Journal of Statistics and Mathematics
ISSN: 0976-8807 & E-ISSN: 0976-8815 Volume 3, Issue 1, 2012, pp.-69-73.
Available online at http://www.bioinfo.in/contents.php?id=85

MAXIMUM LIKELIHOOD ESTIMATION OF CABG PATIENTS BY PARAMETRIC MODELS BASED ON INCOMPLETE AND COMPLETE POPULATION

SALEEM M.[1], ZAFAR MAHMUD[2] AND KHAN K.H.[3]

[1]Centre for Advanced Studies in Pure and Applied Mathematics Bahauddin Zakariya University Multan, Pakistan.
[2]Department of Statistics Bahauddin Zakariya University Multan, Pakistan.
[3]Centre for Advanced Studies in Pure and Applied Mathematics Bahauddin Zakariya University Multan, Pakistan.
*Corresponding Author: Email- colsaleem_2009@yahoo.com, drzafarmahmud@yahoo.com and drkhizar@gmail.com

Received: February 10, 2012; Accepted: March 22, 2012

Abstract- In this paper, a new approach is used for estimation of survival proportions of Coronary Artery Bypass Graft Surgery (CABG) patients by complete population, from its incomplete population. The availability of a complete population may lead to better estimates of the survivor's proportions. Maximum likelihood method, in-conjunction with Davidon-Fletcher-Powell optimization method and Cubic Interpolation method is used in estimation of survivor's proportions of some parametric distributions.
Keywords- CABG Patients, Complete and Incomplete populations, Parametric models (Weibull, Exponential and etc), Maximum likelihood method, Davidon-Fletcher-Powell optimization method and Survivor's proportions.

Citation: Saleem M., Zafar Mahmud and Khan K.H. (2012) Maximum Likelihood Estimation of CABG Patients by Parametric Models Based on Incomplete and Complete Population. Journal of Statistics and Mathematics, ISSN: 0976-8807 & E-ISSN: 0976-8815 Volume 3, Issue 1, pp.-69-73.

Introduction

The coronary Artery Disease (CAD) is a chronic disease, which progresses with age at different rates. CAD is a result of built-up of fats on the inner walls of the coronary arteries. Thus, the size of coronary arteries become narrow and the blood flow to the heart muscles is reduced / blocked. Therefore, the heart muscles do not receive required oxygenated blood, which leads to the heart attack. CAD is a leading cause of death (see William, Stephen, Van -Thomas and Robert [37], John [18], Hansson [13], Axel, Yiwen, Dalit, Veena, Elaine, Catia, Matthew, Jonathan, Edward & Len [3] and Sun and Hoong [33]). The medical scientists; William, Ellis, Josef, Ralph and Robert [37], Heymann [15], Goldstein, Adams, Alberts, Appel, Brass, Bushnell, Culebras, DeGraba, Gorelick & Guyton [11] and Jennifer [17] are of the opinion that CABG is an effective treatment for CAD patients. The medical research organizations like Heart and stroke foundation Canada [14], American Heart Association [2] and Virtual Health Care Team Columbia [35] pointed out that hypertension; diabetes and smoking are main contributing factors for CAD's survivors. Xing, Xiao and Marcello [38] estimated survival after AIDS diagnosis from surveillance data. Collett [7] modelled survival data in medical research The importance of parametric models for analysis of lifetime date has been mentioned by Mann, Schefer and Singpurwala [30], Nelson [31], Cyrus [9], Lawless [26], Klein & Moeschberger [23] and Sridhar and Mun Choon Chan [32].The exponential distribution model has been used by Lee, Kim and Jung [28] in medical research for survival data of patients.The Weibull distribution model has been used for survival analysis by Cohen [6], Gross and Clark [12], Bunday [5], Crow [8], Klein & Moeschberger [23], Lawrencce [27], Abrenthy [1], Hisada & Arizino [16], Lawless [26], David & Mitchel [10] and Lang [24]. In particular, the survival study of chronic diseases, such as AIDS and Cancer, has been carried out by Bain and Englehardt [4], Khan & Mahmud [20 & 21], Klein & Moeschberger [23], Lawless [26] and Swaminathan and Brenner [34] using exponential and Weibull distributions. Lanju & William [25] used Weibull distribution to human survival data of patients with plasma cell and in response-adaptive randomization for survival trials respectively. Lee, Kim and Jung [28] used exponential in medical research for survival data of the patients.

