Statistical Modelling 2010; **10**(**2**): 197–214

Predicting the amount individuals withdraw at cash machines using a random effects multinomial model

Adam R Brentnall[1], Martin J Crowder[2] and David J Hand[1,2]
[1]Queen Mary University of London, Barts and The London School of Medicine and Dentistry, Wolfson Institute of Preventive Medicine, Charterhouse Square, London
[2]Department of Mathematics, Imperial College London, London

Abstract: Retail finance organizations use data on past behaviour to make predictions for customer value management strategies. Random-effects models, where each individual has a behavioural pattern drawn from an overall population distribution, are a natural statistical form in this context. The random effects models in this paper are used to predict how much individuals withdraw at a single cash machine visit. A multinomial distribution is taken for the distribution of amounts and the random effects are modelled by a Dirichlet distribution or the empirical distribution of individual maximum likelihood fits. A third model extends the multinomial distribution by incorporating a form of serial dependence and uses an empirical distribution for the random effects. Several prediction tests on a sample of 5000 UK high-street bank accounts find that the greatest benefit from the models is for accounts with a small number of past transactions; that little information may be lost by binning and that the Dirichlet distribution might overestimate the probability of previously unobserved withdrawal amounts. The empirical distribution of random effects is found to perform well because there are a large number of individual accounts.

Key words: ATM; cash withdrawal; customer; personal banking; prediction; random effects; retail banking

Received November 2007; **revised** May 2008; **accepted** September 2008

1 Introduction

Retail finance organizations collect large amounts of information on their customers in streams of data. These data are often used to help manage risk by developing statistical models to predict future behaviour. This paper is concerned with the prediction of the amounts that individuals withdraw from ATMs (automated teller machines). The predictions may be of use within a decision-making structure, perhaps for customer value management, to trigger appropriate risk management actions or to serve as a basis for ATM replenishment strategies.We show that predictions

Address for correspondence: Dr Adam R. Brentnall, Institute for Mathematical Sciences, Imperial College London, 53 Prince's Gate, South Kensington, London. SW7 2PG, UK. E-mail: a.brentnall@qmul.ac.uk

 10.1177/1471082X0801000205

based solely upon an individual's past behaviour can be improved upon by using information about the population, through the use of a random-effects model.

There do not appear to be statistical models of this form for ATM withdrawal amounts in the literature. The most relevant statistical analysis is for credit card transactions in Hand and Blunt (2001). In the economics literature, there has been some interest in understanding the relationship between ATMs and demand for cash. Amromin and Chakravorti (2007) conducted a linear regression panel data analysis to compare countries, paying particular attention to the growth in debit card point-of-sale transactions. Boeschoten (1998) and Snellman and Viren (2006) had similar goals but developed their conclusions using models of individual customer behaviour. Boeschoten (1998) analysed data from a survey of Dutch consumers between 1990 and 1994 with a deterministic inventory model where individuals were assumed to maintain a level of cash, used at a constant rate, replenishing their stock when it falls below a certain threshold. They found that ATM users typically had a lower stock than non-users because their cost of obtaining cash was lower. The relationship between the cost of obtaining cash and the number of ATMs was modelled in Snellman and Viren (2006) as a deterministic optimization problem, where costs were assumed to be proportional to distance from ATMs. Although these economic papers used models for customer withdrawal amounts and times, they are simplifications to enable conclusions about macro demand for cash and can be improved upon when making predictions at an individual level.

A sample of 5000 accounts making ATM withdrawals is described in Section 2 and four aspects of individual account withdrawal amount patterns are examined: (i) the number of withdrawals, (ii) their distribution, (iii) serial dependence and (iv) temporal effects. The analysis is developed in Section 3, where three versions of a random effects model for individual account withdrawal amounts are outlined and methods for estimation and prediction are described. The first two models assume that withdrawal amounts are independent of their history and use a multinomial distribution. The third model also bins the data but incorporates a form of dependence where withdrawals made in quick succession may be for similar amounts. The distribution of random effects is modelled as a Dirichlet distribution in the first model; in the others, it is left unspecified and approximated by an empirical distribution. The different models are applied to prediction examples in Section 4.

2 Data

The 5000 accounts examined in this report are customers with a UK high-street bank and were selected at random from a list of all those that withdrew cash from a cash machine, topped up a mobile phone using an ATM, or received 'cash-back' over a counter during a 4-month period in 2005. The data recorded the time and amount of each withdrawal, but the different types of withdrawal cannot be distinguished. Brentnall *et al.* (2008) discussed a statistical model for the timings of withdrawals, and in this paper we focus on the amounts.

The notation used for these data is as follows. The jth withdrawal amount from individual $i = 1, \ldots, n$ is represented as w_{ij}, where $j = 1, \ldots, n_i$. Denote the set of distinct withdrawal amounts in the complete data as $(a_1, \ldots, a_M)$, where M is the number of different withdrawal amounts. Then if the same number $m \leq M$ of bins are chosen for each account, the withdrawal-amounts recorded for account i can be represented as $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im})$, where y_{ij} is the number of amounts falling in the jth bin. An alternative for the withdrawal amount record of account i to time t is $z_i(t) = \{z_{i1}, \ldots, z_{in_i(t)}\}$, where $z_{ik} \in (1, \ldots, m)$, m is the number of bins chosen for the distribution of amounts and $n_i(t)$ is the number of withdrawals made to time t. We denote by $\boldsymbol{x}_i(t) = \{x_{i1}, \ldots, x_{in_i(t)}\}$ the withdrawal times.

