

Прогнозирование банкротства средних и малых российских компаний

Тихонова Анна Сергеевна, НИУ-ВШЭ, магистратура МИЭФ
Демешев Борис Борисович, НИУ-ВШЭ, ст. преподаватель

Москва, 2014

Цель

Моделирование вероятности серьёзных финансовых трудностей средних и малых российских непубличных компаний с помощью финансовых и нефинансовых показателей

Задачи

- Сравнение различных методов
- Выявление различий моделей до и после кризиса 2009 года
- Проверка гипотезы о влиянии нефинансовых показателей
- Проверка гипотезы о различиях по отраслям и правовым формам

Основные источники

- Beaver, W. H., (1966), "Financial ratios as predictors of failure"
- Altman, E. I., Sabato, G., (2007), "Modelling credit risk for SMEs"
- Lugovskaya, L., (2010), "Predicting default of Russian SMEs on the basis of financial and non-financial variables"

Отличия от других работ и описание данных

Банкроты: ликвидированные банкроты + добровольно ликвидированные компании

Большой массив данных

База данных: Руслана

Период исследования: 2004 — 2012 годы

От 110 до 56 210 наблюдений в модели

Два типа выборок

- Балансированные
- Небалансированные

Учёт неоднородности выборки

Федеральный округ

Возраст

Отрасли по ОКВЭД:

- Строительство
- Обрабатывающие производства
- Операции с недвижимостью
- Оптовая и розничная торговля

Правовые формы

- ООО
- ЗАО

Размер компаний

- Микро-компании
- Малые компании
- Средние компании

Честные прогнозы

Оценить модель по данным года t , построить прогнозы на год $t + 1$

Уравнение модели: $\mathbb{P}(def_{it+1} = 1) = f_t(x_{it})$

Методы: логит- и пробит-модели, линейный, квадратичный и смешанный дискриминантный анализ, метод опорных векторов, классификационные деревья, алгоритм случайного леса

Базовые версии моделей

Адаптация модели Альтмана: age, iptd, roa, ebta, stdte, liq, sr

Популярность: age, iptd, roa, ebta, wcta, cr, lr, tdta

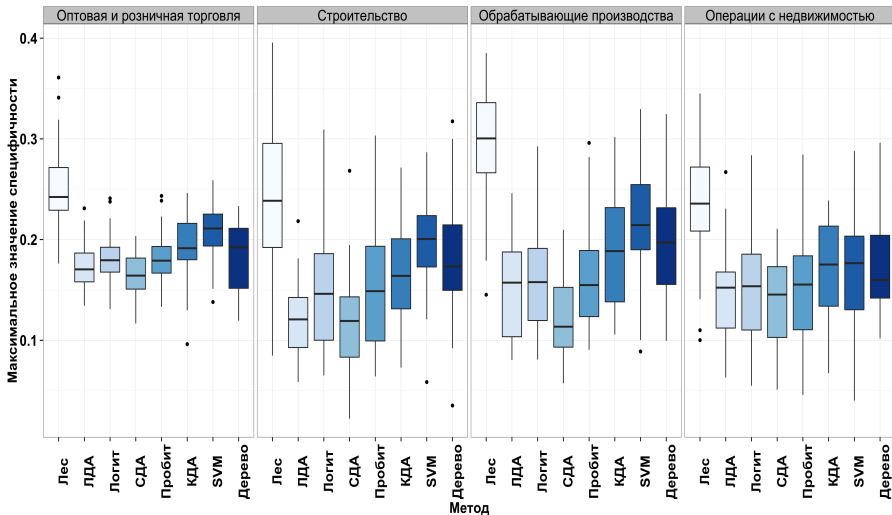
LASSO: age, wcta, gg, nat, ebtm, ltdta

Расширенные версии моделей

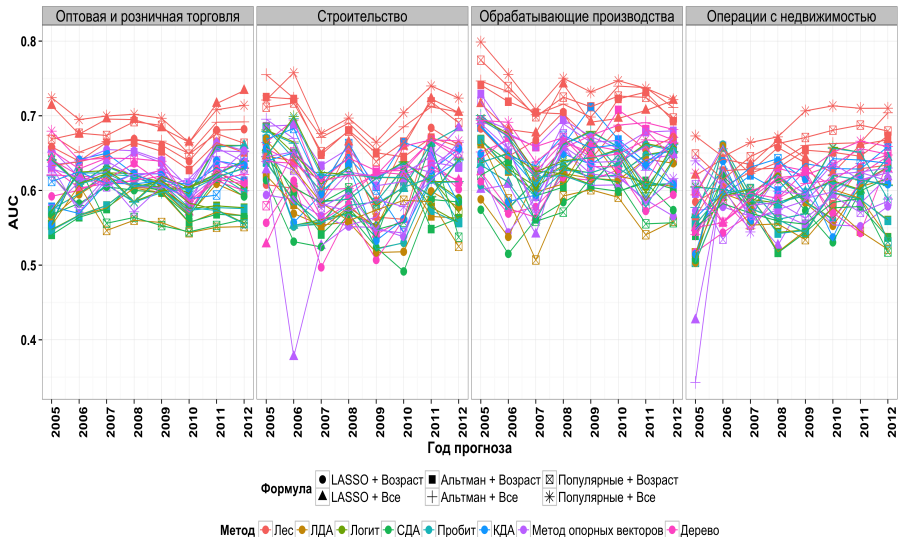
Добавляются переменные: fedreg, lasize, legalform

9 лет × 6 формул × 4 отрасли × 8 методов × 2 типа выборок

Специфичность при чувствительности 0.9 в зависимости от метода



Площадь под ROC-кривой для всех оценённых моделей

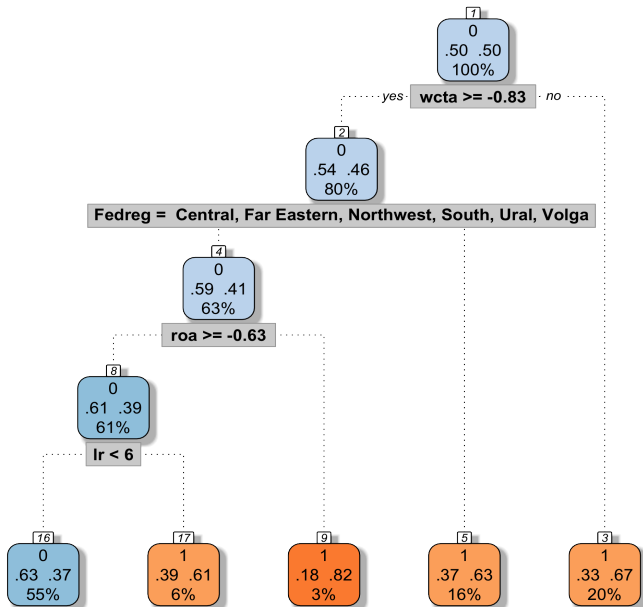


Предельные эффекты в логит-моделях (оптовая торговля 2012 год)

