Measuring Energy Intake via Energy Balance Principle While Accounting for Measurement Error

Danny Ries

Iowa State University

June 22, 2016

The Obesity Epidemic

Over 35% of Americans are obese, and over 75% of men are either overweight or obese. Obesity is linked to many different medical, psychological, emotional, and economic effects such as:

- Type 2 Diabetes
- Coronary Heart Disease
- High Blood Pressure
- Clinical Depression
- Anxiety
- Increased Health Care Costs
- Lost Wages
- Discrimination

The "Fatal Flaw in Obesity Research"

It has been said the "Fatal Flaw in Obesity Research" is our inability to accurately measure how much someone eats (EI) in free living situations

- Current measures of EI; ie. self report, are clouded with (measurement) error
- Garbage in Garbage out
- Tough to understand dietary trends over the years
- Cannot measure adherence to clinically prescribed interventions

This error in measurement extends to EE and body composition, albeit not nearly as severe

2 Remedies to aid in Obesity Research

1. Accurately and efficiently measure Energy Intake (EI)
2. Assess compliance to *2008 Physical Activity Guidelines*

Modeling Energy Balance

Energy Balance

The application of the first law of thermodynamics to nutrition/exercise science:
Change in Energy Stores (ΔES) = Energy Intake (EI) - Energy Expenditure (EE)

where $\Delta \text{ES} = c_1 \frac{\Delta FM}{\Delta T} + c_2 \frac{\Delta FFM}{\Delta T}$

This provides an alternative way to measure EI for an individual

Modeling Energy Balance Cont.

We are now in a situation where we must measure both EE and ΔES in order to calculate EI
$\rightarrow$ But gold standard measures for both exist!

In a world of unlimited resources, researchers needing EI for individuals could use gold standard measures of EE and ΔES and use simple measurement error models

Unfortunately,
DLW $\sim$ \$500/person
DXA $\sim$ \$100/person

Modeling Energy Balance Cont.

There are many other cheaper measures of EE and ΔES, that even when used together to calculate EI, are still more accurate than self-reported EI

Goal: Create a statistical measurement model for gold standard and cheap measurements of both EE and ΔES in order to develop calibration equations for cheap measurements.

This will allow future research to calibrate cheaper measurements (when gold standard measures aren't used) and thus eliminate known biases

Modeling Energy Balance Cont.

- Lots of research has been done for calibrating and evaluating measurement error for EI
- Some research for EE
- Little research for ΔES

To the best of our knowledge, no research has been done in evaluating the measurement error and calibrating measurements jointly via the Energy Balance principle

Energy Balance Study

The Energy Balance Study (EBS) was conducted 2011-2012 at the University of South Carolina

- 430 male and females aged 20-35
- 5 DXA scans, one every 3 months
- Sensewear Armband measuring EE every 3 months (averaged across 10 days)
- Subset of 119 participants receieved DLW at end of 12 months, with additional DXA scan
- Demographic variables

Although these data don't have perfect replicates, it provides a baseline to start our modeling of measurement error in Energy Balance

