# PRACTICAL TREATMENT OF THE MULTICOLLINEARITY: THE OPTIMAL RIDGE METHOD AND THE MODIFIED OLS

©2021 **TYZHNENKO A. G., RYEZNIK Y. V.**

**Tyzhnenko A. G., Ryeznik Y. V.**

**Practical Treatment of the Multicollinearity: The Optimal Ridge Method and the Modified OLS**

*The paper discusses the applicability of the two main methods for solving the linear regression (LR) problem in the presence of multicollinearity – the OLS and the ridge methods. We compare the solutions obtained by these methods with the solution calculated by the Modified OLS (MOLS) [1, 2]. Like the ridge, the MOLS provides a stable solution for any level of data collinearity. We compare three approaches by using the Monte Carlo simulations, and the data used is generated by the Artificial Data Generator (ADG) [1, 2]. The ADG produces linear and nonlinear data samples of arbitrary size, which allows the investigation of the OLS equation's regularization problem. Two possible regularization versions are the COV version considered in [1, 2] and the ST version commonly used in the literature and practice. The performed investigations reveal that the ridge method in the COV version has an approximately constant optimal regularizer $(\lambda_{ridge}^{(opt)} \approx 0.1)$ for any sample size and collinearity level. The MOLS method in this version also has an approximately constant optimal regularizer, but its value is significantly smaller $(\lambda_{MOLS}^{(opt)} \approx 0.001)$. On the contrary, the ridge method in the ST version has the optimal regularizer, which is not a constant but depends on the sample size. In this case, its value needs to be set to $\lambda_{ridge}^{(opt)} \approx 0.1(n-1)$. With such a value of the ridge parameter, the obtained solution is strictly the same as one obtained with the COV version but with the optimal regularizer $\lambda_{ridge}^{(opt)} \approx 0.1$ [1, 2]. With such a choice of the regularizer, one can use any implementation of the ridge method in all known statistical software by setting the regularization parameter $\lambda_{ridge}^{(opt)} \approx 0.1(n-1)$ without extra tuning process regardless of the sample size and the collinearity level. Also, it is shown that such an optimal ridge(0.1) solution is close to the population solution for a sample size large enough, but, at the same time, it has some limitations. It is well known that the ridge(0.1) solution is biased. However, as it has been shown in the paper, the bias is economically insignificant. The more critical drawback, which is revealed, is the smoothing of the population solution – the ridge method significantly reduces the difference between the population regression coefficients. The ridge(0.1) method can result in a solution, which is economically correct, i.e., the regression coefficients have correct signs, but this solution might be inadequate to a certain extent. The more significant the difference between the regression coefficients in the population, the more inadequate is the ridge(0.1) method. As for the MOLS, it does not possess this disadvantage. Since its regularization constant is much smaller than the corresponding ridge regularizer (0.001 versus 0.1), the MOLS method suffers little from both the bias and smoothing of its solutions. From a practical point of view, both the ridge(0.1) and the MOLS methods result in close stable solutions to the LR problem for any sample size and collinearity level. With the sample size increasing, both solutions approach the population solution. We also demonstrate that for a small sample size of less than 40, the ridge(0.1) method is preferable, as it is more stable. When the sample size is medium or large, it is preferable to use the MOLS as it is more accurate yet has approximately the same stability.*

*Tyzhnenko Alexander G.* – Candidate of Sciences (Physics and Mathematics), Associate Professor, Associate Professor of the Department of Mathematics and Mathematcal Methods in Economics, Simon Kuznets Kharkiv National University of Economics (9a Nauky Ave., Khark³v, 61166, Ukraine)

*E-mail:* olersandr.tyzhnenko@m.hneu.edu.ua

*ORCID:* http://orcid.org/0000-0001-8508-7341

*Ryeznik Yevgen V.* – Candidate of Sciences (Physics and Mathematics), Lecturer, Department of Mathematics, Uppsala University (buildings 1, 6 and 7, 1, Lägerhyddsvägen, Uppsala, 75106, Sweden)

*E-mail:* yevgen.ryeznik@math.uu.se

*ORCID:* https://orcid.org/0000-0003-2997-8566

***Тижненко О. Г., Рєзнік Є. В. Практичне рішення проблеми мультиколінеарності: Оптимальний метод рідж-регресії та модифікований метод найменших квадратів***

*У цій статті розглядається придатність двох основних методів для вирішення проблеми лінійної регресії (LR) за наявності мультиколінеарності, а саме OLS, та ridge-методу порівняно з рішеннями модифікованого методу OLS (MOLS) [1; 2], який, як і ridge, забезпечує стабільне рішення на будь-якому рівні колінеарності даних. Порівняння проведено методом Монте-Карло із використанням штучного генератора даних (ADG) [1; 2], який генерує лінійні вибірки даних будь-якого розміру. Використання ADG дозволяє нам дослідити проблему регуляризації рівняння OLS. Було виявлено, що можливі дві версії регуляризації: версія COV, яка була запропонована та досліджена в [1; 2], та версія ST, яка зазвичай використовується в літературі та практичних реалізаціях. Запропоновані дослідження показують, що у версії COV ridge метод має приблизно постійний оптимальний регулятор $(\lambda_{ridge}^{(opt)} \approx 0.1)$ для будь-якого обсягу вибірки та рівня колінеарності. Метод MOLS також має у цій версії приблизно постійний оптимальний регуля-*

тор, але він значно менший за значенням $(\lambda_{MOLS}^{(opt)} \approx 0.001)$. У той же час у загальновживаній версії ridge-методу нам потрібен оптимальний регулятор $\lambda_{ridge}^{(opt)} \approx 0.1(n-1)$, який залежить від обсягу вибірки n і не є константою. Нам було показано в роботі, що версія ST, яка використовується як правило на практиці разом із ridge-методом, при використанні оптимального параметра $\lambda_{ridge}^{(opt)} \approx 0.1(n-1)$, дає строго те саме рішення, що і COV версія хребта з оптимальним регулятором $\lambda_{ridge}^{(opt)} \approx 0.1$ [1, 2]. Це дозволяє використовувати коди ridge-методу у всім відомому статистичному програмному забезпеченні, встановлюючи параметр регуляризації $\lambda_{ridge}^{(opt)} \approx 0.1(n-1)$ без будь-якого процесу налаштування, незалежно від обсягу вибірки та рівня колінеарності. Ми також показуємо, що таке оптимальне рішення ridge(0,1) наближається до рішення в популяції для досить великого обсягу вибірки, але одночасно має деякі проблеми. Той факт, що метод ridge(0,1) дає зміщення, відомий, але це зміщення, як було показано в роботі, є економічно незначущим. Найважливішим виявленим недоліком є згладжування популяційного рішення: ridge-метод значно зменшує різницю між коефіцієнтами регресії популяції. Отже, ridge(0,1) може дати економічно правильний (з правильними ознаками), але певною мірою неадекватний розв'язок. Неадекватність ridge(0,1) виявляється тим більше, чим більша різниця між коефіцієнтами регресії в популяції. Цим недоліком MOLS практично не володіє, оскільки для нього константа регуляризації має набагато менше значення (0,001 проти 0,1). Через це метод MOLS практично мало страждає як від зміщення, так і від згладжування своїх рішень. З практичної точки зору, обидва методи, ridge(0,1) та MOLS, дають тісні стабільні рішення проблеми LR для будь-якого обсягу вибірки та рівня колінеарності, які наближаються до рішень в популяції зі збільшенням обсягу вибірки. У статті також показано, що для малих вибірок менше 40 переважно використовувати ridge(0,1), оскільки він є більш стабільним. Для середніх та великих зразків переважно використовувати MOLS, оскільки він є більш точним із приблизно однаковою стабільністю..

*Ключові слова:* мультиколінеарність, економічна коректність, економічна адекватність, модифіковане правило Крамера, модифіковане OLS, оптимальна ridge-регресія.

*Рис.:* 3. *Табл.:* 5. *Формул:* 14. *Бібл.:* 27.

*Тижненко Олександр Григорович* – кандидат фізико-математичних наук, доцент, доцент, кафедра вищої математики та економіко-математичних методів, Харківський національний економічний університет ім. С. Кузнеця (просп. Науки, 9а, Харків, 61166, Україна)

*E-mail:* olersandr.tyzhnenko@m.hneu.edu.ua

*ORCID:* http://orcid.org/0000-0001-8508-7341

*Рєзнік Євген Володимирович* – кандидат фізико-математичних наук, викладач, кафедра математики, Уппсальський університет (Легерхюддсвеген 1, кор. 1, 6 і 7, Уппсала, 75106, Швеція)

*E-mail:* yevgen.ryeznik@math.uu.se

*ORCID:* https://orcid.org/0000-0003-2997-8566

**Preamble.** It is worth recalling that in practice, researchers continue using only the OLS to solve the LR problems regardless of the VIF factor's value. They only check the OLS solutions obtained just for economic correctness (only the correctness of the regression coefficients signs). If the OLS solution is not correct, it is simply not taken into further consideration. Because of that, a lot of valuable information is lost.

As the authors discovered in the previous works [1; 2], the actual situation might be much worse. The point is that economically correct solutions can be inadequate to the solutions in the population. That is, an approximate OLS solution may be economically correct but quite far from the true solution.

Thus, to date, only the OLS solution with correct signs is the valid approximation of the true solution. Due to this situation, the following two problems arise immediately: 1) discarding the data, which one could use correctly; 2) considering the obtained economically correct OLS solution as the true solution, although it might be far enough from the solution in the population.

This situation has lasted for many decades, even though in 1959-62, A. E. Hoerl proposed a powerful method to approximate the LR problem solution under near-collinearity, namely the *ridge* method. However, the practical application of this method turned out to be limited. It happened due to the significant dependence of the solution on the value of the regularization constant. The search for the optimal regularization constant continues to this day yet without noticeable success. The investigations carried out in this work have shown

that under the commonly used formulation of the regularization problem for the OLS equation, the optimal value of the regularization parameter is not a constant but is proportional to the sample size.

