International Journal of Statistics and Applications 2017, 7(3): 178-185
DOI: 10.5923/j.statistics.20170703.03

The Role of Outliers in Growth Curve Models: A Case Study of City-based Fertility Rates in Turkey

O. Ufuk Ekiz, Meltem Ekiz*

Department of Statistics, Gazi University, Ankara, Turkey

Abstract Fertility rate, the most important indicator of a population growth in a region or a country is defined as the ratio of actual births to number of women in reproductive age group. The purpose of this study is to create growth curve models (GCM) for fertility rates and then make predictions from these models for various age groups in cities of Turkey. To achieve this, we have used the data published online by Turkish Statistical Institute (TUIK). These city-based data consist of number of women with different ages and their healthy births in 2009-2015. However, multiple outliers in the data will negatively impact the estimators and predictions. To overcome these problems outliers should be purified from the data. By using GCM in the analysis of outliers we observed that outlying cities differ from other cities not only in fertility rates but also in socio-economic status. It is also observed that making future predictions with models containing all the cities (some of which are possible outliers) can be misleading in interpreting results.

Keywords Growth curve model, Outlier, Fertility rate

1. Introduction

Fertility rate for a country or any region is introduced as the ratio of number of actual births to number of women childbearing age during a year. Angeles pointed out that the issue of fertility rate and problems addressed within it has been discussing in the past few decades [1]. A high fertility rate is one of the indicators for high population growth which could have adverse effect on the economy of many countries. High fertility poses health risks for children and their mothers detracts from human capital investment, slows economic growth, and exacerbates environmental threats [1]. Moreover, the fertility rate depends on various demographic, social, and economic variables, such as age at marriage, level of educational attainment, socio-economic status, mode of living, active participation in the work force, exposure to contraceptive information and effect of conservative religious practices [2].

In this paper we focused on considering the changes in fertility rates based on time, by fitting a model for various age groups in cities of Turkey. This allows us to make predictions about the future. A way to put this in practice is using growth curve models (GCM). However, existing outliers in the data would impact the model parameter estimates obtained from GCM and employing such a wrong defined model would negatively affect the future predictions. Hence, first outliers should be purified from the data which leads to better parameter estimates as well as more accurate predictions. Cities detected as outliers by an outlier detection method differ by many factors as mentioned above. Thus, in order to avoid problems, it is essential to take into account these factors for cities that causes to contradictions on fertility rates.

The structure of the paper is as follows. In Section 2, we introduced the GCM used to model the changes in fertility rates for various age groups in terms of years. The method to detect the outliers, which are known that have a great impact on parameter estimates and predictions, is given in Section 3. Finally, in Section 4, by considering all cities we estimate the model parameters and make predictions for the future. Next, to emphasize the possible impact of outliers on the parameter estimates and predictions they are removed from the model and both the parameter estimations and predictions are repeated. The results obtained without outliers are given in Section 4.

2. Growth Curve Models (GCM)

The GCM, [3], is defined as

$$Y_{p\times n} = X_{p\times m} B_{m\times r} Z_{r\times n} + \varepsilon_{p\times n}, \qquad (1)$$

where X and Z are known design matrices of rank $m < p$ and $r < n$, respectively, and the regression coefficient B is unknown. Furthermore, the columns of the error matrix are independent p-variate normal with mean 0 and common unknown covariance matrix $\sum_{p\times p} > 0$, that is Y is a

* Corresponding author:
ozmeltem@gazi.edu.tr (Meltem Ekiz)
Published online at http://journal.sapub.org/statistics

matrix- variate normal distribution. Hence, Y ~ $N_{p,n}\left(XBZ,\Sigma,I_n\right)$, where XBZ is the expected value, Σ and I_n are the covariance matrices of Y_{ij} (i fixed and $j=1,...,p$) and Y_i ($i=1,...,n$), respectively, [3]. Usually, p is the number of time points observed on each of n cases. (m-1) is the degree of polynomial in time, and r is the number of treatment groups. Many real life examples of growth application for the GCM were investigated in literature (see, [3], [4], [5], [6]).