The models are based on an initial data analysis, with four summary findings:

1. There is variation in the number of withdrawals made by individual accounts.
2. Individual withdrawal amount distributions often have a low preferred amount of £10–£50 but a long tail into hundreds of pounds.
3. For most accounts, withdrawal amounts seem to be independent of their history but some appear to show serial dependence, where withdrawals made a relatively short time apart are of similar amounts.
4. It is inconclusive as to whether there are temporal effects—if they exist, then they are not very strong and may not add very much to the predictive power of a model.

The first two points are likely to provide lift to prediction from the use of random effects, since those with only a few withdrawals benefit from information about the population, and future events are likely to regress to the population mean because there is variation between individual withdrawal distributions. The findings are next reported in more detail.

2.1 Number and amount of transactions

The distribution of the number of transactions made by individual accounts is shown in Figure 1. There is substantial variability with the median making 16 withdrawals over the 4 months, or about once a week.

Cash withdrawals were usually in multiples of £10 or, less commonly, £5. This is unsurprising since ATMs often dispense cash in these multiples, up to a maximum limit (different for different accounts) into hundreds of pounds. However, some amounts withdrawn are not bank-note multiples. These are due to ATM charges or withdrawals abroad where the amount does not translate into an integer Sterling amount. In the sample, around 8% of withdrawals were non-integer amounts, with a quarter of these made outside the UK. Figure 2 shows a histogram of all amounts withdrawn in the sample. They range from £2.88 to £600.00. The distribution has a low mode and long tail and some bins contain markedly higher frequencies than their neighbours. For example, £300 is preferred to £290 or £310. This type of behaviour

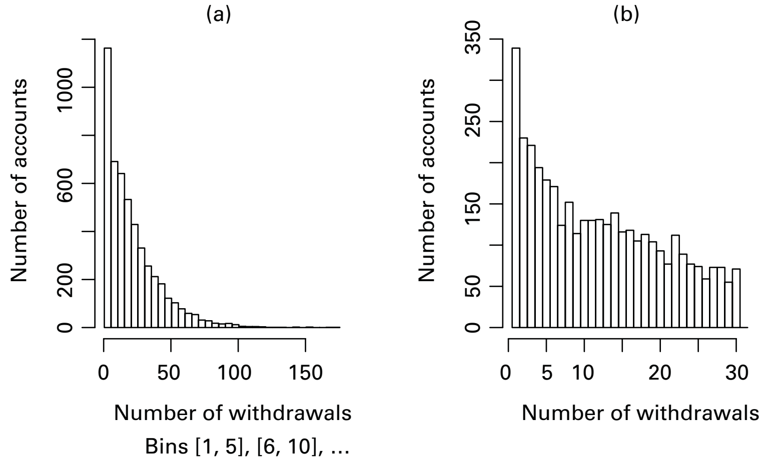

Figure 1 Histograms of the number of withdrawals made by accounts over the 4-month period for (a) all accounts and (b) those with fewer than 31 withdrawals

has also been observed in credit card data (Hand and Blunt, 2001). In the present case, people have a tendency to favour amounts like £50, £100 and £200 because ATMs invite you to take such amounts.

A variety of withdrawal amount distributions are observed by account. Some only withdraw relatively small amounts, others only large amounts; the rest fall in between, often making withdrawals of £10–£50, occasionally with some much

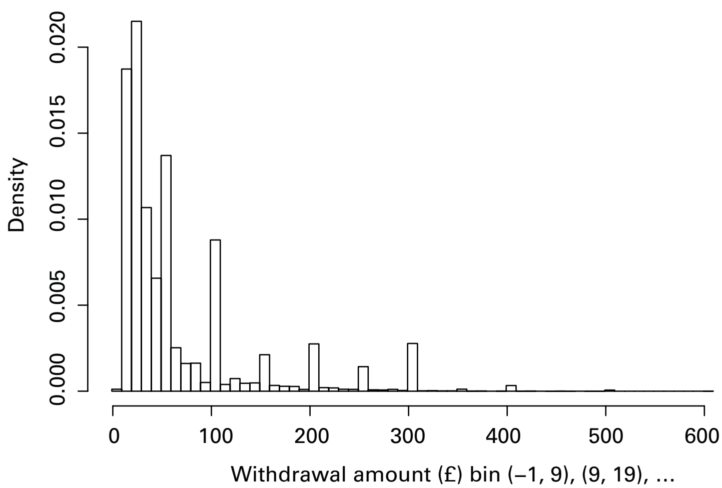

Figure 2 Histogram of amounts withdrawn

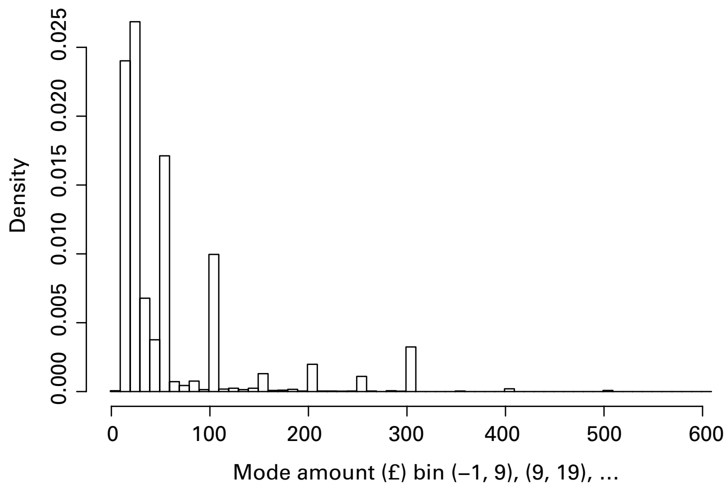

Figure 3 Histogram of individual account modes

larger withdrawal amounts. Figure 3 shows the distribution of individual account modes: most have a relatively low preferred amount. Half the sample take their preferred amount between 29% and 55% of the time, but an additional 9% always take the same amount. A large number of accounts have a long tail: the difference between an account's preferred amount and their maximum withdrawal varied between £0 and £550, with half the sample between £31 and £130.

