

# Statistical Modelling of Distance Associated Marriage Migration with Special Reference to Western Uttar Pradesh (India)

C. B. Gupta<sup>1</sup>, Brijesh P. Singh<sup>2</sup>, Sachin Kumar<sup>1,\*</sup>

<sup>1</sup>Department of Mathematics, Birla Institute of Technology and Science Pilani-Pilani Campus, India

<sup>2</sup>Faculty of Commerce, Banaras Hindu University, Varanasi, India

**Abstract** In the present study, the marriage distance pattern of females from western Uttar Pradesh (U.P.), India is analysed. The suitability of Pareto-exponential function in the study region is tested and two other models viz; Gamma and Weibull distribution with two parameters are proposed. The performance of all the three above models across different categories is compared, and then to observe the effect of various covariates on marriage distance Gamma regression is applied. Based on the study, interesting pattern and relationship through the models applied to the data are found.

**Keywords** Marriage, Migration, Regression, M.S.E., Caste and religion etc

## 1. Introduction

In any society, migrations are caused by many factors including social, economic and political. Among these, marriage is an important source of this movement in India, particularly, north India. It is by far the largest form of migration in India and is close to universal for every female in rural areas. It is observed that it is, predominantly, the females who migrate. Every girl has to move to her in-law's place of residence after marriage. 91.2% females reported marriage as the cause of dislocation of their residence in rural areas ([9]). Among all forms of migration, female marriage centric migration alone constitutes 64.9% share ([2]).

All the studies based on developing countries considered the female migrants passive and tied movers, means they have to move as a result of marriage and social customs prevailed in the region ([5] [14], [1], [3]). Though, net migration is zero, as the out-migrants females are balanced by in-migrant females, but, it leaves a significant socio-economic impact on both the origin and destination place by bringing about social and cultural change. Despite such a high magnitude and growing significance, female marriage migration has been little explored in migration studies. It is to be noted that pattern of female marriage migration has been much skewed depending on the social norms and customs prevailed in a particular society.

During the past few decades, many attempts have been

made to understand the relationship between marriages and distance ([8], [10], [11], [7], [16], [21], [15], [4], [20]). In Indian context, [16] first, proposed a probability model to explain the distribution of distance associated with marriage migration, under the assumptions (i) the number of marriages are proportional to the area of the distance interval  $d_1$  to  $d_2$  i.e. number of marriages with distance interval ( $d_1 < d_2 < D$ ) are proportional to  $\pi(d_2^2 - d_1^2)$  (ii) after the distance  $D$  in a particular direction, the distribution of distance follows an exponential form i.e. the number of marriages in distance interval  $d_1$  to  $d_2$  ( $D < d_1 < d_2$ ) is;

$e^{-\lambda(d_1-D)} - e^{-\lambda(d_2-D)}$ , where  $\lambda$  is the risk parameter.

[21], again, modified the above discussed model as;

If  $M$  is the number of marriages at distance  $r$ , then;

$M < 2\pi r f(r)dr$ , where

$$f(r) = e^{-\lambda(r-D)}, \text{ if } r > D$$

$$= 0, \text{ if } r \leq D$$

[4] proposed the Pareto-exponential function developed by [8] to the data taken from Bangladesh and found it suitable than the models proposed by [16] and [21]. [13] applied Exponential, Pareto and Pareto-Exponential function to Bangladesh data and tested the suitability of three models. Recently, [18] proposed a three parameter Weibull distribution and applied it satisfactorily to the data collected from Bihar. But all the above models suffered from certain drawbacks. Secondly, all the above discussed models were tested on the data taken from eastern Uttar Pradesh, Bihar and Bangladesh, while, our study is based on the data collected from western Uttar Pradesh. This part of the state is quite different from the other parts, in terms of geographical,

\* Corresponding author:

sachin.224@rediffmail.com (Sachin Kumar)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

economic and social aspects. Higher per capita income, availability of fertile agriculture land, growth of industrial units, well developed road and transport facilities and proximity to national capital region set this region apart from the other regions and generated an interest to the researcher to study the marriage-distance pattern of the region. To the best of our knowledge, no study has been undertaken in this region to study the distribution of marriage-distance pattern.

