Working Paper.
ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Paris, France, 28-30 April 2014)

Topic (i): Selective editing / macro editing

Adjusting for Remaining Measurement Error after Selective Editing

Prepared by Thomas Laitila and Anders Norberg, Statistics Sweden
(thomas.laitila@scb.se, anders.norberg@scb.se)

I. Introduction

1. A major activity in a survey is the editing process. Developments of new theories and methods useful for reducing the resources spent on editing are of interest as it may provide with substantial cost savings and increase in timeliness. One alley for finding potential improvements is development of more efficient tools for identification of erroneous observations. Another is to reduce the number of observations edited. Indeed the traditional approach to edit all observations is not generally necessary for appropriate statistical inference.

2. One approach towards a more efficient editing process is selective editing, defined as editing methods where only a subset of the response set is selected for editing (Granquist and Kovar, 1997). Here the leading idea is to spend resources only on those observations which will have potential effects on the estimates. For this selective editing is based on the calculations of “global scores” expressing a combined measure of importance in estimation and suspicion of measurement error. These global scores can then be used for ranking of observations in the response set and those observations with the largest scores are selected for editing. Here either a predetermined number of observations are edited or all observations with a score larger than a threshold are edited.

3. Selective editing is largely based on ad hoc methods developed from pure common sense reasoning, and there is yet no accepted theory developed (de Waal, 2014). In particular it does not rest on randomization theory and it is not possible to use traditional statistical methods for generalizing the results from the edited set to the set of non-edited observations and a corresponding part of the population. Thus, estimators used for calculation of estimates on selective edited data sets are biased due to remaining measurement errors. Ilves and Laitila (2009) and Laitila and Ilves (2012) suggest selective editing based on random selection of units for editing and derives, with respect to measurement errors, unbiased estimators of population parameters.

4. This paper contains a proposal for bias correction of estimators due to remaining errors in selective edited data sets. A model based approach is utilized where observed measurement errors in the edited set of observations are related to their observed scores. Using an estimated model, measurement errors in unedited cases are predicted and summarized up to the population level yielding an indication of the level of estimation error due to unedited measurement errors. The paper provides with i) an argument for a model based approach in editing of sample survey data and ii) a methodology for estimation of measurement error bias. Data from Statistics Sweden are used for an illustration of the theory and methodology.

II. Randomness of measurement errors

5. For simplicity consider a real valued scalar study variable y_k, where *k* denotes the *k*:th unit in e.g. the sample. The literature contains several proposals for measurement error models, where the additive measurement error model is perhaps the most frequently applied. This model is formulated as

$$y_k = \mu_k + \xi_k \tag{1}$$

where μ_k is the true value on the study variable and ξ_k is a random measurement error. This model is here later adapted for the analysis of errors in population parameter estimates due to remaining measurement errors.

6. Usually randomness is made an assumption, but here an argument for treating the measurement error as random is given. The argument also implies distributional properties of measurement errors involving heteroskedasticity and dependence both within and between measured units.

7. The outcome of data collection into recorded values into a data matrix is a result of a complex process involving a large number of factors. Some factors are in control while others are not. Some of the uncontrolled factors, probably most of them, are even unknown. Let all the factors, both controllable and uncontrollable, which have effect on the final recorded value be included in the vector $\mathbf{z}_k$. An example of an element in $\mathbf{z}_k$ is an indicator for data collection mode, which is controllable. Another factor is time available for the respondent to complete the interview or questionnaire, which is not controllable, in general. Other examples of uncontrollable factors are events on individual or group levels which may occur during the data collection period. Events taking place affecting the respondent during an interview can have an effect on responses, while events taking place affecting the interviewer may have an effect on the collection of data from several respondents.

8. Each element in $\mathbf{z}_k$ can take on a set of different values whereby the whole vector can take on values in a universal set $\mathbf{z}_k \in \Omega_z$. For a subset of these values $\mathbf{z}_k \in A_z \subset \Omega_z$ it can be assumed that a correct value is recorded if data collection is made under those conditions. Under other conditions, the recorded value is with error. With this division of the set Ω_z the measurement error can be expressed as

$$\xi_k = \begin{cases} 0 & if \quad \mathbf{z}_k \in A_z \\ g(\mathbf{z}_k) & if \quad \mathbf{z}_k \in \overline{A}_z \end{cases}$$

where $\overline{A}_z = \Omega_z - A_z$ and $g(\cdot)$ is some unknown function. Note that $\mathbf{z}_k$ is not assumed random and the measurement error are made a function of the conditions $\mathbf{z}_k$ without adding random components.

9. The model here developed with the vector, $\mathbf{z}_k$, the subset A_z, and the function $g(\cdot)$ is generic in the sense it applies to all kind of variables measured and all units in the population. Such an encompassing can be made to hold by letting the vector $\mathbf{z}_k$ contain elements relating to the study variable and the population unit, respectively.

10. No assumption of randomness has yet been made, and will not necessarily have to be made either. The data collection process is in large a black-box process where some factors is in control or observable while others are neither. Most factors having effect on the processes of collecting and storing data also evolve or change over time, whereby the outcome, i.e. the measurement error, can vary over time if we would imagine repeated independent measurements.

