Parameter recovery using remotely sensed variables

Supplementary Materials

Jonathan Proctor ⓡ Tamma Carleton ⓡ Sandy Sum

Proctor: Harvard University, jproctor1@fas.harvard.edu. Carleton: University of California, Santa Barbara, tcarleton@ucsb.edu. Sum: University of California, Santa Barbara, sandysum@ucsb.edu.

A Supplementary Tables

Variable	Units	Native resolution
Forest cover	% forest cover	$\sim$ 30 m $\times$ 30 m
Elevation	meters	$\sim$ 611.5 m $\times$ 611.5 m
Population density	log(people per sq. km)	$\sim$ 1 km $\times$ 1 km
Nighttime lights	log(nanoWatts /cm^2/sr)	$\sim$ 500 m $\times$ 500 m
Income	USD per household	census block group
Road length	meters	polyline
Housing price	USD per sq. ft.	geocoded point data

Table A.1: Description of variables obtained from Rolf et al. (2021). Table lists the variables that form the benchmark dataset leveraged throughout this analysis. Each variable is measured both using remote sensing and using ground truth observations. Most of the ground truth data are based on measurements from 2010-2015. Ground-truth values (labels) for these tasks are assembled from publicly available data, and all remotely sensed predictions are made available by the authors at https://codeocean.com/capsule/6456296/tree/v2. More details on the original data are described in Rolf et al. (2021).

B Supplementary Figures

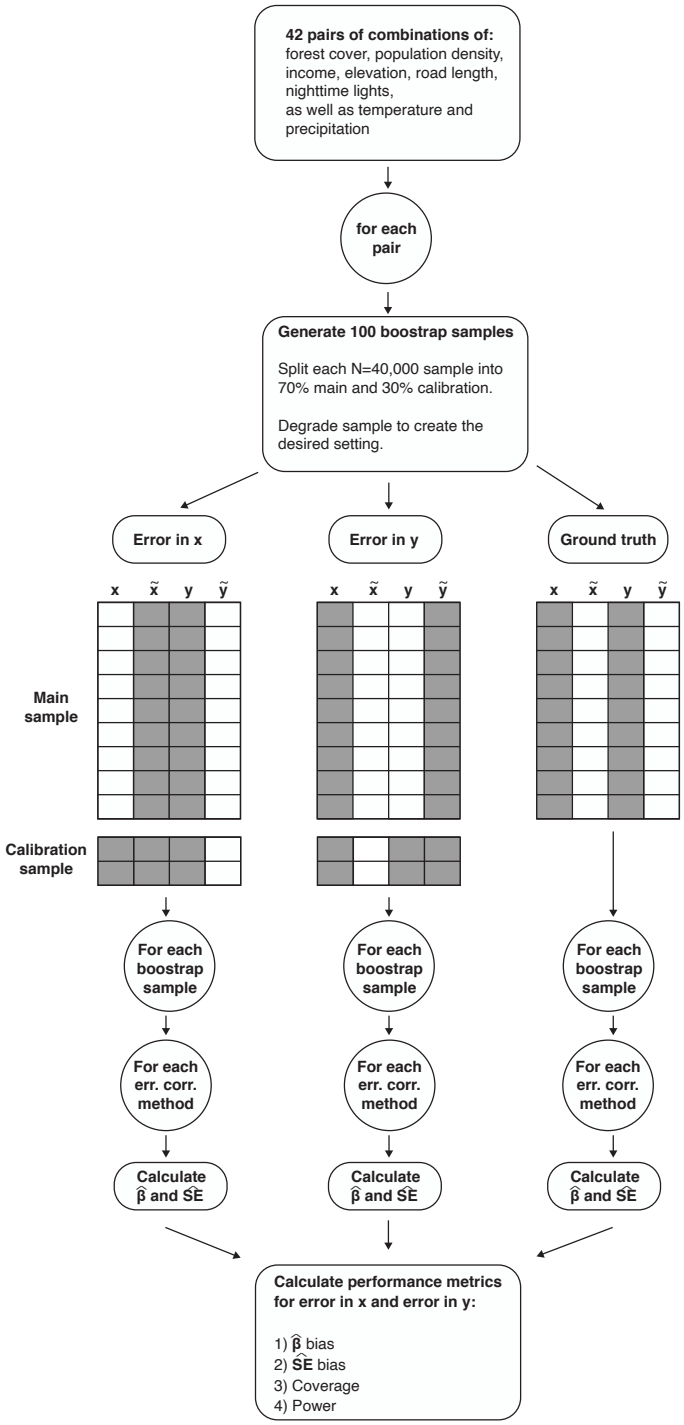

Figure B.1: Experimental design used to evaluate the impact of remotely sensed variables on downstream regression analyses. Schematic outlines experiments used to evaluate the bias, coverage, and power implications of measurement error introduced into regression analysis by remotely sensed variables.

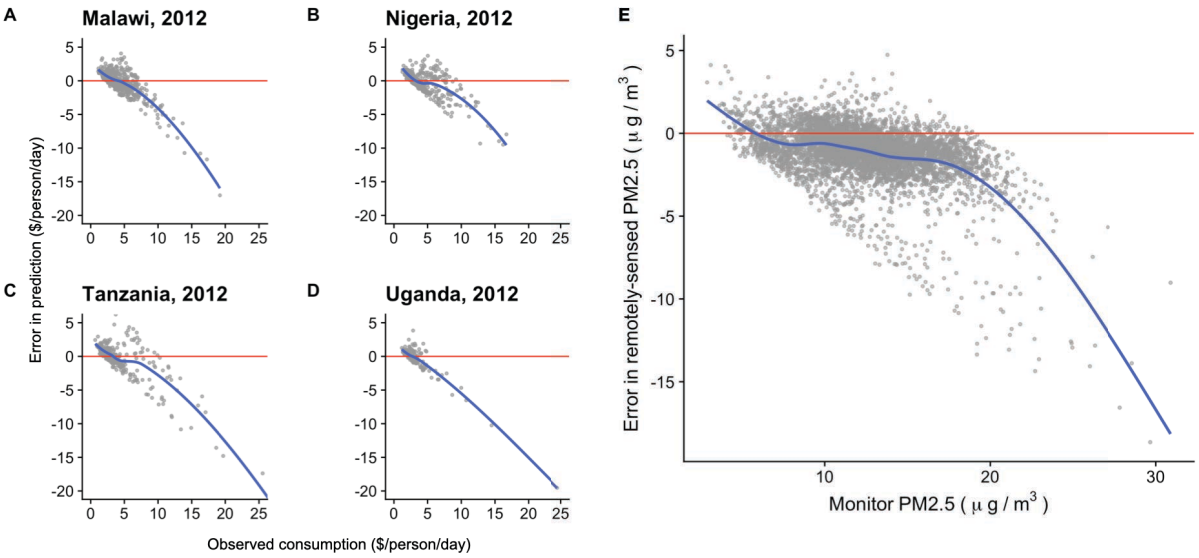

Figure B.2: Remotely sensed consumption from Jean et al. (2016) and remotely sensed $PM_{2.5}$ from Van Donkelaar et al. (2021) display mean-reverting measurement error. Panels A-D show errors in consumption predicted from satellite images (measured in \$/person/day) by Jean et al. (2016) plotted against corresponding ground truth values of consumption for four different countries in Africa. Panel E shows errors in remotely sensed county-level $PM_{2.5}$ (measured in $\mu g/m^3$) from Van Donkelaar et al. (2021) plotted against U.S. EPA monitor-based $PM_{2.5}$ over the years 2001-2007. Downward slopes in all panels indicate mean-reverting measurement error (Bound and Krueger, 1991), consistent with $\lambda < 1$ in Equation 5.

