FACTA UNIVERSITATIS (NIŠ)
Ser. Math. Inform. Vol. 36, No 5 (2021), 1019- 1032
https://doi.org/10.22190/FUMI210406074P
Original Scientific Paper

A NEW UNBIASED ESTIMATOR OF A MULITPLE LINEAR REGRESSION MODEL OF THE CAPM IN CASE OF MULTICOLLINEARITY

Dimitrios Pappas[1] **and Konstantinos Bisiotis**[2]

[1]**Department of Economics, National and Kapodistrian University of Athens, 1 Sofokleous Str, 10559 Athens, Greece**
[2]**Department of Statistics, Athens University of Economics and Business, 76 Patission Str, 10434 Athens, Greece**

Abstract. In this work we propose an unbiased estimator for a multiple linear regression model of the CAPM in the presence of multicollinearity in the explanatory variables. Multicollinearity is a common problem in empirical Econometrics. The existing methods so far have not dealt with cases of perfect multicollinearity. This new optimization method that belongs to the class of unbiased estimators is suitable for cases with strong or perfect multicollinearity, imposes restrictions of the minimizing matrix and produces small standard errors for the estimated parameters. First, we presented the theoretical background of our approach and next we derive an expression for the covariance matrix of estimated coefficients. As an example, we have estimated the basic linear regression model on Apple Inc expected stock returns and we have examined multivariate extensions of this model in the special case of multicollinearity using the proposed method.
Keywords CAPM, Data multicollinearity, Moore-Penrose inverse, MDLUE, Multiple linear regression.

1. Introduction

Multicollinearity is a problem that occurs when we estimate linear or generalized linear models and the independent variables in the regression model are highly

Received April 06, 2021. accepted May 13, 2021.
Communicated by Predrag Stanimirović
Corresponding Author: Dimitrios Pappas, Department of Economics, National and Kapodistrian University of Athens, 1 Sofokleous Str, 10559 Athens, Greece | E-mail: dipappas@econ.uoa.gr
2010 *Mathematics Subject Classification.* Primary 15A09; Secondary 15A10,62P05,62J05

correlated to each other. This situation has as a result unstable estimates of the regression coefficients. The coefficients of the model become very sensitive to small changes in the model. Also, multicollinearity reduces the precision of the estimate coefficients ([1]). In our work we concentrate in cases where we have strong or perfect collinearity between explanatory variables, this means for values of correlation greater than 0.9.

The Capital Asset Pricing Model (CAPM) was introduced by [24] and [15] based on the the work of Markowitz on modern portfolio theory ([17], [16]). The CAPM describes the relationship between expected return and systematic risk for stocks. It is also widely used for pricing risky securities and generating expected returns for assets given the risk of those assets and the cost of the capital. The formula for calculating the CAPM is

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f) \tag{1.1}$$

or else

$$R_i = \alpha_i + \beta_i R_m + \epsilon_i \tag{1.2}$$

where $E(R_i)$ is the expected return of the investment, R_f is the risk-free rate, β_i is the systematic risk given by

$$\beta_i = \frac{cov(R_i, R_m)}{\sigma^2(R_m)}, \tag{1.3}$$

and $E(R_m)$ is the expected return of market. The quantity $E(R_m) - R_f$ is the market risk premium. In equation (2), R_i is the return of asset i, α_i is a constant term, R_m refers to the return of the market and ϵ_i is an error term. The most commonly used estimation method for the CAPM is the ordinary least squares (OLS)([7]).

In [22] the authors derive a multiple linear regression model of the CAPM by examining various explanatory variables that can be added to the basic CAPM for the expected returns on Apple Inc.. Their model, in addition to the market return (S&P500 returns), includes as explanatory variables the average spread and its interaction term with the market return. The average spread is the difference between the daily highest ask price and the lowest bid price divided by the price of the stock at the end of the day.

Various methods have been proposed for dealing with multicollinearity, such as deleting parameters, principal components regression, ridge regression estimation, maximum entropy estimators and shrinkage estimators (e.g. see [23], [13], [20]). The work of [25] introduces the generalized maximum entropy (GME) approach in order to estimate the quantile regression model for CAPM. The OLS method is very sensitive to extreme observations and [7] propose a fuzzy regression method which takes into account possible extreme observations and needs less assumptions from the OLS method. The method that we apply in our work belongs to the class of unbiased estimators, such as the minimum dispersion method (see for example [23]) in contrast to the ridge regression which is a biased estimation method ([26]).

