

DOI <https://doi.org/10.15407/usim.2018.02.087>

УДК 004.942

Н.В. КУЗНЕЦОВА, канд. техн. наук, доцент,
natalia-kpi@ukr.net

П.И. БИДЮК, д-р техн. наук, профессор,
pbidyuke_00@ukr.net

Кафедра математических методов системного анализа института прикладного системного анализа (ИПСА) Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского»

ВЫЯВЛЕНИЕ РИСКОВ МОШЕННИЧЕСТВА В ХОДЕ ТЕНДЕРНЫХ ЗАКУПОК МЕТОДАМИ ТЕОРИИ ВЫЖИВАНИЯ

Исследованы возможности анализа тендерных заявок на платформе ProZorro для определения мошенничества или сговора. Предложено построить модели выживания на основе пропорциональных рисков Кокса и оценки Каплан–Майера, которые позволили определить поведение компаний, подозреваемых в сговоре.

Ключевые слова: тендерные закупки, риски мошенничества, модель пропорциональных рисков Кокса, оценка Каплан–Майера

Введение

Борьба с коррупцией стала одним из основных способов сохранения бюджета стран постсоветского пространства. Непоступления в бюджет, завышение стоимости работ на госзаказы, строительство и ремонты втрисорога — яркий пример несовершенства бюджетного государственного регулирования. Много усилий приложено в Украине для разработки законопроекта «О публичных закупках», принятого в 2016 г., а позже многократно модифицированного и доработанного. Этим законом предполагается, что все договора об услугах и товарах на сумму более 50 тыс. гривень должны осуществляться с использованием электронной системы закупок с целью отбора поставщика товара, услуг и исполнителя работ для заключения договора; заказчики должны

придерживаться принципов осуществления установленных публичных закупок [1]. Однако в украинских реалиях разработка даже самого совершенного закона еще не означает его внедрения и прозрачной работы. Множество искусственных сложностей было на этапе внедрения, но все же была создана специальная электронная платформа «ProZorro» для регулирования и свободного публичного проведения торгов онлайн [2]. Для работы этой системы предполагалась разработка и применение методов интеллектуального анализа данных [3, 4] для проверки и выявления мошеннических действий и, соответственно, автоматического блокирования таких недобросовестных закупок. Статья — о выявлении рисков мошенничества в ходе тендерных закупок с использованием методов теории выживания.

Постановка задачи

Разработать новые поведенческие модели для анализа заявок и закупок на электронной платформе *ProZorro* с целью обнаружения и исследования поведения компаний, предпринимающих мошеннические действия для получения госзаказов. Сравнить предложенные модели с известными и выработать рекомендации для усовершенствования системы и выявления недобросовестной конкуренции.

Анализ существующих подходов к выявлению мошеннических действий на платформе *ProZorro* и постановка задачи моделирования

Обнаружение мошенничества заключается в том, чтобы как можно быстрее выявлять случаи сговора среди многочисленных договоров, заключаемых на электронной платформе. Для решения задачи выявления таких действий в системе реализованы процедуры автоматической классификации тендерных заявок. В целях безопасности ни алгоритмы, ни характеристики множества хранимых на платформе методов, используемых для классификации тендерных заявок, не публикуются в открытом доступе. При этом множество публикаций в СМИ и заявлений общественных организаций о необходимости проведения мониторинга торгов на предмет обнаружения сговора в установленном законом порядке говорит о том, что существующие в системе алгоритмы не справляются с задачей обнаружения мошенничества и необходимы новые приемы и методы для автоматического определения подозрительных заявок.

Одним из самых распространенных способов сговора при проведении публичных закупок является регистрация на платформе так называемых фирм-однодневок, которые подают заявку на участие в тендере, а затем снимаются с торгов в пользу другой компании. Другим способом мошенничества является существование «фирм-марионеток», участвующих в торгах исключительно с определенными

компаниями, состоящими в сговоре для создания видимости конкуренции. При этом стоимость работ и ценовая политика для тендера явно завышена или согласована с другой компанией для обеспечения ей выигрыша. Еще один способ мошенничества — парное участие в торгах одних и тех же компаний, созданных одними и теми же акционерами или их окружением, и соответственно очередное выигрывание тендеров одними и теми же лицами в фирмах с разными названиями.

