Supporting Information. Ogle, K. and J.J. Barber. 2020. Ensuring identifiability in hierarchical mixed effects Bayesian models. *Ecological Applications.*

Appendix S1

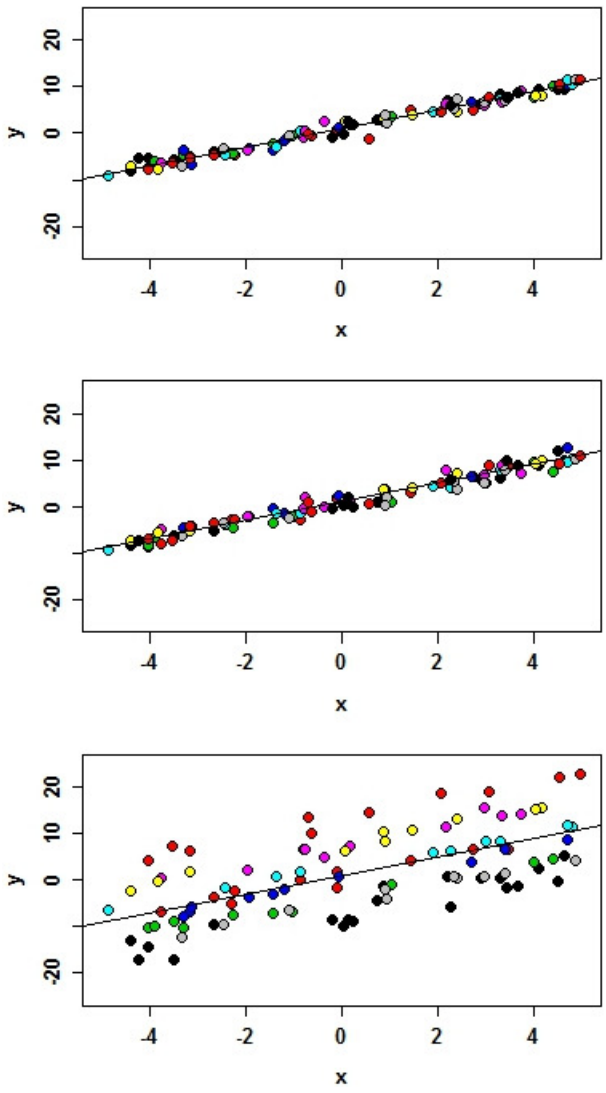

Figure S1. Synthetic data used to illustrate the random effects regression, producing results in Figures 2, 3, and 4. Data were simulated for $J = 10$ random effect levels, with 10 observations each, for a total of $N = 100$. That is, for $y_i \sim Normal(\beta_0 + \beta_1 x_i + \varepsilon_{j(i)}, \sigma^2)$ and $\varepsilon_j \sim Normal(0, \sigma_\varepsilon^2)$, data were generated from true values of $\beta_0 = 1$, $\beta_1 = 2$, $\sigma = 1$, with $x_i \sim Uniform(-5,5)$, and for **(A)** $\sigma_\varepsilon = 0.1$ (top panel), **(B)** $\sigma_\varepsilon = 1$ (middle panel), and **(C)** $\sigma_\varepsilon = 10$ (bottom panel). Points are colored by random effect level.

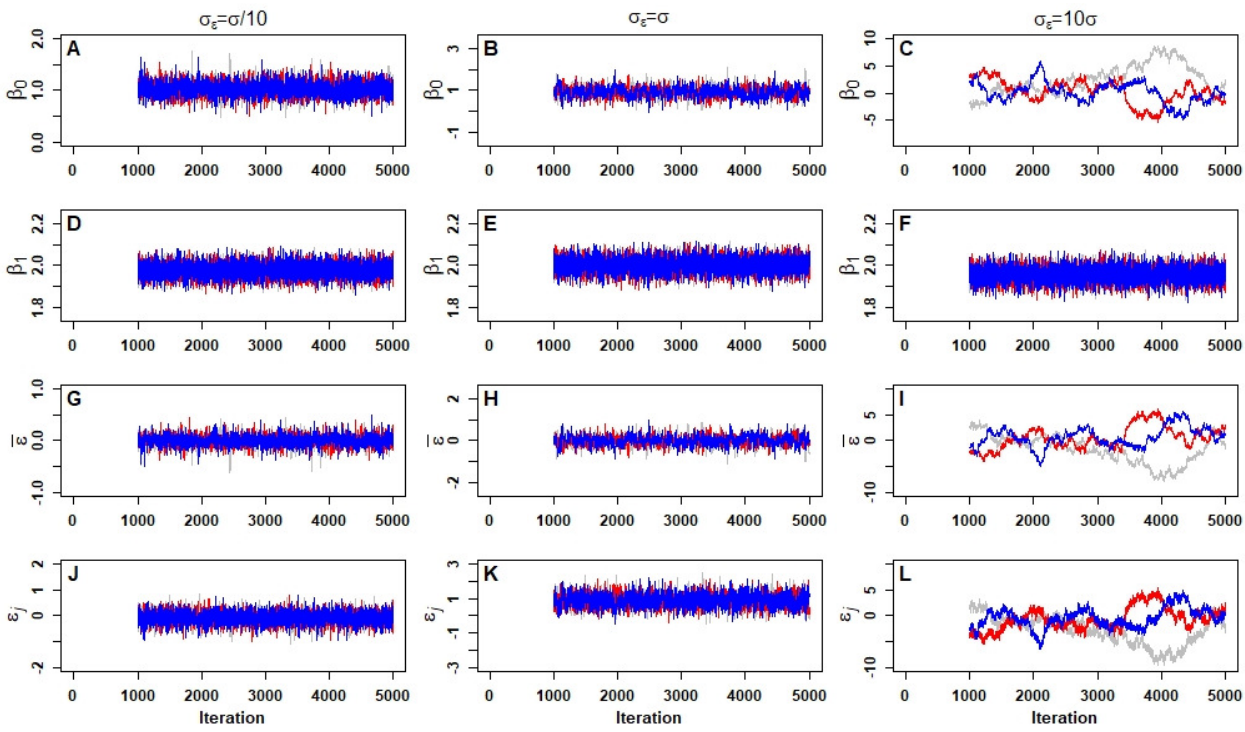

Figure S2. History plots of the MCMC samples for parameters associated with the model in Equation (5), but implemented in OpenBUGS. See Figure 3 in the main manuscript for a full definition of the model, data, and plot details; as for Figure 3, the OpenBUGS simulations are based on the models involving *Gamma*(0.1, 0.1) priors for σ and σ_ε. Here, the history plots start at iteration 1001 given that the adapting period was implemented for the first 1000 iterations.