Also, it has been found that the formulation of the OLS equation regularization method is of fundamental importance. Therefore, the paper considers two formulations of the OLS equation regularization: the COV version (the covariance matrix is used) and the ST version (the standard formulation common in practice and theory up to date).

The paper demonstrates that an approximately constant value of the optimal regularization parameter is observed only in the COV version of the ridge method and has a value of about 0.1. The approximate optimal regularization parameter value of the OLS equation in the ST version equals 0,1 ($n$ –1), where $n$ is the sample size. Herewith, a problem in the COV version with the regularization parameter of 0.1 and a corresponding problem in the ST version with the regularizer of 0,1 ($n$ –1) have coinciding solutions for any sample size and a collinearity level.

**Introduction**. From a mathematical perspective, an LR problem is equivalent to a curve-fitting problem [3–8]. The OLS solves this problem impeccably.

The OLS approach is usually not suitable for obtaining an economically adequate solution to the LR problem [9–14]. The OLS can give an adequate solution to the economic regression problem if the regressors are near-orthogonal. Unfortunately, this is not the case in practice.

The main drawback that prevents the OLS solution from being adequate to an economic problem is the near-collinearity of regressors [9–20]. With the VIF factor increasing, the variability of the OLS solution drastically increases. This issue prevents from getting an adequate economic solution to the regression problem in practice.

There exist different remedies dealing with ill-conditioning and near-collinearity, including various regularization methods, ridge regression [12], omitting variables, grouping variables in blocks, collecting additional data, and so on; see [5; 10–22] among others. However, these remedies may be time-consuming, costly, impossible to achieve, or controversial [23]. Also, the diagnostic tools that signal the presence of near-collinearity are crucial. However, the authors agree with [24] that no signal of multicollinearity exists at all because «multicollinearity is a matter of degree rather than one of kind».

Many years of efforts did not yield any results in the search for a critical level of near-collinearity. It seems that A. C. Harvey in [24] was right that there is no such a critical level at all and the influence of near-collinearity at an OLS solution is a continuous process that depends on many parameters. This issue is also confirmed in [1; 2] and this paper.

Despite the theoretical warnings about the inadmissibility of using the OLS in the presence of near-collinearity of any level, this technique is still in use in practice, in economic and other studies, with attempts to reduce somehow the level of collinearity. As for the appearance of incorrect signs in the OLS solutions, that is, when solutions have no economic (in general, physical) sense, this phenomenon, as shown in [1], is connected with the fundamental property of nonsingular square matrices.

The possible economic inadequacy of OLS solutions, given that the solution has the correct signs (i.e., is economically correct), has been investigated in [2] and in the current paper. It has been shown that large enough random errors in solutions may cause this effect. Numerical studies of the variability in the OLS, MOLS, and ridge solutions allow us to state that economic inadequacy occurs when the coefficient of variation (CV) of a solution is larger than 10% [2].

In this paper, we investigate the problem of the appearance of the incorrect and inadequate OLS solutions to the LR problem in the presence of multicollinearity and compare the OLS with the wide known *ridge* method and the new, providing a stable solution, Modified OLS (MOLS) method proposed in [1; 2].

The MOLS, as well as the *ridge* method, is an approximate method that uses the well-known Tikhonov's regularization principle and a new method of solving the regularized OLS equation called in [1; 2] the *Modified Cramer Rule.*

The Artificial Data Generator (ADG) proposed in [1; 2] allows generating data with different levels of regressors' collinearity. Such data modeling helps to compare the three methods mentioned above to determine their applicability in real economic studies.

It worth mentioning that all considerations regarding the applicability of the OLS, ridge, and MOLS methods for optimal values of the regularization constant cannot be rigorously theoretically proven. Therefore, to demonstrate the ideas, we use the Monte Carlo simulations for linear and nonlinear data generated by the ADG. For this purpose, we use ADG to generate both linear data with a priori equal or different regression coefficients in the population and nonlinear data with and without autocorrelation.

An essential feature of the ADG is the opportunity to estimate the population regression coefficients due to the OLS consistency property. For some values of the ADG parameters, we can estimate the population solution with any accuracy using a sample size large enough, making it possible to compare all the methods in terms of both the bias and accuracy of their solutions for any sample size.

*Regularization of the OLS equation.*

Given the data generated by the ADG, we investigated the regularization problem of the ill-conditioned standardized OLS equation for *beta*-coefficients:

$$Y = \sum_{j=1}^{m} \beta_j X_j + \varepsilon,$$

where $\{Y, X\}$ is the standardized data; $\{y, x\}$ – the observed data.

There are two methods to regularize the OLS equation: first, by the standard method (ST version) in the generally accepted form

$$(X'X + \lambda I)\beta = X'Y, \qquad (1)$$

second, by using the covariance matrix (*COV* version [1; 2]) in the form

$$\left(\frac{1}{n-1}X'X + \lambda I\right)\beta = \frac{1}{n-1}X'Y. \qquad (2)$$

Here, we demonstrate that both the COV version of the *ridge* method and the COV version of the MOLS have approximately constant optimal regularizer: $\lambda_{ridge}^{(opt)} \approx 0.1$ for the *ridge* method, and $\lambda_{MOLS}^{(opt)} \approx 0{,}001$ for the MOLS, which implies that in practice a researcher can use only one regularizer $(\lambda_{ridge}^{(opt)} \approx 0.1)$ in all cases without tuning for the *ridge* method and only one regularizer $(\lambda_{MOLS}^{(opt)} \approx 0.001)$ for the MOLS, and these regularizers are approximately optimal and provide reasonable accuracy.

As for the commonly used ST version, the optimal regularizer in this case is

$$\lambda_{ST}^{(opt)} = (n-1)\lambda_{COV}^{(opt)}.$$

For any computational environment one can use the program implementation of the *ridge* method by setting an optimal regularization constant to $\lambda_{ridge}^{(opt)} = 0.1(n-1)$ for any data sample and can obtain the estimated population solution with a good enough accuracy.

There is also considered here more one serious drawback of the regularization at all, excluding the well-known bias. The investigations have revealed that any regularization procedure tends to smooth the LR solution. The larger the regularization constant, the less the difference between the regression coefficients unequal in the population (and the larger is the bias).

If all regression coefficients are equal in the population, the *ridge*(0.1) method (with optimal regularizer), for example, also gives all the regression coefficients equal but slightly less in absolute value. If the regression coefficients are different in population, then, in addition to the general bias, there the smoothing between the regression coefficients arises. In this case, the larger coefficients decrease in absolute value, and the smaller ones increase, but the sum of the beta coefficients practically does not change.

Change in the relationship between the regression coefficients leads to incorrect interpretation of regressors' influence on the response, that is, to the inadequacy of the economic regression problem's solution. This effect is an inherent part of the regularization process, and the larger the regularization constant, the more significant it is. Then, for the *ridge* method with the optimal regularization constant $(\lambda_{ridge}^{(opt)} \approx 0.1)$, this effect is much more significant than for the MOLS with the optimal regularization constant $\lambda_{MOLS}^{(opt)} \approx 0.001$ that we will demonstrate further.

First, we generate data with a low level of collinearity (VIF~ 20), for which an OLS solution converges fast in probability to a population solution with the sample size increasing. For such data, we can take, for instance, the OLS solution for a sample of size $10^7$ as the population one.

Having in hand the population solution, we investigate the bias and standard deviation of the OLS, MOLS, and *ridge*(0.1) solutions for any sample size (*n*) and collinearity level (VIF). By varying these parameters, we can investigate both the stability of the methods and the economic adequacy of the solutions they provide. Note that we will conduct these studies using the standardized (normalized) data. It is connected with the interesting behavior of the standardized regressor coefficients ($\beta_j$). For example, the sum of beta-coefficients moduli is close to one for all cases observed. Besides, for regressors with the same residual errors in partial regressions with a response, the beta coefficients are the same, regardless of the law of their linear influence on the response. If for every simple regression with a response, the residual errors are approximately similar, and the beta coefficients are different, this implies some autocorrelation in data. As is shown later, the existence of autocorrelation in data greatly impairs the adequacy of the OLS solutions for finite samples.

*Linear and nonlinear data simulation with the ADG.*

The main principles of simulating the linear data by the ADG are outlined in [1; 2]. However, the very definition of the data linearity concept has not been given. By itself, data cannot be linear or nonlinear. However, one may define data linearity when the data is used in a simple LR problem. If the LR problem is being solved, given data $\{y,x\}$:

$$y = b_0 + b_1 x + e,$$

one can calculate the *b*-coefficients estimates $(\hat{b}_0, \hat{b}_1)$ and the linear trend function $\hat{y} = \hat{b}_0 + \hat{b}_1 x$. Then *one can define the observed data as linear if the residual* $\varepsilon = y - \hat{y}$ *follows the normal distribution with a zero mean,* $N(0, \sigma^2)$. *For the multivariate regression problem, each regressor can be checked for linearity either theoretically or graphically to make appropriate economic decisions.*

As we will show below, the point is that the presence of regressors that are not close to linear leads to a significant decreasing in the corresponding regression coefficients in modulus. The researcher has to decide whether to keep or discard a regressor that contributes a little to the response, even though its presence remarkably changes the LR problem solution. Further on, we will touch on some variants of a linear data generation used for investigating the applicability of the OLS, *ridge*, and MOLS in practical research.