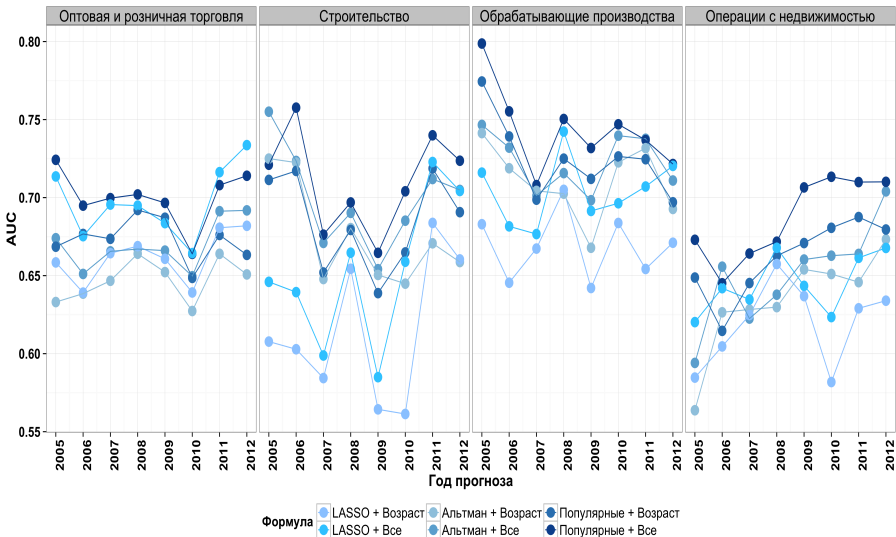
| | Альтман Возраст | Альтман Все | Популярные Возраст | Популярные Все | LASSO Возраст | LASSO Все |
|------------------------|--------------------|-------------------|-----------------------|-------------------|-------------------|-------------------|
| Intercept | 0.037 (0.013)** | 0.148 (0.057)* | 0.053 (0.012)*** | 0.253 (0.063)*** | 0.102 (0.016)*** | 0.122 (0.075) |
| iptd | 0.169 (0.056)** | 0.145 (0.056)** | 0.005 (0.008) | 0.006 (0.011) | | |
| ebta | -0.004 (0.004) | -0.005 (0.005) | -0.002 (0.002) | -0.002 (0.002) | | |
| stdte | 0.000 (0.000)* | 0.000 (0.000)* | | | | |
| roa | -0.009 (0.005)* | -0.009 (0.005) | -0.002 (0.001) | -0.002 (0.001) | | |
| liq | 0.115 (0.034)** | 0.133 (0.036)*** | | | | |
| sr | -0.039 (0.010)*** | -0.042 (0.010)*** | | | | |
| cr | | | 0.000 (0.000) | 0.000 (0.000) | | |
| wcta | | | 0.002 (0.001) | 0.002 (0.001) | 0.007 (0.018) | -0.001 (0.019) |
| lr | | | 0.000 (0.000) | 0.000 (0.000) | | |
| tdta | | | 0.002 (0.001) | 0.002 (0.001) | | |
| gg | | | | | -0.051 (0.008)*** | -0.053 (0.008)*** |
| nat | | | | | 0.000 (0.000) | 0.000 (0.000) |
| ebtm | | | | | -0.102 (0.044)* | -0.078 (0.042) |
| ldta | | | | | -0.319 (0.161)* | -0.381 (0.172)* |
| Age | -0.010 (0.002)*** | -0.010 (0.002)*** | -0.012 (0.002)*** | -0.012 (0.002)*** | -0.014 (0.002)*** | -0.013 (0.002)*** |
| Micro | | -0.243 (0.039)*** | | -0.298 (0.032)*** | | -0.375 (0.035)*** |
| Small | | -0.222 (0.040)*** | | -0.259 (0.034)*** | | -0.307 (0.040)*** |
| Limited liability firm | | 0.003 (0.046) | | -0.062 (0.053) | | 0.173 (0.066)** |
| Far Eastern | | 0.136 (0.037)*** | | 0.081 (0.041)* | | 0.100 (0.046)* |
| North Caucasian | | 0.188 (0.046)*** | | 0.254 (0.046)*** | | 0.258 (0.054)*** |
| Northwest | | 0.139 (0.029)*** | | 0.171 (0.025)*** | | 0.140 (0.030)*** |
| Siberian | | 0.197 (0.035)*** | | 0.268 (0.022)*** | | 0.263 (0.025)*** |
| South | | 0.069 (0.030)* | | 0.100 (0.035)** | | 0.129 (0.039)** |
| Ural | | 0.081 (0.029)** | | 0.106 (0.032)** | | 0.056 (0.037) |
| Volga | | 0.059 (0.023)** | | 0.102 (0.025)*** | | 0.126 (0.029)*** |
| AIC | 5042.551 | 4858.187 | 5025.012 | 4811.855 | 4092.390 | 3888.233 |
| BIC | 5092.284 | 4970.087 | 5080.894 | 4929.827 | 4134.522 | 3990.555 |
| Log Likelihood | -2513.275 | -2411.094 | -2503.506 | -2386.928 | -2039.195 | -1927.117 |
| Deviance | 5026.551 | 4822.187 | 5007.012 | 4773.855 | 4078.390 | 3854.233 |
| Num. obs. | 3702 | 3702 | 3674 | 3674 | 3038 | 3038 |