The aim of the current work is to find a purpose for a new unbiased estimator for a multiple regression model of CAPM in case of strong multicollinearity using Linear Algebra techniques. [12] compares through a simulation study various biased and unbiased alternative estimators to the OLS estimator in the case of collinearity. In regression analysis, least squares estimations assume that explanatory variables are not correlated with each other. In the presence of multicollinearity, inference about the coefficients of regression can be difficult due to instability in the coefficients.

In this work we will apply a solution to a minimization problem for a matrix-valued function under linear constraints, in the case of a singular matrix. The theoretical framework of this method is not new and it is based on the paper [19]. Here we adapt and extend this framework by deriving an expression for the covariance matrix of estimated coefficients. Our method differs from others on the restriction of the minimizing matrix to the range of the corresponding quadratic function. In the case of singular positive matrices, many matrix valued functions are investigated using a partial ordering. Using matrix analysis results, we propose this additional relation as a constraint, by taking advantage of the canonical form related to this class of matrices. Moreover, the singularity of the matrix implies the use of the Moore-Penrose inverse matrix, giving us a unique minimal norm solution to the problem.

This paper is organized as follows: In section 2 we present the data and the multiple regression model and define the special case of multicollinearity. Section 3 introduces the proposed estimation technique in case of multicollinearity and we estimate the covariance matrix of estimated coefficients. Section 4 presents the estimation results for the simple CAPM and the multiple regression models of the CAPM. In addition, the proposed method is tested and compared against another known methods in terms of the standard errors of the estimated coefficients. Finally, concluding remarks appear in section 5.

2. Data and the Multiple Regression Model

The multiple linear regression model of the CAPM that we use has the following form:

$$E(R_A) = \alpha + \beta_1 R_1 + \beta_2 R_2 + \beta_3 R_3 + \beta_4 E(R_m) + \varepsilon_t \tag{2.1}$$

where $E(R_A)$ are the expected daily returns of the asset and $E(R_m)$ are the expected daily market returns. We remind that the OLS estimator for coefficients of the multiple regression $Y = \alpha + \beta X + \varepsilon$ is given by

$$\hat{\beta} = (X^{'}X)^{-1}X^{'}Y \tag{2.2}$$

In the presence of collinearity the quantity $(X^{'}X)$ is not invertible and the estimation of the variance of the coefficient estimates

$$Var(\hat{\beta}) = \sigma^2(X^{'}X)^{-1} \tag{2.3}$$

is problematic. If the quantity $(X'X)$ is not exactly singular but very close to be non-invertible, then the variance will be large. Moreover, if there is not an exact linear relationship among the predictor variables but they are close to each other, then the matrix $(X'X)$ will be invertible but the inverse matrix will have very large entries, due to the very small value of the determinant. If some of the variables are highly correlated then the matrix $(X'X)$ becomes non-orthogonal and as a result the inversion is unstable. As for the OLS solution of the model, the analysis and interpretation of each of the explanatory variables is difficult (see e.g. [13]). Multicollinearity has several effects in a regression model. For example the high variance of coefficients may reduce the precision of the estimation or the estimated coefficients to have the wrong sign. Also, the estimates of the coefficients may be sensitive to a particular set of the data. In our paper we try to overcome the problem of multicollinearity and find an unbiased solution. Since the problem with multicollinearity in multiple regression has infinite solutions, we will choose among them the minimal norm least squares solution, making use of the Moore-Penrose inverse.

For the basic CAPM model we use daily data of Apple Inc. stock returns (APPLE) and the market returns are the S&P500 daily returns (SP500). In the multiple linear regression model, the observed values are the daily expected stock returns of Apple Inc. (APPLE). The explanatory variables are the S&P500 daily returns (SP500), the opening stock price (OPENP), the semi-sum of opening and lower stock price (OPENLOW) of each day and the closing price (CLOSEP). The data are from January 1, 2007 until June 6, 2014.

Table 1 presents some descriptive statistics for our data. The skewness of the data show that they are approximately symmetric. The distributions of the time series Apple Inc. returns and market S&P500 returns have positive excess kyrtosis and are leptokurtic. Also the distributions of the opening stock price, the semi-sum of opening and lower stock price of each day and the closing price are having thinner tails than those of the normal distribution. Table 2 presents the correlation coef-

	APPLE	SP500	OPENP	OPENLOW	CLOSEP
Maximum	0.130	0.11	134.46	132.555	133
Minimum	-0.197	-0.095	11.341	11.438	11.171
Mean	0.0001	0.0002	56.77	56.753	56.770
Median	0.001	0.0007	51.031	51.009	56.736
St. Deviation	0.021	0.014	123.119	35.019	35.007
Skewness	-0.448	-0.315	0.482	0.481	0.481
Kurtosis	9.725	12.511	2.059	2.057	2.057
Range	0.328	0.204	123.119	121.117	121.829