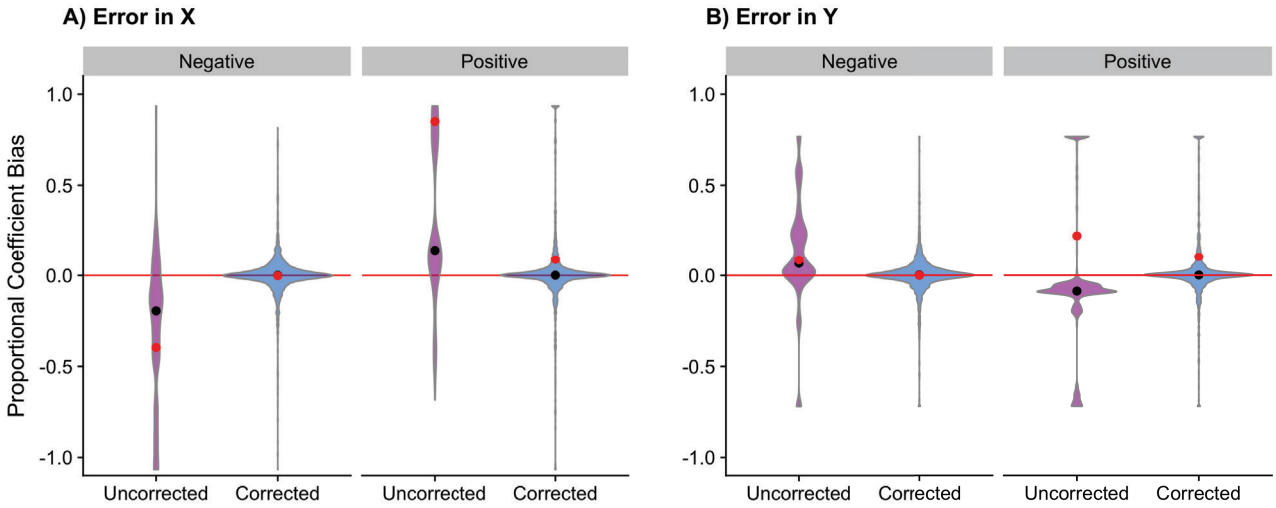

Figure B.3: Remotely sensed variables tend to inflate coefficients in error-in-X models and attenuate coefficients in error-in-Y models. Figure decomposes results from the top row of Figure 4 into coefficient biases in regression models with true slope coefficients less than zero (labeled "Negative") versus those with true slope coefficients greater than zero (labeled "Positive"). Here, proportional bias is signed and is computed as $\frac{\hat{\beta}-\hat{\beta}}{\hat{\beta}}$, following notation from Section 3. Using this definition, attenuation bias would appear as positive bias for true negative coefficients (first column of each panel) and as negative bias for true positive coefficients (second column of each panel). While there is large heterogeneity displayed, uncorrected error-in-X models tend to exhibit coefficient inflation, while uncorrected error-in-Y models tend to exhibit attenuation (shown in purple). In contrast, coefficients corrected using multiple imputation (shown in blue) display minimal bias and exhibit no systematic inflation or attenuation of coefficients. Data for violin plots has been winsorized at the 2.5% and 97.5% levels to cap outliers for visual display purposes only. The unwinsorized mean is indicated by the red circle while the median is indicated by the black circle.

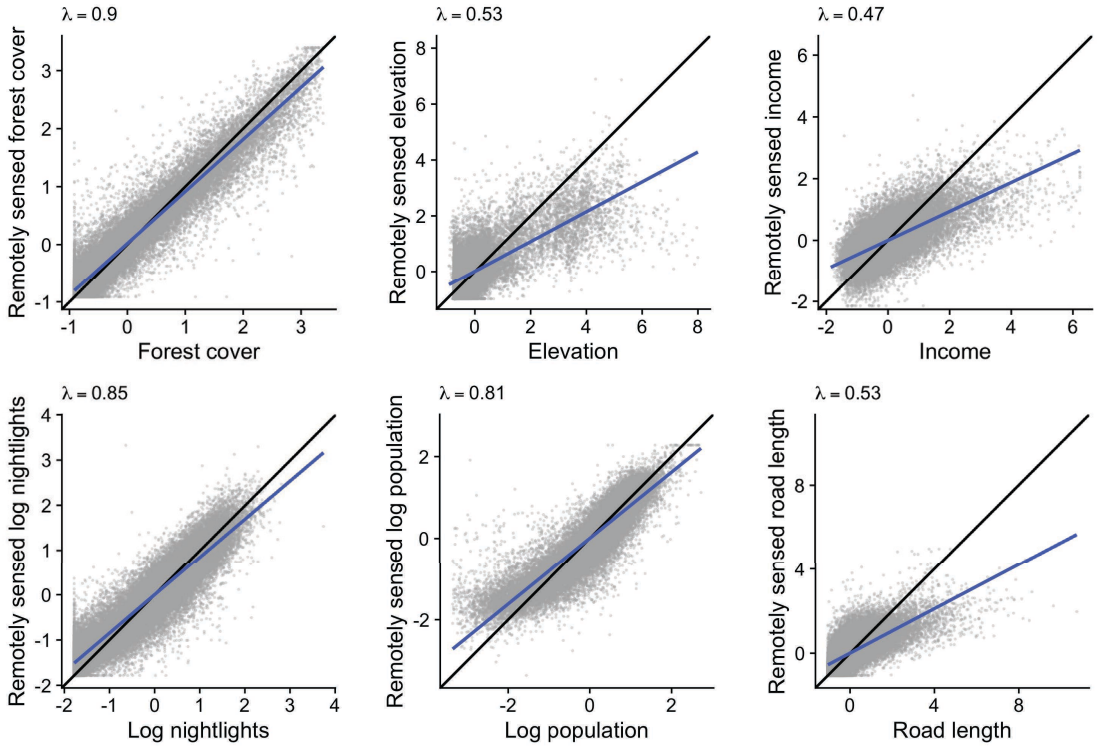

Figure B.4: Remotely sensed predictions of six diverse variables exhibit mean-reverting measurement error. Figure shows the relationship between remotely sensed values ($y-$axis) and ground truth values (x-axis) for all six variables used throughout the analysis and obtained from Rolf et al. (2021). Slope coefficients λ for each subplot correspond to the Equation 5 and are indicated visually by the blue line. The 45 degree line is indicated in black. All estimated values of λ are less than unity, demonstrating that these remotely sensed predictions exhibit mean-reverting measurement error, which contributes to bias in both error-in-X and error-in-Y regression models (see Equation 6).

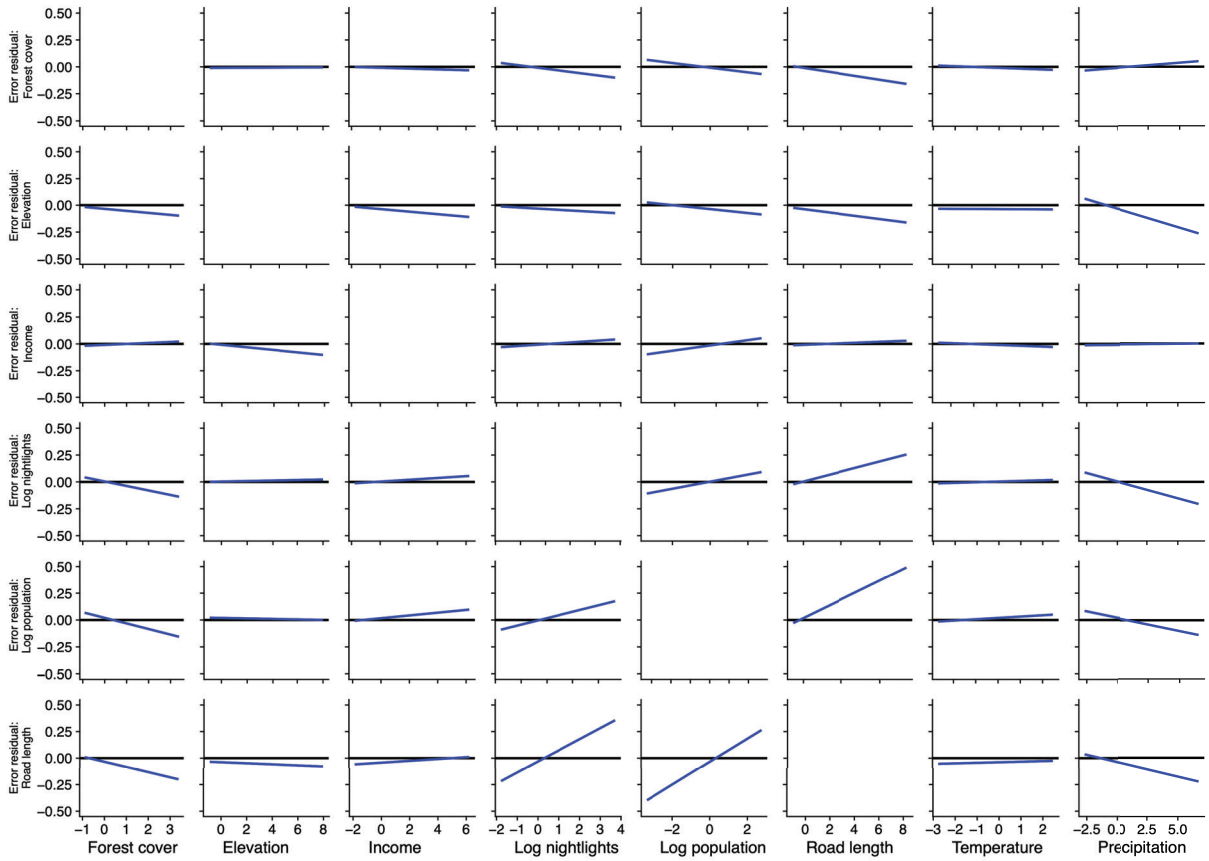

Figure B.5: Relationships between error residuals in one variable and levels of another variable show differential measurement error for many variable pair combinations. Residuals from the linear measurement error model in Equation 5 are plotted on the y-axis against ground truth values of all other variables on the x-axis. Off-diagonal boxes show the covariance between a variable's remotely sensed residual errors and the true value of *another* variable. Any nonzero correlations indicate the presence of differential measurement error in remotely sensed variables (i.e., non-zero values of σ_{yu} or σ_{xu} in Equation 5). Diagonals are omitted as they provide information about λ (shown in Figure B.4), not information about differential error.

