Paper 181-2007

Using SAS® Proc NLMIXED for Robust Regression

Steven A. Gilbert, Rho Inc., Newton, MA
Ling Chen, Genzyme Corporation, Cambridge, MA

ABSTRACT

We show how the NLMIXED procedure can be used to fit linear models (simple linear regression, multiple linear regression, analysis of variance and analysis of covariance) when data are not normally distributed or contaminated with potential outliers. This is accomplished using PROC NLMIXED's capability of fitting user defined distributions through the general log likelihood option in the Model statement. This methodology can offer advantages over the M-estimation option in Proc ROBUSTREG which cannot analyze repeated measures and Proc NPAR1WAY which only provides p-values and cannot adjust for covariates or handle repeated measures. We provide some the basic theory and directions for specifying the likelihoods needed and then demonstrate the usefulness of the method on several examples.

INTRODUCTION

Linear models are often fitted using the REG, GLM, GENMOD, and MIXED procedures. With the exception of the GENMOD procedure the remaining procedures always assume that the data, Y, can be modeled as

Y= mean function + error,

where the errors have a normal distribution. When the data consist of independent observations the procedures will use either ordinary least squares (OLS) or weighted least squares (WLS), if weights are specified in the procedure call. The normality assumption and presence of outliers can be assessed by examining the residuals from the model. All of these procedures supply raw residuals, calculated as the difference between the observed data value and the estimated mean function,

residual=Y-estimated mean function.

The raw residuals can then be examined for non-normality and for outliers by graphical means and with Proc UNIVARIATE.

It is not uncommon to find that the residuals display one of the following: heavier tails than expected for a normal distribution; skewness; or a small number of larger than expected residuals, which may reflect outliers. In this paper we focus on the presence of outliers and heavy tailed residual distributions.

There is no single approved definition of an outlier; we will use the intuitive but imprecise definition that an outlier is an observation that does not fit in with the bulk of the data. In a regression model outlying data points are often identified as observations with residuals that are much larger (in absolute value) than the remaining residuals. Note that residuals are dependent upon the model. It is possible for apparent outliers to be produced by an incorrect model and can be corrected by replacing an incorrect model with the correct model. Another possibility is that the data have been contaminated in some way. For example, in a clinical trial we may need to measure a subject's blood pressure. There can be measurement error due to equipment malfunction or operator error, or the blood pressure may simply be written down incorrectly on the Case Report Form. It is also possible that one or two subjects decide to go on an espresso and cappuccino binge artificially raising their blood pressure.

Heavy tailed distributions are another difficulty that often arises in real world data. The term "heavy tail" is often used in comparison to the relatively "light tail" of a normal distribution. When plotted, the probability density function (pdf) of a heavy tailed distribution will exhibit thicker tails than a normal distribution, that is, the heavy tailed distribution allows for a higher probability of relatively extreme events. Almost all observations generated from a normal distribution will be within plus or minus three standard deviations from the mean (99.8% of the time to be precise). Heavy tailed distribution allow for a larger number of observations far from the mean. Heavy tailed distributions are not uncommon. They may be either an intrinsic characteristic of the data or they may be a byproduct of the analysis method. For an example of the first kind, many types of financial data empirically demonstrate heavy tails (Bouchaud and Potter, 2000). An example where the analyst creates a heavy tailed distribution can be found in clinical trials. Many clinicians like to look at percent changes, 100% X (post baseline- baseline)/baseline. If some baseline values are close to zero, extremely large percent changes (either positive or negative) can result. This is problematic enough, but if the measured variable can only take on positive values, the percent change variable can also be highly skewed.

