

Energy Balance Measurement Error Models

Daniel Ries

November 8, 2016

1 Model Framework

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 (EE) or stored (Δ ES). Fortunately, commonly accepted gold standard measurements exist for both EE (DLW) [1],[2],[3],[4],[5],[6], [7] and Δ ES (DXA) [3],[6],[5]. We will denote the measured average daily EE by DLW for subject i over time period j , W_{ij}^{EE} , and average daily change in energy stores measured by DXA for subject i over time period j , $W_{ij}^{\Delta ES}$. A positive value for Δ ES 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, we will always talk about average kcals per day. We also have cheaper measurements for EE and Δ ES. 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 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 Δ ES, we will keep the notation here general since in any given situation we will be talking only about one specific measurement tool of each. 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 Δ ES for subject i . We define *usual* as a long run average of the true EE and Δ ES. Let X_i^{EE} represent the *usual* daily EE for subject i and $X_i^{\Delta ES}$ represent the *usual* daily Δ ES 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 . Then:

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:

$$f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES} | Z_i, \boldsymbol{\theta}) \quad (1)$$

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 Δ ES for subject i), and covariates Z_i , we believe it is reasonable to 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
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 Δ ES
5. Given only X_i^{EE} and Z_i , W_{ij}^{EE} is independent from $W_{ij}^{\Delta ES}$ and vice versa for Δ ES

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.

Based on the assumptions made in the previous paragraphs, we can specify our full model (1) by:

$$\begin{aligned}
f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES} | Z_i, \boldsymbol{\theta}) &= \int \int f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES}, X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta}) dX_i^{EE} dX_i^{\Delta ES} \\
&= \int_{\mathcal{X}_{es}} \int_{\mathcal{X}_{ee}} f(W_{ij}^{EE} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{wee}) f(W_{ij}^{\Delta ES} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{wes}) \times \\
&\quad f(Y_{ij}^{EE} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{yee}) f(Y_{ij}^{\Delta ES} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{yes}) \times \\
&\quad f(X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta}_x) dX_i^{EE} dX_i^{\Delta ES} \\
&= \int_{\mathcal{X}_{es}} \int_{\mathcal{X}_{ee}} f(W_{ij}^{EE} | X_i^{EE}, Z_i, \boldsymbol{\theta}_{wee}) f(W_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{wes}) \times \\
&\quad f(Y_{ij}^{EE} | X_i^{EE}, Z_i, \boldsymbol{\theta}_{yee}) f(Y_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{yes}) \times \\
&\quad f(X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta}_x) dX_i^{EE} dX_i^{\Delta ES}
\end{aligned} \tag{2}$$

Further, we specify the models for the W 's and Y 's as:

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

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

$$W_{ij}^{EE} = X_i^{EE} + \nu_{ij}^{EE} \quad (5)$$

$$W_{ij}^{\Delta ES} = X_i^{\Delta ES} + \nu_{ij}^{\Delta ES} \quad (6)$$

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 Δ ES, 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 $m(\cdot)$. We cannot directly estimate both within person variability and measurement error simultaneously due to realistic data constraints. 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 Δ ES ($X_i^{\Delta ES}$), respectively, and the cheap measurements are giving biased estimates of (possibly) unknown form.

We will assume subjects in the study will have no influence on each other in terms of consumption or exercise behavior, and thus subjects are assumed to be 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 individual's usual value*, which means once we know an individual's baseline, we expect variation from their to be independent day to day. The likelihood function for an individual can then be written as:

$$L_i(\boldsymbol{\theta}) = \prod_{j=1}^J f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES} | Z_i, \boldsymbol{\theta}) \quad (7)$$

where the RHS is expanded in (2). Given the independence assumptions in the previous paragraph,

the full likelihood is

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

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 Δ ES 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 begining 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.

In our model, we need Δ ES 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 (9). Because we are assuming energy stores are characterized only as either Fat Mass or Fat Free Mass, 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 Δ ES in kcals. Our result for each individual will be either a positive or negative number in kcals.

$$\Delta ES = c_1 \frac{\Delta FM}{\Delta T} + c_2 \frac{\Delta FFM}{\Delta T} \quad (9)$$

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.

The goal in this exploratory exercise is to determine the general form and distribution of the measurements 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 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. Note the armband for EE is not a gold standard measure, but there is absolutely no replication for DLW measurements in this data. 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.

Figure (6) show plots of EE and ΔES faceted by age and BMI. These were measured by DLW and DXA, respectively. A simple linear regression model is fit in blue along with confidence bands to each pair. Looking first at EE vs age, we see a somewhat surprising relationship in that they look almost unrelated. It is misleading though since these data were only collected on individuals aged 20-35, so if we're sampling a larger age base, we would expect there to be more of a (negative) relationship. EE vs BMI shows a moderate, positive relationship. This makes sense because because the majority of calories people burn is through their resting metabolic rate (RMR). Part of this is directly related to the amount of oxygen we consume, and the bigger and more active we are, the more oxygen we consume. The amount of calories burned from physical activity are actually much smaller and this explains the relationship we see in the data.

ΔES vs age shows a weak, negative relationship indicating, at least with the EBS age group, that older people tend to burn more energy than they consume, on average. ΔES vs BMI shows no meaningful relationship. Even though this exploratory analysis shows no relationship between these two variables, we will continue to include BMI as a covariate for ΔES due to its interpretability. Because these plots are based on data from the gold standard measurements, which we are assuming are unbiased for the truth, they give an approximate representation of the true relationship between these variables.

Figure (7) shows plots of EE vs age (left) and BMI (right). EE in these plots is measured by the

armband (cheap measurement). Because all 430 subjects received measurements of EE with the armband, these plots include the entire data set. Looking at EE vs age there appears to be a more obvious negative relationship which we would expect, even in this limited age group. EE vs BMI also shows an obvious relationship, except positive. Again this is what we'd expect as discussed previously.

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 considerable 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 it and the other 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 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 Simulated Data

In this section we describe how we simulate a data set that reflects our problem. 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 need to simulate data for all the components in (1) as well as the latent variables $(X_i^{EE}, X_i^{\Delta ES})$. First we will describe how we simulate the covariate values. Then we will explain how we simulate (X_i^{EE}, X_i^{EI}) . Then we will show how we simulate within person variability. Lastly we will explain

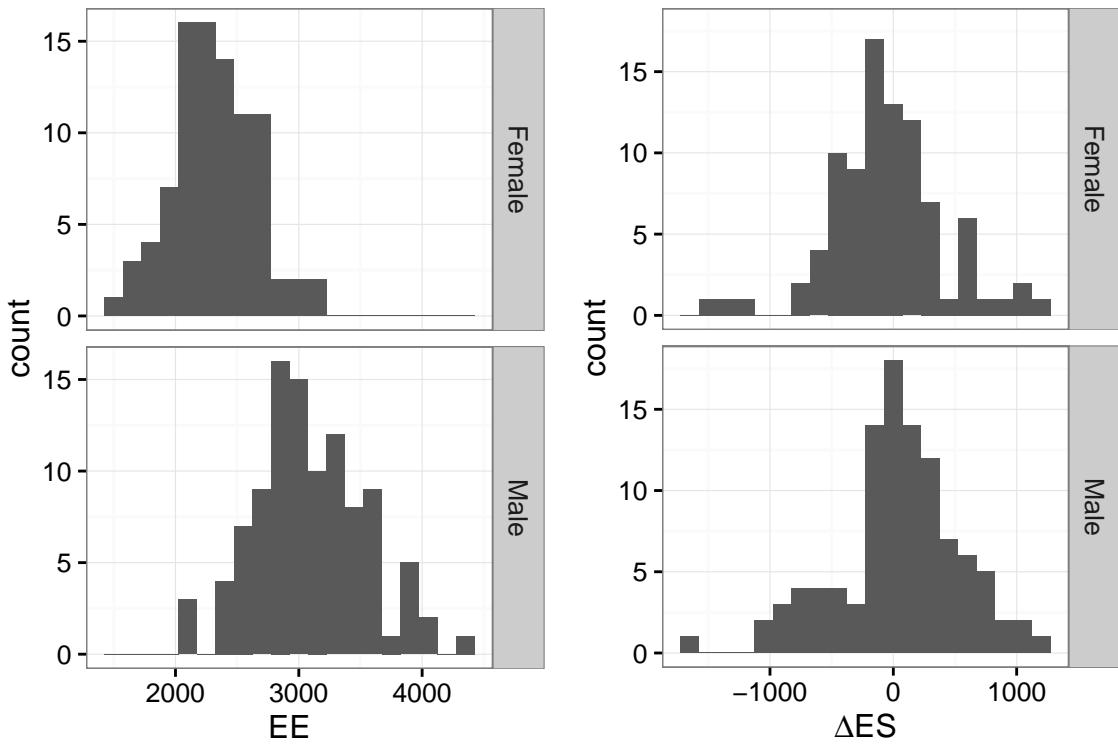


Figure 1: Distribution of EE (left) measured by DLW and Δ ES measured by DXA (right). Females on top, Males on bottom. Includes data from EBS from individuals who received DLW measurements.

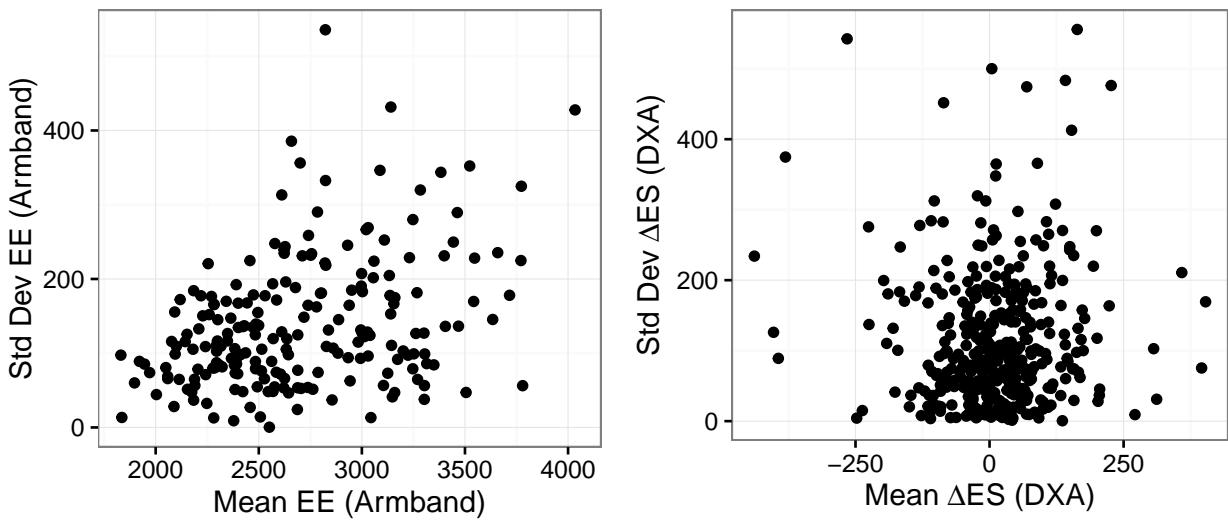


Figure 2: Variance Diagnostic Plots. The left shows the mean EE (Armband) for an individual vs his/her standard deviation of EE based on replicates. The right is the same except for Δ ES measured by DXA.

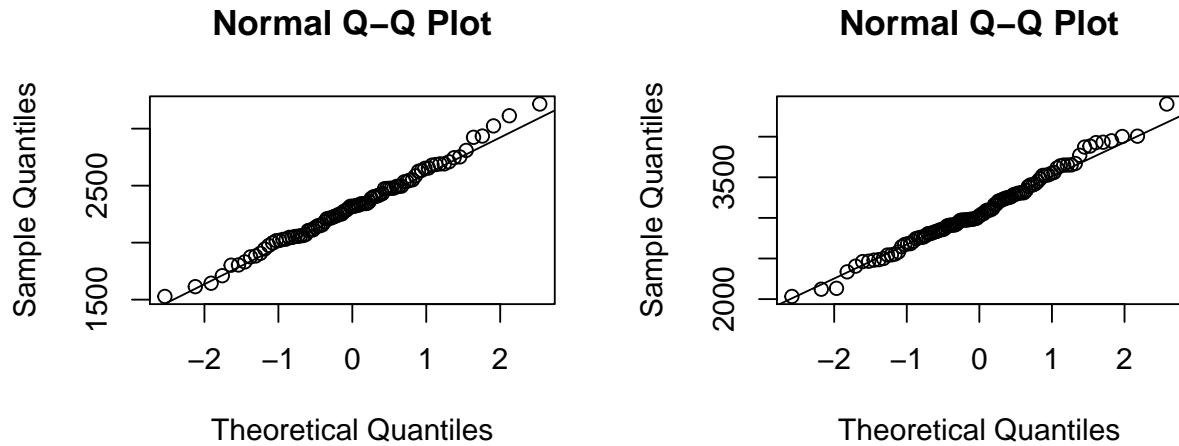


Figure 3: Normal Quantile Plots for EE as measured by DLW for Female (left) and Male (right). Data is from those who received DLW in EBS.

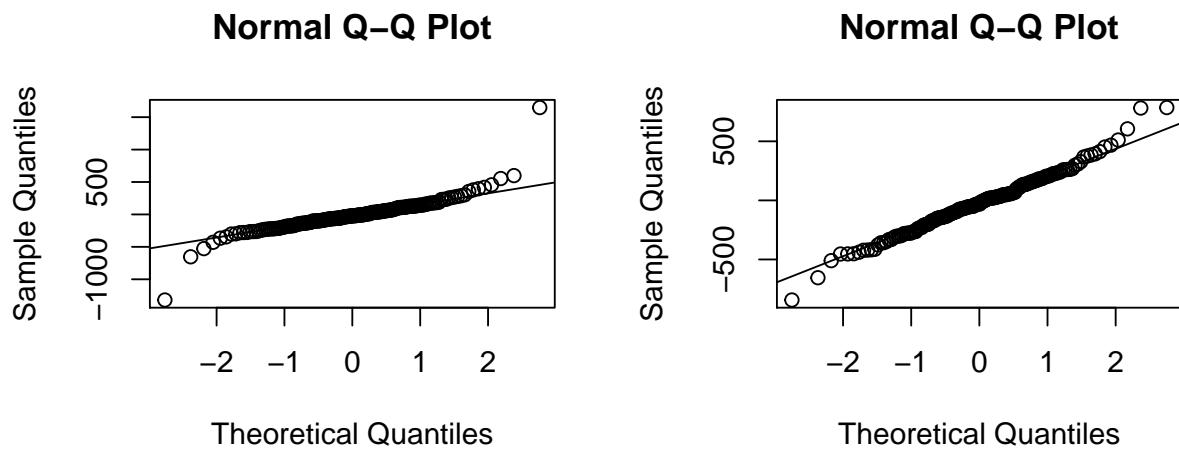


Figure 4: Normal Quantile Plots for differenced Δ ES ((12m DXA - 9m DXA)-(9m DXA - 6m DXA)) as measured by DXA for Female (left) and Male (right). Data is from those who had DXA scans at 6,9,12 months

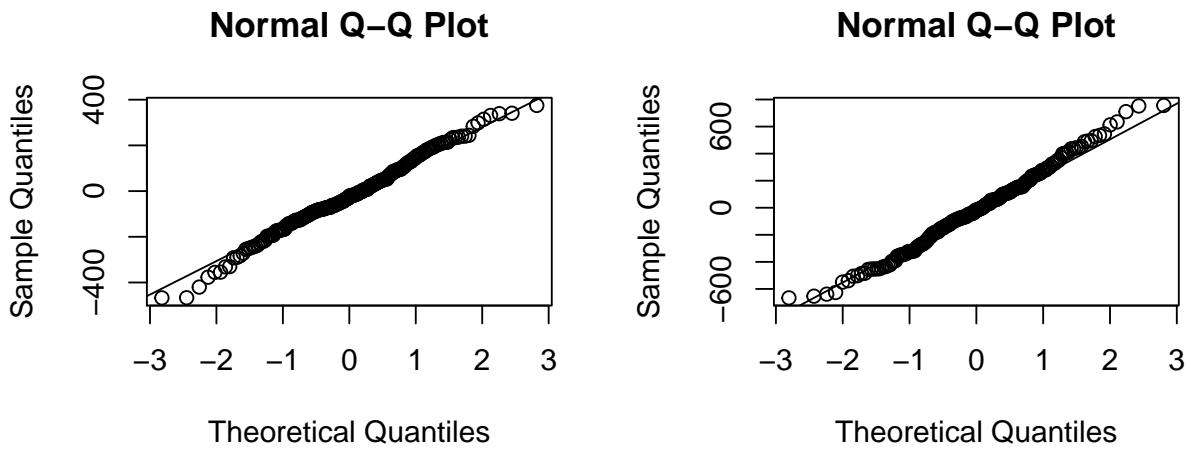


Figure 5: Normal Quantile Plots for differenced EE (12m EE - 9m EE) as measured by armbands for Female (left) and Male (right). Based on entire EBS individuals who remained for entire 12 months

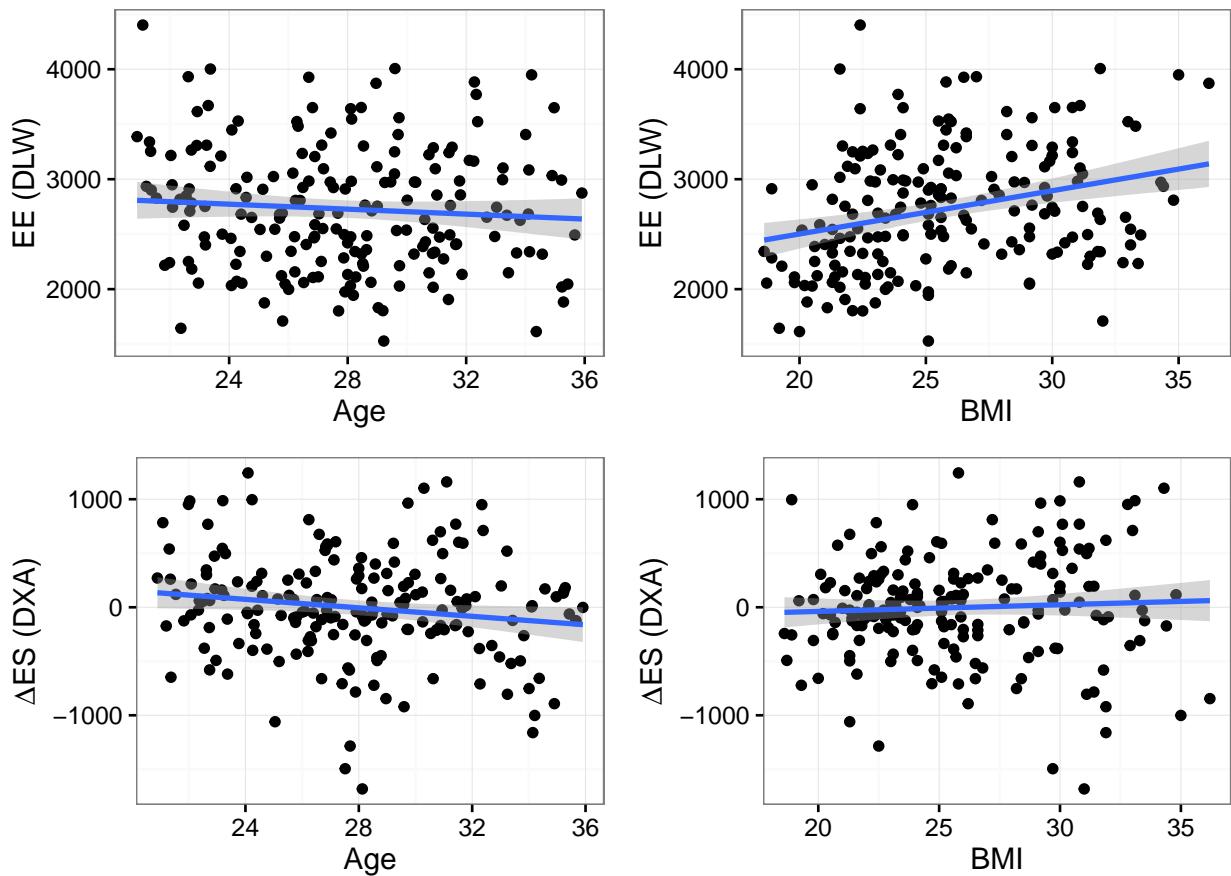


Figure 6: EE (top) and Δ ES (bottom) vs Age (left) and BMI (right). EE measured by DLW and Δ ES measured by DXA. Only uses data from EBS on those who received DLW.

