Orthogonal Projection in Teaching Regression and Financial Mathematics

Farida Kachapova
Auckland University of Technology

Ilias Kachapov

Journal of Statistics Education Volume 18, Number 1 (2010),
www.amstat.org/publications/jse/v18n1/kachapova.pdf

Key Words: Population model of regression; Regression in finance; Optimal portfolio; Minimizing risk.

Abstract

Two improvements in teaching linear regression are suggested. The first is to include the population regression model at the beginning of the topic. The second is to use a geometric approach: to interpret the regression estimate as an orthogonal projection and the estimation error as the distance (which is minimized by the projection). Linear regression in finance is described as an example of practical applications of the population regression model.

The paper also describes a geometric approach to teaching the topic of finding an optimal portfolio in financial mathematics. The approach is to express the optimal portfolio through an orthogonal projection in Euclidean space. This allows replacing the traditional solution of the problem with a geometric solution, so the proof of the result is merely a reference to the basic properties of orthogonal projection. This method improves the teaching of the topic by avoiding tedious technical details of the traditional solution such as Lagrange multipliers and partial derivatives. The described method is illustrated by two numerical examples.

1. Introduction

In this paper we demonstrate how the concepts of vector space and orthogonal projection are used in teaching linear regression and some topics in financial mathematics. Through geometric interpretations the proofs are made shorter and clearer.

In Section 2 we remind the reader of some basic facts from linear algebra about orthogonal projection. In Section 3 we describe the Euclidean space of random variables and discuss meanings of the term "independence" in different contexts.

Sections 4, 6 and 7 describe some improvements in teaching linear regression. Many statistics courses consider regression as fitting lines to data. Modern textbooks, for example Wild and Seber (2000), Chatterjee (2000), Moore and McCabe (2006), teach applied statistics without referring to probability and mathematical foundations of the statistical methods. Computer software is widely used for data analysis, which makes the analysis easier but also turns it into a mysterious process. The population model of regression is sometimes taught in courses on probability theory, for example Hsu (1997), Grimmett and Stirzaker (2004). But this model is rarely considered in statistics courses. These courses teach regression only for samples or briefly mention the population model in a descriptive way. They introduce formulas for estimates of the regression coefficients without considering formulas for the coefficients themselves. This leads to long definitions and tedious proofs (or lack of the proofs). So instead of focussing on the idea of regression the students concentrate on long calculations or specifications of particular computer software.

There are at least two benefits of discussing the population model of regression. The first is that it makes some of the ideas in regression clear because sample estimates are natural analogs of features in the population. The second is that this facilitates discussion of an interesting application in financial mathematics, which will be discussed in Section 5.

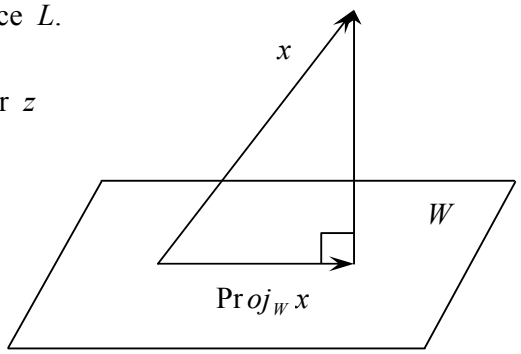
This paper describes a geometric approach in teaching regression when the regression estimate is interpreted as an orthogonal projection and the residual is interpreted as the distance from the projection. The geometric approach is used in textbooks on regression to some extent but the projection there is completely different: it relates to samples and n-dimensional space $\mathbf{R}^n$ while in this paper the projection relates to the population and linear space of random variables.

In Section 8 we describe another application of orthogonal projection in financial mathematics. One of the problems of portfolio analysis is finding an optimal portfolio – the portfolio with the lowest risk for a targeted return. This problem is included in textbooks on financial modelling (Benninga, 2000; Francis and Taylor, 2000), often without mathematical justification of the result. When a mathematical solution is provided, the problem is treated as a minimization problem and the solution is found in coordinate form using Lagrange multipliers and partial derivatives (Teall and Hasan, 2002; Cheang and Zhao, 2004; Kachapova and Kachapov, 2005; Kachapova and Kachapov, 2006). Here we suggest an invariant solution that uses geometric approach to random variables and orthogonal projection in particular.

2. Geometric Facts about Orthogonal Projection

In this section we will fix a Euclidean space L.

Definition 1. Suppose x is a vector in L and W is a linear subspace of L. A vector z is called the **orthogonal projection** of x onto W if
1) $z \in W$ and
2) $(x - z) \perp W$.
Then z is denoted $\mathrm{Proj}_W x$. •



The following is a well-known fact in linear algebra.

