

УДК 519.711.3

М.В. Потанина, ст. преподаватель

Севастопольский национальный технический университет

ул. Университетская, 33, г. Севастополь, Украина, 99053

E-mail: mem@sevgtu.sebastopol.ua

ПОСТРОЕНИЕ МОДЕЛИ ОЦЕНКИ РЕНТАБЕЛЬНОСТИ АВТОТРАНСПОРТНОГО СРЕДСТВА МЕТОДОМ СМЕЩЕННОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ РЕГРЕССИОННЫХ УРАВНЕНИЙ

Рассмотрена методика построения модели оценивания рентабельности автотранспортного средства методом ридж-регрессии на примере предприятия, осуществляющего городские пассажироперевозки. В качестве критерия оптимальности регрессионной модели рентабельности предлагается использовать величину ошибки прогноза в заданной области.

Постановка проблемы

Одним из главных финансовых показателей экономической эффективности предприятия, осуществляющего пассажироперевозки, является рентабельность использования автотранспортных средств. Формула для рентабельности имеет вид:

$$P_{a/c} = \frac{\Pi_{\text{общ}} \cdot 100}{Z_{\text{общ}}}, \quad (1)$$

где $\Pi_{\text{общ}}$ – общая прибыль, $Z_{\text{общ}}$ – общие затраты на содержание автотранспортного средства [1]. Однако не существует рекомендуемой приемлемой методики построения модели оценки этого показателя с учетом неопределенности структуры модели и особенностей специфики самого метода смещенного оценивания.

Анализ последних исследований и публикаций

Уровень рентабельности автотранспортных предприятий Украины, занимающихся пассажироперевозками, традиционно невысок. Экономические расчеты автопредприятий показывают: если ранее, по состоянию на 2007 г., уровень их рентабельности (в разрезе соотношения чистой прибыли к затратам) составлял 6,18%, то сейчас он опустился до отметки ниже 5%. Его можно считать критическим, поскольку стопроцентную техбезопасность гарантирует не менее пятнадцатипроцентная рентабельность.

Отечественное законодательство регламентирует расчет значения показателя рентабельности автотранспортного средства по формуле (1). Здесь не учитывается, что каждое автотранспортное средство характеризуется огромным числом показателей, которые обычно коррелируют между собой. Структура модели данного сложного объекта, которым является автотранспортное средство, не определена. Возникает проблема выбора метода оценивания для построения прогнозирующих регрессионных моделей такого объекта. Было разработано и исследовано множество методов оценивания. Например, метод наименьших квадратов (МНК), взвешенный метод наименьших квадратов, ридж-регрессия и т.д. [2]. Среди линейных по наблюдениям оценок параметров уравнения регрессии наиболее часто употребляют оценки (МНК), которые при выполнении ряда условий статистически оптимальны.

Несмотря на это, при решении практических задач, связанных с обработкой данных экспериментов об объекте управления, точность МНК становится недостаточной. Эффективность оценок повышается при привлечении априорной информации об оцениваемых параметрах уравнений регрессии вероятностного или детерминированного характера. Учет априорной информации, связан, как правило, с получением смещенных оценок, позволяющих уменьшить средний квадрат ошибки предсказания по сравнению с МНК-оценками и тем самым получить значения параметров, наиболее близких к истинным.

Рассмотрим ситуацию, когда имеет место сильная мультиколлинеарность экспериментальных данных, на практике характеризующих любое автотранспортное средство. Она типична для случая обработки данных заранее не спланированного эксперимента и имеет весьма отрицательные последствия для оценивания регрессионных коэффициентов. Существуют различные подходы к решению проблемы мультиколлинеарности. Например, можно перейти к другому рода оценкам, смещенным, позволяющим устранить недостатки МНК-оценок в этой ситуации. К таким оценкам относятся «гребневые» или ридж-оценки [2]. Они являются более устойчивыми, чем оценки МНК и имеют меньшее значение среднеквадратической ошибки прогноза.

Основной, центральной проблемой использования ридж-оценок является выбор параметра регуляризации γ . Главная трудность заключается в том, что и смещение и квадратическая ошибка $R(r)$

зависят от неизвестных истинных значений параметров. Такую процедуру практически невозможно автоматизировать.

Был предложен ряд подходов, позволяющих осуществить выбор параметра регуляризации.