Table 2.1: Descriptive Statistics for Data. The data are the Apple Inc. stock returns (APPLE), the S&P 500 daily returns (SP500), the opening stock price (OPENP), the semi-sum of opening and lower stock price (OPEN_LOW) of each day and the closing price (CLOSEP). The data are from January 1, 2007 until June 6, 2014.

ficients of the explanatory variables in the multiple regression model. The results indicate that there is a strong positive relationship between the explanatory variables except for the market S&P500 returns. For the detection of multicollinearity

	SP500	OPENP	OPENLOW	CLOSEP
SP500	1	0.014	0.018	0.022
OPEN	0.014	1	0.999	0.999
OPENLOW	0.018	0.999	1	0.999
CLOSEP	0.022	0.999	0.999	1

Table 2.2: Correlation coefficients for the explanatory variables: S&P 500 daily returns (SP500), opening stock price (OPENP), semi-sum of opening and lower stock price (OPENLOW) of each day, closing stock price (CLOSEP). The data are from January 1, 2007 until June 6, 2014.

in regression models there are various diagnostic techniques.

In the following part, we will briefly present two of the basic diagnostic tools for collinearity. The first is the Variance Inflation Factor (VIF) which measures the inflation of the parameter estimates being computed for all the explanatory variables in the regression model ([2]). The VIF is given by

$$VIF = \frac{1}{1-\mathcal{R}_i^2}, i = 1, \ldots, p \tag{2.4}$$

where p is the number of explanatory variables and $\mathcal{R}^2$ is the squared multiple correlation coefficient. The VIF has a lower bound value equal to 1 but no upper bound. Higher values signify that it is difficult to define accurately the contribution of the predictor variable to a regression model. Usually values higher than 10 indicate collinearity. Table 2.3 presents the variance inflation factor (VIF) and condition index results of the explanatory variables for the multiple regression model. From the results it is obvious that there exists high collinearity between the opening stock price, the semi-sum of opening and lower stock price of each day and the closing price. Another measure of collinearity is the condition index. The condition index (CI) is the square root of the ratio of each eigenvalue λ to the smallest eigenvalue of X ([6]) and indicates how close the underlying matrix is to a singular matrix. The condition index is defined as

$$C_k = \sqrt{\frac{\lambda}{\lambda_{min}}} \tag{2.5}$$

where λ_{min} is the smallest eigenvalue value of $X^{'}X$. Values between 10 and 30 are a sign of multicollinearity and multicollinearity occurs when the value of the condition indices are greater than 30 ([8]). The results from table 2.3 confirm the existence of collinearity between the explanatory variables except the S&P500 variable.

Table 2.3: Results for the Variance Inflation Factor (VIF) and condition index for the explanatory variables: S&P 500 daily returns (SP500), opening stock price (OPENP), semi-sum of opening and lower stock price (OPENLOW) of each day, closing price (CLOSEP). The data are from January 1, 2007 until June 6, 2014.

Variable	VIF	Cond. Index
S&P 500	0 1.7323	1
OPEN	5.6830e+14	17.330
OPENLOW	2.2710e+15	182.4603
CLOSEP	5.6740e+14	7.8015e+15

3. Constrained matrix optimization

In this section, we will briefly present the basic concepts of the theoretical background of our matrix constrained optimization (MCO) method, for more information see [19]. As discussed previously, the collinearity of the data makes the quantity $(X^{'}X)$ not invertible (or very close to singular) and the estimation of the variance of the coefficient estimates

$$Var(\hat{\beta}) = \sigma^2(X^{'}X)^{-1}$$

is problematic. So, a way to tackle the problem is to use a constrained matrix optimization method, making use of the Moore-Penrose inverse matrix.

