

# A Method for Fitting of Schaefer Model with Autoregressive of Order One

N. OKENDRO SINGH\*

Division of Forecasting and Econometrics Techniques, Indian Agricultural Statistics Research Institute  
Pusa Campus, Library Avenue, New Delhi – 110 012 (INDIA)

## Abstract

Schaefer model is one of the most popularly used surplus production models in fisheries to estimate maximum sustainable yield and the corresponding optimum fishing effort. When we employ nonlinear estimation procedures for estimating the parameters, high parameter correlations are generally observed, which is undesirable. Moreover, catch-effort fisheries data are collected during a certain time period and hence data points are generally correlated among themselves. In the present paper, utility of expected value parameters to make parameter correlations low is highlighted. An explicit expression giving catch-effort relationship for Schaefer model with autoregressive of order one using expected value parameters is also developed. This is illustrated with an example, considering the serially correlated catch-effort data observed from Gobindsagar reservoir.

## Introduction

The abundance of fish stock for a particular water body is a function of interactions between environmental factors and the fish stock properties. The stock tends to stabilize at a particular set of environmental conditions (Gulland, 1977). When the surplus production is not harvested, at the level of maximum fish stock size the addition of recruitment and growth to the stock is just sufficient to compensate for natural mortality and hence, surplus production will be equal to zero (Haddon, 2001). This implies that fishing plans can be expressed in terms of surplus production; they are very flexible and have different variations. Schaefer's surplus production model (Schaefer, 1954) and its extensions dominated research in production models for fisheries. The reason for their widespread use is that, these models have modest data requirements and the input data for these models are a simple time series of catch and effort from fishery, which is readily available for most fisheries. Further, when we are dealing with catch-effort fisheries data, which are observed over continuous time periods, the data points are generally correlated among themselves. Also, Prajneshu and Ravichandran (2003); Prajneshu and Kandala (2005) emphasized the importance of reparameterization of the parameters for fitting of nonlinear surplus production models in fisheries to reduce large correlations among the parameter estimates. Thus, by examining the presence of autocorrelation in the observations, in the present study a method of fitting a Schaefer model with

---

\*Corresponding author. E-mail address: nosingh2000@yahoo.com

serially correlated error structure using expected value parameters is proposed which is illustrated with an example, considering the catch-effort data observed from Gobindsagar reservoir, India.

## Material and Methods

A surplus production model, as described by Schaefer (1954), facilitates estimation of MSY and the optimum fishing effort for harvesting the MSY ( $E_{MSY}$ ). The equilibrium Schaefer model is given by

$$C_t = KE_t \left( 1 - \frac{E_t}{r} \right), \quad \dots \quad (1)$$

where  $C_t$  and  $E_t$  are catch and effort at time  $t$ ;  $r$  is the intrinsic growth rate;  $K$  is the carrying capacity and correspondingly,  $MSY = rK/4$  and  $E_{MSY} = r/2$ .

As we are dealing with time-series data, it is, therefore, required to check for the validity of the above model by examining the independency assumption of error term. The Durbin-Watson test can be employed for the said purpose and is based on the assumption that the errors ( $\varepsilon_t$ 's) follow autoregressive of order one. The corresponding test statistic 'd' is defined as

$$d = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}, \quad 0 \leq d \leq 4. \quad \dots \quad (2)$$

A statistic 'd' value ranges between 0 and 4. A value of 'd' near 2 indicates little autocorrelation; a value toward 0 indicates positive autocorrelation while a value toward 4 indicates negative autocorrelation.

