Model and Survey Performance Measurement by the RSE and RSESP

James R. Knaub, Jr.
Energy Information Administration, EI-53, Forrestal Bldg., Washington DC, 20585

Abstract:
The model-based relative standard error for a superpopulation, the model-based RSESP, found in Knaub (2002, 2003, and 2004), may be used to compare competing regression models for the same data sets, when regressor data are available. This tool may then be used to track survey performance to a degree, indicating some important changes in total survey error. Analyses involving both the RSESP and the usual relative standard error, RSE, may be informative. Here a series of graphical results are presented and interpreted.

Key Words:
regression; total survey error; performance measure; establishment survey; classical ratio estimator; cutoff sampling

Introduction:

This paper illustrates the model-based relative standard error for a superpopulation as described as a measure for comparing alternate linear regression models, primarily in Knaub (2003), and as an indicator of important aspects of total survey error, primarily described in Knaub (2004). The simple random sample analogy would be to estimate variance for a survey, and remove the finite population correction factor. For model-based variance, when calculating these estimates, we simply sum over all N cases, rather than just the N-n unobserved cases. Normally a relative standard error, RSE, is designed to indicate the error that may result as a consequence of estimating for data not collected in a sample when approximating the population. It is impacted by nonsampling error, but only to the degree that it impacts the part of the population not sampled. For a census, the RSE is always zero, no matter how much nonsampling error is present. These RSE estimates are not designed to measure total survey error. The RSESP estimates are impacted by nonsampling more evenly, even for a census. The models being used also impact the RSESP estimates discussed in this paper. Consequently, it is an excellent measure for comparing model performances. As an indicator of total survey error, however, it is weaker, as a high estimated RSESP may not necessarily mean low accuracy. When discussing this measure with other statisticians, John Vetter, formerly with the Energy Information Administration's Survey Methods Group (EIA-SMG), suggested that the best use in that area appeared to be in comparing these estimates from one time period to another. Past cases of suddenly increasing RSE estimates had keyed data investigations within the Energy Information Administration's Electric Power Division (EIA-EPD), and this looked like a good avenue to explore, as RSESP estimates should perform even better at such a task. Thus the graphs below are for monthly survey results.

Note also that RSEs have been of interest to EIA, as in Waugh, Norman and Knaub (2003). Total survey error is a key concept. It can be found in Federal Committee on Statistical Methodology (2001), and is a key concept behind the Encyclopedia of Survey Research Methods (2008).

Another reference of interest is the Model Quality Report (1999). Knaub (2004) page 2 states that the design-based equivalent of the RSESP, for simple random sampling, would be found by removing the finite population correction factor (fpc), a standard term. On page 113 of Model Quality Report (1999) they describe what happens if one were to "…remove finite population corrections from the variance estimator…" when "assessing the variance impact of measurement error." This attempts to account for measurement error, but overestimates it in both the design-based and model-based cases. In the model-based case, different factors that impact the RSESP, one of which is measurement error, are also indicated in Knaub (2004), page 1.

Nearly two decades ago, the EIA-EPD started including a definition of sampling error measurement (EPD called cvs, now RSEs) that was included in the EPM Tech Notes to help customers understand what these variance derived numbers meant. It was explained that these variances are impacted by nonsampling error (such as measurement error). This applied well to a monthly retail electric sales sample, the Form EIA-826, and paying attention to these

RSE estimates helped the EIA-EPD 'clean' data of excess nonsampling error.

As a census is approached, the impact of nonsampling error is diluted in the usual RSE estimates. It is best to know more about the variance due to measurement error. Note that on page 107, Model Quality Report (1999), they have a note on "Variance inflation" in their section on measurement error. This note confirms this author's interpretation of variance as a measure that can be impacted by nonsampling error (such as measurement error). There Model Quality Report (1999) also says that "The variance inflating impact of measurement error is likely to be most important for the largest businesses in the completely enumerated strata. Such businesses do not contribute at all to sampling variance, but random errors in their reporting values may have a significant impact on the total variance of the survey estimates. …" They go on further describing the overall variance, part from sampling error and part from nonsampling error.

In a census (complete enumeration), the sampling variance in the RSE completely ignores the impact of nonsampling error, and is thus 0. That is why one should look at RSESP estimates as well.