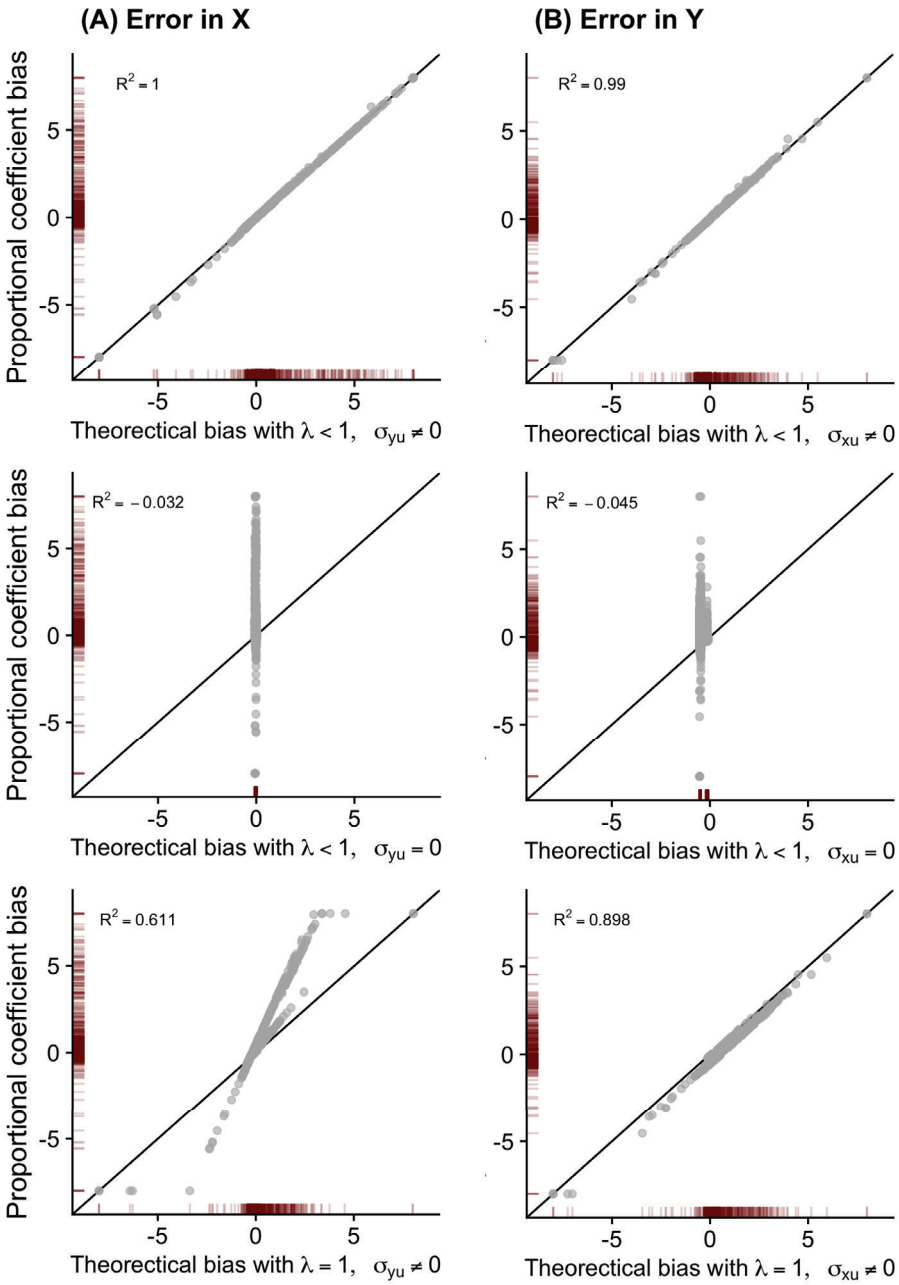

Figure B.6: Decomposing the source of bias introduced by remotely sensed measurements in downstream regression analyses. As detailed in Section 4.2, both mean reverting measurement error and differential measurement error can contribute to bias in error-in-X and error-in-Y regression models. This figure decomposes overall bias to show that differential measurement error is the most important factor in explaining the coefficient biases we recover. Each row plots observed coefficient bias (y-axis) against the theoretically predicted bias that would arise under alternative assumptions about measurement error structure (x-axis). The bottom row assumes no mean reversion, but allows for the presence of differential measurement error. The middle row assumes non-differential measurement error, but allows for mean reversion. The top row is the most general, and allows for both mean reversion and differential measurement error. This top row shows that together, these two forces explain all of the observed coefficient biases we recover across our diverse regression models.

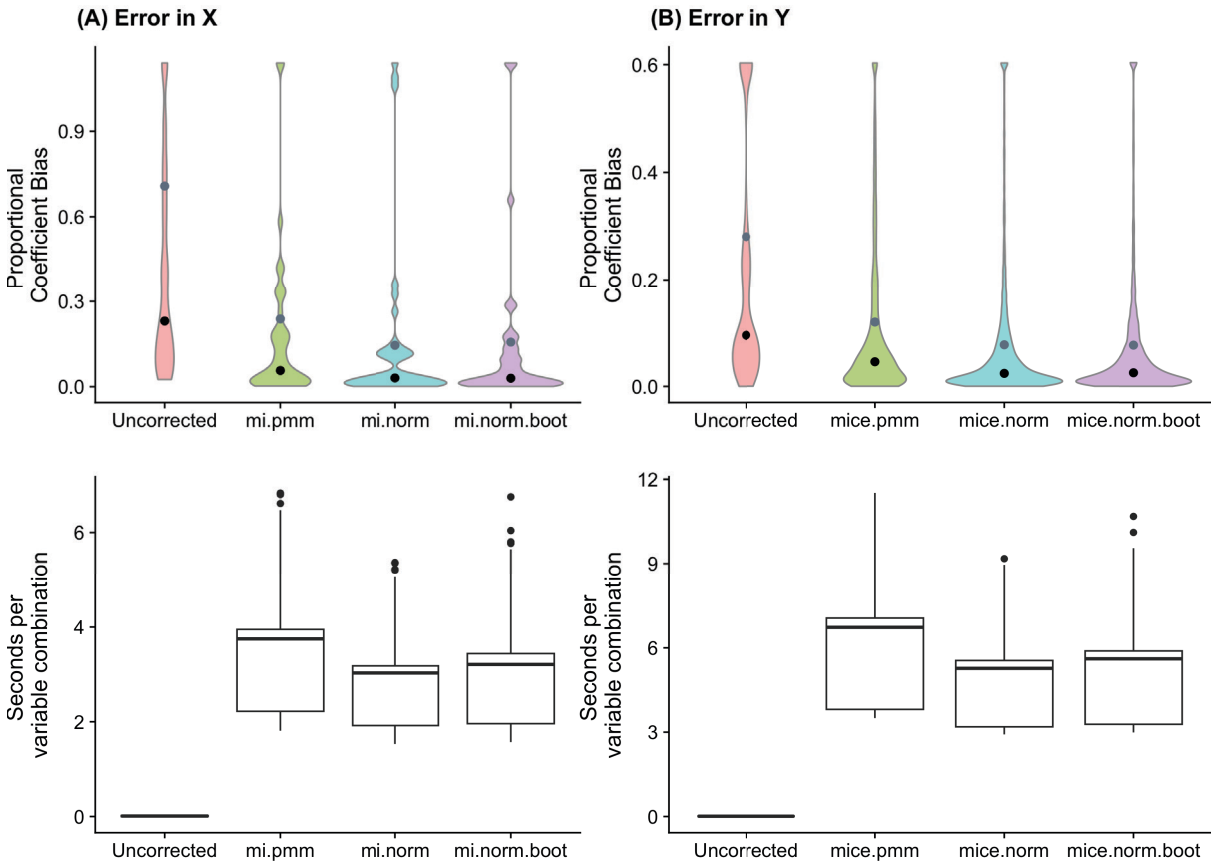

Figure B.7: Proportional coefficient bias and mean computation time for commonly used multiple imputation methods. Figure shows performance and compute time for the three most commonly used multiple imputation methods implemented in the `mice` package in `R`. The top row shows proportional coefficient bias while the bottom row shows the relative time taken to perform the bias correction for a single variable ($N = 40,000$). Results indicate that the Bayesian linear regression specification for multiple imputation (labeled "mi.norm"), which we use throughout our analysis, out-performs predictive mean matching (labeled "mi.pmm") and bootstrap linear regression (labeled "mi.norm.boot"), while being computationally the least demanding. Results shown throughout the main text are from error correction method "mi.norm".