Suppose that $A \in \mathcal{R}^{n\times n}$ is a square matrix with $\mathcal{N}(A)$ and $\mathcal{R}(A)$ its kernel and its range respectively. Also we denote as A' the transpose of the square matrix A. The generalized inverse, also known as the Moore-Penrose inverse of a matrix A is the unique matrix $A^{\dagger}$ satisfying the following four Penrose conditions:

$$AA^{\dagger} = (AA^{\dagger})', \qquad A^{\dagger}A = (A^{\dagger}A)', \qquad AA^{\dagger}A = A, \qquad A^{\dagger}AA^{\dagger} = A^{\dagger}. \tag{3.1}$$

It is easy to see that $AA^{\dagger}$ is the orthogonal projection of $\mathcal{R}^n$ onto $\mathcal{R}(A)$, denoted by P_A, and that $A^{\dagger}A$ is the orthogonal projection of $\mathcal{R}^n$ onto $\mathcal{R}(A^{'})$ noted by $P_{A'}$. It is also well known that $\mathcal{R}(A^{\dagger}) = \mathcal{R}(A^{'})$. For more on the Moore-Penrose inverse, see e.g. [4], [5].
The Moore-Penrose inverse also satisfies the following inequality ([21])

$$\| AA^{\dagger}B - B \|_2 \leq \| AX - B \|_2 \tag{3.2}$$

for all X.

We remind that given a matrix $R \in \mathcal{R}^{M\times M}$, minimizing $W'RW$ with

$$W \in \mathcal{R}^{M\times m}$$

means finding a matrix $\hat{W} \in \mathcal{R}^{M\times m}$ such that the $m \times m$ matrix $(W'RW - \hat{W}'R\hat{W})$ is positive semidefinite for all $W \in \mathcal{R}^{M\times m}$. (The Löwner partial ordering for hermitian nonnegative definite matrices, defined as: $A \geq B$, if $A - B$ is positive semidefinite). See e.g. [3], [10].

3.1. Matrix Optimization and Linear Regression

Next we assume that t $R \in \mathcal{R}^{M\times M}$ is a positive semidefinite symmetric matrix. The main problem is the minimization of $W'RW, \quad W \in \mathcal{R}^{M\times m}$ under the Löwner ordering, when W satisfies a set of linear constraints :

$$S = \{W \in \mathcal{R}^{M\times m} : C'W = F\}$$

with $C \in \mathcal{R}^{M\times n}, \quad F \in \mathcal{R}^{n\times m}$. As a result, we will find a matrix $\hat{W}$ such that $W'RW \geq \hat{W}'R\hat{W}$ for all $W \in S$.
In [9] and [14] where a similar problem is treated the matrix R is assumed to be positive definite. In our work the matrix R is positive semidefinite (therefore singular). The difference in our method is that the matrix W will also satisfy the relation $\mathcal{R}(W) \subseteq \mathcal{R}(R)$ in order to overcome the singularity of R.
In our case the positive semidefinite matrix R is singular, $\mathcal{N}(R) \neq \{0\}$ and therefore we have that $W'RW = 0$ for all matrices W of appropriate dimensions belonging to the set $\mathcal{Z} = \{W : RW = 0\}$ and so, the problem

$$minimize \quad W'RW, W \in S \tag{3.3}$$

has many solutions when $\mathcal{S} \cap \mathcal{Z} \neq \emptyset$.
In other words, since the matrix R is symmetric, we have that $\mathcal{R}(R) = \mathcal{R}(R^{\dagger})$ and therefore we are looking for the minimum of $W'RW$ under the constraints $C'W = F$ and $\mathcal{R}(W) \subseteq \mathcal{R}(R)$.
From Theorem 1 in [19] we have that the minimizing problem in eq. 10 has the unique solution $\hat{W} = R^{\dagger}C[C'R^{\dagger}C]^{\dagger}F$. In the case now that S is empty then the constraint must be replaced by the equation $C'W = F_1 = P_{\mathcal{R}(C'R^{\dagger}C)}F$. The following Corollary is a consequence of the previous result:

Corollary 3.1. *Let $R \in \mathcal{R}^{M\times M}$ a positive semidefinite symmetric matrix, the matrices $W \in \mathcal{R}^{M\times m}, \quad C \in \mathcal{R}^{M\times n}, \quad F \in \mathcal{R}^{n\times m}$ with $m < M, n < M$, and the equation $C'W = F$. The problem:*

$$minimize \quad W'RW, \quad W \in \hat{S}$$

where $\hat{S} = \{W : C'W = P_{\mathcal{R}(C'R^{\dagger}C)}F, such \quad that \quad \mathcal{R}(W) \subseteq \mathcal{R}(R)\}$ has a unique solution among the generalized constrained solutions which is

$$\hat{W} = R^{\dagger}C[C'R^{\dagger}C]^{\dagger}F$$

In many statistical applications as in our case, the matrix R is equal to CC'. In this case, we have the following proposition:

