Barometr Regionalny
Tom 15 nr 3

A Note on the UEK Method

Anna Pajor
Cracow University of Economics, Poland

Abstract

The paper concerns certain pitfalls of using the Moore-Penrose pseudoinverse for estimating regression coefficients in linear regression models when the matrix of explanatory variables has not full column rank. The aim of the paper is to show that in this case estimator of parameters based on the Moore-Penrose pseudoinverse is biased, and the bias leads to biased forecasts.

Keywords: linear regression, MP-inverse matrix, UEK method

JEL: C13, C20

Introduction

The multiple linear regression model is one of the most popular models in statistics. It is used to explain the relationship between a continuous dependent variable and the so-called explanatory (independent) variables. One of the basic assumptions of the linear regression model is that the matrix of explanatory variables ($\mathbf{U}$) has full column rank. This condition is necessary and sufficient for the existence of the inverse of $\mathbf{U}^T\mathbf{U}$, and in consequence — for the existence and uniqueness of solution of the normal equations, based on which the ordinary least squares (OLS) estimator is obtained. In other words, the OLS estimator requires that matrix $\mathbf{U}^T\mathbf{U}$ is non-singular, which is not the case if the matrix of explanatory variables ($\mathbf{U}$) has not full column rank. If matrix $\mathbf{U}^T\mathbf{U}$ is singular, then the normal equations have an infinite number of solutions. In this case (Popławski and Kaczmarczyk 2016) have proposed to select one of them, which is based on the Moore-Penrose pseudoinverse for $\mathbf{U}^T\mathbf{U}$ They have called this procedure “the UEK method” or “the UEK formula.” This formula has been applied by Kawa and Kaczmarczyk (2012), and by Popławski and Kaczmarczyk (2013, 2016). In this short note, we show that the estimator of linear regression coefficients based on the Moore-Penrose pseudoinverse is biased and its bias depends on the unknown regression coefficients. Moreover, any prediction based on Moore-Penrose pseudoinverse is unreliable because of unknown bias.

1 The UEK method within the regression model

Let us formalize the UEK method applied by Popławski and Kaczmarczyk (2016). We assume that an $n \times 1$ vector of observations $\mathbf{K}$ on the dependent variable (also called response variable) satisfies the following equation:

$$\mathbf{K} = \mathbf{UE} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{U}$ is an $n \times m$ matrix of explanatory variables (also called independent variables), $\mathbf{E}$ is an $m \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ is an $n \times 1$ vector of random variables. Moreover, we assume that elements of $\boldsymbol{\varepsilon}$ are uncorrelated with each other, each

* Publication was financed from the funds granted to the Faculty of Finance and Law at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

with mean zero and common variance $\sigma^2 > 0$, that is $E(\boldsymbol{\varepsilon}) = 0$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is an $n \times n$ unit matrix.

It is well known that if $\mathbf{U}$ is a nonstochastic matrix with rank m, then the ordinary least squares estimator, $\hat{\mathbf{E}}$, is the best linear unbiased estimator of $\mathbf{E}$ (Goldberger 1964):

$$\hat{\mathbf{E}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{K}. \tag{2}$$

When the matrix $\mathbf{U}^T\mathbf{U}$ is singular (this occurs, for example, when the number of observations is smaller than the number of regression coefficients), the ordinary least squares estimator, given by the formula (2), does not exist. In this case Popławski and Kaczmarczyk (2016) have proposed to use the Moore-Penrose pseudoinverse of $\mathbf{U}$, and they obtained

