

Научная статья

УДК

JEL: E21, H41, I31, C51

DOI 10.25205/2542-0429-2024-24-2-99-134

Адаптированная гребневая регрессия на основе коэффициента байесовского правдоподобия и её применение в модели оценки влияния здоровья на благосостояние граждан Российской Федерации

Алексей Александрович Заболотский

Институт экономики и организации промышленного производства СО РАН,
Новосибирск, Россия

ieie@inbox.ru, <https://orcid.org/0000-0003-2683-6116>

Аннотация

Сравнивается метод пакетной реализации гребневой регрессии с математической моделью на основе оптимизированного байесовского критерия. Цель – проверить модель адаптированной гребневой регрессии на коррелирующих данных. Предлагаемая модель имеет преимущество, так как, помимо отбора переменных, как это делает классическая гребневая регрессия, позволяет оценить p -values построенной линейной регрессии. Кроме этого, метод позволяет регулировать параметры тьюнинга таких моделей, как гребневая регрессия или лассо, на более глубоком уровне. Методический подход применяется для многомерной выборки, в которой необходимо отсеять наименее значимые факторы, и при этом данные отличаются однородностью и разного рода выбросами. Данные взяты на основе опроса населения Российской Федерации, проведенного Высшей школой экономики.

Ключевые слова

гребневая регрессия, байесовский критерий, влияние здоровья, доход

Финансирование

Статья подготовлена по плану НИР ИЭОПП СО РАН, проект «Региональное и муниципальное стратегическое планирование и управление в контексте модернизации государственной региональной политики и развития цифровой экономики»

Для цитирования

Заболотский А. А. Адаптированная гребневая регрессия на основе коэффициента байесовского правдоподобия и её применение в модели оценки влияния здоровья на благосостояние граждан Российской Федерации // Мир экономики и управления. 2024. Т. 24, № 2. С. 99–134. DOI 10.25205/2542-0429-2024-24-2-99-134

© Заболотский А. А., 2024

Adapted Ridge Regression based on the Bayesian Likelihood Factor and its Application to a Model for Assessing the Impact of Health on the Well-being of Citizens of the Russian Federation

Alexey A. Zabolotsky

Institute of Economics and Industrial Engineering SB RAS,
Novosibirsk, Russian Federation

ieie@inbox.ru, <https://orcid.org/0000-0003-2683-6116>

Abstract

This article is aimed at their comparison of standard Ridge model based on default MSE evaluation of optimal lambda value with Bayesian criteria. Goal is to test adapted Ridge model on highly correlated data. Proposed method, unlike default one, has another advantage aside from conventional factors importance assessment – it gives the way for p-value evaluation based on these Ridge coefficients. Those method gives a chance to tune parameters more deeply. In this case, there is a multidimensional sample in which it is necessary to filter out the least significant factors. At the same time, the data is characterized by homogeneity and various types of outliers. This justifies the choice of the specified methods. Data is based on the poll conducted among the population of the Russian Federation under the guidance of Higher School of Economics.

Keywords

Ridge regression, Bayesian criteria, health impact, income

Funding

The article was prepared according to the research project «Regional and municipal strategic planning and management in the context of modernization of state regional policy and development of the digital economy»

For citation

Zabolotsky A.A. Adapted Ridge Regression based on the Bayesian Likelihood Factor and its Application to a Model for Assessing the Impact of Health on the Well-being of Citizens of the Russian Federation *World of Economics and Management*, 2024, vol. 24, no. 2, pp. 99–134. (in Russ.) DOI 10.25205/2542-0429-2024-24-2-99-134

Введение

Гребневая регрессия, или ридж-регрессия, – это метод, который можно использовать для подбора модели регрессии, когда в данных присутствует мультиколлинеарность [1], шум [2], неоднородности [3], аномалии [4] и нужно отбросить наименее значимые компоненты [5] или сгладить кривую интерполяции [6]. Безусловно, все эти методы требуют адаптации модели гребневой регрессии, однако сама модель, основанная на штрафе за аномальные значения, создает отправную точку исследования «сложных» многомерных данных.

Метод гребневой регрессии представляет собой «метод сжатия», который обычно применяется к регрессионной модели, когда существуют вариации, возникающие в результате коллинеарности предикторов.

В обычной регрессии минимизируется $RSS = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$. (1*)¹

¹ Machine Learning, Dr. S. Jackson, Machine Learning module MATH42815 of the Masters of Data Science course at Durham University. URL: <https://bookdown.org/ssjackson300/Machine-Learning-Lecture-Notes/ridge-regression.html>

где y_i – матрица значений зависимой переменной, $\beta_j x_{ij}$ – матрица независимых переменных и значений уравнений, β_0 – смещение.

Иными словами, в обычной линейной регрессии (OLS) оценки коэффициента β задаются по формуле²

$$\hat{\beta} = (X^T X)^{-1} X^T y. \quad (2^*)$$

Когда предикторы коллинеарные или почти коллинеарные, матрица $X^T X$ в (1*) становится сингулярной или почти сингулярной, тогда обратная матрица чувствительно реагирует на ошибки, что приводит к неустойчивости качества прогнозирования и оценок с помощью такой модели.

В гребневой регрессии подход основан на штрафе за размер вариации коэффициентов регрессии:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 = RSS + \lambda \sum_{j=1}^p \beta_j^2, \quad (3^*)^3$$

Аппроксимация модели Штраф

где λ – ключевой параметр тьюнинга модели, который при значении, равном 0, сводит модель к линейной регрессии. То есть это и есть фактически штаф модели.

Иначе говоря, минимизация – это $\left\{ \frac{\text{аппроксимация}}{RSS} + \text{штраф за размер коэффициентов} \right\}$.

Ридж-регрессия обеспечивает компромисс между смещением и дисперсией модели. При относительно небольшом смещении можно ожидать значительного уменьшения дисперсии и, следовательно, среднеквадратической ошибки:

$$MSE = E(\hat{\beta} - \beta)^2 = (E(\hat{\beta} - \beta))^2 + E(\hat{\beta} - E\hat{\beta})^2 = bias^2 + variance. \quad (4^*)^4$$

Это достигается применением штрафа, как уже показывалось выше (в матричной форме):

$$(y - X\beta)^T (y - X\beta) + \lambda \sum \beta_j^2. \quad (5^*)^5$$

β , который минимизирует функцию потерь, и есть формула ридж-регрессии, оценивающей β :

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y. \quad (6^*)^6$$

² Ruikun Xiao, Reed Coots, Yuzhe Ye. Ridge Regression | R, Group 13: <https://jbhender.github.io/Stats506/F17/Projects/G13/R.html#references>

³ Machine Learning, Dr. S. Jackson, 2023-03-29. URL: <https://bookdown.org/ssjackson300/Machine-Learning-Lecture-Notes/ridge-regression.html>

⁴ Ridge Regression | R, Group 13: Ruikun Xiao, Reed Coots, Yuzhe Ye <https://jbhender.github.io/Stats506/F17/Projects/G13/R.html#references>

⁵ Там же.

⁶ Там же.

Более общий вид: $E(\beta) = (X^T X + \lambda I)^{-1} (X^T X) \beta$ ⁷. При этом, как видно, $X^T X$ заменена на $(X^T X + \lambda I)^{-1}$, что и делает модель более устойчивой.

С геометрической точки зрения лассо и гребневая регрессия накладывают ограничения на веса модели. Можно сказать, что гребневая регрессия ограничивается весами, лежащими внутри гиперболы, радиус которой определяется параметром регуляризации. По аналогии, лассо ограничивает веса ромба таким образом, чтобы они находились внутри этого ромба, размер которого определяется параметром регуляризации. Это делается для фокусировки аппроксимирующего объема до минимального значения, позволяющего «генерализировать» значения. Это меньшее подпространство можно считать менее «сложным», чем полное пространство.

С байесовской точки зрения эта минимизация эквивалентна апостериорному распределению по всем возможным вариантам выбора весов. Например, уравнение колебаний белого шума $\eta(t)$ будет иметь схожий вид:

$$\xi(t) = s(t, \lambda) + \eta(t). \quad (7^*)^8$$

Здесь λ – неизвестная стохастическая величина. $S(t, \lambda)$ – векторные значения аппроксимации сигнала. И лассо, и гребневая регрессия эквивалентны оценке MAP⁹ после размещения априорных значений весов (лассо использует априорный лапласиан, а гребневая регрессия – априорный гауссовский метод). Например, умножение функции правдоподобия на узкую гауссову априорную функцию (что соответствует большому штрафу за гребень) дает более узкую апостериорную функцию. Одна из причин регуляризации – борьба с переобучением и слишком узкое сходжение с отдельной локальной выборкой. Это связано с тем, что менее ограниченное семейство вариаций предоставляет «больше» способов аппроксимации данных, и все более вероятно, что один из них сможет соответствовать случайным колебаниям в обучающем наборе. Для более формального рассмотрения см. компромисс между смещением и дисперсией (дилемма вариаций и смещения¹⁰). Это не обязательно означает, что выбор модели из более ограниченного семейства подойдет. Для достижения качества модели необходимо, чтобы ограниченное семейство действительно содержало репрезентативные модели. Это означает, что нужно выбрать априорное ограничение, которое хорошо соответствует конкретной задаче.

