

УДК 519.234

## ВЛИЯНИЕ СПОСОБА ЭКСТРАПОЛЯЦИЙ МАКРОЭКОНОМИЧЕСКИХ ФАКТОРОВ НА ТОЧНОСТЬ ПРОГНОЗА КРЕДИТНЫХ РИСКОВ



**М. М. Васьковский**

Заведующий кафедрой высшей математики БГУ; старший статистический аналитик в HiQo Solutions Ltd; доктор физико-математических наук.



**А. А. Карпович**

Студентка магистратуры Белорусского государственного университета; специалист по данным в EPAM Systems

Кафедра Высшей математики. Белорусский государственный университет, Республика Беларусь.  
HiQo Solutions, Ltd, USA.  
E-mail: vaskovskii@bsu.by, maxim.vaskovski@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. Доктор физико-математических наук.

### **A. Karpovich**

Студентка магистратуры Белорусского государственного университета, специальность – Алгоритмы и системы обработки больших данных; специалист по данным в EPAM Systems.

**Аннотация.** В данной работе исследуется влияние способа экстраполяции макроэкономических факторов на точность прогноза кредитных потерь в рамках модели DTSM. Целью данной работы было сравнить точность рассчитанных потерь. Использование экстраполяций с помощью решений уравнения Орнштейна-Уленбека позволило получить значительно лучший результат, если сравнить с тем, в котором использовались краткосрочные сценарии – 13 % против 28 %. Но лучший результат удалось получить, используя экстраполяции с помощью рекуррентной нейронной сети LSTM – 11 %.

**Ключевые слова:** кредитные потери, экстраполяция экономических рядов, модель регрессии, стохастические дифференциальные уравнения, рекуррентные нейронные сети.

### **Введение.**

Прогнозирование кредитных рисков является актуальной задачей в банковской сфере сейчас. В данной работе была реализована скоринговая модель выживаемости от дискретного времени (DTSM) для вычисления вероятности дефолта или закрытия аккаунта в данный момент времени в зависимости от таких переменных, как функция, отражающей жизненный цикл аккаунта и зависит от его возраста, от задержки выплаты по кредиту, от года открытия (как от категориальной переменной) и от функции, отражающей текущую экономическую ситуацию.

Для создания модели функции, зависящей от макроэкономических факторов, были использованы общедоступные наборы данных от Freddy Mac [1] и Fannie Mae [2], как было описано в книге [3]. Данные предоставляют информацию о происхождении и производительности по 30-летним ипотечным кредитам с фиксированной ставкой. В дополнение к ежемесячному статусу кредита база данных содержит ряд атрибутов, подходящих для оценки кредитного риска на уровне кредита.

Интерес исследования представляет способ экстраполяции для получения прогнозов. В данной модели вероятность дефолта или закрытия кредита зависит от функции, отражающей

текущее экономическое состояние, которая, в свою очередь, строится с помощью регрессии в зависимости от макроэкономических факторов. В данной работе реализовано несколько подходов к получению прогноза экономической функции: использование краткосрочных сценариев макроэкономических факторов, предоставленных экономическими экспертами, и экстраполяция полученной в результате регрессии функции, второй подход – экстраполяция исходных макроэкономических факторов с помощью решения уравнения Орнштейна-Уленбека первого и второго порядков, рекуррентной нейронной сети архитектуры LSTM.

Использование экстраполяций с помощью решений уравнения Орнштейна-Уленбека позволило получить значительно лучший результат, если сравнить с тем, в котором использовались краткосрочные сценарии – 13 % против 28 %. Но лучший результат удалось получить, используя экстраполяции с помощью рекуррентной нейронной сети LSTM – 11 %.

#### Описание модели.

В данной работе мы будем анализировать функции вероятности дефолта и вероятности закрытия, которые можно вычислить по следующим формулам:

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

$$PA(a, t) = \frac{N_{attr}(a, t)}{N_{act}(a, t-1)}, \quad (2)$$

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

Если сравнить жизненный цикл одного кредитного аккаунта с динамикой лечения некоторого пациента, то можно выделить некоторые важные аналогии: задержки оплат по кредиту сходны с осложнениями в период лечения пациента, погашение кредита аналогично полному выздоровлению, а дефолт по кредитному аккаунту – смерти пациента. Данная концепция легла в основу модели APC (Age-Period-Cohort) для аппроксимации вероятности дефолта и закрытия аккаунта, в которой в качестве основных факторов выступают возраст аккаунта (аналогично возрастным особенностям организма), дата открытия кредита (здесь аналогия проводится с особенностями определенных поколений индивидуумов) и влияние макроэкономических факторов (окружающей среды) в каждый из моментов времени существования аккаунта.

Для  $PD$ ,  $PA$  вычисляется APC декомпозиция, которая зависит также от переменной  $v$  – даты открытия кредита:

$$\text{logit}(PD(a, t)) = F_{PD}(a) + H_{PD}(t) + G_{PD}(v), \quad (3)$$

$$\text{logit}(PA(a, t)) = F_{PA}(a) + H_{PA}(t) + G_{PA}(v), \quad (4)$$

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  $F(a)$  – отражает жизненный цикл аккаунтов, так как зависит напрямую от возраста аккаунта,  $H(t)$  – функция окружающей среды, зависящая от макроэкономических факторов,  $G(v)$  отражает влияние неэкономических факторов.

Из кредитных данных из источников [1, 2] мы будем использовать  $F(a)$ , зависящую от возраста кредита функцию (Рисунок 1).

Далее строится экономическая модель для  $H(t)$ :

$$H(t) = \beta_0 + \sum_{i=1}^n \beta_i f_i(t) + \varepsilon(t), \quad (5)$$

где  $f_i(t)$  ( $i = 1, \dots, n$ ) – некоторый набор макроэкономических факторов,  $\beta_i$  – искомые коэффициенты,  $\varepsilon(t)$  – погрешность. Описание этих этапов анализа было приведено в книге [3].

Макроэкономические факторы опубликованы в открытых электронных ресурсах [4, 5].

