Appendix C

Quality of the Data

Appendix C

Quality of the Data

Introduction

All data collection activities and the population estimates produced from them are subject to a variety of errors. These errors may be broadly classified under two general types, sampling and nonsampling errors.

Sampling errors are defined as the variability in a survey estimator that arises because data used to estimate population values are collected from a sample of units rather than the complete enumeration of the entire population. Each possible sample produces different estimates of population values, depending on the set of respondents that are selected. Consider, for example, a sample of two units from a population comprised of three units. In this example, there exist three possible sample sets of respondents, each of which produces a different estimate of the population total. The difference between the estimate calculated from one of the samples and the population total is referred to as the sampling error.

Nonsampling errors, on the other hand, occur in any data collection activity, whether a sample survey or a complete enumeration of the population. Nonsampling errors may be associated with any part of a survey process except sampling and can include both random and systematic (biasing) errors. Commonly recognized sources of nonsampling error include undercoverage, random and systematic response errors, unit and item nonresponse, data processing errors, and tabulation errors. This appendix describes the effect of both sampling and nonsampling errors on data from the MECS. In addition, the measure for sampling errors of the population estimates are given. More details are presented in the methodological report for the MECS.[44]

Sampling Errors

The estimated values appearing in this report were developed from a sample of the universe of manufacturing establishments and, as a result, will differ from true population values that would be obtained from a complete enumeration of the manufacturing universe. This difference occurs because the MECS sample is only one of a very large number of samples that could have been selected under the same sampling specifications. Each possible sample would yield its own unique estimates of the true population values, with the differences attributable to the particular set of establishments selected into each sample.

One measure of variability due to sampling is the square root of the average of the squared differences between the estimates that would be produced by all possible samples and the mean value of those estimates. This type of measure is commonly known as sampling error. Estimates of the magnitude of these sampling errors based on data from a single sample are provided by a statistic known as the *standard error of an estimate*. There are two different types of statistics presented in this MECS: consumption estimates and population estimates. Each of these types has a different method of computing the standard error of the estimate.

Standard Errors of Consumption Statistics

Standard errors for MECS consumption estimates are directly computed from the reported data by using the formula:

$$S_{\hat{Y}} = \sqrt{\sum_{i=1}^{n} y_i^2 (W_i)(W_i - 1)}\,, \qquad \textbf{(1)}$$

where $\hat{Y} = \Sigma(y_i W_i)$ is the MECS survey estimator of weighted values reported by the i^{th} MECS sample establishment, y_i is the reported value of characteristic Y for the i^{th} MECS sample establishment, W_i is the final adjusted weight used to inflate

[44]*Energy Information Administration, Manufacturing Energy Consumption Survey: Methodological Report 1985*. Although this report describes data quality in the 1985 MECS, much of the discussion still holds for the 1994 MECS.

the sample data to population estimates, and n is the number of MECS respondents. Justification for this formula is found in the MECS methodological report.

Standard Errors of Population Statistics

Standard errors for MECS population estimates are computed directly from the reported data by using the formula[45]:

$$S_{\hat{Y}} = \sqrt{\sum_{i=1}^{n} w_i(w_i - 1)(u_i - p)^2}\,, \qquad \textbf{(2)}$$

where $\hat{Y} = \Sigma(y_i W_i)$ is the MECS survey estimator, w_i is the adjusted sample weight for establishment I, n is the number of respondents in the sample, u_i is equal to 1 if it is in the domain of interest or 0 otherwise for the i^{th} MECS sample establishment, and p is the proportion of the population that has the characteristic of interest

$$p = \frac{\sum_{i=1}^{n} u_i w_i}{\sum_{i=1}^{n} w_i}\,. \qquad \textbf{(3)}$$

Estimates of standard errors have been computed from the MECS sample data for the estimated aggregate values and selected ratios appearing in this report. In the 1985 and 1988 MECS reports, measures of precision were presented separately in the form of relative standard errors (RSE), that is, the standard error divided by the estimated value to which it refers. In this report, as was the method used in the 1991 report, computed RSEs for Tables A1 through A44 are approximated in a two-factor model and are embedded into each table as "row and column factors."

Sampling Error from Generalized Variances

The RSEs computed by using standard errors from equations (**1**) and (**2**) may be efficiently modeled by a generalized variance procedure, which has been successfully used in several complex sample surveys conducted by EIA. This procedure provides a comprehensive means of reporting generalized relative standard errors, which minimizes the publishing space required to present standard errors and eases the reader's use of precision measures. Actual RSEs (by equations **1** and **2**) are used for statistical tests and confidence intervals presented in the text and for determining if a population estimate is too imprecise to publish (RSE greater than 50 percent).

