JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 94, NO. B6, PAGES 7667-7672, JUNE 10, 1989

Comment on "Saugus-Palmdale, California, Field Test for Refraction Error in Historical Leveling Surveys" by R. S. Stein, C. T. Whalen, S. R. Holdahl, W. E. Strange, and W. Thatcher, and Reply to "Comment on 'Further Analysis of the 1981 Southern California Field Test for Levelling Refraction by M. R. Craymer and P. Vaníček' by R. S. Stein, C. T. Whalen, S. R. Holdahl, W. E. Strange, and W. Thatcher"

MICHAEL R. CRAYMER AND PETR VANÍČEK

Geodetic Research Laboratory, Department of Surveying Engineering, University of New Brunswick, Fredericton, Canada

INTRODUCTION

Craymer and Vaníček [1986] performed a multiple linear regression analysis on the levelling data from the 1981 field test for levelling refraction reported by *Stein et al.* [1986]. Our analysis revealed that in addition to differential refraction, turning point settlement and an effect dependent on height differences are also present in the discrepancies of the section height differences from forward and backward runnings. *Stein et al.* [1986] dismiss our results arguing that (1) the independent variables used in a multiple regression analysis must be statistically uncorrelated and that any statistically significant correlation invalidates a multiple regression analysis, (2) our results are not "robust" due to the correlation between two of our independent variables (section height difference *dH* and number of turning points *tp*) when two sections are deleted from the data sample, and (3) our conclusions are "untenable" because *dH* and *tp* are positively correlated, whereas the estimated regression coefficients have opposite signs.

We believe that these comments are unjustified and would have responded to them in our paper had we been given the opportunity to review theirs prior to publication (there was no reference to our work in the manuscript supplied to us by R. Stein prior to publication). We will therefore show here that their arguments are not based on correct statistical theory. In addition, we will also address a few more questions raised in their reply to our rebuttal [*Stein et al.*, this issue].

MULTIPLE REGRESSION AND COLLINEARITY

The problem of linear correlation between two independent variables is well known in multidimensional regression analysis. The terms collinearity, multicollinearity, and ill-conditioning are all used to denote this situation [*Neter and Wasserman*, 1974, p. 339; *Belsley et al.*, 1980, p. 85; *SAS Institute Inc.*, 1985, p. 672]. The latter, however, is preferred since it more precisely describes the problem as a numerical one. In fact, most of the diagnostic tools used for detecting collinearity are from the field of numerical analysis.

It is true that theoretically it becomes difficult to separate the effects due to highly collinear variables in a multiple regression. Collinearity among the independent variables results in a matrix of normal equations whose ill-conditioning simply causes the estimated regression coefficients and their covariance matrix to be numerically unstable [*Belsley et al.*, 1980, p. 96]. In its limiting form, perfectly correlated or collinear variables cause the matrix of normal equations to become singular, and no regular solution can be computed. The presence of correlations among the independent variables shows in the covariance matrix of the estimated multiple regression coefficients. Large correlations will inflate both the diagonal (i.e., errors in regression coefficients) and off-diagonal (i.e., correlations) elements of the covariance matrix [*Belsley et al.*, 1980, p. 115; *Neter and Wasserman*, 1974, p. 341].

The question of concern, however, is whether the magnitude of this correlation is large enough to significantly affect our results (i.e., the estimates of the regression coefficients) and not just whether the correlations among the independent variables are statistically significant, as *Stein et al.* [1986] contend. A review of the literature on this topic will reveal that the independent variables must be highly correlated before instabilities may be expected to occur in the estimates of the regression coefficients (see, e.g., *Neter and Wasserman* [1974, p. 341] and *Belsley et al.* [1980, p. 86]). In their introductory book on regression analysis, *Neter and Wasserman* [1974, p. 341] state

> The fact that some or all independent variables are correlated among themselves does not, in general, inhibit our ability to obtain a good fit nor does it tend to affect inferences about mean responses or predictions of new observations, provided these inferences are made within the region of observations.

In practice, statisticians do not worry about such correlations until they are of the order of about 0.9 [*Belsley et al.*, 1980, p. 94, 153; M.S. Srivastava, Department of Statistics, University of Toronto, personal communication, 1986]. Clearly, this is not a problem in our analysis where correlations are less than 0.5, even when the data sample is altered by deleting sections 57 and 58 as suggested by *Stein et al.* [1986].