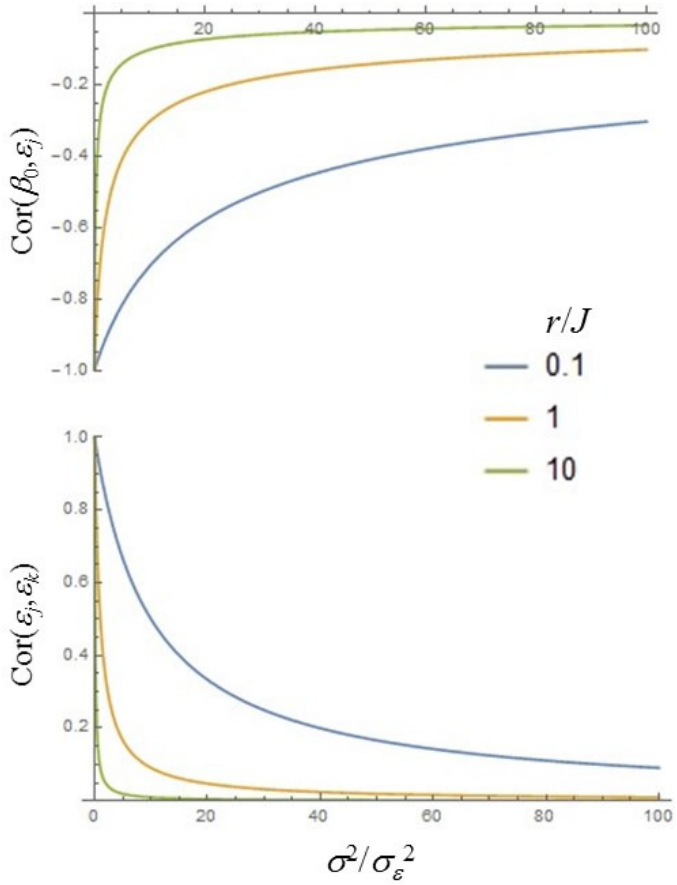

Figure S3. Correlation between the intercept (β_0) and a random effect (ε_j) (top) and between two random effects (ε_j, ε_k), based on a simple model with $\mu_i = \beta_0 + \varepsilon_j$, and with a flat prior on β_0. The correlation function is based on analytical solutions derived by Gelfand et al. (1995), as given in Gilks and Roberts (1996). For simplicity, the correlations are shown for a balanced design whereby the number of replicates (r) from each level of the random effects factor is the same for all levels, and there are J levels of the factor, for a total sample size of $N = rJ$. The correlations decrease as the ratio of the observation variance (σ^2) to the random effects variance (σ_ε^2) increases (σ_ε^2 becomes small relative to σ^2, or as $\sigma^2/\sigma_\varepsilon^2$ increases). Different colored lines correspond to different ratios of r to J.

References

Gelfand, A. E., Sahu, S. K. and Carlin, B. P. 1995. Efficient parametrizations for normal linear mixed models. - Biometrika 82: 479-488.

Gilks, W. R. and Roberts, G. O. 1996. Strategies for improving MCMC. - In: Gilks, W. R., Richardson, S. and Spiegelhalter, D. (eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC.

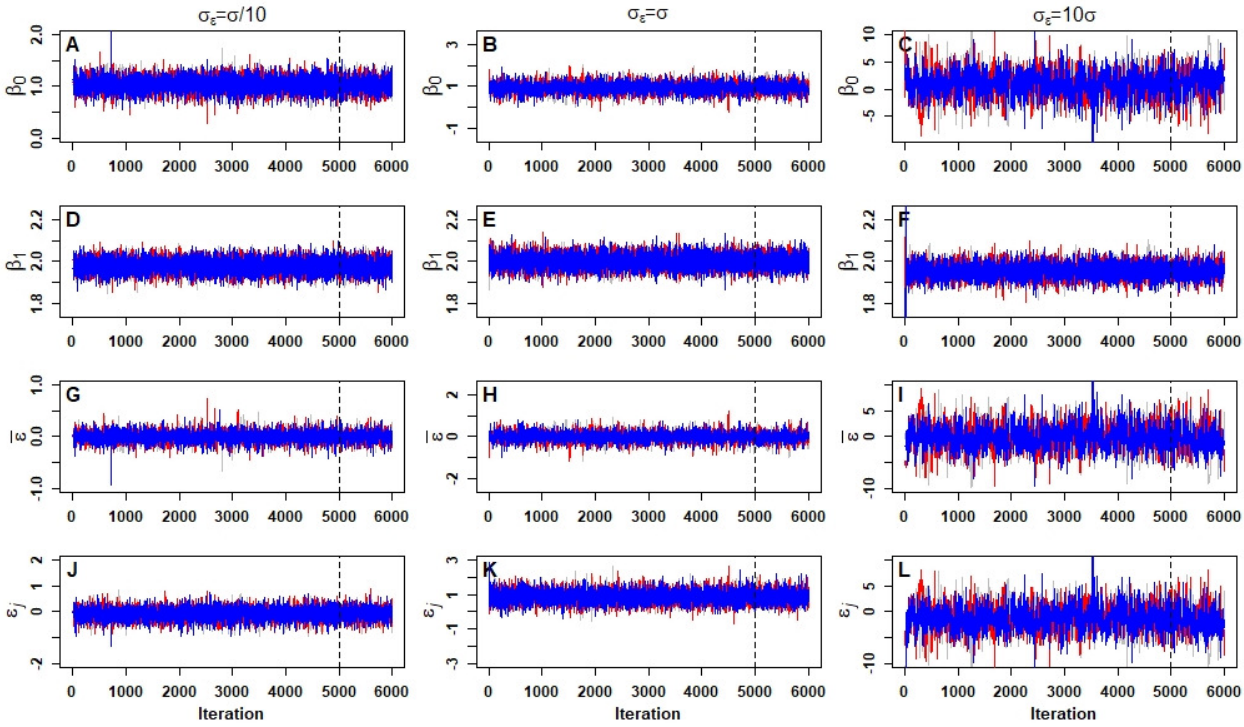

Figure S4. History plots of the posterior samples for parameters associated with the model in Equation (5), but implemented in Stan. See Figure 3 in the main manuscript for a full definition of the model, data, and plot details; as for Figure 3, the Stan simulations are based on the models involving *Gamma*(0.1, 0.1) priors for σ^{-2} and σ_ε^{-2}. Here, the history plots cover iterations 1-6000, where iterations 1-5000 denote the adapting phase (to the left of the vertical dashed line).