2.2 Serial dependence

Graphical exploratory analysis appeared to show a relationship where withdrawals a short time apart were for similar amounts. For instance, one pattern could be a fairly regular pattern of withdrawals between £10 and £50, then two withdrawals minutes apart for amounts into the hundreds of pounds; another could be two transactions within a few hours for £20.

Table 1 provides some evidence of this. It shows that withdrawals less than 1 hour apart are more likely to be for similar amounts than those made more than 1 hour apart. The table also highlights how the difference is not linked to the withdrawal amount itself: smaller and larger amounts are both more likely to follow an hour after a previous withdrawal.

The sort of dependence can be thought of as a sticky key on the ATM. If the individuals return to an ATM shortly after a transaction, their finger is attracted by the same button. If they return later on, the glue has dried up and their withdrawal amount is independent of the history. In Section 4, we compare predictive performance for models which allow for, or ignore, this type of serial dependence.

Table 1 Withdrawals made within 1 hour of a previous one are more likely to be of a similar amount. The table is based on all withdrawal pairs over the 4-month period

Range (£)		% same bin		Number same bin	
(	]	>1 hour	≤1 hour	> 1 hour	≤1 hour
0	13	39.7	68.9	7211	598
13	23	32.2	47.1	6698	476
23	33	18.7	41.0	1940	191
33	43	13.9	36.6	886	118
43	53	30.1	42.0	3976	255
53	73	11.2	33.2	449	78
73	123	26.8	42.7	3048	306
123	173	12.9	36.8	434	100
173	223	16.6	32.4	535	110
223	273	12.6	30.3	197	53
273	∞	32.4	58.0	863	536
Total		27.6	47.5	26237	2821

2.3 Temporal effects

Here we consider whether there are temporal effects for withdrawal amounts. Plots of individual marked point processes do not show any obvious patterns, and neither do histograms of withdrawal amounts by account, split by time of day, day of week, week or month. We can look for trends using more formal statistical tests as follows. We split withdrawals in each account into temporal groups such as day of week, and then test whether there are differences in the withdrawal distribution between the groups. Since the distribution of withdrawal amounts is often in multiples of £10, with features such as inflated multiples of £50, and the number of observations in each group may be small, we use the Kruskal–Wallis non-parametric test (Rice, 1994) for ranked withdrawal amounts. Our null hypothesis is of no difference between the groups. Each account is tested for a difference in group medians and a p-value is calculated. The test is one sided and we might determine small p-values as individually significant if they fall below a certain level γ. In the long run, if there is no difference, we would expect the proportion of accounts with p-values below γ to equal γ. Table 2 shows the relation between significance levels and the proportion of accounts which exceed them. Except for the week grouping, the observed proportions look somewhat higher than γ, and so there is some evidence of temporal effects. However, for those such as the AM/PM group with around twice the expected proportions, it can be shown that the predictive performance of a model may increase for half of those flagged at an α level chosen, but decrease for the other half. That is, if there is a predictive gain to be had, then it is likely to be small, and it is at the expense of increased model complexity for often, sparse data. We therefore do not consider temporal effects in the rest of this paper.

Table 2 Proportion of individual p-values below significance level

	Significance level		
Group	0.01	0.05	0.10
AM/PM	0.03	0.09	0.17
Month	0.02	0.09	0.16
Weekday	0.04	0.10	0.17
Week	0.01	0.04	0.08

3 Models and prediction

In this section, we describe random-effects models for individual account withdrawal amounts. Random effects have been used in other areas of retail banking, including credit scoring (Leonard, 1988, 1993) and portfolio credit risk (Rösch and Scheule, 2003; McNeil and Wendin, 2007).

We first consider how to model withdrawal amount using a parametric model. The description of withdrawal amount distributions in Section 2.1 suggests that most withdrawals for many accounts are in multiples of £10, but some have non-integer values. The distribution may have bumps at multiples of £50, a high-density mode, and be right skewed with a long tail. Further, transaction amounts may be a combination of mobile phone top-ups, cash machine withdrawals and supermarket cash-back transactions. A first approach might be to use a transformation, such as $\log(w_{ij}+1)$ that was suggested by Whittaker *et al.* (2005) for other retail financial data. This does not overcome bunching of values, but might be used in a mixture model such as that used by Hand and Blunt (2001) for an analogous situation. However, we failed to find a simple model whose form was appropriate for all accounts. For this reason, we next outline a multinomial model for amounts, where a suitable choice of bins for individual withdrawal amounts is made.

3.1 Model 1

The distribution of $\boldsymbol{y}_i$, as defined in Section 2, is taken to be multinomial, thus assuming independence of withdrawal amounts. The multinomial probability vector for $\boldsymbol{y}_i$ is denoted by $\boldsymbol{q}_i = (q_{i1}, \ldots, q_{im})$, and the $\boldsymbol{q}_i$ are random effects with some distribution over the population.

