

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***ОЦЕНКА КАЧЕСТВА ИДЕНТИФИКАЦИИ СЛОЖНОГО ОБЪЕКТА УПРАВЛЕНИЯ В УСЛОВИЯХ МУЛЬТИКОЛЛИНЕАРНОСТИ**

*Рассмотрена методика оценки качества процесса идентификации СОУ в условиях мультиколлинеарности. В качестве критерия оптимальности регрессионной модели предлагается использовать величину ошибки прогноза в заданной области. Предложена методика нахождения оптимального параметра регуляризации при смещенном оценивании параметров регрессионных уравнений.*

**Постановка проблемы**

В процессе идентификации СОУ основной задачей, которую необходимо решить, является оценка качества идентификации. Задача значительно усложняется при наличии мультиколлинеарности, которая часто имеет место при функционировании СОУ. Устранение отрицательных последствий мультиколлинеарности частично решает переход к смещенным методам оценивания регрессионных моделей, например, ридж-регрессии. Но при смещенном оценивании параметров центральной проблемой является выбор параметра регуляризации  $r$ . На настоящий момент не существует рекомендуемой приемлемой методики построения модели оценки этого показателя с учетом неопределенности структуры модели и особенностей специфики самого метода смещенного оценивания.

Главная трудность заключается в том, что и смещение и квадратическая ошибка прогноза зависят от неизвестных истинных значений параметров.

В работе рассматривается разработка и исследование метода выбора оптимального параметра регуляризации при структурной идентификации модели СОУ на основе исследования функции потерь как критерия оптимальности при смещенном оценивании регрессионной модели на предмет экстремума.

**Анализ последних исследований и публикаций**

Пусть СОУ описывается выходной переменной  $y$  и вектором независимых входных переменных  $x^T = [x_1, x_2, \dots, x_k]$ , где  $k$  – количество входных переменных. Значение вектора переменных  $x$  можно точно вычислить, измерить, предсказать и задать в эксперименте. Переменная  $y$  является показателем качества работы объекта или целевой функцией. Выбор целевой функции определяется конкретными условиями исследуемого процесса, а также возможностью оперативного измерения, надежностью и доступностью информации, универсальностью и статистической эффективностью. Пусть также имеется априорная информация об областях возможных значений переменных:  $y_1 \leq E\{y\} \leq y_2$ ,  $x \in W$ , где  $W$  – множество возможных значений вектора  $x$ ,  $E\{\cdot\}$  – символ математического ожидания.

Идентификация осуществляется при помощи настраиваемой модели определенной структуры, параметры которой могут изменяться.

Пусть функциональная зависимость между переменными имеет вид

$$y = \eta(x) + \varepsilon, \quad (1)$$

где  $\eta(x)$  – истинная, но неизвестная модель и  $\varepsilon$  независимая случайная величина с нулевым средним и дисперсией  $\sigma^2$ .

Рассматривается случай, когда на классе моделей задана структура, позволяющая ввести частичный порядок на множестве моделей. При таком порядке классы моделей как бы вложены один в другой:  $S_1 \subset S_2 \subset \dots \subset S_q$ ,  $S_j$  – множество моделей класса  $j$ . Так, для линейных по параметрам моделей можно задать структуру в зависимости от количества членов модели. В этом случае каждый класс  $S_j$  задается следующей моделью

$$\eta(\bar{x}, \bar{\alpha}_j) = f^T(\bar{x}) \bar{\alpha}_j \quad (j = 1, 2, \dots, q), \quad (2)$$

где  $f_j^T(\bar{x}) = [f_1(\bar{x}), f_2(\bar{x}), \dots, f_{n_j}(\bar{x})]$  – вектор известных функций от вектора  $\bar{x}$ ,  $n_j$  – число параметров модели  $j$ ,  $q$  – максимально возможный порядок модели,  $\bar{\alpha}_j^T = [\alpha_1, \alpha_2, \dots, \alpha_{n_j}]$  – вектор неизвестных параметров.