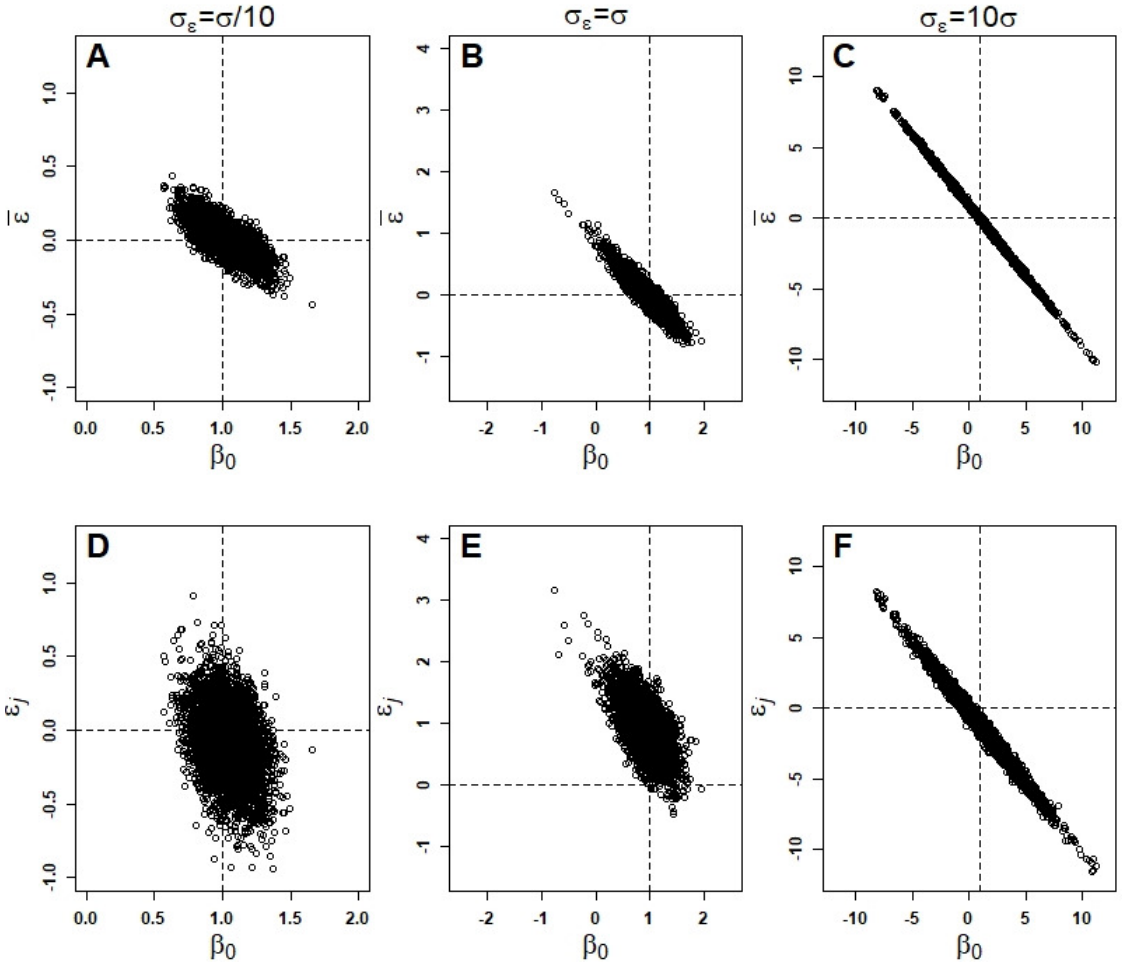

Figure S5. Bivariate scatterplots of the posterior samples obtained from the Stan simulations, after the adapting phase (iterations 5000-1000, one chain), are shown for the random effects mean ($\bar{\varepsilon}$) versus the intercept (β_0) (top row; A, B, and C) and for an individual random effect (ε_j, for $j = 4$) versus β_0 (bottom row; D, E, and F). See Figure 4 in the main text for additional details; the Stan implementation used *Gamma*(0.1, 0.1) priors for σ^{-2} and σ_ε^{-2}, just as in Figure 4. Note that despite the excellent mixing and convergence behavior associated with all three σ_ε scenarios (see Figure S5), the correlation between $\bar{\varepsilon}$ (or ε_j) and β_0 is nearly identical to that produced by the Jags (and OpenBUGS) simulations (compare against Figure 4), where the correlations become stronger as σ_ε increases relative to the observation variance (std. dev., $\sigma = 1$) such that for $\sigma_\varepsilon = 0.1$, **(A)** r = -0.71 and **(D)** r = -0.32; for $\sigma_\varepsilon = 1$, **(B)** r = -0.92 and **(E)** r = -0.60; and, for **(C)** $\sigma_\varepsilon = 10$, r = -1.00 and **(F)** r = -0.98 (F).

