

Correcting Bias When Using Latent Regressors*

Connor T. Jerzak[†]

Stephen A. Jessee[‡]

August 3, 2026

Abstract

Many core concepts in political science are latent and therefore can only be measured with error. Measurement error in a predictor attenuates slope coefficient estimates in regression, often biasing them toward zero. We show that widely used strategies for correcting attenuation bias—including instrumental variables and the method of composition—are themselves biased when applied to latent regressors, often performing worse than simple regression ignoring the measurement error altogether. We derive a correlation-based correction using split-sample measurement strategies. Rather than assuming a particular estimation strategy for the latent trait, our approach is modular and can be easily deployed with a wide variety of latent variable measurement strategies, including additive score, factor, or machine learning models; it requires no joint estimation while yielding consistent slopes under standard assumptions. Simulations and applications show stronger relationships after our correction, sometimes by as much as 50%. Open-source software implements the procedure. Results underscore that latent predictors demand tailored error correction; otherwise, conventional practice can exacerbate bias.

Keywords: Latent variables; Measurement error; Attenuation bias; Identification restrictions

Word count: 5,991

*Forthcoming in *Political Analysis*. Authors are listed in alphabetical order. We thank Devin Caughey, Guilherme Duarte, Robert Kubinec, Jeff Lewis, Umberto Mignozzetti, Jacob Montgomery, Aaron Pancost, Erik Snowberg, Zeynep Somer-Topcu, Chris Warshaw, and participants at MPSA, PolMeth, and the University of Texas Political Methodology Workshop for helpful comments. We thank Major Valls for excellent research assistance. Package link: [GitHub.com/cjerzak/lpmec-software](https://github.com/cjerzak/lpmec-software).

[†]Assistant Professor, Government Department, the University of Texas at Austin, email: connor.jerzak@austin.utexas.edu URL: ConnorJerzak.com ORCID: 0000-0003-1914-8905.

[‡]Professor, Government Department, the University of Texas at Austin, email: sjessee@utexas.edu URL: <https://laits.utexas.edu/~sjessee/> ORCID: 0000-0002-3937-6123.

Introduction

Many key concepts in political science, such as political knowledge, ideology, and democracy, are fundamentally unobservable and can only be estimated based on sets of observed indicators. A large body of research relies on measures of these latent variables as predictors in regression models to test theories and make empirical claims. For example, studies have employed proxies for political knowledge to explain patterns of political participation (Carpini and Keeter, 1996), used estimates of individuals' ideology to predict vote choice and other political behaviors (Tausanovitch and Warshaw, 2013; Jessee, 2009), and used democracy indices to assess the relationship between democratic institutions and economic growth (Acemoglu et al., 2019) or the risk of conflict (Hegre et al., 2001).

Because latent variables are unobserved and must be inferred from indicators, any empirical estimate of them (e.g., an index score or latent-trait estimate) inevitably contains measurement error. While the problems of measurement error and attenuation bias are widely recognized (Fuller, 2009; Carroll et al., 2006; Benoit et al., 2009; Pancost and Schaller, 2022; Fong and Tyler, 2021), the additional challenges posed by latent predictors are often underemphasized in applied work. Indeed, a survey of top political science journals finds that while a growing number of recent articles mention latent variables or measurement error, a small and relatively constant number discuss the two in combination (see Appendix V).

In this paper, we examine the bias introduced by using estimated latent variables as predictors and present methods to mitigate it under specific assumptions. We build on previous work, including Achen (1975) and Ansolabehere et al. (2008), while deriving new results and introducing tools that account for key features of latent variables and their estimation. We show that two commonly used techniques for adjusting regression estimates with latent predictors can and often do exacerbate bias, sometimes markedly. Most notably, the method of composition (Treier and Jackman, 2008), which is the most prominent method used in this area, typically makes the bias due to measurement error *worse* than it would be if the measurement error had simply been ignored altogether.

Building on split-sample methods (Angrist and Krueger, 1995), we use multiple estimates of latent traits based on disjoint sets of indicators to correct standard OLS and 2SLS estimators, which are each biased in opposing directions in the latent predictor case. The method is modular—compatible with a wide range of latent trait estimation approaches and allows for measurement approaches that do not specify a full statistical model for the indicators as a function of the latent trait (e.g., additive/averaging scores, factor models, and machine-learning approaches).

In applications, we show that accounting for latent-predictor measurement error can shift estimates by as much as 50%. But in other cases, measurement error may be so mild that corrections are negligible. We also present theory and simulation results. These show that with latent predictors, the standard OLS estimator (ignoring measurement error in the latent predictor) is biased toward zero, and that previously used methods for addressing measurement error can result in even more severely biased estimates. By contrast, our corrected estimator greatly reduces bias and produces estimates similar to those obtained using the more involved full joint estimation approach. Finally, we provide open-source software—an R package, `lpme`—for implementing these methods, helping applied researchers account for measurement error when using latent predictors in practice.

1 Measurement Error and Bias in Linear Regression

1.1 Standard Classical Measurement Error

We define the *standard (non-latent) measurement-error* case as the scenario where the regressor X has an intrinsic scale but is observed with additive noise. Here, and throughout, our parameter of interest is the slope β_X in a linear regression of an outcome Y on a predictor X :

$$Y_i = \beta_0 + X_i\beta_X + \varepsilon_i. \quad (1)$$

We use β_X to denote the true regression coefficient that we seek to estimate, and X is the true predictor variable, measured without error; i indices are suppressed where unambiguous.

Under the classical measurement error setup, the true predictor X is not directly observed, but we instead observe

$$\tilde{X} = X + U, \quad (2)$$

where $\mathbb{E}[U] = 0$, U is independent of both X and ε , and $\text{Var}(U) = \sigma_U^2$ so that $\text{Var}(\tilde{X}) = \text{Var}(X) + \text{Var}(U)$. When estimating a linear regression using noisy $\tilde{X}$, rather than X itself, as our predictor, the least squares estimate of β_X is biased toward zero according to the standard measurement error attenuation factor (Fuller, 2009; Bound et al., 2001):

$$\text{plim } \hat{\beta}_{\tilde{X}} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2} \beta_X = \lambda \beta_X, \quad (3)$$

where $\lambda \equiv \sigma_X^2 / (\sigma_X^2 + \sigma_U^2)$ is called the attenuation factor.

In the case of multivariate regression that includes covariates C in addition to X , which is measured with error, all slope coefficients will typically be biased unless X is uncorrelated with the other covariates (see Carroll et al., 2006, Chapter 3). The attenuation factor for the slope coefficient on the variable measured with error becomes

$$\lambda_1 = \frac{\sigma_{X|C}^2}{\sigma_{X|C}^2 + \sigma_U^2} \quad (4)$$

where $\sigma_{X|C}^2$ is the residual variance from regressing X on C . Note that $\lambda_1 \leq \lambda$ with the equality holding only when X is uncorrelated with all other covariates. Measurement error in X will also tend to bias coefficient estimates for C , with this bias becoming more severe as the variance of the measurement error in X increases. In the case of non-classical measurement error, these results are important but less straightforward to derive, and often require simulation-based adjustment (see Carroll et al., 2006 p. 49).

