Bayesian Modeling of Uncertainties in End-Use Electricity Consumption Amounts Inferences

Hiroaki Minato
U. S. Energy Information Administration
1000 Independence Ave., SW, Washington, DC 20585

Abstract
Minato (2016) approached the problem of statistical calibration of energy-engineering-expert-model (E3M) estimates of residential electricity end-use consumption amounts with Bayesian multilevel models. The Residential Energy Consumption Survey (RECS), conducted by the U. S. Energy Information Administration (EIA), provided the data on building characteristics as well as energy end choices and uses. With the survey data and weather data, engineering models were formulated to estimate various end-use energy consumption amounts. However, the Bayesian multilevel models did not incorporate the engineering models' estimation errors or the administrative billing data's processing errors. In this paper, we directly model the uncertainties in those errors within the Bayesian framework. Survey weighting errors are also modeled for population inferences.

Key Words: administrative data; Bayesian multilevel measurement error models; end-use energy consumption amounts; Residential Energy Consumption Survey (RECS); Stan; survey weights

Contents

1. Introduction

When there are no direct, accurate, and affordable measurements of end-use energy consumption amounts, we are left to model and estimate the amounts somehow. The U. S. Energy Information Administration (EIA) conducts a national complex-design survey (Residential Energy Consumption Survey (RECS)), collecting data on housing unit characteristics, energy choices, end-use energy equipment, and household energy behavior from sample housing units (U. S. Energy Information Administration, 2016). From energy suppliers, requested are monthly consumption amounts and expenditures data of the survey respondents, which we annualize to produce total fuel consumption amounts for each survey respondent, e.g., annual total electricity consumed. Weather data are also acquired from NOAA (National Oceanic and Atmospheric Administration) and linked to the survey respondents via the closest weather stations (National Oceanic and Atmospheric Administration, 2016). With the survey variables in RECS and the administrative weather data, energy experts construct engineering-based models and produce end-use fuel consumption estimates for each survey respondent, e.g., annual electricity consumed for space heating by the respondent. See Minato (2016) for more details. These expert estimates of end-use fuel consumption amounts are calibrated against the total fuel consumption amount for each survey respondent. Finally, they are weighted by the survey weights to make population inferences. Here, we consider electricity and the following end uses: space heating, air conditioning (A/C), water heating, refrigerators, and others for the calendar year of 2009. The "others" include: lighting, stoves/ovens/stove-ovens, dishwashers, clothes washers and dryers, kitchen appliances, and the other electric end uses that were surveyed in the 2009 RECS. Note, therefore, that there may exist unknown or non-surveyed electricity end uses. A list of end uses covered by the 2009 RECS is given in Appendix A as well as in Minato (2016).

There are many calibration methods. The simplest is a uniform calibration model, i.e., applying one multiplicative factor to all end-use estimates so that they simply sum to the total annual energy consumption value for a given home. Minato (2016) tried a Bayesian multilevel regression modeling approach in differentially calibrating survey-weighted end-use estimates. With the varying differential multiplicative factors $\beta1$, $\beta2$, $\beta3$, $\beta4$, and $\beta5$ for space heating ($x1$), A/C ($x2$), water heating ($x3$), refrigerators ($x4$), and others ($x5$), respectively, and the varying multiplicative Census Region adjustment factor α, Minato (2016) modeled:

$$y_i \sim \text{normal}\big(\alpha_{r[i]}(\beta1_{\text{e}}x1_i + \beta2_{\text{e}}x2_i + \beta3_{\text{e}}x3_i + \beta4_{\text{e}}x4_i + \beta5_{\text{e}}x5_i), \sigma_{(r,\text{e})[i]}\big) > 0,$$

where the second parameter in the normal distribution notation is the standard deviation, i indexes the housing unit respondent, $r[i]$ indicates the Census Region that respondent i belongs to, e is one of the actual values that classify the respondents by end-use combination, and $(r,\text{e})[i]$ identifies a group of respondents that belong to the Census Region r and the end-use combination e. The Census Regions are Northeast Census Region (r = 1), Midwest Census Region (r = 2), South Census Region (r = 3), and West Census Region (r = 4). The priors were give as follows (for e = 13 as an example, which is equivalent to e = 14 in the current paper):

$$\alpha_r \sim \begin{cases} \text{Normal}(1,\ 0.5) > 0, \text{ if } r = \text{Northeast or Midwest} \\ \text{Normal}(1,\ 1) > 0, \text{ if } r = \text{South or West} \end{cases};$$
$$\beta1_{13} \sim \text{Normal}(1,\ 0.5) > 0 \text{ for } x1 \text{ (space heating) when } e = 13;$$
$$\beta2_{13} \sim \text{Normal}(1,\ 0.5) > 0 \text{ for } x2 \text{ (A/C) when } e = 13;$$
$$\beta3_{13} \sim \text{Normal}(1,\ 0.5) > 0 \text{ for } x3 \text{ (water heating) when } e = 13;$$
$$\beta4_{13} \sim \text{Normal}(1,\ 0.5) > 0 \text{ for } x4 \text{ (refrigerator) when } e = 13;$$
$$\beta5_{13} \sim \text{Normal}(1,\ 1) > 0 \text{ for } x5 \text{ (others) when } e = 13; \text{ and}$$
$$\sigma_{r,e} \sim \text{Normal}\big(0,\ |\overline{w\delta_{r,e}}|/6\big) > 0 \text{ for } r \text{ and } e \text{ pairs,}$$

where $|\overline{w\delta_{r,e}}|$ is the absolute value of the observed weighted average of $w_iy_i - w_ix1_i - w_ix2_i - w_ix3_i - w_ix4_i - w_ix5_i$ when the respondent i belongs to the end-use combination e and the Census Region r and has the survey weight of w_i.

One of the next steps suggested in Minato (2016) was the "examination of robustness against bad data, i.e., inaccurate engineering-based end-use estimates" and Minato (2016) proposed to "consider modeling the engineering-based end-use estimate errors and the survey weighting errors more directly as measurement errors".

In this paper, we are going to incorporate the energy-engineering-expert model (E3M) estimation errors and the administrative billing data processing errors. We directly model uncertainties in those errors within the Bayesian framework. Survey weights errors are also modeled for population inferences, and the modeling is done separately from the total and end-use consumption amounts modeling. To summarize, we have the following ingredients and processing steps that lead to population inferences:

I. Survey variables + Administrative weather data => E3M-ing of end-use electricity consumption amounts;
II. E3M estimates of end-use electricity consumption amounts + Administrative annualized total electricity consumption value => Bayesian calibration; and
III. Calibrated E3M end-use electricity consumption estimates + Housing unit survey weights => Population inferences of total and end-use electricity consumption amounts.

2. Ingredients to Population Inferences

For a 2009 RECS sample respondent home i (where i = 1, …, 12,083), let y_i be the raw annual total electricity consumption value, $x1_i$ be the raw annual space heating electricity consumption estimate by the space heating E3M, $x2_i$ be the raw annual A/C electricity consumption estimate by the A/C E3M, $x3_i$ be the raw annual water heating electricity consumption estimate by the water heating E3M, $x4_i$ be the raw annual refrigerators electricity consumption estimate by the refrigerator E3M, and $x5_i$ be the raw annual others electricity consumption estimate by the others E3M. The y_i turns out to be positive for all i in the 2009 RECS sample data, which is not surprising for electricity. Meanwhile, the 2009 RECS survey responses indicate that not every respondent owns and uses all of the five end uses with electricity, which is also non-surprising. For a given respondent, a consumption amount of any non-owned and non-used end use is deduced to be zero, where the information about the ownership and usage comes from the respondent's survey responses themselves. In this sense, accuracy of survey responses is critical for the existential deduction as well as for the end-use consumption estimation by the expert models.

