Published paper

Adcock, C.J. (2007) *Extensions of Stein's lemma for the skew-normal distribution*, Communications in Statistics - Theory and Methods, 36 (9-12), pp. 1661-1671
http://dx.doi.org/10.1080/03610920601126084

An Extension Of Stein's Lemma For The Skew Normal Distribution

By

C J Adcock
The University of Sheffield, UK

Summary

When two random variables have a bivariate normal distribution, Stein's lemma, Stein(1973, 1981), provides, under certain regularity conditions, an expression for the covariance of the first variable with a function of the second. This result plays an important role in the modern theory of finance. When returns on financial assets have a multivariate normal distribution, Kallberg and Ziemba(1983) show that all well behaved utility functions lead to a point on Markowitz' mean-variance efficient frontier, Markowitz(1952). The implication of this result is that, under normality, it is pointless to seek a better utility function. However, it is well known that returns on financial assets are not normally distributed. They exhibit both skewness and kurtosis. As well an appropriate model for the multivariate probability distribution of asset returns, the general issue of the choice of utility function is therefore open. This short paper describes an extension of Stein's lemma for the multivariate skew normal distribution, which was introduced by Azzalini and Dalla Valle(1996). The extension of the lemma shows that, under this distribution, investors who are expected utility maximisers will be located on a single mean-variance-skewness efficient surface, regardless of their choice of utility function.

Keywords: Fubini's theorem, multivariate skew normal distribution, portfolio selection, Stein's lemma, utility function.

Address for correspondence:

C J Adcock
School of Management
University of Sheffield
9, Mappin Street
Sheffield, S1 4DT
UK

Tel: +44 (0)114 222 3402
Fax: +44 (0)114 222 3458
Email: c.j.adcock@shef.ac.uk

1. Introduction

When two random variables, X and Y say, have a bivariate normal distribution, Stein's lemma, Stein(1973), states that, under certain regularity conditions on a function g(), cov{X, g(Y)} = cov(X,Y)E{g'(Y)}. In a later paper, Stein(1981), he provides a proof of an underlying result, namely cov{X, g(X)} = var(X)E{g'(X)}.

These results play an important role in modern finance. The theory of portfolio selection developed by Harry Markowitz, see for example Markowitz(1952) and Markowitz(1987), assumes that investors minimise the variance of portfolio returns subject to achieving a given target expected return. Varying the target expected return generates a set of mean-variance efficient portfolios whose expected returns and variances lie on a parabolic curve known universally in finance as the efficient frontier. This method is equivalent to assuming that investors maximise the expected utility when the utility function used is quadratic in portfolio return. The use of quadratic utility functions in finance, however, is criticized, see for example Pratt(1964), on the grounds that there must be circumstances in which an investor appears to prefer less wealth to more wealth. This criticism, coupled with the natural desire to achieve higher portfolio returns or lower portfolio volatility or both, has lead to the search for what might be called better utility functions. The role that Stein's lemma plays is as follows. When returns on financial assets have a multivariate normal distribution, the consequence of the lemma is that all well behaved utility functions, U(R) say where R denotes the return on a portfolio, will lead to a point on Markowitz' mean-variance efficient frontier. In this context, well behaved means that U() is differentiable at least twice, U'() $\geq$ 0 and U''() $\leq$ 0, and that the expected value of U() exists. This result, which is described in more detail in Kallberg and Ziemba(1983) and which is also credited independently to Rubinstein(1973), means that it is pointless to seek a better utility function. The only issue, when returns are normal, is the choice of location on the efficient frontier. However, it is well known that returns on financial assets are not normally distributed. They exhibit both skewness and kurtosis. As well as the choice of an appropriate model for the multivariate probability distribution of asset returns, the issue of the choice of utility function is therefore open.

The purpose of this short paper is to describe an extension of Stein's lemma for the multivariate skew normal. This multivariate distribution, which was first introduced by Azzalini and Dalla Valle(1996), is an attractive model for applications in finance. Adcock and Shutes(2001) describe some of the theoretical aspects of the model when it is used for portfolio selection and related applications. In a recent working paper, Harvey et al(2002) report an empirical study of portfolio selection for American stocks which uses the distribution. As a coherent multivariate probability distribution, it is suitable for portfolio selection in the presence of skewness and offers a number of different insights into the sources of expected return and risk in a portfolio. It is also a parsimonious model for skewness; a not insignificant advantage if one is considering a portfolio of several hundred stocks. As is shown in section 4, the multivariate skew normal distribution admits a result which is an extension of Stein's lemma. The implication of this is that when returns follow the multivariate skew normal distribution there is a single efficient surface. All investors who are expected utility maximisers will be located on this surface, regardless of their choice of utility function.