If outliers are suspected you have several options. The suspected outliers can be omitted from the model one at a time and then in groups to see their affect on parameter estimates and p-values. (PROC REG offers an impressive array of statistics to help in this task using the INFLUENCE option in the model statement.) As an alternative outliers can remain in the data set but a statistical analysis method insensitive to outliers can be used. The ROBUSTREG procedure in version 9 offers several methods that are insensitive to outliers (see the SAS documentation for PROC ROBUSTREG for further details). ROBUSTREG also incorporates methods for locating outliers using the DIAGNOSTICS option in the MODEL statement. Another option is to use PROC NPAR1WAY to perform a non-parametric analysis of variance. Among the tests available in NPAR1WAY are rank based tests such as the Wilcoxon test, which offers protection against outliers. To see how this is so, note that the Wilcoxon test can be approximated by ranking the observed data from smallest to largest and performing a standard ANOVA on the rank transformed data. If there is a need to adjust for covariates it is possible to perform a rank based analysis of covariance using the RANK procedure and the FREQ procedure (see Stokes, Davis, and Koch 2000). Rank based tests provide protection against outliers. For example, if there are 99 data points that range from 0 to 99 and a hundredth observation with a value of 10,000, it will be ranked 100. Even if the 100th data point has a value of 1,000,000 its rank is still only 100 so that its actual magnitude will not change the inference. We note that there are methods for obtaining parameter estimates from rank based analyses of the residuals of a linear model, R-estimation, (Terpstra and McKean, 2005) but do not pursue those methods further.

A heavy tailed distribution of residuals can be diagnosed with PROC UNIVARIATE. A histogram of the residuals can be produced with an overlaying fitted normal curve and an overlaying kernel density estimate. If the tails of the kernel density estimate are higher than the normal estimate, this may indicate a heavy tailed distribution. Quantile – quantile plots and probability plots can also be created. In a small to medium sized data set it is difficult to distinguish a heavy tailed distribution from a normal distribution contaminated with outliers. In general, a heavy tailed distribution is not as problematic as outliers. In many cases we are concerned with the normality of the parameter estimates in our model, not the residuals. Under a wide variety of conditions some version of the central limit theorem may ensure that the parameter estimates are approximately normally distributed even if the data and the residuals are not normally distributed (see White 2001). However, we may not be able to completely verify all of the relevant conditions and we may not be sure that our sample size is sufficient. In these cases it is often advisable to perform a more robust analysis to see if the results are consistent with the original normal theory based analysis.

ROBUST ALTERNATIVES IN NLMIXED

We consider three classes of error distributions that offer a robust alternative to the normal distribution; t-distributions, generalized error distributions (GED), and Tukey's contaminated normal distribution. These distributions are robust in the sense that outliers have less of an effect on the estimated mean (regression) function than on models based on normal distributions. At an intuitive non-mathematical level the use of heavy tailed distributions allows for a small number of large residuals. Since a normal distribution forces the residuals to be within a few standard deviations of the mean outliers can adversely affect the estimated mean, biasing the estimate, and or greatly inflate the estimated standard error of the mean, losing power and increasing the width of confidence intervals. Coincidently, these same distributions are also useful for heavy tailed data as well.

t-DISTRIBUTIONS

t-distributions offer a family of distribution with heavier tails than the normal, see Figure 1. The top curve represents a normal distribution with standard deviation 1 and mean 0. The curves for the t-distributions are lower in the center and have thicker tails. Thicker tails are associated with lower degrees of freedom. As the degrees of freedom increase the distributions more closely resemble a normal distribution. A t-distribution with an infinite number of degrees of freedom is theoretically equivalent to a normal distribution and in practice a t-distribution with 30 or more degrees of freedom is almost indistinguishable from a normal distribution. A t-distribution with 1 degree of freedom is equivalent to a Cauchy distribution. Therefore, changing the degrees of freedom allows you to see the affect of fitting a heavy tailed Cauchy distribution (1 degree of freedom) or a light tailed normal distribution (large number of degrees of freedom) and anything in between the two extremes.

These distributions have three parameters, one each for the mean, scale, and degrees of freedom. The mean parameter controls where the center of the distribution is located, just as in a normal distribution. The scale parameter helps to control the width of the distribution in conjunction with the degrees of freedom. The variance of a t-distribution with a scale parameter σ and degrees of freedom v is $\sigma^2 \nu/\nu - 2$, when v is greater than 2. If the degrees of freedom are 2 or less the variance is not defined, this is property of a very heavy tailed distribution. An excellent reference for the use of t-distributions for robust modeling is in Lange, Little and Taylor (1989).

