A study on diabetic Patients for Using Multivariate Forecasting Models

S. Poyyamozhi[1] and A. Kachi Mohideen[2].

[1]Assistant Professor and Head, Department of Statistics,
Government Arts College (Autonomous), Kumbakonam – 612 002.

[2]Assistant Professor, Department of Statistics,
Periyar EVR College (Autonomous), Trichy - 620 023.

Abstract

A pilot study utilizing the above multi-sensor data series gadget became finished in 2017 involving 40 diabetic sufferers from the Rajah Muthiah Medical College (RMMC) Chidambaram in collaboration with caregivers and scientific doctors. Overall, 37 of the forty topics inside the original examine were monitored for the entire 72 hour study length. The manner in which subjects are labeled is in no way an indication of the total wide variety of subjects who participated inside the pilot look at (e.G. Subject 196 does no longer imply there have been 196 subjects inside the study). The purpose of have a look at to investigate the drug utilization pattern, expertise, attitude, and practice of diabetic patients in rural community of India. Diabetes mellitus (DM) is an vital public health problem in developed countries and increasingly more also in growing international locations. It is a relatively widely wide-spread situation affecting an envisioned 171 million worldwide. Diabetic expertise and abilities to make adjustment to day by day control of medication, meal plan, exercising and different aspect that impact on blood glucose. An significant range of ARIMA time collection models had been estimated on a affected person-by way of-affected person basis the usage of feasible predicting independent variables, and 3 specific time averaging periods for Example for one of the 27 topics as Full Model output for Subject 84 of the overall model output produced through the SPSS software program. The estimation of parameters is as follows and in step with SPSS's Time Series Algorithm Manual (2010).

1. Introduction

India is a growing united states the diabetes mellitus is a prime scientific and public health hassle. The incidence of diabetes mellitus is major among Indian people. Diabetes mellitus is a chronic incurable condition resulting from received deficiency in production of insulin. Diabetes mellitus is a noticeably widespread situation affecting an estimated 171 million individual worldwide. Diabetes is a metabolic syndrome characterised by way of irrelevant excessive blood glucose bring about the form of both low stage of insulin or in the form of extraordinary resistance to insulin effect coupled with insufficient stage of insulin secretion to

compensate. And also study emphasizes the need for comprehensive diabetes about risk factors, complications, diet control, physical activity, regular checkups and screening will go a long way in achieving better control of diabetes and thus reduce the burden due to diabetes complications. The National Program for Control of Diabetes, with the fundamental purpose of improving the treatment outcomes for patients by providing evidence based guidance to physicians and general practitioners

In this paper represents the diabetic patients had been recruited via advertisements and Rajah Muthiah Medical College (RMMC). Patients were given present certificates as an incentive and token of appreciation A 72-hour tracking duration turned into determined for the pilot study, reflecting the top restriction of the period to which the blood glucose display will be broken continuously. Glucose analyzing is extracting every 10 seconds and recorded in the screen as a five minute average. Different the opposite sensors but, no capabilities existed on the time to transmit these records wirelessly.

This study period protected a four-day period - on the primary day, contributors were installation at the rehab sanatorium at a time handy for them, ensuing in a partial day of tracking. The next two days (day 2 and three) had been full days of monitoring. On the fourth day, they returned to the health facility for debriefing, once more resulting in a partial day of tracking. A - hour prematurely interview on the first day turned into used to hook up the numerous sensors and teach topics on their use. A complete hour of this time turned into spent with the blood glucose monitor, owing to its invasive nature (a sensor is implanted underneath the skin via a nurse).

Patients were given the option of wearing remaining devices on their belt (as shown in Figure 3) in a small "fanny pack" style pouch provided, or in their own purse/bag. Outside of recharging the BlackBerry and GPS receiver, the only time patients were instructed to manually interact with the devices was if the BlackBerry issued a long "buzz". In this case, an automated message would appear on screen (generated by the on-board software) instructing patients to manually turn on one of the sensors which may have inadvertently become disabled, recharge a device, or call a research assistant (which was done by Selecting "OK" and using the BlackBerry; no dialing was needed).

2. Models Description

The following subsections introduce fundamental time-collection ideas that serve as the foundation for information evaluation on this paper inclusive of the multistep method of ARIMA modeling and its fundamental concepts. Much of the technical material within the following by using Box et al. (2008), Sir Bernard Law et al. (2008).

Introduced by Box and Jenkins in 1970, the autoregressive integrated shifting average (ARIMA) model can be finished on one variable, or more than one enter variables without a whole lot trouble inside the shape of preprocessing, as is usually the case with different techniques. It is the reason, therefore, of univariate time-series strategies to statistically measure the diploma of this relationship. The widespread shape of ARIMA p,d,q) is:

$$\nabla^{d} y_t = \mu + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

Where, $\nabla^{d} = (1 - B)^{d}$ $(d - \text{order differencing operator})$

$$\emptyset(B) = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right) (p - order\ AR\ operator)$$

$$\Theta(B) = \left(1 - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_q B^q\right) (the\ q - order\ MA\ operator)$$

Here we utilize Predictive Analytics Software (SPSSv20), which has the capacity to carry out ARIMA - additionally known as Box-Jenkins Models - TF fashions. This software program is capable of carry out a number of automated strategies to help with the modeling method.