В данной работе оцениваются сильно коррелирующие параметры, влияющие на благосостояние граждан. Сложность таких оценок возникает из-за множества коррелирующих и разнородных факторов, что является проблемой [7] для построения модели. Именно поэтому выбрана гребневая регрессия как наиболее соответствующая данному типу задач с отсевом наиболее релевантных метрик и способная оптимизировать баланс величин входных данных путем штрафа.

⁷ Ridge Regression | R, Group 13: Ruikun Xiao, Reed Coots, Yuzhe Ye <https://jbhender.github.io/Stats506/F17/Projects/G13/R.html#references>

⁸ Bayesian interpretation of kernel regularization. URL: https://en.wikipedia.org/wiki/Bayesian_interpretation_of_kernel_regularization

⁹ maximum a posteriori probability (MAP).

¹⁰ Bias–variance tradeoff. URL: https://en.wikipedia.org/wiki/Bias–variance_tradeoff

Также применялись подходы на основе временных лагов здоровья граждан [8], которые выявили взаимосвязи. При таком подходе возникает сложность с выбором временного интервала с опять же коррелирующими лагами, так как здоровье от предыдущих лет, очевидно, влияет на последующие. Кроме этого, в ряде работ находилась взаимосвязь между географией, здоровьем и благополучием [9] и рассматривалась обратная взаимосвязь неравенства, здоровья и благополучия [10]. Во всех этих работах указывается на коррелирующие зависимости независимых переменных, их смешение и вариации, поэтому в нашей работе выбрана именно гребневая регрессия как способ решения данной проблемы наряду с выбором из множества этих переменных.

Описание модели

Модель оценивается на основании данных по мониторингу здоровья и экономического положения граждан РФ за 2015–2021 гг., собранных НИУ ВШЭ на основании опросов населения Российской Федерации¹¹. В данных собрано работоспособное население России – мужчины от 18 до 60 лет и женщины от 18 до 55 лет. После очистки отобраны данные с 3024 наблюдениями и 95 метриками.

Гребневая регрессия, как и лассо, как раз адаптирована к работе с большим количеством метрик сильно коррелирующих данных. В табл. 1 виден высокий коэффициент β , что предполагает плохую аппроксимацию и высокий штраф. Однако нет никакой гарантии, что меньший вес на самом деле лучше. Лассо и гребневая регрессия накладывают штраф на значения. Таким образом, при совпадении типа распределения переменных этот метод себя хорошо показывает. На следующем шаге мы находим лямбда-значение, которое даст наименьшее значение MSE ¹² путем k-fold кросс-вариации.

Таблица 1

Данные модели

Table 1

Summary model

	Length	Class	Mode
100	-none-	numeric	параметр
9500	dgCMatrix	S4	коэффициент бета
100	-none-	numeric	степень свободы
2	-none-	numeric	размерность данных
100	-none-	numeric	коэффициент лямбда
100	-none-	numeric	отклонение
1	-none-	numeric	нулевое смещение
1	-none-	numeric	количество прогонов модели
1	-none-	numeric	коэффициент
1	-none-	logical	тип смещения логический
4	-none-	call	количество вызовов
1	-none-	numeric	константа

¹¹ Данные РЛМС. <https://www.hse.ru/rlms/>
¹² mean squared error - MSE

Как видно, наиболее значимые по модулю – *oper* (были ли операции), *diab* (диабет), *aller* (аллергия), *goodsah* (хорошее здоровье по личному мнению), *other*, *healthp* (проблемы здоровья), *three_conl3* (три хронических с 3 лагами), *skin* (кожные), *uro* (урологические), *onco* (онкология), *five_conl1* (пять хронических с лагом¹), *heart* (сердечно-сосудистые). Следствием этого является то, что зависимая переменная Y в большей степени зависит от этих независимых переменных до оптимизации модели, с одной стороны, оказывающих ключевое влияние, с другой, приводящих к переобучению модели, когда обобщаемые закономерности еще не обработаны алгоритмом Ридж. Когда же они обработаны, то их влияние может быть нивелировано, что будет видно после штрафа, который может нивелировать их влияние, но при этом стабилизировать модель.

Аналогичные вычисления проведены для модели гребневой регрессии, адаптированной под байесовский критерий на основе линейной регрессии. Переменные для модели приведены в табл. 1 Приложения.

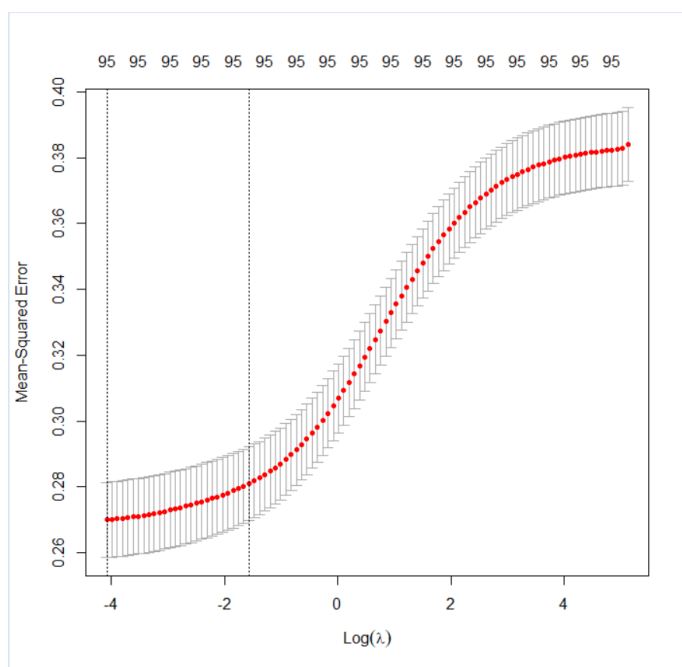


Рис. 1. Значения MSE и оптимальная λ

Fig. 1. MSE values and optimal λ

Полученные результаты – Multiple R-squared (0,346) и Adjusted R-squared (0,3251) показали низкую точность модели. Также здесь наиболее значимые переменные с p -value: $< 2.2e-16$, связанные с географической расположенностью. В OLS-модели, наоборот, они имели наименее значимые веса – уровень ***, *priv*, *north*, *centr*, *volga*, *cauca*, *ural*, *west*, *east*, *medm*, *largef*, *dms* и др. (Приложение,

табл. 2). То есть модель была сильно зашумлена, имела сильно коррелирующие веса других переменных, тем самым подавляла эти переменные до регуляризации. При этом R-квадрат невысокий. Параметр регуляризации (рис. 2) показал высокую степень штрафов. Чем круче «склон» аппроксимации, характеризующий переменную, тем выше штраф и влияние переменной:

$$\|\beta\|_2 = \sqrt{\sum_{j=1}^p \beta_j^2}. \quad (8^*)^{13}$$

На рис. 2 показаны коэффициенты гребневой регрессии для каждого из предикторов, построенные как функция λ . Как λ увеличивается (слева направо по оси X), оценки коэффициента гребневой регрессии сжимаются к нулю. Более узкая часть регуляризованных коэффициентов соответствует большему ограничению и меньшей сложности, поскольку высокая плотность соответствует меньшему набору параметров. Когда коэффициент λ велик, то все оценки коэффициентов гребневой регрессии (ГР) в основном равны нулю (что соответствует нулевой модели, которая не содержит предикторов). Следует отметить, что, хотя оценки ГР коэффициента имеют тенденцию к снижению в целом по мере роста λ , отдельные коэффициенты могут иногда увеличиваться по мере роста λ .¹⁴ Чем круче кривая на данном графике для каждой метрики, тем больший вклад вносит как во влияние на модель, так и на переобучение и взаимное влияние переменных.

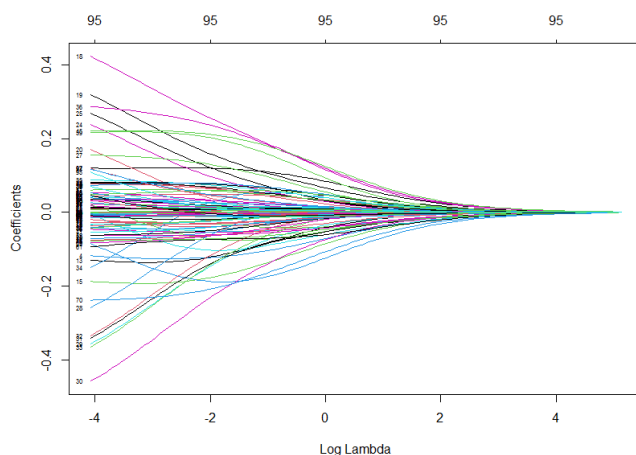


Рис. 2. Кривые регуляризации: чем они менее крутые, тем больший эффект оказывают на Y , и наоборот. Виден целый кластер переменных с относительно низким клоном и, следовательно, высоким уровнем влияния на зависимую переменную

Fig. 2. Regularization curves: the less steep they are, the greater the effect on Y , and vice versa. We can see a whole cluster of variables with a relatively low clone and, therefore, a high level of influence on the dependent variable

¹³ Machine Learning, Dr. S. Jackson, 2023-03-29. URL: <https://bookdown.org/ssjackson300/Machine-Learning-Lecture-Notes/ridge-regression.html>

¹⁴ Lasso and Ridge Regression in Python Tutorial. URL: <https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression>

Анализ остатков модели предсказания (рис. 3) показал неплохие результаты, т. е. предсказательная модель оценки с помощью R-квадрат достаточно точная.