The following code estimates the mean for a sample using a t-distribution:

```
proc nlmixed data=d1;
    parms beta0 10 sigma2 3 dft 5;;
    pi=arcos(-1);
    mu=   beta0;
    z=(y-mu)/sqrt(sigma2);
    logl=lgamma(.5*(dft+1))-.5*log(sigma2)-.5*log(dft*pi)-lgamma(dft/2)
                -( (dft+1)/2 )* log( 1+ (1/dft)*z**2);
    model y ~ general(logl);
run;
```

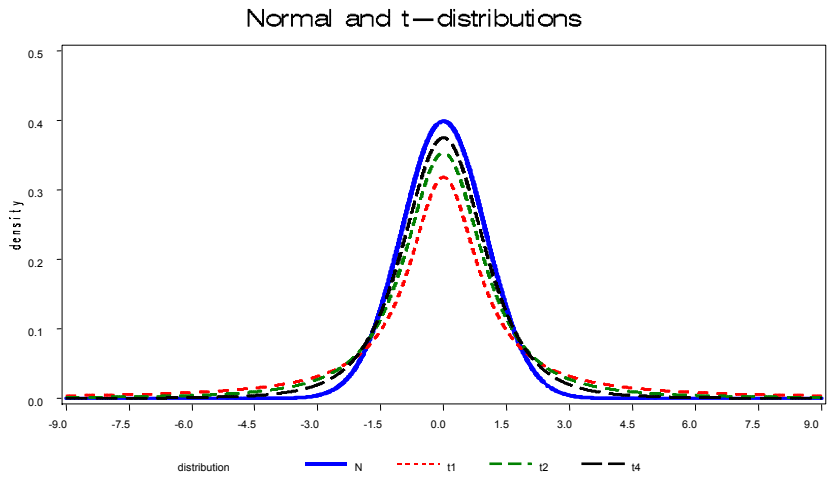
In this code, mu represents the mean, sigma2 is the squared scale parameter, dft is the degrees of freedom, y represents the observed data and beta0 is the single regression parameter. The PARMS statement provides starting estimates for the function optimization routine built in PROC NLMIXED (values from a normal theory based regression are often adequate). It is good practice to run the code several times using different starting values in the PARM statement to make sure that the procedure finds a global maximum for the log likelihood function .

The code can be modified in different ways. A regression model can be estimated by including the covariates and additional regression parameters in the line of code defining the variable mu. For example if the data had three covariates named cov1, cov2, and cov3 we could use:

```
mu=   beta0 + beta1*cov1 + beta2*cov2+beta3*cov3;
```

and the beta parameters beta1, beta2, and beta3 should also be added to the PARMS statement. Another useful modification is to fix the degrees of freedom at a specific value because estimating this parameter can sometimes be difficult. We have used 4 degrees of freedom as a default value.