The estimator used to approximate RSEs is based on a two-factor model. This model-based estimator is given as

$$R\hat{S}E_{i,j} = R_i \bullet C_j\,, \qquad \textbf{(4)}$$

where R_i is the row factor for the i^{th} row and C_j is the column factor for the j^{th} column used to compute the generalized RSE of the sample estimate at the intersection of the i^{th} row and j^{th} column. Since RSEs calculated by this generalized variance technique are approximate, confidence intervals and statistical tests of significance must also be regarded as approximate. See Table C1 for a specific example of computing an approximate RSE. The three boxes that follow give examples of calculating the approximate standard error, calculating the confidence range of the estimate, and measuring statistical significance between any two estimates.

[45]Steven K. Thompson, *Sampling*, (New York: John Wiley & Sons, Inc. 1992), pp. 69-70.

Table C1. Calculation of Generalized Relative Standard Error (RSE) Estimates Taken from Table A1: Total First Use of Energy for All Purposes, 1994: Part 2

(Estimates in Btu or Physical Units)

SIC Code	Industry Group and Industry	Total (trillion Btu)	Net Electricity (million kWh)	Residual Fuel Oil (1000 bbl)	Distillate Fuel Oil (1000 bbl)	Natural Gas (billion cu ft)	LPG (1000 bbl)	Coal (1000 short tons)	Coke and Breeze (1000 short tons)	Other (trillion Btu)	Shipments[g] (trillion Btu)	RSE Row Factors
	RSE Column Factors:	**0.5**	**0.5**	**1.2**	**1.1**	**0.6**	**1.1**	**1.2**	**1.5**	**1.1**	**1.8**	
		Total United States										
20	Food and Kindred Products	1,193	58,004	4,785	3,181	613	W	7,500	W	141	0	**5.7**
2011	*Meat Packing Plants*	53	3,924	138	101	35	W	W	0	2	0	**16.0**
2033	*Canned Fruits and Vegetables*	51	1,432	241	119	42	W	W	0	*	0	**15.7**
2037	*Frozen Fruits and Vegetables*	42	2,901	204	62	27	76	0	0	2	0	**18.0**
2046	*Wet Corn Milling*	173	5,662	169	10	67	9	3,556	6	5	0	**20.4**
2051	*Bread, Cake, and Related Products*	37	2,436	Q	131	26	89	0	0	*	0	**21.8**
2061	*Cane Sugar, Except Refining*	111	W	W	220	2	W	W	0	105	0	**19.1**
2061	*Cane Sugar Refining*	23	125	313	97	16	4	0	0	4	0	**29.6**
2063	*Beet Sugar*	64	425	270	W	18	W	1,790	W	*	0	**3.5**
2075	*Soybean Oil Mills*	57	1,845	147	W	30	W	682	0	3	0	**1.4**
2082	*Malt Beverages*	51	2,311	W	21	21	W	789	0	*	0	**8.1**
21	Tobacco Products	W	842	133	W	W	W	W	0	W	0	**25.5**
22	Textile Mill Products	310	32,614	2,680	1,274	113	999	1,821	0	14	0	**13.1**
23	Apparel and Other Textile Products	W	7,735	W	106	25	W	W	0	W	0	**20.9**
24	Lumber and Wood Products	491	19,836	389	4,314	47	W	W	0	341	0	**9.3**
2421	*Sawmills and Planing Mills, General*	201	6,556	W	1,206	11	W	0	0	160	0	**14.3**
2436	*Softwood Veneer and Plywood*	74	2,517	Q	251	3	168	0	0	61	0	**15.3**
2493	*Reconstituted Wood Products*	98	4,453	198	128	17	W	W	0	60	0	**14.2**
25	Furniture and Fixtures	69	6,590	60	154	23	211	115	0	18	0	**13.4**
2511	*Wood Furniture, Except Upholstered*	24	2,146	47	62	2	59	56	0	13	0	**15.6**
26	Paper and Allied Products	2,665	65,479	27,444	1,564	558	1,334	13,812	0	1,373	0	**3.6**
2611	*Pulp Mills*	W	2,190	3,583	178	21	W	328	0	W	0	**13.5**
2621	*Paper Mills*	1,297	34,419	14,942	746	264	476	8,783	0	612	0	**5.5**
2631	*Paperboard Mills*	954	13,512	7,914	314	194	119	4,552	0	556	0	**2.9**
27	Printing and Publishing	112	17,409	W	264	46	W	0	0	2	0	**10.6**
28	Chemicals and Allied Products	5,328	152,482	17,457	2,324	2,495	435,926	13,239	449	442	166	**6.4**
2812	*Alkalies and Chlorine*	135	13,424	W	53	52	W	W	0	16	0	**9.1**
2813	*Industrial Gases*	104	23,525	0	W	23	W	15	29	1	W	**25.9**
2816	*Inorganic Pigments*	51	2,393	159	W	21	W	W	W	11	0	**10.1**
2819	*Industrial Inorganic Chemicals, nec*	377	42,239	W	264	145	W	W	352	28	*	**14.6**
2821	*Plastics Materials and Resins*	642	16,408	542	153	234	89,084	875	0	26	21	**9.1**
2822	*Synthetic Rubber*	102	2,276	W	W	53	6,899	190	0	9	0	**26.2**
2823	*Cellulosic Manmade Fibers*	28	419	0	23	W	W	W	0	W	0	**1.2**
2824	*Organic Fibers, Noncellulosic*	116	7,093	1,435	88	39	W	W	0	W	W	**8.3**
2861	*Gum and Wood Chemicals*	28	211	*	W	4	W	164	W	19	0	**7.4**
2865	*Cyclic Crudes and Intermediates*	206	4,789	W	204	98	W	152	0	35	0	**19.5**
2869	*Industrial Organic Chemicals, nec*	2,369	18,786	795	319	1,037	282,371	W	W	250	142	**9.8**
2873	*Nitrogenous Fertilizers*	622	3,817	0	30	589	4	0	0	2	0	**14.0**
2874	*Phosphatic Fertilizers*	46	1,131	W	W	13	3	W	0	25	0	**6.4**
2895	*Carbon Black*	80	W	9,344	W	19	W	0	0	*	0	**13.3**

