Validation of 2015 Residential Energy Consumption Survey End-Use Estimates by Bayesian Calibration Models

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

Abstract

The 2015 Residential Energy Consumption Survey (RECS) national end-use estimates were produced in three steps: (1) the annual-level end-use energy consumption models, developed by energy-engineering experts (EEE), were applied to the 2015 RECS data and some weather data, outputting initial EEE end-use estimates; (2) the EEE end-use estimates were calibrated to the fuel-specific total consumption value for each sample respondent, followed by imputation for cases with missing total consumption values; and (3) the calibrated end-use estimates are multiplied by the survey weights to produce the 2015 national end-use estimates. In this paper, we keep (1) and (3) as fixed, but in (2) we implement two simpler but more direct Bayesian calibration methods than the official method in an attempt to increase the validity of the latter method and its published results. Uncertainty of the final official estimates are assessed against the posterior predictive distributions that are derived under one of the Bayesian models.

Key Words: Bayesian multilevel models; energy-engineering-expert end-use energy consumption models; posterior predictive checking; Residential Energy Consumption Survey (RECS); Stan

Contents

1. Introduction

When you have a higher-level estimate, e.g., a national or annual total and also independent lower-level estimates, e.g., state or monthly totals, you may wish to calibrate the latter estimates against the former estimate. Such corrections or adjustments may be necessary for consistency or may just make sense if the higher-level estimate is based on more accurate census or administrative data while the lower-level estimates are based on more uncertain sample or predicted data. Another illustration is that a household may remember its total spending better than itemized amounts for a year. From utility companies, the U. S. Energy Information Administration (EIA) collects monthly electricity billing data for the respondents in the Residential Energy Consumption Survey (RECS). Those administrative billing data are checked, imputed where missing, and annualized to form annual total electricity consumption values for the RECS respondents. On the other hand, EIA utilizes energy-engineering-expert (EEE) knowledges to build a variety of annual-level end-use energy consumption models from RECS's energy and housing-unit characteristics variables and from the administrative weather data variables. For each RECS respondent, these EEE models produce end-use electricity consumption estimates, which are calibrated against the annual total electricity consumption. (Notice that two kinds of administrative data are used for two entirely different purposes—for modeling something directly and for calibrating some modeled output.) Finally, national inferences are drawn with survey weights. Detailed and complete descriptions of the data and methodology for the most recent 2015 vintage are provided at EIA's 2015 RECS Web

page https://www.eia.gov/consumption/residential/data/2015. In particular, end-use energy modeling and estimation methods are described in the report titled "Residential Energy Consumption Survey (RECS) 2015 Consumption and Expenditures Technical Documentation Summary" (U. S. Energy Information Administration, 2018b).

In this paper, two simple Bayesian calibration models are tried in the calibration step—in place of the so-called minimum variance calibration method implemented for the 2015 RECS. The Bayesian models here are not only simple but also simpler in the sense that they are expected to under-fit the data or not to fit every part of the data. It is intentional, as a present goal is not to develop an *individual-level* Bayesian model for this calibration problem but to conduct a quick but meaningful study to check the validity of the official 2015 RECS end-use *inferences*. Also, to make the validation study as impartial as possible, though it still is subjective, the Bayesian calibration models are internally pre-registered. That is, they are built before any data analyses and are unchanged during the data analyses.

2. Bayesian End-Use Estimates Calibration Model I

We consider calibrating the following twenty-five EEE end-use electricity consumption estimates against the total electricity consumption value for each of the 5,686 respondents in the 2015 RECS (which will be indexed by the subscript i later):

$x1$ = Space heating, main and secondary
$x2$ = Air conditioning (central systems and individual units)
$x3$ = Water heating, main and secondary
$x4$ = All refrigerators
$x5$ = Cooking (stoves, cooktops, and ovens)
$x6$ = Clothes dryers
$x7$ = Freezers
$x8$ = Indoor and outdoor lighting
$x9$ = Clothes washers
$x10$ = Dishwashers
$x11$ = All televisions and related peripherals
$x12$ = Computers and related peripherals
$x13$ = Air handlers and boiler pumps used for heating and air handlers used for cooling
$x14$ = Evaporative coolers
$x15$ = Ceiling fans
$x16$ = Floor, whole house, and attic fans
$x17$ = Dehumidifiers
$x18$ = Humidifiers
$x19$ = Small kitchen appliances
$x20$ = Swimming pool pumps and heaters
$x21$ = Hot tub pumps and heaters
$x22$ = Electric vehicles
$x23$ = Electric parts of gas dryers
$x24$ = Devices and purposes surveyed for the sample cases interviewed only in the in-person mode
$x25$ = Unknown devices and purposes

All are estimated in the annual kWh. Note that in the official 2015 RECS end-use microdata and tables a slightly modified set of end uses is used. For example, the end uses $x22$, $x23$, and $x24$ are rolled into the end use $x25$, which is labeled by "Other devices and purposes not elsewhere classified".

The first Bayesian calibration model assumes that the total electricity value y_i is distributed normally with the mean μ_i and the standard deviation σ for the 2015 RECS housing-unit respondents i = 1, …, 5,686:

$$y_i \sim \text{Normal}(\mu_i, \sigma^2),\ i = 1, \ldots, 5{,}686.$$

For each i, the mean μ_i is defined by the linear and additive function:

$$\mu_i = \beta 1 x1_i + \cdots + \beta 25 x25_i,$$

where $x1_i, \ldots, x25_i$ are the twenty-five EEE end-use electricity consumption estimates described in the above for the respondent i. The standard deviation σ is constant over i. This data model is simply a multiple linear regression model without the intercept. In the generalized linear model language, the link function is identity and equals the mean function. $x1_i, \ldots, x25_i$ are given and considered fixed for all i. The total electricity consumption y_i is always positive, but the normal approximation is used and expected to be reasonable as the σ will be made tight around the μ_i by its prior distribution discussed next.

The prior distribution for the regression parameters $\beta 1, \ldots, \beta 25$ is assumed to be a multivariate normal distribution:

$$(\beta 1, \ldots, \beta 25)^{\top} \sim \text{Multi. Normal}((1, \ldots, 1)^{\top}, \Sigma),$$

where the variance-covariance matrix Σ is a 25×25 diagonal matrix with diagonal entries ($0.25^2, \ldots, 0.25^2, 1^2$). $\beta 1, \ldots, \beta 25$ are theoretically all positive, but the multivariate normal approximation is used and their means are set to 1's because there is no prior information about positive or negative biases in $x1_i, \ldots, x25_i$. The standard deviations for $\beta 1, \ldots, \beta 24$ are made equal at 0.25 and rather tight or informative around the mean values of 1's. We are less certain about $\beta 25$ than $\beta 1, \ldots, \beta 24$, however, and its standard deviation is made four times larger at 1. Note that the standard deviations can be interpreted as the coefficients of variation here.