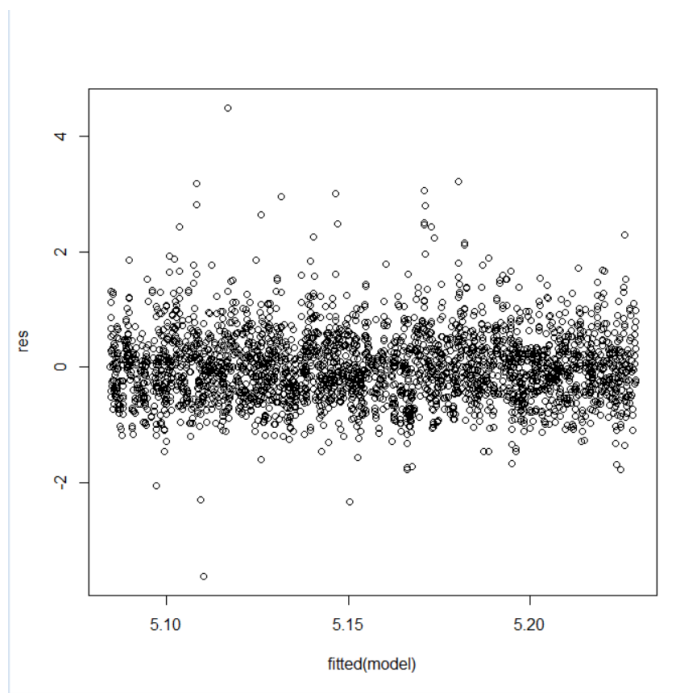


Рис. 3. График остатков

Fig. 3 Residuals graph

Остатки показаны как симметрично распределенные с небольшим отклонением от горизонтальной оси. Однако точность Adjusted R-squared 0.3251 низкая. Поэтому применим алгоритм адаптированной ридж-модели¹⁵, воссозданной алгоритмически из линейной регрессии со строкой ковариатной матрицы, аугментированной матрицей со значениями $\sqrt{\lambda}$ по диагонали [1]. При применении адаптированной модели ridge на базе BIC значений регрессии (см. Приложение, табл. 3) значения R-квадрат значительно улучшились до R-squared: 0.7778. Смысл этой оптимизации заключается в поиске плато стабильности BIC значений модели. Здесь также используется функция предварительной аугментации матрицы ковариации с матрицей, у которой по диагонали значения $\sqrt{\lambda}$ для аппроксимации к оштрафованной гребневой регрессии, которая в результате этих штрафов максимально сближается с исходными данными¹⁶.

¹⁵ Ridge regression in r with p-values and goodness of fit. URL: <https://stats.stackexchange.com/questions/171448/ridge-regression-in-r-with-p-values-and-goodness-of-fit>

¹⁶ На данной предварительной стадии с аугментированной матрицей достигнуты значения Multiple R-squared: 0.7845, Adjusted R-squared: 0.7777

Гребневая регрессия реализуется в виде аугментированной матрицы и матрицы ковариаций, последнее делается для устойчивости к вариациям. Функция воспроизводит гребневую регрессию через матрицу с корнями квадратными значений λ по диагонали [11].

Полученные данные Multiple R-squared: 0.7845, Adjusted R-squared: 0.7778, p -value: $< 2.2e-16$ уже значительно улучшают модель. Как видно, модель оставила значимыми коэффициенты, такие как география (значимые с тремя звездами), добавив *incsec*, *second*, *vocat*, *uni*. То есть наиболее устойчивые показатели связаны именно с географией, страховкой, частично профессией (технические профессии).

Анализ остатков показывает хорошее распределение остатков. Оценка R^2 вызывает доверие.

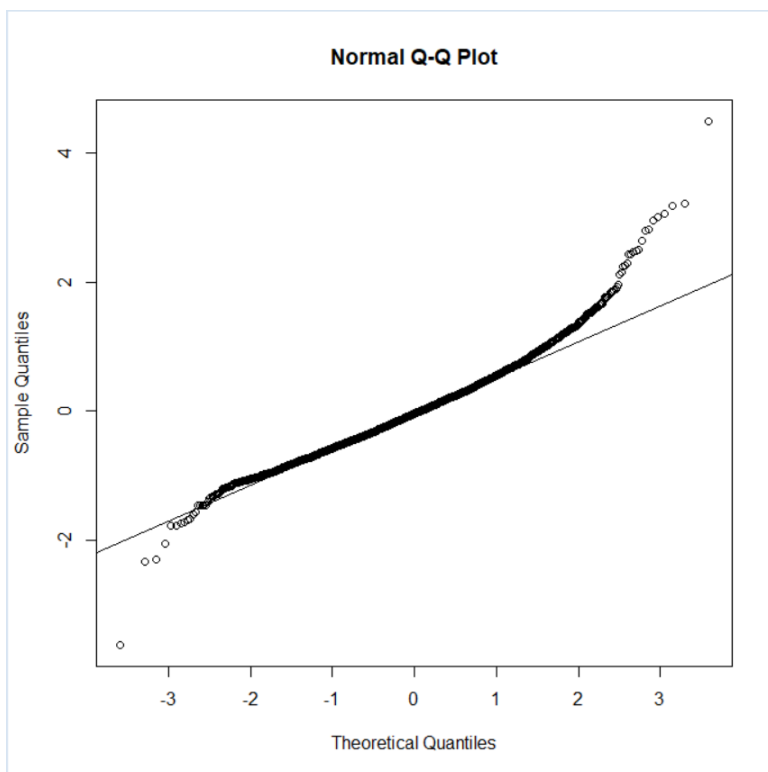


Рис. 4. Графики Q-Q plot

Fig. 4. Q-Q plot graphs

Рис. 5 показывает, что модель хорошо ложится в равномерное распределение $\text{Uniform}(-1, 1)$ или распределения Стьюдента с 2 степенями свободы $> \text{plot}(\text{density}(\text{res}))$.

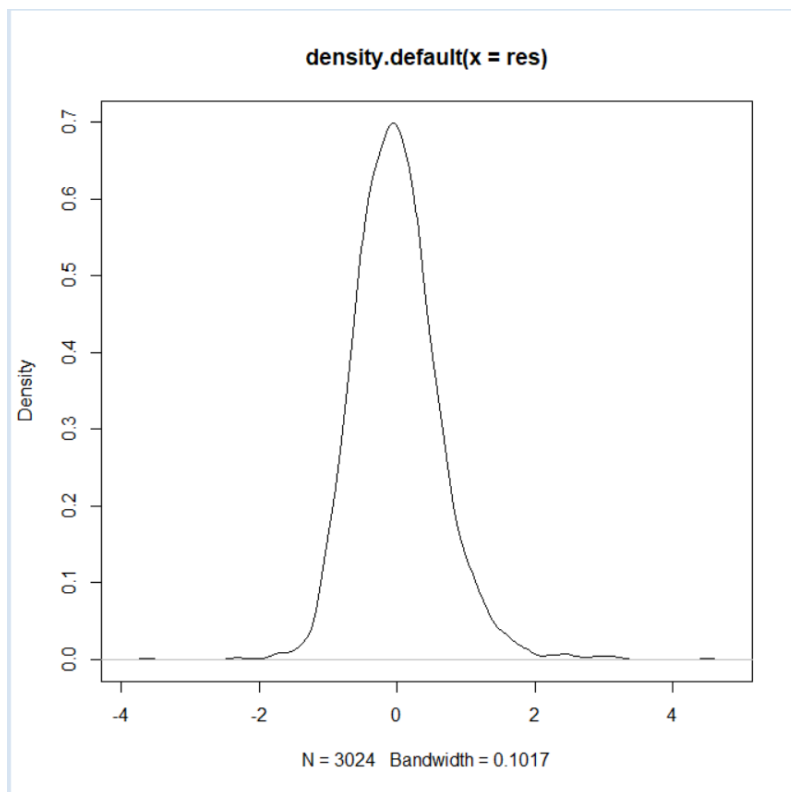


Рис. 5. Распределение данных

Fig. 5. Data distribution

Как видно, данные хорошо просцируются в нормальное распределение, что говорит о высоком влиянии целого набора устойчивых факторов.

