THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

ECONOMICS

ON THE DIFFERENTIATION OF A LOG-LIKELIHOOD FUNCTION USING MATRIX CALCULUS

by

Darrell A Turkington

Business School
The University of Western Australia

DISCUSSION PAPER 11.06

ON THE DIFFERENTIATION OF A LOG-LIKELIHOOD FUNCTION USING MATRIX CALCULUS

Darrell A Turkington

Department of Economics

Business School

The University of Western Australia

DISCUSSION PAPER 11.06

ABSTRACT

Simple theorems based on a mathematical property of $\frac{\partial \text{vec}Y}{\partial \text{vec}X}$ provide powerful tools for obtaining matrix calculus results. By way of illustration, new results are obtained for matrix derivatives involving vecA, vechA, $\bar{v}$(A) and vecX where X is a symmetric matrix. The analysis explains exactly how a log-likelihood function should be differentiated using matrix calculus.

Keywords: Matrix Derivatives, Vecs, Log-Likelihood Function

Corresponding author:-

Darrell Turkington, Department of Economics, University of Western Australia, 35 Stirling Highway, Crawley, Perth, Western Australia 6009, Australia.
Tel: + 61-08-6488-2880; Fax: + 04-6488-1016
Email Address: Darrell.Turkington@uwa.edu.au

Date April 7 2011

1. Introduction[1]

In a recent article Magnus (2010) advocates $\frac{\partial \mathrm{vec}Y}{\partial \mathrm{vec}X'}$ as the concept of the matrix derivative to use for the differentiation of a matrix Y with respect to the matrix X. He does this on the grounds of mathematical correctness and mathematical convenience. As far as the first ground is concerned, Parring (1992) shows that the three concepts of the matrix derivative of Y with respect to X, commonly used in the literature, all qualify as permissible mathematical operators. It depends on the matrix or vector space you are working with and how this space is normed. (For relationships that exist between these three concepts see Turkington (2007)). However, I would agree with Magnus on the second ground. Certainly $\frac{\partial \mathrm{vec}Y}{\partial \mathrm{vec}X'}$ is the most convenient concept of a matrix derivative to work with out of the concepts he considers. Having said that, an argument can be made for working with $\frac{\partial \mathrm{vec}Y}{\partial \mathrm{vec}X}$, which is just the transpose of the concept advocated by Magnus. But this transpose makes a difference in terms of mathematical convenience.

In the next section it will be explained as succinctly as possible why from a practitioner's point of view $\frac{\partial \mathrm{vec}Y}{\partial \mathrm{vec}X}$ has certain mathematical advantages over all the other concepts of matrix derivatives used in the literature. Simple theorems involving this concept will be presented, whose proofs are almost trivial. However, taken together, these theorems provide powerful tools for deriving matrix calculus results.

This is demonstrated both in section 3 and section 4 of the article. In section 3 use is made of these theorems to derive results, some of which are new, for derivatives involving vecA, vechA and $\bar{\mathrm{v}}(A)$ where A is a square matrix. These three vectors are of interest to statisticians. In section 4 the same theorems are used to derive an easy method for obtaining derivatives involving the vecs of symmetric matrices from known matrix calculus results. Again this is of interest to statisticians as covariance matrices appear in log-likelihood functions.

Section 5 brings the analysis together and demonstrates how matrix calculus should be used to correctly differentiate a log-likelihood function. The last section is reserved for a brief conclusion.

[1] The author would like to acknowledge the kind hospitality of Nuffield College, Oxford. Work that lead to this paper was commenced when he was an academic visitor there in 2010. He would like to thank Helmut Lutkepohl for acting as a sounding board for problems whose solutions also contributed to the paper. He would also like to thank his colleague Les Jennings of the Department of Mathematics at the University of Western Australia for helpful suggestions.

2. Theorems involving $\partial \mathrm{vec}Y / \partial \mathrm{vec}X$

The main advantage of using this concept of a matrix derivative can be put succinctly in a few lines. Consider a $m \times 1$ vector $y = (y_1 \quad \dots \quad y_m)'$, a $n \times 1$ vector $x = (x_1 \quad \dots \quad x_n)'$ and ℓ any scalar function. Then using $\frac{\partial \,\mathrm{vec}\, Y}{\partial \,\mathrm{vec}\, X}$ as our concept of a matrix derivative $\frac{\partial \ell}{\partial y} = \left(\frac{\partial \ell}{\partial y_1} \quad \dots \quad \frac{\partial \ell}{\partial y_m} \right)'$ and $\frac{\partial \ell}{\partial x} = \left(\frac{\partial \ell}{\partial x_1} \quad \dots \quad \frac{\partial \ell}{\partial x_n} \right)'$. Suppose $y = Ax$, where A is a matrix of constants, that is, the elements of A are not scalar functions of x. Then,

$$\frac{\partial \ell}{\partial y} = A \frac{\partial \ell}{\partial x},$$

so for this important case *the same functional relation exists between* $\frac{\partial \ell}{\partial y}$ *and* $\frac{\partial \ell}{\partial x}$ *as between y and x.*

Several of the following theorems involving $\frac{\partial \mathrm{vec}Y}{\partial \mathrm{vec}X}$ arise from this notion.