In model 1, we take the $\boldsymbol{q}$-distribution as a Dirichlet distribution with parameter set $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$, and the density of $\boldsymbol{q}_i$ is (Evans *et al.*, 2000)

$$p(\boldsymbol{q}_i; \boldsymbol{\alpha}) = B_m(\boldsymbol{\alpha})^{-1} \prod_{j=1}^{m} q_{ij}^{\alpha_j - 1}, \qquad (3.1)$$

where $B_m(\boldsymbol{\alpha}) = \prod_{j=1}^{m} \Gamma(\alpha_j)/\Gamma(\boldsymbol{\alpha}_+)$, with $\boldsymbol{\alpha}_+ = \sum_{j=1}^{m} \alpha_j$. Then we need to maximize the overall likelihood

$$p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n; \boldsymbol{\alpha}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i; \boldsymbol{\alpha})$$
$$= \prod_{i=1}^{n} \int p(\boldsymbol{y}_i|\boldsymbol{q}_i) p(\boldsymbol{q}_i; \boldsymbol{\alpha}) \boldsymbol{d}\boldsymbol{q}_i, \tag{3.2}$$

where

$$p(\boldsymbol{y}_i|\boldsymbol{q}_i) = \prod_{j=1}^{m} q_{ij}^{y_{ij}}.$$

The integral in (3.2) is $B_m(\boldsymbol{\alpha} + \boldsymbol{y}_i)/B_m(\boldsymbol{\alpha})$. Any suitable optimization routine, such as a quasi-Newton or Nelder–Mead simplex algorithm (see, e.g., Lawless, 1982 Appendix F), might be used to estimate $\boldsymbol{\alpha}$.

3.2 Model 2

In model 2, we again take the distribution of $\boldsymbol{y}_i$ to be multinomial, but now leave the $\boldsymbol{q}$-distribution unspecified. It is estimated using an empirical distribution $(\hat{\boldsymbol{q}}_1, \hat{\boldsymbol{q}}_2, \ldots, \hat{\boldsymbol{q}}_n)$, formed from maximum likelihood fits to each of the n accounts; $\hat{\boldsymbol{q}}_i$ is then just the vector of sample proportions with bin j probabilities estimated as $y_{ij}/\sum_{j=1}^{m} y_{ij}$.

3.3 Model 3

The final model extends the multinomial model for withdrawal amounts by incorporating a form of dependence observed in Section 2.2. For $j = 1, \ldots, m$ a model for the probability of the amount bin $b_i(t)$ chosen by an individual i at time $t > t_p$ is

$$P\{b_i(t) = j|\boldsymbol{u}_i, \boldsymbol{v}_i, b_p, t_p\} = \frac{u_{ij} + I(j = b_p)\exp\{v_{i1} - v_{i2}(t - t_p)\}}{\boldsymbol{u}_{i+} + \exp\{v_{i1} - v_{i2}(t - t_p)\}}, \tag{3.3}$$

where $I(.)$ is the indicator function, b_p is the previous withdrawal bin, t_p is the previous withdrawal time, $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{im})$ and $\boldsymbol{u}_{i+} = \sum_{j=1}^{m} u_{ij}$. The v_{i1} parameter is unconstrained, but v_{i2} is greater than or equal to zero. The model allows for a form of the 'sticky-button' behaviour described in Section 2.2, where withdrawals within a short time period may be in the same bin. It may be interpreted in the following way. Conditional upon a withdrawal occurring at time t, the probability of it being from each bin is linked to a set of parameters $\boldsymbol{u}_i$. If $\exp(v_{i1})$ is zero, then it is equivalent to a multinomial model. The $\boldsymbol{v}_i$ terms have the effect of increasing the probability that the last withdrawal bin b_p is chosen again soon afterwards. The increased probability, determined by v_{i1}, decays exponentially following the previous transaction at t_p at a rate determined by v_{i2}. The other bin probabilities are proportionally adjusted so that they all sum to unity.

The likelihood function $p\{\boldsymbol{z}_i(t)|\boldsymbol{u}_i, \boldsymbol{v}_i, \boldsymbol{x}_i(t)\}$ for individual i, conditional upon the times of withdrawal $\boldsymbol{x}_i(t)$ with $n_i(t) > 1$, is

$$u_{iz_{i1}} \prod_{j=2}^{n_i(t)} P\{b(x_{ij}) = z_{ij}|\boldsymbol{u}_i, \boldsymbol{v}_i, z_{ij-1}, x_{ij-1}\}. \tag{3.4}$$

This can be explored using a general search method, such as a Nelder–Mead simplex algorithm, to obtain maximum likelihood parameters $\hat{\boldsymbol{r}}_i = (\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{v}}_i)$ for each account i. We leave the distribution of these random effects unspecified and approximate it by $(\hat{\boldsymbol{r}}_1, \hat{\boldsymbol{r}}_2, \ldots, \hat{\boldsymbol{r}}_n)$.

3.4 Random effects prediction

Our main reason for developing a model is to estimate $p(\boldsymbol{s}_i|\boldsymbol{y}_i)$, the distribution of the random effect $\boldsymbol{s}_i$ for each account i, in order to predict its future behaviour. For the multinomial Dirichlet model 1, we have $\boldsymbol{s}_i = \boldsymbol{q}_i$ and first compute $\hat{\boldsymbol{\alpha}}$. Since the Dirichlet distribution is conjugate to the multinomial distribution, the posterior $p(\boldsymbol{q}_i|\boldsymbol{y}_i; \hat{\boldsymbol{\alpha}})$ is also a Dirichlet distribution with parameters $(\hat{\alpha}_1 + y_{i1}, \ldots, \hat{\alpha}_m + y_{im})$. When the random effects distribution is unspecified, as in models 2 and 3, we use the empirical distribution $(\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, \ldots, \hat{\boldsymbol{s}}_n)$ (where $\boldsymbol{s}_i$ is equal to $\boldsymbol{q}_i$ for model 2 or $\boldsymbol{r}_i$ for model 3) as an estimate of $p(\boldsymbol{s})$ to calculate $p(\boldsymbol{s}|\boldsymbol{y}_i)$ through Bayes formula:

$$p(\boldsymbol{s}_i|\boldsymbol{y}_i) \propto p(\boldsymbol{y}_i|\boldsymbol{s}_i)p(\boldsymbol{s}_i).$$

Now we may make a model-based prediction of any future random outcome B_i for account i, using the posterior Dirichlet distribution, or for the empirical random effects distribution:

$$p(B_i|\boldsymbol{y}_i) = \sum_{j=1}^{n} p(B_i|\hat{\boldsymbol{s}}_j)p(\hat{\boldsymbol{s}}_j|\boldsymbol{y}_i). \tag{3.5}$$

3.5 Alternative prediction methods

Some other methods that might be used to form predictions include individual multinomial models, empirical distribution functions (EDFs) and the overall mean.

