International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:6, No:7, 2012

A martingale Residual Diagnostic for Logistic Regression Model

Entisar A. Elgmati

***Abstract*—**Martingale model diagnostic for assessing the fit of logistic regression model to recurrent events data are studied. One way of assessing the fit is by plotting the empirical standard deviation of the standardized martingale residual processes. Here we used another diagnostic plot based on martingale residual covariance. We investigated the plot performance under several types of model misspecification. Clearly the method has correctly picked up the wrong model. Also we present a test statistic that supplement the inspection of the two diagnostic. The test statistic power agrees with what we have seen in the plots of the estimated martingale covariance.

***Keywords*—**Covariance, Logistic Model, Misspecification, Recurrent Events

I. Introduction

THE paper studies a diagnostic tool based on the covariance between martingale residuals to assess the goodness of fit of the logistic regression model to recurrent event data. This graphical procedure and the test that supplement it, originally were proposed for additive regression model, introduced by Aalen [1], in [2]. They extended the idea put forwarded by Diggle et al. [3] in the context of the longitudinal data subject to dropout. Diggle et al used a martingale random effects approach for continuous longitudinal data and exploited the uncorrelated increments property for diagnostic purposes. A similar approach was developed in [2] for longitudinal binary and recurrent event data.

Using logistic regression model with recurrent event data was presented in [4]. They showed that logistic regression model gives reasonably similar results to that using additive regression model. Also martingale residuals method based on the standard deviation of the standardized residual processes that have been used to judge the goodness of fit for the additive model are shown to be useful for judging the goodness of fit of the logistic model too.

In this work we will define the covariance diagnostic plot, followed by a simulation study to demonstrate its use with the logistic regression model. A formal test that presented in [2], to supplement the inspection of the covariance diagnostic plot and the standard deviation of the standardized residual processes, is used with a simulation study to check the performance of the test. The paper is concluded by applying the test to two data sets, the Blue Bay and Serrinha data.

Entisar A. Elgmati is with the Depatment of Statistics, Faculty of Science, Tripoli University, Tripoli, Libya (e-mail: eelgmati@hotmail.com).

II. Logistic model

At each time point standard logistic regression model is used. A logistic regression model is fitted for the individuals who are at risk at that time using the following equation

$$\lambda_i(t) = Y(t)U(t) \tag{1}$$

where $U(t) = \frac{e^{\beta(t)X(t)}}{1+e^{\beta(t)X(t)}}$

and $\beta(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_p(t))^T$ are the regression functions that need to be estimated, $X(t)$ is the corresponding covariate vector (including an intercept term) and $Y(t)$ is an at-risk indicator which takes value 1 for individuals that are at risk at time t and value 0 otherwise. Note that for the rest of the paper we will not rewrite the model (1) instead we will write the $U(t)$, since this is the part where the changes in the model are.

Here, $\lambda_i(t)$ is the probability that an individual i has an event at time t conditional on all the information that is available up to that time, i.e. the history or filtration

$$\lambda_i(t) = P(dN_i(t) = 1/F_{t^-}) \tag{2}$$

where $dN_i(t)$ is a variable representing the occurrence or non-occurrence of the event at time t for individual i. So the logistic regression assumes that the logit of the probability of an event at time t, conditional on the history up to time t (i.e. F_{t^-}), is a linear function of the covariates and a constant term specific to that time point.

One method of assessing the fit of the model is to plot the empirical standard deviation of standardized martingale residuals defined as

$$M_i(t) = N_i(t) - \Lambda_i(t) \tag{3}$$

Here $N_i(t)$ is the counting process for the *ith* individual and $\Lambda_i(t)$ its cumulative intensity function at time t (a predictable and non-decreasing function). The estimated martingale residual can be used here with $\Lambda_i(t)$ replaced by its estimate $\hat{\Lambda}_i(t)$

$$\hat{M}_i(t) = N_i(t) - \hat{\Lambda}_i(t) = N_i(t) - \sum_{s\in E, s\le t} \hat{\lambda}_i(s) \tag{4}$$

where E is the set of times when the estimation is possible. Also, since $dN_i(t)$ is a binary process, and $dM(t) = dN(t) - d\Lambda(t)$ then the variance of $dM(t)$ is equal to the variance of $dN(t)$.