Generalized least square estimator of parameter B in the model given in Equation 1 is

$$\hat{B}=(X'X)^{-1}X'YZ'(ZZ')^{-1}. \tag{2}$$

Furthermore, estimation of Σ parameter, $\hat{\Sigma}$, based on $\hat{B}$ is defined as

$$\hat{\Sigma}=\frac{1}{n}(Y-X\hat{B}Z)(Y-X\hat{B}Z)'. \tag{3}$$

Suppose the covariance matrix Σ is of Rao's simple covariance structure (SCR), i.e., $\Sigma=X\Gamma X'+Q\Theta Q$, where both $\Gamma: m\times m$ and $\Theta:(n-m)\times(n-m)$ are unknown positive definite matrices, and $Q\in\varphi$ in which φ is the orthogonal matrix space of X defined by

$$\varphi=\{Q\,|\,Q:p\times(p-m),\ \text{rank}(Q)=p-m,\ X'Q=0\}, \tag{4}$$

[7]. In this case maximum likelihood estimators of parameters B, Γ, and Θ are

$$B=(X'X)^{-1}XYZ'(ZZ')^{-1}, \tag{5}$$

$$\Gamma=1/_n(X'X)^{-1}X'SX'(X'X)^{-1}, \tag{6}$$

$$\hat{\Theta}=1/_n(Q'Q)^{-1}QYY'Q(Q'Q)^{-1}, \tag{7}$$

respectively. Here, $S\equiv Y(I_n-P_{Z'})Y'$ and $P_{Z'}=Z(Z'Z)^{-1}Z'$, [5, 6].

3. Detect outliers in GCM

For detecting outliers in GCM with Rao's SCS, it is stated in [8] that the mean shift regression model (MSRM) can be used as one of the most common outlier-generating models. MSRM, [5, 6],

$$Y_{p\times n}=X_{p\times n}B_{m\times r}Z_{r\times n}+X_{p\times m}\Phi_{p\times n}D_{k\times n}+\varepsilon_{p\times n}, \tag{8}$$

where ε~ $N(0,\ \Sigma,\ I_n)$, Φ is a mean shift parameter, k is number of outliers, and $D=(d_{n-k+1},d_{n-k+2},...,d_n)'$ is a matrix of indicator variables, depends on Rao's SCS $\Sigma=X\Gamma X'+Q\Theta Q'$, where $Q\in\varphi$. The ith column of D' which is denoted by d_i is an n-variate vector with ith component equal to one and others zero, $n-k+1\le i\le n$.

In [9] and [10] it is pointed out that the problem of outlier detection can be reduced to testing whether or not the mean of the population is actually shifted. For the GCM with Rao's SCS, this problem becomes testing if the mean shift parameter in the MSRM is zero. In other words, it is sufficient to test the hypothesis $H_0:\Phi=0$ versus $H_0:\Phi\ne 0$. If the null hypothesis is rejected at size α, then the spurious observation set $Y_\ell=(y_{n-k+1},y_{n-k+2},...,y_n)$ can be declared as k outliers at level α (see [9], p. 28-30 and [10], p. 187-190).

Theorem 1. [5, 6] For the MSRM with Rao's SCS, the likelihood ratio test of level α of $H_0:\Phi=0$ versus $H_0:\Phi\ne 0$ is equivalent to rejecting H_0 if

$$\Lambda_\ell=\det\{I_k-E_\ell'X(X'SX)^{-1}X'E_\ell(I_k-H_\ell)^{-1}\le C_\alpha, \tag{9}$$

where the constant C_α denotes the lower 100α percent critical point of Wilk's distribution $\Lambda(m,n-k-r,k)$ and $H_\ell=Z_\ell'(ZZ')^{-1}Z_\ell$.

Without loss of generality the index set can be assumed to be $\ell=\{n-k+1,\ n-k+2,...,n\}$, therefore the response matrix Y can be partitioned into $Y=(Y_{(\ell)}:Y_\ell)$, where $Y_\ell=(y_{n-k+1},\ y_{n-k+2},...,\ y_n)$. Correspondently, the matrices Z and E are partitioned into $Z=(Z_{(\ell)}:Z_\ell)$ and $E=(E_{(\ell)}:E_\ell)$, respectively. In this theorem $S\equiv Y(I_n-P_Z)Y'$, $P_Z=Z(Z'Z)^{-1}Z'$, and $E=Y(I_n-P_Z)$, [5, 6].

Pan and Fang used $k=1$ and $\ell=\{i\}$, $1\le i\le n$, in Theorem 1, and considered whether or not the ith observation (case) is an outlier [5]. According to one of Wilk's distributional properties (e.g. [11]], the null hypothesis $\Lambda(m,n-1-r,1)$ implies that

$$\frac{n-r-m}{m}\frac{1-\Lambda_i}{\Lambda_i}\sim F_{m,n-r-m} \tag{10}$$

under the null hypothesis H_0. Therefore, the ith individual is declared an outlier if $\frac{n-r-m}{m}\frac{1-\Lambda_i}{\Lambda_i}\ge F_\alpha^*$, where F_α^* is the upper 100α percent critical point of $F_{m,n-r-m}$ distribution, [5].

