

1. INTRODUCTION

We consider robust estimation in linear regression when there is measurement error in the predictors. This paper was motivated by a study of nutrient intake via questionnaires and food records, in which we found that the deletion of a few points had a large effect on a standard measurement error analysis, see Section 3.

Denote the observed response by Y , the true but unobservable covariate by X , and the observable surrogate by W . In our example, we observe replicates of both W and Y . By definition, the i^{th} person has $m_{i,1} + m_{i,2}$ replicate measures of Y and $m_{i,2} + m_{i,3}$ replicate measures of W . These observations include $m_{i,2}$ simultaneous replicates of both Y and W . Thus,

$$\begin{aligned}
 \bullet Y_{ij} &= X_i^T \beta + q_i + r_{ij}; & j &= 1, \dots, m_{i,1}; \\
 \bullet Y_{ij} &= X_i^T \beta + q_i + r_{ij}; & W_{ij} &= X_i + U_{ij}; & j &= m_{i,1} + 1, \dots, m_{i,1} + m_{i,2}; \\
 \bullet W_{ij} &= X_i + U_{ij}; & j &= m_{i,1} + m_{i,2} + 1, \dots, m_{i,1} + m_{i,2} + m_{i,3} = m_i.
 \end{aligned} \tag{1.1}$$

The common dimension of X , W and β is p , and the first element of X and W is 1.0, indicating an intercept. In this model, q_i represents equation error, r_{ij} represents a variance component for a given individual's response, and U_{ij} represents the variance component for measurement error. The true covariate X is not observed. We will also assume that r_{ij} and U_{ik} are independent for $j \neq k$, but we do allow the possibility that r_{ij} and U_{ij} are correlated.

The important classical measurement error model with no replication of Y , some replication of W and independence of the measurement errors for Y and W is included in our paper by taking $m_{i,1} \equiv 1$ and $m_{i,2} \equiv 0$. In all other cases, we assume that $m_{i,2} \geq 2$.

In Section 2.2, we introduce the heteroscedasticity-consistent method of moments estimate defined by Fuller (1987, p. 187), where by heteroscedasticity-consistent we mean that the estimate is consistent and asymptotically normally distributed even when the errors are heteroscedastic, i.e., depend on the unit i . This estimator, which is not robust to outliers or leverage points, is a solution to a set of unbiased estimating equations which depend on within-unit sample variances and covariances. One might attempt to robustify the estimating equations directly, but trying to do so while maintaining consistency for a wide range of models is difficult because of the dependence on the within unit sample variances and covariances. Indeed, we believe that such a straightforward approach would lead to estimates consistent only at the normal error model, see Section 2.3.

We take an entirely different tack, which reduces the problem to one which is easy to solve. In Section 2.3, we show that the estimating equations are *linear*, a previously unknown fact which simplifies the problem greatly. In particular, in Section 2.3 we define a new robust estimator which has bounded influence and bounded leverage, as defined by Hampel, et al. (1986). The estimators of the nonintercept coefficients are consistent even for asymmetric errors. If the errors are symmetrically distributed, the robust estimates are heteroscedasticity consistent as well.

In Section 3 we apply these techniques to some nutrition data.

The literature on robust estimation in measurement error models is fairly small, and none of it applies to the case that there are replicates of W and the errors in Y and W are correlated. Zamar (1989), and Cheng & Van Ness (1989) considered the classical model with known reliability ratio, while Carroll & Gallo (1982) discussed the classical independent error model with replication of the predictors.

2. ESTIMATES WITH REPLICATION IN W AND Y

2.1. Introduction

We assume that the sequences (q_i) , (X_i) and (r_{ij}, U_{ij}) are mutually independent. The first component of X and W equals 1.0.