Для проверки подобных факторов выберем характеристики, доступные на платформе, и влияющие, по мнению авторов, на подтверждение факта мошенничества. Выдвинем гипотезу, что длительность пребывания компании, количество случаев ее участия в торгах и снятия с торгов может быть подтверждением ее манипулятивных действий в системе торгов. Отметим, что из выборки отфильтрованы компании, участвовавшие в конкурсе на платформе менее трех раз, поскольку будем классифицировать именно те компании — постоянные участники тендеров, подозреваемые в сговоре, для выявления этих групп компаний и подачи соответствующих документов на проверку соответствующим контролирующим органам. Проверено так называемое правое цензурирование и отсеяны случаи, которые прекращаются до наступления события [5].

Эмпирическим путем сформирована следующая выборка данных:

- *Wins* — количество выигрышей определенной компании;
- *Losses* — количество проигрышей компании в торгах;
- *Sum_of_deals* — общая сумма выигранных торгов;
- *Participations* — общее количество участия в торгах;
- *Objections* — количество жалоб, поданных данной компанией;
- *Date_start* — дата начала участия в системе торгов;
- *Date_finish* — дата последнего участия в торгах;

- *IdTenderer* — уникальный номер участника тендера;

- *Suspected* — переменная, показавшая, была ли фирма заподозрена в неправомερных сговорах с другими участниками;

- *Churn out* — целевая переменная, равная единице, если компания прекращала участие в торгах в короткий срок, т.е. фиктивная компания, для одного торга.

Компания считалась продолжающей торги с промежутком времени между началом торгов на платформе и временем последнего торга более 60 дней (статистическая средняя продолжительность бизнес-циклов данных компаний на платформе, полученная по официальным данным *ProZorro* [2]).

При решении задачи классификации с использованием нейронных сетей, деревьев решений, логистической регрессии, байесовского классификатора [3, 4, 6] ошибка классификации составляла около 30 процентов, что свидетельствует о недостаточно высокой точности функционирования системы для автоматического выявления мошенничества.

В данной статье предложен подход к анализу поведения компаний, участвующих в торгах, с целью обнаружения определенных закономерностей с учетом фактора времени, что позволит отличить компании-марионетки от реальных участников торгов.

Формализация основных понятий теории выживания для задачи оценки дальнейшей работы компании на платформе

Для моделирования вероятности наступления риска (P) на выборке из n независимых одинаково распределенных величин $\{(Y_1, X_1, \delta_1), \dots, (Y_n, X_n, \delta_n)\}$, случайного вектора $\{Y, X, \delta\}$, где $Y = \min\{T, C\}$ — наблюдаемая зрелость, т.е. период созревания (обслуживания) компании на платформе *ProZorro*, T — время от начала наблюдения до наступления риска, C — время до окончания исследования, $\delta = (T \leq C)$ — индикатор нецензурирования и X — вектор поясняющих ковариант, предполагается незави-

симость величин T и X , а также условная независимость величин T и C для данного X [5].

Тогда условное распределение случайной величины T можно охарактеризовать, используя некоторые общие соотношения в анализе выживаемости. Пусть функция условной выживаемости, $S(t|x)$, условный уровень опасности, $\lambda(t|x)$, условная кумулятивная функция риска, $\Lambda(t|x)$, и условная функция распределения $F(t|x)$, связаны так [8]:

$$S(t|x) = P(T > t | X = x) = \int_t^{\infty} f(u|x) du; \quad (1)$$

$$\lambda(t|x) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t, X = x)}{\Delta t} = \frac{f(t|x)}{S(t|x)}; \quad (2)$$

$$\Lambda(t|x) = \int_0^t \lambda(u|x) du = \int_0^t \frac{f(t|x)}{S(t|x)} du; \quad (3)$$

$$S(t|x) = e^{-\Lambda(t|x)}; \quad (4)$$

$$F(t|x) = 1 - S(t|x), \quad (5)$$

где $f(u|x)$ — функция плотности вероятности. Поскольку время — непрерывная величина, вероятность того, что событие произойдет во время t равна нулю, но можно говорить о вероятности того, что событие произойдет в малый интервал $[t; t + \Delta t]$.

Тогда вероятность наступления риска P во времени t можно оценить как посредством моделей логистической регрессии, так и модели пропорциональных рисков Девида Кокса [7], когда оценка функции выживаемости получена путем решения уравнений частичного правдоподобия. Регрессионная модель Кокса дает оценку вероятности наступления события на основе модели пропорциональных рисков (*proportional hazards model*) \hat{P}^{PHM} с использованием обобщенной линейной модели (*generalized linear model*) \hat{P}^{GLM} с параметрами, оцененными методом максимального правдоподобия.