Ideally, the sum of the annual end-use electricity consumption estimates equals the administrative annualized total electricity consumption value for each home i:

$$y_i = x1_i + x2_i + x3_i + x4_i + x5_i \text{ for each } i.$$

Further, with "correct" survey weights in terms of data-inclusion probabilities, our weighted estimators of population consumption amounts would be unbiased in the Horvitz-Thompson sense.

Realistically, however, there are many sources and types of errors in the E3M end-use electricity consumption estimates, in the annual total electricity consumption values, and in the survey weights.

First, energy-engineering experts could misspecify end-use models. Misspecification errors or approximations may arise due to the limited set of variables available from the survey, as survey interviewing may not allow too technical a question. Within the survey variables, response or measurement errors could occur, and those errors are propagated into the expert end-use models. Similarly with the administrative weather data. Weather measurement instruments may malfunction, and linking of measurement locations, i.e., weather stations, to the sample housing units is only approximate. In this paper, all these errors of different sources and types are combined as one meta-error in the expert modeling and estimation of end-use electricity consumption amounts.

Second, utility companies' administrative billing data may not be perfect, though it is expected to be quite reliable. Metering of electricity consumption amounts may not be exact. There is a chance of administrative (human, mechanical, or eletronic) errors in recording and billing. Wrong billing data for survey respondents could be sent to EIA. Further, as an *annual* total electricity consumption value is required for each survey respondent, imputation of consumption amounts for any missing bills in the year and annualization of multiple billing amounts are both necessary. Those data processing are often model-based or approximate, and the annual total electricity consumption values are subject to these processing errors as well as the administrative errors.

Finally, survey weights are prone to sampling frame errors and survey (or unit) non-responses. Because of the over- or/and under-coverage of the target population, the design weights calculated from a given sampling frame are only approximate. Ineffective survey operations could cause a large number of survey non-responses, which would further compromise validity of the weights. Adjustments of survey weights against these problems are model-based, and the models are "wrong" or at best approximate. All the weighting errors are inherited into the final survey analysis weights.

With all these errors, what we often do or can do is to ignore them or to assume they would just cancel out. However, if we have any information about the size or/and direction of errors, random or systematic, we might as well try using the information to improve our estimation or inferences. And, accepting that such information is uncertain, we try to assess and quantify the degree of uncertainty through some probability models, pursuing naturally a Bayesian approach.

3. Model Specifications and Justifications—An Internal "Preregistration" of Analysis

Schematically, we consider the values in Table 1 as our data. There are 12,083 respondents in the 2009 RECS, all of which say they use electricity for at least one of the end uses: space heating, A/C, water heating, refrigerators, and others. We partially pool them by the so-called end-use combination (indexed by e), which is defined by the existence and absence of the five electricity end uses in Table 2. Note that $2^5 - 1 = 31$ combinations are possible but only 14 are observed in the survey data. For the weights error models, we partially pool the respondents by Census Region (indexed by r), independently from the end-use combination pooling. Our model is multilevel but not hierarchical—one level by end-use group for the total, space heating, A/C, water heating, refrigerators, and others consumption amounts and another level by Census Region for the weights.

The number of respondents in e = 14 (full end-use combination) is 3,331, while that in e = 3 is 1. The numbers of respondents in e = 1, …, 6 are small, as they are those respondents who responded they did not use (electric) refrigerators, which is possible but rare in America. For illustrations, we describe our models for e = 14 in this paper. On the other hands, the numbers of respondents in Census Regions are as follows: r = Northwest: 2,266; r = Midwest: 2,843; r = South: 4,090; and r = West: 2,884.

By internal preregistration, we mean that we first specify all the models for our data and prior knowledges, including our simulation algorithms and strategies, before estimation and inferences, whose framework is also prefixed as Bayesian here.

Table 1: Schematic Data for End-Use Electricity Consumption Amounts Inferences

Housing Unit: i	*Census Region:* r	*Weight:* w	*Total Electricity Consumption:* y	*Expert End-Use Estimates*					*End-Use Combination:* e
				Space Heating: $x1$	*A/C:* $x2$	*Water Heating:* $x3$	*Refrige-rators:* $x4$	*Others:* $x5$	
1	Northeast	w_1	y_1		$x2_1$	$x3_1$	$x4_1$	$x5_1$	12
2	South	w_2	y_2	$x1_2$	$x2_2$	$x3_2$	$x4_2$	$x5_2$	14
3	South	w_3	y_3		$x2_3$	$x3_3$	$x4_3$	$x5_3$	12
4	Midwest	w_4	y_4	$x1_4$			$x4_4$	$x5_4$	7
5	West	w_5	y_5	$x1_5$			$x4_5$	$x5_5$	7
6	Northeast	w_6	y_6		$x2_6$	$x3_6$		$x5_6$	3
…	…	…	…	…	…	…	…	…	…

Table 2: Electricity End-Use Consumption Combinations

End-Use Combination: e	*Space Heating:* $x1$	*A/C:* $x2$	*Water Heating:* $x3$	*Refrigerators:* $x4$	*Others:* $x5$	*Sample Size:* n
1	0	0	0	0	1	7
2	0	0	1	0	1	2
3	0	1	1	0	1	1
4	1	0	0	0	1	2
5	0	1	0	0	1	1
6	1	1	1	0	1	6
7	1	0	0	1	1	437
8	0	0	0	1	1	1,056
9	1	1	0	1	1	1,983
10	0	0	1	1	1	244
11	0	1	0	1	1	3,784
12	0	1	1	1	1	834
13	1	0	1	1	1	395
14	1	1	1	1	1	3,331
All						*12,083*

3.1 Uncertainties in the Annual Total Electricity Consumption Amounts y_i and the Annual End-Use Electricity Consumption Estimates by the E3M estimates $x1_i$, $x2_i$, $x3_i$, $x4_i$, and $x5_i$

For each housing unit i in the illustrative end-use combination e = 14 (the latter subscript is omitted), we set up standard measurement error models (Stan Development Team, 2017b) for the administrative annual total electricity consumption amount y_i and the E3M annual end-use electricity consumption estimates $x1_i$, $x2_i$, $x3_i$, $x4_i$, and $x5_i$. Let y_i' be the true value of y_i and $x1_i', x2_i', x3_i', x4_i'$, and $x5_i'$ be the true values of $x1_i$, $x2_i$, $x3_i$, $x4_i$, and $x5_i$, respectively. The data generating models are:

$y_i \mid y_i',\ \sigma_{y_i} \sim$ Normal$(y_i', \sigma_{y_i}) > 0$, with $\sigma_{y_i} = 0.05 \times y_i'$,
$x1_i \mid x1_i',\ \sigma_{x1_i} \sim$ Normal$(x1_i', \sigma_{x1_i}) > 0$, with $\sigma_{x1_i} = 0.125 \times x1_i'$,
$x2_i \mid x2_i',\ \sigma_{x2_i} \sim$ Normal$(x2_i', \sigma_{x2_i}) > 0$, with $\sigma_{x2_i} = 0.125 \times x2_i'$,
$x3_i \mid x3_i',\ \sigma_{x3_i} \sim$ Normal$(x3_i', \sigma_{x3_i}) > 0$, with $\sigma_{x3_i} = 0.125 \times x3_i'$,
$x4_i \mid x4_i',\ \sigma_{x4_i} \sim$ Normal$(x4_i', \sigma_{x4_i}) > 0$, with $\sigma_{x4_i} = 0.125 \times x4_i'$, and
$x5_i \mid x5_i',\ \sigma_{x5_i} \sim$ Normal$(x5_i', \sigma_{x5_i}) > 0$, with $\sigma_{x5_i} = 0.25 \times x5_i'$.