$$\hat{\mathbf{E}}^+ = \mathbf{U}^+\mathbf{K}, \tag{3}$$

where $\mathbf{U}^+$ is the Moore-Penrose pseudoinverse of matrix $\mathbf{U}$. The expression $\hat{\mathbf{E}}^+$ is the solution with respect to $\mathbf{E}$ of the linear system of equations (so-called the normal equations) $\mathbf{U}^T\mathbf{U}\mathbf{E} = \mathbf{U}^T\mathbf{K}$.[1] Moreover, $\|\hat{\mathbf{E}}^+\|$[2] attains its minimum value over the solution set of the normal equations, $S = \{\mathbf{E}\colon \mathbf{U}^T\mathbf{U}\mathbf{E} = \mathbf{U}^T\mathbf{K}\}$ (the solution which attains its minimum value over the set S is unique) (Harville 2008, 512). For any matrix $\mathbf{U}$: $\mathbf{U}^+ = \mathbf{U}^T(\mathbf{U}\mathbf{U}^T)^+ = (\mathbf{U}^T\mathbf{U})^+\mathbf{U}^T$ (Harville 2008, 510), thus $\hat{\mathbf{E}}^+ = (\mathbf{U}^T\mathbf{U})^+\mathbf{U}^T\mathbf{K}$. Consequently, when matrix $\mathbf{U}^T\mathbf{U}$ is nonsingular $\hat{\mathbf{E}}^+ = \hat{\mathbf{E}}$ (since $(\mathbf{U}^T\mathbf{U})^+ = (\mathbf{U}^T\mathbf{U})^{-1}$).

Let us assume that matrix $\mathbf{U}$ is nonstochastic (deterministic). Thus, when matrix $\mathbf{U}^T\mathbf{U}$ is singular (for whatever reason), it is easy to conclude that $\hat{\mathbf{E}}^+$ is not an unbiased estimator. In fact

$$E(\hat{\mathbf{E}}^+) = E(\mathbf{U}^+\mathbf{K}) = E\left(\mathbf{U}^+(\mathbf{U}\mathbf{E} + \boldsymbol{\varepsilon})\right) = \mathbf{U}^+\mathbf{U}\mathbf{E} \tag{4}$$

and the bias is equal to

$$E(\hat{\mathbf{E}}^+) - \mathbf{E} = (\mathbf{U}^+\mathbf{U} - \mathbf{I})\mathbf{E}. \tag{5}$$

It is also simple to obtain the covariance matrix of $\hat{\mathbf{E}}^+$, which is equal to

$$V(\hat{\mathbf{E}}^+) = E\left(\left[\hat{\mathbf{E}}^+ - E(\hat{\mathbf{E}}^+)\right]\left[\hat{\mathbf{E}}^+ - E(\hat{\mathbf{E}}^+)\right]^T\right) = E\left(\mathbf{U}^+\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T(\mathbf{U}^+)^T\right) = \sigma^2\mathbf{U}^+(\mathbf{U}^+)^T = \sigma^2(\mathbf{U}^T\mathbf{U})^+, \tag{6}$$

since $\hat{\mathbf{E}}^+ - E(\hat{\mathbf{E}}^+) = \mathbf{U}^+\mathbf{K} - \mathbf{U}^+\mathbf{U}\mathbf{E} = \mathbf{U}^+(\mathbf{U}\mathbf{E} + \boldsymbol{\varepsilon}) - \mathbf{U}^+\mathbf{U}\mathbf{E} = \mathbf{U}^+\boldsymbol{\varepsilon}$.

It is easy to conclude that

$$s^2 = \frac{(\mathbf{K} - \mathbf{U}\hat{\mathbf{E}}^+)^T(\mathbf{K} - \mathbf{U}\hat{\mathbf{E}}^+)}{n - tr\left((\mathbf{U}^T\mathbf{U})^+\mathbf{U}^T\mathbf{U}\right)}, \tag{7}$$
[3]

is an unbiased estimator $\sigma^2 = var(\varepsilon_i)$. Of course, in the case of a perfect fit (when $\mathbf{K}$ and $\mathbf{U}\hat{\mathbf{E}}^+$ coincide), the common variance of each random disturbance ε_i for $i = 1, 2, \ldots, n$ cannot be evaluated.