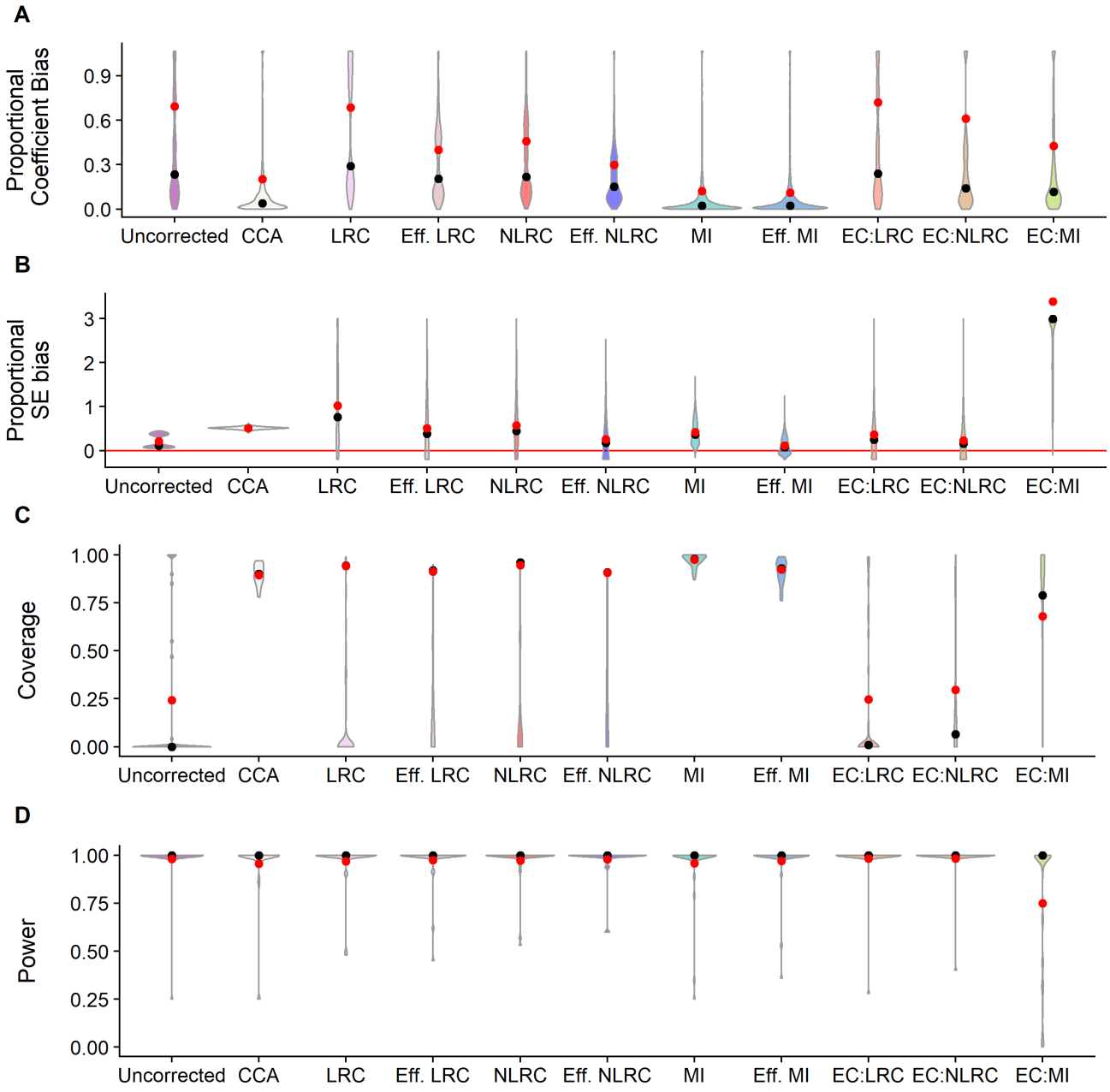

Figure B.8: Performance of alternative error correction models in the error-in-X case. Figure shows proportional coefficient bias, proportional standard error bias, coverage, and power across all regression models tested for each of nine error correction models, one data subsampling approach, and the uncorrected regression approach. Data for violin plots has been winsorized at the 2.5% and 97.5% to cap outliers for visual display purposes only. The unwinsorized mean is indicated by the red circles and the corresponding median is in black. Error correction approaches are detailed in Section C.4.

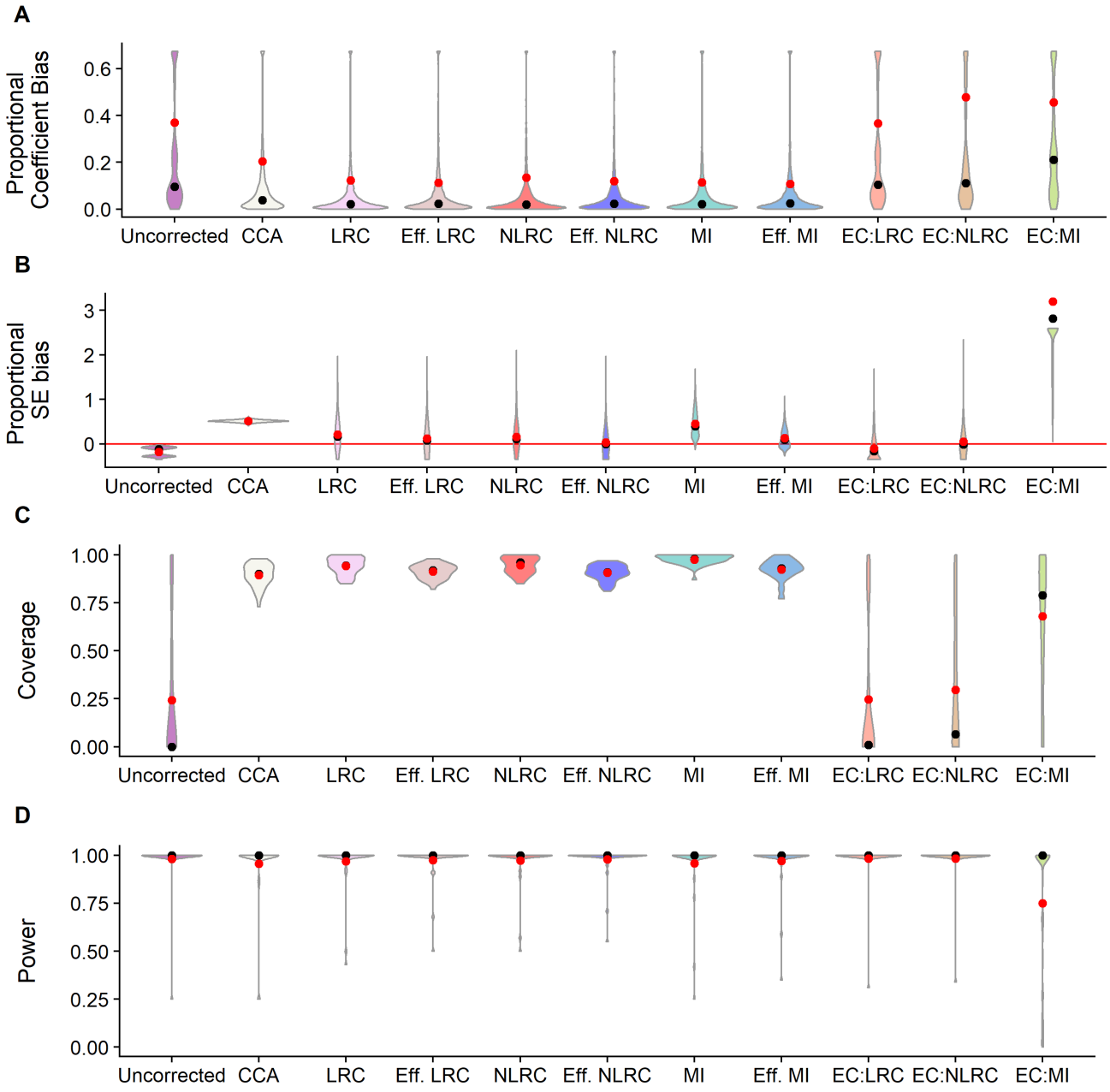

Figure B.9: Performance of alternative error correction models in the error-in-Y case. Figure shows proportional coefficient bias, proportional standard error bias, coverage, and power across all regression models tested for each of nine error correction models, one data subsampling approach, and the uncorrected regression approach. Data for violin plots has been winsorized at the 2.5% and 97.5% to cap outliers for visual display purposes only. The unwinsorized mean is indicated by the red circles and the corresponding median is in black. Error correction approaches are detailed in Section C.4.

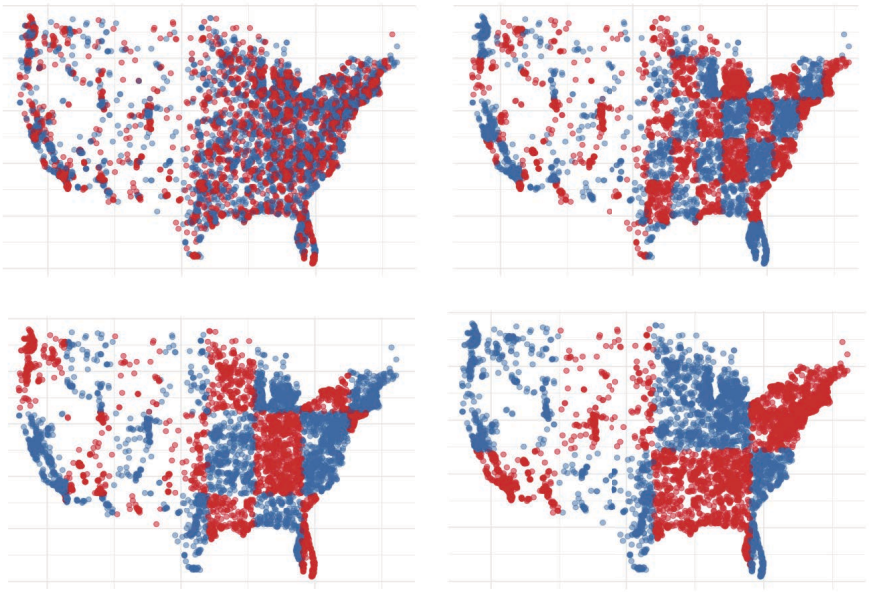

Figure B.10: Illustration of the data availability regime used to test multiple imputation when the calibration sample is spatially separated from the main sample. Figure visually displays how the calibration and main datasets are separated in the spatial experiment outlined in Section 4.4. From top left, clockwise: when separation distance is 1, 4, 8, and 16 degrees. To implement this experiment, we draw the main sample from the red boxes and the calibration sample from the blue boxes. As the width of the grid becomes larger, bias correction using multiple imputation becomes more difficult, as any point in the main sample is, in expectation, farther in physical distance from a point in the calibration sample. For each spatial separation distance, we "jitter" the grid up, down, and left for different bootstrap runs to ensure that results are representative across geographic divisions.