4. Estimation of GCM, Future Predictions and Outliers for City-based Fertility Rates in Reproductive Age-groups in Turkey

Here, we focused on constructing GCM and make future predictions by using them for fertility rates that have a great impact on population growth rate based on cities in Turkey. For this purpose, data which is published online by Turkish Statistical Institute (TUIK) is collected. This data consists of female population and the number of live births for different age groups in 2009-2015. The names and the corresponding numbers of the cities are given in Table 1.

In each city i the value of the dependent variable for jth year of a certain age group is denoted as Y_{ij}, where $i=1,...,81$ and $j=2009,...,2015$. Its value is obtained by ratio of live births to female population for the corresponding age group and city. The X matrix in GCM introduced in Equation 1 for years 2009-2016 and 2016-2020, 2025 are identified as

$$X=\begin{bmatrix}1&1&1&1\\1&2&4&8\\1&3&9&27\\1&4&16&64\\1&5&25&125\\1&6&36&216\\1&7&49&343\end{bmatrix},\quad X_K=\begin{bmatrix}1&8&64&512\\1&9&81&729\\1&10&100&1000\\1&11&121&1331\\1&12&144&1728\\1&17&289&4913\end{bmatrix},$$

respectively. This is due to the fact that a third degree growth curve fits the data more accurately. Furthermore, Z is an n-variant vector with components of one, $Z=I_n$.

Cities in Turkey differ with respect to fertility rates because of their various socio-economic developments.

Hence, cities that differ on socio-economic properties with a great extent than the others may have an adverse impact on model parameter estimates. These observations are noted as outliers in statistics literature [8]. If these outliers are omitted from the data, the model fits the remaining data more accurately. The remaining data set which is not including outliers have similar socio-economic features. For this, determining outliers in each age group is more appropriate.

Let A (outliers) denote a cluster including the order numbers of cities that are detected as outliers for every age group. To see the importance of outliers in future predictions of fertility rates in various age groups first we have estimated $\hat{B}$ and then predicted $\hat{Y}=X_K\hat{B}Z$ and $\hat{Y}=X\hat{B}Z$ for years 2016-2020, 2025 and 2009-2015, respectively. Additionally, same steps are repeated for same years but without outliers in set A. These estimators are denoted as $\hat{B}_{(A)}$, $\hat{Y}_{(A)}=X_K\hat{B}_{(A)}Z$, and $\hat{Y}_{(A)}=X\hat{B}_{(A)}Z$. The results given in Table 2 show that $\hat{B}$, $\hat{Y}=X_K\hat{B}Z$ and $\hat{Y}=X\hat{B}Z$ differ from $\hat{B}_{(A)}$, $\hat{Y}_{(A)}=X_K\hat{B}_{(A)}Z$, and $\hat{Y}_{(A)}=X\hat{B}_{(A)}Z$, respectively, thus outlying points have an impact on future predictions. Building a model that includes cities with high fertility rates (due to socio economic reasons) can be misleading. This makes clear that we have to build models for cities with similar socio-economic features. In this study, the outlying points of each age group are determined as cities with high or low socio-economic indicators. The impact of detected cities (outliers) on growth curves is our interest rather than building models by clustering cities with respect to their socio-economic developments. An evidence of the outliers effect on the estimates of the model parameter $\hat{B}$ could be seen by considering the differences between $\hat{B}$ and $\hat{B}_{(A)}$ (see Table 2).

We next demonstrate figures (Figure 1-9) for each age group to show graphically the variations of fertility rates in years 2009-2015. In the x-axis, numbers 1-7 represents years 2009 to 2015, respectively. y axis represents the fertility rates. Also, in all of the graphical illustrations (a), (b), and (c) are plotted figures for fertility rates of all the 81 cities in Turkey, cities without outliers, and the means of fertility rates of cities with and without outliers, respectively. Growth curve for each city is plotted by a different colour as can be seen in figures (a) and (b). The figures in (c) are the estimates of the means of GCM with respect to with and without outliers and plotted as blue and green curves, respectively. By considering Figure 1c it's observed that means of fertility rates decreases to women grouped as under 15 years. Furthermore, these rates decreases both for 81 cities and for cities with and without outliers numbered as 8-11, 15, and 18. By considering the data it's observed that decreasing is much faster in high socio-economic developed cities. Means of fertility rates for women aged 25-29 seem to be increasing by 2010 (see Figure 4). In this age group cities numbered as 11, 18, 32, 33, 36, 43, 52, 63, 80, and 81 do have a great impact on the means. Though, this increasing structures are visible in Figures 5-7, it's not observed for other ages groups (Figures 2, 3, 8, 9). Furthermore, we obtained that fertility ages in Turkey is generally between ages 25-44. In addition, rate of fertility under age 25 is also decreasing expect for outlying points (cities).