Checking Normality of Measurement Errors

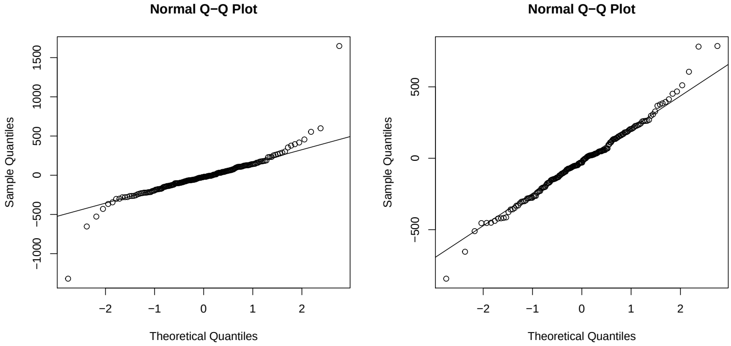


Figure: Differenced DXA ΔES

Checking Normality of Measurement Errors

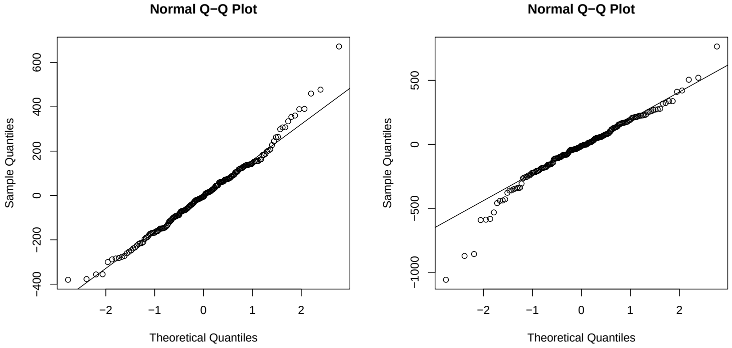


Figure: Differenced SWA EE

Näive Check for Biases

It seems reasonable that cheap measurement tools could be affected by factors other than the *truth*, ie. demographics

$\rightarrow$ Fit the following multiple regression model with the subset data from EBS
SWA EE $= \beta_0 + \beta_1$DLW EE $+ \beta_z$' Demographic Variables $+ \epsilon$

This is not ideal since we would want True EE not DLW EE in the regression, but still provides motivation

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

Coefficient	Estimate	Std Error	P-value
Intercept	878.422	154.113	$<$0.0001
DLW EE	0.558	0.040	$<$0.0001
Age	-7.351	3.999	0.0676
Male	305.582	43.258	$<$0.0001
BMI	14.146	3.988	0.0004

Notation

let i represent individual and j represent replicate number

Observable:

- W_{ij}^{EE} and $W_{ij}^{\Delta ES}$ represent gold standard measures of EE and ΔES
- Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$ represent cheap measures of EE and ΔES
- Z_i represent a $k \times 1$ vector of error free covariates

Latent:

- X_i^{EE} and $X_i^{\Delta ES}$ represent *usual* EE and ΔES

Independence Assumptions

Given X_i^{EE}, Z_i

- Y_{ij}^{EE} are mutually independent for all i, j
- W_{ij}^{EE} are mutually independent for all i, j
- Y_{ij}^{EE} is independent of W_{ij}^{EE} for all i, j
- Y_{ij}^{EE} is independent of $W_{ij}^{\Delta ES}$ and $Y_{ij}^{\Delta ES}$ for all i, j
- W_{ij}^{EE} is independent of $W_{ij}^{\Delta ES}$ and $Y_{ij}^{\Delta ES}$ for all i, j

Same assumptions hold for reverse case (replace EE with ΔES and ΔES with EE)

Model for Observed Variables

$$
\begin{aligned}
Y_{ij}^{EE} &= m_{ee}(X_i^{EE}, Z_i) + \epsilon_{ij}^{EE} \\
Y_{ij}^{\Delta ES} &= m_{es}(X_i^{\Delta ES}, Z_i) + \epsilon_{ij}^{\Delta ES} \\
W_{ij}^{EE} &= X_i^{EE} + \nu_{ij}^{EE} \\
W_{ij}^{\Delta ES} &= X_i^{\Delta ES} + \nu_{ij}^{\Delta ES}
\end{aligned}
$$

$$E(\epsilon_{ij}^{EE}) = E(\epsilon_{ij}^{\Delta ES}) = E(\nu_{ij}^{EE}) = E(\nu_{ij}^{\Delta ES}) = 0$$

Joint Likelihood

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

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

Näive Model

The Näive Model assumes no measurement error in gold standard measurements (Note the part for EE is the same as what we used for our exploratory analysis)

$$(Y_{ij}^{EE}|W_{ij}^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \overset{iid}{\sim} N(\beta_{0,ee} + \beta_{1,ee} W_{ij}^{EE} + \boldsymbol{\gamma_{ee}} Z_i, \sigma^2_{\epsilon^{EE}})$$
$$(Y_{ij}^{\Delta ES}|W_{ij}^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \overset{iid}{\sim} N(\beta_{0,es} + \beta_{1,es} W_{ij}^{\Delta ES} + \boldsymbol{\gamma_{es}} Z_i, \sigma^2_{\epsilon^{\Delta ES}})$$