In the present work, we applied the Pareto-Exponential function to test its suitability in the region and proposed two other probability distributions to describe the behaviour pattern of marriage-distance relationship. The purpose of this work is, therefore, to compare the applicability of basic distribution to the data and to provide a set of logical basis to understand the marriage-distance relationship clearly. Therefore, our proposed work has four objectives to (a) use of Pareto-Exponential function, Gamma distribution and Weibull distribution with two parameters (b) apply these models and to investigate how well these models fit the data (c) comparison of the models (d) to investigate the relationship between distance and various covariates through a Gamma regression model.

## 2. Description of Data

The data used in the present study have been collected from the rural area of district Meerut (U.P.). There has been a rapid spurt of industrial development in and around the city during the past decades and it has become a major hub for various economic, cultural, educational and developmental activities over the years. The baseline survey of nearly 3600 households was undertaken to get the reliable and relevant data pertaining to the problem under study. Following the guidelines from [17] data were collected from three types of villages through a stratified clustered sampling method. These three types of village have been identified as semi-urban, remote and growth centres.

The villages from district Meerut are divided into two groups as per their distance from the boundary of Meerut Nagar Nigam, to constitute two strata. The first stratum consists of the villages less than 8 km. of distance, and are termed as semi-urban villages, while the rest are called as remote villages and belonged to second stratum. For the selection of growth centres, villages, which are located nearby industrial areas and sugar mills, are taken into consideration.

Table 1 displays the frequency distribution of four sets of data, classified according to their religion/category, and shows the number of marriages along with their mean and variances. Distance interval is kept at 15 km due to well-developed road and transport facilities and improved communication system. Another reason for this is attributed to the fact that village and gotra (clan) endogamy is strictly prohibited in this region. Boys and girls from the same village and gotra (clan) are considered to be brothers and

sisters. In majority of cases, villages from same gotra (clan) are situated nearby.

**Table 1.** Frequency distribution of data

Distance (in km.)	No. of marriages (Over all)	No. of marriages (Hindu OBC)	No. of marriages (Hindu SC/ST)	No. of marriages (Muslim)
0-15	127	57	25	30
15-30	196	87	58	37
30-45	151	87	36	21
45-60	97	52	25	12
60-75	74	45	19	9
75-90	43	22	9	9
90-105	15	9	1	4
105-120	7	6	1	-
120-135	6	5	1	-
<b>Total</b>	716	370	175	122
<b>Mean</b>	37.79	40.72	36.09	34.16
<b>Variance</b>	631.73	692.94	525.91	582.27

## 3. Description of Models

**(a). Pareto-Exponential function:** Morril and Pitts (1967) proposed a function as

$$Y = aD^{-b} \quad (1)$$

Where Y is the number of marriages at distance D, but it has the disadvantage of overestimating the closed-in frequencies, so they proposed another Exponential model as;

$$Y = ae^{-bD} \quad (2)$$

and then they combined (i) and (ii) as

$$Y = aD^{-b}e^{-cD} \quad (3)$$

and called as Pareto-Exponential function.

