A Calibration Experiment in a Longitudinal Survey with Errors-in-Variables

Cindy L. Yu and Jason C. Legg

Cindy L. Yu (E-mail: cindyyu@iastate.edu) and Jason C. Legg (Email: jlegg@iastate.edu) are in the Department of Statistics, Iowa State University, Ames, IA 50010.

ABSTRACT

The National Resources Inventory (NRI) is a large-scale longitudinal survey conducted to assess trends and conditions of nonfederal land. A key NRI estimate is year-to-year change in acres of developed land, where developed land includes roads and urban areas. In 2003, a digital data collection procedure was implemented replacing a map overlay. Data from an NRI calibration experiment are used to estimate the relationship between data collected under the old and new protocols. A measurement error model is postulated for the relationship, where duplicate measurements are used to estimate one of the error variances. If any significant discrepancy is detected between new and old measures, some parameters that govern the algorithm under new protocol can be changed to alter the relationship. Parameters were calibrated so overall averages nearly match for the new and old protocols. Analyses on the data after initial parameter calibration suggest that the relationship is a line with an intercept of zero and a slope of one, therefore the parameters currently used are acceptable. The paper also provides models of the measurement error variances as functions of the proportion of developed land, which is essential for estimating the effect of measurement error for the whole NRI data.

KEY WORDS: Area sampling; Generalized least squares; Longitudinal survey; Measurement error.

1. INTRODUCTION

During a long-term monitoring study, advances in theory and methodology for collecting data occur. Changing data collection procedures can reduce measurement error and other nonsampling errors. For example, the introduction of computer assisted self-administered interviewing (CASI) has been shown to increase reporting accuracy for studies involving sensitive topics (Tourangeau and Smith 1996). Since measuring change is one of the primary objectives of longitudinal surveys, the effect of changing survey mode needs to be measured.

Experiments have been built into surveys to estimate the impact of changes to data collection procedures on data. Schräpler et. al. (2006) describe advantages and disadvantages of switching from personal interviewing to CASI interviewing and report on an experiment focusing on irregular observations and nonresponse rates with mixed modes. Schenker et. al (1993) describe an imputation procedure used to generate time series when the codes for the Census industry and occupation questionnaire changed in 1980. The experiment consisted of recoding a subset of 1970 data and modeling the change. A third example is an experiment to estimate the effect of questionnaire design and format change for the race and Hispanic origin questions in the 2000 US Census. Martin et. al. (2005) compared data collected under different formats in 2000.

We provide details of an experiment that accounts for errors-in-variables when replicate observations on a unit are possible. Replicate observations were taken under a new protocol at the same time on a sample of elements. The replication allows estimation of measurement error variances for new and old procedures. The experiment and analysis are described in the context of a protocol change in the National Resources Inventory (NRI). The procedure can be adapted to other longitudinal surveys involving physical measurements. Our study is similar to instrument calibration and measurement error studies such as gage repeatability and reproducibility experiments (Vardeman and VanValkenburg 1999) common in engineering. Our model accounts for time dependency of measurement errors induced by the longitudinal data structure and collection protocol. Many calibration problems rely on a gold standard, but our study permits measurement error in the current measuring device.

Section 2 describes the NRI and the developed land measurement protocols. Section 3 provides details on the measurement error experimental design. Section 4 contains analyses of the relationship between measurements under new and old protocols. Section 5 includes an analysis of measurement error variances under new and old protocols. Section 6 contains some concluding remarks.

2. NRI BACKGROUND

The NRI is a large-scale monitoring program designed to assess status, condition, and trends of soil, water, and related resources on nonfederal land (Nusser and Goebel 1997). Current reports are for the 48 coterminous states. Data are used for evaluation of public agriculture policy and allocation of funds to environmental programs. Much of the NRI data are observed via photograph interpretation. Prior to 2003, photograph interpretation was performed on a transparent overlay on an aerial photograph. After 2003, the photographs were digitized and interpreted on a computer. Along with the change to digital imagery, a new protocol was created for determining area devoted to developed land. In 2003, determinations were made using both protocols on every segment. A calibration study was conducted using 2003 data to assess the impact of the protocol change and whether adjustments to the new protocol are needed. A calibration study was required because determinations under the new protocol were not independent of the previous 2003 determinations and because determinations under both protocols had measurement error.

The NRI survey has a stratified two-stage design with approximately 300,000 area segments in the basic NRI sample. For the central United States excluding Texas and for most western states, the strata are defined by the Public Land Survey (PLS) System. For states under the PLS, a stratum is defined to be a two mile by six mile block, which is one-third of a township. Typically, two half-mile by half-mile blocks, called segments, are selected within a stratum. Within each selected segment, three points are selected using a restricted randomization procedure to ensure geographic spread. Segment level observations are made on the areas devoted to built-up areas, roads, streams, and small water bodies. We refer to structures and the maintained area around structures as *urban land*, roads and railroads as *roads*, and the combination of urban land and roads as *developed land*.

NRI data were collected at 5-year intervals from 1982 to 1997 and yearly starting in 2000. A subset of about 40,000 segments in the 1997 sample are observed every year. The remaining segments are rotated in and out of data collection with about 35,000 segments observed in most years.

The change in the developed land observation protocol deals with assigning area to residences. The protocol for residential areas from 1982 through 2003 involved the data gatherers delineating the area around residences considered as urban. The boundary of an area polygon was delineated using a hand planimeter on a transparent overlay placed over an aerial photograph. Under the new protocol, data gatherers create a cross using a mouse click on the roof of all of the residences on a digital photograph displayed on a computer monitor. A computer program generates a hexagon centered on each cross on the digitized photograph. Two hexagons are linked if the distance between their boundaries is below a specified threshold. If four or more hexagons are linked, the area of the hexagons is considered developed land. An area entirely enclosed by linked hexagons or other delineated built-up areas, called an enclosure, is considered built-up if the enclosed area is below another specified threshold. Roads are delineated by choosing a line thickness and tracing the road or by delineating the area within the road boundary. Nonresidential urban areas are delineated using a digitized vertex version of the old protocol. Any linked residence hexagons within some tolerance of a nonresidential polygon are considered built-up. Small water bodies are delineated like nonresidential urban areas and small streams are delineated like roads. The protocols for collecting road, nonresidential urban area, small water body, and small stream data are the same for previous and current data collection except that delineation is done on a computer rather than on a transparent overlay.

[FIGURE 1 ABOUT HERE]

Figure 1 is output of the program that translates crosses into hexagons and links polygons for an example segment. The light grey linear polygons are delineated roads. The grid of roads in the top of Figure 1 are related to a future residential area. The dark polygons are delineated nonresidential areas. Hexagons are associated with residences with the center of a hexagon being the location where a data collector placed a cross. A circle surrounds each hexagon. Hexagons are linked if the center of one hexagon is in the circle of another hexagon. The set of three residences in the top left of Figure 1 do not contribute to the total built-up area of the segment. If the nonresidential polygon below the three residences was touching one of the hexagons, the three residential hexagons would contribute to the total built-up area. The total area of developed land for the segment in Figure 1 is the sum of the area for roads, dark hexagons, and delineated polygons.