Figure 1

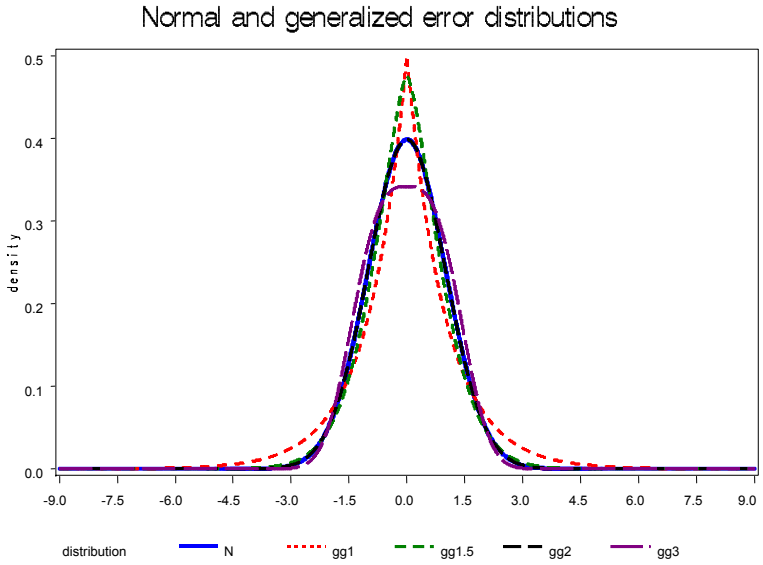


GENERALIZED ERROR DISTRIBUTION

The generalized error distribution (GED) is also known as the exponential power distribution and the Subbotin distribution. We use the parameterization of Mineo and Ruggieri (2005). This distribution also has three parameters, but instead of degrees of freedom there is a shape parameter *p*. When the shape parameter is set to 2 we have the normal distribution. A shape parameter of 1 obtains the Laplace (or double exponential distribution). Unlike the t-distribution we can use shape parameters that greater than 2; that result in distributions that are even more sensitive to outliers than the normal distribution. This is a useful method of assessing the impact of outliers on estimates. Unfortunately, we have not been successful in reliably estimating the shape parameter and instead estimate the mean and scale parameters for a pre-selected shape parameter. Several examples from this family of distributions are plotted along with a normal curve for reference in Figure 2 The normal distribution is the heavy blue line, as the shape parameter varies the tails can be made heavier or lighter than those of a normal distribution. Unlike the t-

distribution, heaver tailed members of this family are more peaked at the center but drop down quicker than a normal distribution. The purple curve with a lower peak than the normal distribution has a shape parameter of 3 and lighter tails than a normal distribution.

Figure 2



Fitting these distributions can be more computationally demanding than fitting t-distributions. We use two versions of the code depending upon the value of the shape parameter. When the shape parameter is greater than 1 the following code can be used:

```
proc nlmixed data=d1;
    parms beta0 0 sigma 40   ;
    pi=arcos(-1);
    p=1.5;
    mu=   beta0;
    a=sqrt(sigma*gamma(1/p)/gamma(3/p));
    z=(y-mu)/a;
    logl1=-lgamma(1+1/p)-log(a)-abs(z)**p;
    model y ~ general(logl1);
    estimate 'scale' sqrt(sigma);
run;
```

When the shape parameter is exactly equal to one and a Laplace distribution is used, numerical difficulties arise because the function is not differentiable at z=0. We approximately fit the function using the following code which includes an arbitrary small constant intended to bound the function away from the non-differentiable point at z=0, see Hitomi (2001):

```
proc nlmixed data=d1;
    parms beta0 10 sigma1 3831   ;
    pi=arcos(-1);
    p=1;
    mu=   beta0;
    a=sqrt(sigma1*gamma(1/p)/gamma(3/p));
    z=(y-mu)/a;
    logl1=-lgamma(1+1/p)-log(a)-sqrt((z)**2+.000001)  **p  ;
    model y ~ general(logl1);
run;
```

Note that we are finding parameters that minimize the sum of the absolute values of the residuals instead of the sum of the squared residuals as is done in least squares and normal theory regression. This type of regression is also known as least absolute value (LAV) regression and can also be performed in SAS IML.

TUKEY'S CONTAMINATED NORMAL DISTRIBUTION

The contaminated normal distribution is a simple mixture of two normal distributions. The pdf can be written as

