

Моделирование работы банкомата

Данная статья посвящена созданию модели банкомата, позволяющей прогнозировать расход денежных средств.

Предыстория такова: представитель филиала одного из достаточно крупных банков указал на потребность в более-менее точном прогнозе расхода денежных средств по банкоматам (с целью оптимизации работы инкассаторской службы) и предложил на коммерческой основе рассмотреть вопрос о возможности создания модели и софта, позволяющего закрыть эту потребность. Как и следовало ожидать, в итоге содействия и заинтересованности со стороны банка особого не было, денег никто никому не заплатил, поэтому позволю себе выложить эту работу на всеобщее обозрение...

Постановка задачи (писал сам «как понял», поэтому мог что-нибудь перевернуть)

Имеется порядка 80 банкоматов, 400 действующих (всего ~ 700) зарплатных проектов и ~160 000 картсчетов. В каждом банкомате 4 кассеты с купюрами разного достоинства, максимальное количество купюр в каждой кассете = 2800 штук (загружено, по разным соображениям, может быть меньше). Критерием необходимости загрузки является 10%-ный остаток купюр в одной из кассет. Максимальная сумма снятия за день (по зарплатным проектам) ограничена 200 тыс., максимальная сумма за транзакцию ограничена количеством купюр, пролезавшей в щель (максимум 40 штук). В реальности крупные суммы в основном снимают по терминалу в банке – есть возможность заказать банкноты определенного достоинства и безопаснее. Возможны ситуации, когда банкомат полностью «разгружается» за сутки; если в течение 3х месяцев «разгрузки» не произошло, банкомат инкассируется принудительно. Банкомат может быть настроен на два варианта алгоритма работы:

- а) чтобы купюр разного номинала было равное кол-во
- б) чтобы сумма была выдана минимальным числом купюр.

В течение рабочего дня возможно инкассировать не более 20 банкоматов (в выходные и праздничные дни инкассация не работает). В 15.00 делается заявка по деньгам на следующий день. С банкоматов, в режиме OnLine, возможно получение любой информации, в том числе по остаткам купюр.

При начислении ЗП, через банк-клиент подаются списки на перечисления, перечисляется сумма на соответствующий корсчет, выполняется проверка (в течение 1,5 часа), если перечисление выполнено после 16-45, то сумма зачисляется на картсчета только на следующий день.

При расходе с картсчета, уменьшается лимит, остаток изменяется только после акцептирования (подтверждения) операции, задержка при этом в отдельных случаях может превышать месяц.

Ставится задача минимизировать число инкассаторских поездок, с приемлемой точностью прогнозируя расход денежных средств и необходимость инкассации на ближайшие несколько дней.

Далее рассматривается подход к созданию параметрической модели банкомата.

Предварительный анализ суммарных показателей

Анализ суммарных показателей показывает, что суммарное (за 2,5 года) поступление средств по зарплатным проектам составляет порядка 95% от суммарного расхода по всем банкоматам.

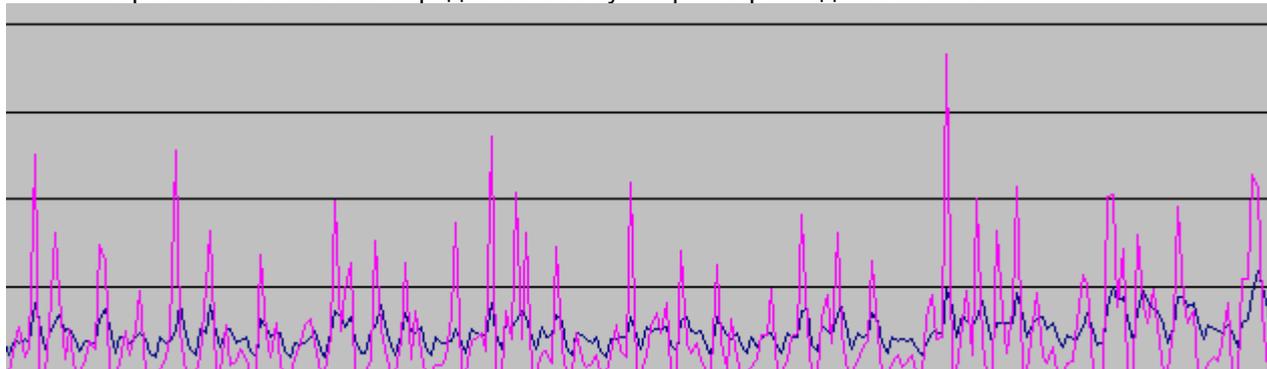


Рисунок 1. Сравнение суммарного поступления (розовый цвет) по зарплатным проектам и суммарного снятия по всем банкоматам (синий). Период - около полугода.

Учитывая превалирующую роль составляющей по зарплатным проектам, необходимо определять и учитывать коэффициенты привязки зарплатных проектов к банкоматам.