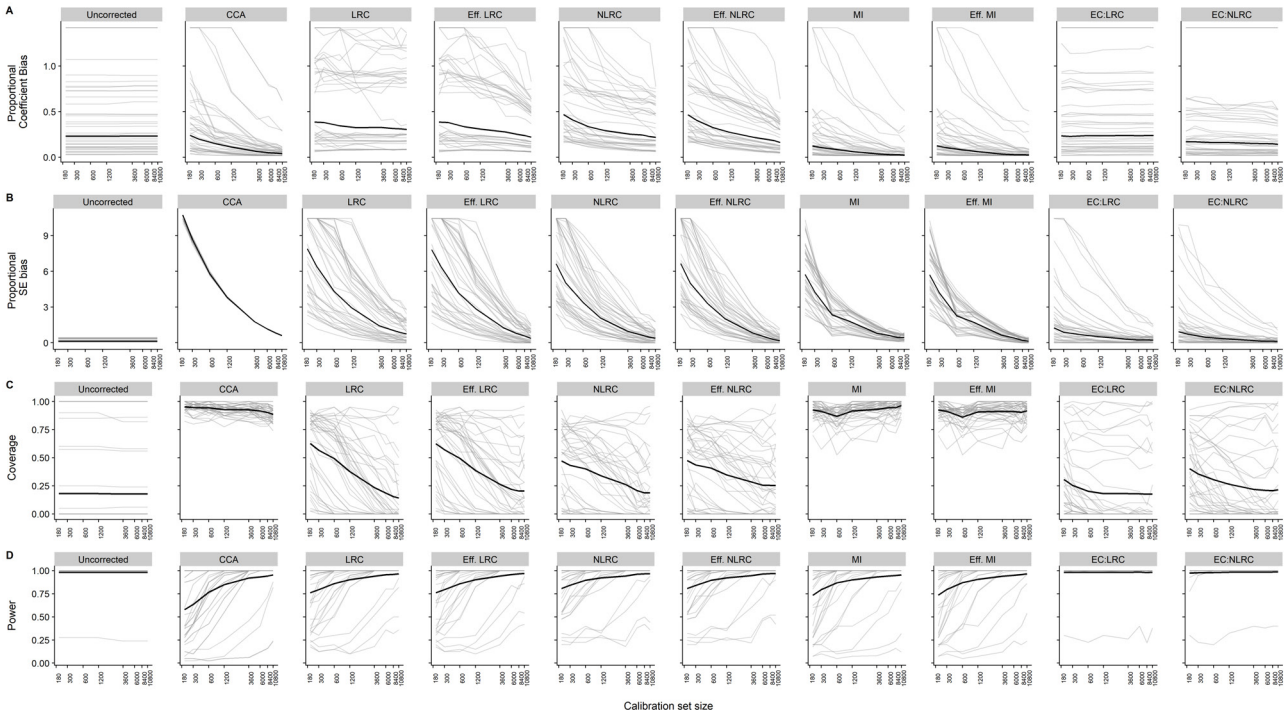

Figure B.11: Effects of calibration sample size on bias, coverage, and power across all alternative error correction techniques (error-in-X case). Proportional coefficient bias, proportional standard error bias, coverage, and power across all models plotted against calibration sample size for each error correction technique. Results can be interpreted as in Figure 5, but include additional error correction approaches beyond multiple imputation. Error correction approaches are detailed in Section C.4.

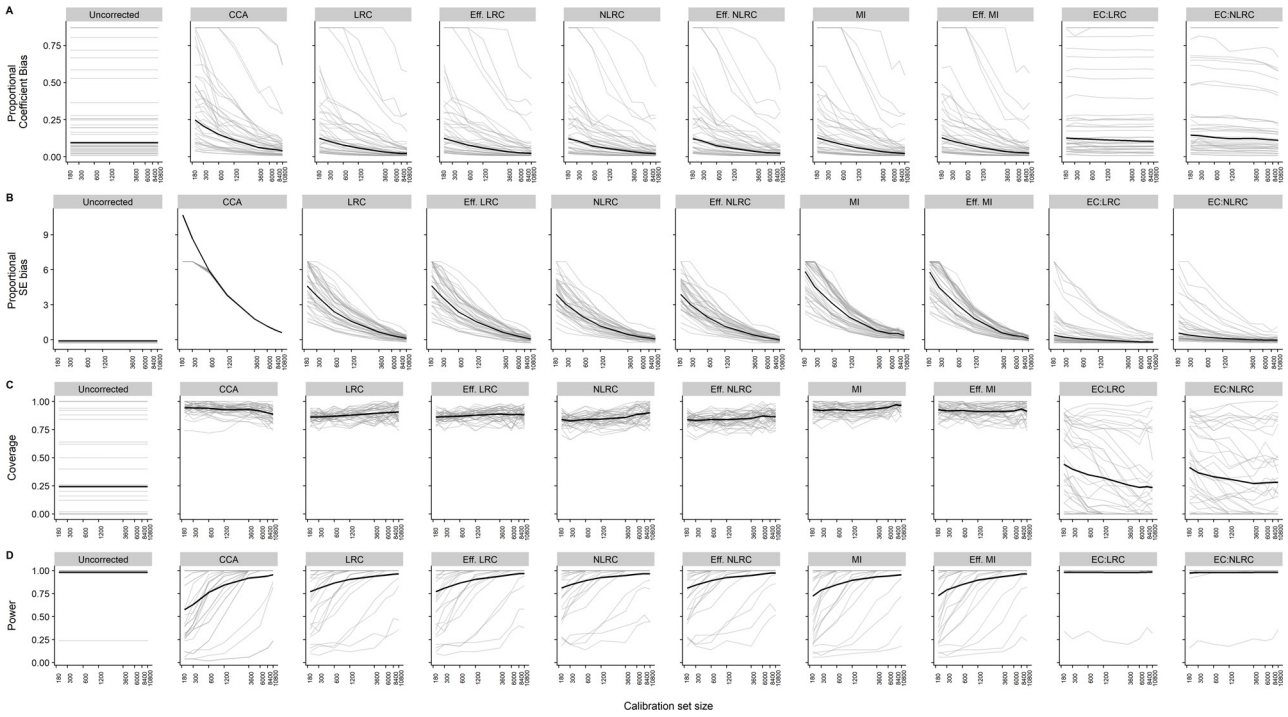

Figure B.12: Effects of calibration sample size on bias, coverage, and power across all alternative error correction techniques (error-in-Y case). Proportional coefficient bias, proportional standard error bias, coverage, and power across all models plotted against calibration sample size for each error correction technique. Results can be interpreted as in Figure 5, but include additional error correction approaches beyond multiple imputation. Error correction approaches are detailed in Section C.4.

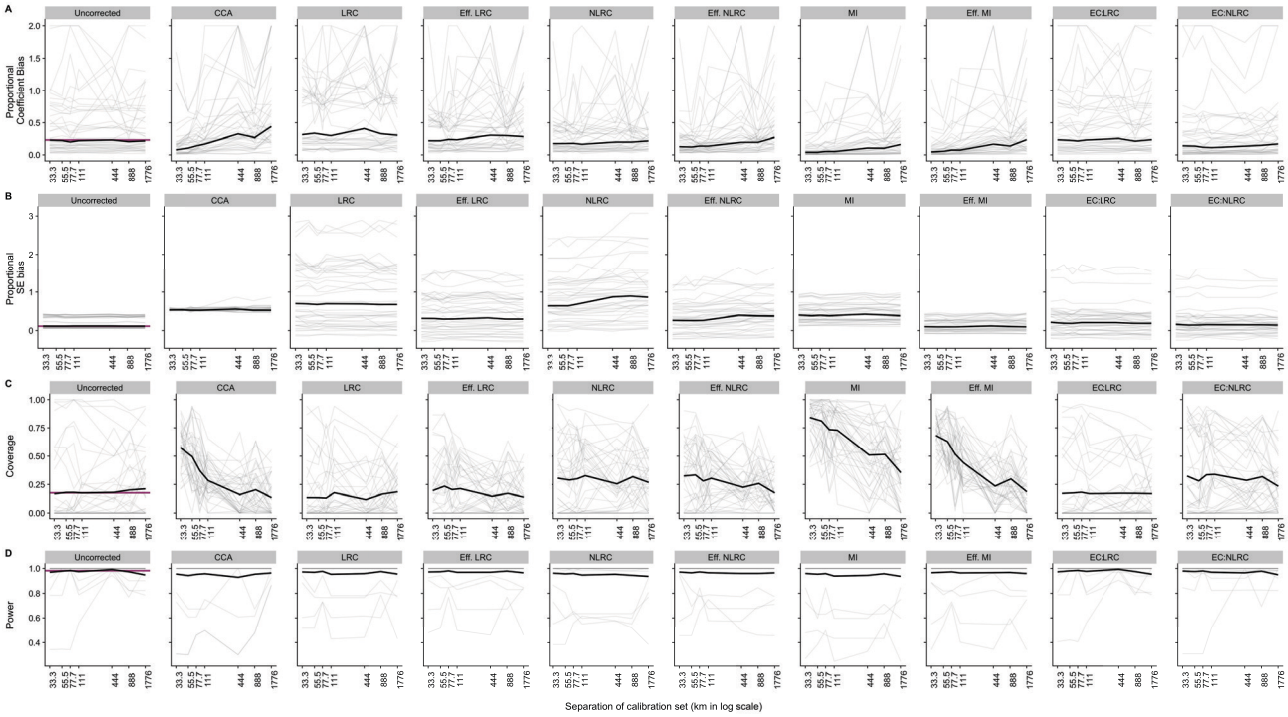

Figure B.13: The effect of distance between calibration and main datasets on the ability of all alternative error correction techniques to correct biases introduced by remotely sensed variables (error-in-X case). Proportional coefficient bias, proportional standard error bias, coverage, and power across all models plotted against distances between calibration sample and main sample for each error correction technique. Results can be interpreted as in Figure 6, but include additional error correction approaches beyond multiple imputation. Error correction approaches are detailed in Section C.4. As in the main text, horizontal purple lines show results for a random sampling of the calibration set (i.e., no spatial separation between calibration and main samples imposed) for the uncorrected model. For completeness we also show here average uncorrected values for each grid size. As expected, these values change very little as the grid size changes. Thus, for simplicity, we show the purple line only in the main text figures.