In this observe a multivariate time-series ARIMA model is hired to keep in mind the effect of independent variable inputs (x, t, i) as well as their preceding disturbances on predicting future values of a dependent variable (y_t) at the same time as controlling for autocorrelation between residuals. The dependent variable tested in detail is the 5-minute common BG degree on a topic-to challenge basis. As such, a transfer function (TF) might be utilized just like a take a look at discuss by means of Helfenstein (1996) where this method become used to have a look at the effect of insulin remedy on BG of a unmarried DM affected human being.

The Identification of ARIMA Parameters (p,d,q) According to Chatfield (1991) a time series is taken into consideration to be deterministic if inherent destiny values are determined by way of a mathematical purpose of its preceding values. Stochastic, or time collection may be explained by way of some probability of distribution and Stationarity or nonstationarity may be decided by using a visual inspection of the time series ACF and PACF residual plots is mentioned in Box (1970b) and Bowerman et al. (1993),

2.1. Auto-covariance and Auto-correlation functions

If the time series is deemed to be stationary, this means the joint probability distribution of any two subsequent random observations of a time series, say for instance, y_t and y_{t+k}, will be the same for any two time periods t and t+k. the covariance between y_t and its value at another time period, y_{t+k} is referred to as autocovarience is discussed by Montgomery (2008) is given by

If the time series is deemed to be desk bound, this means the joint chance distribution of any two next random observations of a time series, say as an instance, y_t and y_{t+k} , might be the equal for any two time durations t and t+k. The covariance among y_t and its cost at over again duration, y_{t+k} is called autocovarience is mentioned through montgomery of alamein (2008) is given by means of

$$\gamma_k = Cov\,(y_t\,, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)]$$

Mean, variance, and auto covariance,

$$\mu_y = E(y) = \int_{-\infty}^{\infty} y f(y) dy$$

$$\sigma_y^2 = Var(y) = \int_{-\infty}^{\infty} (y - \mu_t)^2 f(y) dy$$

$$\rho_k = \frac{e[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]E[(y_{t+k} - \mu)^2]}}$$

$$\rho_k = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} = \frac{\gamma_k}{\gamma_o}$$

$$c_k = \hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k} (y_t - \bar{y})\,(y_{t+k} - \bar{y}), k = 0,1,2, \cdots, k$$

And the autocorrelation function is approximated via the sample autocorrelation feature (as suggested via Box et al. (2008)

$$\gamma_k = \frac{\sum_{t=1}^{T-k}(y_t - \bar{y})\,(y_{t+k} - \bar{y})}{\sum_{t=1}^{T-k}(y_t - \bar{y})^2}$$

Therefore, examining a time series' ACF plot is necessary to ensure stationary.

2.2. Partial autocorrelation function

The PACF serves as a fundamental device of Box-Jenkins time collection evaluation. Used along with the ACF, each may be used to differentiate among lower order and excessive order AR (p) strategies. The PACF works similarly to a partial correlation; where, at A: lags, controls for confounding autocorrelation in intermediate lags. Deriving the partial correlation is supportive in order to understand its source and meaning. To illustrate consider three random variables *X, Y,* and Z Montgomery (2008) illustrates a linear regression of *X on Z and* V on Z as:

$$\widehat{X} = a_1 + b_1 Z \; where \; b_1 = \frac{Cov(Z,X)}{Var(Z)} \;, \qquad \widehat{Y} = a_2 + b_2 Z \; where \; b_2 = \frac{Cov(Z,X)}{Var(Z)}$$

$X^* = X - \widehat{X} = X - (a_2 + b_2 Z\,),\;\; Y^* = Y - \widehat{Y} = Y - (a_2 + b_2 Z\,)\,,$

The partial correlation between X and Y after adjusting for Z can then be defined as the correlation betweenX^* and Y^* ; $corr(X^*,\; Y^*) = corr\big(X - \widehat{X}\,, Y - \widehat{Y}\,\big)$.

The partial autocorrelation function that exists between y_t and $y_{t-k}$is the autocorrelation between y_t and y_{t-k} after adjusting for $y_{t-1}, y_{t-2}, \cdots, y_{t-k+1}$ Therefore, for an AR (p) parameter, PACF between y_t and y_{t-k} for $k > p$ should be equal to zero. The more formal definition, according to Montgomery (2008), is as follows. Considering a time series model (y_t) that is not an AR process. Further consider, for any fixed value of k, the ACF of an AR (p) process is given by

$$\rho(j) = \sum_{t=1}^{k} \emptyset_{tk}\, \rho(j = 1), j = 1,2, \cdots, k$$

To solve for $\emptyset_k$, the equation is, $\emptyset_k = P_k^{-1}\, \rho_k$

For any given k, k=1,2, … , the final coefficient $\emptyset_{kk}$ is referred to as the partial correlation at lag k. for an AR(p) process $\emptyset_{kk} = 0\; for\; k > p$. Thus, when viewing a PACF residual, it is possible to identify when the PACF cuts off at a particular lag, say lag p, for an AR (p). refer to Quenouille (1949), Jenkins (1954, 1956), and Daniels (1956).