The intent of the protocol change is to reduce the measurement error in urban area determinations. Marking residences is a more repeatable process than delineation, because the boundary of a delineated area is subject to data gatherer discretion. Roads and nonresidential urban areas involve a decision on what portion of the land is maintained. Therefore, roads and nonresidential determinations remain at the discretion of the data gatherer. Any change in the measurement error distribution for delineations of roads and nonresidential determinations is due to changes in the quality of data collection materials or the switch to digital data collection.

The data from the NRI calibration experiment are used to estimate the relationship between data collected under the old and new protocols. If the relationship is not a line with an intercept of zero and a slope of one, parameters in the program that translates crosses into areas will be modified. The parameters of the program are the size of the hexagons, the distances needed to link hexagons and polygons, and the number of linked polygons needed to count as built-up land. Adjusting the parameters changes the relationship between observations made under the new and old protocols. Another objective of the experiment is to provide an estimate of the relative contribution of the measurement error variance to the total variance of an estimator of total developed land.

3. EXPERIMENT DESIGN

The calibration experiment was designed with replicates for measurement error variance estimation. The NRI data gatherers have access to previously collected data. Therefore, the measurement error is assumed to be correlated over time. The data collection procedure was designed to reduce the correlation between two observations made on the same segment in 2003. Four people are involved in data collection under the new protocol. The first two people make observations for 2001 using the available 1997 materials. The third person and fourth person make observations for 2003, where the third person uses 2001 materials from the first data gatherer and the fourth person uses 2001 materials from the second data gatherer.

A fifth person made a determination for 2003 under the old protocol. Eight data collectors are grouped together. For each eight segments, four data collectors are randomly selected to work on the first four segments and the complement set of data collectors are assigned to work on the second set of four segments. A Latin square design was used to assign the four segments to the four data collectors such that each data collector performed each of the four observation types once. Some control was made across groups of eight segments to ensure mixing of data collectors into the groups of four. A working assumption under this design is that the two observations under the new protocol made in 2003 are independent and are also independent of the observation made under the old protocol. The independence assumption is justified by the inclusion of the intermediate data collector between the 2003 data collector and the original 1997 data collector.

Photograph interpretation occurs at three Remote Sensing Laboratories (RSLs). The

RSLs are called West, Central, and East. Each RSL collects data on states in the region of the RSL. Training of data gatherers occurs at each RSL. Differences between data collection techniques can arise due to differences in geography and personnel at the RSLs. Therefore, the segment selection and analysis were conducted by RSL region.

Segments for the experiment were selected based on geography and 2003 measurements under the previous protocol. Segments were divided into groups based on the area of developed land and area of small water and small streams. Segments completely covered with water, federal land, or developed land are not interesting for the experiment because the residence protocol need not be applied. Therefore, segments classified as 100% urban, 100% federal, or 100% water were not included in the study. Segments with a change in urban, water, or roads from 2001 to 2003 under the old protocol were selected with certainty for the Central and West RSLs. Segments with change usually have development occuring, which can make implementing the new protocol more challenging than for segments without change. A subset of segments in these categories were selected with certainty for the East RSL. The remaining segments were divided into strata (i) presence of developed land but no change in developed land from, (ii) presence of water but no change in water from and no developed land, (iii) and no water and no developed land in the segment in 2001 and 2003. Within each category, segments were sorted geographically and a systematic sample was selected. The West RSL had 607 segments, the Central RSL had 1055 segments, and the East RSL had 1036 segments, for a total of 2698 segments.

Twenty-seven segments containing federal land were removed from the analysis dataset. Because the boundary of federal land within a segment was not determined in the experiment, it was impossible to determine if the developed land areas were on federal land. Seventy-seven segments where all three calibration experiment observations for 2003 have no developed land were removed from the analysis dataset. Including segments with no developed land or all developed land in the analysis would increase the evidence that observations under the new and old protocol estimate the same quantity, possibly masking some departures near the extremes. Developed land areas were converted into proportions by dividing built-up determinations by digitized segment size.

4. ESTIMATION OF THE MEAN FUNCTION

In this section, we postulate measurement error models to estimate the mean functions of the new and old determinations. We calibrate the new procedure to the target of the old procedure in order to gain consistency in trend estimators. The dataset used for this analysis is from the West RSL and contains 503 segments. Protocol calibration and analysis conducted for the East and Central RSLs gave similar results to the West analysis, but East and Central RSL results are not presented here.

Let Y_{ji} be the proportion of developed land in segment i made by observer j ($j = 1$ or 2) under the new protocol and let X_i be the proportion of developed land in segment i under the old protocol. The mean of X_i is 0.231 (0.011) and the mean of $0.5(Y_{1i} + Y_{2i})$ is 0.232 (0.011). The numbers in parenthesis are standard errors throughout the paper. The correlation between X_i and $0.5(Y_{1i} + Y_{2i})$ is 0.954.

The proposed model for the 2003 data is a segmented linear model with a slope break at 0.5, that is

$$X_i = x_i + u_i, \tag{1}$$

$$Y_{ji} = \eta_0 + \eta_1 x_i v_i + 0.5\eta_1(1 - v_i) + \eta_2(x_i - 0.5)(1 - v_i) + e_{ji}, \tag{2}$$

$$v_i = \begin{cases} 1 & x_i < 0.5 \\ 0 & x_i \geq 0.5, \end{cases} \tag{3}$$

$$x_i \sim (\mu_x, \sigma_x^2), \tag{4}$$

and

$$\begin{bmatrix} u_i \\ e_{1i} \\ e_{2i} \end{bmatrix} | x_i \sim \left[0, \begin{bmatrix} \sigma_{ui}^2 & 0 & 0 \\ 0 & \sigma_{ei}^2 & 0 \\ 0 & 0 & \sigma_{ei}^2 \end{bmatrix} \right], \tag{5}$$

for all i and $j = 1, 2$, where x_i is the long run average of repeated observations under the old protocol for segment i, μ_x and σ_x^2 are the mean and the variance of x_i's selected into the sample, and u_i and e_{ji} are measurement errors on segment i under the old and new protocols, respectively. In this model, the slope before $x = 0.5$ is η_1 and the slope after $x = 0.5$ is η_2. The errors for different segments are assumed to be independent. From the experimental design, we assume e_{1i}, e_{2i} and u_i are conditionally uncorrelated from each other for each segment i. The measurement error variances, σ_{ei}^2 and σ_{ui}^2, represent the variances of errors in repeated measurements on segment i in 2003 under the new protocol and old protocol, respectively. The measurement error variances are likely a function of the proportion of developed land in the segment, x_i. Note that $x = 0.5$ provides a good chance to detect a trajectory change, although other break points would work.