The prior for the scale parameter σ is a usual Gamma distribution but with thick and flat mass over the plausible scale value of y_i or weakly informative:

$$\sigma \sim \text{Gamma}\big(\text{Shape} = 1,\ \text{Inverse Scale} = 0.001\big).$$

A Gamma distribution is specified in order to take advantage of SAS's GENMOD Procedure with BAYES options (SAS Institute, 2009).

There are 26 parameters or priors, which are all conjugate for the likelihood function. Thus, the posterior distribution can be derived mathematically. Numerically, however, SAS uses the conjugacy sampling, that is, SAS samples directly from the target distribution (the multivariate normal posterior distribution), and the sample size of 3,000 is specified. (Otherwise, SAS uses Gibbs sampler with the adaptive rejection Metropolis sampling algorithm by default. MCMC diagnostics available are: Tables of Posterior Autocorrelations, Gelman-Rubin Diagnostics, Geweke Diagnostics, Raftery-Lewis Diagnostics, Heidelberger-Welch Diagnostics, Effective Sample Sizes, and Monte Carlo Standard Errors as well as Plots of Traces, Autocorrelations, and Posterior Distributions.) The SAS codes are provided in Appendix A.

Before graphically describing the posteriors of the regression parameters, their priors are reminded in Figure 1 with the means and the 50% and 95% ranges (the 95% range is omitted for $\beta 25$). (All the parameter plots are produced by the R package ggplot2.)

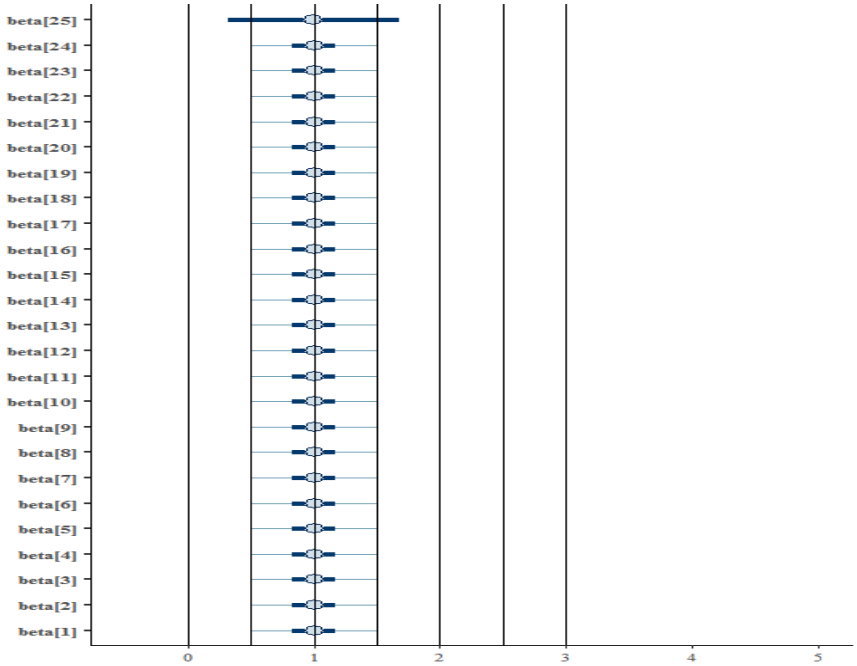

Figure 1: Prior Means with 50% and 95% Ranges of Regression Parameters

The posterior means and the 50% and 95% (equal-tail) credible intervals of the regression parameters are calculated from the 3,000 posterior sample cases and are graphed in Figure 2.

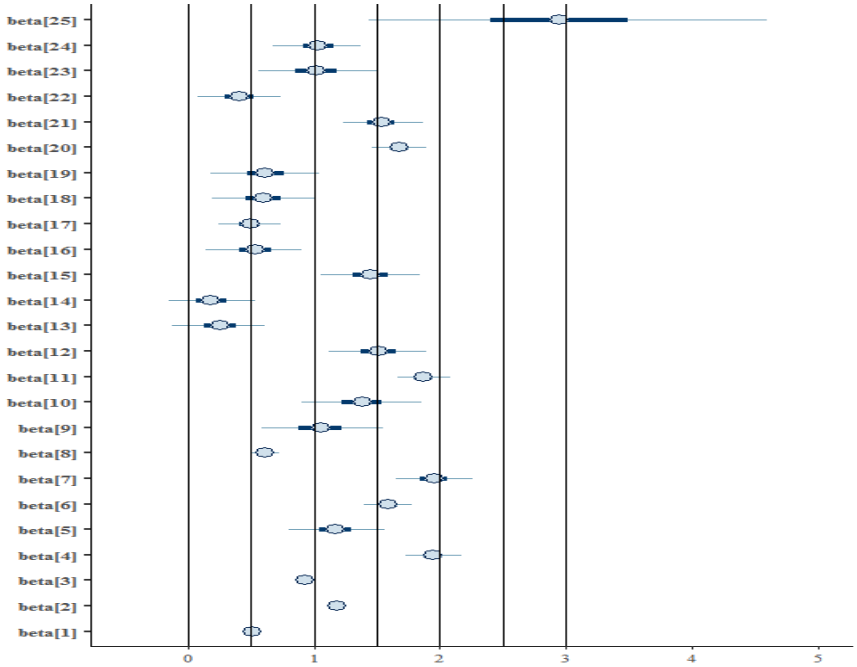

Figure 2: Model I Posterior Means with 50% and 95% (Equal-Tailed) Credible Intervals of Regression Parameters

With the data, some of the posterior means have become smaller than one (e.g., $\beta 1$, $\beta 8$, $\beta 13$, and $\beta 14$), while the others have become larger than one (e.g., $\beta 4$ and $\beta 25$). A few remain around one (e.g., $\beta 3$, $\beta 9$, $\beta 23$, and $\beta 24$). The levels of uncertainty have decreased for most of the regression parameters, that is, their

posterior normal distributions are narrower than the prior counterparts, e.g., we are much less uncertain about $\beta 1$, $\beta 2$, and $\beta 3$. The exception is $\beta 25$, about which we are much more uncertain. This could be due to the bad EEE model estimate of $x25$ or the bad prior, or both. And, the distribution of $\beta 23$ is almost unchanged. Note the lower tails of the posterior distributions of $\beta 13$ and $\beta 14$ cross 0 and recall none of the regression parameters are constrained to be positive.

These posterior means may be interpreted absolutely. For example, $\beta 1$ = 0.5 may suggest that the EEE model estimates of $x1_i$ are overestimated by a factor of 2. However, the posterior means of the regression parameters will be uniformly rescaled for each i so that the sum of calibrated end-use estimates will equal to y_i. Thus, the relative sizes of the posterior means are more important. Recall that our current problem is calibration of the EEE model estimates of end-use consumption values, not estimation of the true regression parameters.

The scale parameter σ is a nuisance parameter, but we note that its simulated posterior Gamma distribution has the mean of 4,451, the minimum of 4,401, and the maximum of 4,590.