In practice EPD has used the RSESP to compare model performances and now as an aid in determining sample sizes. Consider electric revenue and sales data, in the public transportation sector reported as a monthly cutoff sample on Form EIA-826 (and on an annual census as well), or the commercial plant wood-fired generation data from the Form EIA-906/920. In each of those two cases a census will bring the RSE estimates to zero, but in each of these cases, we would still be left with a large estimated RSESP. This tells us that either we have a hard time modeling those data, or there may be a lot of nonsampling error, or both. This leads us to understand that we must pay special attention to nonsampling error in those cases, and that it would be advisable to look for a better model, in case of nonresponse. That is, the RSESP estimates tell us that those are two cases in particular that need attention.

Both Knaub (2004) and Model Quality Report (1999) therefore point out that such a treatment of variance can lead one to overstate measurement error. On the other hand, Nancy Kirkendall and Janice Lent pointed out that systematic error was not included in this focus.

Graphical Examples of the Use of the model-based RSESP –
Comparing Models:

First consider the case of two models (a single regressor model labeled "MOD," and a multiple regression model labeled "MR" – both ratio models as in Knaub (1999), Knaub (2003), and Knaub (2005)) for electric gross generation from wind power from Independent Power Producers (IPPs), in the Pacific Contiguous [Bureau of the] Census Division (California, Oregon, and Washington), as shown in Figures 1, 2, and 3.

Figure 1 – RSE Estimates

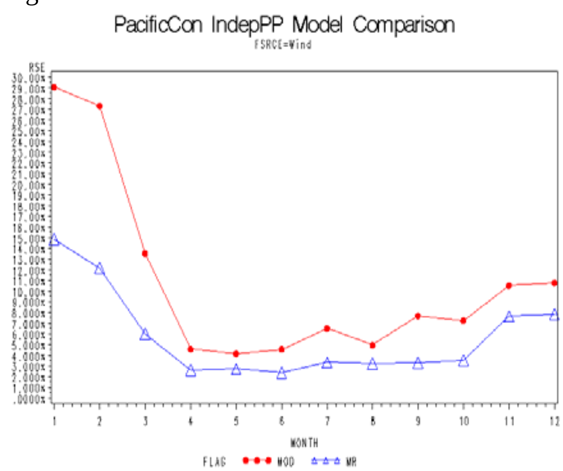


Figure 2 – RSESP Estimates

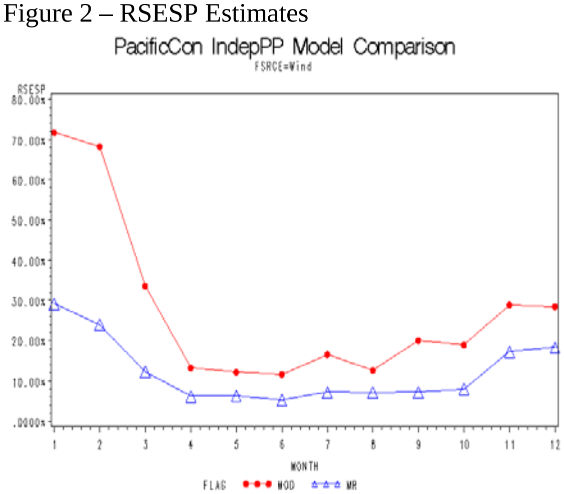
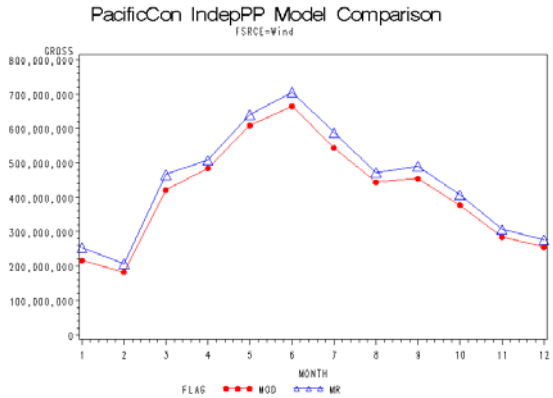


Figure 3 – Estimates of Gross Generation



Note that the estimated RSEs and estimated RSESPs, Figures 1 and 2, follow generally the same pattern. That often occurs. When data are gathered in different groups for purposes of modeling and estimation than they are for purposes of publication (Knaub (1999)), then this may not be true, especially if there is a censused group/stratum involved. Such was the case with Figure 4 (which is also found in Knaub (2007)). There we have a case of investor-owned utilities (IOUs) that are nearly censused for sales of electricity and associated revenue. A peak in nonsampling error among IOUs could increase RSESP estimates without a corresponding increase in RSE estimates.