11. Let $T = [0, \tau]$ denote an interval of possible time points for the start of a data collection process aiming at a measurement y_k. Also, let $\mathbf{z}_k(t)$ denote the value of $\mathbf{z}_k$ at time point $t \in T$. It is then possible to define the proportion of the interval *T* for which $\mathbf{z}_k(t)$ is in the set A_z. Let's denote this

proportion as $p_k(A_z)$. Consider now the random experiment of drawing a number *t* from the uniform distribution over *T*. Then $\mathbf{z}_k = \mathbf{z}_k(t)$ is stochastic and the probability of having a correct measurement $y_k = \mu_k$ equals $\Pr(\mathbf{z}_k \in A_z) = p_k(A_z)$. Furthermore, the average of the values $g(\mathbf{z}_k(t))$ conditionally on $\mathbf{z}_k(t) \in \overline{A}_z$ and $t \in T$ defines the conditional expectation $E(g(\mathbf{z}_k) | \mathbf{z}_k \in \overline{A}_z)$ whereby

$$E(\xi_k) = E(g(\mathbf{z}_k) | \mathbf{z}_k \in \overline{A}_z) \Pr(\mathbf{z}_k \in \overline{A}_z) \quad (2)$$

12. Similarly, the variance of a measurement error and the covariance of errors in two different measurements are

$$V(\xi_k) = E(g(\mathbf{z}_k)^2 | \mathbf{z}_k \in \overline{A}_z) \Pr(\mathbf{z}_k \in \overline{A}_z) - E(\xi_k)^2$$

and

$$Cov(\xi_{1k}, \xi_{2k}) = E(g(\mathbf{z}_{1k}) g(\mathbf{z}_{2k}) | \mathbf{z}_{1k} \in \overline{A}_z, \mathbf{z}_{2k} \in \overline{A}_z) \Pr(\mathbf{z}_{1k} \in \overline{A}_z \cap \mathbf{z}_{2k} \in \overline{A}_z) - E(\xi_{1k}) E(\xi_{2k})$$

respectively. These expressions show that measurement errors are heteroskedastic and the measurement errors among measures from one respondent are dependent in general. Also, for two different units, $k \neq l$,

$$Cov(\xi_k, \xi_l) = E(g(\mathbf{z}_k) g(\mathbf{z}_l) | \mathbf{z}_k \in \overline{A}_z, \mathbf{z}_l \in \overline{A}_z) \Pr(\mathbf{z}_k \in \overline{A}_z \cap \mathbf{z}_l \in \overline{A}_z) - E(\xi_k) E(\xi_l)$$

and measurement errors among different units are also dependent.

13. The complexity of the distribution of measurement error indicated by the results may in some cases present difficulties in applying a proper analysis of data. However, the complexity also presents a potential opportunity as information on the measurement error in one variable is contained in the measurements of other variables within the unit as well as in measurements obtained from other units.

14. The source of randomness of measurement errors are above solely due to the assumption of randomly selecting the time for data collection. In practice such a randomization is not applied. However, the process leading to the time of start of data collection in a survey is complex and by itself a function of many factors, whereby treating it as an outcome of a random trial is a reasonable approximation. It is interesting to note, however, that a treatment of measurement errors within a true randomization theory framework can be achieved by randomly selecting the time of data collection. The results above also hold treating the time as nonrandom and assuming elements of $\mathbf{z}_k$ to be outcomes of random trials.

III. Modelling errors using scores

15. Let Γ_k denote a global score, for unit *k*, and be calculated from a set of local scores $s_{kj} = s_{kj}(\gamma_{kj})$. Here $\gamma_{kj} = y_{kj} - \tilde{y}_{kj}$ where $\tilde{y}_{kj}$ denotes an “anticipated” value of the observation y_{kj}. The difference γ_{kj} can be calculated for all units in the response set and, depending on the quality of $\tilde{y}_k$ (subindex *j* dropped for simplicity), it can be useful for estimating errors in unedited values using the result from an analysis of edited values.

16. Let $\mathbf{x}_k = f(\mathbf{z}_k, \gamma_k, \Gamma_k)$ denote a vector valued function defined for the observable part of $\mathbf{z}_k$, γ_k and the global score Γ_k. A model for analysis of measurement errors is formulated as

$$\xi_k = \mathbf{x}_k^t \beta + \varepsilon_k \quad (3)$$

where $E(\xi_k | \mathbf{x}_k) = \mathbf{x}_k^t \beta$, $\varepsilon_k = \xi_k - \mathbf{x}_k^t \beta$ and $V(\varepsilon_k | \mathbf{x}_k) = \sigma_{\varepsilon k}^2$. Equation (2) shows on a complex relation between the expected value of the error ξ_k and the conditions $\mathbf{z}_k$. The function $f(\cdot)$ may therefore contain main effect terms, interaction terms as well as polynomial terms of its arguments.