Из графика суммарных оборотов (рисунок 1) видно, что:

- Пики снятия хорошо коррелируют с пиками перечисления, что подтверждает превалирующий вклад зарплатных проектов.

- Нет явной зависимости, за исключением уикенд-минимумов, снятия от τ - параметров (параметры, характеризующие дату), что не позволяет сделать достоверный прогноз с помощью модели, представленной в виде $f(\tau_j)$, без учета других факторов.

Оценка влияния остатка по проектам

Если исходить из предположения, что клиент использует стратегию равномерного снятия суммы от зарплаты до зарплаты (точнее – по мере надобности, исходя из индивидуальных потребностей), то сумма снятия будет пропорциональна доступному остатку.

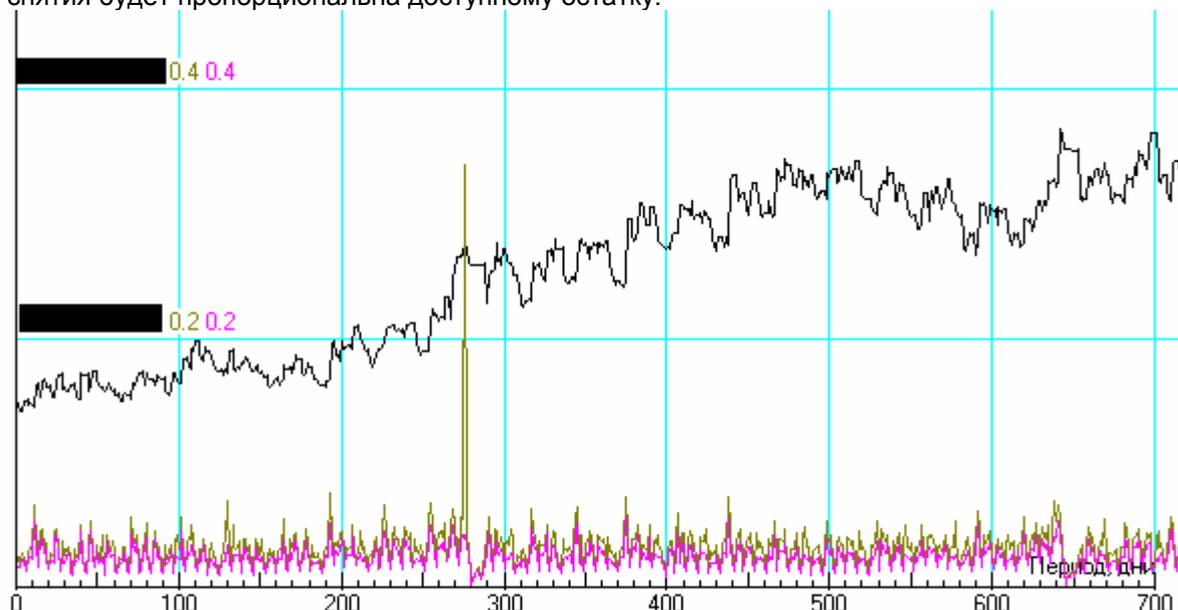


Рисунок 2. Сравнение коэффициентов суммарного снятия и расхода. Период - около двух лет.

- Черная линия – остаток по всем проектам.
- Розовая линия – отношение суммарного снятия к суммарному остатку.
- Коричневая - отношение суммарного расхода к суммарному остатку.

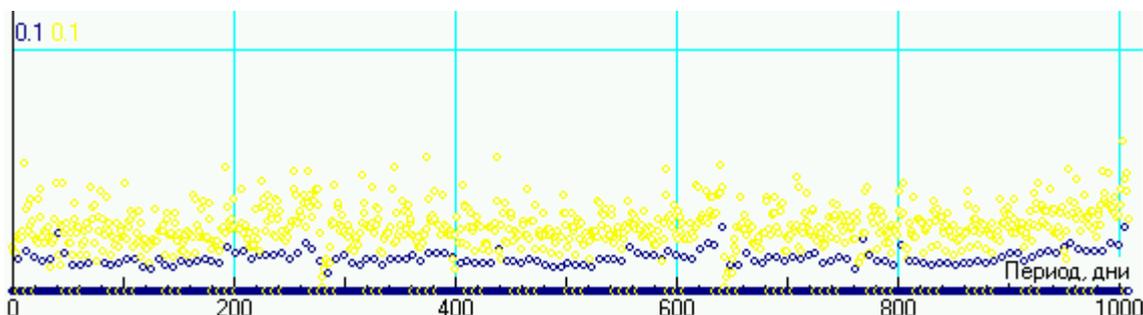


Рисунок 3. Сравнение коэффициентов суммарного снятия для суббот (синий цвет) и остальных дней недели (желтый цвет) . Период - около двух лет.

Анализ суммарных показателей по проектам позволил выявить следующие закономерности:

1) Снятие по банкоматам более равномерно и прогнозируемо, чем полный расход с картсчетов (рисунок 2). Это, судя по всему, является следствием того, что разовые крупные суммы удобнее снимать не через банкоматы.