In [1; 2], we only considered strictly linear regressors with the same statistical parameters, which should give the same regression coefficients and hence the same contribution in the response. The latter follows from the very principle of data generation. We proceed in the ADG from the specification of the response in the form of a pseudo-random vector: $y = randn(n,1)$ in the MATLAB notation. The regressors ($x_j$) we generate with the aid of the simple regression model as

$$x_j = k_j^{-1}(y + d_j \alpha \varepsilon), \qquad (3)$$

where $k_j$ sets the law of nonstochastic linear influence of a regressor on the response: $y = k_j x_j$; $\varepsilon$ sets the stochastic impact of unaccounted factors, which are set as $\varepsilon = randn(n,1)$ with the amplitude $d_j \alpha$. If we write (3) as

$$y = k_j x_j + d_j \alpha \cdot randn(n,1), \qquad (4)$$

we can see that $d_j \alpha \cdot randn(n,1)$ sets a residual error in the simple regression of *y* on $x_j$. The multiplier $d_j \alpha$ in (3) sets the random component in $x_j$ incoherent to *y*. The first component in (3) is coherent to *y* and sets the economic law of influence of the regressor $x_j$ on the response.

Note that with a decrease in α, the incoherent component also decreases in amplitude, leading to an increase in the level of *near*-collinearity of the regressors (VIF) and the equality of the *beta*-coefficients in the population, even though $d_j$-parameters may be different.

In the code, we set: $k_j^{-1} = \tan\left(\theta_j \cdot \dfrac{\pi}{180} 180\right)$, where $\theta_j$ is the angle in degrees between the trend and *OY* axes in the simple regression $x_j$ on *y*. So, for $\theta_j = 1°$, for example, we have a significant influence of $x_j$ on *y* in the regression *y* on $x_j$; for $\theta_j = 89°$, for instance, we have a negligible effect of $x_j$ on *y* in the regression *y* on $x_j$. If we take $\theta_j > 90°$, the sign of the regression coefficient changes, and so on.

If α tends to zero, the regressors $x_j$ become closer to each other, which increases their collinearity, and so their VIF factor. The multiplier $d_j$ allows us to adjust the level of influence of unaccounted factors on the response for each regressor for not very small α (for very small α all $d_j \alpha$ are approximately zero, and a change in $d_j$ does not influence the solution, if $d_j$ is not too large).

So, in the ADG model (3), the multicollinear regulator α and the multiplier $d_j$ are connected such that the difference between the regressors diminishes as α tends to zero. This means

that with an increase in the level of a *near*-collinearity (VIF) in the ADG, the regression coefficients $b_j$ in the linear model become the same in the population. In this work, we use this fact to check the adequacy of the method for solving the LR problem in the presence of a *near*-collinearity.

Particularly, we further use this to investigate the approximation of the population solution under the presence of a *near*-collinearity by the OLS solution with an increase in a sample size. This question is of fundamental importance for studying the methods of solving the LR problem, given the near-collinear data. Thus, for the data generated by ADG, if the OLS solution to the LR problem stabilizes with sample size increasing and does not change, say for $n > 10^4$, then one can consider the OLS solution obtained for $n > 10^6$ or $n > 10^7$ as an approximate solution in the population. This is exactly how we estimate the population solution in this work.

The approach mentioned above allows us to evaluate both the bias and adequacy of the LR problem solution obtained for a finite sample size. Multiple repetitions of solutions for data extracted from the same population (ADG) allows us to determine the coefficients of variation (CV) of solutions for different methods depending on the sample size, which helps the researcher to decide on the applicability of the results obtained by the chosen method in practice. Such a decision will be correct only for strictly linear data. Therefore, before practically solving the LR problem, it is necessary to check the data linearity, at least graphically. The influence of the data nonlinearity is very significant, and we will discuss it below.

The choice of parameters in (3) determines the solution to the LR problem in the population. So, if $d_j$ and $k_j$ in (3) are the same, it is obvious that all regression coefficients will also be equal in the population. This case was investigated in [1; 2]. We used there the COV version (2) of the regularized OLS equation and compared solutions of the OLS, the *ridge*(0.1) method with regularization constant 0.1, and the MOLS with regularization constant 0.001. There, we also used the OLS solution for a large sample of size $n > 10^6$ as the population solution, which helped to determine the bias for each method at different values of the sample size and the coefficients of variation (CV) of the solutions.

The use of the OLS solution for a large sample as an estimate of the population solution is based on the well-known property of consistency of the OLS estimators, see, for example, [9]. When simulating regressors with the same statistical characteristics, the OLS solutions did tend to a certain limit, and for samples larger than $n \approx 10^3$ did not practically change.

To demonstrate the OLS, MOLS, and *ridge* solutions properties in other different cases, we give below their solutions simultaneously for the COV version of the regularized OLS equation (2). The MOLS algorithm for the COV version (2) was described in [1; 2]. Instead of using the *ridge(0.1)* in the COV version (2), one can use any program implementation of the *ridge* method (ST version) with the regularization parameter $\lambda = 0.1(n-1)$ for any sample size $n$ and the collinearity level.

*Some peculiarities of the OLS solutions.*

Everyone knows that the OLS solutions to the LR problems can be incorrect if the regressors are collinear to some degree. It is convenient to measure the degree of collinearity of the data by the VIF indicator. The question of at what value of the VIF the OLS solutions cannot be used remains unexplored to date, apart from the cases when the method gives incorrect solution signs.

The latter also raises the question: if the OLS solution is economically incorrect, can the experimental data still be used for scientific purposes, or should they simply be thrown away? The answer to the last question is usually the advice to increase the experimental dataset. In reality, this is not always possible, and even if it is possible, the question remains to what extent the dataset should be increased.

To clarify these issues, we consider OLS solutions' features and compare them with solutions of the *ridge* method and the MOLS. As the first step, we consider solutions to the LR problem with *four* regressors on strictly linear data generated using the ADG with equal parameters $d_j = 1$ and $\theta_j = 1°$.

In this case, all the population *beta*-coefficients will be $\approx 0.25$ ($\sim 1/m$, for $m = 4$). With the aid of the $10^4$ times the ADG data repeating, we estimate the stability and adequacy of all three methods' solutions for small and large samples of various VIFs.

For $n = 10^3$ and $\alpha = 0.2$ (VIF $\sim 20$), i.e. average sample sizes in the absence of *near*-collinearity (for strictly linear regressors), we obtain for one sample: the OLS *beta*-coefficients can be [0.2477 0.2587 0.2517 0.2516]; for the *ridge*(0.1) solution: [0.2448 0.2478 0.2459 0.2458]; for the MOLS($10^{-3}$) solution: [0.2497 0.2561 0.2519 0.2520]. As we can see, all solutions are close to the population one ($\approx 0.25$). Note that all the values of the beta coefficients obtained by the *ridge*(0.1) method are less than the corresponding coefficients in the MOLS because the regularization constant in the *ridge* method is larger than that in the MOLS. However, the *ridge*(0.1) bias is small enough.

More precisely, the bias of the *ridge*(0.1) method solutions can be calculated by resampling (Tables 1, 2). In the tables, the coefficients of variation (CV) are also shown for all three methods. (The population solution estimation with the aid of the OLS solution for $n = 10^6$ is $\beta = [0.2527; 0.2524; 0.2521; 0.2526]$; for $n = 10^7$ is $\beta = [0.2524; 0.2524; 0.2524; 0.2525]$).

Solutions with a coefficient of variation less than 10% can be considered adequate at the 5% significance level [1; 2]. Table 1 shows that for linear data, the OLS solutions can be considered adequate for sample sizes larger than $10^3$. For sample sizes less than 30, all methods give an inadequate solution. When $n = 30$, only the *ridge*(0.1) method gives adequate solutions. When data are *linear*, and the *regression coefficients are equal in the population*, the *ridge*(0.1) method is the best: it is most stable and gives a small enough bias. For more details, see papers [1; 2].

Modeling nonlinearity with the ADG and analyzing solutions to the LR problem with nonlinear data will be considered in this paper after discussing the optimality of the *ridge*(0.1) and MOLS solutions.

*Ridge and MOLS optimality.*

The main problem of solving the OLS equation by the regularization method is to determine the optimal value of the regularization constant, at which the regression coefficients are stable and sufficiently close to the corresponding values in the population.

**Means and CVs of beta-coefficients,** VIF ~ 20, $10^4$-repetitions

| n | $n = 10^3$ | | | | $n = 30$ | | | |
|---|---|---|---|---|---|---|---|---|
| **betas** | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_3$ |
| The OLS solutions in the absence of collinearity. | | | | | | | | |
| mean | 0.2526 | 0.2523 | 0.2524 | 0.2524 | 0.2516 | 0.2525 | 0.2535 | 0.2524 |
| CV(%) | 5.61 | 5.55 | 5.49 | 5.51 | 36.01 | 35.69 | 36.06 | 35.93 |
| The *ridge*(0.1) solutions in the absence of collinearity. | | | | | | | | |
| mean | 0.2461 | 0.2461 | 0.2461 | 0.2461 | 0.2459 | 0.2461 | 0.2465 | 0.2461 |
| CV(%) | 1.59 | 1.58 | 1.56 | 1.57 | 9.53 | 9.48 | 9.54 | 9.50 |
| The MOLS(0.001) solutions in the absence of collinearity. | | | | | | | | |
| mean | 0.2525 | 0.2523 | 0.2524 | 0.2524 | 0.2519 | 0.2524 | 0.2532 | 0.2523 |
| CV(%) | 3.33 | 3.29 | 3.26 | 3.27 | 18.89 | 18.75 | 18.85 | 18.79 |

Many theoretical works are devoted to this problem, see, for example, [20; 25; 26], from which it can be concluded that there is probably no optimal regularization constant for the standard regularization scheme (ST version) of the OLS equation.

This article investigates this problem using the linear data simulation with the aid of the ADG. Considering that we know with sufficient accuracy the solution in the population, we directly investigated the second norm of the difference between the LR-solution for a sample of size $n$ and the solution in the population.