Following [5] and [6], plotting the empirical standard deviation of standardized martingale residuals will indicate whether the model fits the data well. The closer these are to one, the better the model fits the data. Simulation study and application of this diagnostic to assess the fitted logistic model were presented in [4].

III. Covariance Diagnostic Plot

Another diagnostic tool that fits into the martingale assumption is the covariance diagnostic plot. This diagnostic is based on the idea that, since the martingales have uncorrelated increments then

$$Cov[M_i(t_0), M_i(t)] = Var[M_i(t_0)],\ 1 \le t_0 < t \quad (5)$$

For any fixed time t_0 [3] where $M(t) = N(t) - \Lambda(t)$. Therefore by evaluating the left hand side of (5) at each time point and then plotting these values against t, the plot should be flat if the fitted model is correct. Any departures from a flat line indicate the non suitability of the model being fitted. So when the right model is fitted we expect to get a straight line with zero slop.
More formally, let

$$C(t) = \frac{1}{n}\left[\sum_i M_i(t_0)M_i(t) - n\overline{M}(t_0)\overline{M}(t)\right]$$

Where $\overline{M}(t) = n^{-1}(M_1(t) + ... + M_n(t))$, be the sample covariance of the true martingale residuals for the sampled data and let $\hat{C}(t)$ be the corresponding quantity based on estimated martingale residuals $\hat{M}_i(t)$. We assume t_0 is held fixed and have not explicitly incorporated that into the notation. Then for any $t_0 < t < u$,

$$\hat{C}(u) - \hat{C}(t) = O(\frac{1}{\sqrt{n}})$$

provided the fitted model is correctly specified and estimators are $\sqrt{n}$-consistent.

Next a simulation study will be given to investigate the plot behaviour from fitting different models.

IV. Simulation Study

To demonstrate the use of the covariance plot as a diagnostic tool, a simulation study is carried out. First we looked at the performance of the covariance diagnostic when the true model is correct, i.e. we generate data from a specific model and fit that model only. For example, assume that the true model is of the form (1) with $U(t)$ being

$$K1: U(t) = 0.1 + 0.05x_1 + 0.05x_2$$

where x_1 and x_2 are two time constant covariates. The data were generated with $\tau = 100$ and a sample size of 250. Then the following model is fitted

$$K2: U(t) = \beta_0(t) + \beta_1(t)x_1 + \beta_2(t)x_2$$

Figure 1 shows an average over 100 simulations for the covariance diagnostic plot (dark line) and 10 samples to show sample to sample variability (light lines). From the figure one can notice that the model seems to fit the data very well: the covariances stay constant in time. In addition to this model, we also simulated from other models and fitted the true ones to them. For instance, we simulated from a model with fixed covariates but different covariate structures to Model $K1$ and fitted a fixed covariate model with same number of covariates and same structure.

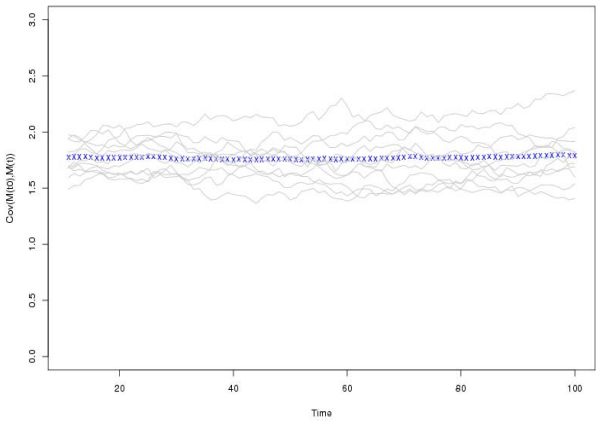

Fig. 1 The average of 100 covariance diagnostic plots for fitting Model $K2$ to simulated data from Model $K1$ (dark line) and 10 samples to show sample to sample variability (light lines).

And we simulated from a model with different dynamic covariates and fitted the true ones. In all of these models, we tried different sample sizes and different parameter values. The covariance and the standard deviation diagnostics for the fitted models behaved very well, with both exhibiting evidence of a good fit (not shown here).