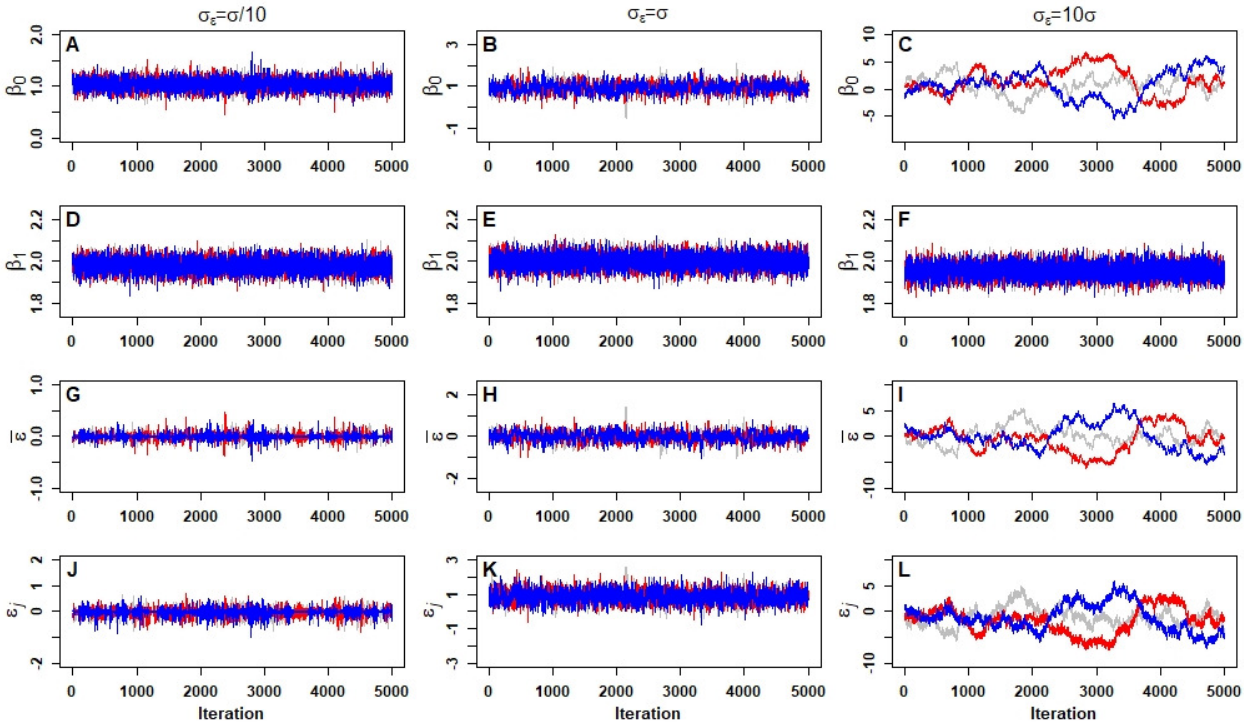

Figure S6. History plots of the MCMC samples for parameters associated with the model in Equation (5); see Figure 3 in the main manuscript for further details. In this example, a folded *Cauchy*(0,1) prior was specified for the random effects standard deviation, σ_ε. As in Figure 3, the simulated data were generated from true values of $\beta_0 = 1$, $\beta_1 = 2$, $\sigma = 1$, and for $\sigma_\varepsilon = 0.1$ (left column), $\sigma_\varepsilon = 1$ (middle column), and $\sigma_\varepsilon = 10$ (right column). The history plots correspond to the overall intercept **(A-C)** β_0, **(D-F)** slope coefficient β_1), **(G-I)** the mean of the random effects, $\bar{\varepsilon}$, and **(J-L)** an individual random effect (ε_j, for $j = 4$). All quantities show excellent mixing and convergence for the scenarios with true $\sigma_\varepsilon = 0.1$ (left column) and $\sigma_\varepsilon = 1$ (middle column), but for large σ_ε ($\sigma_\varepsilon = 10$), $\beta_0, \bar{\varepsilon}$, and ε_j exhibit very poor mixing and lack of convergence by iteration 5000 (C, I, and L, respectively). Differences in mixing and within-chain autocorrelation lead to differences in the number of posterior samples required for inference (see Table S2). Importantly, the history plots behave nearly identical to those shown in Figure 3, which specified a relatively non-informative *Gamma*(0.01, 0.01) prior for the precision, σ_ε^{-2}, indicating that the weakly informative folded *Cauchy*(0,1) prior did not solve the non-identifiability of β_0 and $\bar{\varepsilon}$ or ε_j.

Table S1. Posterior estimates (mean and 95% credible interval [CI]) for the parameters in the random effects linear regression summarized in Equation (5), and based on synthetic data described in Figure 3. The "True value" is the value of the parameter that was used to generate the synthetic data. Red italicized CIs do not contain the true value, which only occurs for three instance of σ_ε. Results are only provided for the original model parameterization, without addressing identifiability issues, implemented in three different software packages: Jags (matches results in Table 1 of the main manuscript for Approach = Orig.), Jags with a folded *Cauchy*(0,1) prior for σ_ε, which matches the results in Table S2 for Approach = Orig.), OpenBUGS, and Stan. With the exception of Jags-FC, all other results are based on models that use *Gamma*(0.1,0.1) priors for the precisions, σ^{-2} and σ_ε^{-2}.

Parameter	True value	Software*	Random effects variance scenario			Total iterations required[&]
			$\sigma_\varepsilon = 0.1\sigma = 0.1$	$\sigma_\varepsilon = \sigma = 1$	$\sigma_\varepsilon = 10\sigma = 10$	
β_0	1	Jags	1.042 (0.779, 1.309)	0.939 (0.439, 1.432)	0.921 (-3.991, 5.801)	830,547
		Jags-FC	1.039 (0.828, 1.249)	0.941 (0.431, 1.458)	0.960 (-3.736, 5.541)	210,604
		BUGS	1.038 (0.777, 1.305)	0.938 (0.432, 1.452)	0.960 (-3.736, 5.541)	1,290,899
		Stan	1.040 (0.772, 1.305)	0.937 (0.419, 1.436)	0.875 (-3.994, 5.628)	25,731
β_1	2	Jags	1.978 (1.914, 2.042)	2.002 (1.935, 2.068)	1.955 (1.886, 2.025)	3,873
		Jags-FC	1.980 (1.916, 2.044)	2.002 (1.935, 2.068)	1.955 (1.886, 2.025)	4,028
		BUGS	1.978 (1.913, 2.040)	2.001 (1.934, 2.066)	1.955 (1.886, 2.025)	4,028
		Stan	1.978 (1.913, 2.042)	2.002 (1.936, 2.067)	1.955 (1.886, 2.025)	7,967
σ	1	Jags	0.935 (0.812, 1.079)	0.959 (0.829, 1.116)	0.995 (0.860, 1.156)	3,906
		Jags-FC	0.933 (0.812, 1.077)	0.959 (0.829, 1.114)	0.995 (0.861, 1.156)	3,865
		BUGS	0.935 (0.811, 1.085)	0.959 (0.828, 1.118)	0.995 (0.861, 1.156)	3,867
		Stan	0.935 (0.812, 1.080)	0.959 (0.829, 1.115)	1.005 (0.862, 1.157)	6,877
σ_ε	Varies (see columns)	Jags	*0.289 (0.152, 0.529)*	0.720 (0.395, 1.261)	7.410 (4.677, 12.305)	3,838
		Jags-FC	0.127 (0.005, 0.381)	0.725 (0.392, 1.264)	7.064 (4.551, 11.395)	5,123
		BUGS	*0.290 (0.152, 0.530)*	0.723 (0.399, 1.270)	7.064 (4.551, 11.395)	7,552
		Stan	*0.290 (0.155, 0.541)*	0.722 (0.398, 1.250)	7.426 (4.646, 12.267)	7,967
$\bar{\varepsilon}$	0	Jags	-0.001 (-0.199, 0.192)	-0.002 (-0.462, 0.466)	-0.024 (-4.902, 4.882)	1,016,392
		Jags-FC	0.000 (-0.110, 0.112)	-0.004 (-0.489, 0.472)	-0.063 (-4.634, 4.627)	198,036
		BUGS	0.002 (-0.189, 0.193)	-0.001 (-0.483, 0.465)	-0.063 (-4.634, 4.627)	1,239,480
		Stan	0.000 (-0.193, 0.193)	0.000 (-0.468, 0.483)	0.022 (-4.742, 4.875)	27,230

*Jags-FC = Jags implementation with folded *Cauchy*(0,1) prior for σ_ε; BUGS = OpenBUGS v3.2.3.

[&]Total iterations required is only provided for the scenario $\sigma_\varepsilon = 10\sigma = 10$, and is based on the Raftery & Lewis diagnostic's (raftery.diag function) estimate of the number of iterations needed to sufficiently sample the (marginal) posterior distribution of each parameter (e.g., considering within-chain autocorrelation of each parameter).

Table S2. Posterior estimates (mean and 95% credible interval [CI]) for the parameters in the random effects linear regression summarized in Equation (5), and based on synthetic data described in Figure 3, and using a folded *Cauchy*(0,1) prior for the random effects standard deviation term (σ_ε). The "True value" is the value of the parameter that was used to generate the synthetic data. Red italicized CIs do not contain the true value, which only occurs for one instance of σ_ε. Approach is indicated by Orig. = original without addressing identifiability issues; HC = hierarchical centering (Solution 1); SZ = sum-to-zero constraints for random effects (Solution 2); and, PS = post-sweeping of random effects (Solution 4). Results for scenario $\sigma_\varepsilon = 10\sigma = 10$ are only shown.

