U.S. NUCLEAR POWER PLANT
PRODUCTION FUNCTIONS: AN ESTABLISHMENT STUDY

Edward C. Kokkelenberg, Department of Economics and School of Management, SUNY Binghamton, Binghamton, NY 13902-6000

KEY WORDS: Panel data, econometrics, Future data unavailability

INTRODUCTION

The U.S. electric utility industry has invested about \$0.5 trillion dollars (constant 1984 overnight construction costs) in 108 currently operable nuclear power reactors that together generated 619 billion kilowatt hours or 23.8 percent of our electricity in 1992, clearly a major sector of the U.S. Economy.[1]

Yet, U.S. nuclear power plants have been, on the whole, a disappointment. This technology was originally expected to continuously generate electricity that would be too cheap to warrant metering. Further, these plants were expected to last 40 years or more before they were to be reconditioned or shut down. Neither very low cost power nor long life seems to have materialized. As a result, no new nuclear power plant has been ordered since 1978 and about 70% of the plants ordered in the last 25 years have been canceled, deferred indefinitely, or rejected.[2]

However, there appears to be considerable variation across plants, or establishments, in their initial construction costs, their operating costs, and their operating rates. Some plants produce power at costs exceeding those of modern coal-fired plants, whereas some other utilities and specific power plants have been quite successful. Several companies for example, are considered as model nuclear power plant operators (e.g. Duke Power). Further, nuclear power offers one substantial hope to slow or even reverse CO2 emissions and global warming; it is considered by many to be a most attractive alternative power source.

Therefore, it would be useful to know the reasons for the specific success and failures. This study of individual establishment efficiencies was undertaken to help determine the causes of the observed variation in nuclear power plant operations.

This present study engages the problem of efficiency using a plant level panel data set, and a production function approach. Most past economic studies of nuclear power plants have been cost function based. However, the assumptions necessary, the lack of relevant price data, and the confounding of "normal" maintenance and operating costs with capital and safety maintenance has led to problematic results. For example, it has been almost impossible to disentangle production costs from costs associated with improving the plant's ability to function or to meet safety requirements.

This stndy uses an establishment level data set that at best will be difficult to assemble for future years. The Energy Information Administration (EIA) will no longer analyze the Federal Energy Regulatory Commission (FERC) data collection forms and report the data used herein.

MODELING NUCLEAR POWER PLANTS[3]

Nuclear power plants operate in a market in which real prices are difficult to observe. This is so because there is no formal market for the risk associated with the operation or failure of nuclear power plants, many of the costs associated with nuclear power plant construction and operation can be either passed through to the rate payer or receive favorable tax treatment, and the nuclear power plant is only part of the generation capacity a utility owns or has access to.

Therefore, in this paper, profit maximizing behavior is assumed, a production function is postulated which needs no prices to estimate, thus, estimation and analysis follow. The production function assumed is a Cobb-Douglas (CD) form which has a long history in production economics, and is relatively simple. However, this form restricts the elasticitics of substitution to unity, and may lead to indeterminate input combinations for the individual establishment.[4]

The general production function may be written as:

$$Q = G_1\,[MW, K, F, Z]\, e^{wV} + G_2(L, N). \qquad (1)$$

Here,

- $Q \equiv$ Millions of kilowatt hours generated.
- $MW \equiv$ Installed "name plate" capacity in megawatts, a scale variable,
- $F \equiv$ Kilograms of fuel burned,
- $K \equiv$ Flow of services from capital stock,
- $Z \equiv$ Operating utilities cumulative nuclear power reactor experience,
- $V \equiv$ Vintage of reactor or plant,
- $L \equiv$ Labor input, and
- $N \equiv$ All other inputs,
- $e \equiv$ The Naperian base

w= A technical efficiency parameter associated with vintage

Note that separability in L and N is assumed because of data limitations.

A Cobb Douglas realization of $G_1(\cdot)$ is then:

(2) $$Q = MW^{\alpha 1} K^{\alpha 2} F^{\alpha 3} Z^{\alpha 4} e^{wV}.$$

Letting lower case letters denote natural logarithms, we can write:

(3a) $$q_t = \alpha_1 m_t + \alpha_2 k_{t-1} + \alpha_3 f_t + \alpha_4 z_{t-1} + wv_t.$$

I assume further that capital stock in place at end of period t-1 proxies for the flow of services from capital in period t (as modified by capitalized rentals), and that the cost of fuel proxies for the quality adjusted quantity of fuel.[5]

DATA

The final data set used in estimation contains 821 observations on 75 plant sites (33 with more than one reactor). The original data consists of about 1000 observations at the power plant level, some with two or more reactors, covering the period 1976 through 1991. The data were assembled and manipulated in LOTUS 123, whereas estimation was performed using Time Series Processor Version 4.2A on SUNY's ES9000. The data was taken from EIA documents, primarily Electric Plant Cost and Power Production Expenses and its predecessors, the Handy-Whitman publication Public Utility Construction Cost Bulletins and from New York State Electric and Gas data sets which were based upon trade data sources.

As usual, detailed data sets at the establishment level are fraught with problems. In some cases of multiple reactor plants, some, but not all, data were available reactor by reactor; in others the data was not. Thus, the data were aggregated so the observational unit was the plant or site, not the reactor. Where necessary, the aggregation was done by weighting by either name plate capacity or by output. Another major complication arose because of capital disallowances. Regulatory agencies have disallowed certain significant capital costs from inclusion in rate bases. In many cases, EIA source documents were adjusted to reflect these disallowances, but as the capital was believed to be still in place, where the dollar value of the disallowance was known, it was restored. Finally, some data was missing. Some times, for example in early years, a plant would sparsely report information. Some times this was then imputed by EIA, or by trade sources. When EIA imputed the observation, it generally was not so stated. When trade sources imputed or estimated the data it is noted and a dummy variable was used as a flag. The variables definitions follow those of the EIA's.[6] The variables in addition to those noted above are:

KT	=	Total capital value (dollars),
H	=	Hours connected to load,
R	=	Capital rents (dollars),
E	=	Annual operation supervision and engineering expenses (dollars),
F	=	Annual fuel cost (dollars), and
C	=	Total other operation and maintenance cost (dollars).

Note all capital values as reported are at end of the reporting year and in historic dollars. All flows (e.g. hours) are values for the reporting year and dollar values are also historic. Deflation of stocks of capital were made using the appropriate regional Handy-Whitman index for capital additions. The characteristics of the data as used in this paper (divided by the scale variable, MW) are noted in Table I.

ESTIMATION

The first estimation procedure was to use ordinary least squares to estimate a variant of equation (3), here noted as:

(4) $$q_t/m_t = a_1 + a_2 k_{t-1}/m_t + a_3 c_t/m_t + a_4 f_t/m_t + a_5 e_t/m_t + a_6 Z_t + a_7 INV_t/m_t + a_8 H_t + w V_t + A'D_t + \mu_t.$$

All terms are as above, except the log of the net real investment in capital stock is denoted inv and c denotes the log of all operating and maintenance costs exclusive of fuel, f, and engineering expenses, e. Joint ownership and other dummies for imputation etc., are denoted by the vector D, while operating hours by H. The plant subscript, i = 1--- 75 is suppressed and μ_t is an error term. Results, indicate a modest overall fit (adjusted R-squared of 0.410) and coefficients that have the expected values and signs.

These regression results were used to reduce the number of variables (those having t-statistics less than one and no theoretic basis for inclusion were dropped) and to identify "influential" observations, using the "hat" matrix test [Mosteller and Tukey, (1977)]. A dummy for each of the 58 influential observations was introduced and the regression rerun. The adjusted R-squared was 0.415.

Again the model was reduced and now reestimated via the panel routine of TSP under the assumption that Panel data should be treated as such [See Nerlove and Balestra (1992)]. The panel routine provides a total OLS estimate, an OLS estimate on the mean value of variables for each establishment, a fixed effects estimate (where the intercept term varies by establishment), and a random effects model estimation (where the model contains the usual stochastic term plus a stochastic component to the intercept). The data can be considered as a drawing from a larger population of nuclear power plants so the random effects model is a theoretically reasonable one.