1.2 Measurement Error in Latent Predictors

When a predictor X is latent and must be estimated, measures of X necessarily include some uncertainty. Additionally, its scale must be defined (*identified*), and as a consequence, regression coefficients are attenuated by a different factor than in the standard measurement error case above (see Table 1). To understand this, note that one reason classical measurement error in X biases

Table 1: *Classical measurement error versus latent-predictor measurement error.*

	Observable predictor with classical error	Estimated latent predictor
TARGET REGRESSOR	X has an intrinsic scale	X has an identified scale, e.g. $\mathbb{E}[X] = 0, \text{Var}(X) = 1$
OBSERVED REGRESSOR	$\tilde{X} = X + U$	$\hat{X} = [\tilde{X} - \text{mean}(\tilde{X})]/\text{sd}(\tilde{X})$
VARIANCE	$\text{Var}(\tilde{X}) = \text{Var}(X) + \text{Var}(U)$	$\text{Var}(\hat{X}) = 1$ by construction
OLS ATTENUATION	β_X is multiplied by $1/(1 + \sigma_U^2)$ when $\text{Var}(X) = 1$	β_X is multiplied by $1/\sqrt{1 + \sigma_U^2}$
CONSEQUENCE FOR IV	IV can correct classical attenuation	IV over-corrects unless the identifying rescaling is addressed

coefficient estimates is that it changes the scale of the regressor, with the variance of the observed $\tilde{X}$ being larger than that of the true X . When estimating latent traits, however—unless the scale is identified such that the variance of the estimates equals the variance of the true X plus the variance of the typically unknown measurement error—then the resulting attenuation bias will differ from the result in Equation 3.

To show how identification rescaling changes measurement-error bias, we start with the noisy measure $\tilde{X}$ in Equation 2 and then impose the usual latent-trait normalization: mean zero, unit variance, and a fixed direction. Together, this identifies the scale in the common one-dimensional setting.¹ As is typically done in latent traits estimation, we assume that the estimates of the latent trait have the same identifying restriction, i.e., $\hat{X}$ has been estimated or transformed so that it also has mean 0 and standard deviation 1 in the sample, with the same directional restriction. The fact that this identification restriction applies both to the true latent trait and to whatever point estimates of the latent trait are used in subsequent analyses is typically overlooked in applied work, but it is central to understanding how measurement error operates in the case of latent predictors.

Formally, we assume that our point estimates of a latent variable predictor are

$$\hat{X} = \frac{\tilde{X} - \text{mean}(\tilde{X})}{\text{sd}(\tilde{X})}, \quad (5)$$

which are the transformed (i.e., identified) set of estimates of the latent trait. Importantly, researchers do not observe the intermediate $\tilde{X}$ in this case, since this would require knowing the measurement error variance. These observed, identified estimates of X , which we call $\hat{X}$, can be represented as an affine transformation of $\tilde{X}$, which implies that the estimated slope coefficient will be changed by a compensating linear transformation.

In the latent-predictor case, identification sets $\sigma_X^2 = 1$. Because the regression uses $\hat{X} =$

¹See Rivers (2003) for a formal discussion.

$\tilde{X}/\sqrt{1+\sigma_U^2}$ up to centering, which affects only the intercept, the slope on $\hat{X}$ equals the slope on $\tilde{X}$ multiplied by $\sqrt{1+\sigma_U^2}$:

$$\text{plim } \hat{\beta}_{\hat{X}} = \frac{1}{\sqrt{1+\sigma_U^2}} \beta_X. \quad (6)$$

In other words, the slope coefficient estimate in the latent predictor case is still attenuated, but not as severely as in the standard case since $0 < \frac{1}{1+\sigma_U^2} < \frac{1}{\sqrt{1+\sigma_U^2}} < 1$ whenever $\sigma_U^2 > 0$.

As shown in Appendix Section A.I.4, the logic of our results below applies to other identifying restrictions (e.g., setting two observations’ latent trait values at -1 and 1, respectively, or even assuming different identifying restrictions for the true latent variable and for its estimates) in a straightforward way. Since mean 0, standard deviation 1 restrictions are common in the literature and also produce easily interpretable measures, we adopt them in all of our discussions below.

2 Existing Approaches and Their Limitations

2.1 No Adjustment

The most common approach to dealing with measurement error in latent predictors has been to simply ignore the problem and use point estimates of the latent trait as the predictor without any further adjustment. In cases where measurement error is small relative to the variance of the predictor, using this standard approach may be reasonable. But with significant measurement error, slope estimates will be biased toward zero, sometimes dramatically so.

2.2 Full Joint Estimation

When attenuation from treating estimated latent scores as error-free is a concern, a natural solution is to estimate the measurement and outcome models jointly (Richardson and Gilks, 1993). Examples include Bayesian factor–outcome models (Kubinec, 2019; Battaglia et al., 2024) and SEM/LISREL-style approaches. Joint estimation is attractive because the outcome and indicators inform inference about the latent trait within a single model.

This joint approach requires a full statistical model for the indicators as a function of the latent trait and also for the outcome as a function of the latent trait. While modern computational tools, including `brms` (Bürkner, 2017), have made some forms of joint modeling easier, estimation can still be somewhat computationally demanding and, more importantly, potentially difficult for “black-box” scores (e.g., PCA, additive indices, or ML/text embeddings) that lack an explicit likelihood or one that can be tractably deployed via MCMC. Because joint estimation is the gold-standard benchmark whenever it is feasible, we treat it as the reference point for evaluating other methods.

2.3 Instrumental Variables

When full joint estimation is infeasible, a standard errors-in-variables workaround is instrumental variables (IV), which aims to recover β_X by isolating variation due to the signal rather than

measurement noise (cf. Gillard (2006)). In the classical setting with $\tilde{X} = X + U$, IV/2SLS corrects attenuation by using an instrument Z that satisfies: (i) $\text{Cov}(X, Z) \neq 0$ (relevance); (ii) $\text{Cov}(Z, U) = 0$; and (iii) $\text{Cov}(Z, \varepsilon) = 0$ (exclusion). Under these conditions, 2SLS effectively leverages the part of the observed regressor that co-moves with Z —the component attributable to X , not U —to identify the target slope.²

This logic is appealing in latent-variable applications because one can often construct internal instruments based on different sets of indicators for the same latent variable. In the classical errors-in-variables case, this “two-measures” strategy can plausibly satisfy the IV conditions: the alternate measure is correlated with X and, if built from independent information, can be approximately uncorrelated with the measurement error in the primary measure (and unrelated to ε).