Finally, the posterior correlation matrix of the regression parameters shows a few negative correlation coefficients, though the variance-covariance matrix Σ of their prior was assumed diagonal. Those less than or equal to -0.2 are red-colored in Table 1.

Table 1: Model I Posterior Correlation Matrix of Regression Parameters

Parameter	*beta[1]*	*beta[2]*	*beta[3]*	*beta[4]*	*beta[5]*	*beta[6]*	*beta[7]*	*beta[8]*	*beta[9]*	*beta[10]*	*beta[11]*	*beta[12]*	*beta[13]*	*beta[14]*	*beta[15]*	*beta[16]*	*beta[17]*	*beta[18]*	*beta[19]*	*beta[20]*	*beta[21]*	*beta[22]*	*beta[23]*	*beta[24]*	*beta[25]*
beta[1]	1																								
beta[2]	0.0	1																							
beta[3]	-0.4	0.0	1																						
beta[4]	0.0	-0.1	0.0	1																					
beta[5]	0.0	0.0	-0.1	-0.1	1																				
beta[6]	0.0	-0.1	-0.2	0.0	-0.1	1																			
beta[7]	0.0	0.0	0.0	-0.1	0.0	-0.1	1																		
beta[8]	-0.1	-0.1	0.0	-0.2	-0.1	-0.1	0.0	1																	
beta[9]	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	1																
beta[10]	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	1															
beta[11]	0.0	-0.1	-0.1	-0.2	0.0	-0.1	-0.1	-0.1	0.0	0.0	1														
beta[12]	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	1													
beta[13]	0.0	-0.2	0.1	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	1												
beta[14]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1											
beta[15]	0.0	-0.2	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	0.0	1										
beta[16]	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	1									
beta[17]	0.0	0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	1								
beta[18]	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	-0.2	1							
beta[19]	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	1						
beta[20]	0.0	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1					
beta[21]	0.0	0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	1				
beta[22]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1			
beta[23]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1		
beta[24]	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	1	
beta[25]	-0.1	-0.1	-0.1	-0.2	-0.1	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	0.1	0.0	0.0	0.0	-0.1	1

3. Bayesian End-Use Estimates Calibration Model II

The second Bayesian model to calibrate the twenty-five EEE end-use electricity consumption estimates against the total electricity consumption value for each respondent i = 1, …, 5,686 assumes that the total electricity value y_i is distributed folded-normally (at 0) with the mean μ_i and the standard deviation $\sigma_{h[i],d[i]}$ that is partially pooled by the housing unit type h and the Census Divisions d:

$$y_i \sim \text{Folded-Normal}\left(\mu_i, \sigma^2_{h[i],d[i]}\right),\ i = 1, \ldots, 5{,}686.$$

Note that unlike in Model I y_i are constrained to be non-negative. ($h[i]$, $d[i]$) indexes the five housing unit types and the ten Census Divisions over i:

h = 1: Mobile home
2: Single-family detached house
3: Single-family attached house
4: Apartment in a building with 2 to 4 units
5: Apartment in a building with 5 or more units

d = 1: New England
2: Middle Atlantic
3: East North Central
4: West North Central
5: South Atlantic
6: East South Central
7: West South Central
8: Mountain North
9: Mountain South
10: Pacific.

For each i, the mean μ_i is defined by the linear and additive function as in the Model I:

$$\mu_i = \beta 1 x 1_i + \cdots + \beta 25 x 25_i,$$

where $x1_i$, …, $x25_i$ are the twenty-five EEE end-use electricity consumption estimates for the respondent i, described in Section 2. The data model is a multilevel model without the intercept. The EEE model estimates $x1_i$, …, $x25_i$ are given and considered fixed for all i as before. The total consumption value y_i is always positive in the sample, and the normal distribution is folded at 0 so that the distribution is constrained to be non-negative.

The prior distributions for the first 24 regression parameters $\beta 1$, …, $\beta 24$ are assumed to be folded-normal distributions:

$$\beta 1, \ldots, \beta 24 \sim \text{Folded-Normal}(1, 0.25^2)$$

and for the last regression parameter $\beta 25$:

$$\beta 25 \sim \text{Folded-Normal}(1, 1^2).$$

These priors are assumed to be independent, and they are essentially identical to the multivariate normal prior in Model I. The only differences are the non-negativity constraints of the distributions. However, these constraints are relatively weak, given the mean and standard deviation hyper-parameters values.

Now, there are 5 × 10 = 50 scale parameters $\sigma_{h[i],d[i]}$, and they are assumed to be distributed folded-normally as:

$$\sigma_{h[i],d[i]} \sim \text{Folded-Normal}\left(s_{h,d},\ s_{h,d}^2\right),$$

where $s_{h,d}$ is the sample standard deviation of $(y_i - (x1_i + \ldots + x25_i))$ in $h[i]$ and $d[i]$ over i. (Note that the hyper-parameters are made data-dependent or specified empirically.) The standard deviation of $\sigma_{h[i],d[i]}$ is set to equal the mean of $\sigma_{h[i],d[i]}$, i.e., the coefficient of variation is specified as 1. This gives a lot of mass to the distribution from 0 to $s_{h,d}$, as the distribution is folded at 0.

Specifying a folded-normal distribution for the likelihood, the regression parameters, or the scale parameters is not possible in SAS GENMOD, nor is the hierarchical modeling. (SAS MCMC allows Bayesian analysis of a linear mixed, generalized linear mixed, or nonlinear mixed model.) With these 75 priors, Stan is used from R. "Stan® is a state-of-the-art platform for statistical modeling and high-performance statistical computation. ... Users specify log density functions in Stan's probabilistic programming language and get:

- full Bayesian statistical inference with MCMC sampling (NUTS, HMC)
- approximate Bayesian inference with variational inference (ADVI)
- penalized maximum likelihood estimation with optimization (L-BFGS)

Stan's math library provides differentiable probability functions & linear algebra (C++ autodiff). Additional R packages provide expression-based linear modeling, posterior visualization, and leave-one-out cross-validation." (Stan Development Team, 2018b). Stan "interfaces with the most popular data analysis languages (R, Python, shell, MATLAB, Julia, Stata) and runs on all major platforms (Linux, Mac, Windows)" and "is freedom-respecting, open-source software (new BSD core, some interfaces GPLv3)" (Stan Development Team, 2018b).

Three independent chains of 1,000 warm-up cases and then 1,000 sample cases (without thinning) are generated by the NUTS (No-U-Turn Sample) of Stan—3,000 posterior sample cases in total (Hoffman and Gelman, 2014). (The RStan codes are provided in Appendix B.) There are no divergent transitions diagnosed. The log posterior distribution looks reasonable. The chains have mixed well according to the $\hat{R}$ statistic and the sample draws are quite independent according to the effective sample sizes (Gelman et al., 2013). The posterior means and the 50% and 95% (equal-tail) credible intervals of the regression parameters are graphed in Figure 3.