Table 1. Cities and numbers attained for them

No	City	No	City	No	City	No	City
1	Adana	21	Bursa	41	Kastamonu	61	Sakarya
2	Adıyaman	22	Denizli	42	Kayseri	62	Samsun
3	Afyonkarahisar	23	Diyarbakır	43	Kilis	63	Siirt
4	Aksaray	24	Düzce	44	Kocaeli	64	Sinop
5	Amasya	25	Edirne	45	Konya	65	Sivas
6	Ankara	26	Elazığ	46	Kütahya	66	Tekirdağ
7	Antalya	27	Erzincan	47	Kırklareli	67	Tokat
8	Ardahan	28	Erzurum	48	Kırıkkale	68	Trabzon
9	Artvin	29	Eskişehir	49	Kırşehir	69	Tunceli
10	Aydın	30	Gaziantep	50	Malatya	70	Uşak
11	Ağrı	31	Giresun	51	Manisa	71	Van
12	Balıkesir	32	Gümüşhane	52	Mardin	72	Yalova
13	Bartın	33	Hakkâri	53	Mersin	73	Yozgat
14	Batman	34	Hatay	54	Muğla	74	Zonguldak
15	Bayburt	35	Isparta	55	Muş	75	Çanakkale
16	Bilecik	36	Iğdır	56	Nevşehir	76	Çankırı
17	Bingöl	37	Kahramanmaraş	57	Niğde	77	Çorum
18	Bitlis	38	Karabük	58	Ordu	78	İstanbul
19	Bolu	39	Karaman	59	Osmaniye	79	İzmir
20	Burdur	40	Kars	60	Rize	80	Şanlıurfa
						81	Şırnak

Table 2. Estimations and future predictions

	With outliers				Without outliers		
Age (Women)	$\hat{B}$	$\hat{Y} = X_K\hat{B}Z$	$\hat{Y} = X\hat{B}Z$	Outliers (City no)	$\hat{B}_{(A)}$	$\hat{Y}_{(A)} = X_K\hat{B}_{(A)}Z$	$\hat{Y}_{(A)} = X\hat{B}_{(A)}Z$
<15	0.2283[a] -0.0712[a] 0.0122[a] -0.0008[a]	0.0000 -0.0000 -0.0001 -0.0002 -0.0003 -0.0016	0.1685[a] 0.1281[a] 0.1018[a] 0.0848[a] 0.0718[a] 0.0578[a] 0.0378[a]	8 9 10 11 15 18 25,33,36 55,59,70	0.1646[a] -0.0459[a] 0.0072[a] -0.0005[a]	0.0109[a] -0.0172[a] -0.0568[a] -0.1109[a] -0.1822[a] -0.8994[a]	0.1254[a] 0.0977[a] 0.0786[a] 0.0651[a] 0.0545[a] 0.0439[a] 0.0303[a]
15-19	0.0433 -0.0054 0.0009 -0.0001	0.0206 0.0152 0.0079 -0.0017 -0.0141 -0.1306	0.0387 0.0354 0.0330 0.0311 0.0293 0.0273 0.0245	8 11 30 40 43 48 55 57,73	0.0395 -0.0049 0.0007 -0.0001	0.0190 0.0144 0.0083 0.0003 -0.0099 -0.1067	0.0353 0.0322 0.0300 0.0282 0.0266 0.0247 0.0223
20-24	0.1282 -0.0070 0.0008 -0.0001	0.0957 0.0907 0.0845 0.0767 0.0671 -0.0201	0.1219 0.1169 0.1129 0.1094 0.1063 0.1032 0.0997	11 15 32 33 40 43 52,56,69 80,81	0.1237 -0.0070 0.0008 -0.0001	0.0920 0.0873 0.0814 0.0740 0.0649 -0.0182	0.1175 0.1125 0.1084 0.1051 0.1020 0.0990 0.0958
25-29	0.1360 -0.0043 0.0012 -0.0001	0.1369 0.1355 0.1321 0.1262 0.1174	0.1328 0.1315 0.1317 0.1328 0.1343	11 18 32 33 36	0.1276 -0.0043 0.0013 -0.0001	0.1307 0.1288 0.1246 0.1174 0.1068	0.1246 0.1236 0.1242 0.1258 0.1278