Source: Energy Information Administration, 1994 Manufacturing Energy Consumption Survey.

Calculating the Approximate Standard Error

RSE Column Factor Natural Gas)
= 0.6
RSE Row Factor (Chemicals and Allied Products)
= 6.4
Approximate RSE (Chemicals and Allied Products, Natural Gas)
= 6.4 · 0.6
= 3.8 percent
Approximate Standard Error (Chemicals and Allied Products, Natural Gas)
= (0.038) · (2,495 billion cubic feet) = 95 billion cubic feet.

Calculating the Confidence Range with Generalized RSEs

Steps to calculate the 95-percent confidence range (that range which includes the true value of the estimate with 95 percent confidence):

1. Multiply the standard error by 1.96;
2. Subtract the result of Step 1 from the given estimate to determine the lower bound of the range;
3. Add the result of Step 1 to the given estimate to determine the upper bound of the range.

Measuring Statistical Significance with Generalized RSEs

Steps to determine if the difference between any two estimates in this report are statistically significant:

1. Calculate the standard error of each estimate;
2. Square the standard error of each estimate;
3. Add the two values from Step 2;
4. Take the square root of the value in Step 3;
5. Multiply the value in Step 4 by 1.96;
6. If the value in Step 5 is less than the difference between the estimates, the difference between the estimates is statistically significant according to the generalized RSEs.

Derivation of Row and Column Factors

Row and column factors are derived by an analysis of variance procedure with the table of RSEs. Although analysis of variance is used to derive row and column effects from which row and column factors are computed, this generalized variance procedure can not be considered an analysis of variance because the primary concern here is to determine model fit rather than to analyze the effects of row and column variables on the RSEs. The two-way model is fit separately for each log transformed RSE table and is consistent for every table in this report. Because of this consistency over all tables, the model can be written in general format as

$$\log(RSE_{i,j}) = m + r_i + c_j + e_{i,j} \,, \qquad \textbf{(5)}$$

where m is the grand mean of $\log(RSE_{i,j})$ of a "balanced" table composed of I non-zero rows and J non-zero columns, r_i is the effect of the i^{th} row, c_j is the effect of the j^{th} row, and $e_{i,j}$ is the error term. Model parameters are fit by the standard formulas for Ordinary Least Squares given by Cochran and Cox.[46] For a given table of log(RSE) estimates, point estimators of model parameters are given as

[46]William G. Cochran and Gertrude M. Cox, *Experimental Design* (2nd ed.),(New York: John Wiley & Sons, Inc. 1957).

$$\hat{m} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\log(RSE_{i,j})}{I \bullet J} = \overline{\log(RSE_{\bullet,\bullet})}$$

$$\hat{r}_i = \frac{\sum_{j=1}^{J}\log(RSE_{i,j})}{J} - \hat{m} = \overline{\log(RSE_{i,\bullet})} - \hat{m} \tag{6}$$

$$\hat{c}_j = \frac{\sum_{i=1}^{I}\log(RSE_{i,j})}{I} - \hat{m} = \overline{\log(RSE_{\bullet,j})} - \hat{m}\,.$$

The row and column factors are then computed by back-transforming the estimated model parameters; that is, by taking the $\log^{-1}$ of the model effects. This transformation yields

$$R_i = \log^{-1}(\hat{m} + \hat{r}_i) = \log^{-1}\left(\overline{\log(RSE_{i,\bullet})}\right)$$

$$C_j = \log^{-1}(\hat{c}_j) = \log^{-1}\left(\overline{\log(RSE_{\bullet,j})} - \overline{\log(RSE_{\bullet,\bullet})}\right)\,. \tag{7}$$

For ease of presentation, the row factor includes the grand mean, m. Because of this factoring, the row factor for the i^{th} row can alternately be expressed as the geometric mean of i^{th} row:

$$R_i = \left(\prod_{j=1}^{J} RSE_{i,j}\right)^{\frac{1}{J}}\,. \tag{8}$$

And, column factors, C_j, for a given table have a geometric mean equal to 1.0.