Because x_i in (3) is not observed an iterative method of estimation was implemented. First a linear model without the slope break was introduced, that is

$$X_i = x_i + u_i, \tag{6}$$

$$Y_{ji} = \beta_0 + \beta_1 x_i + e_{ji}, \tag{7}$$

the moments of the errors are defined in (5), and the x_i sample moments are as defined in (4). To estimate the parameters, we define the observation vector

$$\boldsymbol{Z_i} = (Z_{1i}, Z_{2i}, Z_{3i}) = (X_i, 0.5[Y_{1i} + Y_{2i}], 2^{-0.5}[Y_{1i} - Y_{2i}]). \tag{8}$$

The Z_i representation has a less complex covariance matrix than (X_i, Y_{1i}, Y_{2i}) since the sample variance of Z_{3i} is a direct estimator for the average measurement error under the new protocol and $Y_{1i} + Y_{2i}$ is uncorrelated with $Y_{1i} - Y_{2i}$. Let the sample covariance matrix of $\boldsymbol{Z}$ be

$$\boldsymbol{m} = (n-1)^{-1} \sum_{i \in A} (\boldsymbol{Z}_i - \overline{\boldsymbol{Z}})'(\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}), \tag{9}$$

where A is the set of indices in the calibration sample. Under the model, the sample covariance matrix has expectation

$$E(\boldsymbol{m}) = \begin{bmatrix} \sigma_x^2 + \sigma_{a,u}^2 & \beta_1 \sigma_x^2 & 0 \\ \beta_1 \sigma_x^2 & \beta_1^2 \sigma_x^2 + 0.5\sigma_{a,e}^2 & 0 \\ 0 & 0 & \sigma_{a,e}^2 \end{bmatrix}, \tag{10}$$

where $\sigma_{a,u}^2$ and $\sigma_{a,e}^2$ denote the averages of σ_{ui}^2 and σ_{ei}^2, respectively. The term in the second row and second column of (10), for example, is the expectation of the sample variance of $\beta_0 + \beta_1 x_i + 0.5(e_{1i} + e_{2i})$. Denote the element in row r and column c of matrix $\boldsymbol{m}$ by m_{rc}. By (10), the method of moments estimators are

$$\widehat{\beta}_1 = m_{12}^{-1}(m_{22} - 0.5m_{33}), \tag{11}$$

$$\widehat{\beta}_0 = \overline{Z}_2 - \widehat{\beta}_1 \overline{Z}_1, \tag{12}$$

$$\widehat{\sigma}_x^2 = (m_{22} - 0.5m_{33})^{-1} m_{12}^2, \tag{13}$$

$$\widehat{\sigma}_{a,e}^2 = m_{33}, \tag{14}$$

$$\widehat{\sigma}_{a,u}^2 = m_{11} - \widehat{\sigma}_x^2, \tag{15}$$

and

$$\widehat{\theta} = \widehat{\sigma}_{a,e}^{-2} \widehat{\sigma}_{a,u}^2. \tag{16}$$

The method of moments estimators are derived by solving for the parameters in the equation

$\boldsymbol{m} = E(\boldsymbol{m})$, noting that the number of parameters to be estimated matches the number of nonzero components in (10). The $\widehat{\theta}$ is an estimator of the ratio of the average error variance of the old protocol to the average error variance of the new.

Parameter estimates using estimators (11)-(16) are

$$\begin{array}{rcccccc} (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\sigma}_x^2, \widehat{\sigma}_{a,e}^2, \widehat{\sigma}_{a,u}^2, \widehat{\theta}) = & (0.0011, & 0.998, & 0.0556, & 0.00086, & 0.00498, & 5.8), \\ & (0.0032) & (0.015) & (0.0044) & (0.00022) & (0.00058) & (2.0). \end{array}$$

Standard errors were estimated using a delete-1 segment jackknife, where the jackknife weights are those of simple random sampling. The intercept under this model is not statistically significantly different from 0 and the slope is not statistically significantly different from one. The error variance under the old protocol is estimated to be 5.8 times the error variance under the new protocol.

Using these estimated parameters, x_i was estimated for each observation using an estimated generalized least squares (EGLS) estimator. The vector $(Z_{1i}, Z_{2i} - \widehat{\beta}_0)$ is regressed on $(1, \widehat{\beta}_1)'$ using weights equal to the inverses of $\widehat{\sigma}_{a,u}^2$ and $\widehat{\sigma}_{a,e}^2$, where $(\widehat{\beta}_0, \widehat{\beta}_1)$ are the previously estimated coefficients. That is,

$$\widehat{x}_i = \widehat{w}_1 Z_{1i} + \widehat{w}_2 (Z_{2i} - \widehat{\beta}_0)/\widehat{\beta}_1 \tag{17}$$

where $\widehat{w}_1 = \widehat{\sigma}_{a,u}^{-2}/(\widehat{\sigma}_{a,u}^{-2} + 0.5\widehat{\beta}_1^2\widehat{\sigma}_{a,e}^{-2})$, and $\widehat{w}_2 = 0.5\widehat{\beta}_1^2\widehat{\sigma}_{a,e}^{-2}/(\widehat{\sigma}_{a,u}^{-2} + 0.5\widehat{\beta}_1^2\widehat{\sigma}_{a,e}^{-2})$.

With an estimator for v_i, we can estimate our original segmented model (1) - (5) that allows for a shift in the slope at $x = 0.5$. To ease the computation of the correction matrix when adjusting for the effect of measurement errors in regression estimators, we rewrote the original split line model in terms of X_i as a function of y_i, which is the long run average of repeated observations under the new protocol for segment i. The model is

$$Y_{ji} = y_i + e_{ji}, \tag{18}$$

$$X_i = \delta_0 + \delta_1 y_i v_i + 0.5\delta_1(1 - v_i) + \delta_2(y_i - 0.5)(1 - v_i) + u_i, \tag{19}$$

and the moments of the errors are defined the same as (5). The indicator variable, v_i, is replaced by an estimator $\widehat{v}_i$ by substituting $\widehat{x}_i$ for x_i in (3). The simple regression of Z_{1i} (i.e. X_i) on $G_i = (1, Z_{2i}\widehat{v}_i + 0.5(1-\widehat{v}_i), (Z_{2i} - 0.5)(1-\widehat{v}_i))$ produces biased estimators of the parameters because Z_{2i} is measured with error (Fuller 1987, p. 4). We adjust the equation defining the regression estimators to account for the effect of measurement error. Let A_1 denote the part of the sample where $\widehat{v}_i = 1$ and A_2 denote the part of the sample where $\widehat{v}_i = 0$. The bias corrected regression estimator is