**(b). Gamma distribution:** Gamma distribution is a two parameter family of distribution. Let X be a random variable denoting the distance associated with marriage migration, naturally, X takes only the positive values. It has been observed from the data that initially the number of marriages increases with increase in distance, reaches a peak and then starts declining. There is no demographic rationale behind the models we choose, in fact, the distribution are chosen to have the properties to model data of this kind, like, they must be always positive and skewed to the right. The probability density function of Gamma distribution is given as follows;

$$f(x) = \frac{\beta^\alpha}{\Gamma\alpha} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0, \alpha > 0, \beta > 0$$

where  $\alpha$  is the shape parameter and determines the shape of distribution curve, and  $\beta$  denotes the rate parameter. Mean and variance of the distribution is given as follows;

$$E(x) = \frac{\alpha}{\beta}, \quad V(x) = \frac{\alpha}{\beta^2}$$

**(c). Weibull distribution:** The probability density function of a Weibull distribution is;

$$f(x) = \frac{\alpha}{\lambda} \left(\frac{x}{\lambda}\right)^{\alpha-1} e^{\left(\frac{-x}{\lambda}\right)^{\alpha}}, \quad x \geq 0, \alpha > 0, \lambda > 0$$

Where  $\alpha$  and  $\lambda$  are shape and scale parameter respectively. The mean and variance are given as follows;

$$E(x) = \lambda \Gamma\left(1 + \frac{1}{\alpha}\right),$$

$$V(x) = \lambda^2 \left[ \Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2 \right]$$

### 3.1. Parameter Estimates

The principle of least squares is employed to estimate the parameter of Pareto-Exponential function by taking the logarithmic on both sides. The parameter of Gamma and Weibull distribution are estimated by maximum likelihood estimation method, since this method possesses several desirable properties and advantage over other method of estimation. It is more robust in the sense that it possesses the properties of invariant, asymptotic normality, consistency, sufficiency and minimum variance for large samples. The values of various estimated parameters are shown in table 2. The p-values from table 3 also demonstrate that Gamma distribution and Pareto-Exponential function provides a good fit as compared to other model across all the categories.

### 3.2. Model Accuracy Results

Table 4 displays model accuracy results for all the four models. It is the most popular method of observing the efficiency of model, when several models are in fray. [6] have described several measures of accuracy. Some measures are scale dependent and some are scale independent. Scale independent measures should not be used across different data sets which are not on the same scale. We have used four following measures of accuracy.

**(a) Mean square error (MSE):** It is a scale dependent measure of error.

$MSE = \text{mean} (Z_i - \hat{Z}_i)^2$ , where  $Z_i$  is the observed value and  $\hat{Z}_i$  is the fitted value.

**(b) Root mean square error (RMSE):**  $RMSE = \sqrt{MSE}$ . It has the advantage of being on the same scale as the data, but both MSE and RMSE have the demerit of being more sensitive to the outliers.

**(c) Mean absolute error (MAE):**  $MAE = \text{mean} (|Z_i - \hat{Z}_i|)$ .

**(d) Mean absolute percentage error (MAPE):** MAPE has the advantage of being scale dependent, reliable and easy to calculate. It can be used across different data sets, which are on different scale. The formula for MAPE is given as follows;

$$MAPE = \text{mean} \left( \left| 100 \frac{(Z_i - \hat{Z}_i)}{Z_i} \right| \right)$$

Table 2. Estimated results of parameter

Type of data	Model	Shape ( $\alpha$ )	Scale ( $\lambda$ )	Rate ( $\beta$ )	a	b	c
Over all	G	2.22	-	0.059	-	-	-
	W	1.58	42.26	-	-	-	-
	PE	-	-	-	17.58	-1.18	0.055
Hindu(OBC)	G	2.19	-	0.054	-	-	-
	W	1.60	45.51	-	-	-	-
	PE	-	-	-	8.70	-1.11	0.049
Hindu(SC/ST)	G	2.59	-	0.072	-	-	-
	W	1.68	40.64	-	-	-	-
	PE	-	-	-	1.92	-1.56	0.069
Muslim	G	2.17	-	0.063	-	-	-
	W	1.51	38.14	-	-	-	-
	PE	-	-	-	19.68	-0.36	0.032