The annual total electricity consumption amount y_i for a given i depends on or are conditional on two parameters and is distributed folded-normally with those parameters specifying the mean and standard deviation of the distribution. This gives us the likelihood function. However, we define the standard deviation as a function of the mean, using the coefficient of variation (CV), reducing the number of unknown parameters to one. We use CV to represent the level of our uncertainty throughout this paper. For σ_{y_i}, we choose CV to be 0.05. This says that most or about 95 percent of the measurement errors are assumed to fall within 10 percent of the mean or true value y_i' from the mean, which is also the mode.

Similarly for the E3M annual end-use electricity consumption estimates for a given i, using folded normal distributions. However, for each i, we set CV to be 0.125 for $x1_i', x2_i', x3_i'$, and $x4_i'$ while CV is set to be 0.25 for $x5_i'$. In other words, for space heating, A/C, water heating, and refrigerators, the expert estimation error is assumed to be identical and approximately in the range of ±25 percent of the true end-use consumption amount from the true end-use consumption amount, while for the others end use, the expert error is larger by the factor of two and approximately in the range of ±50 percent of the true end-use consumption amount from the true end-use consumption amount. As we mentioned earlier, the others end use by a respondent i is composed of multiple and specific electricity end uses, which are contained in the housing unit survey questionnaire and whose usages are claimed by the particular respondent. Measurement errors in different end-use consumption amounts in the others end-use group are expected to vary. Based on all these variability and uncertainty, we assume the overall measurement error in $x5_i$ to be twice as large as that in each of $x1_i$, $x2_i$, $x3_i$, and $x4_i$.

The prior distribution of the true total electricity consumption amount y_i' of i is assumed to be a folded normal distribution with the mean defined by the sum of true end-use electricity consumption amounts $x1_i' + x2_i' + x3_i' + x4_i' + x5_i'$ and with the standard deviation of σ_i. We do not specify y_i' to be exactly equal to $x1_i' + x2_i' + x3_i' + x4_i' + x5_i'$ or we do not set σ_i to be exactly 0, mainly because the others end-use group, by design, does not include end uses that are not included in the housing unit survey, as discussed earlier. That is, we expect there may exist some unknown other end uses of electricity in each i.

Also, erroneous omissions or commissions of some end uses are possible due to misreporting by the survey respondent. For example, a respondent might mistakenly reports possession or/and usage of some miscellaneous electric appliances mentioned in the survey. A response error or misreporting could happen not only in the others end-use group but also with any of the main end uses we consider here: space heating, A/C, water heating, and refrigerators. However, we note that we implemented stricter response data quality assurance processes with respect to those main end uses during and after the survey interviewing. Also, identification and correction of possible misreporting of at least space heating and A/C are relatively easy, as their consumption amounts are often seasonal and substantial and any corresponding (usually monthly) electricity consumption billing data could show unexpected patterns in the series of consumption values.

Thus, the *true* (unknown) total and end-use electricity consumption amounts themselves, including zero values, are determined deductively or conditionally with the ownership and usage responses by the respondent. The respondent's true ownership and non-zero usage of each end use may be modeled with some binary outcome probability model. If we condition on this true status, the relation between the true total electricity consumption amount and the sum of the true end-use electricity consumption amounts becomes exact: $y_i' = x1_i' + x2_i' + x3_i' + x4_i' + x5_i'$ (here allowing zero consumption values) for each i. We, however, do not pursue this approach at this time.

The hyper prior distribution of $(x1_i', x2_i', x3_i', x4_i', x5_i')^\top$, which is also the prior distribution relative to $(x1_i, x2_i, x3_i, x4_i, x5_i)^\top$, is assumed to be a folded multivariate normal distribution. Its mean vector is written as $(\mu_{x1'}, \mu_{x2'}, \mu_{x3'}, \mu_{x4'}, \mu_{x5'})^\top$ such that $\mu_{x1'} = \mu_{x2'} = \mu_{x3'} = \mu_{x4'} = \mu_{x5'}$ = 15,800 kWh / 5, where 15,800 kWh is the empirical weighted-average annual total electricity consumption value in the end-use combination e = 14, calculated with the current data y_i, and where 15,800 kWh is uniformly or equally distributed to the five end-use groups. More specifically, for a given end-use combination e, the empirical weighted average is the estimated population total of the annual total electricity consumption amounts divided by the estimated occupied housing unit population size, rounded to the nearest hundred. We make no prior assumptions about the relations among $\mu_{x1'}$, $\mu_{x2'}$, $\mu_{x3'}$, $\mu_{x4'}$, and $\mu_{x5'}$, which is the reason for the uniform or non-informative distribution of 15,800 kWh over the five means. Empirical weighted averages for the other end-use combinations before uniform allocation are given in Table 3.

Table 3: Empirical Weighted-Average Annual Total Electricity Consumption Values

End-Use Combination (e)	*Number of End Uses*	*Weighted Average (kWh)*
1	1	5,800

2	2	2,200
3	3	1,200
4	2	3,500
5	2	1,900
6	4	9,100
7	3	6,600
8	2	5,600
9	4	11,100
10	3	83,00
11	3	9,400
12	4	11,600
13	4	12,300
14	5	15,800

We let Σ be the variance-covariance matrix of the folded multivariate normal distribution of $(x1'_i, x2'_i, x3'_i, x4'_i, x5'_i)^\top$, and we decompose it as follows:

$$\Sigma = \mathbf{\Delta\Omega\Delta} = \mathbf{\Delta L L^\top \Delta},$$

where $\mathbf{\Delta}$ is the diagonal matrix of scales and $\mathbf{\Omega}$ is the correlation matrix that is further Cholesky-decomposed to the lower-triangular matrix **L**. The scale parameters in $\mathbf{\Delta}$ are specified as:

$$\delta_{x1'} = 1 \times \mu_{x1'},$$
$$\delta_{x2'} = 1 \times \mu_{x2'},$$
$$\delta_{x3'} = 1 \times \mu_{x3'},$$
$$\delta_{x4'} = 1 \times \mu_{x4'}, \text{ and}$$
$$\delta_{x5'} = 1 \times \mu_{x5'}.$$

Here, CV equals 1 for each scale parameter, which makes the folded multivariate normal distribution to be rather weakly informative or quite flat over the realistic values of the true end-use electricity consumption amounts. That is, marginally for $x1'_i$, most of the probability mass falls between 0 and $\mu_{x1'} + 2 \times \delta_{x1'}$ = $3 \times \mu_{x1'}$ = 3 × (15,800 kWh / 5) = 9,480 kWh. Likewise for $x2'_i, x3'_i, x4'_i$, and $x5'_i$. (It is certainly possible to make the distribution even flatter. However, we have seen some artefactual convergence problems in the log posterior distribution with the current data, specifically some bimodality in the distribution.)

The lower-triangular matrix **L** of the correlation matrix $\mathbf{\Omega}$ is assumed to be distributed as the LKJ distribution with the shape parameter ν of 2. LKJ stands for Lewandowski, Kurowicka, and Joe, and the distribution was originally developed to generate random correlation matrices for a given dimension (Lewandowski et al., 2009). The shape parameter ν (> 0) controls the distribution of expected correlations among the parameters $\mu_{x1'}$, $\mu_{x2'}$, $\mu_{x3'}$, $\mu_{x4'}$, and $\mu_{x5'}$. The ν value of 1 would specify the uniform distribution over the correlation matrices. The ν value of 2 here weakly gives a little more mass around the identity matrix, i.e., the zero-correlation matrix. (The ν values less than one would shape the distribution to dip around the identity matrix.)

