Supplemental Digital Content

Estimating Long-term Average Household Air Pollution Concentrations from Repeated Short-Term Measurements in the Presence of Seasonal Trends and Crossover

Joshua P. Keller[1] and Maggie L. Clark[2]

[1]Department of Statistics, Colorado State University, Fort Collins, CO, USA

[2]Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, USA

eTable 1: Equations for temporal trends in the simulations.

Label	Temporal Trend
A	$f(t) = 0$
B	$f(t) = 0.5\sin(2\pi t)$
C	$f(t) = -0.1t$
D	$f(t) = -0.5t + 0.5\sin(2\pi t)$
E	$f(t) = -0.5\sin(2\pi t) - 0.1\sin(12\pi t) - 0.2\sin(4\pi(t - 0.2))$
F	$f(t) = 0.2279585 - 0.1\exp(-2\sin(2\pi t)) + 0.05\sin(12\pi t)$

eTable 2: Summary statistics for $PM_{2.5}$ measurements.

	N	Mean	SD	Geo. Mean	GSD	Min	Q25	Median	Q75	Max
Personal										
Traditional	622	141	281	87	2.4	4	50	81	141	5509
Justa	585	83	216	45	2.5	5	27	43	73	3652
Kitchen										
Traditional	629	427	724	183	3.7	4	69	178	440	5520
Justa	578	107	211	58	2.7	5	29	53	103	3358

Geo. Mean = Geometric mean; GSD = geometric standard deviation; N = Number of observations; Q25 = 25th percentile; Q75 = 75th percentile; SD = standard deviation

eTable 3: Parameter estimates from Honduras study mixed models, which included a temporal spline with 6 degrees of freedom.

| Parameter	Personal Measurements	Kitchen Measurements
Intercept (log µg/m3)	4.34 (4.23, 4.45)	5.06 (4.92, 5.21)
Stoves:		
Traditional	Reference	Reference
Justa (log µg/m3)	-0.39 (-0.56, -0.22)	-0.82 (-1.03, -0.61)
Household Variance (σ_A^2)	0.22	0.48
Residual Variance (σ_E^2)	0.56	0.87

eTable 4: Impact of different numbers of repeated measures, subset by consecutive visits, on accuracy of predicted long-term averages of personal $PM_{2.5}$ exposure in Honduras data

| Prediction Type	Data	Correlation	RMSE
Mixed Model	All (6 obs/hh)	(Reference)	(Reference)
	5 obs/hh (Visits 1-5)	0.970	0.120
	5 obs/hh (Visits 2-6)	0.965	0.200
	4 obs/hh (Visits 1-4)	0.917	0.200
	4 obs/hh (Visits 3-6)	0.927	0.228
	2 obs/hh (Visits 3-4)	0.833	0.261

eTable 5: Impact of different modeling choices and numbers of repeated measures on accuracy of predicted long-term averages of kitchen $PM_{2.5}$ concentrations in Honduras data. Correlations and root mean squared error (RMSE) are calculated using the prediction from the mixed model with all data as the truth. obs/hh=Observations per household.

| Prediction Type	Data	Correlation	RMSE
Mixed Model	All (6 obs/hh)	(Reference)	(Reference)
	5 obs/hh	0.993	0.099
	4 obs/hh	0.971	0.197
	2 obs/hh	0.889	0.383
Mixed Model with no Time Spline	All (6 obs/hh)	0.999	0.067
	5 obs/hh	0.991	0.121
	4 obs/hh	0.969	0.209
	2 obs/hh	0.888	0.383
Household Average (by Stove)	All (6 obs/hh)	0.848	0.628
	5 obs/hh	0.833	0.665
	4 obs/hh	0.826	0.684
	2 obs/hh	0.743	0.894
Single Observation	1 obs/hh	0.682	0.599