17. Let $\hat{Y} = \sum_{r(e)} w_k \mu_k + \sum_{r(ue)} w_k y_k$ denote the estimator for given sets of edited *r(e)* and undedited *r(ue)* observations in the response set. Model (3) gives the conditional expectation

$E(\xi_k | \mathbf{x}_k) = \mathbf{x}_k^t \beta$. So, given a sample and a response set, the conditional bias of $\hat{Y}$ due to measurement errors equals

$$Bias(\hat{Y} | r(e), r(ue), \mathbf{x}) = \sum_{r(ue)} w_k \mathbf{x}_k^t \beta \qquad (4)$$

18. An estimator of (4) is obtained by replacing for an estimator $\hat{\beta}$ and, a bias adjusted estimator of the population total is given by

$$\tilde{Y} = \sum_r w_k y_{\bullet k} - \sum_r w_k 1(k \in r(ue)) \mathbf{x}^t \hat{\beta}$$

where $y_{\bullet k} = y_k$ for $k \in r(ue)$ and $y_{\bullet k} = \mu_k$ for $k \in r(e)$. This estimator is unbiased with respect to measurement errors, i.e. $E(\tilde{Y}) = E(\sum_r w_k \mu_k)$, if $E(\hat{\beta} | pq\mathbf{x}) = \beta$, where $pq\mathbf{x}$ denotes conditioning on the sampling design, the response probability function, and the vectors $\mathbf{x}_k$, $k \in r$.

19. Assume the vector $\mathbf{x}_k$ contains variables such that $E(\varepsilon_k \varepsilon_l | \mathbf{x_k}, \mathbf{x_l}) = 0$ when $k \neq l$ and assume $\hat{\beta}$ is unbiased for β. Also, $\mathbf{x}_k$ is assumed to contain the global score used for selection of units to be edited. Then

$$V(\tilde{Y}) = V\left(\sum_r w_k \mu_k\right) + E\left(\sum_r w_k^2 1(k \in r(ue) \sigma_{\varepsilon k}^2\right) + E\left(X_{r(ue)}^t V(\hat{\beta}) X_{r(ue)}\right) \qquad (5)$$

where $V(\hat{\beta})$ is the conditional covariance matrix for $\hat{\beta}$, and $X_{r(ue)} = \sum_r w_k 1(k \in r(ue)) \mathbf{x}_k$. For the first term in the r.h.s. of (5), the unindexed V operator is with respect to the sampling design and the response probability distribution. The expectation operators are with respect to the sampling design, the response probability distribution, and the vectors $\mathbf{x}_k$, $k \in r$.

20. It can be expected that $V(\tilde{Y}) > V(\hat{Y})$ since trading for bias reduction implies a cost in terms of increased variance. However, the first term on the left hand side in (5) is generally smaller than $V(\hat{Y})$ as the error term causing extra variance is removed.

21. An estimator of (5) is not yet developed. With $\tilde{\mu}_k = y_k - 1(k \in r(ue)) \mathbf{x}^{\mathbf{t}} \hat{\beta}$, a potential candidate is

$$\hat{V}(\tilde{Y}) = \hat{V}\left(\sum_r w_k \tilde{\mu}_k\right) + X_{r(ue)}^t \hat{V}(\hat{\beta}) X_{r(ue)}$$

where $\tilde{\mu}_k$ are treated as non-random terms and $\hat{V}(\hat{\beta})$ is an estimator of $V(\hat{\beta})$. Note that the second term on the r.h.s. gives an estimate of the variance of the bias estimator (4), conditionally on $pq\mathbf{x}$. This allows for assessing if the obtained estimate after selective editing is significantly different from the estimate which would be obtained if all units in the response set are edited.

22. One alternative for estimation of β and $V(\hat{\beta})$ is to estimate model (3) with OLS using the edited set of observations *r*(e), and accompany the OLS estimates with a covariance matrix estimator robust against heteroskedasticity. Since the variance of $\hat{\beta}$ affects the variance of the bias adjusted estimator, it may be preferable to use a more efficient estimator of β. One such estimator is the feasible WLS estimator, weighting observations with an estimate of disturbance variance.

23. The model (3) does not account for the dependence among survey question responses within units. This could be done by extending model (3) into a simultaneous system of equations model. Such a system could be estimated using the scores γ_k as instrument variables. Still, an equation of the form (4) would be used for estimation of conditional bias as the errors are not observed for the unedited set.

24. Note that the selection of units for editing through use of global scores does not cause a selection problem of the kind studied by Heckman (1979). Since the global score is one of the elements in $\mathbf{x}_k$, the conditional mean of $\varepsilon_k = \xi_k - \mathbf{x}_k^t \beta$ is zero among units selected for editing.

IV. Numerical example

25. The data set used is from a stratified SRS sample survey on establishments and their salary payments to employees. Here, this data set is merely used for illustration of the theory and the methodology suggested above. The illustration is not an evaluation of the selective editing methodology in the specific application since it would require a much more rigorous treatment.

26. For each establishment a number of variables are measured for each employee, and one of the variables measured is monthly payment. An establishment will be named a unit and data for one employee is named an observation. So here the observation y_{kj} denotes monthly payment to employee *j* at unit *k*. Using data from earlier waves of the study an anticipated value $\tilde{y}_{kj}$ and the difference $\gamma_{kj} = y_{kj} - \tilde{y}_{kj}$ is calculated. Similar calculations are made for the other variables measured and a local score per observation is calculated. The local scores are then aggregated over observations into a global score for the units.

27. In this illustration, the anticipated values for monthly salary are recalculated and may differ from those used in the calculation of the local scores. In case of no earlier observations, means of earlier observations over categories are used as anticipated values. First cross tabulation over four categorical variables is used. These variables describe kind of profession, staff category, salary category, and gender. For observations without an anticipated value after the first cross tabulation, a second cross tabulation is used over three categorical variables. If still missing, cross tabulation is made over two variables.

