

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ДИСКРИМИНАНТНОГО И КЛАСТЕРНОГО АНАЛИЗА СТАТИСТИЧЕСКИХ ЗАВИСИМОСТЕЙ В ЗАДАЧАХ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ КРЕДИТНЫХ РЕЙТИНГОВ

В.И. Малюгин

Белорусский государственный университет,
кафедра математического моделирования и анализа данных
пр. Независимости, 4, 220030, г. Минск, Беларусь
телефон: + (375) 0172095530; факс: + (375) 0172095104; e-mail: Malugin@bsu.by
web: www.fpmi.bsu.by

Для оценки и прогнозирования кредитных рейтингов заемщиков предлагаются алгоритмы, основанные на методах дискриминантного и кластерного анализа статистических зависимостей регрессионного типа. Проводится сравнительный анализ точности классификации заемщиков на основе альтернативных алгоритмов на модельных и реальных данных.

Ключевые слова - дискриминантный и кластерный анализ статистических зависимостей, кредитные рейтинги и кредитный скоринг, многомерные регрессионные модели.

1 ОБЩАЯ ХАРАКТЕРИСТИКА ПРОБЛЕМЫ

Одним из необходимых условий устойчивого функционирования коммерческих банков и банковской системы в целом является эффективное управление различными рисками [1]. Основным видом риска, с которым сталкиваются в своей деятельности банки, является кредитный риск или риск дефолта (*default risk*). На практике управление кредитным риском предполагает решение задач анализа и классификации заемщиков (юридических лиц) по степени кредитоспособности, платежеспособности и финансовой устойчивости. Основной акцент при этом делается на оценке готовности и способности заемщика выполнить кредитные обязательства в полном объеме. Важным инструментом анализа и принятия решений является кредитный скоринг. Обычно используемые в системах кредитного скоринга алгоритмы позволяют классифицировать заемщиков по степени кредитоспособности, разбивая их на группы (классы) заемщиков с различным кредитным рейтингом, на основании информации, касающейся лишь самих заемщиков. При этом не учитываются качество обслуживания ранее полученных кредитов, макроэкономические условия, имеющие место в момент предоставления кредита, а также – возможность изменения этих условий к моменту истечения срока кредита вследствие внешних шоковых воздействий. Очевид-

но, качество обслуживания кредита конкретным заемщиком может существенно зависеть от изменений макроэкономической ситуации. Более того, возможна миграция заемщиков из одной группы надежности в другую. Это обусловлено тем, что такие внешние шоки, как скачкообразные изменения цен на энергоносители, существенные изменения процентных ставок и валютного курса, смена условий налогообложения, ухудшение рыночной конъюнктуры могут оказывать непосредственное влияние на финансовое состояние предприятий. Последствия таких шоковых воздействий для различных типов предприятий-заемщиков могут быть разными, в силу различия их возможностей адаптироваться к внешним воздействиям. Эти возможности могут зависеть от отраслевой принадлежности, уровня технологий, качества менеджмента и т.д.

Для оценки влияния внешних факторов на различные типы заемщиков, а также – на кредитный рынок в целом (например, в смысле увеличения или уменьшения числа проблемных заемщиков, роста кредиторской задолженности в банковской системе и т.д.), предлагается использовать эконометрические модели зависимостей показателей состояния [2] определенных типов заемщиков (т.е. кредитных рейтингов заемщиков) от их индивидуальных показателей (балансовых и показателей его кредитной истории) и макроэкономических показателей. Разработка подобных эконометрических моделей может вестись, как в интересах органов, осуществляющих надзор за кредитным рынком страны, так и самих коммерческих банков.

2 ЗАДАЧИ И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Пусть при оценке (прогнозировании) класса состояния (рейтинга) t -го ($t = 1, \dots, T, \dots$) заемщика на момент окончания срока кредитного договора состояние заемщика описывается случайным вектором признаков $y_t \in \mathbb{R}^{N+M}$, который допускает разбиение на подвекторы:

$$y_t = \begin{pmatrix} x_t \\ z_t \end{pmatrix} \in \mathbb{R}^{N+M}, \quad x_t \in \mathbb{R}^N, \quad z_t \in Z \subset \mathbb{R}^M, \quad (1)$$

где T – общее количество заемщиков из анализируемой выборки; $x_t = (x_{t1}, \dots, x_{tN})'$ – вектор эндогенных (внутренних) переменных (балансовых показателей и показателей кредитной истории заемщика), характеризующих состояние заемщика; $z_t = (z_{t1}, \dots, z_{tM})'$ – вектор экзогенных переменных (факторов), описывающих влияние на состояние заемщиков со стороны общих внешних экономических факторов.