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.

<i>Dependent variable:</i>	
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)
<hr/>	
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

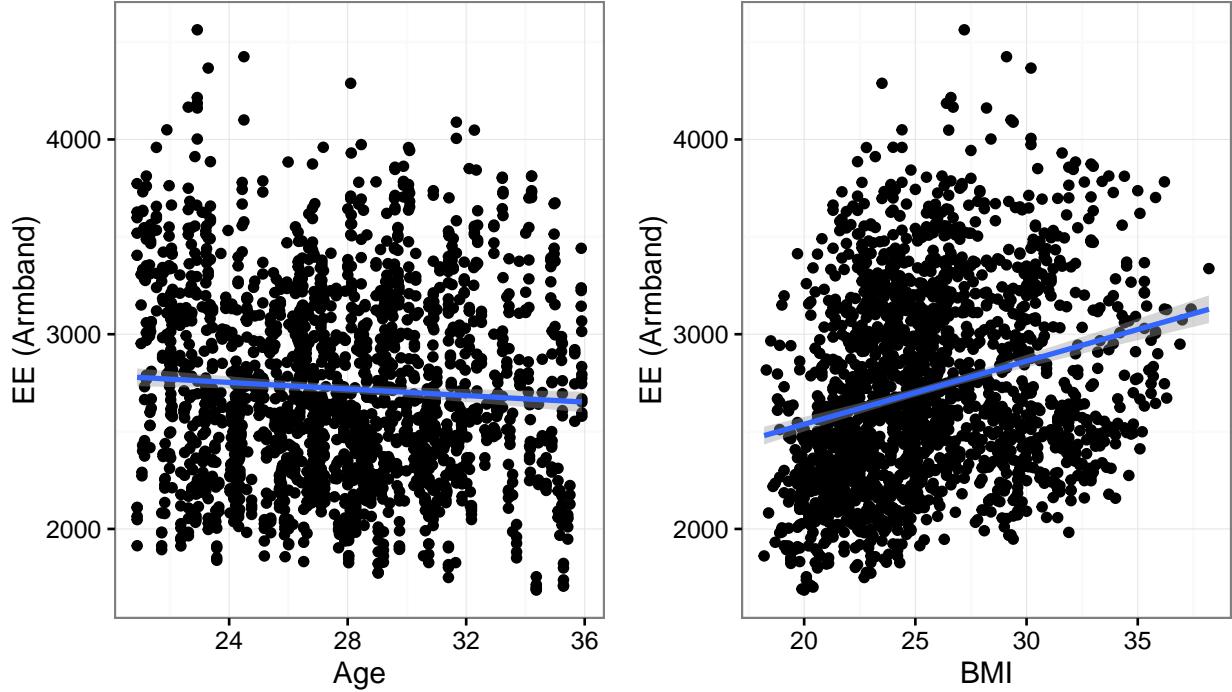


Figure 7: EE vs Age (left). EE vs BMI (right). EE measured by armband (cheap measurement). Includes all EBS data.

how we simulate gold standard and cheap measurements. We will explore measurement errors for the gold standard and cheap measurements in 3 different situations: normal errors, skewed errors, bimodal errors.

3.1 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 simulated from a Normal(27,5) since the average BMI for an American is around 27 [10]. All these values were simulated independently. Let Z be the matrix of dimension 300x3 with all covariate values.

	Mean EE	Mean EI	Std Dev EE	Std Dev EI	Cor(EE,EI)
Component 1	$1500 + \gamma_{ee}Z_{comp,1}$	$1590 + \gamma_{ei}Z_1$	50	70	0.4376
Component 2	$2000 + \gamma_{ee}Z_{comp,2}$	$2090 + \gamma_{ei}Z_2$	50	65	0.4376
Component 3	$2300 + \gamma_{ee}Z_{comp,3}$	$2390 + \gamma_{ei}Z_3$	90	120	0.4376
Component 4	$2600 + \gamma_{ee}Z_{comp,4}$	$2690 + \gamma_{ei}Z_4$	40	50	0.4376
Component 5	$3200 + \gamma_{ee}Z_{comp,5}$	$3290 + \gamma_{ei}Z_5$	90	120	0.4376

Table 2: Mean and Variances for X

3.2 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 . To be clear, within each component (60 simulated pairs of EE and EI), the mean vector (of length 2) is different for every individual simulated pair (X_i^{EE}, X_i^{EI}) , but the scale matrix remains constant. Across components, the mean vectors and the scale matrix changes as well. 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 (8) shows histograms for the latent variables for one simulated data set. Figure (9) gives contours of the bivariate distributions.

3.3 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 terms of (3), (4) this is a working assumption, and we will put the within person variability within

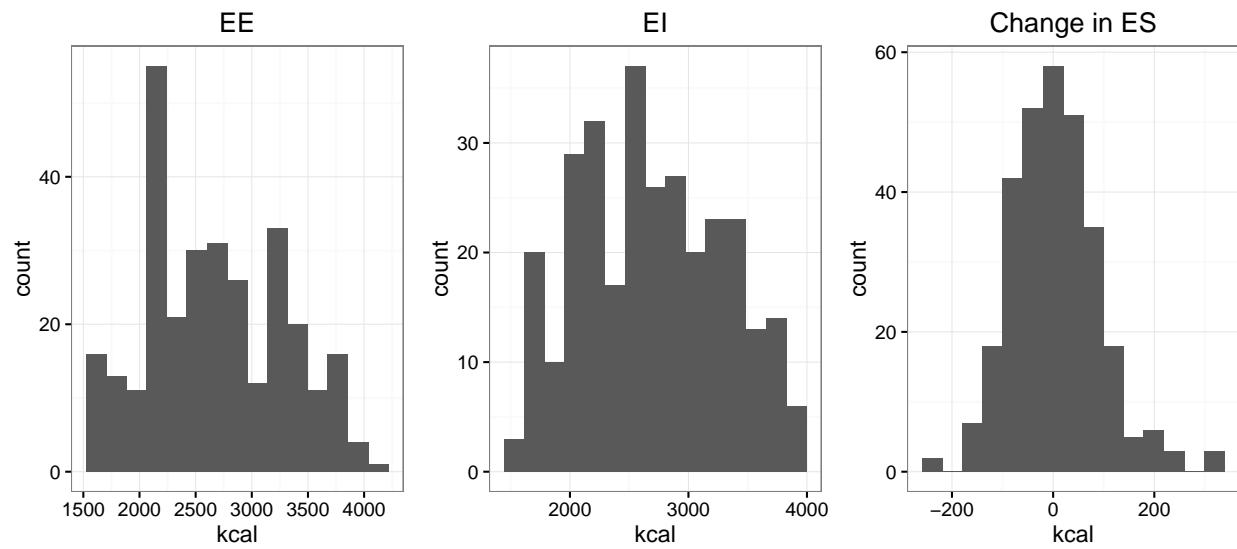


Figure 8: Distribution of simulated latent variables X from one simulated data set

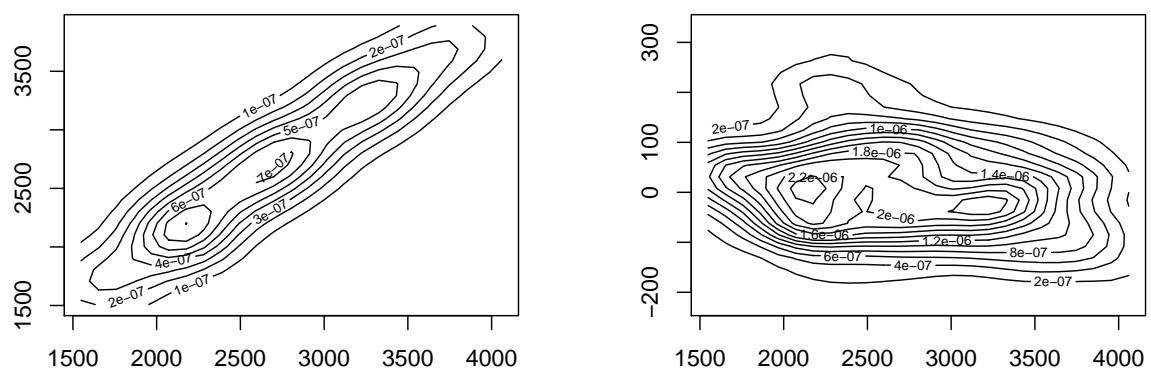


Figure 9: Contour plot os bivariate distributions of one simulated data set for latent variables. X^{EE} vs X^{EI} on left and X^{EE} vs $X^{\Delta ES}$ on right

the function . Although we cannot separate the two types of errors when modeling, we can separate them when we simulate data. For the gold standard measurements, let

$$\begin{aligned}\nu_{ij}^{EE} &= u_{ij}^{EE} + \delta_{ij}^{EE} \\ \nu_{ij}^{\Delta ES} &= u_{ij}^{\Delta ES} + \delta_{ij}^{\Delta ES}\end{aligned}\tag{10}$$

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:

$$\begin{aligned}X_{ij}^{EE} &= X_i^{EE} + \delta_{ij}^{EE} \\ X_{ij}^{\Delta ES} &= X_i^{\Delta ES} + \delta_{ij}^{\Delta ES}\end{aligned}\tag{11}$$

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 because the within person variability for EE and ΔES during time j are clearly related for person i . The δ terms will be assumed to be independent across individuals and within individuals across time as well as with all X . We simulate the within person variability terms $(\delta_{ij}^{EE}, \delta_{ij}^{\Delta ES})$ according to the bivariate distribution below:

$$(\delta_{ij}^{EE}, \delta_{ij}^{\Delta ES}) \stackrel{iid}{\sim} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\delta^{EE}}^2 & \rho_{\delta} \sigma_{\delta^{EE}} \sigma_{\delta^{\Delta ES}} \\ \rho_{\delta} \sigma_{\delta^{EE}} \sigma_{\delta^{\Delta ES}} & \sigma_{\delta^{\Delta ES}}^2 \end{bmatrix} \right) \tag{12}$$

$$\begin{aligned}\sigma_{\delta^{EE}} &= 150 \\ \sigma_{\delta^{\Delta ES}} &= 50 \\ \rho_{\delta} &= -0.266\end{aligned}\tag{13}$$

3.4 Gold Standard Measurements

We assume DLW and DXA are unbiased measurements of EE and Δ ES, respectively. These measurements are simulated according to (5),(6) where we further broke down ν in (10). 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 (14). The skewed normal distribution was introduced by Fernandez and Steel [11]. The pdf of the skewed normal is the limit of (18) 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 (15). 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 (16). 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 Δ ES. Values of the standard deviation parameters are given in (17). We simulate from the skewed normal distribution using the R package `fGarch`. Figure (10) shows density plots of all three distributions using our specified standard deviation and skew parameters.

$$\begin{aligned}u_{ij}^{EE} &\stackrel{iid}{\sim} N(0, \sigma_{\nu^{EE}}^2) \\ u_{ij}^{\Delta ES} &\stackrel{iid}{\sim} N(0, \sigma_{\nu^{\Delta ES}}^2)\end{aligned}\tag{14}$$

$$\begin{aligned}u_{ij}^{EE} &\stackrel{iid}{\sim} SN(0, \sigma_{\nu^{EE}}^2, 10) \\ u_{ij}^{\Delta ES} &\stackrel{iid}{\sim} SN(0, \sigma_{\nu^{\Delta ES}}^2, 10)\end{aligned}\tag{15}$$

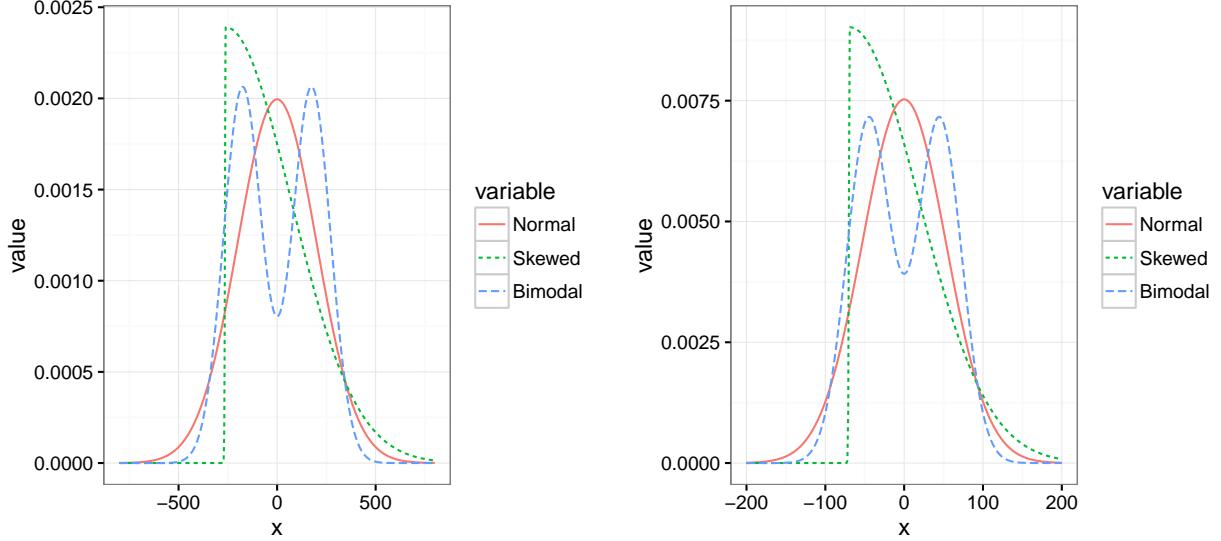


Figure 10: Measurement Error Distributions, EE on left and ΔES on right

$$\begin{aligned} u_{ij}^{EE} &\stackrel{iid}{\sim} 0.5 \times N(-175, \sigma_{\nu^{EE}}^2 - 175^2) + 0.5 \times N(175, \sigma_{\nu^{EE}}^2 - 175^2) \equiv BM(175, \sigma_{\nu^{EE}}^2) \\ u_{ij}^{\Delta ES} &\stackrel{iid}{\sim} 0.5 \times N(-45, \sigma_{\nu^{\Delta ES}}^2 - 45^2) + 0.5 \times N(45, \sigma_{\nu^{\Delta ES}}^2 - 45^2) \equiv BM(45, \sigma_{\nu^{\Delta ES}}^2) \end{aligned} \quad (16)$$

$$\begin{aligned} \sigma_{\nu^{EE}} &= 200 \\ \sigma_{\nu^{\Delta ES}} &= 53 \end{aligned} \quad (17)$$

$$\begin{aligned} p(y|\mu, \sigma, \nu, \gamma) &= \frac{2}{\gamma + \frac{1}{\gamma}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) (\pi\nu)^{1/2}} \sigma^{-1} \times \\ &\left\{ 1 + \frac{(y-\mu)^2}{\nu\sigma^2} \left(\frac{1}{\gamma^2} I_{[0,\infty)}(y-\mu) + \gamma^2 I_{(\infty,0)}(y-\mu) \right) \right\}^{-\frac{(\nu+1)}{2}} \end{aligned} \quad (18)$$

3.5 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 (11) 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:

$$\begin{aligned} e_{ij}^{EE} &\stackrel{iid}{\sim} N(0, \sigma_{\epsilon^{EE}}^2) \\ e_{ij}^{\Delta ES} &\stackrel{iid}{\sim} N(0, \sigma_{\epsilon^{\Delta ES}}^2) \end{aligned} \quad (19)$$

$$\begin{aligned} e_{ij}^{EE} &\stackrel{iid}{\sim} SN(0, \sigma_{\epsilon^{EE}}^2, 10) \\ e_{ij}^{\Delta ES} &\stackrel{iid}{\sim} SN(0, \sigma_{\epsilon^{\Delta ES}}^2, 10) \end{aligned} \quad (20)$$

$$\begin{aligned} e_{ij}^{EE} &\stackrel{iid}{\sim} 0.5 \times N(-350, \sigma_{\epsilon^{EE}}^2 - 350^2) + 0.5 \times N(350, \sigma_{\epsilon^{EE}}^2 - 350^2) \equiv BM(350, \sigma_{\epsilon^{EE}}^2) \\ e_{ij}^{\Delta ES} &\stackrel{iid}{\sim} 0.5 \times N(-190, \sigma_{\epsilon^{\Delta ES}}^2 - 190^2) + 0.5 \times N(190, \sigma_{\epsilon^{\Delta ES}}^2 - 190^2) \equiv BM(190, \sigma_{\epsilon^{\Delta ES}}^2) \end{aligned} \quad (21)$$

$$\begin{aligned} \sigma_{\epsilon^{EE}} &= 380 \\ \sigma_{\epsilon^{\Delta ES}} &= 210 \end{aligned} \quad (22)$$

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 $m_{es}(X, Z) = \frac{1000}{1+e^{-0.04X_i^{\Delta ES}}} - 2000 + X_i^{\Delta ES}$. These functions represent the bias involved in the cheap instruments as a function of $(X^{EE}, X^{\Delta ES})$ and Z . Figure (11) shows $m_{ee}(\cdot)$ on the left and $m_{es}(\cdot)$ on the right both against a $y = x$ line for comparison.

4 Models

In this section we present our model to assess the measurement error and bias involved in this energy balance problem. Before that we present a simpler model as well as a naive model to compare results

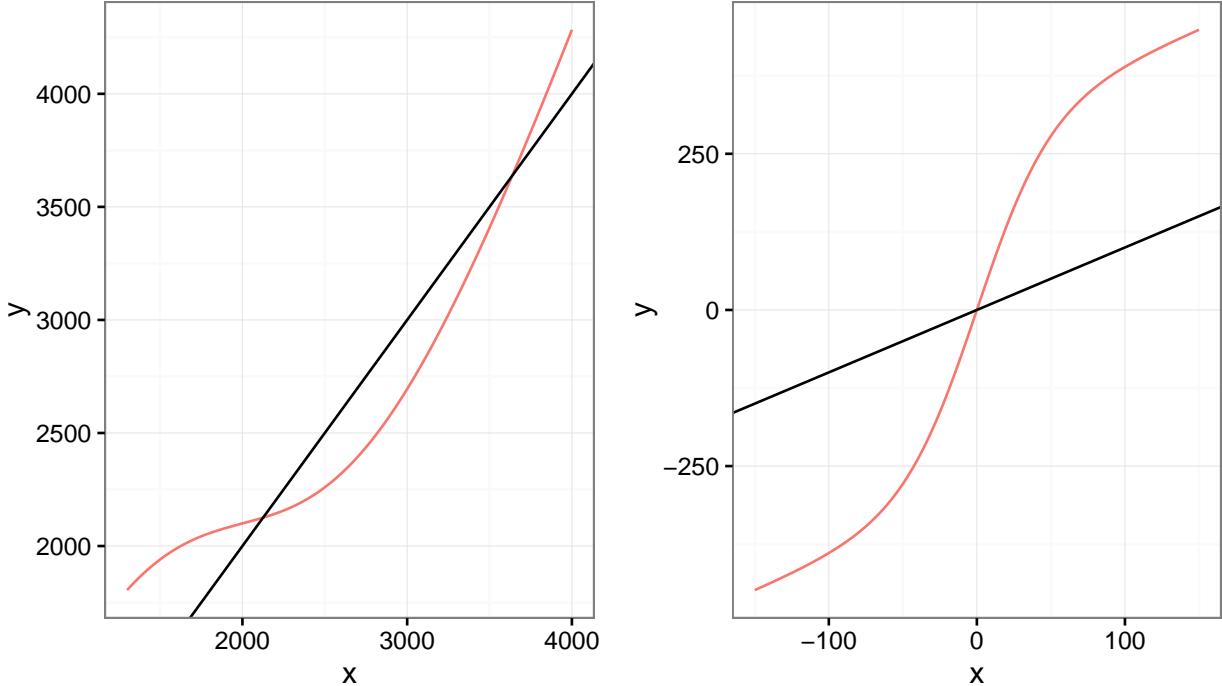


Figure 11: Plot of nonlinear functions $m_{ee}()$ (left) and $m_{es}()$ (right), and $Y=X$ is black for reference to unbiased measurement

with in the simulation study.