28. The total number of observations in the data set is 158 474. Some of these are employees classified as not belonging to the population; this is not a problem for our purposes and the observations are retained in the data set. An observation is edited (all measured variables) if *1)* the global score for the unit exceeds a threshold and *2)* the local score exceeds a threshold. Some observations are edited due to other reasons than potential errors in measured variables. To guarantee having these observations selected for editing, 1 000 000 was added to the local scores. In total there are 29 239 selected for editing. Among the edited observations around 6 000 have missing values on the variable studied, and is here treated as missing values. These observations are here classified as nonresponses, yielding an edited set of 23 353 observations for the analysis.

29. Figure 1 presents a plot of the "v-scores" $\gamma_{kj} = y_{kj} - \tilde{y}_{kj}$ vs. the global scores gs_k for observations in the unedited set with gs_k <1 000 000. The observed range of the v-scores is within -75 000 to 130 000. Some observations in the unedited set are associated with large global scores, meaning that although the global score passes the threshold for selection, the local scores do not. Some of these observations have large v-scores which in turn imply a low importance of the observation. A similar plot (not included) for unedited observations with gs_k ≥1 000 000 shows a similar range of values of v-scores.

30. A corresponding plot of v-scores vs. gs_k for the edited set of observations is shown in Figure 2, where observations are restricted to those with gs_k <1 000 000. The observed range of v-scores and global scores are similar to those observed for the unedited set. Again the plot obtained for observations with gs_k ≥1 000 000 shows a similar range of v-scores (plot not included).

31. Measurement errors in the edit set are in Figure 3 and Figure 4 plotted against the v-scores. In Figure 3, observations with global scores less than 1 000 000 are included, and in Figure 4 those with scores of at least 1 million are included. In both figures, the range of the v-scores are restricted to -75 000

– 130 000. For nonzero measurement errors, both figures show on a positive linear relationship between measurement error and the v-score. The level and slope seem to be the same in both figures. A large fraction of observations selected for editing are without errors, however, emerging as horizontal lines through the scatters of positively related points.

32. The regression vector in (3) is formulated with the v-score $\gamma_{kj} = y_{kj} - \tilde{y}_{kj}$ and its square γ_{kj}^2. Let $SC2_k$ denote a dummy variable for a global score $gs_k \geq$1 000 000. To allow for different relationships between measurement error and v-score for observations with large and small global scores, respectively, the variables $\gamma SC2_{kj} = \gamma_{kj} \cdot SC2_k$ and $\gamma^2 SC2_{kj} = \gamma_{kj}^2 \cdot SC2_k$ are also included in the regression vector, as is the dummy variable $SC2_k$. The global score is in the regression allowed to have effect only if the global score is less than 1 million. This is implemented with the variable $gsm_k = (1 - SC2_k) gs_k$, which is introduced in the regression vector with its square gsm_k^2. Finally, an interaction term between local and global scores is introduced with $\gamma gsm_{kj} = \gamma_{kj} \cdot gsm_k$.

33. The resulting regression model is estimated with feasible WLS using an estimated scedastic function where the log of the residual variance is specified as a linear function of the same variables as above. WLS estimates suggest there is no difference in the relationship between measurement errors and the v-scores for large and small global scores, respectively. The dummy variable for large global scores, $SC2_k$, is also insignificant. The final model estimated is reported in Table 1 together with the estimated scedastic function.

34. By using the estimates in Table 1, the conditional bias (4) is calculated for 77 domains. These bias estimates are presented in Figure 5 where they are plotted against the proportion of edited observations in the domain response sets. Out of the 77 bias estimates, 71 are significant at the 5% significance level using approximate z-score statistics. The figure shows one domain with a particularly large bias estimate, around -40 million SEK. Some domains have bias estimates between -10 and -5 million SEK. Remaining domains have small bias estimates in level terms. The figure also shows on a tendency to negative bias estimates.

35. Relative absolute biases are depicted in Figure 6, where they are plotted against the proportion of edited observations in the domain response set. Here relative bias is defined in relation to the point estimate obtained after selective editing, i.e. $\hat{Y} = \sum_{r(e)} w_k \mu_k + \sum_{r(ue)} w_k y_k$. The plot shows on a small relative bias for most domains, relative biases are less than 5%. For a few domains, relative biases are larger and two estimates are more than 8%. There is a tendency in the figure showing on a decreasing relative bias with the proportion of edited observations, which is to be expected.

IV. Discussion

36. Unattended measurement errors in a data set yields biased estimates unless strong assumptions on the characteristics of the errors are satisfied. The idea of selecting a subset of a sample for editing does leave a large set of observations with potential unattended measurement errors and, this leaves a potential estimation error bias which is of unknown size.

37. This paper suggests a modeling approach where measurement errors detected in the edited data set is related to characteristics of the observations in a regression model. Based on the estimated model mean measurement errors are estimated for unedited observations and summed up to the level of point estimates, which gives an estimate of the estimator bias due to remaining measurement errors.

38. There are several ways in which such bias estimates can be utilized. One alternative is to adjust the estimator by subtracting the bias estimator. This will reduce bias in the estimator but will in most cases also increase its variance. Another option is to use the bias estimates similar to macro editing,

identifying estimates with potentially large errors. Another possible use is the evaluation of pseudo-bias estimates in the process of calibrating algorithms for selective editing.