Finally, back to the standard deviation σ_i of the prior distribution of the true total electricity consumption amount y'_i of i, we assume it is distributed as a folded normal distribution with the mean $\kappa_i = 0$ and the standard deviation $\tau_i = 0.025 \times (x1'_i + x2'_i + x3'_i + x4'_i + x5'_i)$. By $\kappa_i = 0$, we assume y'_i is distributed around $x1'_i + x2'_i + x3'_i + x4'_i + x5'_i$ without any systematic "bias". The CV of 0.025 represents our level of uncertainty in the discrepancy of y'_i from $x1'_i + x2'_i + x3'_i + x4'_i + x5'_i$—with y'_i mostly within ±5 percent of $x1'_i + x2'_i + x3'_i + x4'_i + x5'_i$ from $x1'_i + x2'_i + x3'_i + x4'_i + x5'_i$ for each i.

3.2 Uncertainties in the Survey Weights w_i

For the housing unit i in the Census Region r, the survey weight w_i is distributed folded-normally with the true weight value w_i' as the mean and the standard deviation of σ_{w_i}:

$$w_i \mid w_i', \sigma_{w_i} \sim \text{Normal}(w_i', \sigma_{w_i}) > 0.$$

Again this gives us the likelihood. We specify the prior for the true weight value w_i' by a folded normal distribution with the mean $\mu_{w'[r[i]]}$ and the standard deviation $\sigma_{w'[r[i]]}$, implementing the partial pooling of the respondents by the Census Region r, where $r[i]$ indicates the Census Region to which i belongs:

$$w_i' \mid \mu_{w'[r[i]]}, \sigma_{w'[r[i]]} \sim \text{Normal}(\mu_{w'[r[i]]}, \sigma_{w'[r[i]]}) > 0,$$

where

$$\mu_{w'[\text{Northeast}]} = 9{,}200,$$
$$\mu_{w'[\text{Midwest}]} = 9{,}100,$$
$$\mu_{w'[\text{South}]} = 10{,}300,$$
$$\mu_{w'[\text{West}]} = 8{,}600,$$
$$\sigma_{w'[\text{Northeast}]} = 0.125 \times \mu_{w'[\text{Northeast}]},$$
$$\sigma_{w'[\text{Midwest}]} = 0.125 \times \mu_{w'[\text{Midwest}]},$$
$$\sigma_{w'[\text{South}]} = 0.125 \times \mu_{w'[\text{South}]}, \text{ and}$$
$$\sigma_{w'[\text{West}]} = 0.125 \times \mu_{w'[\text{West}]}.$$

The means $\mu_{w'[r[i]]}$ are given by the "average" weights of the Census Regions. For example, the Northeast Census Region had about 20.8 million occupied housing units in 2009 according to the American Community Survey figures and the 2009 RECS had 2,266 respondents; thus, the average weight is calculated by dividing 20.8 million by 2,266 as 9,179, which is rounded to the nearest hundred as 9,200. The means are more or less in the same magnitude, though South Census Region gets the slightly larger mean of 10,300.

The standard deviations $\sigma_{w'[r[i]]}$ of w_i' represents our prior assessment of how the true weights might be spread around the means, using CV's. Given our choice of stratification, the initial target numbers of completed interviews or of the respondents were fixed for the strata, based on the target level of expected precision or standard error in the sample estimates of key survey variables. There was no oversampling of any particular groups of housing units. Thus, the initial design weights are expected to be similar in size. Possible variation in the weights come from differential survey response propensities/rates and weight adjustments to correct for them. The nonresponse adjustments were model-based and also poststratification-based. CV of 0.125 is specified equally to each Census Region and to be on the conservative side, i.e., to be quite non-informative.

The standard deviation σ_{w_i} of w_i represents our level of uncertainty in each calculated weight w_i. We set for a given i:

$$\sigma_{w_i} = 0.05 \times w_i',$$

with CV of 0.05. In other words, we do not think the weighting error is that large for each i—within about 10 percent of the true weight value w_i' from the true weight value.

Note that the models for survey weights, i.e., the likelihoods and priors, are independent from those for the total and end-use electricity consumption amounts. In fact, all of our likelihoods are defined with the housing-unit-level parameters, though the respondents are partially pooled. Thus, we need conduct posterior distribution simulations for each housing unit. We first group our respondents by the end-use combination and then within the end-use combination we draw posterior samples for the true weight values w_i' as well as for the true total and end-use electricity consumption amounts y_i', $x1_i'$, $x2_i'$, $x3_i'$, $x4_i'$, and $x5_i'$ and for the lower-triangular correlation matrix **L**.

4. Posterior Distributions of Unknown Values

The goal of Bayesian inference is to get a "good" joint posterior distribution of unknown values of our interest. With independence assumptions among the unknowns, we would further like to get a good marginal distribution of each key unknown. We hope the current data would increase our information about the unknowns in the sense that the posterior distributions become improved "priors" distributions or that the posterior distributions are more informative than the prior distributions.

In our current problem, the key unknowns are y_i', $x1_i'$, $x2_i'$, $x3_i'$, $x4_i'$, $x5_i'$, and w_i', which are all independent from each other and from **L**. The joint posterior distributions are specified by the likelihoods and priors provided in the previous section:

$$\mathrm{p}(y_i', x1_i', x2_i', x3_i', x4_i', x5_i', \mathbf{L} \mid y_i, x1_i, x2_i, x3_i, x4_i, x5_i)$$
$$\propto \mathrm{p}(y_i, x1_i, x2_i, x3_i, x4_i, x5_i \mid y_i', x1_i', x2_i', x3_i', x4_i', x5_i', \mathbf{L}) \times \mathrm{p}(y_i', x1_i', x2_i', x3_i', x4_i', x5_i', \mathbf{L})$$
$$= \text{likelihood} \times \text{prior}$$

and

$$\mathrm{p}(w_i' \mid w_i) \propto \mathrm{p}(w_i \mid w_i') \times \mathrm{p}(w_i') = \text{likelihood} \times \text{prior}.$$

The first joint posterior distribution can be factored into independent marginal posterior distributions of y_i', $x1_i'$, $x2_i'$, $x3_i'$, $x4_i'$, $x5_i'$, and **L**.

Now, posterior distributions are not always easy to derive mathematically—in our case, due to the folding of the likelihoods and priors, though they are all based on normal distributions. So, we rely on simulations from the posterior distributions. However, it does not mean that the simulations are always computationally easy or doable. In practice, we would like to get an effective sample efficiently, which concerns the question of how to draw and then select sample draws. We utilize the very effective and efficient algorithm called No-U-Turn Sampler (NUTS) offered in Stan, running it from R (Stan Development Team, 2017a). Basically, NUTS adaptively optimizes the Hamiltonian Monte Carlo method (Hoffman and Gelman, 2014).

For each housing unit i in the given end-use combination e, we have generated 2,000 NUTS cases from each of three chains with 1,000 warm-up cases discarded (without any thinning), getting 3,000 total cases to approximate the posterior distributions. With these samples, the convergence turns out to be very good for each of the parameters in terms of the $\hat{R}$ diagnostic (Gelman et al., 2013), which we compute, using ShinyStan—the interactive graphical R package for Markov chain Monte Carlo diagnostics (Stan Development Team, 2017d). To summarize each posterior distribution, we compute the posterior sample mean, and those posterior means are used to make our population inferences in the next section. The Stan codes for Bayesian modeling (for e = 14) and for the population inferences are provided in Appendix B. The standard deviation and quantiles of each posterior distribution simulation are also available, but they are not discussed here. However, most of the distributions are quite symmetrical, and the standard deviations we have inspected are about five percentage of the respective means or much smaller than them.

We interpret these to suggest that most of the posterior distributions are substantively tight around the means.

The posterior means of the correlation parameters in $\mathbf{L}$ ($\Sigma = \mathbf{\Delta\Omega\Delta} = \mathbf{\Delta LL^{\top}\Delta}$) are also available. With the weakly informative LKJ prior, the data have spoken up rather loudly for these unknown parameters. However, all of the signs of the correlation coefficients turn out to be in the right directions when the sizes of the correlation coefficients are substantively large.