Proposition 3.1. *Let $R \in \mathcal{R}^{M\times M}$ to be a positive semidefinite symmetric matrix and the matrices $W \in \mathcal{R}^{M\times m}, \quad C \in \mathcal{R}^{M\times n}, \quad F \in \mathcal{R}^{n\times m}$ with $m < M, n < M$. Also, suppose that $C'W = F$ and the set $S = \{W : C'W = F, such \quad that \quad \mathcal{R}(W) \subseteq \mathcal{R}(R)\}$ is not empty. Taking as $R = CC'$ then the problem:*

$$minimize \quad W'RW, \quad W \in S$$

has the unique solution

$$\hat{W} = (C')^{\dagger} F$$

Consider a simple linear model

$$y = C\beta + \epsilon \tag{3.4}$$

where $y \in \mathcal{R}^{m\times 1}$ is a vector of observed data, $C \in \mathcal{R}^{m\times p}$ the matrix of p observed covariates, $\beta \in \mathcal{R}^{p\times 1}$ vector of parameters to be estimated and $\epsilon \in \mathcal{R}^{m\times 1}$ the noise vector. When $\text{Rank}(C) = p$, and $C'C$ is nonsingular the inverse $(C'C)^{-1}$ can be computed. Then the Best Linear Unbiased Estimator (BLUE) for β is defined as

$$\hat{\beta} = (C'C)^{-1}C'y = W'y \tag{3.5}$$

In the case when the matrix R is singular, then Theorem 1 in [19] can be applied to the problem of computing the Best Linear Unbiased Estimator in linear models. In this specific case, the size of the matrices are:

$$R \in \mathcal{R}^{M\times M}, \quad W \in \mathcal{R}^{M\times m}, \quad C \in \mathcal{R}^{m\times p}, \quad I \in \mathcal{R}^{m\times m}, m < M$$

In the case when the matrix C is of full rank, then W' is the unique left inverse of C, and therefore, Theorem 1 can be applied to find the optimum matrix $\hat{W}$.

There might be cases, however, where there is a deficiency in rank of the design matrix C. That means that $\text{Rank}(C) = r < p < m$ and thus $C'C$ is singular. When $m < p$ ordinary least squares method (OLS) cannot be used to estimate β in linear model (7). Some methods to overcome this problem have been suggested based on maximum entropy estimation ([11]) or penalized regression ([18]). We will denote any generalized inverse of $C'C$ as $(C'C)^{-}$. In such a case there might be several values of β that lead to same values of $C\beta$. In addition, the estimator is not unbiased anymore, since the condition $W'C = I$ does not necessarily hold. However, let $L'\beta$ be linear functions of β such that $\mathcal{R}(\text{L}) \subset \mathcal{R}(C')$ implying $L = C'A$ for some A. Then if $(C'C)^{-}$ is any generalized inverse of $C'C$ and

$$\hat{\beta} = (C'C)^{-}C'y \tag{3.6}$$

it can be shown ([23] p. 30) that $L'\hat{\beta}$ is the *minimum dispersion unbiased estimator* (MDLUE) of $L'\beta$ with dispersion matrix $\sigma^2 L'(C'C)^{-}L$. In the case when $Rank(C) < p$, then C does not have a left inverse, and therefore the constraint must be slightly modified, as said in Theorem 1 ([19]), since $C'W = I$ does not hold.

Following all the above and using Theorem 1 we will minimize $W'RW$ under the constraint

$$C'W = P_{\mathcal{R}(C'R^{\dagger}C)}, \text{ with } \mathcal{R}(W) \subseteq \mathcal{R}(R)$$

where $P_{\mathcal{R}(C'R^{\dagger}C)}$ is the orthogonal projection on the range of $C'R^{\dagger}C$.

So, from the above discussion and Theorem 1 we have the following Proposition:

Proposition 3.2. *Consider a simple linear model*

$$y = C\beta + \epsilon \tag{3.7}$$

and let $C \in \mathcal{R}^{M\times m}$ *the matrix of* m *observed covariates with* $m < M$, $\beta \in \mathcal{R}^{m\times 1}$ *the vector of parameters to be estimated ,* $R \in \mathcal{R}^{M\times M}$ *a positive semidefinite symmetric matrix such that* $R = E(y - C\beta)(y - C\beta)'$ *. Then,*

$$\hat{\beta}_{MCO} = [R^{\dagger}C[C'R^{\dagger}C]^{\dagger}]'y \tag{3.8}$$

gives a solution which is restricted on a particular set defined by the orthogonal projection, thus giving the minimum dispersion unbiased estimator of any linear combination of β_{MCO}*.*

Moreover, in [23] p. 25, a different way of estimating $\hat{\beta}$ is also presented when C is not of full rank:

$$\hat{\beta}_{RMDLUE} = (C'C)^{-}C'y + (I - (C'C)^{-}C'C)w \tag{3.9}$$

where $(C'C)^{-}$ defined as before and w is an arbitrary vector.
The rationale follows from the fact that the empirical predictor (given as $\hat{y} = C\hat{\beta}$) has the same value for all solutions of β that emerge from $C'C\hat{\beta} = C'y$.