The results of estimating a modified version of (4) as a panel are shown in Table II. A Hausman (1978) test, conducted to test for correlation between the error of the random effects model and the regressors, indicates that I can not reject the null hypothesis of E $(\mu_{it}/X_{it}) = 0$ at the 90% confidence level, [H calc. = 2.836, $\chi^2_{1,\ 10\%} = 2.706$], but can at the 95% level. Further, the results of an F-test indicate that the intercepts are not equal for each establishment [F(652, 158) = 3.996, F crit. (500, 150) = 1.37]. Thus, while all the results of the panel are given in Table II, I discuss only the random effects model. The estimators that are significant are the three cost categories, the years of operations, the vintage, the hours operated, and the "influential" observation dummy. Note that engineering costs (associated with downtimes?), years of operation and vintage are negative; whereas fuel, other costs, and hours operated are positive. Overall, the fit is decent as 49% of the variation in output per megawatt of capacity is explained. In the log form as estimated, the coefficients can be interpreted as elasticities. Thus we can interpret for example, the results that output increases 37.2% for a 100% increase in fuel expenditures. Overall, these results are encouraging.

SUMMARY AND FUTURE WORK

In summary, I have assembled an establishment level data set and explored its economic implications. Theory suggests this data is best modeled using a production function. A simple Cobb Douglas production function is fitted and estimation led to a narrowing of the model. Panel results are encouraging.

Several problems remain. These are the unsatisfactory nature of the current capital stock variable, the residual diagnostics (one suspects serial correlation exists), the estimation procedure, frontier efficiency, and time aggregation. These will be explored in future work, as will more complex functional forms for the production frontier.

On a negative note, I observe that in the future, data availability will be problematic. The EIA will no longer report data collected by FERC and economic studies like this will be hampered.

REFERENCES

Arrow, Kenneth J., H.B. Chenery, B. Minhas, and R.M. Solow, "Capital Labor Substitution and Economic Efficiency," The Review of Economics and Statistics, 1961, 29, 155-73.

Christensen, Laurits R., Dale W. Jorgensen, and Laurence J. Lau, "Conjugate Duality and the Transcendental Logarithmic Production Function," Econometrica July 1971, 39, 255-6.

Cobb, C.W., and (Senator) Paul H. Douglas, "A Theory of Production," American Economic Review Papers and Proceedings, 1928, 28, 139-65.

Diewert, W. Erwin, "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," Journal of Political Economy, 1971, 79, 481-507.

Energy Information Administration, Inventory of Power Plants in the United States 1988, DOE/EIA-0095(88), Washington, D.C.: Superintendent of Documents.

______, Monthly Energy Review, May 1993, DOE/EIA-0035(93/05), Washington, DC: Superintendent of Documents.

______, Electric Plant Cost and Power Production Expenses, 1991, DOE/EIA-0455(91), Washington, D.C.: Superintendent of Documents.

Halter, A.N., H.O. Carter, and J.E. Hocking, "A Note of the Transcendental Production Function," Journal of Farm Economics, 1957, 39, 4, 466-74.

Hewlett, James G., An Analysis of Nuclear Power Plant Operating Costs: A 1991 Update, Energy Information Administration, Washington, D.C., May, 1991.

Whitman Reqnardt and Associates, Handy-Whitman Index of Public Utility Construction Cost, September, 1992.

Hausman, Jerry A., "Specification Tests in Econometrics," Econometrica 46 (1978), 1251-1277.

Kokkelenberg, Edward C., "U.S. Nuclear Power Plant Production Functions," School of Management Working Paper, June, 1993.

Mosteller, F. and J.W. Tukey, Data Analysis and Regression, Reading, Mas: Addition-Wesley 1977.

Nerlove, Marc, and P. Balestra, "Formulation and Estimation of Econometric Models for Panel Data," in Laszlo Matyas and P. Sevestre, eds., The Econometrics of Panel Data, Dordrecht, the Netherlands: Kluwer (1992).

Footnotes

1 As of February 28, 1993, there were 116 nuclear power generating units in all stages of constructions and operation, Monthly Energy Review, Pg. 97.

2 Public utility executives and industry watchers suggest that no utility will order a new nuclear power plant until there is a practical plan to dispose of spent fuel assemblies, and someone (U.S. DOE?) undertakes the construction and operation of a new demonstration plant to serve as a pattern.

3 A detailed discussion of this problem is in Kokkelenberg (1993).

4 See Cobb and Douglas [1928].

5 Three other candidate production functions will be investigated later but are not further considered herein. The first is the Generalized Leontief [Diewert (1971)] the second is the Constant Elasticity of Substitution (CES) production function. [Arrow, Chenery, Minhas and Solow (1961)], and the third transcendental logarithmic (Translog) form (Halter, Carter, and Hocking (1957) and Christensen Jorgensen and Lau (1973).

6 See Electric Plant Cost and Power Production Expenses 1991, and, Hewlitt (1991), Pg. 47ff.

Table I
Data Characteristics

Variable	Mean	Minimum	Maximum	Std. Dev.
Log output	8.15	-8.15	9.02	2.26
Log lagged capital	12.51	5.35	14.56	0.74
Log Engineering Cost	8.88	4.67	11.76	1.09
Log Fuel Costs	10.12	-5.83	12.59	1.48
Log Other Costs	10.64	8.56	12.78	0.68
Years Operated	10.16	1	30	5.84
Hours	6638	-0-	8766	2078

All variables, except dummies (0, 1), years operated, and hours are scaled or divided by the megawatt or "name plate" capacity of the unit.

Table II
Panel Estimation Results
Cobb Douglas Production Function
Log of Kilowatt Hours as Dependent Variable
(Standard Errors in Parentheses)
821 Observations

Variable/ Statistic	Full Data	Between Est. on Means	Fixed Effects	Random Effects
Intercept	-7.346 (1.822)	-8.103 (4.500)	NA	-5.824 (2.367)
Lagged Capital	0.027 (0.137)	0.139 (0.406)	-0.077 (0.165)	-0.032 (0.156)
Cost Excluding Fuel and Engineering	1.123 (0.154)	1.106 (0.399)	1.053 (0.222)	1.073 (0.194)
Fuel	0.317 (0.049)	0.217 (0.116)	0.476 (0.071)	0.372 (0.060)
Engineering	-0.250 (0.090)	-0.216 (0.198)	-0.521 (0.140)	-0.384 (0.118)
Years Operated	-0.105 (0.024)	-0.023 (0.065)	-0.077 (0.033)	-0.094 (0.029)
Vintage	-0.132 (0.029)	-0.071 (0.065)	-0.321 (0.283)	-0.100 (0.039)
Investment	-0.022 (0.015)	-0.076 (0.076)	-0.013 (0.014)	-0.016 (0.014)
Joint Ownership	0.176 (0.190)	-0.191 (0.492)	0.283 (0.200)	0.216 (0.197)
Influential Obs.	0.868 (0.265)	0.949 (0.612)	0.362 (0.295)	0.634 (0.280)
Hours Operated	0.0006 (0.00003)	0.0005 (0.0001)	0.0006 (0.00004)	0.0006 (0.00004)
Standard Error	1.728	0.892	1.521	1.598
R-Squared	0.422	0.445	0.449	0.487

AVERAGE SALARY OF COMMERCIAL EMPLOYEES IN FINLAND IN 1991

Erkki J. Pahkinen,
Statistics Finland, Box 770, FIN-00101 Helsinki,Finland

KEY WORDS: average mean salary, occupational groups, establishment survey, one-stage cluster sampling, deff estimates.

1. Introduction. Each year in August Statistics Finland collects data from business firms in the commerce sector to estimate the average salary of employees in different occupations within this sector. Stratified sampling is used. The main concern in this paper is with the estimation of average salaries from this material with takes the sample design into account. The primary sampling unit in this case is the individual firm, which implies that data on salaries at the employee level are clustered by firms and accordingly that an appropriate design is needed for the calculation of average salaries. The design employed here is stratified one-stage cluster sampling: this is used to calculate the average salaries in the commerce sector as a whole as well as in certain occupational groups within this sector. The results are compared with the figures obtained with three other designs.

2. Sample design

The sampling frame is Statistics Finland's Business Register, which divides business firms in the commerce sector into two subclasses. The first comprises all firms that are members of the Confederation of Commerce Employers LTK (for convenience, LTK firms). From this subclass the Confederation collects census data on salaries in different commercial occupations. The total number of employees in this subclass is 190,217. The average salaries calculated on the basis of complete data sets will be used as a point of reference in subsequent comparisons.

The other subclass comprises firms that are nonmembers of the Confederation of Commerce Employers. From this subclass, Statistics Finland has selected a stratified simple random sample by using the individual firm as the primary sampling unit. The subclass consists of a total 57,762 commercial employees, out of whom 13,987 are included in the present sample. How, then, should the average salaries of different occupations be estimated from this sample?