After looking at the performance of the covariance diagnostic when the fitted model is correct, we looked at other situations where the fitted models are not the models that were used to generate the data. For instance assume that the true model is of the frailty form

$$K3: U(t) = Z[0.1 + 0.05x_1 + 0.05x_2]$$

Here x_1 and x_2 again are two time constant covariates and Z, is a frailty random variable, with gamma distribution with mean and standard deviation equal to one. The data were generated with $\tau = 100$ and a sample size of 250. Then the following models were fitted and the whole simulation process was repeated 100 times.

$$A: U(t) = \beta_0(t) + \beta_1(t)\, x_1 + \beta_2(t)\, x_2$$
$$B: U_{D_\infty}(t) = \beta_0(t) + \beta_1(t)\, x_1 + \beta_2(t)\, x_2 + \beta_3(t) R_1(t)$$
$$C: U_{D_{30}}(t) = \beta_0(t) + \beta_1(t)\, x_1 + \beta_2(t)\, x_2 + \beta_3(t) R_2(t)$$
$$D: U_{D_{20}}(t) = \beta_0(t) + \beta_1(t)\, x_1 + \beta_2(t)\, x_2 + \beta_3(t) R_3(t)$$

Here $R_1(t)$, $R_2(t)$ and $R_3(t)$ are the residuals from regressing different dynamic covariates ($D_1(t), D_2(t)$ and $D_3(t)$) on the other covariates to avoid the distortion in the estimates of the effects of the fixed covariates when a dynamic covariate is included ([5] and [6]). These covariates are

$$D_u(t) = \begin{Bmatrix} \{N(t) - N(t-u)/u & t \geq u \\ N(t)/t & t < u \end{Bmatrix}$$

We make the obvious definition that $D_\infty(t) = N(t)/t$ for all t.

Figure 2 (dark lines) shows the average of 100 covariance diagnostic plots from fitting the above models respectively ($A - D$) to the simulated data. The light lines represent a sample of 10 individual simulations to illustrate any sample to sample variability. We used $t_0 = 10$ in these plots.

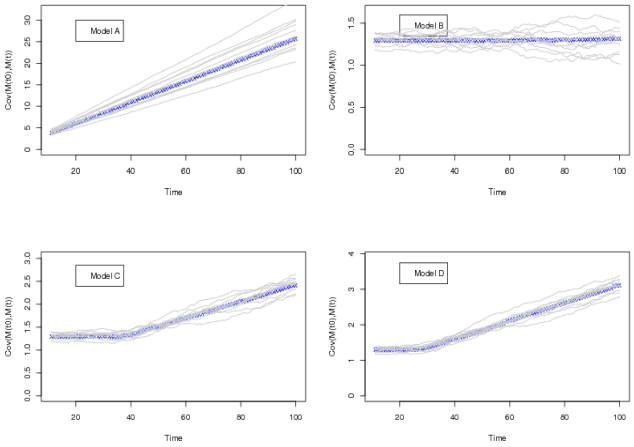

Fig. 2 Blue lines: the average of 100 covariance diagnostic plots for fitting Models A, B, C and D to simulated frailty data ($K3$). Grey lines show a sample of 10 simulations

TABLE II
TEST POWER FOR FITTING MODEL D TO MODEL A, B AND C FOR DIFFERENT SAMPLE SIZES N=250 AND N=500, BASED ON 1000 SIMULATIONS

N=500

	B0=0.1 B1=B2=0.05 B1=B2=0.00833		B0=0.05 B1=B2=0.0125		B0=0.025	
xi	sdt	cov	sdt	cov	sdt	cov
1	100%	33%	100%	100%	100%	100%
0.25	100%	100%	100%	100%	100%	100%
0.09	100%	100%	100%	79%	90%	35%
0.05	100%	90%	87%	30%	60%	15%
0.009	55%	16%	12%	7%	6%	5%
0.0	5%	5%	5%	5%	5%	5%

From the figure one can see that, Models A, C and D, the fixed and dynamic models with $D_2(t)$ and $D_3(t)$ as covariates, did not fit the data well. For Model A the covariance increases linearly as t increases away from t_0. For Models C and D the covariance increases after an initial flat phase. Note that the initial flat phase in Models C and D last up to times 40 and 30 respectively, i.e. 30 and 20 later than $t_0 = 10$. This is because up to this length we have looked all the way back in history so the dynamic model could capture all the frailty and hence provide a good fit to the data. However, Model B the first dynamic model, which always looks all the way back, seems to fit the data very well. Also one can notice that the same conclusions would be drawn from single samples, not just the average.