Since the MECS report presents a variety of energy-related estimates that are unique to certain industries, measures of the precision of population estimates are sometimes equal to zero or are withheld from publication. When an RSE table contains a zero or withheld RSE, the table of RSEs is considered for generalization purposes to be "unbalanced." When the condition of an "unbalanced" table arises, substitute RSE estimates are inserted for these missing elements of the RSE table. Substitution of missing RSEs elements is based on an iterative procedure developed by Cochran and Cox.[47] A detailed description of the automated procedure used to produce the row and column factors appearing in this report can be found in Gargiullo and Goldberg.[48]

Sampling Error of Proportions

The estimates in this report can be used to produce proportion statistics based on the ratio of various estimates reported in the tables. Proportions are not given in the "Detailed Statistics Tables" but can be used to clarify the analysis. A proportion is the statistic of the form

$$\hat{P} = \frac{\hat{Y}}{\hat{X}}\,, \tag{9}$$

[47] Ibid.

[48] P.M. Gargiullo and M.L Goldberg, "A Modified Table Producing Language (TPL) for Producing Tables of Survey Statistics with Variances," Proceedings of the Bureau of the Census Fifth Annual Research Conference (1989).

where $\hat{Y}$ and $\hat{X}$ are survey-based estimates of aggregate parameters Y and X, respectively, and characteristic X "encompasses" characteristic Y ($Y \subset X$). That is, each population element (and, thus, each sample case) that contributes to Y also contributes to X, and the value of X for each element is greater than or equal to the value of Y.

From standard errors given by equation (**1**) that are then generalized by equation (**4**), the approximate RSEs of aggregate statistics can be used to produce an upper bound on the approximate errors for proportions. The straightforward additive error formula shown in equation (**1**) gives rise to a similarly straightforward upper-bound approximation to the error of an estimated proportion. The approximation can be expressed in terms of the generalized RSEs of the aggregate statistics entering into the proportion as

$$\hat{RSE}(\hat{P}) \leq \sqrt{[\hat{RSE}^2(\hat{Y}) \bullet (1 - 2 \bullet \hat{P})] + \hat{RSE}^2(\hat{X})} \,. \qquad \textbf{(10)}$$

Justification for this formula is found in the MECS methodological report.

Sampling Error of Average Values

Estimates in this report can be used to produce ratio statistics when the sample case characteristic (y) is not found in every sample case (x) (i.e. Y is not a subset of X). The ratio is the statistic of the form

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} \,, \qquad \textbf{(11)}$$

where $\hat{Y}$ and $\hat{X}$ are survey-based estimates of average parameters Y and X, respectively, and characteristic X does not "encompass" characteristic Y. That is, each population element (and, thus, each sample case) that contributes to Y may not contribute to X.

From standard errors given by equation (**1**) that are then generalized by equation (**4**), the approximate RSEs of average statistics can be used to produce an upper bound on the approximate errors for proportions. An approximation of the RSE can be expressed by entering the average statistics in the ratio as[49]

$$RSE(\hat{R}) \doteq \sqrt{[RSE^2(\hat{Y}) + RSE^2(\hat{X}) - 2RELCOV(\hat{Y},\hat{X})]} \,, \qquad \textbf{(12)}$$

where RELCOV(y, x) is assumed to be zero. Hence, the RSE(r) is overestimated.

Nonsampling Errors and Bias

Nonsampling errors that affect MECS sample units can be divided into four major categories:

- Operational errors, including editing, coding, and tabulation errors
- Errors of measurement, including a lack of precision by the respondent, failure of the respondent to understand instructions, etc.
- Errors of estimation, including the assumptions underlying the derived values
- Errors of nonobservation, including unit and item nonresponse and noncoverage.

These errors are collectively referred to as nonsampling errors because they are not related to the sampling process and, thus, would be equally likely to occur in a complete census or a sample survey.

[49]M. Hansen, W. Hurwitz, and W. Madlow, "Sample and Survey Methods and Theory, Volume I" (New York: John Wiley & Sons, Inc., 1953), p. 166.