The sampling frame for the present sample is the 1988 Business Register, from which the smallest companies (those employing 1-2 people) have been excluded. This leaves a population of 25,345 companies, which are stratified into five categories by number of employees and also into five categories by branch or line of business, giving a total of 25 strata. Sampling fractions vary by stratum; in some cases all firms are included, in others only part of them. The order in which individual firms appear in the Business Register is then randomized by strata. Next, starting from the top, the required number of units are sampled from each stratum. Insofar as the sampling takes place at the firm level, the sample design may be described as stratified one-stage random sampling without replacement. This type of sampling is below referred to as **STRWOR**. If conclusions were to be drawn for the firm level, then the analysis would be carried out within a stratified simple random sampling design. For example, this sort of sample design is well-suited to the analysis of turnover and similar data at the firm level.

However, the purpose here is to estimate the average salaries of employees in different occupations. This implies a different interpretation of the sample design in that the individual employee who is the unit of analysis proper is not the primary sampling unit. The selection of a certain firm into the sample means that all its employees are also included. Each firm that is selected should therefore be interpreted as a primary sampling cluster the members of which comprise all the firm's employees. This sample design is described as stratified one-stage cluster sampling **STRCLU**. There is only the one single stage in the sampling procedure; that is, the sampling of the firm. Within that firm, then, data are collected on the salaries of all employees. The analysis of employee data in company cluster design is discussed in more detail in the health studies published by the Social Insurance Institution (Lehtonen, 1988).

The specific concern here is with the regular monthly salaries of commercial occupations at the time of measurement in August 1991. These occupations are grouped according to the classification developed by Statistics Finland. Official Statistics regularly publish the average salaries of 22 occupational groups, but some of these categories are so small that for reasons of privacy protection only the job title can be indicated. The focus here is restricted to those occupational groups that occur in at least 50 sampling units or firms. One item that is obviously of special interest is the average salary for the whole commerce sector, which in the present sample design comprises 744 firms or clusters with a total of 13,987 employees. When weighted by the inverse of the sampling fraction, the size of the corresponding population is 57,762 employees - which is the same figure as that given in Official Statistics (see Palkat 1992:25).

3. Mean estimators and their weights

For the present kind of material it is possible to construct different types of mean estimators depending on how far the sample design is taken into account. The following presents four alternatives; the results for each calculation are given later.

In a simple random sampling SRSWOR the firm level is omitted and the sample at the employee level is interpreted as a simple random sample of the employee population. Here the corresponding estimator of average salary is

$$\bar{y}_{srswor} = \sum_h \sum_k \sum_i \frac{\hat{N}}{n} y_{hki} / \hat{N}, \qquad (1)$$

where y_{hki} is the salary of the i^{th} employee in the h^{th} stratum of the k^{th} firm and the joint sample size $n = 13,987$. As we can see, the same inflating weight $\hat{N}/n$ is here applied to all employees; this is the inverse of the sampling fraction. The total number of employees in the population is $\hat{N} = 57,762$, which means that the weight is $\hat{N}/n = 57,762/13,987 = 4.13$. This coefficient could only be justified if the sampling had been carried out at the employee level and if neither stratification nor clustering had been done. In the present case neither of these conditions hold. As a general rule it is possible to calculate this sort of mean estimator even from complex sample designs. Its variance is useful in determining the estimate of the design effect or $deff$, a measure which summarizes the effects of design complexities on the results. The $deff$ is for the mean as a ratio of two variance estimators

$$deff(\bar{y}) = \frac{\hat{v}_{p(s)}(\bar{y})}{\hat{v}_{srswor}(\bar{y})}, \qquad (2)$$

where $p(s)$ refers to the actual sampling design and $\hat{v}_{srswor}(\bar{y})$ is the variance estimate of $\bar{y}$ from theSRSWOR design. If the $deff$ is close to one, the actual sample design has no effect on the standard errors of the estimates. In this case analysis of the material requires no sample design codes but can be carried out by using common software packages such as SAS, SPSS and BMDP. In situations where cluster sampling is used (as in the present case), the $deff$ is typically much higher than one. This means that for purposes of statistical analysis it is necessary to use design codes and to do the actual calculations by using specialized software such as SUDAAN or CARP. The deffs of the calculations performed here are shown in the Tables in the Results section. The figures are much higher than one. In the simple random sampling design the $deff$ coefficient is by definition one ($deff = 1$).

Stratified simple random sampling STRWOR. This design is used by the Finnish Official Statistics. Different weights are chosen for different strata. The estimator of average salary is

$$\bar{y}_{strwor} = \sum_h \frac{\hat{N}_h}{n_h} \sum_k \sum_i y_{hki} / \hat{N}, \qquad (3)$$

where $\hat{N}$ is the sum of weights, or in this case 57,762 employees. This figure corresponds to the number of employees in the sampling frame population and also appears in the Official Statistics. The weight is quite simply $\hat{N}_h/n_h$, or the inverse of the sampling fraction in stratum h. It is noteworthy that the weight remains constant for all employees in the same stratum even if (as indeed is the case in practice) they work for different companies. In stratified sampling the $deff$ should usually be smaller than one ($deff < 1$).

Stratified cluster sampling; clusters of equal size STRCLU, where it is assumed that the clusters are of the same size. Here the mean estimator is the same as that used in the Official Statistics design. However, the designs produce different calculations of standard error, which are used for determining confidence intervals, for instance. In stratified cluster sampling the $deff$ is usually higher than one ($deff \geq 1$) depending on the internal homogeneity of the clusters.

Stratified cluster sampling; clusters of unequal size STRCLUWGHT, where the size of the clusters may vary. This is a very realistic assumption in samples of business firms. The size of firms (i.e., the size of the cluster), measured in terms of the total number of employees, varies considerably.

In this case the design can be taken into account by inferring the mean estimator according to *Horvitz–Thompson* and regarding the relative size of the cluster as sampling weight. Here the relative size of the cluster is measured by the number of employees in the firm m_{hk} divided by the total number of employees in the corresponding stratum M_h. This will yield a cluster weight for a certain firm, and the inverse of this figure is accordingly the raising factor for that particular firm. To match the sum of the weights with the total number of employees within the population of the entire material, this figure must still be divided by the number n_h of sample firms in the stratum. This will give the following mean estimator

$$\bar{y}_{strcluwght} = \sum_h \sum_k \sum_i \frac{M_h}{n_h \times m_{hk}} y_{hki} / \hat{N}. \quad (4)$$

This equation incorporates all information concerning the sample design. It contains the sampling weight that varies from firm to firm and the data obtained from stratification. This design also includes in the divisor the weighted sum of employees, which is the same figure as earlier, 57,762 employees.

4. Results

The sample material has been analysed by the SUDAAN software package which takes into account the sample design (see Shah et al., 1991). The settings have been entered in accordance with the four designs described above. The following shows the calculation routine used in the Official Statistics **STRWOR** design. The only results shown are average salary as well as the average salary of one occupational group.

Table 1. SUDAAN code for an analysis of mean salary under stratified simple random sampling design.

```
               SUDAAN   Survey Data Analysis Software
  Copyright    Research Triangle Institute    February 1993
  Release 6.31
  1      PROC DESCRIPT DATA=WAGE 910S FILETYPE=SAS
             DESIGN=STRWOR MEANS DEFT;
  2      WEIGHT STRWGHT;
  3      NEST STRATUM;
  4      TOTCNT STRSIZE;
  5      VAR WAGE;
  6      SUBGROUP OGROUP;
  7      LEVELS 22;
  8      SETENV LINESIZE=60;
  9      PRINT MEAN SEMEAN DEFFMEAN \STYLE=NCHS;
  Number of observations read : 13987  Weighted count: 57762
  Denominator degrees of freedom : 13967
  Date: 04-19-93   Research Triangle Institute    Page : 1
  Time: 12:26:27                                  Table: 1
  by: Variable, OGROUP.
  ------------------------------------------------------------
  Variable                                              DEFF
       OGROUP                    Mean       SE Mean     Mean
  ------------------------------------------------------------
  WAGE
  All employees               9528.07        54.65      1.72
  of which; shop managers     8503.93       168.90      1.77
```

In this printout the $deffs$ are higher than one even though they should usually be smaller than one in stratified simple random sampling ($deff \leq 1$). One possible explanation is that the stratification has been carried out not at the level of employees but at the level of firms, which means that the raising factor used by Statistics Finland at the employee level has not been consistent with the sample design employed. On the other hand, it was also found when the sample material was examined afterwards that some firms had outgrown their original size category. This is explained by the three-year time lag between the sampling frame (the 1988 Business Register) and the actual sampling date in 1991. For instance, a firm that in the sampling frame is classified as having $5-10$ employees may have grown to employ 22 people; in other words it should actually be in the stratum $20-50$ employees. This, however, provides only a partial explanation for why stratification does not lend itself fully to improving the accuracy of estimation.