Theorem 1. **1)** $\mathrm{Pr}oj_W x$ is the vector in W closest to x and it is the only vector with this property.
2) If $v_1,..., v_n$ is an orthogonal basis in W, then

$$\mathrm{Pr}oj_W x = \frac{(x, v_1)}{(v_1, v_1)} v_1 + ... + \frac{(x, v_n)}{(v_n, v_n)} v_n,$$

where (u, v) denotes the scalar product of vectors u and v. ■

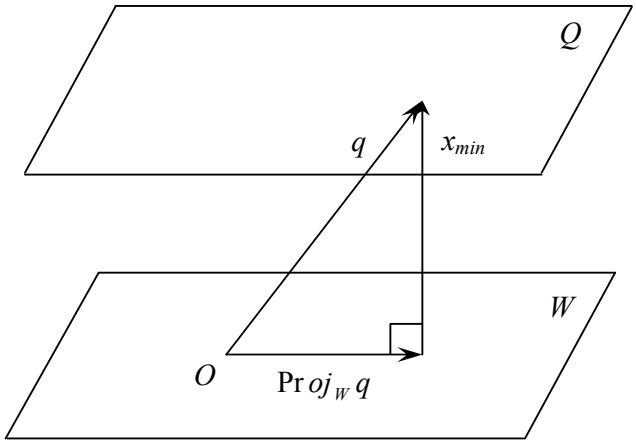
Definition 2. A subset Q of a linear space B is called an **affine subspace** of B if there is $q \in Q$ and a linear subspace W of B such that $Q = \{q + w \mid w \in W\}$. Then W is called the corresponding linear subspace. •

It is easy to check that any vector in Q can be taken as q.

Theorem 2. Let $Q = \{q + w \mid w \in W\}$ be an affine subspace of L. Then the vector in Q with the smallest length is unique and is given by the formula

$$x_{min} = q - \mathrm{Pr}oj_W q.$$

Proof of Theorem 2



Denote $z = \mathrm{Pr}oj_W q$, then $x_{min} = q - z$.

Consider any vector $y \in Q$. For some $w \in W$, $y = q + w$. By Theorem 1.1), z is the vector in W closest to q and $-w \in W$, so we have

$$\| y \| = \| q - (-w) \| \geq \| q - z \| = \| x_{min} \|.$$

The equality holds only when $-w = z$, that is when $y = q + w = q - z = x_{min}$. ■

Since x_{min} is unique, it does not depend on the choice of q.

3. Geometric Approach to Random Variables

3.1. The Vector Space of Random Variables

We will consider random variables on the same probability space. The set H of all random variables with finite variances is a linear space (with obvious operations of addition and multiplication by a number).

We denote $\mu_X = \mathrm{E}(X)$ the expectation of a random variable X, $\sigma_X^2 = \mathrm{Var}(X)$ the variance of X and $\mathrm{Cov}(X, Y)$ the covariance of random variables X and Y.

A scalar product given by $(X, Y) = \mathrm{E}(X \cdot Y)$ makes H a Euclidean space. In this space the length of a vector X is given by $\| X \| = \sqrt{(X, X)} = \sqrt{E(X^2)}$ and the distance between vectors X and Y is given by $d(X, Y) = \| X - Y \|$.

A similar approach is used by Grimmett and Stirzaker (2004, pg. 343-347) but they do not introduce scalar product on random variables. However, the scalar product is very relevant to orthogonal projections and makes proofs shorter.

In simple cases we can construct a basis of the space H. The following example illustrates that.

Example 1. Consider a finite sample space $\Omega = \{\omega_1, ..., \omega_n\}$ with the probabilities of the outcomes $p_i = \mathrm{P}(\omega_i) > 0$, $i = 1, ..., n$. In this case we can introduce a finite orthogonal basis in the Euclidean space H.

For each i define a random variable F_i as follows: $F_i(\omega_j) = \begin{cases} 1 & if \quad j = i, \\ 0 & if \quad j \neq i. \end{cases}$

Then for any random variable X in H,

$$X = \sum_{i=1}^{n} x_i F_i , \qquad (1)$$

where $x_i = X(\omega_i)$.

For any $i \neq j$, $F_i \cdot F_j = 0$ and $(F_i, F_j) = \mathrm{E}(F_i \cdot F_j) = 0$, so

$$F_i \perp F_j . \qquad (2)$$

(1) and (2) mean that $F_1, ..., F_n$ make an orthogonal basis in H and the dimension of H is n.

For any X, Y in H, their scalar product equals $(X, Y) = \sum_{i=1}^{n} p_i x_i y_i$, where $y_i = Y(\omega_i)$.

■

3.2. The Concepts of Independence

The authors think it is worthwhile to discuss the concepts of independence and dependence that students encounter in different contexts. There is independence of events, which we do not use here. For the current topics it is important that the students distinguish between independence/ dependence of random variables, their linear relation and their linear dependence as vectors in H. We will remind the definitions for only two variables for brevity.

Definition 3. Suppose X and Y are elements of H.

1) The random variables X and Y are called **independent** if for any numbers x, y,

$$P(X \le x,\ Y \le y) = P(X \le x) \cdot P(Y \le y).$$

Otherwise the random variables are called **dependent**.

2) The random variables X and Y are said to have a **linear relation** if $Y = \alpha + \beta X$ or $X = \alpha + \beta Y$ for some numbers α, β.
Equivalently it means that there are numbers a, b, not both zero, and a number c, for which $aX + bY = c$.

3) The vectors X and Y in a linear space are called **linearly dependent** if there are numbers a, b, not both zero, for which $aX + bY = 0$. Otherwise the vectors are called **linearly independent**. •

Obviously linear dependence 3) implies linear relation 2), and linear relation 2) implies dependence 1).