$$f(y)= p\varphi((y-mu)/\sigma\_small)+(1-p)\ \varphi((y-u)/\sigma\_large),$$

where φ( ) is the standard normal pdf, mu is the mean, σ_small and σ_large are standard deviations, and p is the mixing proportion. In words for some proportion of the time p, between 0 and 1, observations are drawn from a normal distribution with mean mu and standard deviation σ_small, and the for remaining (1-p) proportion of the time outlying observations are drawn from a normal distribution with mean mu and standard deviation σ_large. An example of PROC NLMIXED code that can be used for fitting this model is:

```
proc nlmixed data=d1 tech=newrap maxit=500 maxfunc=500;
    parms beta0 10 sigmaS2 40  sigmaL2 100  zp 0  ;
    pi=arcos(-1);
    p=exp(zp)/(1+exp(zp));
    mu=   beta0;
    z1=(y-mu)/sqrt(sigmaS2);
    z2=(y-mu)/sqrt(sigmaL2);
    logl=log( p/sqrt(sigmaS2)*exp(-.5*z1**2) +(1-p)/sqrt(sigmaL2)*exp(-.5*z2**2) );
    model y ~ general(logl);
    estimate 'p' p;
    estimate 'sigma small' sqrt(sigmaS2);
    estimate 'sigma large' sqrt(sigmaL2);
run;
```

There are known difficulties fitting mixture models. One such difficulty is that the way we have coded the model does not force sigmaL (σ_large) to be greater than simgaS (σ_small). However, we have used two different starting values in the PARM statement one larger than the other. Another difficulty is that the mixing proportion must be between 0 and 1, in order to assure this we have defined the parameter ZP which can take on any value and transformed it with the code:

```
p=exp(zp)/(1+exp(zp));
```

to force the mixing parameter to be between zero and one. An interesting byproduct of this model is that the value of p, is an estimate of the proportion of outliers in the data set. Note that we have specified the Newton-Raphson optimization technique, `tech=newrap`, (the quasi –Newton algorithm `tech=QUANEW` is the default) and have increased the number of allowable optimization iterations and function calls, `maxit=500 and maxfunc=500`, above the default settings.

EXAMPLES

TOY EXAMPLE

We start with a toy example in order to gain some intuition for how these methods work in a simple example before proceeding to the more complicated examples to follow. For this example we estimate the location/mean from the data:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 300.

The value of 300 is intended to represent a gross outlier. If we are looking to estimate a mean/location parameter we intuitively expect it to be less than 20 if it is going to represent a typical value and it is clearly not equal to zero. We present the results from fitting different alternatives in PROC GLM and PROC NLMIXED in the Table 1.

A few observations are in order. The model fit in GLM using standard normal theory regression for independent data (i.e., ordinary least squares) returns the arithmetic mean of 24.286. The arithmetic mean does not however provide a good location value in this problem since it does not provide a reasonable estimate of the bulk of the data which ranges from 1 to 20. In addition the estimated mean of 24.286 is not considered statistically significantly different from zero, p=0.0947 because the outlier has forced us to fit a normal distribution with a large standard deviation, 63.437, which in turn increases the standard error of the mean and loses statistical power for hypothesis testing. Next, PROC GLM was run again dropping the outlier, the mean estimate of 10.524 and is now highly significant, p<.0001, because the estimated standard deviation (scale parameter) has been greatly reduced to 5.916. For comparison PROC ROBUSTREG was run using M-estimation with the Huber weight function and Huber method of estimating the scale parameter. The estimate of 11.000 is highly significant, p <.0001.

Table 1

Model (Proc)	μ	Standard Error	Prob (μ ne 0)	Scale parameter	Degrees freedom	σ_large	Mixing proportion
Normal (GLM)	24.286	13.843	0.0947	63.437			
Normal without outlier (GLM	10.500	1.323	<.0001	5.916			
M-estimate – Huber (ROBUSTREG)	11.000	1.503	<.0001	7.098			
t-estimated df (NLMIXED)	10.524	1.672	<.0001	25.093	1.429		
t-4 df (NLMIXED)	10.543	1.585	<.0001	37.851	4		
GED p=1 (NLMIXED)	11.000	0.586	<.0001	26.196			
GED p=1.5(NLMIXED)	13.159	6.335	0.0502	44.085			
GED p=3 (NLMIXED)	63.147	17.337	0.0015	80.952			
Contaminated Normal (NLMIXED)	10.516	1.294	<.0001	5.762		174.50	0.9493

The next two models fitted are based on t-distributions: the first model uses an estimated degrees of freedom, 1.429 in this case; and the second model uses a pre-specified 4 degrees of freedom. In both cases the estimates of the mean are very close to 10.5 and have highly significant p-values, <.0001.