To handle a situation when there is an evidence for the presence of autocorrelation, an autoregressive (AR) error term  $\varepsilon_t$  of order one is added to the right hand side of above equation (1):

$$\varepsilon_t = \Phi \varepsilon_{t-1} + u_t; |\Phi| < 1, \quad \dots \quad (3)$$

where  $u_t$ 's are independently and normally distributed with zero mean and constant variance and  $\Phi$  denotes the autoregressive parameter. Incorporating an AR(1) additive error structure, the Schaefer model becomes:

$$C_t = KE_t \left( 1 - \frac{E_t}{r} \right) + \Phi \varepsilon_{t-1} + u_t. \quad \dots \quad (4)$$

### Partial Reparameterization of Schaefer Model with AR(1)

When we deal with nonlinear estimation procedures for estimating the parameters, high parameter correlation may be detected which is indeed undesirable. Ratkowsky (1990) suggested using expected-value parameters for choosing the values of the explanatory variable in such a way so as to make the parameter correlations low. Expected value-parameters should fall within the observed range of the data and not correspond to asymptotes or extrapolations outside the data range because outside the range of the observed data is less efficient. Further, expected-value parameters can be advantageously used, as they are nearly unbiased, normally distributed and with minimum variance estimators.

To be convenient for mathematical notations, the above equation (4) is rewritten in the following form:

$$C = KE \left( 1 - \frac{E}{r} \right) + \Phi \varepsilon. \quad \dots \quad (5)$$

It may not always be necessary to replace all the parameters of a model by expected-value parameters; however, only the offending parameters can be replaced. Thus, to obtain expected-value parameters for  $r$  and  $K$  of the above equation (5), we need to choose values  $E_1$  and  $E_2$  of the explanatory variable  $E$ , within the observed range of  $E$  and correspondingly the values of  $\varepsilon_1$  and  $\varepsilon_2$  of  $\varepsilon$  are to be selected. Then, we get the expected values from equation (5) as follows:

$$C_1 = KE_1 \left( 1 - \frac{E_1}{r} \right) + \Phi' \varepsilon_1, \quad \dots \quad (6)$$

and

$$C_2 = KE_2 \left( 1 - \frac{E_2}{r} \right) + \Phi' \varepsilon_2. \quad \dots \quad (7)$$

From equation (6),  $K$  can be obtained as:

$$K = \frac{(C_1 - \Phi' \varepsilon_1)r}{E_1(r - E_1)} \quad \dots \quad (8)$$

Again from equation (7), the expression for  $r$  is given as:

$$r = \frac{E_2}{\left[ 1 - \frac{(C_2 - \Phi' \varepsilon_2)}{KE_2} \right]} \quad \text{----- (9)}$$

Substituting the expression for K from equation (8) in equation (9) and solving, we get

$$r = \frac{[(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2]}{[(C_1 - \Phi' \varepsilon_1)E_2 - (C_2 - \Phi' \varepsilon_2)E_1]} \quad \text{----- (10)}$$

Also, equations (8) and (10) provide

$$K = \frac{[(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2]}{E_1 E_2 (E_2 - E_1)} \quad \text{----- (11)}$$

Since the parameters r and K are non-negative values, the following conditions should be satisfied: (i)  $(C_1 - \Phi' \varepsilon_1)E_2 > (C_2 - \Phi' \varepsilon_2)E_1$ , and (ii)  $E_2 > E_1$ .

Using equations (10) and (11), equation (5) becomes,

$$C = \left[ \left\{ EE_2(E_2 - E)(C_1 - \Phi' \varepsilon_1) - EE_1(E_1 - E)(C_2 - \Phi' \varepsilon_2) \right\} / E_1 E_2 (E_2 - E_1) \right] + \Phi \varepsilon \quad \text{----- (12)}$$

where  $\Phi'$  is the estimated value of  $\Phi$  obtained by fitting the above equation (4).

Now, this process has eliminated the original parameters r and K, replacing these with new expected-value parameters. Different parameterizations of the same basic model will produce the same goodness-of-fit and the same fitted values. However, they may differ greatly in their estimation behavior (Ratkowsky, 1990). The unknown parameters and autoregressive parameter in the above nonlinear models are estimated using Levenberg-Marquardt method (Seber and Wild, 1989).