The structure of this paper is as follows. Section 2 summaries the multivariate skew normal distribution and such properties that are required in the rest of the paper. Section 3 present some preliminary results that are used in section 4, which contains the extension to the Stein's lemma. Section 5 shows how this is applied to portfolio selection under the multivariate skew normal distribution.

2. The Multivariate Skew Normal Distribution

The multivariate skew normal distribution was introduced by Azzalini and Dalla Valle(1996). It is an extension of the univariate skew normal distribution which was originally due to Roberts(1966) and, separately, O'Hagan and Leonard(1976) and which was developed in articles by Azzalini(1985, 1986). The standard form is obtained by considering the distribution of a random vector, X say, which is defined as:

$$X = Y + \lambda U.$$

The vector Y has a full rank multivariate normal distribution with mean vector μ and variance-covariance matrix Σ. The scalar variable U, which is independent of Y, has a standard normal distribution that is truncated below at zero. The vector λ is a vector of skewness parameters, which may take any real values. For applications in finance, a modification of this distribution is employed, as reported in Adcock and Shutes(2001). The vectors X, Y and λ are defined as above. The scalar variable U has a normal distribution with mean τ and variance 1 truncated below at zero. This modification generates a richer family of probability distributions. In particular, it gives more flexibility in modelling skewness and kurtosis. The idea of adding a skewness shock to a multivariate normally distributed vector is not new. It is suggested in Simaan(1993), which predates Adcock and Shutes. The probability distribution of X is multivariate skew normal with parameters μ, Σ , λ and τ, denoted as $X \sim MSN(\mu, \Sigma, \lambda, \tau)$. The probability density function of this distribution is:

$$f_R(x) = n(x; \mu + \lambda\tau, \Sigma + \lambda\lambda^T)\frac{\Phi(v)}{\Phi(\tau)}, \tag{1}$$

where:

$$v = \alpha\left(\tau + \lambda^T\Sigma^{-1}(x - \mu)\right), \ \alpha = (1 + \boldsymbol{\lambda}^T\Sigma^{-1}\boldsymbol{\lambda})^{-\frac{1}{2}}, \tag{2}$$

and where $\Phi(x)$ is the standard normal distribution function evaluated at x. The notation $n(x; \omega, W)$ denotes the probability density function, evaluated at x, of a multivariate normal distribution with mean vector ω and variance covariance matrix W.

This density function is essentially Azzalini and Dalla Valle's(1996) result with a change of notation and generalization to accommodate a non-zero value of τ. The distribution of any sub-vector of x, including the scalar variable x_i, is of the same form, based upon the corresponding sub-vectors of μ and λ and sub-matrix of Σ.

The vector of expected values and variance covariance matrix of X are, respectively:

$$\begin{aligned} E(X) &= \mu + \lambda\{\tau + \xi_1(\tau)\} = \gamma, \text{say} \\ \operatorname{var}(X) &= \Sigma + \lambda\lambda^T\{1 + \xi_2(\tau)\} = \Omega, \text{say} \end{aligned} , \qquad (3)$$

where the function $\xi_k()$ is defined as:

$$\xi_k(x) = \frac{\partial^k \log \Phi(x)}{\partial x^k}, k = 1,2,...$$

3. Preliminaries

The results that follow are concerned with the scalar random variable V which is defined at (2) above and with expected values of functions of V of the form $h(v) = k(v)\xi_1(v)$. The results use the notation defined above. First:

$$\xi_1(v) f_R(x) = \alpha\xi_1(\tau) n(x;\mu,\Sigma) \qquad (4)$$

The proof is by re-arrangement of the left hand side.