Сравнение моделей и обсуждение данных

Можно сделать выводы по полученным методам линейной регрессии, ридж, ридж для лучшей лямбда и ридж оптимизированной по байесовскому коэффициенту правдоподобия (см. Приложение, табл. 4). Модели имеют отличия, хотя OLS- и BIC-оптимизированная гребневая регрессии имеют наиболее близкие коэффициенты. Сравнение коэффициентов OLS, гребневой регрессии и гребневой регрессии при лучшей лямбда и BIC-оптимизированной гребневой регрессии представлено в таблице ниже. Чем выше коэффициент, тем он оказывает более высокое влияние на модель.

Данные OLS модели в целом хорошо совпали с аппроксимированной гребневой регрессией – `lm1` и `best_model` или 1 и 3 колонки, что говорит об устойчивости модели ридж для этого набора данных. В целом модель оптимизированной гребневой регрессии не так сильно изменила коэффициенты, однако точность выросла

значительно – с 32 до 78 %, при этом оптимизированная модель даст возможность оценить значимость коэффициентов.

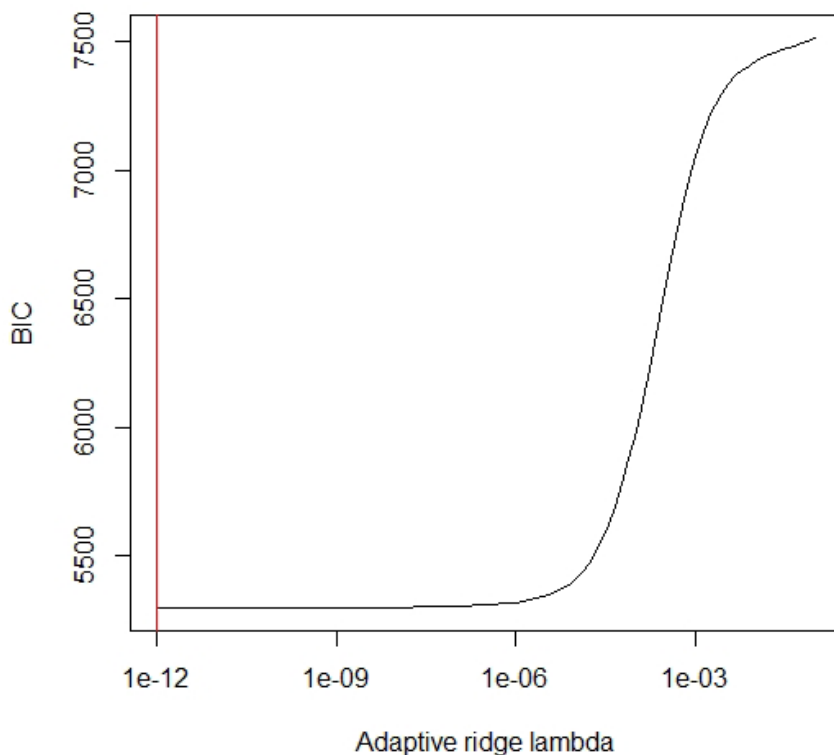


Рис. 6. Значения лямбда для оптимизированной модели
Fig. 6. Lambda values for the optimized model

Видно, что, например, на центральном участке коэффициенты до оптимизации имели «яму». После данные гребневой регрессии приблизились по значению к верхним значениям для линейной регрессии. Помимо байесовского коэффициента, для оценки качества аппроксимации можно было бы использовать и AIC, и Adjusted R2. По умолчанию используется MSE, который дал точность 34 %, однако BIC показал явное преимущество, повысив точность до 78 %.

Выводы

Таким образом, гребневая регрессия показала хорошую устойчивость модели, показав, в отличие от пакетной модели, что при оптимизации через аугментацию с помощью диагональной матрицы с корнями квадратными из значений лямбда¹⁷ регрессия указывает на наиболее значимые коэффициенты модели в ус-

¹⁷ How to derive the ridge regression solution? <https://stats.stackexchange>

ловиях высокой коллинеарности и корреляции данных при точности в 78 против 34 % при пакетной реализации гребневой регрессии. Наиболее значимыми оказались параметры профессии, страхования, географических и региональных локаций, а также возраст и безопасность, ДМС, выявленный диабет, наличие операции, размер предприятия, который влияет на шумность и стрессовость, по всей видимости, прохождение программ чекап, наличие нескольких хронических заболеваний. Устойчивая адаптированная модель не нашла взаимосвязи влияния на здоровье после подавления переобучения. Однако были найдены не прямые взаимосвязи, которые влияют и на здоровье, и на благосостояние. Возможно, это связано с тем, что граждане с вышеуказанными заболеваниями не ставят задачу максимизации дохода или не имеют существенного дохода. То есть можно говорить, что это факторы, подавляющие значимость самой зависимой переменной. В то же время указанные влияющие зависимости географии, ДМС и профессии сами по себе влияют на доход либо имеют взаимное влияние, так как здоровый человек стремится максимизировать доход и меняет локацию и, вероятно, приобретает ДМС. Имеется динамическое влияние переменных, что также порождает сильно коррелирующие зависимости, которые и были подавлены гребневой регрессией. В целом гребневая регрессия, как и лассо и эластик, может показать примерно одинаково хорошие результаты на таких данных.

Комбинированная аугментация данных и оптимизация по коэффициенту BIC улучшили результат и точность модели, однако последнее незначительно улучшило модель в сравнении с предыдущим шагом (аугментации). Иначе говоря, модель не нашла зависимости специализированных факторов, таких как тип заболевания, во влиянии на благосостояние. Более того, самооценки здоровья также не влияли. Скорее всего, это связано с тем, что люди предпочитают работать, несмотря на самочувствие, и не обращаются за помощью.

Приложение

Таблица 1

Коэффициенты β в модели.

Данные для некоторых переменных агрегированы для исключения перекрестного влияния друг на друга [12]

Переменные	Коэффициенты
1	2
X	5,431142
age (возраст)	0,000001
agesq (возраст квадратичный)	-0,003192
incsec (неполное среднее образование)	-0,002315
second (среднее)	-0,119292

Продолжение табл. 1

1	2
vocat (среднее профессиональное)	–0,036295
uni (университет)	–0,047721
goodsah (хорошее здоровье – опрос)	0,117540
badsah (плохое – опрос)	0,077297
mard (в браке)	–0,027795
div (разведн–а)	–0,002589
wid (вдова, вдовец)	–0,006255
town (город)	0,017421
pgt (поселок городского типа)	–0,129602
rural (село)	–0,018973
exper (опыт работы – человек работал)	–0,187111
expnewsq (опфт в квадрате)	0,001555
manag (профессия менеджер)	–0,000266
prof (профессия – топ-менеджеры)	0,425068
tech (технические профессии)	0,321498
clerk (клерк)	0,173139
servshop (работники услуги и продаж)	0,023801
agrfish (сельское хозяйство, рыбное, лесное)	–0,083102
crafttrade (ручной труд)	–0,003620
machop (машинный труд)	0,240468
kids18 (дети есть или нет)	0,270653
priv (частный сектор предприятия)	0,042637
north (регион Север)	0,155652
centr (Центральный регион)	–0,260323
volga (Волжский регион)	–0,358663
cauca (Кавказ)	–0,459713
ural (Урал)	–0,343704
west (запад)	–0,336085
east (восток)	–0,367122
medmf (предприятие среднее – где работает 100–1000 человек)	–0,152154
largef (>1000)	0,087675
oper (операции)	0,286527
diab (диабет)	0,082103

Продолжение табл. 1

1	2
infar (был ли инфаркт)	0,078511
stroke (сердечный приступ)	0,216437
anem (анемия)	–0,070976
disab (инвалидность)	–0,053055
healthp (проблемы здоровья)	–0,061962
checkup (программа чекап)	0,079294
dms (наличие ДМС)	0,008843
heart (сердце)	0,220691
lungs (легкие)	–0,008609
liver (печень)	–0,013602
kidney (почки)	0,032072
gastro (гастроэнтерологические заболевания)	0,036876
spine (позвоночник)	–0,034841
diabSR (самовыявленный диабет)	0,022926
hyper (гипертония)	–0,040201
joint (болезнь суставов)	–0,001217
respir (респираторные)	0,023464
neuro (нервные заболевания)	–0,061758
eye (проблемы со зрением)	–0,074616
aller (аллергия)	0,061106
vein (вены)	0,072826
skin (кожные заболевания)	–0,014311
onco (онкология)	–0,082535
uro (урология)	–0,093978
other (другие)	–0,009152
one_con (1 хроническое)	0,005473
two_con (2 хронических)	–0,006546
three_con (3 хронических)	0,042214
four_con (4 хронических)	0,053623
five_con (5 хронических)	0,013677
satlife (удовлетворенность жизнью по опросу)	–0,019961
nsatcond (не удовлетворен условиями жизни)	0,047844
desind100 (disease index – вес каждого заболевания) – от 0 до 100 для человека	–0,238188