For the ADG sampling with equal parameters $d_j = 1$ and $\theta_j = 1°$ we solve the COV version of the regularized OLS equation (2) by the *ridge*(0.1) method with various values of the regularization constant λ for different sample sizes ($n$) and different VIFs and calculate the Euclidean norm of the deviation of the sample solution from that of the population (*deviation norm* in Figures (1, 2) below).

Figures (1, 2) demonstrate the general situation: the optimal value of the regularization parameter exists but depends on both the sample size and VIF; in principle, for each sample, it is possible to find the exact value of the optimal regularization parameter, but from the figures and more detailed studies it can be concluded that *the deviation of the exact optimal parameter from the value 0.1 insignificantly changes a solution to the LR problem for any sample.*

A similar situation also occurs for the MOLS: the exact value of the optimal regularizer changes for each sample in the vicinity of 0.001, but the solution remains practically un-
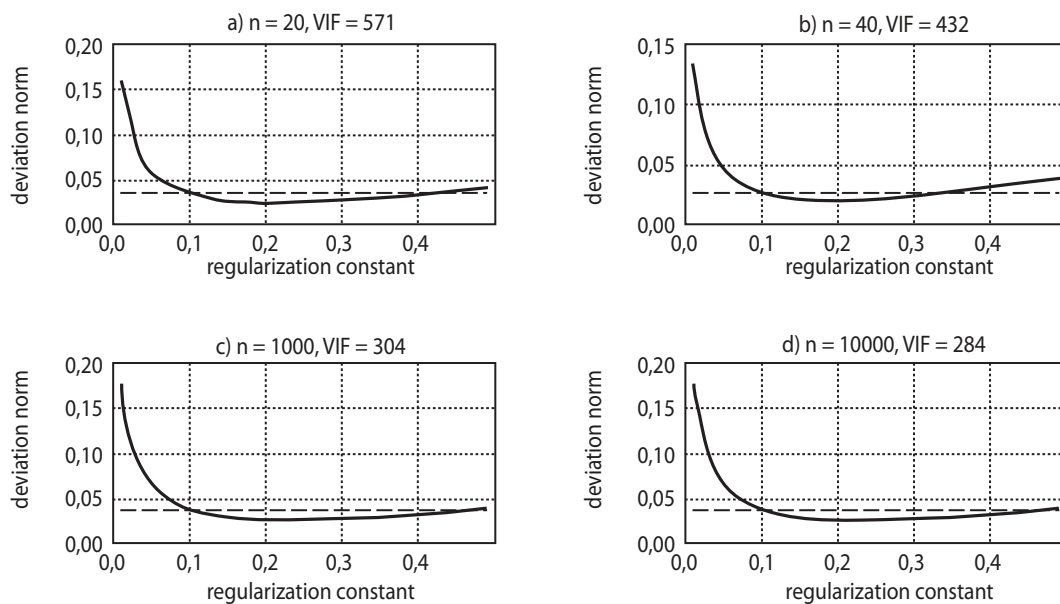


**Figure 1. Deviation of *ridge*(0.1) solutions from the population one
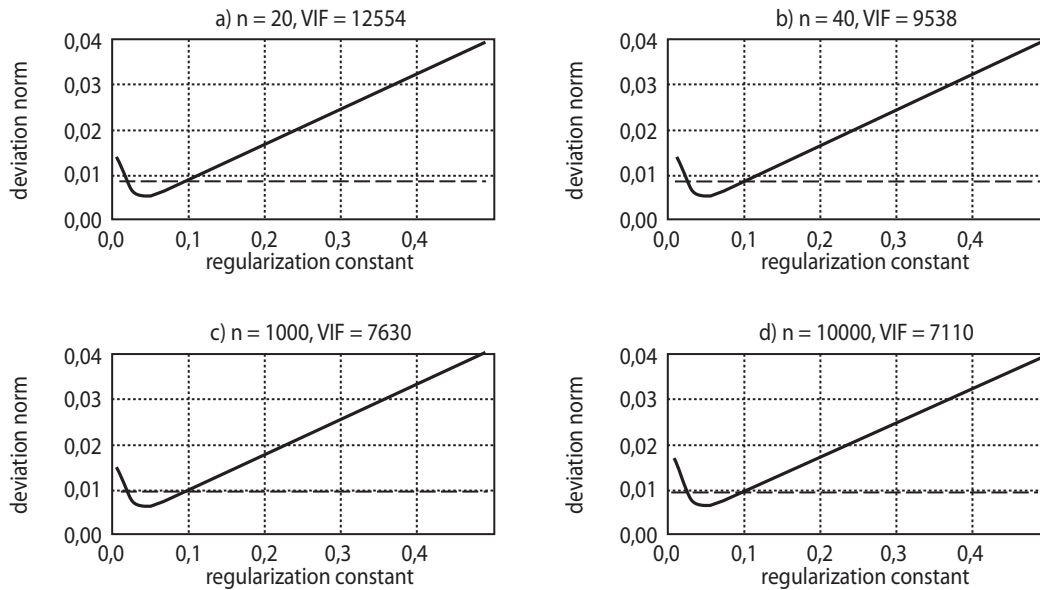for not very strong collinearities** (α = 0,05) **depending on** λ

**Figure 2. Deviation of** *ridge*(0.1) **solutions from the population one for strong enough collinearities** ($\alpha$ = 0,01) **depending on** $\lambda$

changed. Detailed studies of this problem have shown that in practice, it is quite possible to use in all cases the *ridge* method in the COV version with the regularizer equals 0.1 (*ridge*(0.1)), and the MOLS method with the regularizer equals 0.001.

*Linear regressors with a different variance of the noise.*

It was shown above that the *ridge*(0.1) method gives excellent solutions for linear regressors with the same parameters $\theta_j$ and $d_j$. The stability of the method is higher than for MOLS due to a larger regularization constant (0.1 vs. 0.001), and the bias is not significant for economic research. Let us now consider a situation in which the regressors are also strictly linear, but with different laws of influence on the response (different $\theta_j$) and the same amplitude of the normal noise in regressors (the same $d_j$).

To demonstrate the quality of *ridge*(0.1) solutions, we consider, as above, a small sample ($n$ = 30) and show the coefficient of variation (CV) for all three methods. Let us set for this: $\theta$ = [89°; 1°; 89°; 1°], $d$ = [1; 1; 1; 1], $\alpha$ = 0.2 in (3). Note that for $\alpha$ = 0.2 (VIF ~20), the collinearity is practically absent, and the OLS gives a correct and adequate solution to the LR problem, with which we compare the MOLS and *ridge*(0.1) solutions.

With such parameters, the *b*-coefficients in the population will be: [small; large; small; large] according to the $\theta$-values; the *beta*-coefficients will be the same ($\approx$0.25) as we affirmed above.

Since we do not know the regression coefficients in the population, we find their estimate by the OLS solution with $n$ = 10^6 and 10^7. For *b*s we have for $n$ =10^6:
$b$ = [0.1491($b_0$); 0.0481; 14.1879; 0.0482; 14.1721];
for *beta*s we have: $\beta$ = [0.2522; 0.2525; 0.2528; 0.2523].

For $n$ = 10^7, we have for *b*s:
$b$ = [0.1493($b_0$); 0.0043; 14.1981; 0.0043; 14.1792];
for *beta*s we have: $\beta$ = [0.2524; 0.2526; 0.2523; 0.2524]. It is clearly that the *beta*-solutions in the population are the same and close to 0.25.

In practice, the researcher is interested in the statistical properties of an LR problem solution on a one-sample of a specific size. To date, the researcher cannot obtain such information since this requires a large set of samples of a given size from the population, and it is impossible to obtain such one.

In this work, due to the program for generating samples with a given property of solutions in the population (ADG), we can draw a sample from the population an unlimited number of times. This allows us to evaluate the statistical characteristics of a solution to the LR problem by the chosen method (OLS, MOLS, or *ridge* (0.1)) for data of a certain type, for example, for strictly linear data, for near-linear data, or, in general, for nonlinear data.

First, consider the statistical characteristics of solutions to the LR problem for sufficiently small strictly linear samples by various methods (OLS, MOLS, and *ridge* (0.1)). For $n$ = 30, $\alpha$ = 0.2, the same *d*-parameters (all ones), and different $\theta$-parameters ($\theta$ = [89°; 1°; 89°; 1°]), we have the following solutions' means and their CVs using 10^4 repetitions. Mean *beta*-coefficients:

$$\begin{aligned} \text{OLS: } & [0.2516;\ 0.2525;\ 0.2535;\ 0.2524]; \\ \text{MOLS: } & [0.2519;\ 0.2524;\ 0.2532;\ 0.2523]; \\ Ridge(0.1): & [0.2459;\ 0.2461;\ 0.2465;\ 0.2461]. \end{aligned} \quad (5)$$

Coefficients of variation (CVs):

$$\begin{aligned} \text{OLS: } & [36.01;\ 35.69;\ 36.06;\ 35.93]; \\ \text{MOLS: } & [18.89;\ 18.75;\ 18.85;\ 18.79]; \\ Ridge(0.1): & [9.53;\ 9.48;\ 9.54;\ 9.49]. \end{aligned} \quad (6)$$

Comparing with the case of identical laws of influence of regressors on the response (equal $\theta_j$ in Table 1), we see that the bias and the stability of all methods have not changed. Again, the *ridge*(0.1) method is the best despite a small bias.