5. Population Inferences of Annual Total and End-Use Electricity Consumption Amounts in the U. S.

Using the posterior means of the annual total and end-use electricity consumption amounts for each respondent i, we first produce sum-over-respondents estimates at the national level, shown in the "Sample Total (million kWh) : Bayesian" column of Table 5. The "Sum" row gives a simple sum of the five end-use values in the table, while the "Total" row holds the annual total electricity consumption values. The "*(Weight)*" row values are in millions as well, but obviously without the kWh unit; thus, the italicization and brackets.

The sample estimates with the raw end-use estimates, the raw total estimates, and the raw survey weights are also given under the "Sample Total (million kWh) : Raw" column of Table 5. Comparing the Sample Bayesian column against the Sample Raw column, only two end uses show rather significant changes: about 7.3 million kWh decrease in Space Heating and about 12.0 million kWh increase in Others. The Sample Bayesian Total is increased but only by about 4.0 million kWh, which is similar to the level of increase in the Sample Bayesian Sum (4.9 m. kWh). The latter similarity is clear from the fact that the Sum and the Total are in good agreement for the Sample Bayesian estimates (140.9 m. vs. 140.4 m.) as well as for the Sample Raw estimates (136.0 m. vs. 136.4 m.). Recall that the Total and the Sum are based on two independent data sources—administrative billing data vs. expert modeling and estimation with housing units survey data and administrative weather data. Thus, those agreements are quite remarkable.

We also show the Sample Bayesian estimates with some adjustments, which (a) normalize the Bayesian Total estimates so that they sum to the sum of the Raw Total values over the respondents and (b), for each respondent, uniformly calibrate the Bayesian End-Use estimates so that they sum over the end uses to the respondent's Raw Total value. These Adjusted Bayesian Total and End-Use estimates are more conservative than the Bayesian Total and End-Use estimates in the sense that we assume the sum of the Raw Total values over the respondents are correct. Even more conservative estimates could be derived if we assume the Raw Total value is correct for each respondent and uniformly calibrate the Bayesian End-Use estimates to the Raw Total value for each respondent. That is, the administrative billing data and the annualization of the data are assumed to be exactly correct.

The Sample Bayesian Weight sum (109.1 m.) turns out to be about 4.5 million lower than the Sample Raw Weigh sum (113.6 m.). Given the amount of regularization we put in weighting errors (σ_{w_i}= 0.05 × w_i'), the difference, though small in the absolute term, is relatively large. The Sample Bayesian Weight estimates can also be adjusted or normalized so that they sum to the sum of the Sample Raw Weight values over the respondents.

Applying survey weights to individual sample values, we can make population inferences. Here, we look at the national level. The "National Total (billion kWh) : Raw" column shows the population estimates of total and end-use consumption amounts, calculated with the Sample Raw total, end-use, and weight values. We note again that the Sum of the end-use electricity consumption estimates (1,262 b.) is relatively close to the Total electricity consumption estimate (1,286 b.) with the difference of -24 b. The National estimates in Table 4 are graphed in Figure 1 as well.

The "National Total (billion kWh) : Bayesian" column gives the population estimates from the Sample Bayesian total, end-use, and weight estimates. The decrease in Space Heating (from 328 b. to 258 b.) and the increase in Others (from 505 b. to 601 b.) are substantial as in the sample counterparts. With the Bayesian estimates, the Sum of the end-use electricity consumption estimates (1,270 b.) is even closer to the Total electricity consumption estimate (1,267 b.) with the difference of just about 3 b. Comparing to the difference of the differences in Sample Raw and Sample Bayesian, -0.4 m. (= 136.0 m. – 136.4 m.) and 0.5 m. (= 140.9 m. – 140.4 m.), respectively, the difference (-24 b. – 3 b.) is rather large. More specifically, the Bayesian weights seem to keep the National difference (3 b.) under control in terms of the size and sign of the difference, given the Sample difference (0.5 m.), while the Raw weights somewhat blow up the difference between Sum and Total from the Sample difference (-0.4 m.) to the Population difference (-24 b.). In other words, the Bayesian weights seem to contribute in producing such National End-Use estimates and National Total estimate that are more consistent with each other.

The other things we notice, comparing the National Raw and Bayesian estimates of Sum and Total, are that (a) the National Bayesian Sum estimate (1,270 b.) stays larger than the National Raw Sum estimate (1,262 b.), keeping the same directional relation under the sample, although the sum of the Bayesian Weight estimates (109.1 m.) becomes smaller than the sum of the Raw Weight values (113.6 m.) and (b) the National Bayesian Total estimate (1,267 b.) becomes smaller than the National Raw Total estimate (1,286 b.), although the Sample Bayesian Total estimate (140.4 m.) is larger than the Sample Raw Total value (136.4 m.), with the smaller-in-sum Bayesian Weight estimates.

The "National Total (billion kWh) : Bayesian + Adjustments" column is calculated with the Adjusted Bayesian Total and End-Use consumption estimates and the Adjusted Bayesian Weight estimates. The National Sum estimate (1,281 b.) equals the National Total estimate by definition of the adjustments, and it is larger than the National Bayesian Sum estimate (1,270 b.) or the National Bayesian Total estimate (1,267 b.) but falls between the National Raw Sum estimate (1,262 b.) and the National Raw Total estimate (1,286 b.). We also note that, with the adjusted Bayesian estimates, the National Space Heating estimate goes further up, while the National Others estimate goes further down.

Any statistical analysis should include some assessment of uncertainty or variability in the quantities of interest investigated. Here, we provide only intuitive assessment. Since all the Sample Bayesian estimates in Table 4 are based on the linear combinations of posterior means of the total, end-use, and weight values of the individual respondents, their uncertainty is expected to be roughly at the same level as that of the individual respondents, which was about five percentage in terms of CV. The National Bayesian estimates in Table 4 might be less certain due to the multiplication of the Sample Bayesian Total and End-Use estimates and the Sample Bayesian Weight estimates. But, if we note that $\mathrm{CV}(XY) = \sqrt{\mathrm{CV}^2(X)\mathrm{CV}^2(Y) + \mathrm{CV}^2(X) + \mathrm{CV}^2(Y)}$, when X and Y are independent, then with $\mathrm{CV}(X) = \mathrm{CV}(Y) = 0.05$ we can calculate $\mathrm{CV}(XY)$ to be about 0.07. The standard deviation relative to the mean (0.07) is still quite small in our substantive interpretation.

Table 4: Sample and Population Estimates of Annual Total and End-Use Electricity Consumption Amounts

End Use	*Sample Total (million kWh)*			*National Total (billion kWh)*		
	Raw	*Bayesian*	*Bayesian + Adjustments*	*Raw*	*Bayesian*	*Bayesian + Adjustments*
Space Heating	35.3	28.0	25.9	328	253	244
A/C	21.3	21.2	20.8	196	190	194
Water Heating	14.7	14.5	13.8	141	134	136
Refrigerators	9.9	10.3	10.0	92	92	94
Others	54.8	66.8	65.8	505	601	617
Sum	**136.0**	**140.9**	**136.4**	**1,262**	**1,270**	**1,281**
Total	**136.4**	**140.4**	**136.4**	**1,286**	**1,267**	**1,281**
(Weight)	*(113.6)*	*(109.1)*	*(113.6)*	-	-	-