eFigure 1: Root mean squared error (RMSE) of predicted long-term averages from simulation with Design 1 (parallel design with measurements every three months, all households with baseline measurements in same month) and Trend A (no temporal trend). Compared to Figure 2, these results come from a simulation with $\sigma_E^2 = 0.5$ (instead of 1). HH = Household, MM = Mixed model
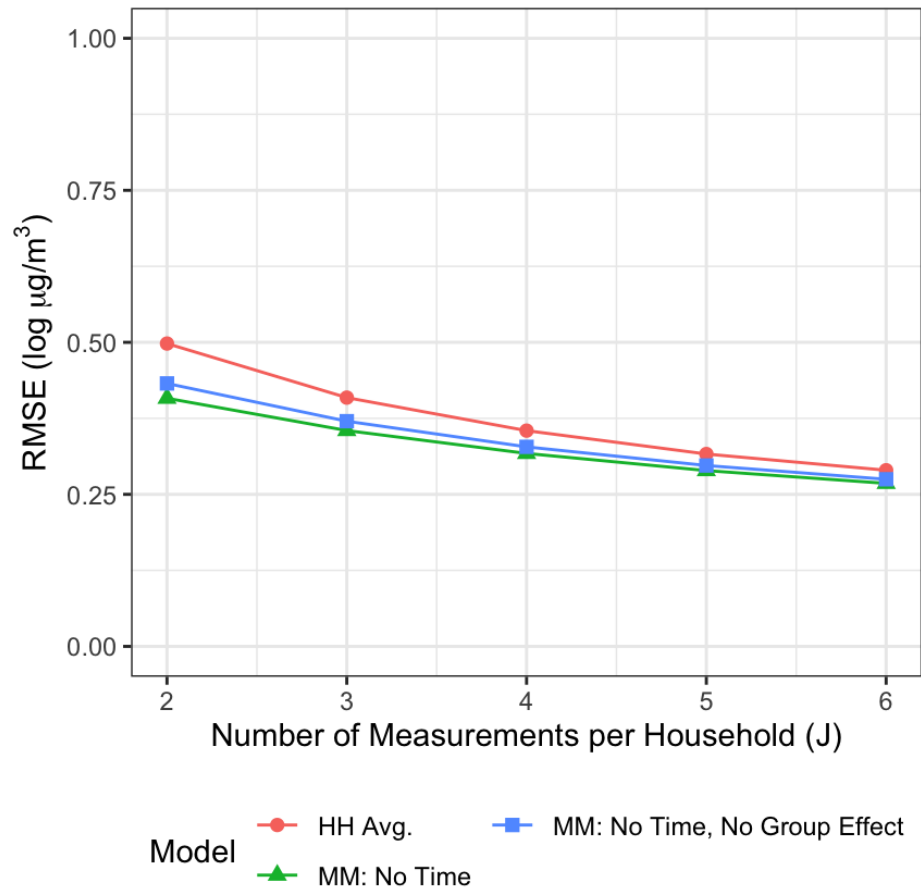
eFigure 2: Absolute root mean squared error (RMSE) of predicted long-term averages