4.1 Model 0

The first model we will consider is what we call Model 0, or the naive 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 measurements of 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 naive model is represented as:

$$Y_{ij}^{EE} = \beta_{0,ee} + \beta_{1,ee} W_{ij}^{EE} + \gamma_{ee} Z_i + \epsilon_{ij}^{EE} \quad (23)$$

$$Y_{ij}^{\Delta ES} = \beta_{0,es} + \beta_{1,es} W_{ij}^{\Delta ES} + \gamma_{es} Z_i + \epsilon_{ij}^{\Delta ES} \quad (24)$$

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 :

$$(Y_{ij}^{EE}|W_{ij}^{EE}, Z_i, \boldsymbol{\theta}_{yee}) \stackrel{iid}{\sim} N(\beta_{0,ee} + \beta_{1,ee}W_{ij}^{EE} + \boldsymbol{\gamma}_{ee}Z_i, \sigma_{\epsilon^{EE}}^2) \quad (25)$$

$$(Y_{ij}^{\Delta ES}|W_{ij}^{\Delta ES}, Z_i, \boldsymbol{\theta}_{yes}) \stackrel{iid}{\sim} N(\beta_{0,es} + \beta_{1,es}W_{ij}^{\Delta ES} + \boldsymbol{\gamma}_{es}Z_i, \sigma_{\epsilon^{\Delta ES}}^2) \quad (26)$$

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.

$$p(\beta_{0,ee}) \sim N(0, 100000) \quad (27) \quad p(\beta_{0,es}) \sim N(0, 100000) \quad (31)$$

$$p(\beta_{1,ee}) \sim N(1, 100000) \quad (28) \quad p(\beta_{1,es}) \sim N(1, 100000) \quad (32)$$

$$p(\gamma_{g,ee}) \sim N(0, 100000), \quad g = 1, 2, 3 \quad (29) \quad p(\gamma_{g,es}) \sim N(0, 100000), \quad g = 1, 2, 3 \quad (33)$$

$$p(\sigma_{\epsilon^{ee}}^2) \sim IG(0.1, 0.1) \quad (30) \quad p(\sigma_{\epsilon^{es}}^2) \sim IG(0.1, 0.1) \quad (34)$$

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

4.2 Model 1

Model 1 is a measurement error model that otherwise takes similar form to the naive 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:

$$(Y_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta}_{yee}) \sim N(\beta_{0,ee} + \beta_{1,ee}X_i^{EE} + \boldsymbol{\gamma}_{ee}Z_i, \sigma_{\epsilon_{EE}}^2) \quad (35)$$

$$(Y_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{yes}) \sim N(\beta_{0,es} + \beta_{1,es}X_i^{\Delta ES} + \boldsymbol{\gamma}_{es}Z_i, \sigma_{\epsilon_{\Delta ES}}^2) \quad (36)$$

$$(W_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta}_{wee}) \stackrel{iid}{\sim} N(X_i^{EE}, \sigma_{\nu_{EE}}^2) \quad (37)$$

$$(W_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{wes}) \stackrel{iid}{\sim} N(X_i^{\Delta ES}, \sigma_{\nu_{\Delta ES}}^2) \quad (38)$$

$$(X_i^{EE}, X_i^{\Delta ES}|\theta_X) \stackrel{iid}{\sim} N\left(\begin{bmatrix} \mu_{EE} \\ \mu_{\Delta ES} \end{bmatrix}, \Sigma_X\right) \quad (39)$$

We specify weak priors for reasons described before:

$$p(\beta_{0,ee}) \sim N(0, 100000) \quad (40) \quad p(\beta_{0,es}) \sim N(0, 100000) \quad (45)$$

$$p(\beta_{1,ee}) \sim N(1, 100000) \quad (41) \quad p(\beta_{1,es}) \sim N(1, 100000) \quad (46)$$

$$p(\gamma_{g,ee}) \sim N(0, 100000), g = 1, 2, 3 \quad (42) \quad p(\gamma_{g,es}) \sim N(0, 100000), g = 1, 2, 3 \quad (47)$$

$$p(\sigma_{\epsilon^{ee}}^2) \sim IG(0.1, 0.1) \quad (43) \quad p(\sigma_{\epsilon^{es}}^2) \sim IG(0.1, 0.1) \quad (48)$$

$$p(\sigma_{\nu^{ee}}^2) \sim IG(0.1, 0.1) \quad (44) \quad p(\sigma_{\nu^{es}}^2) \sim IG(0.1, 0.1) \quad (49)$$

$$p(\Sigma_X) \sim Inv-Wish(I_{2x2}, 3) \quad (50)$$

$$p(\mu_{EE}) \sim N(2400, 100000) \quad (51)$$

$$p(\mu_{\Delta ES}) \sim N(0, 100000) \quad (52)$$

The interpretation of $\beta_{1,\cdot}$ is similar to the naive model, except it is with respect to X^\cdot instead of a gold standard measurement. The interpretation of the γ coefficients is then the bias in EE and Δ ES from the cheap measurements. For example, all else equal, the difference in the cheap measurement of EE between a male and a female is $\gamma_{1,ee}$ kcals. The larger magnitude each γ is, the more bias that instrument has.

4.3 Model DP

Our model to analyze the measurement error and biases involved with various measurement tools for EE Δ ES involves both parametric and nonparametric 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 (2), 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 start by explaining how we will model the gold standard measurements. As shown in (5),(6), we assume measurements W_{ij}^{EE} and $W_{ij}^{\Delta ES}$ are unbiased to X_i^{EE} and $X_i^{\Delta ES}$. We continue with our independence assumptions described earlier. Based on the EBS data and the normal quantile plots generated from it in Figures (3),(4), we feel comfortable assuming the error distributions are normal. Further, we will assume constant variance which seems reasonable at least for DXA as seen in Figure (2); unfortunately we don't have any replicates for DLW at this time. The conditional specifications are then:

$$f(W_{ij}^{EE}|X_i^{EE}, Z_i, \sigma_{\nu^{EE}}^2) \stackrel{iid}{\sim} N(X_i^{EE}, \sigma_{\nu^{EE}}^2) \quad (53)$$

$$f(W_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \sigma_{\nu^{\Delta ES}}^2) \stackrel{iid}{\sim} N(X_i^{\Delta ES}, \sigma_{\nu^{\Delta ES}}^2) \quad (54)$$

$$(55)$$

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_j using free knot cubic B-splines, and model demographic covariates with a linear component. This has three benefits. First, the spline is flexible and can pick up an unknown relationship between X_j 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 important 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. Like in the gold standard case, we assume conditional on X , the cheap measurements are independent of the gold standard measurements and of the other cheap measurements. We are also assuming the cheap measurements are independent across individuals and within individual. We will also assume constant variance for both measurement errors even though (2) shows some signs of nonconstant variance. We do not have data with replicates of cheap Δ ES measures. The conditional specifications are then:

$$f(Y_{ij}^{EE}|X_i^{EE}, Z_i, \sigma_{\epsilon^{EE}}^2, \beta_{ee}) \stackrel{iid}{\sim} N(s(X_i^{EE}; \beta_{ee}) + \gamma_{ee}Z_i, \sigma_{\epsilon^{EE}}^2) \quad (56)$$

$$f(Y_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \sigma_{\epsilon^{\Delta ES}}^2, \beta_{\Delta es}) \stackrel{iid}{\sim} N(s(X_i^{\Delta ES}; \beta_{\Delta es}) + \gamma_{es}Z_i, \sigma_{\epsilon^{\Delta ES}}^2) \quad (57)$$

$$(58)$$

$$s(X_i^{EE}; \beta_{ee}) = \sum_{i=1}^{k_{ee}+3} b_{i,ee}(\mathbf{X}^{EE}) \beta_{i,ee} = B_{ee}(\mathbf{X}^{EE}) \beta_{ee} \quad (59)$$

$$s(X_i^{\Delta ES}; \beta_{\Delta es}) = \sum_{i=1}^{k_{es}+3} b_{i,es}(\mathbf{X}^{\Delta ES}) \beta_{i,es} = B_{es}(\mathbf{X}^{\Delta ES}) \beta_{es} \quad (60)$$

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 in [12] in Appendix A. Note that k_{ee} and k_{es} are random variables because we are using free knot splines.

The modeling approach described here is the same no matter which cheap measurement instruments are used. Because of the conditional independence assumptions, if we wanted to understand the measurement error and biases in multiple instruments, we would perform multiple analyses using

the different cheap measurements while using the same gold standard measurements (assuming they are all from the same study and individuals).

As a final note, we briefly justify our use of B-splines as the nonparametric 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. [13]

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. In any given problem, we can select H such that $\sum_{h=1}^H \pi_h < \epsilon$ for some $\epsilon > 0$. Then we can model the distribution for the latent variables as:

$$X_i^{EE}, X_i^{\Delta ES} \stackrel{iid}{\sim} \sum_{h=1}^H \pi_h N \left(\begin{bmatrix} \mu_{ee} \\ \mu_{es} \end{bmatrix}, \Sigma_x \right) \quad (61)$$

$$V_h \sim \text{Beta}(1, \alpha) \quad (62)$$

$$V_H = 1 \quad (63)$$

$$\pi_h = V_h \prod \ell < h(1 - V_\ell) \quad (64)$$

α helps control how many components are used. This value can be set, but we will put a prior on it to account for uncertainty. In [14] p.553, the authors note that the data can often be informative about this parameter. Before we can write a MCMC sampler we must choose priors for all the unknown parameters. 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.

$$p(\sigma_{\epsilon ee}^2) \sim IG(a_{yee}, b_{yee}) \quad (65)$$

$$p(\sigma_{\nu ee}^2) \sim IG(a_{yes}, b_{yes}) \quad (66)$$

$$p(\gamma_{g,ee}) \sim N(0, C_{\gamma_{ee}}), \quad g = 1, 2, 3 \quad (67)$$

$$p(k_{ee}) \sim Poi(\lambda_{ee}) \quad (68)$$

$$p(r_{ee}) \sim DisUnif(X_1^{EE}, \dots, X_n^{EE}) \quad (69)$$

$$p(\sigma_{\epsilon^{es}}^2) \sim IG(a_{wee}, b_{wee}) \quad (70)$$

$$p(\sigma_{\nu^{es}}^2) \sim IG(a_{wes}, b_{wes}) \quad (71)$$

$$p(\gamma_{g,es}) \sim N(0, C_{\gamma_{es}}), \quad g = 1, 2, 3 \quad (72)$$

$$p(k_{es}) \sim Poi(\lambda_{es}) \quad (73)$$

$$p(r_{es}) \sim DisUnif(X_1^{\Delta ES}, \dots, X_n^{\Delta ES}) \quad (74)$$

$$p(\Sigma_X) \sim Inv-Wish(\psi, d) \quad (75)$$

$$p(\mu_{EE}, \mu_{\Delta ES}) \sim N(M_\mu, C_\mu) \quad (76)$$

$$p(\alpha) \sim Gamma(a_\alpha, b_\alpha) \quad (77)$$

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 β_{ee} and β_{es} , this is because we will not be updating this in a fully Bayesian fashion. We will discuss this more in the RJMCMC section.

4.4 Issues with DP

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, but they are biased measurements of the latent variables unlike the gold standards. During our simulation study, we found that the DP 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 propose using a bivariate normal distribution for the latent variables.

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 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 models 0 and 1, we utilized JAGS to simulate draws from the posterior distribution. 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 $i = 1, \dots, n$ represent the number of individuals, and $j = 1, \dots, J$ 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.
3. For iteration k=1,...,K, sample from its full conditional:

$$(a) \ \{\zeta_i^{(k)} : i = 1, \dots, n\} \mid \cdot \text{ where } P(\zeta_i^{(k)} = h) = \frac{\pi_h^{(k-1)} N(\mu_h^{(k-1)}, \sigma_h^{2(k-1)})}{\sum_{h=1}^H \pi_h^{(k-1)} N(\mu_h^{(k-1)}, \sigma_h^{2(k-1)})}$$

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

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

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

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

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

$$(d) \ \alpha \mid \cdot \sim Gamma(a_\alpha + n_h + 1, b_\alpha - \sum_{r=1}^{n_h-1} \log(1 - V_r))$$

$$(e) \ \{\Sigma_{X,h}^{(k)}, h = 1, \dots, H\} \mid \cdot \stackrel{ind}{\sim} Inv-Wish(d + n_h, \psi + (\mathbf{X}_{i,h}^{(k-1)} - \mu_h^{(k-1)})'(\mathbf{X}_{i,h}^{(k-1)} - \mu_h^{(k-1)}))$$

$$\mathbf{X}_{i,h} \equiv \mathbf{X}_i I(\zeta_i^{(k)} = h)$$

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

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

$$(f) \ \{(\mu_{EE,h}^{(k)}, \mu_{\Delta ES,h}^{(k)}), h = 1, \dots, H\} \mid \cdot \stackrel{ind}{\sim} N(M'_\mu, C'_\mu)$$

$$C'_\mu = (C_\mu^{-1} + n_h \Sigma_{X,h}^{-1(k)})^{-1}$$

$$M'_\mu = C'_\mu (C_\mu^{-1} M + n_h \Sigma_{X,h}^{-1(k)} \bar{\mathbf{X}}^{(k-1)})$$

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

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

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

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

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

- (l) Update $\beta_{ee}, \beta_{es}, k_{ee}, k_{es}, \mathbf{r}_{ee}, \mathbf{r}_{es}$ using RJMCMC described in next section. Calculate $m_{ee}(\mathbf{X}^{\text{EE}(k)}; \beta_{ee}^{(k)})$ and $m_{es}(\mathbf{X}^{\Delta \text{ES}(k)}; \beta_{es}^{(k)})$

- (m) $\gamma_{ee}^{(k)}|\cdot \sim N(M'_{\gamma_{ee}}, C'_{\gamma_{ee}})$
 $\gamma_{es}^{(k)}|\cdot \sim N(M'_{\gamma_{es}}, C'_{\gamma_{es}})$

$$C'_{\gamma_{ee}} = \left(\frac{1}{\sigma_{\epsilon^{EE}}^{2(k)}} Z' Z + \frac{1}{C_{\gamma_{ee}}} I_{3 \times 3} \right)^{-1}$$

$$M'_{\gamma_{ee}} = C'_{\gamma_{ee}} \left(\frac{1}{\sigma_{\epsilon^{EE}}^{2(k)}} Z' (\bar{\mathbf{Y}}_i^{\text{EE}} - m_{ee}(\mathbf{X}^{\text{EE}(k)}; \beta_{ee}^{(k)})) + \frac{m_{\gamma_{ee}}}{C_{\gamma_{ee}}} \right)$$

$$C'_{\gamma_{es}} = \left(\frac{1}{\sigma_{\epsilon^{\Delta ES}}^{2(k)}} Z' Z + \frac{1}{C_{\gamma_{es}}} I_{3 \times 3} \right)^{-1}$$

$$M'_{\gamma_{es}} = C'_{\gamma_{es}} \left(\frac{1}{\sigma_{\epsilon^{\Delta ES}}^{2(k)}} Z' (\bar{\mathbf{Y}}_i^{\Delta \text{ES}} - m_{es}(\mathbf{X}^{\Delta \text{ES}(k)}; \beta_{es}^{(k)})) + \frac{m_{\gamma_{es}}}{C_{\gamma_{es}}} \right)$$

$$\bar{\mathbf{Y}}_i^{\text{EE}} = \frac{1}{J} \sum_{j=1}^J Y_{ij}^{EE}$$

$$\bar{\mathbf{Y}}_i^{\Delta \text{ES}} = \frac{1}{J} \sum_{j=1}^J Y_{ij}^{\Delta ES}$$

There are identity matrices because we are assuming independent priors for the linear regression components.

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 $\beta_{ee}, \beta_{es}, k_{ee}, k_{es}, \mathbf{r}_{ee}, \mathbf{r}_{es}$.

RJMCMC was introduced by Green [15] 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 [16],[17],[18]. We will, for the most part, take the approach of Denison et al. [19]. The approach of Denison et al. is a simple and quick implementation that performs comparably to the fully Bayesian approach of DiMatteo et al. [16] 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

$$\begin{aligned} y_i &= f(x_i; \boldsymbol{\beta}) + \epsilon_i \\ \epsilon_i &\stackrel{iid}{\sim} N(0, \sigma^2) \end{aligned} \tag{78}$$

where the $y_i, i = 1, \dots, n$ are the response variable and $x_i, i = 1, \dots, n$ 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

$$\begin{aligned} b_k &= c \times \min \left(1, \frac{p(k+1)}{p(k)} \right) \\ d_k &= c \times \min \left(1, \frac{p(k)}{p(k+1)} \right) \end{aligned} \tag{79}$$

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 $\mathcal{A} = \{x_i, i = 1, \dots, n : x_i \text{ is not currently a knot or within } \ell + 1 \text{ locations of a current knot}\}$. 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 $\{x_i, i = 1, \dots, n\}$ and the positions of the knots. Then use OLS to estimate the spline parameters β . 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 [19]. 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

$$\alpha = \min(1, \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}) \tag{80}$$

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. To check “convergence” of the chain, Denison et al. suggest monitoring the MSE of the models as defined below. Once the MSE has “settled down”, samples from the posterior should be reasonable.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i; \boldsymbol{\beta}))^2 \quad (81)$$

5.3 Our RJMCMC Implementation

Here we present how we implement our version of RJMCMC for free knot splines. The model (78) that was used to develop the method in [19] is simpler than our model in three respects: (i) our covariate value x_i is a latent variable, (ii) we have an additional linear component to our mean function, (iii) the regression model is part of a larger hierarchical model. The third issue doesn’t present much of a problem due to our conditional independence assumptions.

Issue (i) is not as simple. In the free knot spline setup in Denison et al., they have a regression model of the form (78) 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 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 $m.(X^{(k)}, \boldsymbol{\beta}^{(k)})$ 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 $X^{(k)}$, 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 $X^{(k)}$ 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} .

Our solution to issue (ii) is a simple fix. We fit the free knot spline using the original data Y^{EE} and $Y^{\Delta ES}$ regressed on the B-spline basis matrix generated using the current draw of latent variables X^{EE} and $X^{\Delta ES}$, respectively. Then we calculate $\tilde{Y}^{EE(k)} = Y^{EE} - m_{ee}(X^{EE(k)}, \beta_{ee}^{(k)})$ and $\tilde{Y}^{\Delta ES(k)} = Y^{\Delta ES} - m_{es}(X^{\Delta ES(k)}, \beta_{es}^{(k)})$ where the superscript (k) represents the k^{th} draw from the MCMC. We then update the $\gamma^{(k)}$ coefficients using $\tilde{Y}^{EE(k)}$ and $\tilde{Y}^{\Delta ES(k)}$ as our data.

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

1. Calculate b_k and d_k according to (79)
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 $r_{\cdot}^{(k-1)}$ 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 regression coefficients β^* by using OLS by regressing Y on $B^*(X^{(k)})$
6. Accept proposed knots and coefficients with probability α .

Otherwise set $r^{(k)} = r^{(k-1)}$, $\beta^{(k)} = \beta^{(k-1)}$ and calculate spline basis matrix $B'(X^{(k)})$ using $r^{(k)}$ in order to calculate mean function $m(\cdot)$

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

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

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

$$Z(k) = 2(\ell + 1) + k.(2\ell + 1)$$

$$k. = \text{length}(r^{(k-1)})$$

where the acceptance probabilities $\alpha_{rj.}$ are given in the notation of Denison et al.

6 Simulation Study

In this section we conduct a simulation study to check the performance of the discussed models and modeling assumptions. There are multiple goals for this simulation study. One, the true joint density of $(X_i^{EE}, X_i^{\Delta ES})$ must be specified when we model data (or we can assume error free measurements which we'll see is a bad idea), and since this term is never observed by anyone in any study, it is difficult to select a density for. We want to check the robustness of different choices of bivariate densities. Second, the regression functions m_{ee} and m_{es} (3),(4) are also unknown, again due to the fact we never observe $(X_i^{EE}, X_i^{\Delta ES})$, and therefore don't know the true relationship between X and the cheap measurements. Third, normal errors is a common assumption but can affect predictive performance, and we want to check the robustness to this assumption.

We will present posterior predictive checks to assess the model assumptions and fit, as well as performance measures such as predicted mean squared error (PMSE) for the regression function in question. For the measurement error models, we are interested in the posterior predictive distributions for Y, W, X because we are placing distributional assumptions on all of them. The forms for these distributions are given below. Other posterior summaries will be provided as well. We start with the naive model, then the linear measurement error model, and finally our model.

$$p(Y^*|W, Y, Z) = \int \int p(Y^*|\theta, X, Z)p(\theta, X|Y, W)d\theta dX \quad (82)$$

$$p(W^*|W, Y, Z) = \int \int p(W^*|\theta, X, Z)p(\theta, X|Y, W)d\theta dX \quad (83)$$

$$p(X^*|W, Y, Z) = \int \int p(X^*|\theta)p(\theta, X|Y, W)d\theta dX \quad (84)$$

6.1 Model 0

For both of the simulated data sets, we ran 3 chains of 10,000 iterations each with 2000 burn in. The chains for both of the models had Gelman-Rubin diagnostics of one's for all parameters and the trace plots have good mixing with little autocorrelation.