Предполагается, что заемщики могут находиться в одном из двух классов состояний Ω_0 или Ω_1 , где Ω_0 – класс надежных заемщиков, обладающих достаточно высокой степенью платежеспособности; Ω_1 – класс проблемных заемщиков обладающих более низкой степенью платежеспособности. Номер класса состояния заемщика (или текущего рейтинга) описывается ненаблюдаемой дискретной случайной величиной $d_t \equiv d(t) \in S = \{0, 1\}$ с распределением вероятностей

$$P\{d_t = i\} = \pi_i > 0 (i \in S), \pi_{0t} + \pi_{1t} = 1, \quad (2)$$

параметры $\{\pi_i\} (i \in S)$ соответствуют априорным вероятностям классов состояний для заемщика t . Анализируемая выборка заемщиков в целом характеризуется вектором регрессионных наблюдений $X = (x_t) \in \mathcal{R}^{TN}$, который соответствует вектору значений факторов $Z = (z_t) \in \mathcal{R}^{TM}$, а также – вектором состояний заемщиков $d = (d_1, \dots, d_T)' \in S^T$.

Для произвольного заемщика t между подвекторами составного вектора наблюдений (1) существует статистическая зависимость, которая для класса $d(t) \in S$ описывается моделью многомерной линейной регрессии вида

$$x_t = B_{d(t)} z_t + \xi_{d(t)}, \quad (3)$$

где $B_{d(t)}$ – $(N \times M)$ -матрица коэффициентов регрессии; $\xi_{d(t)}$ – случайные ошибки наблюдения с нулевым математическим ожиданием и невырожденной ковариационной $(N \times N)$ -матрицей $\Sigma_{d(t)}$.

Вектор состояний заемщиков $d = (d_t) \in S^T$ может быть известен (обучающая выборка регрессионных наблюдений X, Z – классифицированная) или не известен (выборка X, Z – не классифицированная).

Имеют место следующие задачи классификации заемщиков, состояние которых описывается моделью регрессии (3): 1) задача дискриминантного анализа многомерных регрессионных наблюдений, если имеется классифицированная обучающая выборка; 2) задача кластерного анализа многомерных регрессионных наблюдений, если имеется неклассифицированная выборка наблюдений; 3) задача непараметрического дискриминантного анализа, случайных векторов $y_t \in \mathcal{R}^{N+M}$ вида (1) в

случае, когда параметрический вид зависимости подвекторов $x_t \in \mathcal{R}^N$ и $z_t \in \mathcal{R}^M$ не известен.

В рамках первой задачи построено подстановочное байесовское решающее правило для некоррелированных [3] и автокоррелированных ошибок наблюдения [4]. Для задачи кластерного анализа предложен алгоритм расщепления смесей распределений регрессионных наблюдений [5]. Для задачи непараметрического дискриминантного анализа предложен непараметрический классификатор случайных векторов с существенно зависимыми компонентами [6]. Для всех построенных решающих правил установлена состоятельность в смысле минимума риска классификации, а также проведено исследование риска (вероятности ошибки) классификации на основе метода асимптотических разложений. Проведен сравнительный анализ точности классификации альтернативных алгоритмов кредитного скоринга на модельных и реальных данных белорусского кредитного рынка [7, 8]. Разработаны компьютерные системы кредитного скоринга физических и юридических лиц.

ЛИТЕРАТУРА

- [1] Малюгин, В. И. Оценка устойчивости коммерческих банков на основе эконометрических моделей с дискретными зависимыми переменными / В. И. Малюгин, Е. В. Пытляк // Банковский Вестник. – 2007. – № 4 (369). – С. 30–36.
- [2] Харин, Ю. С. Эконометрическое моделирование / Ю.С. Харин, В.И. Малюгин, А.Ю. Харин. – Минск: БГУ, 2003. – 313 с.
- [3] Малюгин, В.И. Об оптимальности классификации случайных наблюдений, различающихся уравнениями регрессии / В.И. Малюгин, Ю.С. Харин // Автоматика и телемеханика. – 1986. – №7. – С. 35–46.
- [4] Малюгин, В.И. Дискриминантный анализ многомерных зависимых регрессионных наблюдений в условиях структурной параметрической неоднородности моделей // Информатика. – 2008. – № 3. – С. 17-28.
- [5] Малюгин, В.И. Статистический анализ смесей распределений регрессионных наблюдений // Информатика. – 2008. – С. 79-88.
- [6] Малюгин, В.И. Асимптотический анализ риска непараметрической классификации в случае существенно зависимых признаков / В.И. Малюгин // Вес. Нац. акад. наук Беларусі. Сер. Фіз. Мат. Інфарм. – № 3. – 2009. – С. 10–24.
- [7] Малюгин, В.И. Анализ и прогнозирование кредитного риска на основе эконометрических моделей / В.И. Малюгин, Н.В. Гринь // Экономика, моделирование, прогнозирование. – Вып. 2. – 2008. – С. 260-277.
- [8] Малюгин, В.И. Исследование эффективности алгоритмов классификации заемщиков банков на основе балансовых коэффициентов / В.И. Малюгин, Н.В. Гринь, О.И. Корчагин // Банковский Вестник. – 2009. – № 3 – С. 27–32.