Consider now the case when all $\theta_j$ are equal but $d_j$ are different. In other words, the law of influence on the response is the same for all regressors, but each regressor is subjected to a different effect of unaccounted factors. Let be $\alpha = 0.2$, $n = 30$, $\theta = [1°; 1°; 1°; 1°]$ but $d = [1; 3; 1; 3]$. The *beta*-coefficients now will be: [large; small; large; small]. The calculated mean *beta*-coefficients are:

$$OLS: [0.4517\ 0.0573\ 0.4502\ 0.0576];$$
$$MOLS: [0.4407\ 0.0685\ 0.4402\ 0.0687]; \quad (7)$$
$$Ridge(0.1): [0.3708\ 0.1281\ 0.3707\ 0.1282].$$

The coefficients of variation (CVs) of the *beta*-coefficients:

$$OLS: [0.4517\ 0.0573\ 0.4502\ 0.0576];$$
$$MOLS: [0.4407\ 0.0685\ 0.4402\ 0.0687]; \quad (8)$$
$$Ridge(0.1): [0.3708\ 0.1281\ 0.3707\ 0.1282].$$

We see in (7) that the *ridge*(0.1) estimations of the population solution became more biased. The *beta*-coefficients became smoothed: large coefficients decreased, while smaller ones became significantly larger.

This is especially noticeable for smaller coefficients, which have doubled. Considering that for $\alpha = 0.2$ there is practically no collinearity of data, the MOLS and *ridge*(0.1) solutions should, on average, be close to the OLS solution. This is really so for the MOLS, but this is no longer the case for the *ridge*(0.1) method, as seen in (7).

All three methods provide inadequate solutions for $n = 30$ (CV > 10 %) regarding the adequacy. That is, for $n = 30$ and different $d_j$, all solutions became significantly more variable. This is especially true for smaller values of the regression coefficients. It means that in a real situation, for small samples, the solution to the LR problem may well be economically correct but significantly inadequate for all considered methods.

For larger sample sizes, the situation improves but very slowly for $\alpha = 0.2$ (no near-collinearity). In such a case, only for $n > 1000$ all solutions become adequate. The calculated mean *beta*-coefficients become in this case:

$$OLS: [0.4508\ 0.0573\ 0.4507\ 0.0574];$$
$$MOLS: [0.4441\ 0.0644\ 0.4440\ 0.0645]; \quad (9)$$
$$Ridge(0.1): [0.3741\ 0.1244\ 0.3741\ 0.1244].$$

The coefficients of variation (CVs) of the *beta*-coefficients become:

$$OLS: [3.5743\ 13.8269\ 3.5990\ 13.9954];$$
$$MOLS: [13.96\ 71.756\ 13.97\ 71.54]; \quad (10)$$
$$Ridge(0.1): [1.6169\ 5.0790\ 1.6419\ 5.1409].$$

With the appearance of near-collinearity, the situation changes dramatically for all sample sizes. OLS solutions become more unstable and therefore inadequate, but the *ridge*(0.1) solutions become more stable and adequate. The MOLS solution becomes adequate for even larger sample sizes in this case.

Thus, the most noticeable *drawback* of the *ridge*(0.1) method compared to the MOLS is the smoothing of the regression coefficients. Of course, one can interpret this drawback as an increase in the bias for different *beta*-coefficients in the population, but it is namely a smoothing because, with the same coefficients in the population, the bias of the *ridge*(0.1) solution does not manifest itself noticeably. It is worth noting that only strictly linear data were considered above, although, for nonlinear data, the situation with the adequacy of the *ridge*(0.1) method remains practically the same.

*OLS consistency peculiarities.*

It is proved (see, for example, [9]) that any OLS solution to the LR is consistent. This property was used in [1; 2] and above for not near-collinear data (VIF~20) to estimate the population solution using the OLS solution for a linear sample of size $10^7$. Unfortunately, when modeling data, we cannot always use the OLS solution for large sample sizes to estimate the population's solution.

For the case considered above: $\theta = [1°; 1°; 1°; 1°]$, and $d = [1; 3; 1; 3]$, but with large enough near-collinearity, $\alpha = 0.01$ (VIF~$10^3$), the OLS solution converges to the population solution too slowly (if so at all) to be used in modeling as an estimate of the population solution.

Table 2 shows the OLS, MOLS, and *ridge*(0.1) solutions for large sample sizes generated by another principle than in Table 1.

The first two regressors, $x_1$ and $x_2$ was generated with the aid of the algorithm (3):

$$x_j = k_j^{-1}(Y + d_j \alpha \varepsilon),$$

the latter two ones, $x_3$ and $x_4$ were generated with the aid of the algorithm $x_j = k_j^{-1}(y + d_j \varepsilon)$. The second algorithm differs from the first by the absence of the $\alpha$-parameter, which defines the level of incoherent noise in the regressors. Thus, the regressors $x_3$ and $x_4$ do not depend on $\alpha$ at all, and due to the independence of the regressors, their contributions to the

**Table 2**

**The OLS convergence under a near-collinearity**

| n | $n = 10^6$, $\alpha = 0.05$ (VIF ~ 225) | | | | $n = 10^7$, $\alpha = 0.005$ (VIF ~ $2 \cdot 10^4$) | | | |
|---|---|---|---|---|---|---|---|---|
| **betas** | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_3$ |
| OLS | 0.4393 | 0.4388 | 0.1102 | 0.0128 | 0.4995 | 0.4991 | 0.0013 | 0.0002 |
| MOLS | 0.3162 | 0.3163 | 0.3058 | 0.064 | 0.3178 | 0.3178 | 0.3065 | 0.0599 |
| *Ridge* | 0.2825 | 0.2825 | 0.2624 | 0.1521 | 0.2854 | 0.2854 | 0.2585 | 0.1498 |

response should not depend on the $\alpha$-parameter. Therefore, *beta*-coefficients $\beta_3$ and $\beta_4$ should also not depend on this parameter. Also, since $d_4$ is larger than $d_3$, then $\beta_4$ should be less than $\beta_3$. Since $d_1$, $d_2$ and $d_3$ are equal, then $\beta_1$, $\beta_2$, and $\beta_3$ should also be approximately equal since the solutions to the LR problem should not depend on the regressors' collinearity level due to their independence. It is clear that the above conditions should be met the more accurately the larger the sample size.

All these conditions are strictly fulfilled for the MOLS and *ridge*(0.1) solutions, considering the existing bias and rounding of solutions by these methods. As for the OLS solution, one can see that $\beta_3$ and $\beta_4$ significantly depend on the $\alpha$-parameter. With a decrease in this parameter, these coefficients tend to zero. For example, if $n = 10^7$, = 0.001 (VIF $\sim 5 \cdot 10^5$) the OLS solution is: [0.4999; 0.5001; 0.0001; 0.0000]; the MOLS and *ridge*(0.1) solutions do not change practically: [0.3178; 0.3178; 0.3065; 0.0599]; [0.2854; 0.2854; 0.2584; 0.1498].

This means that in the presence of a near-collinearity, the OLS solutions for extremely large samples may be inadequate to the solution in the population for some data. That is, it cannot be stated that if the sample is large, then we can confidently assume that the OLS provides adequate solutions. An exception is a case of strictly or approximately linear regressors with the same or approximately the same variances of incoherent noise (residual error in partial regressions of $x_j$ on $y$). The linearity can be checked graphically, for instance. The residual errors can be estimated by the solution of all partial regressions of $x_j$ on $y$. If they differ insignificantly, the researcher can use the OLS solution for a large enough sample as the adequate one with confidence.

*Nonlinearity and autocorrelation in data.*

The nonlinearity of the regressor manifests itself in an increase in the variance of the residual error in the problem of the simple linear regression of this regressor on the response. This is equivalent to increasing $d_j$ in algorithm (3), which leads to a decrease in the corresponding *beta*-coefficient. Nonlinearity can be modeled in different ways. In this work, we use a new function for this:

$$P(n,x) = \text{sign}(x)|x|^n . \qquad (11)$$

With the aid of this function, we can add a nonlinearity in the algorithm (3) either by changing the residual error term,

$$x_j = k_j^{-1}(y + d_j \alpha P(n,\varepsilon)) \qquad (12)$$

or by changing the whole regressor,

$$x_j \rightarrow x_j^{(n)} = P(n,x_j). \qquad (13)$$

The use of formula (12) is simply equivalent to an increase in the residual error ($d_j$) in the algorithm (3). At the same time, the use of formula (13) leads to increasing the nonlinearity in the regressor $x_j$. This effect we can see in Fig. 3 a) for $n = 200$, $\alpha = 0.1$, $d_j = 1$, $\theta_j = 1$ and the power $n = 3$ in (11).
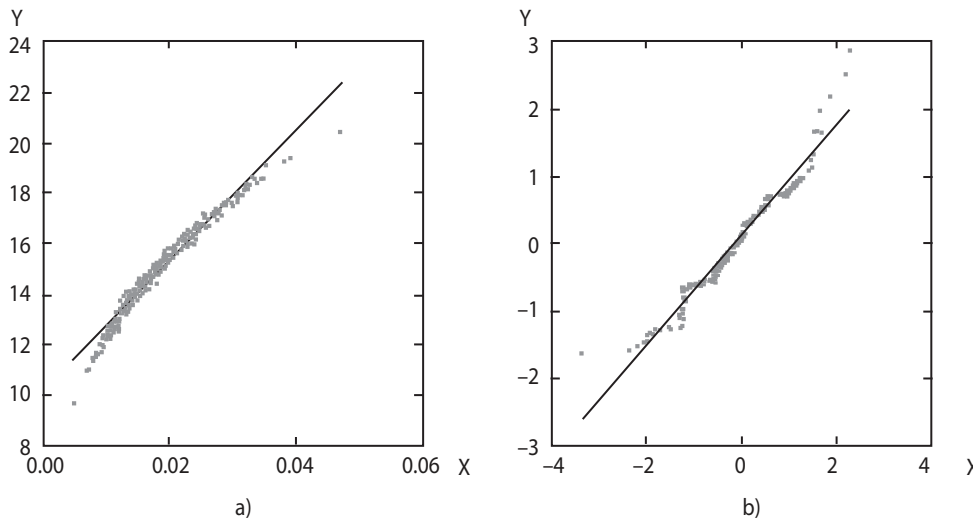


Figure 3. Simple regressions of $y$ on $x_j$.
a) nonlinearity;  b) autocorrelation

The autocorrelation we model in this paper with the aid of the sorted *pseudo*-random variables (*randn.m* in the MATLAB). The data (a regressand and $m$ regressors) that we use below has been generated as follows:

$$y = sort(randn(n,1)); \quad x = sort(randn(n,m)). \quad (14)$$

The variables generated in such a way have significant autocorrelation due to the sorting of random values of the *pseudo*-random number generator. We will not go into the estimation of the autocorrelation level since it is not essential now. Let us just say that the autocorrelation is significant as to the Durbin-Watson test. In Fig. 3 b), we show the simple regression of $y$ on one regressor of $x$ from (14) for $n = 100$.