Tables 3,4,5 give posterior summaries for parameters for the naive model with normal, skewed and bimodal errors, respectively. Overall, results look similar. Under all 3 data sets, the naive model is able to estimate the linear, error free coefficients. Slopes for X^{EE} and $X^{\Delta ES}$ were different for each data set with the bimodal errors have the largest values. The bimodal errors also had lower values for the regression error variances.

Figures 12,13,14 gives quantiles from the posterior predictive distribution for Y for EE and ΔES for simulated data with normal, skewed, and bimodal errors respectively. We computed the posterior predictive distributions using the equation below. The vertical line in each plot indicates the truth in the simulated data. If the plotted distribution covers this truth it indicates the model captures that quantity of interest. If not, it indicates a shortcoming of the model as there is a discrepancy between the truth and the model. Figure 12 shows that all values of EE are being underestimated and not capturing the truth at all for the most part. Values for ΔES actually look okay, but for both EE and ΔES these distributions have very high variance. These issues could be coming

from a variety of places in the model including the mean functions for Y and the distributional assumptions. Since we are in a measurement error situation, we know that using the observed W as the truth is not a good idea, and we should implement a measurement error model. Figures 13,14 shows similar results.

The goal here was to show that assuming values are measured with no error when error is in fact present, can cause serious problems, in addition to poor performance in terms of predicted mean squared error (PMSE). We realize the issue of not having an acceptable mean structure could confound this conclusion, which is why we present a ME model with the same mean structure as this model next.

$$p(Y^*|Y, W) = \int p(Y^*|W, \theta)p(\theta|W, Y)d\theta \quad (85)$$

	2.5%	25%	50%	75%	97.5%	Truth
$\beta_{0,ee}$	277.71	462.09	555.58	650.60	830.17	
$\beta_{1,ee}$	0.74	0.78	0.79	0.81	0.85	
$\beta_{0,es}$	-307.64	-160.92	-90.00	-20.16	117.63	
$\beta_{1,es}$	3.38	3.60	3.71	3.82	4.03	
σ_{yee}	466.10	483.20	492.71	502.63	522.21	
σ_{yes}	327.54	339.66	346.24	353.04	366.78	
$\gamma_{1,ee}$	240.54	294.45	322.45	350.14	403.34	300
$\gamma_{2,ee}$	7.64	12.75	15.47	18.27	23.78	14
$\gamma_{3,ee}$	-14.23	-10.11	-7.86	-5.64	-1.51	-7
$\gamma_{1,es}$	-277.99	-241.98	-222.91	-204.06	-167.10	-200
$\gamma_{2,es}$	5.37	9.17	11.18	13.17	17.39	8
$\gamma_{3,es}$	-8.14	-4.99	-3.33	-1.76	1.23	-5

Table 3: Parameter estimates for Model 0 for normal errors with correct specification of within person errors

6.2 Model 1

Like in Model 0, we ran 3 chains of 10,000 iterations each with 2000 burn in and all parameters had Gelman-Rubin diagnostics of less than 1.02 and trace plots indicated good mixing.

Tables 6,7,8 give posterior summaries for parameters for the simple model under normal, skewed, and bimodal errors respectively. Regression coefficients look similar based on both types of errors.

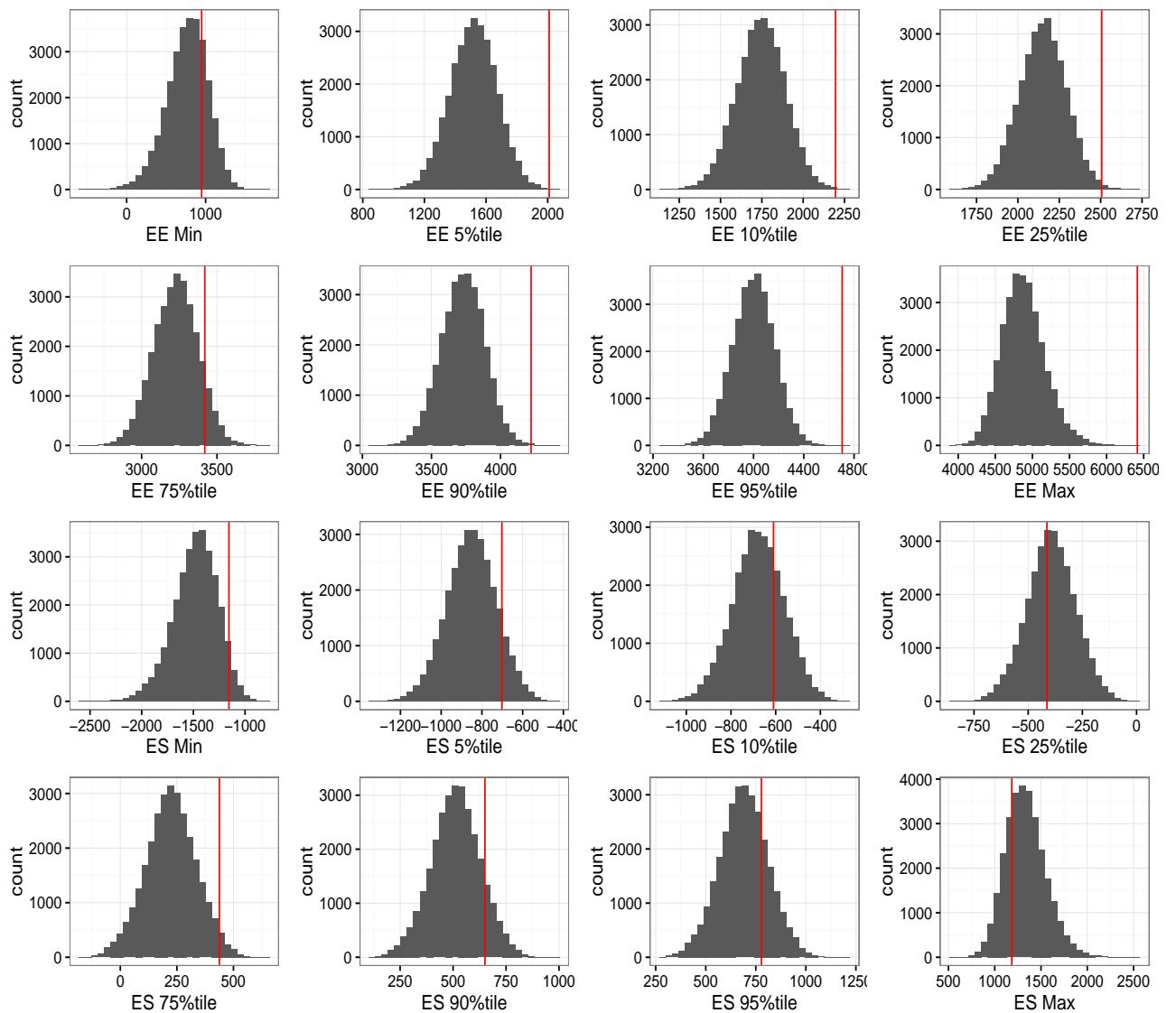


Figure 12: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model 0 with Normal Errors with correct specification of within person errors

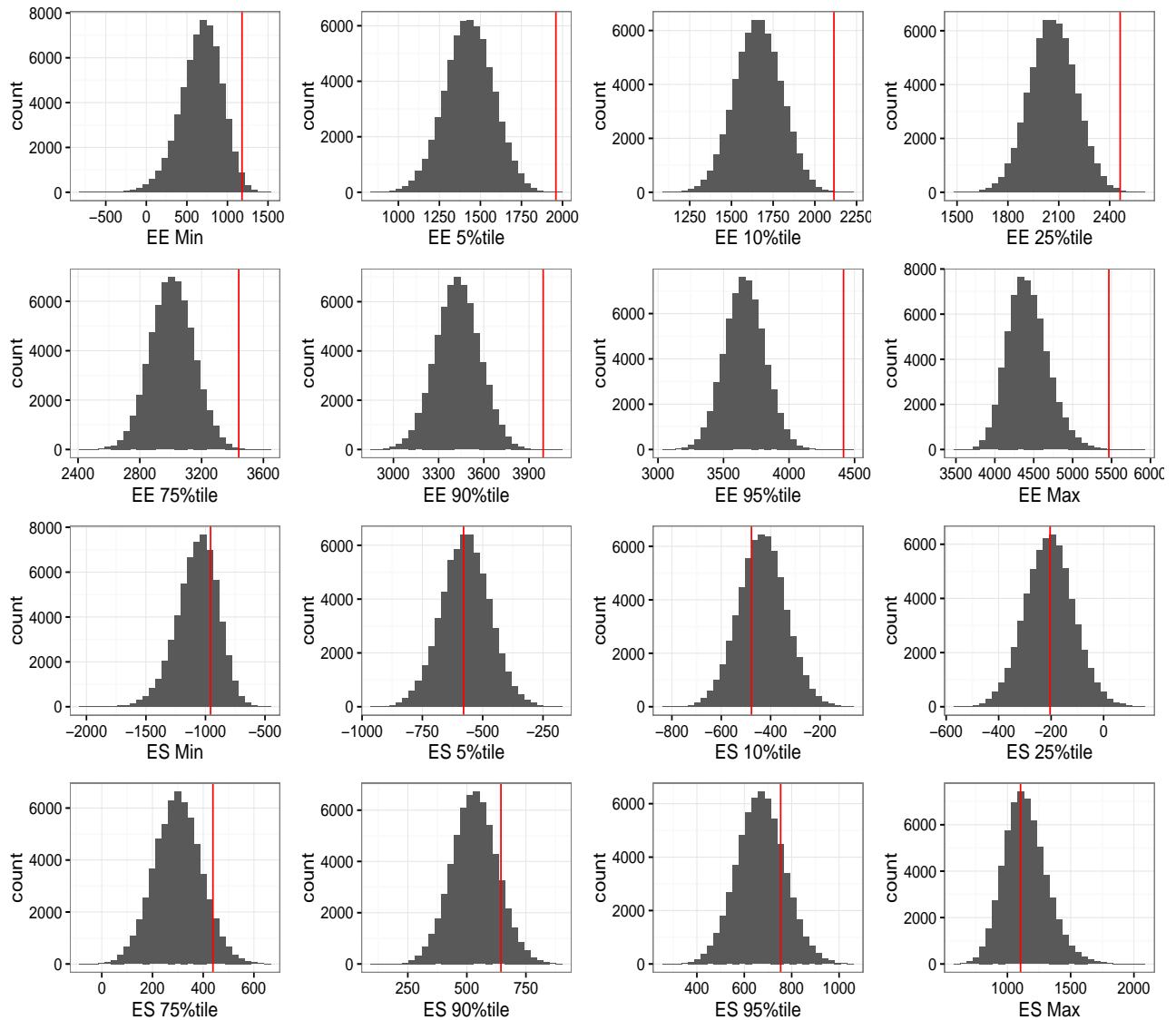


Figure 13: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model 0 with Skewed Normal Errors with correct specification of within person errors

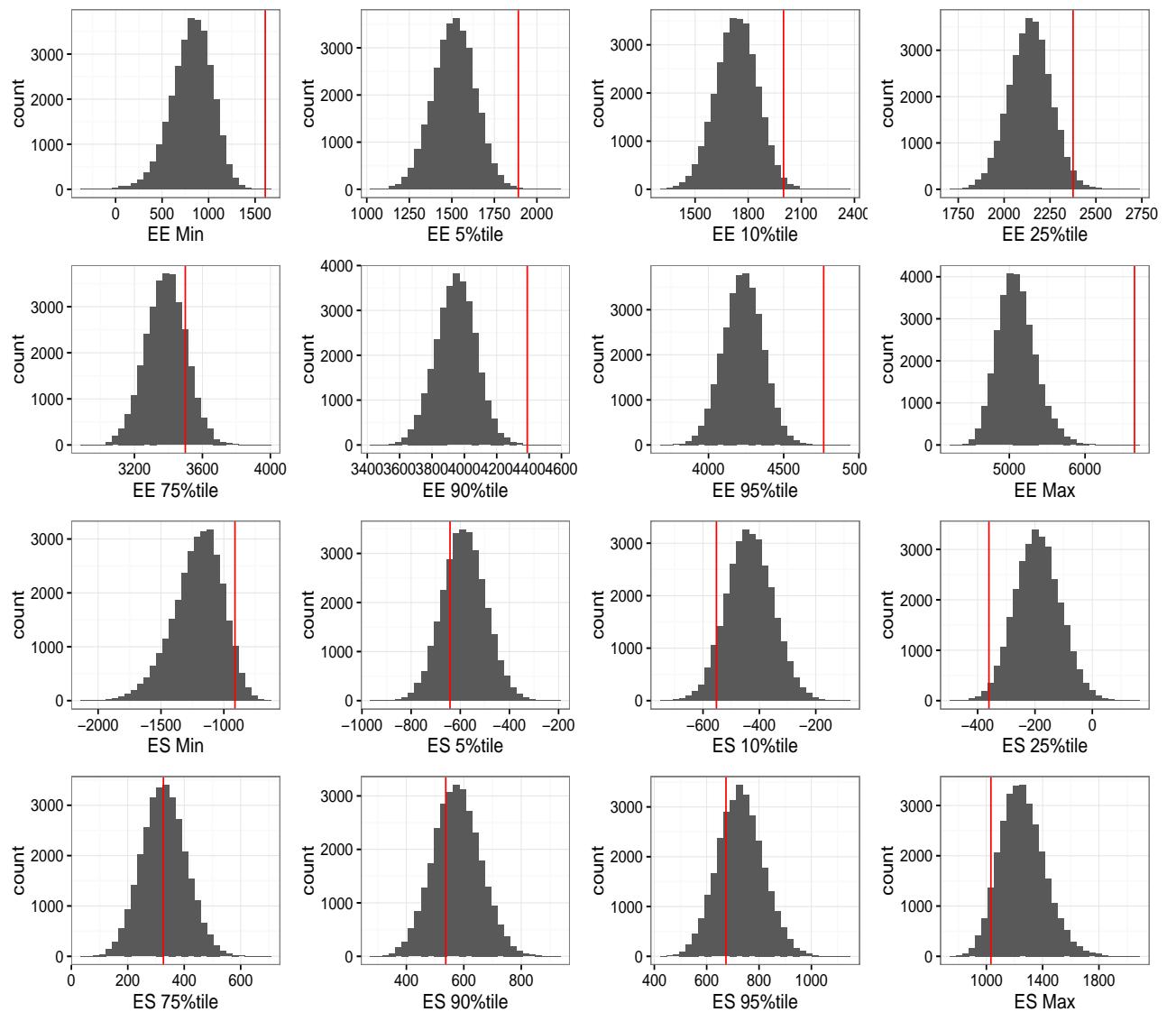


Figure 14: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model 0 with Bimodal Errors with correct specification of within person errors

	2.5%	25%	50%	75%	97.5%	Truth
$\beta_{0,ee}$	39.20	228.72	329.47	429.62	622.31	
$\beta_{1,ee}$	0.91	0.94	0.96	0.98	1.02	
$\beta_{0,es}$	-182.65	-46.24	20.78	85.59	210.74	
$\beta_{1,es}$	2.66	2.86	2.97	3.07	3.27	
σ_{yee}	455.71	472.67	481.86	491.63	511.03	
σ_{yes}	312.59	324.02	330.35	336.95	350.18	
$\gamma_{1,ee}$	188.54	241.12	268.70	296.06	349.57	300
$\gamma_{2,ee}$	2.31	7.72	10.40	13.25	18.55	14
$\gamma_{3,ee}$	-16.59	-12.25	-10.05	-7.91	-3.56	-7
$\gamma_{1,es}$	-286.35	-251.71	-233.56	-215.41	-180.68	-200
$\gamma_{2,es}$	1.96	5.69	7.53	9.32	13.10	8
$\gamma_{3,es}$	-7.56	-4.67	-3.17	-1.67	1.27	-5

Table 4: Parameter estimates for Model 0 for skewed errors with correct specification of within person errors

	2.5%	25%	50%	75%	97.5%	Truth
$\beta_{0,ee}$	-214.68	-60.93	15.11	93.78	236.57	
$\beta_{1,ee}$	0.99	1.02	1.03	1.05	1.07	
$\beta_{0,es}$	-57.37	31.59	81.77	131.86	232.59	
$\beta_{1,es}$	3.22	3.36	3.44	3.51	3.65	
σ_{yee}	339.11	351.89	358.93	366.09	380.58	
σ_{yes}	231.48	239.98	244.66	249.57	259.42	
$\gamma_{1,ee}$	205.14	244.14	264.19	284.56	323.20	300
$\gamma_{2,ee}$	9.12	13.16	15.22	17.28	21.15	14
$\gamma_{3,ee}$	-14.28	-11.31	-9.66	-8.08	-5.03	-7
$\gamma_{1,es}$	-226.15	-200.38	-186.83	-173.27	-147.33	-200
$\gamma_{2,es}$	0.78	3.58	5.02	6.38	8.94	8
$\gamma_{3,es}$	-7.18	-5.05	-3.97	-2.86	-0.66	-5

Table 5: Parameter estimates for Model 0 for bimodal errors with correct specification of within person errors

The only time the truth isn't covered by the 95% credible interval is for gender in Δ ES under the skewed errors.

Figures 15 and 18 give posterior predictive checks for the latent variables X^{EE} and $X^{\Delta ES}$. Both sets of plots look okay except for the minimum and maximum values. This is somewhat surprising considering this model uses only a single bivariate normal distribution to model the latent variables. One issue, however, is the amount of variance in these plots. For a given percentile the posterior predictive distribution gives a wide range of values and for the minimum values of EI and EE, it gives values close to -1000, which makes no sense in the context of our problem.

Figures 16 and 19 give posterior predictive checks for the gold standard measurements W for normal and skewed errors respectively. Figure 16 shows no signs of issues which is not surprising since this

model assumes normal measurement errors. Figure 19 also looks good except there are some extreme outliers in Δ ES measures.

Figures 17 and 20 give posterior predictive checks for the gold standard measurements W for normal and skewed errors respectively. Figure 17 shows no major signs of problems, which like before, is expected because the truth is normal errors. The variance for the percentiles is still large though. Figure 20 shows in the small percentiles of EE as well as the 75th and maximum. For Δ ES, only the 5th and 90th percentiles look okay. One culprit for this is misspecification of the measurement errors, but this could be compounded by assuming a linear structure and/or the normality assumption on the latent variables.

PMSE for this simulated data and these fitted models is given in Table 12.