Сложный объект управления характеризуется большим числом показателей, которые часто коррелируют между собой. Структура модели такого рода объектов обычно не определена. Возникает проблема выбора метода оценивания для построения прогнозирующих регрессионных моделей сложных

объектов. На настоящий момент разработано и исследовано множество методов оценивания. Например, метод наименьших квадратов (МНК), взвешенный метод наименьших квадратов, ридж-регрессия и т.д. [1]. 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.

В работе рассматривается ситуация, когда имеет место сильная мультиколлинеарность экспериментальных данных. 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 другого рода оценкам, смещенным, позволяющим устранить недостатки МНК-оценок в этой ситуации. К таким оценкам относятся «гребневые» или ридж-оценки [1]. Они являются более устойчивыми, чем оценки МНК и имеют меньшее значение среднеквадратической ошибки прогноза.

Запишем регрессионную модель в виде

$$Y = F\alpha + \varepsilon, \quad (3)$$

где  $\alpha$  – вектор-столбец неизвестных регрессионных коэффициентов,

$F$  – матрица функций от управляемых переменных,

$\varepsilon$  – вектор-столбец остатков или случайное возмущение.

МНК-оценки имеют вид:

$$\hat{\alpha} = (F^T F)^{-1} F^T Y. \quad (4)$$

Тогда для стабилизации оценок МНК при решении системы нормальных уравнений вместо матрицы  $F^T F$  предлагается использовать матрицу  $(F^T F + rI_n)$ , где  $r$  – малое положительное число, добавленное к диагональным элементам матрицы  $F^T F$ , так называемый коэффициент регуляризации, а  $I_n$  – единичная матрица [3]. Такие оценки получили название ридж-оценок и могут быть записаны в виде:

$$\tilde{\alpha} = (F^T F + rI_n)^{-1} F^T Y. \quad (5)$$

Основной, центральной проблемой использования ридж-оценок является выбор параметра регуляризации  $r$ . Главная трудность заключается в том, что и смещение и квадратическая ошибка прогноза  $R(r)$  зависят от неизвестных истинных значений параметров. Такую процедуру практически невозможно автоматизировать.

Был предложен ряд подходов, позволяющих осуществить выбор параметра регуляризации.

Херл и Кеннард [2] показали, что квадратическая ошибка  $R(r)$  для регуляризованной оценки будет всегда меньше, чем для МНК-оценки, если выполняется неравенство

$$0 < r < \sigma^2 / \beta_{\max}^2, \quad (6)$$

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

Теоболд [2] вывел достаточные условия минимизации для более общей функции потерь  $R_\omega(r) = E(\alpha(r) - \alpha)^T \Omega(\alpha(r) - \alpha)$ , где  $\Omega$  – произвольная функция. Эти условия задаются неравенством:

$$0 < r < 2\sigma^2 / \alpha^T \alpha. \quad (7)$$

Обечейн [2] получил необходимые и достаточные условия выполнения неравенства  $R(r) < R(0)$ , т.е.

$$0 < r < 2/|\chi_k|, \quad (8)$$

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

Все эти подходы требуют знания истинных значений дисперсии  $\sigma^2$  и вектора неизвестных коэффициентов  $\alpha$ . Подстановка вместо значений их оценок делает выбор  $r$  стохастическим и не гарантирует снижение среднеквадратической ошибки.

Херл и Кеннард, исходя из формулы (8), предлагают выбирать коэффициент регуляризации равным

$$r = s_m^2 \hat{\beta}_{\max}^2, \quad (9)$$

где  $\beta_{\max}$  – максимальный по модулю элемент вектора  $\beta = V\alpha$ , а  $s^2$  – оценка  $\sigma^2$ . Видно, что в таком случае величины  $\sigma^2$  и  $\beta^2$  в (9) 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 решаются вопросы оценки качества процесса идентификации СОУ в условиях мультиколлинеарности путем построения его прогнозирующей модели методом смещенного оценивания параметров регрессионных уравнений. При этом предлагается методика выбора параметра регуляризации  $\gamma$  на основе минимума среднеквадратической ошибки прогноза, полученной по этой модели. Приводится алгоритм исследования функции потерь на экстремум, доказательство существования локального минимума, возможности вычисления оптимального параметра регуляризации.