Multinomial projections use the maximum likelihood fit $\hat{\boldsymbol{q}}_i = \boldsymbol{y}_i/n_i$ to each individual only.

Predictions about withdrawal amounts W using individual EDFs are based on

$$Pr(W < a) = 1/n_i \sum_{j=1}^{n_i} I(w_{ij} < a).$$

The overall mean pools together all the data. So if the data are binned, then the overall probability $\hat{q}_j$ of the jth bin is $(1/\sum_{i=1}^{n} n_i)\sum_{i=1}^{n} y_{ij}$ and $\hat{\boldsymbol{q}}_i = \hat{\boldsymbol{q}}$ for all i.

4 Model testing

In this section, four prediction tests are made on the models in order to gain insight into their relative performance. The fitting period is 3 months, and predictions are compared to the first withdrawal in the last month. Two choices are made for model bins in UK£ units. In the first, the break points are (19, 29, 99), i.e., there are four bins (0, 19], (19, 29], (29, 99], (99, ∞); in the second there are 11 bins with break points (13, 23, 33, 43, 53, 73, 123, 173, 223, 273). A third bin structure is used for model 2, with one bin for each distinct withdrawal amount in the sample, and this is denoted as model 2b.

4.1 Predicting a binned range

The first test uses four bins to predict the probability that each individual account withdraws in the (£19, £29] range. The baseline performance measure is the proportion of observed withdrawals in the bin. This is a reasonable baseline to use because the bin is common for many accounts: most accounts have a low mode, and £20 was withdrawn around 22% of the time overall as shown in Figure 2. Before presenting the prediction results, two aspects of the maximum likelihood fits are reported.

The first observation concerns the degree to which the serial dependence in model 3 occurs in the population. Some accounts are found to have a large boost in probability relative to the proportion of time the £20 bin was observed (i.e., multinomial fit), some small and some have none: there were 41% whose probability was more than their multinomial fit by less than 0.01, 1 minute after a withdrawal, but 21% whose fitted probability of withdrawing from the £20 bin was more than 0.99, 1 minute after, compared to just one person with such a probability 1 month after. This suggests that the sort of dependence suggested by the data is significant for some, but not common to the entire population.

The second observation concerns the Dirichlet distribution used by model 1. Figure 4 compares the marginal random effects distribution estimates for the (£19, £29) bin between models 1 and 2. The mismatch at the lower end is because some individuals do not make any withdrawals in a given bin over the time period considered. When the total number of withdrawals by an individual is low, the number of empirical distribution zeros may be more than appropriate, and one might

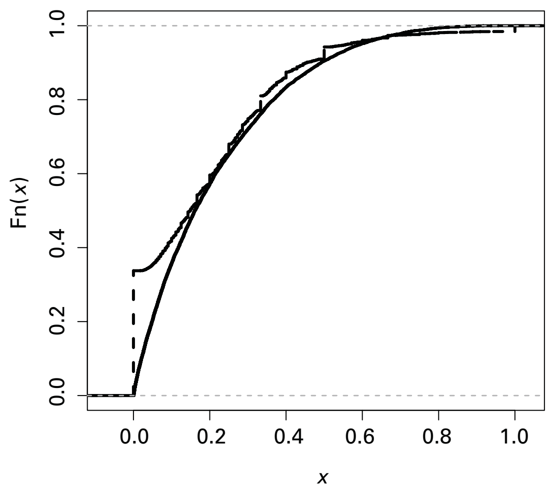

Figure 4 Comparison between marginal Dirichlet model distribution function (smooth curve) and EDF. For each account, x is the proportion of withdrawal amounts in bin 2 ($x = \hat{q}_{i2}$).

argue that model 1 appropriately smooths this out. For example, if only two withdrawals have been made, at least two out of the four bins are empty; indeed, 1100 accounts out of the 5000 made fewer than four withdrawals in the time period. Alternatively, the disparity might be a goodness-of-fit test failure for model 1: perhaps fewer accounts are likely to withdraw from certain bins than it can allow for. The effect on prediction is shown in Figure 5. Here, model 1 shrinks predictions more

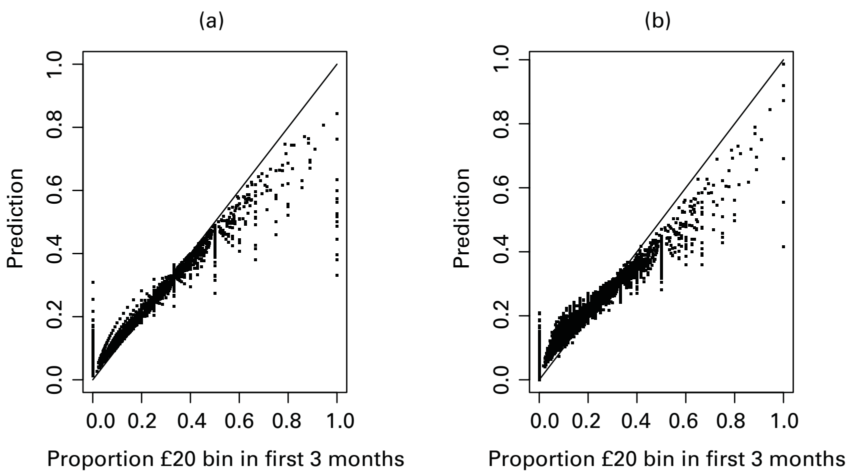

Figure 5 Shrinkage in (a) model 1 (Dirichlet) and (b) model 2 (empirical) multinomial random effects model prediction