Окончание табл. 1

1	2
goodsahl1 (хорошее здоровье в группе опроса по возрасту – 1 лаг назад для исключения взаимного влияния переменных)	0,000050
goodsahl2 (хорошее здоровье в группе опроса по возрасту – 2 лага назад)	–0,012454
goodsahl3 (хорошее здоровье в группе опроса по возрасту – 3 лага назад)	0,011141
badsahl1 (плохое здоровье по опросу – 1 лаг назад)	–0,027492
badsahl2 (плохое здоровье по опросу – 2 лага назад)	–0,077938
badsahl3 (плохое здоровье по опросу – 3 лага назад)	–0,041940
desind100l1 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 1	–0,031452
desind100l2 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 2	0,000064
desind100l3 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 3	–0,002099
one_conl1 (одно хроническое заболевание – 1 лаг в прошлом году для удаления причинности)	–0,002787
one_conl2 (одно хроническое заболевание – 2 лага в прошлом году для удаления причинности)	–0,001406
one_conl3 (одно хроническое заболевание – 3 лага в прошлом году для удаления причинности)	0,040717
two_conl1 (два хронических заболевания – 1 лаг)	0,028415
two_conl2 (два хронических заболевания – 2 лага)	–0,015590
two_conl3 (два хронических заболевания – 3 лага)	–0,005625
three_conl1 (три хронических заболевания лаг 1)	0,036892
three_conl2 (три хронических заболевания лаг 2)	–0,047681
three_conl3 (три хронических заболевания лаг 3)	0,056640
four_conl1 (четыре хронических заболевания – 1 лаг) (четыре хронических заболевания – лагов 1)	0,075391
four_conl2 (четыре хронических заболевания – лагов 2)	0,025081
four_conl3 (четыре хронических заболевания – лагов 3)	0,050503
five_conl1 (пять хронических заболеваний – 1 лаг)	0,127341
five_conl2 (пять хронических заболеваний – 2 лага)	–0,009192
five_conl3 (пять хронических заболеваний – 3 лага)	0,123786

Таблица 2

Коэффициенты регрессии на базе ридж-неоптимизированной модели

Переменные	Estimate	Std. Error	t value	Pr (> t)	Значи- мость
1	2	3	4	5	6
(Intercept)	5,024e+00	5,878e-01	8,547	2,00E-16	***
X	1,348e-06	8,824e-07	1,528	0,126726	
age (возраст)	-8,092e-03	1,229e-02	-0,658	0,510424	
agesq (возраст ква- дратичный)	3,182e-03	1,442e-02	0,221	0,825294	
incsec (неполное среднее образова- ние)	4,002e-01	5,156e-01	0,776	0,437753	
second (среднее)	4,821e-01	5,149e-01	0,936	0,349174	
vocat (среднее про- фессиональное)	4,648e-01	5,151e-01	0,902	0,366873	
uni (университет)	6,217e-01	5,153e-01	1,207	0,227717	
goodsah (хорошее здоровье – опрос)	7,920e-02	2,509e-02	3,157	0,001613	**
badsah (плохое – опрос)	-3,189e-02	7,906e-02	-0,403	0,686661	
mard (в браке)	-2,130e-03	5,721e-02	-0,037	0,970303	
div (разведн-а)	-9,590e-03	6,342e-02	-0,151	0,879818	
wid (вдова, вдовец)	2,614e-02	8,276e-02	0,316	0,752087	
town (город)	-1,138e-01	2,503e-02	-4,546	5,70e-06	***
pgt (поселок город- ского типа)	1,455e-03	4,169e-02	0,035	0,972157	
rural (село)	-1,714e-01	2,923e-02	-5,864	5,02e-09	***
exper (опыт работы – человек работал)	2,069e-03	3,559e-03	0,581	0,561098	
expnewsq (опфт в квадрате)	-2,017e-03	1,165e-02	-0,173	0,862591	
manag (профессия менеджер)	5,389e-01	5,714e-02	9,430	<	2,00E- 16
prof (профессия – топ-менеджеры)	4,377e-01	5,411e-02	8,089	8,71e-16	***
tech (технические профессии)	2,799e-01	5,159e-02	5,425	6,27e-08	***

Продолжение табл. 2

1	2	3	4	5	6
clerk (клерк)	1,148e-01	6,116e-02	1,877	0,060575	,
servshop (работники услуги и продаж)	1,857e-02	5,112e-02	0,363	0,716382	
agrfish (сельское хозяйство, рыбное, лесное)	1,092e-01	3,268e-01	0,334	0,738184	
crafttrade (ручной труд)	3,466e-01	5,249e-02	6,603	4,75e-11	***
machop (машинный труд)	3,761e-01	5,163e-02	7,284	4,15e-13	***
kids18 (дети есть или нет)	3,787e-02	2,730e-02	1,387	0,165594	
priv (частный сектор предприятия)	1,605e-01	2,102e-02	7,635	3,04e-14	***
north (регион Север)	-3,686e-01	6,327e-02	-5,826	6,31e-09	***
centr (Центральный регион)	-4,722e-01	4,278e-02	-11,039	2,00E-16	***
volga (Волжский регион)	-5,756e-01	4,335e-02	-13,276	2,00E-16	***
cauca (Кавказ)	-4,627e-01	4,728e-02	-9,786	2,00E-16	***
ural (Урал)	-4,532e-01	4,341e-02	-10,441	2,00E-16	***
west (запад)	-4,854e-01	4,667e-02	-10,400	2,00E-16	***
east (восток)	-2,624e-01	5,192e-02	-5,054	4,60e-07	***
medmf (предприятие среднее – где работает 100–1000 человек)	9,092e-02	2,468e-02	3,684	0,000233	***
largef (>1000)	2,994e-01	3,580e-02	8,365	2,00E-16	***
oper (операции)	8,236e-02	6,653e-02	1,238	0,215852	
diab (диабет)	7,834e-02	6,952e-02	1,127	0,259884	
infar (был ли ин-фаркт)	2,040e-01	1,474e-01	1,384	0,166560	
stroke (сердечный приступ)	-7,390e-02	1,520e-01	-0,486	0,626957	
anem (анемия)	-5,303e-02	6,941e-02	-0,764	0,444935	
disab (инвалидность)	-6,451e-02	1,011e-01	-0,638	0,523445	

Продолжение табл. 2

1	2	3	4	5	6
healthp (проблемы здоровья)	7,838e-02	2,423e-02	3,235	0,001229	**
checkup (программа чекап)	6,703e-03	2,128e-02	0,315	0,752794	
dms (наличие ДМС)	2,163e-01	3,754e-02	5,761	9,21e-09	***
heart (сердце)	-1,679e-02	6,441e-02	-0,261	0,794323	
lungs (легкие)	-3,172e-02	5,570e-02	-0,569	0,569098	
liver (печень)	3,141e-02	5,917e-02	0,531	0,595549	
kidney (почки)	3,794e-02	5,774e-02	0,657	0,511119	
gastro (гастроэнтерологические заболевания)	-4,482e-02	4,883e-02	-0,918	0,358732	
spine (позвоночник)	1,732e-02	5,188e-02	0,334	0,738468	
diabSR (самовыявленный диабет)	-5,316e-02	6,271e-02	-0,848	0,396663	
hyper (гипертония)	-1,632e-03	4,984e-02	-0,033	0,973888	
joint (болезнь суставов)	2,355e-02	5,216e-02	0,451	0,651731	
respir (респираторные)	-6,962e-02	5,183e-02	-1,343	0,179242	
neuro (нервные заболевания)	-8,306e-02	7,169e-02	-1,159	0,246712	
eye (проблемы со зрением)	5,116e-02	4,966e-02	1,030	0,303043	
aller (аллергия)	5,708e-02	5,069e-02	1,126	0,260240	
vein (вены)	-2,351e-02	5,398e-02	-0,435	0,663286	
skin (кожные заболевания)	-9,152e-02	7,190e-02	-1,273	0,203178	
onco (онкология)	-1,135e-01	1,571e-01	-0,722	0,470098	
uro (урология)	-1,421e-02	8,656e-02	-0,164	0,869589	
other (другие)	-1,276e-02	1,473e-01	-0,087	0,931007	
one_con (1 хроническое)	1,659e-03	4,885e-02	0,034	0,972916	
two_con (2 хронических)	6,143e-02	8,532e-02	0,720	0,471594	
three_con (3 хронических)	8,006e-02	1,241e-01	0,645	0,518992	