Parameter	True value	Approach	Random effects variance scenario: $\sigma_\varepsilon = 10\sigma = 10$		
			Posterior mean	Central 95% CI	Total iterations required*
β_0	1	Orig.	0.960	(-3.736, 5.541)	210,604
		HC	0.889	(-3.787, 5.516)	4,197
		SZ	0.896	(0.697, 1.091)	3,897
		PS	0.898	(0.703, 1.096)	3,834
β_1	2	Orig.	1.955	(1.886, 2.025)	4,028
		HC	1.955	(1.887, 2.025)	3,867
		SZ	1.955	(1.885, 2.025)	3,897
		PS	1.955	(1.885, 2.024)	4,095
σ	1	Orig.	0.995	(0.861, 1.156)	3,865
		HC	0.994	(0.860, 1.153)	4,197
		SZ	0.994	(0.859, 1.156)	3,802
		PS	0.994	(0.861, 1.153)	3,865
σ_ε	10	Orig.	7.064	(4.551, 11.395)	5,123
		HC	7.050	(4.550, 11.436)	5,254
		SZ	*5.646*	*(3.630, 9.205)*	4,913
		PS	7.170	(4.582, 11.616)	4,996
$\bar{\varepsilon}$	0	Orig.	-0.063	(-4.634, 4.627)	198,036
		HC	0.007	(-4.621, 4.681)	4,302
		SZ	0	0	N/A
		PS	0	0	N/A

* See Table S1 for definition of total iterations required.

Section S1

The simulated data used to generate Figures 3-5 and Table 1 are provided here, on Github, and the simulated data are plotted in Figure S1. The data were simulated according to the model described by Equation (5) and as summarized in the legend of Figure 3. The data are in the "Data S1.RData" object and are loaded into R using the command:

```
> load("Data S1.RData")
```

This loads four variables: x, y1, y2, and y3. Each variable is of dimension 10 rows (observation) by 10 columns (group), such that each dataset contains 10 observations per 1 of 10 group levels.

The covariate (predictor variable) data are contained in x. Three different sets of response variables were simulated based on different values set for the among group variance (σ_ε) relative to the within group variance (σ, which was set to 1, $\sigma = 1$): (1) small σ_ε ($\sigma_\varepsilon = 0.1$), y1; (2) medium σ_ε ($\sigma_\varepsilon = 1$), y2; and, (3) large σ_ε ($\sigma_\varepsilon = 10$), y3.

The data are, as contained in Data S1:

x:

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	2.191	-2.321	4.384	-1.427	3.307	2.948	-3.854	-1.093	0.114	-4.030
2	4.105	1.439	-3.499	-3.166	4.811	-3.788	0.884	-3.354	3.440	-0.629
3	-4.030	-3.779	3.997	3.411	2.988	-0.357	-3.169	2.961	0.234	-3.167
4	-4.392	-0.116	1.038	4.701	2.264	-1.976	4.159	4.860	0.039	4.543
5	0.740	-0.864	-3.318	-3.291	-4.873	-0.795	0.865	0.912	4.510	4.950
6	-2.661	3.468	-2.265	-0.072	-0.888	3.341	0.058	2.409	-0.216	-3.548
7	3.293	-0.105	-1.428	-1.942	-2.435	-0.789	1.448	2.311	3.650	2.069
8	0.877	2.724	-4.039	-3.140	4.705	3.713	2.397	3.387	-3.514	3.061
9	4.611	-2.237	-3.910	-1.221	1.892	2.176	4.025	-2.459	2.258	-0.700
10	2.896	-2.686	-1.044	2.679	-1.386	0.173	-4.414	0.886	-4.243	0.574

y1:

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	7.073	-4.138	10.035	-3.441	7.372	5.871	-7.618	-0.623	2.602	-7.748
2	9.372	5.118	-6.118	-4.941	10.299	-6.373	3.407	-7.065	7.576	-0.596
3	-5.387	-6.372	7.742	7.186	6.561	2.575	-5.203	6.848	1.818	-5.453
4	-8.185	0.429	3.620	10.362	6.113	-3.615	8.018	11.440	-0.122	10.608
5	2.978	0.378	-4.781	-3.508	-9.010	-0.838	3.047	2.344	9.475	11.476
6	-3.812	7.600	-4.249	1.042	0.556	6.733	2.589	7.468	-0.948	-6.475
7	8.461	0.864	-2.129	-3.248	-4.757	0.620	3.908	5.359	8.646	4.552
8	2.726	5.144	-7.289	-6.740	11.402	8.970	4.671	7.191	-5.702	7.706
9	9.470	-4.620	-5.881	-1.566	4.712	6.479	8.234	-3.356	5.966	-0.311
10	6.281	-4.622	-0.453	6.739	-2.945	1.789	-7.050	3.870	-5.291	-1.045

y2:

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	5.103	-2.889	7.370	-0.392	7.712	6.968	-5.464	-2.368	1.850	-6.812
2	8.770	3.091	-6.191	-5.107	10.367	-4.795	3.885	-6.157	9.768	-1.121
3	-8.684	-8.067	9.624	9.682	4.976	-0.101	-5.146	4.955	0.110	-4.237
4	-8.178	1.962	1.069	12.597	4.136	-2.019	9.953	10.362	0.276	9.043
5	1.101	-2.729	-5.932	-4.495	-9.305	-0.506	3.724	1.922	11.848	10.803
6	-5.099	8.530	-4.359	2.179	-1.301	8.889	1.395	3.638	-0.423	-7.303
7	6.123	1.770	-3.590	-1.952	-3.848	1.882	4.064	5.371	8.936	4.981
8	1.653	6.588	-8.367	-4.294	9.660	7.054	7.152	7.870	-6.166	8.721
9	10.038	-2.664	-6.742	-1.550	4.381	7.734	9.325	-3.607	5.918	0.870

```
10      5.615   -3.364  -2.261  6.468   -1.261  1.068   -7.367  0.417   -7.306  0.470

y3:
        V1      V2      V3      V4      V5      V6      V7      V8      V9      V10
1       0.850   -5.057  4.555   -3.017  8.297   15.552  -0.188  -6.600  -8.382  4.135
2       2.619   4.112   -8.714  -6.974  11.490  0.372   8.341   -12.416 -1.718  10.134
3       -14.417 -6.901  3.951   6.717   8.199   5.013   1.771   0.708   -8.838  6.069
4       -12.870 1.872   -0.859  8.748   6.091   2.233   15.536  4.173   -10.005 22.162
5       -4.442  0.015   -10.279 -7.789  -6.524  6.535   10.281  -4.077  -0.247  22.823
6       -9.511  6.742   -7.502  0.876   1.940   13.924  6.072   0.382   -8.505  7.405
7       0.314   -1.766  -7.176  -3.889  -1.814  6.627   10.858  0.810   -1.338  18.609
8       -1.464  6.422   -10.345 -5.867  11.573  14.072  13.241  1.525   -17.126 19.085
9       5.297   -2.318  -9.910  -2.153  5.902   11.325  15.255  -9.587  -5.755  13.600
10      0.542   -3.805  -6.929  3.757   0.661   7.105   -2.466  -2.002  -17.147 14.330
```

