Bayesian Semi-Parametric Energy Balance Measurement Error Models

Daniel Ries

January 16, 2017

1 Introduction

We are interested in formulating a model for one's energy balance by using energy expenditure (EE) and changes in energy stores (Δ ES) while accounting for dependence between the two and measurement error. We can then use this to calculate one's energy intake (EI) via the energy balance equation. Recall that measuring one's EI is an incredibly hard task, and as such, there is no true accepted gold standard. Because of this, we are using the energy balance principle which states that any energy consumed during a fixed period of time must either be burned or stored. Fortunately, commonly accepted gold standard measurements exist for both EE (DLW) [1],[2],[3],[4],[5],[6], [7] and ΔES (DXA) [3],[6],[5]. Unfortunately, these measures are expensive and are more involved than most would like. As we discussed earlier, there are cheaper alternative measures for both EE and ΔES for which researchers would rather be able to use. Our goal is to model energy balance by using both gold standard measures and cheap measures with the end goal of evaluating the error present in measurements and ultimately calibrating cheaper measurements, so in future studies, researchers can calibrate their measurements of EE and/or ΔES if they are not using a gold standard measurement.

In this chapter we take a Bayesian semi-parametric approach to modeling this problem. We will make distributional assumptions about error terms, but we try to be flexible in modeling the true relationship between cheap measurements and the truth. This analysis requires replicate observations on gold standard and cheap measurements on each individuals, which is expensive and time

1

consuming in terms of data collection, but this method also provides a detailed solution that accounts for all forms of uncertainty. We use data from the Energy Balance Study [8] to assist in constructing a model.

2   EBS Data

The Energy Balance Study [8] was conducted in 2011 and 2012 at the University of South Carolina. Over the span of a year for an individual, measurements of EI, EE, and ΔES were taken on 430 men and women between the ages of 21 and 35 and BMI's ranging from 20 to 35. Study participants received 5 DXA scans, one at baseline, and one every 3 months thereafter. To measure EI, multiple 24 hour telephone recalls were administered, a set at baseline and then a set every 3 months up to a year. Average values were recorded every 3 months. Lastly, EE was measured with Sensewear Armbands over ten day periods, again one at baseline, and then a 10 day period every 3 months after for up to a year. During the assessment period every three months, other measures such as weight, BMI, hip and waist circumference were taken as well. On a subset of the participants ($n = 191$), DLW was administered after their 12 months of the study was complete. At the beginning end of the two week period required for DLW, DXA scans were taken. All EE and EI measurements were recorded in kcals and were averaged over the assessment period to a single, daily average. DXA gives measurements of fat mass and fat free mass in kg at a certain point in time.

2.1   Calculation of ΔES

In our model, we need ΔES to be in terms of kcals, positive or negative. To convert DXA measurements of fat mass and fat free mass to kcals, we use equation (1). Because we are assuming energy stores are characterized only as either fat mass (FM) or fat free mass (FFM), this equation is exact if we know the values of c_1 and c_2. Recall these coefficients represent the energy in kcals per kg of FM and FFM, respectively. It was discussed previously that there is no exact agreed upon set of values for these terms, so at this point for our exploratory exercise we will let $c_1 = 9500$ and $c_2 = 1100$ like [3] uses. Then we can calculate the change in FM and change in FFM because we have 2 DXA measurements 2 weeks apart on a subsample. Then we divide these by the exact change in time (14 days $\pm$ 3 days) and multiply by c_1 and c_2 to get ΔES in kcals. Our result for each

individual will be either a positive or negative number in kcals.

$$\Delta ES = c_1 \frac{\Delta FM}{\Delta T} + c_2 \frac{\Delta FFM}{\Delta T} \qquad (1)$$

2.2 Replicate Measurements

Because of the way these data were collected, we do not technically have replicates on individuals for the quantities we're interested in. We can only consider these data as having a replicate on an individual if we assume there's no systematic changes/differences over the 3 month span between measurements; ie. individuals don't change their exercise or consumption habits over the 3 month period of time. This would be a rather extreme assumption because individuals know they are involved in a health study and thus have some conscious or subconscious reason to change over the time. Also 3 months is a rather long period of time when talking about exercise and diet (different seasons, holidays, etc.). Because of these reasons, we are weary to make this assumption to estimate model parameters. However, we will make this assumption now for exploratory purposes only.

2.3 Characteristics of EBS Data

The goal in this exploratory exercise is to determine the general form and distribution of the variables as well as check some common measurement error model assumptions. We also want to look at the underlying relationships between measurements as well as with demographic factors. In Figure 1 we plot the distribution of EE (left) and ΔES (right) as measured by gold standard instruments faceted by gender. Values for these plots were from the subsample receiving DLW only. EE by DLW is slightly skewed right with males having systematically greater EE. ΔES by DXA is fairly symmetric and has no obvious systematic differences between genders.

In Figure 2 we have diagnostic plots to check whether constant variance would be a reasonable assumption for these specific measures of EE and ΔES, respectively. Both these plots have the mean based on replicates for an individual on the x axis and that same individual's standard deviation based on replicates on the y axis. The left plot is for EE as measured by the Sensewear Armband and the right plot is for ΔES as measured by DXA. Looking at the EE plot, we see some

signs of nonconstant variance, with variability increasing with EE. The ΔES plot doesn't show much in terms of nonconstant variability. This is again based on the subsample.

Figure 3 shows normal quantile plots for DLW measured EE split by gender, Females on left and Males on right. Note this is not the usual measurement error diagnostic plot suggested in Carroll et al. [9] since it is based on actual EE values and not differenced values, however we don't have replicates for DLW measurements. By splitting this measure by gender, we get approximately normally distributed values as opposed to a distribution with fatter tails. Because DLW is the gold standard for EE, or what we referred to as W^{EE}, we have at least some empirical evidence of normality for these measurement errors.

Figure 4 shows normal quantile plots for differenced DXA measurements of ΔES split by gender, Females on left and Males on right. We calculated the differenced ΔES for 6 month DXA to 9 month DXA and 9 month DXA to 12 month DXA and took the difference. Both plots look similar and show small signs of non normality in the tails, but nothing too extreme. Figure 5 shows normal quantile plots for differenced armband measured EE split by gender.We calculated the differenced EE by taking the 12 month measurement minus the 9 month measurement. Females on left and Males on right. Again, there are no gross departures from normality.

We also wanted to check for evidence of systematic bias in the cheap measurement of EE, which based on the EBS data is the Sensewear Armband. Using only the subset of individuals who received DLW at the end of the study, we ran a multiple regression using DLW measured EE, age, BMI, and gender as covariates. Regression results are given in Table (1). Unsurprisingly, the coefficient on DLW measured EE was highly significant and close to 0.55. A simple linear regression of EE as measured by armband on EE as measured by DLW results in a slope of 0.78 which is still considerably less than 1, which would indicate an unbiased estimator. The coefficients on gender and BMI also were highly significant as indicated by small p-values and large effect sizes (305 for males and 14 for BMI). Age was only weakly significant, but again the EBS study only included a small age group and this effect could change with a larger population. We also like the interpretability of the coefficients and because of that we will include age as well as a linear bias term in our model. We understand that we are doing this regression using DLW measured EE and not *true, usual* EE as specified in (3),(4). Obviously we can never observe the truth and the DLW measurement is the next best, not to mention we are considering DLW to be an unbiased measurement. This simple
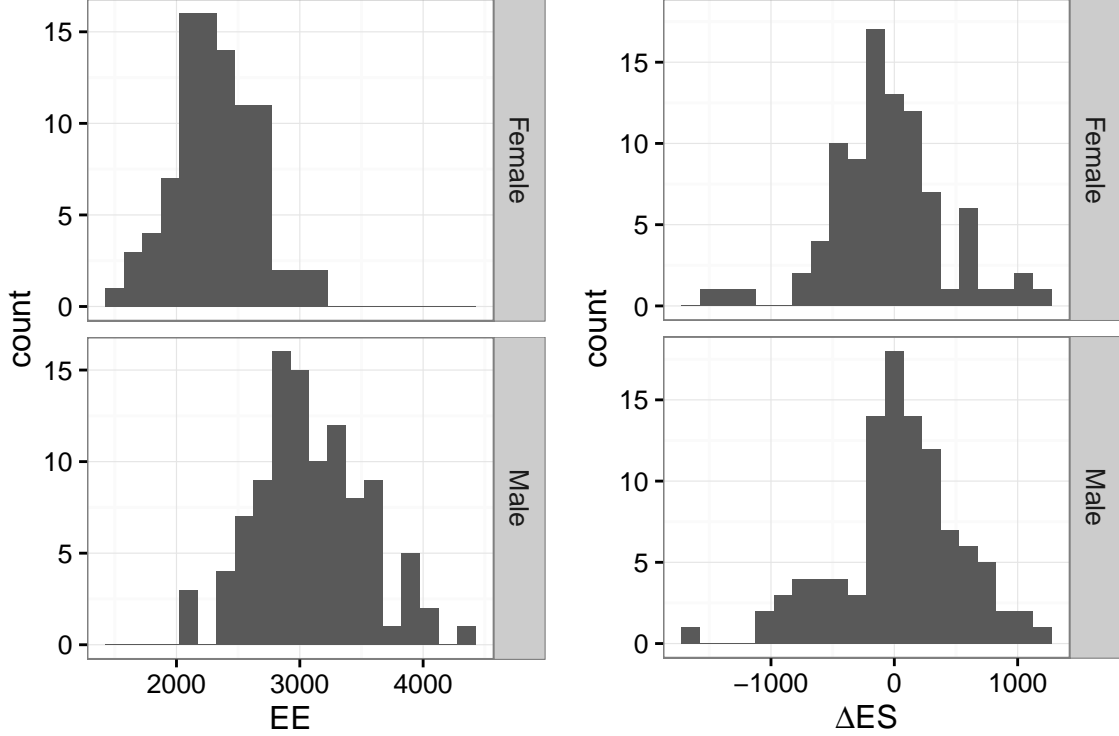
Figure 1: Distribtion of EE (left) measured by DLW and ΔES measured by DXA (right). Females on top, Males on bottom. Includes data from EBS from individuals who receieved DLW measurements.

regression model provides evidence that systematic biases exist in cheap measurements of EE, and we will make the assumption that the same holds true for cheap ΔES measurements.

3 Methodology

3.1 Notation

We will denote the measured average daily EE by DLW for subject i over time period j, W_{ij}^{EE}, and average daily ΔES measured by DXA for subject i over time period j, $W_{ij}^{\Delta ES}$. A positive value for ΔES indicates that more food was consumed than burned. We are averaging these values because DLW gives an estimate of EE over the course of approximately 14 days, so to keep things simple and comparable, we will always talk about average kcals per day.

We also have cheaper measurements for EE and ΔES. We introduced some of these commonly used methods in the first chapter. When collecting data on a large population, it makes more sense to collect these cheaper measurements on most of the subjects. However, they are cheaper for a
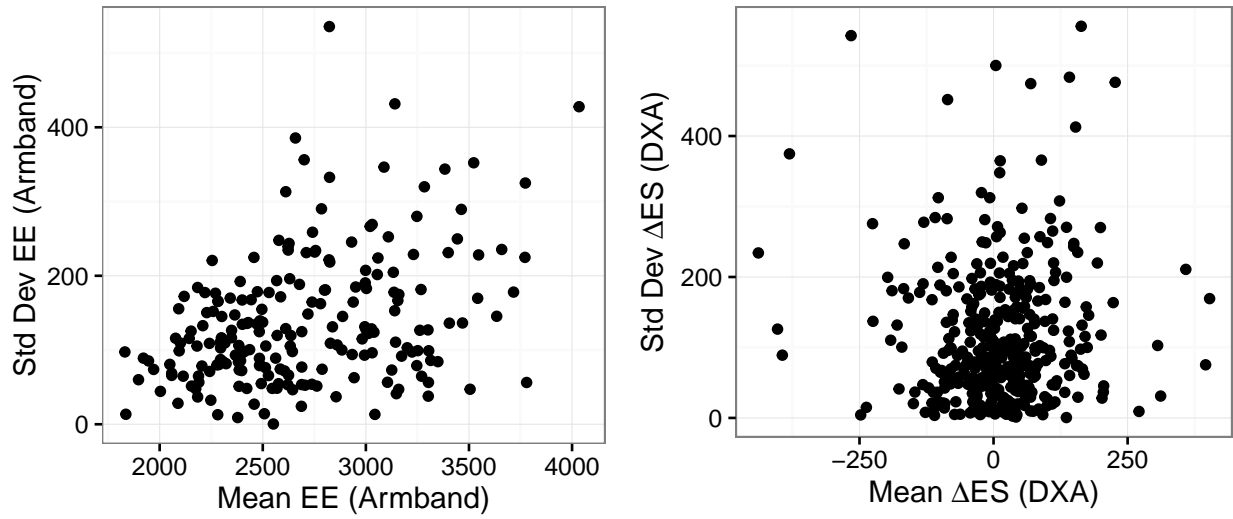
Figure 2: Variance Diagnostic Plots. The left shows the mean EE (Armband) for an individual vs his/her standard deviation of EE based on replicates. The right is the same except for ΔES measured by DXA.
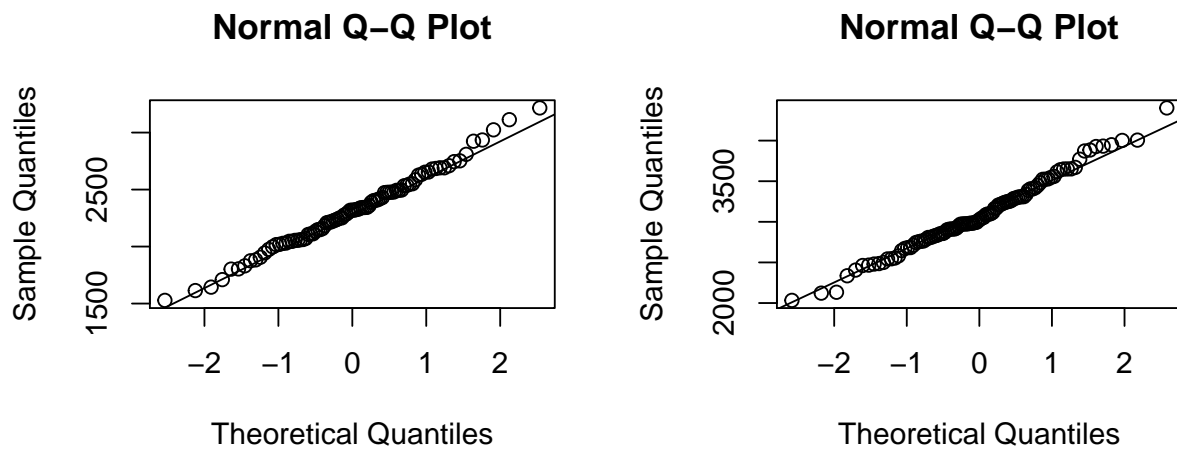

Figure 3: Normal Quantile Plots for EE as measured by DLW for Female (left) and Male (right). Data is from those who recieved DLW in EBS.