Collinearity is more reliably identified by examining the eigenvalues of the matrix of normal equations. In the presence

Paper number 89JB00356.
0148-0227/89/89JB-00356$02.00

of collinearity, some of the eigenvalues will be very small. *Belsley et al.* [1980] assess "smallness" using condition numbers defined as the square root of the ratio of the largest and smallest eigenvalues. When this number is very large, the problem is ill-conditioned and collinearity is said to exist. Based on experimental evidence, *Belsley et al.* [1980, pp. 105, 112, 153] associate weak dependencies with condition numbers around 5 or 10 and moderate to strong relations with condition indices of 30 to 100. Only when condition indices are of the order of 100 do they consider any great potential harm to the regression estimates. Our results exhibit no large condition numbers. In fact the largest condition number is 4 (obtained when sections 57 and 58 are omitted), typical of only a weak dependency. Thus the comments by *Stein et al.* [1986] regarding the incorrect application of multiple linear regression in the presence of moderate correlation are erroneous as they now realize.

INFLUENTIAL OBSERVATIONS

In their second comment, *Stein et al.* [1986] argue that by removing "just" one and two sections, the correlation between the section height difference dH and number of turning points tp becomes statistically significant. They conclude from this that our results are not "robust." However, as explained above, it is not the existence of statistically significant correlations but their magnitude that influences the regression solution by causing numerical instabilities in their estimation. As we have shown, there is no evidence of significant collinearity or ill-conditioning in our results even when the two disputed sections are deleted from the data sample.

Stein et al. [this issue] further argue that the regression coefficient for dH depends on the two sections in question. This is simply not true. We have already shown in our paper that removal of section 57 (or any section for that matter) from the multiple regression analysis does not change the values of the estimates at any reasonable significance level. Only the variances and correlations among the regression coefficients actually become larger. The absolute value of correlation between the regression coefficients for dH and dt increase from 0.61 to 0.77. The variances and correlations become even larger when both sections 57 and 58 are removed from the sample: the absolute value of correlation between regression coefficients for dH and dt increases to 0.87. Although the deletion of both sections 57 and 58 increases the level of multicollinearity and ill-conditioning in the regression estimation, the actual regression coefficients do not change greatly contrary to the claim by *Stein et al.* [this issue]. This is in complete agreement with the passage quoted from *Draper and Smith* [1981, p. 170] by *Stein et al.* [this issue] which states that the fit of the model (i.e., regression coefficients and not just their statistical significance) must be greatly affected by the deletion of one or two observations before they can be considered for removal. Thus there is little evidence to support the removal of the two sections in dispute.

The deletion of sections 57 and 58 from the analysis should also be resisted from the physical point of view. It is precisely these sections (the only ones proceeding downhill) that enable the regression analysis to reliably resolve the effects due to dH and tp. Although more such observations would have been better, the lack of abundance of levelling downhill should not be taken as justification for removing those we have. These sections effectively act as a basis for comparison with the uphill levelling. Calling these sections "outliers" is also incorrect ("influential" observations is used in conventional statistical terminology) and falsely implies they are contaminated by blunders or gross errors: there are no blunders, the levelling simply went downhill! We wish to make the point that we would consider it most improper to eliminate any event from a given sample unless it can be shown conclusively that a blunder (mistake) occurred in the levelling observations. Each event represents a unique set of circumstances that a successful model must be capable of explaining.

The statement by *Stein et al.* [1986] that "just" one or two sections would be removed is also misleading. In fact, each section is composed of hundreds of observations of the levelling rods! Summing together the number of setups in these sections, one finds that "only" one or two sections actually amounts to 46 or 78 setups (184 or 312 individual observations), respectively! Contrary to what Stein et al. would have us believe, this is indeed a significant portion of the sample which should not be dismissed so lightly. Moreover, every such "event" in levelling is a result of a series of measurements subjected to repeated internal testing to prevent any blunders from occurring. *Stein et al.* [this issue] contend that this testing is performed only for individual setups and that a disturbance of an individual instrument setup or turning point will not be detectable. This is only partially not true, for an error caused by a disturbance of the instrument would be fully apparent in the differencing of the high-low scale rod readings. Although a disturbance of a turning point will indeed go undetected in this check, it is precisely these systematic errors that we are accounting for and successfully resolving in our regression model.

