

УДК 519.234

ПРИМЕНЕНИЕ СКОРИНГОВЫХ МОДЕЛЕЙ ВЫЖИВАЕМОСТИ К ПРОГНОЗИРОВАНИЮ ДОХОДНОСТИ БАНКОВ В ОБЛАСТИ АВТОМОБИЛЬНОГО КРЕДИТОВАНИЯ



М. М. Васьковский
заведующий кафедрой высшей математики в Белорусском государственном университете;
старший статистический аналитик в HiQo Solutions Ltd;
доктор физико-математических наук.



А. В. Кулешова
статистический аналитик в HiQo Solutions Ltd

Кафедра высшей математики
Белорусский государственный университет, Республика Беларусь
HiQo Solutions, Ltd, USA
E-mail: vaskovskii@bsu.by, alina.kuleshova@hiqo-solutions.com

M. Vaskouski

Заведующий кафедрой высшей математики в Белорусском государственном университете; старший статистический аналитик в HiQo Solutions Ltd. Эксперт в теории стохастических дифференциальных уравнений и их приложений в финансовой математике. Доктор физико-математических наук.

A. Kuleshova

Статистический аналитик в HiQo Solutions Ltd. Эксперт в области разработки наукоёмкого программного обеспечения для анализа финансовых и экономических процессов.

Аннотация. В работе рассматривается применение дискретных моделей выживаемости (DTSM) для моделирования доходности на примере автомобильных кредитов США. В основе рассматриваемой модели лежат Age-Period-Cohort декомпозиции условных вероятностей, определяющих текущее состояние кредитного аккаунта. Важную роль при построении модели играет экономическое моделирование на основе использования реальных экономических данных по ключевых макроэкономическим факторам. Предлагаемый подход учитывает также индивидуальные характеристики кредитных аккаунтов посредством построения скоринговых моделей для условных вероятностей состояний кредитных аккаунтов.

Ключевые слова: кредитные потери, доходность, стохастические дифференциальные уравнения, регрессионные модели.

Введение

Скоринговые модели выживаемости (Discrete Time Survival Models) успешно применяются к анализу и прогнозированию кредитных потерь банков [1, 2]. Важнейшее преимущество моделей DTSM заключается в том, что эти модели позволяют проводить анализ на уровне отдельных категорий аккаунтов (например, аккаунтов, открытых в определенный период времени, или имеющих высокие риски). В настоящей работе предлагается приложение моделей DTSM к прогнозированию доходности банков. В настоящей статье мы рассматриваем кредиты на автомобили. Однако предлагаемые методы

с незначительными модификациями работают и для других типов кредитов (ипотечные, потребительские и пр.).

В статье приводится детальное описание модели DTSM, формулы для вычисления доходности на основе модели DTSM, а также обсуждаются методы валидации результатов моделирования. Приводятся практические результаты, полученные в процессе моделирования. Для моделирования и прогнозов использовались достаточно большие выборки автомобильных кредитных займов США.

Основные результаты

В основе моделей DTSM лежат Age-Period-Cohort декомпозиции (далее APC-декомпозиции) для уровней дефолта и погашения займов

$$PD(a, t) = \frac{N_{def}(a, t)}{N_{act}(a, t - 1)},$$

$$PA(a, t) = \frac{N_{attr}(a, t)}{N_{act}(a, t - 1)},$$

где N_{def} , N_{act} , N_{attr} – числа дефолтных, активных и закрытых аккаунтов, a – возраст кредитного займа (в месяцах), t – текущая дата наблюдения.

Компоненты APC-декомпозиции находятся с помощью следующей биномиальной регрессионной модели:

$$\text{logit}(PD(a, t)) = F(a) + H(t) + G(v) + \varepsilon(a, t),$$

где $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$ – обратная к логистической функции $g(z) = \frac{1}{1+e^{-z}}$, $z \in R$, функция $F(a)$ является жизненным циклом займа, $H(t)$ – макроэкономическая составляющая (далее функция энвайронмента), а функция $G(v)$ отражает влияние неэкономических факторов (далее функция кредитного риска), v – дата открытия займа, $\varepsilon(a, t)$ – ошибка.

В силу линейной зависимости аргументов a , v , t метод наименьших квадратов оказывается неприменимым, поэтому для поиска декомпозиций применяют методы регуляризации или байесовские методы на основе симуляции случайных блужданий цепей Маркова [3]. А для решения проблемы однозначной идентификации функций $F(a)$, $H(t)$, $G(v)$ предполагают, что функция кредитного риска $G(v)$ не содержит линейного тренда, а энвайронмент $H(t)$ не содержит линейного тренда и имеет нулевое среднее.