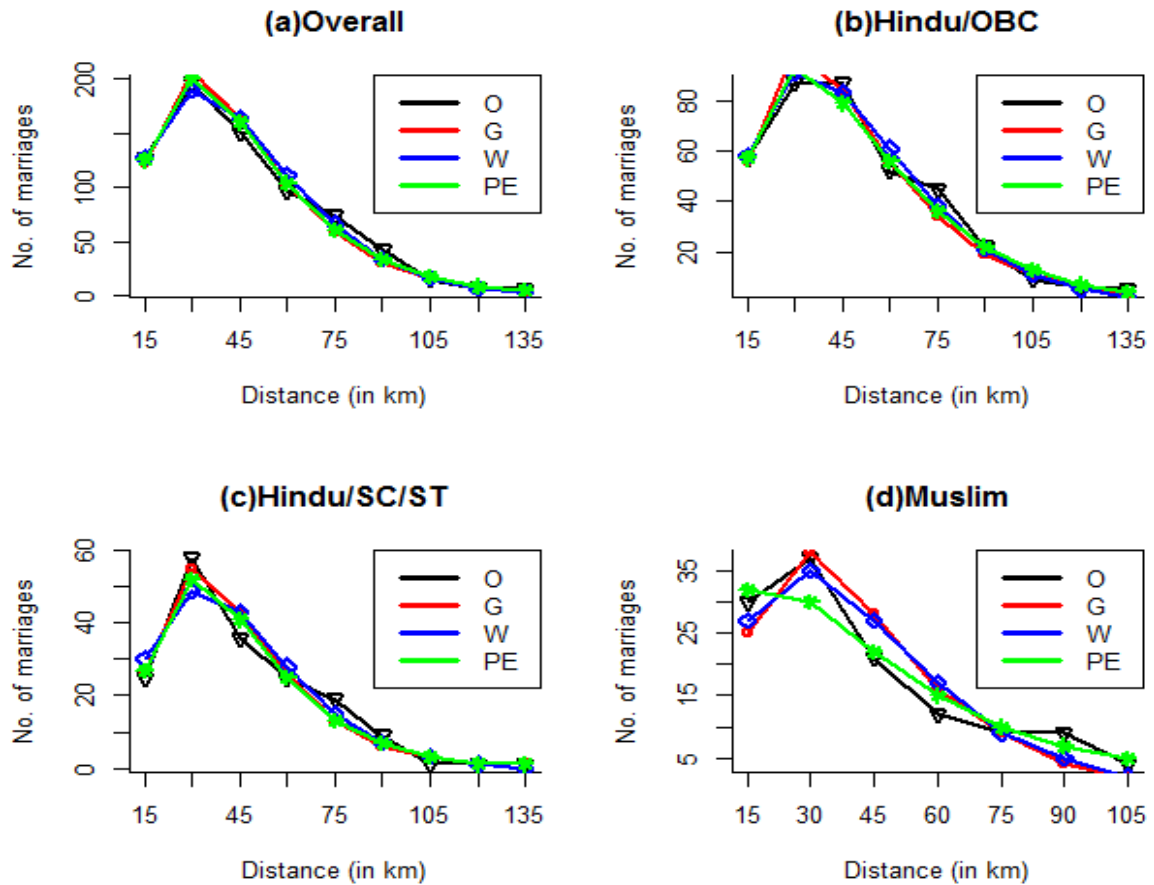
Table 3. Goodness of fit measures

Type of data	Model	Chi-square	P-value	AIC	BIC	Loglikelihood
Over all	G	10.77	0.06	6426.13	6435.27	-3211.07
	W	7.22	0.20	6437.50	6446.65	-3216.75
	PE	7.39	0.19	-	-	-
Hindu(OBC)	G	6.94	0.23	3380.57	3388.40	-1688.29
	W	5.70	0.34	3379.10	3386.93	-1687.55
	PE	5.36	0.37	-	-	-
Hindu(SC/ST)	G	6.56	0.16	1539.81	1546.14	-767.91
	W	5.11	0.16	1546.46	1552.79	-771.23
	PE	5.59	0.13	-	-	-
Muslim	G	11.94	0.008	1075.50	1081.10	-535.75
	W	8.39	0.04	1080.44	1086.05	-538.22
	PE	3.72	0.29	-	-	-