	With outliers				**Without outliers**		
Age (Women)	$\hat{B}$	$\hat{Y} = X_K\hat{B}Z$	$\hat{Y} = X\hat{B}Z$	**Outliers (City no)**	$\hat{B}_{(A)}$	$\hat{Y}_{(A)} = X_K\hat{B}_{(A)}Z$	$\hat{Y}_{(A)} = X\hat{B}_{(A)}Z$
		0.0122	0.1359 0.1369	43 52,63 80,81		-0.0182	0.1296 0.1308
30-34	0.0938 -0.0071 0.0028 -0.0002	0.0997 0.0906 0.0746 0.0505 0.0167 -0.3438	0.0893 0.0891 0.0918 0.0959 0.1002 0.1031 0.1035	2 14 18 27 32 33 36,40,52 80,81	0.0850 -0.0057 0.0024 -0.0002	0.0948 0.0880 0.0756 0.0563 0.0290 -0.2686	0.0815 0.0817 0.0845 0.0885 0.0928 0.0960 0.0971
35-39	0.0505 -0.0053 0.0017 -0.0001	0.0499 0.0455 0.0377 0.0256 0.0086 -0.1775	0.0468 0.0456 0.0462 0.0478 0.0498 0.0513 0.0516	2 11 18 36 40 43 63,69,71 80,81	0.0430 -0.0050 0.0016 -0.0001	0.0450 0.0412 0.0339 0.0226 0.0063 -0.1739	0.0395 0.0386 0.0395 0.0414 0.0437 0.0456 0.0462
40-44	0.0173 -0.0025 0.0004 -0.0000	0.0157 0.0166 0.0174 0.0180 0.0183 0.0112	0.0153 0.0140 0.0133 0.0132 0.0134 0.0140 0.0148	2 11 15 17 18 28 32,33,36,55 63,71,80,81	0.0102 -0.0011 0.0002 -0.0000	0.0116 0.0125 0.0132 0.0139 0.0143 0.0120	0.0093 0.0089 0.0088 0.0090 0.0095 0.0101 0.0108
45-49	0.0028 0.0004 -0.0002 0.0000	0.0017 0.0021 0.0030 0.0043 0.0062 0.0278	0.0030 0.0030 0.0027 0.0024 0.0020 0.0018 0.0016	11 14 17 18 33 36 52,55,63 71,80,81	0.8975[a] 0.1584[a] -0.0632[a] 0.0052[a]	0.0008 0.0010 0.0014 0.0019 0.0027 0.0110	0.0010 0.0010 0.0009 0.0009 0.0008 0.0007 0.0007
≥ 50	0.1328[a] -0.0299[a] 0.0023[a] 0.0000[a]	0.0533[a] 0.0673[a] 0.0870[a] 0.1124[a] 0.1438[a] 0.3932[a]	0.1053[a] 0.0826[a] 0.0648[a] 0.0521[a] 0.0444[a] 0.0420[a] 0.0449[a]	11 15 18 33 34 36 55,63,71 80,81	0.5180[a] -0.1115[a] 0.0099[a] -0.0001[a]	0.2063[a] 0.2407[a] 0.2893[a] 0.3516[a] 0.4269[a] 0.9774[a]	0.4163[a] 0.3337[a] 0.2697[a] 0.2236[a] 0.1949[a] 0.1828[a] 0.1868[a]

[a] 1.0e-003

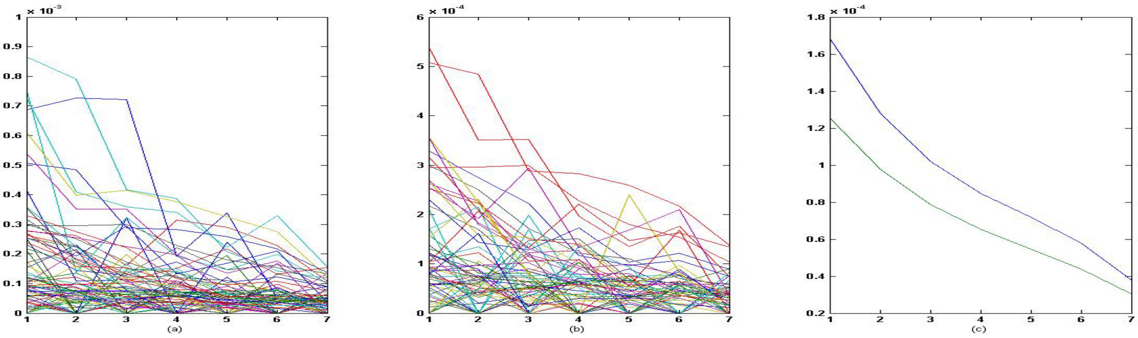

Figure 1. City-based plots on; (a) fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged under 15 in 2009-2015