Аналогичная модель строится для нахождения APC-декомпозиций переменной PA .

Модели строились на данных, содержащих полный жизненный цикл. Таким образом, для построения прогнозов нет необходимости экстраполировать функции жизненного цикла для переменных PD , PA .

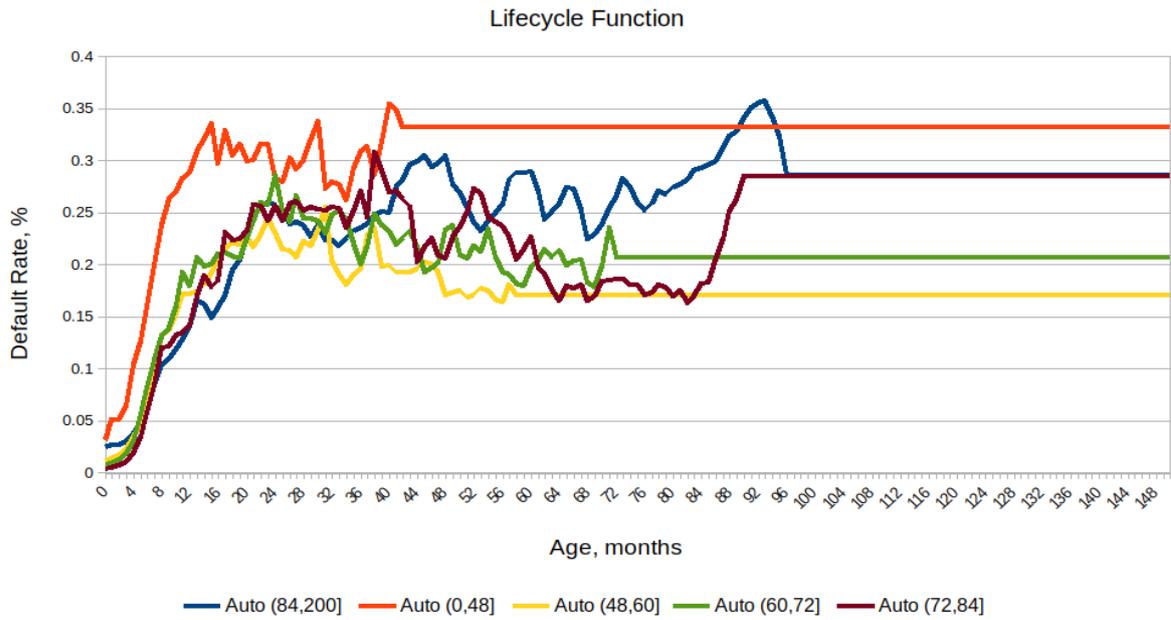


Рисунок 1. Жизненные циклы PD, сегментированные по сроку выдачи займа

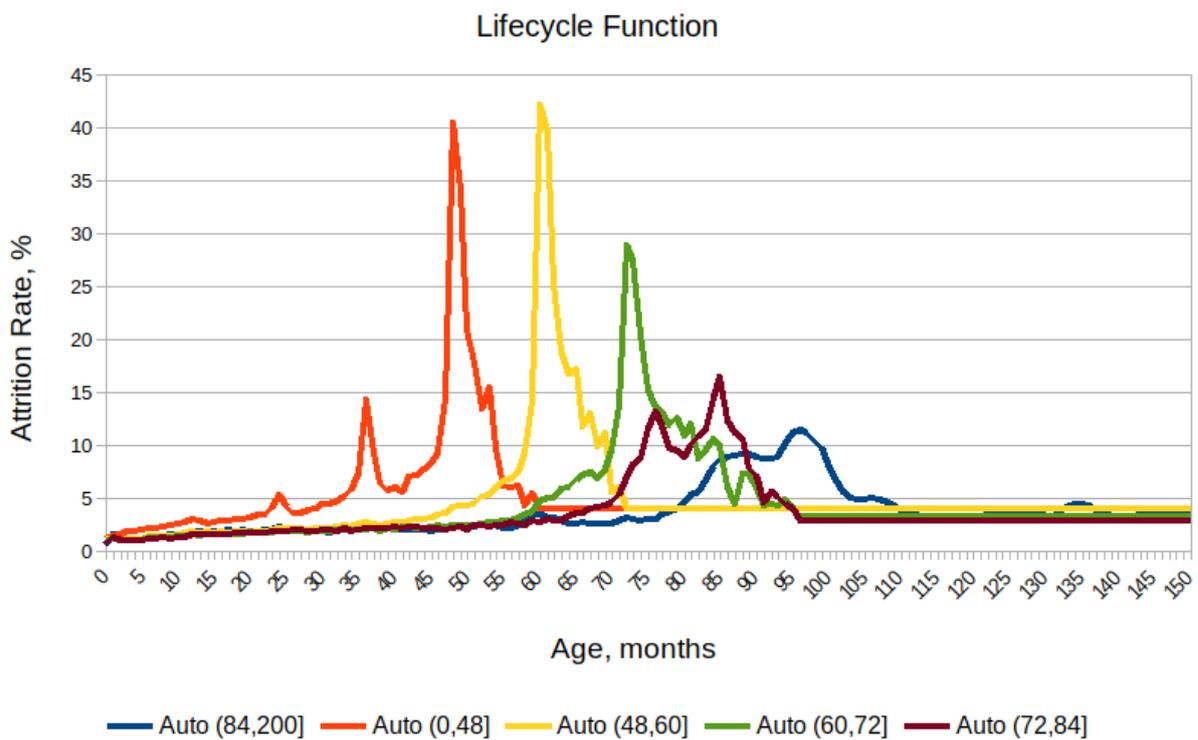


Рисунок 2. Жизненные циклы PA, сегментированные по сроку выдачи займа

Одна из наиболее сложных проблем заключается в нахождении реалистичных экстраполяций функции энвайронмента $H(t)$.

На основе гауссовской регрессии строится экономическая модель для функции энвайронмента $H(t)$:

$$H(t) = \beta_0 + \sum_{i=1}^n \beta_i f_i(t) + \varepsilon(t),$$

где $f_i(t)$ ($i = 1, \dots, n$) – некоторый набор макроэкономических факторов, β_i – искомые коэффициенты, $\varepsilon(t)$ – ошибка.

Следующие наблюдения имеют важное значение при построении экономических моделей: 1) динамика изменения экономических показателей существенно их абсолютных значений, поэтому вместо исходных макроэкономических факторов используются дискретные аналоги производной (Diff) и логарифмической производной (LogRatio); 2) показатели кредитных рисков определяются состоянием экономики с учетом некоторого запаздывания.

Для PD получается следующий результат

Таблица 1. Лучшее приближение для PD

Переменная	Преобразование,	Оптимальное значение lag	Оптимальное значение win	Коэффициент,
House Price Index	LogRatio	9	18	-1.31
Real DPI	LogRatio	0	12	-7.13