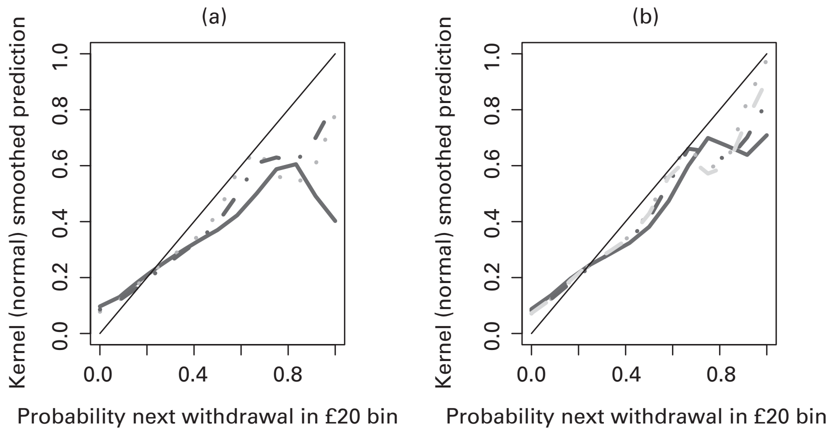

Figure 6 Predictive performance of a multinomial projection (solid line), model 1 (dashed-dotted line), model 2 (dotted line) and model 3 (dashed line) for those with at least one withdrawal in the last month and the number in the first month at least (a) one withdrawal (3778 accounts) and (b) nine withdrawals (2753 accounts)

for those who only withdrew amounts in the £20 bin during the first 3 months (the far right side of the plot), or made no withdrawals in the £20 bin (the far left side of the plot). That is, the accounts that only made withdrawals in the £20 bin are predicted lower probabilities under model 1; and the accounts that did not make any withdrawals in the £20 bin are predicted higher probabilities under model 1.

It is worth noting that the shrinkage observed in Figure 5 is the main reason for applying the random-effects models. The prediction equation (3.5) means that, e.g., when the proportion of the £20 bin in the first 3 months is 1.0, the associated prediction is not automatically 1.0. The different levels of shrinkage relate to the number of withdrawals made by the individual, and the population has a greater effect on those with fewer withdrawals in the first 3 months.

The predictive performance is assessed in Figure 6. This shows (normal) kernel-smoothed lines (with bandwidth $h = 1/12$) from regressing the binary outcomes against predictions. This approach to viewing performance of such a situation is recommended by Copas (1983). Two charts are presented because model 3 requires more observations in order to be fitted, and only those accounts with more than eight withdrawals are used. The plots show that the predictions are ordered sensibly for all prediction methods. In Figure 6a, the random-effects models show slightly better performance for predictions between 0.4 and 0.6 to the proportion of previous withdrawals in the bin because their lines of fit are closer to the ideal 45° line, and 52% of the model 1 predictions are less than 0.2, 39% between 0.2 and 0.4 and 8% between 0.4 and 0.6. However, Figure 6b suggests that the improvement may not be significant as the number of withdrawals observed for each individual account increases. In this example, by the time it is possible to fit model 3, it may not be worth the effort.

4.2 Predicting binned distributions

The next test compares four- and 11-bin forecast distributions using the Brier (1950) and log scoring rules. The Brier score assigns a loss of $1/n \sum_{i=1}^{n} \sum_{j=1}^{m} (q_{ij} - c_{ij})^2$, where q_{ij} is the predicted j-bin probability for individual i, and $c_{ij} = 1$ if the next withdrawal from individual i was in bin j and 0 otherwise. The log score is $-1/n \sum_{i=1}^{n} \log q_{ij}$. Because they have a unique minimum for the true underlying data generating mechanism, we can use them to compare model performance. Roulston and Smith (2002) and Selten (1998) have different reasons for preferring the log or Brier score as a measure of performance, and some other scores are discussed in Gneiting and Raftery (2007).

Table 3 shows the results for the three models (partial shrinkage), a multinomial projection (no shrinkage) and the overall mean (full shrinkage). Differences are tested by using a non-parametric bootstrap, resampling the individual score differences 2000 times.

For the four-bin approach, the null hypothesis of no difference between Brier scores for the multinomial forecast and models 1 and 2 has respective single p-values of < 0.001 and 0.013, while that between model 1 and 2 has a p-value of 0.054. There is not enough evidence to reject the null hypothesis for other combinations. Similar results are found using the log score, suggesting that there is little benefit from the more complicated model 3, and that model 2 performs the best of all.

For the 11-bin approach under the Brier score, only the model 2 prediction is significantly better than the multinomial projection (p-value < 0.001); but model 1 has the minimum log score and here all the observed differences are significant with p-values < 0.001. The reason is related to the goodness-of-fit issue noted in Section 4.1. The log score penalizes a few near-zero prediction probabilities more than the Brier score, and model 1 avoids this by increasing the probabilities at bins with no previous transactions more than the other models. The small number of these predictions dominates the score.

The results suggest that model 2 may be a better choice than model 3 when interest is in predicting a binned distribution. However, if several bins are used and it is important not to assign very small probabilities to events, then model 1 may be preferred.