Normal Q–Q Plot

Sample Quantiles

Theoretical Quantiles

Normal Q–Q Plot

Sample Quantiles

Theoretical Quantiles
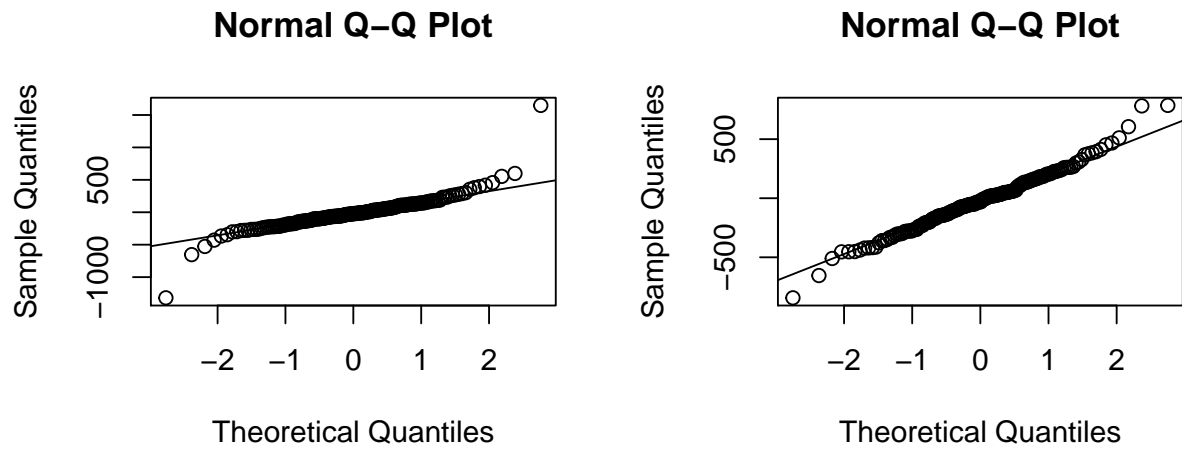
Figure 4: Normal Quantile Plots for differenced ΔES ((12m DXA - 9m DXA)-(9m DXA - 6m DXA)) as measured by DXA for Female (left) and Male (right). Data is from those who had DXA scans at 6,9,12 months

Normal Q–Q Plot

Sample Quantiles

Theoretical Quantiles

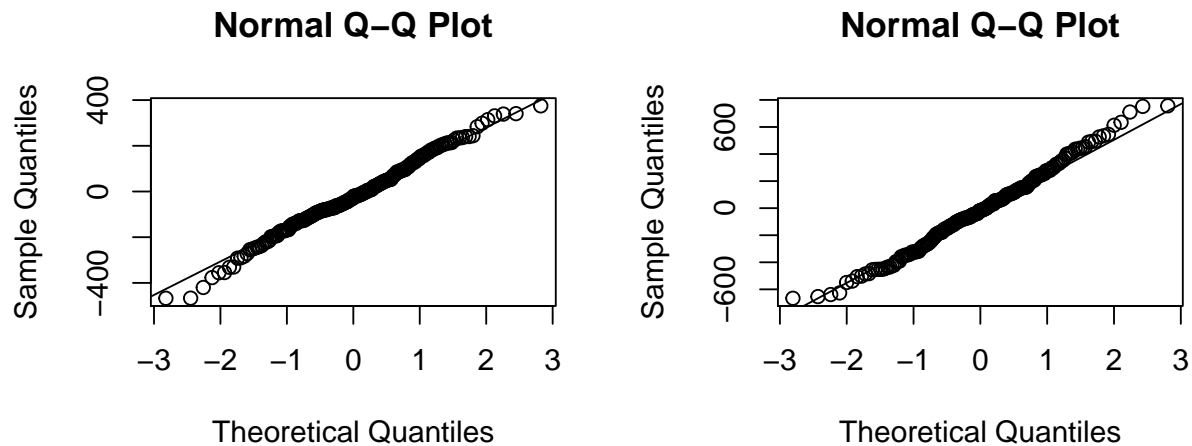**Normal Q–Q Plot**

Sample Quantiles

Theoretical Quantiles

Figure 5: Normal Quantile Plots for differenced EE (12m EE - 9m EE) as measured by armbands for Female (left) and Male (right). Based on entire EBS individuals who remained for entire 12 months

Table 1: Regression of Sensewear Armband EE on DLW EE, Age, BMI, Gender using EBS data. Results show systematic biases could exist in cheap EE measurements.

	Dependent variable:
	Energy_expenditure
TDEE	0.559***
	(0.040)
age	−7.351*
	(4.000)
genderMale	305.583***
	(43.259)
BMI	14.146***
	(3.989)
Constant	878.423***
	(154.113)
Observations	191
R^2	0.830
Adjusted R^2	0.826
Residual Std. Error	206.875 (df = 186)
F Statistic	226.938*** (df = 4; 186)
Note:	*p<0.1; **p<0.05; ***p<0.01

reason, and ultimately, we would like to use our model assess the measurement error involved and calibrate these cheaper measurements to make them more accurate and precise. Although there are multiple different cheap ways to measure EE and ΔES, we will keep the notation here general since in any given situation we will be talking only about one specific measurement tool of each. Therefore we will not talk about specific measurement tools here, rather we lump them together as cheap measurements. We will denote the measured average daily EE by a cheaper method for subject i over time period j, Y_{ij}^{EE}, and average daily change in energy stores measured by a cheaper method for subject i over time period j, $Y_{ij}^{\Delta ES}$. In application, it will be clear in every situation which tool was used to measure the Y's.

Lastly, the values which we never observe are the *usual* EE and ΔES for subject i. We define *usual* as a long run average of the true EE and ΔES. Let X_i^{EE} represent the *usual* daily EE for subject i and $X_i^{\Delta ES}$ represent the *usual* daily ΔES for subject i. Note there is considerable within person variability to this *usual* value. However, this value is of not much interest itself. We refer to the *truth* for individual i during time period j as being individual i's *usual* value plus their within person variation during time period j.

3.2 (In)Dependence Assumptions

Our observed data for subject i over time j will consist of the values $(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES}, Z_i)$ where Z_i is a vector of (assumed) perfectly measured covariates for subject i. A joint probability model for these observables is represented as:

$$f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES} | Z_i, \boldsymbol{\theta}) \tag{2}$$

Because of all the dependencies involved, we will deconstruct this via conditional independence assumptions. Given the value X_i^{EE} (*usual* daily EE for subject i) and $X_i^{\Delta ES}$ (*usual* daily ΔES for subject i), and covariates Z_i, we assume that:

1. Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$ are independent of each other and

2. Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$ are each independent from both W_{ij}^{EE} and $W_{ij}^{\Delta ES}$

3. W_{ij}^{EE} and $W_{ij}^{\Delta ES}$ are independent of each other

   Additionally, given that the above assumptions are reasonable, we expand these by additionally assuming

4. Given only X_i^{EE} and Z_i, Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$ are independent of each other and from both W_{ij}^{EE} and $W_{ij}^{\Delta ES}$ and vice versa for ΔES

5. Given only X_i^{EE} and Z_i, W_{ij}^{EE} is independent from $W_{ij}^{\Delta ES}$ and vice versa for ΔES

This is because (1): given the true values X and covariates Z, knowing a cheap measurement one will give you no more information about the cheap measurement of the other, so long as it is not self-administered. And (2): once we know the truth X, having an unbiased measurement of X will not provide any more information to the cheap, biased measurement of X. DLW and DXA are conducted by separate personnel and these are done separately from those administering the cheap measurement methods. (3) follows the same lines as (1), and again, DXA and DLW are performed by different people.

We are comfortable with conditional independence assumptions (1)-(3) and their validity will not be discussed again. The last two conditional independence assumptions are not as clear cut. It does seem possible that the true change in body composition (ΔES) could affect how well an activity monitor records EE, even after knowing the true EE and covariates. It also seems possible, but slightly less likely that the same is true for the gold standard measurements. We need to remember this assumption when making conclusions later on.

3.3 Model Likelihood

We further specify the models for the W's and Y's as:

$$Y_{ij}^{EE} = m_{ee}(X_i^{EE}, Z_i) + \epsilon_{ij}^{EE} \tag{3}$$

$$Y_{ij}^{\Delta ES} = m_{es}(X_i^{\Delta ES}, Z_i) + \epsilon_{ij}^{\Delta ES} \tag{4}$$

$$W_{ij}^{EE} = X_i^{EE} + \nu_{ij}^{EE} \tag{5}$$

$$W_{ij}^{\Delta ES} = X_i^{\Delta ES} + \nu_{ij}^{\Delta ES} \tag{6}$$

where $m_{ee}(\cdot)$ and $m_{es}(\cdot)$ are assumed to be general, smooth functions. ν_{ij}^{EE} and $\nu_{ij}^{\Delta ES}$ represent the measurement error for subject i during time period j using DLW and DXA, respectively, and within person variability. For now, we let ϵ_{ij}^{EE} and $\epsilon_{ij}^{\Delta ES}$ represent the measurement error for subject i during time period j using cheap measurements of EE and ΔES, respectively, and within person variability. This is a working assumption since the within person variability is in reality not an additive error, rather it is added to the usual and realized through the functions $m_{\cdot}(\cdot)$. We assume the expected values of all ϵ, ν are 0. This implies we are assuming DLW and DXA are providing unbiased estimates for the *usual* daily EE (X_i^{EE}) and ΔES ($X_i^{\Delta ES}$), respectively, and the cheap measurements are giving biased estimates of (possibly) unknown form.

We assume subjects are independent of each other. Additionally, conditional on each subjects' respective *usual* value X_i, replicate measurements within subject are independent. Independence within individuals is justified because it is conditional *on each individuals usual value*, which means once we know an individual's baseline, we assume their day to day variation to be independent from each other. One can imagine where this assumption might fail, for example, if we measured EE and ΔES only in December we might expect the day to day variation to during that time to be consistently negative (less exercise than *usual*) for EE and positive for ΔES (gaining weight), so we will keep in mind this is a simplifying but not always justified assumption. The likelihood function for individual i can then be written as:

$$
\begin{aligned}
L_i(\boldsymbol{\theta}) &= \prod_{j=1}^{J} f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES} | Z_i, \boldsymbol{\theta}) \\
&= \int_{\mathcal{X}_{es}} \int_{\mathcal{X}_{ee}} \prod_{j=1}^{J} f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES}, X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta}) dX_i^{EE} dX_i^{\Delta ES} \\
&= \prod_{j=1}^{J} \int_{\mathcal{X}_{es}} \int_{\mathcal{X}_{ee}} f(W_{ij}^{EE} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wee}}) f(W_{ij}^{\Delta ES} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wes}}) \times \\
&\quad f(Y_{ij}^{EE} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yee}}) f(Y_{ij}^{\Delta ES} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \times \\
&\quad f(X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta_x}) dX_i^{EE} dX_i^{\Delta ES} \\
&= \prod_{j=1}^{J} \int_{\mathcal{X}_{es}} \int_{\mathcal{X}_{ee}} f(W_{ij}^{EE} | X_i^{EE}, Z_i, \boldsymbol{\theta_{wee}}) f(W_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wes}}) \times \\
&\quad f(Y_{ij}^{EE} | X_i^{EE}, Z_i, \boldsymbol{\theta_{yee}}) f(Y_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \times \\
&\quad f(X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta_x}) dX_i^{EE} dX_i^{\Delta ES}
\end{aligned}
\tag{7}
$$

where the RHS is expanded in (7). Given the independence assumptions in the previous paragraph, the full likelihood is

$$
L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta})
\tag{8}
$$

3.4   Näive Model

The first model we will consider is what we call the näive model. This model will assume no measurement error in the gold standard instrument exists, thus DLW and DXA give error-free measurements of X_i^{EE} and $X_i^{\Delta ES}$, respectively. We also assume the cheap measurements Y are linearly related to the *truth* and of error free covariates. Based on empirical evidence from EBS data, gender, BMI, and age all had some effect on the cheap measurement of EE. Mathematically, the näive model is represented as:

$$Y_{ij}^{EE} = \beta_{0,ee} + \beta_{1,ee}W_{ij}^{EE} + \boldsymbol{\gamma_{ee}}Z_i + \epsilon_{ij}^{EE} \tag{9}$$

$$Y_{ij}^{\Delta ES} = \beta_{0,es} + \beta_{1,es}W_{ij}^{\Delta ES} + \boldsymbol{\gamma_{es}}Z_i + \epsilon_{ij}^{\Delta ES} \tag{10}$$

where the $\beta_{1,\cdot}$ terms represents the relationship between cheap measurements and the *truth* EE and ΔES and the β_0 terms represent systematic biases. $\boldsymbol{\gamma_\cdot} = (\gamma_{1,\cdot}, \gamma_{2,\cdot}, \gamma_{3,\cdot})$ and $\gamma_{1,\cdot}$ is the coefficient for gender, $\gamma_{2,\cdot}$ is the coefficient for BMI, and $\gamma_{3,\cdot}$ is the coefficient for age. We take the standard approach and assume Normal distributions for the error terms, as well as assuming independence within and across individuals conditional on subject i:

$$(Y_{ij}^{EE}|W_{ij}^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \overset{iid}{\sim} N(\beta_{0,ee} + \beta_{1,ee}W_{ij}^{EE} + \boldsymbol{\gamma_{ee}}Z_i, \sigma_{\epsilon EE}^2) \tag{11}$$

$$(Y_{ij}^{\Delta ES}|W_{ij}^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \overset{iid}{\sim} N(\beta_{0,es} + \beta_{1,es}W_{ij}^{\Delta ES} + \boldsymbol{\gamma_{es}}Z_i, \sigma_{\epsilon \Delta ES}^2) \tag{12}$$

Priors are assumed independent and are assigned below. We chose forms for priors to be conjugate or conditionally conjugate for ease of implementation but also to incorporate weak information.

$$p(\beta_{0,ee}) \sim N(M_{\beta_{0,ee}}, C_{\beta_{0,ee}}) \tag{13}$$
$$p(\beta_{0,es}) \sim N(M_{\beta_{0,es}}, C_{\beta_{0,es}}) \tag{17}$$

$$p(\beta_{1,ee}) \sim N(M_{\beta_{1,ee}}, C_{\beta_{1,ee}}) \tag{14}$$
$$p(\beta_{1,es}) \sim N(M_{\beta_{1,es}}, C_{\beta_{1,es}}) \tag{18}$$

$$p(\gamma_{g,ee}) \sim N(M_{\gamma_{ee}}, C_{\gamma_{ee}}),\ g = 1, 2, 3 \tag{15}$$
$$p(\gamma_{g,es}) \sim N(M_{\gamma_{es}}, C_{\gamma_{es}}),\ g = 1, 2, 3 \tag{19}$$

$$p(\sigma_{\epsilon ee}^2) \sim IG(a_{yee}, b_{yee}) \tag{16}$$
$$p(\sigma_{\epsilon es}^2) \sim IG(a_{yes}, b_{yes}) \tag{20}$$

$\sigma_{\epsilon ee}^2$ and $\sigma_{\epsilon es}^2$ represent the measurement error and within-person variability as measured in the cheap EE and ΔES tools, respectively.