Operational Errors

It is felt that operational errors are not a major concern for the estimates included in this report. The quality control procedures that were employed for check-in, editing, coding, and keying the returned questionnaires (Appendix B) are standard procedures that are in place at the Bureau of the Census and have withstood the test of time. Data tabulations were independently verified by comparing marginal totals in tables generated from files supplied to EIA with corresponding totals generated directly from microdata files maintained at the Census Bureau.

Errors of Measurement

Errors of measurement are a concern in any data collection activity. The survey results for the MECS were subjected to extensive computer editing procedures which were specifically designed to detect errors of measurement. Establishments that failed these tests for response reasonableness and consistency were contacted again by analysts familiar with manufacturing processes and energy use. Major errors, including omissions and misreporting by orders of magnitude, were corrected. No editing procedure is capable of identifying all measurement errors, however, and some small errors will remain. To the extent that these errors are due to random, rather than systematic, misjudgments, they are compensated in the aggregate totals presented in this report, and it is believed that there are few large systematic biases that result from them.

Errors of Estimation

Errors of estimation of energy consumption could have resulted from the assumptions that underlie the derived values (see Appendix B), and the estimates of the consumption of onsite- and offsite-produced fuels and raw material inputs could be biased as a result of such errors. For example, the derivation logic makes the assumption that energy produced onsite at a manufacturing establishment is considered first as a shipped product, second as a feedstock, and lastly as a fuel. If that logic does not hold, derived estimate values will be misapportioned. However, considering the mechanisms required to produce energy onsite, it is highly probable that this logic accurately represents manufacturers. These nonsampling errors, if present, are relevant only for tables in this report that are based on derived values. Estimates based upon reported values would not be subject to this potential source of bias.

Errors of Nonobservation

Finally, several potential sources of nonsampling error and bias result from errors of nonobservation. As described in Appendix B, the 1994 MECS represents, in terms of sampling coverage, the mail frame of the 1994 ASM or 98 percent of the manufacturing universe, which is consistent with the 1991 MECS.

Unit Nonresponse

Even though the MECS is a legislatively mandated survey and sampled establishments are given sufficient opportunity and time to respond, nonresponse occurs in the MECS and is accounted for in a nonresponse adjustment of sampling weights presented in Appendix B. Clearly, had these adjustments not been performed, the estimates produced from only the responding establishments would not have been representative of the target universe for the MECS. Such estimates would have been biased. Adjusting the sampling weights to reflect the target universe is an attempt to mitigate the potential effects of such a bias.

Adjustment factors are calculated for each of the 72 published strata to account for the variation of nonresponse between strata. Each stratum represents a relatively homogeneous subgrouping of establishments with respect to primary product output and level of fuel consumption.

Implicit in that procedure is the assumption that primary product output and level of fuel consumption are highly correlated with energy expenditure patterns, so that the establishments within a stratum would also be homogeneous with respect to the quantities, types, and shares of energy consumed as fuels and for nonfuel purposes. Also, the weight adjustment method assumes that the relationship between survey variables of interest and the control variable used for constructing the adjusted sample weight is the same for the population covered by MECS respondents within an adjustment stratum as it is for the rest of the population within that stratum.

To the extent that the nonresponding establishments within the adjustment stratum share the energy expenditure patterns of the responding establishments within the strata, the resulting adjustments to the MECS estimates will tend to be minimally biased. If, on the other hand, the energy expenditure patterns of the responding and nonresponding establishments differ substantially, the resulting adjustments are potentially biased, and the overall estimates may not accurately represent the originally targeted MECS universe.

Item Nonresponse

Item nonresponse is the type of nonresponse that occurs when an item (or several items) is missing in an otherwise completed questionnaire. In 1994, MECS expanded its collection of establishments' characteristics. MECS now collects economic information, such as floorspace, motor purchase evaluations, reasons for fuel switching, energy management activities, and energy efficient technologies. Although a Response Analysis Survey of 1991 respondents indicated that a record-keeping system tracked these items, some establishments did not have a record-keeping system which would have enabled them to respond to these types of questions. To the extent that information systems excluded economic items, the MECS incurred item nonresponse.

Some surveys impute values for item nonresponse. The MECS did not impute for most of these missing items. The complexities and inherent heterogeneity of manufacturing establishments prevent the use of imputation techniques for most of these missing items. However, budget restrictions and timeliness issues halted respondent recontacts before item nonresponse could be eliminated. Hence, tables in this report have nonresponse columns or rows to reveal the extent of item nonresponse; for example, the floorspace data in Table A7.