2.2. Heteroscedasticity-Consistent Moments Estimates

Let $\Omega_{X,i}$, $\Omega_{U,i}$ and $\Omega_{rU,i}$ be the covariance matrices of X_i , U_{ij} and the joint covariance of r_{ij} and U_{ij} , respectively. The errors (q_i) can also be heteroscedastic. Define \bar{Y}_i and \bar{W}_i to be the averages of all the Y 's and W 's within the i^{th} sampled unit. Let π_i be fixed observation weights which satisfy the conditions of Theorem 3.1 of Fuller (1987). For example, if there were no measurement error in W and no equation error, then $\pi_i = (m_{i,1} + m_{i,2})^{-1}$ would be the appropriate weights from Gauss-Markov theory. Let $\hat{\Omega}_{U,i}$ and $\hat{\Omega}_{rU,i}$ be unbiased estimates of $\Omega_{U,i}$ and $\Omega_{rU,i}$ respectively. For example, these estimates could be those generated by a components of variance analysis, the former being the sample covariance matrix within the i^{th} individual of the W_{ij} for $j = m_{i,1} + 1, \dots, m_i$, and the latter the sample covariance of (Y_{ij}, W_{ij}) for $j = m_{i,1} + 1, \dots, m_{i,1} + m_{i,2}$. For the classical model with replication only of the predictor and no correlation of the errors, $\Omega_{U,i} \equiv 0$, $m_{i,1} \equiv 1$ and $m_{i,2} \equiv 0$; set $\hat{\Omega}_{U,i} \equiv 0$.

The estimate of $\Omega_{rU,i}$ discussed above relies only on those observations for which we can observe both Y and W . As such, it may be inefficient if either $m_{i,1}$ or $m_{i,3}$ are not small compared to $m_{i,2}$. In our example, $m_{i,1} = m_{i,3} = 0$. Let

$$A_{n,i,1} = \left\{ \overline{W}_i \overline{W}_i^T - (m_i - m_{i,1})^{-1} \widehat{\Omega}_{U,i} \right\};$$

$$A_{n,i,2} = \left\{ \overline{W}_i \overline{Y}_i - \frac{m_{i,2}}{(m_{i,1} + m_{i,2})(m_i - m_{i,1})} \widehat{\Omega}_{rU,i} \right\}.$$

Then the errors-in-variables estimate is (Fuller, 1987, Section 3.1)

$$\widehat{\beta}_M = \left(\sum_{i=1}^n \pi_i A_{n,i,1} \right)^{-1} \sum_{i=1}^n \pi_i A_{n,i,2}, \quad (2.1.a)$$

$$\text{i.e., the solution to } 0 = \sum_{i=1}^n \pi_i (A_{n,i,2} - A_{n,i,1} \beta) \quad (2.1.b)$$

Note that this estimate is not robust. In the terms $A_{n,i,1}$, one can see sensitivity to leverage and outlying replicates in the surrogates, while in the terms $A_{n,i,2}$, one sees sensitivity to outlying mean responses as well as replicates in the response. Small sample corrections to (2.1.a) as in Fuller (1987, p. 193) may be employed to ensure that the inverted matrix is positive definite.

2.3. Robust Moment Estimates

One might attempt to define robust estimates by directly “robustifying” the unbiased estimating equations (2.1.b). The usual device is to bound the effects of each term in the sum; see for example Carroll & Ruppert (1988, Chapter 6). We believe that such an attempt would fail to produce an estimator which is consistent under the broad conditions for consistency that our estimator achieves. The problem, of course, is that (2.1.b) involves sample variances and covariances, so that making any bounded estimating equation to have mean zero will become difficult if not hopeless unless one makes strong distributional assumptions, e.g., normality. If this were the case, then one would have defined a “robust” estimate consistent only at a specific model.

In contrast, we will derive a new and useful representation (2.2) below for the moments-based estimate, showing that it is a linear function of the data in a particularly convenient form. From this, we will define a simple, intuitively appealing estimator which is consistent for homoscedastic models, as well as consistent for heteroscedastic models with symmetric errors.

Using the previous definitions of $\widehat{\Omega}_{U,i}$ and $\widehat{\Omega}_{rU,i}$, it is possible to write (2.1.a) as the solution to a simple unbiased estimating equation. Define $d_i(j, k) = 1$ if $j \geq m_{i,1} + m_{i,2} + 1$ or if $k \leq m_{i,1}$, and $d_i(j, k) = m_{i,2}/(m_{i,2} - 1)$ otherwise. Further, define

$$c_i = \{(m_{i,1} + m_{i,2})(m_i - m_{i,1})(m_i - m_{i,1} - 1)\}^{-1};$$

$$H_{i,1}(\beta) = c_i \sum_{j=m_{i,1}+1}^{m_i} \sum_{k=1, k \neq j}^{m_{i,1}+m_{i,2}} \sum_{l=m_{i,1}+1, l \neq j}^{m_i} d_i(j, k) W_{ij} (Y_{ik} - W_{il}^T \beta). \quad (2.2)$$