Theorem 1

Let x be a $n \times 1$ vector whose elements are distinct. Then

$$\frac{\partial x}{\partial x} = I_n .$$

Proof

Clearly

$$\frac{\partial x}{\partial x} = \left(\frac{\partial x_1}{\partial x} \quad \dots \quad \frac{\partial x_n}{\partial x} \right) = \left(e_1^n \quad \dots \quad e_n^n \right) = I_n \quad ,$$

where e_j^n is the j^{th} column of I_n.

Theorem 2

Suppose x and y are two column vectors and $y = Ax$ where A is a matrix of constants.

Let z be a column vector. Then

$$\frac{\partial z}{\partial y} = A\frac{\partial z}{\partial x} \quad .$$

<u>Proof</u>

We know that for any scalar ℓ,

$$\frac{\partial \ell}{\partial y} = A\frac{\partial \ell}{\partial x} \quad .$$

Write

$$z = \begin{pmatrix} z_1 & \dots & z_p \end{pmatrix}' \quad .$$

Then

$$\frac{\partial z}{\partial y} = \begin{pmatrix} \frac{\partial z_1}{\partial y} & \dots & \frac{\partial z_p}{\partial y} \end{pmatrix} = \begin{pmatrix} A\frac{\partial z_1}{\partial x} & \dots & A\frac{\partial z_p}{\partial x} \end{pmatrix} = A\begin{pmatrix} \frac{\partial z_1}{\partial x} & \dots & \frac{\partial z_p}{\partial x} \end{pmatrix} = A\frac{\partial z}{\partial x} \quad .$$

* * * *

<u>Theorem 3</u>

Suppose x and y are two column vectors such that

$$y = Ax$$

where A is a matrix of constants and the elements of x are distinct. Then

$$\frac{\partial y}{\partial x} = \left(\frac{\partial x}{\partial y}\right)' \quad .$$

<u>Proof</u>

Using the advocated concept of a matrix derivative $\frac{\partial y}{\partial x} = A'$. But from theorem 2

$$\frac{\partial z}{\partial y} = A\frac{\partial z}{\partial x}$$

for any vector z. Taking $z = x$ gives

$$\frac{\partial x}{\partial y} = A\frac{\partial x}{\partial x}$$

and as the elements of x are distinct by theorem 1, the derivative $\frac{\partial x}{\partial x}$ is the identity matrix so

$$\frac{\partial x}{\partial y} = A = \left(\frac{\partial y}{\partial x}\right)' .$$

Taking transposes gives the result.

* * * *

In using the recommended concept of a matrix derivative a backward chain rule applies (see Turkington (2004)) which is just the transpose of the chain rule reported by Magnus (see Magnus (2010)). That is, if y is a vector function of u and u is a vector function of x, so $y = y(u(x))$ then

$$\partial y = \frac{\partial u}{\partial x}\frac{\partial y}{\partial u} .$$

Using this result gives us the following theorem.

<u>Theorem 4</u>

For any vectors x and y

$$\frac{\partial y}{\partial x} = \frac{\partial x}{\partial x}\frac{\partial y}{\partial x} .$$

<u>Proof</u>

Write $y = y(x(x))$ and apply the backward chain rule.

* * * *

3. Theorems concerning derivatives involving vecA, vechA and $\bar{v}(A)$

Let $A = \{a_{ij}\}$ be a $n \times n$ matrix and partition A into its columns so $A = \begin{pmatrix} a_1 & \dots & a_n \end{pmatrix}$ where a_j is the j^{th} column of A for $j = 1, \dots, n$. Then vecA is the $n^2 \times 1$ vector given by

$\text{vec}A = \begin{pmatrix} a_1' & \dots & a_n' \end{pmatrix}'$, that is, to form vecA we stack the columns of A underneath each other. VechA is the $\frac{1}{2}n(n+1) \times 1$ vector given by

$$\text{vech}A = \begin{pmatrix} a_{11} & \dots & a_{n1} & a_{22} & \dots & a_{n2} & \dots & a_{nn} \end{pmatrix}' \quad .$$

That is, to form vechA we stack the elements of A on and below the main diagonal one underneath the other. The vector $\bar{v}(A)$ is the is the $\frac{1}{2}n(n-1) \times 1$ vector given by

$$\bar{v}(A) = \begin{pmatrix} a_{11} & \dots & a_{n1} & a_{32} & \dots & a_{n2} & \dots & a_{nn-1} \end{pmatrix}' .$$

That is, we form $\bar{v}(A)$ by stacking the elements of A below the main diagonal, one beneath the other. These vectors are important for statisticians and econometricians. If A is a covariance matrix then vecA contains the variances and covariances but with the covariances duplicated. The vector vechA contains the variances and covariances without duplication and $\bar{v}(A)$ contains the covariances without the variances.