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, \cdot $p < 0.1$

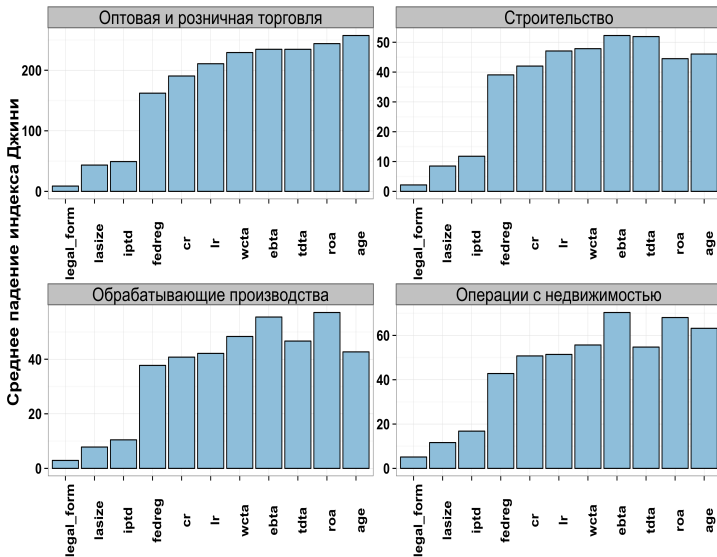
Дерево для отрасли торговли в 2012 году



Площадь под ROC-кривой для алгоритма случайного леса



Важность переменных в алгоритме случайного леса в 2012 году



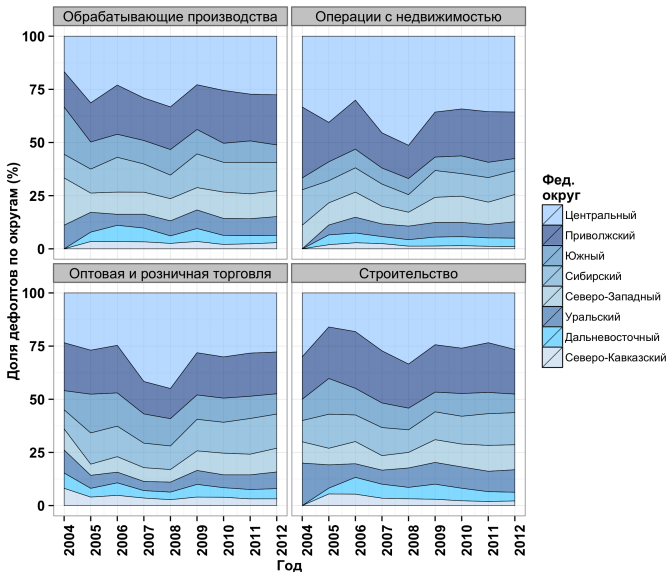
- 1 Лучшим методом для всех отраслей оказался случайный лес.
- 2 Лучшая модель содержит все нефинансовые переменные и финансовые переменные, отобранные по упоминаемости в работах.
- 3 Добавление федерального округа и размера компании улучшает модели.
- 4 Увеличение возраста влияет отрицательно на вероятность банкротства.
- 5 Среди финансовых переменных важными оказались показатели рентабельности, ликвидности и финансового рычага.
- 6 Вид организационной формы не влияет на вероятность банкротства.
- 7 Результаты для балансированных и небалансированных выборок схожи.
- 8 Обнаружены структурные сдвиги.
- 9 Отрасли существенно отличаются по влиянию отдельных переменных и по структуре временной зависимости.
- 10 Существенно отличается качество прогнозов по обучающей выборке и вне неё.

Спасибо за внимание!

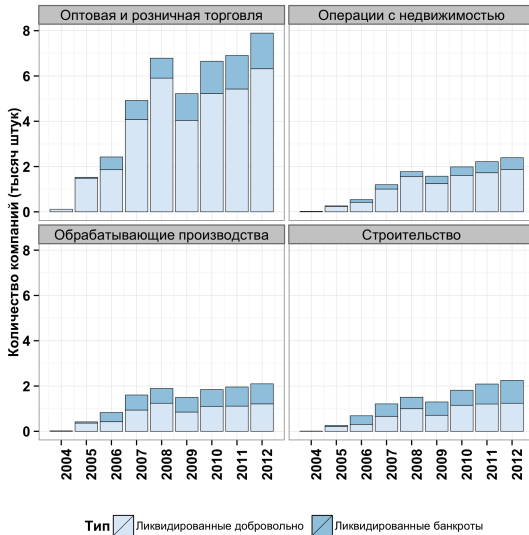
Деление компаний по размеру в России и Европейском союзе

| Размер | Количество работников | | Выручка | |
|---------|-----------------------|---------|------------------|--------------------------|
| | Европейский союз | Россия | Европейский союз | Россия |
| Микро | 1—10 | 1—15 | ≤ €2 млн | ≤ €1.4 млн (RUB 60 млн) |
| Малое | 11—50 | 16—100 | ≤ €10 млн | ≤ €9.6 млн (RUB 400 млн) |
| Среднее | 51—250 | 101—250 | ≤ €50 млн | ≤ €24 млн (RUB 1 млрд) |

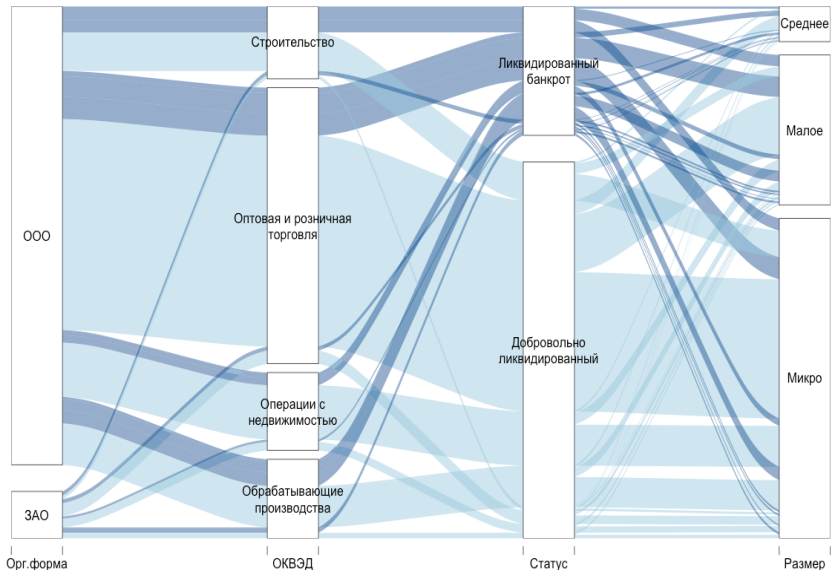
Распределение дефолтных компаний по федеральным округам



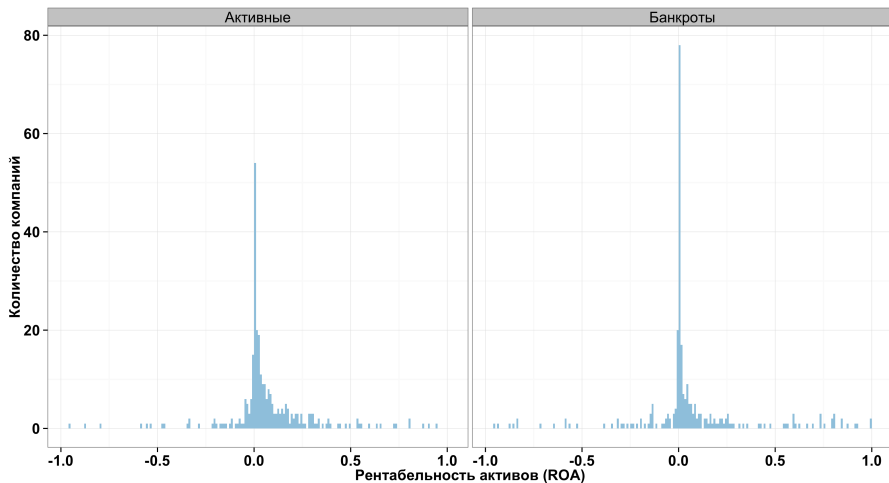
Распределение дефолтных компаний по типу дефолта