The data were simulated in R v3.5.1 using the following code:

```
# Generate synthetic datasets associated with three different
# random effects variance terms.
set.seed(19275)
y1 = array(data = NA, dim=c(10,10))
y2 = array(data = NA, dim=c(10,10))
y3 = array(data = NA, dim=c(10,10))
x = array(data = NA, dim=c(10,10))
mu = array(data = NA, dim=c(10,10))
# Small random effects variance, sig.eps = 0.1
eps1 = rnorm(n=10,mean=0,sd=0.1)
eps1 = eps1 - mean(eps1)
# Medium random effects variance, sig.eps = 1
eps2 = rnorm(n=10,mean=0,sd=1)
eps2 = eps2 - mean(eps2)
# Large random effects variance, sig.eps = 10
eps3 = rnorm(n=10,mean=0,sd=10)
eps3 = eps3 - mean(eps3)

for(j in 1:10){
  for(r in 1:10){
    x[r,j] = runif(1,min=-5,max=5)
    # Mean with true intercept = 1 and true slope = 2:
    # mu[r,j] = 1 + 2*x[r,j] + eps[j]
    y1[r,j] = rnorm(n=1, mean = 1+2*x[r,j] + eps1[j], sd = 1)
    y2[r,j] = rnorm(n=1, mean = 1+2*x[r,j] + eps2[j], sd = 1)
    y3[r,j] = rnorm(n=1, mean = 1+2*x[r,j] + eps3[j], sd = 1)
  }
}

save(x,y1,y2,y3,file = "Data S1.RData")
```

Section S2

The model code for implementing the random effect regression summarized in Equation 5 in the main text is available here, on Github, as "Code A2.R," which includes an R script for loading and preparing data, for specifying the Bayesian models, and for implementing the models in JAGS. The code does not implement any reparameterization or computational solutions to solving the non-identifiabilty of the overall intercept and random effects terms, but it does include options for specifying a relatively non-informative *Gamma* prior or a weakly informative folded-*Cauchy* prior for the random effects variance-related parameter. The model is applied to the three datasets contained in Section S1 (via the "Data S1.RData" object posted on Github). The results from this model are illustrated in Figures 3, 4, and 5A, and in Table 1 (Approach = "Orig.").

The models were implemented in R v3.5.1, via RStudio v1.1.456, and used JAGS 4.3.0.

Code A2 contains code for:

1) Loading the data (in "Data S1.RData")
2) Specifying the random effects Bayesian regression (definition of mod.string)
3) Updating the JAGS model via jags.model for each of the 3 datasets
4) Computing the Raftery diagnostics to determine the number of MCMC iterations that are required for posterior summary statistics, and to evaluate convergence of the MCMC chains.
5) Updating the JAGS model object via coda.samples to obtain a sufficient number of samples (iterations) from the posterior
6) Visualizing the MCMC samples via mcmcplot (e.g., history plots, etc.)
7) Computing posterior summary statistics for model parameters given the converged samples.

The model, as specified in model string (mod.string) will be of most interest to readers, and should be contrasted against the models in Code A3, also on Github, which shows how to recode the model to implementing the identifiability solutions.

The random effect Bayesian regression code for implementation in JAGS or OpenBUGS (as provided in Code A2) is:

```
model{
  for(j in 1:10){
    for(r in 1:10){
      # Likelihood of (stochastic) data:
      y[r,j] ~ dnorm(mu[r,j],tau)
      # Mean model for a random effects regression:
      mu[r,j] <- b0 + b1*x[r,j] + eps[j]
    }
    # Zero-centered hierarchical prior for random effects:
    eps[j] ~ dnorm(0,tau.eps)
  }
  # Relatively non-informative priors for root nodes:
  b0 ~ dnorm(0,1E-6)
  b1 ~ dnorm(0,1E-6)
```

```
  tau ~ dgamma(0.1, 0.1)
  sig <- 1/sqrt(tau)

  # Conjugate gamma prior for precision:
  tau.eps ~ dgamma(0.1, 0.1)
  sig.eps <- 1/sqrt(tau.eps)

  # Folded Cauchy(0,A^2,df=1) prior for standard deviation:
  # alpha ~ dnorm(0,1)
  # tau.temp ~ dgamma(a,b)
  # sig.temp <- 1/sqrt(tau.temp)
  # sig.eps <- abs(alpha)*sig.temp
  # tau.eps <- pow(sig.eps,-2)

  # Compute mean of random effects (to illustrate non-identifiability)
  mean.eps <- mean(eps[])
}
```

Section S3

The model code for implementing Solutions 1 (hierarchical centering, HC), 2 (sum-to-zero, SZ), and 4 (post-sweeping, PS) with the random effect regression summarized in Equation 5 in the main text is available here, on Github, as "Code A3.R," which includes an R script for loading and preparing data, for specifying the Bayesian models, and for implementing the models in JAGS. The models are applied to the three datasets contained in Data S1. Results from these model are illustrated in Figures 5B (HC), 5C (SZ), 5D (PS), and in Table 1.

The models were implemented in R version 3.5.1, via RStudio version 1.1.456, and used JAGS version 4.3.0.

Code A3 contains code for:

1) Loading the data (in "Data S1.RData")
2) Specifying the random effects Bayesian regression using the three different solutions (definition of mod.string.hc, mod.string.sz, and mod.string.ps)
3) Initializing the JAGS models via jags.model and updating the jags.model objects using coda.samples for each of the three model implementation versions.
4) Visualizing the MCMC samples via mcmcplot (e.g., history plots, etc.)
5) Computing the Raftery diagnostics to determine the number of MCMC iterations that are required for posterior summary statistics, and to evaluate convergence of the MCMC chains.
6) Computing posterior summary statistics for model parameters given the converged samples.

The models, as specified in the model strings (mod.string.XX), will be of most interest to readers, and should be contrasted against the model in Code A2, which represents the original, non-identifiable implementation.

