An Evaluation of the 2002 Manufacturing Energy Consumption Survey Measure of Size

William R. Gifford[1]
Energy Information Administration[1]

Abstract

The Manufacturing Energy Consumption Survey (MECS) is an establishment survey which collects data on energy consumption, expenditures and other energy related topics from the manufacturing sector. The MECS frame is based off of the business register, which is managed by the US Census Bureau. The MECS takes advantage of the establishment-level information that comes with sharing a frame with the manufacturing section of the Economic Census. While the MECS is fortunate to have a measure of size assigned to each element in its frame based on data recorded from those actual elements, analysis is needed for understanding how the measure of size compares with the energy consumption measures collected in the MECS. In this paper the 2002 MECS measure of size will be compared with consumption measure values graphically.

Keywords: measure of size evaluation, Manufacturing Energy Consumption Survey

1. Introduction

The Manufacturing Energy Consumption Survey (MECS) is an establishment survey which collects data on energy consumption, expenditures, and other energy related information from US manufacturing establishments quadrennially. The MECS is a product of the Energy Information Administration (EIA), the independent statistical branch of the US Department of Energy.

The MECS takes advantage of the establishment-level information that comes with sharing a frame with the manufacturing section of the Economic Census. Cost of electricity and cost of fuels is recorded for all establishments in the MECS frame that respond to the Economic Census. The MECS uses the sum of these two costs as the measure of size for establishments or it is imputed*. The MECS then assigns a probability of inclusion to the sample that is proportional to the measure of size for each establishment.

2. Motivation for a Measure of Size Study

Analysis is needed to evaluate how the measure of size relates to the measures of energy consumption collected. Without understanding how they relate to each other there is risk that MECS consumption statistics are flawed. This could occur if weights have been assigned incorrectly to the establishments. The MECS uses modified Horovitz-Thompson Estimators[2], for which sampled units are weighted by the inverse of their inclusion probabilities. These probabilities are constructed from the establishment measures of size. It is desired that for each pair of establishments, (i,j) within stratum k of the MECS frame and for energy consumption measure C,

$$\frac{MOS(i,k)}{C(i,k)} = \frac{MOS(j,k)}{C(j,k)} \qquad (1)$$

where MOS(i,k) and C(i,k) are the measure of size and consumption value for establishment i.
The assigned weight for establishment i in stratum k is a modified version[2] of the following weight formula:

$$w(i,k) = \frac{\sum_{j\in\Omega_k} MOS(j,k)}{MOS(i,k)} \qquad (2)$$

If (1) is true, then

$$w(i,k) = \frac{\frac{MOS(i,k)}{C(i,k)}\sum_{j\in\Omega_k} C(j,k)}{MOS(i,k)} = \frac{\sum_{j\in\Omega_k} C(j,k)}{C(i,k)} \qquad (3)$$

which is the "correct" weight for establishment i in stratum k for consumption measure C. Of course such a perfect measure of size does not exist, so the goal is for

$$\frac{MOS(i,k)}{C(i,k)} \approx \frac{MOS(j,k)}{C(j,k)} \qquad (4)$$

If (4) is true, then if C(i,k) were to be plotted against the MOS(i,k) for the set of establishments in stratum k, the resulting scatter plot would show an approximately linear relationship between the consumption measure and the measure of size.

If this study reveals that this is not the case, it would be desired to find a function, *f*, such that

$$\frac{f(MOS(i,k))}{C_k(i,k)} \approx \frac{f(MOS(j,k))}{C_k(j,k)} \qquad (5)$$

This function could be applied to the original measure of size for each establishment and the resulting set of values could serve as the new measures of size for the establishments in stratum k.

3. Methods

The first step will be to look at measure of size plotted against the unweighted consumption measure values for the 2002 MECS. Looking at the relationship graphically will either confirm that there is a linear relationship or if not, it will hopefully help in determining the kind of transformation that will be needed to achieve a linear relationship. Separate scatter plots will be generated for each {fuel, industry group} combination that will be considered.

Energy Sources	Industry
Net Electricity	Food
Residual Fuel Oil	Paper
Distillate Fuel Oil	Petroleum and Coal
Natural Gas	Chemicals
LPG/NGL	Primary Metals
Coal	Other Manufacturing
Combined Other	Industries
Total Energy	Total Manufacturing

A total of (8 energy sources) x (7 industry groups) = 56 scatter plots will be generated.

Each scatter plot will be of establishments at the national level, rather than by Census Region. The 2006 MECS sample will not be stratified by Census Region, so looking at the plots at a national level will be more useful in designing the 2006 MECS. Restricting to the national level will also reduce the number of plots to look at. It is understood that the cost of energy varies by region. In combining the regions the following problem can occur: Suppose establishments A and B are in the same industry group but different regions. Suppose they consume exactly the same amount of fuel. If fuel costs more for establishment A than for establishment B, then the MOS for establishment A will be greater than the MOS for establishment B. While this is certainly not a desired result, it just means that establishments in regions where energy is more expensive than in other regions will be assigned slightly higher initial probabilities of selection. This translates to establishments in the Northeast Census Region will be slightly more represented in the 2006 MECS than they should. Since there is no perfect set of adjustment factors to "level" the regional differences in cost of energy, no regional adjustments will be done.