Херл и Кеннард [3] показали, что квадратическая ошибка $R(r)$ для регуляризованной оценки будет всегда меньше, чем для МНК-оценки, если выполняется неравенство:

$$0 < r < \sigma^2 / \beta_{\max}^2, \quad (2)$$

где $\beta = V\alpha$, а V – ортогональная $(k \times k)$ -матрица, составленная из собственных векторов информационной матрицы $X^T X$; α – неизвестные значения регрессионных коэффициентов; $\beta_{\max} = (|\beta_1|, |\beta_2|, \dots, |\beta_k|)$.

Теоболд [3] вывел достаточные условия минимизации для более общей функции потерь $R_\omega(r) = E(\alpha(r) - \alpha)^T \Omega (\alpha(r) - \alpha)$, где Ω – произвольная функция. Эти условия задаются неравенствами:

$$0 < r < 2\sigma^2 / \alpha^T \alpha. \quad (3)$$

Обечейн [3] получил необходимые и достаточные условия выполнения неравенства $R(r) < R(0)$:

$$0 < r < 2 / |\chi_k|, \quad (4)$$

где χ_k – отрицательное собственное число матрицы $(X^T X)^{-1} - (\alpha \alpha^T / \sigma^2)$.

Все эти подходы требуют знания истинных значений дисперсии σ^2 и вектора неизвестных коэффициентов α . Подстановка вместо значений их оценок делает выбор r стохастическим и не гарантирует снижение квадратической ошибки.

Херл и Кеннард, исходя из формулы (3) предлагают выбирать коэффициент регуляризации равным:

$$r = s_m^2 / \beta_{\max}^2, \quad (5)$$

где $\beta_{\max}$ – максимальный по модулю элемент вектора $\beta = V\alpha$, а s^2 – оценка σ^2 . Видно, что в таком случае величины σ^2 и β^2 в (4) просто заменяются их оценками. Херл и Кеннард предлагают добавлять к разным диагональным элементам $X^T X$ различные добавки – $r_{ii} = s^2 / \beta_i^2$, регуляризующие параметры, но при этом не всегда происходит существенное улучшение. К сожалению, в этом подходе также требуется информация, которая редко имеется в распоряжении исследователя.

Цель статьи

В настоящей работе предлагается модифицированная методика построения прогнозирующей модели оценки рентабельности автотранспортного средства методом смещенного оценивания параметров регрессионных уравнений. При этом предлагается использовать выбор параметра регуляризации r на основе минимума среднеквадратической ошибки прогноза, полученной по этой модели.

Изложение основного материала исследования с полным обоснованием полученных научных результатов

В виде приложения разработанной методики рассматривалась проблема получения наилучшей прогнозирующей модели вычисления рентабельности при использовании автотранспортного средства среднего класса типа «Эталон» на разных маршрутах предприятия-перевозчика. В качестве Y рассматривается показатель рентабельности выбранного транспортного средства. В качестве X использовались данные, характеризующие это транспортное средство: протяженность маршрута, количество автобусов на маршруте по расписанию, количество автобусов на маршруте по списку, продолжительность оборотного рейса на маршруте, эксплуатационная скорость, интервал движения в часы пик и т.д. Все X в рассматриваемом примере, в силу своей специфики, являются сильно коррелированными, количество их достаточно велико, что делает невозможным точное определение тех параметров, которые должны присутствовать в прогнозируемой модели, а также значительно снижает ее прогнозирующие способности.

Данная методика использует преимущество ридж-регрессии оперировать со смещенными оценками коэффициентов прогнозирующей модели. При этом полученная модель объекта дает наименьшую ошибку прогноза в прогнозируемой области.

Планирование вычислительного эксперимента при этом для параметрической идентификации математической модели объекта исследования удовлетворяет ограничениям, накладываемым мерой корреляции экспериментальных данных.

В качестве критерия оптимальности регрессионной модели выберем величину ошибки прогноза, полученной по этой модели. Будем предполагать, что модель используется для целей прогноза величины y в заданной области значений величины x .