However, applying this standard IV logic directly to identified latent scores $\hat{X}$ is generally inappropriate. Unlike $\tilde{X}$, $\hat{X}$ is a standardized (identified) transformation of $\tilde{X}$: identification forces $\text{Var}(\hat{X})$ to match $\text{Var}(X)$ by construction. As a result, unadjusted IV tends to “correct” as if operating on $\tilde{X}$ even though the analyst is regressing on $\hat{X}$. Thus, unadjusted IV is biased away from zero: for any nonzero β_X ,

$$\text{plim } \hat{\beta}_{\text{IV}, \hat{X}^{(1)} | \hat{X}^{(2)}} = \beta_X \sqrt{1 + \sigma_U^2},$$

so IV inflates the coefficient magnitude whenever $\sigma_U^2 > 0$ (Appendix I).

2.4 Method of Composition (MOC)

A different—and in applied latent-trait work, common—approach is the method of composition (MOC), which aims to propagate uncertainty from the measurement stage, where X is estimated based on a set of observed indicators W , into the second-stage regression. In the two-stage MOC implementation commonly applied with latent predictors, one first draws a latent-score vector from $p(X | W)$, then re-anchors it to a common identified scale and regresses Y on the resulting standardized draw. We denote the pre-standardized draw by $\tilde{X}_{\text{MOC}}^{(t)}$ and the standardized regressor by $\hat{X}_{\text{MOC}}^{(t)}$. Crucially, these draws condition on W but not on Y , even though Y contains information about X whenever $\beta_X \neq 0$.³

Many high-impact works have used MOC with the goal of appropriately accounting for the uncertainty in latent predictors (e.g. Bernhard et al., 2020; Caughey and Warshaw, 2018; Caughey and Warshaw, 2019; Jessee, 2010; Kstellec et al., 2015; Lax et al., 2019; Tai et al., 2024). The Varieties of Democracy (V-Dem) project even provides, in addition to point estimates of their democracy measures, a series of posterior samples so that researchers can employ MOC in subsequent analyses (Pemstein et al., 2018).⁴

²Although we focus on measurement error here, IV can also address simultaneity and omitted-variable bias (see Pancost and Schaller, 2022).

³Treier and Jackman (2008, p. 216) state this as the assumption, $p(x | Z) = p(x | w, y, Z)$; here their Z corresponds to our indicator matrix W and, suppressing additional covariates, this is the restriction $p(X | W, Y) = p(X | W)$.

⁴Our critique concerns using MOC to fit second-stage regression models with latent predictors. V-Dem’s use of MOC to combine individual measures into more general indices and account for the resulting uncertainty (McMann et al., 2016) does not necessarily suffer from the same issues.

More concretely, MOC proceeds according to the following steps for posterior draws $t = 1, \dots, T$:

1. Fit the measurement model using W alone and retain pre-standardized posterior latent-trait draws from $p(X | W)$, denoted $\tilde{X}_{\text{MOC}}^{(t)}$ for $t = 1, \dots, T$.
2. Re-anchor each posterior draw to a common identified scale before outcome analysis, since the latent trait is identified only up to location, sign, and scale; denote the standardized draw by $\hat{X}_{\text{MOC}}^{(t)}$.
3. For each draw, estimate $Y_i = \beta_0^{(t)} + \beta_{\text{MOC}}^{(t)} \hat{X}_{\text{MOC},i}^{(t)} + \varepsilon_i$, recording the slope estimate $\hat{\beta}_{\text{MOC}}^{(t)}$ and its estimated variance-covariance matrix $\hat{V}^{(t)}$.
4. Summarize either the draw-specific slopes $\{\hat{\beta}_{\text{MOC}}^{(t)}\}_{t=1}^T$ or coefficient draws from a normal approximation centered at $\hat{\beta}_{\text{MOC}}^{(t)}$ with covariance $\hat{V}^{(t)}$.

This two-stage sampling-and-refitting procedure is typically justified as a way to “propagate” measurement-stage uncertainty forward into the regression and avoid overstating precision. However, when the regressor is measured with error, the dominant first-order problem is attenuation bias: measurement error pulls slope estimates toward zero.

Indeed, this use of MOC averages over a set of latent-score draws, each of which will typically contain more random variation than the posterior-mean latent scores. In the stylized additive model of Appendix I, the point estimates are equal to the true latent traits plus some measurement error, and the individual draws from the posterior distribution for X given W are approximated as pre-standardized draws around the point estimates:

$$\tilde{X}_{\text{MOC}}^{(t)} = \tilde{X} + \xi^{(t)} = X + U_{\text{MOC}} + \xi^{(t)},$$

where $\tilde{X} = X + U_{\text{MOC}}$ is the pre-standardized point estimate and $\xi^{(t)}$ is an additional source of mean-zero draw variation.

Appendix I formalizes this intuition in a stylized additive-error representation in which a standardized MOC draw can be written as

$$\hat{X}_{\text{MOC}}^{(t)} = \frac{X + U_{\text{MOC}} + \xi^{(t)}}{\sqrt{1 + \sigma_{U,\text{MOC}}^2 + \sigma_{\xi}^2}},$$

with $\xi^{(t)}$ mean-zero and orthogonal to X , U_{MOC} , and ε . In that setting, MOC draw-level regressions are attenuated by $1/\sqrt{1 + \sigma_{U,\text{MOC}}^2 + \sigma_{\xi}^2}$ rather than $1/\sqrt{1 + \sigma_{U,\text{MOC}}^2}$. Thus, when posterior draws are used in this two-stage way, draw-level variation can worsen attenuation relative to using posterior means. In the multivariate case, this additional draw-level noise can also exacerbate biases on the slope estimates for other predictors that are correlated with X .

2.5 Multiple Over-Imputation

A compromise between MOC and full joint estimation is multiple over-imputation (Rubin, 1988; Blackwell et al., 2017). Multiple over-imputation proceeds in this setting as follows:

1. Fit the measurement model to W alone and obtain an approximation to $p(X_i | W)$, e.g., via $\mu_i = \mathbb{E}(X_i | W)$ and $\sigma_i = \text{SD}(X_i | W)$.
2. Form an imputation data set containing the observed outcome Y and a latent-score column. The latent-score entries are over-imputed rather than fixed at μ_i , with (μ_i, σ_i) supplied as unit-specific priors (Y included in the imputation model).
3. For each completed data set $r = 1, \dots, M_{\text{imp}}$, re-anchor the completed latent-score vector $X_{\text{imp}}^{(r)}$ to the identified latent scale, fit $Y_i = \beta_0^{(r)} + \beta_X^{(r)} X_{\text{imp},i}^{(r)} + \varepsilon_i$, and pool the resulting slope estimates across completed data sets.

The second and third steps make multiple over-imputation more similar to full joint estimation MOC because the outcome can enter the imputation step. It remains distinct from full joint estimation, however, because the original measurement model is not re-estimated jointly with the outcome equation; instead, the first-stage measurement distribution is approximated through unit-specific summaries supplied to the imputation model. Consequently, the completed-data regressions should generally be interpreted as estimates from a staged approximation to $p(X | W, Y)$, not as draws from the full joint posterior $p(\beta_0, \beta_X, X | W, Y)$, except under strong specification and compatibility conditions. We therefore include multiple over-imputation as a comparison method in the simulations below, with details in Appendix A.III.5.