MECS has a full reporting of energy consumption from all responding establishments. However, economic variables (such as value of shipments and value added) have historically been obtained from the Annual Survey of Manufactures (ASM). The change in sampling frames from the ASM to the CM had an indirect result on item nonresponse; that is, some establishments in the MECS sample did not have economic data (value of shipments and value added) because they were not included in the 1994 ASM sample. One approach considered by EIA was to query these establishments on their shipments. Unfortunately, duplication issues with the ASM precluded the MECS from querying establishments on their shipments and value added. Hence, these establishments did not have the ability to report their economic data.

Unlike the establishment characteristics, like floorspace, value of shipments and value-added estimates were imputed for all MECS establishments that were not in the sample. Of the 19,292 MECS sample respondents, 13,011 cases have ASM-reported value of shipments data and ASM-reported value-added data. Of the remaining cases, we only have the total number of employees and payroll from the Standard Statistical Establishment List (SSEL). Using aggregate payroll data, by industry (at the four-digit SIC level), value of shipments and value added were imputed by forward indexing. Using value added as an example, this imputation technique is expressed as

$$VA_{imp} = \left(\frac{SW_{94}}{SW_{92}}\right) \cdot (VA_{92}) \ , \tag{13}$$

where SW_{94} is payroll from the 1994 SSEL and SW_{92} and VA_{92} are payroll and value added from the 1992 Census of Manufactures (CM). Value of shipments was imputed analogously.

The imputation model was evaluated by testing how well it predicted known 1994 ASM value-added data. Matching was done for SIC industries 2000 through 3999. If an establishment was indicated as out-of-business, the observation was deleted. If the ratio SW_{94}/SW_{92} was less than 1/8 or more than 8, the data were considered either as outliers or as recording errors. Such observations (696 out of 52,227) were removed prior to analysis.

A paired comparison t-test was done for the percent difference between VA_{imp} and actual VA_{94} to determine if the mean percent difference was statistically significant from zero. At a 95-percent confidence level, the null hypotheses (the mean percent difference between VA_{imp} and actual VA_{94} is zero) could not be rejected. Hence, there was no difference statistically between Va_{imp} and VA_{94}. A t-test using value of shipments yielded the same conclusion.

Table C2 provides the counts, percents, and weighted percents for the imputed estimates of value of shipments and value added by industry.

More detailed information on sources of nonsampling error in the MECS can be found in the methodological report.

Table C2. Number of Establishments and Weighted and Unweighted Percentages of the Imputed Value of Shipments and Value Added