Regardless as to whether A is symmetric or not, the elements in vechA and $\bar{v}(A)$ are distinct. The elements in vecA are distinct provided A is not symmetric. If A is symmetric the elements of vecA are not distinct. So from theorem 1 we have

$$\frac{\partial \text{vech}A}{\partial \text{vech}A} = I_{\frac{1}{2}n(n+1)} \qquad \text{for all A}$$

$$\frac{\partial \bar{v}(A)}{\partial \bar{v}(A)} = I_{\frac{1}{2}n(n-1)} \qquad \text{for all A}$$

$$\frac{\partial \text{vec}A}{\partial \text{vec}A} = I_{n^2} \qquad \text{provided A is not symmetric.}$$

What $\frac{\partial \text{vec}A}{\partial \text{vec}A}$ is in the case where A symmetric is discussed in section 4.

Regardless of the nature of A, it is well known that there exist $\frac{1}{2}n(n+1)\times n^2$ and $\frac{1}{2}n(n-1)\times n^2$ zero-one matrices L_n and $\bar{L}_n$ respectively, such that

$$L_n \text{vec}A = \text{vech}A$$

and

$$\bar{L}_n \text{vec}A = \bar{v}(A) \quad .$$

If A is symmetric then

$$N_n \text{vec}A = \text{vec}A$$

where $N_n = \frac{1}{2}(I_{n^2} + K_{nn})$ and K_{nn} is a commutation matrix, so for this case

$$L_n N_n \text{vec}A = \text{vech}A$$

and

$$\bar{L}_n N \text{vec}A = \bar{v}(A).$$

The matrices L_nN_n and $\bar{L}_n N_n$ are not zero-one matrices. However, along with L_n and $\bar{L}_n$, they form a group of matrices known as elimination matrices. The difference in the operation of L_n and L_nN_n on vecA is this. The matrix L_n chooses a_{ij} for i > j for vechA directly from vecA, whereas L_nN_n recognises that A is symmetric and forms a_{ij} for vechA using $a_{ij} = \frac{a_{ij} + a_{ji}}{2}$.

For special cases there exist zero-one matrices called duplication matrices which take us back from vechA and $\bar{v}(A)$ to vecA. If A is symmetric there exists a $n^2 \times \frac{1}{2}n(n+1)$ zero-one matrix D_n such that

$$D_n \text{vech}A = \text{vec}A \, .$$

If A is strictly lower triangular then

$$\bar{L}_n{}' \bar{v}(A) = \text{vec}A \, .$$

For an excellent discussion of the special matrices associated with vecA, vechA and $\bar{v}(A)$ and their properties see Magnus (1988).

Consider ℓ any scalar function. Then the same relationships exist between $\frac{\partial \ell}{\partial \text{vec}A}$, $\frac{\partial \ell}{\partial \text{vech}A}$ and $\frac{\partial \ell}{\partial \bar{v}(A)}$ as exist between vecA, vechA and $\bar{v}(A)$ respectively.

Thus for general A

$$\frac{\partial \ell}{\partial \text{vech}A} = L_n \frac{\partial \ell}{\partial \text{vec}A}$$

$$\frac{\partial \ell}{\partial \bar{v}(A)} = \bar{L}_n \frac{\partial \ell}{\partial \text{vec}A} \quad .$$

For symmetric A

$$\frac{\partial \ell}{\partial \text{vech}A} = L_n N_n \frac{\partial \ell}{\partial \text{vec}A} \tag{1}$$

$$\frac{\partial \ell}{\partial \bar{v}(A)} = \bar{L}_n N_n \frac{\partial \ell}{\partial \text{vec}A} \quad .$$

$$\frac{\partial \ell}{\partial \text{vec}A} = D_n \frac{\partial \ell}{\partial \text{vech}A}$$

and for A a strictly lower triangular matrix

$$\frac{\partial \ell}{\partial \text{vec}A} = \bar{L}_n' \frac{\partial \ell}{\partial \bar{v}(A)} \quad .$$

Using the theorems of section 3 we can prove the following results.