The corresponding calculations have also been performed for the three other designs, which have different weightings of employees and which take the sample design into account to different extents. The most important of these designs is quite obviously STRCLUWGHT, which incorporates all the information concerning the sample design. The results of the calculations based on these sample designs can also be compared with the statistics on average salaries in the member companies of the Confederation of Commerce Employers, which represent a total of 190,217 employees. The statistics compiled by the Confederation can be regarded as a complete enumeration of the population concerned: they collected data on the salaries of all employees in their member firms in August 1991. In the Table below, these data are shown on the line 'CENSUS LTK'. The Statistics Finland sample specifies the estimated number of employees at 57,762, which means that the figure for the whole sector in August 1991 would have been $57,762 + 190,217 = 247,979$ full-time employees.

Table 2. Average salary of service sector employees in 1991 on the basis of different criteria in August 1991.

Sample design	Weighted sample size	Average salary	Standard error	deff
SRSWOR WGHT=NO	57762	10 458:-	44:-	1.00
STR WGHT=KERR	57762	9 528:-*)	55:-	1.72
STRCLU WGHT=KERR	57762	9 528:-	60:-	2.10
STRCLU WGHT=CLUWGHT	57762	9 402:-	66:-	2.58
CENSUS LTK REGISTER	190217	9 098:-	..	..

*) Published by Statistics Finland

As we can see, the unweighted calculation or the **SRSWOR** design gives the highest average salary at FIM 10,458. On the other hand, it also has the lowest standard error with an s.e. mean of FIM 44. In other designs the average salary approximates the reference figure under 'CENSUS LTK REGISTER', which is FIM 9,098. Since this is the exact figure for the subpopulation, it obviously contains no standard error. The design that comes closest to the reference figure gives **STRCLUWGHT** FIM 9,402, which is more sensitive than any other option to the sample design and which also has different weights for different firms. The deffs increase in direct proportion to the amount of information injected into the calculation design. This is partly related to the fact that the primary sampling unit has been the firm, but the weighting is done at the lower, employee level.

5. Comparison of the results

Moving on to look at average salaries in selected commercial occupational groups, the following compares the figures of three of the above sources: the Confederation of Commerce Employers registry data, the Official Statistics calculations based on **STRWOR** and finally the **STRCLUWGHT** figures. The comparison covers the biggest occupational categories on which data have been obtained from at least 50 companies.

Table 3. Average salaries in different occupational groups in August 1991: LTK member companies and the Statistics Finland sample.

	AVERAGE SALARY IN AUGUST 1991		
	CENSUS	FINSTAT'S SAMPLE	
OCCUPATIONAL GROUP	LTK MEMBER	STRCLUWGHT	OFF.STAT.
Total	9098	9402	9528
7 Shop managers	9582	8835	8504
10 Service station workers	6893	6977	6905
13 Shop assistants	..	..	..
14 Cleaners	6840	5417	5386
15 Warehouse workers	7106	7112	7082
16 Warehouse supervisors	..	..	..
17 Van\Lorry drivers	7804	7138	7231
25 Forwarders	8944	12866	13635
39 Other branches	8407	7656	7748
41 Upper white-collar	15131	14432	14395
42 Office management	19212	19659	19774
43 Office supervisors	13969	15000	15117
44 Clerical staff	8881	10157	10151
49 Motor-transport workers	9593	7917	7871

There are clear differences between the figures based on the census data and the sample compiled by Statistics Finland. However, since these differences only occur in a small number of occupational groups, it would seem useful to look more closely at the internal compatibility of occupational classifications used in different statistical sources. On average, the results from the **STRCLUWGHT** design come closer to the census figures than the Official Statistics results which use the **STRWOR** design. The use of a complete design significantly increases the standard errors of average salary estimates. One possible reason for this is that during the time lag between the compilation of the sampling frame and the sampling date, firms have have moved up or down from their original size category but retained the weight of that stratum. This was evident in the $deffs$ in the sample design employed by Statistics Finland ($deff = 1.72$). The effect is further accentuated with the use of firm-specific weights, which increases the $deffs$ and accordingly the standard errors of average salaries.

5. Conclusions

This paper presented four different statistical analyses of the same material by varying just one parameter: the degree to which the sample design is taken into account. It was discovered that in the estimation of average salaries for different occupational groups, the sample design used by Official Statistics should be interpreted as stratified one- stage cluster sampling in which business firms are the primary sampling clusters. Since the size of these clusters varies, it is also necessary to adjust the raising factors so that the weighting can be done at firm level. The estimate of average salary produced by this sort of design came closest to the reference figure based on census data. The relatively high deff estimates of the cluster designs ($2.10 \leq deff < 2.58$) lend further support to the argument that there is a powerful cluster effect that must be taken into consideration in the calculation of average salaries in business firms. Cluster effect here means that employees working in a certain occupation within the same firm (say, shop assistants) have more or less the same salary, whereas their salary is clearly different from the average pay of shop assistants in other firms. This observation also supports the view that the calculation of average salaries should use weights at the cluster level, as we have done here. Another factor that speaks in favour of cluster-level weights is the wide range of variation in firm (cluster) size. The most natural way to do this is to apply $Horvitz - Thompson$ estimators.

Acknowledgements. My thanks are due to Ph.D. Risto Lehtonen of the Finnish Social Incurance Institution, Mr. Seppo Kouvonen of the Finnish Statistical Office and Mrs. Anja Roikonen of the University of Jyväskylä who have helped me with their advises and contributions to fullfill this paper.

References:

Lehtonen,R. (1988). *The execution of the National Occupational Health Care Survey.* Helsinki: Publications of the Social Insurance Institution, Finland, M:64 (In Finnish with English summary).

Shah, Babubhai V., Barnwell, Beth G., Hunt, P. Nileen, and LaVange, Lisa M. (1991). SUDAAN User's Manual, Release 5.50. Research Triangle Institut, Research Triangle Park, NC, 27709.

Statistics Finland (1992). *Wages and Salaries in Service Sector in August 1991* (in Finnish). Tecnical Report. Palkat 1992:25. Helsinki.

ALTERNATIVE TO THE ITERATED REWEIGHTED LEAST SQUARES METHOD: APPARENT HETEROSCEDASTICITY AND LINEAR REGRESSION MODEL SAMPLING

James R. Knaub, Jr.
U.S. Department of Energy, EI-521, Washington, D.C. 20585

KEY WORDS: Model failure, weighted least squares

Abstract:
Knaub (1991) and Knaub (1992) introduced a method to measure heteroscedasticity which may often be superior to the usual Iterated Reweighted Least Squares (IRLS) method. This new method relies on decomposing heteroscedastic error into random and nonrandom components. A graphical analysis is then employed. This procedure, in conjunction with the IRLS method and other possible considerations appears to be useful in identifying certain substantial model failures. Establishment surveys are investigated here.

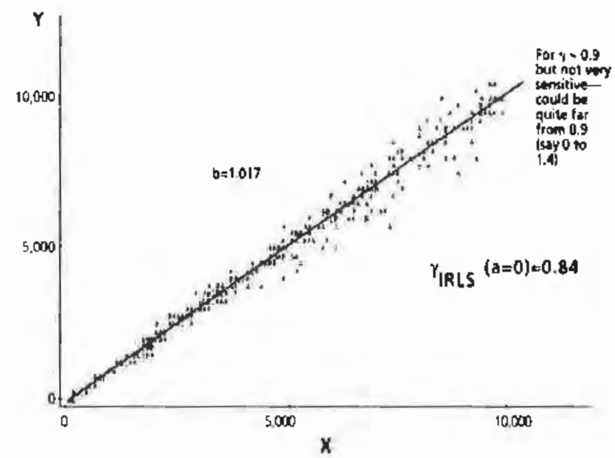
NEW METHOD TO MEASURE HETEROSCEDASTICITY:

Royall (1970) examined the usefulness of linear regression models for estimating totals. Zero-intercept modeling was emphasized. Totals are estimated by summing the observations and adding an estimate of the remainder. Observations not found in the sample (sampling and imputation being somewhat synonymous here) are modeled as $y_i = bx_i + x_i^{\gamma} e_{o_i}$. (Note that when gamma is 0.5, the result is the usual ratio estimator.) The IRLS method could be used to estimate gamma, but an alternative method will be described here. For this alternative approach to estimating gamma, each error is considered as a product of a random component, e_o, and a nonrandom component, x^{γ}. If a linear model with zero intercept and heteroscedasticity such that $\gamma = w$ is appropriate, then if all $y_i - bx_i$ (error) values are divided by x_i^w the resulting error components should be nearly homoscedastic with expected value approximately zero. If a homoscedastic, linear regression is fit to the absolute values of these error components on the vertical axis and the original regressor is still on the horizontal axis, then the slope should be near zero, as there should be no growth trend, either positive or negative, as x increases. Next, if these slopes are plotted on a vertical axis against gamma values on the horizontal axis, then the points where the plotted line crosses the horizontal axis are of interest. To make these points more obvious and graphs easier to manage, the absolute values of the slopes are plotted against gamma. The shapes of these plots are studied as they relate to model failure. Note that this method could easily be used to study other functional forms for the nonrandom error component.