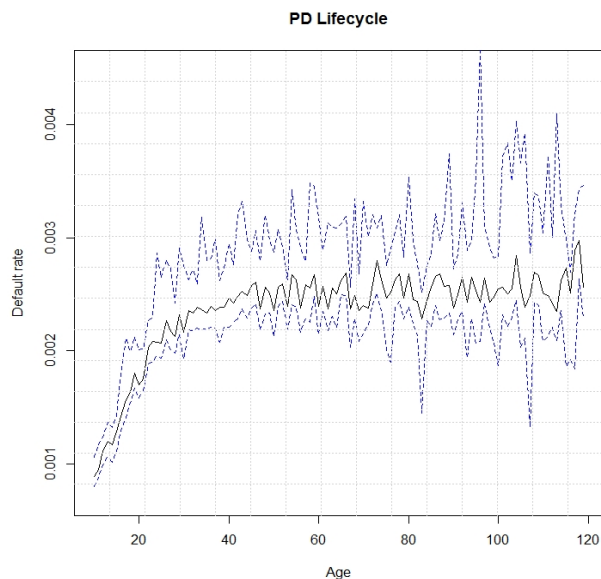


Рисунок 1. F(a)

Приведем графики построенной экономической модели  $H(t)$  для двух случаев – отражающим рейтинг дефолта и закрытия кредитных аккаунтов (Рисунок 2 и 3 соответственно):

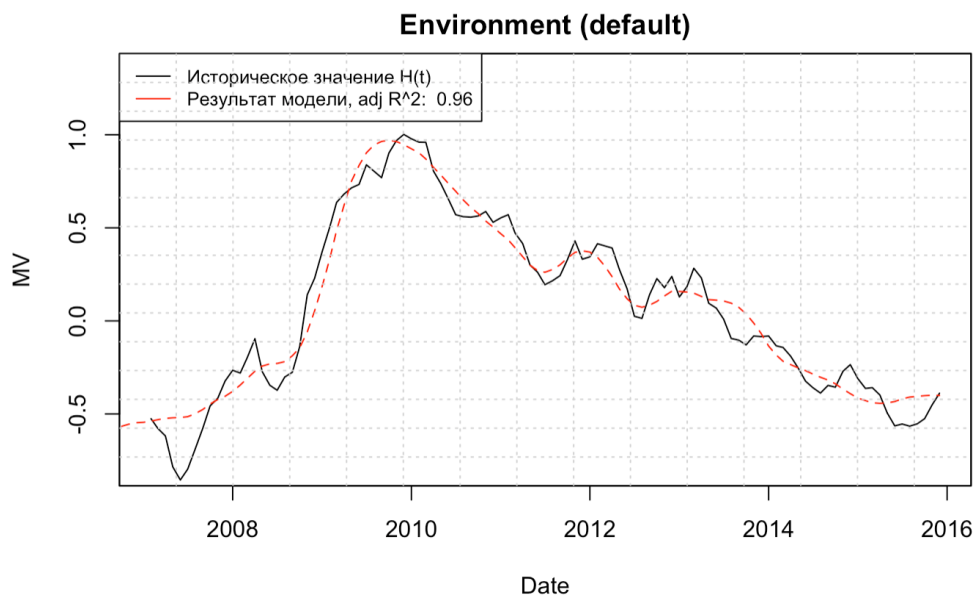


Рисунок 2. Модель для  $H(t)$  для дефолта

На следующем этапе мы строим скоринговую модель DTSM (Discrete Time Survival Model) переменных  $PD, PA$ . Будем считать, что винтажную функцию,  $G(v)$ , можно аппроксимировать категориальной переменной, отражающей год открытия аккаунта, и задержкой выплат по кредиту (в месяцах). Тогда для  $h = 1, 2, \dots$  получаем следующее представление:

$$\begin{aligned} \text{logit}(PD_i(a, t, h)) = & F_{PD}(a) + \widehat{H}_{PD}(t) + b_0(h) + \sum_j b_j(h) \text{Vintage.annual}_{j,i} + \\ & + \sum_j c_j(h) \text{Delinquency}_{j,i} + \varepsilon_i(h), \end{aligned} \quad (6)$$

где  $PD_i(a, t, h)$  – условная вероятность ухода в дефолт  $i$ -ого аккаунта на момент времени  $t$ , который был активным в момент времени  $t - 1$ ; категориальная переменная  $Vintage.annual_j$  отвечает году открытия аккаунта, скоринговый фактор  $Delinquency_j$  обозначает просрочку платежа в месяцах; коэффициенты  $c_j(h)$ ,  $b_j(h)$  являются искомыми коэффициентами скоринговой модели;  $\varepsilon_i(h)$  – погрешность. Аналогично для  $PA$  [6].