39. The use of global scores in selection of observations for editing is an important aspect for the modeling approach suggested here. First, it is necessary to include the global scores among the regressor variables. If excluded model estimates are biased and inconsistent due to a sample selection effect. Secondly, local scores are not perfectly correlated with global scores. This provides with local scores with a range including unedited observations. Thus, the problem of having to use extrapolation for estimation of mean measurement error may not be a major issue. This also suggests replacing the global score with all local scores used in its definition.

40. Finally, although the theory presented in Section 2 is not of a direct practical importance, it is essential for the theoretical motivation behind the suggested modeling approach. The idea of regressing measurement errors on scores and other variables is straightforward. Now, applying such a modeling approach implicitly assumes the measurement errors as outcomes of random trials. Such an assumption must be followed by a strong argument, especially when it comes to the field of official statistics. The theory in Section 2 presents such an argument and is the foundation for the modeling approach proposed in this paper.

References

De Waal, T. (2014). Selective Editing: A Quest for Efficiency and Data Quality, *Journal of Official Statistics*, 29:4, 473-488.

Granquist, L. and J.G. Kovar, (1997). Editing of Survey Data: How Much is Enough?, In: *Survey Measurement and Process Quality*, L.E. Lyberg, P. Biemer, M. Collins, E.D. de Leeuw, C. Dippo, N. Schwarz and D. Trevin (eds.), Wiley, Hoboken, NJ, pp. 416-435.

Heckman, J.J. (1979). Sample Selection Bias as a Specification Error, *Econometrica* 47:1, 153-161.

Ilves, M. and T. Laitila (2009. Probability-Sampling Approach to Editing, Austrian Journal of Statistics, 38:3, 171-182.

Laitila, T. and M. Ilves (2012). Probability Editing, Paper prepared for the UNECE Work Session on Statistical Data Editing, Oslo, Norway, 24-26 September, 2012.

Table 1: WLS estimates of measurement error model (3) and OLS estimates of function for $\sigma^2_{\varepsilon k}$.

	Measurement error model (3)		$\log \sigma^2_{\varepsilon k} = \mathbf{x}_k^t \alpha$	
Variable	Estimate	St.err	Estimate	St.err
Constant	22.6	47.1	16.2	0.035
γ_{kj}	0.498	0.004	-0.189×10^{-4}	0.016×10^{-4}
γ^2_{kj} ($\times10^{-5}$)	0.146	0.018	0.904×10^{-4}	0.0245×10^{-4}
$SC2_k$	-	-	-0.804	0.041
$\gamma SC2_{kj}$	-		-0.249×10^{-4}	0.020×10^{-4}
$\gamma^2 SC2_{kj}$	-		0.864×10^{-9}	0.036×10^{-9}
gsm_k	20.1	4.59	0.0018	0.0016
gsm^2_k	-0.022	0.007	-0.106×10^{-4}	0.027×10^{-4}
γgsm_{kj} ($\times10^{-3}$)	0.993	0.028	-0.142×10^{-3}	0.039×10^{-3}

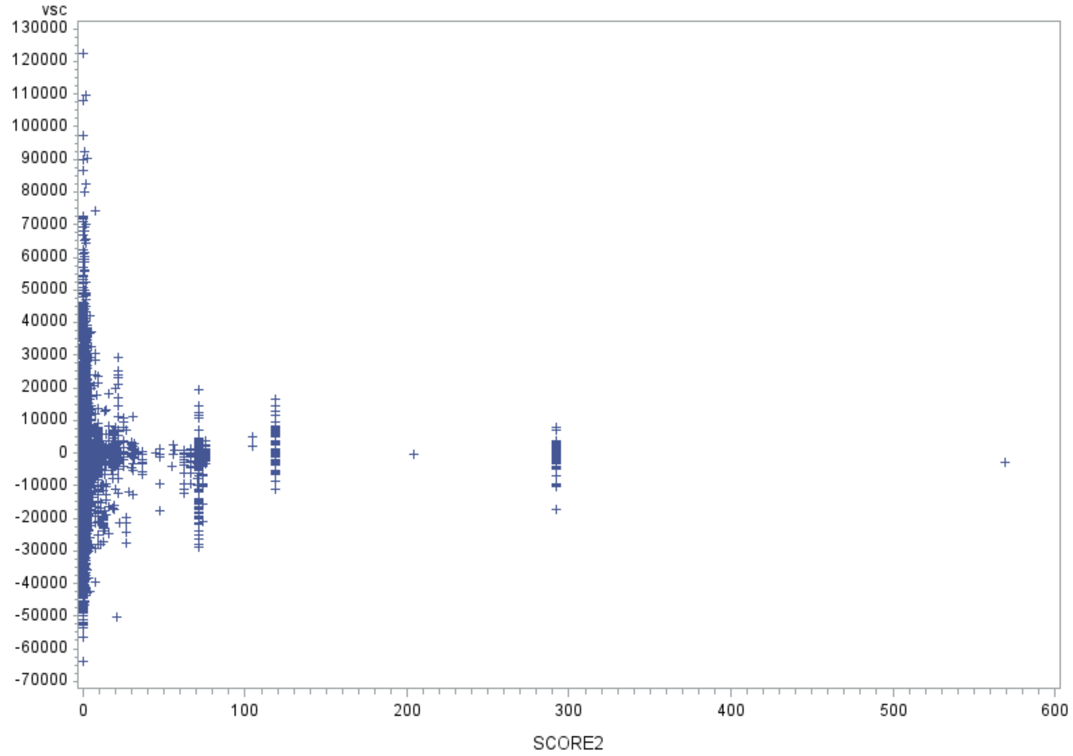

Figure 1: Plot of v-scores $\gamma_{kj} = y_{kj} - \tilde{y}_{kj}$ (vsc) vs. global scores gs_k (score2) for unedited set with score2<1 000 000.