The cross correlation coefficient at lag k is estimated by

$R_{xy}(k) = \frac{C_{xy}(k)}{S_x S_y}$

Where,

$$C_{xy}(k) = \begin{cases} \frac{1}{n} = \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t-k} - \bar{y}), & k = 0,1,2,\cdots \\ \frac{1}{n} = \sum_{t=1}^{n+k} (y_t - \bar{y})(x_{t-k} - \bar{x}), & k = -1,-2,\cdots \end{cases}$$

$S_x = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2}$, $S_y = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2}$

The cross correlation function is not symmetric about k = 0. Approximate standard error of $r_{xy}(k)$ is

$$se\left(r_{xy}(k)\right) \cong \sqrt{\frac{1}{n - |k|}}, k = 0, \pm 1, \pm 2, \cdots$$

The standard error is also based on the assumption that the series are not cross correlated and one of the series is white noise. (The general formula for the standard error can be found in Box et al. (2008)). In a time series with y_1, y_2, $\cdots$ y_t where we are interested in the percentage change in y_T is,

$x_t = \frac{100(y_t - y_{t-1})}{y_{t-1}}$

The approximate percentage change in y_t can be calculated from the differences of the log-transformed series $x_t \cong 100[\text{In}(y_t) - \text{In}(y_{t-1})]$ because

$$100[\text{In}(y_t) - \text{In}(y_{t-1})] = 100 \text{ In}\left(\frac{y_t}{y_{t-1}}\right) = 100 \text{ In}\left(\frac{y_{t-1(y_t - y_{t-1})}}{y_{t-1}}\right)$$

$$= 100 \text{ In}\left(1 + \frac{x_t}{100}\right) \cong x_t$$

This is accomplished by subtracting each datum in a time series from its predecessor. That involves applying the difference operator to the original time series in order to obtain a new time series is,

$$x_t = y_t - y_{t-1} = \nabla_{y_t}$$

The Second difference is,

$$x_t = \nabla^2{}_{y_t} = \nabla\left(\nabla_{y_t}\right) = (1 - B)^2 y_t = (1 - 2B + B^2) = y_t - 2_{y_{t-1}} + y_{t-2}$$

Generally, powers of the backward difference operator and the backshift operator are defined as,

$B^d y_t = y_{t-d}$

$\nabla^d = (1 - B)^d$

Differencing directs autocorrelation toward 0 or beyond in a negative direction. When single differencing results in an autocorrelation spike in an ACF residual plot >.5, over-differencing has occurred. Parameter estimates are derived from two possible algorithms discussed at length here as they are beyond the scope of this work, a more detailed review can be found in Yaffee (2000). Results are compared for optimal goodness-of-fit. The parameter estimates should be of reasonable magnitude, and statistically significant with t-ratios > 1.96. Non significant parameters from the model. The general steps for the estimation of parameters are as follows and according to SPSS's Time Series Algorithm Manual (2010):

2.3. Diagnostic Statistics and Sum of Square Errors Mean Square Errors

The model-produced output can then be compared to target values allowing a measure of distance to be calculated. The error functions that follow are derived from Yaffee (2000) were used to determine the best fitting models based on time interval (5, 15 or 30 minute) and predictive independent variables.

While not used explicitly for measuring goodness-of-fit or forecast accuracy in this paper, these measures are used to calculate other substantiating explanatory forecast accuracy measures: root mean square error and stationary r-squared. Sum of square errors (SSE) and mean square errors (MSE) may be used however, to attain a sense of dispersion error. The squares for the entire forecast or time series in order to obtain the SSE:

This modeling produced output can then be in comparison to target values allowing a degree of distance to be calculated. The functions are comply with are derived from Yaffee (2000) have been used to determine the pleasant fitting models primarily based on time c language (5, 15 or 30 minute) and predictive impartial variables.

While no longer used explicitly for measuring goodness-of-match or forecast accuracy in this paper, those measures are used to calculate different substantiating explanatory forecast accuracy measures: root mean square and r-squared. Sum of square errors (SSE) and suggest rectangular errors (MSE) may be used however, to attain a feel of dispersion errors. The squares for the whole forecast or time series to be able to attain the SSE and MSE,

$$\sum_{t=1}^{T} e_t^2$$

$$\sqrt{\sum_{t=1}^{T} \frac{e_t^2}{T-k}}$$

And also root mean square error (RMSE):

$\sum_{t=1}^{T} \frac{e_t^2}{T-k}$

A model with a lower RMSE indicates a good fit. The RMSE serves as an indicator of the difference between predicted and actual values.

Stationary R- Squared

Selecting a model that maximizes the R^2 is the same as choosing the model that minimizes the sum of the foundation imply square error. Large values of the R^2 imply a terrific healthy with the historical facts. However, for the reason that root suggest square blunders usually decreases as parameters are introduced to the modeling, depending totally at the R^2 price to pick out a forecasting model of nice-match helps applying more parameters than are vital to obtain a terrific forecast,

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

Related to the R-squared statistic,

$$SR^2 = \left(1 - \frac{k}{T}\right) r^2$$

The Ljung-Box statistic is done in addition to examining ACF and PACF residuals for spike that may indicate an erroneously inflated Q statistic.

$$Q(k) = n\,(n+2) \sum_{k=1}^{k} r_k^2/(n-k)$$

3. Results and Discussion

In this paper, the main results of subject-by-subject time series modeling analysis are presented. The response of each subject's BG (dependent variable) to possible predictive inputs (independent variables) is assessed using the Linear Transfer Function Modeling Method

described as a background to this, subject profiles and related descriptive statistics are first presented. Overall, 37 of the 40 subjects in the original study were monitored for the full 72 hour study period. The manner in which subjects are labeled is by no means an indication of the total number of subjects who participated in the pilot study (e.g. Subject 196 does not mean there were 196 subjects in the study). The reason why this labeling method was used is unknown to the author. Two subjects ceased monitoring after day 2 (providing 1.5 days monitoring), and one after day 3 (providing 2.5 days of monitoring), all reporting irritation with the tape associated with the blood glucose monitor as the reason. Additionally, four of the first eight subjects GPS data were deemed unusable - owing to a technical problem. Closer examination of the data prior to this thesis revealed a further 6 subjects whose data contained missing values that prevented time series modeling. Four of these had missing food diary data, one had missing heart-rate and accelerometer data, and one had missing activity diary data. This left a total of 27 subjects available for analysis in this paper.