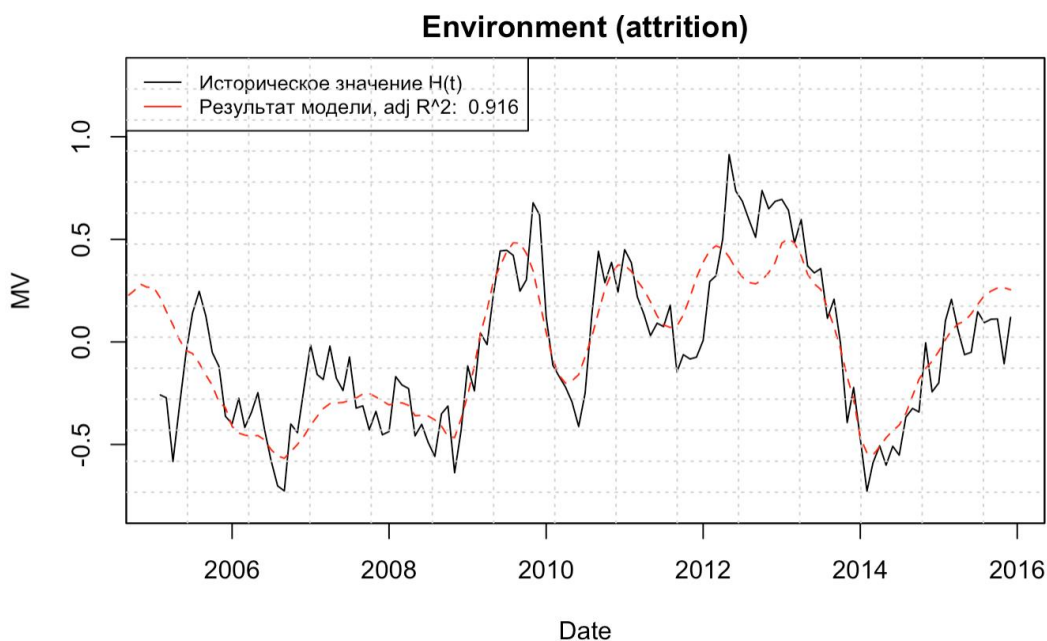


Рисунок 3. Модель для  $H(t)$  для закрытия аккаунтов

Если взять один аккаунт на фиксированный момент времени  $t$ , то на момент времени  $t + 1$  возможно 3 исхода: аккаунт уйдет в дефолт с вероятностью  $PD(t + 1)$ , закроется с вероятностью  $PA(t + 1)$  и останется открытым с вероятностью  $PAct(t + 1, id)$  (при этом будет выплачен платеж за текущий месяц). Аккаунт останется открытым, только при условии, что он был открыт на момент времени  $t$ , а значит вероятность этого события будет равна:

$$PAct(t+1, id) = PAct(t, id) * (1 - PA(t+1, id) - PD(t+1, id)). \quad (7)$$

Таким образом, мы можем рассчитать баланс аккаунта на текущий момент времени с помощью формулы:

$$Balance(t, id) = Balance(t-1, id) - DefaultBalance(t, id) - AttritionBalance(t, id) - PAct(t-1, id) * (1 - PA(t, id) - PD(t, id)) * PrincipalPayment(t), \quad (8)$$

где

$$DefaultBalance(t, id) = PD(t) * Balance(t-1) \quad (9)$$

$$AttritionBalance(t, id) = PA(t) * Balance(t-1). \quad (10)$$

Основной целью является прогнозирование потерь по активным на некоторую дату балансам, для этого служит следующая формула:

$$LossRate = \frac{\sum_{id, t} Default.Balance(id, t)}{\sum_{id} Balance(id)}, \quad (11)$$

где  $t$  изменяется от начала прогноза до конца жизни аккаунта,  $Balance(id)$  – сумма активного баланса на начало прогноза.

### Описание сценариев и их экстраполяция с помощью уравнений Орнштейна-Уленбека.

Макроэкономические сценарии для данной работы были взяты из отчетов компании Consensus Economics, которая предоставляет квартальные сценарии на следующий двухлетний период, поэтому в работе используются реальные экономические предположения, имеющиеся на тот момент. Из представленных факторов для моделирования жизненного цикла аккаунтов ипотечных кредитов были учтены следующие факторы: реальный валовой внутренний продукт, индекс доходности, уровень безработицы и индекс цен на жилье. Эти сценарии представляют средние прогнозы для двадцати четырех выдающихся экономистов или институтов экономического прогнозирования [7]. При построении экономической модели на основании этих макроэкономических факторов для функции среды использована PLS-регрессия вместе обычной регрессии для решения проблемы мультиколлинеарности, так как PLS-регрессия не требует выполнения условия линейной независимости векторов.

Кроме использования макроэкономических сценариев, подход, реализованный в данной работе, – экстраполяция макроэкономических факторов без использования макроэкономических сценариев. Для этого будем использовать решения уравнения Орнштейна-Уленбека первого и второго порядков и рекуррентные нейронные сети архитектуры LSTM.

Рассмотрим одномерное уравнение Орнштейна-Уленбека:

$$dx(t) = \theta(\mu - x(t))dt + \sigma dW(t). \quad (12)$$

Решение уравнения Орнштейна-Уленбека первого порядка.

$$x(t) = e^{-\theta(t-\tau)}x_0 + \mu(1 - e^{-\theta(t-\tau)}) + \sigma \int_{\tau}^t e^{-\theta(t-s)} dW(s) [8]. \quad (13)$$

Так же на основании отмеченного в предыдущем пункте свойства интеграла Ито, математическое ожидание решения будет.

$$E(x(t)) = x_0 e^{-\theta(t-\tau)} + \mu(1 - e^{-\theta(t-\tau)}). \quad (14)$$

Теперь рассмотрим следующую систему уравнений (уравнение Орнштейна-Уленбека второго порядка):

$$\begin{cases} dx(t) = (\theta_1(\mu - x(t)) + v(t)) dt \\ dv(t) = -\theta_2 v(t) dt + \sigma_2 dW(t) + \sigma_1 dB^H(t), \end{cases} \quad (15)$$

где  $B^H(t)$  – дробное броуновское движение с индексом Херста  $H > \frac{1}{2}$ ,

$$x_0 = x(0), v_0 = v(0). \quad (16)$$

Решением данной системы является процесс  $x(t)$ , задаваемый равенством:

$$x(t) = \mu + C_1 e^{-\theta_1 t} + C_2 e^{-\theta_2 t} + \frac{1}{\theta_1 - \theta_2} \int_0^t (e^{-\theta_2(t-s)} - e^{-\theta_1(t-s)}) dM(s), \quad (17)$$

где  $M(t) = \sigma W(t) + \sigma_1 B^H(t)$ , а интеграл в правой части – траекторный интеграл Янга. Математическое ожидание  $x(t)$ :  $E(x(t)) = \mu + C_1 e^{-\theta_1 t} + C_2 e^{-\theta_2 t} \xrightarrow{t \rightarrow \infty} \mu$  [9].