Example 2. Let us look at the variables $F_1,..., F_n$ described in Example 1. They are dependent as random variables (Definition 3.1) because $P(F_1 = 1, F_2 = 1) = 0$ and

$P(F_1 = 1) \cdot P(F_2 = 1) = P(\omega_1) \cdot P(\omega_2) = p_1 \cdot p_2 \neq 0.$

Next, the random variables $F_1,..., F_n$ have a linear relation (Definition 3.2) because

$F_1 +...+ F_n = 1$. Indeed, for any ω_i, $(F_1 +...+ F_n)(\omega_i) = F_i(\omega_i) = 1.$

Finally, $F_1,..., F_n$ are linearly independent as vectors (Definition 3.3) because if

$a_1F_1 +...+ a_nF_n = 0$, then for any $i = 1,..., n$, $0 = (a_1F_1 +...+ a_nF_n)(\omega_i) = a_iF_i(\omega_i) = a_i$. ■

To add to the students' confusion, there are also terms "independent variable" and "dependent variable" in statistics, which are not strictly defined but intuitively understood.

In statistics we also talk about linear dependence of random variables measured by their correlation coefficient ρ_{XY}. This term "linear dependence" is never properly defined but the theory states that the linear dependence between X and Y gets weaker as ρ_{XY} approaches 0 and does not exist when $\rho_{XY} = 0$.

This linear dependence has an interesting geometric analogy for variables with 0 expectations. If X and Y are such variables in H, then $(X, Y) = \mathrm{Cov}(X, Y)$,

$\| X \| = \sigma_X$, $\| Y \| = \sigma_Y$, and for the angle θ between vectors X and Y,

$$\mathrm{Cos}\ \theta = \frac{(X,Y)}{\| X \| \cdot \| Y \|} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \rho_{XY}.$$

It is easy to prove the following facts using properties of the correlation coefficient as a measure of linear dependence between random variables.

Theorem 3. Suppose X and Y are variables in H with 0 expectations, and θ is the angle between them as vectors. Then

1) Cos $\theta = \rho_{XY}$,

2) the random variables X and Y have a linear relation if and only if the angle θ equals 0° or 180° ;

3) if X and Y are independent, then they are orthogonal;

4) there is no linear dependence between random variables X and Y if they are orthogonal vectors;

5) with angle θ getting closer to 90° there is less linear dependence remaining between X and Y. ■

4. The Population Model of Regression and a Geometric Approach to Teaching Regression

The idea of regression is clear and simple when it is applied to random variables and expressed in terms of a population, not samples. Regression apparently means "estimating an inaccessible random variable Y in terms of an accessible random variable X" (Hsu, 1997), that is finding a function $f(X)$ "closest" to Y. We call this the population model of regression. $f(X)$ can be restricted to a certain class of functions, the most common being the class of linear functions. We describe "closest" in terms of the distance d defined in Section 3.1.

Theorem 1.1) states that $\mathrm{Proj}_W Y$ is the vector in W closest to Y. Choosing different W's we can get different types of regression: simple linear, multiple linear, quadratic, polynomial, etc.

Theorem 4. The conditional expectation $\mathrm{E}(Y \mid X)$ is the function of X closest to Y. ■

This is based on the following fact:

$$\mathrm{E}(Y \mid X) = \mathrm{Proj}_W Y \quad \text{for} \quad W = \{ f(X) \mid f\colon \mathbf{R} \to \mathbf{R} \text{ and } f(X) \in H \}.$$

Grimmett & Stirzaker (2004) on pg. 346 prove the fact by showing that $\mathrm{E}(Y \mid X) \in W$ and that for any $h(X) \in W$, $\mathrm{E}[(Y - \mathrm{E}(Y \mid X)) \cdot h(X)] = 0$, that is $Y - \mathrm{E}(Y \mid X) \perp h(X)$.

Theorem 5 (**simple linear regression**). If $\sigma_X \neq 0$, then the linear function of X closest to Y is given by

$$\alpha + \beta X, \text{ where } \begin{cases} \beta = \dfrac{Cov(Y, X)}{\sigma_X^{\,2}}, \\ \alpha = \mu_Y - \beta \mu_X. \end{cases} \qquad (3)$$

■

Corollary. If $\hat{Y} = \alpha + \beta X$ is the best linear estimator of Y from Theorem 5, then the residual $\varepsilon = Y - \hat{Y}$ has the following properties:

1) $\mu_\varepsilon = 0$, **2)** $Cov(\varepsilon, X) = 0$.

■

Thus, according to the Corollary, the residuals (estimation errors) equal 0 on average and are uncorrelated with the predictor X; this is another evidence that $\hat{Y}$ is the best linear estimator of Y.

Geometric proofs of Theorem 5 and its Corollary will be given in Section 6.

5. Application in Financial Mathematics

Regression is often taught only for samples, without even considering the population model for random variables. Perhaps some people believe that the population model is not used in real life. We use the following application of regression in finance to demonstrate to students that the population regression model is a practical concept.
Portfolio analysis is the part of financial mathematics that studies the world of N fixed assets $A_1, A_2, \ldots, A_N$, and their combinations called portfolios.
The % return (we will omit % for brevity) from an investment is treated as a random variable. We will identify any portfolio x with its return and also denote the return by x.