Table 1 presents basic information of these 27 subjects, including age, gender (1 = male, 2 = female), weight, waist size, type of DM, year of DM diagnosis, HBA1C, total number of BG readings, average BG value and the standard deviation of BG. Participants were split evenly by gender, ages ranged between 32 and 74 with a mean age of 56. Participants weighed between 45 and 147 kilograms, with a mean weight of 86.5 kilograms. Average BG for each subject over the course of the study ranged from 5.1 - 13.6 mmol/L. The overall average across all subjects being 7.3 mmol/L, indicating a wide variation, with some subjects demonstrating extremely high BG and other near normal (nondiabetic) BG. The standard deviation of the mean BG ranged from 0.71 - 5.28 with a mean of 2.01, indicating that some subjects have little BG variation while others BG seem to be far less stable.

Table 2 displays the drug type, action and specific name of insulin medication used by each subject. Three types of insulin medication were used amongst subjects: insulin, an insulin sensitizer, and an insulin releaser – insulin sensitizers allow the body to respond more normally to insulin secretion and insulin releasers stimulate endogenous release of insulin. Specific types of insulin medication used by subjects included: Actos, Diabeta Diamicron, Humalin, Humalog, Lantus, Levemir, Novarapid, and Metaformin; each having a different response time varying between slow, medium and rapid. Subjects may have used one, or a combination of different medications

Table-1 Subject profile

User_Id	Hours of Continuous BG Data	Age	Gender	Weight (Kg)	Waist (cm)	Diabetes Type	Diagnosis Year	HBAIC	Blood Clucose(mmol/L)		
61	69.6	64	1	123.1	123	2		8.1	835	9.41	2.12
66	71.8	45	2	106.5	124	2	2003	7.1	861	11.43	3.5
70	101.3	49	2	55	71	2	2002	6.4	1216	7 28	1.25
73	71.8	67	2	59 1	88	2	1997	6.4	861	8.66	2 4
74	71.8	64	1	91.8	100	1	1997	7 8	861	13.59	5.28
77	72.2	69	1	94.5	113	2	2006	8.1	866	6.74	2.27
80	70.8	44	2	59	74	1	1996	7 2	849	7.31	4 7
81	68.6	56	1	104 5	117	2	2016	6.5	823	8.75	153
84	71.7	49	2	67 3	94	2	2012	6.2	860	6 12	1.53
86	73.3	42	1	94.8	112	2	2016	7.9	879	5.13	1.41
87	73.2	52	2	66.8	88	2	2016	5.8	878	7.38	1.81
88	73.0	59	2	53 6	75	2	2016	10	876	6.47	1 6
90	63.3	60	1	99.1	107	2	2016	6.5	765	6 63	1.54
91	70.3	69	2	67.7	92 5	2	2016	5 1	843	6 74	1.74
92	67.4	64	2	82 3	93	2	2008	6.8	809	6.45	1.59
93	72.1	35	2	84.5	97 5	2	1995	5.6	865	5.78	0.75
153	63.3	50	2	129 5	125	Pre	2016	6.6	760	6.65	1.1
163	70.1	55	2	85.9	106	2	2016	5.2	841	7.52	1.54
167	71.3	50	2	127 3	117	2	2016	5.3	856	5.36	0 71
170	49.8	46	1	77 7	93	2	2016	6.2	597	6.41	1.41
171	67.7	66	2	45	83	2	2003	7.7	812	6.48	2.86
172	68.3	68	1	86.4	115	2	2003	7.2	819	6 91	2.87
173	72.0	48	1	93 6	105	2	2006	7.6	864	7.15	1.41
175	75.2	32	1	71	91	1	2006	.	902	6.96	3.57
176	71.5	68	1	90	113	2	2012	.	858	5 78	0 86
177	72.5	51	2	147	109.5	2	2012	.	870	7.61	2.4
196	70.0	65	1	90 5	71	2	2005	.	840	6.12	2.07

Table 2 provides a summary of total subject food and activity data over the course of the 72-hour monitoring period. Carbohydrate consumption ranged from 330-1854 grams with a study

average of 642 grams. Sugar consumption ranged from 46-335 grams with a study average of 187 grams. Total calorie consumption ranged from 2476-14055 with a study average of 5424. Activity diary data also varied widely amongst subjects with the most time spent at home followed by work/school. Total exercise varied between 0-6.8 hours with an average of 2.3 hours. Notice also the number of subjects who did not exercise or exercised very little. Lastly, sugar consumption for subject 84 appears to be unusually high and could be the result of a coding error.

Table 2 the total subject food and activity data over the course of the 72-hour monitoring period.