In many practical applications estimation of β relies on either formulas (3.9) or (3.6). In the next section we will use Proposition 3.2 and hence the result given by eq.(3.8) in order to solve the multicollinearity problem, finding a unique MDLUE solution among the infinite solutions that this problem admits. Our solution gives a model similar to the one found using eq.(3.9) with differences in the coefficients due to the different choice of the unique solution among the infinite ones.
The variance of all the estimated coefficients using the MCO approach is given by

$$\begin{aligned} V(\hat{\beta}_{MCO}) &= V([R^{\dagger}C[C'R^{\dagger}C]^{\dagger}]'y) \\ &= ([R^{\dagger}C[C'R^{\dagger}C]^{\dagger}]')V(y)([R^{\dagger}C[C'R^{\dagger}C]^{\dagger}]')' \\ &= ([R^{\dagger}C[C'R^{\dagger}C]^{\dagger}]')\sigma^2[R^{\dagger}C[C'R^{\dagger}C]^{\dagger}] \end{aligned}$$

Also we have that, since R is symmetric, so $(R^{\dagger})' = R^{\dagger}$:

$$\begin{aligned} ([R^{\dagger}C[C'R^{\dagger}C]^{\dagger}]')([R^{\dagger}C[C'R^{\dagger}C]^{\dagger}]')' &= ([R^{\dagger}C[C'R^{\dagger}C]^{\dagger}]')(R^{\dagger}C[C'R^{\dagger}C]^{\dagger}) \\ &= ([C'R^{\dagger}C]^{\dagger})'(R^{\dagger}C)'R^{\dagger}C[C'R^{\dagger}C]^{\dagger} \\ &= ([C'R^{\dagger}C]')^{\dagger}C'R^{\dagger}[R^{\dagger}C[C'R^{\dagger}C]^{\dagger} \\ &= [C'R^{\dagger}C]^{\dagger}C'R^{\dagger}R^{\dagger}C[C'R^{\dagger}C]^{\dagger} \end{aligned} \tag{3.10}$$

Denote with $K = C'R^{\dagger}$ then equation (3.10) becomes

$$\begin{aligned} [C'R^{\dagger}C]^{\dagger}C'R^{\dagger}R^{\dagger}C[C'R^{\dagger}C]^{\dagger} &= (KC)^{\dagger}KK'(KC)^{\dagger} \\ &= A^{\dagger}KK'A^{\dagger} \end{aligned} \tag{3.11}$$

where $A = KC$. As a result the variance of the estimated coefficient $\hat{\beta}_{MCO}$ is given by

$$V(\hat{\beta}_{MCO}) = \sigma^2 A^{\dagger} K K' A^{\dagger} \tag{3.12}$$

The standard errors are estimated by taking the square root of the diagonal of $V(\hat{\beta}_{MCO})$.

4. Estimation Results

In this study, we apply our matrix constrained optimization method to two multivariate regression models. The first (Multiple Regression Model I) contains all ofl the explanatory variables we previous mentioned. The second model (Multiple Regression Model II) excludes the variable with the lowest correlation coefficient, the S&P 500 market returns and as a result we have a model with strong correlation between the regressors, almost equal to one. As said above, in order to compare our method, the regression coefficients have been also estimated using the MDLUE method presented in [23]. Table 4.1 reports the results for the basic CAPM for the Apple Inc. stock returns and the S&P500 expected returns as the market returns. The resulting simple CAPM has the following form

$$E(R_{APPLE}) = 0.0008 + 0.9568(SP500)$$

Table 4.1: Capital Asset Pricing Model (CAPM) coefficients. The table presents the value of the coefficients for the Capital Asset Pricing Model with an intercept and one explanatory variable, the S&P 500.