Table 3 Brier and log scores for (a) four-bin models with more than eight transactions and (b) 11-bin models with more than 12 transactions

	(a)		(b)	
Method	Brier	Log	Brier	Log
Model 1	0.6011	1.0972	0.7849	1.9240
Model 2	0.5993	1.0965	0.7810	2.0302
Model 3	0.6024	1.0977	0.7835	2.0589
Multinomial projection	0.6046	∞	0.7857	∞
Overall mean	0.7327	1.3248	0.8668	2.1732

Table 4 Percentage of predictions which fall outside upper γ confidence level for accounts with a transaction in the last month and (a) at least one transaction in the first 3 months (3778) and (b) more than 12 transactions in the first 3 months (2100)

		Level γ (%)				
	Model	10.0	5.0	2.5	1.0	0.1
(a)	EDF	10.8	8.3	7.5	7.2	7.1
	1	4.3	2.5	1.7	1.2	1.0
	2	8.9	5.3	3.8	2.7	1.7
	2b	9.1	5.8	4.0	3.3	2.4
(b)	EDF	9.3	5.7	4.2	3.7	3.6
	1	5.1	2.6	1.6	0.9	0.7
	2	9.0	5.3	4.0	3.0	1.7
	2b	9.0	5.5	4.0	3.4	2.8
	3	9.2	5.6	4.1	3.1	1.7

4.3 Predicting upper quantiles

Extreme percentile points are also of interest to banks. For example, define x_γ by $P(W > x_\gamma) = \gamma$ for some specified (small) probability γ, where W is the withdrawal amount; so x_γ represents an unusually large withdrawal amount. Then, predictions may be made based on an estimate of this probability and then used within a fraud detection strategy. A test of the random-effects model is to choose different values of γ and compare the proportion of accounts whose next withdrawal is beyond the upper γ percentile. If the model is accurate, the observed proportion should equal γ.

The results are presented in Table 4 for 11-bin models. When bins are used, the percentile points are taken as the first bin k with cumulative distribution function (i.e., $\sum_{j=1}^{k} P(q_{ij}|\mathbf{y}_i)$) greater than or equal to $1-\gamma$. In this comparison, model 2 performs noticeably better than use of individual EDFs, especially for those with not many transactions. There is not a clear difference between models 2, 2b and 3. Model 1 does worse than the others for the higher γ levels, but better for the lower ones.

The difference between model 1 and the others is again related to the goodness-of-fit issue discussed in Section 4.1. The mismatch at the higher levels of γ shows that the probabilities assigned to the bins with no previous withdrawal may be too high for model 1. However, it does better at the lower γ levels, and so may be more useful for predictions in this range.

4.4 Predicting the mean

The last test compares predicted mean withdrawal amounts using mean square error (MSE), where in models 1, 2 and 3 the mean is calculated by assigning the 11 bins

Table 5 MSE of predictions for accounts with a transaction in the last month and (a) at least one transaction in the first 3 months (3778) and (b) more than 12 transactions in the first 3 months (2100)

Method	(a)	(b)
Last withdrawal	7020	6946
Mean of last five	5060	4982
Overall Mean	4673	4364
Model 1	5040	4432
Model 2	4517	4323
Model 2b	4556	4358
Model 3	—	4298

their median values. The results are shown in Table 5. Differences, tested using a non-parametric bootstrap with 5000 resamples, reveal several aspects.

Firstly, a bin structure is no obstacle to predicting the mean when 11 bins are used: there is no significant difference between models 2, 3 and 2b.

Secondly, the number of transactions by account appears to have a bearing on predictive performance of the models. The differences in column (b) between models 2, 2b, 3 and the baseline of the mean from all previous withdrawals are not individually significant. The equivalent differences in column (a) are significant. It therefore appears that the largest gain in predictive performance occurs for those with a small number of recent transactions.

Thirdly, model 1 performs badly, and is significantly worse as a predictor than the mean of past transactions. This is another consequence of the goodness-of-fit issue discussed in the prediction examples earlier.

Overall, then, the results show that using all information on past behaviour is better, in an MSE sense, than only using the most recent five, and using information on everyone through models 2, 2b and 3 is better than just using information from the individual when they do not have a large number of past withdrawals.

5 Conclusion

In this paper, we considered the prediction of the individual withdrawal amounts using a sample of 5000 UK high-street bank accounts. Three random-effects models were developed using features suggested by an initial examination of the data. All the models assumed no strong, consistent temporal effects. Models 1 and 2 also made the simplifying assumption that withdrawal amounts are independent of their history, but a form of serial dependence was allowed in model 3. We found some evidence from data analysis to suggest that the model assumptions were reasonable for many individuals. This section concludes by reviewing the findings from the prediction tests and identifies some possible future avenues for research.

5.1 Findings

The relative performance of the models was examined, with several findings.

1. Prediction improvement is possible using random-effects models, but when there is a large amount of data for each customer, it may be sufficient to use simple summary statistics. This depends on the type of prediction, e.g., the proportion of times a withdrawal had been made in a common binned range is a good predictor for those with several transactions, but using an empirical distribution for upper quantile points is less robust.
2. Several of the tests showed that the Dirichlet distribution for the distribution of multinomial parameters (model 1) has different predictions to an empirical distribution (model 2). Specifically, model 1 gives higher probabilities to previously unobserved bin values.
3. Choice of model 1 or 2 may depend on the performance measure. Model 2 does better for an 11-bin distribution under a log scoring rule test, and for predictions of small upper quantile points. Model 1 is superior in Brier score tests for four and 11-bin distributions, it is more accurate across a range of percentile points, and it has a lower MSE when predicting the mean amount.
4. Model 3 may be less useful than model 2 because it cannot be applied to accounts with fewer withdrawals than parameters, and predictive performance is no better.
5. Little information is lost by binning in the cases considered: the predictive performance of 11-bin models was comparable with empirical distributions in Sections 4.3 and 4.4.