3.5   Linear Measurement Error Model (LMEM)

The Linear Measurement Error Model, or what we'll refer to as the LME model, is a measurement error model that otherwise takes similar form to the näive model. Instead of assuming the gold standard measurements are error free, we allow them to follow the form of (5),(6). We assume the measurement error terms follow independent normal distributions. We justify this assumption in the development of our full model. We also need to specify a bivariate distribution for the *usual*, $(X_i^{EE}, X_i^{\Delta ES})$. The common approach is to assume a Normal distribution, which we do here. In mathematical notation, the model is represented as:

$$(Y_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \sim N(\beta_{0,ee} + \beta_{1,ee}X_i^{EE} + \boldsymbol{\gamma_{ee}}Z_i, \sigma^2_{\epsilon_{EE}}) \tag{21}$$

$$(Y_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \sim N(\beta_{0,es} + \beta_{1,es}X_i^{\Delta ES} + \boldsymbol{\gamma_{es}}Z_i, \sigma^2_{\epsilon_{\Delta ES}}) \tag{22}$$

$$(W_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta_{wee}}) \overset{iid}{\sim} N(X_i^{EE}, \sigma^2_{\nu_{EE}}) \tag{23}$$

$$(W_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wes}}) \overset{iid}{\sim} N(X_i^{\Delta ES}, \sigma^2_{\nu_{\Delta ES}}) \tag{24}$$

$$(X_i^{EE}, X_i^{\Delta ES}|\theta_X) \overset{iid}{\sim} N\left(\begin{bmatrix} \mu_{EE} \\ \mu_{\Delta ES} \end{bmatrix}, \Sigma_X\right) \tag{25}$$

We choose the priors as:

$$p(\beta_{0,ee}) \sim N(M_{\beta_{0,ee}}, C_{\beta_{0,ee}}) \tag{26} \qquad p(\beta_{0,es}) \sim N(M_{\beta_{0,es}}, C_{\beta_{0,es}}) \tag{31}$$

$$p(\beta_{1,ee}) \sim N(M_{\beta_{1,ee}}, C_{\beta_{1,ee}}) \tag{27} \qquad p(\beta_{1,es}) \sim N(M_{\beta_{1,es}}, C_{\beta_{1,es}}) \tag{32}$$

$$p(\gamma_{g,ee}) \sim N(M_{\gamma_{ee}}, C_{\gamma_{ee}}),\ g = 1,2,3 \tag{28} \qquad p(\gamma_{g,es}) \sim N(M_{\gamma_{es}}, C_{\gamma_{es}}),\ g = 1,2,3 \tag{33}$$

$$p(\sigma^2_{\epsilon^{ee}}) \sim IG(a_{yee}, b_{yee}) \tag{29} \qquad p(\sigma^2_{\epsilon^{es}}) \sim IG(a_{yes}, b_{yes}) \tag{34}$$

$$p(\sigma^2_{\nu^{ee}}) \sim IG(a_{wee}, b_{wee}) \tag{30} \qquad p(\sigma^2_{\nu^{es}}) \sim IG(a_{wes}, b_{wes}) \tag{35}$$

$$p(\Sigma_X) \sim Inv - Wish(\psi, d) \tag{36}$$

$$p(\mu_{EE}, \mu_{\Delta ES}) \sim N(M_\mu, C_\mu) \tag{37}$$

3.6   Spline Measurement Error Model (SMEMDP)

Our model to analyze the measurement error and biases involved with various measurement tools for EE and ΔES involves both parametric and non-parametric components. The goal of this approach is to be able to have interpretable parameters, but also to have a model that is flexible enough to handle relationships that are unknown. Following (7), we explain how we are going to model every portion to have in the end, a joint probability model. We will now be more specific than in (3)-(5).

we follow the same construction of the LMEM for the modeling of the gold standard measurements as being unbiased to the *usual* and having normally distributed measurement errors:

$$f(W_{ij}^{EE}|X_i^{EE}, Z_i, \sigma_{\nu EE}^2) \overset{iid}{\sim} N(X_i^{EE}, \sigma_{\nu EE}^2) \tag{38}$$

$$f(W_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \sigma_{\nu \Delta ES}^2) \overset{iid}{\sim} N(X_i^{\Delta ES}, \sigma_{\nu \Delta ES}^2) \tag{39}$$

$$\tag{40}$$

Next we explain how we will model the cheap measurements Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$. Similar to above, we start with the structure in (3),(4). We have reason to believe and have empirical evidence that cheap instruments are not only noisy, but biased as well. Our goal here is to understand both the biases as functions of *usual* value and demographic covariates, as well as the measurement error in the instruments themselves. We propose modeling the cheap measurements as a semi-parametric regression problem. Specifically, model the functions $m_{\cdot}()$ using free knot cubic B-splines, and model demographic covariates with a linear component. We need to have monotone functions, but this only requires the spline coefficients to be nondcreasing [10], ie. $\beta_1 \leq \beta_2 \leq ... \leq \beta_k$ which is a simple linear programming problem. Our approach has three benefits. First, the spline is flexible and can pick up an unknown relationship between $X^{\cdot}$ and the cheap measurement of the same, which is

important because we never observe the truth and therefore don't have a reasonable guess as to the function form of the relationship. Second, the use of free knot splines eliminates the need for us to specify the number and position of the knots. We use Reversible Jump MCMC (RJMCMC) to determine the number and position of knots. This means we treat the number of knots in each regression equation and their knot locations as random variables. Third, the linear component for the covariates allows for an easy interpretation of the parameters and thus the biases involved in the instrument. This is convenient because the covariates we are using are standard to collect in any exercise, diet, or health study and they could help better calibrate the cheap instruments. We will make a working assumption of constant variance for all measurement errors at this time. Based on the above, the model specification is then:

$$f(Y_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \stackrel{iid}{\sim} N(s(X_i^{EE}; \boldsymbol{\beta_{ee}}) + \boldsymbol{\gamma_{ee}} Z_i, \sigma^2_{\epsilon EE}) \tag{41}$$

$$f(Y_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \stackrel{iid}{\sim} N(s(X_i^{\Delta ES}; \boldsymbol{\beta_{\Delta es}}) + \boldsymbol{\gamma_{es}} Z_i, \sigma^2_{\epsilon \Delta ES}) \tag{42}$$

$$\tag{43}$$

$$s(X_i^{EE}; \boldsymbol{\beta_{ee}}) = \sum_{i=1}^{k_{ee}+3} b_{i,ee}(\mathbf{X^{EE}})\beta_{i,ee} = B_{ee}(\mathbf{X^{EE}})\boldsymbol{\beta_{ee}} \tag{44}$$

$$s(X_i^{\Delta ES}; \boldsymbol{\beta_{\Delta es}}) = \sum_{i=1}^{k_{es}+3} b_{i,es}(\mathbf{X^{\Delta ES}})\beta_{i,es} = B_{es}(\mathbf{X^{\Delta ES}})\boldsymbol{\beta_{es}} \tag{45}$$

where $B_{ee}()$ and $B_{es}()$ are $n \times (k_{ee} + 3)$ and $n \times (k_{es} + 3)$ B-spline basis matrices that can be constructed using the recursion specified in [11]. k_{ee} and k_{es} are the number of knots for the EE and ΔES splines, respectively. Note that k_{ee} and k_{es} are random variables because we are using free knot splines.

As a final note, we briefly justify our use of B-splines as the non-parametric regression function. There are many different types of splines, but we picked B-splines because in similar problems it has been shown that they are numerically more stable, which can have major effects on outcomes as compared in Sarkar et al. [12]

We allow more flexibility in the distribution of the latent variables $(X_i^{EE}, X_i^{\Delta ES})$ by specifying a Dirichlet process mixture prior for them. This allows the data to speak for itself which is ideal since we are dealing with latent variables. The density of $(X_i^{EE}, X_i^{\Delta ES})$ can then modeled as an infinite mixture of normals. Then we can model the distribution for the latent variables as:

$$X_i^{EE}, X_i^{\Delta ES}|\zeta_i = h \overset{iid}{\sim} N\left(\begin{bmatrix} \mu_{ee,h} \\ \mu_{es,h} \end{bmatrix}, \Sigma_h\right) \tag{46}$$

$$\zeta_i \overset{iid}{\sim} Cat(H, \boldsymbol{\pi}) \tag{47}$$

$$V_h \sim \text{Beta}(1, \alpha) \tag{48}$$

$$V_H = 1 \tag{49}$$

$$\pi_h = V_h \prod_{\ell < h}(1 - V_\ell) \tag{50}$$

α helps control how many components are used. We choose to set α to be 1 so that two observations have equal chance of being in the same clusters as they do different clusters, *apriori*. ζ_i represents that observation i came from group h. $Cat(H, \boldsymbol{\pi})$ is a categorical random variable such that $P(\zeta_i = h) = \pi_h, h \leq H$. In any given problem, we can select H such that $\sum_{h=1}^{H} \pi_h < \epsilon$ for some $\epsilon > 0$. To help simplify the construction of the sampler, we choose conditionally conjugate priors when available. We also let all priors be independent of each other.

$$p(\sigma_{\epsilon ee}^2) \sim IG(a_{yee}, b_{yee}) \tag{51}$$
$$p(\sigma_{\epsilon es}^2) \sim IG(a_{wee}, b_{wee}) \tag{55}$$

$$p(\sigma_{\nu ee}^2) \sim IG(a_{yes}, b_{yes}) \tag{52}$$
$$p(\sigma_{\nu es}^2) \sim IG(a_{wes}, b_{wes}) \tag{56}$$

$$p(k_{ee}) \sim Poi(\lambda_{ee}) \tag{53}$$
$$p(k_{es}) \sim Poi(\lambda_{es}) \tag{57}$$

$$p(r_{ee}) \sim DisUnif(X_1^{EE}, ..., X_n^{EE}) \tag{54}$$
$$p(r_{es}) \sim DisUnif(X_1^{\Delta ES}, ..., X_n^{\Delta ES}) \tag{58}$$

$$p(\Sigma_h) \sim Inv - Wish(\psi, d), h = 1, ...H \tag{59}$$

$$p(\mu_{EE,h}, \mu_{\Delta ES,h}) \sim N(M_\mu, C_\mu), h = 1, ...H \tag{60}$$

The prior on k_{ee} and k_{es} deals with the number of knots and thus the complexity of the splines. Although we do not know the true form, we do not anticipate it to be highly complex, so we would like to use as few knots as possible. r_{ee} and r_{es} are the knot locations. Our discrete uniform prior on these means that knots can only occur at the latent values of $(X_i^{EE}, X_i^{\Delta ES})$. This was done largely for computational convenience as we could have assigned a continuous prior, but we don't believe this will adversely affect the problem because the latent $(X_i^{EE}, X_i^{\Delta ES})$ will be updated every MCMC iteration. Notice that we have not placed priors on the spline regression coefficients $\boldsymbol{\beta_{ee}}$ and $\boldsymbol{\beta_{es}}$, or the linear regression coefficients $\boldsymbol{\gamma_{ee}}$ and $\boldsymbol{\gamma_{es}}$; this is because we will not be updating this in a fully Bayesian fashion. We will discuss this more in the RJMCMC section.

4 Simulated Data

In this section we describe how we simulate a data set that reflects our problem in order to perform a simulation study. Our simulated data needs to be sufficiently complex and incorporate dependence in order to represent the distributions of true EE and EI, as well as gold standard measurements and cheap measurements. We try to simulate data that looks similar to the EBS data. We need to simulate data for all the components in (2) as well as the latent variables $(X_i^{EE}, X_i^{\Delta ES})$. We will explore measurement errors for the gold standard and cheap measurements in 3 different situations: normal errors, skewed errors, bimodal errors.

4.1 Demographic Covariates

For this simulation, we used three covariates: gender, age, BMI. Using a total sample size of 300, we sampled 300 Bernoulli(0.5) to determine gender. Age was determined as a Uniform(20,40). The age range was chosen because the Energy Balance Study [8], which has related data used this age range, so it will be easier to compare real data with this simulated data. BMI for an individual was

	Mean EE	Mean EI	Std Dev EE	Std Dev EI	Cor(EE,EI)
Component 1	$1500 + \gamma_{ee}Z_{comp,1}$	$1630 + \gamma_{ei}Z_1$	50	70	0.4376
Component 2	$1900 + \gamma_{ee}Z_{comp,2}$	$2030 + \gamma_{ei}Z_2$	80	95	0.4376
Component 3	$2100 + \gamma_{ee}Z_{comp,3}$	$2180 + \gamma_{ei}Z_3$	60	90	0.4376
Component 4	$2900 + \gamma_{ee}Z_{comp,4}$	$3030 + \gamma_{ei}Z_4$	400	410	0.4376
Component 5	$3200 + \gamma_{ee}Z_{comp,5}$	$3330 + \gamma_{ei}Z_5$	220	250	0.4376

Table 2: Mean and Variances for X

simulated from a Normal(27,5) since the average BMI for an American is around 27 [13]. All these values were simulated independently. Let Z be the matrix of dimension 300×3 with all covariate values.

4.2 Latent Variables

We simulate (X_i^{EE}, X_i^{EI}) from a mixture of 5 bivariate t-distributions. 60 observation pairs will be simulated from 5 different bivariate t's each. Simulation was done using the `rmvt` function in the `mvtnorm` package in `R`. The mean of the 2-dimensional vector for each of the 5 t-distributions will be a constant plus a linear combination of the covariates described above. The scale matrix for each of the 5 t-distributions will consist of constants for standard deviations and correlation chosen empirically. The degrees of freedom will remain equal to 5 throughout.

We let the correlation between EE and EI be the same for every component of the mixture. This value was determined empirically through the EBS data. Table (2) shows the values of the mean and standard deviation for EE and EI for every component. The coefficients for the covariates will remain constant across components. The values used for the vector $\gamma_{ee} = (300, 14, -7)$ and $\gamma_{es} = (-200, 8, -5)$ for gender, BMI and age, respectively. $Z_{comp,k}$ is a 60x3 matrix of covariate values corresponding to the 60 individuals in component k. Once 300 data pairs (X_i^{EE}, X_i^{EI}) are simulated according to this procedure, we can then calculate $X_i^{\Delta ES}$ using the energy balance equation. Figure 6 shows histograms for the latent variables for one simulated data set. Figure 7 gives contours of the bivariate distributions.

4.3 Within Person Variability

Recall in (3), (4), (5), (6) what the error terms represent: measurement error in the instrument as well as within-person variability (an individual's day to day variation from his *usual*). Again, in
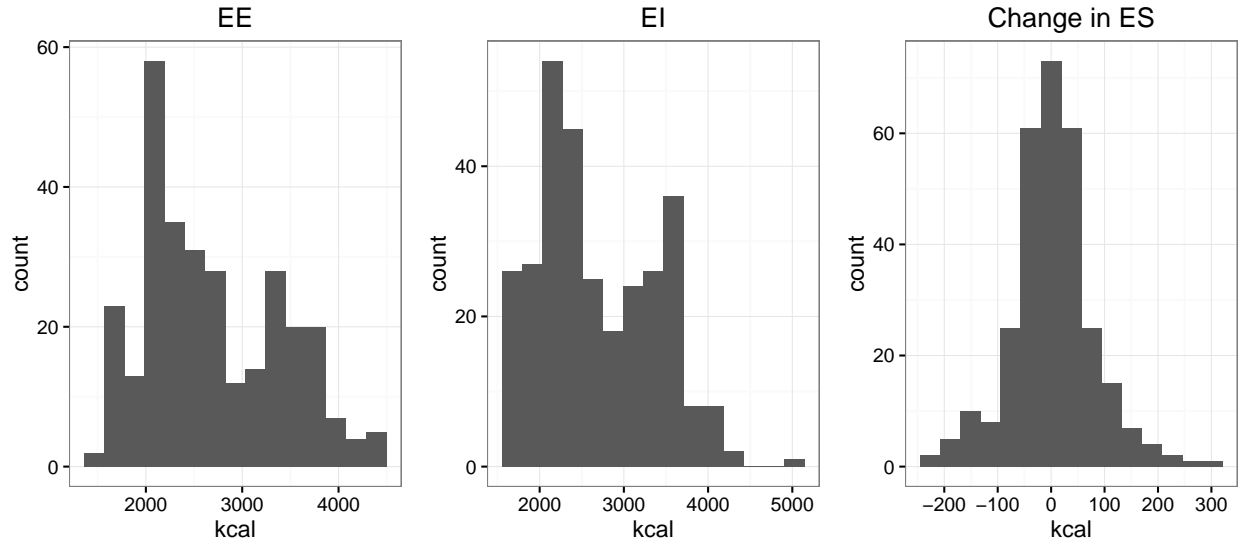
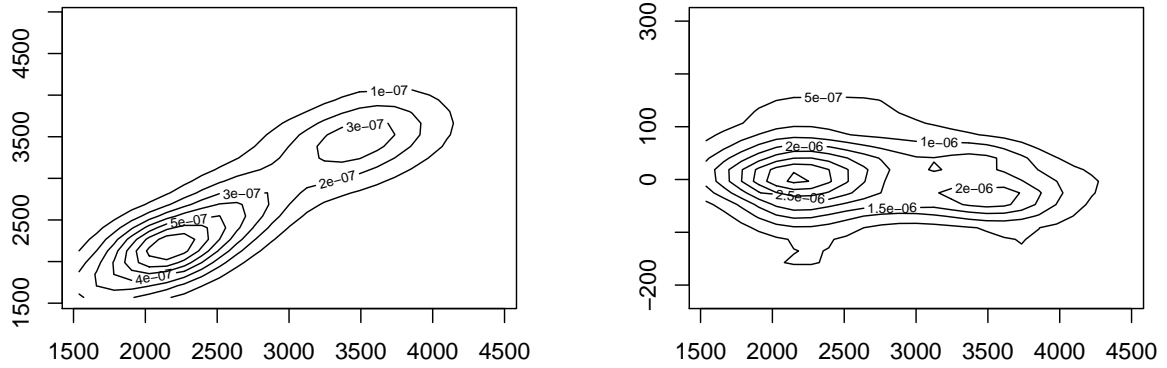Figure 6: Distribution of simulated latent variables X from one simulated data set

Figure 7: Contour plot os bivariate distributions of one simulated data set for latent variables. X^{EE} vs X^{EI} on left and X^{EE} vs $X^{\Delta ES}$ on right

terms of (3), (4) this is a working assumption, and here we will correctly put the within person variability within the function. For the gold standard measurements, let

$$
\begin{aligned}
\nu_{ij}^{EE} &= u_{ij}^{EE} + \delta_{ij}^{EE} \\
\nu_{ij}^{\Delta ES} &= u_{ij}^{\Delta ES} + \delta_{ij}^{\Delta ES}
\end{aligned}
\tag{61}
$$

where u^{EE} represents the measurement error in DLW and $u^{\Delta ES}$ represents the measurement error in DXA. δ_{ij}^{EE} represents the within person variation in EE for person i during time period j from his true mean and $\delta_{ij}^{\Delta ES}$ similarly represents the within person variation in ΔES for person i during time period j from his true mean. For the cheap measurements there is a slightly different setup. The within person variability gets added to each individuals *usual* values of EE and ΔES and thus is affected by the functions $m.(\cdot)$. Therefore we add these within person variation terms δ to the *usual* values we simulated X to get:

$$
\begin{aligned}
X_{ij}^{EE} &= X_i^{EE} + \delta_{ij}^{EE} \\
X_{ij}^{\Delta ES} &= X_i^{\Delta ES} + \delta_{ij}^{\Delta ES}
\end{aligned}
\tag{62}
$$

where X_{ij} will be the term in the functions $m.(\cdot)$ to simulate the cheap measurements.