<u>Theorem 5</u>

$$\frac{\partial \text{vec}A}{\partial \text{vech}A} = D_n' \qquad \text{if A is symmetric}$$

$$\frac{\partial \text{vec}A}{\partial \text{vech}A} = L_n \qquad \text{if A is not symmetric.}$$

Proof

If A is symmetric $\mathrm{vecA} = D_n \mathrm{vechA}$ and the result follows. For the case where A is not symmetric consider

$$\mathrm{vechA} = L_n \mathrm{vecA} \quad .$$

By theorem 2 we have that for any vector z

$$\frac{\partial z}{\partial \mathrm{vechA}} = L_n \frac{\partial z}{\partial \mathrm{vecA}} \quad .$$

Taking z = vecA gives

$$\frac{\partial \mathrm{vecA}}{\partial \mathrm{vechA}} = L_n \frac{\partial \mathrm{vecA}}{\partial \mathrm{vecA}}$$

and as A is not symmetric the elements of vecA are distinct, so by theorem 1

$$\frac{\partial \mathrm{vecA}}{\partial \mathrm{vecA}} = I_{n^2}$$

and

$$\frac{\partial \mathrm{vecA}}{\partial \mathrm{vechA}} = L_n \, .$$

*　　*　　*　　*

Theorem 6

$$\frac{\partial \mathrm{vechA}}{\partial \mathrm{vecA}} = D_n \qquad \text{if A is symmetric}$$

$$\frac{\partial \mathrm{vechA}}{\partial \mathrm{vecA}} = L_n' \qquad \text{if A is not symmetric.}$$

Proof

A trivial application of theorem 3.

*　　*　　*　　*

The method used in theorem 5 can also be used to quickly derive results about elimination matrices, duplication matrices and the matrix N_n. Consider for example the case where A is a symmetric $n \times n$ matrix so

$$L_n N_n \mathrm{vec}A = \mathrm{vech}A \quad .$$

By theorem 2 for any vector z

$$\frac{\partial z}{\partial \mathrm{vech}A} = L_n N_n \frac{\partial z}{\partial \mathrm{vec}A} \quad .$$

Take z = vechA. Then

$$\frac{\partial \mathrm{vech}A}{\partial \mathrm{vech}A} = L_n N_n \frac{\partial \mathrm{vech}A}{\partial \mathrm{vec}A} = L_n N_n D_n$$

by theorem 6.

But as the elements of vechA are distinct

$$\frac{\partial \mathrm{vech}A}{\partial \mathrm{vech}A} = I_{\frac{1}{2}n(n+1)},$$

so

$$L_n N_n D_n = I_{\frac{1}{2}n(n+1)} \quad . \tag{2}$$

4. Theorems concerning derivatives involving vecX where X is symmetric

Consider X a $n \times n$ symmetric matrix and let x = vecX. Then the elements of x are not distinct and one of the implications of this is that

$$\frac{\partial x}{\partial x} \neq I_{n^2} \quad .$$

Consider the 2×2 case. Then

$$X = \begin{pmatrix} x_{11} & x_{21} \\ x_{21} & x_{22} \end{pmatrix}$$

and $x = \begin{pmatrix} x_{11} & x_{21} & x_{21} & x_{22} \end{pmatrix}'$, so

$$\frac{\partial x}{\partial x} = \begin{pmatrix} \frac{\partial x_{11}}{\partial x} & \frac{\partial x_{21}}{\partial x} & \frac{\partial x_{21}}{\partial x} & \frac{\partial x_{22}}{\partial x} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad .$$

Clearly this matrix is not the identity matrix. What it is, is given by the following theorem whose proof again calls on our results of section 3.

<u>Theorem 7</u>

Let X be a $n \times n$ symmetric matrix. Then

$$\frac{\partial \text{vec}X}{\partial \text{vec}X} = D_n D_n' \quad .$$

<u>Proof</u>

As X is a $n \times n$ symmetric matrix

$$\text{vec}X = D_n \text{vech}X$$

so it follows from theorem 2 that for any vector z

$$\frac{\partial z}{\partial \text{vec}X} = D_n \frac{\partial z}{\partial \text{vech}X} \quad .$$

Take z = vecX so

$$\frac{\partial \text{vec}X}{\partial \text{vec}X} = D_n \frac{\partial \text{vec}X}{\partial \text{vech}X} = D_n D_n' \tag{3}$$

by theorem 5.