Variable	Estimation
Intercept	0.0008
S&P500	0.9568

Table 4.2 presents the coefficients of the multiple linear regression in case of multicollinearity for the proposed constrained matrix optimization method (MCO, eq. (16)) and the MDLUE proposed by [23] (RMDLUE, eq. (17)) along with their standard errors. The multiple linear regression model under MCO is the following:

$$\begin{aligned} E(R_{APPLE}) = {} & 0.0012 + 0.5671(SP500) - 0.0324(OPENP) \\ & + 0.0005(OPENLOW) + 0.0334(CLOSEP) \end{aligned}$$

We compare the performance of our approach with the RMDLUE method in terms of the standard errors of the estimated regression coefficients. One of the problems of multicollinearity is that affects the standard error of the parameter estimators. For the Multiple Regression Model I the standard errors for the coefficients estimated with the MCO method are smaller in most of the cases than

MCO	Coefficients	Std Error	t value	p-value
Intercept	0.0012	0.0015	0.3257	0.7447
S&P500	0.5671	0.0493	4.1584	< 0.00001
OPEN	-0.0324	0.0015	-11.9939	< 0.00001
OPENLOW	0.0005	2.6597e-05	20.6638	< 0.00001
CLOSEP	0.0334	0.0016	12.3642	< 0.00001
RMDLUE	Coefficients	Std Error	t value	p-value
Intercept	0.0015	0.7277	2.2479	< 0.00001
S&P500	0.6885	4.7211e-06	28.34	< 0.00001
OPEN	-0.1933	0.0082	18.09	< 0.00001
OPENLOW	0.3558	0.0088	-14.64	< 0.00001
CLOSEP	-0.1625	0.0825	24.47	< 0.00001

Table 4.2: Multiple Regression Model I.Parameter estimates for the matrix constrained optimization method (MCO) and the minimum dispersion linear unbiased method (RMDLUE) and their standard errors.

those from the RMDLUE approach. The t-value column represents whether the estimated coefficients of the variables in the multiple regression model are statistically significant. Also, in the table we present the p-values, the probability that the variable in the model is not significant. The reported p-values are low, which means that the variables are statistically significant. The results in table 7 for the Multiple Regression Model I show that the relationship between the S&P 500 market returns and the Apple Inc returns is positive and the value of regression coefficient is 0.5671. This means that an increase in the S&P 500 daily market returns lead to an increase in the Apple Inc returns. The same behavior happens between the Apple Inc returns the semi-sum of opening and lower stock price (OPENLOW) of each day and the closing stock price (CLOSEP). In contrast the relationship between the opening stock price (OPENP) and the Apple Inc returns is negative.

Table 4.3 now presents the coefficients of the multiple linear regression in case of multicollinearity for the constrained matrix optimization method (MCO) and the RMDLUE method if we exclude the variable of the stock market returns. The standard errors for the OPEN, OPENLOW and CLOSEP coefficient are smaller than those estimated with the RMDLUE approach. The results for both regression models indicate that the MCO method could be a good alternative when someone wants to obtain estimates with small standard errors and the variable that appear to have strong collinearity are all of interest. As previously, the reported p-values indicate that the variables are statistically significant.

MCO	Coefficients	Std Error	t value	p-value
Intercept	0.0003	0.0015	0.0489	0.961004
OPEN	-0.0433	0.0014	-12.2014	< 0.00001
OPENLOW	0.0007	3.0788e-05	20.9787	< 0.00001
CLOSEP	0.0488	0.0014	12.6021	< 0.00001
RMDLUE	Coefficients	Std Error	t value	p-value
Intercept	0.0011	6.7648e-04	1.3351	0.182006
OPEN	-0.1624	0.009	-29.91	< 0.00001
OPENLOW	0.2850	0.619e-05	54.12	< 0.00001
CLOSEP	-0.1226	0.0078	22.57	< 0.00001

Table 4.3: Multiple Regression Model II. Parameter estimates for the matrix constrained optimization method (MCO) and the minimum dispersion linear unbiased method (RMDLUE) and their standard errors. The multiple linear regression model excludes the S&P500 factor.

5. Concluding Remarks

In this research, we refer to the multicollinearity issue of a multiple regression problem. Various techniques have been proposed in order to overcome this problem such as ridge regression or delete the factors that are collinear. The matrix constrained optimization method that we proposed is an unbiased estimator that can be applied in situations where exists strong or perfect collinearity and we can not delete any collinear factor because this may affect the interpretation of the model results and the factor is important for the analysis. Also, we obtain an expression for the variance-covariance matrix of the estimated coefficients.
The method is applied in a special case of a multiple linear regression model which is an extension of the Capital Asset Pricing Model (CAPM). The matrix constrained optimization method is implemented in two multiple regression models. The difference between these models is that the first includes an explanatory variable with low correlation which in the second model,this factor is excluded. The results are compared with another unbiased linear estimator, the MDLUE, in terms of standard errors of the estimated parameters. We have mentioned that this technique is appropriate when high levels of correlations exist among the regressors, and there is a need for an unbiased estimator. In this case, the solution of deleting the factors with high collinearity may not be feasible because of the importance of the factors in the regression model.

