

APPLICATIONS OF MATHEMATICS IN ECONOMETRICS

KAI LI

1. Introduction. In a broad view, econometrics is the science and art of analyzing economic data using economic theory and statistical techniques [3]. For instance, econometricians are interested in exploring the wage gap among different education levels so that policymakers can adjust policies effectively. They also investigate returns on strategies used by central banks to decide whether to follow implied economic theory or not. Lastly, they predict critical macroeconomic variables such as interest rates, inflation rates, and GDP to track the health of a nation's economy [7].

I am interested in econometrics since it is both a theoretical subject that focuses on economic theory and empirical analysis that uses data sets to test economic theory. Econometrics is cross-disciplinary, and its methods utilize undergraduate-level statistics and data analysis techniques. Thus, familiarity with calculus, matrix algebra, and statistics greatly help to understand mathematics applications in econometrics.

From an empirical perspective, econometricians use data to test a theory or estimate a relationship by constructing a formal economic model [7]. One problem that occurs during the stage of collecting data is to identify various data types. The three main types are cross-sectional data, time series data, and panel/longitudinal data [3].

Cross-sectional data is the most fundamental type and will be used to formulate an economic model in the paper. Though some econometric methods can be applied directly to many different types of data sets without any modifications, some data sets' unique characteristics must be accounted for or should be exploited [7]. In that sense, we have to be careful identifying appropriate methods corresponding to the data sets, as it is a significant procedure in econometric analysis.

The goal of the paper is to specify an econometric model using applied mathematics concepts. Specifically, I will be focusing on the linear regression model, using the method of least squares, because this model readily leads to further mathematical treatment and provides good approximations to otherwise complicated regression equations [1]. The least squares method is a curve fitting technique suggested by Legendre [1] and published by Gauss early in the nineteenth century [6].

Once an econometric model has been constructed, hypothesis testing of interest can be stated in terms of unknown parameters. Forecasting in either testing a theory or studying a policy's impact becomes probable in later econometric analysis [7].

2. The Method of Least Squares (LS) in Linear Regression. Consider a probabilistic linear multiple regression model of a sample with n observations:

$$(1) \quad Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, 2, \dots, n,$$

where ϵ_i is a random error with $E[\epsilon_i] = 0$ and $\text{Var}[\epsilon_i] = \sigma^2$ [6]. $Y_i, x_{i1}, \dots, x_{ik}$ are called the regressand and regressors, respectively, and are determined by the economic theory, intuition and data considerations [3, 7]. β_0, \dots, β_k are regression coefficients. The method of LS finds all coefficients such that regression line provides the best fit to data. Denote the vertical distance from each observed point to estimated regression line by residual e_i . If the model is correct, e_i 's are estimates of ϵ_i 's. The LS estimate of β 's, denoted by $\hat{\beta}$'s, can be obtained by minimizing the following equation¹:

$$(2) \quad Q = \sum e_i^2 = \sum [y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})]^2.$$

¹I will refer to \sum as $\sum_{i=1}^n$ for all summations in the paper.

3. Extreme Values of a Function. Advanced calculus helps to minimize Q in Eq. (2). Based on Theorem 11.6.3 from [2], we differentiate partially with respect to the β 's, and equate these $k + 1$ partial derivatives to zero. After simplifications, we are able to solve the following $k + 1$ normal equations, and get the desired $\hat{\beta}$'s:

$$(3) \quad \begin{aligned} \sum y_i &= n\beta_0 + \left(\sum x_{i1}\right)\beta_1 + \cdots + \left(\sum x_{ik}\right)\beta_k \\ \sum x_{ij}y_i &= \left(\sum x_{ij}\right)\beta_0 + \left(\sum x_{ij}x_{i1}\right)\beta_1 + \cdots + \left(\sum x_{ij}x_{ik}\right)\beta_k, \quad j = 1, 2, \dots, k. \end{aligned}$$

It can be shown that Q is a convex function of β 's. By Theorem 11.6.6 in [2], the first partials of Q with respect to β_0, \dots, β_k equal to zero yield a global minimum. That is, the $\hat{\beta}$'s are indeed the coefficients for our econometric linear regression model.

4. The Big Picture for LS: Projection. The method of least squares, Eq. (1), can be represented more conveniently using matrix notation: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Denote

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

However, it is not always possible to find a perfect regression line for data points. To verify, $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$ has no solution if the nullspace of \mathbf{X} is $\{\mathbf{0}\}$ [5]. In this case, $\|\mathbf{e}\|$ can be minimized and $\hat{\boldsymbol{\beta}}$ can be found using the Least Squares approximation. Otherwise, the linear regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ is a perfect fit with residual $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0}$, which exists in rare and idealized situations [6]. In econometrics, due to the existence of other likely factors that impact the regressand, the observations generally do not fall exactly on the population regression line [3].

Therefore, suppose that no exact solution exists. The cross-product matrix $\mathbf{X}^T\mathbf{X}$ is invertible, square, and symmetric, having the same nullspace as \mathbf{X} [4]. This implies that solving $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$, equivalently solving Eq. (3), gives the projection $\mathbf{p} = \mathbf{X}\hat{\boldsymbol{\beta}}$ of \mathbf{y} onto the column space of \mathbf{X} that fully determines the best vector $\hat{\boldsymbol{\beta}}$ [5]. That is to say, the solution of Eq. (3) is given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, which can be easier to solve compared with solving Eq. (3). Hence, $\hat{\boldsymbol{\beta}}$ minimizes $Q = \|\mathbf{e}\|^2 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$ with residual satisfying $\mathbf{X}^T\mathbf{e} = \mathbf{0}$. After replacing $\hat{\boldsymbol{\beta}}$ with coefficients in our econometric model, we can proceed to do hypothesis testing and make predictions as the future steps in econometric analysis.

REFERENCES

- [1] I. MILLER AND M. MILLER, *John E. Freund's Mathematical Statistics with Applications*, Pearson, Boston, MA, USA, 8 ed., 2014.
- [2] J. S. PETROVIC, *Advanced Calculus*, CRC Press, Boca Raton, FL, USA, 2 ed., 2020.
- [3] J. H. STOCK AND M. W. WATSON, *Introduction to Econometrics*, Pearson, New York, NY, USA, 4 ed., 2019.
- [4] G. STRANG, *Linear Algebra and Its Applications*, Thomson Brooks/Cole, Belmont, CA, USA, 4 ed., 2006.
- [5] G. STRANG, *Introduction to Linear Algebra*, Wellesley-Cambridge Press, Wellesley, MA, USA, 5 ed., 2016.
- [6] A. C. TAMHANE AND D. D. DUNLOP, *Statistics and Data Analysis: From Elementary to Intermediate*, Prentice Hall, Upper Saddle River, NJ, USA, 2000.
- [7] J. M. WOOLDRIDGE, *Introductory Econometrics: a Modern Approach*, Cengage, Boston, MA, USA, 7 ed., 2020.