Linear Measurement Error Model

This is a basic modification to the Näive model when there is measurement error in a covariate

$$
\begin{aligned}
(Y_{ij}^{EE} | X_i^{EE}, Z_i, \boldsymbol{\theta_{yee}}) &\sim N(\beta_{0,ee} + \beta_{1,ee} X_i^{EE} + \boldsymbol{\gamma_{ee}} Z_i, \sigma^2_{\epsilon_{EE}}) \\
(Y_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) &\sim N(\beta_{0,es} + \beta_{1,es} X_i^{\Delta ES} + \boldsymbol{\gamma_{es}} Z_i, \sigma^2_{\epsilon_{\Delta ES}}) \\
(W_{ij}^{EE} | X_i^{EE}, Z_i, \boldsymbol{\theta_{wee}}) &\overset{iid}{\sim} N(X_i^{EE}, \sigma^2_{\nu_{EE}}) \\
(W_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wes}}) &\overset{iid}{\sim} N(X_i^{\Delta ES}, \sigma^2_{\nu_{\Delta ES}}) \\
(X_i^{EE}, X_i^{\Delta ES} | \theta_X) &\overset{iid}{\sim} N\left(\begin{bmatrix} \mu_{EE} \\ \mu_{\Delta ES} \end{bmatrix}, \Sigma_X \right)
\end{aligned}
$$

Extending the Linear Model

We would like to relax the assumption that the relationship between a cheap measurement and *usual* EE and ΔES is linear

We propose using free knot splines to model the relationship between cheap and *usual*

- Allows for a flexible nonlinear relationship
- No need to specify number or location of knots
- If using Reversible Jump MCMC, incorporates uncertainty in spline selection

Free Knot Spline Model

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

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

In order to ensure the functions are monotone, constrain
$\beta_{1,ee} \leq \beta_{2,ee} \leq \dots \leq \beta_{k_{ee}+3,ee}$

MCMC Algorithms

1. The Näive model and Linear Measurement Error Model were fit using JAGS
2. The Spline Measurement Error Model required Reversible Jump MCMC (RJMCMC) and was written in C++ via `Rcpp`

Gibbs Sampler

Priors were chosen to be conjugate to help simplify the sampler

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

1. $\{\Sigma^{(k)}\} \mid \cdot \overset{ind}{\sim}$
   $Inv-Wish(d+n, \psi + (\mathbf{X_i^{(k-1)}} - \mu^{(\mathbf{k-1})})'(\mathbf{X_i^{(k-1)}} - \mu^{(\mathbf{k-1})}))$
2. $\{(\mu^{(k)}_{EE}, \mu^{(k)}_{\Delta ES})\} \mid \cdot \overset{ind}{\sim} N(M'_\mu, C'_\mu)$
   $C'_\mu = (C_\mu^{-1} + n\Sigma^{-1(k)})^{-1}$
   $M'_\mu = C'_\mu(C_\mu^{-1}M + n\Sigma^{-1(k)}\bar{\mathbf{X}}^{(\mathbf{k-1})})$
   $\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X_i}$
3. Update $\mathbf{X_i}$ with a random walk
   $\{\mathbf{X_i^{(k)}} : i = 1, ..., n\} \mid \cdot$ for $i = 1, ...n$ sample $\mathbf{X_i^*}$ from $N(\mathbf{X_i^{(k)}}, C_{X_i})$
   Set $\mathbf{X_i^{(k)}} = \mathbf{X_i^*}$ with probability α_{X_i}, otherwise set $\mathbf{X_i^{(k)}} = \mathbf{X_i^{(k-1)}}$
   $\alpha_{X_i} = min\left(1, \frac{f(\mathbf{X_i^*}|\cdot)}{f(\mathbf{X_i^{(k-1)}}|\cdot)}\right)$

4. Update $\boldsymbol{\beta_{ee}}, \boldsymbol{\beta_{es}}, k_{ee}, k_{es}, r_{ee}, r_{es}, \boldsymbol{\gamma_{ee}}, \boldsymbol{\gamma_{es}}$ using RJMCMC described next. Calculate $m_{ee}(\mathbf{X^{EE(k)}}; \boldsymbol{\beta_{ee}^{(k)}})$ and $m_{es}(\mathbf{X^{\Delta ES(k)}}; \boldsymbol{\beta_{es}^{(k)}})$
5. $\sigma_{\epsilon^{EE}}^{2(k)}|\cdot \sim IG(a_{yee} + J \times \frac{n}{2}, b_{yee} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(Y_{ij}^{EE} - m_{ee}(X_i^{EE(k)}; \boldsymbol{\beta_{ee}^{(k)}}) - \boldsymbol{\gamma_{ee}}^{(k)} Z_i)^2)$
6. $\sigma_{\epsilon^{\Delta ES}}^{2(k)}|\cdot \sim IG(a_{yes} + J \times \frac{n}{2}, b_{yes} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(Y_{ij}^{\Delta ES} - m_{es}(X_i^{\Delta ES(k)}; \boldsymbol{\beta_{es}^{(k)}}) - \boldsymbol{\gamma_{es}}^{(k)} Z_i)^2)$
7. $\sigma_{\nu^{EE}}^{2(k)}|\cdot \sim IG(a_{wee} + J \times \frac{n}{2}, b_{wee} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(W_{ij}^{EE} - X_i^{EE(k)})^2)$
8. $\sigma_{\nu^{\Delta ES}}^{2(k)}|\cdot \sim IG(a_{wes} + J \times \frac{n}{2}, b_{wes} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(W_{ij}^{\Delta ES} - X_i^{\Delta ES(k)})^2)$