For a portfolio x, $\mu_x = \mathrm{E}(x)$ is called the expected return and the variance $\sigma_x^2 = \mathrm{Var}(x)$

is used as a measure of risk.

Despite the tradition to use capital letters for random variables, in portfolio analysis it is more suitable to use small letters for portfolio returns.

All portfolios are regressed to the *market portfolio m*, the portfolio containing every asset with the weight proportional to its market value. Thus, for any portfolio x the following is true.

1) $x = \alpha_x + \beta_x m + \varepsilon$. The regression line $\alpha_x + \beta_x m$ is the linear function of m closest to x. So the coefficient β_x is the average rate of change of x's return with respect to the market return.

2) For the residual ε, $\mu_\varepsilon = 0$ and $Cov(\varepsilon, m) = 0$.

3) The variance σ_x^2 represents the total risk of portfolio x;

$\beta_x^2 \sigma_m^2$ is the *systematic risk* (or *market risk*) of x, the risk that affects most investments;

σ_ε^2 is the *unsystematic risk* of x, the risk that affects only a small number of investments.

The total risk of x is a sum of the systematic risk and unsystematic risk:

$$\sigma_x^2 = \beta_x^2 \sigma_m^2 + \sigma_\varepsilon^2 .$$

Indeed, $\sigma_x^2 = Var(x) = Var(\alpha_x + \beta_x m + \varepsilon) = Var(\beta_x m + \varepsilon) =$

$= \beta_x^2 Var(m) + Var(\varepsilon) + 2\beta_x Cov(\varepsilon, m) = \beta_x^2 \sigma_m^2 + \sigma_\varepsilon^2$, since $Cov(\varepsilon, m) = 0$.

4) The coefficient $\rho_{x\,m}$ of correlation between the x's return and the market return

has the following interpretation.

Theorem 6. ρ_{xm}^2 is the proportion of the systematic risk of portfolio x.

Proof of Theorem 6

$$\frac{systematic\ risk\ of\ x}{total\ risk\ of\ x} = \frac{\beta_x{}^2\sigma_m{}^2}{\sigma_x{}^2} = \beta_x{}^2\frac{\sigma_m{}^2}{\sigma_x{}^2} = \text{(by Theorem 5)} =$$

$$= \left(\frac{Cov(x,m)}{\sigma_m^2}\right)^2 \cdot \frac{\sigma_m^2}{\sigma_x^2} = \left(\frac{Cov(x,m)}{\sigma_m\sigma_x}\right)^2 = \rho_{xm}^2 \ . \qquad \blacksquare$$

Clearly $\beta_x{}^2 = \frac{\beta_x{}^2\sigma_m{}^2}{\sigma_m{}^2} = \frac{systematic\ risk\ of\ x}{risk\ of\ the\ market\ portfolio}$. So the beta coefficient is used for ordinal ranking of assets according to their systematic risk. In particular, an asset with $\beta > 1$ is called an **aggressive** asset (it is more volatile than the market portfolio), and an asset with $\beta < 1$ is called a **defensive** asset (it is less volatile than the market portfolio).

The following table shows the estimated beta coefficients of some companies in June 2006:

Company Name	**Beta coefficient**
Coca-Cola Company	0.36
Honda Motor Company	0.65
Toyota Motor Corporation	0.70
Telecom Corporation NZ	0.74
Vodafone Group PLC	1
Harley-Davidson	1.03
Sony Corporation	1.09
Microsoft Corporation	1.11
Boeing	1.11
Hilton Hotels	1.14
McDonald's Corporation	1.19
General Motors	1.32
Apple Computer	1.53
Nokia	1.79
Ford Motor Company	1.84
Xerox Corporation	1.92
Yahoo! Inc.	2.50

5) All this leads to a very important model in finance – the **capital asset pricing model**: $\mu_x = f + \beta_x(\mu_m - f)$. Here f is the risk-free return and μ_m is the expected return of the market portfolio. For more details on regression in finance see Kachapova and Kachapov (2006).

6. Geometric Proofs

Advanced textbooks on regression (Seber, 1980; Saville and Wood, 1996; Freund, 1998; Seber and Lee, 2003; Chiang, 2003; Dowdy, 2004) contain mathematical proofs of the regression model that use the geometric approach with samples rather than the population. Such proofs involve coordinates, matrices and partial derivatives and either are very long or contain significant gaps. So it would be beneficial for the students to provide shorter proofs. The authors believe that the geometric proof below for the coefficients of simple linear regression is shorter and conceptually clearer than the usual proofs minimising mean-square error.

Proof of Theorem 5

Denote $W = \{ a + b X \mid a, b \in \mathbf{R} \}$. $\Pr oj_W Y \in W$, so $\Pr oj_W Y = \alpha + \beta X$ for some $\alpha, \beta \in \mathbf{R}$. We just need to show that α and β are given by the formula (3).