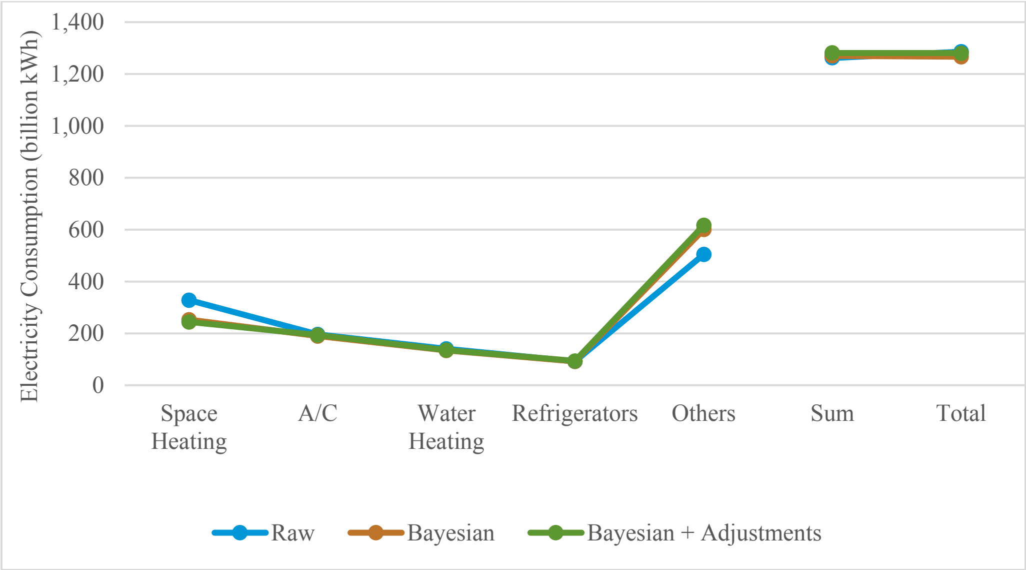

Figure 1: National Estimates of Total and End-Use Electricity Consumption Amounts (in billion kWh)

6. Future Work

Although we have used non-informative or weakly informative priors, mainly to regulate the posterior distributions, the effects of different priors are not thoroughly investigated. Also, as mentioned earlier, the modeling of the uncertainty in the end-use consumption amounts could be improved against the E3M end-use estimates, if we explicitly model the uncertainty in the existence or absence of each end use for each respondent or the uncertainty in the response accuracy of each end use by each respondent. Finally, apart from model checking and improvement, it is important for us continually to evaluate and advance survey measurements, administrative data quality, data processing methods, sampling designs, and weighting procedures so that substantive questions of our interest may be better addressed statistically and scientifically.

Appendix A. List of End Uses Covered in the 2009 RECS

- Lighting
- Space heating
- Space cooling (A/C)
- Water Heating
- Major appliances (Refrigerators; Freezers; Stoves, ovens, and stove-ovens; Dishwashers; Clothes washers; Clothes dryers)
- Kitchen appliances (Microwaves; Coffee makers; Toasters)
- Miscellaneous electric loads (Dehumidifiers; Humidifiers; Ceiling fans; Computers and monitors; Printers, Fax's, and copiers; Home network equipment; Televisions; Set top boxes; DVD's, VCR's, and combo DVD/VCR units; Video game consoles; Rechargeable tools and electronics)
- Others (Pool and spa heaters; Well pumps; Automobile block/battery heaters)