#### Изложение основного материала исследования с полным обоснованием полученных научных результатов

В виде приложения разработанной методики рассматривалась проблема получения наилучшей прогнозирующей модели сложного объекта управления. Все  $X$  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.

Соответствие настраиваемой модели объекту оценивается критерием качества идентификации, который представляет собой средние потери настраиваемой модели

$$I(\alpha) = \int (y - \varphi(x, \alpha))^2 P(x, y) dx dy, \quad (10)$$

где  $P(x, y)$  – совместная плотность распределения переменных  $x$  и  $y$ . Чем меньше средние потери, тем выше качество идентификации [3].

Если  $x$  – неслучайная 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:

$$I(\alpha) = \frac{1}{N} \sum_{u=1}^N \int (y - \eta(x_u))^2 P(y | x_u) dy + \frac{1}{N} \sum_{u=1}^N (\eta(x_u) - \varphi(x_u, \alpha))^2. \quad (11)$$

Первая часть этой формулы связана с дисперсией величины  $\epsilon$ , а вторая – с качеством предложенной модели. Если дисперсия  $\sigma^2$  постоянна, то величина среднего риска задается формулой:

$$I(\alpha) = \sigma^2 + \frac{1}{N} \sum_{u=1}^N (\eta(x_u) - \varphi(x_u, \alpha))^2. \quad (12)$$

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:

$$L(j) = \int_{W_1} E(y - \eta_j(x, \alpha_j))^2 dx / \int_{W_1} dx, \quad (j = 1, 2, \dots, q). \quad (13)$$

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

$$L(j) = \sigma^2 + \int_{W_1} (\eta(x) - \eta_j(x, \alpha_j))^2 dx / \int_{W_1} dx. \quad (14)$$

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 оценка  $\alpha_j$ :

$$\hat{\alpha}_j = \Phi_j(X, Y), \quad (15)$$

где  $\Phi_j$  – функция от  $Y$ , плана  $D$  и класса моделей  $S_j$ .

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

$$\hat{\alpha}_j = (F_j^T F_j + r I_n)^{-1} F_j^T Y, \quad (16)$$

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  $x \in [-1, 1]$ :

$$L(q) = E \int_{W_1} x^T (\alpha - \hat{\alpha}(r)) (\alpha - \hat{\alpha}(r))^T x dx / \int_{W_q} dx = \sigma^2 sp X B J B X^T + \alpha^T (I_q - X^T X B) J (I_q - B X^T X) \alpha, \quad (17)$$

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

Формула (17) 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 – влияние смещения оценок, а вторая часть  $\sigma^2 sp X B J B X^T$  содержит дисперсию ошибки. Очевидно, что можно найти такое значение параметра регуляризации, при котором можно существенно уменьшить квадратическую ошибку оценок за счет их небольшого смещения.

Автором разработана итерационная процедура поиска оптимального параметра регуляризации  $\gamma$  в условиях мультиколлинеарности и определения неизвестных параметров регрессионной модели. Предполагается, что если функция потерь в случае ридж-регрессии  $ER(r)$  имеет локальный минимум, то он и соответствует оптимальному значению параметра регуляризации.

Для того, чтобы исследовать функцию потерь  $ER(r)$  на локальный экстремум, необходимо ее дифференцировать и приравнять нулю. Потом, найдя корни, определить, какой экстремум получится в этой точке (минимум или максимум). Если функция меняет знак в области экстремума с «–» на «+», то это минимум, если наоборот, то это максимум.