For $\varepsilon = Y - \Pr oj_W Y = Y - (\alpha + \beta X)$, we have $\varepsilon \perp 1$ and $\varepsilon \perp X$, since $1, X \in W$.

So $(\varepsilon, 1) = 0$ and $(\varepsilon, X) = 0$, $(\alpha + \beta X, 1) = (Y, 1)$ and $(\alpha + \beta X, X) = (Y, X)$, which leads to a system of linear equations:

$$\begin{cases} E(\alpha + \beta X) = E(Y) \\ E(\alpha X + \beta X \cdot X) = E(Y \cdot X) \end{cases} \quad \text{and} \quad \begin{cases} \alpha + \beta \mu_X = \mu_Y \\ \alpha \mu_X + \beta E(X^2) = E(Y \cdot X) \end{cases}.$$

The solution of the system is given by (3). ■

The following corollary of Theorem 5 was stated in Section 2.

Corollary. If $\hat{Y} = \alpha + \beta X$ is the best linear estimator of Y from Theorem 5, then the residual $\varepsilon = Y - \hat{Y}$ has the following properties:

$$\textbf{1)}\ \mu_\varepsilon = 0, \quad \textbf{2)}\ Cov(\varepsilon, X) = 0.$$

Proof of Corollary

1) $\varepsilon \perp 1$, so $E(\varepsilon) = 0$.

2) $\varepsilon \perp X$, so $E(\varepsilon \cdot X) = 0$ and $Cov(\varepsilon, X) = E(\varepsilon \cdot X) - E(\varepsilon) \cdot E(X) = 0$. ■

The Corollary leads to a different interpretation of regression. It is useful in teaching students who are not familiar with the concept of orthogonal projection. The random variable Y is divided into two parts:

$$Y = f(X) + \varepsilon$$

and it is required that $\mu_\varepsilon = 0$ and $Cov(\varepsilon, X) = 0$. The part $f(X)$ is called the regression estimate of Y.

Theorem 5a. The regression estimate of Y in the class of all linear functions of X is given by the formula (3).

Proof of Theorem 5a

$Y = \alpha + \beta X + \varepsilon$. $\mathrm{Cov}(Y, X) = \mathrm{Cov}(\alpha, X) + \beta\,\mathrm{Cov}(X, X) + \mathrm{Cov}(\varepsilon, X) =$

$= 0 + \beta\,\sigma_X^2 + 0 = \beta\,\sigma_X^2$. So $\beta = \dfrac{Cov(Y, X)}{\sigma_X^2}$.

$\mu_Y = \alpha + \beta\,\mu_X + \mu_\varepsilon = \alpha + \beta\,\mu_X$, so $\alpha = \mu_Y - \beta\,\mu_X$. ■

Unlike Theorem 5, Theorem 5a is not stated in terms of "closest" object. But it has a very easy proof.

7. Linear Regression for Samples

After the population regression model is introduced we create the statistics regression model as a sample estimate. We follow the common pattern in estimation theory when a population object is estimated from a sample. For example, the population mean μ is estimated by a sample mean $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$. Similarly the equation $Y = \alpha + \beta X + \varepsilon$ of the simple linear regression is estimated from a sample by the equation $Y = a + b X + e$, where a, b, and e are sample estimates of α, β, and ε respectively. Substituting the corresponding sample estimates for the parameters in (3), we get formulas for the coefficients a and b:

$$\begin{cases} b = \dfrac{s_{yx}}{s_x^2}, \\ a = \bar{y} - b\,\bar{x}, \end{cases}$$

where $\bar{x}$, $\bar{y}$, s_x^2 and s_{yx} are the sample estimates of μ_X, μ_Y, σ_X^2 and $Cov(Y, X)$ respectively.

8. Optimal Portfolio of Financial Assets

We fix N financial assets $A_1, A_2, \ldots, A_N$.

8.1. Modelling Portfolios as Random Variables

Notations

r_k denotes the return of asset A_k; $\mu_k = \mathrm{E}(r_k)$, $\sigma_k^2 = \mathrm{Var}(r_k)$ and $\sigma_{kj} = \mathrm{Cov}(r_k, r_j)$.

$U = \begin{bmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{bmatrix}$ is the column of ones of length N;

$M = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_N \end{bmatrix}$ is the vector of the expected asset returns, where not all μ_k are the same;

$S = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2N} \\ \dots & \dots & \dots & \dots \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{NN} \end{bmatrix}$ is the covariance matrix of the asset returns.

We will assume that all expected returns $\mu_1, \dots, \mu_N$ are defined and the covariance matrix S exists with $\det S > 0$. The positivity of the determinant of S is equivalent to the fact that the random variables $r_1, \dots, r_N$ do not have a linear relation (recall Definition 3.2). This also means that $r_1, \dots, r_N$ are linearly independent as vectors.

For a portfolio x, x_k denotes the proportion of the value of asset A_k in the portfolio's total value (negative x_k means short sales).

Theorem 7. For a portfolio x,

1) $x = \sum_{k=1}^{N} x_k r_k$;

2) $x_1 + \dots + x_N = 1$. ■

Let us consider the set K of all linear combinations of $r_1, \dots, r_N$. Apparently K is a linear subspace of the Euclidean space H of random variables described in Section 3.