2) Суммарный остаток за исследуемый период меняется достаточно сильно, в то же время коэффициент снятия (отношение суммарного снятия к суммарному остатку) колеблется примерно возле одного и того же значения ($\sim 0,024$ для будней и воскресений и примерно в 1,5-2 раза меньше для суббот, рисунок 3).

3) Набор τ -параметров сводится к признаку суббота/остальные дни недели, стабильной зависимости от остальных (день недели, месяца и период года) практически не наблюдается, разве что характерный всплеск и следующий за ним минимум под Новый Год. Статистический разброс составляет около 50% процентов, распределение коэффициента снятия можно представить в виде би-нормальной зависимости (рисунок 4). На рисунке 2 можно наблюдать ежемесячные локальные пики, но они, скорее всего, обусловлены увеличением перечислений в эти дни и строгой зависимости от даты не имеют.

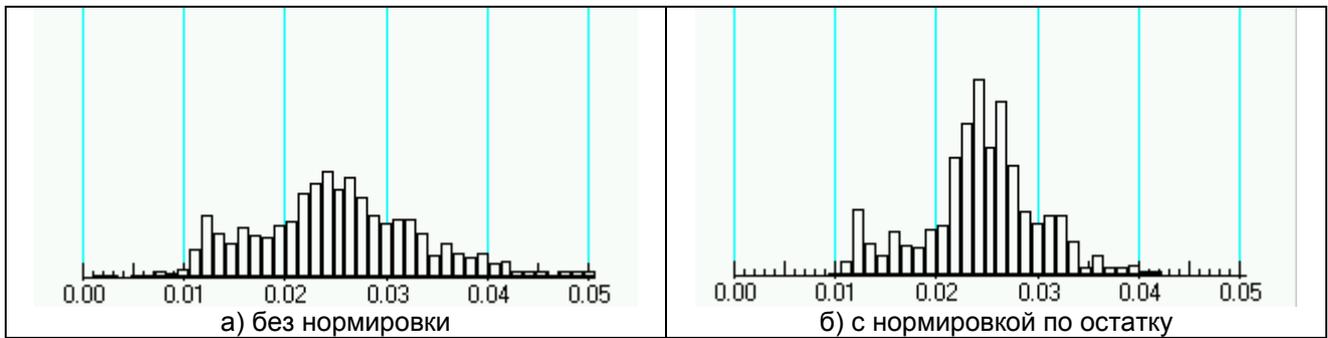


Рисунок 4. Статистика распределения коэффициента суммарного снятия.
Период – около 2х лет. Распределение хорошо описывается суммой двух нормальных распределений (на субботы и остальные дни)

Анализ распределения коэффициентов снятия для отдельных проектов показывает (рисунки 5, 6), что проекты можно разделить на две составляющие, «спящие», по которым расхода практически нет либо он не зависит от остатка и «активные», снятие по которым хорошо коррелирует с остатком на счете.

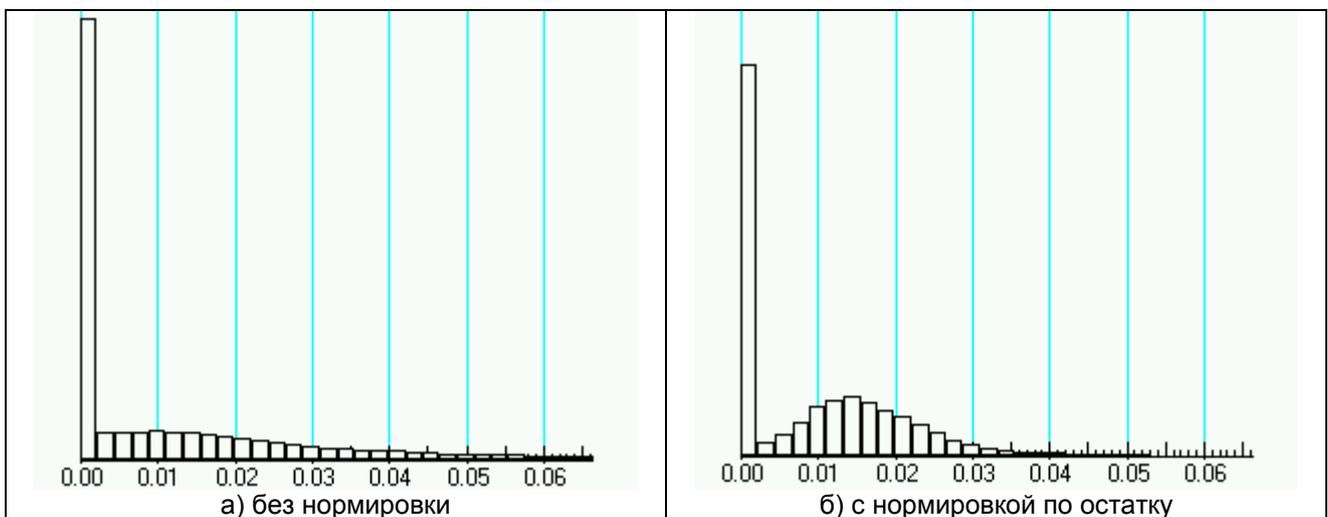


Рисунок 5. Статистика распределения частных коэффициентов снятия. Период – около 2х лет.