Secondly, using equation (4) gives:

$$E[k(V)\xi_1(V)] = \alpha\xi_1(\tau) E_Z[k\{\alpha\tau + (1-\alpha^2)^{\frac{1}{2}} Z\}] \qquad (5)$$

where E_Z denotes expectation over the standard normal distribution. To prove this, the integrand in the expectation may be written as:

$$\alpha\xi_1(\tau) k(v) n(x;\mu,\Sigma)$$

Under the $N(\mu, \Sigma)$ distribution, the scalar variable:

$$(\lambda^T \Sigma^{-1} \lambda)^{-\frac{1}{2}} \lambda^T \Sigma^{-1} (x - \mu),$$

is distributed as N(0,1). Noting that V may be written as:

$$V = \alpha\tau + (1-\alpha^2)^{\frac{1}{2}} Z\,,$$

completes the proof.

The following standard result is reproduced here for convenience. It is essentially a result described in Azzalini and Dalla Valle(1996), with a modification to accommodate the use of non-zero value of τ. If X is partitioned into sub-vectors X_1 and X_2, with corresponding partitions for λ and Σ, the conditional distribution of X_1 given $X_2 = x_2$ is MSN(μ_C, Σ_C, λ_C, τ_C) where:

$$\tau_C = \alpha_C\{\tau + \lambda_2^T \Sigma_{22}^{-1}(x_2 - \mu_2)\}$$

$$\lambda_C = \alpha_C(\lambda_1 - \Sigma_{12}\Sigma_{22}^{-1}\lambda_2)$$

$$\Sigma_C = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\mu_C = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\alpha_C = (1 + \lambda_2^T \Sigma_{22}^{-1}\lambda_2)^{-\frac{1}{2}},$$

and the conditional expected value of X_1 given $X_2 = x_2$ is:

$$E[X_1 \mid x_2] = \mu_1 + \lambda_1\tau + \Psi(x_2 - \mu_2 - \lambda_2\tau) + \lambda_C\xi_1(\tau_C)$$

$$\Psi = (\Sigma_{12} + \lambda_1\lambda_2^T)(\Sigma_{22} + \lambda_2\lambda_2^T)^{-1}.$$

When X_1 and X_2 are both scalars, the expected value of X_1 given $X_2 = x_2$ is:

$$E[X_1 \mid x_2] = m_1 + l_1 t + y_C\left(x_2 - m_2 - l_2 t\right) + l_C x_1(t_C)\,, \qquad (6)$$

where:

$$\tau_C = \alpha_C\{\tau + (\lambda_2 / \sigma_2^2)(x_2 - \mu_2)\}$$

$$\lambda_C = \alpha_C\{\lambda_1 - (\sigma_{12} / \sigma_2^2)\lambda_2\} \tag{7}$$

$$\psi_C = \alpha_C^2(\sigma_{12} + \lambda_1\lambda_2) / \sigma_2^2$$

$$\alpha_C = \{1 + (\lambda_2^2 / \sigma_2^2)\}^{-\frac{1}{2}}.$$

Finally, for the univariate case, the first derivative of the probability density function with respect to x_2 is:

$$f'(x_2) = -\frac{\alpha_C^2(x_2 - \mu_2 - \lambda_2\tau)}{\sigma_2^2} f(x_2) + \frac{\lambda_2\alpha_C^2}{\sigma_2^2} n(x_2; \mu_2, \sigma_2^2) \tag{8}$$

4. Extension of Stein's Lemma

When X_1 and X_2 have a bivariate skew normal distribution and assuming that the expected values and derivatives exist, the covariance between X_1 and a function g() of X_2 is given by:

$$\text{cov}\{X_1, g(X_2)\} = (\sigma_{12} + \lambda_1\lambda_2)E\{g'(X_2)\} + \lambda_1\xi_1(\tau)[E_N\{g(X_2)\} - E\{g(X_2)\}], \tag{9}$$

where E_N denotes expectations taken over the distribution $N(\mu,\sigma^2)$.