Notice that RSE estimates in Figure 1 vary between models used, from values of approximately 1 to 14 percent. The estimates of gross generation also vary between models, but do they vary by a reasonable amount? Estimated RSESPs may be used to compare model performances, but the estimated gross generations here would only vary because of differences in RSE estimates. The same nonsampling error would be present in both gross generation estimates. To see if Figure 3 appears reasonable, consider the following where estimated totals and their estimated standard errors are considered for two models, "A" and "B":

Compatibility of Graphs Comparing Gross Generation Estimates from Two Different Models to Graphs Comparing Their Estimated RSEs:

$$\frac{\hat{T}_A - \hat{T}_B}{\hat{T}_A} = u\left[\frac{\hat{\sigma}_{TA}}{\hat{T}_A} - \frac{\hat{\sigma}_{TB}}{\hat{T}_B}\right]$$

roughly translates as $\frac{\Delta T}{T} = u\,\Delta RSE$ or, approximately, $\frac{dT}{d\sigma_T} = u$. So, if $T = \hat{T} + z\,\hat{\sigma}_T$, then $\frac{dT}{d\sigma_T} = z$, and in that case *u* = *z*, as indicated above. (If *u* is N(0,1) then *u* is expected to fall between -1 and 1 with probability approximately 0.68.)

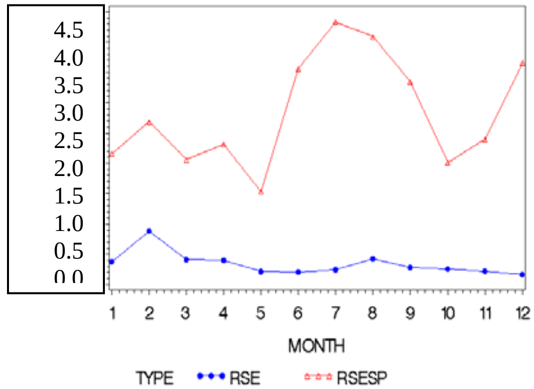
Note that $\frac{\hat{\sigma}_{TB}}{\hat{T}_B}$ is not the same as $\frac{\hat{\sigma}_{TB}}{\hat{T}_A}$, necessary in making $\frac{dT}{d\sigma_T} = u$, and therefore, as $\left|\hat{T}_A - \hat{T}_B\right|$ becomes larger, the above holds less well. Thus, the probability that *u* falls between -1 and 1 becomes smaller as variance increases. If *u* once again can be assumed to be between -1 and 1 in about half of all instances, then we still can assume that $\left|\frac{\hat{T}_A - \hat{T}_B}{\hat{T}_A}\right|$ is greater than $\left|\frac{\hat{\sigma}_{TA}}{\hat{T}_A} - \frac{\hat{\sigma}_{TB}}{\hat{T}_B}\right|$ in about half of all instances, and *vice versa*. There should not be too many cases where $\left|\dfrac{\hat{T}_A - \hat{T}_B}{\hat{T}_A}\right|$ is much smaller than $\left|\dfrac{\hat{\sigma}_{TA}}{\hat{T}_A} - \dfrac{\hat{\sigma}_{TB}}{\hat{T}_B}\right|$.

Thus Figure 3, in comparison with Figure 1, may not be unreasonable, although the large change in the difference of RSE estimates between early and later months is not reflected. (Other data examined had shown smaller differences in gross generation estimates than might be expected, which prompted this 'reality check,' but may be explained by the removal of certain data from consideration, brought to the author's attention by Joel Douglas (EIA-EPD, formerly SAIC).)