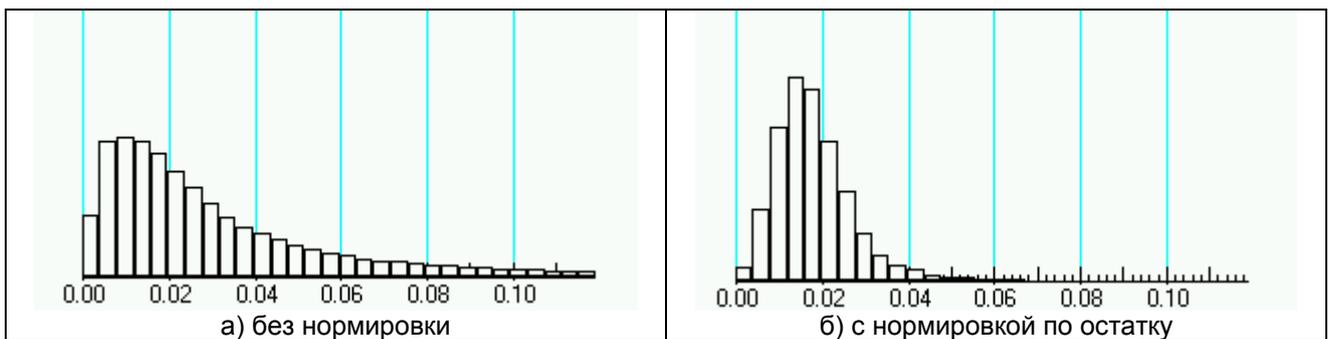


Рисунок 6. Статистика распределения частных коэффициентов снятия (без учета нулевых коэффициентов). Период – около 2х лет. Распределение хорошо описывается распределением Вейбула.

Несмотря на высокую дисперсию коэффициента снятия, использование его для прогноза дает неплохую точность. В прогнозе на рисунке 7 коэффициент снятия на весь период использовался в виде константы, поэтому видны его «новогодние» отклонения в виде всплесков фактического расхода. Также хорошо заметно повышенное снятие со счетов в связи с резким ростом курса доллара в декабре 2008 г.

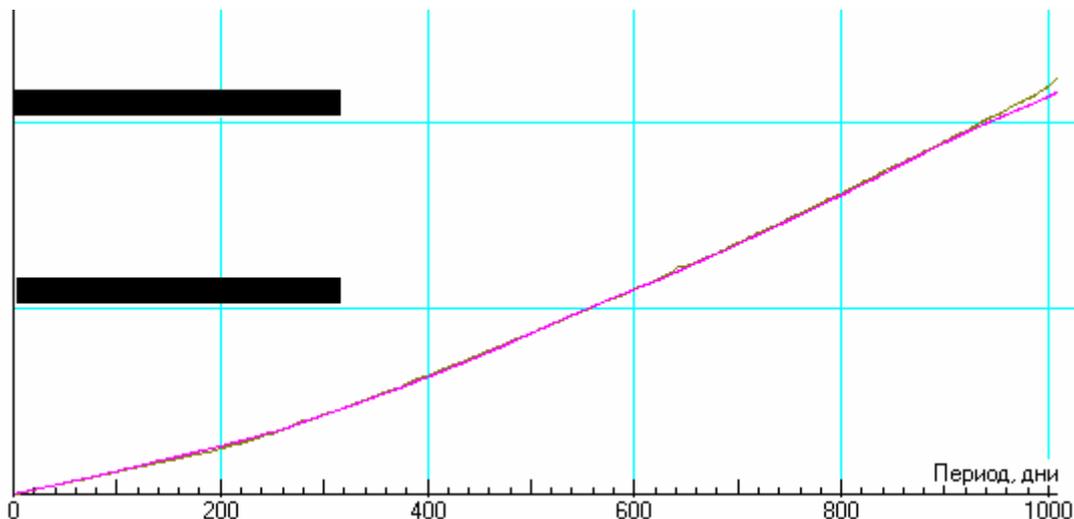


Рисунок 7. Сравнение суммарного снятия по факту (коричневая линия) и прогнозу (розовая линия). Период около 2,5 лет.

Краткие выводы:

1) Точность прогнозирования суммарного расхода, используя коэффициент снятия, уже на период около недели составляет более 95% (рисунок 7). Точность можно повысить, если использовать различные коэффициенты для субботы и остальных дней и учитывать изменение среднего значения (например, вызванного изменением макроэкономической ситуации).