RJMCMC

RJCMCMC was introduced by Green (1995) as a way to "jump" between different models and dimensions as a means for model determination. Denison introduced a simple and effective implementation of a Bayesian free-knot spline

We extend method of Denison (1998) in two ways:

1. The spline covariate is a latent variable that is also sampled from
2. There are additional linear components to the mean function

Our algorithm is run independently for EE and ΔES regression functions due to conditional independence, so we will let $(\cdot)$ be a placeholder

$\mathcal{A}$=$\{x_i, i = 1, \ldots n : x_i$ is not currently a knot or within $\ell + 1$ locations of a current knot$\}$

1. Calculate $b_k = c \times min\left(1, \frac{p(k+1)}{p(k)}\right)$
   $d_k = c \times min\left(1, \frac{p(k)}{p(k+1)}\right)$
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 and linear regression coefficients $\beta^*_\cdot, \gamma^*_\cdot$ by using OLS by regressing $Y^\cdot$ on $B^*_\cdot(X^{\cdot(k)}) + Z$
6. Accept proposed knots and coefficients with probability α. Otherwise set $r_\cdot^{(k)} = r_\cdot^{(k-1)}$, $\boldsymbol{\beta}_\cdot^{(\boldsymbol{k})} = \boldsymbol{\beta}_\cdot^{(\boldsymbol{k-1})}$, $\boldsymbol{\gamma}_\cdot^{(\boldsymbol{k})} = \boldsymbol{\gamma}_\cdot^{(\boldsymbol{k-1})}$

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

$$Z(k) = 2(\ell + 1) + k_\cdot(2\ell + 1)$$
$$k_\cdot = length(r_\cdot^{(k-1)})$$

7. calculate mean function $m(X^{\cdot(k)}; \boldsymbol{\beta}^{(\boldsymbol{k})})$ using spline basis matrix

Simulation Study

We performed a simulation study to assess the performance of the models we discussed in terms of parameter estimation and prediction (for calibration)

Simulated 200 data sets for both 2 and 4 replicates per individual. Number of individuals was set to be 300. Three different measurement errors were used for each set of replicates:

- Normal
- Skew-Normal
- Bimodal (50-50 mixture of two Normals)

For each simulated data set, we ran the MCMC for 12,000 iterations, using the first 2000 as burn in

	σ_{yee}		σ_{yes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2
Mean Est	477.65	473.42	347.49	354.85	254.67	248.66	14.88	14.03	-4.14	-5.29	-200.53	-199.39	7.91	8.31	-4.9
Std Err	17.82	19.24	9.43	6.80	43.65	36.10	4.33	3.62	3.37	3.06	28.19	22.83	2.90	2.13	2.35
Bias	72.15	67.92	13.49	20.85	-45.33	-51.34	0.88	0.03	2.86	1.71	-0.53	0.61	-0.09	0.31	0.06
Truth	405.50	405.50	334.00	334.00	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.0

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2
Mean Est	444.34	446.63	320.41	338.26	255.70	255.85	69.18	71.81	249.50	240.63	14.30	13.67	-4.50	-5.28	-199
Std Err	16.84	14.22	10.76	7.53	10.74	6.33	2.27	1.56	43.44	37.02	4.25	3.60	3.38	3.04	28.3
Bias	38.84	41.13	-13.59	4.26	5.70	5.85	-3.68	-1.05	-50.50	-59.37	0.30	-0.33	2.50	1.72	0.73
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2
Mean Est	393.69	400.55	313.47	331.79	246.81	248.93	67.61	71.04	293.16	294.61	14.17	14.11	-6.86	-6.78	-200
Std Err	11.52	8.29	12.00	8.27	8.78	5.79	2.32	1.66	36.16	26.19	3.50	2.42	2.86	2.16	26.8
Bias	-11.81	-4.95	-20.53	-2.21	-3.19	-1.07	-5.25	-1.82	-6.84	-5.39	0.17	0.11	0.14	0.22	-0.0
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200