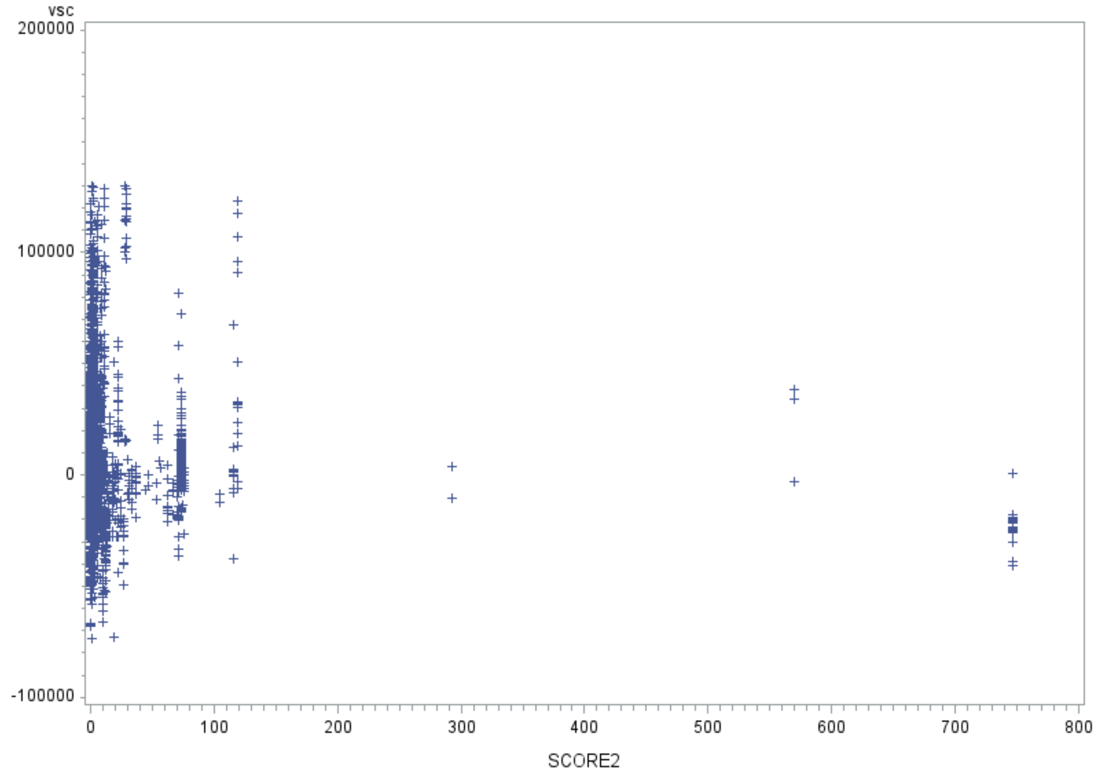

Figure 2: Plot of v-scores $\gamma_{kj} = y_{kj} - \tilde{y}_{kj}$ (vsc) vs. global scores gs_k (score2) for edited set with score2<1 000 000.