for mixed models that ignore time ("MM: No Time"), mixed models that adjusts for time ("MM: Time Spline"), and household averages ("HH Average") in the simulation of Design 1. Panels correspond to the different trends in Figure 1 and for various numbers of repeated measurements per household. This is a companion figure to Figure 3 that depicts relative RMSE.
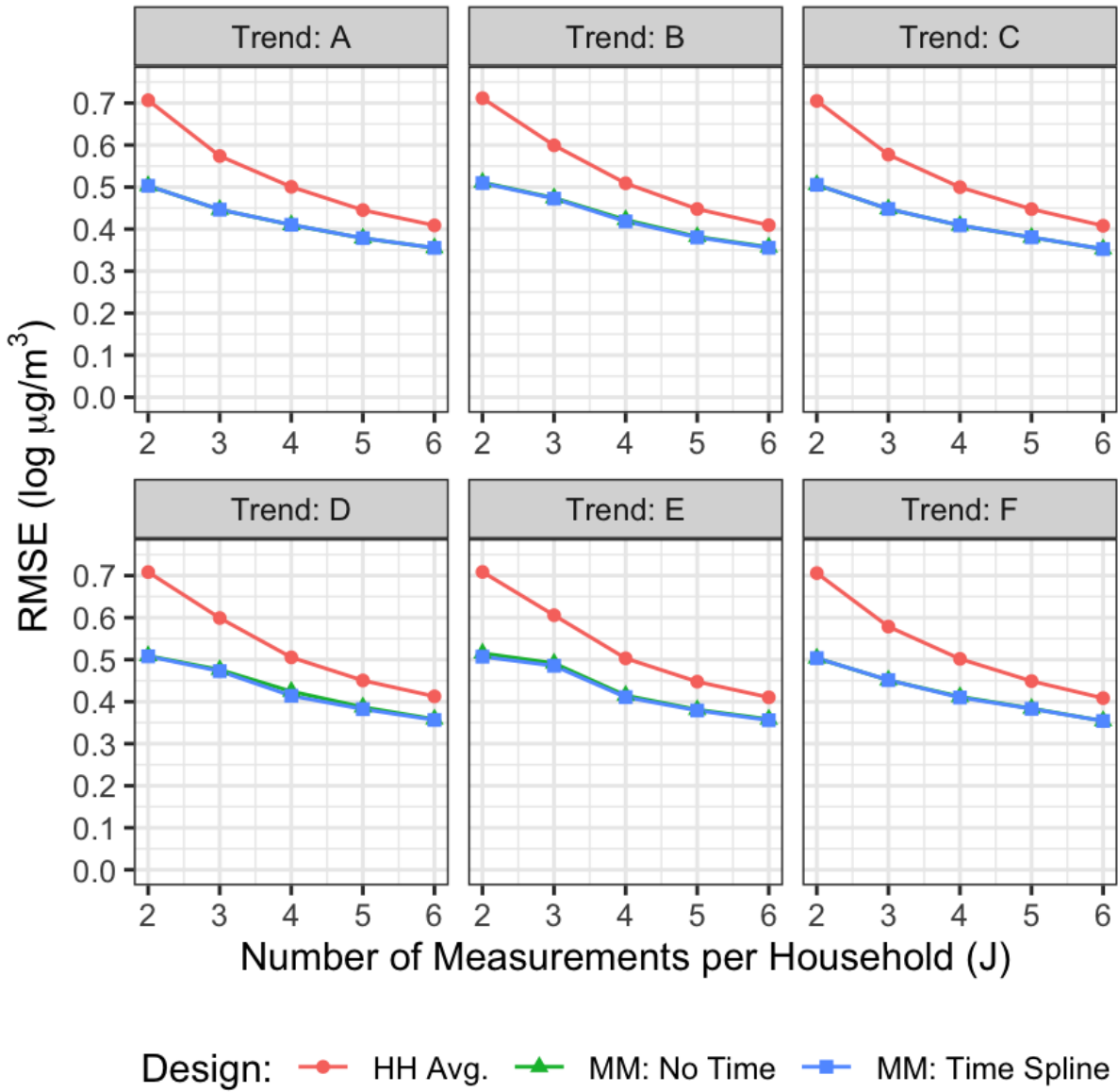
eFigure 3: Absolute mean squared error (MSE) of predicted long-term averages for mixed models that