3 A Correlation-Corrected Estimator Using Split Indicators

Taken together, these limitations—naïve IV’s tendency to over-correct once identification rescaling is imposed, and MOC’s tendency to add noise and worsen attenuation—motivate a different tack: estimate the latent-scale measurement error and then undo the identification-aware attenuation implied by Equation 6. To do this, we need a way to estimate σ_U^2 , since the attenuation factor is determined by this measurement-error variance. Importantly, this is measurement error variance on the scale of the true latent trait X , not the transformed (standardized) scale for the estimates of the latent trait that we can observe.

Indeed, one way to learn about σ_U^2 —and therefore about the latent-predictor attenuation factor in Equation 6—is to obtain (at least) two independent measurements of the same latent trait X on the *latent scale* (i.e., before the identification/standardization step that produces $\hat{X}$). Intuitively, if two measurements of X give similar values for each observation, then the measurement error variance σ_U^2 must be small; if the two measurements often disagree substantially, then σ_U^2 must be larger.

In indicator-based latent-trait settings, we can usually construct such repeated measures by partitioning the indicators into two disjoint sets and estimating the trait separately within each set. For example, with 10 indicators for a latent trait X , we could use 5 indicators to obtain estimates $\hat{X}^{(1)}$ and the other 5 to obtain estimates $\hat{X}^{(2)}$. If we split the indicators evenly, it may be reasonable to assume that the two resulting estimates have the same measurement-error variance.⁵ Following the notation above, let $\tilde{X}^{(1)}$ and $\tilde{X}^{(2)}$ denote the two *unobserved* (unidentified) measurements on

⁵For an odd number of indicators, consider all partitions where the two sets differ by one item.

the same scale as the true latent trait:

$$\tilde{X}^{(1)} = X + U^{(1)} \quad \text{and} \quad \tilde{X}^{(2)} = X + U^{(2)},$$

where we assume $U^{(1)}$ and $U^{(2)}$ each have mean zero and the same variance σ_U^2 , and are independent of each other and of X .⁶ Finally, let $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ denote the corresponding *identified* versions of these split-indicator estimates—i.e., the quantities we actually observe after imposing the usual mean-zero, unit-variance normalization (consistent with Equation 5). In this section, σ_U^2 denotes the common error variance of the split-specific unstandardized measures $\tilde{X}^{(1)}$ and $\tilde{X}^{(2)}$; when the full-indicator score is discussed separately, we write $\sigma_{U,\text{full}}^2$.

Although we do not observe $\tilde{X}^{(1)}$ or $\tilde{X}^{(2)}$ directly, their correlation is identical to the observed split-score correlation, $\text{Cor}(\hat{X}^{(1)}, \hat{X}^{(2)})$, because each observed split score is an affine (mean/variance) transformation of its latent-scale counterpart. The idea of using correlation between multiple sets of estimates to correct for measurement error has been around since at least Spearman (1904). In that context, we can readily show that

$$\begin{aligned} \rho \equiv \text{Cor}(\hat{X}^{(1)}, \hat{X}^{(2)}) &= \text{Cor}(\tilde{X}^{(1)}, \tilde{X}^{(2)}) = \frac{\text{Cov}(\tilde{X}^{(1)}, \tilde{X}^{(2)})}{\sqrt{\sigma_{\tilde{X}^{(1)}}^2 \sigma_{\tilde{X}^{(2)}}^2}} \\ &= \frac{\text{Cov}(X + U^{(1)}, X + U^{(2)})}{\sqrt{(\sigma_X^2 + \sigma_U^2)^2}} \\ &= \frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2}. \end{aligned} \tag{7}$$

In our setup, $\sigma_X^2 = 1$ by identification so that $\rho = 1/(1 + \sigma_U^2)$. Thus ρ equals the classical attenuation factor λ only in this standardized equal split-error special case.

In our latent variable predictor setup, the attenuation factor for the slope coefficient is $\frac{1}{\sqrt{1 + \sigma_U^2}}$ (see Equation 6). In the standard observable variable case, it is $\frac{1}{1 + \sigma_U^2}$ (see Equation 3). This suggests using the square root of the sample correlation between the two separate estimates of X to correct attenuation in the estimated slope coefficient, β_X .⁷ To do this, we first estimate the standard linear regression coefficient using $\hat{X}^{(1)}$ as our predictor to obtain $\hat{\beta}_{\hat{X}^{(1)}}$ and then correct this estimate:

$$\hat{\beta}_{\hat{X}^{(1)}}^* = \frac{\hat{\beta}_{\hat{X}^{(1)}}}{\sqrt{\text{Cor}(\hat{X}^{(1)}, \hat{X}^{(2)})}}. \tag{8}$$

We define this square-root correction for partitions with a positive estimated split-half correlation; nonpositive split-half correlations are flagged or discarded as weak-measurement partitions.

This correction will rely on several maintained assumptions. First, the target parameter is the coefficient on the identified latent trait in the linear regression of Y on X , with $\mathbb{E}[X] = 0$ and

⁶This independence assumption is similar to the conditional independence assumption commonly used in latent traits measurement strategies, including many item response models.

⁷In some situations, this split-half may not be appropriate. For example, if indicators are designed to tap different aspects of a latent variable rather than with internal consistency in mind (e.g., Gosling et al., 2003), it may be preferable to consider test-retest reliability.

$\text{Var}(X) = 1$. Second, each split-indicator estimate can be represented, before standardization, as $\tilde{X}^{(j)} = X + U^{(j)}$ for $j \in \{1, 2\}$, where the split-specific errors are mean-zero and independent of X and of the regression disturbance. Third, the two split-specific errors are independent of one another, which is the split-indicator analog of conditional independence assumptions common in latent-trait models. Fourth, for the simplest derivation, the split-specific error variances are equal; Appendix II evaluates the sensitivity of the estimator to violations of this equality. Finally, the split estimates must be sufficiently correlated for the correction to be empirically informative since low split-half correlations are a special case of the “weak instruments problem” and should be interpreted as a diagnostic indicating that the indicators contain limited information about the latent trait. Under these assumptions, we find Proposition 1.