	2.5%	25%	50%	75%	97.5%	Truth
$\beta_{0,ee}$	83.43	286.31	392.49	497.99	719.08	
$\beta_{1,ee}$	0.80	0.84	0.86	0.88	0.92	
$\beta_{0,es}$	-320.83	-158.99	-78.38	5.59	178.58	
$\beta_{1,es}$	4.14	4.47	4.66	4.85	5.23	
μ_x^{EE}	2617.25	2670.55	2698.99	2727.88	2781.14	
$\mu_x^{\Delta ES}$	-7.77	-1.29	2.16	5.63	12.07	
σ_{yee}	431.23	449.77	459.89	469.98	490.46	
σ_{yes}	303.30	316.12	323.31	330.85	345.57	
σ_{wee}	239.31	252.22	259.48	267.14	282.67	
σ_{wes}	65.14	68.04	69.60	71.22	74.43	
σ_{xee}	652.14	689.61	710.28	731.30	775.42	
σ_{xes}	65.97	71.11	73.95	76.92	82.93	
ρ	-0.24	-0.15	-0.11	-0.06	0.02	
$\gamma_{1,ee}$	231.70	288.52	317.27	345.91	401.34	300
$\gamma_{2,ee}$	5.89	11.80	14.86	17.65	22.50	14
$\gamma_{3,ee}$	-14.59	-9.96	-7.65	-5.30	-1.35	-7
$\gamma_{1,es}$	-289.47	-245.90	-222.68	-199.28	-153.92	-200
$\gamma_{2,es}$	3.59	8.33	10.79	13.19	17.77	8
$\gamma_{3,es}$	-9.11	-5.50	-3.49	-1.61	2.09	-5

Table 6: Parameter estimates for Model 1 for normal errors with correct specification of within person errors

6.3 Model BVN

Table 12 gives PMSE for both the EE and Δ ES regression functions under normal, skewed, and bimodal errors. Not only does ignoring the measurement error give bad parameter estimates and posterior predictive checks, but also its prediction is poor. Our model that has splines in its

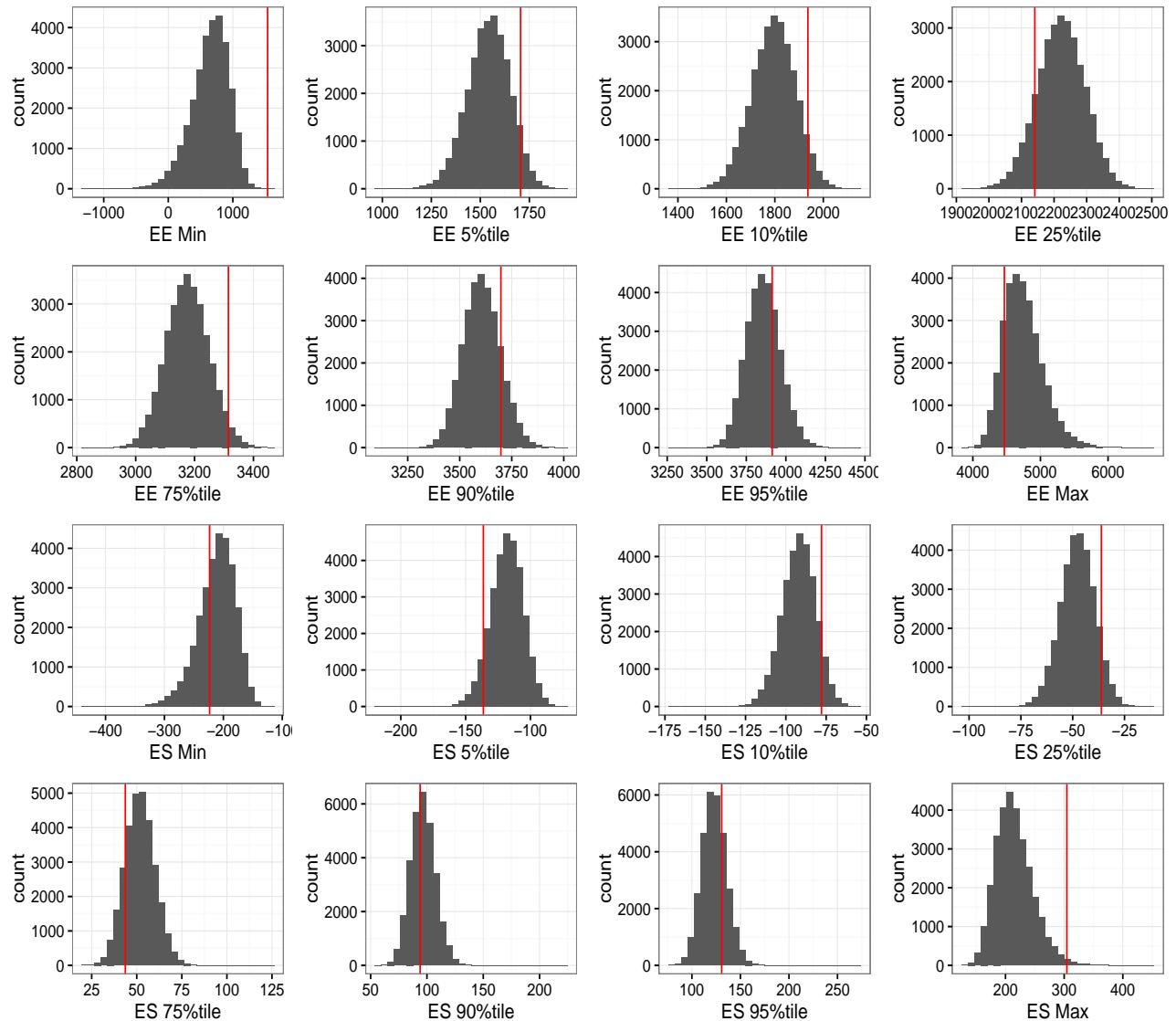


Figure 15: Posterior Predictive Discrepancy Measures For X^{EE} and $X^{\Delta ES}$ for Model 1 with Normal Errors

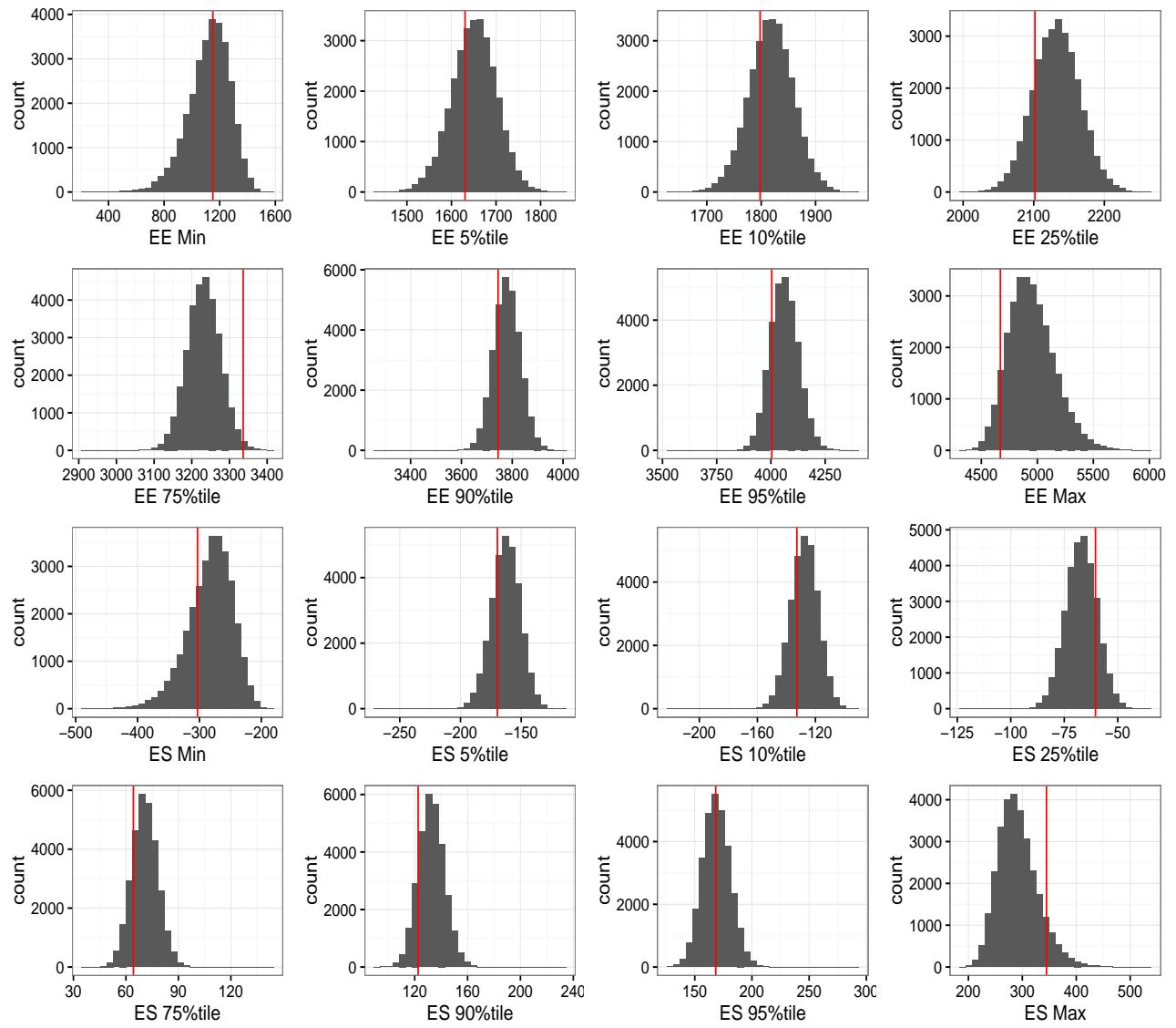


Figure 16: Posterior Predictive Discrepancy Measures For W^{EE} and $W^{\Delta ES}$ for Model 1 with Normal Errors

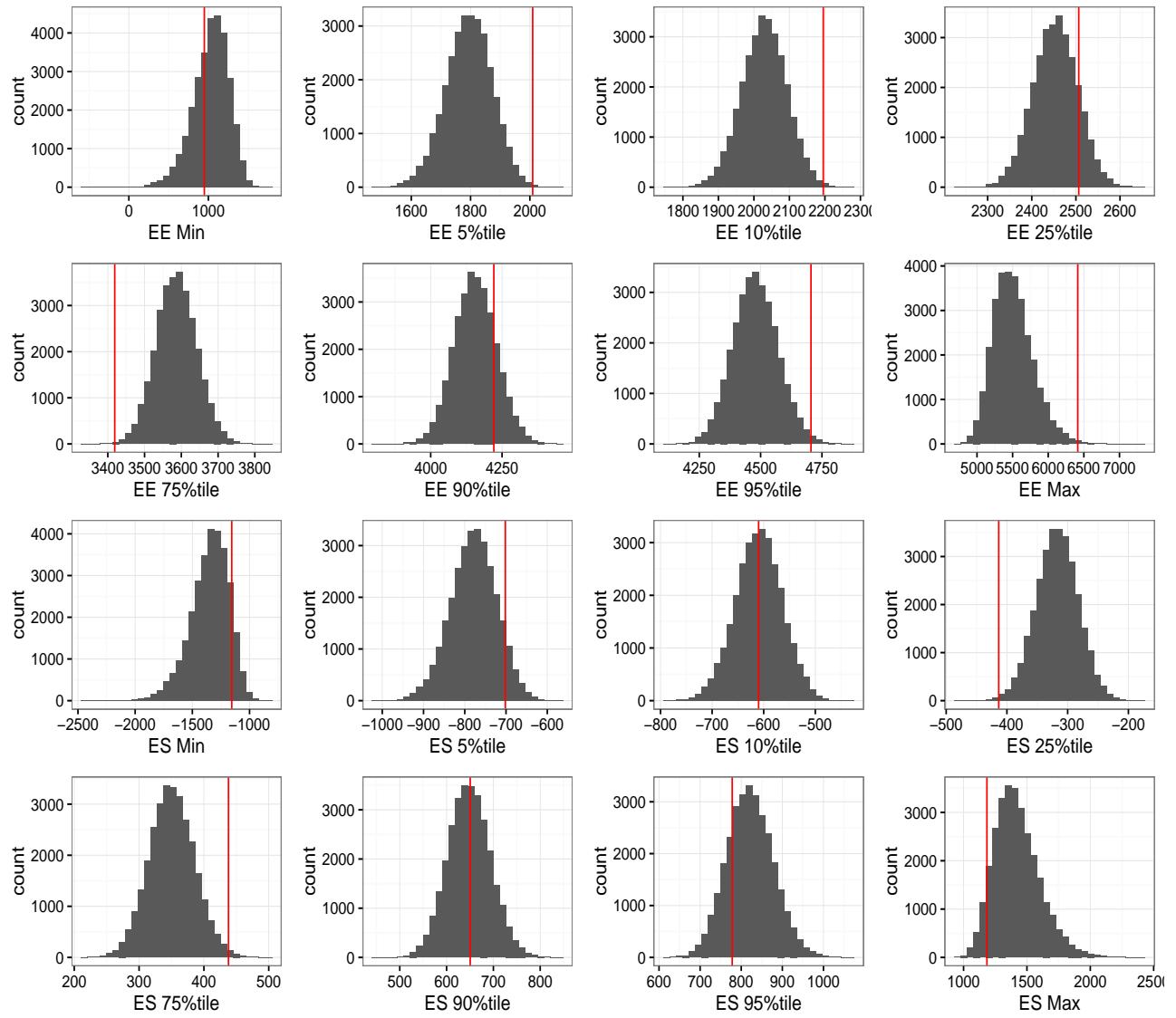


Figure 17: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model 1 with Normal Errors

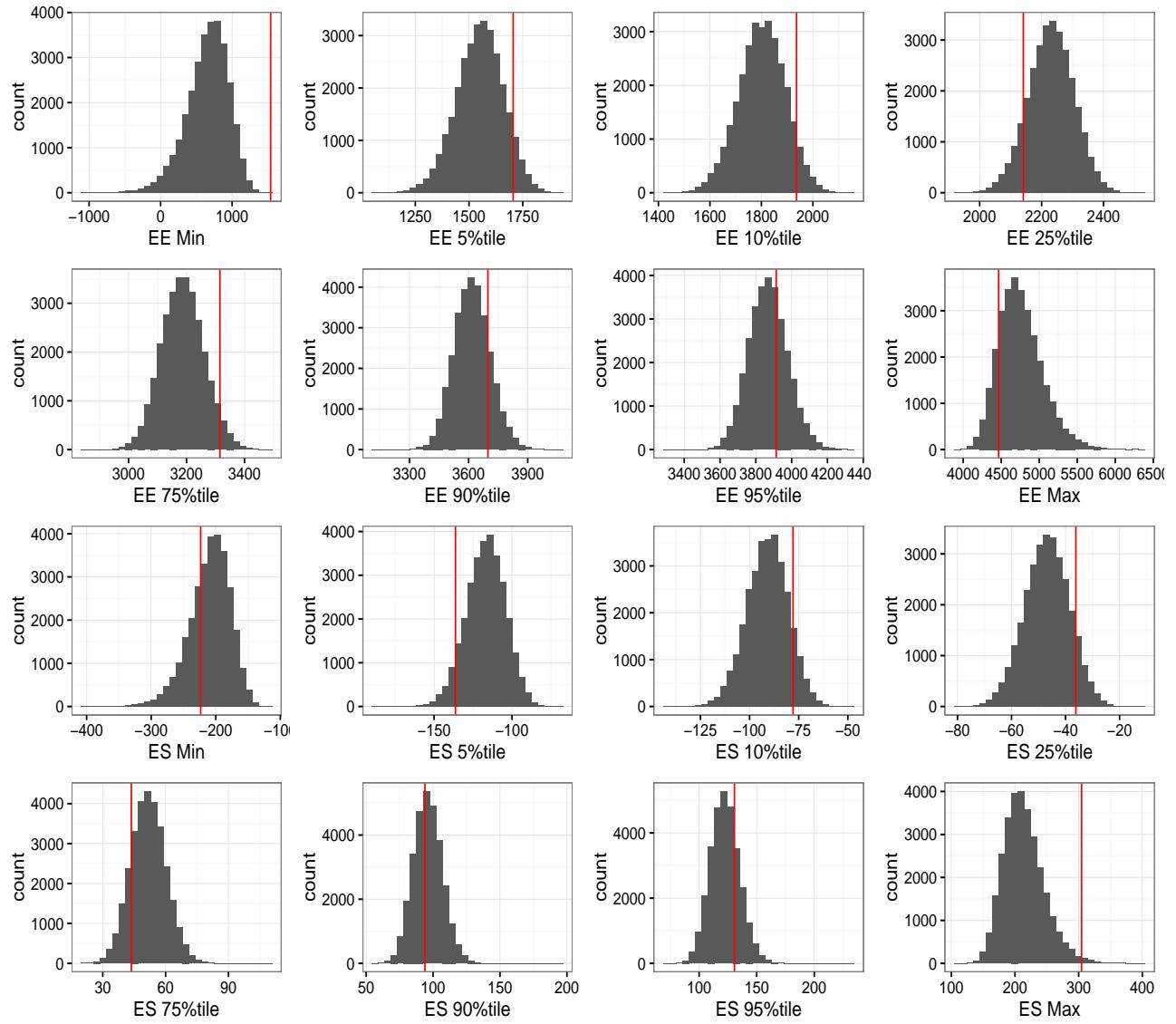


Figure 18: Posterior Predictive Discrepancy Measures For X^{EE} and $X^{\Delta ES}$ for Model 1 with Skewed Normal Errors

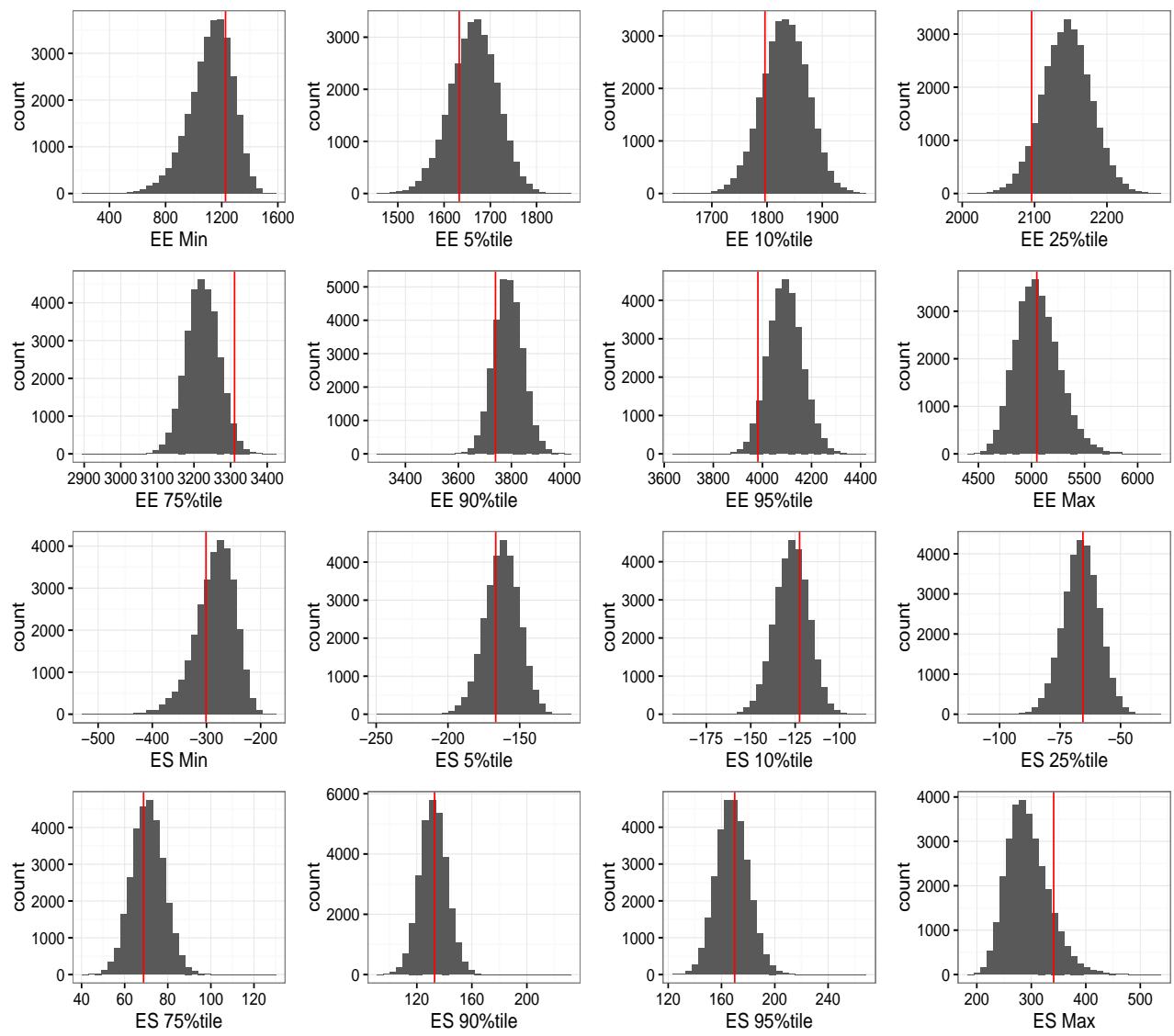


Figure 19: Posterior Predictive Discrepancy Measures For W^{EE} and $W^{\Delta ES}$ for Model 1 with Skewed Normal Errors