### ***Measures of Model Adequacy***

It is generally assessed by the following statistics:

(i) Root Mean Square Error,

$$RMSE = \left[ \frac{\sum_{t=1}^n (C_t - \hat{C}_t)^2}{n} \right]^{\frac{1}{2}},$$

(ii) Mean Absolute Error,

$$MAE = \frac{\sum_{t=1}^n |C_t - \hat{C}_t|}{n}, \quad t = 1, 2, \dots, n,$$

where  $n$  is the number of observations;  $\hat{C}_t$  is the predicted fish catch at time  $t$ .

The better fitted-model will have the lower value of the above statistics. Further, residual analysis is required to check the assumptions made for the model to be developed. Thus, independence assumption of the residuals needs to be tested. To test the independence assumption of residuals run test procedure is available in the literature (Ratkowsky, 1990). However, the normality assumption is not so stringent for selecting nonlinear models because their residuals may not follow normal distribution.

To illustrate the above methodology, data on fishing effort and its corresponding catch in Gobindsagar reservoir during 1974-75 to 1989-90 observed by Kaushal et al. (2006) is considered. The Statistical Package for Social Sciences (SPSS) 12.0 version has been used for fitting of the above models. Different sets of initial parameter values were tried to meet the global convergence criterion for best fitting of the nonlinear models.

## Results and Discussion

The MSY and its corresponding optimum efforts of different forms of Schaefer model are computed and the results of the fitted models are presented in Table 1. The MSY value (in tons) estimated by Schaefer model is 808 and its corresponding optimum efforts (number of gill nets) is 1,406 respectively. The randomness assumption does not follow since the run test  $|Z|$  value (2.303) is greater than 1.96 of normal distribution at 5% level of significance. However, the normality assumption regarding the error term in catch is met for the model since Shapiro-Wilk test p-value of the fitted model is 0.123. Further, Durbin-Watson statistic has been calculated and the statistic value is toward zero and hence positive autocorrelation is suspected. The Schaefer model is refitted incorporating the AR(1) error structure and the results are again shown in Table 1. Further, run test and Shapiro-Wilk test are employed to the residuals of catch. Here, the run test  $|Z|$  value and Shapiro-Wilk test p-value of the refitted model indicate that both the randomness and normality

assumptions are satisfied. The Durbin-Watson statistic value calculated from the refitted model is very closed to 2 i.e. the presence of autocorrelation is negligible. Also, a significant improvement in RMSE and MAE values is seen in the refitted model as compared to the original Schaefer model. However, the correlation matrix given in Table 2 shows that the extreme value of correlation coefficient between parameters 'r' and 'K' i.e.  $\rho(r, K) = -0.979$ . It indicates that the two parameters are not estimated independently, while the values of  $\rho(r, \Phi)$  and  $\rho(K, \Phi)$  are acceptable. For getting a possible solution to this, it has been attempted to fit equation (12), which is derived from equation (5) with expected-value parameters for 'r' and 'K', since they are the possible offending parameters of the model, while the parameter  $\Phi$ , is being kept unchanged.

**Table 1.** Summary statistics for fitting of different forms of Schaefer model to catch-effort data observed from Gobindsagar reservoir.

	Schaefer Model	Schaefer Model with AR(1)	Reparameterization of Schaefer Model with AR(1)
<b>Parameters</b>			
K	1.149 *(0.425)	1.449 (0.379)	-
r	2811.288 (3041.811)	1758.330 (669.831)	-
C <sub>1</sub>	-	-	530.661 (46.355)
C <sub>2</sub>	-	-	735.204 (56.723)
$\Phi$	-	0.630 (0.247)	0.630 (0.247)
<b>Statistics</b>			
MSY (in tons)	808	637	637
E <sub>MSY</sub> (no. of gill nets)	1406	879	879
Durbin-Watson statistic	0.843	1.753	1.753
<b>Model adequacy</b>			
RMSE	140.292	114.455	114.455
MAE	123.802	96.767	96.767
<b>Residual analysis</b>			
Run test $( Z )$	2.303	0.259	0.259
Shapiro-Wilk test p-value	0.123	0.231	0.231

\*The values given in parentheses are the corresponding asymptotic standard errors.