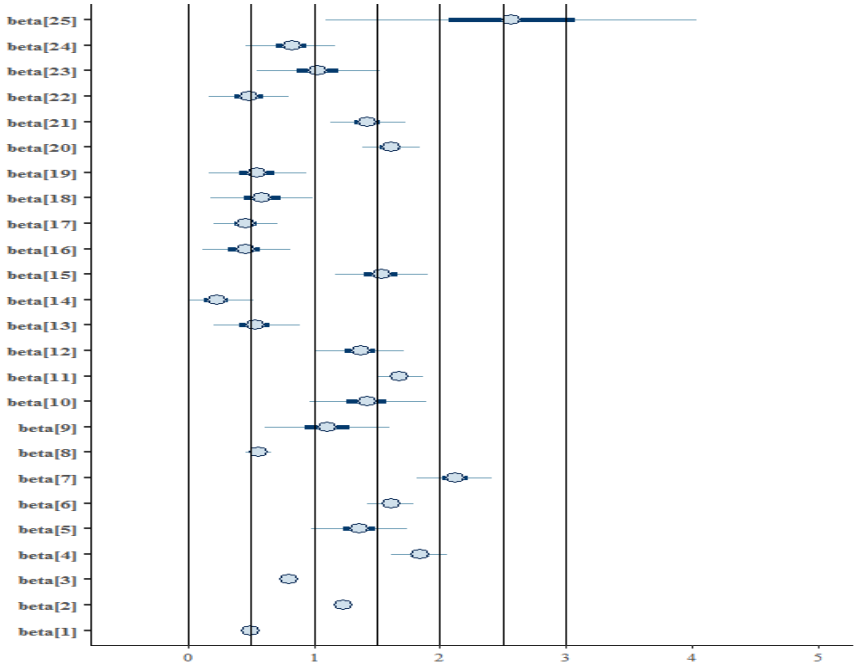

Figure 3: Model II Posterior Means with 50% and 95% (Equal-Tailed) Credible Intervals of Regression Parameters

As a whole, these posteriors turn out to be quite similar to those derived by Model I in Figure 2. Although there is some wiggling or shifting of the posterior means, the posterior spreads show similar patterns. Also, all posterior means except that of $\beta 24$ stay on the same sides with respect to the unit line. Again, these regression parameters are better to be interpreted in the relative magnitudes.

The posterior distributions of the partially pooled scale parameters $\sigma_{h[i],d[i]}$ are plotted in Figure 4, where the vertical line is drawn at the posterior mean 4,451 of the scale parameter σ in Model I.

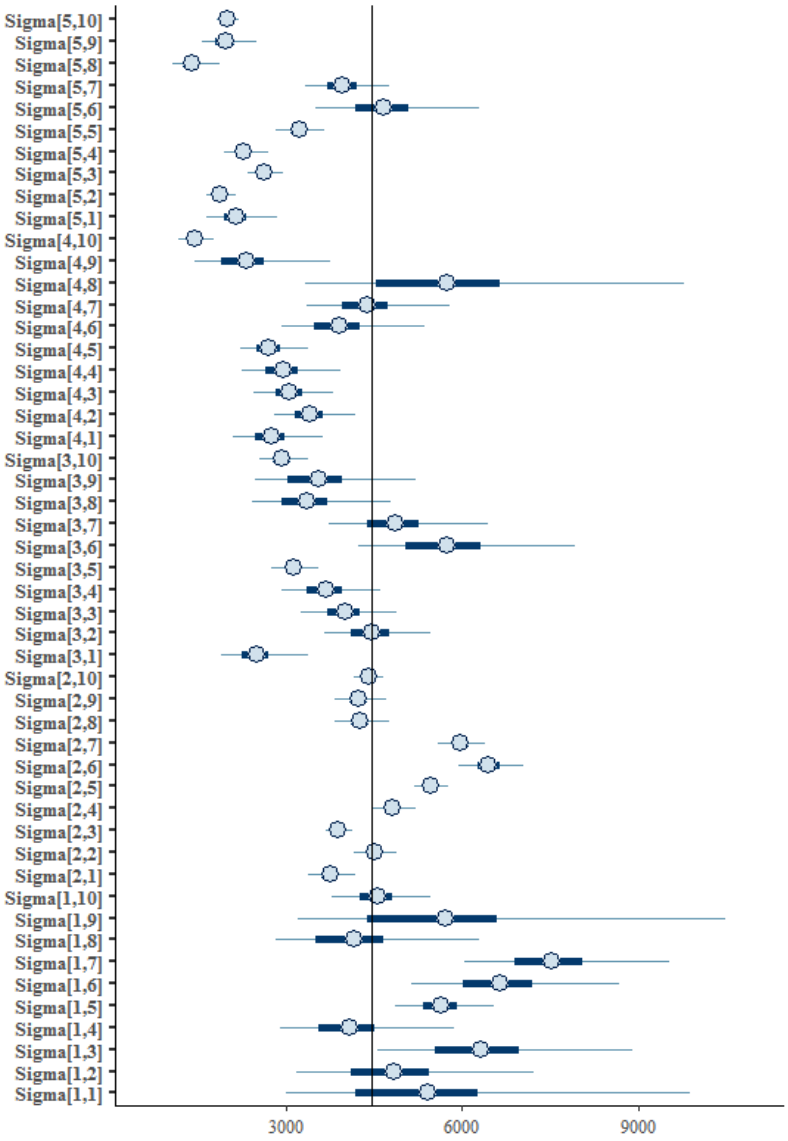

Figure 4: Model II Posterior Means with 50% and 95% (Equal-Tailed) Credible Intervals of Scale Parameters $\sigma_{h[i],d[i]}$

These posteriors are everywhere with varying uncertainty levels, but quite a few of them have tight distributions. Further pooling or regularization of the scale parameters may be possible for the housing unit types 2 (Single-family detached house) and 5 (Apartment in a building with 5 or more units).

The priors of the regression parameters were assumed to be independent, which, of course, implies their correlation coefficients were zeroes. However, the posterior correlation matrix of the regression parameters shows a few negative correlation coefficients as in Model I. The coefficients less than or equal to -0.2 are red-colored in Table 2, most of which overlap with those in Table 1.