$$\left[\widehat{\delta}_0, \widehat{\delta}_1, \widehat{\delta}_2\right]' = (G'G - C)^{-1}(G'Z_1), \tag{20}$$

where

$$C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & C_1 & 0 \\ 0 & 0 & C_2 \end{bmatrix}, \tag{21}$$

$$C_1 = \sum_{i \in A_1} 0.25(Y_{1i} - Y_{2i})^2, \tag{22}$$

and

$$C_2 = \sum_{i \in A_2} 0.25(Y_{1i} - Y_{2i})^2. \tag{23}$$

See Fuller (1996, p. 103). The simple regression of Z_{1i} on G_i would have a denominator of $G'G$ in (20), which contains Z_{2i}^2 terms. The expectation of Z_{2i}^2 is

$$\begin{aligned} E(Z_{2i}^2) &= E[y_i + 0.5(e_{1i} + e_{2i})]^2 \\ &= E(y_i^2) + 0.5\sigma_{ei}^2. \end{aligned} \tag{24}$$

The terms C_1 and C_2 remove the $0.5\sigma_{ei}^2$ terms from the expectation of the denominator of the regression equation since $E[0.25(Y_{1i} - Y_{2i})^2] = 0.5\sigma_{ei}^2$. The resulting regression coefficient has approximately the expectation of regressing Z_{1i} on $(1, y_i\widehat{v}_i + 0.5(1-\widehat{v}_i), (y_i - 0.5)(1-\widehat{v}_i))$.

Under the model described in (18) and (19), the sample covariance of $\boldsymbol{Z}$ can be used to estimate the average error variances once the regression coefficients are obtained. The estimator for $\sigma^2_{a,e}$ is m_{33} from (14). An estimator for $\sigma^2_{a,u}$ is obtained by combining estimators from A_1 and A_2.

To estimate the error variances, let $\boldsymbol{m_v}$ be the sample covariance matrix of (Z_1, Z_2) for data with $\widehat{v}_i = 1$ and $\boldsymbol{m_{1-v}}$ be the covariance sample covariance matrix of (Z_1, Z_2) for data with $\widehat{v}_i = 0$. The expectations of the sample covariance matrices are

$$E\{\boldsymbol{m_v}\} = \begin{bmatrix} \delta_1^2\sigma^2_{vy} + n_1^{-1}\sum\limits_{i\in A_1}\sigma^2_{ui} & \delta_1\sigma^2_{vy} \\ \delta_1\sigma^2_{vy} & \sigma^2_{vy} + \sum\limits_{i\in A_1}(2n_1)^{-1}\sigma^2_{ei} \end{bmatrix} \tag{25}$$

and

$$E\{\boldsymbol{m_{1-v}}\} = \begin{bmatrix} \delta_2^2\sigma^2_{1-v,y} + n_2^{-1}\sum\limits_{i\in A_2}\sigma^2_{ui} & \delta_2\sigma^2_{1-v,y} \\ \delta_2\sigma^2_{1-v,y} & \sigma^2_{1-v,y} + \sum\limits_{i\in A_2}(2n_2)^{-1}\sigma^2_{ei} \end{bmatrix}, \tag{26}$$

where n_1 is the size of A_1, n_2 is the size of A_2, σ^2_{vy} is the variance of y_i in A_1, and $\sigma^2_{1-v,y}$ is the variance of y_i in A_2. Method-of-moments estimators for σ^2_{vy} and $\sigma^2_{1-v,y}$ are

$$\widehat{\sigma}^2_{vy} = \widehat{\delta}_1^{-1} m_{v,12} \tag{27}$$

and

$$\widehat{\sigma}^2_{1-v,y} = \widehat{\delta}_2^{-1} m_{1-v,12}, \tag{28}$$

where $(\widehat{\delta}_1, \widehat{\delta}_2)$ is the estimator from (20). An estimator for $\sigma^2_{a,u}$ is

$$\widehat{\sigma}^2_{a,u} = (n_1 + n_2)^{-1}(n_1\{m_{v,11} - \widehat{\delta}_1^2\widehat{\sigma}^2_{vy}\} + n_2\{m_{1-v,11} - \widehat{\delta}_2^2\widehat{\sigma}^2_{1-v,y}\}). \tag{29}$$

Estimates for the parameters of model (18)-(19) are

$$
\begin{aligned}
(\widehat{\delta}_0, \widehat{\delta}_1, \widehat{\delta}_2, \widehat{\sigma}^2_{a,e}, \widehat{\sigma}^2_{a,u}, \widehat{\theta}) = \quad & (0.0012, \quad 0.985, \quad 1.045, \quad 0.00086, \quad 0.00454, \quad 5.3), \\
& (0.0033) \quad (0.027) \quad (0.064) \quad (0.00022) \quad (0.00077) \quad (1.7),
\end{aligned}
$$

where standard errors were computed using a delete-1 segment jackknife.

The intercept $\widehat{\delta}_0$ is not statistically significantly different from zero and both the slopes, $\widehat{\delta}_1$ and $\widehat{\delta}_2$, are not statistically significantly different from one. The estimated $\widehat{\sigma}^2_{a,u}$ and $\widehat{\theta}$ are smaller than the corresponding estimates from model (6)-(7), but the difference in estimates is not large.

We computed an approximate F-test of

$$H_o : (\delta_0, \delta_1, \delta_2) = (0, 1, 1) \tag{30}$$

versus

$$H_a : (\delta_0, \delta_1, \delta_2) \neq (0, 1, 1). \tag{31}$$

The test statistic was

$$(3SSF)^{-1} 497(SSR - SSF), \tag{32}$$

where SSF is the residual sum of squares from estimating the split line model of (18)-(19) and SSR is the residual sum of squares from fitting the model with the constraints of the null hypothesis. The denominator degrees of freedom of 497 is $n - 6$, where the 6 is the penalty for estimating $(\delta_0, \delta_1, \delta_2, \sigma^2_x, \sigma^2_{a,e}, \sigma^2_{a,u})$ and three is for difference in number of parameters between the full and reduced models. The F statistic is 0.52, which when compared to F distribution with 3 and 497 degrees of freedom results in a p-value of 0.67. Therefore, we accepted the reduced model of

$$Y_{ji} = y_i + e_{ji}, \tag{33}$$

$$X_i = y_i + u_i. \tag{34}$$

Figure 2 contains both the fitted split line (solid) and a $(0, 0)$ to $(1, 1)$ reference line (dashed). We divide the data set into 10 bins with an equal number of observations up to rounding from data sorted by $\widehat{x_i}$ values. The bins are useful for display and provide groups for determining linear departures other than a split at $x_i = 0.5$. Figure 2 shows the mean of Z_2 versus Z_1 in each bin. The binned means lie closely around the lines, indicating the reduced model $((\delta_0, \delta_1, \delta_2) = (0, 1, 1))$ suffices for describing the data.