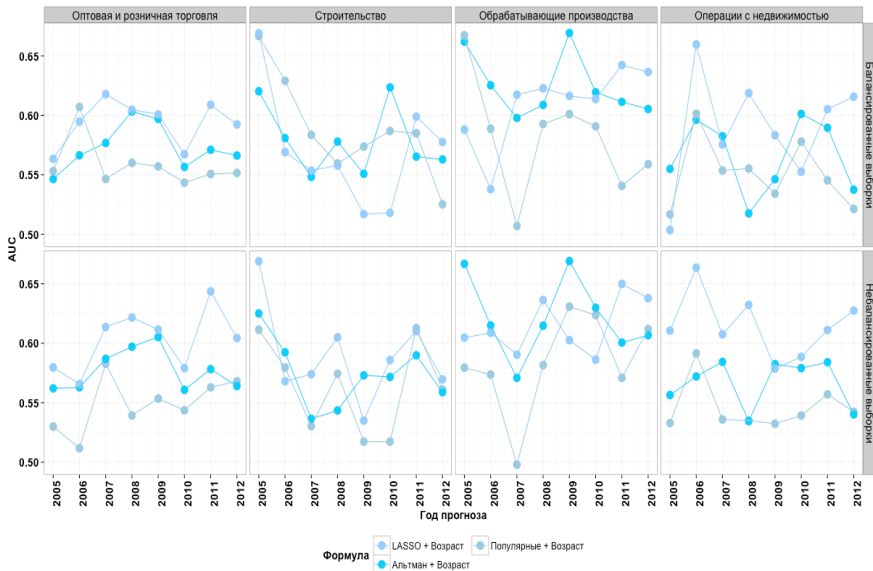
Характеристики дефолтных компаний



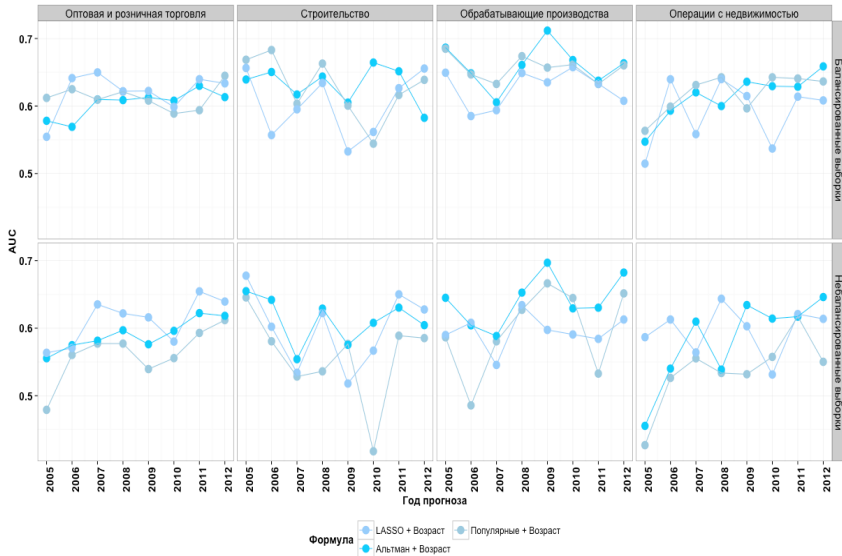
Распределение рентабельности активов (roa) для отрасли торговли в 2012 году



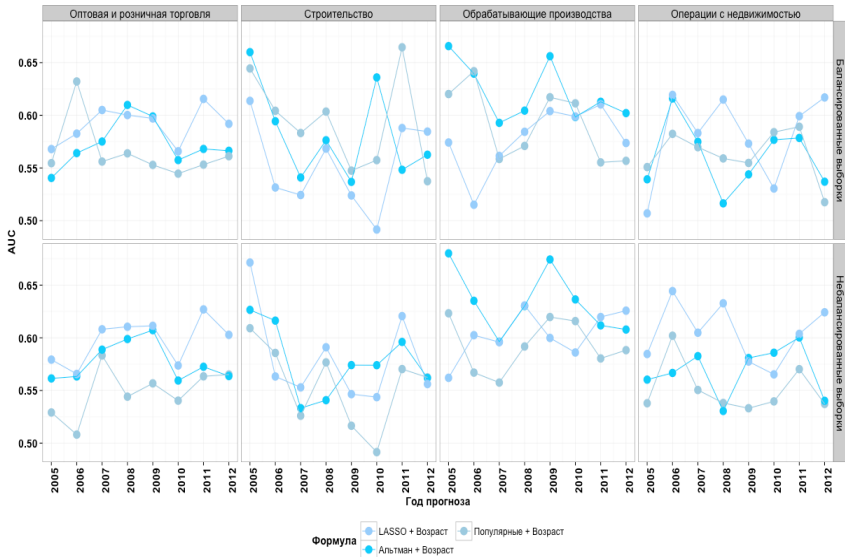
Площадь под ROC-кривой для линейного дискриминантного анализа



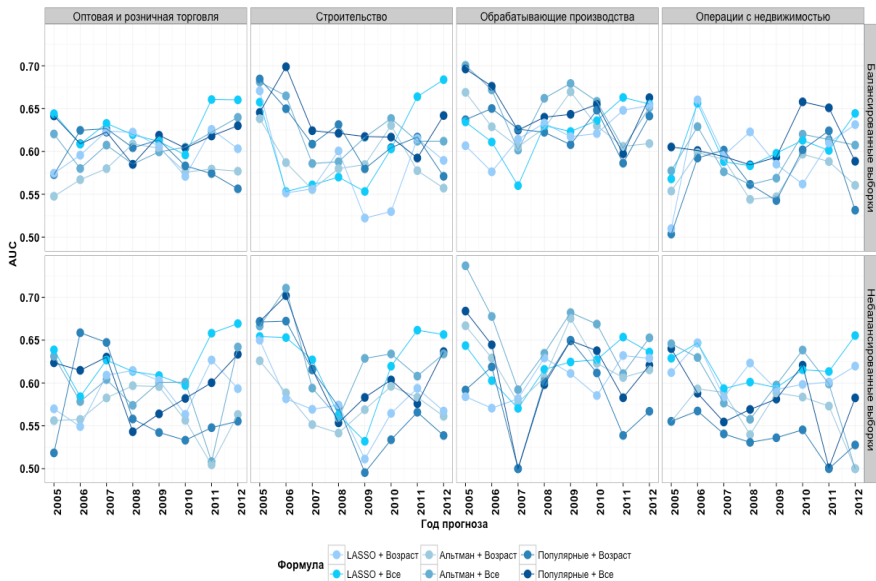
Площадь под ROC-кривой для квадратичного дискриминантного анализа