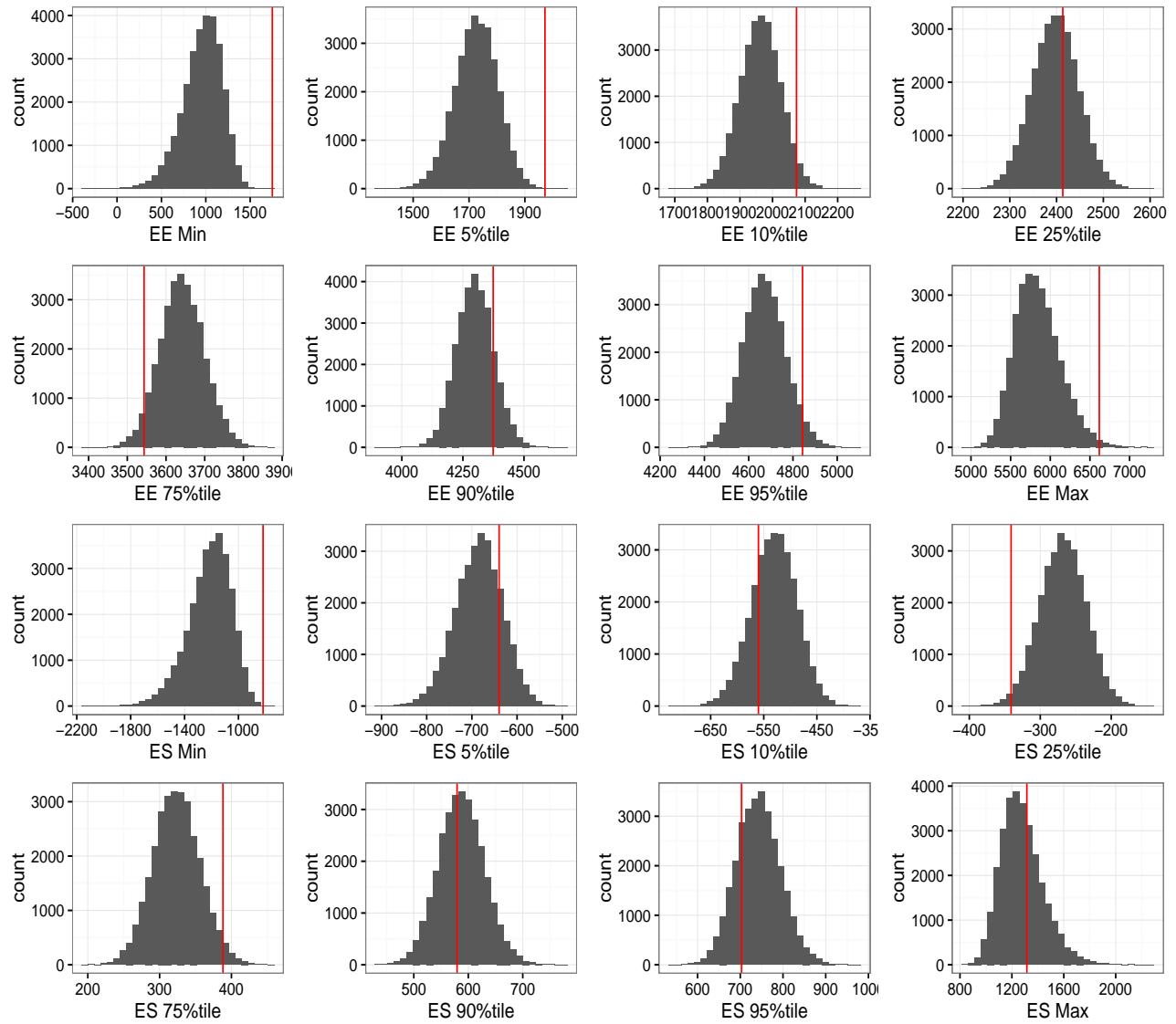


Figure 20: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model 1 with Skewed Normal Errors

	2.5%	25%	50%	75%	97.5%	Truth
$\beta_{0,ee}$	-136.25	64.01	164.19	272.76	478.66	
$\beta_{1,ee}$	0.98	1.02	1.04	1.06	1.10	
$\beta_{0,es}$	-223.94	-63.11	9.64	87.16	227.35	
$\beta_{1,es}$	3.32	3.62	3.79	3.97	4.33	
μ_x^{EE}	2626.86	2680.08	2708.93	2737.61	2792.61	
$\mu_x^{\Delta ES}$	-7.53	-0.91	2.51	5.97	12.54	
σ_{yee}	404.01	422.30	432.51	442.74	463.17	
σ_{yes}	294.31	306.63	313.48	320.62	334.84	
σ_{wee}	247.61	260.74	268.09	275.86	291.15	
σ_{wes}	65.28	68.23	69.86	71.55	74.98	
σ_{xee}	653.95	690.93	711.49	733.06	776.22	
σ_{xes}	65.12	70.54	73.42	76.36	82.30	
ρ	-0.28	-0.20	-0.15	-0.11	-0.02	
$\gamma_{1,ee}$	181.87	236.10	265.11	293.78	348.64	300
$\gamma_{2,ee}$	0.52	6.32	9.30	12.20	17.84	14
$\gamma_{3,ee}$	-16.95	-12.98	-10.62	-8.17	-3.49	-7
$\gamma_{1,es}$	-290.13	-249.26	-228.14	-206.89	-165.11	-200
$\gamma_{2,es}$	1.69	5.58	7.74	9.95	13.94	8
$\gamma_{3,es}$	-8.47	-5.00	-3.20	-1.38	1.88	-5

Table 7: Parameter estimates for Model 1 for skewed errors with correct specification of within person errors

regression function whose knots can vary iteration to iteration performs much better than just a linear model.

7 Calibration

One of the main goals of this research was to provide the best possible estimates of EE and Δ ES 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 Δ ES and return a more accurate estimate along with measures of uncertainty.

Once we have fit our model to a specific set of cheap measurement tools for EE and Δ ES, we can use our model to calibrate measurements from those cheap measurements in the future. That is, given an estimate of EE or Δ ES 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

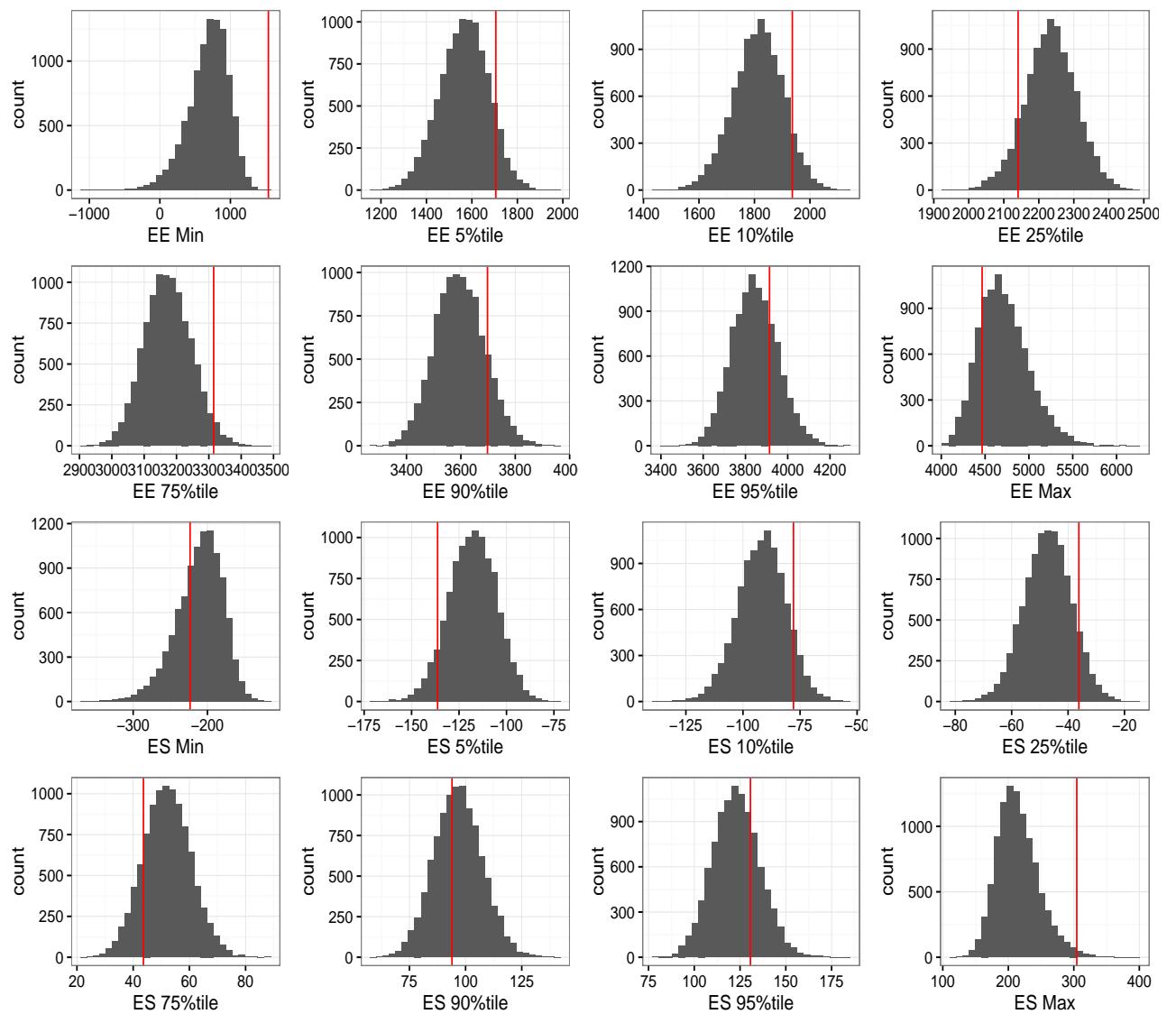


Figure 21: Posterior Predictive Discrepancy Measures For X^{EE} and $X^{\Delta ES}$ for Model BVN with Normal Errors

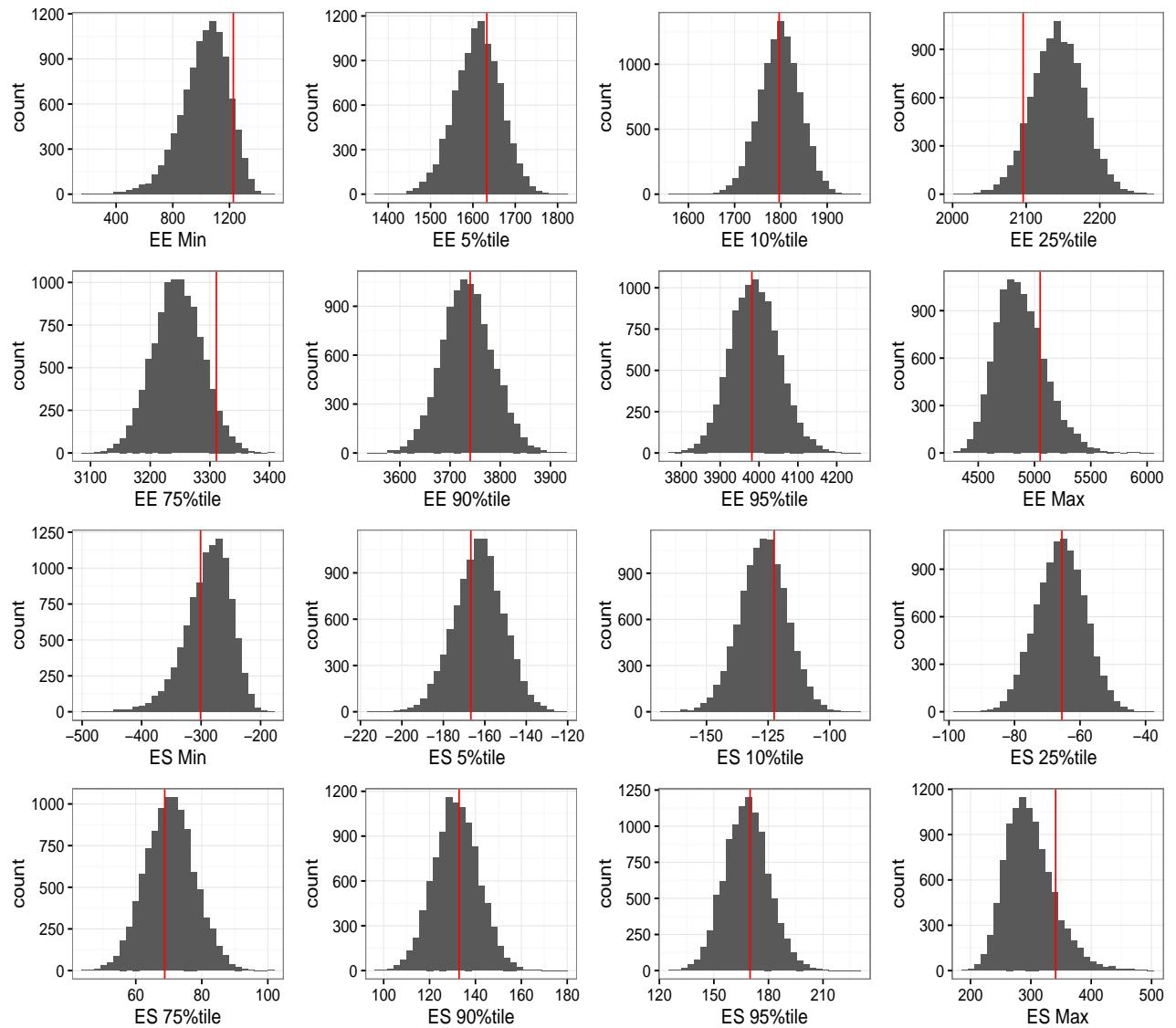


Figure 22: Posterior Predictive Discrepancy Measures For W^{EE} and $W^{\Delta ES}$ for Model BVN with Normal Errors

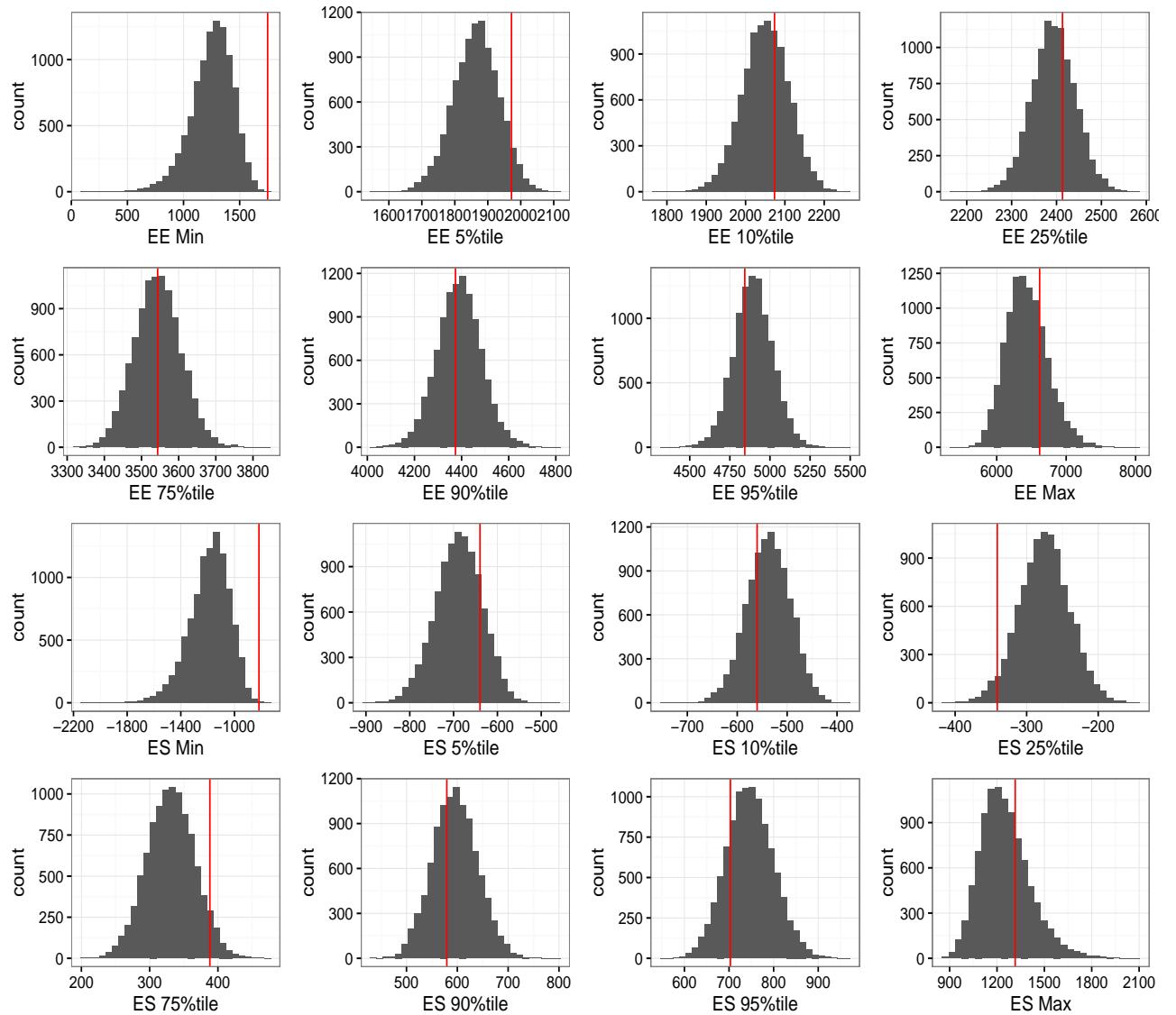


Figure 23: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model BVN with Normal Errors

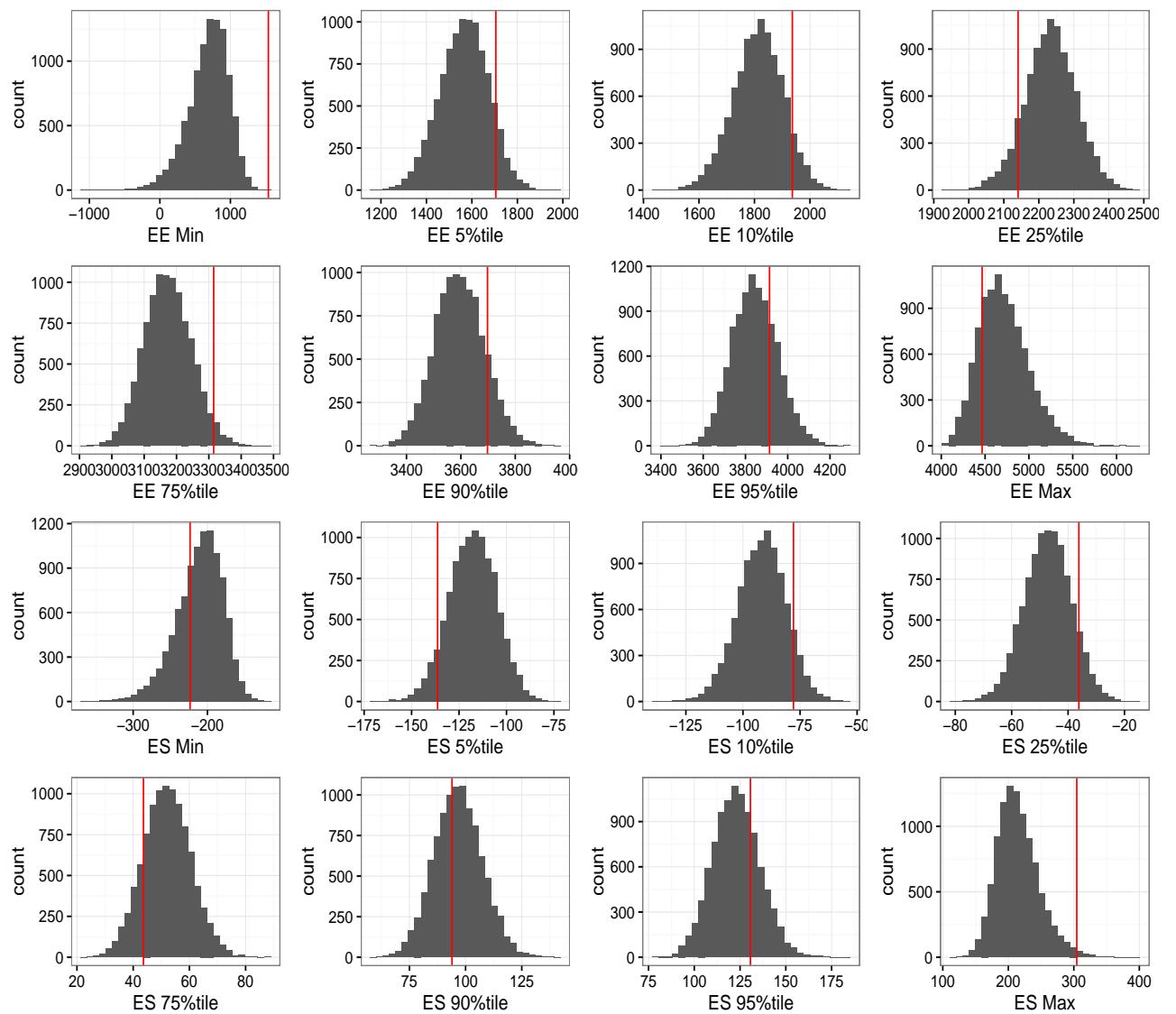


Figure 24: Posterior Predictive Discrepancy Measures For X^{EE} and $X^{\Delta ES}$ for Model BVN with Skewed Normal Errors