2) Для прогноза расхода по отдельным проектам, как показывает сравнение графиков на рисунках 4-6, коэффициенты снятия будут существенно различаться, дисперсия их распределения будет несколько выше. Однако (если не принимать во внимание «нерегулярные» проекты), точность предсказания расхода на срок более недели может быть даже выше, чем для суммарного расхода, ибо для проекта будет использоваться его собственный, а не усредненный коэффициент.

Оценка влияния перечислений по проектам

Если исходить из предположения, что клиент использует стратегию снятия всей суммы через некоторое время после перечисления, то распределение снятия средств, поступивших по i -ому

траншу, будет подчиняться экспоненциальной зависимости $\frac{dS_{iz}}{dt} = Kp * e^{-Kp(t-T_{iz})}$.

Судя по тому, что на рисунке 8 пики прогноза и факта хорошо коррелируют, для определенной части картсчетов эта зависимость действительно имеет место быть. Но не для всех – поэтому прогноз получается существенно завышенным.

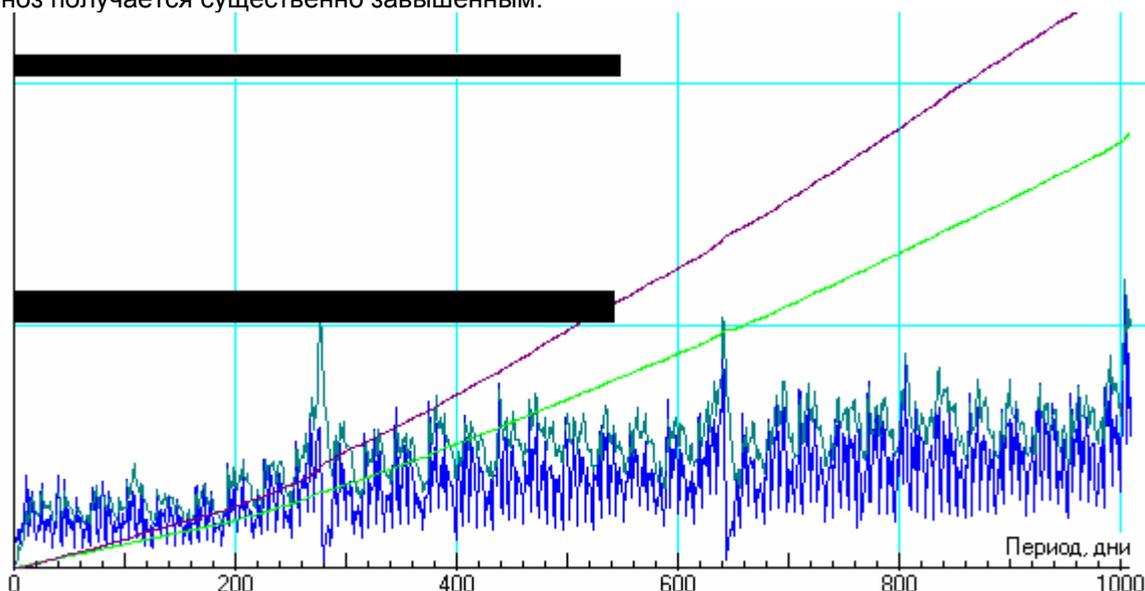


Рисунок 8. Сравнение суммарного и ежедневного снятия и прогноза, использующего экспоненциальную зависимость от даты поступления. Период около 2,5 лет. $Kp=1$.

- Светло-зеленая линия: суммарное снятие
- Сиреневая линия: суммарный прогноз
- Синяя линия: ежедневное снятие
- Темно-зеленая линия: ежедневный прогноз

Оценка влияния расхода по чужим проектам

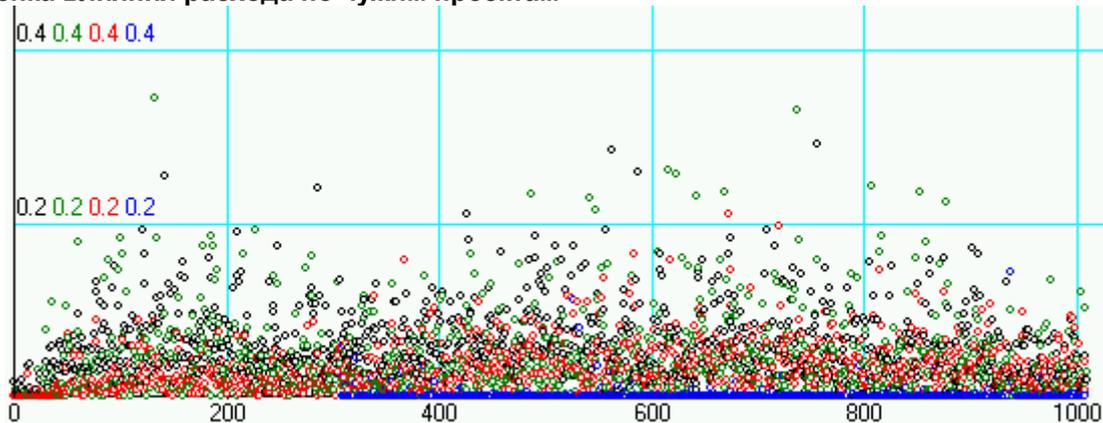


Рисунок 9. Сравнение коэффициента расхода по чужим проектам для 4х банкоматов. Период около 2,5 лет.