Straightforward calculations show that $H_{i,1}(\beta) = A_{n,i,2} - A_{n,i,1}\beta$, so that $\widehat{\beta}_M$ solves

$$0 = \sum_{i=1}^n \pi_i H_{i,1}(\beta). \quad (2.3)$$

Equation (2.2) is useful because it lays out the effects of outliers and leverage. Indeed, except for the summations in (j, k, l) , (2.3) is similar in form to the normal equations for ordinary least squares estimation without measurement error. As such, (2.3) is easily “robustified”. In what follows, we construct a bounded influence robust M–estimate of β from within the Mallows class, see Simpson, et al. (1992). This last paper considers ordinary linear regression and constructs high breakdown Mallows one–step estimates, giving careful attention to the details required for proving asymptotic normality. Instead of retracing their details and arguments, we will only outline the estimates used and state the appropriate limit distributions.

Symmetry of the error distributions is *not* required. Any lack of symmetry in the errors simply means that the intercept parameter depends on the method of estimation, but all slope parameters are consistently estimated, see Simpson, et al. (1992).

In ordinary linear regression, the Mallows estimates are constructed with three components: (i) an estimate of scale σ ; (ii) an odd function $\psi(\cdot)$ used to control outliers in the response; and (iii) a weighting function $h(\cdot)$ used to control outliers in the design. In our problem, the same three components are required. We will require an estimate of scale σ for the “errors” $Y_{ik} - W_{il}^T \beta$. The Mallows class also controls outliers in the predictors W_{ij} by means of a weighting function $h(\cdot)$, see Section 2.5. Details of implementation, i.e., formulae used in our example, are given in Section 3.

To estimate the scale, we will use an even function χ such that $\int \chi(u)\phi(u)du = 0$, where $\phi(\cdot)$ is the standard normal density function. Define the estimate $(\widehat{\beta}_R, \widehat{\sigma})$ of (β, σ) to be a solution to

the equations:

$$0 = \sum_{i=1}^n \pi_i H_{i,2}(\beta, \sigma), \quad \text{where } H_{i,2}^T(\beta, \sigma) = \{\ell_{i,1}^T(\beta, \sigma), \ell_{i,2}(\beta, \sigma)\}^T; \quad (2.4)$$

$$\ell_{i,1}(\beta, \sigma) = c_i \sum_{j=m_{i,1}+1}^{m_i} \sum_{l=m_{i,1}+1, l \neq j}^{m_i} \sum_{k=1, k \neq j}^{m_{i,1}+m_{i,2}} d_i(j, k) h(W_{ij}) W_{ij} \psi\left(\frac{Y_{ik} - W_{il}^T \beta}{\sigma}\right); \quad (2.5)$$

$$\ell_{i,2}(\beta, \sigma) = c_i \sum_{j=m_{i,1}+1}^{m_i} \sum_{l=m_{i,1}+1, l \neq j}^{m_i} \sum_{k=1, k \neq j}^{m_{i,1}+m_{i,2}} \chi\left(\frac{Y_{ik} - W_{il}^T \beta}{\sigma}\right). \quad (2.6)$$

The theory derived by Simpson, et al. (1992) is applicable not to solutions to (2.4), which need not be unique, but instead to one-step approximate solutions to (2.4) starting from an $n^{1/2}$ -consistent estimate of (β, σ) and using the method of scoring or the Newton-Raphson method. We use the latter method because it is heteroscedasticity consistent. The normal limit distribution and estimated covariance matrix of the estimates are then easy to derive. Specifically, the estimated covariance matrix of $(\hat{\beta}_R, \hat{\sigma})$ is

$$\begin{aligned} \text{cov}\left(\hat{\beta}_R, \hat{\sigma}\right) &= n^{-1} \hat{B}_3^{-1} \hat{B}_4 \hat{B}_3^T; \\ \hat{B}_3 &= n^{-1} \sum_{i=1}^n \pi_i \left\{ \frac{\partial}{\partial(\beta^T, \sigma)} H_{i,2}(\hat{\beta}_R, \hat{\sigma}) \right\}; \\ \hat{B}_4 &= (n-p)^{-1} \sum_{i=1}^n \pi_i^2 \left\{ H_{i,2}(\hat{\beta}_R, \hat{\sigma}) H_{i,2}(\hat{\beta}_R, \hat{\sigma})^T \right\}. \end{aligned} \quad (2.7)$$