*   *   *   *

The fact that in the case where X is a $n \times n$ symmetric matrix $\frac{\partial \text{vec}X}{\partial \text{vec}X} = D_n D_n'$ means that all the usual rules of matrix calculus, regardless of what concept of a matrix derivative one is using, do not apply for vecX where X is symmetric. However theorem 4, coupled with theorem 7, provides a quick and easy method for finding the results for this case using known matrix calculus results.

Consider again x = vecX with X a symmetric matrix. Let $\frac{\phi y}{\phi x}$ denote the matrix derivative we would get if we differentiated y with respect to x using the concept of differentiation advocated but *ignoring the fact that X is a symmetric matrix.* Then the full import of theorem 4 for this case is given by the equation

$$\frac{\partial y}{\partial x} = \frac{\partial x}{\partial x}\frac{\phi y}{\phi x} \quad . \tag{4}$$

Combining Eqs. (3) and (4) give the following theorem.

<u>Theorem 8</u>

Consider y = y(x) with x = vecX and X is a $n \times n$ symmetric matrix. Let $\frac{\phi y}{\phi x}$ denote the derivative of y with respect to x obtained when we ignore the fact that X is a symmetric matrix. Then

$$\frac{\partial y}{\partial x} = D_n D_n' \frac{\phi y}{\phi x} \quad . \tag{5}$$

*   *   *   *

A few examples will suffice to illustrate the use of this theorem. (For the rules referred to in these examples see Turkington (2004), Lutkepohl (1996) or Magnus and Neudecker (1999)).

For x with distinct elements and A a matrix of constants we know that

$$\frac{\partial x'Ax}{\partial x} = 2(A + A')x \quad .$$

It follows that when x = vecX and X is a $n \times n$ symmetric matrix

$$\frac{\partial x'Ax}{\partial x} = 2D_n D_n'(A + A')x \quad .$$

For X non-singular but non-symmetric matrix

$$\frac{\partial |X|}{\partial \text{vec}X} = |X| \text{vec}(X^{-1})'$$

so for X non-singular but symmetric

$$\frac{\partial |X|}{\partial \text{vec}X} = |X| D_n D_n{}' \text{vec}X^{-1} \quad .$$

For X a $n \times n$ non-symmetric matrix, A and B matrices of constants

$$\frac{\partial \text{vec}AXB}{\partial \text{vec}X} = B \otimes A'$$

so for X a $n \times n$ symmetric matrix

$$\frac{\partial \text{vec}AXB}{\partial \text{vec}X} = D_n D_n{}' (B \otimes A') \quad .$$

All results using either $\frac{\partial \text{vec}Y}{\partial \text{vec}X}$ or $\frac{\partial \text{vec}Y}{\partial \text{vec}X'}$ (in which case we have to take transposes) can be adjusted in this way to allow for the case where X is a symmetric matrix.

5. The Matrix Differentiation of a Log-Likelihood Function.

Suppose we are dealing with a statistical model that has a log-likelihood function $\ell(\theta)$ where θ is a vector containing the parameters of the model. Then we can always partition θ as $\theta = (\delta' \, v')'$ where $v = \text{vech}\Sigma$ and Σ is a covariance matrix associated with the model. The problem is that $\ell(\theta)$ is never expressed in terms of v. Rather it is written in terms of Σ. The question then is how do we form $\frac{\partial \ell}{\partial v}$. The results of the previous section allow us to do this. As Σ is a symmetric matrix and assuming it is $n \times n$ we have from theorem 8 that

$$\frac{\partial \ell}{\partial \text{vec}\Sigma} = D_n D_n{}' \frac{\phi \ell}{\phi \text{vec}\Sigma} \quad .$$

But from Eq. (1) we also have

$$\frac{\partial \ell}{\partial v} = L_n N_n \frac{\partial \ell}{\partial \text{vec}\Sigma} \quad ,$$

so

$$\frac{\partial \ell}{\partial v} = L_n N_n D_n D_n{}' \frac{\phi \ell}{\phi \text{vec}\Sigma} = D_n{}' \frac{\phi \ell}{\phi \text{vec}\Sigma}$$

as by Eq. (2) $L_n N_n D_n = I_{\frac{1}{2}n(n+1)}$. Our method then is to differentiate the log likelihood function with respect to $\mathrm{vec}\,\Sigma$ ignoring the fact that Σ is symmetric. Then premultiply the result by $D_n{}'$. Note that from theorem 5, $\frac{\partial \mathrm{vec}\Sigma}{\partial \mathrm{v}} = D_n{}'$ so we could write if we like that

$$\frac{\partial \ell}{\partial \mathrm{v}} = \frac{\partial \mathrm{vec}\Sigma}{\partial \mathrm{v}} \frac{\phi \ell}{\phi \mathrm{vec}\Sigma}$$

which resembles a backward chain rule. This is approach was taken by Turkington (2004).