UserID	Carbohydrates consumed (Total, grams)	Sugar consumed (Total grams)	Calories consumed (Total)	Insulin taken in last 5 minutes (yes/no)		Total hours in diary	Activity time (hours), all types)	Trip time (hours, (all types)	Exercise time (hours) (all types)	Time spent with people (hours)	Activity-At home time (hours)	Activity-Work/school time (hours)	Activity-Shopping time (hours)	Tnp-Automobile time (hours)
61	628 27	240 94	4669 94	778	85	75 8	73 6	22	4 4	22	62 0	00	1 2	22
66	73146	202 46	5474 17	839	22	76 4	69 9	65	29	34 1	55 5	00	0 4	22
70	505 95	115 18	4967 42	1184	32	95 1	93 1	20	4 4	20 0	46 2	17 4	35	19
73	450 77	113 3	2476 24	853	8	1014	99 6	1 9	00	9 3	61 2	00	19	19
74	856 7	335 53	8403 97	838	23	57 1	47 4	97	4 3	52	30 1	10 9	0 4	28
77	454 68	46 74	5073 23	859	7	75 6	72 8	28	20	0 3	55 9	78	0 3	26
80	576 48	200 63	5126 45	829	20	72 5	70 4	21	08	85	60 2	0 0	22	21
81	439 75	133 46	3363 96	819	4	72 2	64 3	79	21	53	38 4	19 5	0 1	58
84	8419	0	6868 4	852	8	63 1	59 8	33	48	61	33 0	11 5	0 1	21
86	602 13	190 64	5175 43	879	0	84 2	77 6	66	68	73	43 0	50	13	3 2
87	453 75	177 46	3995 54	878	0	101 1	96 1	50	39	53	38 2	27 2	00	09
88	43169	137 22	5691 14	876	0	102 1	97 3	48	55	56	47 8	12 0	0 2	07
90	450 34	120 84	3247 32	748	17	65 2	619	33	00	1 7	35 9	22 7	0 2	30
91	364 62	128 66	2534 98	843	0	73 6	72 1	1 5	00	25	59 0	00	0 4	14
92	687 34	235 08	4949 45	772	37	63 7	616	2 1	03	18	58 4	00	2 4	21
93	875 19	315 09	5775 34	859	6	70 4	64 3	6 1	20	79	34 3	24 5	0 5	0 3
153	734 52	193 53	5542 43	752	8	64 1	58 9	52	22	6 3	30 7	17 9	3 4	0 0
163	747 03	220 03	7942 67	841	0	73 2	67 6	56	1 4	13 0	49 1	4 6	20	4 2
167	577 21	137 02	7298 01	846	10	610	59 3	16	20	10	46 8	00	1 5	1 1
170	330 94	63 87	3350 63	597	0	90 0	89 4	06	0 1	10	45 9	00	0 0	00
171	464 03	148 5	4331 78	779	33	74 8	72 0	28	07	10 0	54 5	00	00	28
172	656 71	197 75	5943 99	798	21	716	65 5	61	14	97	44 7	4 6	0 1	58
173	741 6	139 09	5374 27	860	4	72 9	65 3	76	28	25 4	28 6	13 1	30	72
175	1854 08	616 6	14055 1	878	24	96 1	88 6	75	33	21	45 5	3 4	5 1	53
176	476 05	144 78	3626 29	853	5	84 1	80 6	35	22	11 7	55 7	0 0	2 2	10
177	597 54	195 19	5359 09	870	0	71 8	66 8	50	00	79	62 6	0 0	20	50
178	502 25	166 98	4066 16	859	11	67 8	66 6	1 3	00	0 0	30 7	10 5	0 5	1 3

196	963 7	264 52	7197 49	817	23	68 9	63 0	60	4 1	3 2	57 8	0 0	16	38

An extensive number of ARIMA time series models were estimated on a patient-by-patient basis using possible predicting independent variables, and three different time averaging intervals: 5, 15, and 30-minute. Several 15 and 30- minute models were considered adequate, however, after reviewing model statistics and goodness-of-fit measures, 5-minute interval models performed best overall while testing for the predictability of all the independent variables presented in Table 5. Overall, the 5-minute interval models suggested the 'optimal' averaging period for all subjects based on having produced: 1) the most significant BG predicting variables; and *2)* the lowest standard deviation (RMSE) of dependent series differences from its model-predicted level. A summary of the different time interval models' performance is shown in Table 4 (n = 28).

Table 3 Model Comparison

	5 min	15 min	30 min
PREDICTORS	2.17	1.25	0.792
RSME	0.181	0.509	0.609

Table 3 demonstrates the average decrease by subject in the number of BG predicting variables when modeling with larger time intervals. Following this trend is a decrease in the ability to accurately predict future values of BG indicated by larger RMSE values at larger time intervals.

In this study, a multivariate time-series model is employed to examine the effect of a wide range of independent variable inputs, as well as their previous disturbances, on BG, the dependent variable. To do so, a TF using the independent/explanatory variables presented in Table 4 is employed. All variables are measured with respect to their values over the course of 5 minutes and the preceding measurement of BG. For example, "Calories_5m" indicates the number of calories consumed in the 5 minutes period prior to the given BG measurement; "Event23_time_5m" indicates how many of the previous 5 minutes were spent shopping. Vmag is a measurement of the intensity of physical activity and is calculated by taking the square root of the sum of acceleration of the x, y and z-axes. Given past research and expectations, variables reflecting food intake, insulin medication and physical activity are included in the analysis.

Example Full Model output for Subject 84 of the full model output for one of the 27 subjects as produced by the SPSS software. More specifically from Table 6, the model

parameters (p,d,q) for the dependent variable BG are presented. AR(2) indicates the BG value of the current time period is regressed upon the previous two values of itself plus some random error. The differencing and moving average, D(2) and MA(9) indicate that a differencing order of 2 and moving average involving 9 lags were required to achieve stationarity and eliminate autocorrelation between residuals.