Table 2: Model II Posterior Correlation Matrix of Regression Parameters

Parameter	*beta[1]*	*beta[2]*	*beta[3]*	*beta[4]*	*beta[5]*	*beta[6]*	*beta[7]*	*beta[8]*	*beta[9]*	*beta[10]*	*beta[11]*	*beta[12]*	*beta[13]*	*beta[14]*	*beta[15]*	*beta[16]*	*beta[17]*	*beta[18]*	*beta[19]*	*beta[20]*	*beta[21]*	*beta[22]*	*beta[23]*	*beta[24]*	*beta[25]*
beta[1]	1.0																								
beta[2]	0.0	1.0																							
beta[3]	-0.3	-0.1	1.0																						
beta[4]	0.0	-0.2	0.0	1.0																					
beta[5]	-0.1	0.0	-0.2	-0.1	1.0																				
beta[6]	0.0	-0.1	-0.2	-0.1	-0.1	1.0																			
beta[7]	-0.1	0.0	0.0	0.0	0.0	-0.1	1.0																		
beta[8]	-0.1	0.0	0.0	-0.2	0.0	-0.1	0.0	1.0																	
beta[9]	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	1.0																
beta[10]	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	-0.1	0.0	1.0															
beta[11]	0.0	-0.1	-0.1	-0.2	0.0	-0.1	0.0	-0.1	0.0	0.0	1.0														
beta[12]	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	-0.1	0.0	-0.1	-0.1	1.0													
beta[13]	0.1	-0.2	0.0	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	-0.1	1.0												
beta[14]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0											
beta[15]	0.0	-0.2	0.0	0.0	0.0	-0.1	-0.1	-0.1	0.0	0.0	-0.1	0.0	0.0	0.0	1.0										
beta[16]	-0.1	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	1.0									
beta[17]	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	1.0								
beta[18]	0.0	0.0	0.1	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	-0.1	1.0							
beta[19]	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0	1.0						
beta[20]	0.0	-0.2	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0					
beta[21]	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	-0.1	-0.1	1.0				
beta[22]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	1.0			
beta[23]	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
beta[24]	0.0	0.1	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	
beta[25]	-0.1	-0.1	-0.1	-0.3	-0.1	0.1	0.0	-0.1	0.0	0.1	-0.2	-0.2	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	0.1	0.0	0.0	0.0	-0.1	1.0

Recall that Model II has 50 parameters, while Model I has 26 parameters. Is Model II more complex or does it fit the data better? It is not clear. Gelman (2018) says, "With unregularized estimation such as least squares or maximum likelihood, adding parameters to a model (or making a model more complex) leads to overfitting. With regularized estimation such as multilevel modeling, Bayesian inference, lasso, deep learning, etc. etc., the regularization adds complexity but in a way that reduces the problem of overfitting. So traditional notions of model complexity and tradeoffs are overturned."

4. Comparisons of National Inferences of 2015 Electricity End-Use Consumption Amounts

A main checking point of the above calibration models is the national inference of the 2015 electricity consumption by the twenty-five end uses. To derive final end-use estimates from Models I and II, a ratio adjustment is applied to the calibrated EEE end-use estimates so that the final end-use estimates sum to the total for each i while keeping the calibration factors ratios constant—i.e., uniform rescaling of the posterior regression parameters in each i with the raw EEE end-use estimates being kept fixed. If the sum of the calibrated end-use estimates is greater than the total, the estimates are scaled down; otherwise, scaled up. The ratio-adjustment or uniform-rescaling factor varies from respondent to respondent and between the two Bayesian calibration models. Note that the survey weights are not constant for the respondents. Thus, the derivation of the national end-use estimates is not easily tractable conceptually, though simple computationally. Table 3 compares (1) the 2015 RECS national estimates of the twenty-five end-use electricity consumption amounts with the 50% confidence intervals, (2) the Model I national estimates by the posterior means with the 50% (equal-tailed) credible intervals, and (3) the Model II national estimates by the posterior means with the 50% (equal-tailed) credible intervals. The estimates are rounded to the nearest billions in kWh.

Table 3: Comparisons of National Inferences of 2015 Electricity End-Use Consumptions in Billion kWh

National	*RECS Method*			*Bayesian Model I*			*Bayesian Model II*		
	Estimate	*50% lower bound*	*50% upper bound*	*Mean*	*25th percentile*	*75th percentile*	*Mean*	*25th percentile*	*75th percentile*
$x1$: Space heating	187	182	192	180	179	181	180	179	181
$x2$: Air conditioning	214	210	219	210	210	211	212	211	212
$x3$: Water heating	173	170	176	173	172	173	173	172	173
$x4$: Refrigerators	89	88	89	106	106	106	105	105	105
$x5$: Cooking	18	18	18	21	21	21	21	21	21
$x6$: Clothes dryers	57	57	58	65	65	65	65	65	65
$x7$: Freezers	20	20	21	25	25	25	25	25	25
$x8$: Lighting	131	129	132	136	135	136	136	135	136
$x9$: Clothes washers	6	6	6	7	7	7	7	7	7
$x10$: Dish Washers	7	7	7	9	9	9	9	9	9
$x11$: TV and related	87	87	88	101	101	100	101	101	100
$x12$: PC and related	31	30	31	36	37	36	36	37	36

$x13$: Furnace fans	31	30	31	32	32	32	32	32	32
$x14$: Evaporative coolers	3	3	4	4	4	4	4	4	4
$x15$: Ceiling fans	23	23	23	25	26	25	25	25	25
$x16$: Other fans	18	18	18	20	21	20	20	21	20
$x17$: Dehumidifiers	15	15	16	16	16	16	16	16	16
$x18$: Humidifiers	8	7	8	8	8	8	8	8	8
$x19$: Small kitchen appliances	24	23	24	27	28	27	27	28	27
$x20$: Swimming pools	14	13	14	15	15	15	15	15	15
$x21$: Hot tubs	7	6	7	8	8	7	7	8	7
$x22$: Electric vehicles	2	2	2	2	2	2	2	2	2
$x23$: Electric parts of gas dryers	1	1	1	1	1	1	1	1	1
$x24$: CAPI Others	23	23	23	26	26	25	25	25	25
$x25$: Unknowns	79	76	82	15	15	14	14	15	14

The final national end-use consumption estimates are quite similar between Models I and II, which is expected given the similarities in the posteriors of the regression parameters. And, they are close to the final national end-use consumption estimates of the 2015 RECS. All these estimates are rather precise at least at the national level. So, we focus on the point estimates or the mean estimates. (Note that in Models I and II some of the 25th percentile estimates are larger than the mean or 75th percentile estimates. Intuitively, for a given case, the three ratio adjustments make the final end-use estimates sum to the same total. So, if one ratio adjustment (say, the 25th) makes one end-use estimate larger than what the other ratio adjustment (say, the 75th) does, it has to make the other end-use estimate(s) smaller than what the other ratio adjustment does to those end-use estimate(s). That is, these ratio adjustments are zero-sum.)

In Figure 5, the final national end-use consumption estimates of the 2015 RECS, Model I, and Model II are plotted together. The discrepancy in $x25$ (Unknowns) between the 2015 RECS estimate and those by Models I and II is very visible and substantively large. The latter estimates are quite a bit lower than the former. Although both Models I and II have produced posterior means of $\beta 25$ that are almost three (while the prior mean was 1), we started with a relatively small EEE model estimate of $x25$, compared to the other end uses, and the uniform rescaling could not take advantage of the differential calibration of the EEE model end-use estimates. As a result, the final national estimate of $x25$ by Model I or II is likely to be underestimated and became in par with other "miscellaneous" end-use estimates. Recall that large uncertainty remained in the posterior of $\beta 25$ by Model I or II, but the final national estimate of $x25$ appears quite certain. In the sense, we are certainly wrong about our national inference of $x25$, and there is some room to improve the calibration model for the EEE model estimate of $x25$ or/and the rescaling method for $\beta 25$, if not the EEE model estimate of $x25$ itself.