GRAPHS ILLUSTRATING MODEL FAILURE:

In the following, linear, zero-intercept modeling is demonstrated under various conditions. The second graph in each pair shows the absolute values of the 'slopes' of interest (for the new method of estimating gamma) on the y-axis, ABSYRL, against gamma on the x-axis.

Figure 1. Here the model fits very well. The "V" shaped graph is typical when this is the case.



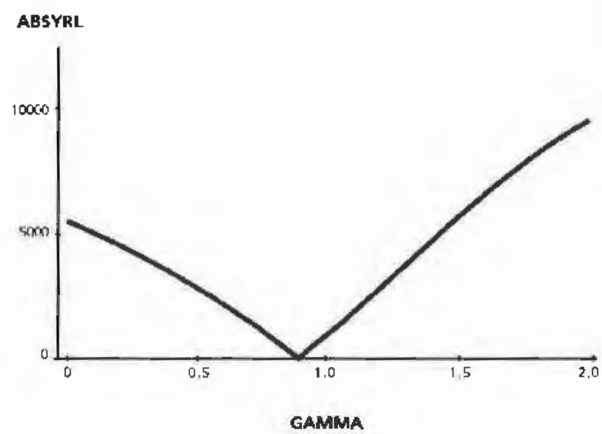
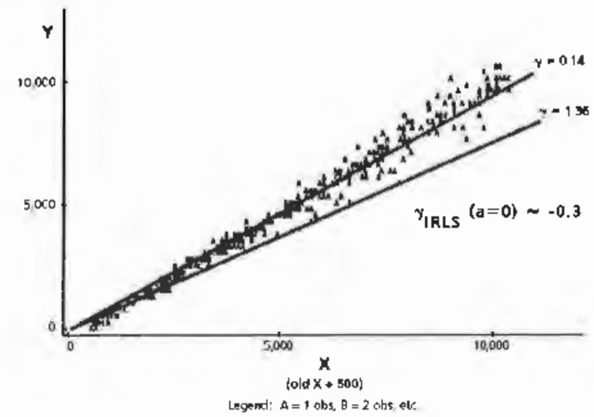


Figure 2. Here a non-zero intercept is introduced. The IRLS method, assuming a zero-intercept linear model, imposes a line between the origin and the 'largest' of the data points resulting in a negative estimate of gamma. Note that the two values for gamma identified by the new methodology correspond to the two lines shown on the raw data graph. (Gamma = 1.36 was weakly indicated. In other cases a 'weakly' indicated gamma value may be closer to zero, and may be the best estimate of gamma to use.)



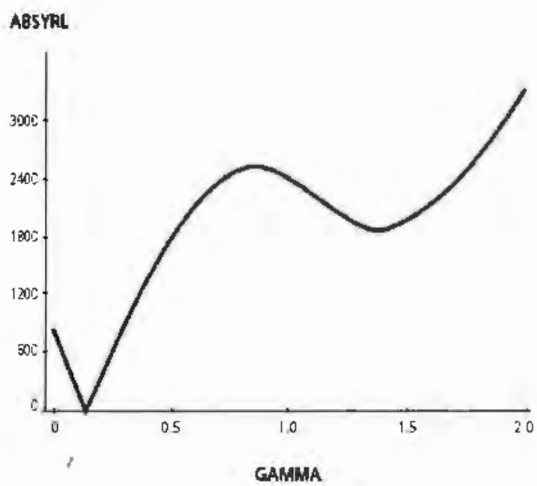
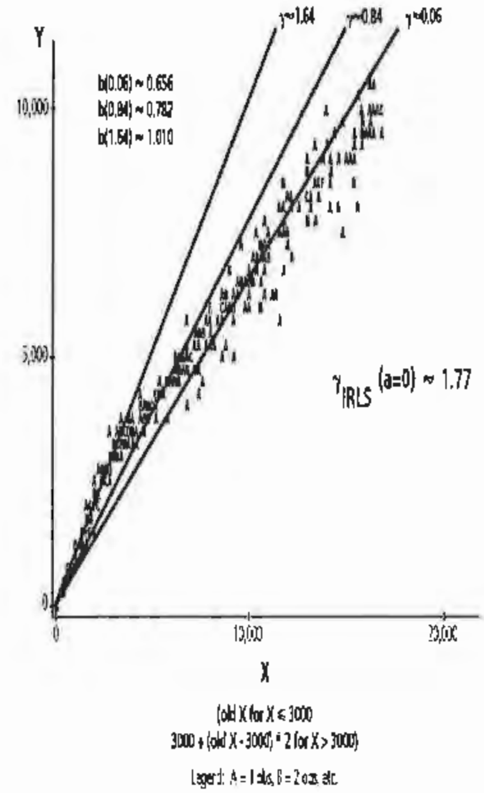


Figure 3. In this case, the data are contrived to show an obvious nonlinear relationship. Note that the three values for gamma indicated by the new methodology correspond to the three lines shown on the raw data graph. Gamma = 1.64 is nearly what is obtained for this particular case using the IRLS method with intercept zero.



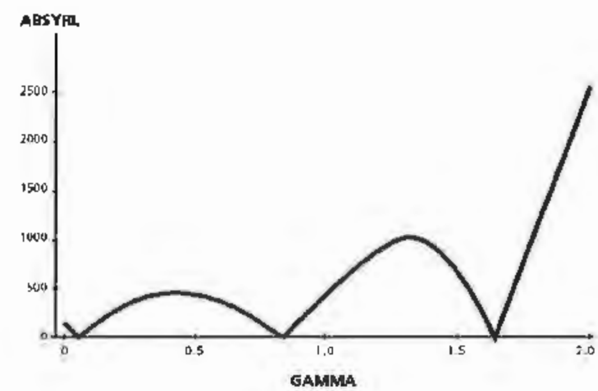


Figure 4. a) A small sample size coupled with a possible outlier problem is shown here. b) Here, another possible outlier is shown.

a) b)