Model Description

			Model Type
Model ID	Blood Glucose(mml/l; with Imputed values)	Model_1	ARIMA(2,2,9)

Model Statistics

Model	Number of predictors	Model Fit statistics		Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RSME	Statistics	DF	Sig.	

Model Statistics

Model	Number of predictors	Model Fit statistics		Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RSME	Statistics	DF	Sig.	
Blood Glucose (mmol/L; with imputed values)	1	0.389	0.092	25.868	13	0.018	5

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Blood Glucose (mmol/L;)	No Trans formation	AR	Lag 1	-505	0.034	-14.7	0.000
			Lag 2	-245	0.061	-4.0	0.000
		Difference		2			
		MA	Lag 2	-0.223	0.68	-3.3	0.001
			Lag 7	0.131	0.035	3.8	0.000
			Lag 9	0.087	0.035	2.5	0.014
Average Vmag for Time period	No Trans formation	Delay		3			
		Numerator	Lag 0	-124	0.036	-3.4	0.001
		Difference		2			

Indicates one (1) BG predicting variable (Vmag, which the software produces in a separate ARIMA Model Parameters table) the software found to be significant. The stationary R^2 value serves as a criterion when comparing other competing models and selecting a forecasting model of best-fitted the value presented here translates to the model being able to explain about 39% of the observed variation in the time series. The RMSE serves as an indicator of the difference between predicted and actual values. The Ljung-Box statistic should have significance levels < 0.05 (within the 95% confidence interval) for the time series under analysis. If The autocorrelation is within these bounds, it is not considered to be statistically different from zero and the time series is deemed to be stationary (Ljung et al. 1978). In this case, the significance value of 0.018 does not violate this assumption; additionally, a check of ACF and PACF plots verifies that the assumption of stationarity is upheld.

The t-test results for the dependent variable BG and any predicting variable significant at the 95% confidence level. Significant AR (e.g. AR1 for the first-order autoregressive component $p = 1$), and MA (e.g. MAI for a first-order moving average component where $q = 1$) estimates of these components that is AR or MA and any predicting independent variable reveal which variables meaningfully contribute to predicting future values of the dependent variable with non-significant variables being excluded - this is similar to significance testing of b coefficients for ordinary regression models. The Predicted values are calculated using the linear TF ARIMA equation, repeated here for convenience.

The model fit statistic RMSE, The software calculates the RMSE based on solving for the abovementioned equation to determine predicted values by inputting the significant variable and Adding a TF adds continuous variables to the right-hand side of the time series equation. The objective of adding transfer functions is to see how the independent variable influences the dependent variable rather than simply observing how previous values of the dependent variable are related to itself. The transfer function equation contained a polynomial numerator in the form $\omega(B)_t = \omega_{0,t} - \omega_{1,t}B - \omega_{2,t}B^2 \cdots - \omega_{h,t}B^s$. The numerator parameters are used to establish the magnitude of the effect of a predictor variable, X_t, on the output, y_t.

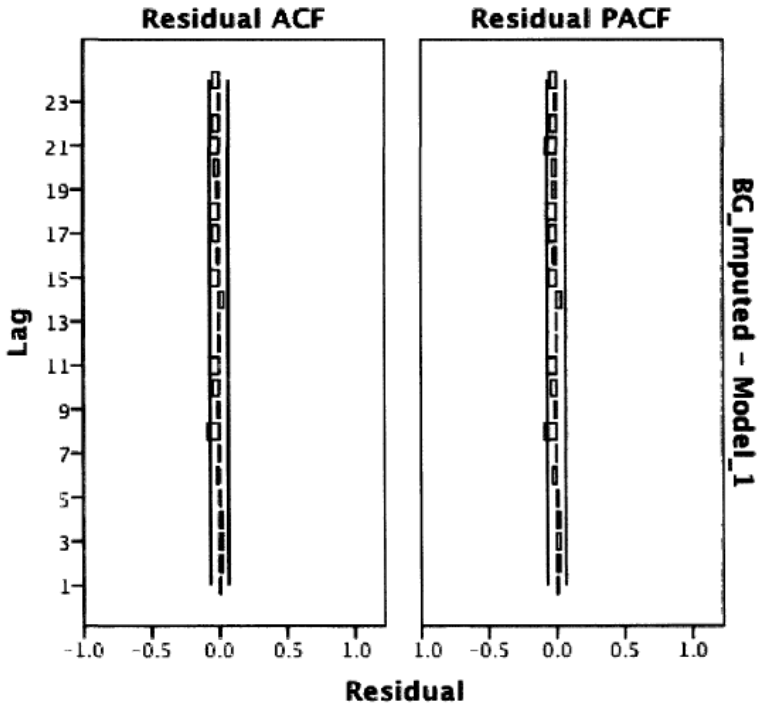

Figure 1 ACF and PACF residual plots for subject 84

Figure 2 shows the ACF and PACF plots for subject 84. Visual examination of ACF and PACF plots serves as a supplementary check to the Ljung- Box statistic to ensure that there's no autocorrelation among residuals in the time series exists. Multiple large spikes past the confidence c program language period restriction might indicate the existence of autocorrelation requiring the re-parameterization of the model - that may be a re-estimate of the (p, d, q) parameters to make certain the assumption upheld. In this example, a spike at lag 8 in each the ACF and PACF plots and lag 21 within the PACF plot appears to have befell randomly with the aid of hazard as residual normality (randomness among residuals) exists for all other lags, thereby upholding the assumption of stationarity rendering the model acceptable. The combination of the numerator and denominator and assessment of the respective CCF therefore decide/affirm the direction of the TF impact on BG, that is, whether or not or not the predicting variable outcomes in an increase or lower in BG (+ or -). Results in imply that the independent predicting variable Vmag has a reducing have an effect on on BG after 3 lags (15 min). This is tested by using examining the CCF of BG and Vmag.