One alternative is to specify a much larger mean in the prior for $\beta 25$, e.g., $\beta 25 \sim$ Folded-Normal$(5, 5^2)$. The standard deviation can be made smaller if one has stronger information. Also, the scale parameter σ or $\sigma^2_{h[i],d[i]}$ could be regularized to be so small that the residuals $y_i - \mu_i$ are pushed to zeros (or less than 0.5 kWh).

Another alternative is, instead of the ratio adjustments, to accept the calibrated end-use estimates of $x1$, …, $x24$ as they are and to add the residual $y_i - (\beta 1 x1_i + \cdots + \beta 25 x25_i)$ to $\beta 25 x25_i$ in each i. That is, $x25$ absorbs not only the consumption by unknown end uses but also the EEE model estimation errors for all the end uses $x1$, …, $x25$. This latter alternative may not work well at the respondent level, but the national inferences would provide more model-dependent estimates that may be more directly checked and improved.

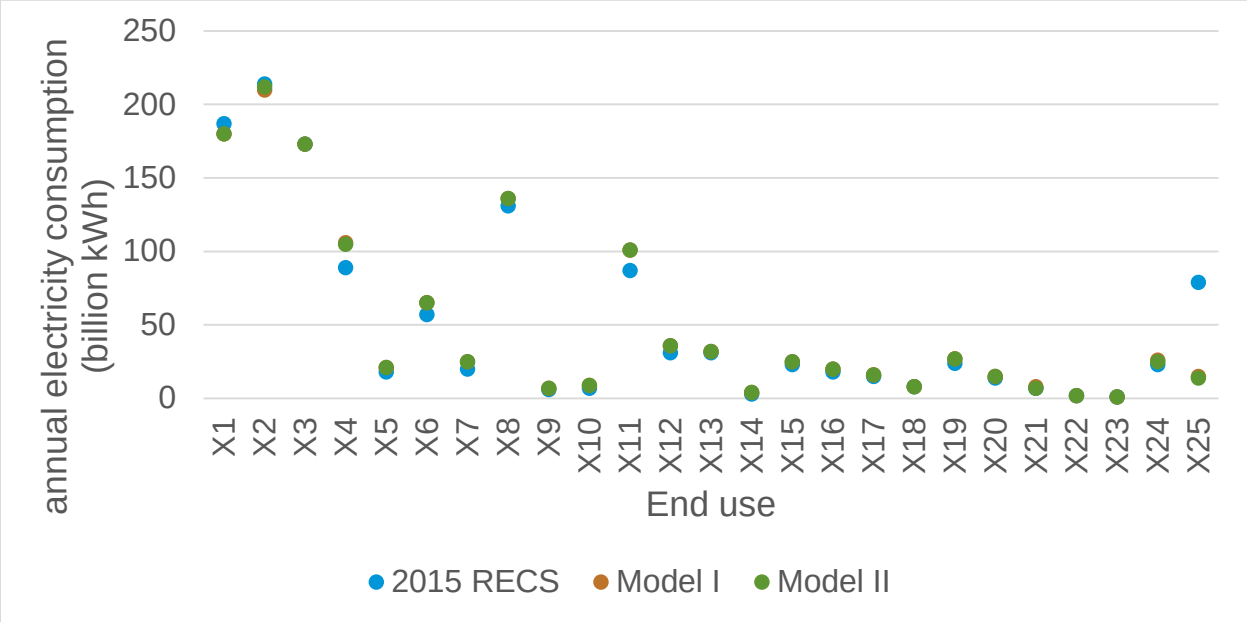

Figure 5: Final National End-Use Consumption Estimates by the 2015 RECS, Model I, and Model II

The discrepancy in $x25$ (Unknowns) between the 2015 RECS estimate and that by Model I or II must go somewhere, but it does not go to any particular end uses. It is spread over all the end uses roughly in proportion to the consumption. In the absolute term, higher national estimates by Model I or II are observed for $x4$ (Refrigerators), $x6$ (Clothes dryers), and $x11$ (TV and related). Meanwhile, Models I and II provide slightly lower estimates for $x1$ (Space heating) and $x2$ (Air conditioning) than the 2015 RECS.

Our conclusion about the 2015 RECS national end-use estimates is not binary—valid vs. not valid. We have tried two simple Bayesian calibration methods and they have led to the final national end-use estimates that are quite understandable and mostly consistent with the 2015 RECS estimates. For that reason, we consider this study has helped us increase our confidence in the 2015 RECS national end-use estimates.

4. Posterior Predictive Checking

In a Bayesian analysis, the major output is a posterior distribution of some unknown parameter or quantity given some data. Further with the posterior, however, one can predict future data y^{rep}:

$$p(y^{rep}|y) = \int p(y^{rep}|\Theta)p(\Theta|y)d\Theta,$$

where $p(\Theta|y)$ is the posterior distribution of the parameter set Θ and is proportional to the likelihood times the prior $p(y|\Theta)p(\Theta)$. Θ contains $\beta 1, \ldots, \beta 25$ and σ in case of Model I and $\sigma_{h[i],d[i]}$ in case of Model II.

The idea is to assess or measure any systematic discrepancies between y and y^{rep} in order to improve the model adequacy.

Here, we compare the 2015 RECS national end-use estimates with the replicated national end-use estimates from the posteriors of Model II *before* the ratio adjustments, i.e., with the posterior predictive distributions of the Model-II-calibrated national end-use estimates. Thus, although a ratio adjustment is not applied to each end use of i, the survey weight is used. Because of the omission of ratio adjustments, the checking of the Bayesian calibration model is rather direct. In Figure 6, we graphically conduct such posterior predictive checking for, e.g., $x1$ (Space heating) with the R package ggplot2.