The next group of three models is based on the generalized error distribution. The first model using a shape parameter of 1 is equivalent to either using a Laplace distribution or minimizing the total sum of absolute deviations provides a location estimate of 11.000 (the median of the data set) with a highly significant p-value of <0.0001, and a scale parameter of 26.196. When p=1.5 the estimate increases to 13.159 and barely misses significance, p=0.0502 because the scale parameter has increased to 44.085. Lastly we fit a model with a shape parameter of 3 which is more sensitive to outlier than the normal theory model. The mean is now estimated at 63.147, statistically significant, p= 0.0015 despite the large scale parameter estimate of 80.952. Demonstrating significance is affected by both bias and variance.

The last model fit is the contaminated normal. We expect this model to perform well since the data was generated in manner similar to the assumptions of this model. The only difference is that the bulk of the data were generated uniformly over the interval 1 to 20 instead of from a normal distribution and a single outlier was added. The estimates of the mean, standard error and scale are closer to those of the estimate based on omitting the outlier than any of the other models. The mixing proportion of 0.9493 representing the proportion of data from the uncontaminated portion of the model is very close to actual proportion of 20/21=0.9523.

GROWTH STUDY

The growth study data of De Long and Summers (1991) is included as an example in the PROC ROBUSTREG documentation. The data set includes information on economic growth for 61 countries from 1960 through 1985. De Long and Summers fitted the following model using ordinary least squares:

GDP= beta0+beta1 x LFG + beta2 x GAP + beta3 x EQP + beta4 x NEQ + error

"where the response variable is the growth in gross domestic product per worker (GDP) and the regressors are labor force growth (LFG), relative GDP gap (GAP), equipment investment (EQP), and non-equipment investment. " (SAS example 62.3). The data set was analyzed again by Zaman, Rousseeuw, and Orhan (2001) using least trimmed squares (LTS) a robust method now available in PROC ROBUSTREG. They demonstrated that the observation for Zambia is an outlier and that it adversely affected the statistical significance of the regression coefficient for NEQ. Table 2 presents the results using ROBUSTREG and NLMIXED.

Table 2

Model (Procedure)	Intercept (p-value)	LFG (p-value)	GAP (p-value)	EQP (p-value)	NEQ (p-value)	Scale parameter (df*)
Normal (REG)	-0.0143 (0.1697)	-0.0298 (0.8811)	0.0203 (0.0313)	0.2654 (0.0002)	0.0624 (0.0787)	0.01306
Normal without Outlier (REG)	-0.0222 (0.0210)	0.0446 (0.8022)	0.0245 (0.0043)	0.2824 (<.0001)	0.0849 (0.0092)	0.01160
LTS (ROBUSTREG)	-0.0222 (0.0175)	0.0446 (0.8013)	0.0245 (0.0029)	0.2824 (<.0001)	0.0849 (0.0069)	0.0116
M-estimate – Huber (ROBUSTREG)	-0.0204 (0.0452)	0.0515 (0.7933)	0.0233 (0.0104)	0.2844 (<.0001)	0.0758 (0.0280)	0.0124
t-estimated df * (NLMIXED)	-0.0229 (0.0182)	0.0889 (0.6154)	0.0237 (0.0021)	0.2897 (<.0001)	0.0851 (0.0121)	0.0098 (4.73)
t-4 df * (NLMIXED)	-0.0236 (0.0076)	0.1010 (0.5391)	0.0239 (0.0014)	0.2926 (<.0001)	0.0869 (0.0079)	0.0094 (4)
GED p=1 (NLMIXED)	-0.0248 (<.0001)	0.1325 (<.0001)	0.0228 (<.0001)	0.3094 (<.0001)	0.0887 (<.0001)	0.01271
GED p=1.5 (NLMIXED)	-0.0211 (0.0150)	0.0680 (0.7044)	0.0221 (0.0010)	0.2864 (<.0001)	0.0819 (0.0128)	0.01230
GED p=3 (NLMIXED)	-0.0025 (0.8348)	-0.1572 (0.4618)	0.0141 (0.2463)	0.2634 (<.0001)	0.0269 (0.4357)	0.01330
Contaminated Normal (NLMIXED)	-0.0213 (0.0243)	0.0452 (0.7972)	0.0238 (0.0041)	0.2821 (<.0001)	0.0816 (0.0111)	0.01084/0.02836#

*df=degrees of freedom, # σ_small / σ_large.