Appendix B. Stan Codes for Bayesian Modeling (for e = 14) and for Population Inferences

```
data {
  int<lower = 1> N;                       // Number of respondents/observations
  int<lower = 1> n_eu;                    // Number of end uses
  int<lower = 1> n_reg;                   // Number of Census Regions
  int<lower = 1, upper = n_reg> reg[N];  // Census Region identifier
  int<lower = 1> su_id[N];               // SU_ID
  vector<lower = 0>[N] x1;               // Space heating electricity consumption estimates by the experts
  vector<lower = 0>[N] x2;               // A/C electricity consumption estimates by the experts
  vector<lower = 0>[N] x3;               // Water heating electricity consumption estimates by the experts
  vector<lower = 0>[N] x4;               // Refrigerators electricity consumption estimates by the experts
  vector<lower = 0>[N] x5;               // Others end-use electricity consumption estimates by the experts
  vector<lower = 0>[N] y;                // Annual total electricity consumption amounts from the administrative billing data
  vector<lower = 0>[N] w;                // Survey weights
}

transformed data {
  real<lower = 0> y_sum;            // Sum of the annual total electricity consumption amounts over respondents
  real<lower = 0> w_sum;            // Sum of the survey weights over respondents

  vector<lower = 0>[n_eu] mu_tx;    // The mean vector of the priors for the true end-use consumption amounts
  vector<lower = 0>[n_eu] L_sigma;  // The scale or SD vector of the true end-use consumption amounts

  vector<lower = 0>[n_reg]  mu_tw;  // The mean of the prior for the true survey weight by Region
  vector<lower = 0>[n_reg] sig_tw;  // The SD of the prior for the true survey weight by Region

  mu_tx[1] = 0.2 * 15800;           // (1 / n_eu) of the average annual total electricity consumption value in e = 14,
  mu_tx[2] = 0.2 * 15800;           // (1 / n_eu) based on the published estimate, rounded to the hundredth
  mu_tx[3] = 0.2 * 15800;           // (1 / n_eu)
  mu_tx[4] = 0.2 * 15800;           // (1 / n_eu)
  mu_tx[5] = 0.2 * 15800;           // (1 / n_eu)

  L_sigma[1] = 1 * mu_tx[1];        // tx1: CV = 1
  L_sigma[2] = 1 * mu_tx[2];        // tx2: CV = 1
  L_sigma[3] = 1 * mu_tx[3];        // tx3: CV = 1
  L_sigma[4] = 1 * mu_tx[4];        // tx4: CV = 1
  L_sigma[5] = 1 * mu_tx[5];        // tx5: CV = 1

  mu_tw[1] = 9200;                  // Northeast = N1 / n1 = 20.8 m / 2266 = 9179, rounded to the hundredth
  mu_tw[2] = 9100;                  // Midwest   = 25.9 m / 2843 = 9110
  mu_tw[3] = 10300;                 // South     = 42.1 m / 4090 = 10293
  mu_tw[4] = 8600;                  // West      = 24.8 m / 2884 = 8599

  sig_tw[1] = 0.125 * mu_tw[1];     // CV = 0.125
  sig_tw[2] = 0.125 * mu_tw[2];     // CV = 0.125
  sig_tw[3] = 0.125 * mu_tw[3];     // CV = 0.125
  sig_tw[4] = 0.125 * mu_tw[4];     // CV = 0.125

  y_sum = 0;
```

```
  w_sum = 0;

  for (i in 1:N) {
    y_sum = y_sum + y[i];
    w_sum = w_sum + w[i];
  }
}
parameters {
  vector<lower = 0>[N] ty;              // Latent/unknown true values of annual total electricity consumption amount
  matrix<lower = 0>[N, n_eu] tx;        // Latent/unknown true values of annual end-use electricity consumption amounts
  cholesky_factor_corr[n_eu] L_Omega;  // The lower triangular matrix of the correlation matrix of tx's
  vector<lower = 0>[N] tw;              // Latent/unknown true values of survey weights
}

transformed parameters {
  vector<lower = 0>[N] sig_model;     // The SD of ty
  vector<lower = 0>[N]       sig_y;   // The SD of the prior for the observed y
  matrix<lower = 0>[N, n_eu] sig_x;   // The SD's of the priors for the observed x1, x2, x3, x4, and x5
  vector<lower = 0>[N]       sig_w;   // The SD of the prior for the observed w
  sig_model   = 0.025 * (col(tx, 1) + col(tx, 2) + col(tx, 3) + col(tx, 4) + col(tx, 5));
                                      // The omission/commission error rate of 2.5% of the sum of the "true" end uses: CV = 0.025

  sig_y       = 0.05  * ty;           // The measurement error level in y:        CV = 0.05
  sig_x[ , 1] = 0.125 * col(tx, 1);   // The expert estimation error level in x1: CV = 0.125
  sig_x[ , 2] = 0.125 * col(tx, 2);   // The expert estimation error level in x2: CV = 0.125
  sig_x[ , 3] = 0.125 * col(tx, 3);   // The expert estimation error level in x3: CV = 0.125
  sig_x[ , 4] = 0.125 * col(tx, 4);   // The expert estimation error level in x4: CV = 0.125
  sig_x[ , 5] = 0.25  * col(tx, 5);   // The expert estimation error level in x5: CV = 0.25
  sig_w       = 0.05  * tw;           // The weighting error level:               CV = 0.05
}

model {
  L_Omega ~ lkj_corr_cholesky(2);  // Cholesky decomposition (lower triangular matrix) of the correlation matrix
                                   // = L_Omega * L_Omega' ~ lkj_corr(2)
  for (i in 1:N) {  // The means of the true end-use consumption amounts are assumed to be correlated.
    tx[i, ] ~ multi_normal_cholesky(mu_tx, diag_pre_multiply(L_sigma, L_Omega));
  }

  x1 ~ normal(col(tx, 1), col(sig_x, 1));  //sig_x[ , 1] are transformed parameters or functions of tx[ , 1]
  x2 ~ normal(col(tx, 2), col(sig_x, 2));  //sig_x[ , 2] are transformed parameters or functions of tx[ , 2]
  x3 ~ normal(col(tx, 3), col(sig_x, 3));  //sig_x[ , 3] are transformed parameters or functions of tx[ , 3]
  x4 ~ normal(col(tx, 4), col(sig_x, 4));  //sig_x[ , 4] are transformed parameters or functions of tx[ , 4]
  x5 ~ normal(col(tx, 5), col(sig_x, 5));  //sig_x[ , 5] are transformed parameters or functions of tx[ , 5]

  ty ~ normal(col(tx, 1) + col(tx, 2) + col(tx, 3) + col(tx, 4) + col(tx, 5), sig_model);

  y ~ normal(ty, sig_y);  // Observed values which depend on the latent true values
                          // sig_y are transformed parameters or functions of ty

  for (i in 1:N) {
    tw[i] ~ normal(mu_tw[reg[i]], sig_tw[reg[i]]);
```

```
  }

  w ~ normal(tw, sig_w);  // sig_w are transformed parameters or functions of tw
}

generated quantities {
  vector<lower = 0>[N] x1_est_raw;  // Sample estimates of x
  vector<lower = 0>[N] x2_est_raw;
  vector<lower = 0>[N] x3_est_raw;
  vector<lower = 0>[N] x4_est_raw;
  vector<lower = 0>[N] x5_est_raw;
  vector<lower = 0>[N] x_est_raw;   // Sum of sample estimates of x's

  vector<lower = 0>[N] y_est_raw;
  vector<lower = 0>[N] w_est_raw;
  real<lower = 0> y_est_raw_sum;    // Sum of the raw estimates of y
  real<lower = 0> w_est_raw_sum;    // Sum of the raw estimates of w
  vector<lower = 0>[N] y_est_adj;   // Adjusted or normalized y_est_raw, which sum to the sum of (observed) y
  vector<lower = 0>[N] w_est_adj;   // Adjusted or normalized w_est_raw, which sum to the sum of (observed) w

  vector<lower = 0>[N] x1_est_cal;  // Uniformly calibrated sample estimates of x so that x1_est_cal + ... = y_est_raw_adj
  vector<lower = 0>[N] x2_est_cal;
  vector<lower = 0>[N] x3_est_cal;
  vector<lower = 0>[N] x4_est_cal;
  vector<lower = 0>[N] x5_est_cal;

  vector<lower = 0>[N] x1_pop_raw;  // Weighted population estimates with raw estimates of x and w
  vector<lower = 0>[N] x2_pop_raw;
  vector<lower = 0>[N] x3_pop_raw;
  vector<lower = 0>[N] x4_pop_raw;
  vector<lower = 0>[N] x5_pop_raw;
  vector<lower = 0>[N] y_pop_raw;   // Weighted population estimates with raw estimates of y and w

  vector<lower = 0>[N] x1_pop_cal;  // Calibrated weighted population estimates with calibrate estimates of x
  vector<lower = 0>[N] x2_pop_cal;  // and adjusted estimates of w
  vector<lower = 0>[N] x3_pop_cal;
  vector<lower = 0>[N] x4_pop_cal;
  vector<lower = 0>[N] x5_pop_cal;
  vector<lower = 0>[N] y_pop_adj;   // Adjusted weighted population estimates with adjusted estimates of y and adjusted estimates of w

  corr_matrix[n_eu] Omega_x;        // The correlation matrix of the means of the true x's
  cov_matrix[n_eu]  Sigma_x;        // The covariance matrix of the means of the true x's

  x1_est_raw = col(tx, 1);
  x2_est_raw = col(tx, 2);
  x3_est_raw = col(tx, 3);
  x4_est_raw = col(tx, 4);
  x5_est_raw = col(tx, 5);
  x_est_raw  = x1_est_raw + x2_est_raw + x3_est_raw + x4_est_raw + x5_est_raw;
  y_est_raw  = ty;
  w_est_raw  = tw;
```

```
  y_est_raw_sum = 0;
  w_est_raw_sum = 0;
  for (i in 1:N) {
    y_est_raw_sum = y_est_raw_sum + y_est_raw[i];
    w_est_raw_sum = w_est_raw_sum + w_est_raw[i];
    }

  y_est_adj = y_est_raw * (y_sum / y_est_raw_sum);  // Adjust y_est_raw so that they sum to the sum of (observed) y (vector * scalar)
  w_est_adj = w_est_raw * (w_sum / w_est_raw_sum);  // Adjust w_est_raw so that they sum to the sum of (observed) w

  x1_est_cal = x1_est_raw .* (y_est_adj ./ x_est_raw);  // Uniform calibration of end uses so that the sum of the calibrated end-use
  x2_est_cal = x2_est_raw .* (y_est_adj ./ x_est_raw);  // estimates = the adjusted y estimate
  x3_est_cal = x3_est_raw .* (y_est_adj ./ x_est_raw);  // Elementwise multiplication (similar efficiency to loop)
  x4_est_cal = x4_est_raw .* (y_est_adj ./ x_est_raw);
  x5_est_cal = x5_est_raw .* (y_est_adj ./ x_est_raw);

  x1_pop_cal = w_est_adj .* x1_est_cal;  // The calibrated population estimates of x, i.e., with the uniformly calibrated
  x2_pop_cal = w_est_adj .* x2_est_cal;  // sample estimates of x and the adjusted weight estimates
  x3_pop_cal = w_est_adj .* x3_est_cal;
  x4_pop_cal = w_est_adj .* x4_est_cal;
  x5_pop_cal = w_est_adj .* x5_est_cal;
  y_pop_adj  = w_est_adj .* y_est_adj;

  x1_pop_raw = w_est_raw .* x1_est_raw;  // The raw population estimates of end uses, i.e., with the raw sample estimates of end uses
  x2_pop_raw = w_est_raw .* x2_est_raw;  // and the raw weight estimates
  x3_pop_raw = w_est_raw .* x3_est_raw;
  x4_pop_raw = w_est_raw .* x4_est_raw;
  x5_pop_raw = w_est_raw .* x5_est_raw;
  y_pop_raw  = w_est_raw .* y_est_raw;   // The raw population estimates of y, i.e., with the raw sample estimates of y and the raw
                                         // weight estimates

  Omega_x = multiply_lower_tri_self_transpose(L_Omega);
  Sigma_x = quad_form_diag(Omega_x, L_sigma);
}
```

References

Alvarez I., Niemi, J., and Simpson, M. (2014), "Bayesian Inference for a Covariance Matrix," arXiv:1408.4050v2 [stat.ME].