In our opinion no event should be eliminated unless there is good physical evidence to show that it is affected by blunders. A proposed model should be capable of modelling all events. If the data do not fit a model, the model should be modified, not the data. By eliminating one, two, four, six events (when does one stop?), one can statistically prove almost anything.

CORRELATION AND REGRESSION COEFFICIENTS

In their final comment, *Stein et al.* [1986] state that positive correlation between dH and tp makes our conclusion that the regression estimates have opposite signs "untenable." This is true only for simple linear regressions (i.e., regressions using only one independent variable) or for multiple regressions where the independent variables are completely uncorrelated with each other. It is not generally true for multiple regressions when the independent variables are correlated even moderately. The correlation coefficients are based on individual simple regressions of the dependent variable on each independent variable and not on a multiple regression with all variables. In a multiple regression the sign of the regression coefficients do not depend solely on the correlation coefficients; other correlations must also be taken into account.

It is relatively easy to prove this mathematically. Consider, as an example, a linear regression (including the intercept term) of a dependent variable z on two independent variables x and y. Let r_{zx} and r_{zy} represent the correlation coefficients between z and the independent variables x and y, respectively, and r_{xy} the correlation coefficient between the independent variables themselves. The regression coefficients b_x and b_y may then be expressed in terms of the correlation coefficients as [Edwards, 1979, p. 45]

$$b_x = \frac{s_z\,(r_{zx} - r_{zy}\, r_{xy})}{s_x\,(1 - r_{xy}^2)}$$

$$b_y = \frac{s_z\,(r_{zy} - r_{zx}\, r_{xy})}{s_y\,(1 - r_{xy}^2)}$$

where s_x, s_y, and s_z are the sample standard deviations of the two independent variables x and y and of the dependent variable z, respectively. Clearly, the three quantities r_{xy}, r_{zx}, and r_{zy} govern the sign of b_x and b_y (note that s_x, s_y, s_z, and r_{xy}^2 are all positive), not just the sign of r_{xy} as implied by *Stein et al.* [1986]. For example, when r_{xy} is positive and r_{zx} and r_{zy} are of opposite signs, the regression coefficients (b_x and b_y) will also be of opposite signs. In our regression model with three independent variables (two of which are moderately correlated) and no intercept, the expressions for the regression estimates in terms of correlations will be even more complicated (each involving four correlation coefficients) and difficult to predict.

Clearly, one cannot predict the signs of the regression coefficients from their correlations with the dependent variable alone as *Stein et al.* [1986, this issue] would have us believe. Our results are obtained from the popular SAS software, which has been thoroughly tested and routinely used for such problems [see *SAS Institute Inc.*, 1985]. Moreover, these results agree exactly with those from our own independently developed software used by *Craymer and Vaníček* [1986]. The fact remains that the regression coefficients for *dH* and *tp* are statistically significant and of opposite sign.

PARTIAL CORRELATION COEFFICIENTS

In their reply, *Stein et al.* [this issue] also argue that because of the lack of correlation between the discrepancy (dependent variable) and *dH* and *tp* (independent variables), these effects should be left out of the model. What they fail to mention, however, is that the effects are indeed statistically significant in a multiple regression. This oversight results from their apparent misunderstanding of the correlation coefficient and its use. Throughout their reply they use both the terms correlation and partial correlation for the same statistic. The actual statistic used by them is the correlation coefficient, not the partial correlation coefficient as they claim!

The correlation coefficient is not based on a multiple regression but only on a simple regression of the dependent variable on a single independent variable without any consideration of the others. It measures the reduction of the total variance of the dependent variable when only one specific variable alone is entered in the model. The partial correlation coefficient, on the other hand, is based on a multiple regression. It is a measure of the reduction in the total variance that results when each independent variable is entered into the model in sequence [see *Draper and Smith*, 1981, p. 265]. Only when the independent variables are completely uncorrelated are the simple correlation and partial correlation coefficients the same.

In our model, two of the independent variables are moderately correlated. Thus, contrary to *Stein et al.* [this issue], the correlation coefficients between the dependent and independent variables cannot properly describe the effect of each independent variable on the full model. We would expect the correlations for *dH* and *tp* to be small, as *Stein et al.* [this issue] have found, due to the much larger effect of refraction which "swamps" the others. Examining the actual partial correlation coefficients (the values listed in Table 1 of *Stein et al.* [this issue] are just correlations), one finds that for the turning point and height difference variables in our model, the partial correlation coefficients are 0.13 and 0.28, respectively (see Table 1), not -0.06 and -0.08 as Stein et al. claim.