SIC Code	Industry Groups and Industry	Value of Shipments: Number of Establishments	Value of Shipments: Percent	Value of Shipments: Weighted Percent	Value Added: Number of Establishments	Value Added: Percent	Value Added: Weighted Percent
2011	Meat Packing Plants	10	1.2	4.2	10	1.1	4.6
2033	Canned Fruits and Vegetables	47	10.1	19.2	47	9.5	17.7
2037	Frozen Fruits and Vegetables	18	5.2	21.3	18	3.0	17.7
2046	Wet Corn Milling	12	3.6	3.6	12	3.2	3.2
2051	Bread, Cake, and Related Products	24	2.3	5.7	24	2.0	5.3
2061	Cane Sugar, Except Refining	26	36.7	36.7	26	35.1	35.1
2062	Cane Sugar Refining	4	6.8	6.8	4	7.0	7.0
2063	Beet Sugar	4	W	W	4	W	W
2075	Soybean Oil Mills	10	2.4	2.4	10	3.0	3.0
20	Balance of Food and Kindred Products	322	6.1	17.4	322	3.9	13.2
21	Tobacco Products	7	1.3	1.5	7	1.2	1.4
22	Textile Mill Products	139	6.3	14.5	139	6.7	18.0
23	Apparel and Other Textile Products	247	7.9	17.2	247	8.7	21.9
2421	Sawmills and Planing Mills, General	177	19.0	38.3	177	20.9	48.1
2436	Softwood Veneer and Plywood	5	2.8	3.9	5	2.8	4.0
2493	Reconstituted Wood Products	13	5.9	13.4	13	4.5	11.0
24	Balance of Lumber and Wood Products	313	14.8	42.5	313	17.2	49.7
2511	Wood Furniture, Except Upholstered	38	10.3	32.7	38	8.9	27.1
25	Furniture and Fixtures	131	9.3	35.2	131	9.0	37.7
2611	Pulp Mills	9	6.5	6.5	9	5.6	5.6
2621	Paper Mills	61	4.2	4.2	61	4.2	4.2
2631	Paperboard Mills	64	6.0	6.0	64	7.3	7.3
26	Balance of Paper and Allied Products	240	10.5	24.6	240	10.0	23.9
27	Printing and Publishing	612	9.9	35.2	612	8.7	32.9
2812	Alkalies and Chlorine	18	24.8	24.8	18	33.2	33.2
2813	Industrial Gases	24	9.4	11.8	24	8.8	10.8
2816	Inorganic Pigments	33	12.0	12.0	33	11.3	11.3
2819	Industrial Inorganic Chemicals, nec.	42	6.6	13.6	42	3.6	9.7
2821	Plastics Materials and Resins	22	4.8	5.2	22	3.8	4.4
2822	Synthetic Rubber	15	6.6	6.6	15	6.1	6.1
2823	Cellulosic Manmade Fibers	2	W	W	2	W	W
2824	Organic Fibers, Noncellulosic	16	3.3	3.3	16	2.8	2.8
2861	Gum and Wood Chemicals	19	12.9	12.9	19	12.8	12.8
2865	Cyclic Crudes and Intermediates	17	7.9	10.2	17	9.2	10.8
2869	Industrial Organic Chemicals, nec.	46	6.1	8.7	46	6.9	9.9
2873	Nitrogenous Fertilizers	54	19.9	19.9	54	17.2	17.2
2874	Phosphatic Fertilizers	27	17.5	17.5	27	3.5	3.5
287	Balance of Agricultural Chemicals	9	1.9	19.4	9	1.0	7.6
2895	Carbon Black	7	W	W	7	W	W
28	Balance of Chemicals and Allied Products	114	2.1	11.8	114	1.6	9.5
2911	Petroleum Refining	46	3.0	3.0	46	3.9	3.9
29	Balance of Petroleum and Coal Products	158	22.1	29.3	159	24.1	34.1
3011	Tires and Inner Tubes	1	W	W	1	W	W
308	Miscellaneous Plastics Products, nec.	353	8.8	23.8	353	8.5	24.0
30	Balance of Rubber and Misc. Plastics Products	135	15.0	39.6	135	13.8	37.4
31	Leather and Leather Products	48	9.1	29.9	48	9.8	31.5
3211	Flat Glass	11	1.9	1.9	11	2.3	2.3
3221	Glass Containers	2	W	W	2	W	W
3229	Pressed and Blown Glass, nec.	4	0.9	3.3	4	1.0	3.6
3231	Glass Products Made from Purchased Glass	28	5.4	15.0	28	4.7	5.3
3241	Cement, Hydraulic	23	16.3	17.6	23	14.5	15.7
3274	Lime	57	77.1	77.1	57	77.4	77.4
3296	Mineral Wool	2	W	W	2	W	W
32	Balance of Stone, Clay, and Glass Products	384	19.4	44.8	384	18.7	46.6
3312	Blast Furnaces and Steel Mills	59	2.7	2.7	59	2.0	2.0
3313	Electrometallurgical Products	17	61.6	61.6	17	73.5	73.5
331	Balance of Blast Furnace and Basic Steel Products	49	12.5	35.0	49	14.3	42.3
3321	Gray and Ductile Iron Foundries	34	7.0	17.5	34	7.1	18.3

See footnotes at end of table.

Table C2. Number of Establishments and Weighted and Unweighted Percentages of the Imputed Value of Shipments and Value Added (Continued)

SIC Code	Industry Groups and Industry	Value of Shipments			Value Added		
		Number of Establishments	Percent	Weighted Percent	Number of Establishments	Percent	Weighted Percent
3331	Primary Copper	3	W	W	3	W	W
3334	Primary Aluminum	3	W	W	3	W	W
3339	Primary Nonferrous Metals, nec	38	16.2	16.7	38	13.9	13.3
3353	Aluminum Sheet, Plate, and Foil	12	2.5	2.5	12	2.4	2.4
33	Primary Metal Industries	218	11.0	23.5	218	10.7	25.7
34	Fabricated Metal Products	529	8.4	30.7	529	8.9	32.8
357	Computer and Office Equipment	9	0.2	5.2	9	0.2	7.1
35	Balance of Industrial Machinery and Equipment	545	5.0	27.2	545	5.6	28.9
36	Electronic and Other Electric Equipment	141	1.5	14.9	141	1.5	15.5
3711	Motor Vehicles and Car Bodies	2	W	W	2	W	W
3714	Motor Vehicle Parts and Accessories	36	1.0	8.8	36	1.2	10.1
37	Balance of Transportation Equipment	76	1.1	11.2	76	1.2	12.9
3841	Surgical and Medical Instruments	20	1.7	12.3	20	1.4	10.6
38	Balance of Instruments and Related Products	102	2.4	16.2	102	2.1	16.3
39	Miscellaneous Manufacturing Industries	141	10.2	44.2	141	9.5	43.8
	Total	6,284	4.8	17.8	6,284	4.7	19.5

W = Withheld to avoid disclosing data for individual establishments.
Source: Energy Information Administration, Office of Energy Markets and End Use, Energy End Use and Integrated Statistics Division, Form EIA-846, "1994 Manufacturing Energy Consumption Survey."