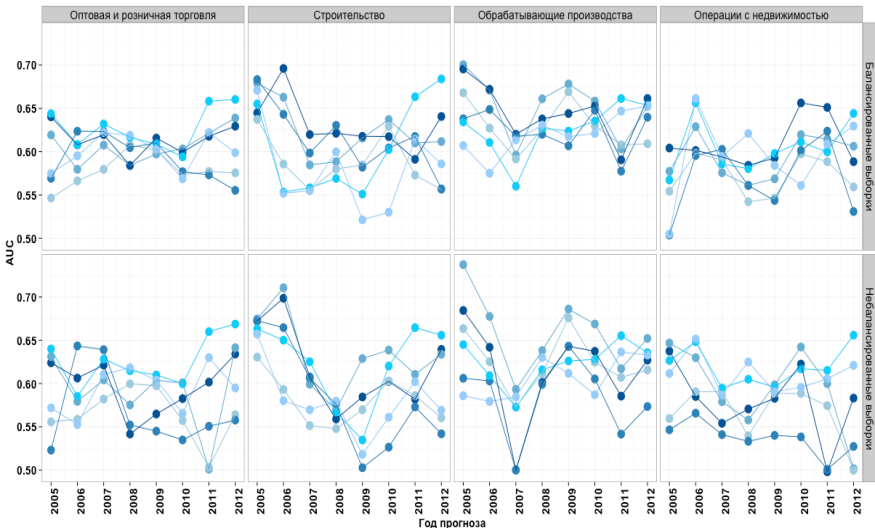
Площадь под ROC-кривой для смешанного дискриминантного анализа



Площадь под ROC-кривой для логит-модели



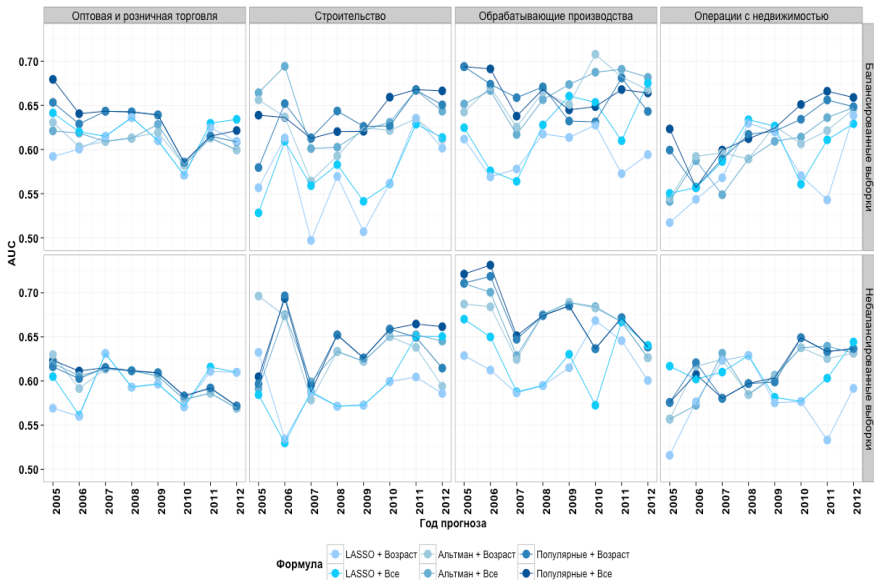
Площадь под ROC-кривой для пробит-модели



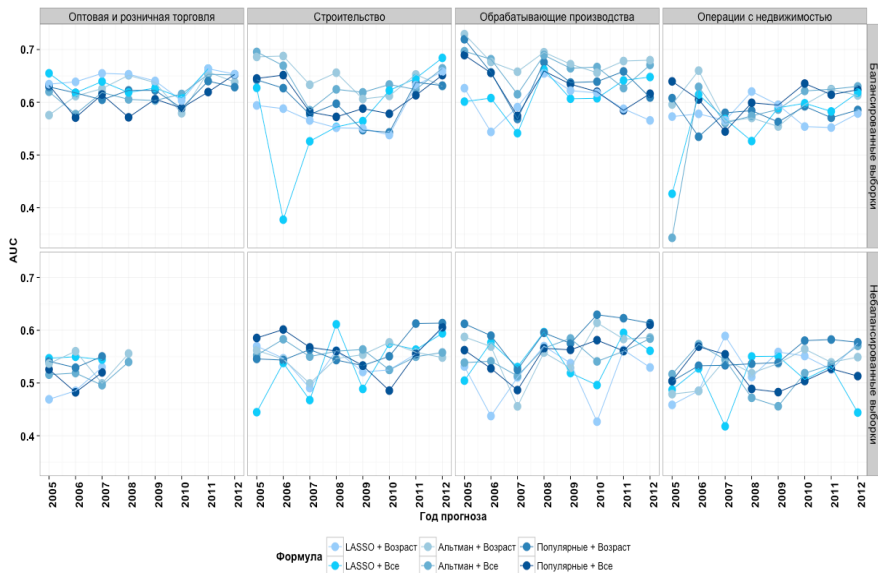
Формула

- LASSO + Возраст
- LASSO + Все
- Альтман + Возраст
- Альтман + Все
- Популярные + Возраст
- Популярные + Все

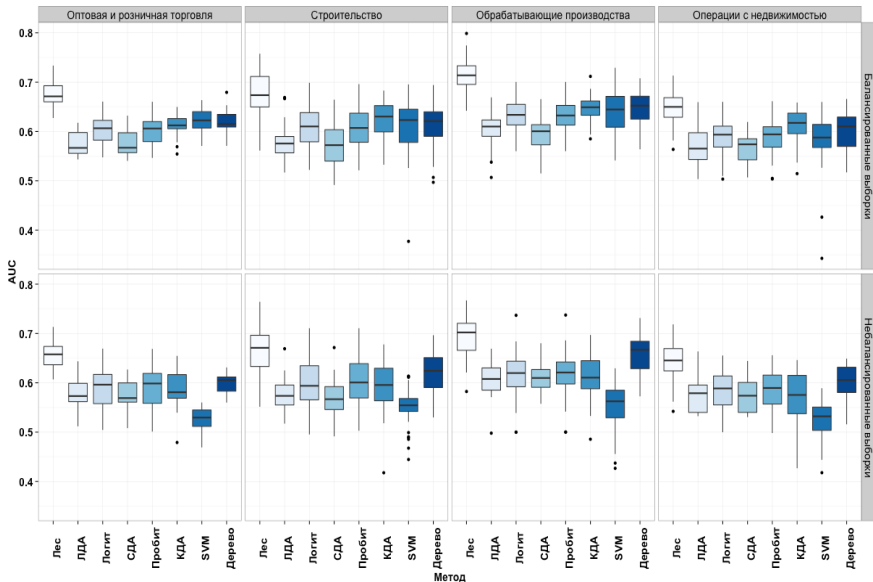
Площадь под ROC-кривой для классификационного дерева