Продолжение табл. 2

1	2	3	4	5	6
four_con (4 хронических)	5,015e-02	1,627e-01	0,308	0,757965	
five_con (5 хронических)	3,024e-02	2,281e-01	0,133	0,894521	
satlife (удовлетворенность жизнью по опросу)	4,097e-02	2,146e-02	1,909	0,056368	,
nsatcond (не удовлетворен условиями жизни)	-2,401e-01	2,546e-02	-9,430	2,00E-16	***
desind100 (disease index – вес каждого заболевания) – от 0 до 100 для человека	NA	NA	NA	NA	
goodsahl1 (хорошее здоровье в группе опроса по возрасту – 1 лаг назад для исключения взаимного влияния переменных)	-1,722e-02	2,520e-02	-0,684	0,494329	
goodsahl2 (хорошее здоровье в группе опроса по возрасту – 2 лага назад)	1,318e-02	2,508e-02	0,525	0,599328	
goodsahl3 (хорошее здоровье в группе опроса по возрасту – 3 лага назад)	-2,449e-02	2,443e-02	-1,002	0,316238	
badsahl1 (плохое здоровье по опросу – 1 лаг назад)	-8,667e-02	8,927e-02	-0,971	0,331733	
badsahl2 (плохое здоровье по опросу – 2 лага назад)	-3,337e-02	7,177e-02	-0,465	0,641977	
badsahl3 (плохое здоровье по опросу – 3 лага назад)	-2,310e-02	6,821e-02	-0,339	0,734862	

Продолжение табл. 2

1	2	3	4	5	6
desind10011 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 1	4,251e–03	5,351e–03	0,794	0,427068	
desind10012 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 2	–4,535e–03	4,330e–03	–1,047	0,295056	
desind10013 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 3	–7,501e–03	4,225e–03	–1,776	0,075900	,
one_con11 (одно хроническое заболевание – 1 лаг в прошлом году для удаления причинности)	–3,432e–02	5,023e–02	–0,683	0,494430	
one_con12 (одно хроническое заболевание – 2 лага в прошлом году для удаления причинности)	5,951e–02	4,413e–02	1,349	0,177587	
one_con13 (одно хроническое заболевание – 3 лага в прошлом году для удаления причинности)	6,568e–02	4,241e–02	1,549	0,121534	
two_con11 (два хронических заболевания – 1 лаг)	–8,312e–02	8,959e–02	–0,928	0,353572	
two_con12 (два хронических заболевания – 2 лага)	2,578e–02	7,703e–02	0,335	0,737879	
two_con13 (два хронических заболевания – 3 лага)	1,105e–01	7,320e–02	1,510	0,131165	

Окончание табл. 2

1	2	3	4	5	6
three_conl1 (три хронических заболевания – лаг 1)	–1,541e–01	1,306e–01	–1,180	0,238186	
three_conl2 (три хронических заболевания – лаг 2)	1,102e–01	1,100e–01	1,002	0,316525	
three_conl3 (три хронических заболевания – лаг 3)	1,879e–01	1,066e–01	1,763	0,077942	,
four_conl1 (четыре хронических заболевания – 1 лаг) (четыре хронических заболевания – лагов 1)	–1,024e–01	1,706e–01	–0,600	0,548632	
four_conl2 (четыре хронических заболевания – лагов 2)	1,265e–01	1,455e–01	0,869	0,384824	
four_conl3 (четыре хронических заболевания – лагов 3)	2,772e–01	1,428e–01	1,942	0,052268	,
five_conl1 (пять хронических заболеваний – 1 лаг)	–1,854e–01	2,309e–01	–0,803	0,422146	
five_conl2 (пять хронических заболеваний – 2 лага)	2,193e–01	2,003e–01	1,095	0,273617	
five_conl3 (пять хронических заболеваний – 3 лага)	3,267e–01	1,949e–01	1,677	0,093712	

Таблица 3

Коэффициенты регрессии на базе гребневой модели,
оптимизированной по BIC «плато»
и аугментацией (генерация значений по диапазону) матрицы ковариации

	Estimate	Std. Error	t value	Pr (> t)	
1	2	3	4	5	6
(Intercept)	0,039370	0,051840	0,759000	0,447665	
X	0,000002	0,000001	2,101000	0,035736	*
age (возраст)	0,038170	0,010990	3,475000	0,000519	***
agesq (возраст ква- дратичный)	−0,049030	0,013000	−3,773000	0,000164	***
incsec (неполное среднее образование)	4,260000	0,244700	17,405000	0,000000	***
second (среднее)	4,347000	0,242100	17,957000	0,000000	***
vocat (среднее про- фессиональное)	4,327000	0,243000	17,806000	0,000000	***
uni (университет)	4,483000	0,243700	18,396000	0,000000	***
goodsah (хорошее здоровье – опрос)	0,085470	0,024990	3,421000	0,000632	***
badsah (плохое – опрос)	−0,022880	0,078760	−0,291000	0,771417	
mard (в браке)	0,040990	0,056770	0,722000	0,470337	
div (разведн–а)	0,027170	0,063040	0,431000	0,666548	
wid (вдова, вдовец)	0,069210	0,082300	0,841000	0,400428	
town (город)	−0,118100	0,024930	−4,735000	0,000002	***
pgt (поселок город- ского типа)	0,006427	0,041530	0,155000	0,877026	
rural (село)	−0,162600	0,029100	−5,587000	0,000000	***
exper (опыт работы – человек работал)	0,000220	0,003539	0,062000	0,950403	
exrnewsq (опфт в квадрате)	0,003310	0,011590	0,286000	0,775256	
manag (профессия менеджер)	0,566700	0,056840	9,970000	0,000000	***
profg (профессия – топ-менеджеры)	0,467500	0,053790	8,691000	0,000000	***
tech (технические профессии)	0,308100	0,051290	6,007000	0,000000	***

Продолжение табл. 3

1	2	3	4	5	6
clerk (клерк)	0,141300	0,060860	2,322000	0,020322	*
servshop (работники услуги и продаж)	0,052900	0,050770	1,042000	0,297509	
agrfish (сельское хозяйство, рыбное, лесное)	0,107900	0,325600	0,331000	0,740469	
crafttrade (ручной труд)	0,381000	0,052140	7,307000	0,000000	***
machop (машинный труд)	0,407900	0,051310	7,951000	0,000000	***
kids18 (дети есть или нет)	0,045090	0,027190	1,658000	0,097323	,
priv (частный сектор предприятия)	0,167600	0,020920	8,008000	0,000000	***
north (регион Север)	-0,346700	0,062980	-5,506000	0,000000	***
centr (Центральный регион)	-0,445500	0,042500	-10,482000	0,000000	***
volga (Волжский регион)	-0,551600	0,043100	-12,798000	0,000000	***
cauca (Кавказ)	-0,441500	0,047040	-9,385000	0,000000	***
ural (Урал)	-0,421700	0,043090	-9,787000	0,000000	***
west (запад)	-0,463500	0,046430	-9,982000	0,000000	***
east (восток)	-0,244200	0,051680	-4,726000	0,000002	***
medmf (предприятия среднее – где работает 100–1000 человек)	0,095080	0,024580	3,868000	0,000112	***
largef (>1000)	0,302200	0,035660	8,473000	0,000000	***
oper (операции)	0,084970	0,066280	1,282000	0,199938	
diab (диабет)	0,091270	0,069240	1,318000	0,187568	
infar (был ли инфаркт)	0,228800	0,146800	1,558000	0,119282	
stroke (сердечный приступ)	-0,059610	0,151500	-0,394000	0,693967	
anem (анемия)	-0,048520	0,069150	-0,702000	0,482899	
disab (инвалидность)	-0,062130	0,100700	-0,617000	0,537368	
healthp (проблемы здоровья)	0,076150	0,024140	3,155000	0,001619	**

Продолжение табл. 3

1	2	3	4	5	6
checkup (программа чекап)	0,008016	0,021200	0,378000	0,705376	
dms (наличие ДМС)	0,216000	0,037400	5,776000	0,000000	***
heart (сердце)	-0,025500	0,064160	-0,398000	0,691027	
lungs (легкие)	-0,027140	0,055490	-0,489000	0,624844	
liver (печень)	0,020590	0,058940	0,349000	0,726886	
kidney (почки)	0,029240	0,057510	0,508000	0,611240	
gastro (гастроэнтерологические заболевания)	-0,047380	0,048650	-0,974000	0,330182	
spine (позвоночник)	0,002187	0,051660	0,042000	0,966238	
diabSR (самовыявленный диабет)	-0,060990	0,062470	-0,976000	0,328969	
hyper (гипертония)	-0,010340	0,049650	-0,208000	0,835008	
joint (болезнь суставов)	0,031360	0,051960	0,603000	0,546255	
respir (респираторные)	-0,074780	0,051630	-1,448000	0,147589	
neuro (нервные заболевания)	-0,086030	0,071430	-1,204000	0,228508	
eye (проблемы со зрением)	0,045220	0,049470	0,914000	0,360793	
aller (аллергия)	0,049910	0,050490	0,989000	0,322984	
vein (вены)	-0,038830	0,053750	-0,722000	0,470085	
skin (кожные заболевания)	-0,101700	0,071620	-1,420000	0,155587	
opco (онкология)	-0,123200	0,156500	-0,787000	0,431114	
uro (урология)	-0,022320	0,086230	-0,259000	0,795812	
other (другие)	-0,026740	0,146800	-0,182000	0,855434	
one_con (1 хроническое)	0,014290	0,048640	0,294000	0,768880	
two_con (2 хронических)	0,077510	0,084980	0,912000	0,361801	
three_con (3 хронических)	0,096490	0,123600	0,780000	0,435239	
four_con (4 хронических)	0,071530	0,162100	0,441000	0,659059	