The partial correlation coefficient is also not without its problems however. In particular, the coefficients depend on the order in which the independent variables are entered into the model [*Edwards*, 1979, p. 49]. A better statistic is the type II partial correlation coefficient used by SAS (the former is referred to as type I). This is a measure of the reduction in the total variance that results from the addition of an independent variable when all others are included (conversely, it can also be described as a measure of the increment in the total variance when a variable is removed from the full model) [*SAS Institute Inc.*, 1985, pp. 9 and 660]. In our model, we get values of 0.27 and 0.28 for type II partial correlations corresponding to *dH* and *tp*, respectively. When using correct statistical procedures, we therefore find that there is indeed a strong and statistically significant correlation when the other variables in the model are also considered. The correlation coefficients quoted by *Stein et al.* [this issue] are meaningless in multiple regression problems and explain their erroneous conclusions regarding the significance of the turning point settlement and height difference variables.

It is important to realize that partial correlations are used in the context of multiple regressions only as a crude indicator of which variable to select next in a forward stepwise approach to building the regression model (see, e.g., *Draper and Smith* [1981, pp. 307-310]). They are not used to test whether a specific variable is statistically significant and should remain in the model. The variable with the largest (not necessarily statistically significant) partial correlation is simply the next one to add into the model. Once a new variable is found, it is then tested to see if it results in a statistically significant reduction in the total variance of the dependent variable.

The test used for assessing the statistical significance of an independent variable is the partial F test which is identical to the t test on the individual regression coefficients [*Draper and Smith*, 1981, pp. 101-102]. This test is mandatory in multiple regression problems as it accounts for all the independent variables, whereas the correlation coefficients (simple or partial) do not. We have given the confidence levels for these tests [*Craymer and Vaníček*, 1986] and show that the estimated regression coefficients in our model (i.e., refraction, turning point settlement, and height difference) are all statistically significant the the 95% level (see Table 1).

Finally, we wish to point out that contrary to the assertions in Table 1 of *Stein et al.* [this issue], *Craymer and Vaníček* [1986] never used partial correlations. We consider it improper for them to attribute these quantities to us, especially when their values are wrong! We therefore computed the correct values and give them here in our Table 1. *Craymer and Vaníček* [1986] did not include them because we did not use a forward stepwise approach to build our model. Instead we used a backward stepwise method [see *Draper and Smith*, 1981, pp.

TABLE 1. Comparison of Multiple Linear Regressions on Discrepancy Between Forward and Backward Section Runnings With and Without an Intercept "int"

Model	Cumulative Magnitude	Regression Coefficients	Standard Deviation	Partial F Test	Type I Partial Correlation	Type II Partial Correlation	Total F Test and R^2
			All data used				
ref	13.6 mm	4.6×10^{-5} mm/m^3 °C	0.5×10^{-5}	99.99%	0.77	0.77	99.99%
tp	20.3	0.014 mm/tp	0.006	96	0.13	0.27	0.61
dH	-14.2	-0.026 mm/m	0.012	97	0.28	0.28	
int	25.3mm	0.421 mm/section	0.169	98%	n/a	n/a	99.99%
ref	13.5	4.6×10^{-5} mm/m^3 °C	0.5×10^{-5}	99.99	0.76	0.77	0.60
dH	-15.7	-0.029 mm/m	0.012	98	0.31	0.31	
			Sections 57 and 58 omitted				
ref	12.2 mm	4.5×10^{-5} mm/m^3 °C	0.5×10^{-5}	99.99%	0.77	0.75	99.99%
tp	9.2	0.007 mm/tp	0.011	45	0.05	0.08	0.59
dH	-6.6	-0.011 mm/m	0.022	37	0.07	0.07	
int	18.1 mm	0.313 mm/section	0.234	81%	n/a	n/a	99.99%
ref	12.3	4.5×10^{-5} mm/m^3 °C	0.5×10^{-5}	99.99	0.76	0.76	0.59
dH	-11.8	-0.020 mm/m	0.019	70	0.14	0.14	

"*tp*" represents the turning point argument used in modelling the settlement effect, "*dH*" the height difference, and "*ref*" refraction. See *Craymer and Vaníček* [1986] for the derivation of these arguments. Note that the partial correlation coefficients given here are correct. The values quoted by *Stein et al.* [this issue] are actually simple correlation coefficients. Note also that for the models with two sections deleted, the cumulative effects of the variables will be smaller.