2.4. Heteroscedastic Consistency and Independence

We have noted that the moments estimate (2.4) has the property that it is consistent even when the errors are heteroscedastic. Exactly the same property hold for the robust estimates if the errors are symmetrically distributed. For asymmetric homoscedastic errors, formal Taylor series expansions of (2.4) indicate that (2.7) still holds. However, for heteroscedastic errors, our calculations (not given here) indicate that symmetry may be necessary for the estimating equations to be unbiased. More precisely, we have been unable to show that (2.5) has mean zero for heteroscedastic errors unless they are also symmetric. In the presence of asymmetry, heteroscedasticity-consistent estimators require that one *model* the heteroscedasticity, even in cases where the X_i 's can be observed without error, see Welsh, et al. (1991).

In Section 2.1, we stated the assumption that the errors be independent of the (X_i) . This can be weakened considerably, namely to the assumption that (2.2) and (2.5) have mean zero. For

example, consider error models of the form $q_i = H_1(X_i)q_{i*}$, $r_{ij} = H_2(X_i)r_{ij*}$ and $U_{ij} = H_3(X_i)U_{ij*}$. If (X_i) , (q_{i*}) and (r_{ij*}, U_{ij*}) are mutually independent with mean zero, then (2.2) has mean zero. If in addition the errors are symmetric, then (2.5) has mean zero. In both cases, the asymptotic results apply without change.

2.5. Design Weight Functions

The weight functions $h(W_{ik})$ given in (2.5) are defined as in Simpson, et al. (1992). Write $W_{ik}^T = (1, Z_{ik}^T)$, where $\dim(Z_{ik}) = p-1$. Let μ_Z and C_Z be robust estimates of center and covariance for the Z_{ik} . We will leave these estimates unspecified for now, although in Section 3 we give details of estimation for our example. Further discussion is given by Simpson, et al. (1992). Define the weights as follows:

$$h(Z_{ik}) = \{\psi(c_{ik})/c_{ik}\}^a;$$

$$c_{ik} = \{(Z_{ik} - \mu_Z)^T C_Z^{-1} (Z_{ik} - \mu_Z)/(p-1)\}^{1/2}.$$

The case that $a = 1$ is the usual Mallows method, whereas the case $a = 2$ has been proposed by Simpson, et al. (1992) to insure robustness of standard error estimates.

3. EXAMPLE

3.1. Introduction

As a numerical example, we consider nutrition data from the pilot phase of the Women's Health Trial; see Henderson, et al. (1990). We focus on dietary intake of major nutrients, including caloric intake, saturated fat intake and protein intake. We will analyze (i) the % calories from fat; (ii) $\log(1 + \text{saturated fat})$; (iii) $\log(1 + \text{protein})$. The data transformation for (ii) and (iii) was used to stabilize, in part, some extreme heteroscedasticity. We are assuming that model (1.1) applies to the transformed data.

The design follows that of Section 1, with the model given by (1.1). For each of (i)–(iii), the response Y is the intake of that nutrient estimated from a semiquantitative food frequency questionnaire of food intake, see Block, et al. (1986). The surrogate W is intake estimated from food records. Measurements of Y and W were obtained at two time periods, the first one year into the study, the second a further year later. We are assuming that the time lag between interviews

is sufficiently long that interview–reinterview correlation is not a problem, an assumption which seems reasonable in the context. In the notation of Section 3, $m_{i,1} = 0$, $m_{i,3} = 0$ and $m_{i,2} = 2$.

We *define* X to be the expectation of nutrient intake from food records, so that the surrogate W is unbiased for X . Generally, a detailed food record is thought to be more reliable than questionnaires, so that this definition makes some operational sense. In general, some such assumption is required to obtain an identified model in replicated observation studies. While possible biases in the food records are of serious practical interest, examination of this issue is beyond the scope of this paper.

This is a simple linear regression problem, and principal interest lies in the following question:

- Is the slope less than 1? If so, this indicates a differential reporting bias.