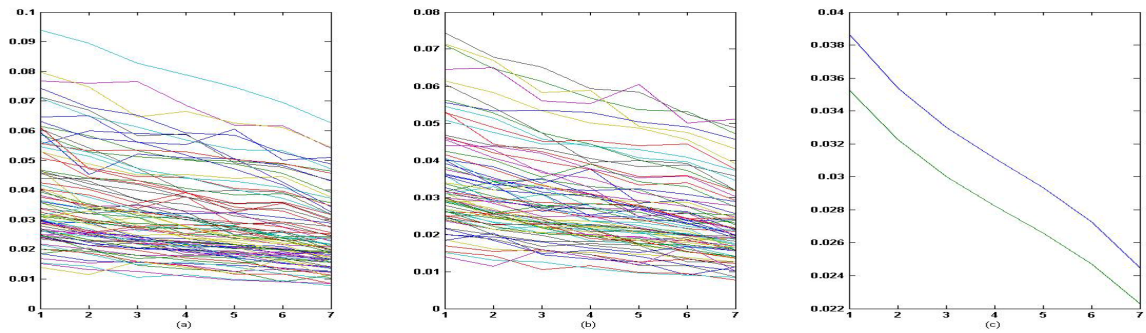

Figure 2. City-based plots on; (a) fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged 15-19 in 2009-2015

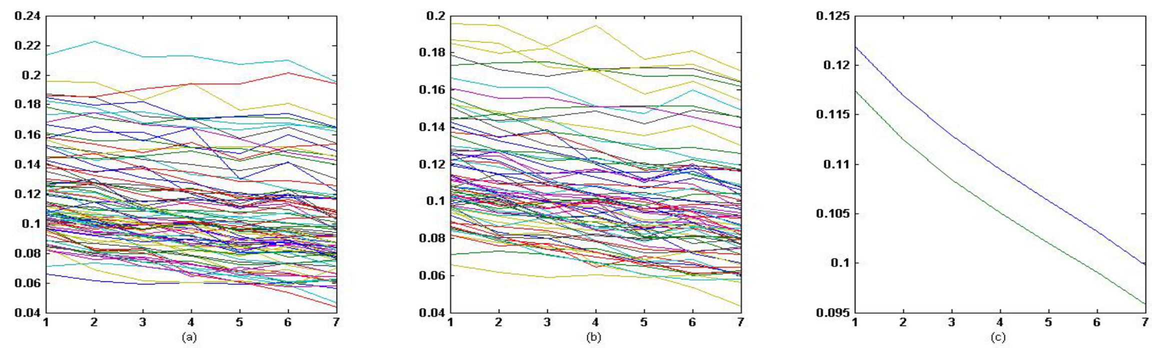

Figure 3. City-based plots on; (a) Fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged 20-24 in 2009-2015

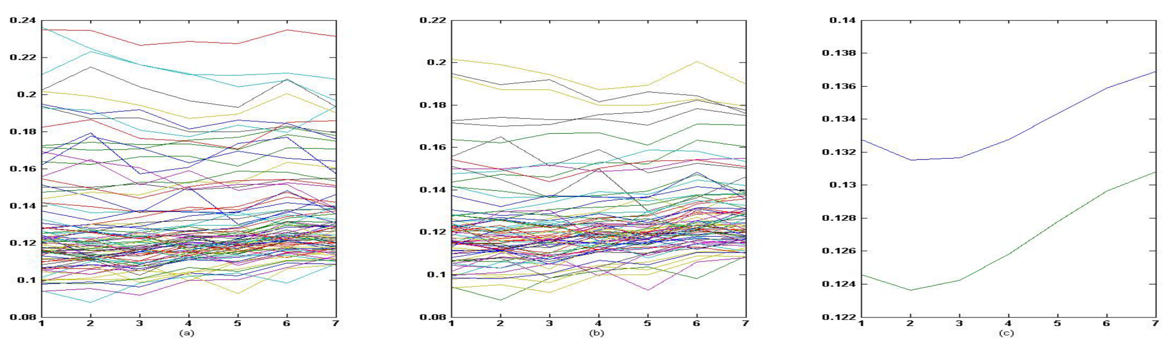

Figure 4. City-based plots on; (a) Fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged 25-29 in 2009-2015

