ELEMENTARY MATRIX MAGIC

In my introductory course on mathematical economics, I no longer cover matrix algebra in lecture or ask students to master the techniques for the exams. However, many students have requested that some coverage of matrix methods be included. I used to spend three weeks on it, but eliminated coverage because use of the matrix tools in the economic literature continues to be relatively limited, and indeed in other classes at GMU. Matrix algebra would be better covered in the second math econ course. Since we have not taught the second math economic course for several years, I thought that some condensed coverage could be managed in a hand out.

There are a variety of economic and econometric analyses where matrix algebra and matrix calculus continues to be used. (1) Matrix algebra can be used to solve linear demand systems for a price vector that would clear all markets simultaneously. (2) Matrix forms of the first and second order conditions can be used to characterize a function's maxima and minima, and to determine the concavity of the objective functions and or constraints in cases where many control variables are involved. (3) Most statistical methods in economics have been developed for linear estimators which can be very easily written in matrix/vector forms. (4) Comparative statics of models with more than one first order condition.

I. The Essentials of Matrix Algebra

A. A matrix is a compact way of representing a series of numbers, or several series of numbers.

i. For example, suppose there are three variables: X1, X2, X3, and you have four observations of each. One can denote matrix X as:

ii. $$X = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \\ 4 & 8 & 12 \end{vmatrix}$$

iii. The values of X1 are in the *first column* X1 = 1, 2, 3, 4; those of X2= 2, 4, 6, 8 are in the second column, and those of X3 = 3, 6, 9, 12 are in the third.

iv. This representation is most useful in cases where each *row* represents a point described by a specific combination of X1, X2 and X3. The first point in the illustration would be (1, 2, 3) the second would be (2, 4, 6) , the third (3, 6, 9) and the forth (4, 8, 12).

v. Such a representation is, of course, very useful when one collects data about some event, or time period, for several variables: money supply, national income, interest rates, etc.

B. Each element of a matrix is given an address based on its row and column. The convention is (row, column) or RC as in my own initials.

i. In the matrix above, x_{13} represents the number (or other expression) that is in the first row and third column of the matrix (here, x_{13} =3).

ii. x_{43} is the number or expression in the fourth row and third column (x_{43}= 12).

C. There are a variety of operations and concepts that are very useful in matrix algebra:

i. Square Matrix: a matrix that has the same number of rows and columns

ii. Identity Matrix: a square matrix that has "1" down the diagonal and zeros every where else (every I_{jj} = 1, and every I_{jk} = 0 for j <> k)

iii. Matrix addition: two matrices can be added together if they have the same dimensions (the same number of rows and the same number of columns). The sum of two such matrices is simply the matrix whose elements is made up of the sum of the elements in each respective cell of the matriced added. That is: A + B = C where $c_{ij} = a_{ij} + b_{ij}$

iv. Scalar Matrix Multiplication: Multiplying any matrix by a "scalar" (e. g. a single real number) yields a matrix whose elements are the elements or the original matrix multiplied by the scalar. (Let k be a scalar and A be a matrix, then kA = B implies that $b_{ij} = ka_{ij}$)

v. Matrix Multiplication: is a bit more complicated convention than the previous ideas: Matrix multiplication multiplies the rows of the first matrix by the columns of the second, (again it is RC) and adds up these number or expressions to form elements of the new matrix. Note that matrix multiplication requires the first matrix to have the same number of columns and the second matrix has rows.

a. (So if A is an "m by n" matrix and then B must be a "n by x" matrix for multiplication to be possible. The new matrix will be a "m by x" matrix, that is to say it will have the same number of rows as the first matrix has columns and the same number of rows as the second matrix has rows)

b. Let C = AB then $c_{kj} = \sum_{i=1}^{n} a_{ki} * b_i$

c. It is clear that in most cases, matrix multiplication is best left to computers, since there are many opportunities to make minor arithmetic errors, but it is none-the-less useful to perform one or two by hand. For example

$$let; A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} and; B = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}; then; AxB = 2+8+18 = 28 \equiv C$$

d. Note that A is a 1x3 matrix, and B is a 3x1 matrix, the result C is a 1x1 matrix.

e. Note, also, that the product of a matrix and the appropriately sized Identity matrix is always the original matrix. AI = A

vi. The **inverse** of a matrix A is denoted A^{-1} and is the matrix that A can be "post" multiplied by to obtain the identity matrix. AxA^{-1}= I.

a. There are a variety of requirements for the a matrix to have an inverse, for example it has to be a square matrix, and has to have at least some non zero elements.

b. If a square matrix does not have an inverse it is *singular*.

c. For example, if the rows (columns) of a matrix are linear combinations of other rows (columns) the matrix will be singular.

d. Inverses, although cumbersome to calculate, are very useful in performing matrix algebra. (For example suppose you know that BA = C and you want to solve for B. Post-multiplying by the inverse of A yields BAA^{-1}= BI = B = CA^{-1})

vii. The **transpose** of matrix A is the matrix that you get by switching the rows and columns of A. Let A^T be the transpose of a then, $a^T_{ij} = a_{ji}$.

a. Note that one can always multiply A by its transpose since the transpose will have the same number of columns as A has rows. (That is if A is a NxM matrix then A^T will be a MxN matrix.)