5.2 Extensions

The comparison between models 1 and 2 shows that the empirical random-effects approach can be useful. This may be because the data consist of a large number of accounts, here we used 5000, but we note that in general banks may have millions.

The main problem with model 1 is that the Dirichlet distribution overestimates the probability associated with previously observed withdrawal amounts. This arises because some individuals only take a small number of different amounts, and so one possible avenue of future work is to find a simple extension of the Dirichlet distribution. One such approach is suggested by the fact that 9% of accounts always withdraw the same amount. We considered a model where

$$p(\boldsymbol{q}_i; \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) = (1-\beta)p_1(\boldsymbol{q}_i; \boldsymbol{\alpha}) + \beta p_2(\boldsymbol{q}_i; \boldsymbol{\gamma}) \tag{5.1}$$

and

$$p_2(\boldsymbol{q}_i; \boldsymbol{\gamma}) = \begin{cases} \gamma_k, & \text{if } q_k = 1, \\ 0, & \text{otherwise,} \end{cases}$$

with $p_1(\boldsymbol{q}_i;\boldsymbol{\alpha})$ as equation (3.1), and parameters $0 < \beta < 1$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$ such that $\sum_{j=1}^{m} \gamma_j = 1$. That is, the model is a mixture between a proportion of customers β who always withdraw from the same bin with an overall distribution linked to $\boldsymbol{\gamma}$ and $1 - \beta$ who follow a Dirichlet distribution with parameter $\boldsymbol{\alpha}$.

The likelihood for an individual i is

$$p(\boldsymbol{y}_i|\boldsymbol{q}_i;\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) = (1 - \beta)B_m(\boldsymbol{\alpha} + \boldsymbol{y}_i)/B_m(\boldsymbol{\alpha}) + \beta g(\boldsymbol{y}_i;\boldsymbol{\gamma}),$$

where

$$g(\boldsymbol{y}_i;\boldsymbol{\gamma}) = I\left[\left\{\sum_{j=1}^{m} I(y_{ij} > 0)\right\} = 1\right]\sum_{j=1}^{m} \gamma_j I(y_{ij} > 0)$$

adds γ_k to the equation when bin k is the only one observed in $\boldsymbol{y}_i$. Then the overall likelihood is

$$p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n;\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i;\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}).$$

Using the same bins as in Section 4, the fits to the first 3 months using a Nelder–Mead simplex algorithm estimated $\hat{\beta}$ to be 0.037 with four bins, and to be 0.011 under 11 bins. These are small adjustments and so predictive performance is likely to be very similar to model 1. For example, the mismatch between the marginal Dirichlet distribution and empirical distribution shown at the lower end of Figure 4 is improved slightly by adding approximately 0.037 to the origin of the Dirichlet curve. The small difference is probably because those who only withdraw one amount are not the only reason for the inflated number of zeros observed in bins relative to a Dirichlet distribution, e.g. some only withdraw two or three different amounts.

This analysis highlights a major benefit of using model 2: there is no requirement to explore possible parametric forms for the random effects distribution. Given this, future work is likely to focus on extensions to model 2. In particular, if improvements are most for those with fewer withdrawals, then it is undesirable to only predict for those with more observations than parameters, as in model 3 when explicit maximum likelihood estimates were not available. A better approach might be to predict individuals with a small number of withdrawals by using the empirical random effects distribution from those that can be fitted. Future work might investigate this possibility.

Acknowledgements

We thank an associate editor and two anonymous referees for helping to improve this paper through their comments and suggestions. The first author's work was supported by EPSRC grant number EP/D505380/1 while at Imperial College London, and the work of the third author was partially supported by a Royal Society Wolfson Research Merit Award.

References

Amromin E and Chakravorti S (2007) Debit card and cash usage: a cross-country analysis. Technical report, Federal Reserve Bank of Chicago.

Boeschoten WC (1998) Cash management, payment patterns and the demand for money. *De Economist*, **146**, 117–42.

Brentnall AR, Crowder, MJ and Hand DJ (2008) A statistical model for the temporal pattern of individual automated teller machine withdrawals. *Applied Statistics*, **57**, 43–59.

Brier GW (1950) Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**, 1–3.

Copas JB (1983) Plotting p against x. *Applied Statistics*, **32**, 25–31.

Evans M, Hastings N and Peacock B (2000) *Statistical distributions*. New York: Wiley-Interscience.

Gneiting T and Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–78.

Hand DJ and Blunt G (2001) Prospecting for gems in credit card data. *IMA Journal of Management Mathematics*, **12**, 173–200.

Lawless JF (1982) *Statistical models and methods for lifetime data*. New York: John Wiley & Sons.

Leonard KJ (1988) *Credit scoring via linear logistic models with random parameters*. PhD thesis, Concordia University.

Leonard KJ (1993) Empirical Bayes analysis of the commercial loan evaluation process. *Statistics & Probability Letters*, **18**, 289–96.

McNeil AJ and Wendin JP (2007) Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance*, **14**, 131–49.

Rice JA (1994) *Mathematical statistics and data analysis*. Belmont, CA: Duxbury Press.

Rösch D and Scheule H (2003) Forecasting retail portfolio credit risk. *Journal of Risk Finance*, **5**, 16–32.

Roulston MS and Smith LA (2002) Evaluating probabilistic forecasts using information theory. *Monthly Weather Review*, **130**, 1653–60.

Selten R (1998) Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, **1**, 43–61.

Snellman H and Viren M (2006) ATM networks and cash usage. Technical Report 21, Bank of Finland.

Whittaker J, Whitehead C and Somers M (2005) The neglog transformation and quantile regression for the analysis of a large credit scoring database. *Applied Statistics*, **54**, 863–78.