A pair  $E_1=594$  and  $E_2=813$  correspondingly,  $\varepsilon_1 = -62.59$  and  $\varepsilon_2 = 161.41$ , give the best result in terms of least correlation coefficient and the required conditions for 'r' and 'K' to be positive are also satisfied.

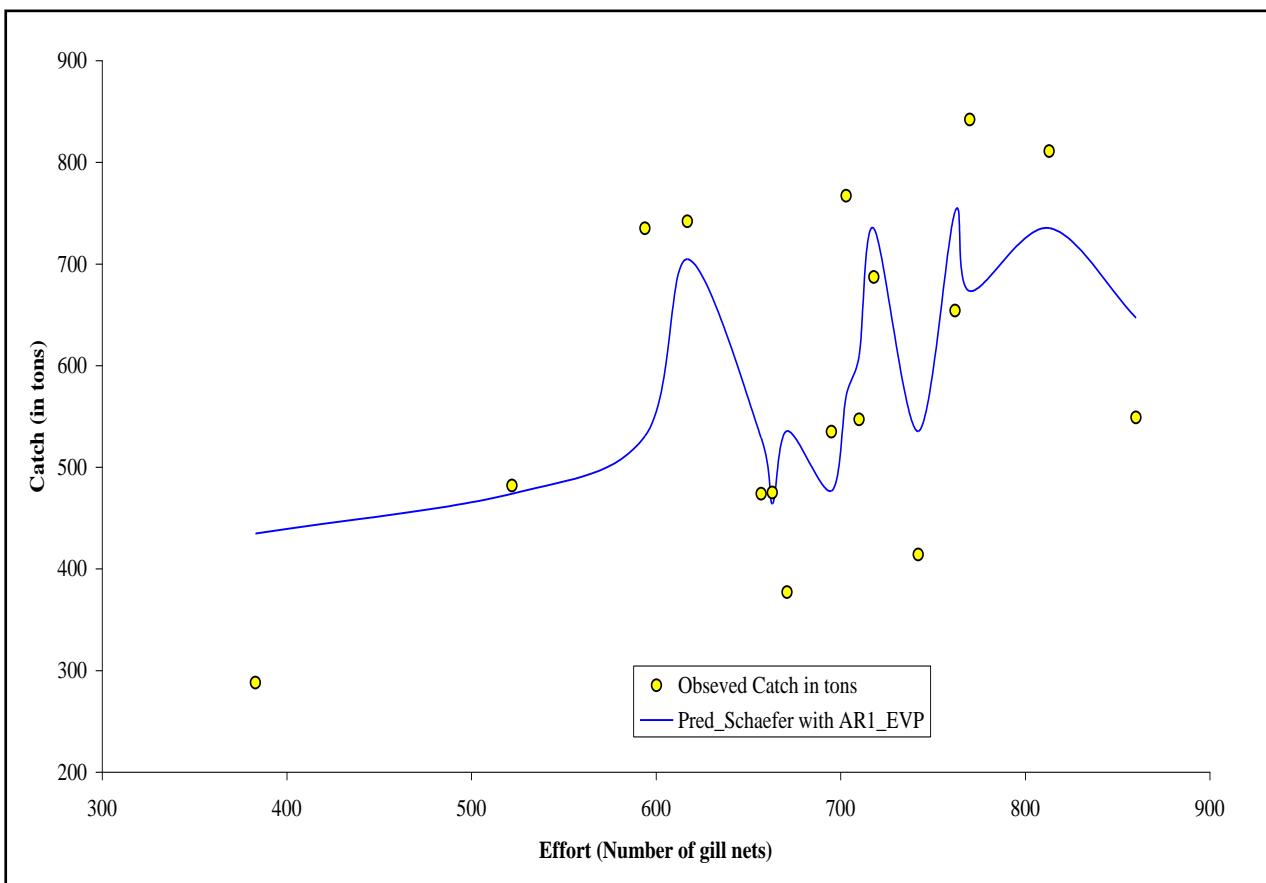
**Table 2.** Asymptotic correlation matrix of the parameter estimates after fitting of Schaefer model with AR(1)