Proposition 1 (Consistency of the split-indicator correction). *Suppose $\{(Y_i, X_i, U_i^{(1)}, U_i^{(2)})\}_{i=1}^n$ are i.i.d. with finite second moments and*

$$Y_i = \beta_0 + X_i\beta_X + \varepsilon_i, \quad \mathbb{E}[X_i] = 0, \quad \text{Var}(X_i) = 1, \quad \text{Cov}(X_i, \varepsilon_i) = 0.$$

For a fixed, sign-aligned split of the indicators, suppose

$$\tilde{X}_i^{(j)} = X_i + U_i^{(j)}, \quad j \in \{1, 2\},$$

where $\mathbb{E}[U_i^{(j)}] = 0$, $\text{Var}(U_i^{(1)}) = \text{Var}(U_i^{(2)}) = \sigma_U^2$, $U_i^{(1)}$ and $U_i^{(2)}$ are independent of one another, and $(U_i^{(1)}, U_i^{(2)})$ are independent of (X_i, ε_i) . Let $\hat{X}_i^{(j)}$ denote the sample-standardized version of $\tilde{X}_i^{(j)}$, with directions aligned to the target latent direction and across splits. Let $\hat{X}^{(s)} = (\hat{X}_1^{(s)}, \dots, \hat{X}_n^{(s)})^\top$ for $s \in \{1, 2\}$. If the population split-half reliability $\rho = \text{Cor}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)})$, equivalently the probability limit of $\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})$, is positive, then

$$\hat{\beta}_{\hat{X}^{(1)}}^* = \frac{\hat{\beta}_{\hat{X}^{(1)}}}{\sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}} \quad \text{satisfies} \quad \text{plim} \hat{\beta}_{\hat{X}^{(1)}}^* = \beta_X.$$

The correction is applied on the event $\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)}) > 0$, whose probability tends to one under $\rho > 0$. The same result holds using $\hat{X}^{(2)}$ as the predictor. See Appendix I for proof.

Here and below (suppressing again dependence on i in notation), we use the asterisk superscript to denote our correlation-corrected estimators. Because we have no reason to privilege either split-half estimate, we compute both $\hat{\beta}_{\hat{X}^{(1)}}^*$ and $\hat{\beta}_{\hat{X}^{(2)}}^*$ and average them within a partition.⁸ This averaging reduces variability from the choice of partition.⁹ Averaging across the two split-half estimates also makes the method less sensitive to unequal error variances. If one split yields a correction that is too small, the other tends to yield a correction that is too large. As we show in Appendix II, this averaging mitigates but does not eliminate imbalance bias when $\sigma_{U,1}^2 \neq \sigma_{U,2}^2$; over plausible ranges of $\sigma_{U,1}^2$ and $\sigma_{U,2}^2$, the remaining multiplicative factor appears close to 1.

⁸We do not use $\hat{X}$ —the point estimates of the latent trait using all of the indicators—because our correction factor is estimated from the split-half reliability of $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$, not the full-indicator score; applying the same correction to $\hat{X}$ would require additional assumptions or estimation.

⁹Our averaging approach, and in particular the equivalent averaged instrumental variables (IV) approach we describe in the next section, shares commonalities with Gillen et al., 2019.

Finally, because there are typically multiple ways to partition the indicators into two equally sized sets, we define $\hat{\beta}^*$ as the median over partitions after first averaging $\hat{\beta}_{\hat{X}^{(1)}}^*$ and $\hat{\beta}_{\hat{X}^{(2)}}^*$ within each partition. The median is less sensitive to unstable partitions with low split-half correlations. For a fixed finite set of valid partitions, the median-aggregated estimator remains root- n consistent when all partition-specific estimators share the same probability limit, but its limiting distribution is generally the median of a joint Gaussian vector rather than Gaussian. We therefore use a row bootstrap to capture uncertainty in both the coefficient and the correction factor. Because the finite-partition median is nonsmooth, additional complexities are involved (see Appendix I), but Winsorized means are another reasonable, mathematically smoother aggregation approach, yielding similar performance (Appendix VIII).

The same split-indicator approach above can be applied to the IV framework. In our case, with two different measures of the latent trait, one split estimate can be used as the regressor and the other as the instrument. As shown above, standard IV adjustments for measurement error will *over-correct* for attenuation, resulting in estimates that are inflated, i.e., biased away from zero by a factor of $\sqrt{1 + \sigma_U^2}$. This suggests using the correlation between the two sets of latent trait estimates to adjust the standard IV estimate. In contrast to the bivariate OLS case considered above, in the IV case we multiply, rather than divide, our slope estimate by $\sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}$ for partitions with positive estimated split-half correlations.

In fact, as shown in Appendix I, the correlation-corrected IV and correlation-corrected OLS estimators are equivalent in the bivariate case. Despite this equivalence, one advantage of the IV approach is its ease of extension to multivariate regression (see Appendix I). In multivariate regressions, the error-prone variable may be the main predictor or merely a “control”; in either case, failing to account for its measurement error can bias the estimated coefficients on all predictors in the regression.

Appendix I formalizes this extension. In brief, consider a regression of Y on a vector of latent predictors X and observed covariates C . Standard IV correction will produce consistent slope estimates for C , but estimates for β_X will be biased away from zero. Appendix Section A.I.2 derives this result formally, showing that the population IV coefficient on the standardized latent predictor converges to $D\beta_X$ rather than β_X , where

$$D = \text{diag}(d_1, \dots, d_q), \quad d_k = \sqrt{1 + \sigma_{U,k}^2}.$$

If each latent predictor has two split-indicator estimates, the first split can be used as the endogenous regressor and the second as its instrument, with C included as exogenous controls in both stages. Under the same split-error independence assumptions, outcome-disturbance exogeneity, and standard rank conditions, the uncorrected IV coefficient on each standardized latent predictor is inflated by the corresponding scale factor d_k . Multiplying the IV coefficient for latent component k by $\sqrt{\widehat{\text{Cor}}(\hat{X}_{\cdot k}^{(1)}, \hat{X}_{\cdot k}^{(2)})}$ therefore recovers the coefficient on the identified latent scale when the componentwise split-half correlations are positive.

As with other IV applications, estimates can become unstable when the split-half correlations are close to zero or nonpositive. Therefore, we recommend conducting standard diagnostics. In the bivariate case, this includes the first-stage F -statistic from regressing $\hat{X}^{(1)}$ on $\hat{X}^{(2)}$ and any observed covariates. With multiple latent components, researchers should inspect componentwise first-stage diagnostics, e.g., for each latent component k , the partial first-stage F -statistic from

regressing $\hat{X}_{\cdot k}^{(1)}$ on $\hat{X}_{\cdot k}^{(2)}$ and C , or an appropriate multivariate weak-IV diagnostic when $q > 1$. In most latent variable applications, the indicators would be expected to have strong correlations since they are assumed to be driven by the same latent variable, so these problems seem unlikely.

4 Illustration: Political Knowledge Predicting Duty to Vote

To make the identification and measurement-error issues concrete, we use the 2024 ANES to compare several approaches in a familiar latent-predictor setting. The latent predictor is political knowledge, measured with a four-item battery covering federal spending, Senate term length, and party control of the House and Senate; the outcome is a seven-point item measuring whether respondents view voting as a duty or a choice. We code knowledge as the number of correct answers and standardize it to mean zero and variance one, matching the identifying restriction used above. Table 2 shows estimates of β_X using several different approaches. The all-item unadjusted OLS estimate is 0.419. As shown in Appendix Figure A.IV.1, the uncorrected OLS slope estimates increase as more indicators are used—a pattern consistent with attenuation.