To summarize, when matrix $\mathbf{U}^T\mathbf{U}$ is singular, $\hat{\mathbf{E}}^+$ is a biased estimator of $\mathbf{E}$, and its bias depends on the unknown vector of parameters, $\mathbf{E}$. Moreover, in that case, information from outside the sample must be added to the sample information in order to estimate all regression coefficients (Zellner 1996, 75). It is worth to note that if matrix $\mathbf{U}^T\mathbf{U}$ is nonsingular, then the OLS estimator based on the Moore-Penrose pseudoinverse is an unbiased estimator of $\mathbf{E}$, because, as it was mentioned above, $\hat{\mathbf{E}}^+ = \hat{\mathbf{E}}$.

2 Prediction With MP-pseudoinverse

One of the main purposes of estimating the vector of parameter in equation (1) is to make predictions of the “future” value of $\mathbf{K}$ associated with some values of $\mathbf{U}$ not observed in the sample. Suppose that the value of the explanatory variable vector is $\tilde{\mathbf{U}}$. It may be a newly observed value $(n + 1)$ or a hypothetical value. We want to predict the value of K_{n+1} conditional on $\tilde{\mathbf{U}}$. Such prediction is

1. If matrix $\mathbf{U}$ is not of rank m, then matrix $\mathbf{U}^T\mathbf{U}$ is singular and the linear system of equations $\mathbf{U}^T\mathbf{U}\mathbf{E} = \mathbf{U}^T\mathbf{K}$ has an infinite number of solutions.
2. $\|\cdot\|$ denotes the (usual) norm of a vector.
3. The symbol $tr(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$.

usually based on the assumption that the linear regression model still holds in the prediction period, that is $K_{n+1} = \tilde{\mathbf{U}}\mathbf{E} + \varepsilon_{n+1}$, where ε_{n+1} represents the stochastic disturbance term in the forecast period, and it is assumed that $E(\varepsilon_{n+1}) = 0$, $var(\varepsilon_{n+1}) = \sigma^2$ and $cov(\varepsilon_{n+1}, \varepsilon_i) = 0$ for $i = 1, 2, \ldots, n$.

The expected value of K_{n+1} is equal to $E(K_{n+1}) = \tilde{\mathbf{U}}\mathbf{E}$. Note that although the estimator $\hat{\mathbf{E}}^+$ is biased, $\mathbf{U}\hat{\mathbf{E}}^+$ is an unbiased estimator of $\mathbf{UE}$. In fact, we can write

$$\begin{aligned}(8)\qquad E(\mathbf{U}\hat{\mathbf{E}}^+) &= E\left(\mathbf{U}(\mathbf{U}^T\mathbf{U})^+\mathbf{U}^T\mathbf{K}\right) = E\left(\mathbf{U}(\mathbf{U}^T\mathbf{U})^+\mathbf{U}^T(\mathbf{UE} + \boldsymbol{\varepsilon})\right) = \\ &= E\left(\mathbf{U}(\mathbf{U}^T\mathbf{U})^+\mathbf{U}^T\mathbf{UE}\right) + E\left(\mathbf{U}(\mathbf{U}^T\mathbf{U})^+\mathbf{U}^T\boldsymbol{\varepsilon}\right) = \\ &= E(\mathbf{UU}^+\mathbf{UE}) + \mathbf{U}(\mathbf{U}^T\mathbf{U})^+\mathbf{U}^T E(\boldsymbol{\varepsilon}) = \\ &= \mathbf{UE},\end{aligned}$$

since $\mathbf{UU}^+\mathbf{U} = \mathbf{U}$. Unfortunately, this property of $\hat{\mathbf{E}}^+$ cannot be applied to prediction problems for $\tilde{\mathbf{U}} \neq \mathbf{w}^T\mathbf{U}$, where $\mathbf{w}$ is an $n \times 1$ vector of known elements. If one assumes that the point predictor is $\tilde{K} = \tilde{\mathbf{U}}\hat{\mathbf{E}}^+$, analogous to that in the linear regression models, then one obtains a biased predictor of $\tilde{\mathbf{U}}\mathbf{E}$. The expected value of the discrepancy between the forecast and actual values (between $\tilde{K}$ and K_{n+1}, respectively) is

$$(9)\qquad E(\tilde{K} - K_{n+1}) = E(\tilde{\mathbf{U}}\hat{\mathbf{E}}^+ - \tilde{\mathbf{U}}\mathbf{E} - \varepsilon_{n+1}) = \tilde{\mathbf{U}}E(\hat{\mathbf{E}}^+ - \mathbf{E}) = \tilde{\mathbf{U}}(\mathbf{U}^+\mathbf{U} - \mathbf{I})\mathbf{E}.$$