305-307; *Neter and Wasserman*, 1974, p. 386; *SAS Institute Inc.*, 1985, p. 765]. In this approach, all variables are included in the multiple regression model. The partial F test is then used to check the statistical significance of each variable. The one with the least significant partial F is omitted from the model and new partial F computed. This is repeated until the partial F for all variables are statistically significant.

In spite of *Stein et al.*'s [this issue] arguments, the fact remains that the regression coefficients in our model are statistically significant. This is proof positive that a multiple linear association exists between the discrepancy and the independent variables describing refraction, turning point settlement, and an effect dependent on height difference.

Intercept Versus Settlement

Stein et al. [this issue] also question the absence of the intercept (or absolute term) in our regression model. We have not taken the "blind" regression approach of simply adding parameters into our model that may have no physical meaning. Instead, we have used an approximation approach where we develop a model based on established physical principles. We know of no such principle which would argue for including such a constant effect in our model. The presence of a constant term makes no physical sense to us in the context of levelling. Although the intercept is statistically insignificant when included in the model, as suggested by *Stein et al.* [this issue], a closer look at the results reveals a strong collinear relation (correlation of -0.9) between the intercept and *tp*. This causes the estimation to become ill-conditioned and, as explained earlier, inflates the variances of the estimates making them appear statistically insignificant. Consequently, Stein et al.'s conclusions based on this near singular model are statistically unsupportable.

Clearly, these highly collinear variables (intercept and *tp*) are accounting for the same effect in the model, and only one of them should be used. Further evidence for this can be found by comparing our regression model (without the intercept) with one where the intercept is substituted for *tp* (see Table 1). From the results it can be seen that both models give statistically identical numerical results; either the intercept or *tp* can be used without significantly influencing the other regression coefficients.

Having to choose between either the intercept or the turning point variable, we have taken the later for the simple reason that we have a physical explanation for its existence. As stated by *SAS Institute Inc.* [1985, p. 712], "Effective model building requires substantive theory to suggest relevant predictors and plausible functional forms of the model." *Craymer and Vaníček* [1986] have shown that one must expect the presence of a settlement effect. Moreover, the magnitude of the effect agrees well with independent field tests as reported by *Craymer and Vaníček* [1985] and *Vaníček et al.* [1985]. On the other hand, we cannot conceive of any reason why the intercept should be present.

Trying to justify the inclusion of the intercept, *Stein et al.* [this issue] define the intercept as the "mean divergence per section." However, the proper definition for the intercept is simply the "*y* intercept" where "*y*" is the axis for the dependent variable (discrepancy). In other words, it is the value of the discrepancy when the independent variables (refraction, turning points, and height difference) are all zero. Although we could argue for a discrepancy when we have zero *dH* and refraction, we cannot imagine any physical justification for the existence of a discrepancy when there are zero turning points (i.e., no levelling)! We think it makes much more sense to constrain the intercept to zero in the model which must have, by definition, a zero discrepancy for zero turning points.

Other Comments by Stein et al.

Stein et al. [this issue] state that *tp* and *dH* explain 9% of the variance in the discrepancy, whereas a much greater proportion (61%) is explained when refraction is also included. They claim, without explanation, that this is due to *tp* and *dH* having correlations of opposite sign to their corresponding regression coefficients. This is completely irrelevant. The reason why *tp* and *dH* explain much less of the variation than refraction does is because the later effect is much larger and has been purposely amplified by the design of the experiment. We would then naturally expect refraction to be the dominate effect in the variance of the discrepancy. Furthermore the 9% reduction in variance is indeed a statistically significant amount with a confidence level of 99.99% [see *Craymer and Vaníček*, 1986, Table 2].

Stein et al. [this issue] also argue that a simple regression on refraction alone explains 56% of the variance, almost the same amount as our model. This too is misleading since this simple model included the intercept which, as we have shown above, is accounting statistically for the same basic effect as *tp*. This simple regression model is statistically equivalent to a multiple regression model with the refraction and *tp* variables but without the intercept.