The code for implementing the models, for each solution, in JAGS or OpenBUGS (as provided in Code A3) is:

```
############################################################
### JAGS (OpenBUGS) code for hierarchical centering (hc).
############################################################

model{
  for(j in 1:10){
    for(r in 1:10){
      y[r,j] ~ dnorm(mu[r,j],tau)
      # Mean model differs from the original, non-identifiable model:
      mu[r,j] <- b0.eps[j] + b1*x[r,j]
    }
    # Hierarchical prior for random effects, centered on overall intercept:
    b0.eps[j] ~ dnorm(b0,tau.eps)
    # If interested in monitoring the random effects, can
    # compute as a deviation term (but, not necessary)
    eps[j] <- b0.eps[j] - b0
  }
```

```
  # Relatively non-informative priors for root nodes:
  b0 ~ dnorm(0,1E-6)
  b1 ~ dnorm(0,1E-6)
  tau ~ dgamma(0.1, 0.1)
  sig <- 1/sqrt(tau)
  # Mean of random effects (not necessary):
  mean.eps <- mean(eps[])

  # Conjugate gamma prior for random effects precision:
  tau.eps ~ dgamma(0.1, 0.1)
  sig.eps <- 1/sqrt(tau.eps)

  # Folded Cauchy(0,A^2,df=1) prior for standard deviation:
  # alpha ~ dnorm(0,1)
  # tau.temp ~ dgamma(a,b)
  # sig.temp <- 1/sqrt(tau.temp)
  # sig.eps <- abs(alpha)*sig.temp
  # tau.eps <- pow(sig.eps,-2)
}


############################################################
### JAGS (OpenBUGS) code with sum-to-zero constraints (sz).
############################################################

model{
    for(j in 1:10){
      for(r in 1:10){
        y[r,j] ~ dnorm(mu[r,j],tau)
        # Mean model the same as the original, non-identifiable model:
        mu[r,j] <- b0 + b1*x[r,j] + eps[j]
      }
    }

    # Hierarchical prior for all but one random effect:
    for(j in 1:9){
      eps[j] ~ dnorm(0,tau.eps)
    }
    # Remaining random effect = minus sum of other random effects:
    eps[10] <- -sum(eps[1:9])

    # Relatively non-informative priors for root nodes:
    b0 ~ dnorm(0,1E-6)
    b1 ~ dnorm(0,1E-6)
    tau ~ dgamma(0.1, 0.1)
    sig <- 1/sqrt(tau)
    # Mean of the random effects (not necessary), which will be exactly zero:
    mean.eps <- mean(eps[])

    # Conjugate gamma prior for random effects precision:
    tau.eps ~ dgamma(0.1, 0.1)
    sig.eps <- 1/sqrt(tau.eps)

    # Folded Cauchy(0,A^2,df=1) prior for standard deviation:
    # alpha ~ dnorm(0,1)
    # tau.temp ~ dgamma(a,b)
    # sig.temp <- 1/sqrt(tau.temp)
```

```
    # sig.eps <- abs(alpha)*sig.temp
    # tau.eps <- pow(sig.eps,-2)
  }


############################################################
### JAGS (OpenBUGS) code for post-sweeping of random
### effects (ps).
############################################################

model{
    for(j in 1:10){
      for(r in 1:10){
        y[r,j] ~ dnorm(mu[r,j],tau)
        # Mean model same as original, non-identifiable model:
        mu[r,j] <- b0 + b1*x[r,j] + eps[j]
      }
      # Zero-centered hierarchical prior for random effects, just
      # as used in the original model (do not monitor or report these):
      eps[j] ~ dnorm(0,tau.eps)
      # Compute identifiable random effects (monitor and report these,
      # if desired):
      eps.star[j] <- eps[j] - mean.eps
    }
    # Prior for non-identifiable intercept (don't monitor or report)
    b0 ~ dnorm(0,1E-6)
    # Compute identifiable intercept (monitor and report this):
    b0.star <- b0 + mean.eps
    # Relatively non-informative priors for other root nodes:
    b1 ~ dnorm(0,1E-6)
    tau ~ dgamma(0.1, 0.1)
    sig <- 1/sqrt(tau)
    # Mean of non-identifiable random effects (required)
    mean.eps <- mean(eps[])
    # Mean of identifiable random effects (not required) (will be zero).
    mean.eps.star <- mean(eps.star[])

    # Conjugate gamma prior for random effects precision:
    tau.eps ~ dgamma(0.1, 0.1)
    sig.eps <- 1/sqrt(tau.eps)

    # Folded Cauchy(0,A^2,df=1) prior for standard deviation:
    # alpha ~ dnorm(0,1)
    # tau.temp ~ dgamma(a,b)
    # sig.temp <- 1/sqrt(tau.temp)
    # sig.eps <- abs(alpha)*sig.temp
    # tau.eps <- pow(sig.eps,-2)
  }
```

Section S4

The model code for implementing Solutions 1 (hierarchical centering, HC), 2 (sum-to-zero, SZ), and 4 (post-sweeping, PS) with the nested random effects regression summarized in Equation 11 in the main text is available here, on Github, as "Code A4.R," which includes an R script for loading and preparing data, for specifying the Bayesian models, and for implementing the models in JAGS. The models are applied a pseudo dataset that is contained within the Code A4 R script file.

The models were implemented in R version 3.5.1, via RStudio version 1.1.456, and used JAGS version 4.3.0.

Code A4 contains code for:

1) Defining the data (data.eq11 as a list)
2) Specifying the nested random effects Bayesian regression model using the three different solutions (definition of mod.string.hc, mod.string.sz, and mod.string.ps)
3) Initializing the JAGS models via jags.model and updating the jags.model objects using coda.samples for each of the three model implementation versions.
4) Visualizing the MCMC samples via mcmcplot (e.g., history plots, etc.)
5) Computing the Raftery diagnostics to determine the number of MCMC iterations that are required for posterior summary statistics, and to evaluate convergence of the MCMC chains.
6) Computing posterior summary statistics for model parameters given the converged samples.

The code for implementing the models, for each solution, in JAGS or OpenBUGS (as provided in Code A4) is given below. Note that relatively non-informative *Gamma* priors are used for the random effects precision terms, but the model code can be easily modified to include other priors, such as the folded-Cauchy priors from Code A2 and Code A3.

```
############################################################
### JAGS (OpenBUGS) code for hierarchical centering (hc).
############################################################

model{
  for(i in 1:N){
    y[i] ~ dnorm(mu[i],tau)
    # Mean model differs from the original specification in Eq 11 to reflect
    # hierarchical centering:
    mu[i] <- b0.eps[plot[i],wshed[i]] + b1*x[i]
  }
  # Hierarchical priors for random effects, centered on overall intercept
  # (b0):
  for(s in 1:Nwater){
    for(p in 1:Nplots){
      # Plot within watershed effects:
      b0.eps[p,s] ~ dnorm(b0.gamma[s], tau.eps)
    }
```

```
    # Watershed effects:
    b0.gamma[s] ~ dnorm(b0, tau.gamma)
    }

  # Relatively non-informative priors for root nodes:
  b0 ~ dnorm(0,1E-6)
  b1 ~ dnorm(0,1E-6)
  tau ~ dgamma(0.1, 0.1)
  tau.eps ~ dgamma(0.1, 0.1)
  tau.gamma ~ dgamma(0.1, 0.1)
  sig <- 1/sqrt(tau)
  sig.eps <- 1/sqrt(tau.eps)
  sig.gamma <- 1/sqrt(tau.gamma)
}


############################################################
### JAGS (OpenBUGS) code with sum-to-zero constraints (sz).
############################################################

model{
  for(i in 1:N){
    y[i] ~ dnorm(mu[i],tau)
    # Mean model as per the original specification in Eq 11:
    mu[i] <- b0 + b1*x[i] + eps[plot[i], wshed[i]] + gamma[wshed[i]]
  }
  # Zero-centered priors for random effects with sum-to-zero constraint
  # Plot within watershed effects:
  for(s in 1:Nwater){
    for(p in 1:(Nplots-1)){
      eps[p,s] ~ dnorm(0, tau.eps)
    }
    # Sum-to-zero occurs for plots within each watershed:
    eps[Nplots,s] <- -sum(eps[1:(Nplots-1),s])
  }
  # Watershed random effects:
  for(s in 1:(Nwater-1)){
    gamma[s] ~ dnorm(0,tau.gamma)
  }
  # Sum-to-zero constraint:
  gamma[Nwater] <- -sum(gamma[1:(Nwater-1)])

  # Relatively non-informative priors for root nodes:
  b0 ~ dnorm(0,1E-6)
  b1 ~ dnorm(0,1E-6)
  tau ~ dgamma(0.1, 0.1)
  tau.eps ~ dgamma(0.1, 0.1)
  tau.gamma ~ dgamma(0.1, 0.1)
  sig <- 1/sqrt(tau)
  sig.eps <- 1/sqrt(tau.eps)
  sig.gamma <- 1/sqrt(tau.gamma)
}
```

```
#################################################################
### JAGS (OpenBUGS) code for post-sweeping of random effects (ps).
#################################################################

model{
    for(i in 1:N){
      y[i] ~ dnorm(mu[i],tau)
      # Mean model as per the original specification in Eq 11:
      mu[i] <- b0 + b1*x[i] + eps[plot[i], wshed[i]] + gamma[wshed[i]]
    }
    # Zero-centered priors for non-identifiable random effects
    # Plot within watershed effects:
    for(s in 1:Nwater){
      for(p in 1:Nplots){
        # Non-identifiable random effect (don't monitor):
        eps[p,s] ~ dnorm(0, tau.eps)
        # Identifiable plot random effect (monitor this):
        eps.star[p,s] <- eps[p,s] - ave.eps[s]
      }
      # Mean plot-level random effects within each watershed:
      ave.eps[s] <- mean(eps[,s])
    }
    # Watershed random effects:
    for(s in 1:Nwater){
      # Non-identifiable random effect:
      gamma[s] ~ dnorm(0,tau.gamma)
      # Identifiable random effect (monitor this):
      gamma.star[s] <- gamma[s] + ave.eps[s] - ave.gamma - ave.ave.eps
    }
    # Mean watershed random effect:
    ave.gamma <- mean(gamma[])
    # Mean overall plot random effect:
    ave.ave.eps <- mean(ave.eps[])

    # Relatively non-informative priors for root nodes:
    # Non-identifiable intercept
    b0 ~ dnorm(0,1E-6)
    # Identifiable intercept (monitor this)
    b0.star <- b0 + ave.gamma + ave.ave.eps
    b1 ~ dnorm(0,1E-6)
    tau ~ dgamma(0.1, 0.1)
    tau.eps ~ dgamma(0.1, 0.1)
    tau.gamma ~ dgamma(0.1, 0.1)
    sig <- 1/sqrt(tau)
    sig.eps <- 1/sqrt(tau.eps)
    sig.gamma <- 1/sqrt(tau.gamma)
  }
```