Экстраполяции на основе решения стохастического дифференциального уравнения Орнштейна-Уленбека для каждого макроэкономического фактора осуществлялось отдельно. Так как такие факторы, как реальный располагаемый доход (Real DPI), реальный ВВП (Real GDP), уровень цен на жилье (House Price Index), имели ярко-выраженный тренд, то, прежде чем

применять экстраполяцию к этим временным рядам, необходимо их преобразовать таким образом, чтобы избавиться от тренда. В этой работе к рядам была применена функция LogRatio –  $y(t) = \log\left(\frac{x(t)}{x(t-1)}\right)$ , а к прогнозам, построенным по преобразованному ряду, была применена обратная функция преобразования. В это время экстраполяция фактора уровня безработицы (Unemployment Rate) строилась по исходному ряду. Параметры  $\theta_i$  были найдены путем оптимизации по MSE на последних 24 месяцах, в которые были известны значения реальных показателей, для каждой даты начала прогноза, чтобы смоделировать реальное построения форекаста для текущей даты. Для наглядности, продемонстрируем экстраполяции макроэкономических факторов с помощью решения уравнения Орнштейна-Уленбека 2-ого порядка путем оптимизации параметров  $\theta_i$  (рис. 4.):

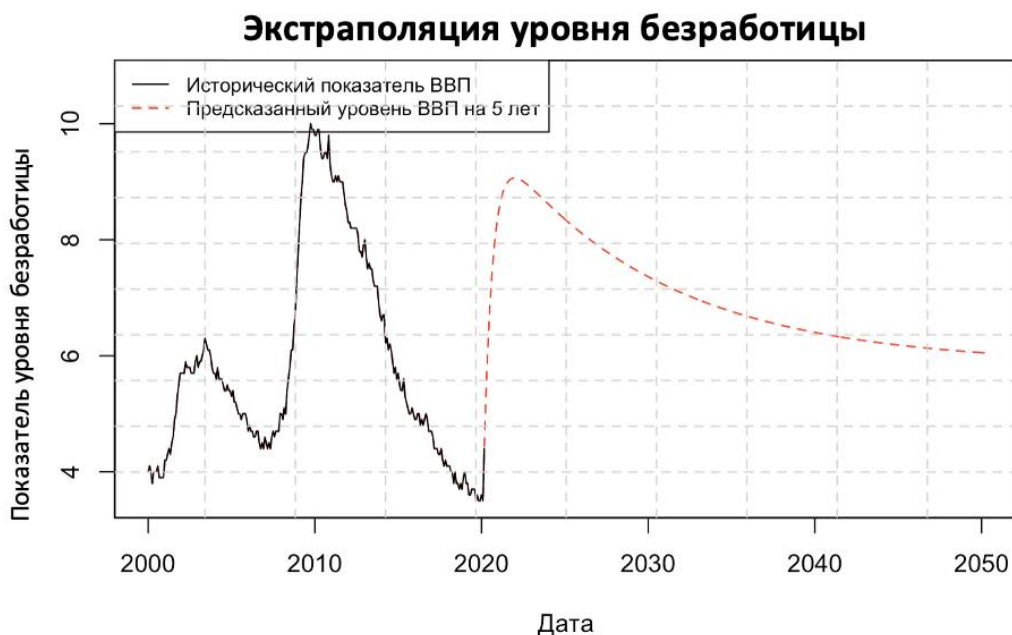


Рисунок 4. Экстраполяция текущего уровня безработицы

Для примера была взята история по безработице в США до марта 2020 года включительно. В данном случае значения  $\theta_i$ , полученные в результате аппроксимации на основе метрики MSE за последние два года, приняли значения  $\theta_1 = 0.101$  и  $\theta_2 = 1.873$ . Для наглядности приведем так же прогноз, начиная с какой-то даты в прошлом, чтобы можно было сравнить с реальной историей. Возьмем, например, апрель 2017 в качестве начала экстраполяции (рис. 5.):

В данном случае значения  $\theta_i$  получились совершенно иными:  $\theta_1 = 0.102$ ,  $\theta_2 = 1.253$ . Как мы видим, действительно скорость стремления к среднему значению намного меньше, чем в случае, представленном на рисунке 4 (минимальный из параметров задает скорость сходимости к среднему значению). Также стоит отметить, что экстраполяция сильно зависит от двух последних значений временного ряда, которые задают направление этой самой экстраполяции.

Продemonстрируем также экстраполяцию, построенную в результате обучения нейронной сети LSTM (архитектура данной сети описана в источнике [10], реализация взята стандартная, включенная в пакете sklearn языка Python) (рисунок 6.).

Как видим, результат экстраполяции с помощью рекуррентной нейронной сети с архитектурой LSTM (долгой краткосрочной памяти), включает больше колебаний, но не сходится к среднему.

Рассмотрим также, как строятся экстраполяции, учитывая тренд. Возьмем экстраполяцию с помощью решения уравнения Орнштейна-Уленбека второго порядка для показателя текущего ВВП. Построим экстраполяцию по преобразованному фактору, полученному после применения функции LogRatio, к которой потом была применена обратная функция для возвращения к исходному ряду (рисунок 7.):