Для любой модели можно определить величину среднего риска:

$$I(\alpha) = \int (y - \varphi(x, \alpha))^2 P(x, y) dx dy, \quad (6)$$

где $P(x, y)$ – совместная плотность распределения переменных X и Y . Задача имеет различные решения в зависимости от области прогнозирования. Если X – не случайная векторная переменная, принимающая только дискретные значения x_u ($u=1,2,\dots,N$), что имеет место в случае заранее спланированного эксперимента, то величину среднего риска можно записать в форме:

$$I(\alpha) = \frac{1}{N} \sum_{u=1}^N \int (y - \eta(x_u))^2 P(y | x_u) dy + \frac{1}{N} \sum_{u=1}^N (\eta(x_u) - \varphi(x_u, \alpha))^2. \quad (7)$$

Первая часть этой формулы связана с дисперсией величины y , а вторая с качеством предложенной модели. Если дисперсия σ^2 постоянна, то величина среднего риска задается формулой:

$$I(\alpha) = \sigma^2 + \frac{1}{N} \sum_{u=1}^N (\eta(x_u) - \varphi(x_u, \alpha))^2. \quad (8)$$

Если зависимость $\varphi(x, \alpha)$ определяется на основе выборочных значений Y , то эта величина тоже будет случайной и необходимо рассматривать математическое ожидание среднего риска.

Пусть цель моделирования заключается в получении оптимальной модели для лучшего предсказания функции отклика Y в области W_1 , тогда функцию потерь для модели j можно будет записать в виде:

$$L(j) = \int_{W_1} E(y - \eta_j(x, \alpha_j))^2 dx / \int_{W_1} dx, \quad (j=1,2,\dots,q). \quad (9)$$

Здесь вычисляется среднеквадратическая ошибка прогноза для каждой модели j области W_1 , Y определяется как возможная величина зависимой переменной в точке X . Если дисперсия σ^2 является константой, то функцию потерь для модели можно записать в виде:

$$L(j) = \sigma^2 + \int_{W_1} (\eta(x) - \eta_j(x, \alpha_j))^2 dx / \int_{W_1} dx. \quad (10)$$

Для того чтобы выбрать аппроксимирующую модель, необходимо иметь значения экспериментальных данных в каждой точке $x_i^T = (x_{1i}, x_{2i}, \dots, x_{ki})$ ($i=1,2,\dots,N$). Пусть матрица $D^T = [x_1, x_2, \dots, x_N]$ – матрица плана размера $N \times k$ и $Y^T = [y_1, y_2, \dots, y_N]$ – вектор N наблюдений над функцией отклика.

Пусть

$$\hat{\alpha}_j = \Phi_j(X, Y) \quad (11)$$

будет оценкой α_j , где Φ_j функция от Y , плана D и класса моделей S_j .

Если модель линейная по параметрам и используется метод ридж-регрессии, тогда

$$\hat{\alpha}_j = (F_j^T F_j + r I_n)^{-1} F_j^T Y, \quad (12)$$

где $F_j^T = [f_j(x_1), f_j(x_2), \dots, f_j(x_N)]$, матрица значений вектора f_j в N экспериментальных точках, r – малое положительное число, добавленное к диагональным элементам матрицы $F_j^T F_j$, так называемый коэффициент регуляризации, а I_n – единичная матрица. Предполагается, что матрица F_j имеет полный ранг.

Рассмотрим выражение для функции потерь в случае ридж-регрессии для модели q по области W_1 , где $x \in [-1, 1]$.

$$\begin{aligned} L(q) &= E \int_{W_1} x^T (\alpha - \hat{\alpha}(r)) (\alpha - \hat{\alpha}(r))^T x dx / \int_{W_q} dx \\ &= \sigma^2 sp X B J B X^T + \alpha^T (I_q - X^T X B) J (I_q - B X^T X) \alpha, \end{aligned} \quad (13)$$

где $B = (X^T X + rI_q)^{-1}$ и $J = \int_{W_1} x x^T dx / \int_{W_1} dx$.

Формула (13) состоит из двух частей. Первая, содержащая истинные коэффициенты $\alpha^T (I_q - X^T X B) J (I_q - B X^T X) \alpha$, это есть влияние смещения оценок, а вторая часть $\sigma^2 \text{sp} X B J B X^T$ содержит дисперсию ошибки. Если изобразить графики каждой части отдельно, то видна следующая закономерность. С ростом параметра регуляризации кривая влияния смещения оценок монотонно возрастает, причем, чем больше коэффициенты, тем рост кривой происходит быстрее. Кривая суммы дисперсий оценок коэффициентов монотонно убывает.