In this study parametric models (Exponential, Weibull, Log-logistic, Gompertz and Logistic distributions) are used to estimate survival proportions for CABG patients. William, Ellis, Josef, Ralph and Robert [37] in 1995 carried out the survival study on incomplete population (progressive censoring of type 1) of CABG patients comprising 2011 patients using Kaplan Meier method [19]. This study also comprises the same data set of 2011 patients. The patients were grouped as male, female, age (non modifiable factors) and hypertension & diabetes (modifiable factors). The patients were undergone through a first re-operation at Emory University hospitals from 1975 to 1993. In this paper, 12 years survival data of same CABG patients, age group under 50 years is considered. Khan, Saleem and Mahmud [22] proposed a procedure, to make an incomplete population (*IP*) a complete population (*CP*). In this paper, the survivor proportions are obtained for the CABG patients of parametric models using Maximum likelihood method, in- conjunction with Davidon-Fletcher-Powell optimization method and Cubic Interpolation method. A subroutine for maximizing log-likelihood function of each model is developed in FORTRAN program to obtain the estimates of the parameters of the model. The survival proportions of *CP* and *IP* for age group under 50 years of CABG patients by the models are presented and their differences are observed.

The method proposed by Kaplan Meier [21] and latter discussed by William [36] in 1995 and Lawless [26] is : $S(t)=\prod_{j:t_j<t}(1-\frac{d_j}{n_j})$, where d_j and n_j are the number of items (individuals / patients) failed (died individuals) and number of individuals at risk at time t_j, that is, the number of individuals survived and uncensored at time t_{j-1}. This method does not take into account the censored individuals c_j completely and thus the analysis is performed on incomplete population (IP). As Khan, Saleem and Mahmud [22] proposed that the censored individuals c_j could be taken into account. The inclusion of splitted censored individuals, c_j proportionally $\left[(1-\frac{d_j}{n_{j-1}-c_j})\times c_j \ and\ (\frac{d_j}{n_{j-1}-c_j})\times c_j\right]$ into known survived, n_j and died individual's d_j respectively makes the population complete. Thus the survival analysis may be performed on the complete population (CP). Klein & Moeschberger [23] and Lawless [26] proposed likelihood function for a survival model, in the presence of censored data. The maximum likelihood method works by developing a likelihood function based on the available data and finding the estimates of parameters of a probability distribution that maximizes the likelihood function. This may be achieved by using iterative method: see Bunday & Al-Mutwali [5] and Khan & Mahmud [20]. The likelihood function for all observed died and censored individuals is of the form: $L(t;\underline{\theta})=\prod_{i=1}^{n}[f(t_i;\underline{\theta})]^{f_{t_i}}\prod_{i=1}^{n}[S(t_i;\underline{\theta})]^{c_{t_i}}$,where f_{t_i} & c_{t_i} are the number of died & censored individuals in interval *i* each of length *t,* $f(t;\underline{\theta})$ is *pdf* in a parametric model with survivor function, $S(t;\underline{\theta})$ & hazard function, $h(t;\underline{\theta})$ *and* $\underline{\theta}$ is vector of parameters say $\underline{\theta}=(\alpha,\beta)$ of the model . To obtain maximum likelihood estimates of parameters of a parametric model using DFP optimization method, we take negative log on both the sides of above equation and therefore by setting $l=-\ln(L(t;\underline{\theta}))$, we get:

$$l=-\sum_{i=1}^{n}f_{t_i}\ln f(t_i;\underline{\theta})-\sum_{i=1}^{n}c_{t_i}\ln\left(S(t_i;\underline{\theta})\right)$$

or

$$=-\sum_{i=1}^{n}f_{t_i}\ln h(t_i;\underline{\theta})-\sum_{i=1}^{n}(f_{t_i}+c_{t_i})\ln\left(S(t_i;\underline{\theta})\right)$$

(As $f(t_i;\underline{\theta})=h(t_i;\underline{\theta})\,S(t_i;\underline{\theta})$) Where, the first sum is for failure and the second sum is for all censored individuals.

Setting $N_{t_i}=(f_{t_i}+c_{t_i})$, where N_{t_i} represents total no of individuals at time t_i we get:

$$l=-\sum_{i=1}^{n}f_{t_i}\ln h(t_i;\underline{\theta})-\sum_{i=1}^{n}(N_{t_i})\ln\left(S(t_i;\underline{\theta})\right) \quad (1)$$

In this study time is partitioned into intervals, which are of unit length *t* starting from zero. Moreover, failures and censoring of the patients occur in each interval *i* of equal length of time *t*, $i=1,2,...,12.$

For complete population the term for censored observations is dropped from the likelihood function.