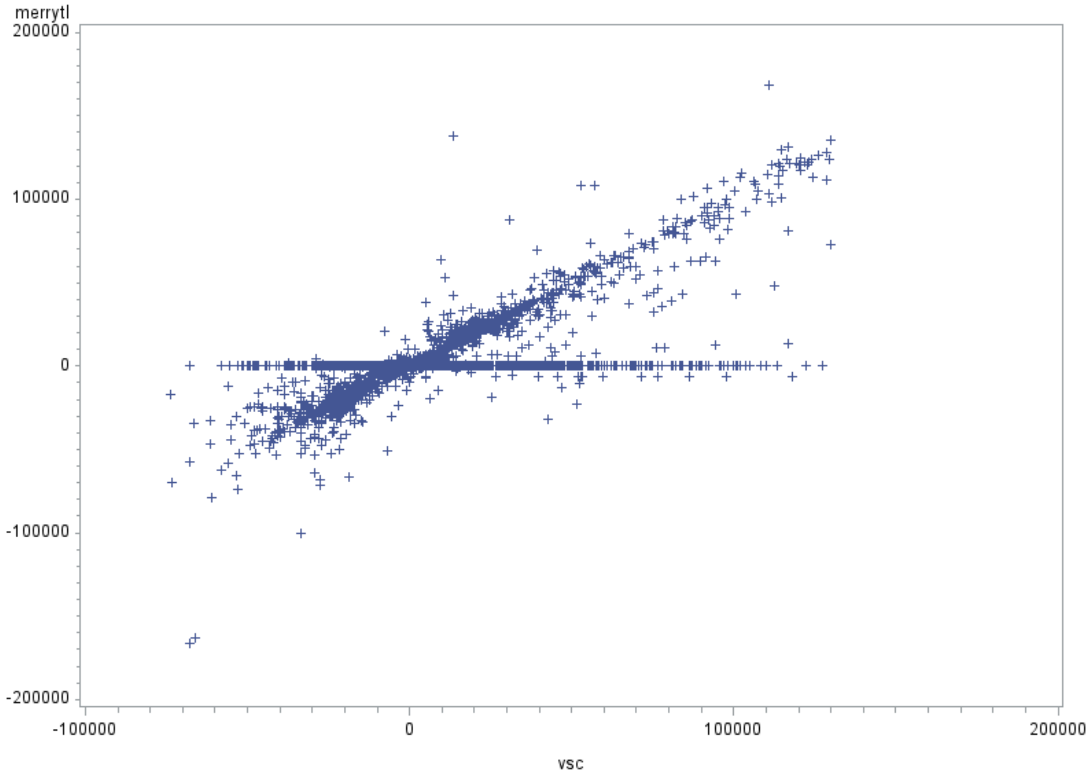

Figure 3: Plot of measurement errors (merrytl) vs. v-scores $\gamma_{kj} = y_{kj} - \tilde{y}_{kj}$ (vsc) in edited set (observations with gs_k < 1 000 000 and -75 000 < vsc < 130 000 included).

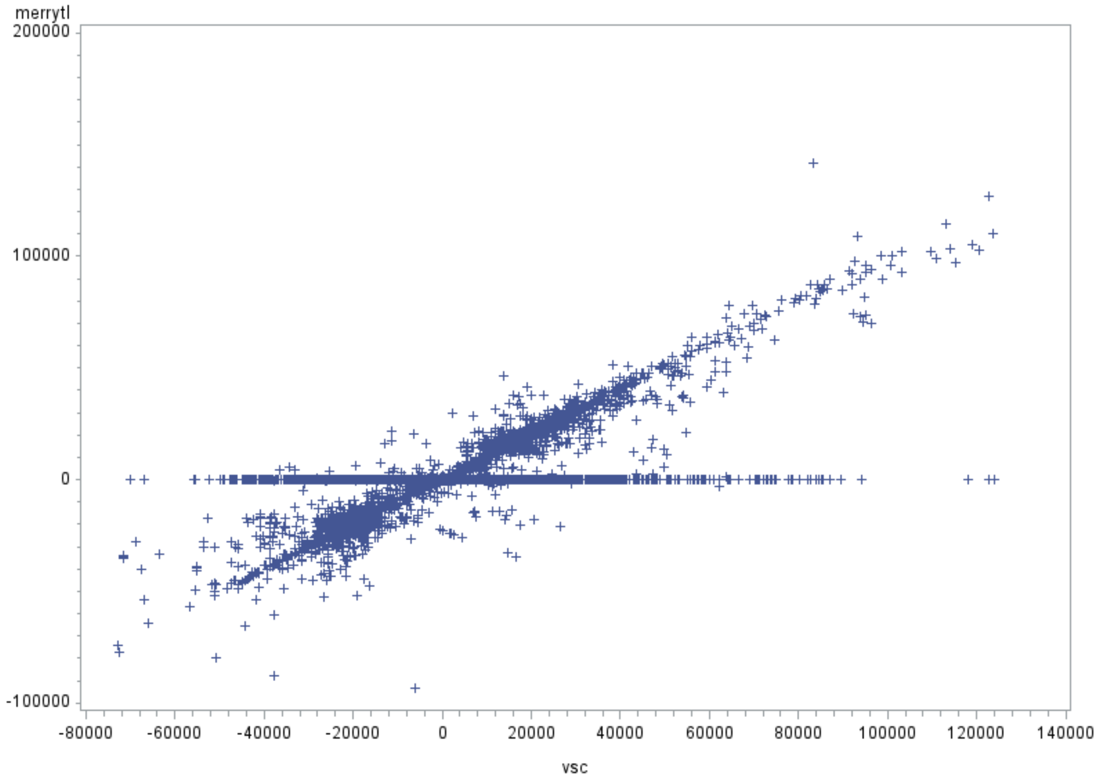

Figure 4: Plot of measurement errors (merrytl) vs. v-scores $\gamma_{kj} = y_{kj} - \tilde{y}_{kj}$ (vsc) in edited set (observations with $gs_k \geq$ 1 000 000 and -75 000 < vsc < 130 000 included).