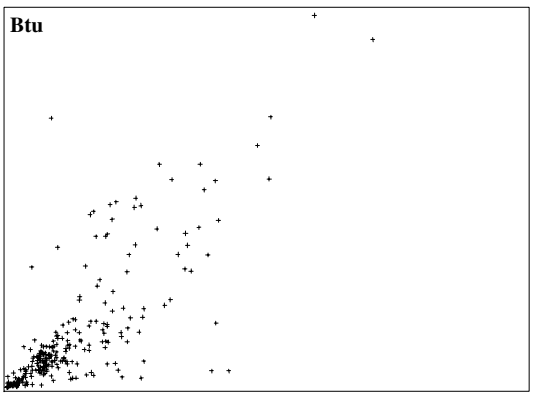
It is expected that the MOS for all industry group by fuel combinations will be highly correlated with both net electricity and natural gas. This is because the MOS = cost of electricity + cost of fuel for each establishment, with those costs coming from the most recent Annual Survey of Manufacturers or Economic Census.

For other combinations, it is hoped that there is a linear relationship between the MOS and the consumption values, but it is expected that this will not be the case for all combinations. For plots with nonlinear relationships, the shape of the plots will be examined and we will look for commonality. Transformations on the MOS will be considered to linearize these nonlinear relationships. If alternative establishment-level data is available that could be used as a MOS in certain cells, it will be considered.

4. Results

As expected, there was an approximately linear relationship between the MOS and Btu of natural gas used as a fuel, net electricity, and total fuels for each of the industry groups considered. The following scatter plot is of total Btu consumed as a fuel by MOS for an undisclosed industry group.

Figure 1. Fuel Use by Measure of Size for Industry group Z, All fuels



This is a typical example from this group of plots. While it is certainly not a perfect linear relationship, it is clear that Btu consumed is positively correlated with MOS.

For other fuels, except for some exceptions, the plots did not suggest any kind of positive linear relationship. In fact, there did not appear to be a relationship of any

kind. Each of the following figures is of the same industry x fuel type combination. The Btu range (y-axis) in Figure 3 is from the origin to the dashed line in Figure 2.

Figure 2. Fuel Use by Measure of Size for Industry Group X, Energy Source Y

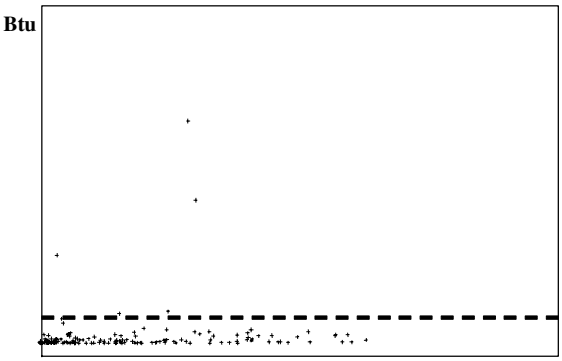
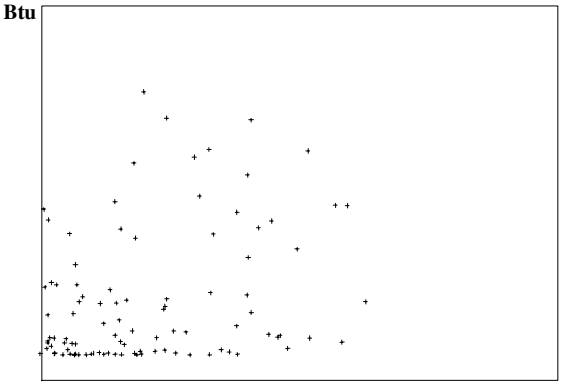


Figure 3. Fuel Use by Measure of Size for Industry Group X, Energy Source Y



The coefficients of determination, R^2, for the plots are not presented in this paper. It is clear that the slope of the regression line for Figure 1 would be significantly greater than zero. It is not clear that the slope of the regression line in Figure 2 or Figure 3 is greater than zero, but it is clear that Btu consumed is not very highly correlated with the MOS. Also, Figure 2 suggests that there would not be a constant variance in the residuals had a regression line been fitted, violating an assumption of the ordinary least squares regression model.

5. Actions Taken and Conclusions

The biggest concern with the kind of relationship seen in figure two is with establishments that have low to moderate MOS values with extremely high Btu consumed values. If these kinds of establishments are selected, their weights will be far higher than they should be, so the already extremely high consumption values may end up raising the estimate of total Btu consumed for the strata by a non-trivial amount.

For each of the fuels for which the Btu consumed in 2002 did not have a strong linear relationship with the MOS, the top consuming establishments were flagged and matched to the 2006 MECS sample frame if they were still part of that frame. Those establishments were then selected with certainty. While this in no way prevents the same undesirable situation from occurring again with the 2006 MECS sample, it does reduce the number of establishments in the frame that are likely to have high consumption values with low to moderate MOS values.

This simple evaluation of the MECS measure of size demonstrates that just because a MOS is constructed from establishment level data, it may not be correlated with the collected values from the survey for some measures. The relationship between the MOS and the measured values should be studied so that appropriate actions can be taken to keep the MOS from compromising the quality of the estimates of the survey in these situations.

References

1. William Gifford, Mathematical Statistician, Energy Information Administration, US Department of Energy, 1000 Independence Ave SW, Washington, DC 20585, email: William.Gifford@eia.doe.gov

2. 2002 Manufacturing Energy Consumption Survey Methodology and Data Quality: Survey Design, Implementation, and Estimates http://www.eia.doe.gov/emeu/mecs/mecs2002/methodology_02/meth_02.html