We consider now the influence of the nonlinearity and autocorrelation on the LR problem solutions for the OLS, MOLS, and *ridge*(0.1) method by comparison with the solutions for strictly linear regressors. Here, we should consider that in the algorithm (14) all beta coefficients in the population should be the same, but in the scheme (3) it depends on the choice of parameters.

For Table 3, we take $\alpha = 0.1$ and $d_j = 1$ in (3) for the first two regressors ($x_1$ and $x_2$) and apply the function (10) to $x_1$ with $n = 3$ (nonlinearity as in Fig. 3a). For the second two regressors, we apply the scheme (14). Since we add the nonlinearity to $x_1$, the value of $\beta_1$ should be less than that of $\beta_2$ in the population; both values, $\beta_3$ and $\beta_4$, should be strictly equal in the popula-

tion because all regressors in (14) are statistically equal. Due to the properties of the algorithm (3), $\beta_2$ should be approximately the same as $\beta_3$ and $\beta_4$.

To this combined data, we apply all three methods mentioned above to verify their adequacy. Table 3 shows the results.

**Table 3**

**The OLS convergence under the nonlinearity and autocorrelation**

| n | $n = 10^6$, $\alpha = 0.1$ (VIF ~ $3 \cdot 10^5$) | | | | $n = 10^7$, $\alpha = 0.1$ (VIF ~ $2 \cdot 10^6$) | | | |
|---|---|---|---|---|---|---|---|---|
| **betas** | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_3$ |
| OLS | 0.0042 | 0.0002 | 0.3900 | 0.6057 | -0.0000 | 0.0000 | 0.3020 | 0.6980 |
| MOLS | 0.1327 | 0.2840 | 0.2928 | 0.2925 | 0.1322 | 0.2848 | 0.2925 | 0.2925 |
| *Ridge* | 0.1866 | 0.2464 | 0.2726 | 0.2725 | 0.1861 | 0.2468 | 0.2726 | 0.2726 |

From this table, we can see that the OLS solutions may not converge to the population solution. The OLS solution tends to the solution [0; 0; 0.3; 0.7] as the sample size increases, which is entirely inconsistent with what is described above. It should be: $\beta_1 < \beta_2$; $\beta_1 \approx \beta_3 = \beta_4$.

For greater clarity of the situation with OLS solutions, we consider repeating the samples $10^4$ times to assess the adequacy of the OLS, MOLS, and *ridge*(0.1) solutions for samples smaller than in Table 3.

For this, we calculate the mean and coefficient of variation (CV) of these three methods' solutions for different sample sizes in Table 4.

It worth noting some features of the data generated in the latter case. With an increase in *n*, the values of $\beta_1$ and $\beta_2$ change smoothly and stabilize around the values of 0.1322 and 0.2848 for the MOLS and *ridge*(0.1) solutions, which one can see in Table 3 and Table 4. The level of near-collinearity does not significantly affect their values, as it should be.

At the same time, the OLS $\beta_1$ and $\beta_2$ simply tend to zero, and their variability clearly increases under the growth of the VIF. The values of $\beta_3$ and $\beta_4$ change smoothly and stabilize around 0.2925 for the MOLS solutions and around 0.2726 for the *ridge*(0.1) solutions, as shown in Tables 3, 4. It seems that in the OLS $\beta_3$ and $\beta_4$ tend to 0.3 and 0.7, respectively.

**Table 4**

**The OLS, MOLS, and *ridge*(0.1) stability demonstration**

| n | $n = 40$, VIF ~ 20 | | | | $n = 10^2$, VIF ~ 150 | | | |
|---|---|---|---|---|---|---|---|---|
| **betas** | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_3$ |
| OLS mean | 0.2171 | 0.3440 | 0.2244 | 0.2244 | 0.1522 | 0.4856 | 0.1841 | 0.1824 |
| MOLS mean | 0.2306 | 0.3317 | 0.2242 | 0.2236 | 0.2094 | 0.3295 | 0.2326 | 0.2332 |
| *ridge* mean | 0.2367 | 0.2852 | 0.2314 | 0.2312 | 0.2235 | 0.2886 | 0.2337 | 0.2339 |
| OLS CV (%) | 48.22 | 38.85 | 85.86 | 86.01 | 57.81 | 28.14 | 81.11 | 81.33 |
| MOLS CV (%) | 33.86 | 28.35 | 37.22 | 37.64 | 27.23 | 19.99 | 21.77 | 21.49 |
| *ridge* CV (%) | 17.59 | 16.33 | 19.68 | 19.89 | 12.01 | 10.59 | 11.06 | 10.94 |
| n | $n = 10^3$, VIF ~ $10^3$ | | | | $n = 10^4$, VIF ~ $5 \cdot 10^3$ | | | |
| OLS mean | 0.0396 | 0.1544 | 0.4080 | 0.3992 | 0.0055 | 0.0232 | 0.4842 | 0.4873 |
| MOLS mean | 0.1427 | 0.2842 | 0.2878 | 0.2879 | 0.1333 | 0.2846 | 0.2922 | 0.2921 |
| *ridge* mean | 0.1918 | 0.2532 | 0.2667 | 0.2667 | 0.1868 | 0.2476 | 0.2719 | 0.2719 |
| OLS CV (%) | 172.34 | 35.47 | 51.07 | 52.37 | 516.04 | 36.32 | 44.29 | 44.04 |
| MOLS CV (%) | 13.09 | 6.07 | 3.14 | 3.14 | 4.12 | 2.03 | 0.98 | 0.99 |
| *ridge* CV (%) | 5.29 | 3.27 | 2.08 | 2.11 | 1.78 | 1.00 | 0.55 | 0.55 |

That is, it is clearly seen that for these data the OLS solutions are not adequate to the solutions in the population, in any case, up to $n = 10^7$, although they are economically correct.

This example shows that the OLS can give inadequate solutions even for large samples. However, the researcher can consider that the OLS solution is correct and consistent since it

seems it is approaching a certain limit with $n$ increasing (in this case, to [0; 0; 0.3; 0.7]). However, this limit can be very far from the true value in the population, as it is in our last example. At the same time, in all the examples considered by us, the MOLS and *ridge*(0.1) methods give stable solutions that are sufficiently close to each other and the solutions in the population.

Another conclusion, which follows from Table 4, concerns the adequacy of the LR problem's solution for small samples. From the calculations, it follows that the inclusion of a nonlinear regressor or a regressor with autocorrelation in the model significantly increases the variability of the solution for all methods. As follows from Table 4, in the presence of nonlinearity and (or) autocorrelation in regressors, the appearance of adequate MOLS and ridge solutions is shifted towards large sample sizes. Table 4 demonstrates the effect of nonlinearity and autocorrelation on the stability of the solution to the LR problem. In contrast to strictly linear data, for which the solutions of the LR problem are adequate, starting with sample sizes greater than 10-30 ([1; 2] and Table 1 in this paper), the presence of nonlinearity or autocorrelation of data can lead to the inadequacy of solutions for sample sizes less than 1000. This

casts doubt on the need to retain explicitly nonlinear regressors or regressors with the autocorrelation in a linear model.

To demonstrate the stability of the *ridge*(0.1) and MOLS(0.001) solutions, consider the solution of the LR problem for actual data, which is considered a benchmark for the presence of strong near-collinearity (and the presence of autocorrelation in one regressor).

*Real data.*

Let us consider the LR problem's solution for the real data [9, p. 1150, Table F4.2 (The Longley data)]: $y$ – Employment; $x_1$ – GNP deflator; $x_2$ – GNP; $x_3$ – Armed Forces; $x_1$ – Year. The sample size is small enough ($n = 16$), VIF ≈ 150.

As one may conclude from [1; 2], we can expect a large variability of solutions to the LR problem for the OLS for such a small sample. In this article, we will not touch on the issue of resampling the real data, but to demonstrate the stability properties of all mentioned methods for solving the LG problem, we will apply a simple method of discarding the last observation (the «leave one out» method), as done in [9, p. 131]. Table 5 shows the OLS, *ridge* (0.1), and MOLS solutions for the whole data and without the last observation.

**Table 5**

**The OLS, MOLS, and *ridge*(0.1) solutions for the whole Longley's data and that without the last observation**

|  | OLS | | *ridge*(0.1) | | MOLS | |
|---|---|---|---|---|---|---|
| n | 16 | 15 | 16 | 15 | 16 | 15 |
| $b_0$ | 1169088 | 1459415 | -350821 | -341932 | -410474 | -407674 |
| $b_1$ | -19.7681 | -181.1230 | 87.7240 | 82.8068 | 91.6909 | 91.8554 |
| $b_2$ | 0.0644 | 0.0911 | 0.0135 | 0.0145 | 0.0129 | 0.0131 |
| $b_3$ | -0.0101 | -0.0749 | 0.2067 | 0.1941 | 0.1525 | 0.1450 |
| $b_4$ | -576.4643 | -721.7561 | 205.3931 | 200.9221 | 235.8999 | 234.4268 |

For the OLS, the solutions are strictly the same as in [9, p. 131], and we demonstrate them only for convenience. In [9, p. 131], the author draws attention to the large percentage of changes in the regression coefficients (600, 800) but says nothing about the economic incorrectness of the OLS solution at all.