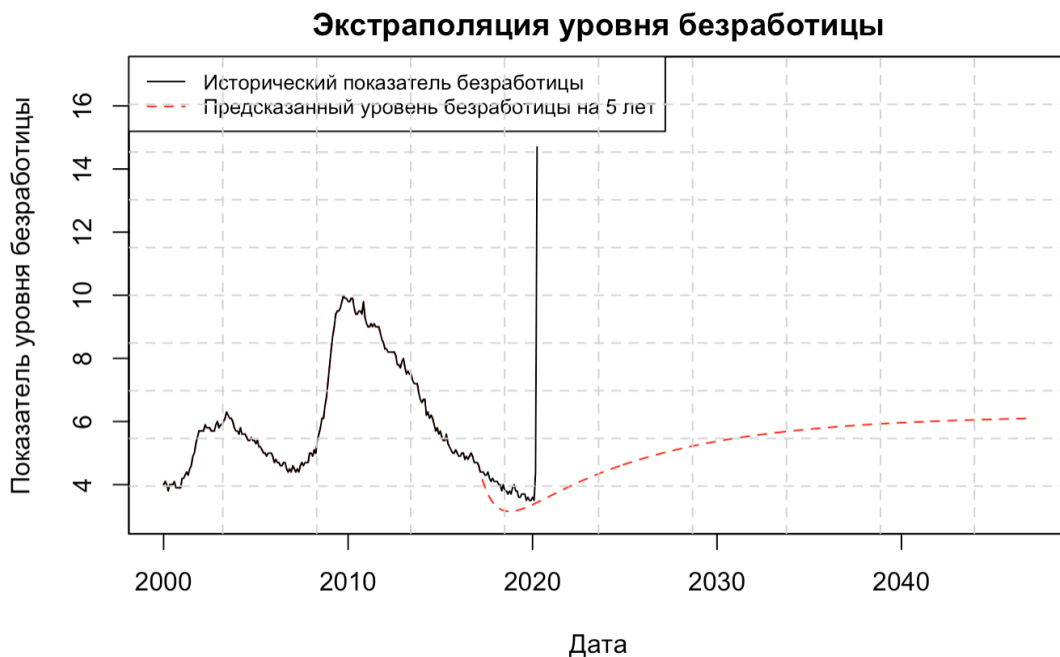


Рисунок 5. Экстраполяция уровня безработицы, начиная с апреля 2017 года



Рисунок 6. Экстраполяция текущего уровня безработицы, используя модель LSTM

Как и в случае с безработицей, построим прогноз, взяв апрель 2017 в качестве начала прогноза (рисунок 8.).

Значения  $\theta_i$  при построении этой экстраполяции:  $\theta_1 = 0.987$ ,  $\theta_2 = 2.99$ . Учитывая, что при оптимизации были выбраны пороги для максимальных возможных значений в 1 и 3 соответственно, такие параметры могут свидетельствовать о том, что при применении функции LogRatio и экстраполяции полученного ряда, функция стремится сойтись к среднему с максимальной скоростью.

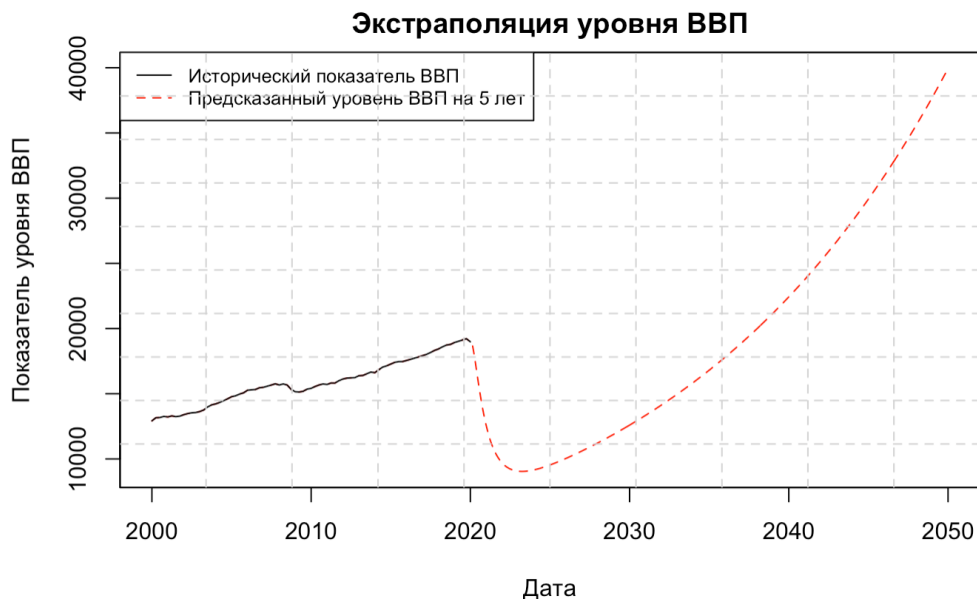


Рисунок 7. Экстраполяция текущего уровня ВВП с помощью решения уравнения Орнштейна-Уленбека 2-ого порядка

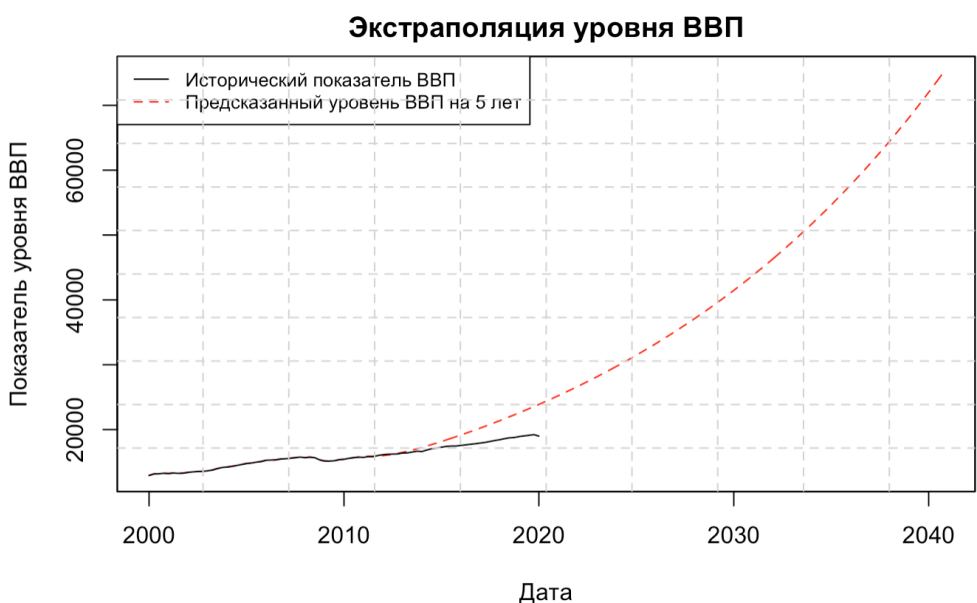


Рисунок 8. Экстраполяция уровня ВВП с апреля 2017 года

### Практические результаты.

Коэффициенты обобщенной линейной регрессии, полученные в результате построения скоринга (для каждого значения горизонта была построена линейная регрессия для значений вероятности дефолта). Продемонстрируем, что они действительно сходятся с изменением горизонта, что позволяет при горизонте  $>12$  (то есть для моментов времени, большим начала прогноза более, чем на год) использовать коэффициенты регрессии для горизонта = 12 (рис. 9.):

Далее, используя полученную модель скоринга, мы построили прогноз для каждого аккаунта, пример для одного аккаунта представлен ниже (прогноз строился для сценария макроэкономических факторов на последний квартал 2011 года) (таблица 1.).