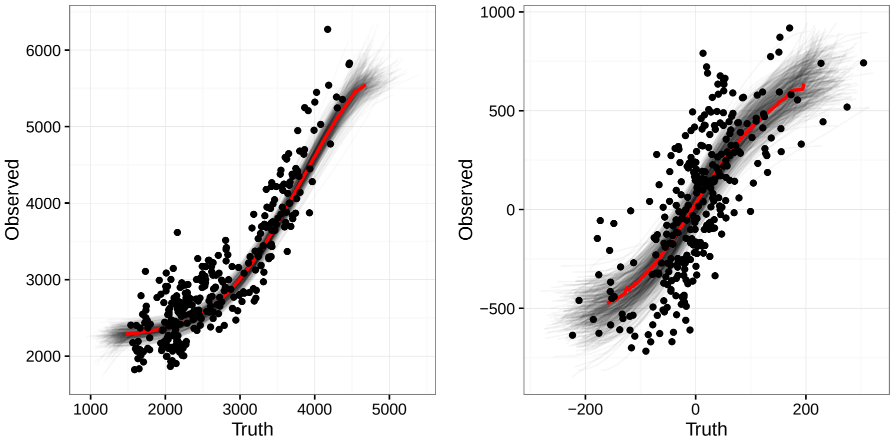
	Replicates	Naive	Linear	SMEMN	SMEMDP
Normal	2	144869.7	113975.56	71308.89	-
	4	100096.78	75098.25	35902.73	35602.62
Skewed	2	138895.41	101586.02	70058.3	-
	4	91110.6	63178.21	35820.27	35449
Bimodal	2	93857.63	46486	42968.21	-
	4	84228.86	35665.26	30936.62	28335.85

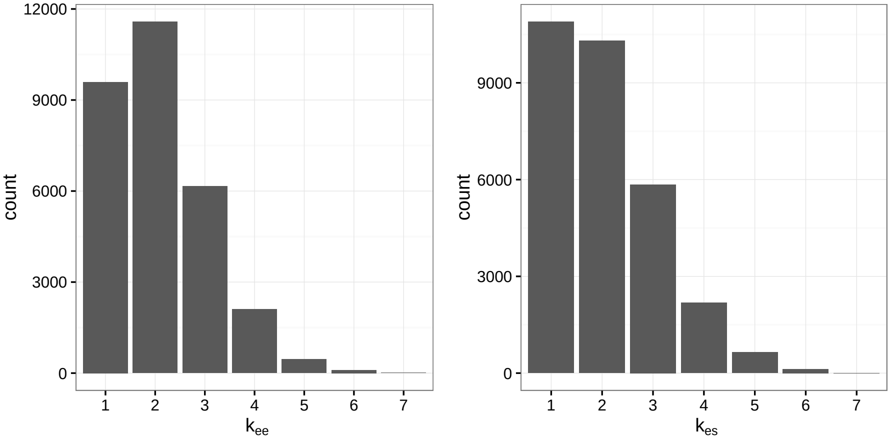
Table: PMSE for EE Regression

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

Table: PMSE for ΔES Regression

Example of one data set




count
12000
9000
6000
3000
0
1 2 3 4 5 6 7
k_{ee}
count
9000
6000
3000
0
1 2 3 4 5 6 7
k_{es}

Calibration

The major goal of this model is to be able to give calibration equations for cheap measurements
$\Rightarrow$ in the future if no gold standard measurements are taken, can correct cheap measurements using these calibration equations

Our goal is to get an estimate of the truth using cheap measurement of the truth y and covariates Z:

$$X_{calibrated} = s^{-1}(y - \gamma' Z)$$

Doing this for both EE and ΔES

Cannot take inverse of spline (which changes dimensions), so we will find its inverse numerically

Use `optimize` wrt x on the value $|f(x) - y^*|$ where $f()$ represents the regression function and y^* is the observed cheap measurement minus the vector of coefficients γ multiplied by the individuals covariate values Z