The claim by *Stein et al.* [this issue] that less than 0.7% of the variation is explained by *tp* is also without merit as this argument is again based on a simple correlation, not a partial correlation. As explained above, it does not account for the effect of the other variables, particularly *dH* which is correlated with *tp*. The actual reduction in variance explained by each of the variables in a multiple regression is, by definition, given by the square of the partial correlation coefficients. For *tp* and *dH*, we find that they explain 7% and 8% of the variation in the discrepancy, respectively, when the other effects are included in the model (i.e., type II partial correlations).

Finally *Stein et al.* [this issue] claim it is "impermissible" that only 7% of the observation fall within the $\pm 1\ \sigma$ error envelope for the trend on *dH*. However, they fail to explain why or what the implications are. What is impermissible is the direct comparison of "apples" (observations) and "oranges" (regression parameters); these are completely different quantities with different stochastic properties which cannot be directly compared. The $\pm 1\ \sigma$ envelope is for the trend, not for the observations as Stein et al. imply. All one can say, based on the data sample, is that 67% of the time the trend on *tp* (not the observations) will fall within this confidence region.

Differential Rod Scale Error

We wish to take this opportunity to amend one of our conclusions. In our paper, we attributed the dependence of the F-B discrepancy on the section height difference to a differential rod scale error (by differential rod scale error we mean the difference in rod scale when observing different parts of the rod). Over the steady sloping terrain of this field test, the short (SSL) and long sight length (LSL) runnings consistently observe the middle and end parts of the rod scales, respectively. Any difference in rod scale between the middle and end of the rods would, we thought, show up in the discrepancy.

It now seems that we were incorrect in linking this effect directly with the section height difference (*dH*). Prior to this comment/reply exchange, R.S. Stein [U.S. Geological Survey, Menlo Park, California, personal communication, 1987] notified us that in this experiment the SSL running was not consistently used in either the F or B directions. Thus the error would have different signs depending on whether the SSL runnings were in the F or B direction. Realizing this, we developed a new argument to properly account for a possible differential rod scale error.

Our original expression describing the differential rod scale effect on the F-B discrepancy is written as [*Craymer and Vaníček*, 1986, equation (21)]

$$(\lambda_F - \lambda_B)\, dH = d\lambda\, dH$$

where λ_F and λ_B are the rod scale errors in the F and B runnings, respectively, $d\lambda$ is the differential rod scale error, and *dH* is the average section height difference. When the SSL coincides with the F running the error $d\lambda_F$ is given by

$$d\lambda_F = (\lambda_S - \lambda_L)$$

where the subscripts S and L refer to the SSL and LSL runnings, respectively. However, when the SSL coincides with the B running, the error $d\lambda_B$ becomes

$$d\lambda_B = (\lambda_L - \lambda_S) = -d\lambda_F$$

In order to resolve this sign difference, we simply use a new variable equal to *dH* when SSL coincides with the F running and -*dH* when it coincides with the B running. We therefore end up with new differential rod scale variable whose coefficient c_λ is

$$c_\lambda = d\lambda = (\lambda_S - \lambda_L)$$

When we consider this new argument in our multiple regression model, its coefficient is not statistically significant at any reasonable significance level. However, the difference in height (*dH*), turning point, and refraction arguments remain as statistically significant as before. Although we cannot justifiably attribute the coefficient for *dH* to a differential rod scale error, the statistical dependence of the F-B discrepancy on *dH* remains very real. *Stein et al.*[this issue] think this is "at odds" with our treatment of the intercept; we point out that we have an alternate variable (*tp*) that explains the same basic effect as the intercept but is also physically meaningful. Unfortunately, we have no such alternate variable for *dH*. Clearly, the physical explanation will require further thought.

Conclusions

Despite the attempts by *Stein et al.* [1986, this issue] to discredit our analysis and conclusions, the fact remains that the regression coefficients for the effects of refraction, turning point settlement, and an effect dependent on section height difference are statistically significant. We have shown that the intercept accounts for the same basic effect as the turning point variable. Having the choice of one or the other, we have chosen the later as we have a good physical explanation for it.

The deletion of two sections from the data sample does not greatly affect the magnitudes or signs of the settlement and height difference effects even though their statistical significance is reduced. We agree that these sections are important in strengthening the resolution of these effects, however, they actually represent a total of 312 individual observations which must not be dismissed lightly as "just one or two" observations.