- Общедоступный в центре города (черный цвет)
- Общедоступный (зеленый цвет)
- Магазин в центре города (красный цвет)
- На заводе (синий цвет)

Судя по рисунку 9, до 10% суммарного расхода банкомата составляет расход по чужим карточкам. Есть исключения, например, на заводе чужие карточки, судя по всему, изредка используют лишь командированные (опять же, непонятно, что подразумевается под чужими – банк может быть тот же самый, но филиал, контролирующий картсчета находится в другом городе).

Сумма расхода по чужим проектам является менее стабильным параметром, чем ее отношение к суммарному расходу. Это вполне объяснимо – локальные и глобальные факторы, оказывающие влияние на расход средств, с большой долей вероятности воздействуют и на чужие зарплатные проекты.

Методика расчета

Опишем модель, позволяющую делать прогноз расхода по каждому банкомату, используя следующие условные обозначения:

b – индекс по банкоматам

z – индекс по зарплатным проектам (параметр с индексом z подразумевает суммирование по всем картсчетам проекта)

i – индекс по перечислениям

T_{iz} – момент перечисления i -ого транша по z -ому проекту

τ_j – набор параметров, зависящих от даты и влияющих на расход (сезон, месяц, день недели, праздник и т.д.)

\sum_t – суммирование за период (не менее 2х месяцев)

Определять какие-либо коэффициенты с точностью до конкретного человека смысла не имеет (люди могут уходить на больничный, в отпуск, менять свои предпочтения в течении года и т.д.), говорить о какой-то стабильности можно в лучшем случае только в целом по конкретному ЗП проекту. Другими словами, предпочтения клиентов (по выбору банкомату, по выбору стратегии снятия) являются вероятностными факторами и говорить о каких-то статистических закономерностях имеет смысл только на больших выборках. Соответственно, уровень детализации при определении коэффициентов регрессии должен быть ограничен зарплатными проектами.

При изменении локальных (с точки зрения окружения в месте расположения банкомата) и глобальных (макроэкономическая ситуация и т.п.) факторов, коэффициенты регрессии будут меняться. Из этого вытекает, что статистически определяемые параметры модели требуется регулярно пересчитывать. Учитывая большое количество параметров, оперативный расчет и корректировку параметров можно будет выполнить только в автоматическом режиме.

Исходя из общих соображений, более-менее регулярные (т.е. поддающиеся прогнозированию) стратегии снятия сумм для конкретного зарплатного проекта ограничены двумя рассмотренными выше вариантами:

1. Равномерное снятие, исходя из имеющихся потребностей, то есть, пропорционально остатку на счете (или лимиту). Например, банкомат рядом, смысла снимать всю ЗП нет, в итоге в течение месяца она расходуется равномерно, крупные суммы (на дорогие покупки) в целом по проекту также снимаются статистически равномерно.

2. Снятие всей суммы сразу, через определенный промежуток времени после перечисления. Суммарно по проекту, с учетом распределения промежутка времени, это приводит к экспоненциальной зависимости для каждого перечисления.

Совместное использование этих двух зависимостей позволит разделить стратегии снятия и выделить «спящие» (см. рисунки 5 и 6) проекты и картсчета.

Прогноз расхода по банкомату за день D складывается как сумма по прогнозам для зарплатных проектов, с учетом коэффициентов привязки:

$$R^b(D) = (1 + K\psi^b) \sum_z q_z^b R_z \quad (1)$$

Снятие по чужим картсчетам учитывается в виде множителя $(1 + K\psi^b)$, а не слагаемого, так как выше уже было показано, что сумма расхода по чужим проектам является менее стабильным параметром, чем ее отношение к суммарному расходу.

$$\text{Здесь } K\psi^b = \frac{\sum_t S^b - \sum_{t,z \in Z} S_z^b}{\sum_{t,z \in Z} S_z^b} \quad \text{коэффициент снятия по чужим проектам} \quad (2)$$

Здесь S^b - снятие по банкомату в целом, S_z^b - для конкретного проекта по этому банкомату.

Коэффициенты привязки зарплатного проекта к банкомату определяются как

$$q_z^b = \frac{\sum_t S_z^b}{\sum_t S_z} \quad (3)$$

Прогноз снятия по конкретному проекту за день D определяется как функция от текущего остатка O_z и предыдущих перечислений P_i^z

$$R_z(D) = f_z(D) \left(h^z Ks^z O^z + (1 - h^z) Kp^z \sum_{i \in z} P_i^z e^{Kp^z (T_i^z - D)} \eta(D - T_i^z) \right) \quad (4)$$

здесь $\eta(x) = \{0, x < 0; 1, x \geq 0\}$

различие коэффициентов снятия для разных дней учитывается в виде

$$f_z(D) = \sum_{d=1}^7 Kd^z \delta(d - D) \quad (5)$$

здесь $\delta(x) = \{0, x < 0; 1, x = 0\}$

$$\text{где коэффициенты дня недели } Kd^z(D) = \frac{\sum_t S^z \delta(d - D)}{\sum_t S^z} \quad (6)$$

Влияние даты учитывается для проектов, а не для банкоматов потому, что похожие стратегии снятия и режим работы владельцев картсчетов более характерны для проекта, чем для банкомата, кроме того, исходя из формулы (1), предполагается, что расход по банкомату не является независимым параметром, а зависит от расхода по проектам.