Real GDP	LogRatio	0	12	-7.92

Проверялись комбинации от 1 до 3 макроэкономических факторов и преобразований из следующего списка: Real GDP (LogRatio), Real DPI (LogRatio), Unemployment rate (Diff, LogRatio), House Price Index (LogRatio), 5-year Treasury yield (Diff), 10-year Treasury yield (Diff), BBB corporate yield (Diff), Commercial Real Estate Price Index (LogRatio), CPI inflation rate (Diff), Market Volatility Index (LogRatio), Dow Jones Total Stock Market Index (LogRatio). При этом тестировались значения запаздывания от 0 до 12 и ширина окна от 1 до 24.

Макроэкономические данные США являются общедоступными [4].

Для нахождения экстраполяций экономического приближения $\hat{H}(t)$ энвайронмента $H(t)$ общепринятым является следующий метод. Используя ежегодный краткосрочный прогноз по макроэкономическим показателям, выпускаемый Федеральной Резервной Системой США [5], осуществляется экстраполяция функции $\hat{H}(t)$ на двухлетний период, начиная с момента t_0 , используя известные коэффициенты β_i . Для получения дальнейшей экстраполяции на период существования аккаунтов, активных на момент t_0 , используется идея возврата к среднему на основе модели Васичека: искомая экстраполяция функции $\hat{H}(t)$ задается на основе одномерного уравнения Орнштейна-Уленбека [6, 7]

$$dx(t) = \theta(\mu - x(t))dt + \sigma dW(t), \quad (1)$$

где μ – долгосрочное среднее, $\theta > 0$ – параметр, характеризующий скорость возврата к среднему значению, $\sigma > 0$ – параметр волатильности, $W(t)$ – винеровский процесс.