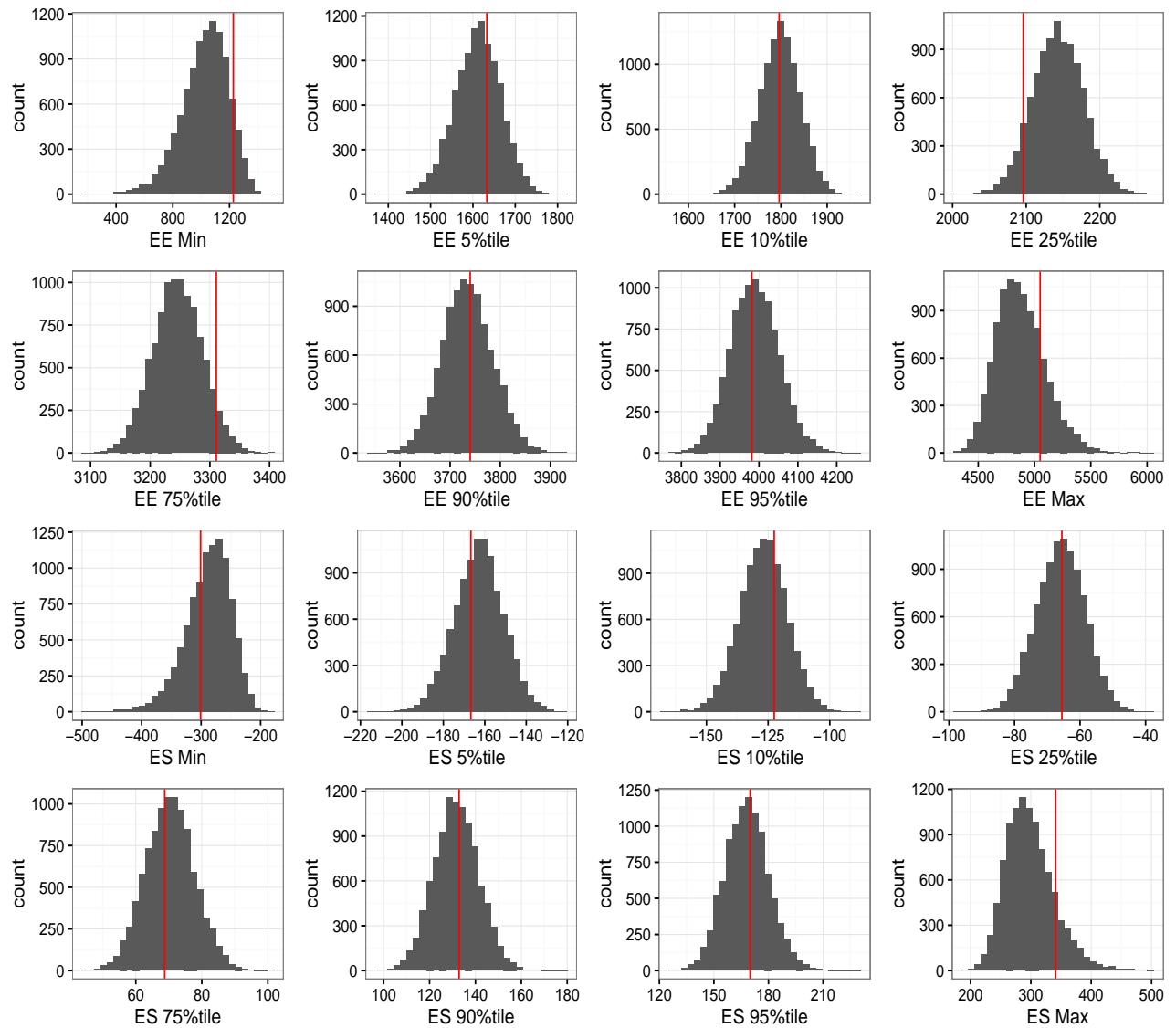


Figure 25: Posterior Predictive Discrepancy Measures For W^{EE} and $W^{\Delta ES}$ for Model BVN with Skewed Normal Errors

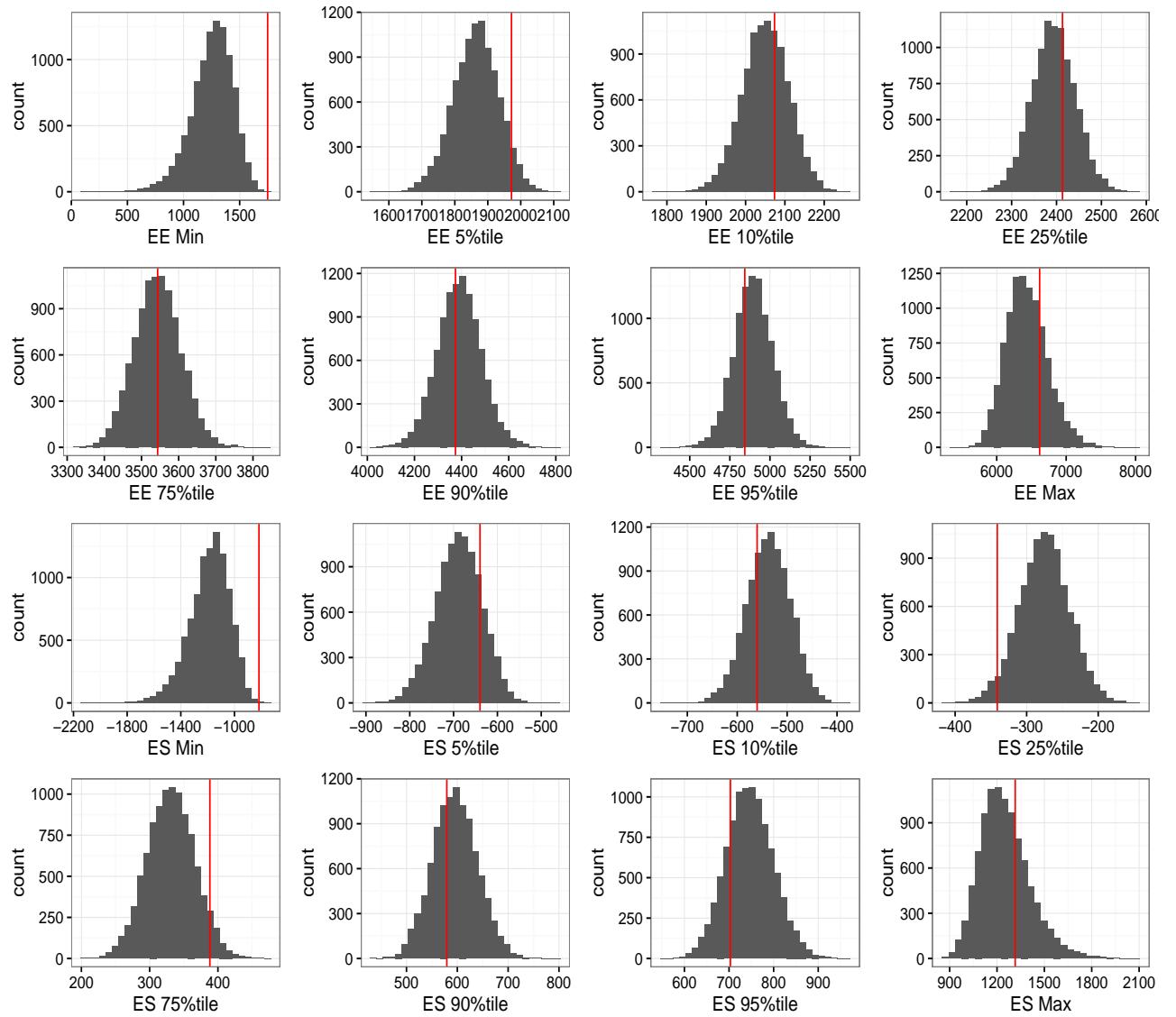


Figure 26: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model BVN with Skewed Normal Errors

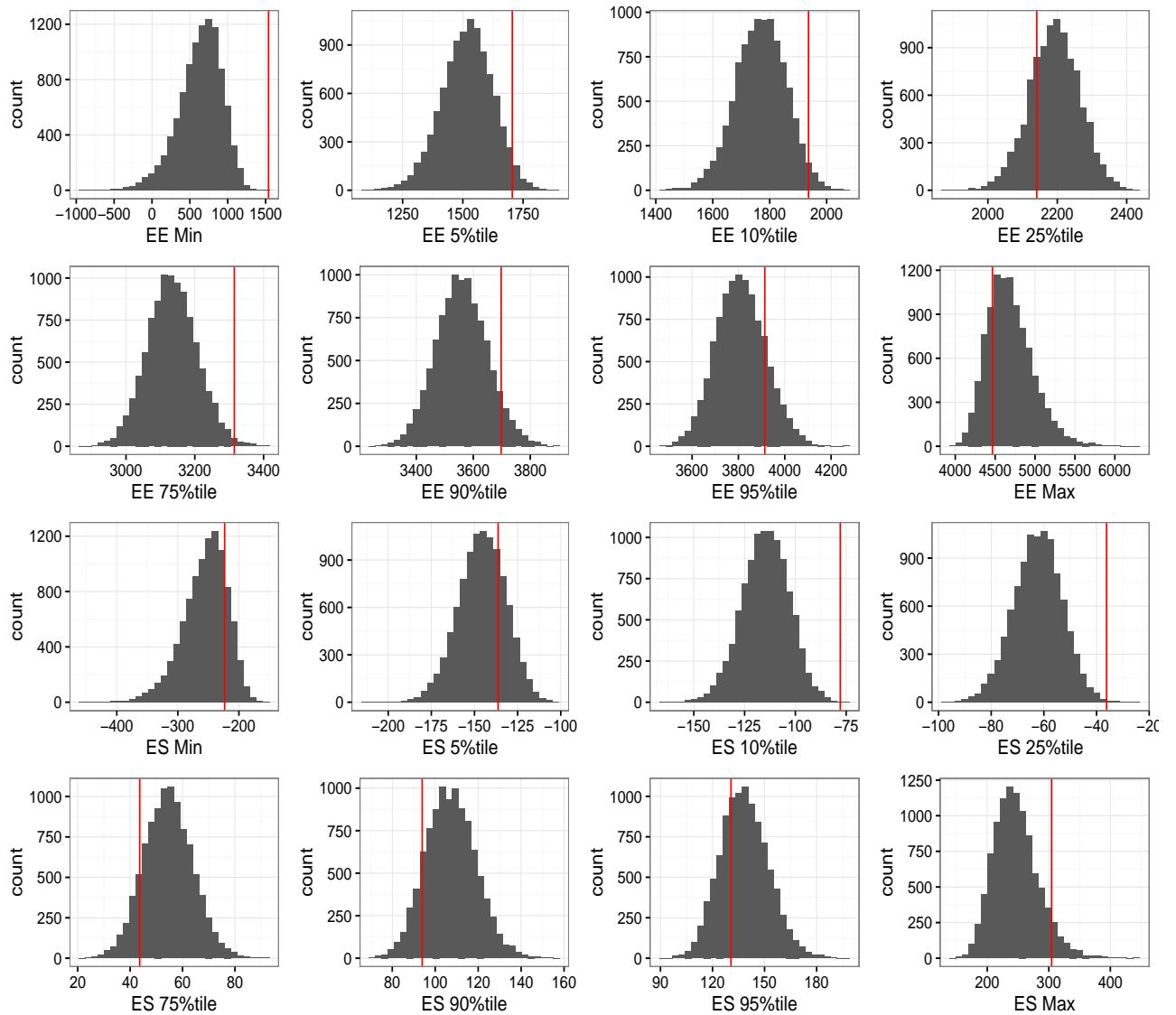


Figure 27: Posterior Predictive Discrepancy Measures For X^{EE} and $X^{\Delta ES}$ for Model BVN with Bimodal Normal Errors

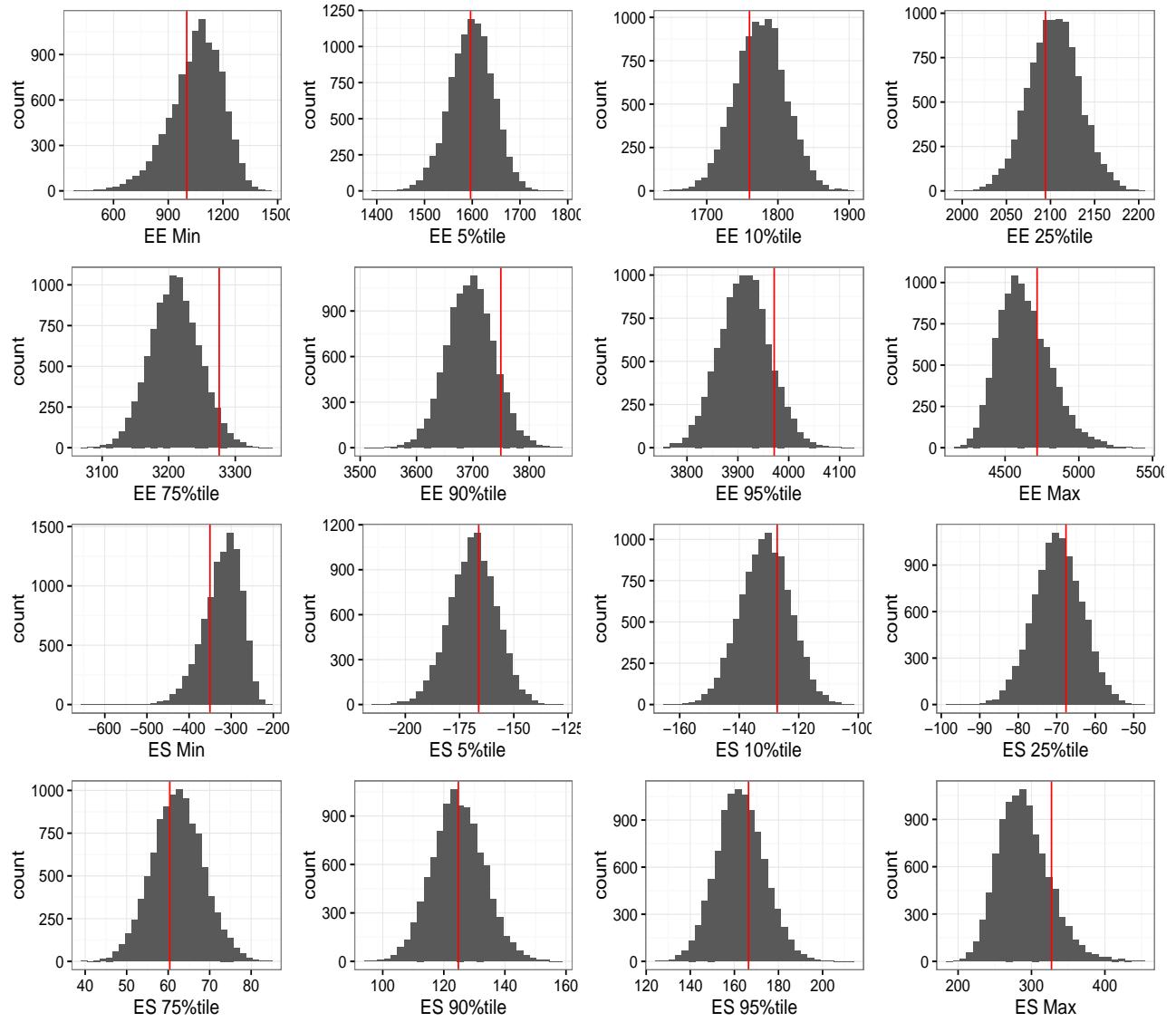


Figure 28: Posterior Predictive Discrepancy Measures For W^{EE} and $W^{\Delta ES}$ for Model BVN with Bimodal Normal Errors

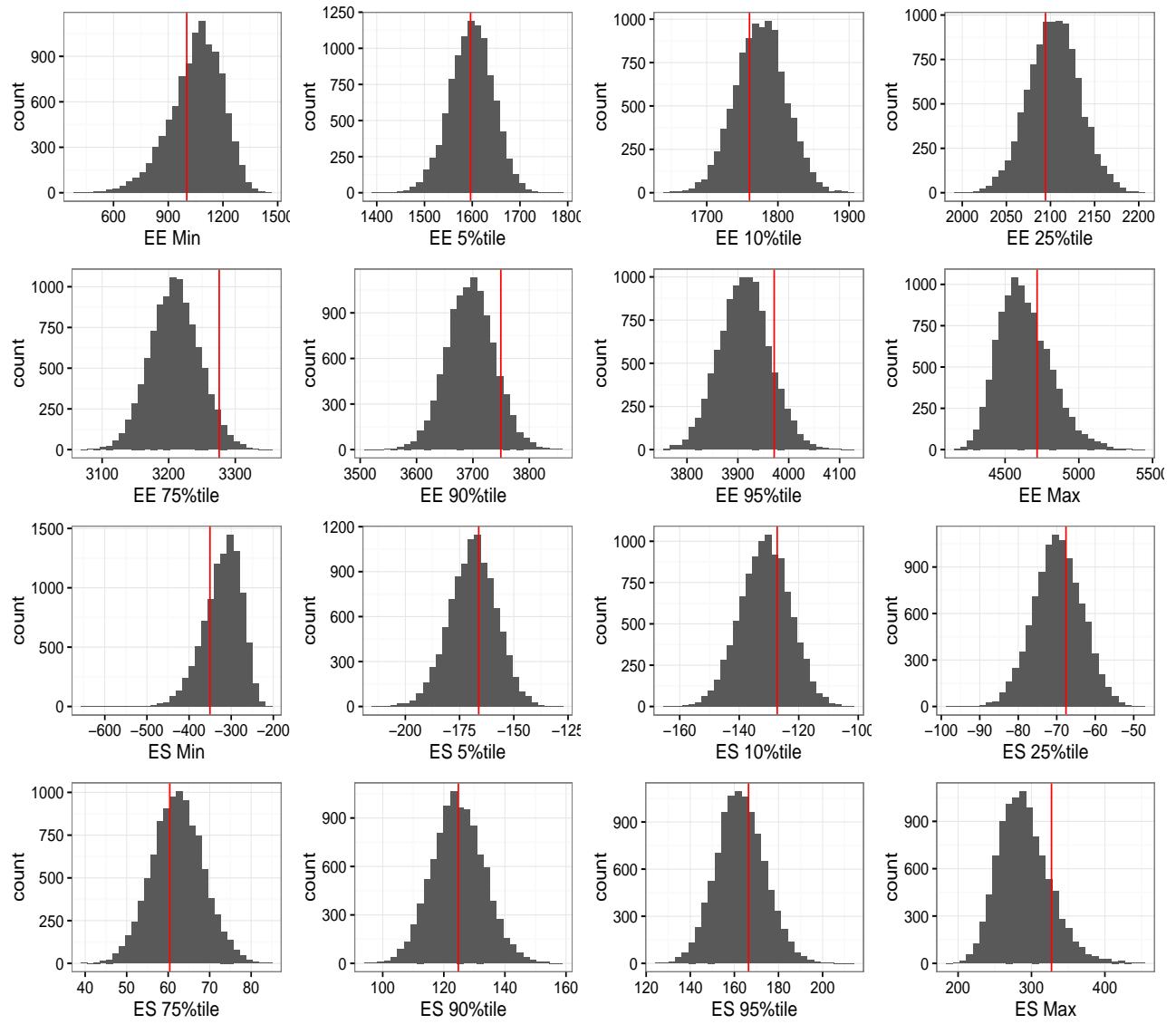


Figure 29: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model BVN with Bimodal Normal Errors