The pairs $(\delta_{ij}^{EE}, \delta_{ij}^{\Delta ES})$ will be simulated jointly but independently across time and individual. We simulate the within person variability terms $(\delta_{ij}^{EE}, \delta_{ij}^{\Delta ES})$ according to the bivariate distribution below:

$$
(\delta_{ij}^{EE}, \delta_{ij}^{\Delta ES}) \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\delta EE}^2 & \rho_\delta \sigma_{\delta EE} \sigma_{\delta \Delta ES} \\ \rho_\delta \sigma_{\delta EE} \sigma_{\delta \Delta ES} & \sigma_{\delta \Delta ES}^2 \end{bmatrix} \right)
\tag{63}
$$

$$\sigma_{\delta EE} = 150$$

$$\sigma_{\delta \Delta ES} = 50 \tag{64}$$

$$\rho_\delta = -0.266$$

4.4 Gold Standard Measurements

We assume DLW and DXA are unbiased measurements of EE and ΔES, respectively. These measurements are simulated according to (5),(6) where we further broke down ν in (61). The u term represents the measurement error components we still need to specify and δ represents the within person component of the error which already described how to simulate. We will assume the u terms are independent within and across individuals as well as mutually with all δ and X.

From these simulated values, we then get simulated gold standard data $W_{ij}^{EE}, W_{ij}^{\Delta ES}$ via (5),(6). We will simulate measurement errors for the gold standard measurements (and cheap measurements) from a total of three different distributions: normal, skewed normal, and mixture of two normals that is bimodal and centered around 0. The specification for the normal errors are given in (65). The skewed normal distribution was introduced by Fernandez and Steel [14]. The pdf of the skewed normal is the limit of (69) as ν goes to ∞. γ is the skewness parameter, for $\gamma > 1$, this results in a right skewed distribution and $\gamma < 1$, a left skewed distribution. The measurement error for this distribution is specified by (66). The final distribution for measurement errors which we explore is a bimodal mixture of two normal distributions. We constructed it such that it is symmetric and centered around 0. This distribution is specified by (67). Parameters were chosen such that the means of all error distributions are 0, and the variances for each distribution is the same within EE within ΔES. Values of the standard deviation parameters are given in (68). We simulate from the skewed normal distribution using the R package `fGarch`. Figure (8) shows density plots of all three distributions using our specified standard deviation and skew parameters.

$$
\begin{aligned}
u_{ij}^{EE} &\overset{iid}{\sim} N(0, \sigma_{\nu EE}^2) \\
u_{ij}^{\Delta ES} &\overset{iid}{\sim} N(0, \sigma_{\nu \Delta ES}^2)
\end{aligned}
\tag{65}
$$

$$
\begin{aligned}
u_{ij}^{EE} &\overset{iid}{\sim} SN(0, \sigma_{\nu EE}^2, 10) \\
u_{ij}^{\Delta ES} &\overset{iid}{\sim} SN(0, \sigma_{\nu \Delta ES}^2, 10)
\end{aligned}
\tag{66}
$$
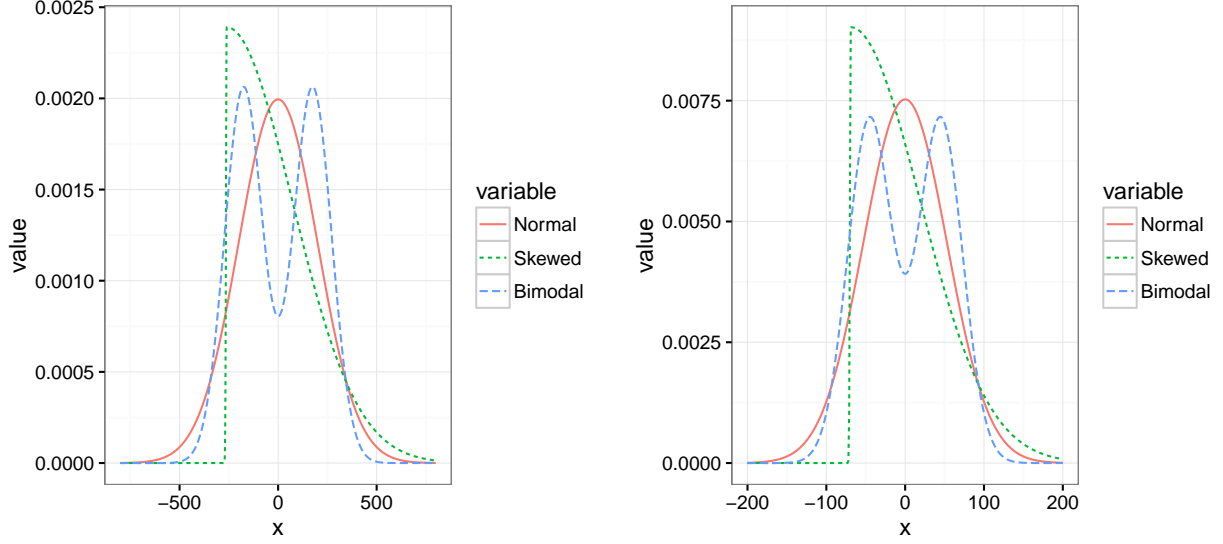
Figure 8: Measurement Error Distributions, EE on left and ΔES on right

$$u_{ij}^{EE} \stackrel{iid}{\sim} 0.5 \times N(-175, \sigma_{\nu EE}^2 - 175^2) + 0.5 \times N(175, \sigma_{\nu EE}^2 - 175^2) \equiv BM(175, \sigma_{\nu EE}^2)$$

$$u_{ij}^{\Delta ES} \stackrel{iid}{\sim} 0.5 \times N(-45, \sigma_{\nu \Delta ES}^2 - 45^2) + 0.5 \times N(45, \sigma_{\nu \Delta ES}^2 - 45^2) \equiv BM(45, \sigma_{\nu \Delta ES}^2)$$

(67)

$$\sigma_{\nu EE} = 200$$

$$\sigma_{\nu \Delta ES} = 53$$

(68)

$$p(y|\mu, \sigma, \nu, \gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{1/2}} \sigma^{-1} \times$$

$$\left\{ 1 + \frac{(y-\mu)^2}{\nu\sigma^2} \left(\frac{1}{\gamma^2} I_{[0,\infty)}(y-\mu) + \gamma^2 I_{(\infty,0)}(y-\mu) \right) \right\}^{-\frac{(\nu+1)}{2}}$$

(69)

4.5   Cheap Measurements

We simulate data for the cheap measurements according to (3), (4). The error terms in those equations represents measurement error, regression error, and within person variability. Equation (62) shows how we break down the error term from (3),(4). We use the same simulated values of δ

that we used for the gold standard measurements.

We simulate the measurement error terms e in a similar fashion as the last section. Again we will simulate three different data sets, one with normal errors, one with skewed normal errors, and one with bimodal errors. We assume that the e terms are independent within and across subjects as well as mutually independent with all δ, X and Z terms. We specify these errors similar to the previous section:

$$e_{ij}^{EE} \overset{iid}{\sim} N(0, \sigma_{\epsilon EE}^2)$$
$$e_{ij}^{\Delta ES} \overset{iid}{\sim} N(0, \sigma_{\epsilon \Delta ES}^2) \tag{70}$$

$$e_{ij}^{EE} \overset{iid}{\sim} SN(0, \sigma_{\epsilon EE}^2, 10)$$
$$e_{ij}^{\Delta ES} \overset{iid}{\sim} SN(0, \sigma_{\epsilon \Delta ES}^2, 10) \tag{71}$$

$$e_{ij}^{EE} \overset{iid}{\sim} 0.5 \times N(-350, \sigma_{\epsilon EE}^2 - 350^2) + 0.5 \times N(350, \sigma_{\epsilon EE}^2 - 350^2) \equiv BM(350, \sigma_{\epsilon EE}^2)$$
$$e_{ij}^{\Delta ES} \overset{iid}{\sim} 0.5 \times N(-190, \sigma_{\epsilon \Delta ES}^2 - 190^2) + 0.5 \times N(190, \sigma_{\epsilon \Delta ES}^2 - 190^2) \equiv BM(190, \sigma_{\epsilon \Delta ES}^2) \tag{72}$$

$$\sigma_{\epsilon EE} = 380$$
$$\sigma_{\epsilon \Delta ES} = 210 \tag{73}$$

Different from the gold standard measurements which we assume to be unbiased to the truth, we want to add an element of bias to the cheap measurements. This comes through the functions m_{ee} and m_{es}. For this simulated data, we let $m_{ee}(X, Z) = 2X_i^{EE} - \frac{4000}{1 + e^{-0.002X_i^{EE} - 2200}}$ and $m_{es}(X, Z) = \frac{1000}{1 + e^{-0.04X_i^{\Delta ES}}} - 2000 + X_i^{\Delta ES}$. We then add $Z_i \boldsymbol{\gamma}$. to the simulated cheap measures of EE and ΔES. This represents the bias involved in the cheap instruments as a function of $(X^{EE}, X^{\Delta ES})$ and Z. Figure 9 shows $m_{ee}(\cdot)$ on the left and $m_{es}(\cdot)$ on the right both against a $y = x$ line for comparison.

5 Estimation

We will be taking a Bayesian approach to this problem, and as such, our goal is to estimate the joint posterior distribution of all parameters and latent variables involved. In our case the posterior
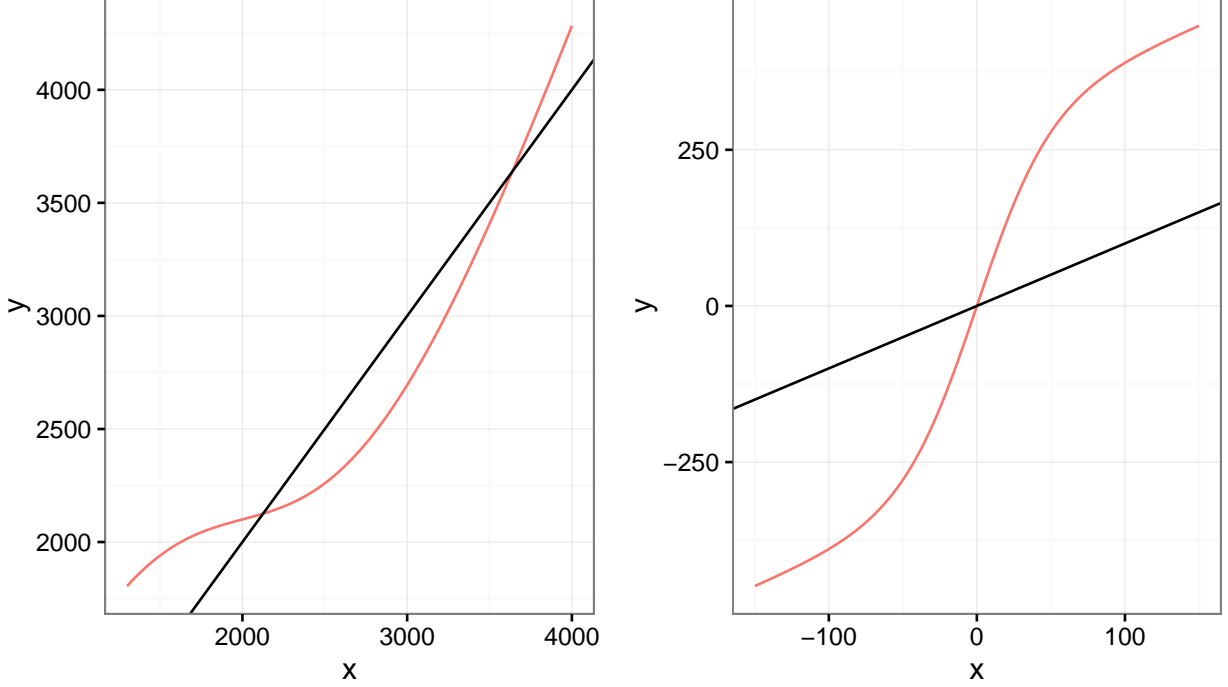
Figure 9: Plot of nonlinear functions $m_{ee}()$ (left) and $m_{es}()$ (right), and Y=X is black for reference to unbiased measurement

is $p(\boldsymbol{\theta}, X^{EE}, X^{\Delta ES}|W^{EE}, W^{\Delta ES}, Y^{EE}, Y^{\Delta ES}, Z)$. Due to the complexity of our problem and the models we will formulate, we will not be able to analytically derive the posterior distribution. Instead, we will use Markov Chain Monte Carlo (MCMC) methods. For the näive LME models, we utilized JAGS to simulate draws from the posterior distribution. This was easy to implement and was relatively quick to sample. In order to estimate free knot splines, which allows for dimension change, we must use Reversible Jump MCMC which requires a more complex sampler. We will first present our Gibbs sampler with the reversible jump as a single step, and in the next subsection we will describe the reversible jump in more detail.

5.1 Gibbs Algorithm

Here we present our Gibbs algorithm to sample from the joint posterior distribution.

1. Set starting values. Let $i = 1, ..., n$ represent the number of individuals, and $j = 1, ..., J$ represent the number of replicates for each individual. Assume for simplicity each individual has the same number of replicates.