Again, if $\mathbf{U}^+\mathbf{U} \neq \mathbf{I}$, the expected discrepancy between the forecast and actual values depends on the unknown vector $\mathbf{E}$. In consequence, the prediction based on $\hat{\mathbf{E}}^+$ (which then is different from $\hat{\mathbf{E}}$) is unreliable, because it is biased. From the definition of $\hat{\mathbf{E}}^+$ and assumption that the random disturbances are uncorrelated we can obtain the variance of the prediction error: $var(\tilde{K} - K_{n+1}) = \sigma^2\left[\tilde{\mathbf{U}}(\mathbf{U}^T\mathbf{U})^+\tilde{\mathbf{U}}^T + \mathbf{I}_m\right]$. Of course, if $(\mathbf{U}^T\mathbf{U})^+ = (\mathbf{U}^T\mathbf{U})^{-1}$ the variance of the prediction error is equal to that obtained using the OLS estimator. However, in the case of a perfect fit, as in Popławski and Kaczmarczyk (2016), the variance of ε_i, σ^2, cannot be evaluated.

3 Numerical illustration

Let us consider the following example to illustrate consequences of the application of the UEK formula. A dependent variable is assumed to be generated by $\mathbf{K} = \mathbf{UE} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ has the normal distribution with mean 0 and variance 1. Moreover: $\mathbf{U} = [2\ \ 7\ \ 5]$, $\mathbf{E} = [2\ \ 3\ \ 5]^T$, $\boldsymbol{\varepsilon} = -0{,}1$ (the value of the observation drawn from the normal distribution), and then $\mathbf{K} = [49{,}9]$.[4] In this case the vector of parameters, $\mathbf{E}$, is known, but suppose that we want to estimate the value of $\mathbf{E}$ using observations of $\mathbf{U}$ and $\mathbf{K}$. First of all, in our example matrix $\mathbf{U}^T\mathbf{U}$ is singular as in Popławski and Kaczmarczyk (2016). The estimate of $\mathbf{E}$ with the use of MP-pseudoinverse is $\hat{\mathbf{E}}^+ \approx [1{,}2795\ \ 4{,}4782\ \ 3{,}1987]^T$; this in turn implies that $\hat{\mathbf{K}}^+ = \mathbf{U}\hat{\mathbf{E}}^+ = [49{,}9]$.

Let us now consider $\tilde{\mathbf{U}} = [4\ \ 12\ \ 3]$ and $K_2 = \tilde{\mathbf{U}}\mathbf{E} + 0 = [59]$. Using the UEK formula for prediction K_2 we get the following result: $\tilde{\mathbf{K}} = \tilde{\mathbf{U}}\hat{\mathbf{E}}^+ \approx [68{,}4526]$. Hence, the forecast error is equal to $\tilde{\mathbf{K}} - K_2 \approx 9{,}4526$. Suppose now that the units of measurement of explanatory variables are changed. Change of the scale of the variables results in a corresponding change in the scale of the coefficients. The first component of $\mathbf{U}$ is divided by 1000, and the first component of $\mathbf{E}$ is just multiplied by 1000. In turn, the second component of $\mathbf{U}$ is divided by 100, and the second component of $\mathbf{E}$ is multiplied by 100. We get exactly the same value for $\mathbf{K}$. Now, in our example $\mathbf{U} = [0{,}002\ \ 0{,}07\ \ 5]$, thus $\mathbf{E} = [2000\ \ 300\ \ 5]^T$, and $\mathbf{K} = [49{,}9]$. The estimate of $\mathbf{E}$ using MP-pseudoinverse matrix is now as follows: $\hat{\mathbf{E}}^+ \approx [0{,}004\ \ 0{,}1397\ \ 9{,}978]^T$, which in turn implies $\hat{\mathbf{K}}^+ = \mathbf{U}\hat{\mathbf{E}}^+ = [49{,}9]$. Similarly, the units of measurement of $\tilde{\mathbf{U}}$ are changed: $\tilde{\mathbf{U}} = [0{,}004\ \ 0{,}12\ \ 3]$. Also, in that case $K_2 = \tilde{\mathbf{U}}\mathbf{E} = [59]$. Using the UEK formula for prediction of K_2 yields the following result: $\tilde{\mathbf{K}} = \tilde{\mathbf{U}}\hat{\mathbf{E}}^+ \approx [29{,}9509]$. Hence, the error of the prediction is equal to $-29{,}0491$. As we can see, forecasting performance can depend on the units of measurement of explanatory variables.