[FIGURE 2 ABOUT HERE]

Collectively, the result of the F-test and the evidence in Figure 2 suggest that the relationship between data collected under the old and new protocols is a line with an intercept of zero and a slope of one. Under the reduced model, the estimated average error variances can be obtained using Equations (13) to (15), where the coefficients β_0 and β_1 in (10) are replaced by 0 and 1, respectively. The estimates are

$$
\begin{aligned}
(\widehat{\sigma}_x^2, \widehat{\sigma}_{a,e}^2, \widehat{\sigma}_{a,u}^2, \widehat{\theta}) = \quad & (0.0555, \quad 0.00086, \quad 0.00498, \quad 5.8), \\
& (0.0042) \quad (0.00022) \quad (0.00058) \quad (2.0).
\end{aligned}
$$

A large difference between the mean of Z_1 and the mean of Z_2 within a bin indicates a lack of fit for the corresponding region on the line. We tested whether the mean of Z_1 is statistically significantly different from the mean of Z_2 within each bin using an approximate t-test (Table 1). The t-statistics were constructed as bias adjusted Beale ratios to account for skewness (Tin 1965). The t-tests provide evidence that the observations under the new and old protocol differ for segments with little developed land. However, misfitting the function near $x_i = 0$ will have a small effect on total estimates. Part of the difference between new and old protocols is attributable to the bias in the calibration sample selection. Segments without developed land in 2003 under the old protocol were not selected unless they contained water features. Therefore, the occurrence of a segment without developed land under the old protocol and developed land under the new protocol is much less frequent in our sample than the occurrence of a segment with developed land under the old protocol and no developed land under the new protocol. This fact can be expected to bias the estimated relationship between old and new protocol observations when x_i is near 0. Further, the differences in the small bins are primarily due to differences in road areas and boundary urban areas. The new protocol is the same as the old protocol for road measurements, and hence, changing the computer program will not affect the differences due to roads. Urban areas on the boundary of the segment present a problem for the new protocol, since data collectors no longer create polygons for residential areas or note houses outside of a segment. Also, recall that observations with zero on both determinations are not included in the analysis. The binned mean tests are mostly in agreement with the conclusion from the split line F-test. Therefore, parameters in the program that translates crosses into areas under new protocol are accepted for the West RSL.

[TABLE 1 ABOUT HERE]

5. ESTIMATION OF THE VARIANCE FUNCTION

5.1 Estimating the Variance Function

The calibration experiment provides the opportunity to estimate the effect of measurement error on NRI estimators. In order to extend the variance results to a larger set of data than the calibration data set, we need a functional form for the measurement error variance. The reason for this requirement is that the calibration experiment sample is partially a purposive sample of NRI segments. If the measurement error variance is a function of x_i, then the estimates of the average variances depend on the set of x_i chosen for the calibration experiment. Modeling the variance functions is difficult due to a few extreme differences between observations made on the same segment. Model assumptions presented below are made to construct estimators of the measurement error variance functions. However, the assumptions are not believed to be true for all of the data, nor would many standard diagnostic procedures be possible to check the validity of assumptions.

The expectations of the squared deviations Z_{3i}^2 and $(Z_{1i} - Z_{2i})^2$ were estimated as a function of x_i. Two assumptions are put on the functional form of variances. One constraint is that the functions are symmetric around 0.5. The underlying assumption is that delineation of developed land when the true proportion is 40% is of the same level of difficulty as when the true proportion is 60%. In other words, the delineation of an area in a particular segment requires the same effort as the delineation of the complement of the area. The second assumption is that the measurement error variance function for the new protocol is proportional to the measurement error variance function for the old protocol. A plot of $Z_{3i}^{-2}(Z_{1i} - Z_{2i})^2$ versus x_i is flat except near zero and one, providing evidence for the second modeling assumption.

Plots of the sample variances of Z_{3i}^2 and $(Z_{1i} - Z_{2i})^2$ against the means of Z_{3i}^2 and $(Z_{1i} - Z_{2i})^2$ using the bins of Figure 2 show that the variance of the squared deviations increases as the mean of the squared deviations increases. A working assumption for modeling is that the variances of centered Z_{3i}^2 and $(Z_{1i} - Z_{2i})^2$ are proportional to $[E(Z_{3i}^2)]^2$ and $[E(Z_{1i} - Z_{2i})^2]^2$, respectively. This working assumption is that of a constant coefficient of variation model, commonly used to model data with increasing variances. The constant coefficient of variation assumption is used to provide weights for generalized nonlinear least squares estimation.

Initial models were fit to the squared deviations. The distribution of the squared deviations conditional on x are highly right skewed. Due to the skewness, the fitted functions were poorly estimated. Transformations of the data were explored to find a suitable transformation. The square root transformation of squared deviations decreased the effect of skewness in the data enough to make the least squares solution reasonable. The working models in the transformed scale are

$$E|Z_{3i}| = \gamma_0 + \gamma_1(0.5^{2.5} - |x_i - 0.5|^{2.5}) := g_i, \tag{35}$$

and

$$E|Z_{1i} - Z_{2i}| = \kappa(\gamma_0 + \gamma_1(0.5^{2.5} - |x_i - 0.5|^{2.5})) = \kappa g_i, \tag{36}$$

which are symmetric around 0.5 and proportional to each other. In fitting the model, the estimated proportion, $\widehat{x}_i$, was used as a proxy for x_i. Since the model is not linear in coefficients, the Gauss-Newton algorithm was used to obtain the nonlinear generalized least squares fit. The estimating equations were weighted by an initial estimate of g_i and κ. The use of weights comes from the constant coefficient of variation working assumption. The 2.5 power was determined primarily by comparing the fit across powers $p = K/2$ for $K = 1, 2, \ldots$. The residual mean squared error was used as a fit statistic. The distributions of the absolute deviations are well approximated by the distributions of a multiple of χ_1^2 random variables. Therefore, we compared the residual mean squared errors to 2, the variance of a χ_1^2 random variable. The 2.5-power model gave the residual mean squared error of 2.06, which was the mean squared error closest to 2 for the powers we considered and the behavior of the standardized residuals was similar across values of $\widehat{x}_i$. The estimated coefficients of the variance functions and their delete-1 segment jackknife standard errors are

$$\begin{aligned}(\widehat{\kappa}, \widehat{\gamma}_0, \widehat{\gamma}_1, \widehat{\theta}) = \quad & (3.55, \quad 0.00212, \quad 0.129, \quad 5.8), \\ & (0.31) \quad (0.00030) \quad (0.012) \quad (2.7).\end{aligned}$$