Use k draws of linear and spline coefficients, knot locations, and latent variables from MCMC to get k calibrated estimates of x

For r = 1,...R

1. Calculate $y_i^* = y_i - \gamma^{(r)'} Z_i$, where Z_i are the covariate values for individual i
2. Use `optimize` for the function $|f_i(x) - y_i^*|$ to choose the value of x that will minimize the aforementioned criteria, call this $x_{i,calibrated}$. $f_i(x)$ is the predicted value of y_i for the given value x using the MCMC draw for the spline coefficients $\beta.^{(r)}$, latent variables $(X^{EE(r)}, X^{\Delta ES(r)})$, and knot locations $(r_{ee}^{(r)}, r_{es}^{(r)})$ from the r^{th} draw of the chain

3 example calibrations

	Lower	Median	Upper	Observed	Truth
EE	2574.18	2666.00	2736.39	3028.89	2199.25
	3452.51	3525.18	3619.08	4119.26	3588.12
	2571.99	2665.46	2744.65	2555.86	2643.14
ΔES	25.15	42.35	60.57	142.30	64.17
	-104.21	-82.93	-63.90	-405.74	-21.08
	-8.41	3.91	17.83	96.06	-0.48

Table: 95% credible interval for calibration estimate for cheap measurements for Skewed Errors

Data Collection

We will be collecting the necessary data on $\sim$30 individuals over the next few months to aid in the development of this model

Discussion

We developed a Bayesian semi-parametric measurement error model for energy balance measurements

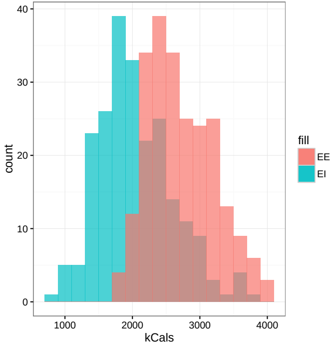
- Constructed a RJMCMC algorithm for free knot splines as a function of latent variables along with a linear component
- Developed calibration algorithm to correct for biases in cheap measurements
- Simulation study shows predictive power of spline model
- Data analysis (coming soon!)

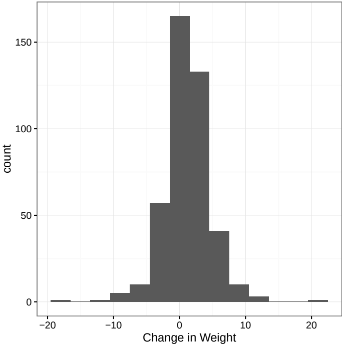
Further Considerations

There are a few ways our model can be improved upon/extended

- Allowing for non-constant variance in measurement errors
- Relaxing normality assumption of measurement errors
- Correctly specifying within person variability for cheap measurements
- Fully Bayesian estimation of spline (and linear component) parameters

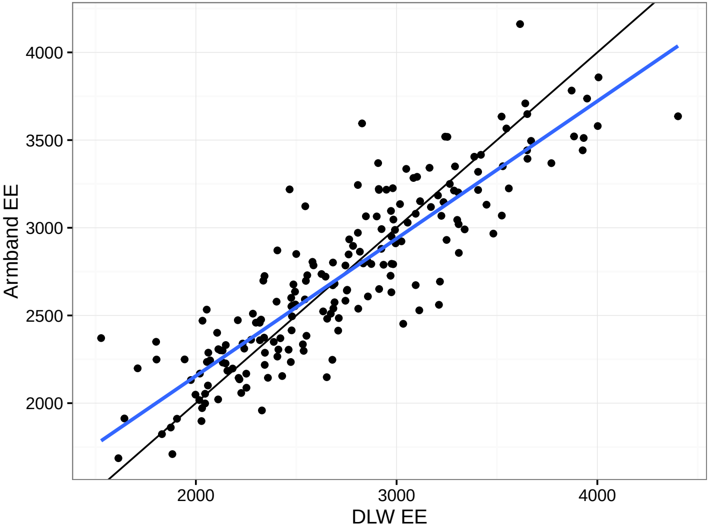
We can create energy!





*Data from Energy Balance Study

Armband vs DLW



Modeling Energy Balance

Measurement Methods

Measurement of EE and ΔES is less noisy than EI and true gold standard measurements exist

- Pedometer
- Consumer grade wearables
- Sensewear Armband
- DLW (gold standard)
- Body Weight
- Calipers
- BodPod
- Bioelectrical impedence
- DXA (gold standard)

Notice how none of these methods require self report?