We note that the results from the PROC REG fit without the outlier, the LTS fit in PROC ROBUSTREG, the t-distribution fit with estimated degrees of freedom and the contaminated normal fit are all similar.

REPEATED MEASURES CLINICAL TRIAL

For this example we simulate a data set from a two arm clinical trial with 100 patients in each arm. The primary endpoint is the change from baseline (CFB) for a laboratory value measured at two time points. The model includes d the baseline value as a covariate. The hypothesis that we test is whether the drug response averaged over the two time points is different for the two drugs. The data are generated from the model

CFB=Patient_i + .2 x baseline + treatment*time effect + error.

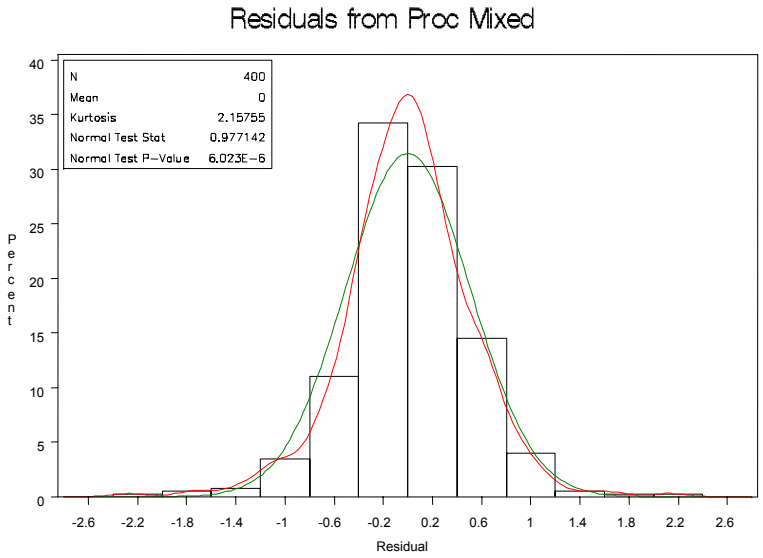
The term Patient_i represents a random intercept term that allows for dependence between the two measurements on each patient, for this simulation they are drawn from a normal distribution with mean 1 and variance 0.5. The baseline values are considered fixed, and were simulated from a normal distribution with mean 5 and variance 1.5625. The drug by time effect were set at

Drug/time	Time 1	Time 2
Drug 1	0.40	0.35
Drug 2	0.20	0.15

The drug effect is therefore (.40+.35)/2-(.20+.15)/2=.20. The error terms were drawn from a t-distribution with 3 degrees of freedom and a scale parameter of 1 .4142. We present the results from a typical simulation. The residuals were examined in PROC UNIVARIATE. A histogram with overlying density estimates and the results of a normality test are displayed in Figure 3. The p-value for testing non-normality is highly significant and the density estimates as well as the Q-Q plot (not shown) suggest heavy tails.

PROC NLMIXED can be used to fit a random effects model by including a normally distributed random intercept for each patient while still allowing for a non-normal error distribution. The random intercept creates a positive correlation between observations on a patient. In PROC MIXED a random statement with a slightly different syntax can be used for the same purpose. In addition, the same correlation structure could have also been modeled in PROC MIXED using a REPEATED statement and specifying TYPE=CS, for compound symmetry.