b. Multiplying a matrix by its transpose, AA^T , is roughly the "equivalent" of squaring the original matrix.

viii. The **determinant** of a matrix, $|A|$, is a unique scalar that is computed by multiplying various element pairs and adding those products up *in a specific pattern*. The determinant of a two by two matrix is easy to calculate, $|A| = a_{11}a_{22} - a_{12}a_{21}$

a. For example if $A = \begin{vmatrix} 3 & 1 \\ 3 & 7 \end{vmatrix}$ then $|A| = (3)(7) - (1)(3) = 18$

b. (Evaluating large matrices is more cumbersome, and, cases beyond 3x3 matrices are best left to mathematical computer packages like Mathematica, Maple, Gause, MathCad etc. See Chaing, pg 96-7, for the Laplace expansion method which is fairly easy to apply to a 3x3 matrix..)

II. Extensions of Matrix Algebra Used in the Mathematics of Optimization

i. A **Hessian matrix** is a square matrix formed by placing the second derivatives of a function in matrix form. Second derivatives of all control variables will lie along the diagonal and cross partials in the off diagonal positions. (see Chaing pg 333)

a. The Hessian matrix is widely used to determine or characterize the concavity or convexity of an objective function with more than one control variable.

b. For example if $Z = f(X_1, X_2)$ the Hessian matrix is $\begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$ where f_{11} is the second partial derivative with respect to X_1 and F_{12} is the second cross derivative e. g. what you get when you differentiate Z with respect to X_1 and then differentiate that result with respect to X_2.

c. Z will be strictly concave if $f_{11} < 0$ and $f_{11}f_{22}-f_{12}f_{21} > 0$, the latter is the determinant of the Hessian.

d. (The Hessian matrix is said to be "**negative definite**" in this case.)

e. Hessians and bordered Hessians (used for Lagrangian second order conditions) are not widely used outside of graduate micro text books any more, as the convention now is to assume that the functions of interest have Concave or Convex shapes as "necessary, " "convenient," or seems plausible for the case at hand.

A. Cramer's rule is a very useful method for solving a simultaneous system of equations for one of the variables of interest.

i. Cramer's rule is used for the multivariate version of the implicit function differentiation rule. It is allows one to compute derivatives (comparative statics) of implicit functions that characterize solutions to optimization problems with more than one control variable. (Lecture 4 focused on the implicit function theorem.)

ii. **Cramer's rule** is described in detail by Chaing on page 108/9, and on 210/2 as applied in the implicit function theorem.

iii. Given an equation $Ax = d$ where A is an $n \times n$ matrix, and X is a nx1 vector, x can be solved for as: $x = A^{-1}d$ (which is obtained by ordinary linear algebra: premultiplying both sides by A^{-1}).

iv. The solution for a single element of the X vector can be computed via Cramer's rule (see Chaing pg 109) as follows:

v. $x_j = \left[\frac{1}{|A|}\right] \begin{bmatrix} a_{11} & \dots & d_1 & a_{1n} \\ \dots & .. & d_2 & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & \dots & d_n & a_{nn} \end{bmatrix}$ where the d vector has been inserted in the J^{th} column of

A. The solution for x_j is the determinant of the new matrix divided by the determint of the original matrix.

vi. Obviously, computing this becomes a task for computers once n exceeds 3. But for many small economics models with 2x2 or 3x3 "A" matrices, Cramer's rule provides nice mathematical solutions for variables and/or derivatives of economic interest.

vii. Of course, there are other models in which a system of linear equations has to be solved for a single variable. Cramer's rule is often applied in macro economics and occassionally in industrial organization (Harberger) and in computable general equilibrium models.

III. Application to Econometrics

A. Derivation of the Regression Estimator: The most widely used bit of matrix algebra in economics characterizes the line, Y = XB, which minimizes the sum of squared residuals through N k+1-dimensional points (observations): $\mathbf{B} = (X^TX)^{-1} X^TY$

i. The OLS estimater is calculated by applying ordinary calculus to matrices.

a. The linear relationship of interest can be written as $\mathbf{\Psi} = X\mathbf{B}$ where $\mathbf{\Psi}$ is a tx1 vector of estimated "Y" observations, **B** is a kx1 vector of coefficients, and X is a txk vector of the k independent variable observations.

b. The residuals generated by an estimate of **B** are simply: $R = Y - \mathbf{\Psi} = Y - \mathbf{B}X$ and the squared residuals are in matrix notation
$R^2 = (Y - \mathbf{XB})(Y - \mathbf{XB})^T = YY^T - 2X^TYB^T + XBX^TB^T$

ii. Choose B to minimize *R*.

a. This can be done by differentiating R^2 with respect to B and setting the result equal to zero. $R^2_B = -2X^TY + 2X^TXB = 0$ at $\mathbf{B}^*$

b. This implies that $2X^TXB = 2X^TY$

c. Dividing by 2 and pre multiplying by X^TX inverse yields $B = (X^TX)^{-1}X^TY$

d. Note that this estimator will minimize the sum of squared residuals regardless of the error distribution, because that is how B is derived. (*The same estimator can be derived in other ways using somewhat different assumptions.*)