In conclusion, it is important to realize that one must expect the presence of rod settlement in the F-B discrepancies between the forward and backward runnings of a section as shown by us and others [e.g., *Craymer and Vaníček*, 1985; *Anderson*, 1983; D.S. Schneider, Federal Office of Topography, Bern, Switzerland, personal communication, 1981]. In fact, our results agree extremely well with those obtained by *Anderson* [1983] in a completely independent experiment. Rod settlement does not appear to affect the results obtained by *Stein et al.* [1986] only because they are using discrepancies that do not retain a unique relation with the direction of levelling and thus tend to randomize the systematic effects. In fact, we observe that our estimates of the refraction effect agree better with that obtained by *Stein et al.* [1986] when the other effects (rod settlement and *dH*) are taken into account.

It is also just as important to remember that the rod settlement effect on the average section height difference will cancel when the number of setups in the forward and backward runnings are balanced and the levelling is performed in a procedurally consistent manner. Any imbalance in the number of setups or a change in field procedure will result in an accumulation of the rod settlement effect in the average height differences. In practice, field procedures require the number of setups to be precisely balanced in order to cancel this effect. The levelling experiment reported on here is not typical as it was designed with the intention of amplifying the refraction effect and, consequently, the imbalance in number of setups.

On the other hand, a steady trend between the section discrepancies and turning points indicates that the settlement effect is relatively constant. If this is so, the effect will indeed cancel in the averaging of the elevation differences. If the trend is not well defined, the settlement effect will tend to randomize, and thus one cannot be sure that it will cancel in the average elevation difference. Therefore one should actually look for this steady trend as a sign that the field procedures are being performed in a consistent manner and that all is going well. If the number of setups is not balanced but the levelling has been performed consistently, it may even be possible to reduce the rod settlement effect using an a posteriori estimate of the settlement effect from a multiple regression analysis.

REFERENCES

Anderson, E.G., The effect of vertical motion of the level and rods in precise levelling, paper presented at the 43rd Annual Meeting, Am. Congr. on Surv. and Mapp., Washington, D.C., March 13-18, 1983.

Belsley, D.A., E. Kuh, and R.E. Welsch, *Regression Diagnostics*, John Wiley, New York, 1980.

Craymer, M.R., and P. Vaníček, An investigation of systematic errors in Canadian levelling lines, proceedings of the Third International Symposium on the North American Vertical Datum, IAG, U.S. Department of Commerce, Rockville, Md., April 21-26, pp. 441-450, 1985.

Craymer, M.R., and P. Vaníček, Further analysis of the 1981 southern California field test for levelling refraction, *J. Geophys. Res.*, *91*(B9), 9045-9055, 1986.

Draper, N.R., and H. Smith, *Applied Regression Analysis*, John Wiley, New York, 1981.

Edwards, A., *Multiple Regression and the Analysis of Variance and Covariance*, W.H. Freeman, New York, 1979.

Neter, J., and W. Wasserman, *Applied Linear Statistical Models*, Richard D. Irwin, Homewood, Ill., 1974.

SAS Institute Inc., *SAS User's Guide: Statistics*, Cary, N.C., 1985.

Stein, R.S., C.T. Whalen, S.R. Holdahl, W.E. Strange, and W. Thatcher, Saugus-Palmdale, California field test for refraction error in historical leveling surveys, *J. Geophys. Res.*, *91*(B9), 9031-9044, 1986.

Stein, R.S., C.T. Whalen, S.R. Holdahl, W.E. Strange, and W. Thatcher, Reply to "Comment on 'Saugus-Palmdale, California, field test for refraction error in historical leveling surveys' by R.S. Stein et al." by M.R. Craymer and P. Vaníček, and comment on "Further analysis of the 1981 southern California field test for leveling refraction by M.R. Cramer and P. Vaníček," *J. Geophys. Res.*, this issue.

Vaníček, P., G.H. Carrera, and M.R. Craymer, Corrections for systematic levelling errors in the Canadian levelling network, Contract Rep. 85-001, Geod. Surv. of Can., Surv. and Mapp. Branch, Energy, Mines and Resour., Ottawa, Ont., 1985.

M.R. Craymer and P. Vaníček, Geodetic Research Laboratory, Department of Surveying Engineering, University of New Brunswick, P.O. Box 4400, Fredericton, N.B., Canada E3B 5A3.

(Received November 24, 1986;
revised February 8, 1989;
accepted February 15, 1989.)