Решение уравнения (1) с начальным условием $x(\tau) = x_0$ задается соотношением

$$x(t) = x_0 e^{-\theta(t-\tau)} + \mu(1 - e^{-\theta(t-\tau)}) + \sigma \int_{\tau}^t e^{-\theta(t-s)} dW(s),$$

а в качестве экстраполирующей кривой выбирается математическое ожидание решения

$$E(x(t)) = x_0 e^{-\theta(t-\tau)} + \mu(1 - e^{-\theta(t-\tau)}),$$

где $\tau = t_0 + 24$ – момент начала долгосрочной экстраполяции.

На следующем этапе мы строим скоринговые модели. Для каждого значения горизонта прогноза $h = 1, 2, \dots$ построим следующую биномиальную логистическую регрессионную модель

$$1_{def,i}(t) = g(F(a_i) + \hat{H}(t) + X_i(t-h)B(h) + \varepsilon_i(t,h)),$$

где $1_{def,i}(t)$ – индикаторная функция состояния дефолта займа с номером i в момент времени t , a_i – возраст займа i в момент времени t , $X_i(t-h)$ – матрица значений скоринговых факторов займа i в момент времени $t-h$, $B(h)$ – вектор коэффициентов скоринговой модели, $\varepsilon_i(t,h)$ – ошибка.

На практике скоринговые модели строят для значений h от 1 до 12 (так называемые поведенческие скоринговые модели), поскольку матрица $B(h)$ близка к стационарной матрице для достаточно больших h . Кроме того, дополнительно строится аналогичная скоринговая модель, в которую входят лишь скоринговые факторы, не изменяющиеся со временем (так называемая скоринговая модель происхождения).

Аналогично строятся скоринговые модели для погашенных займов.

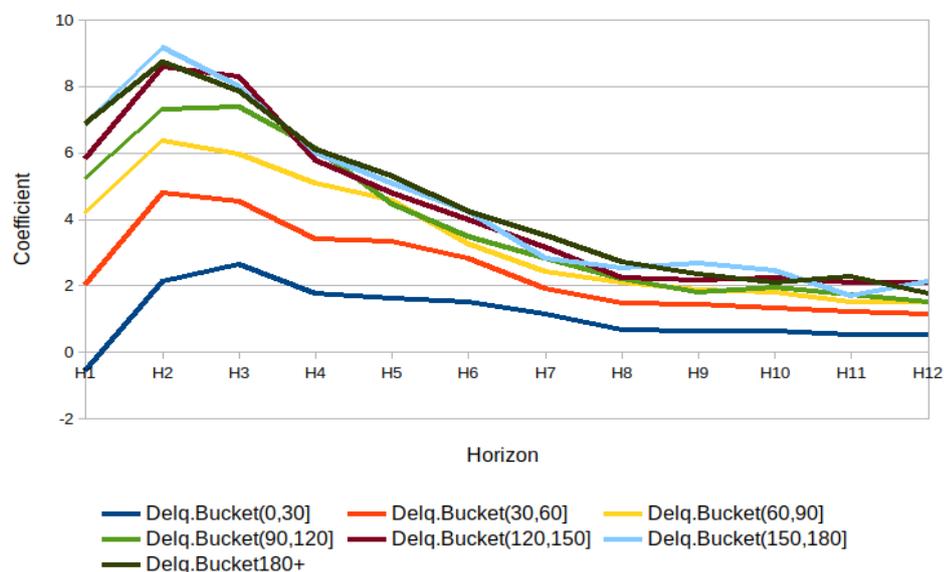


Рисунок 4. Регрессионные коэффициенты текущей задержки платежа по займу

Опишем процедуру нахождения прогноза для вероятностей статусов займов и ожидаемых значений баланса. Через $Default(h)$, $Attrition(h)$, $Active(h)$ обозначаем вероятности того, что кредитный займ находится в статусе «дефолт», «выплаченный», «активный» соответственно. Для каждого значения $h = 1, 2, \dots$ используем рекуррентные формулы:

$$Default(h) = PD(h) * Active(h - 1),$$

$$Attrition(h) = PA(h) * Active(h - 1),$$

$$Active(h) = Active(h - 1) - Default(h) - Attrition(h),$$

$$Out. Bal(h) = Out. Bal(h - 1) - Scheduled. Principal. Payment(h),$$

где $PD(h)$ есть результат применения поведенческой скоринговой модели с горизонтом $\{h, 12\}$, если возраст займа на начало прогноза составляет не менее 6 месяцев, и результат применения скоринговой модели происхождения, если возраст займа на начало прогноза составляет менее 6 месяцев; $PA(h)$ определяется аналогичным образом из скоринговых моделей для погашенных займов; $Scheduled. Principal. Payment(h)$ – предписанный основной платеж по займу на момент времени $t = t_0 + h$, где t_0 – календарная дата, предшествующая дате начала прогноза, $Out. Bal(h)$ – ожидаемое значение невыплаченного баланса по займу на момент времени $t = t_0 + h$.

Для определения доходности по займу нам понадобятся следующие характеристики: $Open. Balance$ – величина займа, APR – годовая процентная ставка на дату открытия займа, LTV – доля стоимости автомобиля, покрываемая займом, LGD – доля дефолтного баланса, неподлежащая возврату, $Cost. Funds$ – стоимость фондов, $Mgmt. Expences$ – расходы на управление, $Depr. Rate$ – процент амортизации, $NewCar. Depr. Rate$ – корректирующий коэффициент амортизации для новых автомобилей.

Процент доходности займа на горизонт h определяется по формуле

$$Yield(h) = \frac{Interest. Payment(h) - Losses(h) - Expences(h)}{Open. Balance},$$

где $Interest. Payment(h)$ – ожидаемый платеж по процентам, вычисляемый по формуле

$$Interest. Payment(h) = Active(h) * Out. Bal(h - 1) * APR/12,$$

$Expences(h)$ – месячные затраты на содержание кредита, оцениваемые по формуле

$$Expences(h) = Active(h) * Out. Bal(h - 1) * Cost. Funds/12 + \\ + Active(h) * Open. Balance * Mgmt. Expences/12,$$

$Losses(h)$ – ожидаемые потери по кредиту вследствие дефолта, определяемые соотношением

$$Losses(h) = LGD * Default(h) * (Out. Bal(h - 2) - Collateral. Val(h)),$$

$Collateral. Val(h)$ – стоимость обеспечения, вычисляемая по формуле

$$Collateral. Val(h) = \frac{Open. Balance * (1 - Depr. Rate)^{t_0+h-v} * (1 - NewCar. Depr. Rate)}{LTV},$$

v – дата открытия займа.

Исторические значения доходности *Historical.Yield* вычисляются по тем же формулам.

Процент доходности займа за весь период жизни определяется по следующей формуле

$$Yield = Historical.Yield + \sum_h Yield(h).$$

Отметим, что все вычисления производятся на уровне индивидуальных кредитных займов, а величины *LGD*, *Cost.Funds*, *Mgmt.Expences*, *Depr.Rate*, *NewCar.Depr.Rate* являются средними по индустрии. Индикатором окончания вычислений для определенного займа является окончание срока выдачи кредита или зануление невыплаченного баланса.

Для валидации построенных экономических и скоринговых моделей проводились тесты статистической значимости коэффициентов, корректность знаков коэффициентов относительно соответствующих однофакторных моделей.

Для оценки качества построенной модели доходности был проведен “sensitivity” тест.

В рамках “sensitivity” теста строился прогноз доходности для базового (Base) и неблагоприятного (Severe) экономических сценариев. Результаты представлены на графике.