Barnard, J., McCulloch, R., and Meng, X.-L. (2000), "Modelling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage," *Statistica Sinica*, 10, 1281-1311.

Betancourt, M. J., and Girolami, M. (2015), "Hamiltonian Monte Carlo for Hierarchical Models," in Current Trends in Bayesian Methodology with Applications, eds. S. K. Upadhyay, U. Singh, D. K. Dey, and A. Loganathan, Boca Raton, FL: Chapman & Hall/CRC, pp. 79-101.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017), "Stan: A probabilistic Programming Language," *Journal of Statistical Software*, 76, 1-32.

Carpenter, B., Hoffman, M. D., Brubaker, M., Lee, D., Li, P., and Betancourt, M. (2015), "The Stan Math Library: Reverse-Mode Automatic Differentiation in C++," arXiv:1509.07164 [cs.MS].

Chen, Q., Gelman, A., Tracy, M., Norris, F. H., and Galea, S. (2015), "Incorporating the Sampling Design in Weighting Adjustments for Panel Attrition," *Statistics in Medicine*, 34, 3637-47.

Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., and Dorie, V. (2015), "Weakly Informative Prior for Point Estimation of Covariance Matrices in Hierarchical Models," *Journal of Educational and Behavioral Statistics*, 40, 136-157.

Gelman, A. (1995), "Method of Moments Using Monte Carlo Simulation," *Journal of Computational and Graphical Statistics*, 4, 36-54.

Gelman, A. (2006a), "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do," *Technometrics*, 48, 432-435.

Gelman A. (2006b), "Prior Distributions for Variance Parameters in Hierarchical Models," *Bayesian Analysis*, 3, 515-533.

Gelman, A. (2007), "Struggles with Survey Weighting and Regression Modeling" (with discussion), *Statistical Science*, 22, 153-164.

Gelman, A. (2009), "Constructing a Scaled-Inverse-Wishart Prior Distribution That is Informative on the Variance Parameters," Statistical Modeling, Causal Inference, and Social Science Blog, 17 August. Available at *http://andrewgelman.com/2009/08/17/constructing_a*.

Gelman, A. (2010), "Bayesian Statistics Then and Now," Discussion of "The Future of Indirect Evidence," by B. Efron, *Statistical Science*, 25,162-165.

Gelman, A. (2011), "The Pervasive Twoishness of Statistics; in Particular, the Sampling Distribution and the Likelihood are Two Different Models, and That is a Good Thing," Statistical Modeling, Causal Inference, and Social Science Blog, 20 June. Available at *http://andrewgelman.com/2011/06/20/the sampling di 1*.

Gelman, A. (2014), "How Do We Choose Our Default Methods?," in Past, Present, and Future of Statistical Science, eds. X. Lin, C. Genest, D. L. Banks, G. Molenberghs, D. W. Scott, and J.-L. Wang, Boca Raton, FL: Chapman & Hall/CRC, pp. 291-299.

Gelman, A. (2016), "The Problems With *P*-Values are not Just With *P*-Values," Supplemental Material to "The ASA's Statement on p-Values: Context, Process, and Purpose" by R. L. Wasserstein and N. A. Lazar, *The American Statistician*, 70, 129-133. Available at *http://amstat.tandfonline.com/doi/suppl/10.1080/00031305.2016.1154108*.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian Data Analysis* (3rd ed.), Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., and Hill, J. (2006), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, New York, NY: Cambridge University Press.

Gelman, A., Hwang, J., and Vehtari, A. (2014), "Understanding Predictive Information Criteria for Bayesian Models," *Statistics and Computing*, 24, 997-1016.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y. S. (2008), "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models," *Annals of Applied Statistics*, 2, 1360-1383.

Gelman, A., Lee, D., and Guo, J. (2015), “Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization,” *Journal of Educational and Behavior Science*, 40, 530-543.

Gelman, A., Meng, X. L., and Stern, H. S. (1996), Posterior Predictive Assessment of Model Fitness via Realized Discrepancies” (with discussion), *Statistica Sinica*, 6, 733-807.

Gelman, A., and Robert, C. (2013), “Not Only Defended but Also Applied: The Perceived Absurdity of Bayesian Inference” (with discussion), *The American Statistician*, 67, 1-5.

Gelman, A., and Shalizi, C. (2012), “Philosophy and the Practice of Bayesian Statistics” (with discussion), *British Journal of Mathematical and Statistical Psychology*, 66, 8-38.

Gelman, A., and Vehtari, A. (2014), “Bootstrap Averaging: Examples Where it Works and Where it Doesn't Work,” Comment on “Estimation and Accuracy After Model Selection,” by B. Efron, *Journal of the American Statistical Association*, 109, 1015-1016.

Hoffman, M. D., and Gelman, A. (2014), “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo,” *Journal of Machine Learning Research*, 15, 1351-1381.

Lewandowski, D., Kurowicka, D., Joe, H. (2009), “Generating Random Correlation Matrices Based on Vines and Extended Onion Method,” *Journal of Multivariate Analysis*, 100, 1989-2001.

Liu, Y., Gelman, A., and Zheng, T. (2013), “Simulation-Efficient Shortest Probability Intervals,” Technical Report, Columbia University, Dept. of Statistics.

McElreath, R. (2016), *Statistical Rethinking: A Bayesian Course with Examples in R and Stan,* Boca Raton, FL: Chapman & Hall/CRC.

Minato, H. (2016), “Statistical Calibration of Energy-Engineering End-Use Electricity Consumption Estimates: A Bayesian Multilevel Model Approach,” in 2016 Proceedings of the American Statistical Association, Government Statistics Section [CD-ROM], Alexandria, VA: American Statistical Association: 1475-1499.

National Oceanic and Atmospheric Administration (2016), *Quality Controlled Local Climatological Data (QCLCD)*. Available at *http://www.ncdc.noaa.gov/qclcd*.

Polson, N. G., and Scott, J. G. (2012), “On the Half-Cauchy Prior for a Global Scale Parameter,” *Bayesian Analysis*, 7, 1-16.

Shirley, K. E., and Gelman, A. (2014), “Hierarchical Models for Estimating State and Demographic Trends in US Death Penalty Public Opinion,” Journal of the Royal Statistical Society, *Series A*, 178, 1-28.

Stan Development Team (2017a), *RStan: The R interface to Stan* (version 2.15.1). Available at *http://mc-stan.org/users/interfaces/rstan.html*.

Stan Development Team (2017b), *Stan Modeling Language Users Guide and Reference Manual* (version 2.15.0). Available at *http://mc-stan.org/users/documentation/index.html*.

Stan Development Team (2017c), *Stan Wiki: Prior Choice Recommendations*. Available at *https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations*.

Stan Development Team (2017d), *shinystan: Interactive Visual and Numerical Diagnostics and Posterior Analysis for Bayesian Models* (version 2.3.0). Available at *http://mc-stan.org/users/interfaces/shinystan*.

Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A. and Tuerlinckx, F. (2011), “Visualizing Distributions of Covariance Matrices,” Technical Report, Columbia University, Dept. of Statistics.

U. S. Energy Information Administration (2015), *An Assessment of Interval Data and Their Potential Application to Residential Electricity End-Use Modeling*. Available at *http://www.eia.gov/consumption/residential/reports/smartmetering*.

U. S. Energy Information Administration (2016), *2009 RECS Survey Data*. Available at *http://www.eia.gov/consumption/residential/data/2009*.

Wang, W., and Gelman, A. (2014), “Difficulty of Selecting Among Multilevel Models Using Predictive Accuracy,” *Statistics and Its Interface*, 7: 1-8.