G- Gamma distribution, W- Weibull distribution, PE-Pareto-Exponential function

**Table 4.** Comparison of accuracy measures

Type of data	Model	MSE	RMSE	MAE	MAPE
Over all	G	80.67	8.98	7.33	14.09
	W	60.44	7.77	6.00	12.71
	PE	46	6.78	5.56	13.63
Hindu(OBC)	G	37.00	6.08	4.33	14.24
	W	19.78	4.45	3.56	16.36
	PE	23.89	4.89	3.67	13.20
Hindu(SC/ST)	G	12.44	3.53	2.67	33.50
	W	21.00	4.58	3.67	45.58
	PE	15.57	3.95	3.29	20.84
Muslim	G	17.14	4.14	3.43	27.37
	W	13.43	3.66	3.14	25.73
	PE	9.86	3.14	2.43	16.10

**Figure 1.** Observed (O) and expected fitted frequencies graph from models selected

In table 4, accuracy results of all the four measures are given for each data set. The model with the least error is preferred. It is evident from the table that Weibull model and Pareto-Exponential model fit the data with minimum error for majority of data sets. Though, Gamma is also very close. Both these models display substantial degree of improvement over the Gamma model.

### 3.3. Model Validation

Cross validity prediction power (CVPP) is employed to

check the stability of model over the population. CVPP is given as;

$$\rho^2 = 1 - \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)}(1-R^2)$$

where  $n$  denotes the number of cases,  $k$  is the number of parameters in the model and  $R$  is correlation coefficient between observed and expected frequencies. The absolute value of difference between  $\rho^2$  and  $R^2$  of the model is equal to (1-shrinkage) ([19]). From table 5, it is clear that all

the fitted models are nearly 100% stable over the population with high coefficient of determination ( $R^2$ ).

**Table 5.** Comparison of coefficient of determination and stability

Type of data	Model	n	k	$R^2$	Stability
Over all	G	716	2	0.9874648	0.9999119
	W	716	2	0.9867133	0.9999066
	PE	716	3	0.9908194	0.9999095
Hindu(OBC)	G	370	2	0.9689695	0.9995813
	W	370	2	0.9808836	0.9997379
	PE	370	3	0.9749148	0.9995176
Hindu(SC/ST)	G	175	2	0.9638016	0.9989391
	W	175	2	0.9360231	0.998125
	PE	175	3	0.9469642	0.9978034
Muslim	G	113	2	0.8901265	0.9953286
	W	113	2	0.9022034	0.995842
	PE	113	3	0.9296636	0.9957561

#### 4. Determinants of Distance through a Gamma Regression Model

Table 6 depicts the regression estimates of relative risk of distance associated with marriage migration. The outcome of multivariate analysis shows that majority of the predictor variables are statistically significant at 5% level of significance both in unadjusted and adjusted models, among

those female education, family size and age at marriage are most prominent. The secondary and higher educated females are at greater risk (54% and 39% respectively) of marrying at larger distance, as compared to lower educated females. The females, whose fathers are engaged in labour work, are at lesser risk, while the females whose fathers are employed somewhere are at greater risk relative to females belonging to peasantry family. Another significant covariate is family size of females. Analysis results show that females from medium and large family are at greater risk of finding their mate in long distance relative to their small family counterpart. Surprisingly, Hindu females have lesser probability of finding their grooms in shorter distance than Muslim females. Residents of remote villages and growth centres have higher risk of migration at longer distance. It is also found that higher the age at marriage, higher is the chance of finding the suitable mate in longer distance. To investigate the effect of education and family size on the marriage distance two additional models are constructed in table 7 (model 2 & 3). In model 2, female education and in model 3 both female education and family size were dropped to observe the effect of these two covariates. Table 7 results indicate the prominence of both variables in determining the marriage distance. All these calculations were done with the help of R-3.2.2 version ([12]) on a Window 10 platform.