A simple example illustrates this method. Magnus and Neudecker (1980) consider a sample of size m from a n dimensional distribution of a random vector y with mean vector μ and a positive definite covariance matrix Σ. The parameters of this model are $\theta = (\mu' \mathrm{v}')'$ where $\mathrm{v} = \mathrm{vech}\Sigma$ and the log likelihood function, apart from a constant, is

$$\ell(\theta) = -\frac{1}{2}\mathrm{m}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\Sigma^{-1}\mathrm{Z}$$

where

$$\mathrm{Z} = \sum_{i=1}^{m} (\mathrm{y}_i - \mu)(\mathrm{y}_i - \mu)' \quad .$$

Now

$$\frac{\phi \ell}{\phi \mathrm{vec}\Sigma} = -\frac{1}{2}\mathrm{m}\frac{\phi \log|\Sigma|}{\phi \mathrm{vec}\Sigma} - \frac{1}{2}\frac{\phi}{\phi \mathrm{vec}\Sigma}\mathrm{tr}\Sigma^{-1}\mathrm{Z}$$

and

$$\frac{\phi \log|\Sigma|}{\phi \mathrm{vec}\Sigma} = \mathrm{vec}\Sigma^{-1} \quad .$$

Using the backward chain rule

$$\frac{\phi \mathrm{tr}\Sigma^{-1}\mathrm{Z}}{\phi \mathrm{vec}\Sigma} = \frac{\phi \mathrm{vec}\Sigma^{-1}}{\phi \mathrm{vec}\Sigma}\frac{\phi \mathrm{tr}\Sigma^{-1}\mathrm{Z}}{\phi \mathrm{vec}\Sigma^{-1}} = -(\Sigma^{-1} \otimes \Sigma^{-1})\mathrm{vec}\mathrm{Z}$$

so

$$\frac{\phi \ell}{\phi \mathrm{vec}\Sigma} = -\frac{1}{2}\mathrm{m}\mathrm{vec}\Sigma^{-1} + \frac{1}{2}(\Sigma^{-1} \otimes \Sigma^{-1})\mathrm{vec}\mathrm{Z} = \frac{1}{2}(\Sigma^{-1} \otimes \Sigma^{-1})\mathrm{vec}(\mathrm{Z} - \mathrm{m}\Sigma)$$

and

$$\frac{\partial \ell}{\partial v} = \frac{1}{2} D_n' (\Sigma^{-1} \otimes \Sigma^{-1}) \mathrm{vec}(Z - m\Sigma)$$

which is the same result Magnus and Neudecker obtained using differentials.

Conclusion

It goes without saying that the correct use of matrix calculus to differentiate a log likelihood function is of great interest to a statistician who wants to apply classical statistical procedures centred around the likelihood function. Once the method is understood using matrix calculus in these procedures, it is no more difficult than the use of ordinary calculus in every day mathematical problems. Moreover, there is no need to first resort to matrix differentials as advocated by Magnus and Neudecker (1999). Rather, using rules which are generalizations of the product rule and chain rule of ordinary calculus, one can easily derive the derivatives required in classical statistics using either $\frac{\partial \mathrm{vec} Y}{\partial \mathrm{vec} X'}$, as advocated by Magnus, or $\frac{\partial \mathrm{vec} Y}{\partial \mathrm{vec} X}$.

References

Lutkepohl, H., (1996). *Handbook of Matrices*. Wiley, New York.

Magnus, J., (1988). *Linear Structures*. Oxford University Press, New York.

Magnus, J., (2010) On The Concept of Matrix Derivative. *Journal of Multivariate Analysis,* **101**, 2200-6

Magnus, J, and Neudecker, H., (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd ed. John Wiley and Sons, Chichester, New York (paperback).

Magnus, J and Neudecker, H., (1980), The Elimination Matrix: Some Lemmas and Applications, *SIAM Journal on Algebraic and Discrete Methods*, **4**, 422-49

Parring, A.M., (1992) About the Concept of the Matrix Derivative. *Linear Algebraic and its Application*, **176**, 223-35

Turkington, D.A., (2004) *Matrix Calculus and Zero-One Matrices, Statistical and Econometric Applications*. Cambridge University Press, New York, (paperback).

Turkington, D.A., (2007) Generalized Vec and Rvec Operators and Different Concepts of Matrix Differentiation, Working Paper, Department of Economics, University of Western Australia.