Parameter(s)	K	r	$\Phi$
K	1.000		
r	-0.979	1.000	
$\Phi$	0.311	-0.305	1.000

In practice,  $E_1$  and  $E_2$  are being chosen in such a way that they are not close to each other. The corresponding values of C i.e.,  $C_1=735$  and  $C_2=811$  are taken as initial values for computation of the final estimates of the parameters  $C_1$  and  $C_2$ . The parameter estimates are presented in Table 1. Now, the correlation coefficients are well acceptable as the correlation coefficients between the parameters are very low as presented in Table 3. Thus, we can say that the parameters are estimated nearly independently. The values of RMSE and MAE calculated from the original model of equation (4) and from the transformed model with the expected value parameters of equation (12) have remained the same as given in Table 1. The graph of fitted model along with observed catch is also depicted in Fig 1. A perusal of the estimates of MSY for different forms of Schaefer model reveal that a simple Schaefer model has slightly over-estimated the MSY and optimum effort values as compared to the values of MSY (637 tons) and optimum effort (879 number of gill nets) estimated by the Schaefer model with AR(1).

**Table 3.** Asymptotic correlation matrix of the parameter estimates after fitting of Schaefer model with AR(1) using expected-value parameters.

Parameter(s)	$C_1$	$C_2$	$\Phi$
$C_1$	1.000		
$C_2$	-0.215	1.000	
$\Phi$	0.270	-0.209	1.000



**Fig. 1.** Fitted Schaefer model with AR(1) to the dataset of catch-effort observed from Gobindsagar reservoir using expected-value parameters.

## Acknowledgements

The author is thankful to Dr. P. C. Mahanta, Director, Directorate of Coldwater Fisheries Research, Bhimtal for his support and guidance. He also sincerely expresses his gratitude to Dr. Prajneshu, Head of Biometrics Division, IASRI, New Delhi for providing a fruitful scientific discussion especially on expected-value parameters, which leads to the present study.

## References

Gulland, J.A. 1977. The problems of population dynamics and contemporary fisheries management. In: Fish population dynamics: the implication for management, Vol. XI (ed. J.A. Gulland), pp. 383-406. John Wiley, Chichester.

Haddon, M. 2001. Modelling and quantitative methods in fisheries, pp. 406. Chapman & Hall, Boca Raton, Fla.

Kaushal, D.K., V.K. Sharma, M.P.S. Kohli, B.C. Jha and H.P. Singh. 2006. Production dynamics of Gobindsagar reservoir on river Sutlej, Himachal Pradesh. In: Fisheries and aquaculture in Indus river region (ed. M.P. Singh Kohli), pp. 100-113. Indian Society of Fisheries Professionals, Mumbai, India.

Prajneshu and V.M. Kandala. 2005. Fitting of nonlinear Schaefer model. Indian Journal of Fisheries 52:33-36.

Prajneshu and S. Ravichandran. 2003. A method for fitting of nonlinear Fox model in fisheries. Indian Journal of Animal Sciences 73:329-331.

Ratkowsky, D.A. 1990. Handbook of non-linear regression models, pp. 30-42. Marcel Dekker, New York.

Schaefer, M.B. 1954. Some aspects of the dynamics of populations important to the management of the commercial fisheries. Bulletin of Inter-American Tropical Tuna Commission 1:27-56.

Seber, G.A.F. and C.J. Wild. 1989. Nonlinear regression. John Wiley, New Jersey. 768 pp.