Пропорциональные риски Кокса

Модель Кокса [7] основана на предположении, что функцию условного риска (или функция

интенсивностей отказов, на техническом сленге), можно факторизировать, т.е. представить в виде произведения двух функций:

$$\hat{\lambda}(t|x) = \hat{\lambda}_0(t) \exp(x^T \hat{\beta}), \quad (6)$$

где $\hat{\lambda}_0(t)$ — оценка базовой функции уровня риска $\lambda_0(t)$, а $\hat{\beta}$ представляет собой оценку вектора параметров β .

Таким образом, в предположении о существовании модели пропорциональных рисков, P оценивается так:

$$\begin{aligned} \hat{P}^{PHM}(t|x) &= \frac{\hat{F}_{\hat{\beta}}(t+b|x) - \hat{F}_{\hat{\beta}}(t|x)}{1 - \hat{F}_{\hat{\beta}}(t|x)} = \\ &= 1 - \frac{\hat{S}_{\hat{\beta}}(t+b|x)}{\hat{S}_{\hat{\beta}}(t|x)}, \end{aligned} \quad (7)$$

где

$$1 - \hat{F}_{\hat{\beta}}(t|x) = \hat{S}_{\hat{\beta}}(t|x) = \exp(-\hat{\Lambda}(t|x)). \quad (8)$$

Метод оценивания для этой модели состоит из двух шагов [5]. На первом шаге интегральная функция базового риска $\Lambda_0(t)$ оценивается так:

$$\hat{\Lambda}_0(t) = \frac{\sum_{i=1}^n 1\{Y_i \leq t, \delta_i = 1\}}{\sum_{j=1}^n 1\{Y_j \geq Y_i\}}. \quad (9)$$

Тогда параметр $\hat{\beta}^{PHM}$ определяется по формуле

$$\hat{\beta}^{PHM} = \arg \max_{\beta} L(\beta), \quad (10)$$

где частичная функция правдоподобности задается выражением

$$L(\beta) = \prod_{i=1}^n \frac{\exp(x_i^T \beta)}{\left(\sum_{j=1}^n 1\{Y_j > Y_i\} \exp(x_j^T \beta) \right)}. \quad (11)$$

Таким образом, оценка условной интегральной функции риска вычисляется по формуле

$$\hat{\Lambda}(t|x) = \int_0^t \hat{\lambda}(s|t) ds = \exp(x^T \hat{\beta}^{PHM}) \hat{\Lambda}_0(t). \quad (12)$$

Функция риска отражает интенсивность, с которой состояние в течение времени t , стремится к прекращению. Условная интегральная функция риска в точке $t|x$ получается сложением

значений функции риска для всех возможных моментов прекращения, меньших t .

Оценка Каплан–Мейера

Эта оценка Каплан–Мейера (КМ) служит обобщением эмпирической функции выживания и учитывает цензурированные наблюдения. Формула КМ для вероятности выживания в определенное время ограничивается произведением характеристик, соответствующих индивидам, оставшимся в живых после времени t_i [8].

Для расчета оценки КМ все наблюдения сортируются в порядке возрастания времени их жизни. Первое вхождение начинается с нуля. Вероятность выживания к этому времени равна единице. Дальнейшие наблюдения исключаются в момент времени их гибели (возможно, в результате цензуры). Множество под риском, которое обозначается $R(t_i)$, — это количество всех индивидов, доживших хотя бы до времени t_i .

Основная идея расчета представляется формулой [5]

$$\hat{S}(t_j) = \prod_{i=1}^j \hat{P}(T > t_i | T \geq t_i), \quad (13)$$

с учетом того, что

$$\hat{P}(T > t_i | T \geq t_i) = \frac{n_i - d_i}{n_i}, \quad (14)$$

где n_i — количество наблюдений в множестве риска ($n_i = |R(t_i)|$); d_i — количество субъектов, погибших в момент t_i ($d_i = |D_i|$).