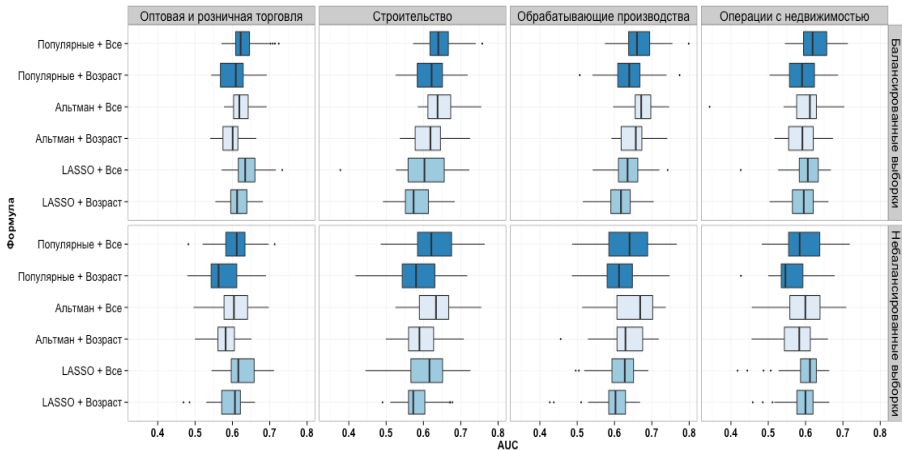
Площадь под ROC-кривой для метода опорных векторов



Площадь под ROC-кривой в зависимости от метода



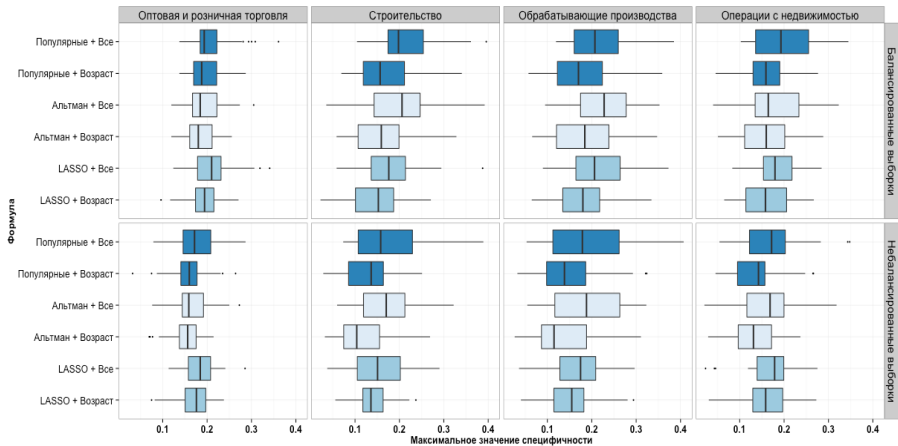
Площадь под ROC-кривой в зависимости от формулы



Средние значения AUC по методам и моделям

| Критерий выбора финансовых переменных | Нефинансовые переменные | Среднее значение AUC в зависимости от метода и формулы | | | | | | | |
|----------------------------------------------------------------|----------------------------|--------------------------------------------------------|---------------|--------|------------------|------------------|------------------------------|------------------|------------------|
| | | ЛДА | КДА | СДА | Логит | Пробит | Метод опорных векторов | Дерево | Лес |
| Балансированные выборки | | | | | | | | | |
| Альтман и Сабато | Возраст Все | 0.5855 | 0.6275 | 0.5840 | 0.5929 0.6200 | 0.5917 0.6192 | 0.6363 0.6232 | 0.6234 0.6284 | 0.6666 0.6855 |
| Популярность | Возраст Все | 0.5697 | 0.6288 | 0.5781 | 0.6025 0.6268 | 0.6004 0.6253 | 0.6106 0.6134 | 0.6338 0.6405 | 0.6858 0.7104 |
| LASSO | Возраст Все | 0.5906 | 0.6099 | 0.5755 | 0.6001 0.6170 | 0.5988 0.6159 | 0.6014 0.5958 | 0.5880 0.6040 | 0.6412 0.6747 |
| Среднее значение AUC внутри метода | | 0.5819 | 0.6221 | 0.5792 | 0.6099 | 0.6086 | 0.6135 | 0.6197 | 0.6774 |
| Небалансированные выборки | | | | | | | | | |
| Альтман и Сабато | Возраст Все | 0.5849 | 0.6064 | 0.5893 | 0.5828 0.6179 | 0.5841 0.6197 | 0.5443 0.5402 | 0.6279 0.6282 | 0.6538 0.6829 |
| Популярность | Возраст Все | 0.5604 | 0.5634 | 0.5619 | 0.5690 0.6027 | 0.5680 0.6024 | 0.5657 0.5415 | 0.6302 0.6331 | 0.6774 0.7076 |
| LASSO | Возраст Все | 0.6054 | 0.6006 | 0.5963 | 0.5950 0.6214 | 0.5975 0.6229 | 0.5266 0.5303 | 0.5938 0.6065 | 0.6134 0.6517 |
| Среднее значение AUC внутри метода | | 0.5836 | 0.5901 | 0.5825 | 0.5981 | 0.5991 | 0.5414 | 0.6200 | 0.6645 |
| Среднее значение AUC внутри метода по всем выборкам | | 0.5827 | 0.6061 | 0.5809 | 0.6040 | 0.6038 | 0.5775 | 0.6199 | 0.6709 |