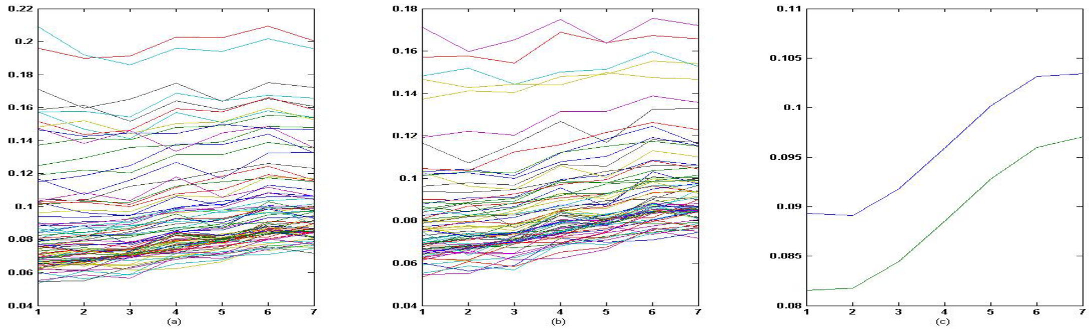

Figure 5. City-based plots on; (a) Fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged 30-34 in 2009-2015

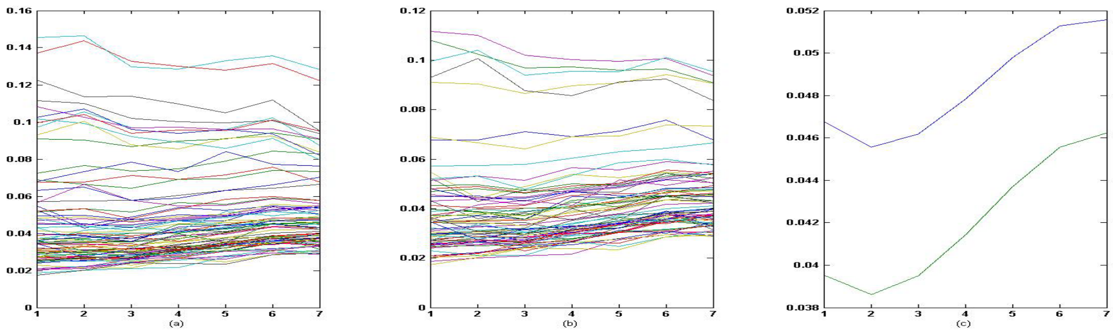

Figure 6. City-based plots on; (a) Fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged 35-39 in 2009-2015

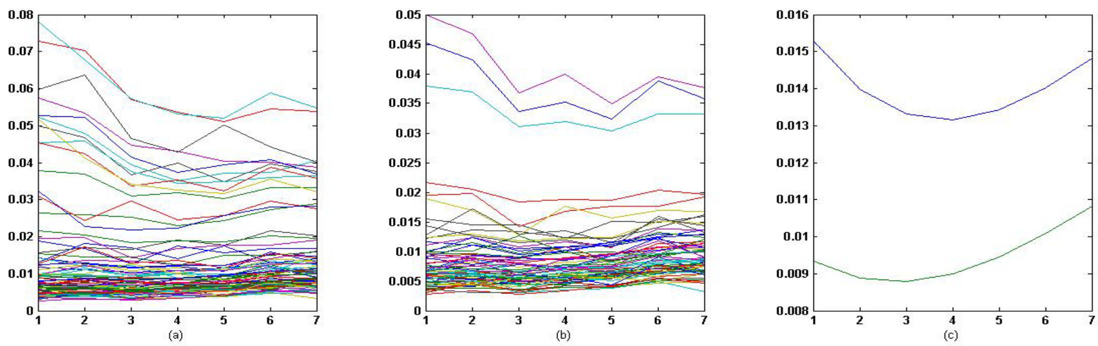

Figure 7. City-based plots on; (a) Fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged 40-44 in 2009-2015

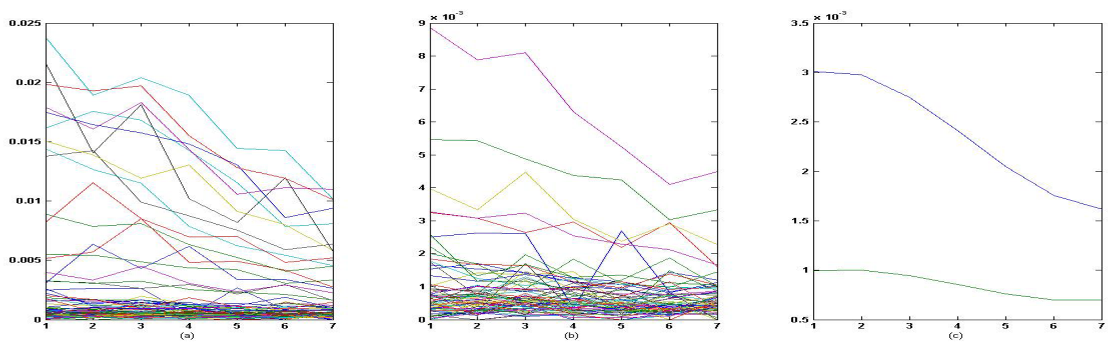

Figure 8. City-based plots on; (a) Fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged 45-49 in 2009-2015