Продолжение табл. 3

1	2	3	4	5	6
five_con (5 хронических)	0,070440	0,227200	0,310000	0,756509	
satlife (удовлетворенность жизнью по опросу)	0,041300	0,021380	1,931000	0,053532	,
nsatcond (не удовлетворен условиями жизни)	–0,230500	0,025340	–9,094000	0,000000	***
desind100 (disease index – вес каждого заболевания) – от 0 до 100 для человека	NA	NA	NA	NA	
goodsahl1 (хорошее здоровье в группе опроса по возрасту – 1 лаг назад для исключения взаимного влияния переменных)	–0,008467	0,025080	–0,338000	0,735712	
goodsahl2 (хорошее здоровье в группе опроса по возрасту – 2 лага назад)	0,007859	0,024980	0,315000	0,753029	
goodsahl3 (хорошее здоровье в группе опроса по возрасту – 3 лага назад)	–0,012710	0,024300	–0,523000	0,601050	
badsahl1 (плохое здоровье по опросу – 1 лаг назад)	–0,091020	0,088940	–1,023000	0,306206	
badsahl2 (плохое здоровье по опросу – 2 лага назад)	–0,030190	0,071500	–0,422000	0,672882	
badsahl3 (плохое здоровье по опросу – 3 лага назад)	–0,023600	0,067950	–0,347000	0,728376	
desind100l1 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 1	0,005293	0,005330	0,993000	0,320816	

Продолжение табл. 3

1	2	3	4	5	6
desind10012 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 2	–0,004118	0,004314	–0,955000	0,339881	
desind10013 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 3	–0,007342	0,004209	–1,744000	0,081180	,
one_conl1 (одно хроническое заболевание – 1 лаг в прошлом году для удаления причинности)	–0,042160	0,050030	–0,843000	0,399531	
one_conl2 (одно хроническое заболевание – 2 лага в прошлом году для удаления причинности)	0,057440	0,043960	1,307000	0,191433	
one_conl3 (одно хроническое заболевание – 3 лага в прошлом году для удаления причинности)	0,070570	0,042250	1,670000	0,094945	,
two_conl1 (два хронических заболевания – 1 лаг)	–0,081140	0,089250	–0,909000	0,363357	
two_conl2 (два хронических заболевания – 2 лага)	0,017180	0,076740	0,224000	0,822848	
two_conl3 (два хронических заболевания – 3 лага)	0,110300	0,072930	1,513000	0,130487	
three_conl1 (три хронических заболевания лаг 1)	–0,167000	0,130100	–1,284000	0,199323	

Окончание табл. 3

1	2	3	4	5	6
three_conl2 (три хронических заболеваний лаг 2)	0,101300	0,109600	0,924000	0,355355	
three_conl3 (три хронических заболеваний лаг 3)	0,184500	0,106200	1,738000	0,082332	,
four_conl1 (четыре хронических заболеваний – 1 лаг) (четыре хронических заболеваний лагов 1)	–0,114000	0,170000	–0,670000	0,502647	
four_conl2 (четыре хронических заболеваний – лагов 2)	0,119000	0,145000	0,821000	0,411688	
four_conl3 (четыре хронических заболеваний – лагов 3)	0,273200	0,142200	1,921000	0,054805	,
five_conl1 (пять хронических заболеваний – 1 лаг)	–0,207000	0,230000	–0,900000	0,368312	
five_conl2 (пять хронических заболеваний – 2 лага)	0,206200	0,199600	1,033000	0,301688	
five_conl3 (пять хронических заболеваний – 3 лага)	0,320400	0,194100	1,650000	0,098980	,

Таблица 4

Сравнение значений регрессии

Модель	OLS	Гребневая регрессия	Гребневая регрессия при лучшей лямбда	BIC-оптимизированная гребневая регрессии
1	2	3	4	5
	lm1	cv_model	best_model	Ridge_BIC
(Intercept)	5,023971	5,295742	5,431142	0,039370
X	0,000001	0,000002	0,000001	0,000002

Продолжение табл. 4

1	2	3	4	5
age (возраст)	–0,008092	–0,002230	–0,003192	0,038170
agesq (возраст квадратичный)	0,003182	–0,002523	–0,002315	–0,049030
incsec (неполное среднее образование)	0,400182	–0,114670	–0,119292	4,260000
second (среднее)	0,482086	–0,050829	–0,036295	4,347000
vocat (среднее профессиональное)	0,464828	–0,052840	–0,047721	4,327000
uni (университет)	0,621743	0,116083	0,117540	4,483000
goodsah (хорошее здоровье – опрос)	0,079197	0,056205	0,077297	0,085470
badsah (плохое – опрос)	–0,031894	–0,036597	–0,027795	–0,022880
mard (в браке)	–0,002130	0,001751	–0,002589	0,040990
div (разведен–а)	–0,009590	–0,001514	–0,006255	0,027170
wid (вдова, вдовец)	0,026145	–0,001134	0,017421	0,069210
town (город)	–0,113781	–0,103220	–0,129602	–0,118100
pgt (поселок городского типа)	0,001455	–0,035698	–0,018973	0,006427
rural (село)	–0,171396	–0,158237	–0,187111	–0,162600
exper (опыт работы – человек работал)	0,002069	0,000594	0,001555	0,000220
expnewsq (опыт в квадрате)	–0,002017	0,001441	–0,000266	0,003310
manag (профессия менеджер)	0,538863	0,218416	0,425068	0,566700
prof (профессия – топ-менеджеры)	0,437682	0,128598	0,321498	0,467500
tech (технические профессии)	0,279876	0,021066	0,173139	0,308100
clerk (клерк)	0,114822	–0,075731	0,023801	0,141300
servshop (работники услуги и продаж)	0,018573	–0,186516	–0,083102	0,052900
agrfish (сельское хозяйство, рыбное, лесное)	0,109244	–0,100262	–0,003620	0,107900
crafttrade (ручной труд)	0,346603	0,073581	0,240468	0,381000

Продолжение табл. 4

1	2	3	4	5
machop (машинный труд)	0,376087	0,092484	0,270653	0,407900
kids18 (дети есть или нет)	0,037867	0,047811	0,042637	0,045090
priv (частный сектор предприятия)	0,160477	0,113947	0,155652	0,167600
north (регион Север)	−0,368561	−0,026396	−0,260323	−0,346700
centr (Центральный регион)	−0,472235	−0,103807	−0,358663	−0,445500
volga (Волжский регион)	−0,575591	−0,179398	−0,459713	−0,551600
cauca (Кавказ)	−0,462715	−0,105919	−0,343704	−0,441500
ural (Урал)	−0,453249	−0,074556	−0,336085	−0,421700
west (запад)	−0,485403	−0,095749	−0,367122	−0,463500
east (восток)	−0,262364	0,047740	−0,152154	−0,244200
medmf (предприятие среднее – где работает 100–1000 человек)	0,090916	0,066970	0,087675	0,095080
largef (>1000)	0,299420	0,210668	0,286527	0,302200
oper (операции)	0,082358	0,071438	0,082103	0,084970
diab (диабет)	0,078339	0,055937	0,078511	0,091270
infar (был ли инфаркт)	0,203966	0,182730	0,216437	0,228800
stroke (сердечный приступ)	−0,073905	−0,042367	−0,070976	−0,059610
anem (анемия)	−0,053027	−0,033054	−0,053055	−0,048520
disab (инвалидность)	−0,064509	−0,059084	−0,061962	−0,062130
healthp (проблемы здоровья)	0,078379	0,060586	0,079294	0,076150
checkup (программа чекап)	0,006703	0,003778	0,008843	0,008016
dms (наличие ДМС)	0,216259	0,197432	0,220691	0,216000
heart (сердце)	−0,016793	−0,009402	−0,008609	−0,025500
lungs (легкие)	−0,031719	0,009576	−0,013602	−0,027140
liver (печень)	0,031412	0,015340	0,032072	0,020590
kidney (почки)	0,037943	0,013813	0,036876	0,029240