Application

We present detail application of above methodology for one of the parametric models considered (Weibull distribution). Same procedure is followed for Exponential, Log-logistic, Gompertz and Logistic models. The probability density function (*pdf*) of Weibull distribution is: $f(t;\underline{\theta})=\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$, where $\underline{\theta}$ is vector of parameters *a* and *b* ; *a* is a scale parameter and *b* is a shape parameter; *a* ,*b* and *t* > 0.

The survival and hazard functions of Weibull distribution are:

$S(t;\underline{\theta})=e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$ and $h(t;\underline{\theta})=\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}$

For incomplete population replacing values of the survival and hazard functions of Weibull distribution in equation (1), we get

$$l=-\sum_{i=1}^{n} f_{t_i}\ln\left[\left(\frac{\beta}{\alpha}\right)\left(\frac{t_i}{\alpha}\right)^{\beta-1}\right]-\sum_{i=1}^{n} N_{t_i}\ln\left(e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}\right)$$

or

$$=-(F)\ln\left(\frac{\beta}{\alpha}\right)-(\beta-1)\sum_{i=1}^{n} f_{t_i}\ln\left(\frac{t_i}{\alpha}\right)+\sum_{i=1}^{n} N_{t_i}\left(\frac{t_i}{\alpha}\right)^{\beta} \quad (2)$$

where, $F=\sum_{i=1}^{n} f_{t_i}$ is the total number of failures in a given time.

Differentiating (2) with respect to α and β and simplifying we get

$$\frac{\partial l}{\partial \alpha}=F\left(\frac{\beta}{\alpha}\right)-\left(\frac{\beta}{\alpha}\right)\sum_{i=1}^{n} N_{t_i}\left(\frac{t_i}{\alpha}\right)^{\beta} \quad (3)$$

$$\frac{\partial l}{\partial \beta}=-\frac{F}{\beta}-\sum_{i=1}^{n} f_{t_i}\ln\left(\frac{t_i}{\alpha}\right)+\sum_{i=1}^{n} N_{t_i}\left(\frac{t_i}{\alpha}\right)^{\beta}\ln\left(\frac{t_i}{\alpha}\right) \quad (4)$$

By using (2), (3) and (4) in the DFP optimization method, we find the parameters estimates for which value of the likelihood function is maximum. For complete population we drop the term for censored observations from likelihood function. Same procedure is followed for Exponential, Log-logistic, Gompertz and Logistic models. FORTRAN program for the parameters estimation of each model is developed. The optimal estimates of the scale and shape / location parameters (*a* and β respectively) of exponential, Weibull, log-logistic, gompertz and logistic distributions distribution using *CP* and *IP* of age group under 50 years are obtained by maximizing the log-likelihood function t-ratios of the parameters are given in parenthesis. The values of parameters estimates, t-tarios, log-likelihood function and variance-covariance matrix are given in table: [*see Table 1,2 and 3*]

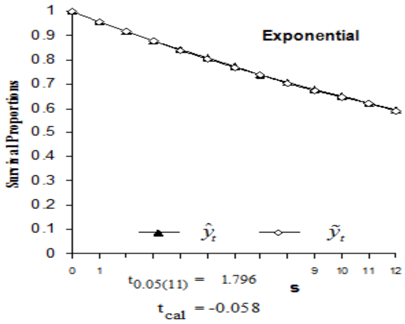
Table 1-

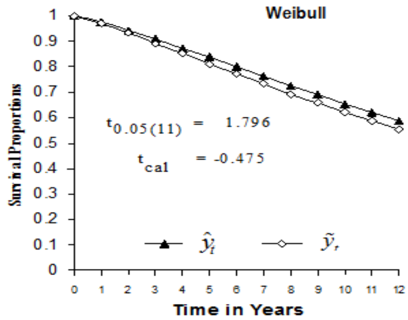
Parameter	Exponential Distribution			
	CP		IP	
	Estimates / (t-ratio)	Gradients	Estimates / (t-ratio)	Gradients
a	4.31×10^{-2} (2.37)	2.11×10^{-7}	4.37×10^{-2} (1.6)	3.48×10^{-7}
Log-Likelihood	418.5939		210.6487	
Variance	3.31×10^{-4}		7.49×10^{-4}	