4. [In the journal European practice of number notation is followed — for example, 36 333,33 (European style) = 36 333.33 (Canadian style) = 36,333.33 (US and British style). — Ed.]

The numerical illustration also shows that, when the matrix $\mathbf{U}^T\mathbf{U}$ is singular, MP-pseudoinverse estimates of the vector $\mathbf{E}$ cannot be interpreted as OLS estimates—i.e., as the estimate of ceteris paribus effects of explanatory variables on the expected value of the dependent variable. In other words, the components of $\hat{\mathbf{E}}^+$ do not have useful interpretations, and therefore the UEK method is useless.

Conclusions and Remarks

When the matrix $\mathbf{U}^T\mathbf{U}$ is singular, the estimator based on MP-pseudoinverse is inappropriate for at least two reasons. First, the estimator $\hat{\mathbf{E}}^+$ is biased, and its bias depends on the unknown true value of $\mathbf{E}$. Second, the point prediction based on this estimator can depend on the units of measurement of explanatory variables. Moreover, when the estimator based on MP-pseudoinverse provides a perfect fit to the data, the estimator of error variance—given by equation (7)—is equal to zero, although the random vector $\boldsymbol{\varepsilon}$ is non-degenerate (not equal to 0 with probability of 1) and estimates of the parameter vector are not equal to its real value. In consequence, the ideal fitness in the sample does not lead to good forecasting performance.

When in the linear regression model the matrix $\mathbf{U}^T\mathbf{U}$ is singular, an unbiased estimator of all coefficients (parameters) cannot be obtained based on information known only from the data. One of possible solutions of this problem is to use additional information which comes from outside of the sample—e.g., an individual investigator's information (subjective but not unfounded beliefs about what the true value of $\mathbf{E}$ is likely to be, prior to looking at the data) as in the Bayesian inference (Zellner 1996).

References

Goldberger, A.S. 1964. *Econometric Theory, A Wiley Publication in Applied Statistics*. New York: J. Wiley.

Harville, D.A. 2008. *Matrix Algebra from a Statistician's Perspective*. 2nd ed. New York: Springer.

Kawa, B., and B. Kaczmarczyk. 2012. "Wskaźnikowy i modelowy sposób oceny sytuacji finansowej Gminy Kraków na koniec 2011 roku." *Zeszyty Naukowe — Polskie Towarzystwo Ekonomiczne* (13): 199–213.

Popławski, Ł., and B. Kaczmarczyk. 2013. "Możliwości zastosowania metody UEK dla gmin województwa świętokrzyskiego — na przykładzie danych za 2006 rok." *Journal of Agribusiness and Rural Development* 4 (30): 207–218.

———. 2016. "Wartość przestrzeni publicznej jako kategoria dobra wspólnego na przykładzie województwa małopolskiego." *Barometr Regionalny. Analizy i Prognozy* 14 (4): 163–170.

Zellner, A. 1996. *An Introduction to Bayesian Inference in Econometrics*. Reprint. New York – Chichester: John Wiley.