**Table 6.** Gamma regression estimates: Model 1

Unadjusted Gamma model Response variable-Distance 95%					Adjusted Gamma model Response variable-Distance 95%			
Explanatory Variables	P-value	Exp(B)	C.I. lower	C.I. upper	P-value	Exp(B)	C.I. lower	C.I. upper
<b>Education</b>								
Primary		1.00				1.00		
Secondary	.00	1.52	.31	.53	.00	1.39	.21	.45
Higher	.00	1.79	.45	.71	.00	1.54	.28	.57
<b>Occupation</b>								
Agriculture		1.00				1.00		
Labour	.02	0.88	-.24	-.02	.39	.95	-.16	.06
Employee	.00	1.32	.15	.42	.01	1.18	.04	.30
<b>Family size</b>								
Small		1.00				1.00		
Medium	.00	1.46	.21	.55	.00	1.36	.14	.47
Large	.00	1.62	.32	.64	.00	1.52	.26	.57
<b>Age at marriage</b>								
<=20		1.00				1.00		
21-24	.01	1.18	.03	.31	.03	1.16	.01	.29
Above 24	.00	1.58	.32	.60	.00	1.41	.19	.48
<b>Religion</b>								
Hindu		1.00				1.00		
Muslim	.07	0.89	-.25	.011	.28	1.08	-.06	.22
<b>Type of village</b>								
Semi urban		1.00				1.00		
Remote	.02	1.14	.02	.24	.07	1.11	-.005	.21
Growth centres	.38	1.06	-.07	.19	.18	1.09	-.04	.21
<b>Constant</b>					.00	15.33	2.56	2.90

AIC: 6300.5

**Table 7.** Gamma regression estimates: Model 2 Model 3

Response variable-Distance 95%					Response variable-Distance 95%			
Explanatory Variables	P-value	Exp(B)	C.I. lower	C.I. upper	P-value	Exp(B)	C.I. lower	C.I. upper
<b>Education</b>								
Primary								
Secondary		-				-		
Higher		-				-		
<b>Occupation</b>								
Agriculture		1.00				1.00		
Labour	.06	.90	.00	.004	.09	.90	-.21	.05
Employee	.001	1.25	.00	.35	.00	1.27	.11	.38
<b>Family size</b>								
Small		1.00						
Medium	.00	1.46	.00	.54		-		
Large	.00	1.60	.00	.63		-		
<b>Age at marriage</b>								
<=20		1.00				1.00		
21-24	.04	1.15	.00	.29	.03	1.17	.01	.30
Above 24	.00	1.46	.00	.53	.00	1.48	.24	.54
<b>Religion</b>								
Hindu		1.00				1.00		
Muslim	.74	.98	.00	.12	.80	1.02	-.12	.16
<b>Type of village</b>								
Semi urban		1.00				1.00		
Remote	.07	1.11	.00	.21	.02	1.14	.02	.24
Growth centres	.22	1.08	.00	.21	.29	1.07	-.06	.20
<b>Constant</b>	.00	19.11	2.75	3.15	.00	27.66	3.17	3.47

AIC: 6337.4

AIC: 6364.3

## 5. Conclusions

In this paper, we studied the distribution pattern of distance associated with marriage migration of western U.P. by using the Pareto-Exponential, Gamma and Weibull models. From the above discussion and analysis, it can be concluded that Pareto-exponential and Gamma models fit the data satisfactorily across all categories. Though, Pareto-Exponential seems to give much better fit, but Weibull model, with two parameters, also gives quite a better approximation. We have also found a larger marriage field in this region relative to previous studies of eastern U.P., Bihar and Bangladesh. Rigorous developmental activities in this region might be the reason for the larger marriage field. Gamma regression analysis provides an insight of how the various covariates affect the decision of finding a suitable mate. Dominance of Hindu population and scattered Muslim population might be the reason for Hindu females to migrate at short distance as compared to Muslim females. All the variables selected affect the distance associated with marriage migration.