Without going into details, it can be seen that all regression coefficients must be positive since all variables increase with time. That is, the OLS solution for this problem cannot be used in any approach. At the same time, the *ridge*(0.1) and the MOLS provide economically correct solutions that differ little from each other. As it should be, the *ridge*(0.1) method gives solutions with smaller absolute values due to the larger regularization constant, but the difference is small. The small coefficients' values ($b_2$, $b_3$), which are larger than that of the MOLS, are due to the rounding property of the *ridge*(0.1) method.

As in [9], let us estimate the stability of the *ridge*(0.1) and MOLS methods using the percentage change in the regression coefficients when discarding the last observation, which we calculate as follows: $\Delta b = \left| b_{16} - b_{15} \right| / \left| b_{16} \right|$. For $b_1$ we have for

the *ridge*(0.1), $\Delta b_1 \approx 5.6\,\%$; for the MOLS, $\Delta b_1 \approx 0.18\,\%$. For $b_2$ we have for the *ridge*(0.1), $\Delta b_2 \approx 7.4\,\%$; for MOLS, $\Delta b_2 \approx 1.6\,\%$. For $b_3$ we have for the *ridge*(0.1), $\Delta b_3 \approx 6.1\,\%$; for the MOLS, $\Delta b_3 \approx 4.9\,\%$. For $b_4$ we have for the *ridge*(0.1), $\Delta b_4 \approx 2.2\,\%$; for the MOLS, $\Delta b_4 \approx 0.6\,\%$. If these estimates were statistical, then, as shown in [1; 2], values of the stability indicator (CV) less than 10% correspond to a 95% probability that the solution is adequate. Based on this, we can assume that both methods (MOLS and *ridge*(0.1)) give adequate solutions for the Longley data.

However, judging by the results obtained in Table 4, the presence of nonlinearity or autocorrelation in regressors leads to a significant increase in the variability of solutions to the LR problem. Therefore, it should be recognized that the LOO method is yet not equivalent to a repetition of samples for estimating the coefficient of variability (CV) of solutions to the LR problem.

Yes, indeed, the LOO method testifies to the stability of the solution but does not say anything about its adequacy.

It is clear that a stable and mathematically correct solution to the LR problem reflects the real situation, but this situation strongly depends on the sample size. As follows from the calculations given in Table 4, one should not delude ourselves about the adequacy of the solutions to the LR problem obtained in Table 5 for the Longley data because of the very small sample size ($n = 16$).

If we examine Longley's data graphically, we will see that the regressors $x_1$, $x_2$, and $x_4$ are practically linear. As for $x_3$, the presence of significant autocorrelation is clearly visible. It is this regressor that is responsible for the potentially high variability of the LR solution. Let us say, for economic reasons, it is permissible to exclude $x_3$ from the model. A new solution will be (VIF ≈ 130):

$$OLS: \left[b_0 = 1.1579e+06; \ b_1 = -21.2785; \ b_2 = 0.0642; \ b_4 = -570.6641\right];$$
$$MOLS: \left[b_0 = -4.0311e+05; \ b_1 = 96.3755; \ b_2 = 0.0130; \ b_4 = 232.0640\right];$$
$$Ridge(0.1): \left[b_0 = -3.4517e+05; \ b_1 = 92.3378; \ b_2 = 0.0138; \ b_4 = 202.4829\right].$$

If we compare this solution with that in Table 4 (with the presence of $x_3$), we will see that the regression coefficients for the MOLS and *ridge*(0.1) have practically not changed. However, the closeness of the data to linear in this case guarantees the closeness of the solution for $n = 16$ to the solution in the population even for such a small sample, as shown in [1; 2].

Strictly speaking, in the case of one sample, we can only say that both methods give fairly stable and close solutions, although they are obtained by different methods. *Since both mathematical methods are correct and solve the same matrix equation but with different regularization constants, the proximity of solutions to each other and their stability guarantees the proximity of these solutions to the population's solution for a large enough sample.*

Using the ADG, it is easy to show, and this can be seen from Table 1, that in the absence of significant collinearity (α = 0.2), the OLS solution almost exactly coincides with the MOLS solution, but the *ridge*(0.1) solution is slightly smaller in absolute value. With an increase in the data collinearity, the OLS solution deteriorates for different data types in different ways, but the proximity between the MOLS and *ridge*(0.1) solutions remains. *This gives the right to believe that the similarity of the MOLS and ridge solutions guarantees their closeness to the solution in the population for a large enough sample. In any case, this issue is confirmed by the numerous model experiments carried out by the authors, and we can think so until the opposite is proved.*

Depending on the problem being solved, the researcher can choose which method to use. The *ridge*(0.1) method is preferable, in our opinion, for small samples of a size less than 40, since in this case, it has higher stability of the solutions that is important for small samples. Herewith, you will have to come to terms with the inevitable rounding of solutions. The studies carried out guarantee that there will be no qualitative error, and a small quantitative error in the regression coefficients should not, as in our mind, affect the results of economic inferences. For samples of a size larger than 40, one should, in our opinion, use the MOLS(0.001) as more precise.

More reasonably, this problem will be considered in subsequent works using the observed data resampling methods.

**Summarizing** the above, we note that the comparison of the widely known *ridge*(0.1) method [12] with the new MOLS(0.001) method [1; 2] allows us to assert that the problem of solving the linear regression problem in the presence of a near-collinearity has been practically solved in [12; 1; 2] and this paper.

First, two regularization versions of the OLS equation are considered, COV and ST. In the COV version, the well-known ridge method has an approximately constant optimal regularization parameter $\lambda_{opt} \approx 0.1$. In the commonly used ST version, the optimal regularizer had to be $\lambda_{opt} \approx 0.1(n-1)$, where $n$ is the sample size. The determination of the optimal regularizer for the ridge method finally opens up the possibility of its widespread use in practice fifty years after its theoretical presentation by the author.

Second, a new method for solving the regularized OLS equation, the MOLS, has been proposed in [1; 2], which gives a stable solution to the ill-conditioned OLS equation in standardized variables for any conditioning level with the aid of the Modified Cramer Rule [1, App. A], which is fundamentally different from the Cramer Rule used in the *ridge* method [12] but gives practically the same solution to the LR problem for any sample size and collinearity level (only for standardized data).

Third, both methods, *ridge*(0.1) and MOLS, mathematically correctly solve the same regularized OLS equation, but with different regularization constants, 0.1 and 0.001, and give stable and close solutions for any degree of data collinearity for any sample sizes.

Fourth, both methods, *ridge*(0.1) and MOLS give close solutions that both tend approximately to the population solution for any ADG-artificial data with the sample size increasing.

All the above give reasons to believe that both methods (*ridge*(0.1) and MOLS) are approximately unbiased (the MOLS to a slightly greater extent than the ridge) and stable (the *ridge*(0.1) to a somewhat greater extent than the MOLS). They converge pretty fast to a population solution (in probability). This one depends on the degree of linearity of the regressors.

With this, the following features of the LR solution were clarified. The *ridge*(0.1) solutions are smoother than the solutions of the MOLS. The more the solutions (the regression coefficients) differ from each other in the population, the smoother is the *ridge*(0.1) solution. However, it is worth noting that the difference between the *ridge*(0.1) and MOLS solutions does not significantly affect the economic inferences as to our mind. In general, both methods can be used in economic research, but one should keep in mind that the *ridge*(0.1) method is preferable for very small samples in terms of stability. For medium and large samples, the MOLS(0.001) is preferable due to its higher accuracy with practically equally acceptable stability.

Both methods are more adequate to the population solution for small sample sizes if the regressors' collinearity level increases. At the same time, the adequacy of the OLS solution decreases.

If the VIF of data is small, about 10-20, the MOLS and ridge(0.1) methods practically give the same solutions as the OLS. On the other side, the adequacy of all solutions is very low due to a significant influence of the uncounted stochastic factors, which decrease the VIF and, at the same time, the absolute value of some regressors' coefficients (those, in which the stochastic noise is considerable).

As for the practical use of the *ridge* method, we note that, as shown in the article, it is applicable only in the COV-version of the *ridge* method with a constant regularizer equal to 0.1. In the standard version of the *ridge* method, which is usually used in packages, the regularizer depends on the sample size. Therefore, before using the ridge method code, it is necessary to check which version is used. The easiest way is to solve the LR problem for the well-known Longley data with the regularizer 0.1 and check the solution against the one given in this article (Table 5). If the solutions do not meet, one should replace the regularizer with $0.1(n$-$1)$.

In conclusion, note that:

1) both methods (MOLS and ridge(0.1)) are sufficiently optimal to stop looking for the exact value of the optimal regularizer;

2) the application of both methods give very close solutions, which confirm their closeness to the solution in the population;

3) both methods make it possible not to discard strongly correlated regressors that generate large values of the VIF factor;

4) the presence of strongly correlated regressors significantly increases the degree of adequacy of solutions to the LR problem for small samples.

### LITERATURE

**1.** Tyzhnenko A. G. A new stable solution to the linear regression problem under multicollinearity. *Economics of Development.* 2018. Vol. 2 (86). P. 89–99. URL: http://www.ed.ksue.edu.ua/ER/knt/ee182_86/e182tyz.pdf

**2.** Tyzhnenko A. G., Ryeznik Y. V. Ordinary List Squares: The Adequacy of Linear Regression Solutions under Multicollinearity and without it. *The Problems of Economy.* 2019. Vol. 1 (39). P. 217–227.