ignore time ("MM: No Time"), mixed models that adjusts for time ("MM: Time Spline"), and household averages ("HH Average") in the simulation of Design 2. Panels correspond to the different trends in Figure 1 and for various numbers of repeated measurements per household. This is a companion figure to Figure 3 that depicts relative MSE.
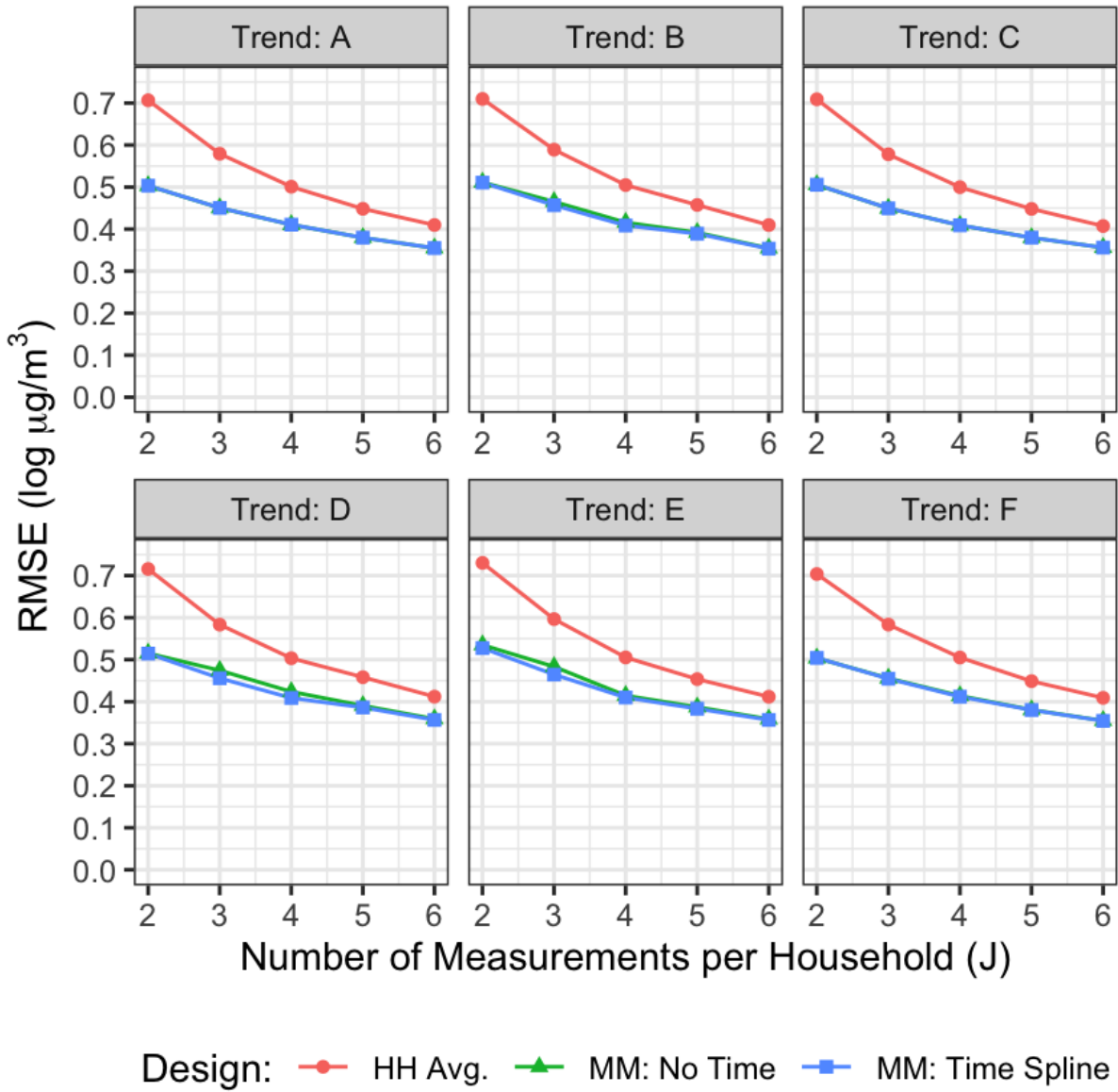
eFigure 4: Cumulative distribution function for predictions of long-term average personal exposure in the Honduras case study.