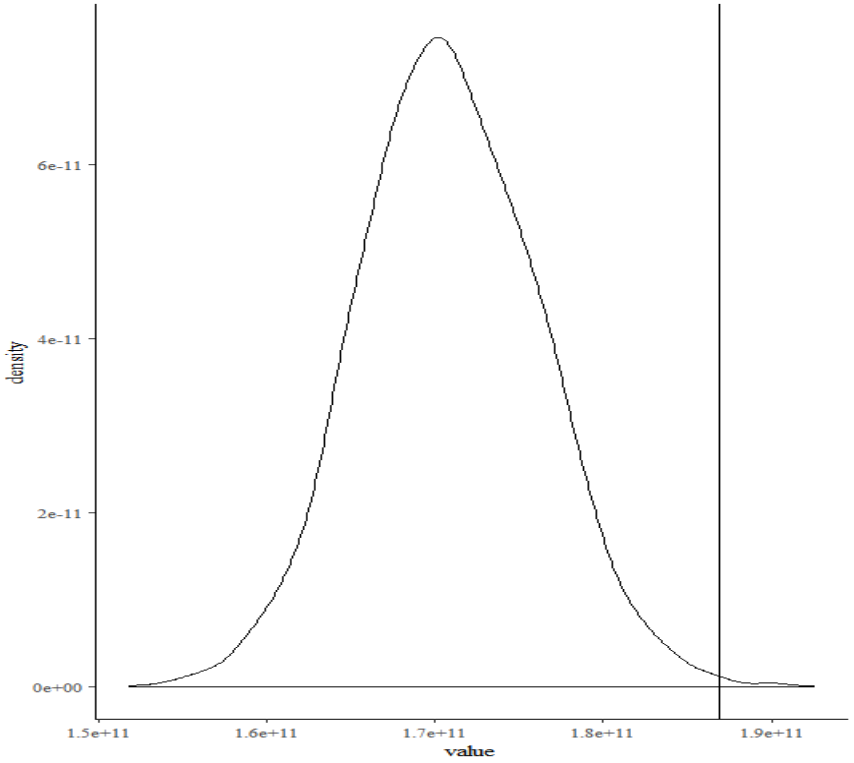

Figure 6: Posterior Predictive Distribution of Model II Calibrated National Estimate of $x1$ (Space Heating) with Corresponding 2015 RECS Estimate at Vertical Line

We can see that the Model II calibration model does not predict the 2015 RECS estimate of $x1$ (Space heating) very well. The posterior mean of the regression parameter $\beta 1$ was about 0.5, and maybe it was too low. The ratio adjustment of $\beta 1$ has managed to make ends meet.

Finally, Figure 7 graphs the density of the 2015 RECS national estimates of $x2$ (Air conditioning) overlaid with fifty replicate densities from Model II (without ratio adjustments) with the R package bayesplot.

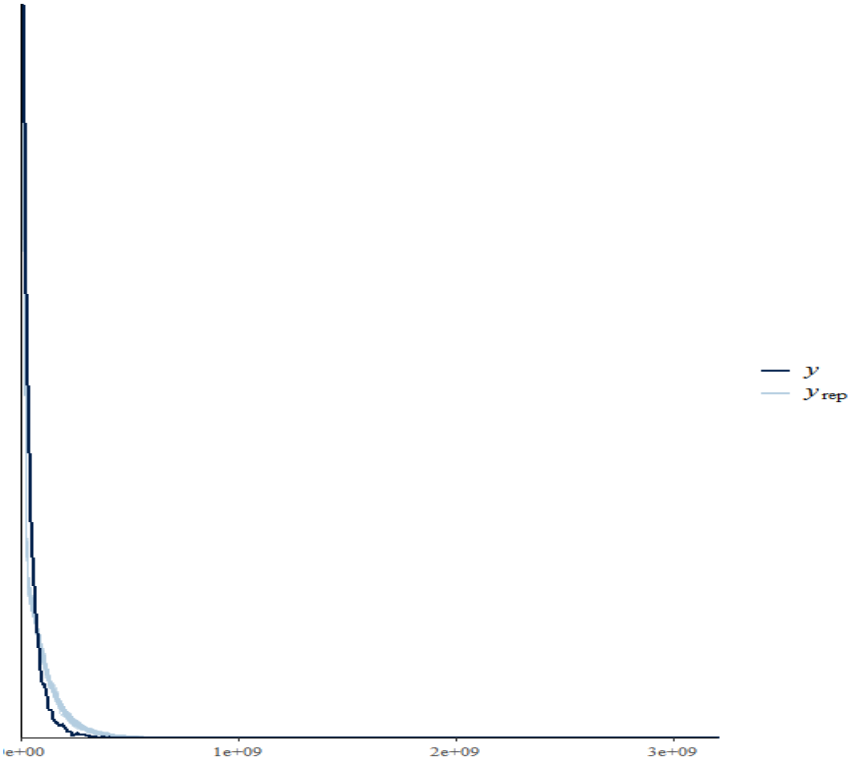

Figure 7: Density of 2015 RECS National Estimate of $x2$ (Air conditioning) Labelled by y and Fifty Replicate Densities from Model II (Without Ratio Adjustments) Labelled by y_{rep}

The band of replicate densities does not quite cover the density of the 2015 RECS national estimate of $x2$ (Air conditioning), though they are close. Recall that the posterior mean of the regression parameter $\beta 2$ was about 1.2. The overestimation by Model II was corrected by the ratio adjustment as well in this example.

The calibration models can be improved to describe better the relative relationships among the EEE model end-use estimates, as briefly discussed in the previous section. At this time, however, the ratio adjustments seem to play a bigger role than the calibration models in adjusting the EEE model end-use estimates. Perhaps, the calibration models can presuppose some systematic differences in $(y_i - (x1_i + \cdots + x25_i))$ or in $(y_i - (\beta 1 x1_i + \cdots + \beta 25 x25_i))$, whose prior mean was earlier assumed to be zero. In other words, we might restart by assuming there is some consumption in y_i that is not accounted for by any of $x1_i$, …, $x25_i$ and we might explicitly include some intercepts in the mean function μ_i. For example, we could specify:

$$\mu_i = \alpha_{h[i],d[i]} + \beta 1 x1_i + \cdots + \beta 25 x25_i,$$

where i and $h[i], d[i]$ are defined as before. The priors for the intercepts could be:

$$\alpha_{h[i],d[i]} \sim \text{Normal}\left(m_{h,d},\ s_{h,d}^2\right),$$

where $m_{h,d}$ and $s_{h,d}$ are respectively the sample average and the sample standard deviation of $(y_i - (x1_i + \ldots + x25_i))$ in $h[i]$ and $d[i]$ over i, cf. the Model II priors for $\sigma_{h[i],d[i]}$. Note that this normal distribution is not constrained to be positive, as the differences and even $m_{h,d}$ can be positive or negative. Then, the scale parameter σ can just regulate the residuals to be mostly within (0 ± 0.5) or rounding errors:

$$\sigma \sim \text{Folded-Normal}(0, 0.25^2).$$

The level of pooling of the intercepts controls the level of data fitting or estimation by the calibration models. If the intercept is made unique to each i, the estimation is basically customized for i.

These formulations help us identify and estimate any biases in y_i, but if we are not allowed to modify y_i, then those biases would have to be re-distributed back to $\beta 1 x 1_i, \ldots, \beta 25 x 25_i$ at the end. In a larger context of measuring errors, modeling errors, and weighting errors, Minato (2017) conducts some Bayesian analysis.