Blood Glucose (mmol / L; with imputed values using LINT(sensor Glucose)) with Average Vmag for time period using average possible accelerometer readings 5m ago

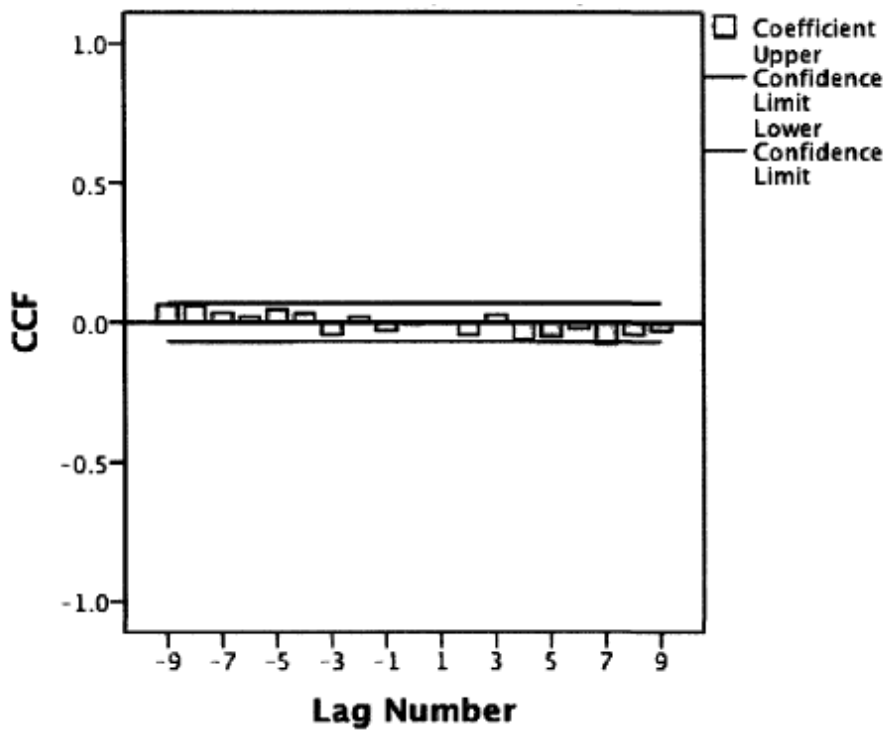

Figure 2 Cross Correlation Function of blood glucose

The overall modeling results for subject 84, as presented in Tables 3,4, Figures 1and 2, demonstrate Vmag as having a diminishing effect on BG after 15 minutes plus some degree of unaccounted for error. the outputs for all individual subject models (in other words, important statistics presented in the full example output from Tables 3,4 and visual verification of results from Figure 2) have been consolidated and presented in Table 4. Each row in the table contains significant modeling components pertaining to one subject. The first row (p, d, q) indicates the AR (p), differencing (d) and MA (q) parameters in addition to any transformation performed and following the Stationary R2, RMSE and Ljung-Box (labeled L-B) statistics and parameter estimates of predicting input variables tested for significance amongst subjects. Numbers in bold represent predicting variables that had a significant effect on BG with the time delay of their impact in parenthesis. A positive number indicates the predicting variable causing an increase in BG, whereas a negative (-) sign indicates the predicting variable causing a decrease in BG.

Table- 4 The represent predicting variables that had a tremendous effect on BG with the time

User_id	(p, d, q)	Predic Input	SR2	RMSE	L-B	Sugar	Curb	Cal	Insulin	Meds *	Vmag	PA	Social	Shop	Dine	Auto	Public Transit	Mean home km
61	(1,1,1)	3	620	085	272	-	.002	-	.05	DM	-	-	-	-	-	-	-	.018
66	(2,1,6)SR	2	386	356	008	-	-	-	-	LMN	-	-	-	-	-	.001	.001	-
70	(0,1,1)	7	247	215	029	-	.023	-	-	M	.345	.001	.001	.001	-	.001	-	-.057
73	(1,1,6)	0	527	258	352	-	-	-	-	AM	-	-	-	-	-	-	-	-
74	(1,1,2)	8	608	282	365	.018	.022	-.	-	HnHg	-	-	-.01	1 00	-	.939	.971	.025)
77	(0,2,13)	0	628	152	149	-	-	.774	-	LN	-	-	-	-	-	-	-	-
80	(1,1,0)SR	4	241	293	177	-	-	-	-	HnN	-	-	.607	-	-.001	.001	-	.015
81	(0,1,2)	3	502	138	000	-	.020	.001	-	M	-	-	-	-	-	-	-	.030
84	(2,2,9)	1	389	092	018	-	-	-	-	M	.345	-	-	-	-	-	-	-
86	(2,1,2)	1	589	089	415	-	.947	-	-		-	-	-	-	-	.001	-	-
87	(0,2,15)	4	723	214	007	-	-	-	-	-	.764	-.01	-	-	-	-	-	.036
88	(1,1,3)	4	616	174	000	.016	.893	-	-	-	.387	-	-	-	-	-	-	-
90	(2,1,2)	0	675	144	295	-	-	-	-	M	-	-	-	.001	-	-	-	-
91	(2,1,12)	3	629	172	043	-	-	.865	-	M	-	-	-.03	-	-	-	-	.166
92	(1,1,14)	0	627	140	021	-	-	-	-	M	-	-	-	-	-	-	-	-
93	(1,1,13)	3	262	153	026	-	-	-	-	-	.181	-.01	-	-	-	-	-	-
153	(1,1,8)	2	433	133	035	-	.004	-	-	M	-	-	-	.002	-	-.001	-	.085
163	(2,1,0)	1	182	244	222	-	-	-	-	-	-	-	-	-	-	-	-	-
167	(0,1,4)	1	622	090	055	-	-	-	-	DM	.738	-	-	-	-	-	-	-
170	(2,1,2)	2	712	152	008	-	-	-	-	DMA	-	-.01	-.88	-	-	-	-	-
171	(1,1,13)N	1	476	217	357	.008	-	-	-	M	-	-	-	-	-	-	-	-
172	(0,2,2)	3	735	114	241	-	-	-	-	Ln	.087	.01	-	-	.740	-	-	-
173	(0,1,4)N	5	619	148	002	-.007	-	-	-	M	-	. 0 1	.001	-	-	.001	-	-
175	(1,1,3)SR	4	681	237	229	-	.706	-.610	.28	-	.129	-	-	.001	-	-	-	-
176	(2,1,0)	2	579	099	078	-	.007	-	-	-	-	-	-	-	-	-	.001	-
177	(0,2,9)	3	793	163	031	-	- .060	-.010	-	-	.616	-	-	-	-	-	-	-
178	(2,1,3)N	1	361	220	006	-	-	-	-	AM	-	-	-	-	-	-.001	-	-
196	(1,1,10)	0	636	129	000		-	-	-	M	-	-	-	-	-	-	-	-