Продолжение табл. 4

1	2	3	4	5
gastro (гастроэнтерологические заболевания)	–0,044822	–0,026109	–0,034841	–0,047380
spine (позвоночник)	0,017324	0,013002	0,022926	0,002187
diabSR (самовыявленный диабет)	–0,053161	–0,014028	–0,040201	–0,060990
hyper (гипертония)	–0,001632	–0,013336	–0,001217	–0,010340
joint (болезнь суставов)	0,023547	0,009738	0,023464	0,031360
respir (респираторные)	–0,069624	–0,041668	–0,061758	–0,074780
neuro (нервные заболевания)	–0,083064	–0,055864	–0,074616	–0,086030
eye (проблемы со зрением)	0,051156	0,053840	0,061106	0,045220
aller (аллергия)	0,057079	0,074641	0,072826	0,049910
vein (вены)	–0,023506	–0,007846	–0,014311	–0,038830
skin (кожные заболевания)	–0,091518	–0,054346	–0,082535	–0,101700
onco (онкология)	–0,113456	–0,071808	–0,093978	–0,123200
uro (урология)	–0,014213	–0,008516	–0,009152	–0,022320
other (другие)	–0,012757	0,011225	0,005473	–0,026740
one_con (1 хроническое)	0,001659	–0,006924	–0,006546	0,014290
two_con (2 хронических)	0,061428	0,018278	0,042214	0,077510
three_con (3 хронических)	0,080058	0,025197	0,053623	0,096490
four_con (4 хронических)	0,050153	0,009562	0,013677	0,071530
five_con (5 хронических)	0,030239	–0,028216	–0,019961	0,070440
satlife (удовлетворенность жизнью по опросу)	0,040971	0,060042	0,047844	0,041300
nsatcond (не удовлетворен условиями жизни)	–0,240131	–0,187575	–0,238188	–0,230500

Продолжение табл. 4

1	2	3	4	5
desind100 (disease index – вес каждого заболевания) – от 0 до 100 для человека	NA	0,000079	0,000050	NA
goodsahl1 (хорошее здоровье в группе опроса по возрасту – 1 лаг назад для исключения взаимного влияния переменных)	–0,017223	0,001805	–0,012454	–0,008467
goodsahl2 (хорошее здоровье в группе опроса по возрасту – 2 лага назад)	0,013176	0,008965	0,011141	0,007859
goodsahl3 (хорошее здоровье в группе опроса по возрасту – 3 лага назад)	–0,024492	–0,018608	–0,027492	–0,012710
badsahl1 (плохое здоровье по опросу – 1 лаг назад)	–0,086666	–0,068602	–0,077938	–0,091020
badsahl2 (плохое здоровье по опросу – 2 лага назад)	–0,033370	–0,032981	–0,041940	–0,030190
badsahl3 (плохое здоровье по опросу – 3 лага назад)	–0,023101	–0,018904	–0,031452	–0,023600
desind100l1 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 1	0,004251	–0,000018	0,000064	0,005293
desind100l2 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 2	–0,004535	–0,000083	–0,002099	–0,004118
desind100l3 (disease index – вес каждого заболевания) – от 0 до 100 для человека – лаг 3	–0,007501	–0,000105	–0,002787	–0,007342

Продолжение табл. 4

1	2	3	4	5
one_conl1 (одно хроническое заболевание – 1 лаг в прошлом году для удаления причинности)	–0,034325	0,004724	–0,001406	–0,042160
one_conl2 (одно хроническое заболевание – 2 лага в прошлом году для удаления причинности)	0,059509	0,023710	0,040717	0,057440
one_conl3 (одно хроническое заболевание – 3 лага в прошлом году для удаления причинности)	0,065685	0,007245	0,028415	0,070570
two_conl1 (два хронических заболевания – 1 лаг)	–0,083124	–0,003890	–0,015590	–0,081140
two_conl2 (два хронических заболевания – 2 лага)	0,025783	–0,016473	–0,005625	0,017180
two_conl3 (два хронических заболевания – 3 лага)	0,110533	–0,004370	0,036892	0,110300
three_conl1 (три хронических заболевания – лаг 1)	–0,154050	–0,017350	–0,047681	–0,167000
three_conl2 (три хронических заболевания – лаг 2)	0,110219	0,011287	0,056640	0,101300
three_conl3 (три хронических заболевания – лаг 3)	0,187891	0,010785	0,075391	0,184500
four_conl1 (четыре хронических заболевания – 1 лаг) (четыре хронических заболевания – лагов 1)	–0,102367	0,028337	0,025081	–0,114000
four_conl2 (четыре хронических заболевания – лагов 2)	0,126483	–0,007999	0,050503	0,119000

Окончание табл. 4

1	2	3	4	5
four_conl3 (четыре хронических заболеваний – лагов 3)	0,277196	0,035997	0,127341	0,273200
five_conl1 (пять хронических заболеваний – 1 лаг)	–0,185368	0,001712	–0,009192	–0,207000
five_conl2 (пять хронических заболеваний – 2 лага)	0,219347	0,037751	0,123786	0,206200
five_conl3 (пять хронических заболеваний – 3 лага)	0,326735	0,018212	0,121261	0,320400

Список литературы

1. **Cule E., Vineis P., De Iorio M.** Significance testing in ridge regression for genetic data // BMC bioinformatics. 2011. Vol. 12, № 1. P. 1–15. DOI: 10.1186/1471-2105-12-372

2. **Zhang S. et al.** Kernel ridge regression for general noise model with its application // Neurocomputing. 2015. Vol. 149. P. 836–846.

3. **Cawley G. C. et al.** Heteroscedastic kernel ridge regression // Neurocomputing. 2004. Vol. 57. P. 105–124.

4. **Burnaev E., Nazarov I.** Conformalized kernel ridge regression // 2016 15th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2016. P. 45–52.

5. **Marquardt D. W., Snee R. D.** Ridge regression in practice // The American Statistician. 1975. Vol. 29, № 1. P. 3–20.

6. **Exterkate P.** Model selection in kernel ridge regression // Computational Statistics & Data Analysis. 2013. 68. Vol. 1–16.

7. **Garau M. et al.** Is the link between health and wealth considered in decision making? Results from a qualitative study // International journal of technology assessment in health care. 2015. Vol. 31, № 6. P. 449–456.

8. **Ghimir U.** The impact of health on wealth: empirical evidence // Available at SSRN 3754401. 2020.

9. **Shi X.** The health-wealth nexus for the elderly: Evidence from the booming housing market in China // Labour Economics. 2022. Vol. 78. P. 102247

10. **Bristow P.** Can Health Outcomes and Inequalities be Improved While Containing Costs? The Role of Voluntary Health Insurance in Universal Health Care Systems in the OECD. – 2023.

11. **Liu X. Q., Gao F.** Linearized ridge regression estimator in linear regression // Communications in Statistics-Theory and Methods. 2011. Vol. 40, № 12. P. 2182–2192.

12. **Канева М. А., Байдин В. М.** Неравенство в доходе и самооценка здоровья в России // ЭКО. 2019. № 12. С. 105–123. DOI: 10.30680/ECO0131-7652-2019-12-105-123

References

1. **Cule E., Vineis P., De Iorio M.** Significance testing in ridge regression for genetic data. *BMC bioinformatics*, 2011, vol. 12, № 1, pp. 1–15. DOI: 10.1186/1471-2105-12-372
2. **Zhang S. et al.** Kernel ridge regression for general noise model with its application. *Neurocomputing*, 2015, vol. 149, pp. 836–846.
3. **Cawley G. C. et al.** Heteroscedastic kernel ridge regression. *Neurocomputing*, 2004, vol. 57, pp. 105–124.
4. **Burnaev E., Nazarov I.** Conformalized kernel ridge regression. *15th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 2016, pp. 45–52.
5. **Marquardt D. W., Snee R. D.** Ridge regression in practice. *The American Statistician*, 1975, vol. 29, № 1, pp. 3–20.
6. **Exterkate P.** Model selection in kernel ridge regression. *Computational Statistics & Data Analysis*, 2013. vol. 68, pp. 1–16.
7. **Garau M. et al.** Is the link between health and wealth considered in decision making? *Results from a qualitative study. International journal of technology assessment in health care*, 2015, vol. 31, № 6, pp. 449–456.
8. **Ghimir U.** The impact of health on wealth: empirical evidence. Available at SSRN 3754401. 2020.
9. **Shi X.** The health-wealth nexus for the elderly: Evidence from the booming housing market in China. *Labour Economics*, 2022, vol. 78, p. 102247
10. **Bristow P.** Can Health Outcomes and Inequalities be Improved While Containing Costs? The Role of Voluntary Health Insurance in Universal Health Care Systems in the OECD. 2023.
11. **Liu X. Q., Gao F.** Linearized ridge regression estimator in linear regression. *Communications in Statistics-Theory and Methods*, 2011, vol. 40, № 12, pp. 2182–2192.
12. **Kaneva M. A., Baidin V. B.** Income Inequality and Self-Assessed Health in Russia. *ECO*, 2019, no. 12, pp. 105–123. DOI: 10.30680/ECO0131-7652-2019-12-105-123 (in Russ.)

Сведения об авторе

Заболотский Алексей Александрович, кандидат экономических наук, научный сотрудник отдела регионального и муниципального управления Института экономики и организации промышленного производства СО РАН

Information about the Author

Alexey A. Zabolotsky, Candidate of Economic Sciences, Researcher, Department of Regional and Municipal Management, Institute of Economics and Industrial Engineering of the Siberian Branch of the Russian Academy of Sciences

*Статья поступила в редакцию 09.04.2024;
одобрена после рецензирования 25.05.2024; принята к публикации 25.05.2024*

*The article was submitted 09.04.2024;
approved after reviewing 25.05.2024; accepted for publication 25.05.2024*