Коэффициент снятия берется как усредненное отношение снятия по проекту к остатку по нему:

$$Ks^z = \frac{1}{t} \sum_t \frac{S^z}{O^z} \quad (7)$$

$h^z \in [0..1]$ определяет соотношение стратегий снятия для конкретного проекта.

Параметр Kp^z имеет величину порядка 1 – большие его значения приводят к 100% снятию всего поступления, а маленькие сглаживают зависимость, тем самым теряется смысл в использовании зависимости от прихода. На первом этапе Kp^z можно принять равным 1.

Таким образом, для первого приближения модели, можно считать, что только коэффициенты h^z не определяются явным образом, а требуют вычисления (например, с использованием метода наименьших квадратов).

С учетом того, что в предоставленных данных встречаются ситуации расхода при нулевом остатке или сумма расхода превышает остаток, можно сделать вывод, что точность прогноза дополнительно можно увеличить, используя следующие подходы:

- приняв во внимание время зачисления;
- ориентируясь не на остаток не счете, а на текущий лимит;
- проанализировав распределение расхода и лимита на зарплатную и овердрафтную часть;
- проведя анализ отдельных характерных зарплатных проектов.

Покупюрный прогноз расхода

Рассматривая необходимость покупюрного прогноза и, соответствующего усложнения модели поведения банкомата, хотелось бы отметить, что сам по себе, покупюрный прогноз не поможет решить задачу минимизации инкассаторских поездок:

Во первых, учитывая достаточно большую дисперсию ежедневного расхода и существенное влияние вероятностного фактора, даже суммарный расход будет даваться с определенной погрешностью, а для покупюрного требуется учитывать распределение выдачи еще и по суммам (на что тоже накладывается существенный дополнительный вероятностный фактор). Учитывая достаточно высокий порог требуемой достоверности, покупюрный прогноз может принести эффект только на некоторых банкоматах, с достаточно низким влиянием вероятностного фактора.

Во вторых, минимизация инкассаторских поездок может быть достигнута только при минимизации остатка в инкассируемых банкоматах. Учитывая, что требованием инкассации является снижение *количества* купюр в одном из лотков до критического, минимизация остатка достигается только при *сбалансированном количестве купюр во всех кассетах банкомата*. Исходя из вышесказанного, гораздо больший эффект принесет перевод банкоматов в режим «выдача равного количества купюр каждого достоинства».

Таким образом, оптимизация баланса покупюрного остатка сводится к анализу статистики снятых сумм и настройку банкомата на оптимальный набор [достоинства] купюр.

Изменение режима работы банкоматов

При изменении режима работы одного или нескольких банкоматов, например, при выводе на профилактику или при вводе в эксплуатацию нового банкомата, коэффициенты регрессии в той или иной степени могут меняться для всех банкоматов. В то же время, этот выход на стационарный режим будет достаточно плавным и растянутым по времени.

Как было показано выше, спрогнозировать изменение коэффициентов регрессии с приемлемой точностью нереально, предсказание перераспределения коэффициентов q_z^b тоже является нетривиальной задачей (уникальной для каждого конкретного случая), поэтому, при изменении режима работы, параметры модели предполагается постоянно уточнять на основе «родной» статистики. Для начального приближения, коэффициенты привязки предполагается брать как усредненные по относительному расстоянию до соседних банкоматов. Однако, следует понимать, что до выхода на стационарный режим, прогноз по этим коэффициентам будет слегка завышен. При превышении реального расхода над суммой прогноза, следует переключаться на автоопределение по статистическим данным.

Приложение 1. Используемые данные

Запрашивали одно, получили совсем другое... Требования (даже согласованные), касающиеся формата предоставления данных нагло игнорировались, каждый «запрос» приходилось ожидать от 2х недель до нескольких месяцев, поэтому мы в итоге плюнули на все и стали работать с тем, что было. Некоторую информацию так и не предоставили – скорее всего, строить большие запросы и склеивать базы было лень. В общем, это наглядно продемонстрировало уровень заинтересованности и позволило оценить объем саботажа (если бы проект утвердили).