Graphical Examples of the Use of the model-based RSESP –
Tracking Variance:

In Figure 4, for electric sales, estimated RSEs and estimated RSESPs are compared on the same graph for one model. If there is a problem with IOUs as described above, then scatterplots of data used in regressions for those points that are impacted should help resolve this. A future step will be to have programming continued to allow us to 'click' on points on a graph such as in these figures, which would take us to appropriate scatterplots, and further, to 'click' on points on those scatterplots that will take us to Respondent Contact Reports (RCRs) that will tell us if information is already available regarding data that appear anomalous.

Figure 4 - For New England Residential Electricity Sales Estimates in 2005



In Figure 5, results separately estimated by plant types, fuel types, and geographic regions are shown for grand total gross generation, by month, from a multiple regression, ratio-type estimator. Gross generation estimates, shown in red 'dots,' show summer increases in both 2005 and 2006. (Yearly sources for regressor data were upgraded, but nearly the same results occurred when the regressor data were kept the same throughout.) The RSESP estimates show some oddly high months in 2005, and the RSE estimates do this in 2006. This is very odd and means that further investigation is warranted. It is begun here, and shown in succeeding figures. ("USAllPP Sector" is used to designate that all plant type groups were accumulated, and "FSRCE"designates "fuel sources.")

Figure 5

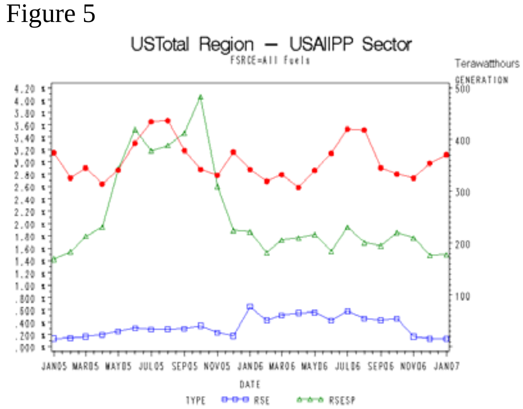


Figure 6 shows us that the RSESP anomalies of Figure 5 may not be explained by looking further at gross generation from coal, but the odd RSE estimates may be investigated further for this fuel-type.

Figure 6

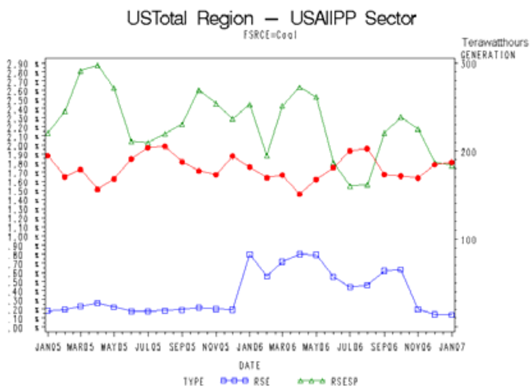


Figure 7 shows that natural-gas fired electric generation may be the place to look for an explanation of the oddly high RSESP estimates in parts of 2005.

Figure 7

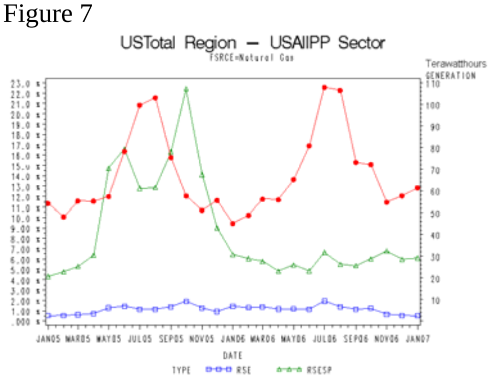
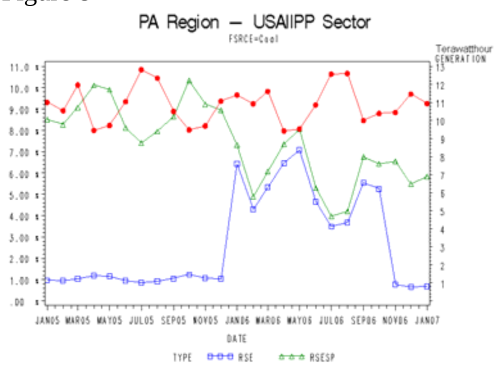