Специфичность при чувствительности 0.9 в зависимости от формулы

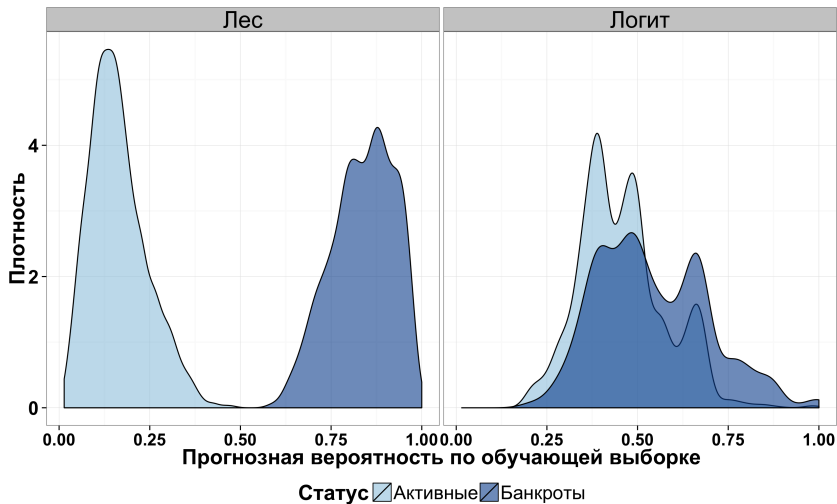


Логит-модели для отрасли торговли в 2012 году

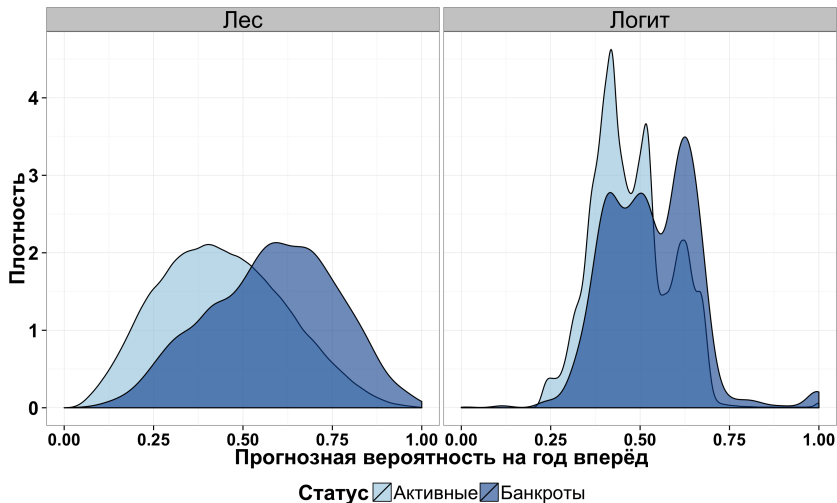
| | Альтман Возраст | Альтман Все | Популярные Возраст | Популярные Все | LASSO Возраст | LASSO Все |
|-------------------------------|--------------------|-------------------|-----------------------|-------------------|-------------------|-------------------|
| <i>Intercept</i> | 0.150 (0.056)** | 0.642 (0.248)** | 0.216 (0.050)*** | 1.101 (0.252)*** | 0.426 (0.064)*** | 0.549 (0.299) |
| <i>iptd</i> | 0.693 (0.353)* | 0.630 (0.327) | 0.022 (0.031) | 0.026 (0.043) | | |
| <i>ebta</i> | -0.016 (0.020) | -0.020 (0.021) | -0.009 (0.009) | -0.008 (0.009) | | |
| <i>stdte</i> | 0.000 (0.000)* | 0.000 (0.000)* | | | | |
| <i>roa</i> | -0.039 (0.022) | -0.038 (0.021) | -0.007 (0.004) | -0.008 (0.004) | | |
| <i>liq</i> | 0.472 (0.144)** | 0.580 (0.149)*** | | | | |
| <i>sr</i> | -0.162 (0.040)*** | -0.184 (0.041)*** | | | | |
| <i>cr</i> | | | 0.002 (0.001) | 0.002 (0.001) | | |
| <i>wcta</i> | | | 0.008 (0.005) | 0.008 (0.005) | 0.029 (0.072) | -0.007 (0.076) |
| <i>lr</i> | | | 0.001 (0.002) | 0.001 (0.002) | | |
| <i>tdta</i> | | | 0.007 (0.005) | 0.008 (0.005) | | |
| <i>gg</i> | | | | | -0.212 (0.032)*** | -0.240 (0.034)*** |
| <i>nat</i> | | | | | 0.000 (0.000) | 0.000 (0.000) |
| <i>ebtm</i> | | | | | -0.425 (0.175)* | -0.351 (0.166)* |
| <i>ltdta</i> | | | | | -1.331 (0.644)* | -1.713 (0.687)* |
| <i>Age</i> | -0.040 (0.007)*** | -0.042 (0.008)*** | -0.048 (0.007)*** | -0.050 (0.008)*** | -0.057 (0.009)*** | -0.056 (0.009)*** |
| <i>Micro</i> | | -1.090 (0.135)*** | | -1.228 (0.139)*** | | -1.591 (0.165)*** |
| <i>Small</i> | | -0.965 (0.135)*** | | -1.129 (0.138)*** | | -1.378 (0.161)*** |
| <i>Limited liability firm</i> | | 0.013 (0.210) | | -0.272 (0.212) | | 0.779 (0.265)** |
| <i>Far Eastern</i> | | 0.706 (0.184)*** | | 0.329 (0.170) | | 0.404 (0.190)* |
| <i>North Caucasian</i> | | 1.074 (0.250)*** | | 1.137 (0.257)*** | | 1.133 (0.288)*** |
| <i>Northwest</i> | | 0.702 (0.113)*** | | 0.707 (0.111)*** | | 0.571 (0.128)*** |
| <i>Siberian</i> | | 1.024 (0.102)*** | | 1.150 (0.105)*** | | 1.105 (0.114)*** |
| <i>South</i> | | 0.328 (0.143)* | | 0.409 (0.148)** | | 0.528 (0.165)** |
| <i>Ural</i> | | 0.389 (0.136)** | | 0.434 (0.135)** | | 0.225 (0.148) |
| <i>Volga</i> | | 0.275 (0.104)** | | 0.413 (0.104)*** | | 0.510 (0.121)*** |
| <i>AIC</i> | 5042.551 | 4858.187 | 5025.012 | 4811.855 | 4092.390 | 3888.233 |
| <i>BIC</i> | 5092.284 | 4970.087 | 5080.894 | 4929.827 | 4134.522 | 3990.555 |
| <i>Log Likelihood</i> | -2513.275 | -2411.094 | -2503.506 | -2386.928 | -2039.195 | -1927.117 |
| <i>Deviance</i> | 5026.551 | 4822.187 | 5007.012 | 4773.855 | 4078.390 | 3854.233 |
| <i>Num. obs.</i> | 3702 | 3702 | 3674 | 3674 | 3038 | 3038 |