V. Testing

In this section a simple test statistic that supplements the inspection of plots is presented. Since, when the right model is fitted we expect to have the standard deviation of the standardized residual processes close to one at all times and to get straight line with zero slope for the covariance diagnostic, thus we used a simple Z -test to test whether the successive lag one differences in both plots are zero. For the covariance method (cov) using the first differences is valid because they are nearly independent (see [2] for the justification of the test). Note that the proof outlined in [2] does not assume the differences have equal variances but does make use of their asymptotic independence. However, this is not the case for the standard deviation method (sdt), where the test did not work. This is because lag one differences are correlated. However, a bigger lag were experimented and found that if lag-5 differences for the sdt procedure was taken, then the test seems to work well. This is of course empirical. The following section describes a simulation study to check the performance and the power of the test in several situations.

A. Simulation Study

In this simulation study all our results are based on 1000 simulations in each case. We simulate and fit the correct model first to check the performance of the test. Then different simulation studies are performed. Checking the method for the correct model was done by simulating and fitting different models with different covariate sets. Assume that the true models are

$$A: U(t) = 0.1 + 0.05x_1 + 0.05x_2$$
$$B: U(t) = 0.1 + 0.05x_1 + 0.05y_1$$
$$C: U(t) = 0.1 + 0.05x_1 + 0.05y_1^2$$

Where x_1 and x_2 are binary covariates, and y_1 is standard normal covariate. Then the following models are fitted

$$A1: U(t) = \beta_0(t) + \beta_1(t)x_1 + \beta_2(t)x_2$$
$$B1: U(t) = \beta_0(t) + \beta_1(t)x_1 + \beta_2(t)y_1$$
$$C1: U(t) = \beta_0(t) + \beta_1(t)x_1 + \beta_2(t)y_1^2$$
$$D: U(t) = \beta_0(t) + \beta_1(t)x_1$$

The test performed well for the two diagnostic tools (the standard deviation of the standardized residual processes and the covariance) with empirical test size close to the nominal 5% when the models $A1, B1$ and $C1$ were fitted to data generated from A, B and C models respectively for different sample sizes ($n = 250$ and $n = 500$). Table I gives the test power when Model D (missing covariate) is fitted to data generated from A, B and C models. Note that the test power increases when the sample size increases. For instance, we found 33% power using the covariance test (COV) and up to 93% in the standard deviation test (sdt) for sample size 250 which increased to 100% for the sdt test and 61% for the COV test for sample of size 500. Furthermore, we checked the test size when the covariates are not independent, fitting Model D to Model A with x_1 and x_2 are two binary dependent variables. We found that power goes down to 30% for the sdt test and to 11% for the COV test for a sample of size 250 and increases up to 54% for sdt and up to 15% for a sample of size 500.

Next we simulated from a logistic model with a frailty variable that has gamma distribution with mean one and variance ξ, a sample of size 500, $\tau = 100$ time points and two time constant binary covariates with $U(t) = Z[\beta_0 + \beta_1 x_1 + \beta_2 x_2]$. Table II shows the test power for different values of ξ and different number of events determined by β_0, β_1 and β_2.

TABLE III
P-VALUES FOR TESTING BOTH DIAGNOSTIC FOR THE TWO DATA SETS (BLUE BAY AND SERRINHA DATA).

		Prevalence		Incidence	
Data	Model	sdt	cov	sdt	cov
Blue bay	Fixed	0.000	0.000	0.000	0.000
	dynamic	0.000	0.452	0.228	0.661
Serrinha	Fixed	0.000	0.000	0.000	0.000
	dynamic	0.000	0.721	0.035	0.922

From the table one can see that both tests have a good power to detect the wrong model even for a small value of ξ, with the test based on the standard deviation of the standardized processes (sdt) being more powerful. Power is very high for even very small amounts of frailty for these recurrent event data. As expected the number of events per person affects the test power. When the number of events is decreased the test power dropped. To check the effect of the sample size n, the above simulation study was repeated at sample size 250 (not shown here). Both tests still have very good power to detect the wrong model. However, the power of the two tests is lower at the smaller sample size, as expected. Again the sdt test in particular has very high power to detect even small amounts of ignored frailty. Also decreasing the number of events per individual, decreases the power of the test dramatically and vice versa.