Figure 8 is an example of one of several individual States' estimates for coal-fired electric gross generation that show a particular problem for January 2006 to October 2006 results. The RSE estimates are drastically larger there, and the RSESP estimates are generally somewhat smaller, so that they are nearly equal. When RSE estimates and RSESP estimates are the same, then there have to be no observed data as part of the estimated (sub)total. All of it would have to be from imputed/predicted numbers from the generally broader groups used for model/estimation purposes. The author was told that data were revised for January through October 2006. Perhaps they were mistakenly designated as "add-ons" (only representing themselves – see Knaub (2002)) in a number of cases, or perhaps there was some other data processing error or other nonsampling error. At any rate, a few key States seem to have been impacted. Pennsylvania was one of them, as shown in Figure 8.

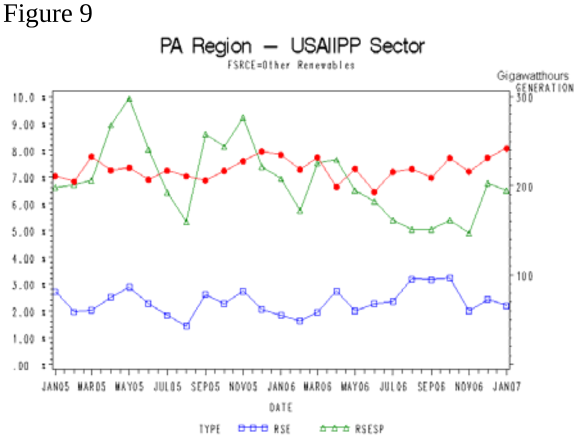
Figure 8



.

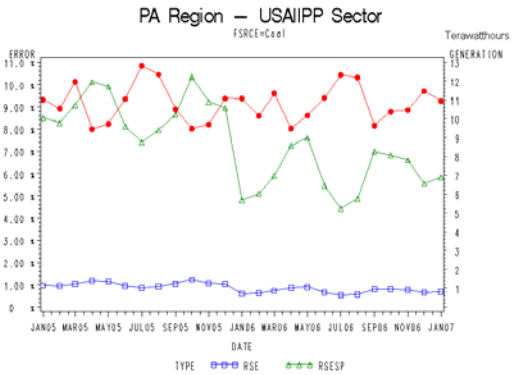
In Figure 9, also for Pennsylvania, but for electric generation from "Other Renewable" sources ('other' than hydroelectric), conditions appear more nearly normal, but the RSE estimates and RSESP estimates do approach each other in June, July and August of 2006, and stay close through October 2006. (Note that generation is in gigawatthours.)

Figure 9



Some considerable time later, after accomplishing substantial data editing, and a data processing change was made, Figure 8 was substantially changed. (See Figure 10 below.) However, Figure 9 did not appear to change at all.

Figure 10



Conclusions:

Model Comparisons**:** In a number of cases, graphs of RSESP estimates have shown which among competing models works best for a given data set. Usually RSE estimate graphs would arrive at the same results, but not always. This could be because of differently defined subsets of data to which each model is applied.

Tracking Variance**:** Possibly when there are a few observed data points with relatively large standard errors of the individual prediction errors (STDI in SAS PROC REG – Maddala (1992) refers to the "variance of the prediction error," which would be the square of the STDI), those contributions to standard error for the RSE will be diluted more in some samples than others. The RSESP is better suited to tracking such variances.

Acknowledgments:
Thanks to John Vetter, Joel Douglas, Nancy Kirkendall, and others for helpful conversations, and to Joel Douglas for graphical support. Anything lacking is on my account.

References:

Encyclopedia of Survey Research Methods (2008), Editor: Paul J. Lavrakas, Sage Publications, to appear in February 2008. (Cutoff sampling entry by J.R. Knaub, Jr.)