2. Set b=0. Run the MCMC below B times. Update tuning parameters C_{X_i}, covariance for

random walk by setting $C_{X_i} = \frac{2.4^2}{2} \frac{1}{B-1} \sum_{b=1}^{B} (X_i^{EE(b)} - \bar{X}^{EE})(X_i^{\Delta ES(b)} - \bar{X}^{\Delta ES})$. Repeat desired number of times, then throw away draws and continue using new tuning values. (See Gelman et al. [15], pg. 290)

3. For iteration k=1,...,K, sample from its full conditional:

(a) $\{\zeta_i^{(k)} : i = 1, ..., n\} \mid \cdot$ where $P(\zeta_i^{(k)} = h) = \frac{\pi_h^{(k-1)} N(\mu_h^{(k-1)}, \Sigma_h^{(k-1)})}{\sum_{h=1}^{H} \pi_h^{(k-1)} N(\mu_h^{(k-1)}, \Sigma_h^{(k-1)})}$

(b) $\{V_h^{(k)} : h = 1, ..., H - 1\} \mid \cdot \overset{ind}{\sim} Beta(1 + n_h, \alpha + n_h')$

Set $V_H^{(k)} = 1$

$n_h = \sum_{i=1}^{n} I(\zeta_i^{(k)} = h)$

$n_h' = \sum_{h'=h+1}^{H} n_{h'}$

(c) Calculate $\{\pi_i^{(k)} : i = 1, ..., n\} = V_i^{(k)} \prod_{\ell < i} (1 - V_\ell^{(k)})$

(d) $\{\Sigma_h^{(k)}, h = 1, ..., H\} \mid \cdot \overset{ind}{\sim} Inv - Wish(d + n_h, \psi + (\mathbf{X_{i,h}^{(k-1)}} - \mu_\mathbf{h}^{(k-1)})'(\mathbf{X_{i,h}^{(k-1)}} - \mu_\mathbf{h}^{(k-1)}))$

$\mathbf{X_{i,h}} \equiv \mathbf{X_i} I(\zeta_i^{(k)} = h)$

$\mathbf{X_i} = (X_i^{EE}, X_i^{\Delta ES})$

$\mu_\mathbf{h} = (\mu_{EE,h}, \mu_{\Delta ES,h})$

(e) $\{(\mu_{EE,h}^{(k)}, \mu_{\Delta ES,h}^{(k)}), h = 1, ..., H\} \mid \cdot \overset{ind}{\sim} N(M_\mu', C_\mu')$

$C_\mu' = (C_\mu^{-1} + n_h \Sigma_h^{-1(k)})^{-1}$

$M_\mu' = C_\mu'(C_\mu^{-1} M + n_h \Sigma_h^{-1(k)} \bar{\mathbf{X}}^{(k-1)})$

$\bar{\mathbf{X}} = \frac{1}{n_h} \sum_{i=1}^{n} \mathbf{X_{i,h}}$

(f) Update $\mathbf{X_i}$ with a random walk

$\{\mathbf{X_i^{(k)}} : i = 1, ..., n\} \mid \cdot$ for $i = 1, ...n$ sample $\mathbf{X_i^*}$ from $N(\mathbf{X_i^{(k)}}, C_{X_i})$

Set $\mathbf{X_i^{(k)}} = \mathbf{X_i^*}$ with probability α_{X_i}, otherwise set $\mathbf{X_i^{(k)}} = \mathbf{X_i^{(k-1)}}$

$\alpha_{X_i} = min\left(1, \frac{f(\mathbf{X_i^*}|\cdot)}{f(\mathbf{X_i^{(k-1)}}|\cdot)}\right)$

(g) Update $\boldsymbol{\beta_{ee}}, \boldsymbol{\beta_{es}}, k_{ee}, k_{es}, r_{ee}, r_{es}, \boldsymbol{\gamma_{ee}^{(k)}}, \boldsymbol{\gamma_{es}^{(k)}}$ using RJMCMC described in next section. Calculate $m_{ee}(\mathbf{X^{EE(k)}}; \boldsymbol{\beta_{ee}^{(k)}})$ and $m_{es}(\mathbf{X^{\Delta ES(k)}}; \boldsymbol{\beta_{es}^{(k)}})$

(h) $\sigma_{\epsilon EE}^{2(k)}|\cdot \sim IG(a_{yee} + J \times \frac{n}{2}, b_{yee} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(Y_{ij}^{EE} - m_{ee}(X_i^{EE(k)}; \boldsymbol{\beta_{ee}^{(k)}}) - \boldsymbol{\gamma_{ee}}^{(k)}Z_i)^2)$

(i) $\sigma_{\epsilon \Delta ES}^{2(k)}|\cdot \sim IG(a_{yes} + J \times \frac{n}{2}, b_{yes} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(Y_{ij}^{\Delta ES} - m_{es}(X_i^{\Delta ES(k)}; \boldsymbol{\beta_{es}^{(k)}}) - \boldsymbol{\gamma_{es}}^{(k)}Z_i)^2)$

(j) $\sigma_{\nu EE}^{2(k)}|\cdot \sim IG(a_{wee} + J \times \frac{n}{2}, b_{wee} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(W_{ij}^{EE} - X_i^{EE(k)})^2)$

(k) $\sigma_{\nu \Delta ES}^{2(k)}|\cdot \sim IG(a_{wes} + J \times \frac{n}{2}, b_{wes} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(W_{ij}^{\Delta ES} - X_i^{\Delta ES(k)})^2)$

5.2 Reversible Jump MCMC

In order to complete our Gibbs algorithm in the previous section, we need to describe how we estimate the parameters of the free knot spline using RJMCMC. We first present background on estimating free knot splines using RJMCMC. At the end of the section, we present the missing piece in the above algorithm that explains how we sample $\boldsymbol{\beta_{ee}}, \boldsymbol{\beta_{es}}, k_{ee}, k_{es}, r_{ee}, r_{es}, \boldsymbol{\gamma_{ee}^{(k)}}, \boldsymbol{\gamma_{es}^{(k)}}$.

RJMCMC was introduced by Green [16] as a way to jump between parameter spaces having different dimensions. It was developed as a Bayesian approach for model determination. Model selection and model averaging can be implemented straightforward through the use of this sampler. Although we are not using RJMCMC for the usual type of model selection *per se*, we are using it to select one piece of our model; that is the spline component of our regression. As discussed earlier, we need to specify the position and number of knots for our spline, but since we have no reasonable guess as to what the error free values would look like, we want to account for that uncertainty by treating the position and number of knots as random variables. RJMCMC is the tool that allows us to do this in a Bayesian framework via free knot splines.

There have been a number of different approaches to Bayesian estimation of free knot splines using RJMCMC [17],[18],[19]. We will, for the most part, take the approach of Denison et al. [20]. The approach of Denison et al. is a simple and quick implementation that performs comparably to the

fully Bayesian approach of DiMatteo et al. [17] for smooth and not highly complex functions when an appropriate adjustment is made (that was pointed out in DiMatteo et al). This is okay for us since we do not expect the function to be highly complex, and this will be shown through our choice of prior for the number of knots for the splines.

Consider the situation where we are interested in estimating a regression spline in the following setup

$$
\begin{aligned}
y_i &= f(x_i; \boldsymbol{\beta}) + \epsilon_i \\
\epsilon_i &\overset{iid}{\sim} N(0, \sigma^2)
\end{aligned}
\tag{74}
$$

where the $y_i, i = 1, ...n$ are the response variable and $x_i, i = 1, ...n$ is the covariate. We do not want to specify the number or location of the knots. Under the specifications of Denison et al., in RJMCMC for free knot splines there are three possible transitions:

1. Birth of a knot

2. Death of a knot

3. Movement of a knot

where in either of the first two transitions the dimension is changing. Which transition that is proposed depends upon the prior for the number of knots k. Using the notation of Denison et al., let the prior probability for k knots be $p(k)$. Then the probability of attempting a birth step, death step or move step are b_k, d_k, η_k, respectively, where $\eta_k = 1 - b_k - d_k$, and

$$
\begin{aligned}
b_k &= c \times min\left(1, \frac{p(k+1)}{p(k)}\right) \\
d_k &= c \times min\left(1, \frac{p(k)}{p(k+1)}\right)
\end{aligned}
\tag{75}
$$

where $0 \leq c \leq 1/2$ to ensure $b_k + d_k \leq 1$. If the birth step transition is chosen, a new knot is proposed from the set of available new knots. Denison et al. puts a discrete uniform prior on knot locations so that only observed locations can be knot locations. A proposed new knot is chosen at random from the set $\mathcal{A} = \{x_i, i = 1, ...n : x_i$ is not currently a knot or within $\ell + 1$ locations of a current

knot}. ℓ is the order of the splines, so in our case 3 since we're using cubic splines. If the death step transition is chosen, a current knot is picked at random and removed. If the transition step is chosen, a current knot is picked at random and moved to a new location at random from the set $\mathcal{A}$. Once new knots are chosen, construct a spline basis matrix using the observed locations $\{x_i, i = 1, ... n\}$ and the positions of the knots. Then use OLS to estimate the spline parameters $\boldsymbol{\beta}$. Although we could put priors on the spline regression parameters, this adds significant computational burden and results have been similar when comparing estimation with OLS versus fully Bayesian [20]. This is our reason for not having priors on the spline regression parameters in an earlier section. The acceptance probability for each step is of the form given by Green

$$\alpha = min(1, \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}) \tag{76}$$

Denison et al. gives simplified acceptance probabilities for each of the three transitions under Poisson prior for number of knots and discrete uniform prior for knot locations. In the regression setup in Denison et al., they then sample variance components using a Gibbs step.

5.3 Our RJMCMC Implementation

Here we present how we implement our version of RJMCMC for free knot splines. The model (74) that was used to develop the method in [20] is simpler than our model in three respects: (i) our regression model is part of a larger hierarchical model, (ii) we have an additional linear component to our mean function, (iii) our covariate value x_i is a latent variable . The first issue doesn't present much of a problem due to our conditional independence assumptions. For issue (ii) we propose to update the linear coefficient parameters at the same time as the spline coefficient parameters using OLS. We understand this is not a fully Bayesian approach, but we anticipate results to be similar with much less computational burden.

Issue (iii) is not as simple. In the free knot spline setup in Denison et al., they have a regression model of the form (74) where y is the observed response and x is the observed covariate. In our case we do not observe x, but rather it is a latent variable that we must sample from using the Gibbs algorithm described in the previous section. Because we have to sample our x's through the

29

Gibbs algorithm, the basis matrix has to be adjusted every iteration. We calculate the basis matrix using the current latent variables and the proposed knots, if we accept this proposal then the knots, spline coefficients, and thus estimate of $m.(X^{\cdot(k)}, \boldsymbol{\beta}^{(k)}_{\cdot})$ are all updated. The bigger problem is what happens when we reject the proposal. Keeping the current knot locations is fine, keeping the current spline regression parameters is fine, but keeping the current predictions doesn't make sense because they are based on $X^{\cdot(k)}$, which is being updated every iteration. What we propose is if we reject the the RJ proposal, we keep the current knots and spline coefficients, calculate a new basis matrix based on the current values of $X^{\cdot(k)}$ and the current knots, then calculate the predicted values by multiplying the basis matrix by the current spline coefficients. An additional complication with X being random instead of fixed has to do with the birth step and the set of available knots to choose from $\mathcal{A}$. We want to pick randomly from the set of X values which we drew in the current iteration of the MCMC, but we do not want a knot to be bigger or smaller than the maximum or minimum value, respectively. This could occur for example, if a new knot is proposed at one of the largest values and it is accepted, then the next iteration through the random walk, the X values could change such that the largest knot is larger than the largest X, which causes issues when constructing the basis matrix. To avoid this issue, we exclude the 3 smallest and 3 largest values of X from $\mathcal{A}$. We don't expect this to be a major issue because we are less concerned with flexibility of the function at its extremes.

This algorithm is run independently for EE and ΔES regression functions due to conditional independence. Therefore let $(\cdot)$ denote the appropriate sub/superscript for EE or ΔES. The reversible jump step that goes into our overall Gibbs algorithm is then:

1. Calculate b_k and d_k according to (75)

2. Select birth, death, or move step with probabilities $b_k, d_k, 1 - b_k - d_k$ respectively

3. **Knot Changes**

   If birth step:

   Select a new knot location at random from the set $\mathcal{A}$ and join with current knots $r^{(k-1)}_{\cdot}$ to create the proposed knot locations $r^*_{\cdot}$

If death Step:

Sample one knot location from $r_{\cdot}^{(k-1)}$ at random and remove it.

If move step:

Sample one knot location from $r_{\cdot}^{(k-1)}$ at random, and change it to a new knot location at random from the set $\mathcal{A}$

4. Calculate the spline basis matrix $B_{\cdot}^{*}(X^{\cdot(k)})$ using $X^{\cdot(k)}$ and proposed knot locations $r_{\cdot}^{*}$

5. Calculate proposed spline and linear regression coefficients $\boldsymbol{\beta}_{\cdot}^{*}, \boldsymbol{\gamma}_{\cdot}^{*}$ by using OLS by regressing $Y^{\cdot}$ on $B_{\cdot}^{*}(X^{\cdot(k)}) + Z$

6. Accept proposed knots and coefficients with probability α.

Otherwise set $r_{\cdot}^{(k)} = r_{\cdot}^{(k-1)}, \boldsymbol{\beta}_{\cdot}^{(k)} = \boldsymbol{\beta}_{\cdot}^{(k-1)}, \boldsymbol{\gamma}_{\cdot}^{(k)} = \boldsymbol{\gamma}_{\cdot}^{(k-1)}$

$$\alpha_{birth} = min\left(1, \text{Likelihood ratio} \times \frac{n - Z(k)}{n}\right)$$

$$\alpha_{death} = min\left(1, \text{Likelihood ratio} \times \frac{n}{n - Z(k)}\right)$$

$$\alpha_{move} = min\left(1, \text{Likelihood ratio}\right)$$

$$Z(k) = 2(\ell + 1) + k_{\cdot}(2\ell + 1)$$

$$k_{\cdot} = length(r_{\cdot}^{(k-1)})$$

7. calculate mean function $m_{\cdot}(X^{\cdot(k)}; \boldsymbol{\beta}_{\cdot}^{(k)})$ using spline basis matrix $B_{\cdot}'(X^{\cdot(k)})$ and $r_{\cdot}^{(k)}$

6 Simulation Study

In this section we conduct a simulation study to check the performance of the discussed models and modeling assumptions. We are interested in the predictive performance of the models as the

31

main goal is measurement tool calibration. We are also interested to see what happens when the errors are not normally distributed which is why we present two alternatives error distributions. We present performance measures such as predicted mean squared error (PMSE) for the regression function in question as well as posterior means and their variability for parameters of interest.

6.1 Setup

We simulated 200 data sets each for normal, skewed, and bimodal errors for both 2 and 4 replicates per individual. The number of individuals is set to be 300. Preliminary analysis showed the number of replicates per individual having more influence than number of individuals.

Although we'd like to be as flexible as possible with our distributional assumptions on the bivariate latent variables, we also want a model that is stable in its estimates given the data constraints of our application. It would be difficult to get more than 2 replicate measurements on an individual for our gold standard measurements. For 3 replicate measurements in the DLW case, this amounts to three 2-week periods along with a washout time in between. In the case of DXA scans, this is 6 scans in a short period of time which can raise flags due to radiation concerns. This is exemplified by the high cost associated with these methods (mainly DLW). Cheap measurement tools are not as much of a concern. During our simulation study, we found that the Dirichlet Process prior on the latent variables was producing unstable results in parameter estimates and acceptance rates of the random walk Metropolis-Hastings under 2 replicate observations per person. The number of individuals played a much smaller role. Results were stable when 4 replicates per person were used. Results were similar when we used mixtures of normals instead of a DP. This presents an issue in our analysis since it is unrealistic to get that number of replicates in application. Because of this issue, we will fit a fourth model using a bivariate normal distribution for the latent variables instead of the Dirichlet Process while still using splines for the regression functions. We refer to this model as SMEMN. The MCMC has a minor change in the Gibbs step.

We set the values of the hyperparameters as follows: $M_{\beta_{0,ee}} = M_{\beta_{0,es}} = 0, C_{\beta_{0,ee}} = C_{\beta_{0,es}} = 100000, M_{\beta_{1,ee}} = M_{\beta_{1,es}} = 1, C_{\beta_{1,ee}} = C_{\beta_{1,es}} = 100000, M_{\gamma_{ee}} = M_{\gamma_{es}} = 0, C_{\gamma_{ee}} = C_{\gamma_{es}} = 100000, a_{yee} = a_{yes} = a_{wee} = a_{wes} = b_{yee} = b_{yes} = b_{wee} = b_{wes} = 0.1, \psi = I_{2\times2}, d = 3, M_\mu = (2400, 0), C_\mu = diag(100000, 100000), \lambda_{ee} = \lambda_{es} = 1$. We ran the MCMC for 12,000 iterations, using the first 2000 as burn in.