B. Application

In [2] the diagnostic plots for the two real data sets ,Serrinha and Blue Bay data (details of the two data sets are available [7], [8], and [9]), from fitting additive model were given. Using logistic model to fit the data gave similar plots (sdt and COV plots) for both data sets. Here we use our proposed test to assess whether various fitted logistic models are consistent with the data. Table III shows the P-values for the tests. From the table one can see that for both data sets, Serrinha and Blue Bay data, both tests were highly significant for the fixed effects only models: there is strong evidence against the models which ignore previous events. This supports the conclusions from inspection of diagnostic plots (not shown here). There is no such evidence of misspecification once the dynamic covariates are included in the COV test; the result has high P-value for both data sets and for prevalence and incidence. But this is not the case with the sdt test where one can notice that the test is highly significant when we fit a dynamic model to the prevalence data for both data sets. It is also significant for the incidence model in the Serrinha data. Recall, however that our choice of lag-5 differences is based on simulations and the test is not fully theoretically justified. Hence we need to be cautious in interpretation.

VI. Conclusion

One way of assessing the fit of logistic regression model to recurrent event data is by plotting the empirical standard deviation of the standardized residual processes. In this paper we used another graphical diagnostic plot based on martingale residual covariances. Also A formal test to supplement the inspection of the two diagnostic plots (standard deviation of the standardized residual processes and the covariance) for checking the model adequacy was also presented. Furthermore we looked at the expected covariance behaviour under several misspecified models, increasing with t when the wrong model is fitted and being flat with zero slope for the right model. These results were also supported by the test. Also we tested our fitted models, i.e. fixed and dynamic models, for the Blue Bay and Serrinha data and found using the COV test that the dynamic models cannot rejected.

The standard deviation of standardized residual procedure has the advantage that we know the estimates should be close to one under the model, whereas all we know for the covariance diagnostic is that there should be no trends with time. On the other hand the covariance diagnostic also has the advantage of asymptotically independent increments and hence the availability of the test statistic. One should keep in mind that the test has been justified for the additive intensity model where the martingale theory is well founded. However, the diagnostic plot and the test seem to work well with logistic regression model.

Acknowledgment

My thanks go to Rosemerire Fiaccone and Mauricio Barreto for providing the diarrhoea data.

References

[1] O. Aalen, A model for non-parametric regression analysis of counting processes. Lecture Notes in Statistics, no. 2, pp. 1-25, 1980.

[2] E. Elgmati, D. Farewell, and R. Henderson, "A martingale residual diagnostic for longitudinal binary or recurrent event data," *Lifetime Data Analysis*, vol. 16, pp. 118-135, 2010.

[3] P. Diggle, D. Farewell, and R. Henderson, "Analysis of longitudinal data with drop-out: objective, assumptions and a proposal (with discussion)," *Journal of the royal statistical society*: *Series C (Applied Statistics*), vol. 56, pp.499-550, 2007.

[4] E. Elgmati, " Logistic regression model versus additive model for recurrent event data," in Proc. 3rd ICMLC, Singapore, Feb. 2011.

[5] J. Fosen, O. Borgan, H. Weedon-Fekjaer, and O. Aalen, "Dynamic analysis of recurrent event data using the additive hazard model," *Biometrical Journal*, vol. 48, pp. 381-398, 2006.

[6] O. Aalen, J. Fosen, H. Wedon-Fekjaer, O. Borgan, and E. Husebye, "Dynamic analysis of multivariate failure time data," *Biometrics*, vol. 60, pp. 764-773, 2004.

[7] O. Borgan, R. L. Fiaccone, R. Henderson, M. L. Barreto, "Dynamic analysis of recurrent event data with missing observations, with application to infant diarrhoea in Brazil ," *Scandinavian Journal of Statistics* , vol. 34, pp. 53–69, 2007.

[8] E. Elgmati, R. L. Fiaccone, R. Henderson, and M. Mohammadi, "Frailty modeling for clustered recurrent incidence of diarrhea," *Statistics in Medicine*, vol. 27, pp. 6489-6504, 2008.

[9] M. Barreto, L. Santos, A. Assis, M. Araujo, G. Farenzena, P. Santos, and R. Fiaccone, " Effect of vitamin A supplementation on diarrhoea and acute lower –respiratorytract infections in young children in Brazil," Lancet, vol.344, pp. 228-231, 1994.