Comparability of MECS Estimates with Other Series

The Energy Information Administration (EIA) collects data from two distinct sources that, in their entirety, provide a comprehensive picture of energy production, marketing, and consumption in the United States.[50] One set of surveys is directed to the suppliers and marketers of specific fuels (including electricity). The second group of surveys collects comprehensive energy consumption and related data directly from end-use consumers. The MECS is a member of the latter group.

Because there is a seeming correspondence between energy supplied and energy consumed, it is tempting to compare or merge their results. However, there are important differences between the supplier and end-user surveys that need to be taken into account in doing such comparisons or other analyses.

An Overview of EIA Surveys

The End-User Surveys

The overall purpose of the end-user surveys is to provide comprehensive baseline data on energy consumption and related characteristics for major sectors of the U.S. economy. Accordingly, the end-user surveys are conducted for the manufacturing sector, commercial buildings, residential households, and residential transportation. These surveys collect data directly from samples of the energy-consuming units comprising those sectors. The results of these end-user surveys are available in a variety of EIA publications.

[50]Descriptions of all EIA data collection activities are included in Energy Information Administration, *Directory of Energy Data Collection Forms*, DOE/EIA-0449(90) (Washington, DC, January 1991).

The Supplier Surveys

The EIA conducts numerous supplier surveys. The overall purpose of these surveys is to measure the quantity of a specific fuel produced and/or supplied to the market, along with other information related to the fuel's production and supply. The results of these surveys are published in several EIA reports.[51]

Combined Results of the Supplier Surveys

In addition to supporting fuel-specific publications of EIA, the results of the supplier surveys are combined to produce estimates of total energy consumption by consuming sector. The consuming sectors consist of the commercial, residential, industrial, transportation, and electric utilities sectors. The resulting combined estimates are published by EIA in the *Monthly Energy Review* (MER), the *State Energy Data Report* (SEDR), and the *Annual Energy Review* (AER).

Defining the Industrial Sector

In general, the "industrial sector" is defined as consisting of manufacturing, mining, construction, agriculture, fisheries, and forestry. The approximate SIC equivalent of the industrial sector includes major group codes 01 through 39.[52] There are a few definitional irregularities, however, that preclude a perfect mapping of the supplier surveys to that range of SIC codes. As pointed out in the MER,

> . . . although end-use allocations are made according to [the sector definitions] as closely as possible, some data are collected by using different classifications. For example, data on agricultural use of natural gas are collected and reported in the commercial sector rather than the industrial sector. Since agricultural use of natural gas cannot be identified separately, it is included in the commercial sector....[rather than the industrial sector.][53]

Comparing the MECS and Industrial Sector Estimates

The MECS produces four separate estimates of manufacturing energy consumption, which are: (1) Total First Use of Energy for All Purposes (Table A1), (2) Total First Use of Energy for Nonfuel Purposes (Table A3), (3) Total Inputs of Energy for Heat, Power, and Electricity Generation (Table A4), and (4) Total Consumption of Offsite-Produced Energy for Heat, Power, and Electricity Generation (Table A5). The differences among those estimates are discussed in detail elsewhere in Appendix B of this report.

The combined estimates for the industrial sector published in SEDR are conceptually similar to the MECS estimates of Total First Use of Energy For All Purposes, because both series measure energy consumption as a fuel and as a raw material or feedstock.

A Final Observation

Many of the substantial differences between the MECS estimates of first use of energy and the combined estimates resulting from the supplier surveys can be reconciled by carefully reviewing the coverage and definitions of the data series involved. It should be emphasized that the differences are not an indication of the relative strengths or weaknesses of either series. Rather, the differences in the estimates simply reflect the differences in the *intents* of the end-user surveys and the supplier surveys. The overall purpose of the end-user surveys is to provide baseline energy consumption and related characteristics data for various groups of end users (manufacturers, residential housing and transportation, and commercial buildings). The overall purpose of the supplier surveys, on the other hand, is to provide baseline data on the production and supply of various fuels. To reiterate, data users should be extremely wary of attempting to compare or combine the results of the end-user and supplier surveys without careful attention to the origins and purposes of the different estimates. The details of a study comparing the 1991 MECS and Industrial Sector estimates can be found in Appendix D in *Manufacturing Energy*

---

[51]For a complete list of publications, see Energy Information Administration, *EIA Publications Directory 1996*, DOE/EIA-0149(96) (Washington, DC, May 1997).

[52]See Office of Management and Budget, *Standard Industrial Classification Manual 1987* (Washington, DC, 1987).

[53]Energy Information Administration, *Monthly Energy Review, September 1993*, p. 40.

Consumption Survey, 1991 (DOE/EIA-0512(91)). The values are somewhat different in 1994 than in 1991, but the major relationships and their explanations remain appropriate.