Acknowledgemnts

The first author acknowledges that this research was conducted during his postdoctoral program entitled "Confronting multicolinearity problems in CAPM- Arbitrage applications using pseudoinverses" at the Department of Economics, National and Kapodistrian University of Athens.

REFERENCES

1. M. P. Allen: *Understanding regression analysis*, Springer Science & Business Media,(2004).
2. A. Bager, M. Roman, M. Algedih, B. Mohammed: *Addressing multicollinearity in regression models: a ridge regression application*,(2017).
3. J. K. Baksalary,F. Pukelsheim: *On the Löwner, minus, and star partial orderings of nonnegative definite matrices and their squares* Linear Algebra and its Applications, **151** (1991) 135-141.
4. A. Ben-Israel,T. N. Greville: *Generalized inverses: theory and applications* (Vol. 15).Springer Science & Business Media,(2003).
5. S. L. Campbell,C. P. Meyer: *Generalized Inverses of Linear Transformations*.Dover. New York,(1991).
6. C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, ... S. Lautenbach: *Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography*, **36** (1) (2013), 27-46.
7. Y. Du,Q. Lu: *Estimate Beta Coefficient of CAPM Based on a Fuzzy Regression with Interactive Coefficients* Journal of Applied Mathematics and Physics, **3** (06) (2015), 664.
8. <. Friendly, E. Kwan: *Where's Waldo? Visualizing collinearity diagnostics* The American Statistician, **63** (1) (2009), 56-65.
9. O. L. Frost: *An algorithm for linearly constrained adaptive array processing*, Proceedings of the IEEE, **60** (8) (1972), 926-935.
10. N. Gaffke,O. Krafft: *Matrix inequalities in the Löwner ordering* Modern Applied Mathematics (1982) 595-622.
11. A. Golan, G. Judge, D. Miller: *Maximum entropy econometrics: Robust estimation with limited data*,(1997).
12. G. Khalaf: *A comparison between biased and unbiased estimators in Ordinary least squares regression*, Journal of Modern Applied Statistical Methods, **12** (2) (2013), 17.
13. G. Khalaf, M. Iguernane: *Multicollinearity and a ridge parameter estimation approach*, Journal of Modern Applied Statistical Methods **15** (2) (2016), 25.
14. O. Krafft: *A matrix optimization problem*, Linear Algebra and its Applications **51** (1983), 137-142.
15. J. Lintner: *Security prices, risk, and maximal gains from diversification*, The journal of finance **20** (4) (1965), 587-615.
16. H. Markowitz: *Portfolio Selection*, The Journal of Finance **7** (1) (1952), 77–91.
17. H. Markowitz: *Portfolio Selection: Efficient Diversification of Investments*, New York: JohnWiley and Sons, Inc.,(1959).
18. B. D. Marx,P. H. Eilers: *Generalized linear regression on sampled signals and curves: a P-spline approach*, Technometrics **41** (1) (1993), 1-13.
19. D. Pappas, A. Perperoglou: *Constrained matrix optimization with applications*, Journal of Applied Mathematics and Computing **40** (1) (2012), 357-369.
20. Q. Paris: *Multicollinearity and maximum entropy estimators*, Economics Bulletin **3** (11) (2001), 1-9.

21. R. Penrose: *A generalized inverse for matrices*, In Mathematical proceedings of the Cambridge philosophical society (Vol. 51, No. 3, pp. 406-413). Cambridge University Press,(1955).

22. S. Qemo,E. Elsaid: *Statistical Modelling of the Capital Asset Pricing Model (CAPM)*, Accounting and Finance Research **7** (146) (2018), 03.

23. R. C. Rao, H. Toutenburg: *Linear Models: Least Squares and Alternatives*, 5-21,(1999).

24. W. F. Sharpe: *Capital asset prices: A theory of market equilibrium under conditions of risk*, The journal of finance, **19** (3) (1964), 425-442.

25. W. Yamaka, K. Autchariyapanitkul,P. Meneejuk, S. Sriboonchitta: *Capital asset pricing model through quantile regression: an entropy approach*, Thai Journal of Mathematics (2017), 53-65.

26. Y. Zakari, S. A. Yau, U. Usman: *Handling multicollinearity; a comparative study of the prediction performance of some methods based on some probability distributions*, Annals. Computer Science Series (2018), 15-21.