Figure 3



The code for fitting a t-distribution is displayed below:

```
proc nlmixed data=drugtrial;
    parms beta0 1.4439 bg1 .1543 bg2 .2943 bg3 -.01256 beta1 -.2643
                s2u .5787 sigma2 .4075 dft 5 to 30 by 5;
    bg4=0;***by parameterization;
    y=diff;**response variable in data set is named diff;
    pi=arcos(-1);
    mu=   beta0+bg1*g1+bg2*g2+bg3*g3+bg4*g4+
          beta1*base +u1;
    z=(y-mu)/sqrt(sigma2);
    logl=lgamma(.5*(dft+1))-.5*log(sigma2)-.5*log(dft*pi)-lgamma(dft/2)
                -( (dft+1)/2 )* log( 1+ (1/dft)*z**2);
    model y ~ general(logl);
    estimate 'drug' (bg1+bg2 +2*beta0)/2 -(bg3+2*beta0)/2;
    random u1 ~ normal(0,s2u) subject=id;
run;
```

The variables BG1 through BG4 are flags for the four drug by time combinations. The ESTIMATE statement defines the contrast of interest and the RANDOM statement defines the random patient effect U1 (note U1 is included in the definition of the mean parameter MU), where each patient has a unique ID variable.

We present the results for the contrast of interest for several different fitting methods in Table 3.

Table 3

Model (Procedure)	Estimate	Standard Error	p-value
Normal (MIXED)	0.02306	0.1251	0.0667
t-estimated (3.625 df *) (NLMIXED)	0.02613	0.1157	0.0250
GED p=1 (NLMIXED)	0.2379	0.1191	0.0471
GED p=3 (NLMIXED)	0.2142	0.1288	0.0979
Contaminated Normal (NLMIXED)	0.2746	0.1176	0.0205

*df=degrees of freedom

All of the procedures with heavier than normal tails declared the difference significant at the 0.05 level while the light tailed normal and GED with shape parameter 3 failed to find significance due to a slightly higher standard error.

CONCLUSION

PROC NLMIXED provides a convenient method to address non-normality in linear models. However, care needs to be taken in verifying the results of the NLMIXED procedure. The procedure should be started from a variety of initial parameter values to ensure a global maximum likelihood estimate has been found. Based on our limited empirical experience the family of t-distributions provides the most stable estimates in a minimal amount of computing time. The great flexibility of PROC NLMIXED can also allow these results be extended to non-linear regression models, flexible fitting methods such as splines, and repeated measures data

REFERENCES

Bouchad, J-P, Potters, M (2000), *Theory of Financial Risks*, Cambridge University Press, Cambridge.

Stokes, M.E., Davis, C.S, and Koch, G.G. (2000), Categorical Data Analysis Using The SAS System, 2nd ed. SAS Institute Inc., Cary, NC.

White, H. (1999), Asymptotic Theory for Econometricians, Revised Edition, Academic Press, San Diego.

Lange, K.L., Little, R.J.A, and Taylor J.M.G. (1989) "Robust Statistical Modeling Using the t Distribution". *Journal of the American Statistical Association*, **84**, 881-896.

Mineo, A.M., Ruggieri, M (2005), "A Software Tool for the Exponential Power Distribution: The normalp package". *Journal of Statistical Software*, **12**

Terpstra, J.T., McKean, J.W. (2005),"Rank-Based Analyses of Linear Models Using R". *Journal of Statistical Software*, **14**

Zaman, A., Rousseeuw, P.J., Orhan, M. (2001), "Econometric Applications of High-Breakdown Robust Regression Techniques," Econometrics Letters, **71**, 1-8.

De Long, J.B., Summers, L.H. (1991), "Equipment Investment and Economic Growth," *Quarterly Journal of Economics*, **106**, 445-501.

Hitomi, K., Kagihara, M. (2001), "Calculation Method for Nonlinear Dynamic Least-Absolute Deviations Estimator", *J. Japanese Statistical Society*, **31**, 39-51

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Steven Gilbert
Rho, Inc.
199 Wells Avenue, Suite 302
Newton, MA 02459
E-mail: sgilbert@rhoworld.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.