Theorem 8. **1)** $r_1, \dots, r_N$ is a basis in K.

2) The dimension of K equals N. ■

According to Theorems 7 and 8, any portfolio x is a vector in the N-dimensional Euclidean space K and can be represented as a column of its coordinates in the basis $r_1, \dots, r_N$:

$x = \begin{bmatrix} x_1 \\ \dots \\ x_N \end{bmatrix}$.

Theorem 9. For any portfolios $x = \begin{bmatrix} x_1 \\ \dots \\ x_N \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix}$ the following hold:

1) $\mu_x = \mathrm{E}(x) = \mu_1 x_1 + \dots + \mu_N x_N = x^{\mathrm{T}} \cdot M$;

2) $\sigma_x^2 = \mathrm{Var}(x) = x^{\mathrm{T}} \cdot S \cdot x$;

3) $\mathrm{Cov}(x, y) = x^{\mathrm{T}} \cdot S \cdot y$. ■

8.2. The Problem of Optimal Portfolio

Suppose an investor wants a certain return c from a portfolio. Usually one can choose between many portfolios with the same expected return. Clearly the investor wants to pick the portfolio with the lowest risk. We will call it the optimal portfolio. Risk is measured with variance. Thus, we get the following definition.

Definition 4. The **optimal portfolio for targeted return** c is the portfolio with the smallest variance among the portfolios with the same expected return c. •

Thus, the optimal portfolio minimizes risk for a targeted return.

A vector $x = \begin{bmatrix} x_1 \\ ... \\ x_N \end{bmatrix}$ is a portfolio with expected return c if it satisfies the following two conditions:

$$\begin{cases} x_1 + ... + x_N = 1 \\ \mu_x = c \end{cases}$$

These conditions can be written in matrix form:

$$\begin{cases} x^T \cdot U = 1 \\ x^T \cdot M = c \end{cases} \qquad (4)$$

Thus, in mathematical form the problem of optimal portfolio can be written as follows:

$$\begin{cases} Var(x) \to \min \\ x^T \cdot U = 1 \\ x^T \cdot M = c \end{cases}.$$

8.3. Geometric Solution of the Problem of Optimal Portfolio

Denote Q the set of all solutions of the system of linear equations (4) and W the set of all solutions of the corresponding homogeneous system:

$$\begin{cases} x^T \cdot U = 0 \\ x^T \cdot M = 0 \end{cases} \qquad (5)$$

Clearly W is a linear subspace of K of dimension N–2 (note that U and M are not proportional) and Q is an affine subspace of K with W as the corresponding linear subspace. That is, if q is any solution of (4), then $Q = \{q + w \mid w \in W\}$.
Here we assume $N \geq 3$. The case $N = 2$ is trivial: in this case there is only one portfolio with the expected return c.

Theorem 10. The optimal portfolio for the targeted return c is unique and is given by the formula

$$x_{min} = q - \mathrm{Proj}_W q,$$

where q is any solution of (4) and W is the set of all solutions of (5).

Proof of Theorem 10
For any solution x of (4) we have $\| x \|^2 = (x, x) = \mathrm{E}(x^2) = \sigma_x^2 + \mu_x^2 = \mathrm{Var}(x) + c^2$.

Since $\mu_x = c$ is fixed, the solution of (4) with the smallest variance is the same as the solution of (4) with the smallest length. So Theorem 2 can be applied. ■

Theorem 11. Suppose q is a solution of (4) and $v_1,..., v_{N-2}$ is an orthogonal system of vectors, each of which is a solution of (5). Then
1) the optimal portfolio for the targeted return c is given by the formula

$$x_{min} = q - \sum_{k=1}^{N-2} \frac{(q, v_k)}{(v_k, v_k)} v_k ;$$

2) for any vector y, $(y, v_k) = \mathrm{Cov}(y, v_k)$, $k = 1,..., N-2$.
Proof of Theorem 11
1) The dimension of W is $N-2$, so the orthogonal system $v_1,..., v_{N-2}$ makes a basis in W, and the formula follows from Theorem 10 and Theorem 1.2).

2) Since each v_k is a solution of (5), $\mathrm{E}(v_k) = v_k^{\mathrm{T}} \cdot M = 0$.

$(y, v_k) = \mathrm{E}(y \cdot v_k) = \mathrm{Cov}(y, v_k) + \mathrm{E}(y) \cdot \mathrm{E}(v_k) = \mathrm{Cov}(y, v_k) + \mathrm{E}(y) \cdot 0 = \mathrm{Cov}(y, v_k)$.
■

Example 3. Three assets have expected returns of 2%, 1% and 1% respectively and covariance matrix $S = \begin{bmatrix} 3 & 3 & -1 \\ 3 & 5 & -1 \\ -1 & -1 & 1 \end{bmatrix}$. If the targeted return is 3%, find

a) the portfolio of these assets with the lowest risk and
b) its variance.

Solution

a) $N = 3$. $M = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, $c = 3$. The system (4) has the form $\begin{cases} x_1 + x_2 + x_3 = 1 \\ 2x_1 + x_2 + x_3 = 3 \end{cases}$.