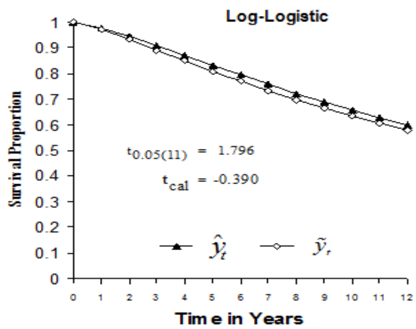
The estimated value of β is greater than 1 (using Weibull distribution) which indicates increasing failure rate with the increase in age. The *t*-ratios in the parenthesis, indicates that estimates of the parameters are significant at 5% level of significance. The positive or negative values of co-variances indicates that the movements of $\hat{\alpha}$ and $\hat{\beta}$ are in the same or opposite direction respectively.

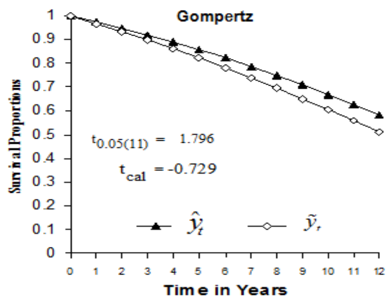
Survival Proportions by the parametric models (Exponential, Weibull, Log-logistic, Gompertz and Logistic distributions) $\hat{y}_t$ and $\tilde{y}_t$ of *CP* and *IP* respectively of age group less than 50 years (Male and Female CABG Patients) and respective graphs are presented as under table 4.

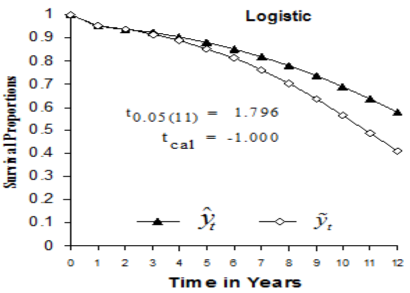
Graphs of the Survival Proportions by models (Exponential, Weibull, Log-logistic, Gompertz and Logistic distributions) $\hat{y}_t$ and $\tilde{y}_t$ of *CP* and *IP* respectively of Age Group under 50 years,











Discussion

The graph of survival proportions (obtained by using exponential distribution) of *CP* and *IP* indicates that for the age group under 50 years of CABG patients, the difference between survival proportions of *CP* and *IP* is almost ignorable. The graph of survival proportions (obtained by using Weibull, log-logistic, gompertz and logistic distributions) of *CP* and *IP* indicates that for the age group under 50 years of CABG patients, the difference between survival proportions of *CP* and *IP* is small at the start, the difference is continuously but slowly increasing, whereas the survival proportions of *IP* of the CABG patients using Weibull, log-logistic, gompertz and logistic distributions are slightly lower then those of *CP*. The differences between the means μ_C and μ_I of survival proportions (obtained by using Weibull, log-logistic, gompertz and logistic distributions) of *CP* and *IP* respectively of age group under 50 years of CABG patients, are tested using *t*-statistic under the null hypothesis H_o : $\mu_I = \mu_C$, against an alternative hypothesis H_1 : $\mu_I > \mu_C$, by using one sided upper tailed test (see Lind, Marchall and Mason [29]). The values of *t*-statistic calculated are 0.058, -0.475, -0.390 and -1.000 respectively, as shown in the respective graphs, when compared with $t_{0.05(11)} = 1.796$, suggest that H_o is accepted which means that the differences between the means of *CP* and *IP* of age group under 50 years are statistically insignificant at 5% level of significance.

Conclusion

The differences between the means of survival proportions obtained by Exponential, Weibull, Log-logistic, Gompertz and Logistic distributions for *CP* and *IP*, of age group under 50 years of CABG patients are statistically insignificant at 5% level of significance. However, the data is best modelled by Exponential and Weibull survival models.