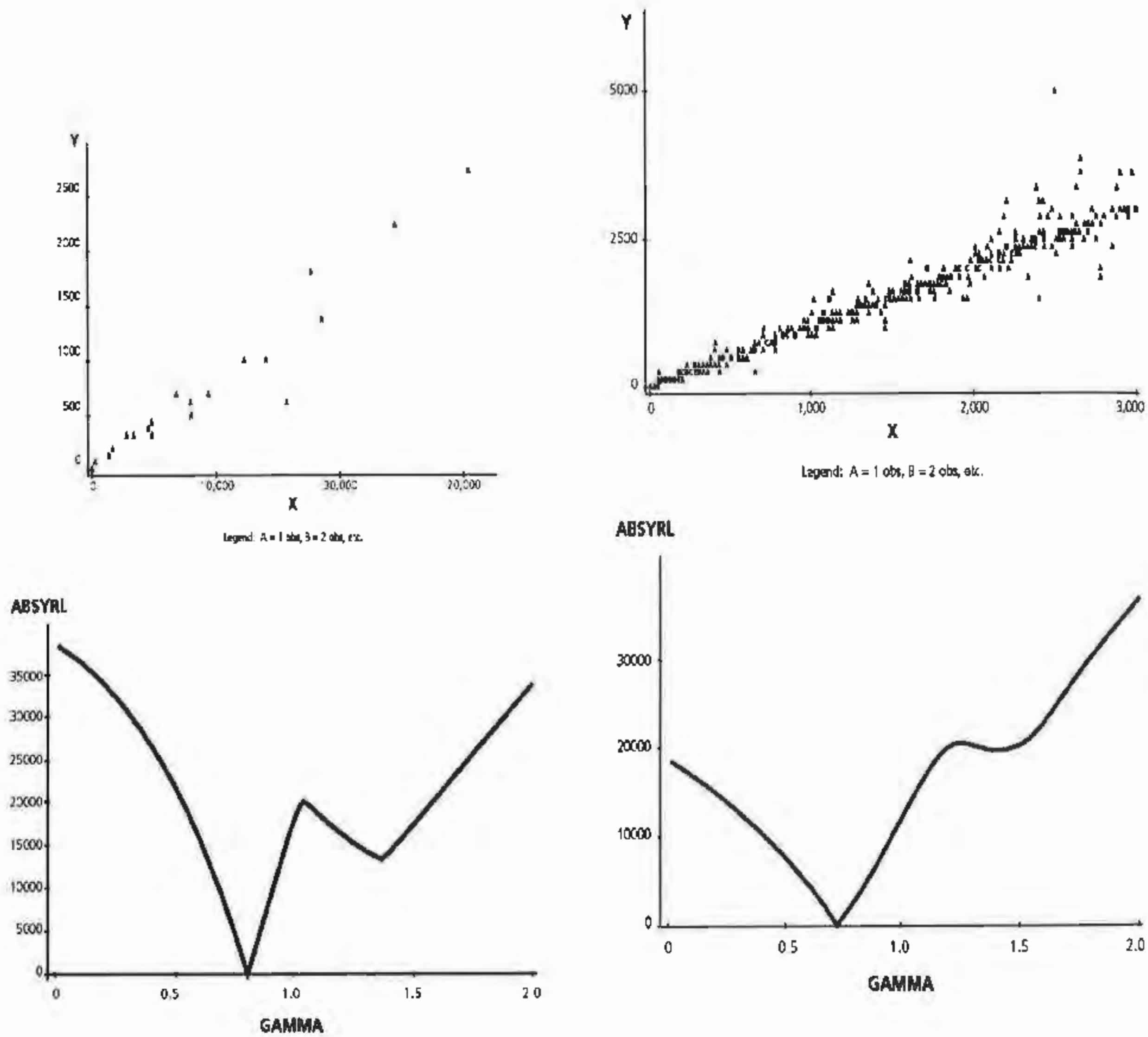


Figure 5. If a few observations are much larger than the others, then the ABSYRL vs. Gamma plot tends to often be 'flat' past the estimated 'best' value of gamma.

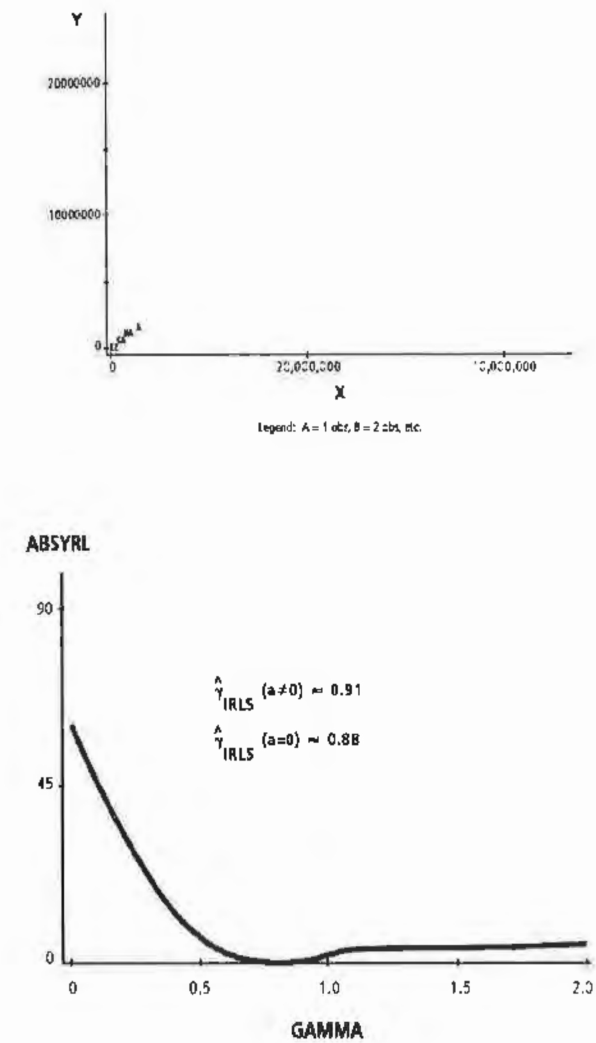
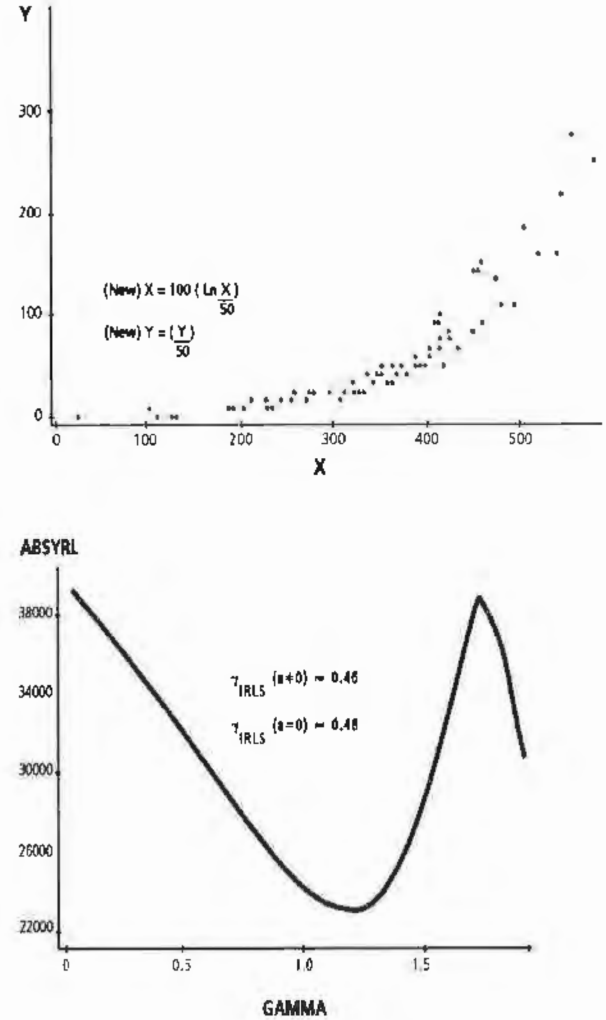


Figure 6. Here is another example dealing with nonlinearity.



OBSERVATIONS, NOTES AND CONCLUSIONS:
The preceding graphs illustrate that this new method of measuring heteroscedasticity may also indicate when substantial model failure has occurred. However, looking for a "V" shape is somewhat subjective. How unlike a "V" may such a graph be, and under what conditions? This is analogous, to an extent, to hypothesis testing. If the sample size dependency of results is ignored, a poorly informed decision may be reached. This problem is common to most hypothesis testing (Knaub 1987), other than sequential hypothesis tests. Hypothesis tests dealing with heteroscedasticity suffer from this problem, unless some sensitivity

analysis is included, thus increasing the importance of this graphical analysis and other considerations. Other information, such as a comparison of robust and nonrobust variance estimates, may help when studying model failure.

Also, this method may be superior to the IRLS method, and alternative functional forms for the nonrandom component of error may be easily substituted. In cases such as those shown in Figures 3 and 4, the IRLS method may perform particularly poorly. (The IRLS method sometimes has difficulty with convergence. Also, it may often be relatively sensitive to the addition or deletion of a single response when sample sizes are small.) My new method, however, may yield multiple solutions, as in Figure 3: possibly one biased toward the points near the origin, one toward larger values, and one may be a more reasonable compromise. Perhaps a gamma value slightly larger than 0.5 may be best, especially in the case of substantial nonlinearity. Here, if one of the 'best' gamma values indicated by my method is between 0.5 and 1.0, then one could subjectively choose a gamma value between that and 0.5 (and/or try a transformation). When there is substantial nonlinearity, it seems that this inflates the apparent heteroscedasticity, as might be expected. It may be difficult to identify an appropriate linear transformation, and simply using the linear model with gamma = 0.5 often performs very well. If a transformation is used, a step function seems to work best as log and power transformations on the regressor alone tend to put curves not only where needed, but also, where they are not wanted. Transforming the dependent variable when model sampling to estimate totals may cause unwarranted difficulties.

