consumers, using the quantitative model to provide qualitative information in two ways.

Firstly, fitted models will inform decision-making mechanisms on how consumers behave under normal circumstances. For instance, someone with an account with a weekly profile shifted towards the weekend might reasonably be expected to be working between Monday and Friday. A propensity to make several withdrawals in quick succession might suggest that an account makes use of the mobile top-up option before or after withdrawing cash, and that the consumer appears willing to make multiple uses of ATM machines. These sorts of qualitative descriptions can be used to drive population segmentation and hence to guide customer-management strategies.

Secondly, predictions about the temporal process of ATM withdrawals for individual accounts may be used to guide decisions about account interventions that a bank may wish to make. That is, the model might be used within a decision-making framework, where rules based on the model guide the times and kinds of interventions. The following areas might benefit from such predictions:

- Attrition: the predicted number of withdrawals might be used to determine when and what type of intervention might be made in order to retain customer loyalty.
- Customer management: behaviour might change towards a more favourable profile, and a bank might wish to change its treatment of the consumer based upon a customer-value management strategy.
- Marketing I: predictions might be used to filter individuals who are more likely to respond to some forms of marketing than others.
- Marketing II: if customer’s predicted behaviour appears to be significantly different from previously then circumstances might have changed in the customer’s life; for example, employment status might have changed. A number of interventions might be targeted to such an account, e.g. mortgage marketing.
- Risk management: when a customer’s predicted behaviour appears significantly different from that previously observed, appropriate risk-aversion interventions might be made; these might relate to fraud.

The models could be used to trigger interventions by monitoring behaviour and choosing actions depending on whether the account behaves as predicted, or not. In the case when variation in behaviour is within the sort of range which the model expects, predictions could be used to form groups of customers who may benefit from targeted actions. For example, one rule used by fraud detection systems is to flag withdrawals made in quick-succession. This rule might only be applied to accounts the model would not expect to behave in such a manner. In the case when an account behaves differently from expected, the change itself could act as a trigger for intervention. For example, if the overall rate of cash withdrawals has decreased more than we might expect from the model, then the cause might be investigated in order to reduce risk of attrition.
The model may also be used as a description of the population of accounts that a bank holds in two broad ways.

- As a snapshot of current behaviour.
  Distributions of the profile-usage across weeks, or between-weeks might be grouped and Customer Relationship Management (CRM) techniques used to best serve the different segments. Figures 4 shows 25 fitted weekly profiles. The plots show considerable variation between individuals, and a range of different behavioural profiles.

- Examining changing behavioural profiles through time.
  At a population level this change may be related to macro-economic or other factors. The structure of the model should be unaffected because the change will be taken into account in the random-effects distributions. Thus, the model may be used to quantify the changes in consumer ATM withdrawal behaviour.

5 Demonstration

To demonstrate the utility the statistical model may bring, we next describe how the methods of estimation and prediction have been used on the five hundred accounts with the largest number of withdrawals in our random sample. Only five hundred have been used due to difficulties in fitting to accounts with the number of observations near to, or less than the number of parameters in the model. Estimation is carried out using the first three-months data, and prediction is made for the last month.

5.1 Estimation

As outlined in Section 4.3, our estimation methodology is to fit each individual account separately using the likelihood equation (2), and to use the empirical distribution of the many fits as the random-effects distribution. The likelihood equation requires numerical evaluation of an integral of the rate function and we have used a trapezoid rule. The likelihood surface may be explored using any unconstrained optimisation routine to estimate parameters. We chose a Nelder-Mead Simplex algorithm (a routine called “nmsmax” in GNU Octave, available from http://octave.sourceforge.net/).

Fits from the estimation procedures showed a range of patterns of behaviour. Figure 4 shows 25 time-of-day profiles. In this selection some are more likely to make ATM withdrawals in the afternoon, others the evening, and for some their rate is spread out fairly evenly throughout the entire day. It is also informative to examine the fitted self-exciting component terms in Figure 5. Each point represents a fitted account. Those to the right of zero on the x-axis have their rate function boosted following a transaction, those to the left are deflated. The speed at which the self-exciting behaviour dies down is also variable. For the accounts at the top of the y-axis the excitation lasts a short time, while for others towards the bottom the rate is excited for longer periods following a withdrawal. For example, a value of 4 on the y-axis would scale an excitation \( \hat{u}_6 \) on the x-axis by around 0.96 one minute after the last transaction, 0.01 after one hour, and zero after one day. A value of -2 on the y-axis would scale the
Figure 4: Some within-day Beta weight functions from 25 individual fits

Figure 5: Self-exciting parameters from 500 individual fits. Only absolute values of $\hat{u}_6$ (x-axis) less than 10 have been plotted. Of those not on this plot, 125 were negative to a minimum $-2.637 \times 10^{307}$, and 3 were positive to a maximum 27.355.
Table 1: Comparison between observed number of zero withdrawals in the next one to four weeks and predicted Expectation and Standard Deviations (SD) from the point process model, and a Poisson model.