$q = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$ is a solution of this system (we assign an arbitrary value to one of the variables, e.g. take $x_3 = 0$, and solve for the other two variables).

Similarly we find a solution v of the homogeneous system $\begin{cases} x_1 + x_2 + x_3 = 0 \\ 2x_1 + x_2 + x_3 = 0 \end{cases}$:

$v = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$.

By Theorem 11, $x_{min} = q - \dfrac{(q,v)}{(v,v)} v$, $(q, v) = \text{Cov}\,(q, v)$ and $(v, v) = \text{Var}\,(v)$.

So $(q, v) = q^{\mathrm{T}} \cdot S \cdot v = \begin{bmatrix} 2 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 3 & 3 & -1 \\ 3 & 5 & -1 \\ -1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = -2;$

$(v, v) = v^{\mathrm{T}} \cdot S \cdot v = \begin{bmatrix} 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 3 & -1 \\ 3 & 5 & -1 \\ -1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = 8;$

$$x_{min} = q - \frac{(q,v)}{(v,v)} v = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} - \frac{-2}{8} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1.25 \\ 0.25 \end{bmatrix}, \quad x_{min} = \frac{1}{4} \begin{bmatrix} 8 \\ -5 \\ 1 \end{bmatrix}.$$

b) The variance of x_{min} equals $x_{min}{}^{\mathrm{T}} \cdot S \cdot x_{min} = 4.5$. ■

Example 4. Four assets have expected returns of 1%, 2%, 1% and 2% respectively and covariance matrix $S = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 3 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix}$. If the targeted return is 3%, find

a) the portfolio of these assets with the lowest risk and
b) its variance.

Solution

a) $N = 4$. $M = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}$, $c = 3$. The system (4) has the form $\begin{cases} x_1 + x_2 + x_3 + x_4 = 1 \\ x_1 + 2x_2 + x_3 + 2x_4 = 3 \end{cases}$.

$q = \begin{bmatrix} -1 \\ 2 \\ 0 \\ 0 \end{bmatrix}$ is a solution of this system (we assign arbitrary values to two variables, e.g. take $x_3 = x_4 = 0$, and solve for the other two variables).

Similarly we find a solution v_1 of the homogeneous system $\begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ x_1 + 2x_2 + x_3 + 2x_4 = 0 \end{cases}$:

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}.$$

Next we need to find a solution $v_2 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ of the same system that is orthogonal to v_1.

Thus, $0 = (v_1, v_2) = \text{Cov}(v_1, v_2) = v_1^{\mathrm{T}} \cdot S \cdot v_2 = \begin{bmatrix} 1 & 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 3 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} =$

$= x_1 - 2x_3 + 2x_4$.

So v_2 is found as a solution of the system $\begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ x_1 + 2x_2 + x_3 + 2x_4 = 0 \\ x_1 - 2x_3 + 2x_4 = 0 \end{cases}$: $v_2 = \begin{bmatrix} 2 \\ 3 \\ -2 \\ -3 \end{bmatrix}$.

By Theorem 11, $x_{min} = q - \frac{(q, v_1)}{(v_1, v_1)} v_1 - \frac{(q, v_2)}{(v_2, v_2)} v_2$,

$$(q, v_1) = \text{Cov}(q, v_1) = q^{\mathrm{T}} \cdot S \cdot v_1 = \begin{bmatrix} -1 & 2 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 3 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} = -1;$$

$$(q, v_2) = \text{Cov}(q, v_2) = q^{\mathrm{T}} \cdot S \cdot v_2 = \begin{bmatrix} -1 & 2 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 3 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \\ -2 \\ -3 \end{bmatrix} = 10;$$

$$(v_1, v_1) = \text{Var}(v_1) = v_1^{\mathrm{T}} \cdot S \cdot v_1 = \begin{bmatrix} 1 & 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 3 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} = 3;$$

$$(v_2, v_2) = \mathrm{Var}\,(v_2) = {v_2}^{\mathrm{T}} \cdot S \cdot v_2 = \begin{bmatrix} 2 & 3 & -2 & -3 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 3 & -1 \\ 1 & 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \\ -2 \\ -3 \end{bmatrix} = 24.$$

$$\text{So} \quad x_{min} = q - \frac{(q, v_1)}{(v_1, v_1)} v_1 - \frac{(q, v_2)}{(v_2, v_2)} v_2 = \begin{bmatrix} -1 \\ 2 \\ 0 \\ 0 \end{bmatrix} - \frac{-1}{3} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} - \frac{10}{24} \begin{bmatrix} 2 \\ 3 \\ -2 \\ -3 \end{bmatrix} = \begin{bmatrix} -1.5 \\ 0.75 \\ 0.5 \\ 1.25 \end{bmatrix},$$

$$x_{min} = \frac{1}{4} \begin{bmatrix} -6 \\ 3 \\ 2 \\ 5 \end{bmatrix}.$$

b) The variance of x_{min} equals ${x_{min}}^{\mathrm{T}} \cdot S \cdot x_{min} = 1.5$. ■

9. Discussion

Most students want to know reasons for the formulas and equations that they study, so omitting all the proofs can be frustrating to students with sufficient mathematical background to understand them. When proofs are presented, the authors suggest making them short and clear. Using the approach described we justify basic formulas for regression and at the same time avoid lengthy and tedious proofs. Clearly this is not applicable to the statistical inference for regression where tedious proofs are hard to simplify.