The data consist of records of 86 individuals who were controls in the clinical trial and had complete data records. These records were carefully screened by nutritionists for inconsistencies prior to the statistical analysis of the data; no statistical screening for outliers was performed. As will be seen below, the nutritionists’ screening work did not eliminate certain observations which were influential in a case–deletion sense.

3.2. Methods Used

The methods used were as follows. Throughout, $\pi_i \equiv 1$. In (2.5), the function ψ is the redescending trisquared function $\psi(v) = v \{1 - (v/b)^2\}^3$ for $|v| \leq b$ and $= 0$ otherwise. Here, b is a tuning constant selected by the user.

The design weights $h(W_{ik})$ were formed as follows. In this particular instance, $W_{ik}^T = (1, Z_{ik})$ for a scalar Z_{ik} . We estimated the center μ_Z and scale σ_Z of the Z_{ik} by Huber’s Proposal 2 (Huber, 1981, p. 147), where $\sigma_Z = C_Z^{1/2}$ in the notation of Section 2.5. Then if $c_{ik} = (Z_{ik} - \mu_Z) / \sigma_Z$, following Section 2.5, the design weights are defined by

$$h(W_{ik}) = \{\psi(c_{ik}) / c_{ik}\}^a. \tag{3.1}$$

We choose $a = 1$ and $a = 2$ in our analyses.

The estimate of σ used in (2.5) was defined as follows. Let $\widehat{\beta}$ be the current estimate of β in the iterative solution to (2.5), and let $r_{ikl} = Y_{ik} - W_{il}^T \widehat{\beta}$. Actually, because $m_i = m_{i,2} = 2$, in this particular instance we will have $k = l$, but the general form is useful to write down. Then $\widehat{\sigma}$ is the solution to (2.6) with $\chi(v) = \psi^2(v) - (n - 2) \int \psi^2(x) \phi(x) dx / n$.

3.3. Numerical Results

In both (3.1) and (2.5), we used $b = 6$ and $b = 8$. In our analysis, writing $\beta^T = (\beta_0, \beta_1)$, we tested the hypothesis that $\beta_1 = 1$ by the usual asymptotic Wald test method. Table 1 lists the slope estimates, estimated standard errors, and the one-sided p-values for these tests. “Moments” refers to the estimate (2.1), while “M(b)” refers to the solutions to (2.4) obtained for a given b .

To check robustness, we deleted single cases and plotted the data to see if there were any unusual observations. From these analyses we concluded that there are two leverage values for % calories from fat, as well as a response which seems unusually small. Protein has a response at the highest level of \bar{Z} which seems too high, in addition to two unusually low values of \bar{Y} and one possible leverage value at the low end of the scale. Saturated fat has a clump of three observations with unusually low values of \bar{Z} . As stated above, these data passed through a careful screening by nutritionists, so they are merely unusual and not obviously in error. We decided to reanalyze the data having deleted the unusual points, the results given in Table 1. For % calories from fat, we first deleted the two potential leverage points, then the potential response outlier, and then all three.

The changes in the analysis for % calories from fat depending on which observations are deleted are the most striking, although there are noticeable instabilities in the moments estimates for the other two variables. The robust estimate M(6) with $a = 2$ yields more stable joint p-values and slope estimates, although it is of course not completely immune to the group deletion effect.

4. DISCUSSION

Fuller’s moment estimates are applicable in model (1.1). They are consistent and asymptotically normal with estimable covariance matrices even when the error variances are heteroscedastic. Their major drawback is their sensitivity to outliers, which we have illustrated in the example.

We have investigated two major cases: (i) the classical case of uncorrelated errors with each surrogate replicated; and (ii) the correlated case when (Y, W) is replicated at least once. We first rewrote Fuller’s estimates in these cases as solutions to unbiased estimating equations. The representation then led us to define bounded influence estimates. These latter estimates were noted to be consistent and asymptotically normally distributed for the heteroscedastic symmetric case or

in general for homoscedastic errors.

ACKNOWLEDGEMENTS

The work of Carroll and Eltinge was supported by the National Institutes of Health and partially completed during Carroll's visit to Purdue University's Center for Decision Sciences. Ruppert's work was partially supported by the Mathematical Sciences Institute at Cornell University and by a grant from the National Science Foundation. We thank Larry Freedman for many helpful discussions, and Yasuo Amemiya for comments on an early draft.