Поскольку визуальное сравнение кривых выживания для различных видов не всегда удобно, то предлагается использовать такие статистические критерии [9]:

$$\text{LogRank} = \frac{\left(\sum_{j=1}^r (d_{1j} - e_{1j}) \right)^2}{\text{var} \left(\sum_{j=1}^r (d_{1j} - e_{1j}) \right)}, \quad (15)$$

где e_{1j} — количество случаев первого вида на j -м временном интервале; r — количество временных промежутков. Этот критерий подобен χ^2 и используется, когда функции рис-

ка одинаковы, а уровень риска — константа во времени.

$$Wilcoxon = \frac{\left(\sum_{j=1}^r n_j (d_{1j} - e_{1j}) \right)^2}{\text{var} \left(\sum_{j=1}^r n_j (d_{1j} - e_{1j}) \right)} \quad (16)$$

этот критерий чувствителен к образцам, которые цензурируются. Его рекомендуется использовать в ранних периодах исследования.

Моделирование участников тендеров моделью Каплан–Майера

Используем некоторые предположения из теории выживания для моделирования нашей задачи в контексте участия и продолжительности пребывания компании на платформе *ProZorro* в условиях конкуренции, выживания и выигрыша в сравнении с другими компаниями.

Непараметрическая модель КМ позволяет оценить изменение функция выживания (т.е. дальнейшего участия компании в тендерах в следующие моменты) со временем. Поскольку функция выживания остается на высоком уровне, то можно сделать вывод, что больше половины компаний остаются в системе торгов на период более 60 дней (рис. 1).

Как видно из графика функции риска (рис. 2), наибольшая угроза выхода компаний — в первые 20 дней и после 40 дня (в среднем соответствует одному–двум торгам в этот период). Для ответа на ключевой вопрос исследования также были построены функции выживания для двух групп — подозреваемых в сговоре фирм (*suspected* = 1) и тех, которые не были заподозрены. Количественное распределение фирм по стратам (группам компаний) приведено в табл. 1, а в табл. 2 показаны значения статистических критериев для сравнения: *WilCoxon* и *Logrank*.

Как видим, согласно построенной модели, компании, не заподозренные в сговорах, в среднем имеют более высокую вероятность «выживания на платформе», т.е. они в меньшем количестве покидают площадку для торгов (рис. 3). Этот печальный факт подтвер-

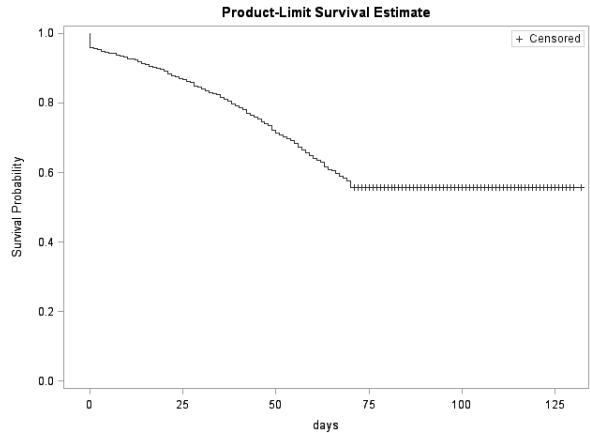


Рис. 1. График вероятности выживания в целом компаний на платформе *ProZorro*, (*Survival Probability* — вероятность выживания, *days* — количество дней)

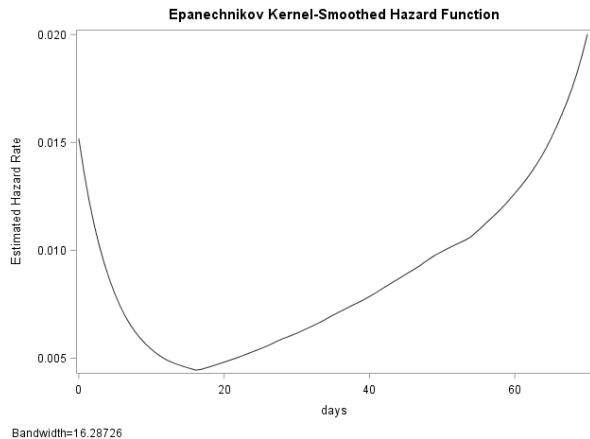


Рис. 2. График функции риска (*Estimated Hazard Rate* — оценочная степень риска)

ждает недееспособность текущей системы обнаружения сговоров на площадке.

На графиках функций риска, построенных для страт подозрительных и нормально классифицированных компаний (рис. 4), вероятность ухода подозреваемой компании с торгов снижается после 20-дневного периода, и она в три–четыре раза ниже, чем для обычной компании.