When considering the population regression model and applying orthogonal projection we clarify the main idea of regression as the estimation of Y in terms of X. This is a logical way to teach regression that we believe improves the students' critical thinking and conceptual knowledge of regression as a complement to the procedural knowledge provided in traditional statistics courses.

The suggested teaching approach requires the students to have some mathematical background. Firstly, they need to have some intuition about such concepts of linear algebra as scalar product, orthogonal projection, length and distance (the last three concepts are intuitively clear for two-dimensional and three-dimensional spaces). Secondly, the students need to be familiar with such concepts of probability theory as random variables, variance and covariance, and have basic skills in applying them. Therefore, the suggested approach can be effective only in the statistics courses, which are part of university courses in quantitative areas, such as mathematical studies, physics and engineering. The authors believe that these students are capable of abstract thinking and will appreciate the logic and structure of this approach.

The authors have used the approach described to teach regression in courses on statistics, probability theory and financial mathematics at the Auckland University of Technology (New Zealand) and the Moscow Technological University (Russia). Though a formal statistical analysis of the results is yet to be done, our case studies show that the students gained a better understanding of the concept of regression, regression formulas and their logical connections.

In particular, they demonstrated an understanding that a regression line constructed from empirical data, is only an approximation of the true relationship between two variables and that a different set of data may lead to a different approximation of the same relationship.

The second part of the paper (Section 8) describes a geometric approach to teaching the problem of optimal portfolio in a university course on financial mathematics. The courses on financial mathematics develop financial theories using mathematical techniques of calculus, probability theory, stochastic differentiation and integration. Here we apply the geometric technique of orthogonal projection to the problem of optimal portfolio. The use of the orthogonal projection makes the reasoning for the problem short and invariant, while the traditional solution for the optimal portfolio is long and involves coordinates and partial derivatives. The new method helps the students to concentrate on meaningful modelling instead of tedious technical details. This is especially helpful for the students whose calculus technique is weak. Clearly the described approach is suitable mostly in the university mathematical courses, since the students need to have a reasonable mathematical background. Case studies at the same universities show that the students understand the topic better and learn it faster with the new approach. The approach described also demonstrates links between different areas of mathematics and helps the students to see practical applications of the abstract concepts of vector space and orthogonal projection.

References

Benninga, S. (2000). *Financial modelling*. Massachusetts Institute of Technology.

Chatterjee, S. (2000). *Regression analysis by example.* (3rd ed.). New York: Wiley.

Cheang, G.H. & Zhao, Y. (2004). *Calculus and matrix algebra for finance*. (2nd ed.). McGraw-Hill.

Chiang, C.L. (2003). *Statistical methods of analysis.* River Edge, N.J.: World Scientific.

Dowdy, S.M. (2004). *Statistics for research.* (3rd ed.). Hoboken, N.J.: Wiley-Interscience.

Francis, J.C. & Taylor R.W. (2000). *Investments*. (2nd ed.) New York: McGraw-Hill.

Freund, R.J. (1998). *Regression analysis: statistical modelling of a response variable.* San Diego: Academic Press.

Grimmett, G. & Stirzaker, D. (2004). *Probability and random processes.* (3rd ed.). New York: Oxford University Press.

Hsu, H.P. (1997). *Probability, random variables, and random processes.* New York: McGraw-Hill.

Kachapova, F. & Kachapov, I. (2006). *Mathematical approach to portfolio analysis.* Auckland: Maths Ken.

Kachapova, F. & Kachapov, I. (2005) Calculating optimal portfolios of three financial assets. *Mathematics and Informatics Quarterly*. No. 3/4, Vol. 15, pp. 100-107.

Moore, D.S. & McCabe, G.P. (2006). *Introduction to the practice of statistics.* (5th ed.). New York: W.H.Freeman and Company.

Saville, D.J. & Wood, G.R. (1996). *Statistical methods. A geometric primer.* New York: Springer-Verlag.

Scheffé, H. (1959). *The analysis of variance.* John Wiley & Sons.

Seber, G.A.F. (1980). *The linear hypothesis: a general theory.* (2th ed.). Charles Griffin & Company Ltd.

Seber, G.A.F. & Lee, A.J. (2003). *Linear regression analysis.* (2th ed.). John Wiley & Sons.

Teall, J.L. & Hasan, I. (2002). *Quantitative methods for finance and investment*. Oxford: Blackwell.

Wild, C.J. & Seber, G.A.F. (2000). *Chance encounters.* John Wiley & Sons.

Dr Farida Kachapova
Auckland University of Technology
School of Computing and Mathematical Sciences
Private Bag 9206
Auckland 1142
New Zealand
farida.kachapova@aut.ac.nz
Ph. +64 9 9219999 ext. 8248

Dr Ilias Kachapov
13 Fairlands Ave
Waterview
Auckland 1026
New Zealand
kachapov@xtra.co.nz
Ph. +64 9 8282007