REFERENCES

- Block, G., Hartman, A. M., Dresser, C. M. et al. (1986). A data based approach to diet questionnaire design and teaching. *Am. J. Epid.*, **124**, 453–469.
- Carroll, R. J. & Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman & Hall, London.
- Carroll, R. J. & Gallo, P. P. (1982). Some aspects of robustness in functional errors-in-variables regression models. *Comm. Statist., Ser. A*, **11**, 2573-2585.
- Cheng, C. & van Ness, J. (1990). Bounded influence errors in variables regression. *Contemporary Mathematics: Statistical Analysis of Measurement Error Models and Applications*, P. Brown and W. Fuller, editors. American Mathematics Society, Providence.
- Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.
- Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- Henderson, M. M., Kushi, L. H., Thompson, D. J., et al. (1990). Feasibility of a randomized trial of a low fat diet for the prevention of breast cancer: dietary compliance in the Women's Health Trial Vanguard Study. *Preventive Medicine*, **19**, 115–133.
- Simpson, D. G., Ruppert, D. & Carroll, R. J. (1992). On one-step GM-estimates and stability of inferences in linear regression. *J. Am. Statist. Assoc.*, to appear.
- Welsh, A. H., Carroll, R. J. & Ruppert, D. (1991). Fitting heteroscedastic regression models. Preprint.
- Zamar, R. H. (1989). Robust estimation in the errors in variables model. *Biometrika*, **76**, 149–154.

ROBUST LINEAR REGRESSION IN REPLICATED MEASUREMENT ERROR MODELS

R. J. Carroll & J. L. Eltinge
Department of Statistics
Texas A&M University
College Station, TX 77843-3143

D. Ruppert
School of Operations Research
and Industrial Engineering
Cornell University
Ithaca, NY 14853

SUMMARY

We propose robust and bounded influence methods for linear regression when some of the predictors are measured with error. We address the important special case that the surrogate predictors are replicated, and that the measurement errors in response and predictors are correlated. The robust methods proposed are variants of the so-called Mallows class of estimates. The resulting estimators are easily computed and have a simple asymptotic theory. An example is used to illustrate the results.

Keywords: ASYMPTOTIC THEORY; BOUNDED INFLUENCE ESTIMATORS; ERRORS IN VARIABLES; LEVERAGE; MALLOWS ESTIMATES.

TABLE 1: PARAMETER ESTIMATES

The significance levels of the test for slope = 1 and the joint test for slope = 1, intercept = 0 are denoted “p” and “Joint”, respectively. For % Calories from Fat, the labels “reduced data (1), (2), and (3)” mean that we deleted (1) the two smallest values of \bar{Z} , (2) the smallest value of \bar{Y} , and (3) all three. For saturated fat and protein, the reduced data are described in the text.

% Calories from Fat						
Method	Full Data			Reduced Data(1)		
	Slope	s.e.	p	Slope	s.e.	p
Moments	0.67	0.12	0.01	0.78	0.17	0.10
M(6), $a = 1$	0.72	0.14	0.02	0.72	0.15	0.03
M(6), $a = 2$	0.73	0.14	0.03	0.71	0.15	0.03

% Calories from Fat						
Method	Reduced Data(2)			Reduced Data(3)		
	Slope	s.e.	p	Slope	s.e.	p
Moments	0.62	0.10	0.00	0.69	0.15	0.02
M(6), $a = 1$	0.69	0.13	0.01	0.68	0.14	0.01
M(6), $a = 2$	0.70	0.13	0.01	0.68	0.14	0.01

Log(1 + Saturated Fat)						
Method	Full Data			Reduced Data		
	Slope	s.e.	p	Slope	s.e.	p
Moments	0.74	0.18	0.08	0.95	0.26	0.42
M(6), $a = 1$	0.83	0.20	0.20	0.91	0.26	0.37
M(6), $a = 2$	0.84	0.21	0.23	0.88	0.27	0.33

Log(1 + Protein)						
Method	Full Data			Reduced Data		
	Slope	s.e.	p	Slope	s.e.	p
Moments	0.80	0.24	0.21	0.66	0.23	0.07
M(6), $a = 1$	0.70	0.22	0.09	0.61	0.20	0.03
M(6), $a = 2$	0.66	0.21	0.05	0.59	0.20	0.02