Далее проведен анализ компаний, участвующих в торгах по маркеру дисквалификации. Так, в выборке присутствуют компании, которые не были дисквалифицированы ни разу (значение *suspected* < 1), были дисквалифицированы дважды из-за неправильно оформлен-

Таблица 1. Количественное соотношение цензурированных и нецензурированных данных

Summary of the Number of Censored and Uncensored Values					
Страта	Подозрительные	Всего	Выбывшие	Цензурированные	Процент цензурированных
1	0	3872	1725	2147	55,45
2	1	94	32	62	65,96
Всего		3966	1757	2209	55,70

Таблица 2. Сравнение статистических критериев по стратам функций выживания

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	3,4498	1	0,0633
Wilcoxon	3,0405	1	0,0812
-2Log(LR)	5,7826	1	0,0162

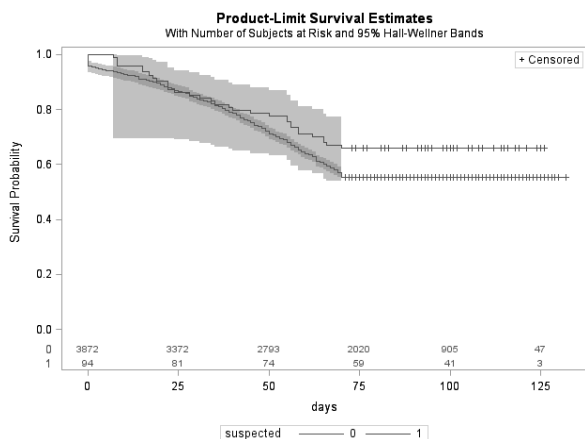


Рис. 3. График функции выживания по стратам (ноль — нормальные компании, единица — заподозренные в мошенничестве (suspected), Survival Probability — функция выживания)

ных или несоответствующих документов, были дисквалифицированы более трех раз из-за подозрения в мошеннических действиях.

На графике (рис. 5) видна разница между поведением групп, не дисквалифицированных, и дисквалифицированных два, три или более раз. Вероятность выживания больше в группах, которые чаще были дисквалифицированы. Поскольку в рамках большинства схем сговоров фигурирует механизм, когда фирмы-марионетки участвуют в торгах, а потом дисквалифицируются и тендер выигрывает компания-заговорщик, данный график скорее подтверждает предложенную гипотезу, что

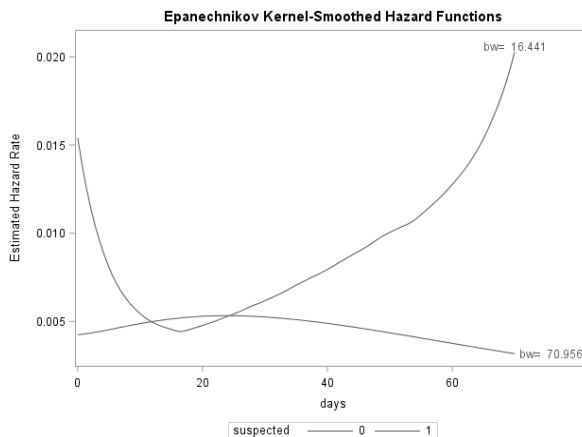


Рис. 4. Графики функций риска по стратам

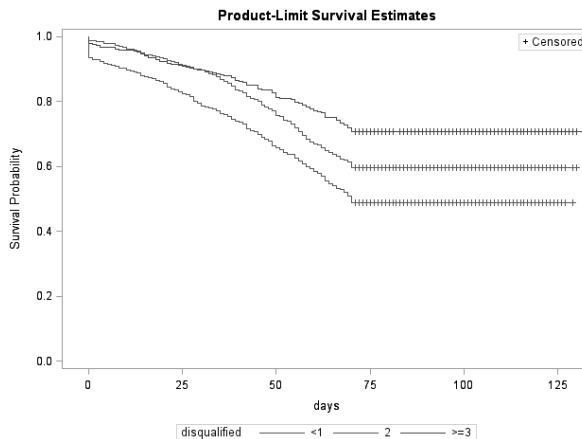


Рис. 5. График функции выживания компаний в зависимости от флага дисквалификации (disqualified — количество дисквалификаций)

фирмы-марионетки активно используются на платформе и отсутствует механизм выявления и отстранения таких фирм от участия в торгах.

Посредством критерия Wilcoxon осуществлена проверка различий в дисквалифицированных стратах.