Section S5

The model code for implementing Solutions 2 (sum-to-zero, SZ) and 4 (post-sweeping, PS) with the Bayesian regression involving crossed random effects, as summarized in Equation 14 in the main text is available here, on Github, as "Code A5.R," which includes an R script for loading and preparing data, for specifying the Bayesian models, and for implementing the models in JAGS. The models are applied a pseudo dataset that is contained within the Code A5 R script file.

The models were implemented in R version 3.5.1, via RStudio version 1.1.456, and used JAGS version 4.3.0.

Code A5 contains code for:

1) Defining the data (data.eq14 as a list)
2) Specifying the Bayesian regression model with cross random effects using the two different solutions (definition of mod.string.sz and mod.string.ps)
3) Initializing the JAGS models via jags.model and updating the jags.model objects using coda.samples for each of the three model implementation versions.
4) Visualizing the MCMC samples via mcmcplot (e.g., history plots, etc.)
5) Computing the Raftery diagnostics to determine the number of MCMC iterations that are required for posterior summary statistics, and to evaluate convergence of the MCMC chains.
6) Computing posterior summary statistics for model parameters given the converged samples.

The code for implementing the models, for each solution, in JAGS or OpenBUGS (as provided in Code A5) is given below. Note that relatively non-informative *Gamma* priors are used for the random effects precision terms, but the model code can be easily modified to include other priors, such as the folded-Cauchy priors from Code A2 and Code A3.

```
############################################################
### JAGS (OpenBUGS) code with sum-to-zero constraints (sz).
### Code uses the indexing variables of plot and date.
############################################################

model{
    for(i in 1:N){
      y[i] ~ dnorm(mu[i],tau)
      # Mean model as per the original specification in Eq 14:
      mu[i] <- b0 + b1*x[i] + eps[plot[i]] + gamma[date[i]]
    }
    # Zero-centered priors for random effects with sum-to-zero constraints
    # Plot random effects:
    for(p in 1:(Nplots-1)){
      eps[p] ~ dnorm(0, tau.eps)
    }
    # Sum-to-zero constraint for plots:
    eps[Nplots] <- -sum(eps[1:(Nplots-1)])
```

```
    # Date random effects:
    for(d in 1:(Ndates-1)){
      gamma[d] ~ dnorm(0,tau.gamma)
    }
    # Sum-to-zero constraint for dates:
    gamma[Ndates] <- -sum(gamma[1:(Ndates-1)])

    # Relatively non-informative priors for root nodes:
    b0 ~ dnorm(0,1E-6)
    b1 ~ dnorm(0,1E-6)
    tau ~ dgamma(0.1, 0.1)
    tau.eps ~ dgamma(0.1, 0.1)
    tau.gamma ~ dgamma(0.1, 0.1)
    sig <- 1/sqrt(tau)
    sig.eps <- 1/sqrt(tau.eps)
    sig.gamma <- 1/sqrt(tau.gamma)
  }


############################################################
### JAGS (OpenBUGS) code for post-sweeping of random effects.
############################################################

model{
    for(i in 1:N){
      y[i] ~ dnorm(mu[i],tau)
      # Mean model as per the original specification in Eq 11:
      mu[i] <- b0 + b1*x[i] + eps[plot[i]] + gamma[date[i]]
    }
    # Zero-centered priors for non-identifiable random effects
    # Plot random effects:
    for(p in 1:Nplots){
      # Non-identifiable random effect (don't monitor)
      eps[p] ~ dnorm(0, tau.eps)
      # Identifiable random effect (monitor this)
      eps.star[p] <- eps[p] - ave.eps
    }
    # Mean plot-level random effect:
    ave.eps <- mean(eps[])

    # Date random effects:
    for(d in 1:Ndates){
      # Non-identifiable random effect
      gamma[d] ~ dnorm(0,tau.gamma)
      # Identifiable random effect (monitor this)
      gamma.star[d] <- gamma[d] - ave.gamma
    }
    # Mean date random effect:
    ave.gamma <- mean(gamma[])

    # Relatively non-informative priors for root nodes:
    # Non-identifiable intercept
    b0 ~ dnorm(0,1E-6)
    # Identifiable intercept (monitor this)
    b0.star <- b0 + ave.gamma + ave.eps
    b1 ~ dnorm(0,1E-6)
```

```
  tau ~ dgamma(0.1, 0.1)
  tau.eps ~ dgamma(0.1, 0.1)
  tau.gamma ~ dgamma(0.1, 0.1)
  sig <- 1/sqrt(tau)
  sig.eps <- 1/sqrt(tau.eps)
  sig.gamma <- 1/sqrt(tau.gamma)
}
```