The proof is in two stages. The first is to take expectations over the distribution of X_1 given X_2 and then to take expectations over the distribution of X_2. After the first stage, the required covariance may be written using (6) as:

$$\begin{aligned} \text{cov}\{X_1, g(X_2)\} &= E_{X_2}[\{\psi_C(X_2 - \mu_2 - \lambda_2\tau) + \lambda_C\xi_1(\tau_C) - \lambda_1\xi_1(\tau)\}g(X_2)] \\ &= T_1 + T_2 + T_3, \text{ say.} \end{aligned} \tag{10}$$

The expected value of T_3 is:

$$E(T_3) = -\lambda_1\xi_1(\tau)E\{g(X_2)\}.$$

For term T_2, the integrand in the expected value integral is:

$$\lambda_C\xi_1(\tau_C)g(x_2)f(x_2),$$

where f() denotes the probability density function of X_2. Using equation (4) this is:

$$\alpha_C \lambda_C \xi_1(\tau) g(x_2) n(x_2, \mu_2, \sigma_2^2).$$

Hence, on using the definitions in equation (7):

$$E(T_2) = \alpha_C^2 \{\lambda_1 - (\sigma_{12} / \sigma_2^2)\lambda_2\} \xi_1(\tau) E_N\{g(X_2)\},$$

where, as above, E_N denotes expectation over the normal distribution $N(\mu,\sigma^2)$. For term T_1, integration by parts is used. The integral in the first part vanishes. The integrand in the second part is:

$$\psi_C(\sigma_2^2 + \lambda_2^2) n(x_2; \mu + \lambda\tau, \sigma_2^2 + \lambda_2^2) \left\{ g'(x_x) \frac{\Phi(\tau_C)}{\Phi(\tau)} + g(x_2) \frac{\alpha_C \lambda_2}{\sigma_2^2} \frac{\phi(\tau_C)}{\Phi(\tau)} \right\}$$

Denoting the two terms as T_{11} and T_{12} and using the definition of ψ_C at equation (7) gives:

$$E(T_{11}) = (\sigma_{12} + \lambda_1\lambda_2) E\{g'(X_2)\}.$$

Using equations (4) and (7) , the integrand for term T_{12} is:

$$\frac{(\sigma_{12} + \lambda_1\lambda_2)\lambda_2 \alpha_C^2 \xi_1(\tau)}{\sigma_2^2} g(x_2) n(x_2; \mu_2, \sigma_2^2),$$

and so:

$$E\{T_{12}\} = \frac{(\sigma_{12} + \lambda_1\lambda_2)\lambda_2 \alpha_C^2 \xi_1(\tau)}{\sigma_2^2} E_N\{g(x_2)\}.$$

Combining this with $E\{T_2\}$ above gives the result at equation (9). When λ is a zero vector this reduces to Stein's result.

Stein(1981) presents a more rigorous proof of the key component of his lemma. If X_2 is $N(\mu_2, \sigma_2^2)$ then lemma 1 of Stein(1981) states that $\mathrm{cov}(X_2, g(X_2)) = \sigma_2^2 E\{g'(X_2)\}$. This result is proved using Fubini's theorem and requires in essence that the function g() is differentiable and that $E\{|g(X_2)|\}$ is bounded. For the multivariate skew normal distribution, the analogous result is:

$$\mathrm{cov}\{X_2, g(X_2)\} = (\sigma_2^2 + \lambda_2^2) E\{g'(X_2)\} + \lambda_2 \xi_1(\tau)[E_N\{g(X_2)\} - E\{g(X_2)\}]. \tag{11}$$

The proof of this follows the same steps in the proof of lemma 1 of Stein(1981). It may be shown that the expected value of $g'(X_2)$ is given by:

$$E\{g'(X_2)\} = -\int_{-\infty}^{\infty} f'(x_2)\{g(x_2) - g(0)\}dx_2 .$$

Using the expression for f'() at equation (7) gives the result at (11) Application of (11) to relevant terms on the right hand side of equation (10) gives (9).

An interesting corollary arises from the fact that this result holds for all values of σ_2. When σ_2 = 0, the variable X_2 is, apart from a shift of location, proportional to an N(τ,1) variable truncated below at zero. Calling this variable U, as in section 2, the corollary is $\mathrm{cov}\{U, g(U)\} = E[g'(U)] + \xi_1(\tau)[E_N\{g(U)\} - E\{g(U)\}]$.