6.2 Results

Tables 3,4,5 give the results for 200 Monte Carlo samples for normal, skewed, and bimodal errors, respectively. The asterisk next to the truth for the measurement error with respect to the cheap measurements is to indicate this is a Monte Carlo approximation to the truth. Recall we included within person variation in the functions $m_.(\cdot)$, but in our model we use the working assumption that the error term accounts for both within person variability and measurement error. Because we can't directly extract the value from the function, we approximate it by generating 10,000 data sets and removing the mean function from the cheap observations, and then calculating the standard deviation of the residual. We then averaged those standard deviation estimates to get the one reported in the table.

Across all models and error types, the linear coefficients are estimated largely without bias. This is not too surprising since these covariates are measured without error. The biases and standard errors are slightly smaller for models SMEMN and SMEMDP, however. All 3 measurement error models perform comparably for assessing the measurement error in the gold standard instruments. They are all biased low for the bimodal errors. This is true for the measurement error for the cheap measurements as well. The näive model sees this too for bimodal errors, although not as extreme. The SMEMN and SMEMDP models have similar results for the measurement error of cheap measurements that are good for EE and ΔES for normal errors, but biased low for ΔES for both skewed and bimodal errors. Both the näive model and the linear measurement error model have estimated measurement error standard deviations for the cheap measurement that are larger than the truth under normal errors and skewed errors for EE. We have evidence that for a grosser departure from normality (bimodal distribution) in terms of measurement error can cause issues estimating variance components, although there was less issues with the skewed errors.

Tables 6,7 give the mean PMSE for each model under each type of error for both EE and ΔES for 2 and 4 replicates. There is a consistent decreasing pattern from left to right in terms of the models. First, the näivemodel does much worse than the same model which accounts for measurement error. The näive model and linear measurement error model perform much worse than the models with free knot splines in terms of PMSE. There's not a large difference between the SMEMN and SMEMDP model in terms of PMSE, but the SMEMDP model generally does better. The question is whether the small improvement is worth the increase in model complexity. It is the authors' opinion

that the answer is no for two reasons: (i) our main focus with this model is calibrating the cheap measurements and not necessarily inference at the latent variable level, and (ii) this only applies in this situation when we have 4 replicates, which for gold standard measurements, is unrealistic. Given the latent data was non-normal, a question might be how poorly does a bivariate normal fit these latent data? We address this question in an example in the next section. Because the main focus is to calibrate cheap measurements as best as possible, the simulation results are promising for our model.

	σ_{yee}		σ_{yes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	477.65	473.42	347.49	354.85	254.67	248.66	14.88	14.03	-4.14	-5.29	-200.53	-199.39	7.91	8.31	-4.94	-5.13
Std Err	17.82	19.24	9.43	6.80	43.65	36.10	4.33	3.62	3.37	3.06	28.19	22.83	2.90	2.13	2.35	1.73
Bias	72.15	67.92	13.49	20.85	-45.33	-51.34	0.88	0.03	2.86	1.71	-0.53	0.61	-0.09	0.31	0.06	-0.13
Truth	405.50	405.50	334.00	334.00	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	444.34	446.63	320.41	338.26	255.70	255.85	69.18	71.81	249.50	240.63	14.30	13.67	-4.50	-5.28	-199.27	-198.49	7.91	8.29	-4.91	-5.15
Std Err	16.84	14.22	10.76	7.53	10.74	6.33	2.27	1.56	43.44	37.02	4.25	3.60	3.38	3.04	28.31	22.96	2.91	2.12	2.30	1.71
Bias	38.84	41.13	-13.59	4.26	5.70	5.85	-3.68	-1.05	-50.50	-59.37	0.30	-0.33	2.50	1.72	0.73	1.51	-0.09	0.29	0.09	-0.15
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	393.69	400.55	313.47	331.79	246.81	248.93	67.61	71.04	293.16	294.61	14.17	14.11	-6.86	-6.78	-200.07	-200.86	8.00	8.04	-4.95	-5.26
Std Err	11.52	8.29	12.00	8.27	8.78	5.79	2.32	1.66	36.16	26.19	3.50	2.42	2.86	2.16	26.86	19.44	2.49	2.06	2.11	1.45
Bias	-11.81	-4.95	-20.53	-2.21	-3.19	-1.07	-5.25	-1.82	-6.84	-5.39	0.17	0.11	0.14	0.22	-0.07	-0.86	0.00	0.04	0.05	-0.26
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

	σ_{yee}	σ_{yes}	σ_{wee}	σ_{wes}	$\gamma_{1,ee}$	$\gamma_{2,ee}$	$\gamma_{3,ee}$	$\gamma_{1,es}$	$\gamma_{2,es}$	$\gamma_{3,es}$
Replicates	4	4	4	4	4	4	4	4	4	4
Mean Est	400.13	331.38	248.94	70.85	297.62	14.16	-7.11	-198.04	8.12	-4.94
Std Err	8.25	8.43	6.05	1.62	26.53	2.72	2.08	18.11	1.85	1.64
Bias	-5.37	-2.62	-1.06	-2.01	-2.38	0.16	-0.11	1.96	0.12	0.06
Truth	405.50	334.00	250.00	72.86	300.00	14.00	-7.00	-200.00	8.00	-5.00

Table 3: Summary of Simulation under Normal Errors for Näive, LME, SMEMN, SMEMDP Models, respectively

Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	473.87	466.80	311.66	317.05	255.46	250.72	14.08	13.23	-5.05	-6.04	-197.98	-200.46	7.79	7.91	-4.98	-4.92
Std Err	17.36	13.06	9.08	6.64	40.99	32.70	3.99	3.53	3.38	3.02	24.90	19.73	2.45	1.78	2.28	1.67
Bias	68.37	61.30	-22.34	-16.95	-44.54	-49.28	0.08	-0.77	1.95	0.96	2.02	-0.46	-0.21	-0.09	0.02	0.08
Truth	405.50	405.50	334.00	334.00	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	432.67	435.88	289.34	304.25	254.13	255.00	69.67	71.68	251.72	244.04	13.38	12.88	-5.46	-6.10	-196.76	-200.05	7.82	7.89	-4.96	-4.94
Std Err	16.70	12.10	9.90	7.05	11.99	6.27	2.63	1.80	41.34	33.01	3.95	3.50	3.35	3.00	25.32	19.53	2.39	1.77	2.25	1.67
Bias	27.17	30.38	-44.66	-29.75	4.13	5.00	-3.20	-1.18	-48.28	-55.96	-0.62	-1.12	1.54	0.90	3.24	-0.05	-0.18	-0.11	0.04	0.06
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	393.56	402.91	282.68	298.37	247.00	248.83	68.93	71.23	306.58	305.65	14.27	14.44	-6.95	-7.14	-197.44	-198.85	7.76	8.04	-4.92	-4.86
Std Err	13.33	9.56	10.96	7.47	9.54	6.33	2.56	1.69	36.09	27.06	3.35	2.57	2.77	2.31	24.41	17.81	2.42	1.73	2.10	1.47
Bias	-11.94	-2.59	-51.32	-35.63	-3.00	-1.17	-3.93	-1.63	6.58	5.65	0.27	0.44	0.05	-0.14	2.56	1.15	-0.24	0.04	0.08	0.14
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

	σ_{yee}	σ_{yes}	σ_{wee}	σ_{wes}	$\gamma_{1,ee}$	$\gamma_{2,ee}$	$\gamma_{3,ee}$	$\gamma_{1,es}$	$\gamma_{2,es}$	$\gamma_{3,es}$
Replicates	4	4	4	4	4	4	4	4	4	4
Mean Est	403.10	298.16	248.70	71.11	313.22	14.13	-7.26	-199.31	8.10	-4.85
Std Err	8.25	7.01	6.67	1.57	27.80	2.37	2.13	17.70	1.54	1.54
Bias	-2.40	-35.84	-1.30	-1.75	13.22	0.13	-0.26	0.69	0.10	0.15
Truth	405.50	334.00	250.00	72.86	300.00	14.00	-7.00	-200.00	8.00	-5.00

Table 4: Summary of Simulation under Skewed Errors for Näive, LME, SMEMN, SMEMDP Models, respectively

	σ_{yee}		σ_{yes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	342.63	344.52	233.52	246.69	227.08	221.48	12.15	12.22	-5.61	-5.52	-198.19	-199.20	8.01	8.13	-4.84	-4.99
Std Err	12.61	15.26	6.43	5.24	32.41	29.92	3.28	3.21	3.05	2.94	17.40	15.19	1.64	1.59	1.49	1.23
Bias	-62.87	-60.98	-100.48	-87.31	-72.92	-78.52	-1.85	-1.78	1.39	1.48	1.81	0.80	0.01	0.13	0.16	0.01
Truth	405.50	405.50	334.00	334.00	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	264.16	265.06	220.65	237.60	207.41	207.86	52.92	56.11	216.24	203.84	11.50	11.52	-5.75	-5.38	-201.48	-201.44	7.90	8.07	-4.74	-4.92
Std Err	15.07	8.86	7.31	5.56	11.92	8.43	1.85	1.26	32.90	30.36	3.20	3.25	3.07	2.91	17.18	15.06	1.66	1.57	1.49	1.23
Bias	-141.34	-140.44	-113.35	-96.40	-42.59	-42.14	-19.94	-16.75	-83.76	-96.16	-2.50	-2.48	1.25	1.62	-1.48	-1.44	-0.10	0.07	0.26	0.08
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	257.20	256.24	217.62	235.09	182.30	189.38	51.86	55.39	220.14	211.17	12.25	11.96	-5.90	-5.81	-201.34	-200.20	7.87	7.98	-5.08	-4.95
Std Err	11.41	10.47	7.51	5.46	8.22	5.74	1.87	1.19	30.56	32.63	3.24	2.84	2.86	2.47	17.33	13.18	1.60	1.28	1.46	1.13
Bias	-148.30	-149.26	-116.38	-98.91	-67.70	-60.62	-21.01	-17.47	-79.86	-88.83	-1.75	-2.04	1.10	1.19	-1.34	-0.20	-0.13	-0.02	-0.08	0.05
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

	σ_{yee}	σ_{yes}	σ_{wee}	σ_{wes}	$\gamma_{1,ee}$	$\gamma_{2,ee}$	$\gamma_{3,ee}$	$\gamma_{1,es}$	$\gamma_{2,es}$	$\gamma_{3,es}$
Replicates	4	4	4	4	4	4	4	4	4	4
Mean Est	251.97	235.27	191.80	55.25	218.08	12.43	-5.67	-202.35	8.01	-5.14
Std Err	12.67	5.58	6.59	1.30	37.21	2.84	2.34	12.93	1.27	1.13
Bias	-153.53	-98.73	-58.20	-17.62	-81.92	-1.57	1.33	-2.35	0.01	-0.14
Truth	405.50	334.00	250.00	72.86	300.00	14.00	-7.00	-200.00	8.00	-5.00

Table 5: Summary of Simulation under Bimodal Errors for Näive, LME, SMEMN, SMEMDP Models, respectively

	Replicates	Naive	Linear	SMEMN	SMEMDP
Normal	2	144869.7	113975.56	71308.89	-
	4	100096.78	75098.25	35902.73	35602.62
Skewed	2	138895.41	101586.02	70058.3	-
	4	91110.6	63178.21	35820.27	35449
Bimodal	2	93857.63	46486	42968.21	-
	4	84228.86	35665.26	30936.62	28335.85

Table 6: PMSE for EE Regression

	Replicates	Naive	Linear	SMEMN	SMEMDP
Normal	2	61567.63	43557.94	38272.77	-
	4	36238.56	24760.97	20654.15	20490.67
Skewed	2	49161.19	35783.01	31818.48	-
	4	27917.97	19980.43	17095.63	16969.31
Bimodal	2	24583.02	18758.07	17057.83	-
	4	15772.86	11376.04	10362.42	10263.01

Table 7: PMSE for ΔES Regression

6.3 Closer Look at One Simulated Dataset

In this section we provide an example analysis using one of the simulated data sets. This gives an idea of the type of results for an individual data set. We provide posterior quantiles to understand uncertainty in estimates. We also provide posterior predictive assessments for our modeling assumptions. The posterior predictive distributions of interest are given in (77),(78),(79). For the measurement error models, we are interested in the posterior predictive distributions for Y, W, X because we are making distributional assumptions on all of them. The forms for these distributions are given below.

$$p(Y^*|W,Y,Z) = \int \int p(Y^*|\theta,X,Z)p(\theta,X|Y,W)d\theta dX \tag{77}$$

$$p(W^*|W,Y,Z) = \int \int p(W^*|\theta,X,Z)p(\theta,X|Y,W)d\theta dX \tag{78}$$

$$p(X^*|W,Y,Z) = \int \int p(X^*|\theta)p(\theta,X|Y,W)d\theta dX \tag{79}$$

For these, we are interested in whether our model can capture the true quantiles (min, 0.05, 0.1, 0.25, 0.75, 0.9, 0.95, max) of the bivariate latent vector and for both the cheap and gold standard

measurements. To get a distribution of values for our quantiles of interest to compare to the truth, we simulate new data based on MCMC iterations. For example, to calculate distributions for the posterior quantiles for the bivariate latent variable X, for each iteration from the MCMC, we take the sampled values of parameters for the joint distribution of X and simulate a new set of X values according to how it is specified in the model, and from that new set of values we calculate the quantiles of interest. Doing this for many MCMC draws, we get distributions for the quantiles of interest. We then compare these distributions to the truth, and if the truth lies well within the bounds it shows the model accurately captures that feature of the data. If the truth lies outside the bounds, it indicates a lack of fit for that component. Because these distributions for quantiles can have large spreads, we also need to pay attention to their ranges as a check for precision. In this section we will only analyze a single data set with 2 replicates per individual with model SMEMN fit. For brevity, we will only present results under one dataset with skewed errors.

Running 3 chains of 12,000 iterations with 2000 burn in results in Gelman-Rubin diagnostics of less than 1.04 for all parameters. Trace plots for parameters show no signs of non-convergence. Due to the adaptive tuning during burn in, the acceptance rates for the latent variables hovers between 17% and 40% for the 300 individuals.

Table 8 give posterior quantiles for this model as well as the truth. These regression coefficients have nice interpretation as they are the linear component of the mean function. $\gamma_{1,ee}$ represents the change in reported EE, all else held equal, for the device user being male. In our example, the truth is 300, which means for two identical people except one male, one female, the cheap measurement tool is biased such that it will record the male as having an EE 300 higher than the female. $\gamma_{2,ee}$ represents the effect of BMI on the device, all else held equal. If BMI increases by 1, then the mean EE reported by the device increases by $\gamma_{2,ee}$. $\gamma_{3,ee}$ represents the effect of age, and the interpretation is similar. Again, we could have included many other/different covariates, and their respective interpretations would be easy as well.

Figure 10 shows plots of posterior predictive assessment for the latent variables skewed errors. Under model SMEMN, there is much less flexibility than with the Dirichlet Process. Some of that is seen in this figure, as the minimum EE is not captured and has a large amount of variability. Maximum ΔES is also only barely covered. Otherwise, these diagnostics look surprisingly good given the truth is clearly non-normal. This result is encouraging since we were concerned about the performance

of a single bivariate normal distribution versus the Dirichlet Process for the bivariate latent vector. We constructed the same diagnostic plot where the Dirichlet Process was used and unsurprisingly it looked better in terms of coverage and precision, however, not by a lot though. However, this was only under 4 replicates because of the issues with 2 replicates described earlier.

Figure 11 show the same plots except for gold standard measurements. Even under skewed errors, there doesn't seem to be any glaring issues as the truth is captured by the distribution. The only concern would be the amount of variability in the distribution for the minimum and maximum values. Figure 12 show the same plots except for cheap measurements. The issue here is with capturing the minimum value as recorded by cheap measurement. Otherwise, the remaining plots look good, again even under skewed error terms. This gives some confidence in an amount of robustness to misspecification of measurement error distribution.