На графике показана соответствующая опасность выхода с торгов для различных типов фирм. Как видим, для компаний, которые можно считать марионетками, существует две стратегии: или они создаются на один день (выходят из игры в тот же день торгов и не возвращаются), или они постоянно прикрывают мошенников.

Таблица 3. Сравнения статистик по тесту Wilcoxon

Adjustment for Multiple Comparisons for the Wilcoxon Test				
Strata Comparison		p-Values		
disqualified	disqualified	Chi-Square	Raw	Scheffe
1,0000	2,0	63,7528	<,0001	<,0001
1,0000	3,0000	109,7	<,0001	<,0001
2,0	3,0000	0,6076	0,4357	0,7380

Построены также модели пропорциональных рисков Кокса в SAS Enterprise Guide [9] и сравнены функции риска и функции выживаемости с течением времени (рис. 7). Графики показывают, что функция выживаемости для компаний существенно падает до 70 дня, это именно и связано с исчезновением с торгов компаний-однодневок. Хотя следует отметить также и отток компаний, которые впервые поучаствовали в торгах, столкнулись с фактом непрозрачности проведения тендеров или просто сговором, и просто приняли решение больше не участвовать в подобных торгах. Поскольку участие в тендерных закупках требует существенной подготовки большого количества документов и времени, то трата времени впустую также нецелесообразна.

График функции риска, который достаточно долго колеблется в течение первых 60 дней, показывает зависимость и рискованность для компаний, успевших поучаствовать только в одних торгах, а дальше прекратили участие в них. Также четко виден пик для компаний, которые служили марионетками и уходили после одноразового фиктивного участия. Интересен факт изменения функций выживания во времени для компаний, в зависимости от факта дисквалификации. Поскольку функция выживания значительно выше для компаний, дисквалифицированных три и более раз, а потом дважды, то это говорит не о случайном факте, а о подтверждении нашей теории, что на платформе зарегистрированы и продолжают участвовать компании, пребывающие в сговоре, и сама система никоим образом не ограничивает их участие в торгах. К сожалению, этот факт ставит под сомнение возможность проведения прозрачных торгов в принципе.

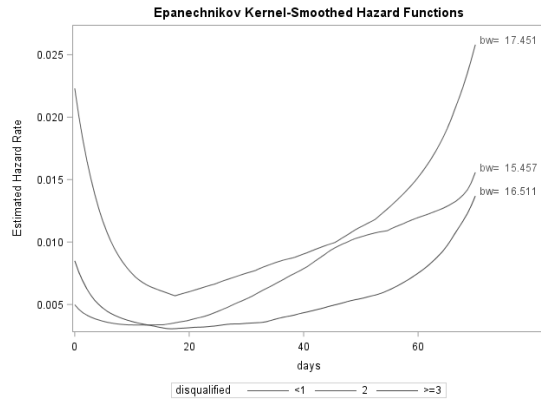


Рис. 6. График функции уровня риска в зависимости от классификации по признаку дисквалификации с торгов

Таблица 4. Анализ оценок максимального правдоподобия

Параметр	DF	Оценка параметра	Станд. ошибка	Chi-Square	Pr > ChiSq	Уровень риска
losses	1	0,05430	0,01203	20,3790	<,0001	1,056
objections	1	-0,13225	0,03631	13,2655	0,0003	0,876
participations	1	-0,15657	0,01051	222,0473	<,0001	0,855
wins	0	0

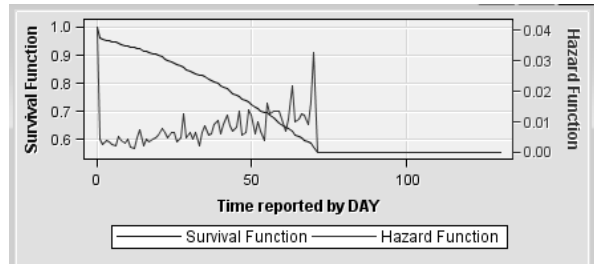


Рис. 7. Изменение графика функции риска и функции выживания в зависимости от количества дней пребывания в системе торгов (Survival Function — функция выживания, Hazard Function — функция риска)

Еще одним интересным фактом, подтверждающим эту теорию, оказалось количество торгов, в которых такие компании участвуют, оно значительно выше среднего, что свидетельствует о частых фактах сговора на платформе.