## ACKNOWLEDGEMENTS

One of the authors (Sachin Kumar) is highly indebted to UGC (BSR), who provided us with the financial assistance through their reference letter number F.7-293/2010(BSR) as dated on November 2013 for carrying out this work.

## REFERENCES

- [1] Bonney, N., and Love, J., 1991. Gender and migration: Geographical mobility and the wife's sacrifices. *Sociology Review*, 39, 335-348.
- [2] Census of India, 2001.
- [3] Fincher, R., 1993. Gender relation and the geography of migration (commentary). *Environment and Planning A*, 25, 1703-1705.
- [4] Hossain, M.Z., Some demographic models and their applications with reference to Bangladesh (Unpublished

- doctoral dissertation). Banaras Hindu University, Varanasi, India, 2000.
- [5] Houstoun, M.F., Kramer, R.G., and Barrett, J.M., 1984. Female predominance of immigration to the United States since 1930: A first look. *International Migration Review*, 18, 908-963.
  - [6] Hyndman R.J., and Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
  - [7] Libbee, M.J., and D.E. Sopher, Marriage and migration in rural India, in Kosinski, L.A. and R.M. Prothero (eds.) *People on the move: Studies on International Migration*, London: Methwen & company, 1975.
  - [8] Morril, R.L., and F.R. Pitts, 1967. Marriage, migration and the mean information field. *Annals of the Association of American Geographers*, 57, 401-422.
  - [9] National Sample Survey Office (2007-08). *Migration in India*, Report number 470, NSSO 64th round, Government of India: Ministry of Programme Implementation.
  - [10] Perry, P., 1969. Marriage distance relationship in north Otago 1975-1914. *New Zealand's Geographers*, 25, 36-43.
  - [11] Perry, P.J., 1969. Working class isolation and mobility in rural Dorset 1837-1936: A study of marriage distances. *Transactions of the Institute of British Geographers*, 46, 121-141.
  - [12] R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. URL <http://www.R-project.org>.
  - [13] Rahman, M.M., Akter, S., and Rahman, A., 2010. Distance associated with marriage migration in a northern and southern region of Bangladesh: An empirical study. *Journal of Biosocial Science*, 42, 577-586.
  - [14] Rosenzweig, M.R., and Stark, O., 1989. Consumption, smoothing, migration and marriage: Evidence from rural India. *Journal of Political Economy*, 9, 905-926.
  - [15] Samuel, M.J., Patterns of female migration, In Maithili Viswanathan (eds.), *Women and Society*, 4, Print well, Jaipur, India 1994.
  - [16] Sharma, L., A study of the pattern of out-migration from rural areas (Unpublished doctoral dissertation). Banaras Hindu University, Varanasi, India, 1984.
  - [17] Singh R.B., Appendix: Rural Development and Population Growth – A Sample Survey 1978. (Unpublished doctoral dissertation), Banaras Hindu University, Varanasi, India 130-146, 1986.
  - [18] Singh, N.K., and Singh, Brijesh P., 2015. Study of distance associated with marriage migration. *International Journal of Mathematics and Computer Applications Research*, 5, 111-116.
  - [19] Stevens, J., *Applied multivariate statistics for the Social Sciences* (third edition). Lawrence Erlbaum Associates Inc. Publishers: New Jersey, 1996.
  - [20] Yadava, K.N.S., Srivastava, S., and Islam, S., 2002. Distribution of distance associated with marriage migration. *International Journal of Statistical Sciences*, 1, 49-54.
  - [21] Yadava, K.N.S., Raju, K.N.M., and Yadava, G.S., 1988. On the distribution of distance associated with marriage migration in rural areas of Uttar Pradesh, India. *Rural Demography*, 15, 7-18.