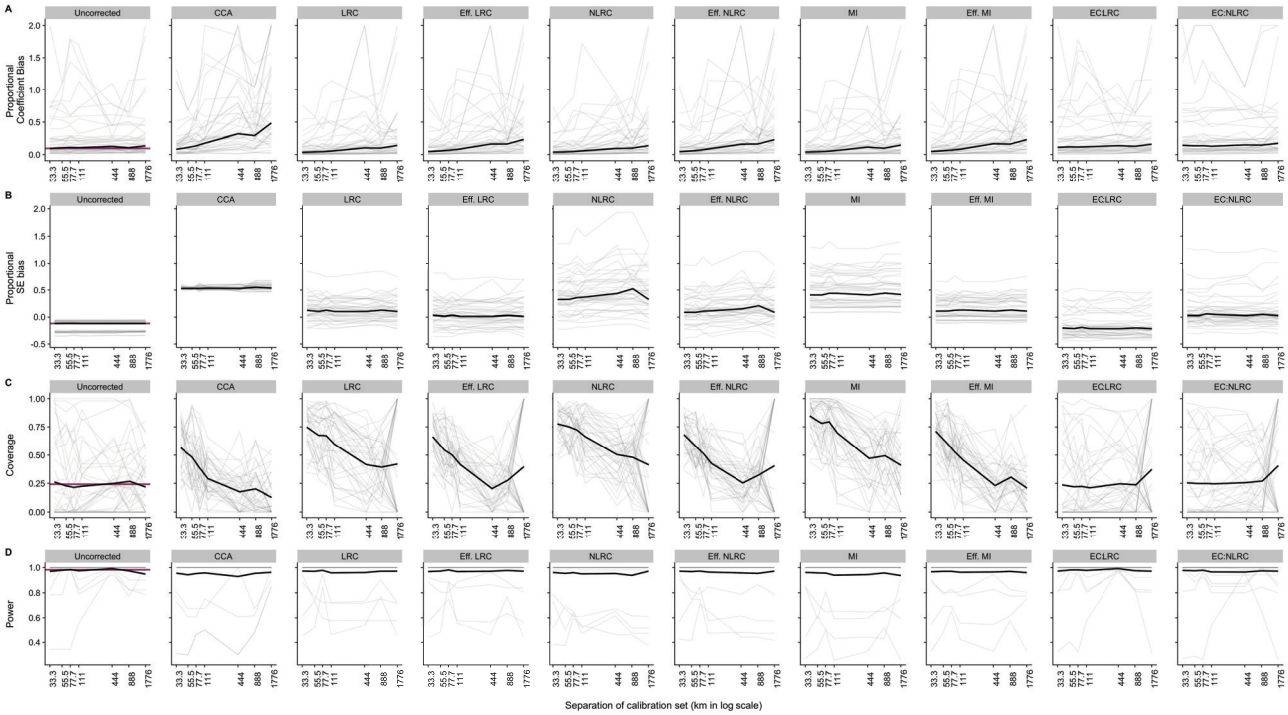

Figure B.14: The effect of distance between calibration and main datasets on the ability of all alternative error correction techniques to correct biases introduced by remotely sensed variables (error-in-Y case). Proportional coefficient bias, proportional standard error bias, coverage, and power across all models plotted against distances between calibration sample and main sample. Results can be interpreted as in Figure 6, but include additional error correction approaches beyond multiple imputation. Error correction approaches are detailed in Section C.4. As in the main text, horizontal purple lines show results for a random sampling of the calibration set (i.e., no spatial separation between calibration and main samples imposed) for the uncorrected model. For completeness we also show here average uncorrected values for each grid size. As expected, these values change very little as the grid size changes. Thus, for simplicity, we show the purple line only in the main text figures.

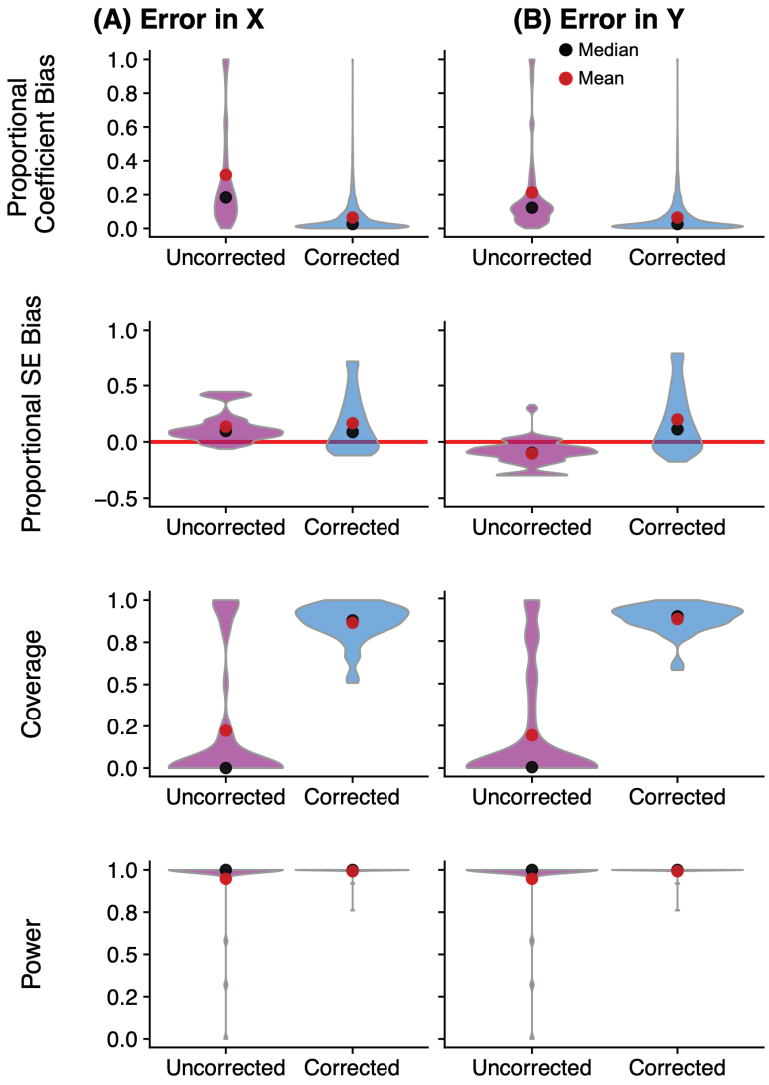

Figure B.15: Bias, coverage, and power for regression models using remotely sensed variables when the estimating equation includes spatial fixed effects. Figure shows analogous results to Figure 4, but here all regression models include state-level fixed effects (see text in Section 4.6 for details). Data for violin plots has been winsorized at the 2.5% and 97.5% to cap outliers for visual display purposes. The unwinsorized mean is indicated by the red circles and the corresponding median is indicated by the black circles.