Целью построения прогноза на какую-то конкретную дату является получение одного численного значения – значения Loss Rate – потерь по текущим открытым кредитам. Для удобства сравнения прогнозов, вычислим также прогноз потерь и исторические потери в течение первых 12 месяцев.



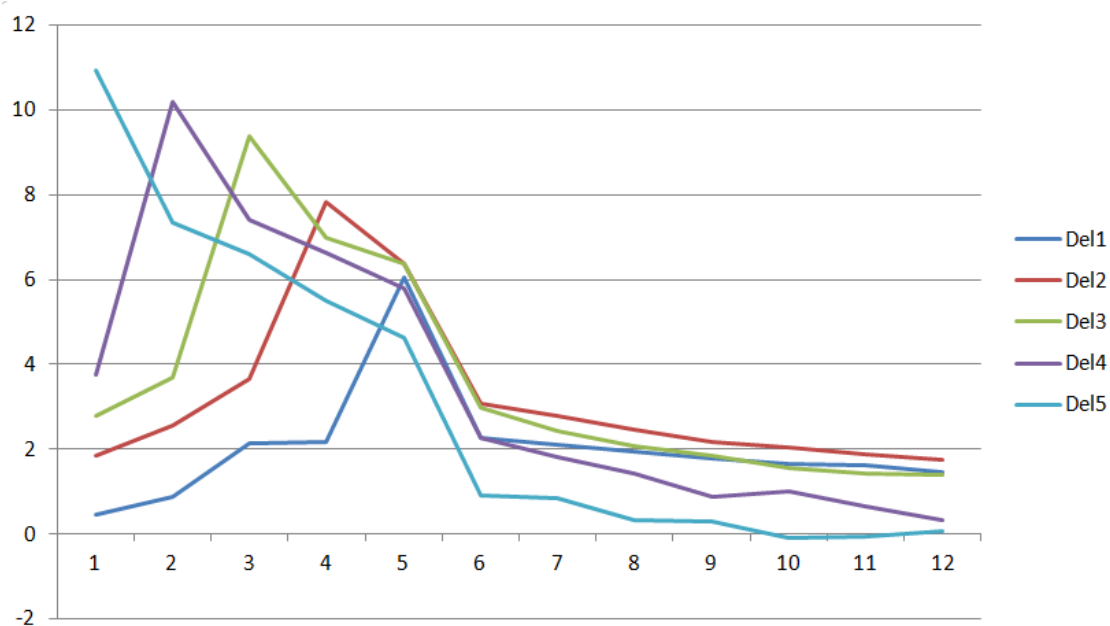


Рисунок 9. Коэффициенты регрессии при переменной месячной задержки выплат

Таблица 1. Прогноз для одного аккаунта

	Date	Vintage	Orig. Balance	Balance	a	PD	PA	PAct	Default. Balance	Attrition Balance	Principal Payment
1	Sep-11	Jan-99	82000	64323.54	152	0.00E+00	0	1	0	0	161.2614
2	Oct-11	Jan-99	82000	63172.34	153	4.42E-05	0.01538	0.98	2.8418	989.582	161.2614
3	Nov-11	Jan-99	82000	61941.4	154	3.33E-05	0.01698	0.97	2.1062	1072.765	162.2189
4	Dec-11	Jan-99	82000	60090.93	155	4.09E-05	0.02737	0.94	2.5323	1695.239	163.182
...											
194	Oct-27	Jan-99	82000	580.32	345	6.89E-04	0.01196	0.09	0.435	7.553	502.5154
195	Nov-27	Jan-99	82000	1010.36	346	6.89E-04	0.01196	0.09	0.3997	6.941	505.499
196	Dec-27	Jan-99	82000	0	347	6.89E-04	0.01196	0.08	0.365	6.34	508.5003

Таким образом, просуммировав дефолтный баланс и разделив его на сумму активных балансов на дату начала прогноза по каждому из полученных прогнозов, получили следующие прогнозы (здесь для каждой из дат, когда она выбирается в качестве начала прогноза, посчитаны ожидаемые потери за все время, а также за первые 12 месяцев) (таблица 2.):

Колонка Loss. Rate. All отвечает предсказанной доле потерь по всем активным аккаунтам до конца их срока, Loss. Rate.12 – предсказанной доле потерь за первые 12 месяцев с даты начала прогноза, а Loss. Rate.12. Hist – за исторические потери за первые 12 месяцев с даты начала прогноза.

Аналогичный вид (как в таблице 1 и 2) имели и прогнозы, построенные по макроэкономическим факторам, сэкстраполированным с помощью решений уравнения Орнштейна-Уленбека разных порядков.

Продemonстрируем на графике сравнение:

- исторических потерь;
- потерь, спрогнозированных с использованием краткосрочных сценариев макроэкономических факторов, а далее экстраполяции полученной функции окружающей среды решением уравнения Орнштейна-Уленбека 2-го порядка;
- потерь, спрогнозированных без использования сценариев, строя экстраполяции каждого из макроэкономических факторов с помощью решения уравнения Орнштейна-Уленбека 2-го порядка, а затем уже по этим экстраполяциям была построена функция окружающей среды;
- потерь, построенных аналогично, без использования сценариев, но для продления факторов используя решения уравнения Орнштейна-Уленбека 1-ого порядка.