In order to estimate $\widehat{\theta}$, the ratio of the error variance in the previous protocol to the error variance of the current protocol, the variance functions in (35) and (36) were converted back to squared scale. We ratio adjusted the fitted functions so that the average of the squared fitted functions is the same as the average of Z_{3i}^2 and $(Z_{1i} - Z_{2i})^2$. Let

$$R_1 = \left(\sum_{i=1}^{n} \widehat{g}_i^2\right)^{-1} \sum_{i=1}^{n} Z_{3i}^2 \tag{37}$$

and

$$R_2 = \left(\sum_{i=1}^{n} \widehat{\kappa}^2 \widehat{g}_i^2 \right)^{-1} \sum_{i=1}^{n} (Z_{1i} - Z_{2i})^2, \tag{38}$$

where $\widehat{g}_i^2$ is the square of the estimated function in (35). Estimators for the mean of the squared deviations are

$$\widehat{E}(Z_{3i}^2) = \widehat{\sigma}_{ei}^2 = R_1 \widehat{g}_i^2 \tag{39}$$

and

$$\widehat{E}(Z_{1i} - Z_{2i})^2 = 0.5\widehat{\sigma}_{ei}^2 + \widehat{\sigma}_{ui}^2 = R_2 \widehat{\kappa}^2 \widehat{g}_i^2. \tag{40}$$

An estimator of the ratio of measurement error variances, $\theta = \sigma_{ei}^{-2} \sigma_{ui}^2$, is

$$\widehat{\theta} = R_1^{-1} R_2 \widehat{\kappa}^2 - 0.5, \tag{41}$$

which is derived by solving (39) and (40) for $\widehat{\sigma}_{ei}^{-2} \widehat{\sigma}_{ui}^2$. Estimators (39) and (40) can also be derived under the working assumptions of $V(|Z_{3i}|) = R_1^{\star} g_i^2$ and $V(|Z_{1i} - Z_{2i}|) = R_2^{\star} \kappa^2 g_i^2$, where $R_1^{\star}$ and $R_2^{\star}$ are constants. Equation (39) and (40) are the estimators that would be derived by replacing the terms in

$$E(Z_{3i}^2) = (E|Z_{3i}|)^2 + V(|Z_{3i}|) \tag{42}$$

and

$$E[(Z_{1i} - Z_{2i})^2] = (E|Z_{1i} - Z_{2i}|)^2 + V(|Z_{1i} - Z_{2i}|) \tag{43}$$

with the corresponding moment estimators, where $\widehat{V}(|Z_{3i}|) = \widehat{R}_1^{\star} \widehat{g}_i^2$ and $\widehat{V}(|Z_{1i} - Z_{2i}|) = \widehat{R}_2^{\star} \widehat{\kappa}^2 \widehat{g}_i^2$.

Standard errors were computed using a delete-1 segment jackknife variance estimator. The estimated θ of 5.8 is near the estimate using the average variances when fitting the mean function earlier. The estimated variance of the ratio of variances is not well estimated in any of our results due to the skewness in the distribution, which explains the discrepancy between standard errors of θ estimators. In order to see the two fits of (39) and (40) on the squared scale, we plot the fitted functions of squares and standardize them to the same scale (Figure 3). The fitted functions track the binned means of Z_{3i}^2 and $(Z_{1i} - Z_{2i})^2$ well for low proportions of developed land area. The functions slightly underestimate the average measurement error variance for low proportions of developed land and overestimate for large proportions of developed land. The eighth binned mean of Z_{3i}^2 is far smaller than expected under the model. However, fitting the anomaly and the other binned means would require a much more complicated functional form. Overall, the model fits the data reasonably well on the squared scale for a relatively simple functional form and the model furnishes adequate results.

[FIGURE 3 ABOUT HERE]

5.2 Discussion About the Estimation of the Variance Function

The parameter estimators for the variance functions do not include adjustments for bias caused by measurement error. Adjustments would involve specifying higher order moments for e_{ji} and u_i. The estimated variance function can be used to compute a regression bias adjusted estimator of x_i. See Carroll and Stefanski (1990) and Fuller (1987, pp. 20-25). From (17), an approximation to the centered variance of $\widehat{x}_i$ is

$$\widehat{V}\{\widehat{x}_i - x_i\} = \widehat{w}_1^2 \widehat{\sigma}_{u,i}^2 + \widehat{w}_2^2 \widehat{\sigma}_{e,i}^2 / 2, \tag{44}$$

where we know that $\widehat{w}_1$ and $\widehat{w}_2$ are $O_p(n^{-1/2})$ estimators of their associated constants. A regression adjusted estimator of x_i is

$$\tilde{x}_i = \overline{Z}_1 + (\widehat{\sigma}_x^2 + \widehat{V}\{\widehat{x}_i - x_i\})^{-1} \widehat{\sigma}_x^2 (\widehat{x}_i - \overline{Z}_1), \tag{45}$$

where $\overline{Z}_1$ is the average of X_i. Estimator (45) is a shrinkage of $\widehat{x}_i$ toward the mean of X_i. If the estimate for measurement error variance is zero, then $\widehat{x}_i$ is the same as $\tilde{x}_i$. Estimator

(45) uses an assumption of normally distributed data.

The parameters of (35) and (36) were estimated using $\tilde{x}_i$ in place of $\widehat{x}_i$ in the nonlinear generalized least squares. The estimates using $\tilde{x}_i$ are

$$\begin{aligned}(\widehat{\kappa}, \widehat{\gamma}_0, \widehat{\gamma}_1) = \quad & (3.55, \quad 0.00211, \quad 0.129), \\ & (0.31) \quad (0.00030) \quad (0.012),\end{aligned}$$

which are very close to the estimates using $\widehat{x}_i$. The congruence of estimates indicates that the effect of measurement error on the variance function estimators is small.

With the functional form of the measurement error variance, we are able to extend the variance results to a larger set of data than calibration data. Define equation (39) as $f(x_i) = \sigma^2_{ei} = R_1 g^2(x_i)$. In order to estimate σ^2_{ei} in the future using an observation under the new protocol, Y_i, we expand $f(Y_i)$ around x_i. By taking expectations, we obtain

$$\sigma^2_{ei} = \frac{E(f(Y_i))}{1 + 0.5 f''(\xi_i)}, \tag{46}$$

where ξ_i is between x_i and Y_i. An estimator of σ^2_{ei} is

$$\widehat{\sigma}^2_{ei} = \frac{f(Y_i)}{1 + 0.5 f''(Y_i)}, \tag{47}$$

where ξ_i in the denominator is replaced by Y_i.