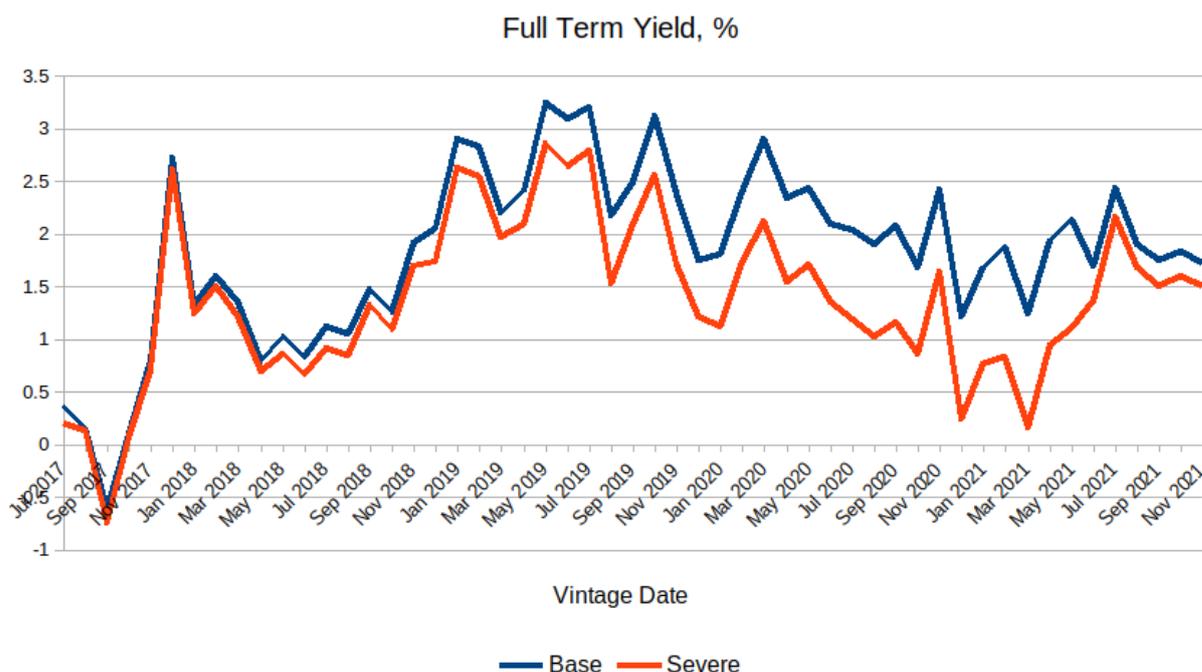


Рисунок 5. Прогноз доходности для для базового (Base) и неблагоприятного (Severe) экономических сценариев, сегментированный по дате выдачи займа

Список использованных источников

- [1] Breeden, J.L. Living with CECL: Mortgage Modeling Alternatives. – Prescient Models LLC, 2018. – 203 с.
- [2] Breeden, J.L. Current expected credit loss procyclicality: it depends on the model / J.L. Breeden, M. Vaskouski // Journal of Credit Risk. — 2020. — Vol. 16, № 1. — P. 27–48.
- [3] Schmid, V.J., Held, L. Bayesian age-period-cohort modeling and prediction – BAMP // Journal of Statistical Software. 2007. – Vol. 21, №8 – P. 1-15.
- [4] Источник баз данных с макроэкономическими факторами за последние 30 лет [1]. [Электр. ресурс]: - <https://fred.stlouisfed.org/>

[5] Источник баз данных с макроэкономическими сценариями [1]. [Электр. ресурс]: - <https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20220210a1.pdf>

[6] Леваков, А.А., Васьюковский, М.М. Стохастические дифференциальные уравнения и включения. – Минск: БГУ, 2019. – 495 с.

[7] Оксендаль, Б. Стохастические дифференциальные уравнения. Введение в теорию и приложения: Пер. с англ. – М.: Мир, ООО «Издательство АСТ», 2003. – 408 с.

APPLICATION OF DISCRETE TIME SURVIVAL MODELS TO YIELD MODELING FOR AUTO CREDIT LOANS

M. VASKOUSKI

Head of the Department of Higher Mathematics at the Belarusian State University; senior statistical analyst at HiQo Solutions Ltd; Doctor of Science in Physics and Mathematics.

A. KULESHOVA

Statistical analyst at HiQo Solutions Ltd

*Department of Higher Mathematics
Belarusian State University, Republic of Belarus
HiQo Solutions, Ltd, USA
E-mail: vaskovskii@bsu.by, alina.kuleshova@hiqo-solutions.com*

M. Vaskouski

Head of the Department of Higher Mathematics at the Belarusian State University, senior statistical analyst at HiQo Solutions Ltd; Doctor of Science in Physics and Mathematics. Expert in theory of stochastic differential equations and its applications in financial mathematics.

A. Kuleshova

Statistical analyst at HiQo Solutions Ltd. Expert in development of knowledge-intensive software for analysis of financial and economics processes

Abstract. We consider an application of discrete time survival models (DTSM) to yield modeling for auto credit loans. The core of the model are the Age-Period-Cohort decompositions of the conditional probabilities of the accounts statuses. Economic modeling based on using of the real macroeconomic factors plays a significant role. The described approach takes into account individual account characteristics, which are involved into score models for the conditional probabilities of the accounts statuses.

Keywords: credit losses, yield, stochastic differential equations, regression models.