<table>
<thead>
<tr>
<th></th>
<th>Week 1</th>
<th></th>
<th>Week 2</th>
<th></th>
<th>Week 3</th>
<th></th>
<th>Week 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>(a) Accounts with $\hat{u}_6 &gt; 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>54</td>
<td>31</td>
<td>23</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>48 (5.2)</td>
<td>23 (3.9)</td>
<td>13 (3.1)</td>
<td>8 (2.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson</td>
<td>35 (4.9)</td>
<td>11 (3.1)</td>
<td>4 (1.9)</td>
<td>1 (1.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) Accounts with $\hat{u}_6 \leq 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>65</td>
<td>31</td>
<td>22</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>47 (6.0)</td>
<td>11 (3.2)</td>
<td>3 (1.8)</td>
<td>1 (1.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson</td>
<td>69 (7.0)</td>
<td>20 (4.2)</td>
<td>7 (2.5)</td>
<td>2 (1.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

excitation by close to to 0.99 one hour following the last transaction, 0.87 after one day and 0.39 after one week. The different types of behaviour observed in the plot are compatible: some people are likely to make another withdrawal following a first, but others are less likely.

5.2 Prediction

The estimates found above are put to use in prediction as an empirical distribution of the random effects. Here we use the theory from Section 4.4 to predict the probability of zero withdrawals in the next one to four weeks, and monitor individual account-behaviour using the cumulative rate function.

For each individual we have calculated the probability of zero withdrawals in the next four weeks. This is compared to the observed number to examine the predictive performance of the model in Table 1. It shows the overall expected number of accounts with zero withdrawals based upon the model, and those observed, for (a) the 148 accounts with positive self-exciting coefficient estimates $\hat{u}_6$ and (b) the 352 accounts with negative $\hat{u}_6$. The expectation and variance are obtained by assuming independence by account. For the first group, relative to a Poisson model (fitted to each account and projected forward), the model prediction is closer to the observed number of zero-withdrawing accounts. This is because the fitted rate will be less than a Poisson model unless an event has occurred recently. For the second group, the Poisson prediction is closer than the model to the number observed. This might be due to over fitting: perhaps we ought to constrain $u_6 \geq 0$. However, we might also expect the model to constantly under-predict the total number of accounts that make no withdrawals because there will be some who stop using their accounts. That is, the assumed model of a constant rate through time scaled for the calendar and past events is inappropriate for those who have closed their account, gone on holiday and stopped withdrawing cash or undergone other behavioural state changes.

In Section 4.4 the cumulative rate function was outlined as a way to monitor behaviour and help determine whether a behavioural-state change has occurred.
Figure 6: Monitoring behaviour using the cumulative rate function
We demonstrate this by choosing a single account. Figure 6 shows some plots for our selected account. The top left plot shows cumulative rate functions during the estimation time period. Visually, these seem close. The top-right plot compares the predicted and observed cumulative rate functions. The two diverge and the Martingale residual plot in the bottom right suggests that this is probably a significant change. In the long run we would expect the distribution of standardised residuals to be Normally distributed with mean zero and variance one (from the Martingale Central Limit theorem). The plot of the standardised residual shows departure from this limit, leading one to suspect a change in behaviour over this period. Note how a Kolmogorov-Smirnov goodness-of-fit test using transformed times, shown in the bottom left corner, does not detect this departure (the 95% significance points plotted are based on a simulation with 10,000 repeats). From this analysis we might conclude that behaviour was different in the fitting and prediction periods, and some behavioural-state change occurred. Reasons for this would be speculative, but they might be linked to a previous burst, around the 30 to 50 day mark.

6 Conclusion

This paper has considered the development of a point process model for ATM withdrawal times. We see it as an example of the sort of phenomenological statistical model that may be developed in the personal finance sector. The main reason for such an approach is to overcome problems that can occur with purely empirical models, where performance can degrade as correlations in the data change over time.

Exploratory analysis on a random sample of 1,000 accounts with a UK highstreet bank was used to identify some features of ATM withdrawals. A point-process model for ATM withdrawal times was chosen based upon the observed features and consideration of a number of modelling approaches. Estimation, prediction and computational issues were discussed, and an unusual approach to prediction was proposed, for reasons given. Goodness-of-fit and prediction tests made on the model seemed to indicate it to be a satisfactory representation of reality. The model may be used to predict behaviour for an individual, assess when state-changes in individual behaviour have occurred, and as a description of behaviour for a portfolio of accounts.

The work in this paper may be extended by using other data that are available to banks. Covariates might be included in the form of the rate function to improve predictive performance, and other aspects of behaviour could be jointly modelled. For example, a marked point process model could be constructed using withdrawal-amount data, and a spatial-temporal model developed using the postcode location of ATMs. The model for the temporal pattern of ATM withdrawals might serve as a starting point for both.

Acknowledgements

This work is supported by EPSRC grant number EP/D505380/1 awarded to the Quantitative Financial Risk Management Centre at Imperial College London.
References