## ANNEXURE

### Derivation for Partial Reparameterization of Schaefer Model with AR (1):

Schaefer model with autoregressive of order one is of the following form:

$$C = KE \left( 1 - \frac{E}{r} \right) + \Phi \varepsilon \quad \text{----- (i)}$$

To obtain expected-value parameters for 'r' and 'K' of the above equation (i), the values  $E_1$  and  $E_2$  of the explanatory variable  $E$ , within the observed range of  $E$  are to be chosen and correspondingly, the values of  $\varepsilon_1$  and  $\varepsilon_2$  of  $\varepsilon$ . Then, we can get the expected values from equation (i) as follows:

$$C_1 = KE_1 \left( 1 - \frac{E_1}{r} \right) + \Phi' \varepsilon_1, \quad \text{----- (ii)}$$

and

$$C_2 = KE_2 \left( 1 - \frac{E_2}{r} \right) + \Phi' \varepsilon_2. \quad \text{----- (iii)}$$

where  $\Phi'$  is the estimated value of  $\Phi$  obtained by fitting the Schaefer model with AR(1) before reparameterization.

From equation (ii), we get

$$K = \frac{(C_1 - \Phi' \varepsilon_1)r}{E_1(r - E_1)} \quad \text{----- (iv)}$$

Again from equation (iii), we have

$$\begin{aligned}
 \left(1 - \frac{E_2}{r}\right) &= \frac{(C_2 - \Phi' \varepsilon_2)}{KE_2} \\
 \Rightarrow \frac{E_2}{r} &= 1 - \frac{(C_2 - \Phi' \varepsilon_2)}{KE_2} \\
 \Rightarrow r &= \frac{E_2}{\left[1 - \frac{(C_2 - \Phi' \varepsilon_2)}{KE_2}\right]} \quad \text{----- (v)}
 \end{aligned}$$

Substituting the expression for K from equation (iv) in equation (v), we get

$$\begin{aligned}
 r &= \frac{E_2}{1 - \frac{(C_2 - \Phi' \varepsilon_2)}{E_2 \left[ \frac{(C_1 - \Phi' \varepsilon_1)r}{(r - E_1)E_1} \right]}} \\
 \Rightarrow r &= \frac{E_2^2 [(C_1 - \Phi' \varepsilon_1)r / (r - E_1)E_1]}{E_2 [(C_1 - \Phi' \varepsilon_1)r / (r - E_1)E_1] - (C_2 - \Phi' \varepsilon_2)} \\
 \Rightarrow r &= \frac{(C_1 - \Phi' \varepsilon_1)r E_2 - (C_2 - \Phi' \varepsilon_2)(r - E_1)E_1}{(r - E_1)E_1} = \frac{(C_1 - \Phi' \varepsilon_1)r E_2^2}{(r - E_1)E_1} \\
 \Rightarrow (C_1 - \Phi' \varepsilon_1)r E_2 - (C_2 - \Phi' \varepsilon_2)(r - E_1)E_1 - (C_1 - \Phi' \varepsilon_1)E_2^2 &= 0 \\
 \Rightarrow r [(C_1 - \Phi' \varepsilon_1)E_2 - (C_2 - \Phi' \varepsilon_2)E_1] &= [(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2] \\
 \Rightarrow r = \frac{[(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2]}{[(C_1 - \Phi' \varepsilon_1)E_2 - (C_2 - \Phi' \varepsilon_2)E_1]} \quad \text{----- (vi)} &
 \end{aligned}$$

Similarly, substituting the expression for r from equation (vi) in equation (iv), we get

$$K = \frac{(C_1 - \Phi' \varepsilon_1)}{E_1} \left\{ \frac{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2}{(C_1 - \Phi' \varepsilon_1)E_2 - (C_2 - \Phi' \varepsilon_2)E_1} \right\}$$