References

[1] Abernathy R.B. (1998) *The New Weibull Handbook.* 3*rd ed*.
[2] American Heart Association Dallas, Texas (2007) *Heart Disease and Stroke Statistics*.
[3] Axel Vise, Yiwen Zhu, Dalit May, Veena Afza, Elaine Gong, Catia Attanasio, Matthew J. Blow, Jonathan. C. Cohen, Edward M.R. bin & Len A. Pennacchio (2010) *Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice*.
[4] Bain L.J. and Englehardt M. (1991) *Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods,* 2*nd ed*.
[5] Bunday B.D. and Al-Mutwali I.A (1981) *IEE Trans*, *Reliability*, 30(4), 367-339.
[6] Cohen A.C. (1965) *Technometrics*., 7(4), 579-588.
[7] Collett D. (1995) *Modelling Survival Data in Medical Research*.
[8] Crow L.H. (1982) *Technometrics.*, 24(1), 67-72.
[9] Cyrus R. Mehta (1981) *Biometrika* 68(3), 669-675.
[10] David G. Kleinbaum, Mitchel Klein (2005) *Survival analysis: a self-learning text*.
[11] Goldstein L., Adams R., Alberts M.L. Appel L. Brass, Bushnell C., Culebras A., DeGraba T., Gorelick P., Guyton J. (2006) *American Journal of Ophthalmology*, 142(4), 716-716.
[12] Gross A.J. and Clark V.A. (1975) *Survival Distribution: Reliability Applications in the Biomedical Sciences Wiley*.
[13] Hansson Göran K.M.D. (2005) *Inflammation, Atherosclerosis, and Coronary Artery Disease*. 352(16), 1685-1695.
[14] Heart and Stroke Foundation Canada (1997) *Heart Disease and Stroke Statistics*.
[15] Heymann C., Von Heymann (2002) *Journal of Cardiothoracic and Vascular Anesthesia*, 16(5), 615-616.
[16] Hisada & Arizino (2002) *IEEE Transactions*, 51(3), 331-336.
[17] Jennifer Heisler R.N. (2008) *After Coronary Artery Bypass Graft Surgery-Recovering From Open Heart Surgery*.
[18] John H. Lemmer (2003) *Hand Book of Patient Care in Cardiology Surgery; Lippincott Williams & Wilkins*.
[19] Kaplan E.L. Meier P. (1958) *J. American. Assoc*., 53, 457-481.
[20] Khan K.H. and Mahmud Z. (1999) *Weibull Distribution Model for The Breast Cancer Survival Data Using Maximum Likelihood Method*.
[21] Khizar H., Khan M. Saleem and Zafar Mahmud (2011) *Survival Proportions of CABG Patients: A New Approch*. 3(3).
[22] Klein P.J. and Moeschberger L.M. (1997, 2003) *Survival Analysis Techniques for Censored and Truncated Data*.
[23] Lang Wu (2010) *Mixed effects models for complex data*.
[24] Lanju Zhang and William F. Rosenberger (2007) *Journal of the Royal Statistical Society: Series C,* 56(2), 153-165.
[25] Lawless Jerald F. Lawless (1982, 2003) *Statistical Models and Methods for Lifetime Data, John Wiley and Sons*.
[26] Lawrencce M. Leemis (1995) *Reliability Probabilistic Modela and Statistical Methods*.
[27] Lee Jaeyong, Kim Jinseog and Jung Sin-Ho (2006) *Lifetime Data Analysis,* 13(1).
[28] Lind A. Douglas, Marchall G. William and Mason D. Robert (2002)*Statistical techniques in Business and Economics*.
[29] Maan N.R., Schafer R.E. and Singpurawalla N.D. (1974) *Method for Statistical Analysis of reliability and Lifetime Data*.
[30] Nelson W. (1982) *Applied Life Data Analysis*.
[31] Sridhar K.N., Mun Choon Chan (2009) *Modeling link lifetime data with parametric regression models in MANETs. IEEE.* 13 (12).

[32]Sun Zhonghua and Hong Ng. Kwan (2011) *World J. Cardiol.* 3 (9), 303-310.
[33]Swaminathan R. and Brenner H. (1998,2011) *Statistical methods for Cancer Survival Analysis.,* 1&2.
[34]William S., Weintraub M.D., Ellis Jonees L., Josef M. Craver, Ralph Grossedwald, Robert A. Guyton (1995,1997) *American Heart Association*.
[35]William S., Weintraub M.D. Stephen D. Clements M.D., Van-Thomas Crisco L., Robert A. Guyton (2003) *American Heart Association*.
[36]Xing Ming Tu, Xiao-Li Meng, Marcello Pagano (1993) *Journal of the American Statistical Association*.