Заключение

Анализ данных, проведенный методами нейронных сетей, логистической регрессии и деревьев решений, показал результаты ошибоч-

ной классификации тендерных заявок на предмет мошенничества на уровне 30 процентов, что имеет достаточно высокое значение, и свидетельствует о несовершенстве применяемых на платформе алгоритмов, и в целом неспособности системы автоматически обнаруживать и блокировать сговоры на этапе подачи заявок на публичное проведение тендеров. Предложенный подход к анализу тендерных заявок на основе теории выживания для прогноза появления компаний, — настоящих участников тендеров, а также для выявления фактов мошенничества или сговора на платформе *ProZorro*, построенные функции риска и функции выживаемости компаний во времени позволяют определить и классифицировать компанию на предмет мошеннических действий до ее непосредственного участия в следующих торгах. Это видно по поведе-

нию компании во времени, изменении функции риска, а также по поведению функции риска в зависимости от характеристики снятия с торгов.

Следует отметить, что предложенный подход к оцениванию поведения компаний, участвующих в тендерах, более универсален для обнаружения мошенничества, поскольку повторение мошенниками действий реального поведения честных компаний невозможен и нецелесообразен для идеи мошенничества. Такой анализ позволяет выявить не только компании-однодневки, разово участвующие в торгах, но и определить компании, фиктивно участвующие в тендерах, оценив их функцию выживания во времени, а также по количеству торгов, в которых они ранее участвовали, сколько раз были выбраны и сколько раз компании были дисквалифицированы.

СПИСОК ЛИТЕРАТУРЫ

1. Закон України Про публічні закупівлі. — Верховна Рада України, <http://zakon2.rada.gov.ua/laws/show/922—19>
2. *ProZorro*: публічні закупівлі, <https://prozorro.gov.ua/>
3. Зайченко Ю.П. Основи проєктування інтелектуальних систем, К.: Слово, 2006, 352 с.
4. Чубукова И.А. Data Mining, М.: Бинум ЛБЗ, 2008, 384 с.
5. Kuznetsova N.V., Bidiuk P.I. Modeling of credit risks on the basis of the theory of survival // J. of Automation and Information Sciences, 2017, **49**, Issue 11, P. 11–24.
6. Бидюк П.И., Романенко В.Д., Тимошук О.Л. Аналіз часових рядів, К.: Політехніка, 2013, 600 с.
7. Cox D.R., Society S.B. Regression models and life-tables // Methodological, 2007, **34**, N 2, P. 187–220.
8. Cao R., Vilar J.M., Devia A. Modelling consumer credit risk via survival analysis // SORT 33 (1) Jan.—June 2009, P. 3–30.
9. Allison P.D. Survival Analysis Using SAS: A Practical Guide, Cary, NC: SAS Institute Inc, 2010, 324 p.

Поступила 16.04.2018

REFERENCES

1. *Zakon* Ukraine Pro Publichni Zakupivli, <http://zakon2.rada.gov.ua/laws/show/922—19>, (In Ukrainian).
2. *ProZorro*: publichni Zakupivli., <https://prozorro.gov.ua/>, (In Ukrainian).
3. Zaychenko U.P. Osnovi proektuvannia intelektualnih system, Kyiv, Slovo, 2006. 352 p., (In Ukrainian).
4. Chubukova I.A. Data Mining, Moscow, Binom LBZ, 2008, 384 p. (In Russian).
5. Kuznetsova N.V., Bidiuk P.I. Modeling of credit risks on the basis of the theory of survival. Journal of Automation and Information Sciences, Volume 49, Issue 11, 2017, P. 11–24.
6. Bidiuk P.I., Romanenko V.D., Tymoshuk O.L. Analiz chasovih riadiv [Time series analysis], Kyiv, Polytechnika, 2013. 600 p. (In Ukrainian).
7. Cox D.R., Society S.B. Regression models and life-tables // Methodological, **34**, N 2. (2007), P. 187–220.
8. Cao R., Vilar J.M., Devia A. Modelling consumer credit risk via survival analysis. SORT 33 (1) January-June 2009, p. 3–30.
9. Allison P.D. Survival Analysis Using SAS: A Practical Guide, Second Edition, Cary, NC: SAS Institute Inc., 2010. 324 p.

Received 16.04.2018

Natalia Kuznietsova, PhD in Technical Sciences, Associate Professor,
natalia-kpi@ukr.net

Petro Biduyk, Doctor of Technical Sciences, Professor,
pbidyuke_00@ukr.net.