In Kirkendall (1992), Knaub (1990), and Knaub (1992), there are other indications of the usefulness of gamma = 0.5 as possibly a compromise value. Some degree of model failure is inevitable, but considerations discussed here may help to determine the level of reaction required. When sampling only a few of the largest establishments, as in Knaub (1992), Table 1, gamma may not be well estimated, even if it is nearly uniform throughout the range of establishments in the universe. For imputation, gamma may be well estimated, but enough model failure will reduce its relevance. Graphs in the previous section illustrated this. (Note that although it is usually best to estimate parameters with a different set of data than is used for estimation of totals or whatever is of interest, this does not seem to be practical here.) In Figure 3, it is also apparent that there is something else to consider. If one is estimating totals or means, then gamma between 0.5 and 0.84 is reasonable if the unsampled or 'missing' responses could occur for any x_i, but if a projection for a larger x value is wanted, this would be different.

For the case of a substantially nonzero intercept, this is perhaps the easiest problem to detect and to correct. Examining the ABSYRL vs. gamma portion of Figure 2, there is a decided disturbance in the right side of that graph. (Note that all ABSYRL vs. gamma plots in this paper show gamma from 0 to 2.)

A 'large' difference between IRLS (a=0) results and those from the new method may also be an indication of model failure. Also, if one does not restrict the intercept to zero (i.e., "a" not necessarily equal to zero), then this is another way to examine that sort of model failure.

Note that when there are a few observations much larger than the others (see Figure 5 for example), gamma = 0.5, as noted above, may be a 'good compromise' as the largest observations may have a different degree of heteroscedasticity and/or other model failure than the smallest observations. In this case, the ABSYRL vs. gamma plot ordinarily shows, as in Figure 5, a 'flattened' right end of the plot (when the x-axis shows gamma up to gamma=2). Perhaps a better way to manage some such cases would be to consider the largest observations as a separate, censused stratum and then to use a model for the remainder of the population without using those relatively very large observations in any calculations involving the stratum not censused.

The plot used in this new method for estimating gamma is volatile with respect to small sample sizes. See Figure 4a. However, when sample sizes are 'large' and the variance is substantial, a "V" shaped ABSYRL vs. gamma plot, such as the one shown as part of Figure 1, may result, even when nonlinearity seems obvious. However, such a case may illustrate some robustness for the linear regression, zero-intercept model.

It is also important to note that, especially when the model fits very well, estimates of totals are often not very sensitive to changes in gamma. It can be seen that totals would not be influenced on average by gamma as long as there is symmetry in the distribution of data points about the fitted line. If the assumptions of linearity and a zero-intercept are reasonable and sample sizes are sufficient, then it would seem that near symmetry would be the usual case. Thus with large sample sizes when model failure is not a problem, the estimation of totals seems to be largely insensitive to gamma. This has appeared to be the case in practice.

Perhaps, the real usefulness of weighted least-squares regression is in not allowing data points with the largest variances to have undue influence on fitting the model. From Willett and Singer (1988), page 236, "Whether the investigator wishes to downplay the importance of data points that are intrinsically more variable..., or simply to decrease the effect on fit of remote data points...," one can see that this is simply a matter of confidence in the data. However, Figure 3 shows that the IRLS method can be very vulnerable to model failure. Therefore, it may not estimate very well what the weights on the largest observations should be.

In Knaub (1992), it is noted that a robust variance estimator, V_D (Royall and Cumberland (1978, 1981)), does not really do much better than the more model dependent V_L. Perhaps, as mentioned earlier, however, another consideration when studying model failure should be to examine how closely estimates from these two estimators tend to be. (Note, however, that cv estimates do not generally appear to be as good as the estimates of the totals themselves.)

If all of the above considerations are employed, then the use of a linear, zero-intercept model for estimating totals from establishment sample surveys which concentrate on the largest possible respondents has, in practice, been shown to be very useful for energy data. The variety of applications and possible applications using energy data is substantial. There are cases where regressor data would be the same variate for a previous period, and those where the regressor is a different variate entirely. As with all establishment surveys, the data are highly skewed, and one may look at only the largest few possible respondents, or perhaps, in the case of imputation, look at data from all but a very few establishments. Even with this variety of circumstances, the estimated gamma values, by either the new method or the IRLS (zero-intercept) method are very often found to be between 0.5 and 1.0 for electric power data.

Acknowledgements:

Thanks to Channele Donald for preparation of the manuscript and to the DOE graphics department and others for support and influence. Hopefully, the notes on hypothesis testing will answer questions raised at the conference.

References:

Kirkendall, N. J. (1991), Presentation to ASA Committee on Energy Statistics, Oct. 1991.

Kirkendall, N. J. (1992), "When Is Model-Based Sampling Appropriate for EIA Surveys?" *Proceedings of the Section on Survey Research Methods*, Amer. Stat. Assoc., pp.637-642.

Knaub, J. R., Jr. (1987), "Practical Interpretation of Hypothesis Tests," letter, *The American Statistician*, August 1987, Vol. 41, No. 3, pp. 246-247.

Knaub, J. R., Jr. (1990), "Some Theoretical and Applied Investigations of Model and Unequal Probability Sampling For Electric Power Generation and Cost," *Proceedings of the Section on Survey Research Methods*, Amer. Stat. Assoc., pp. 748-753.

Knaub, J. R., Jr. (1991), "Some Applications of Model Sampling to Electric Power Data," *Proceedings of the Section on Survey Research Methods*, Amer. Stat. Assoc., pp. 773-778.

Knaub, J. R., Jr. (1992), "More Model Sampling and Analyses Applied to Electric Power Data," *Proceedings of the Section on Survey Research Methods*, Amer. Stat. Assoc., 876-881.

Royall, R. M. (1970), "On Finite Population Sampling Theory Under Certain Linear Regression Models," *Biometrika*, 57, pp. 377-387.

Royall, R. M., and W. G. Cumberland (1978), "Variance Estimation in Finite Population Sampling" *JASA*, 73, pp. 351-358.

Royall, R. M., and W. G. Cumberland (1981), "An Empirical Study of the Ratio Estimator and Estimators of its Variance," *JASA*, 76, pp. 66-88.

Willett, John B., and Singer, Judith D., (1988), "Another Cautionary Note About R^2: Its Use in Weighted Least-Squares Regression Analysis," *The American Statistician*, August 1988, Vol 42, No. 3, pp. 236-238.

ESTIMATION OF MODEL PARAMETERS ACROSS MULTIPLE YEARS: APPLICATIONS USING SURVEY DATA

Gerald Whittaker, Hisham S. El-Osta, and Mary C. Ahearn, U.S.D.A., Economic Research Service
Gerald Whittaker, ERS, 1301 New York Ave. NW, Rm 937, Washington, D.C., 20005

KEY WORDS: time series, complex survey data

Economists commonly use regression analysis to estimate model parameters. The analysis is complicated significantly when the data are from a complex survey design because the estimation of the variance of parameters is not simple. Software, such as PC CARP (Fuller, et.al., 1986, Schnell et.al., 1988), exists to account for the complex sample design in variance estimation of a single equation regression model. Software is limited for other more complex regression models, such as a model which pools cross-section and time-series data, when the data have a complex sample design. Pooling of cross-section and time-series data is of interest when sample sizes are small or when it is necessary to control for annual changes, and when cross-section parameters are believed to remain constant over time.

In the absence of the "ideal" software tools for data with a complex sample design, an economist faces several undesirable choices, such as ignoring the complex sample design or limiting the analysis to a single year. This paper presents the results of a comparison of variances between a pooled model and models based on single years of complex survey data using a new approach suggested by Kott (1992). We draw on two distinct examples to gauge the stability of our findings.

Methodology

Kott suggested a method for estimating parameters across multiple years when the data have a complex sample design (1992). He states that there are two conceptual issues which must be resolved before combining data across years. The first issue involves the definition of the target population. We use data from USDA's annual Farm Costs and Returns Survey (FCRS), whose target population is all establishments sold or normally would have sold at least $1,000 of agricultural products during the year. This definition encompasses the constant transition in the actual farm population. Consequently, the target population of the survey changes from year to year as people enter and leave farming. For any particular year, the target population is a finite population. However, a target population for a single year can also be viewed as an infinite conceptual population of farms, of which the farms in the finite population for any year can be viewed as a single sample drawn from that population.

The second issue to be resolved is the nature of the parameters to be estimated. If the economist is concerned with estimation of a parameter for a single year, combination of the data with other years to increase sample size makes little sense. However, if the economist is interested in the relationship, regardless of year(s), then pooling years may be justified.