Table 2: *Political knowledge and duty-to-vote estimator comparisons.*

Measure	Estimator	Slope
Standardized % Correct	OLS	0.419
	Uncorrected IV	1.143
	Correlation-Corrected	0.620
IRT	OLS	0.420
	Uncorrected IV	1.139
	MOC	0.200
	Correlation-Corrected	0.624
	Full Joint	0.668

Note: $N = 3,059$. Slopes are from regressions of the seven-point duty-to-vote measure on standardized political knowledge estimates; higher outcome values indicate a stronger sense that voting is a duty.

Applying the split-indicator correction to the four knowledge items yields $\hat{\beta}^* = 0.62$, roughly 50% larger than the unadjusted estimate. The result is similar when knowledge is estimated with an item-response model rather than an additive score: the uncorrected OLS slope estimate is 0.42 and the correlation-corrected estimate is 0.624. By contrast, the MOC yields a slope of 0.20, while uncorrected split-indicator IV produces estimates above 1 for both approaches of estimating knowledge.

5 Comparing the Performance of Estimators with Simulation

To compare the performance of our correlation-corrected estimator with other approaches, in this section, we conduct Monte Carlo simulations, generating simulated data (W and Y) with parameter values set to point estimates from real data. Specifically, we rely on the 2016 Cooperative

Congressional Election Study (CCES) ($N=38,851$), a survey that includes 30 agree/disagree policy items that we use as indicators to estimate respondent ideology (X). We assume a commonly used item response model

$$\Pr(W_{ij} = 1) = \Phi(X_i\eta_j + \alpha_j), \quad (9)$$

and our dependent variable (Y) in the regression model is reported presidential vote (0=Democratic, 1=Republican; other responses dropped).

Although the type of data and the approach to estimating the latent trait here are quite common in the literature, both the number of observations (N) and the number of indicators (M) in the CCES are much larger than in most latent trait applications in political science and related literatures. This allows us to assess the performance of estimators across a range of N and M values, from small to large. Specifically, we set parameter values equal to estimates from the full CCES data. We then simulate data for $N \in \{500, 1000, 5000\}$ and $M \in \{6, 12, 30\}$, sampling observations and indicators randomly for each value of N and M . We then estimate β_X with these simulated datasets using each of several approaches discussed above (see Appendix III for more details).

Figure 1 summarizes the simulation results. Uncorrected OLS is attenuated toward zero, and the bias is largest when the latent predictor is estimated from few indicators. The uncorrected IV estimator (not shown; see Appendix Figure A.III.6) moves in the opposite direction, over-correcting, often dramatically, because it ignores the identifying rescaling of the latent scores. MOC also performs poorly in this design because it fits outcome regressions to noisier posterior draws of the latent predictor. MO performs well with a large number of indicators, but struggles somewhat with smaller M . By contrast, the split-indicator correction has low bias across the simulated values of N and M and performs similarly to the full joint benchmark. The gains are most substantively important in the common setting where latent traits are measured with modest numbers of indicators. Details are in Appendix A.III.1–A.III.5, which also reports simulation results with multiple latent predictors and covariates, where the same patterns hold (with the correction reducing bias and RMSE relative to uncorrected OLS).

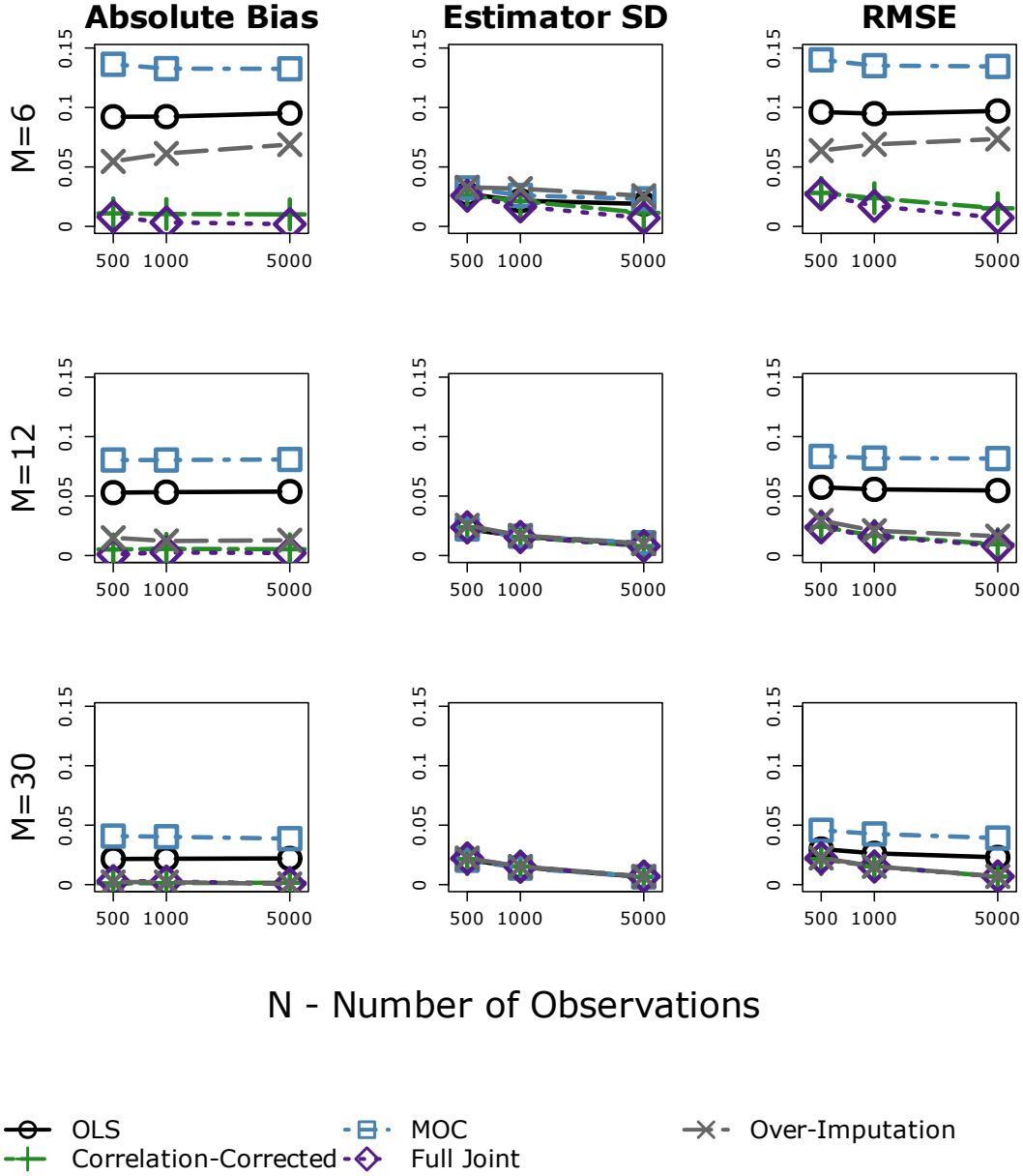


Figure 1: Simulation results varying N and M . Results compare OLS, method of composition (MOC), over-imputation, full joint estimation, and the split-indicator correlation-corrected estimator across the simulation designs described in Section 5.