Research Goal

Research Outcomes

Our overall goal is to provide a cheap, easy, noninvasive *prediction of EI for an individual in free living situations*

To achieve our goal:

- Statistically assess the measurement error in various instruments of both EE and ΔES
- Calibrate cheaper measures so they provide reasonable accuracy without the expense and expertise required for DLW and DXA

Data Required

Because we are assessing the measurement error in these instruments, we need replicate measures of all EE and ΔES measurements on an individual that accounts for all measurement error.

This is a little tricky...

- Replicates measurements of EE for an individual
- Replicate measurements of ΔES for an individual
- Demographic covariates for an individual

Luckily, we have a wealth of data from the Energy Balance Study to help propose a model and empirically check assumptions.

We can further specify forms for each component:

$$Y_{ij}^{EE} = m_{ee}(X_i^{EE}) + \gamma_{ee} Z_i + \epsilon_{ij}^{EE} \tag{1}$$

$$Y_{ij}^{\Delta ES} = m_{es}(X_i^{\Delta ES}) + \gamma_{es} Z_i + \epsilon_{ij}^{\Delta ES} \tag{2}$$

$$W_{ij}^{EE} = X_i^{EE} + \nu_{ij}^{EE} \tag{3}$$

$$W_{ij}^{\Delta ES} = X_i^{\Delta ES} + \nu_{ij}^{\Delta ES} \tag{4}$$

- Because we never can observe $(X_i^{EE}, X_i^{\Delta ES})$, it is difficult to say the functional relationship (m_{ee} and m_{es}) between it and cheap measurements Y^{EE} and $Y^{\Delta ES}$
- Because of this, we choose to model m_{ee} and m_{es} with B-splines
- Must specify number of knots k and knot locations $\zeta_1, ...\zeta_k$

$$m_{ee}(X_i^{EE}) = \sum_{i=1}^{k_{ee}} b_{i,ee}(\zeta_{i,ee})\beta_{i,ee} = B_{ee}(\zeta_{ee})\beta_{i,ee} \tag{5}$$

$$m_{es}(X_i^{\Delta ES}) = \sum_{i=1}^{k_{es}} b_{i,es}(\zeta_{i,es})\beta_{i,es} = B_{es}(\zeta_{es})\beta_{i,es} \tag{6}$$

We let k and $\zeta_1, ...\zeta_k$ vary according to the data

Specifying the Likelihood

$$\epsilon_{ij}^{EE} \overset{iid}{\sim} N(0, \sigma^2_{i,yee}) \tag{7}$$

$$\epsilon_{ij}^{\Delta ES} \overset{iid}{\sim} N(0, \sigma^2_{i,yes}) \tag{8}$$

$$\nu_{ij}^{EE} \overset{iid}{\sim} N(0, \sigma^2_{i,wee}) \tag{9}$$

$$\nu_{ij}^{\Delta ES} \overset{iid}{\sim} N(0, \sigma^2_{i,wes}) \tag{10}$$

Latent Variable Likelihood

The final component to have a joint likelihood specified is the bivariate latent variable. Because we *never ever* observe these values, we must specify its distribution carefully

$$(X_i^{EE}, X_i^{\Delta ES}) \overset{iid}{\sim} \sum_{h=1}^{\infty} \pi_h N(\mu_h, \Sigma_h) \tag{11}$$

$$\pi_h \overset{iid}{\sim} Stick(\alpha) \tag{12}$$

$$\sum_{h=1}^{\infty} \pi_h = 1 \tag{13}$$

$$\pi_h = V_h \prod_{\ell<h} (1 - V_h) \tag{14}$$

$$V_H = 1 \tag{15}$$

$$V_h \sim Beta(1, \alpha), h < H \tag{16}$$

Priors

Need to assign priors for unknown parameters

$$\boldsymbol{\theta} = (\{\sigma_{i,yee}\}_{i=1}^{n}, \{\sigma_{i,yes}\}_{i=1}^{n}, \{\sigma_{i,wee}\}_{i=1}^{n}, \tag{17}$$

$$\{\sigma_{i,wes}\}_{i=1}^{n}, \{\Sigma_i\}_{i=1}^{n}, \{\gamma_{i,yee}\}_{i=1}^{p}, \tag{18}$$