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, \cdot $p < 0.1$

Сравнение алгоритма случайного леса и логит-модели на выборке 2011 года



Сравнение алгоритма случайного леса и логит-модели на выборке 2012 года



Финансовый рычаг

$$gg = \frac{\text{Non-current liabilities} + \text{loans}}{\text{Total equity}}$$

$$ltda = \frac{\text{Long-term debt}}{\text{Total assets}}$$

$$stdte = \frac{\text{Short-term debt}}{\text{Total equity}}$$

$$tdta = \frac{\text{Total debt}}{\text{Total assets}}$$

Ликвидность

$$liq = \frac{\text{Cash and cash equivalent}}{\text{Total assets}}$$

$$sr = \frac{\text{Total equity}}{\text{Total assets}}$$

$$wcta = \frac{\text{Working capital}}{\text{Total assets}}$$

$$cr = \frac{\text{Current assets}}{\text{Current liabilities}}$$

$$lr = \frac{\text{Current assets} - \text{stocks}}{\text{Current liabilities}}$$

Рентабельность

$$roa = \frac{\text{Net income}}{\text{Total assets}}$$

$$ebta = \frac{\text{EBIT}}{\text{Total assets}}$$

$$ebtm = \frac{\text{EBIT}}{\text{Turnover}}$$

Обслуживание долга

$$iptd = \frac{\text{Interest paid}}{\text{Total debt}}$$

$$nat = \frac{\text{Turnover}}{\text{Total equity} + \text{Non-current liabilities}}$$

Медианное абсолютное отклонение

$$MAD = 1.4826 \cdot \text{median}_i |x_i - \text{median}_j(x_j)|$$

Линейный дискриминантный анализ

$$f(x_i, z_i | \text{def}_i = 0) \sim N(\mu_0, \Sigma)$$

$$f(x_i, z_i | \text{def}_i = 1) \sim N(\mu_1, \Sigma)$$

Квадратичный дискриминантный анализ

$$f(x_i, z_i | \text{def}_i = 0) \sim N(\mu_0, \Sigma_0)$$

$$f(x_i, z_i | \text{def}_i = 1) \sim N(\mu_1, \Sigma_1)$$

Дискриминантный анализ смеси нормальных распределений

$$f(x_i, z_i | \text{def}_i = 0) \sim p_{0a}N(\mu_{0a}, \Sigma) + p_{0b}N(\mu_{0b}, \Sigma) + p_{0c}N(\mu_{0c}, \Sigma)$$

$$f(x_i, z_i | \text{def}_i = 1) \sim p_{1a}N(\mu_{1a}, \Sigma) + p_{1b}N(\mu_{1b}, \Sigma) + p_{1c}N(\mu_{1c}, \Sigma)$$

Метод опорных векторов

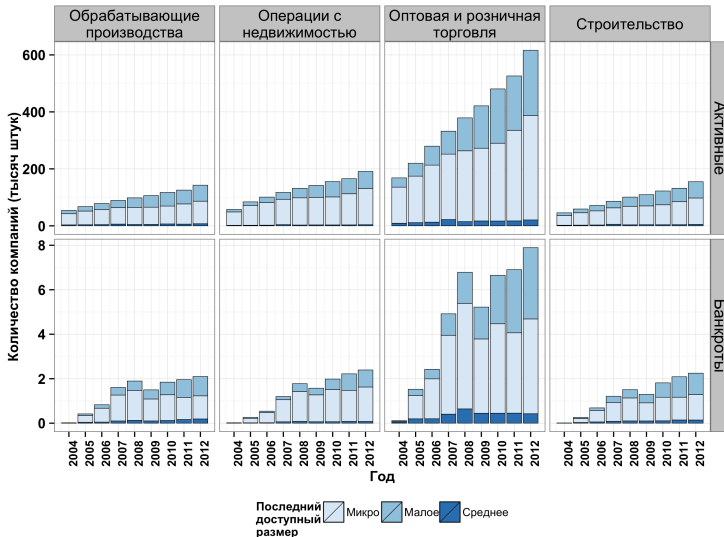
$$Q = \frac{2}{h^2} + \frac{C}{h} \sum_i d_i$$

$$k(r, r') = \exp(-\sigma|r - r'|^2)$$

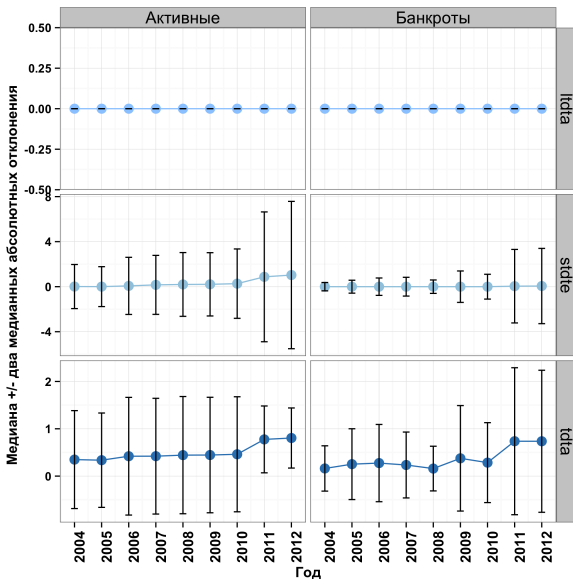
Индекс Джини

$$I_G(\text{Tree}) = \sum_i \frac{n_i}{n} 2h_i(1 - h_i)$$

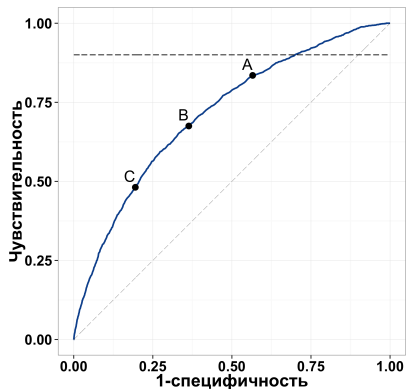
Распределение компаний по размеру и отраслям



Медианы некоторых показателей финансового рычага



ROC-кривая для алгоритма случайного леса



| Прогнозируемый исход | Фактический исход | |
|----------------------|---------------------|---------------|
| | Положительный | Отрицательный |
| | Порог равен 0.4 (A) | |
| Положительный | 1 526 (TP) | 150 936 (FP) |
| Отрицательный | 311 (FN) | 115 915 (TN) |
| Чувствительность | 0.8307 | |
| Специфичность | 0.4344 | |
| | Порог равен 0.5 (B) | |
| Положительный | 1 255 (TP) | 97 050 (FP) |
| Отрицательный | 582 (FN) | 169 801 (TN) |
| Чувствительность | 0.6832 | |
| Специфичность | 0.6363 | |
| | Порог равен 0.6 (C) | |
| Положительный | 884 (TP) | 51 945 (FP) |
| Отрицательный | 953 (FN) | 214 906 (TN) |
| Чувствительность | 0.4812 | |
| Специфичность | 0.8053 | |

Чувствительность: $TPR = \frac{TP}{TP+FN}$

Специфичность: $TNR = \frac{TN}{TN+FP}$

Площадь под ROC-кривой — AUC