Произведя необходимые преобразования и вычисления, приходим к выводу, что функция потерь в случае ридж-регрессии имеет локальный минимум, что доказывает существование оптимального параметра регуляризации. И его значение можно вычислить по формуле:

$$r = \sigma^2 sp (X J X^T) / \alpha^T J X^T X \alpha. \quad (18)$$

Можно также определить необходимые и достаточные условия существования локального экстремума. Очевидно, что кроме  $\gamma$  это выражение зависит от следующих факторов:

- от плана эксперимента;
- от истинных коэффициентов; при больших коэффициентах затруднительно определить, будет ли вообще существовать локальный минимум;
- от дисперсии ошибки.

Следовательно, в дальнейшем можно ввести ограничения на получение локального минимума.

#### Пример

Пусть имеется класс вложенных моделей  $S_j$ :

Модель 1:  $y_1 = \alpha_0 + a_1 x_1$ .

Модель 2:  $y_2 = \alpha_0 + a_1 x_1 + \alpha_2 x_2$ .

Модель 3:  $y_3 = \alpha_0 + a_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3$ .

Присутствует сильная мультиколлинеарность. Первой в имитационном эксперименте анализируется максимальная модель из класса:  $y_3 = \alpha_0 + a_1x_1 + \alpha_2x_2 + \alpha_3x_3$ , коэффициенты модели берутся истинные, дисперсия  $\sigma^2=1$ . Параметр регуляризации  $r$  меняется от нуля до двух с шагом 0,1. Рассчитаем значение функции потерь для каждого  $r$ .

Пусть матрица плана эксперимента  $X_i$  имеет вид:

$$X1 = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix}; \quad X2 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -1 & 1 \\ -1 & -1 \end{bmatrix}; \quad X3 = \begin{bmatrix} -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \end{bmatrix}.$$

Разработано программное обеспечение на языке m-файлов MATLAB, в котором с помощью итерационной процедуры находится  $r$ , при котором производная  $\frac{dEI}{dr}$  обращается в ноль (это доказывает существование локального минимума) и вычисляется значение функции потерь в этой точке. Кроме этого рассчитывается значение оптимального  $r$  с помощью формулы (18). Полученные результаты подтверждают, что функция потерь имеет локальный минимум, которым и является  $r_{\text{опт}}$ . В таблице 1 представлены результаты нахождения оптимального значения параметра регуляризации  $r_{\text{опт}}$ , значений соответствующих ему функций потерь в зависимости от плана эксперимента. В качестве иллюстрирующего примера на рисунке 1 изображен график зависимости функции потерь от значения коэффициента регуляризации  $r$  для плана X1.

Таблица 1– Результаты имитационного эксперимента для модели 3

Вид плана	Значение $r_{\text{опт}}$	Значение $EI_{\text{min}}$
X1	1,00	0,267
X2	0,6	0,360
X3	0,5	0,071

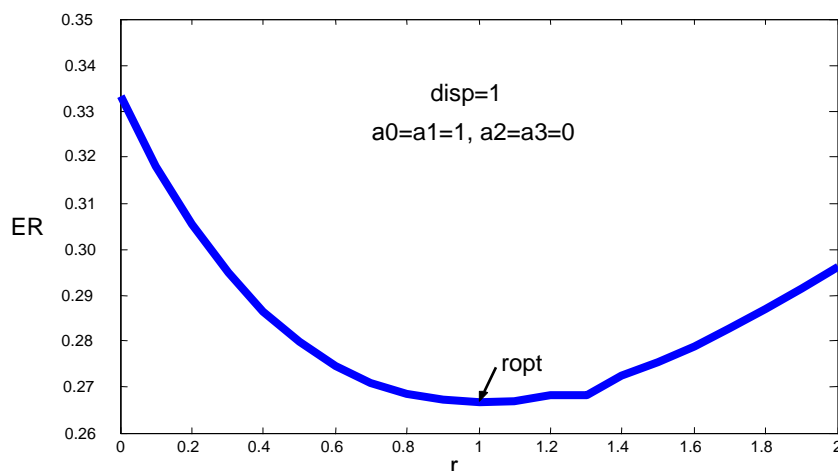


Рисунок 1 – График зависимости функции потерь от значения коэффициента регуляризации  $r$  для плана X1