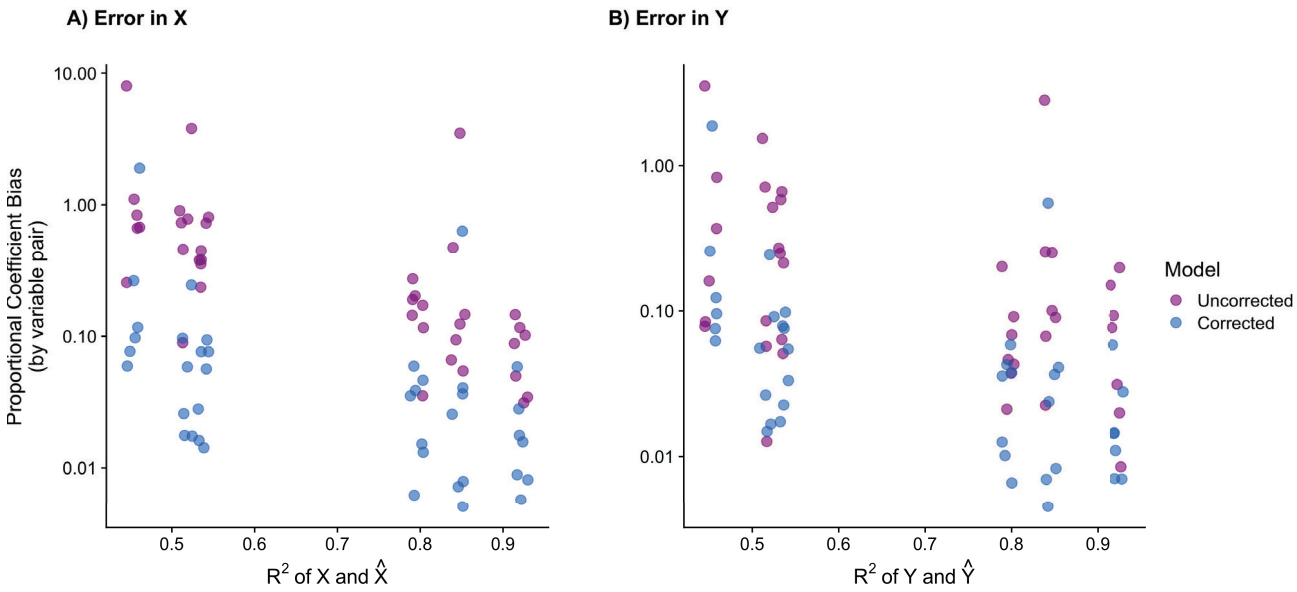

Figure B.16: Remotely sensed variables with higher accuracy exhibit lower coefficient bias in downstream regressions. Figure shows the proportional coefficient bias in uncorrected (purple) and corrected via multiple imputation (blue) downstream regression models plotted against the R^2 of the original remote sensing prediction. Multiple imputation consistently lowers coefficient bias for all regression models, but as R^2 decreases, the ability to correct for bias falls. Variable pairs with less than 0.5 R^2 suffer from large ($\geq$10%) bias in uncorrected models, and in some cases large biases remain even after correction. Points are jittered horizontally for visualization due to sets of variable pairs sharing the same R^2.

C Supplementary Discussions

C.1 Performance metrics

We use four performance metrics to compare uncorrected and corrected regression models to those using ground truth data. First, we compute bias in the regression coefficient:

$$\textit{bias in the regression coefficient} = \left| \frac{(\hat{\beta} - \hat{\beta}_{\tilde{z}})}{\hat{\beta}} \right|, \tag{S1}$$

where $\hat{\beta}$ is the regression coefficient estimated from Equation 1 using ground truth data, and $\hat{\beta}_{\tilde{z}}$ is the error-in-X or error-in-Y coefficient estimated *either* directly from Equation 2 (or the analogous error-in-Y regression) or from a version of Equation 2 that uses an error correction technique, such as multiple imputation, to correct for bias. We consider proportional bias to account for the different strengths of relationships between pairs of variables, and we consider absolute bias because the sign of β varies across variable pairs.

Second, we compute bias in standard error estimates:

$$\textit{bias in standard errors} = \frac{SE(\hat{\beta}_{\tilde{z}}) - SE(\hat{\beta})}{SE(\hat{\beta})}, \tag{S2}$$

where variables are defined as in Equation S1. Note that Equation S2 can be either positive or negative, reflecting overly conservative or overly precise standard errors in a model using remotely sensed data. Finally, we compute two statistics that combine the estimated coefficients and their uncertainty: *coverage* is calculated as the likelihood that a regression model using remotely sensed data recovers a 95% confidence interval containing the ground truth point estimate; and *power* is calculated as the likelihood that a regression model using remotely sensed data rejects a null hypothesis of no relationship between two variables when the ground truth regression also rejects this null (i.e., $p<0.05$).

C.2 Derivation of biases in the linear measurement error model

Here we derive the biases introduced by measurement error under the linear measurement error model, a general error model that encompasses both the familiar classical measurement error model as well as the Berkson error model as special cases (Keogh et al., 2020). We demonstrate how various restrictions imposed on this general measurement error model change the nature of bias, both in the error-in-X and error-in-Y cases.

As in the main text, we consider a simple linear regression framework in which the coefficient of interest is the slope parameter β (we suppress subscripts throughout this

section for parsimony):

$$y = \alpha + \beta x + \varepsilon$$

The "true" value of β is that which would be recovered from a regression with no measurement error in either x or y.

C.2.1 Errors in independent variables (error-in-X)

Under the linear measurement error model, the error-prone variable (in our case, remotely sensed predictions) is written as an affine function of the accurately measured variable (in our case, ground truth observations), as follows:

$$\tilde{x} = \theta + \lambda x + u,$$

where $\tilde{x}$ represents the imperfect measurements of x and u represents residual, mean zero measurement errors. With this definition, we can write the expectation of the slope parameter estimated using remotely sensed $\tilde{x}$ in place of x as:

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_{\tilde{x}}] &= \frac{\sigma_{\tilde{x}y}}{\sigma_{\tilde{x}}} = \frac{\sigma_{(\theta+\lambda x+u,y)}}{\sigma_{\theta+\lambda x+u}} \\
&= \frac{\lambda\sigma_{xy} + \sigma_{yu}}{\sigma_{\theta+\lambda x+u}} \\
&= \frac{\lambda\sigma_{xy} + \sigma_{yu}}{\lambda^2\sigma_x + \sigma_u} && \text{plug in } \sigma_{xy} = \beta\sigma_x \\
&= \frac{\lambda\beta\sigma_x + \sigma_{yu}}{\lambda^2\sigma_x + \sigma_u} \\
&= \beta \underbrace{\frac{\lambda\sigma_x}{\lambda^2\sigma_x + \sigma_u}}_{\substack{\text{random and mean-} \\ \text{reverting error}}} + \underbrace{\frac{\sigma_{yu}}{\lambda^2\sigma_x + \sigma_u}}_{\text{differential error}} .
\end{aligned} \tag{S3}$$

Equation S3 shows that under the *general* linear measurement error model, there are three components contributing to bias for the error-in-X case. First, random error u causes attenuation of β through the "reliability ratio" $\frac{\sigma_x}{\sigma_x+\sigma_u}$ within the first term. Second, differential measurement error causes bias through a nonzero covariance σ_{yu} in the second term. Finally, the relationship between $\tilde{x}$ and x itself, captured through λ and displaying mean reversion when $\lambda < 1$, will introduce bias through both the first and second terms. The direction of bias under this error model is ambiguous and will depend on the relative magnitudes of λ, σ_u, and σ_{yu}.

In Figure B.6, we decompose Equation S3 into its component parts to assess which features drive observed biases. To do so, we derive bias under two less general forms of Equation S3. First, under the *non-differential* linear measurement error model the

covariance σ_{yu} is assumed to be zero and $\mathbb{E}[\hat{\beta}_{\tilde{x}}]$ simplifies to:

$$\mathbb{E}[\hat{\beta}_{\tilde{x}}] = \beta \frac{\lambda \sigma_x}{\lambda^2 \sigma_x + \sigma_u}. \tag{S4}$$

Note that mean reversion with $\lambda < 1$ biases β upward, while the reliability ratio causes familiar attenuation bias. Whether $\hat{\beta}_{\tilde{x}}$ is inflated or attenuated under this error model depends on which of these two forces dominates. If measurement error is Berkson, λ and the reliability ratio balance one another such that no bias arises in the error-in-X case (Carroll et al., 2006).

Second, under a *differential classical* measurement error model, $\lambda = 1$ and $\mathbb{E}[\hat{\beta}_{\tilde{x}}]$ becomes:

$$\mathbb{E}[\hat{\beta}_{\tilde{x}}] = \beta \frac{\sigma_x}{\sigma_x + \sigma_u} + \frac{\sigma_{yu}}{\sigma_x + \sigma_u} \tag{S5}$$

While the reliability ratio attenuates β, σ_{yu} can be greater or less than zero, leading to theoretically ambiguous bias.

In Figure B.6, we plot observed proportional bias on the y-axis against theoretical bias following Equations S3 (top row), Equation S4 (middle row), and Equation S5 (bottom row). These findings show that the full flexibility of Equation S3 is needed to fully explain the patterns of bias we observe in remotely sensed variables.

C.2.2 Errors in dependent variables (error-in-Y)

Under the linear measurement error model for the error-in-Y case, we write:

$$\tilde{y} = \theta + \lambda y + v$$

where all terms are as defined above. With this definition, the expectation of the slope parameter estimated using $\tilde{y}$ instead of y is:

$$\begin{aligned} \mathbb{E}[\hat{\beta}_{\tilde{y}}] &= \frac{\sigma_{x\tilde{y}}}{\sigma_x} = \frac{\sigma_{(x,\theta+\lambda y+v)}}{\sigma_x} \\ &= \frac{\lambda \sigma_{xy} + \sigma_{xu}}{\sigma_x} \\ &= \lambda\beta + \frac{\sigma_{xu}}{\sigma_x} \qquad \text{plug in } \sigma_{xy} = \beta\sigma_x \end{aligned} \tag{S6}$$

Here, there are two components contributing to bias for the error-in-Y version of the general linear measurement error model. First, differential measurement error causes bias through a nonzero covariance σ_{xu} in the second term in Equation S6. Second, the relationship between $\tilde{y}$ and y itself, captured through λ, will introduce attenuation bias through the first term when $\lambda < 1$ under mean reversion. The direction of bias under

this error model is ambiguous and will depend on the relative magnitudes of λ and σ_{xu}.

As above for the error-in-X case, we consider bias under two more restrictive error models in order to decompose overall bias into its component parts. First, under the *non-differential* linear measurement error model, the covariance σ_{xu} is assumed to be zero and $\mathbb{E}[\hat{\beta}_{\tilde{y}}]$ simplifies to:

$$\mathbb{E}[\hat{\beta}_{\tilde{y}}] = \lambda\beta, \tag{S7}$$

where it is clear that mean reverting measurement error $\lambda < 1$ will lead to attenuation of slope coefficients for the error-in-Y case.