Appendix Table A.IV.1 shows the percentage change in coefficient estimates when moving from standard OLS to our corrected estimator, for a finer appendix grid of N and M values. While these differences are relatively small when using 30 indicators, we see changes around 20% when $M = 12$, 25% when $M = 10$, and 35% when $M = 6$. Because estimating latent traits from smaller indicator sets is far more common in practice than using 30 or more indicators, these results for M between 6 and 12 are often the empirically relevant regime.

Appendix III, Section A.III.3, reports a coverage analysis: the corrected estimator improves

coverage over unadjusted approaches, though it undercovers for small M ; for large N and M , it reaches nominal coverage, unlike the unadjusted alternatives.

6 Guidance, Limitations, Extensions

These results show that latent predictors create distinctive measurement-error problems: standard regressions attenuate slopes, MOC here worsens bias, and standard IV tends to overcorrect. Our correlation-corrected estimator reduces bias, performs comparably to full joint estimation, and works with any latent-trait estimator, including methods without explicit models or likelihoods.

Applied researchers using latent predictors should check standard estimates against our corrected estimates. In some cases, we may expect attenuation bias to be negligible. For example, estimates of legislator ideology based on hundreds of roll call votes tend to be estimated extremely precisely, at least conditional on the assumed measurement models (Lewis and Poole, 2004). NOMINATE legislator ideology scores estimated separately with the first and second sessions of the 117th U.S. Senate are correlated at 0.99, implying that standard OLS regressions using NOMINATE scores would be attenuated by less than one percent. Appendix A.IV.2 presents an application of democracy predicting economic growth that shows less than a 1% change in the estimated coefficient after using our corrected estimator.

The estimated coefficients on other latent predictors may show non-trivial, albeit not huge, changes when applying our corrected methods. For example, Ansolabehere et al. (2008) finds that correlations between various issue scales when estimated from the two halves of ANES items range from 0.61 to 0.84. This implies that standard OLS slope estimates in these situations may need to be adjusted upwards in magnitude by between 9 and 28 percent.

But in other applications for latent predictors, attenuation bias may be much more severe. It is quite common for latent traits to be estimated based on a small number of indicators. The political knowledge example presented above demonstrates that correcting standard regression estimates using our proposed method can yield a roughly 50% increase in the estimated slope coefficient. With 10–20 policy questions for ideology, the corrected estimates are often 10–25% larger than standard estimates. When using 6 indicators, correcting estimates can increase their magnitude by over one-third.

Future work should optimize indicator splits to reduce estimator variance, and extend the correction beyond linear models. ■

References

- Acemoglu, Daron et al. (2019). “Democracy Does Cause Growth”. In: *Journal of Political Economy* 127.1, pp. 47–100.
- Achen, Christopher H. (1975). “Mass Political Attitudes and the Survey Response”. In: *American Political Science Review* 69.4, pp. 1218–1231. DOI: [10.2307/1955282](https://doi.org/10.2307/1955282).
- Angrist, Joshua D and Alan B Krueger (1995). “Split-sample Instrumental Variables Estimates of the Return to Schooling”. In: *Journal of Business & Economic Statistics* 13.2, pp. 225–235.
- Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton university press.
- Ansolabehere, Stephen, Jonathan Rodden, and James M Snyder (2008). “The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting”. In: *American Political Science Review* 102.2, pp. 215–232.
- Battaglia, Laura et al. (2024). “Inference for Regression with Variables Generated by AI or Machine Learning”. In: *arXiv preprint arXiv:2402.15585*.
- Benoit, Kenneth, Michael Laver, and Slava Mikhaylov (2009). “Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions”. In: *American Journal of Political Science* 53.2, pp. 495–513.
- Bernhard, Michael et al. (2020). “Parties, Civil Society, and the Deterrence of Democratic Defection”. In: *Studies in Comparative International Development* 55, pp. 1–26.
- Blackwell, Matthew, James Honaker, and Gary King (2017). “A Unified Approach to Measurement Error and Missing Data: Overview and Applications”. In: *Sociological Methods & Research* 46.3, pp. 303–341.
- Bound, John, Charles Brown, and Nancy Mathiowetz (2001). “Measurement Error in Survey Data”. In: *Handbook of Econometrics*. Vol. 5. Elsevier, pp. 3705–3843.
- Bürkner, Paul-Christian (2017). “brms: An R Package for Bayesian Multilevel Models Using Stan”. In: *Journal of Statistical Software* 80.1, pp. 1–28. DOI: [10.18637/jss.v080.i01](https://doi.org/10.18637/jss.v080.i01).
- Carpini, Michael X Delli and Scott Keeter (1996). *What Americans Know About Politics and Why It Matters*. Yale University Press.
- Carroll, Raymond J et al. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC.
- Caughey, Devin and Christopher Warshaw (2018). “Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014”. In: *American Political Science Review* 112.2, pp. 249–266.
- (2019). *Public Opinion in Subnational Politics*.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers (2004). “The Statistical Analysis of Roll Call Data”. In: *American Political Science Review* 98.2, pp. 355–370.
- Fong, Christian and Matthew Tyler (2021). “Machine Learning Predictions as Regression Covariates”. In: *Political Analysis* 29.4, pp. 467–484.
- Fuller, Wayne A (2009). *Measurement error models*. John Wiley & Sons.
- Gillard, Jonathan W (2006). “An Historical Overview of Linear Regression with Errors in Both Variables”. In: *Math. School, Cardiff Univ., Wales, UK, Tech. Rep.*
- Gillen, Ben, Erik Snowberg, and Leeat Yariv (2019). “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study”. In: *Journal of Political Economy*