5. Application To Portfolio Selection

To exemplify the lemma, consider portfolio selection. A portfolio is a set of investment weight or proportions $\{w_i\}$, i = 1(1)n, defined such that an investor invests $100w_i$% of wealth in asset i. It is conventionally assumed that the weights sum to one. If the return on asset i is denoted by the random variable R_i, i = 1(1)n, then return on the portfolio with weights $\{w_i\}$ is:

$$R_p = \sum_{i=1}^{n} w_i R_i = w^T R ,$$

where w and R are vectors of length n containing the investment weights and asset returns, respectively. For portfolio selection, the investor conventionally chooses the portfolio weights to maximise the expected utility of portfolio return. For a general utility function, $U(R_p)$ say, the expected utility is

$$\int \ldots \int U(r_p) f(r) dr = \Xi(w), \text{say}.$$

Following Kallberg and Ziemba(1983), the investor who is an expected utility maximiser solves:

$$\max_w \Xi(w) - \eta(u^T w - 1),$$

where, in the equation above, u is a vector of length n containing ones and η is the Lagrange multiplier of the budget constraint. Ignoring this constraint for simplicity, the first order conditions for the weight for asset i are:

$$\frac{\partial \Xi}{\partial w_i} = \int \ldots \int r_i U'(r_p) f(r) dr .$$

This may be written as:

$$E(R_i)E\{U'(R_p)\} + \int \ldots \int \{r_i - E[R_i]\}U'(r_p)f(r)dr = \text{cov}\{R_i, U'(R_p)\} + E(R_i)E\{U'(R_p)\}$$

When returns follow the multivariate skew normal distribution, application of the extension to Stein's lemma gives the vector of first order conditions for all assets:

$$\gamma E(U') + \left(\Sigma + \lambda\lambda^T\right) w\, E(U'') + \lambda\xi_1(\tau)\{E_N(U') - E(U')\}.$$

This may be re-expressed in terms of Ω, the VC matrix of returns as:

$$\gamma E(U') + \Omega w\, E(U'') + \lambda[\xi_1(\tau)\{E_N(U') - E(U')\} - \xi_2(\tau)E(U'')].$$

This equation is the same for all investors, except for the scalar quantities which are functions of certain expected values of U' and U''. When all elements of the vector λ are equal to zero, asset returns have a multivariate normal distribution and Kallberg and Ziemba's(1983) result is obtained. That is, the portfolios of all investors who are expected utility maximisers are located on Markowitz' mean-variance efficient frontier. Under the multivariate skew normal distribution, investors' portfolios are located on the mean variance-skewness-efficient surface.

References

Adcock, C. J. and Shutes, K. (2001) Portfolio Selection Based on the Multivariate-Skew Normal Distribution, in Financial Modelling, A Skulimowski(Editor), Krakow: Progress & Business Publishers.

Azzalini, A. (1985) A Class of Distributions Which Includes The Normal Ones, Scand. J. Statist, **12**, 171-178.

Azzalini, A. (1986) Further Results on a Class of Distributions Which Includes The Normal Ones, Statistica, **46**, 199-208.

Azzalini, A. & Dalla Valle. A. (1996) The Multivariate Skew-normal Distribution, Biometrika, **83**, 715-726.

Kallberg, J. G. and Ziemba, W. T. (1983) Comparison of Alternative Functions in Portfolio Selection Problems, management Science, 11, p1257 – 1276.

O'Hagan, A. and Leonard, T. (1976) Bayes Estimation Subject to Uncertainty About Parameter Constraints, Biometrika, **63**, 201-203.

Markowitz, H. M.(1952) Portfolio Selection, Journal of Finance, 7, No. 1 p77 – 91.

Markowitz, H. M. (1987) Mean Variance Analysis in Portfolio Choice and Capital Markets, Blackwell, Oxford.

Pratt, J. W. (1964) Risk Aversion in the Small and in the Large, Econometrica, 32, p122 – 136.

Roberts, C. (1966) A Correlation Model Useful In The Study of Twins, Journal of the American Statistical Association, **61**, p1184-1190.

Rubinstein, M. E.(1973) A Comparative Statics Analysis of Risk Premiums, Journal of Business, **46**, p605-615.

Simaan, Y. (1993) Portfolio Selection and Asset Pricing- Three Parameter Framework. Management Science, 1983. **39**(5), 568-587.

Stein, C. M. (1973) Estimation of the Mean of a Multivariate Normal Distribution, Proceedings of the Symposium on Asymptotic Statistics.

Stein, C. M. (1981)Estimation of the Mean of a Multivariate Normal Distribution, Annals of Statistics, **9**, p1135-1151.