Kott proposes two approaches to estimation across years, one for finite population parameter estimates and one for model parameter estimates. It is the model parameters, relating to the underlying conceptual population, that we wish to estimate in this study. The model parameter estimation procedure is used because we wish to learn about the infinite conceptual population, of which the observed farms are simply realizations.

The procedure for combining annual data from the FCRS for estimation using PC CARP follows:

(1) Begin with a data set with several years of observations, including a variable YEAR to denote the year of an observation.

To accommodate the survey design, PC CARP requires variables for stratum identification, cluster (primary sampling unit), and weight (where weight is the inverse of selection probability). The variable names here follow Kott and the FCRS.

(2) Create stratum identification SUSTRM for observations on the list frame:

SUSTRM = 100 + STATE and

SUSTRM = 200 + STATE

for observations on the area frame.

(3) Create cluster identification PSU:

PSU = SEGMENT,

for the list frame, and

PSU = STRATA * 10,000 + SEGMENT,

for the area frame, where SEGMENT is the PSU for FCRS and STRATA is the stratum identifier.

(4) Sort the data by SUSTRM and PSU.

Kott remarks that the variance estimators using this procedure will tend to be conservative because not all the variance reducing potential of the stratification on the list or area side is being captured.

Applications

The first application concerns the effect of government farm commodity programs on the net cash farm income of U.S. farms. The question has been posed by a former assistant secretary of agriculture and seems a simple one: "Did government programs improve farm income?" However, the measurement of this effect is quite difficult, given a multitude of programs, constant incremental changes in the programs, and indirect benefits received by farms which do not even participate in the programs. Farm income is highly variable, primarily due to weather, but prices are also quite volatile. We therefore needed several years' data to try and control for these effects. Coincident with the preliminary issues discussed above, our objective was the relationship involving the farm population, that is, the conceptually infinite population realized by our observations.

Compared with quite a number of simultaneous equation models which are usually used to answer this type of question, but with little statistical basis, we chose ordinary least squares as our estimation procedure. We regressed net farm income on total value of production (TOTVPRD), a production efficiency measure (EFFIC), a binary variable representing participation in commodity programs (PARTIC), regional indicator variables (REG_), fraction of value of production from program crops (FRAC1), and fraction of value of production from non-program crops and livestock (FRAC2). Three years of data were used, 1987-1989.

The second application used logistic regression on a model of off-farm labor participation by farm operators. Most of the studies cited in the literature have used cross-sectional data. Heckman and Macurdy note that the major limitation of using cross-sectional data is the inability to analyze the response of labor to life-time variation in costs and opportunities due to children, unemployment of the spouse, and general business cycle variation. All these factors are considered influential in the timing of labor force participation and not the volume supplied to the market. Hence, when annual data are used, the labor force participation decisions are implicitly assumed symmetrical in that the factors affecting labor force participation have equal but opposite effects on the probability of nonparticipation (Gould and Saupe). The use of panel or longitudinal data is therefore preferred since it will allow for the incorporation of the dynamic aspects of labor force participation. In addition, a single year's data is limited in the variation of local labor market conditions, since many of the sample farms may be located in a single labor market area.

The dependent variable used in the off-farm labor participation model is binary and was coded 1 if farm operators had participated in off-farm employment, zero otherwise. The explanatory variables are operator's age (OPAGE) and the square of operator's age (OPAGESQ) to allow for life-cycle effects, operator's education (OPEDUC), size of operator's household (HHSIZE), unearned income from off the farm (OTHINC) including transfers (i.e. social security, pensions, etc.) and nonfarm rental income, value of machinery per acre (KACRE), and a government payment dummy variable (GOVT). Variables that characterize local areas are also considered and they include county unemployment rate (UNEMP); percent of county income from agriculture (AGRIN); county employment change (EMPCHG); and percent of county employment in manufacturing (MANUF), construction (CONST), services (SERV), and wholesale and retail trade (TRADE). We used three years of data for this application, 1987-89.

Results and Conclusions

The over-all tests of parameter significance were significant in both applications. In the net farm income models, the dummy variables associated with regions were mostly not significant in the estimations using annual data. This result would not be accepted by the agricultural policy community, as it is known that there are important regional differences in farm profitability and level of government payments. However, most of the regional variables were significant in the pooled estimation.

In the off-farm labor supply models, the majority of the farm and household variables were significant. This was true for both the single year models and the pooled models. For the characteristics of the local labor markets, four of the eight variables were significant in the pooled model, compared to one in the single year models.

Figures 1-2 show the estimated standard error for each variable for the annual fits and for the pooled data for both applications. For every variable, pooling of the data resulted in a substantial (at least 20 percent) reduction in variance from the largest annual error.

Our results indicate that, in the cases where pooling is desirable, Kott's suggested approach to variance estimation is preferred to other alternatives, such as using small samples of a single year's data.

References

Gould, Brian W. and William E. Saupe (1989). Off-Farm Labor Market Entry and Exit. American Journal of Agricultural Economics, 71, 960-969.

Heckman, James J. and Thomas E. Macurdy (1980). A Life Cycle Model of Female Labor Supply. Review of Economic Studies, XLVII, 47-74.

Kott, Phillip S. (1992). A Note on Estimating Population and Model Parameters Across Multiple Years of FCRS Data Using SUDAAN. Unpublished manuscript, October.

Fuller, W.A., W. Kennedy, D. Schnell, and H.J. Park (1986). PC CARP, Statistical Laboratory, Iowa State University.

Schnell, D., W.J.Kennedy, G. Sullivan, H.J. Park and W.A. Fuller (1988). Personal Computer Variance Software for Samples Surveys. Survey Methodology, 14, 59-69.

Figure 1. Farm income, comparison of standard errors, by variable, by year

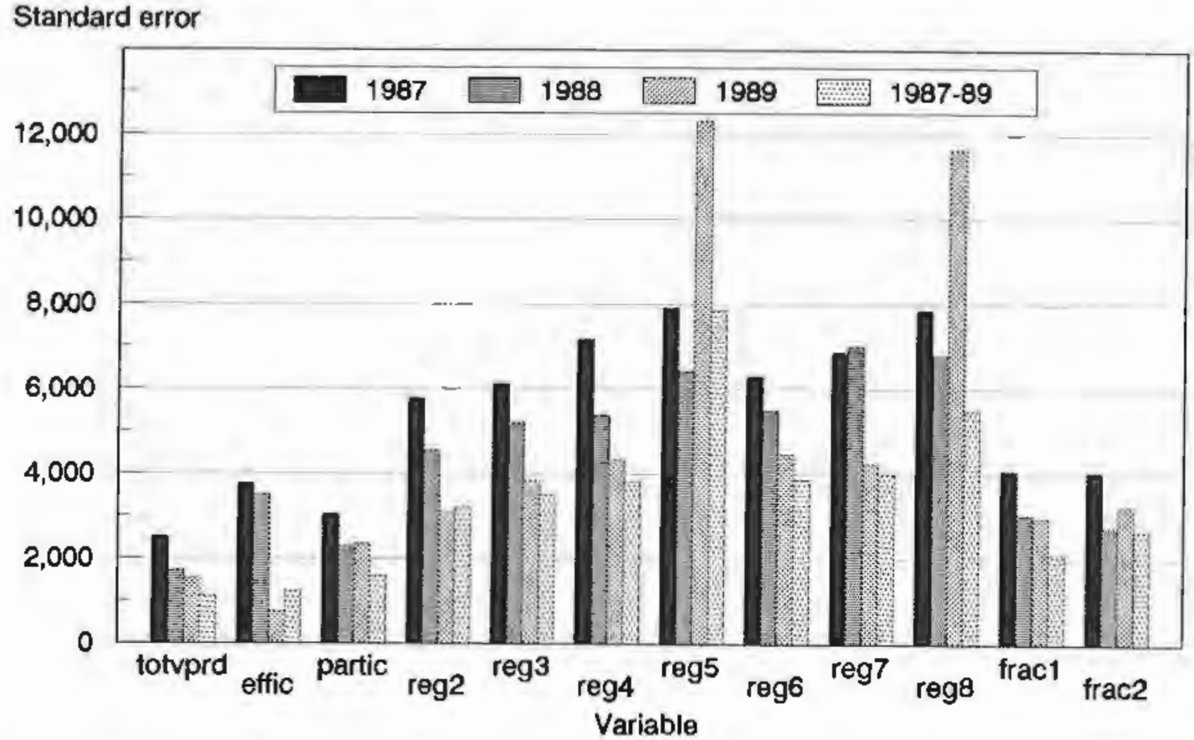


Figure 2. Off-farm labor, comparison of standard errors, by variable, by year