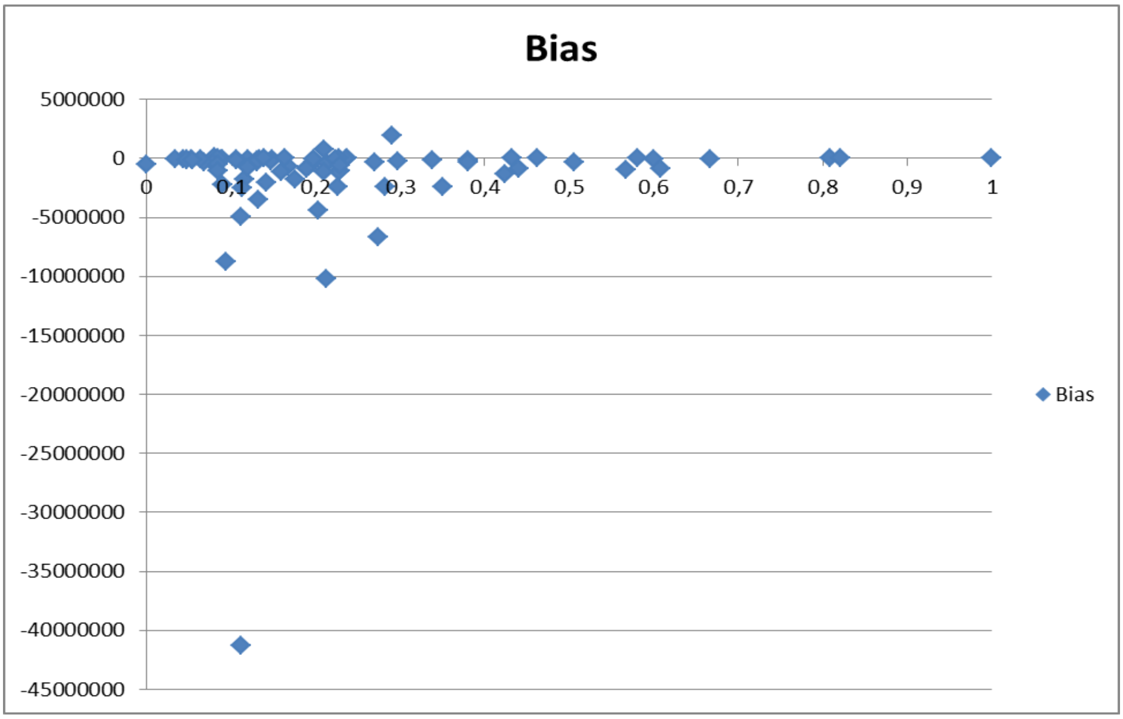

Figure 5: Plot of estimated bias (SEK) due to remaining measurement errors after SE, plot vs. proportion of edited observations in domain. (Proportion w.r.t. number of obs. in response set.)

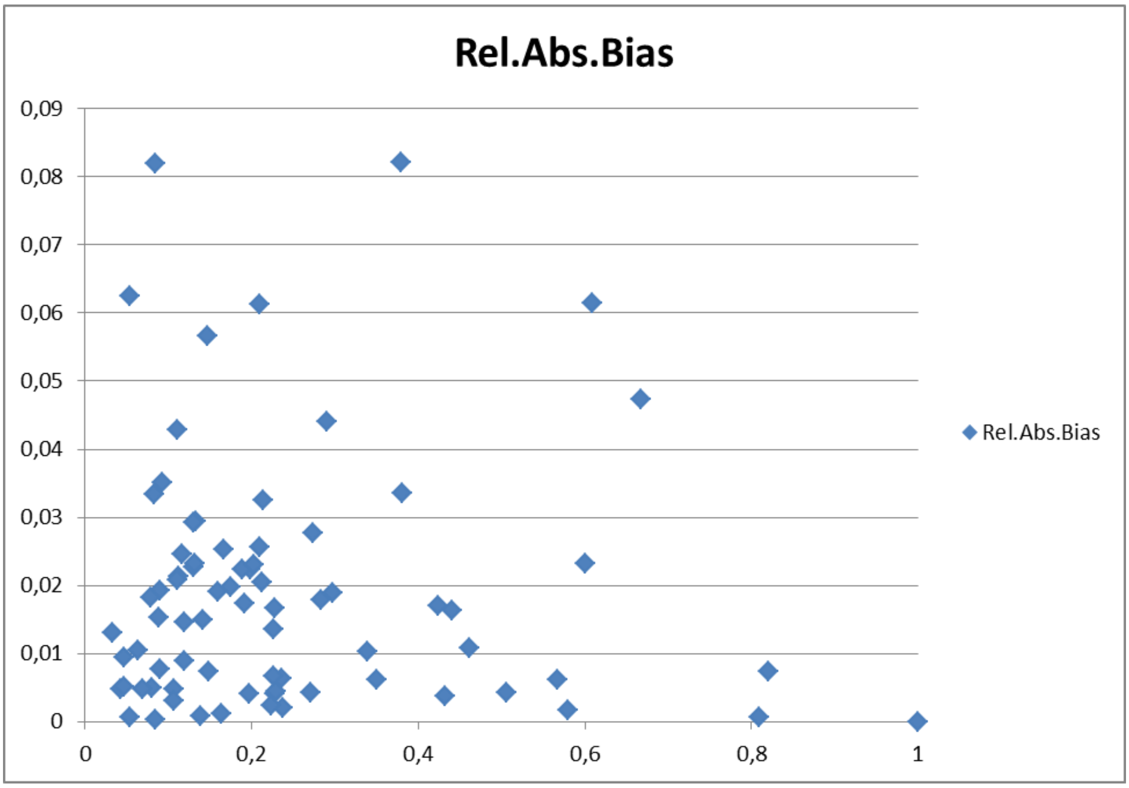

Figure 6: Plot of estimated relative absolute bias of domain estimates vs. proportion of edited observations. (Proportion w.r.t. number of obs. in response set.)