The $\widehat{\sigma}^2_{ei}$ can be use to estimate the fraction of the variance due to measurement error in a total estimator. Let the estimator of developed land in a particular state be $\widehat{T}_L = \sum_i w_i S_i Y_i$, where S_i is the digitized area size for segment i and w_i are design weights from NRI sampling. The variance of measurement error in the acres of developed land for a segment of size S_i is $S_i^2 \sigma^2_{ei}$. Let $\widehat{Var}(\widehat{T}_L)$ be an estimator of the variance of $\widehat{T}_L$, where the finite population correction is ignored. The proportion of the total variance of the estimator attributed to measurement error under the new protocol can be estimated with

$$\frac{\sum_i w_i^2 S_i^2 \hat{\sigma}_{ei}^2}{\widehat{Var}(\widehat{T}_L)}. \tag{48}$$

Also, the $\widehat{\sigma}_{ei}^2$ can be used to correct the bias in the regression estimation. The correction is similar to that of the bias correction in the mean function estimation.

6. DISCUSSION

The parameters used to translate marked residences into developed area have been adjusted during data collection and analysis. The parameters used in this article for the West RSL provide encouraging results that estimators under the new and old protocol coincide within an acceptable tolerance. Adjustments to the protocol have been attempted for the discrepancy between measurements when the proportion of developed land is very low. Specifically the number of linked houses needed to count toward developed land were reduced to three and the distance for linking houses was reduced. However, the adjustments did not solve the lack of fit problem near $x_i = 0$. Analysis related to the effect of segment size and regional differences was conducted with separate parameter estimation for small, medium, and large segments and for mountain, Pacific northwest, and arid regions. Regional effects were small and segment size did not impact results. Similar procedures are used to examine the relationship between observations under the new and old protocols for the Central and East RSLs. Different program parameters were set for the Central and East RSLs.

ACKNOWLEDGEMENTS

The authors thank Cooperative Agreement No. 68-3A75-4-122 between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University and thank Wayne A. Fuller for comments and guidance during

analysis and writing.

REFERENCES

Carroll, R. J., and Stefanski, L. A.(1990), "Approximate quasi-likelihood estimation in models with surrogate predictors," *Journal of the American Statistical Association*, 85, 652-663.

Fuller, W. A. (1987), *Measurement Error Models*, New York: John Wiley and Sons Inc.

Martin, E., de la Puente, M., and Bennett, C.(2005), "The effects of questionnaire and content change on race data: results of a replication of 1990 race and origin questions," *US Bureau of the Census: Research Report Series, Survey Methodology #2005-05*.

Nusser, S. M., and Goebel, J. J. (1997), "The National Resources Inventory: a long-term multi-resource monitoring programme," *Environmental and Ecological Statistics*, 4, 181-204.

Schenker, N, Treiman, D. J., and Weidman, L. (1993), "Analyses of public use Decennial Census data with multiply imputed industry and occupation codes," *Applied Statistics*, 42, 545-556.

Schräpler, J.-P., Schupp, J., and Wagner, G. G. (2006), "Changing from PAPI to CAPI: A longitudinal study of mode-effects based on an experimental design," *DIW Berlin, German Institute for Economic Research*.

Tin, M. (1965), "Comparison of some ratio estimators," *Journal of the American Statistical Association*, 60, 294-307.

Tourangeau, R., and Smith, T. W. (1996), "Asking sensitive questions: the impact of data collection mode, question format, and question content," *Public Opinion Quarterly*, 60, 275-304.

Vardeman, S. B., and VanValkenburg, E. S. (1999), "Two-way random effects analyses and gauge R&R studies," *Journal of the American Statistical Association*, 41, 202-211.

Table 1. Approximate t-test for the Differences Between Z1 and Z2 Over Ten Bins

Bins	1	2	3	4	5
mean of Z_1	0.008	0.027	0.045	0.083	0.109
mean of Z_2	0.004	0.020	0.037	0.076	0.127
t-value	4.31	2.70	1.82	1.17	-2.21
Bins	6	7	8	9	10
mean of Z_1	0.172	0.249	0.350	0.487	0.775
mean of Z_2	0.187	0.255	0.355	0.495	0.756
t-value	-1.49	-0.32	-0.25	-0.55	1.45

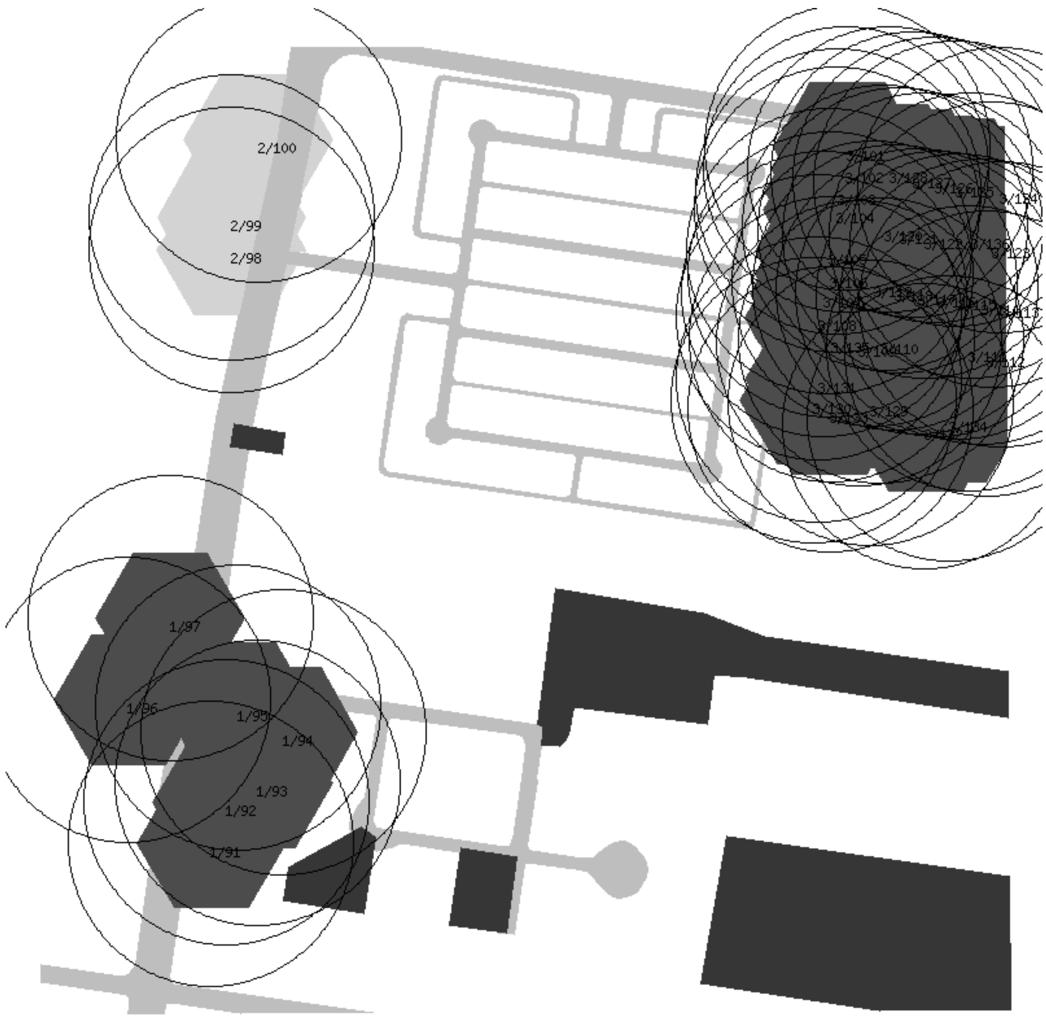

Figure 1. Example segment with hexagons and delineated built-up polygons. Hexagons are centered on residence, road polygons appear as lines, and nonresidential areas are irregular shaped polygons.