- 1) Сгруппированные посуточно данные по суммарному (по всем банкоматам) расходованию средств и поступлению средств по зарплатным проектам
 - Дата
 - Сумма перечисления по зарплатным проектам за сутки (по всем зарплатным проектам)
 - Сумма выдачи средств через банкоматы (суммарно по всем банкоматам)
- 2) Данные о перечислениях
 - Зарплатный проект
 - Дата зачисления
 - Сумма перечисления
- 3) Расход по банкоматам для собственных зарплатных проектов
 - Банкомат
 - Дата снятия
 - Зарплатный проект
 - Сумма снятия

- 4) Остаток по картсчетам, суммарно по проектам. Судя по тому, что воскресное сальдо совпадает с субботним, остатки предоставлены по состоянию на конец дня.
- Зарплатный проект
 - Сальдо по проекту
 - Дата
- 5) Полный расход в разрезе банкоматов (в том числе по чужим проектам)
- Банкомат
 - Дата снятия
 - Сумма снятия
- 6) Полный расход (по банкоматам, терминалам, магазинам и т.д.) в разрезе зарплатных проектов.
- Дата
 - Зарплатный проект
 - Сумма минусования

Выводы

Анализ предоставленных данных показал возможность прогнозирования расхода по банкоматам, основываясь на коэффициентах привязки зарплатных проектов к банкоматам и регрессионной модели, использующей две стратегии снятия средств – **снятие по мере возможности** (экспоненциальная зависимость от момента поступления средств) и **снятие по мере потребности** (линейная зависимость от остатка).

Однако, предоставленные данные были не достаточно полны и точны для получения приемлемой (на малых интервалах) точности при определении коэффициентов регрессии модели.

Поступления осуществляются по зарплатным проектам, однако, в случае, если человек работает в нескольких местах, на его картсчет могут поступать средства по разным зарплатным проектам. Таких случаев не очень много, и погрешность, на мой взгляд, не такая критичная, однако, для корректного построения модели, требуется вести учет остатков на картсчетах в разрезе зарплатных проектов, снятие/расход с такого картсчета уменьшает остаток по проектам пропорционально остатку на момент снятия.

Обе используемые стратегии снятия опираются на понятие доступного остатка на момент снятия, однако было предоставлено сальдо по зарплатным проектам лишь на конец дня. Два раза предоставлялись данные – за пересекающийся период оба раза цифры были разные, как-то это настораживает :).

Не совсем понятно, как определялась привязка к зарплатному проекту (см. выше), но, самое главное, что при определении коэффициентов снятия к остаткам на счете, по некоторым проектам были парадоксальные случаи – снималось на несколько порядков больше, чем остаток по проекту.

Возможных причин несколько:

- 1) Снятие происходит в тот же день, что и перечисление
- 2) Остаток на картсчете не равен доступному остатку «все операции по карте на счетах отражаются с запозданием в среднем от 1 до 3 дней, но бывают и до 45», лимит на момент снятия неизвестен.
- 3) Опять же, неизвестно, снимают люди в обычном режиме или по овердрафту.

По некоторым проектам, в первую очередь где стоят «родные» банкоматы, заметно несоответствие пиков перечисления и снятия. Либо это обусловлено спецификой предприятия (нет возможности снять вечером), либо зачисление происходит только на следующий день.

В итоге:

- Не было корректной информации в разрезе зарплатных проектов
- Не было информации о доступном остатке на момент снятия
- Не было информации о доступном овердрафте и снятии в разрезе овердрафта
- Не было информации о времени зачисления при перечислении (момента, с которого средства становятся доступными).

Натолкнувшись пару раз на коэффициент снятия по проекту на 2-3 порядка больше единицы, стало ясно, что с такими «грязными» данным придется либо корректировать для каждого банкомата коэффициенты модели вручную, подгоняя под факт, либо обходиться более простой моделью (может быть менее точной, но зато нивелирующей погрешность предоставленных данных).

Целью данной статьи является исключительно описание подхода к построению модели. Поэтому дополнительных конкретных данных и графиков, подтверждающих выводы, здесь не приводится.

Информация, которая может составлять коммерческую тайну, также не разглашается (тем более, что данные нам предоставили с нормировкой на некоторый коэффициент).

Поскольку стало ясно, что реализация данного проекта на достойном уровне весьма проблематична (особой заинтересованности у тех лиц, кто что-то решал, не было – оно и понятно, кризис, да и возни и затрат много для такого «венчурного» проекта), я завязываю с этой задачей.

В принципе, задача решаема в масштабах всего банка, но не в это время и не в этой стране :). Более логично такие задачи решать на уровне поставщиков процессинговых систем, ибо как только встает вопрос о внесении каких-либо дополнений в используемый стандартный софт (даже если это технически возможно), энтузиазм моментально улетучивается.

Ну и, до полного счастья, встает вопрос заблаговременного получения (с помощью организационных мер или технически) информации о перечислениях. Т.е. для тех проектов, где «снятие по мере возможности» превалирует над «снятием по мере потребности», достоверный прогноз возможен только на тот период, на который предоставлена информация о планируемых перечислениях.

(С) Сергей Пашков, 2009
s300@ngs.ru, ICQ UIN 117631343