Очевидно, что чем больше дисперсия и меньше истинные коэффициенты регрессии, тем больше величина  $r$  оптимального, т.е. значение оптимального параметра регуляризации прямо пропорционально дисперсии и обратно пропорционально квадрату коэффициентов истинной модели. Найти  $r$  оптимальное возможно не для всех планов, так как вывод формулы для  $r$  является оптимизационной задачей. Для планов, точки которых сильно отличаются от граничных, предлагается использовать итерационную процедуру.

Рассмотрим подтверждение того, что разработанная процедура нахождения оптимального коэффициента регуляризации может работать для неполных моделей из заданного класса на примере. На рисунке 2 показана зависимость функции потерь от параметра регуляризации для всех вложенных моделей.

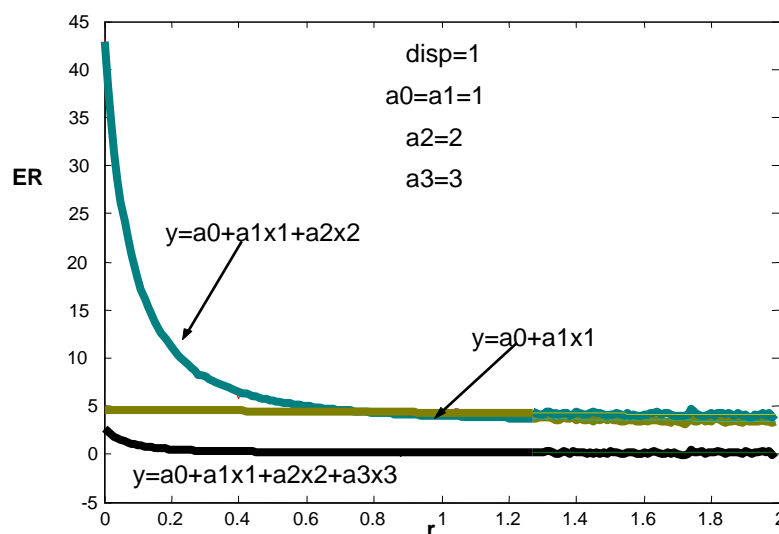


Рисунок 2 – Зависимость функции потерь от параметра регуляризации  $r$  для всего класса вложенных моделей

График подтверждает правильность разработанной процедуры для нахождения  $r_{opt}$  для любой модели из вложенного класса. Значение параметра регуляризации  $r_{opt}$  получается для разных моделей разное.

В результате применения разработанной методики проведено исследование функции потерь сложного объекта управления для оценки существования локального минимума, что доказало и позволило определить значение оптимального параметра регуляризации при смещенном оценивании параметров регрессионных уравнений. После проведенных имитационных экспериментов видно, что полученное в результате использования методики значение оптимального параметра регуляризации  $r_{opt}$  практически не отличается от экспериментального. Это подтверждает правильность разработанной методики.

Таким образом, можно прийти к выводу, что в рассматриваемой задаче достигнута поставленная цель, т.е. повышено качество прогнозирующей модели в условиях мультиколлинеарности с учетом неопределенности структуры модели и метода смещенного оценивания параметров модели, что свидетельствует о корректности использованных и разработанных методов управления, созданного программного обеспечения на языке m-файлов MATLAB, реализующего предложенную методику.

#### Перспективы дальнейших исследований

Предметом дальнейших исследований предполагается изучение взаимосвязи между построением модели СОУ методом смещенного оценивания регрессионных уравнений и критерием регуляризации Ивахненко А.Г.

#### 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 17.12.2008 г.