Figure 13 gives one of the main desired outcomes of this study. This figure plots the fitted spline between the truth and the observed cheap measurement. The plot on the left is for the cheap EE measurement, on the right is for ΔES. The points plotted are from the individual simulated data where the y value is the mean of the two replicates. The red line is the mean spline function from this data set and we also randomly selected 500 MCMC iteration draws for the spline, and plotted them behind the mean. This is of high interest to us because it shows the relationship between what we observe, and what we are ultimately interested in knowing. To calibrate the cheap measurements, we would use the information from this plot as well as the posteriors for the γ. coefficients that refer to demographic variables. Figure 14 gives the distribution of the number of knots for the spline for both the EE and ΔES splines. This shows the splines are not overly complex as they typically use 4 or less knots.

7 Calibration

One of the main goals of this research was to provide the best possible estimates of EE and ΔES using cheap measurement tools. The path we took to create those estimates was by modeling the bias and error in the cheap measurements by using replicates of unbiased gold standard measurements. In this section we describe how we can use our measurement error model to calibrate what a cheap measurement reports for either EE or ΔES and return a more accurate estimate along with measures
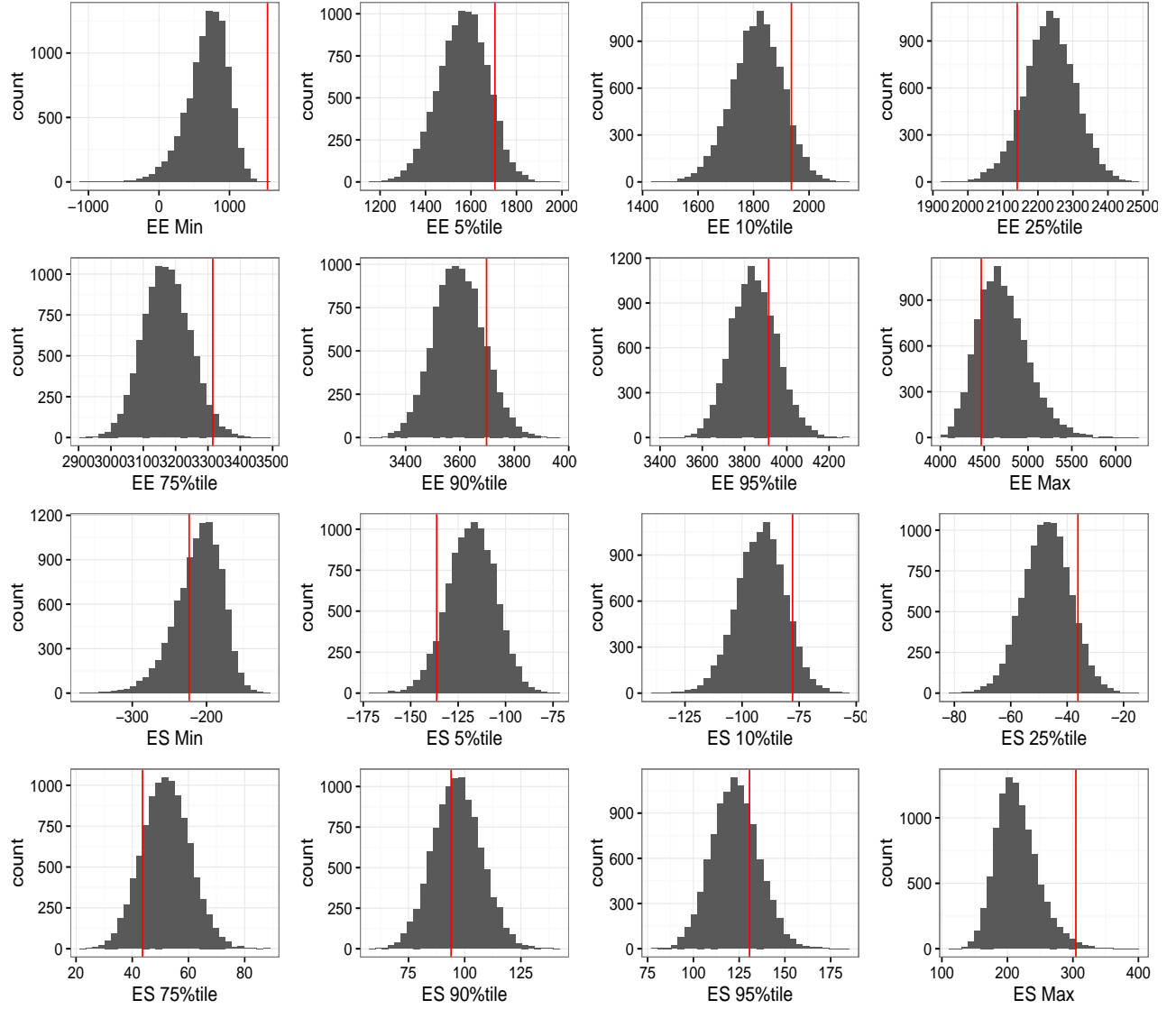
Figure 10: Posterior Predictive Discrepancy Measures For X^{EE} and $X^{\Delta ES}$ for Model SMEMN with Skewed Normal Errors
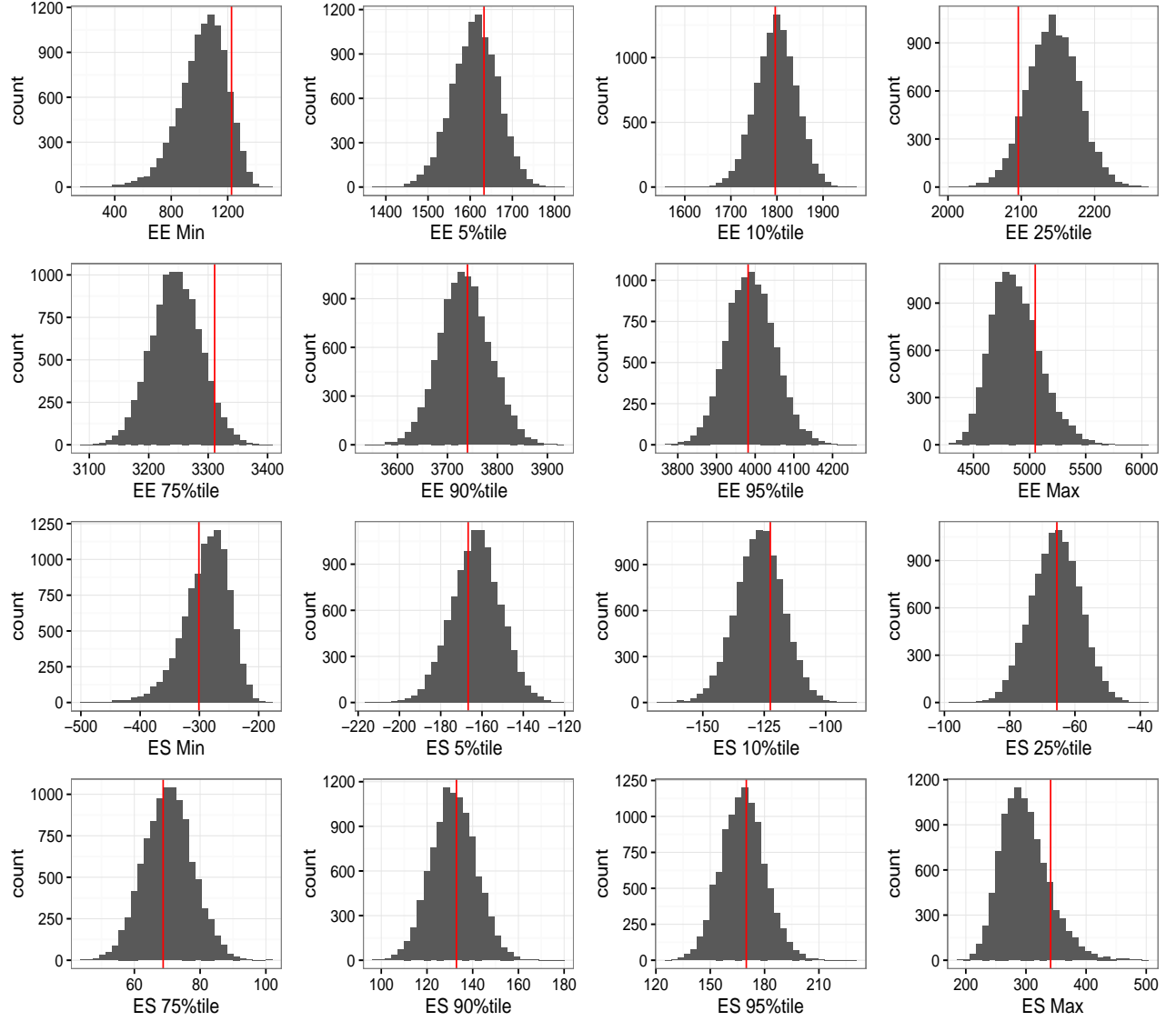
Figure 11: Posterior Predictive Discrepancy Measures For W^{EE} and $W^{\Delta ES}$ for Model SMEMN with Skewed Normal Errors
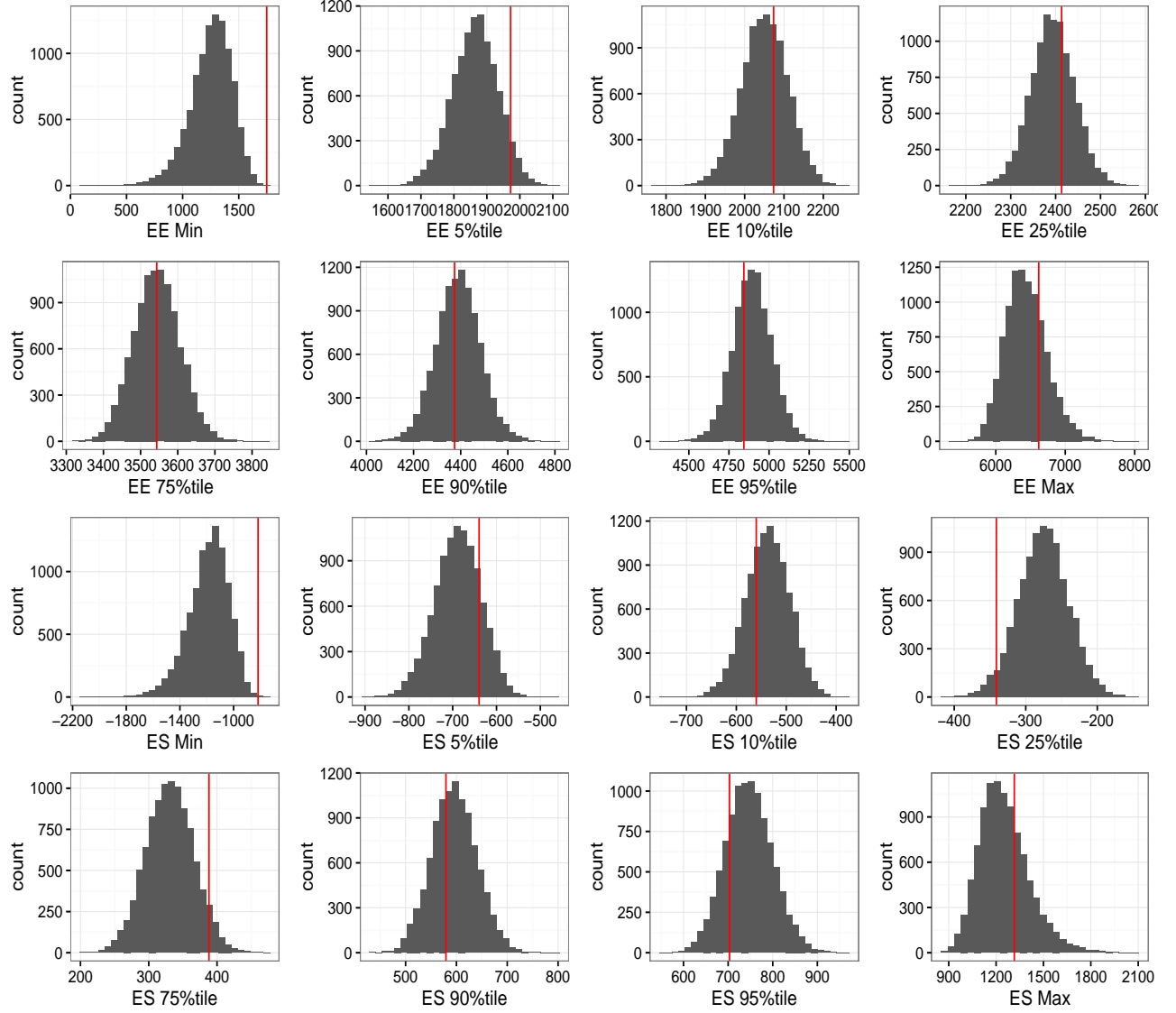
Figure 12: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model SMEMN with Skewed Normal Errors
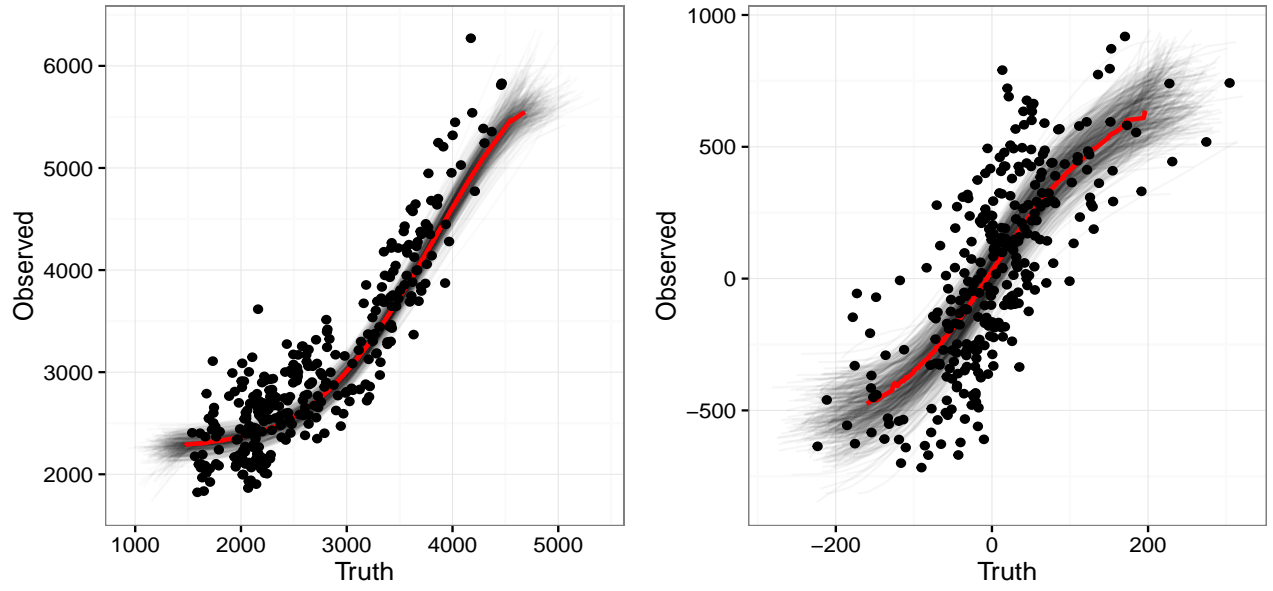
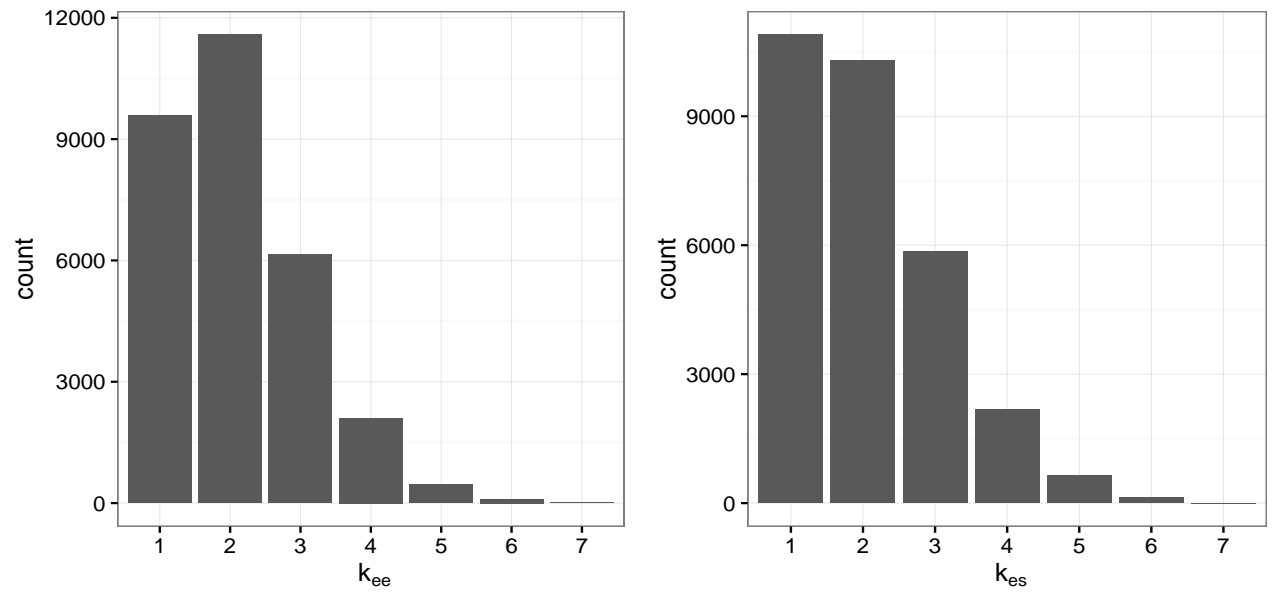Figure 13: Spline function for Model SMEMN with Skewed Errors