Actos(A), Diabeta(D), Humalin(Hn), Humalog(Hg), NovaRapid(N), Metamorfin(M), Levenur(L), Lantus(Ln)

Stationary R^2 values range from 0.182 - 0.793 with an average of 0.539, where the higher the value the greater the model explanation of variance. RMSE values varied from 0.085 - 0.356 with an average of 0.175, where the lower the value, the greater the models ability to predict for future values of BG. The Ljung- Box values ranged from 0.000 - .415, with an average of 0.123. Values less than 0.05 (within the 95% confidence limit) indicate residual randomness and stationarity amongst the time series. In cases where the Ljung-Box value is greater than 0.05 a visual inspection of ACF and PACF plots verified that indeed these values occurred by chance and that residual normality exists for remaining lags, upholding the assumption of stationarity, thus, deeming the models to be acceptable.

Generally, no two models performed identically. That is, all models presented unique statistical values and *(p, d, q)* parameters suggesting each subject having unique BG fluctuation and

correlating factors. Note particularly, the difference in AR(p) values by subject, referring to the regressed nature (or relatedness) of previous BG values to current BG values indicating perhaps a unique difference in subjects metabolic response to BG change. as well as two basic groups can first be differentiated and those whose BG appears to be little influenced by outside factors and thus appear to have their BG largely under control, versus those whose BG is sensitive to a variety of factors, and thus appear to have less control of their BG. For instance, subjects 86, 88 and 91 who are not medicating appear to demonstrate an increase in BG shortly after eating. Other subjects (61, 70, 74, 81, 153, 176, and 177) appear to be able to control for BG, likely with the help of exogenous insulin that appears to even decrease BG shortly after eating. Also interesting is subject 80 - the only subject to use insulin medication that appears to have an increase in BG shortly after eating and may have something to do with the combination of insulin medication they are using, as they are the only one to use Humalin and Humalog.

4. Conclusion

The exploratory nature of these consequences, and a small pilot examine from a single rehabilitation facility of the pattern of sufferers is small, the number of observations was massive, and greater than 800 measurements of BG and predictor variables in keeping with subject (or >22,000 total observations). This would seem to warrant as a minimum a few exploratory evaluation to study if any patient-with the aid of affected person variability exists despite the fact that now not consultant of all diabetic patients, to test the possible significance of regular and new explanatory variables, to check the usefulness of the evaluation approach, and to at the least offer a way to evaluate the capacity usefulness of large samples inside the future.

References

1. Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, et al.(2011) Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of MedicalResearch-INdia DIABetes (ICMR-INDIAB) study. Diabetologia. 54:3022.
2. Geneva: World Health Organization (2006). World Health Organization. Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycemia: Report of a WHO/IDF Consultation; p. 39.
3. Viswanathan V, Shobhana R, Snehalatha C, Seena R, Ramachandran A (1999). Need for education on footcare in diabetic patients in India. J Assoc Physicians India. ;47:1083–5.
4. Diabetes Control and Complications Trial Research Group. Effect on intensive treatment of diabetes on the development and progression of long term complications in insulin dependent diabetes. N Engl J Med 1993; 329:977-86

5. Most RS, Sinnock P. The epidemiology of lower extremity amputations in diabetic individuals. Diabetes Care 1993;6:87-91
6. Ravi, K., Rajasekaran, S., Subramanian, S., 2005. Antihyperlipidemic effect of Eugenia jambolana seed kernel on streptozotocin induced diabetes in rats. Food and Chemical Toxicology 43, 1433 – 1439
7. Kar, A., Choudhary, B.K., Bandyopadhyay, N.G., 2003. Comparative evaluation of hypoglycaemic activity of some Indian medicinal plants in alloxan diabetic rats. Journal of Ethnopharmacology 84, 105 – 108.