$$\{\gamma_{i,yes}\}_{i=1}^{p}, \{\beta_{i,yee}\}_{i=1}^{k_{ee}}, \{\beta_{i,yes}\}_{i=1}^{k_{kes}}, k_{ee}, k_{es}, \tag{19}$$

$$\{\zeta_{i,ee}\}_{i=1}^{k_{ee}}, \{\zeta_{i,es}\}_{i=1}^{k_{es}}) \tag{20}$$

Assume independent priors (for now)

$$p(\sigma_{i,yee}) \overset{iid}{\sim} C^{+}(0,1) \qquad (21)$$
$$p(\sigma_{i,yes}) \overset{iid}{\sim} C^{+}(0,1) \qquad (22)$$
$$p(\sigma_{i,wee}) \overset{iid}{\sim} C^{+}(0,1) \qquad (23)$$
$$p(\sigma_{i,wes}) \overset{iid}{\sim} C^{+}(0,1) \qquad (24)$$
$$p(\Sigma_i) \overset{iid}{\sim} \text{Inv-Wish}(I_{2x2}, 3) \qquad (25)$$

$$p(\gamma_{i,ee}) \overset{iid}{\sim} N(w_{i,ee}, B_{i,ee}) \qquad (26)$$
$$p(\gamma_{i,es}) \overset{iid}{\sim} N(w_{i,es}, B_{i,es}) \qquad (27)$$
$$p(\beta_{i,ee}) \overset{iid}{\sim} N(v_{i,ee}, C_{i,ee}) \qquad (28)$$
$$p(\beta_{i,es}) \overset{iid}{\sim} N(v_{i,es}, C_{i,es}) \qquad (29)$$
$$p(k_{ee}) \sim Poi(a_{ee}) \qquad (30)$$
$$p(k_{es}) \sim Poi(a_{es}) \qquad (31)$$

$$p(\zeta_1, \ldots, \zeta_{k_{ee}} | k_{ee}) \sim DUnif(x_1^{EE}, \ldots, x_n^{EE}) \qquad (32)$$
$$p(\zeta_1, \ldots, \zeta_{k_{es}} | k_{es}) \sim DUnif(x_1^{\Delta ES}, \ldots, x_n^{\Delta ES}) \qquad (33)$$

Priors likely to change as we elicit expert information and use of past data

Estimation

Since we are taking a Bayesian approach, the posterior distribution $p(\boldsymbol{\theta}|Y, W, Z)$ gives us everything we need for inference

- Using Bayes Rule:

$$p(\boldsymbol{\theta}|Y, W, Z) = \frac{p(Y, W|\boldsymbol{\theta}, Z)p(\boldsymbol{\theta})}{\int p(Y, W, \boldsymbol{\theta}|Z)d\boldsymbol{\theta}} \tag{34}$$

- Integral is impossible to evaluate analytically, use MCMC to simulate draws from joint posterior
- Will use Gibbs Sampler to update parameters
- Reversible Jump MCMC step necessary for B-splines since number of knots k and knot locations $\zeta_1, ...\zeta_k$ are random variables and therefore dimension of posterior is allowed to change
- Implement in R/C++

Model Assessment and Comparason

We will assess the fit of our model through the use of the posterior predictive distributions and relevant discrepancy measures $D()$

$$p(Y^*|W, Y, Z) = \int\int p(Y^*|\boldsymbol{\theta}, X, Z)p(\boldsymbol{\theta}, X|Y, W)d\boldsymbol{\theta}dX \tag{35}$$

$$p(W^*|W, Y, Z) = \int\int p(W^*|\boldsymbol{\theta}, X, Z)p(\boldsymbol{\theta}, X|Y, W)d\boldsymbol{\theta}dX \tag{36}$$

$$p(X^*|W, Y, Z) = \int\int p(X^*|\boldsymbol{\theta})p(\boldsymbol{\theta}, X|Y, W)d\boldsymbol{\theta}dX \tag{37}$$

For each simulated replicate data set (for each Y, W, X) calculate $D()$. Compare to $D()$ from true data

Model Comparason

Although I presented only one specific model here, there are simplifying (and more complicating) assumptions we could (and will) make. To compare models we can use:

- DIC
- Bayes Factors
- PMSE of EI — this is not straightforward
- Parsimony and Practical Interpretation