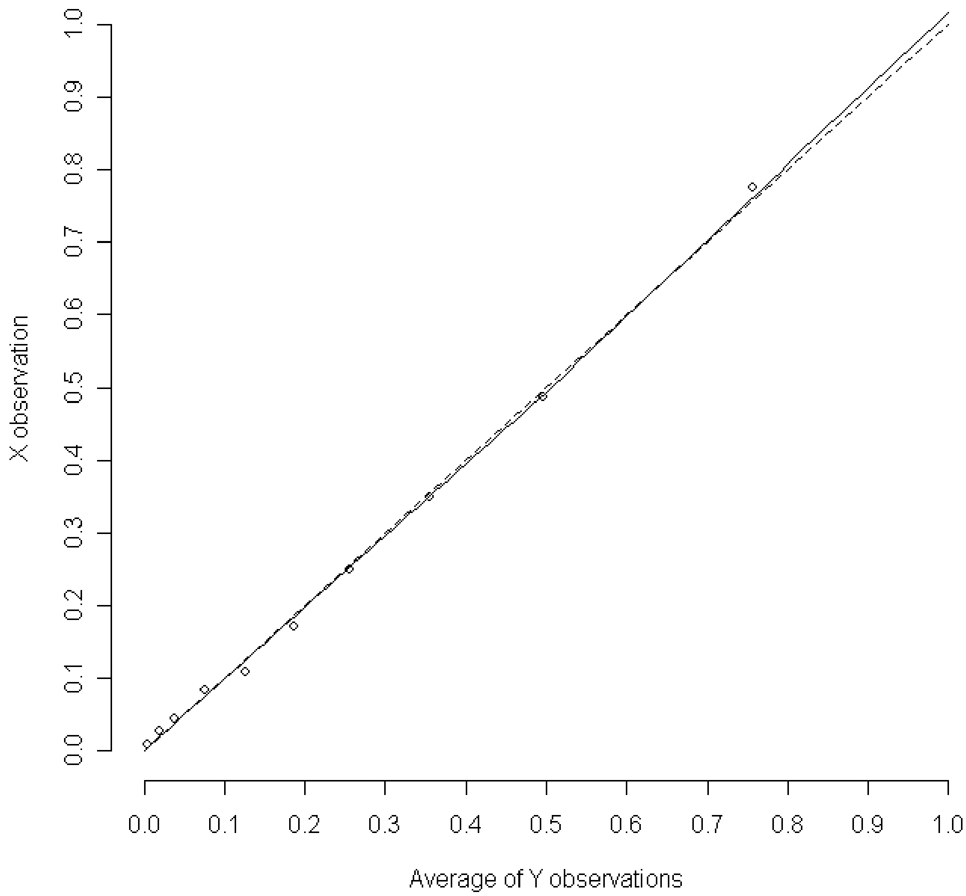

Figure 2. Fitted split line model with binned Z1 and Z2 means. The solid line is the fit of the split line model and the dashed line is a straight line from (0,0) to (1,1).

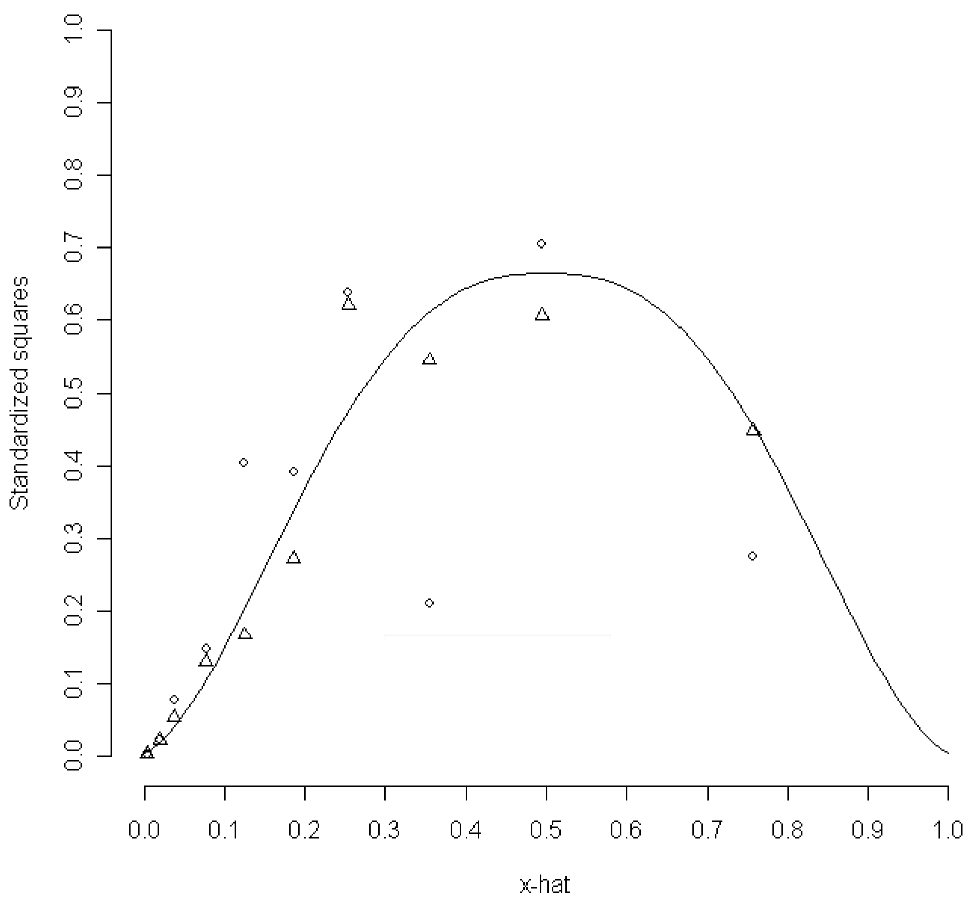

Figure 3. Fitted Variance Function of 2.5 power with binned standardized Z_3^2 and $(Z_1 - Z_2)^2$ means. Z_3^2 means are plotted with circles and $(Z_1 - Z_2)^2$ means are plotted with triangles.