Таблица 2. Прогноз кредитных потерь

	SnapDate	Loss.Rate.All	Loss.Rate.12	Loss.Rate.12.Hist
1	Jul-07	0.173469	0.01463	0.017128
2	Oct-07	0.184527	0.016691	0.020124
3	Jan-08	0.290248	0.022199	0.024409
4	Apr-08	0.2273	0.019584	0.032812
5	Jul-08	0.235889	0.019577	0.044
6	Oct-08	0.25384	0.01998	0.061873
7	Jan-09	0.268172	0.022286	0.076566
8	Apr-09	0.272447	0.02982	0.091217
9	Jul-09	0.284885	0.060389	0.094526
10	Oct-09	0.28939	0.110335	0.088584
11	Jan-10	0.284566	0.140264	0.080393
12	Apr-10	0.253644	0.128989	0.068391
13	Jul-10	0.22456	0.102072	0.063354
14	Oct-10	0.208251	0.084455	0.061242
15	Jan-11	0.201589	0.075604	0.058858
16	Apr-11	0.187124	0.06624	0.055169
17	Jul-11	0.17539	0.055001	0.054388
18	Oct-11	0.16908	0.047052	0.049553
19	Jan-12	0.16996	0.043676	0.048718
20	Apr-12	0.160341	0.04114	0.048221
21	Jul-12	0.15693	0.03803	0.045089
22	Oct-12	0.160226	0.04134	0.04593
23	Jan-13	0.157966	0.03936	0.040827
24	Apr-13	0.154608	0.036205	0.036497
25	Jul-13	0.151062	0.031088	0.030383
26	Oct-13	0.152233	0.029113	0.024956
27	Jan-14	0.151663	0.02587	0.024167
28	Apr-14	0.147419	0.022716	0.022137
29	Jul-14	0.14423	0.020376	0.02184
30	Oct-14	0.148432	0.020729	0.020257

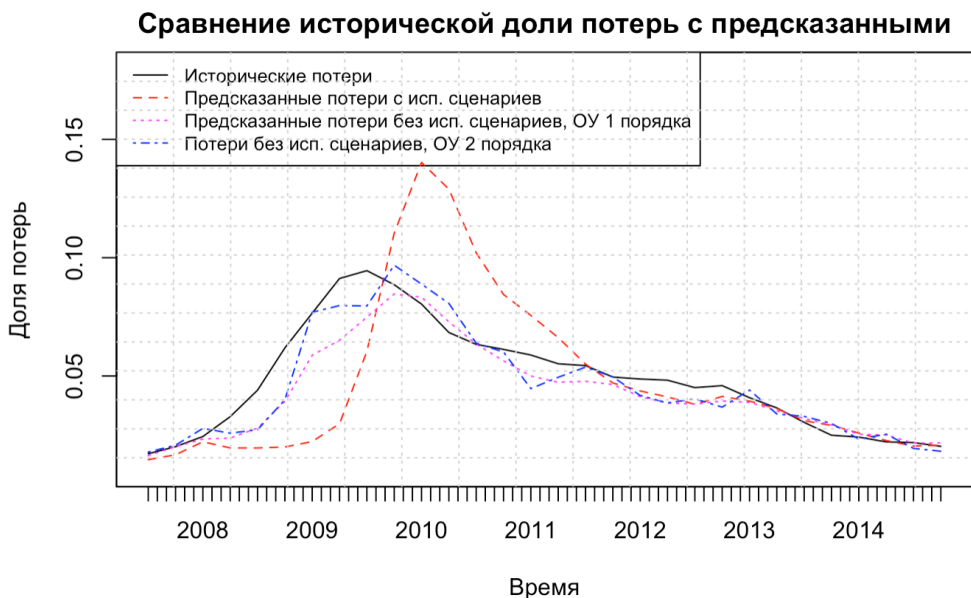


Рисунок 10. Историческая и спрогнозированные доли потерь

Отдельно сравним результаты, полученные с помощью экстраполяции макроэкономических факторов на основе модели LSTM (рекуррентной нейронной сети) (рис. 11.).