Очевидно, что можно найти такое значение параметра регуляризации, при котором можно существенно уменьшить квадратическую ошибку оценок за счет их небольшого смещения. Автором статьи разработана итерационная процедура поиска оптимального параметра регуляризации r в условиях мультиколлинеарности и определения неизвестных параметров регрессионной модели. На рисунке 1 показан график зависимости функции потерь от значения коэффициента регуляризации r . Здесь кривая MNK – значение функции потерь, рассчитанное с помощью метода наименьших квадратов, RIDG – значение функции потерь, рассчитанное с помощью ридж-регрессии, $L1(r)$ – кривая, зависящая от дисперсии коэффициентов, $L2(r)$ – кривая, зависящая от смещения коэффициентов. Пересечение кривых $L1(r)$ и $L2(r)$ дает значение оптимального параметра регуляризации r_{opt} . [2]. При этом видно, что рассчитанное в результате использования модифицированной методики значение оптимального параметра регуляризации r_{opt} практически не отличается от экспериментального. Что подтверждает правильность разработанной методики. Разработано программное обеспечение на языке m-файлов MATLAB, реализующее предложенную модифицированную методику.

В результате применения разработанного модифицированной методики удалось оптимизировать первоначальную структуру модели, правильно произвести ее спецификацию. Показано, что в рассматриваемой задаче достигнута поставленная цель – повышено качество прогнозирующей модели, с учетом неопределенности структуры модели и метода смещенного оценивания параметров модели, что свидетельствует о корректности использованных и разработанных методов управления, созданного программного обеспечения. Результативность данной методики подтвердили имитационные эксперименты.

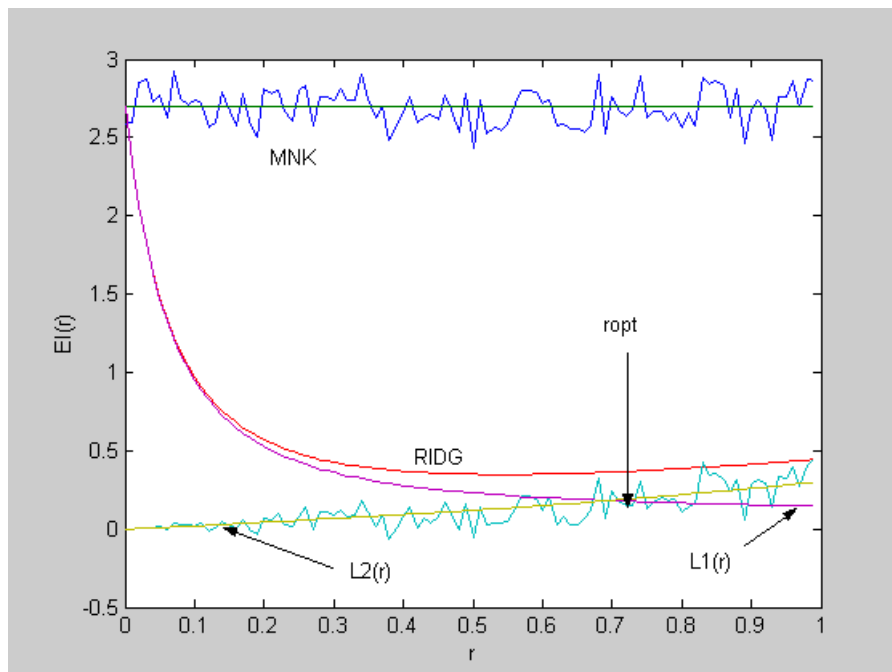


Рисунок 1 – График зависимости функции потерь от значения коэффициента регуляризации r

После проведенных исследований получены значимые коэффициенты регрессии, модель является адекватной, значение коэффициента детерминации равно 84 %. Таким образом, можно сделать вывод, что в результате использования разработанной модифицированной методики, удалось построить модель оценки рентабельности автотранспортного средства, которую можно эффективно использовать для дальнейшей оптимизации и прогнозирования данного показателя.

Перспективы дальнейших исследований

Предметом дальнейших исследований является изучение взаимосвязи между построением модели оценки рентабельности автотранспортного средства методом смещенного оценивания регрессионных уравнений и критерием селекции моделей C_p Малоуса.

Библиографический список

1. Замков О.О. Математические методы в экономике / О.О. Замков. — М.: Финансы и статистика, 1997. — 248 с.
2. Вучков И. Прикладной регрессионный анализ / И. Вучков. — М.: Финансы и статистика, 1987. — 239 с.
3. Бородич С.А. Эконометрика: учеб. пособие для вузов / С.А. Бородич. — Минск: Новое знание, 2001. — 416 с.

Поступила в редакцию 16.05.2008