Table 2-

Parameters	Weibull Distribution CP Estimates/ (t-ratio)	Gradients	IP Estimates/ (t-ratio)	Gradients	Log-logistic Distribution CP Estimates/ (t-ratio)	Gradients	IP Estimates/ (t-ratio)	Gradients
a	19.918 (10.31)	1.5×10^{-9}	18.574 (5.88)	-1.277×10^{-9}	1.3727 (11.06)	-2×10^{-8}	1.2785 (8.95)	1.02×10^{-7}
β	1.253 (10.63)	4.43×10^{-8}	1.197 (8.34)	-4.196×10^{-8}	6.23×10^{-2} (9.85)	2.90×10^{-6}	6.49×10^{-2} (6.57)	-9.88×10^{-7}
Log-Likelihood	415.9276		209.6061		418.3742		146.3356	
Variance-Covariance Matrix	$\begin{pmatrix} 3.735 & -0.131 \\ -0.131 & 1.39\times10^{-2} \end{pmatrix}$		$\begin{pmatrix} 9.981 & -0.329 \\ -0.329 & 2.06\times10^{-2} \end{pmatrix}$		$\begin{pmatrix} 1.54\times10^{-2} & 3\times10^{-4} \\ 3\times10^{-4} & 4\times10^{-5} \end{pmatrix}$		$\begin{pmatrix} 2.04\times10^{-2} & 8.74\times10^{-4} \\ 8.74\times10^{-4} & 9.77\times10^{-5} \end{pmatrix}$	

Table 3-

Parameters	Gompertz Distribution CP Estimates/ (t-ratio)	Gradients	IP Estimates/ (t-ratio)	Gradients	Logistic Distribution CP Estimates/ (t-ratio)	Gradients	IP Estimates/ (t-ratio)	Gradients
a	9.46×10^{-2} (3.15)	-5.29×10^{-4}	9.11×10^{-2} (1.88)	2.50×10^{-8}	3.2030 (12.55)	3.26×10^{-11}	3.2865 (11.65)	-1.46×10^{-7}
β	2.41×10^{-2} (4.4)	-3.351×10^{-3}	3.07×10^{-2} (4.01)	2.31×10^{-7}	0.2411 (12.06)	-5.66×10^{-10}	0.3034 (9.11)	5.19×10^{-7}
Log-Likelihood	413.2877		208.974		431.1049		224.8817	
Variance-Covariance Matrix	$\begin{pmatrix} 9\times10^{-4} & -1\times10^{-4} \\ -1\times10^{-4} & 3\times10^{-5} \end{pmatrix}$		$\begin{pmatrix} 2.36\times10^{-3} & -3.07\times10^{-4} \\ -3.07\times10^{-4} & 5.84\times10^{-5} \end{pmatrix}$		$\begin{pmatrix} 6.51\times10^{-2} & 4.70\times10^{-3} \\ 4.70\times10^{-3} & 4.00\times10^{-4} \end{pmatrix}$		$\begin{pmatrix} 7.96\times10^{-2} & 7.56\times10^{-3} \\ 7.56\times10^{-3} & 1.11\times10^{-3} \end{pmatrix}$	

Table 4-

Years (t)	Exponential		Weibull		Log- logistic		Gompertz		Logistic	
	$\hat{y}_t$	$\tilde{y}_t$	$\hat{y}_t$	$\tilde{y}_t$	$\hat{y}_t$	$\tilde{y}_t$	$\hat{y}_t$	$\tilde{y}_t$	$\hat{y}_t$	$\tilde{y}_t$
0	1	1	1	1	1	1	1	1	1	1
1	0.957	0.957	0.976	0.97	0.978	0.971	0.975	0.968	0.951	0.952
2	0.917	0.916	0.945	0.933	0.946	0.932	0.948	0.934	0.938	0.936
3	0.878	0.877	0.911	0.893	0.909	0.890	0.920	0.899	0.923	0.915
4	0.841	0.839	0.874	0.853	0.871	0.849	0.889	0.862	0.904	0.888
5	0.806	0.803	0.837	0.812	0.832	0.808	0.857	0.823	0.881	0.854
6	0.772	0.769	0.800	0.772	0.794	0.770	0.823	0.782	0.853	0.813
7	0.739	0.736	0.763	0.733	0.758	0.733	0.787	0.740	0.820	0.762
8	0.708	0.704	0.727	0.694	0.722	0.698	0.749	0.696	0.782	0.703
9	0.678	0.674	0.691	0.657	0.689	0.665	0.710	0.651	0.738	0.636
10	0.649	0.645	0.656	0.621	0.657	0.635	0.669	0.605	0.688	0.563
11	0.622	0.618	0.621	0.586	0.627	0.606	0.627	0.559	0.634	0.487
12	0.596	0.591	0.588	0.553	0.599	0.579	0.584	0.512	0.577	0.412