DOI: 10.32983/2222-0712-2019-1-217-227

**3.** Seber G. A. F. LR Analysis. New York : Wiley-Blackwell, 1977. 456 p.

**4.** Seber G. A. F. LR Analysis. New York : Wiley, 2003. 341 p.

**5.** Spanos A. Probability Theory and Statistical Inference: econometric modeling with observational data. Cambridge : Cambridge University Press, 1999. 401 p.

**6.** Gujarati D. N. Basic econometrics. New York : McGraw-Hill, 2002. 526 p.

**7.** Wooldridge J. M. Introductory Econometrics: Modern Approach. Ohio : South-Western, 2009. 633 p.

**8.** Baltagi B. Econometrics. New York : Springer, 2011. 812 p.

**9.** Greene W. H. Econometric Analysis. New York : Pearson, 2012. 1211 p.

**10.** Draper N. R., Smith H. Applied Regression Analysis. New York : Wiley. 1966. 445 p.

**11.** Farrar D., Glauber R. R. Multicollinearity in regression Analysis: The problem revisited. *Review of Economics and Statistics.* 1967. Vol. 49. P. 92–107.

**12.** Hoerl A. E., Kennard R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics.* 1970. Vol. 12 (1). P. 55–67.

**13.** Marquardt D. V. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics.* 1970. Vol. 12. P. 591–612.

**14.** Blanchard O. J. Comment. *Journal of Business and Economic Statistics.* 1987. No. 5. P. 449–451.

**15.** Adkins L. C., Hill R. C. Collinearity // Companion in Theoretical Econometrics / ed. by Badi Baltagi. Oxford : Blackwell Publishers, Ltd, 2001. Ð. 256–278.

**16.** Belsley D. A., Kuh K., Welsh R. E. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York : Wiley, 1980. 292 ð.

**17.** Belsley D. A. Demeaning conditioning diagnostics through centering. *The American Statistician*. 1984. Vol. 38 (2). P. 73–77.

**18.** Rao C. R., Toutenberg H. Linear Models: Least Squares and Alternatives. New York : Springer, 1999. 301 ð.

**19.** Spanos A., McGuirk A. The Problem of Near-Multicollinearity Revisited: erratic vs. systematic volatility. *Journal of Econometrics.* 2002. Vol. 108. P. 365–393.

**20.** Kabanichin S. I. Definitions and Examples of Inverse and Ill-posed Problems. *J. Inv. Ill-Posed Problems.* 2008. Vol. 16. P. 317–357.

**21.** Adkins L. C., Waters M. S., Hill R. C. Collinearity Diagnostics in gretl, Economics Working Paper Series 1506. Oklahoma : Oklahoma State University, Department of Economics and Legal Studies in Business, 2015. 452 p.

**22.** Fox J. Applied regression analysis, linear models, and related methods. Thousand Oaks, CA : Sage Publications, 1997. 742 p.

**23.** Maddalla G. S. Introduction to Economics. New York : Macmillan, 1992. 396 p.

**24.** Harvey A. C. Some Comments on Multicollinearity in Regression. *Applied Statistics*. 1977. Vol. 26 (2). P. 188–191.

**25.** Tikhonov A. N. On the stability of inverse problems. *Doklady Acad. Sci. USSR.* 1943. Vol. 39. P. 176–179.

**26.** Tikhonov A. N., Arsenin, V. Y. Solutions of Ill-Posed Problems. New York : Winston & Sons, 1977. 287 p.

**27.** Dougherty C. Introduction to Econometrics. New York : Oxford University Press, 1992. 402 p.

### REFERENCES

Adkins, L. C., and Hill, R. C. "Collinearity". In *Companion in Theoretical Econometrics*, 256-278. Oxford: Blackwell Publishers, Ltd, 2001.

Adkins, L. C., Waters, M. S., and Hill, R. C. *Collinearity Diagnostics in gretl, Economics Working Paper Series 1506*. Oklahoma: Oklahoma State University, Department of Economics and Legal Studies in Business, 2015.

Baltagi, B. *Econometrics*. New York: Springer, 2011.

Belsley, D. A. "Demeaning conditioning diagnostics through centering". *The American Statistician*. vol. 38 (2) (1984): 73-77.

Belsley, D. A., Kuh, K., and Welsh, R. E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley, 1980.

Blanchard, O. J. "Comment". *Journal of Business and Economic Statistics*, no. 5 (1987): 449-451.

Dougherty, C. *Introduction to Econometrics*. New York: Oxford University Press, 1992.

Draper, N. R., and Smith, H. *Applied Regression Analysis*. New York: Wiley, 1966.

Farrar, D., and Glauber, R. R. "Multicollinearity in regression Analysis: The problem revisited". *Review of Economics and Statistics*, vol. 49 (1967): 92-107.

Fox, J. *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage Publications, 1997.

Greene, W. H. *Econometric Analysis*. New York: Pearson, 2012.

Gujarati, D. N. *Basic econometrics*. New York: McGraw-Hill, 2002.

Harvey, A. C. "Some Comments on Multicollinearity in Regression". *Applied Statistics*, vol. 26 (2) (1977): 188-191.

Hoerl, A. E., and Kennard, R. W. "Ridge regression: Biased estimation for nonorthogonal problems". *Technometrics*, vol. 12 (1) (1970): 55-67.

Kabanichin, S. I. "Definitions and Examples of Inverse and Ill-posed Problems". *J. Inv. Ill-Posed Problems*, vol. 16 (2008): 317-357.

Maddalla, G. S. *Introduction to Economics*. New York: Macmillan, 1992.

Marquardt, D. V. "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation". *Technometrics*, vol. 12 (1970): 591-612.

Rao, C. R., and Toutenberg, H. *Linear Models: Least Squares and Alternatives*. New York: Springer, 1999.

Seber, G. A. F. *LR Analysis*. New York: Wiley, 2003.

Seber, G. A. F. *LR Analysis*. New York: Wiley-Blackwell, 1977.

Spanos, A. *Probability Theory and Statistical Inference: econometric modeling with observational data*. Cambridge: Cambridge University Press, 1999.

Spanos, A., and McGuirk, A. "The Problem of Near-Multicollinearity Revisited: erratic vs. systematic volatility". *Journal of Econometrics*, vol. 108 (2002): 365-393.

Tikhonov, A. N. "On the stability of inverse problems". *Doklady Acad. Sci. USSR*, vol. 39 (1943): 176-179.

Tikhonov, A. N., and Arsenin, V. Y. *Solutions of Ill-Posed Problems*. New York: Winston & Sons, 1977.

Tyzhnenko, A. G. "A new stable solution to the linear regression problem under multicollinearity". Economics of Development. 2018. http://www.ed.ksue.edu.ua/ER/knt/ee182_86/e182tyz.pdf

Tyzhnenko, A. G., and Ryeznik, Y. V. "Ordinary List Squares: The Adequacy of Linear Regression Solutions under Multicollinearity and without it". *The Problems of Economy*, vol. 1 (39) (2019): 217-227.

DOI: 10.32983/2222-0712-2019-1-217-227

Wooldridge, J. M. *Introductory Econometrics: Modern Approach*. Ohio: South-Western, 2009.

## Appendix A

The MATLAB code (mcr.m) for the regularized OLS equation (2) for the standardized regression model $\{y, x\}$ – the observed data; $\{Y, X\}$ – the standardized data. $\beta$ – the *beta*-coefficients.

$$\left(\frac{1}{n-1}X'X + \lambda I\right)\beta = \frac{1}{n-1}X'Y, \qquad (2)$$

$$A = \frac{1}{n-1}X'X, \quad B = \frac{1}{n-1}X'Y.$$

$$beta = mcr(A, B, 1e-3).$$

### 1. The Modified Cramer Rule

function $beta = mcr(A, B, lambada)$

```
n=length(B);H1=A'*A;B1=A'*B; E=eye(size(H1));
H2=lambda.*E+H1;X=zeros(n,1);
for i=1:n; ti=[1:i-1 i+1:n]; H3=H2(ti,ti);H3=H3\eye
(size(H3));D=0; D1=0;
for k=1:n; tk=[1:k-1 k+1:n]; D=D+(-1)^(i+k)*B1(k)*
det(H3*H2(tk,ti));
D1=D1+(-1)^(i+k)*H2(k,i)*det(H3*H2(tk,ti)); end;
X(i)=D/D1; end; beta=X';
```

### 2. The COV-version of the MOLS method
$(\lambda_{opt} = 0.001):$

```
function [b,beta]=RegMOLS(y,x,lambda)
[n,m]=size(x); my=mean(y);mx=mean(x); sy=std(y);
sx=std(x); EE=[y x];
EE=standard(EE); % Standardization
YY=EE(:,1); XX=EE(:,2:m+1); BB=XX'*YY; BB=BB/
(n-1);
AA=XX'*XX; AA=AA/(n-1);
beta=mcr(AA,BB,lambda); b=(beta.*sy)./sx; b0=my-
mx*b'; b=[b0 b]; b=b'; beta=beta';
```

### 3. The COV-version of the *ridge* method
$(\lambda_{opt} = 0.1):$

```
function [b,beta]=RegRidge(Y,X,lambda)
[n,m]=size(X); my=mean(Y);mx=mean(X); sy=std(Y);
sx=std(X);
EE=[Y X]; EE=standard(EE); % Standardization
YY=EE(:,1); XX=EE(:,2:m+1); BB=XX'*YY; BB=BB/
(n-1);
AA=XX'*XX; AA=AA/(n-1);
AA=AA+lambda*eye(size(AA));
beta=AA\BB; b=(beta.*sy)./sx'; b0=my-mx*b; b=[b0 b'];
b=b';
```

### 4. Standardization code:
```
function E=standard(E)
[n,m]=size(E); E=E-ones(n,m)*diag(mean(E));
E=E*diag(1./std(E));
```