### Сравнение доли потерь с предсказанными по ОУ-2, LSTM

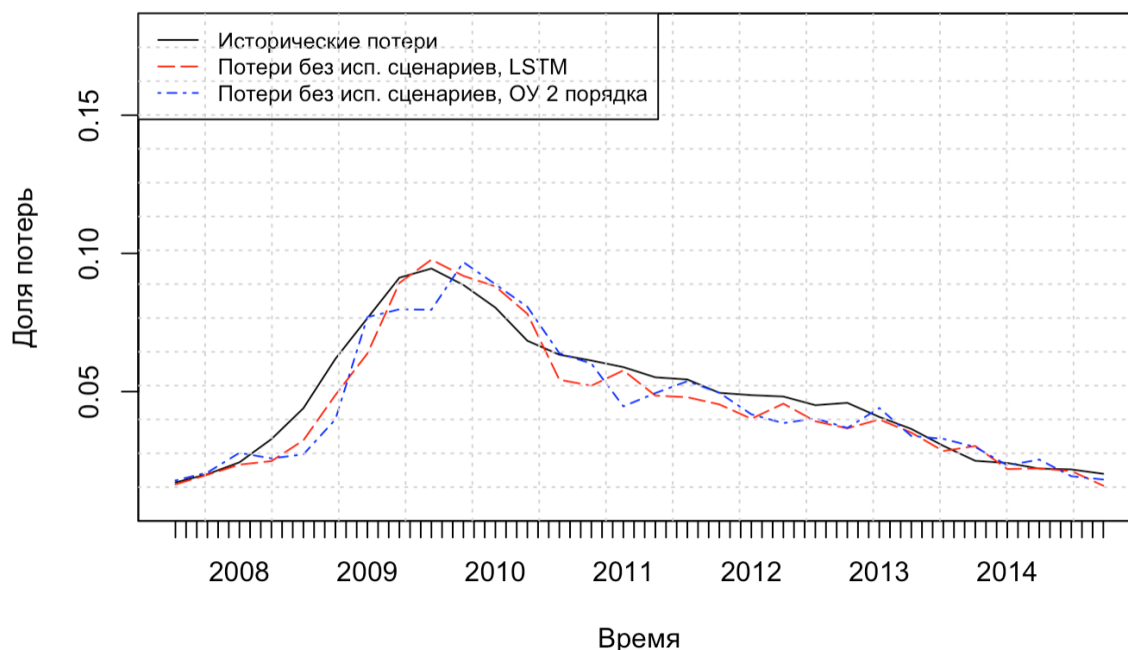


Рисунок 11. Сравнение с прогнозами, основанными на использовании LSTM

Как можно видеть на рисунке 10, на котором представлено сравнение основных результатов, построенная модель SECL в целом дает приемлемый результат с погрешностью около 5 % в условиях кризиса. При этом использование поквартальных сценариев для построения предсказаний позволяет получить оценку потерь по открытым кредитным аккаунтам в условия кризиса, во-первых, с задержкой около года, а во-вторых, с наибольшей погрешностью (пик оценки примерно на 5 % больше пика исторических потерь).

Приведем значения метрики  $\frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i}{y_i} - 1 \right| * 100$  – средней относительной ошибки (домноженной на 100 для получения средней ошибки в процентах) – для прогнозов.

Значение метрики для прогноза, построенного на основе экстраполяции факторов решением уравнения Орнштейна-Уленбека 1-ого порядка – 12.5053 %. Значение метрики для прогноза, построенного с помощью решения уравнения 2-ого порядка – 12.28091 %.

Как мы видим, среднее значение разности квадратов предсказаний и исторических потерь ниже для прогнозов, построенных с помощью экстраполяций 2-ого порядка.

Также приведем значение средней относительной ошибки для прогнозов, построенных с помощью краткосрочных сценариев: 27.08764 %. Разница такого прогноза и прогноза, построенного на экстраполяции макроэкономических факторов – около одного порядка, что позволяет сделать вывод, что модель SECL, построенная на основе экстраполяции макроэкономических факторов решением уравнения Орнштейна-Уленбека, позволяет построить более точный прогноз ожидаемых потерь по кредитам.

На рисунке 11 приведено сравнение прогнозов, построенных на основе экстраполяции макроэкономических факторов с помощью решения стохастического уравнения Орнштейна-Уленбека 2-го порядка, а также с помощью рекуррентной нейронной сети LSTM. На основании метрики – для прогнозов в первом случае она равна 12.29 %, а во втором – 10.54 % – можно сделать вывод, что экстраполяция на основе рекуррентной нейронной сети с архитектурой LSTM (долгой краткосрочной памяти) позволяет получить более точную аппроксимацию для ожидаемых потерь по кредитным аккаунтам. Также нужно отметить важное преимущество этой модели – в условиях экономического кризиса пик потерь предсказан максимально близко к историческому, что может позволить банкам прогнозировать ожидаемые потери точнее.

### Список литературы

- [1] Кредитные данные США . [Электр. ресурс]: – <http://www.freddiemac.com/>.
- [2] Кредитные данные США . [Электр. ресурс]: – <http://www.fanniemae.com/>.
- [3] Breeden, J. L. Living with CECL: Mortgage Modeling Alternatives. – Prescient Models LLC, 2018. – 203 с.
- [4] Источник баз данных с макроэкономическими факторами за последние 30 лет [1]. [Электр. ресурс]: – <https://fred.stlouisfed.org/>.
- [5] Источник баз данных с макроэкономическими факторами за последние 30 лет [1]. [Электр. ресурс]: – <https://www.federalreserve.gov/newsevents/pressreleases/bcreg20180201a.htm>.
- [6] Леваков, А. А., Васьковский, М. М. Стохастические дифференциальные уравнения и включения. – Минск: БГУ, 2019. – 496 с.
- [7] 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.
- [8] Оксендаль, Б. Стохастические дифференциальные уравнения. Введение в теорию и приложения: Пер. с англ. – М.: Мир, ООО «Издательство АСТ», 2003. – 408 с.
- [9] Vaskouski, M. Applications of second order Ornstein Uhlenbeck stochastic processes to credit risk modeling. // Big Data and Advanced Analytics. – Minsk, 2020. – Part 1. – P. 105-111.
- [10] Hochreiter, S. Long Short-term Memory // Neural Computation. – 1997. – Vol.9, № 8 – 33 с.

## INFLUENCE OF EXTRAPOLATION METHODS OF MACROECONOMIC FACTORS ON THE ACCURACY OF FORECASTING CREDIT RISKS

**M. VASKOUSKI**

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

**A. KARPOVICH**

*Master's student at Belarusian State University;  
Data Scientist at EPAM Systems*

*Department of Higher Mathematics  
Belarusian State University, Republic of Belarus  
HiQo Solutions, Ltd, USA  
E-mail: [vaskovskii@bsu.by](mailto:vaskovskii@bsu.by), [maxim.vaskouski@hiqo-solutions.com](mailto:maxim.vaskouski@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 Sciences Physics and Mathematics. I am an expert in theory of stochastic differential equations.*

**A. Karpovich**

*Master's student at Belarusian State University, speciality – Algorithms and systems for processing big data. Works as Data Scientist at EPAM Systems.*

**Abstract.** This paper examines the influence of the method of extrapolating macroeconomic factors on the accuracy of forecasting credit losses within the DTSM model. When constructing a forecast of credit losses, various approaches were applied for extrapolation of economic factor - using short-term scenarios, extrapolations using the solution of the Ornstein-Uhlenbeck equation of the first and second orders, a recurrent neural network of the LSTM architecture. The aim of this work was to compare the accuracy of the calculated losses. The use of extrapolations using solutions to the Ornstein-Uhlenbeck equation yielded a significantly better result when compared with the one in which short-term scenarios were used - 13% versus 28%. The best result was obtained using extrapolations using the recurrent neural network LSTM - 11%.

**Keywords:** credit losses, extrapolation of economic series, regression model, stochastic differential equations, recurrent neural networks.