Appendix A. SAS Codes for Bayesian Model I

```
data input025_1;
length _NAME_ $ 3 _TYPE_ $ 4;
input  _NAME_ $   _TYPE_ $    x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15 x16 x17 x18 x19 x20 x21 x22 x23 x24 x25;
datalines;
.   MEAN 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
x1  COV  0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x2  COV  0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x3  COV  0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x4  COV  0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x5  COV  0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x6  COV  0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x7  COV  0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x8  COV  0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x9  COV  0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x10 COV  0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x11 COV  0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0 0
x12 COV  0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0 0
x13 COV  0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0 0
x14 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0 0
x15 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0 0
x16 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0 0
x17 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0 0
x18 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0 0 0
x19 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0 0
x20 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0 0
x21 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0 0
x22 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0 0
x23 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625 0
x24 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0625
x25 COV  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
;

proc genmod data = recs2015w;
model y = x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15 x16 x17 x18 x19 x20 x21 x22 x23 x24 x25
  /dist = normal link = identity noint;
bayes
  seed        = 2017
  outpost     = postsurg
  coeffprior  = normal(input = input025_1)
  scaleprior  = gamma(shape = 1 iscale = 0.001)
  diagnostics = (autocorr ess Geweke Heidelberger mcerror Raftery Gelman(nchain = 3))
  thinning    = 1
  nbi         = 1000
  nmc         = 3000
  statistics (alpha = 0.05 percent = 2.5, 5, 25, 50  75, 95, 97.5) = (corr cov summary interval);
store model_store;
code file = 'model_store.sas';
run;
```

Appendix B. RStan Codes for Bayesian Model II

```
stanmodelcode <- "
data {
  int<lower = 0>          N;                    // The number of respondents (housing units)
  int<lower = 1>          K;                    // The number of end uses
  int<lower = 1>          TYPEHUQ[N];           // The housing type variable
  int<lower = 1>          H;                    // The number of housing types in TYPEHUQ
  int<lower = 1>          DIVISION[N];          // The Census Division variable
  int<lower = 1>          D;                    // The number of Census Divisions in DIVISION
  matrix<lower = 0>[N, K] x;                    // The EEE model end-use estimates as the predictors
  vector<lower = 0>[N]    y;                    // The annualized total energy consumption (kWh)
  matrix<lower = 0>[H, D] sd_y_minus_x_sum_by;  // The sample std. dev. of (y - the sum of x) in [h, d]
}
parameters {
  vector<lower = 0>[K]    beta;   // The regression coefficients, completely pooled
  matrix<lower = 0>[H, D] Sigma;  // The regression error terms, partially pooled
}
transformed parameters {  // Linking the predictors to the mean of the distribution of y
  vector<lower = 0>[N] mu;
  for (n in 1:N) {
    mu[n] = x[n] * beta;
  }
}
model {
  for (k in 1:(K - 1)) {
    beta[k] ~ normal(1, 0.25);  // The priors of the regression coefficients for x1, .., x24
  }
  beta[K] ~ normal(1, 1);       // The prior of the regression coefficient for x25
  for (h in 1:H) {
    for (d in 1:D) {
      Sigma[h, d] ~ normal(sd_y_minus_x_sum_by[h, d], sd_y_minus_x_sum_by[h, d]);
    }  // The folded-normal distributions with the sample standard deviation of the residuals
  }
for (n in 1:N) {
  y[n] ~ normal(mu[n], Sigma[TYPEHUQ[n], DIVISION[n]]);  // The data model or likelihood
  }
}
generated quantities {  //The posterior predictive distributions of y, beta1x1, ..., beta25x25
  real y_rep[N];
  matrix[N, K] x_rep;
  for (n in 1:N) {
    y_rep[n] = normal_rng(mu[n], Sigma[TYPEHUQ[n], DIVISION[n]]);
    if (y_rep[n] < 0) y_rep[n] = 0;  // Forcing y_rep to be non-negative
    for (k in 1:K) {
      x_rep[n, k] = normal_rng(x[n, k] * beta[k], Sigma[TYPEHUQ[n], DIVISION[n]]);
      if (x_rep[n, k] < 0 || x[n, k] == 0) x_rep[n, k] = 0;
    }  // Forcing x_rep to be nonnegative or 0 when x is 0
  }
}
"
```

```
set.seed(20170801)
fit <- stan(model_code = stanmodelcode,
            model_name = "EUMEC",
            iter       = 2000,
            warmup     = 1000,
            thin       = 1,
            chains     = 3,
            verbose    = false,
            control    = list(max_treedepth = 10, adapt_gamma = 0.05, adapt_delta = 0.8, adapt_kappa = 0.75),
            init       = "random",
            algorithm  = c("NUTS", "HMC", "Fixed_param")
)
```

Acknowledgements

I would like to thank Thomas Leckey, the Assistant Administrator of Energy Statistics at the U. S. Energy Information Administration, for his valuable comments to improve the paper, while any errors are all mine.

References

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].

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2018), "Visualization in Bayesian workflow," arXiv:1709.01449v5 [stat.ME].

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

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. (2018), "Comments on Limitations of Bayesian Leave-One-Out Cross-Validation for Model Selection," Statistical Modeling, Causal Inference, and Social Science Blog, 5 June, Comment. Available at *http://andrewgelman.com/2018/06/05/comments-limitations-bayesian-leave-one-cross-validation-model-selection/#comment-757695*.

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., 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.

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.

Kennedy, M. C., and O'Hagan, A. (2001), "Bayesian Calibration of Computer Models," Journal of the Royal Statistical Society, *Series B*, 63, 425-464.

Minato, H. (2017), "Bayesian Modeling of Uncertainties in End-Use Electricity Consumption Amounts," 2017 Proceedings of the American Statistical Association, Government Statistics Section [CD-ROM], Alexandria, VA: American Statistical Association: 3131-3150.

SAS Institute (2009), "The GENMOD Procedure," in SAS/STAT(R) 9.2 User's Guide (2nd ed.). Available at *https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_genmod_sect012.htm*.

Stan Development Team (2017a), *ShinyStan: Interactive Visual and Numerical Diagnostics and Posterior Analysis for Bayesian Models* (version 2.4.0). Available at *http://mc-stan.org/users/interfaces/shinystan*.

Stan Development Team (2017b), *Stan Modeling Language Users Guide and Reference Manual* (version 2.17.0). Available at *http://mc-stan.org/users/documentation/index.html*.

Stan Development Team (2018a), *RStan: The R interface to Stan* (version 2.17.3). Available at *http://mc-stan.org/users/interfaces/rstan.html*.

Stan Development Team (2018b), *Stan*. Available at *http://mc-stan.org*.

Stan Development Team (2018c), *Stan Wiki: Prior Choice Recommendations*. Available at *https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations*.

U. S. Energy Information Administration (2018a), *2015 RECS Survey Data*. Available at *https://www.eia.gov/consumption/residential/data/2015*.

U. S. Energy Information Administration (2018b), *Residential Energy Consumption Survey (RECS) 2015 Consumption and Expenditures Technical Documentation Summary*. Available at *https://www.eia.gov/consumption/residential/reports/2015/methodology_c&e*.

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