ECONOMICS DISCUSSION PAPERS 2009		
DP NUMBER	**AUTHORS**	**TITLE**
09.01	Le, A.T.	ENTRY INTO UNIVERSITY: ARE THE CHILDREN OF IMMIGRANTS DISADVANTAGED?
09.02	Wu, Y.	CHINA’S CAPITAL STOCK SERIES BY REGION AND SECTOR
09.03	Chen, M.H.	UNDERSTANDING WORLD COMMODITY PRICES RETURNS, VOLATILITY AND DIVERSIFACATION
09.04	Velagic, R.	UWA DISCUSSION PAPERS IN ECONOMICS: THE FIRST 650
09.05	McLure, M.	ROYALTIES FOR REGIONS: ACCOUNTABILITY AND SUSTAINABILITY
09.06	Chen, A. and Groenewold, N.	REDUCING REGIONAL DISPARITIES IN CHINA: AN EVALUATION OF ALTERNATIVE POLICIES
09.07	Groenewold, N. and Hagger, A.	THE REGIONAL ECONOMIC EFFECTS OF IMMIGRATION: SIMULATION RESULTS FROM A SMALL CGE MODEL.
09.08	Clements, K. and Chen, D.	AFFLUENCE AND FOOD: SIMPLE WAY TO INFER INCOMES
09.09	Clements, K. and Maesepp, M.	A SELF-REFLECTIVE INVERSE DEMAND SYSTEM
09.10	Jones, C.	MEASURING WESTERN AUSTRALIAN HOUSE PRICES: METHODS AND IMPLICATIONS
09.11	Siddique, M.A.B.	WESTERN AUSTRALIA-JAPAN MINING CO-OPERATION: AN HISTORICAL OVERVIEW
09.12	Weber, E.J.	PRE-INDUSTRIAL BIMETALLISM: THE INDEX COIN HYPTHESIS
09.13	McLure, M.	PARETO AND PIGOU ON OPHELIMITY, UTILITY AND WELFARE: IMPLICATIONS FOR PUBLIC FINANCE
09.14	Weber, E.J.	WILFRED EDWARD GRAHAM SALTER: THE MERITS OF A CLASSICAL ECONOMIC EDUCATION
09.15	Tyers, R. and Huang, L.	COMBATING CHINA’S EXPORT CONTRACTION: FISCAL EXPANSION OR ACCELERATED INDUSTRIAL REFORM
09.16	Zweifel, P., Plaff, D. and Kühn, J.	IS REGULATING THE SOLVENCY OF BANKS COUNTER-PRODUCTIVE?
09.17	Clements, K.	THE PHD CONFERENCE REACHES ADULTHOOD
09.18	McLure, M.	THIRTY YEARS OF ECONOMICS: UWA AND THE WA BRANCH OF THE ECONOMIC SOCIETY FROM 1963 TO 1992
09.19	Harris, R.G. and Robertson, P.	TRADE, WAGES AND SKILL ACCUMULATION IN THE EMERGING GIANTS
09.20	Peng, J., Cui, J., Qin, F. and Groenewold, N.	STOCK PRICES AND THE MACRO ECONOMY IN CHINA
09.21	Chen, A. and Groenewold, N.	REGIONAL EQUALITY AND NATIONAL DEVELOPMENT IN CHINA: IS THERE A TRADE-OFF?

ECONOMICS DISCUSSION PAPERS 2010		
DP NUMBER	**AUTHORS**	**TITLE**
10.01	Hendry, D.F.	RESEARCH AND THE ACADEMIC: A TALE OF TWO CULTURES
10.02	McLure, M., Turkington, D. and Weber, E.J.	A CONVERSATION WITH ARNOLD ZELLNER
10.03	Butler, D.J., Burbank, V.K. and Chisholm, J.S.	THE FRAMES BEHIND THE GAMES: PLAYER'S PERCEPTIONS OF PRISONER'S DILEMMA, CHICKEN, DICTATOR, AND ULTIMATUM GAMES
10.04	Harris, R.G., Robertson, P.E. and Xu, J.Y.	THE INTERNATIONAL EFFECTS OF CHINA'S GROWTH, TRADE AND EDUCATION BOOMS
10.05	Clements, K.W., Mongey, S. and Si, J.	THE DYNAMICS OF NEW RESOURCE PROJECTS A PROGRESS REPORT
10.06	Costello, G., Fraser, P. and Groenewold, N.	HOUSE PRICES, NON-FUNDAMENTAL COMPONENTS AND INTERSTATE SPILLOVERS: THE AUSTRALIAN EXPERIENCE
10.07	Clements, K.	REPORT OF THE 2009 PHD CONFERENCE IN ECONOMICS AND BUSINESS
10.08	Robertson, P.E.	INVESTMENT LED GROWTH IN INDIA: HINDU FACT OR MYTHOLOGY?
10.09	Fu, D., Wu, Y. and Tang, Y.	THE EFFECTS OF OWNERSHIP STRUCTURE AND INDUSTRY CHARACTERISTICS ON EXPORT PERFORMANCE
10.10	Wu, Y.	INNOVATION AND ECONOMIC GROWTH IN CHINA
10.11	Stephens, B.J.	THE DETERMINANTS OF LABOUR FORCE STATUS AMONG INDIGENOUS AUSTRALIANS
10.12	Davies, M.	FINANCING THE BURRA BURRA MINES, SOUTH AUSTRALIA: LIQUIDITY PROBLEMS AND RESOLUTIONS
10.13	Tyers, R. and Zhang, Y.	APPRECIATING THE RENMINBI
10.14	Clements, K.W., Lan, Y. and Seah, S.P.	THE BIG MAC INDEX TWO DECADES ON AN EVALUATION OF BURGERNOMICS
10.15	Robertson, P.E. and Xu, J.Y.	IN CHINA'S WAKE: HAS ASIA GAINED FROM CHINA'S GROWTH?
10.16	Clements, K.W. and Izan, H.Y.	THE PAY PARITY MATRIX: A TOOL FOR ANALYSING THE STRUCTURE OF PAY
10.17	Gao, G.	WORLD FOOD DEMAND
10.18	Wu, Y.	INDIGENOUS INNOVATION IN CHINA: IMPLICATIONS FOR SUSTAINABLE GROWTH
10.19	Robertson, P.E.	DECIPHERING THE HINDU GROWTH EPIC