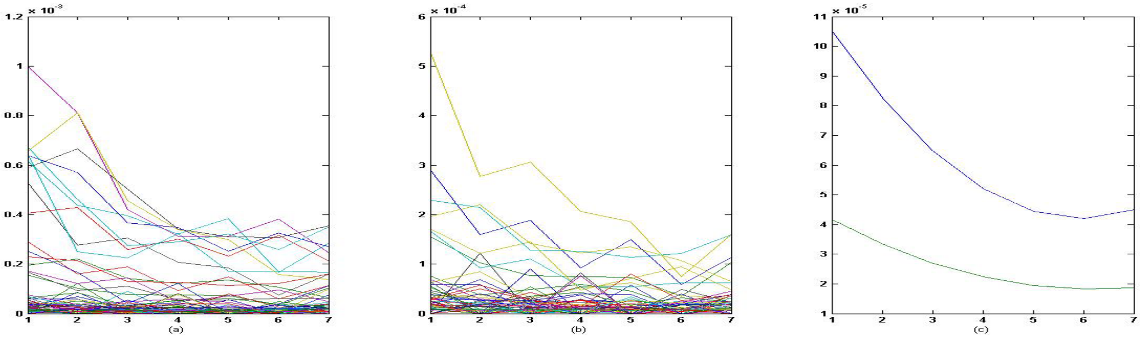

Figure 9. City-based plots on; (a) Fertility rates, (b) fertility rates without outliers, and (c) means of growth curves with respect to "with" and "without" outliers, to women aged ≥ 50 in 2009-2015

5. Conclusions and Discussion

In this study we focused on prediction of future fertility rates by GCM for various age groups in cities of Turkey. Unfortunately, outliers have a great impact on the model parameter estimates and also on future predictions. The effectiveness of these outliers is clearly observed when they are not used in future predictions of fertility rates that are known as the most using indicators of a population growth. Outlying points are found as cities that are weak with respect to socio-economic developments but, have high population rates in contrast to other cities. For this reason, build a GCM that is sensitive to the presence of outliers will be result as estimate the mean of the fertility rates much higher than they are in real.

REFERENCES

[1] L. Angeles, “Children and life satisfaction,” J. Happiness Stud., vol. 11, pp. 523-538, 2010.

[2] M. Atif, I. M. Janjua, A. Riasat, Dr. S. Akhtar, and G. Riasat, “The effect of socio-economic and demographic factors on fertility behaviour in Faisalabad (Pakistan),” GJHSS-Sociology, Economics & Political Science, vol. 12, pp. 7-11, 2012.

[3] R. F. Potthoff and S. N. Roy, “A generalized multivariate analysis of variance model useful especially for growth curve problems,” Biometrika, vol. 51, pp. 313-326, 1964.

[4] E. M. Keramidas and J. C. Lee, “Forecasting technological substitutions with concurrent short time series,” Journal of the American Statistical Association, vol. 85, pp. 625-632, 1990.

[5] J. X. Pan and K. T. Fang, Growth Curve Models and Statistical Diagnostics, Springer Science & Business Media, Springer-Verlag New York, 2002.

[6] J. X. Pan, “Discordant outlier detection in the growth curve model with Rao's simple covariance structure,” Statistics & Probability Letters, vol. 69, pp. 135-142, 2004.

[7] C. R. Rao, “Least squares theory using an estimated dispersion matrix and its application to measurement of signals,” In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, Univ. of California Press, pp. 355-372, 1967.

[8] V. Barnett and T. Lewis, Outliers in Statistical Data, New York, Wiley & Sons, 1984.

[9] R. D. Cook and S. Weisberg, Residuals and Influence in Regression, New York, Chapman and Hall, 1982.

[10] S. Chatterjee and A. S. Hadi, Sensitivity Analysis in Linear Regression, Canada, John Wiley & Sons, 2009.

[11] G. A. F. Seber, Multivariate Observations, New York, John Wiley & Sons, 1984.