Federal Committee on Statistical Methodology (2001). Measuring and Reporting Sources of Errors in Surveys. Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 31). http://www.fcsm.gov/01papers/SPWP31_final.pdf

Knaub, J.R., Jr. (1999), "Using Prediction-Oriented Software for Survey Estimation," InterStat, August 1999, http://interstat.statjournals.net/, partially covered in "Using Prediction-Oriented Software for Model-Based and Small Area Estimation," in ASA Survey Research Methods Section proceedings, 1999, and partially covered in "Using Prediction-Oriented Software for Estimation in the Presence of Nonresponse," presented at the International Conference on Survey Nonresponse, 1999.

Knaub, J.R., Jr. (2002), "Practical Methods for Electric Power Survey Data," InterStat, July 2002, http://interstat.statjournals.net/. (Note another version in ASA Survey Research Methods Section proceedings, 2002.)

Knaub, J.R., Jr. (2003), "Applied Multiple Regression for Surveys with Regressors of Changing Relevance: Fuel Switching by Electric Power Producers," InterStat, May 2003, http://interstat.statjournals.net/. (Note another version in ASA Survey Research Methods Section proceedings, 2003.)

Knaub, J.R., Jr. (2004), "Modeling Superpopulation Variance: Its Relationship to Total Survey Error," InterStat, August 2004, http://interstat.statjournals.net/. (Note another version in ASA Survey Research Methods Section proceedings, 2004.)

Knaub, J.R., Jr. (2005), "Classical Ratio Estimator," InterStat, October 2005, http://interstat.statjournals.net/.

Knaub, J.R., Jr. (2007), "Cutoff Sampling and Inference," InterStat, April 2007, http://interstat.statjournals.net/.

Maddala, G.S. (1992), *Introduction to Econometrics*, 2nd ed., Macmillan Pub. Co.

Model Quality Report (1999), Model Quality Report in Business Statistics, Volume I, Theory and Methods for Quality Evaluation, and Volume IV, Guidelines for Implementation of Model Quality Reports, General Editors: Pam Davies, Paul Smith at http://www.ams.ucsc.edu/~draper/bergdahl-etal-1999-v1.pdf

Waugh, S., Norman, K. and Knaub, J. (2003) "Proposed EIA Guidance on Relative Standard Errors (RSEs)," Presentation to the ASA Committee on Energy Statistics, October 17, 2003, http://www.eia.doe.gov/smg/asa_meeting_2003/fall/files/rseguidance.pdf

Appendix: Notes: When RSESP Estimates are Close to RSE Estimates

I. Calculation of RSE estimates:

Estimate the variance of the data in each estimation group by the relationship of observed data to regressor data – excludes add-ons.

Estimate variance of totals in each publication group by summing information over the N-n (minus add-ons) cases in each publication group (in separate parts corresponding to each estimation group – usually only one – that is at least partially contained in the publication group).

Thus, in each publication group, for RSEs, we sum over N-n-(add-ons) cases.

II. Calculation of RSESP estimates:

Estimate the variance of the data in each estimation group by the relationship of observed data to regressor data – excludes add-ons.

Estimate variance of totals in each publication group by summing information over the N (minus add-ons) cases in each publication group (in separate parts corresponding to each estimation group – usually only one – that is at least partially contained in the publication group).

Thus, in each publication group, for RSESPs, we sum over N-(add-ons) cases.

III. In each publication group, the same data from the N-n non-add-ons that correspond to data not collected in the sample would be used in both I and II. Thus the only way for RSE estimates and RSESP estimates for the same publication group to be equal would be for the n sample observations (besides add-ons) to contribute nothing to either the RSE estimates or the RSESP estimates. By definition, it contributes nothing to the RSE estimates. Further, add-ons do not contribute anything to variance or variance of totals, but observed values that are used in models do.

So, for RSESP – RSE $\longrightarrow$ 0 we have to have $\sum_n \ldots \longrightarrow$ 0, where $\sum_n \ldots$ represents the appropriate summation over n (or all separate summations for each part of the publication group under different estimation groups, still totaling n cases).

Also $\sum_n \ldots \longrightarrow$ 0 when $n \longrightarrow 0$. If observed values used in models become designated as add-ons, this happens.

This is why it seems possible that $n \longrightarrow 0$ for some 2006 monthly data described above, including for coal-fired production in some States using large amounts of coal. However, there may be other ways for $\sum_n \ldots \longrightarrow$ 0 to occur.