Second, under a *differential classical* measurement error model, we assume $\lambda = 1$ and $\mathbb{E}[\hat{\beta}_{\tilde{y}}]$ becomes:

$$\mathbb{E}[\hat{\beta}_{\tilde{y}}] = \beta + \frac{\sigma_{xu}}{\sigma_x}, \tag{S8}$$

where bias is determined by the covariance σ_{xu}.

As discussed above, Figure B.6 plots observed proportional bias on the y-axis against theoretical bias, following Equations S6 (top row), Equation S7 (middle row), and Equation S8 (bottom row) for the error-in-Y case. These findings show that for error-in-Y, just like error-in-X, the flexibility of Equation S3 is needed to fully explain the patterns of bias we observe in remotely sensed variables.

C.3 Calibration sample separation experiments

To evaluate the ability of multiple imputation and other bias correction methods to improve parameter recovery when the calibration sample is spatially distant from the main sample, we design an experiment based on an experimental design in (Rolf et al., 2021) where we evaluate models using main and calibration samples that are increasingly far away from each other in space. Specifically, we create grids over the continental U.S. with side lengths of $\delta \in [0.2,\ 0.3,\ 0.4,\ 0.7,\ 1,\ 4,\ 8,\ 16]$ degrees latitude and longitude. We then use this grid to divide the main and calibration sample into spatially separate sets by randomly sampling the main and calibration sets from boxes that are not adjacent (width-wise and height-wise) within the grid, creating a checkerboard pattern as shown in Figure B.10. As δ increases, the calibration set becomes on average further away from the main sample. This separation makes it more difficult for each bias correction method to correct for bias, as observations in the calibration set are now less likely to be similar to those in the main sample. To minimize the noise from any specific placement of the grid, we move or "jitter" the grid by shifting it down, up, or left by half the width of the grid, before running the analysis for each bootstrap run. Average results are taken over these differently jittered bootstrap runs similarly to the rest of the analysis.

C.4 Alternative bias correction methods

In this section, we briefly describe the various bias correction methods that we evaluate in our setting, and their respective assumptions. For reference, comprehensive reviews of statistical methods of regression bias correction include Fuller (1995); Carroll et al. (2006); Freedman et al. (2008) and Keogh and White (2014). Throughout, we denote observations in the calibration and main dataset set with c and m subscripts. We describe all error correction methods for the error-in-X case; error-in-Y methods are analogous.

1. **Complete Case Analysis (CCA)**
   Complete case analysis estimates regression parameters directly in the calibration set, ignoring the main dataset and therefore any remotely sensed data entirely. This approach is analogous to "benefits transfer" in economics (Boutwell and Westra, 2013), as it simply applies a regression parameter estimated in one sample to a new context without any adjustment. To implement CCA, the regression coefficient of interest is simply estimated using only the ground truth values available in the calibration dataset. All data in the main sample, and therefore all remotely sensed predictions, are dropped. For CCA to provide an unbiased estimate of the true parameter of interest in the main sample, the calibration dataset must have the same relationship between y and x as the main sample.

2. **Single imputation linear regression calibration (LRC)**
   Regression calibration has been one of the most commonly-used methods in the measurement error model literature (Freedman et al., 2008) because of its simplicity. To implement linear regression calibration, the relationship between a ground truth variable, its remotely sensed counterpart, and the ground truth regressor (error-in-Y case) or regressand (error-in-X case) is estimated in the calibration sample and then used to make predictions of the true variable in the main sample. In a second stage, the *predicted* ground truth values are then used to estimate the regression model in the main sample. This is the simplest form of imputation, but has been shown to perform poorly when measurement error is non-classical (Cole, Chu and Greenland, 2006). It also does not carry uncertainty from the calibration step into the final estimates, which can lead to overly-precise parameter estimates.

   The method is implemented as follows:

   (a) Estimate a linear model of $x_c = \delta_{RC} + \gamma_{RC}\tilde{x}_c + \phi_{RC}y_c + e_{RC}$ in the calibration sample.

   (b) Predict $\widehat{x_{RC}}$ using $\widehat{\delta_{RC}}$, $\widehat{\gamma_{RC}}$, $\widehat{\phi_{RC}}$ as well as observations of $\tilde{x}_m$ and y_m in the main sample.

(c) Estimate the linear regression in the main sample treating $\widehat{x_{RC}}$ as you would if it were not measured with error: $y_m = \alpha + \beta_{RC}\widehat{x_{RC}} + \epsilon$

(d) Obtain $\widehat{\beta_{RC}}$ as the linear regression calibration corrected parameter.

3. **Single imputation efficient linear regression calibration (Eff. LRC)**
This method is implemented identically to linear regression calibration above, but in step (c), the calibration set is appended to the main sample before estimation. This yields an estimate of $\widehat{\beta_{ERC}}$, which is an inverse-variance-weighted average of the estimate of the coefficient in the calibration set and the estimate of $\widehat{\beta_{RC}}$ using linear regression calibration (Freedman et al., 2008; Keogh and White, 2014).

4. **Single imputation nonlinear regression calibration (NLRC)**
Nonlinear regression calibration is an extension of single imputation linear regression calibration wherein the first stage model is a nonlinear and flexible model. We use a random forest in our implementation, but other nonlinear methods are possible. The method is implemented as follows:

(a) Estimate a nonlinear model of $x_c = \boldsymbol{\delta_{RF}}(\tilde{x}_c, y_c)$ in the calibration sample.

(b) Predict $\widehat{x_{RF}}$ using $\boldsymbol{\delta_{RF}}$ as well as observations of $\tilde{x}_m$ and y_m in the main sample.

(c) Estimate the linear regression treating $\widehat{x_{RF}}$ as you would if it were not measured with error: $y_m = \alpha + \beta_{RF}\widehat{x_{RF}} + \epsilon$

(d) Obtain $\widehat{\beta_{RF}}$ as the nonlinear regression calibration corrected parameter.

5. **Single imputation efficient nonlinear regression calibration (Eff. NLRC)**
This method is the same as above for single imputation nonlinear regression calibration, but in step (c), the calibration set is appended to the main sample before estimation.

6. **Single imputation external linear regression calibration (EC: LRC)**
In external calibration, the dependent variable is missing in the calibration set (for the error-in-X case). For the error-in-Y case, the independent variable is missing in the calibration set. The method is implemented similarly to linear regression calibration, but data from the dependent variable is omitted in step (a):

(a) Estimate a linear model $x_c = \delta_{ExRC} + \gamma_{ECRC}\tilde{x}_c + e_{ECRC}$ in the calibration sample, noting that y_m is omitted here as compared to method 2 (LRC).

(b) Predict $\widehat{x_{ECRC}}$ using $\widehat{\delta_{ECRC}}$, $\widehat{\gamma_{ECRC}}$ as well as observations of $\tilde{x}_m$ in the main sample.

(c) Estimate the linear regression treating $\widehat{x_{ECRC}}$ as you would if it were not measured with error: $y_m = \alpha + \beta_{ECRC}\widehat{x_{ECRC}} + \epsilon$

(d) Obtain $\widehat{\beta_{ECRC}}$ as the external linear regression calibration corrected parameter.

7. **Single imputation external nonlinear regression calibration (EC: NLRC)**
This approach is the same as above, but uses a nonlinear function (random forest) to model the first stage (a).

8. **Multiple imputation: Bayesian linear model (MI)**
Multiple imputation is described in detail in the main text. In general, this method replaces each "missing" or, in our case, mis-measured, value $\tilde{x}$ with a vector of $K > 1$ possible imputed values. There are many methods to do this imputation, and here we lay out the Bayesian linear model approach, which allows for parameter uncertainty in the first stage. We show in Figure B.7 that our main results do not depend on this particular approach to imputation. Under the Bayesian linear model imputation, the procedure (adapted by van Buuren (2012) from Rubin (1987)) is as follows:

(a) Fit a linear regression model of the ground-truth x to the remotely-sensed variable $\tilde{x}$ and to y using the calibration sample. Obtain the estimated parameters $\hat{\delta}$, $\hat{\gamma}$, $\hat{\phi}$.

(b) Calculate the covariance matrix $S = \widehat{\sum}_{\tilde{x},y}$.

(c) Using the estimated parameters, $\hat{\delta}$, $\hat{\gamma}$, $\hat{\phi}$, and the covariance matrix, $\widehat{\sum}_{\tilde{x},y}$, draw an estimate of the set of three coefficients for each imputation $k \in \{1, 2, ..., K\}$ from the standard (three dimensional) multivariate Gaussian distribution.

(d) Calculate $\hat{x}^k = \hat{\delta}^k + \hat{\gamma}^k\tilde{x}_m + \hat{\phi}^k y_m$, which is the imputed value for the k^{th} imputation.

(e) Perform the linear regression analysis K times using the K values of $\hat{x}^k$. Obtain K estimates of $\hat{\beta}$. Pool these together using Rubin's Rule (see Rubin (1987)) to obtain one final multiple imputation estimate of the parameter, $\widehat{\beta_{MI}}$.

9. **Efficient multiple imputation: Bayesian linear model (Eff. MI)**
This method is the same as above for multiple imputation, but in step (e), the calibration set is appended to the main sample before estimation.