10.20	Stevens, G.	RESERVE BANK OF AUSTRALIA-THE ROLE OF FINANCE
10.21	Widmer, P.K., Zweifel, P. and Farsi, M.	ACCOUNTING FOR HETEROGENEITY IN THE MEASUREMENT OF HOSPITAL PERFORMANCE
10.22	McLure, M.	ASSESSMENTS OF A. C. PIGOU'S FELLOWSHIP THESES
10.23	Poon, A.R.	THE ECONOMICS OF NONLINEAR PRICING: EVIDENCE FROM AIRFARES AND GROCERY PRICES
10.24	Halperin, D.	FORECASTING METALS RETURNS: A BAYESIAN DECISION THEORETIC APPROACH
10.25	Clements, K.W. and Si. J.	THE INVESTMENT PROJECT PIPELINE: COST ESCALATION, LEAD-TIME, SUCCESS, FAILURE AND SPEED
10.26	Chen, A., Groenewold, N. and Hagger, A.J.	THE REGIONAL ECONOMIC EFFECTS OF A REDUCTION IN CARBON EMISSIONS
10.27	Siddique, A., Selvanathan, E.A. and Selvanathan, S.	REMITTANCES AND ECONOMIC GROWTH: EMPIRICAL EVIDENCE FROM BANGLADESH, INDIA AND SRI LANKA

ECONOMICS DISCUSSION PAPERS 2011		
DP NUMBER	**AUTHORS**	**TITLE**
11.01	Robertson, P.E.	DEEP IMPACT: CHINA AND THE WORLD ECONOMY
11.02	Kang, C. and Lee, S.H.	BEING KNOWLEDGEABLE OR SOCIABLE? DIFFERENCES IN RELATIVE IMPORTANCE OF COGNITIVE AND NON-COGNITIVE SKILLS
11.03	Turkington, D.	DIFFERENT CONCEPTS OF MATRIX CALCULUS
11.04	Golley, J. and Tyers, R.	CONTRASTING GIANTS: DEMOGRAPHIC CHANGE AND ECONOMIC PERFORMANCE IN CHINA AND INDIA
11.05	Collins, J., Baer, B. and Weber, E.J.	ECONOMIC GROWTH AND EVOLUTION: PARENTAL PREFERENCE FOR QUALITY AND QUANTITY OF OFFSPRING
11.06	Turkington, D.	ON THE DIFFERENTIATION OF THE LOG LIKELIHOOD FUNCTION USING MATRIX CALCULUS
11.07	Groenewold, N. and Paterson, J.E.H.	STOCK PRICES AND EXCHANGE RATES IN AUSTRALIA: ARE COMMODITY PRICES THE MISSING LINK?
11.08	Chen, A. and Groenewold, N.	REDUCING REGIONAL DISPARITIES IN CHINA: IS INVESTMENT ALLOCATION POLICY EFFECTIVE?
11.09	Williams, A., Birch, E. and Hancock , P.	THE IMPACT OF ON-LINE LECTURE RECORDINGS ON STUDENT PERFORMANCE
11.10	Pawley, J. and Weber, E.J.	INVESTMENT AND TECHNICAL PROGRESS IN THE G7 COUNTRIES AND AUSTRALIA