Figure 14: Number of knots for Model SMEMN with Skewed Errors

	2.5%	25%	50%	75%	97.5%	Truth
σ_{yee}	368.55	384.27	392.91	402.25	422.86	422*
σ_{yes}	290.23	302.64	309.63	316.91	331.80	344*
σ_{wee}	237.48	249.19	255.82	262.64	276.41	250
σ_{wes}	64.86	67.77	69.41	71.11	74.52	72.862
$\gamma_{1,ee}$	268.58	294.32	308.09	321.41	347.99	300
$\gamma_{2,ee}$	6.00	8.38	9.64	10.89	13.28	14
$\gamma_{3,ee}$	-13.79	-11.69	-10.60	-9.54	-7.47	-7
$\gamma_{1,es}$	-256.60	-234.20	-222.84	-211.52	-188.25	-200
$\gamma_{2,es}$	4.32	6.70	7.92	9.19	11.60	8
$\gamma_{3,es}$	-7.18	-5.17	-4.15	-3.07	-1.00	-5

Table 8: Parameter estimates for Model SMEMN for Skewed errors with correct specification of within person errors

of uncertainty.

Once we have fit our model to a specific set of cheap measurement tools for EE and ΔES, we can use our model to calibrate measurements from those cheap measurements in the future. That is, given an estimate of EE or ΔES from a cheap measurement tool and some demographic information, we can return a better estimate of the truth as well as a credible interval that shows the uncertainty in the estimate. Calibration for our models simply amounts to finding the inverse of the fitted models in (3), (4) as a function of Y instead of X, and Z are still measurement error free covariates as shown below.

$$X_{calibrated} = m^{-1}(y - \boldsymbol{\gamma}'Z) \tag{80}$$

We cannot find the inverse of our regression function in closed form so we will find it numerically instead. To do so, we use `optimize` in `R` for the function $|f(x) - y^*|$ where $f()$ represents the regression function and y^* is the observed cheap measurement minus the vector of coefficients $\boldsymbol{\gamma}$ multiplied by the individuals covariate values Z. We specify a number of posterior draws of regression coefficients and latent variables R, to then get a set of predicted values for X. We use each of these posterior draws of regression coefficients, latent variables, and spline functions to get a calibrated value. Using this set of calibrated values, we can then get a point estimate for the value of X as well as a credible interval using percentiles from the set of estimates. The procedure for getting this distribution for the calibration for individual i is as follows:

For r = 1,...R

1. Calculate $y_i^* = y_i - \boldsymbol{\gamma}^{(r)'} Z_i$, where Z_i are the covariate values for individual i

2. Use `optimize` for the function $|f_i(x) - y_i^*|$ to choose the value of x that will minimize the afore-mentioned criteria, call this $x_{i,calibrated}$. $f_i(x)$ is the predicted value of y_i for the given value x using the MCMC draw for the spline coefficients $\boldsymbol{\beta}.^{(r)}$, latent variables $(X^{EE(r)}, X^{\Delta ES(r)})$, and knot locations $(r_{ee}^{(r)}, r_{es}^{(r)})$ from the rth draw of the chain

As an example, we want to calibrate three cheap measurements each from a different individual using Model SMEMN. We randomly select 3 individuals from the same data set used earlier to give results for model SMEMN. Individual 1 is male, BMI of 28.6, age 20.5; individual 2 is female, BMI of 21.5, age 30.1 and individual 3 is male, BMI 38.6 and age 22.8. Observed cheap measurements for these individuals, their true values, as well as 95% credible intervals for their calibrated truth under normal errors are given in Table 9. Figures 15 gives histograms of 1000 calibrated draws for each individual for EE and ΔES measurements under skewed errors.

Looking at the table and figure, one can see that the calibration helps pull the cheap measurement closer to the truth. One can also see that sometimes the cheap measurement is actually close to the truth. In all cases here, the calibration helped bring the cheap measurement very close to the truth or at least improved the estimate. Running this on many of the simulated individuals had similar results. The only situations where the cheap measurement gave a closer value to the truth compared to the calibrated value was when the cheap measurement and truth were very close together. This was not a common occurrence. What's even more exciting about this is that we have measures of uncertainty for these calibration estimates immediately, and thus we can get credible intervals easily.

	Lower	Median	Upper	Observed	Truth
EE	2567.87	2661.31	2745.54	3028.89	2424.49
	3453.99	3526.34	3626.37	4119.26	3588.12
	2576.33	2666.58	2754.15	2555.86	3467.12
ΔES	25.19	42.11	59.99	142.30	-18.80
	-104.40	-82.15	-63.68	-405.74	-21.08
	-8.65	3.98	16.49	96.06	-65.17

Table 9: 95% credible interval for calibration estimate for cheap EE measurements for Skewed Errors
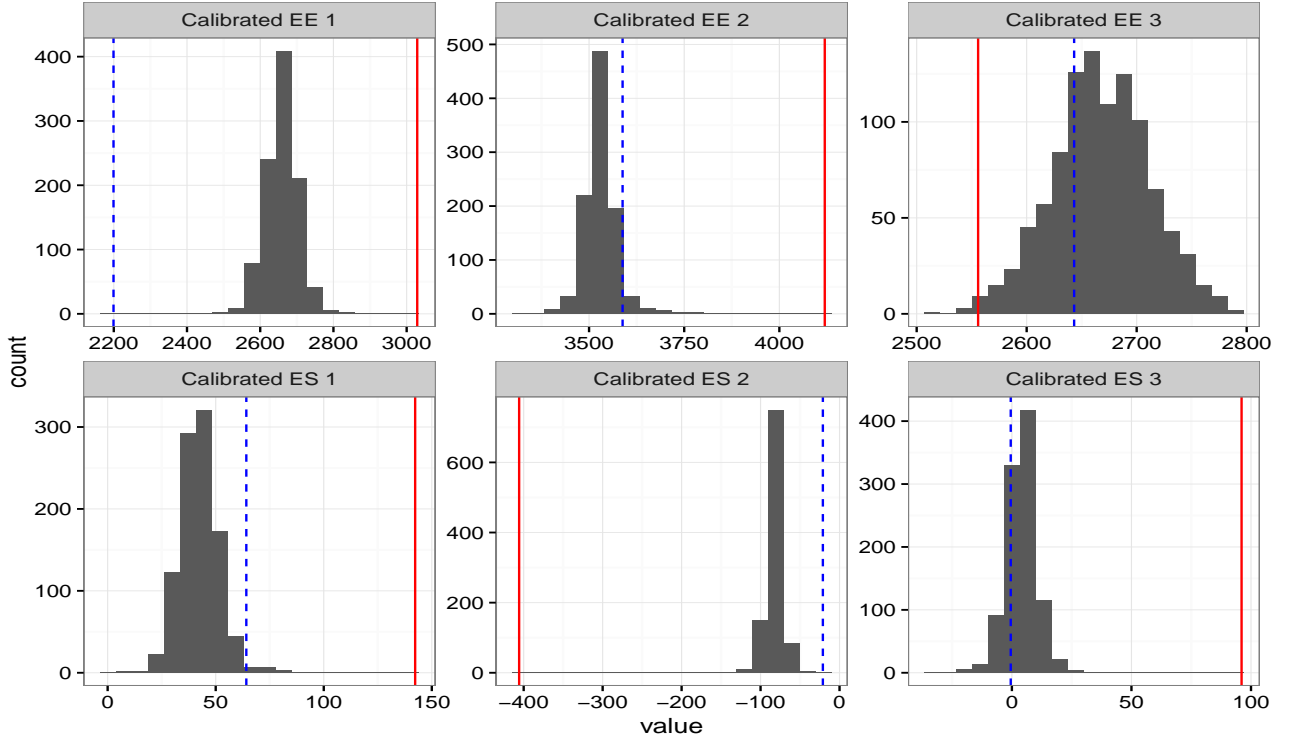
Figure 15: Posteriors of calibrated observations. Solid vertical line shows observed value from cheap measurement and dashed vertical line shows truth.

8 Discussion

In this chapter we presented a semi-parametric model for the measurement of energy balance via EE and ΔES. We assume gold standard measurements are unbiased to the truth while cheap measurements are some functions of the truth which we use splines to evaluate, plus a linear combination of demographic, error free covariates. This allows a very flexible relationship between a cheap measurement and its unobserved truth. We assumed normality for measurement error distributions which was justified through the use of diagnostics recommended by Caroll et al. We assumed gold standard measurements and cheap measurements were conditionally independent within and between each other given the latent vector $(X^{EE}, X^{\Delta ES})$. We modeled the latent vector $(X^{EE}, X^{\Delta ES})$ with both a bivariate normal distribution as well as a Dirichlet process. Although the Dirichlet process was flexible and a much weaker assumption, it required more replicate observations (mainly on gold standard measurements) than is feasible in order to give stable results. The normality assumption was able to give stable and surprisingly good results given the true structure of the latent variables. Because this model was able to give stable and pleasing results under the constraint of only two replicates of gold standard measurements, we believe this is the best model for this specific appli-

cation unless more than 2 replicates per person are available. The resulting estimates and PMSE show the superiority of our model to a simpler linear measurement error model and the näive model that doesn't take measurement error into consideration.

The main motivation for constructing this model was to model the error and bias in measurements so one can calibrate a cheap measurement tool's output in order to make it closer to the truth. We presented a simple way to do this calibration given a cheap measurement for EE and ΔES and values for gender, BMI, and age. Using a Bayesian approach we are easily able to get a posterior distribution for the calibrated estimate which provides a measure of uncertainty. Our example shows that in some cases this calibrated estimate could be far from the observed cheap measurement.

Further considerations to the model would include allowing for nonconstant variance. At the time we have limited data from a limited number of EE and ΔES measurement tools, but it is possible nonconstant variance will be an issue for some and not others. We saw some signs of this already as the Sensewear armband showed signs of nonconstant variance while DXA did not. In this research we opted to assume constant variance for simplicity, but an extension to allow for this would not be difficult. Other measurement error free covariates could be easily added to the regression functions in (3),(4) if desired. Furthermore, we are considering both within person variability and measurement error to be additive in the cheap measurement case which is of course not true, it has a complicated relationship through our estimated functions $m_.(\cdot)$.

References

[1] Y.T. Lagerros and P. Lagiou. Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *European Journal of Epidemiology*, 205:353–62, 2007.

[2] C. Bouten, W. Verboeket-van de Venne, K.R. Westerterp, M. Verduin, and J.D. Janssen. Daily physical activity assessment: comparison between movement registration and doubly labeled water. *Journal of Applied Physiology*, 81:1019–26, 1996.

[3] Diana M. Thomas, Corby K. Martin, Steven Heymsfield, Leanne M. Redman, Dale A. Schoeller, and James A. Levine. A simple model predicting individual weight change in humans. *Journal of Biological Dynamics*, 5:579–99, 2011.

[4] Kevin D Hall and Carson C Chow. Estimating changes in free-living energy intake and its confidence interval. *American Journal of Clinical Nutrition*, 94:66–74, 2011.

[5] L Anne Gilmore, Eric Ravussin, George Bray, Hongmei Han, and Leanne Redman. An objective estimate of energy intake during weight gain using the intake-balance method. *American Journal of Clinical Nutrition*, 100:806–12, 2014.

[6] Arjun Sanghvi, Leanne M. Redman, Corby K. Martin, Eric Ravussin, and Kevin D. Hall. Validation of an inexpensive and accurate mathematical method to measure long-term changes in free-living energy intake. *American Journal of Clinical Nutrition*, 102:353–8, 2015.

[7] Susan B. Racette, Sai Krupa Das, Manjushri Bhapkar, Evan C. Hadley, Susan B. Robers, Eric Ravussin, Carl Pieper, James P. DeLany, William E. Kraus, James Rochon, and Leanne M. Redman. Approaches for quantifying energy intake and %calorie restriction during calorie restriction interventions in humans: the multicenter calerie study. *Am J Physiol Endorinol Metab*, 302:441–8, 2011.

[8] Gregory A. Hand, Robin P. Shook, Amanda E. Paluch, Meghan Baruth, Patrick Crowley, Jason R. Jaggers, Vivek K. Prasad, Thomas G. Hurley, James R. Hebert, Daniel P. O'Connor, Edward Archer, Stephanie Burgess, and Steven N. Blair. The energy balance study: The design and baseline results for a longitudinal study of energy balance. *Research Quarterly for Exercise and Sport*, 84:275–86, 2013.

[9] Raymond J. Caroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceani. *Measurement Error in Nonlinear Models*. Chapman and Hall/CRC, 2006.

[10] Samiran Sinha, Bani K. Mallick, Victor Kipnis, and Raymond J. Carroll. Semiparametric bayesian analysis of nutritional epidemiology data in the presence of measurement error. *Biometrics*, 66:444–54, 2010.

[11] Kunio Takezawa. *Introduction to Nonparametric Regression*. Wiley Series in Probability and Statistics, 2006.

[12] Abhra Sarkar, Bani K. Mallick, and Raymond J. Carroll. Bayesian semiparametric regression in the presence of conditionally heteroscedastic measurement and regression errors. *Biometrics*, 70:823–34, 2014.

[13] CDC. Cdc website. `https://www.cdc.gov/nchs/data/nhanes/databriefs/adultweight.pdf`. Accessed: 2016-10-10.

[14] Carmen Fernandez and Mark F. J. Steel. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93:359–71, 1998.

[15] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2014.

[16] Peter J. Green. Reversible jump markov chain monte carlo computation and and bayesian model determination. *Biometrika*, 82:711–32, 1995.

[17] Ilaria DiMatteo, Christopher R. Genovese, and Robert E. Kass. Bayesian curve-fitting with free-knot splines. *Biometrika*, 88:1055–71, 2001.

[18] Mary J. Lindstrom. Bayesian estimation of free-knot splines using reversible jumps. *Computational Statistics and Data Analysis*, 41:255–69, 2002.

[19] Matthew S. Johnson. Modeling dichotomous item responses with free-knot splines. *Computational Statistics and Data Analysis*, 51:4178–92, 2007.

[20] D.G.T Denison, B.K. Mallick, and A.F.M. Smith. Automatic bayesian curve fitting. *Journal of the Royal Statistical Society*, 60:333–50, 1998.