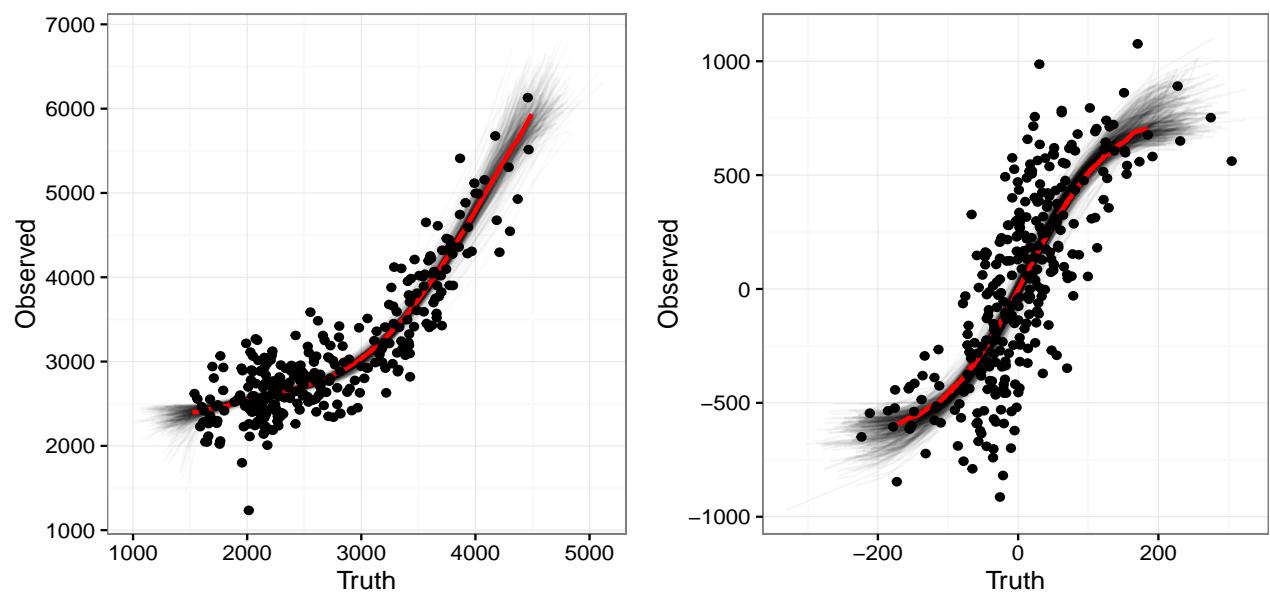


Figure 30: Spline function for Model BVN with Normal Errors

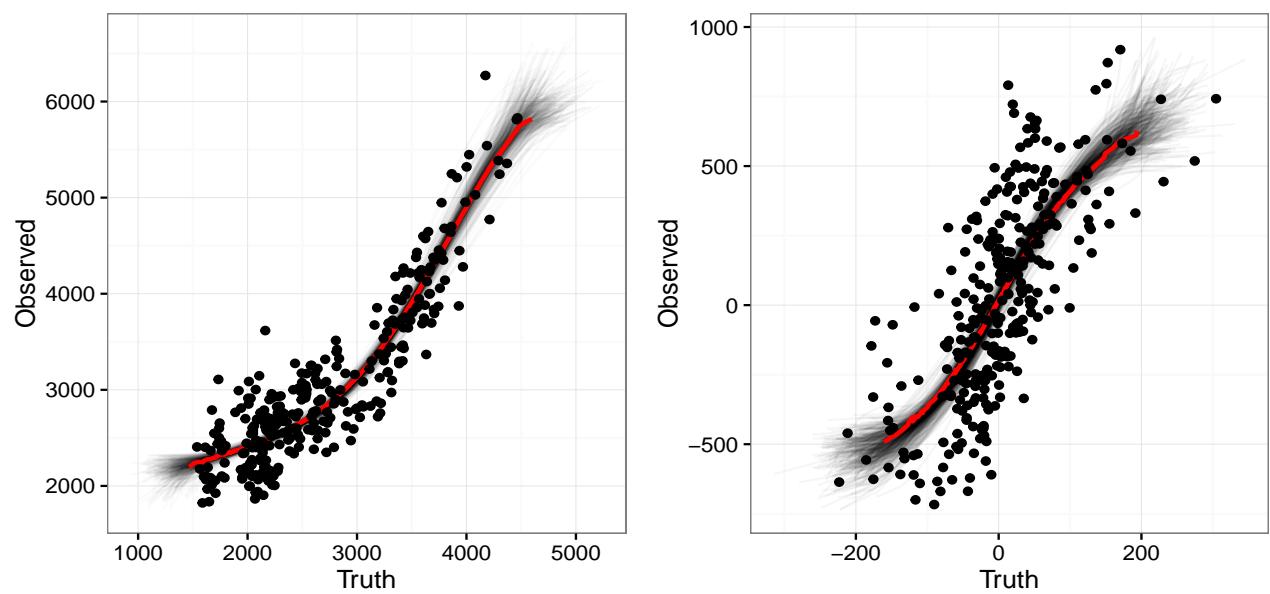


Figure 31: Spline function for Model BVN with Skewed Errors

	2.5%	25%	50%	75%	97.5%	Truth
$\beta_{0,ee}$	-341.79	-200.87	-116.19	-39.05	136.63	
$\beta_{1,ee}$	1.05	1.08	1.10	1.11	1.14	
$\beta_{0,es}$	-110.42	14.56	83.97	146.56	256.27	
$\beta_{1,es}$	3.44	3.63	3.74	3.84	4.05	
μ_x^{EE}	2586.74	2641.17	2669.76	2698.30	2753.33	
$\mu_x^{\Delta ES}$	-14.49	-7.48	-3.78	-0.15	6.79	
σ_{yee}	261.87	278.59	287.30	296.20	314.10	
σ_{yes}	225.11	234.57	239.61	244.85	255.51	
σ_{wee}	184.23	196.59	203.62	210.91	225.74	
σ_{wes}	49.52	51.64	52.84	54.11	56.59	
σ_{xee}	668.24	704.13	724.60	745.56	788.54	
σ_{xes}	78.68	83.62	86.39	89.34	95.14	
ρ	-0.06	0.02	0.06	0.11	0.18	
$\gamma_{1,ee}$	194.85	233.93	254.71	275.39	315.39	300
$\gamma_{2,ee}$	7.67	12.10	14.25	16.33	19.90	14
$\gamma_{3,ee}$	-14.85	-11.73	-10.11	-8.38	-4.85	-7
$\gamma_{1,es}$	-236.86	-205.53	-188.75	-171.75	-139.94	-200
$\gamma_{2,es}$	0.30	3.10	4.84	6.56	10.24	8
$\gamma_{3,es}$	-7.81	-5.19	-3.74	-2.24	0.48	-5

Table 8: Parameter estimates for Model 1 for bimodal errors with correct specification of within person errors

in (3), (4) as a function of Y instead of X , and Z are still measurement error free covariates as shown below.

$$X_{calibrated} = m^{-1}(y - \gamma Z) \quad (86)$$

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 value for which we want a calibrated value for. We input a specified number of posterior draws of regression coefficients and latent variables to then get a set of predicted values for X . 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.

As an example, we want to calibrate three cheap measurements each from a different individual. Let individual 1 be female, BMI of 18, age 25; let individual 2 be male, BMI of 32, age 20 and individual 3 be male, BMI 27 and age 34. Observed cheap measurements for these made up individuals as well

	2.5%	25%	50%	75%	97.5%	Truth
μ_x^{EE}	2610.19	2663.99	2691.59	2719.67	2772.07	
$\mu_x^{\Delta ES}$	-8.11	-1.42	2.08	5.61	12.38	
σ_{yee}	381.01	396.92	405.71	414.93	435.80	
σ_{yes}	296.81	309.75	316.87	324.23	339.56	
σ_{wee}	236.55	248.84	255.74	262.91	277.50	
σ_{wes}	64.41	67.17	68.73	70.33	73.51	
σ_{xee}	637.76	673.81	694.27	715.80	759.92	
σ_{xes}	67.03	72.23	75.17	78.12	84.09	
ρ	-0.23	-0.15	-0.11	-0.06	0.02	
$\gamma_{1,ee}$	178.48	244.17	278.90	314.08	382.01	300
$\gamma_{2,ee}$	-0.05	4.81	7.34	9.87	14.70	14
$\gamma_{3,ee}$	-17.51	-13.32	-11.09	-8.84	-4.59	-7
$\gamma_{1,es}$	-296.12	-241.05	-212.36	-183.20	-126.47	-200
$\gamma_{2,es}$	3.02	6.97	9.05	11.17	15.32	8
$\gamma_{3,es}$	-10.58	-6.88	-4.94	-3.05	0.63	-5

Table 9: Parameter estimates for Model BVN for Normal errors with correct specification of within person errors

as 95% credible intervals for their calibrated truth under normal errors are given in Table 13. Figures 33 gives histograms of 500 calibrated draws for each individual for EE and Δ ES measurements under normal errors. Results for skewed and bimodal were very similar and omitted.

8 Discussion

In this chapter we presented a semi-parametric model for the measurement of EE and Δ ES. 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.

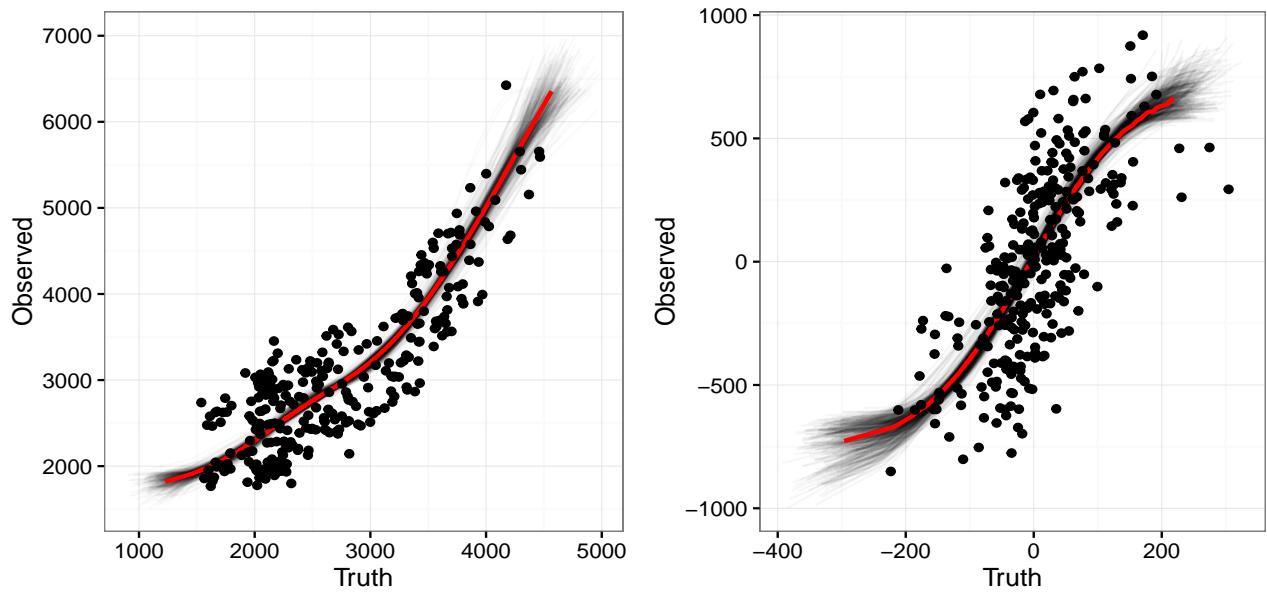


Figure 32: Spline function for Model BVN with Bimodal Errors

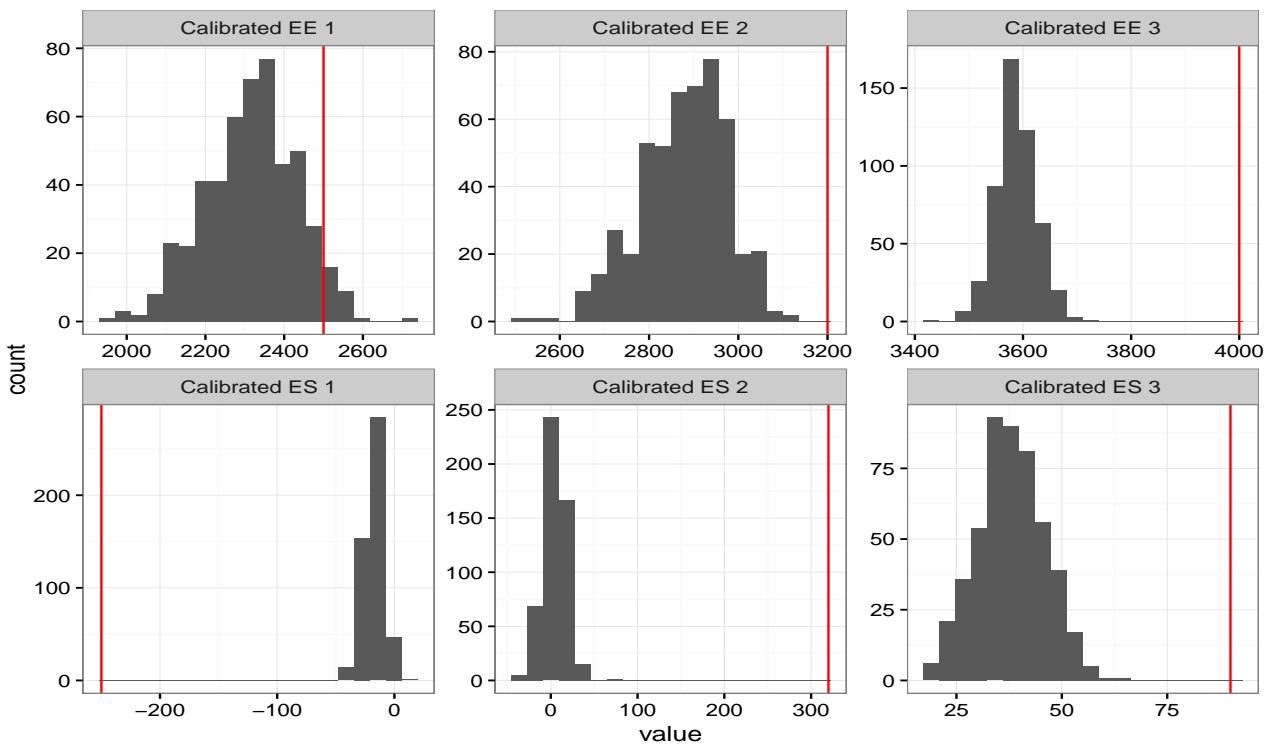


Figure 33: Posteriors of calibrated observations. Vertical line shows observed value from cheap measurement.

	2.5%	25%	50%	75%	97.5%	Truth
μ_x^{EE}	2622.55	2675.43	2703.64	2732.01	2786.64	
$\mu_x^{\Delta ES}$	-7.62	-0.93	2.59	6.10	12.84	
σ_{yee}	371.38	387.03	395.75	405.05	425.11	
σ_{yes}	291.18	303.53	310.39	317.55	331.87	
σ_{wee}	241.21	253.60	260.49	267.51	281.58	
σ_{wes}	65.05	67.98	69.56	71.24	74.65	
σ_{xee}	641.81	679.00	699.07	720.28	764.17	
σ_{xes}	65.68	70.97	73.92	76.89	82.68	
ρ	-0.28	-0.20	-0.15	-0.11	-0.02	
$\gamma_{1,ee}$	150.70	216.27	251.30	285.24	352.67	300
$\gamma_{2,ee}$	0.29	5.23	7.76	10.27	15.22	14
$\gamma_{3,ee}$	-17.58	-13.31	-11.06	-8.85	-4.57	-7
$\gamma_{1,es}$	-297.09	-244.54	-217.35	-190.06	-138.10	-200
$\gamma_{2,es}$	2.31	6.06	8.01	9.99	13.75	8
$\gamma_{3,es}$	-9.18	-5.74	-3.93	-2.15	1.27	-5

Table 10: Parameter estimates for Model BVN for Skewed errors with correct specification of within person errors

Because this model was able to 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 application. The resulting estimates, posterior predictive checks, and PMSE show the superiority of our model to a simpler linear measurement error model and the naive 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 measuremet for EE and Δ ES 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 Δ ES 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 SWA 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. Revisting the conditional independence assumptions could provide

	2.5%	25%	50%	75%	97.5%	Truth
μ_x^{EE}	2581.07	2634.84	2662.65	2690.48	2743.29	
$\mu_x^{\Delta ES}$	-14.55	-7.47	-3.70	0.01	6.92	
σ_{yee}	245.77	261.27	269.43	277.84	297.13	
σ_{yes}	223.45	232.56	237.53	242.90	253.34	
σ_{wee}	176.09	186.78	192.79	199.36	213.55	
σ_{wes}	48.83	50.96	52.15	53.36	55.79	
σ_{xee}	646.03	682.15	702.09	722.74	764.86	
σ_{xes}	79.09	84.14	86.92	89.80	95.74	
ρ	-0.06	0.02	0.06	0.10	0.18	
$\gamma_{1,ee}$	125.13	175.72	201.12	226.66	276.14	300
$\gamma_{2,ee}$	4.54	8.05	9.86	11.66	15.18	14
$\gamma_{3,ee}$	-16.95	-13.86	-12.24	-10.63	-7.58	-7
$\gamma_{1,es}$	-239.57	-198.67	-177.29	-155.91	-115.00	-200
$\gamma_{2,es}$	1.66	4.61	6.19	7.71	10.72	8
$\gamma_{3,es}$	-7.02	-4.31	-2.90	-1.49	1.20	-5

Table 11: Parameter estimates for Model BVN for Bimodal errors with correct specification of within person errors

	Normal Error		SN Error		Bimodal Error	
	EE	ΔES	EE	ΔES	EE	ΔES
Naive Model	152644	57783	148914	50034	102024	22353
Simple Model	121452	42395	103687	39299	55922	19885
Our Model	75488	38350	73637	37313	47398	18940

Table 12: Mean PMSE for all models under 3 different error distributions using correct within person variability specification

additional improvements, especially the last two. We modeled the bivariate latent variables jointly but random variables Y^{EE} and W^{EE} were assumed to be conditionally independent of $X^{\Delta ES}$ and vice versa, but there could be a relationship there to take advantage of. We performed the same analysis assuming complete independence of EE and ΔES and there was much more uncertainty in estimates, we anticipate modeling more of the dependence will only help. We assumed normality for our measurement errors which based on diagnostics from the EBS, looked good, however there are more flexible options available, for example Sarkar et al. We are not able to split within person variability and measurement error explicitly which would be desireable for fully understanding the random error associated with a measurement device. 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 $m(\cdot)$. These last issues will be difficult to fix because we would need additional replicate data and there's no way to exactly decouple the relationship within the B-spline.

	2.5%	50%	97.5%	Observed
EE	2088.69	2323.70	2521.56	2500.00
	2672.75	2887.67	3043.73	3200.00
	3513.16	3584.89	3663.53	4000.00
Δ ES	-34.70	-16.38	-1.54	-250.00
	-21.77	5.32	29.48	320.00
	22.71	37.43	53.35	90.00

Table 13: 95% credible interval for calibration estimate for cheap EE measurements for Normal Errors

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 Endocrinol Metab*, 302:441–8, 2011.
- [8] Gregory A. Hand, Robin P. Shook, Amanda E. Paluch, Meghan Baruth, Patrick Crowley, Jason R. Jagers, 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. Carroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. *Measurement Error in Nonlinear Models*. Chapman and Hall/CRC, 2006.
 - [10] CDC. Cdc website. <https://www.cdc.gov/nchs/data/nhanes/databriefs/adultweight.pdf>. Accessed: 2016-10-10.
 - [11] 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.
 - [12] Kunio Takezawa. *Introduction to Nonparametric Regression*. Wiley Series in Probability and Statistics, 2006.
 - [13] 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.
 - [14] 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.
 - [15] Peter J. Green. Reversible jump markov chain monte carlo computation and and bayesian model determination. *Biometrika*, 82:711–32, 1995.
 - [16] Ilaria DiMatteo, Christopher R. Genovese, and Robert E. Kass. Bayesian curve-fitting with free-knot splines. *Biometrika*, 88:1055–71, 2001.
 - [17] Mary J. Lindstrom. Bayesian estimation of free-knot splines using reversible jumps. *Computational Statistics and Data Analysis*, 41:255–69, 2002.
 - [18] Matthew S. Johnson. Modeling dichotomous item responses with free-knot splines. *Computational Statistics and Data Analysis*, 51:4178–92, 2007.
 - [19] 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.