Mathematical Methods of System Analysis department of the Institute for Applied Systems
Analysis NTUU «Igor Sikorsky KPI», Peremohy ave., 37, Kyiv, 03056, Ukraine

IDENTIFICATION OF THE FRAUD RISK DURING TENDER PURCHASES BY SURVIVAL THEORY METHODS

Introduction. Transparency in public procurement is a vivid information for foreign investors and the international community, which ensures an equal access and opportunities for all companies in the market, indicating a lack of corruption and self-regulation of the market. For Ukraine, where a special platform for open online purchases was created, the task of verifying and detecting conspiracies or opaque contracts is rather relevant, as media are increasingly accused of opacity and conspiracy in conducting such tenders.

The purpose of the article is to develop the new models for the company's identification and classification in order to identify unfair competition and conspiracy.

Methods. Standard methods of data showed rather high error (at 30% level), and therefore it was proposed to construct the survival models based on the proportional Cox risks and Kaplan-Mayer estimates.

Results. Survival functions and risk functions allowed us to determine behavior of fraud suspected companies in time, as well as depend on the time spent on the platform, the number of applications submitted and rejected on the tenders. Separately, there were built survival functions for the companies in the number of rejected tenders, and a tendency for a longer stay on the platform of companies that had certain arrangements with each other and participated in tenders was determined.

Conclusions. The proposed approach to the analysis and classification of companies on the basis of the survival models showed high results of the suspicious companies detection and significantly higher probability of survival and subsequent work in time compared with the companies real participants of the bidding. Higher probabilities of companies' survival are also observed if they have a large number of cases in bidding, and have repeatedly disqualified from bidding. This is a confirmation that the models correctly identify the suspicious companies, which, unfortunately, are present and remain active on the Ukrainian tender market.

Keywords: tender purchases, fraud risks, Cox proportional risk model, Kaplan—Mayer estimation.

Н.В. Кузнецова, канд. техн. наук, доцент,

П.І. Бідюк, д-р техн. наук, професор,

Кафедра математичних методів системного аналізу Інститут прикладного
системного аналізу (ІПСА) Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського» (КПІ)

ВИЯВЛЕННЯ РИЗИКІВ ШАХРАЙСТВА В ХОДІ ТЕНДЕРНИХ ЗАКУПІВЕЛЬ МЕТОДАМИ ТЕОРІЇ ВИЖИВАННЯ

Вступ. Прозорість проведення публічних закупівель є важливою інформацією для іноземних інвесторів та міжнародної спільноти, яка підтверджує, що в країні забезпечується рівний доступ і однакові можливості для всіх компаній на ринку, і свідчить про відсутність корупції та саморегуляцію ринку. Для України, де була створена спеціальна платформа для відкритих онлайн-закупівель, задача перевірки і виявлення змов або непрозорих договорів є досить актуальною, оскільки все частіше лунають звинувачення у ЗМІ про непрозорість і змови при проведенні таких тендерів.

Мета статті — розробити нові моделі для виявлення та класифікації компаній, що беруть участь у тендерах, з метою уникнення недобросовісної конкуренції і змови.

Методи. Стандартні методи інтелектуального аналізу даних на основі нейронних мереж, дерев рішень, байєсівського класифікатора та логістичної регресії показали досить високу похибку, а тому запропоновано побудувати моделі виживання на основі пропорційних ризиків Кокса та оцінки Каплана—Майєра.

Результати. Функції виживання та функції ризику дозволили визначити поведінку компаній, підозрюваних у змові, у часі, а також виявляли залежність від часу перебування на платформі кількості поданих та відхилених заявок на тендерах. Окремо були побудовані функції виживання для компаній за кількістю відхилених тендерів, та визначено тенденцію довшого перебування на платформі компаній, які мали певні домовленості між собою та брали участь у тендерах.

Висновки. Запропонований підхід до аналізу та класифікації компаній на основі моделей виживання показав високі результати виявлення підозрілих компаній та суттєво вищі значення ймовірності з виживання та подальшої роботи в часі порівняно з компаніями — реальними учасниками торгів. Вищі ймовірності виживання були помічені також і для компаній, що часто брали участь у торгах, та навіть неодноразово були дискваліфіковані, що є підтвердженням правильного визначення підозрілих компаній, присутніх і активних на українському ринку тендерних закупівель.

Ключові слова: *тендерні закупівлі, ризики шахрайства, модель пропорційних ризиків Кокса, оцінка Каплана–Майєра.*