$$\begin{aligned}
& \frac{1}{\left\{ \frac{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2}{(C_1 - \Phi' \varepsilon_1)E_2 - (C_2 - \Phi' \varepsilon_2)E_1} - E_1 \right\}} \\
\Rightarrow K &= \frac{(C_1 - \Phi' \varepsilon_1)}{E_1} \left\{ \frac{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2}{(C_1 - \Phi' \varepsilon_1)E_2 - (C_2 - \Phi' \varepsilon_2)E_1} \right\} \\
&\quad \left\{ \frac{(C_1 - \Phi' \varepsilon_1)E_2 - (C_2 - \Phi' \varepsilon_2)E_1}{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2 - (C_1 - \Phi' \varepsilon_1)E_1 E_2 + (C_2 - \Phi' \varepsilon_2)E_1^2} \right\} \\
\Rightarrow K &= \frac{(C_1 - \Phi' \varepsilon_1)}{E_1} \left\{ \frac{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2}{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1 E_2} \right\} \\
\Rightarrow K &= \frac{(C_1 - \Phi' \varepsilon_1) \left[ (C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2 \right]}{(C_1 - \Phi' \varepsilon_1) \left( E_1 E_2^2 - E_1^2 E_2 \right)} \\
\Rightarrow K &= \frac{\left[ (C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2 \right]}{E_1 E_2 (E_2 - E_1)} \quad \text{----- (vii)}
\end{aligned}$$

As the parameters  $r$  and  $K$  are positive values, the following conditions should be satisfied:

$$(1) (C_1 - \Phi' \varepsilon_1)E_2 > (C_2 - \Phi' \varepsilon_2)E_1 \text{ and } (2) E_2 > E_1.$$

Now, using equations (vi) and (vii) in equation (i), we get

$$\begin{aligned}
C &= \left\{ \frac{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2}{E_1 E_2 (E_2 E_1)} \right\} E \\
&\quad \left[ 1 - \frac{E}{\left\{ \frac{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2}{(C_1 - \Phi' \varepsilon_1)E_2 - (C_2 - \Phi' \varepsilon_2)E_1} \right\}} \right] + \Phi \varepsilon \\
\Rightarrow C &= \left\{ \frac{(C_1 - \Phi' \varepsilon_1)E_2^2 - (C_2 - \Phi' \varepsilon_2)E_1^2}{E_1 E_2 (E_2 E_1)} \right\} E
\end{aligned}$$

$$\begin{aligned}
 & \left[ \frac{(C_1 - \Phi'\varepsilon_1)E_2^2 - (C_2 - \Phi'\varepsilon_2)E_1^2 - E\{(C_1 - \Phi'\varepsilon_1)E_2 - (C_2 - \Phi'\varepsilon_2)E_1\}}{(C_1 - \Phi'\varepsilon_1)E_2^2 - (C_2 - \Phi'\varepsilon_2)E_1^2} \right] + \Phi\varepsilon \\
 \Rightarrow C &= \left[ \frac{(C_1 - \Phi'\varepsilon_1)EE_2^2 - (C_2 - \Phi'\varepsilon_2)EE_1^2 - (C_1 - \Phi'\varepsilon_1)E^2E_2 - (C_2 - \Phi'\varepsilon_2)E^2E_1}{E_1E_2^2 - E_1^2E_2} \right] + \Phi\varepsilon \\
 \Rightarrow C &= \left[ \frac{(C_1 - \Phi'\varepsilon_1)EE_2(E_2 - E) - (C_2 - \Phi'\varepsilon_2)EE_1(E_1 - E)}{E_1E_2(E_2 - E_1)} \right] + \Phi\varepsilon \\
 \Rightarrow C &= [ \{EE_2(E_2 - E)(C_1 - \Phi'\varepsilon_1) - EE_1(E_1 - E)(C_2 - \Phi'\varepsilon_2)\} / E_1E_2(E_2 - E_1) ] + \Phi\varepsilon
 \end{aligned}$$

----- (viii)

Thus, the original parameters 'r' and 'K' are replaced by new parameters which by virtue of being expected-value parameters. Although the above expressions are more cumbersome in appearance than the original expressions, parameterizations with expected-value parameters offer many advantages.