- 127.4, pp. 1826–1863. DOI: [10.1086/701681](https://doi.org/10.1086/701681). eprint: <https://doi.org/10.1086/701681>. URL: <https://doi.org/10.1086/701681>.
- Gosling, Samuel D, Peter J Rentfrow, and William B Swann Jr (2003). “A Very Brief Measure of the Big-Five Personality Domains”. In: *Journal of Research in Personality* 37.6, pp. 504–528.
- Hegre, Havard et al. (2001). “Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816-1992”. In: *The American Political Science Review* 95.1, pp. 33–48.
- Hu, Yingyao and Susanne M Schennach (2008). “Instrumental variable treatment of nonclassical measurement error models”. In: *Econometrica* 76.1, pp. 195–216.
- Jessee, Stephen A (2009). “Spatial Voting in the 2004 Presidential Election”. In: *American Political Science Review* 103.1, pp. 59–81.
- (2010). “Partisan Bias, Political Information and Spatial Voting in the 2008 Presidential Election”. In: *The Journal of Politics* 72.2, pp. 327–340.
- Kastellec, Jonathan P et al. (2015). “Polarizing the Electoral Connection: Partisan Representation in Supreme Court Confirmation Politics”. In: *The Journal of Politics* 77.3, pp. 787–804.
- Kubinec, Robert (July 2019). *Generalized Ideal Point Models for Robust Measurement with Dirty Data in the Social Sciences*. DOI: [10.31219/osf.io/8j2bt_v1](https://doi.org/10.31219/osf.io/8j2bt_v1). URL: osf.io/8j2bt_v1.
- Lax, Jeffrey R, Justin H Phillips, and Adam Zelizer (2019). “The Party or the Purse? Unequal Representation in the US Senate”. In: *American Political Science Review* 113.4, pp. 917–940.
- Lewbel, Arthur (2012). “Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models”. In: *Journal of business & economic statistics* 30.1, pp. 67–80.
- Lewis, Jeffrey B and Keith T Poole (2004). “Measuring Bias and Uncertainty in Ideal Point Estimates Via the Parametric Bootstrap”. In: *Political Analysis* 12.2, pp. 105–127.
- Martin, Andrew D, Kevin M Quinn, and Jong Hee Park (2011). “MCMCpack: Markov chain monte carlo in R”. In: .
- McMann, Kelly M et al. (2016). “Strategies of Validation: Assessing the Varieties of Democracy Corruption Data”. In: *V-Dem Working Paper* 23.
- Pancost, N Aaron and Garrett Schaller (2022). “Measuring Measurement Error”. In: *Available at SSRN* 4045772.
- Pemstein, Daniel et al. (2018). “The Varieties of Democracy Measurement Model: Latent Variable Analysis for Cross-national and Cross-temporal Expert-coded Data”. In: *V-Dem Working Paper* 21.
- Politis, Dimitris N and Joseph P Romano (1994). “Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions”. In: *The Annals of Statistics*, pp. 2031–2050.
- Richardson, Sylvia and Walter R. Gilks (1993). “Conditional Independence Models for Epidemiological Studies with Covariate Measurement Error”. In: *Statistics in Medicine* 12.18, pp. 1703–22.
- Rivers, Douglas (2003). “Identification of Multidimensional Item-response Models”. In: *Typescript. Department of Political Science, Stanford University*.
- Rubin, Donald B (1988). “An overview of multiple imputation”. In: *Proceedings of the survey research methods section of the American statistical association*. Vol. 79. Citeseer, p. 84.
- Spearman, C (1904). “The Proof and Measurement of Association between Two Things”. In: *The American Journal of Psychology* 15.1, pp. 72–101.
- Stefanski, Leonard A and Dennis D Boos (2002). “The Calculus of M-estimation”. In: *The American Statistician* 56.1, pp. 29–38.

- Tai, Yuehong ‘Cassandra’, Yue Hu, and Frederick Solt (2024). “Democracy, Public Support, and Measurement Uncertainty”. In: *American Political Science Review* 118.1, pp. 512–518. DOI: [10.1017/S0003055422000429](https://doi.org/10.1017/S0003055422000429).
- Tanner, M.A. (2012). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Series in Statistics. Springer New York. ISBN: 9781468401929. URL: <https://books.google.com/books?id=d1ThBwAAQBAJ>.
- Tausanovitch, Chris and Christopher Warshaw (2013). “Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities”. In: *The Journal of Politics* 75.2, pp. 330–342.
- Treier, Shawn and Simon Jackman (2008). “Democracy as a Latent variable”. In: *American Journal of Political Science* 52.1, pp. 201–217.

Contents

1	Measurement Error and Bias in Linear Regression	3
1.1	Standard Classical Measurement Error	3
1.2	Measurement Error in Latent Predictors	3
2	Existing Approaches and Their Limitations	5
2.1	No Adjustment	5
2.2	Full Joint Estimation	5
2.3	Instrumental Variables	5
2.4	Method of Composition (MOC)	6
2.5	Multiple Over-Imputation	7
3	A Correlation-Corrected Estimator Using Split Indicators	8
4	Illustration: Political Knowledge Predicting Duty to Vote	12
5	Comparing the Performance of Estimators with Simulation	12
6	Guidance, Limitations, Extensions	15
	Appendix I: Technical Derivations and Estimator Properties	19
	Appendix II: Evaluating the Sensitivity of the Corrected Estimator to Assumption Violations	32
	Appendix III: Additional Simulation Details & Results	36
	Appendix IV: Additional Empirical Results	48
	Appendix V: Measurement Error and Latent Predictors in the Political Science Literature	52
	Appendix VI: A <code>1pmec</code> Package Tutorial	53

Appendix I: Technical Derivations and Estimator Properties

This section illustrates how the standard IV correction for measurement error yields estimates biased upward in magnitude in the case of latent predictors. It also demonstrates how, in the split-indicator setup, the correlation between $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ can be used to adjust the IV estimator to provide consistent estimates. Finally, it is shown that the correlation-corrected IV estimate is equivalent to the correlation-corrected OLS estimate in the bivariate case before moving on to an analysis of multivariate corrected IV, a stylized analysis of the Method of Composition, and a discussion of alternative identification restrictions.

A.I.1 Consistency of the Bivariate Corrected Estimator

Assume the true latent trait X has zero mean and unit variance. Additionally, assume the split-specific measurement errors $U^{(1)}$ and $U^{(2)}$ are mean-zero, mutually independent, possess a common variance σ_U^2 , and are independent of both X and the regression error ε .

For a fixed split, write $\tilde{X}^{(j)} = X + U^{(j)}$ for $j \in \{1, 2\}$. After aligning each split to the target latent direction, consider the population-standardized representation

$$\hat{X}^{(j)} = \frac{\tilde{X}^{(j)}}{\sqrt{1 + \sigma_U^2}}.$$

The implemented scores in a given sample of units $i \in \{1, \dots, N\}$ use sample centering and scaling,

$$\hat{X}_i^{(j)} = \frac{\tilde{X}_i^{(j)} - \bar{\tilde{X}}^{(j)}}{s_{\tilde{X}^{(j)}}}.$$

Under the maintained i.i.d. finite-second-moment assumptions, $\bar{\tilde{X}}^{(j)} \xrightarrow{p} 0$ and $s_{\tilde{X}^{(j)}} \xrightarrow{p} \sqrt{1 + \sigma_U^2}$, so replacing the sample-standardized score by the population-standardized expression changes the relevant OLS slopes and split-half correlations only by $o_p(1)$, meaning their difference converges in probability to zero as the sample size approaches infinity. Then (assuming the bivariate OLS regression model $Y = \beta_0 + X\beta_X + \varepsilon$):

$$\begin{aligned} \text{plim } \hat{\beta}_{\hat{X}^{(1)}} &= \frac{\text{Cov}(Y, \hat{X}^{(1)})}{\text{Var}(\hat{X}^{(1)})} = \text{Cov}\left(\beta_0 + X\beta_X + \varepsilon, \frac{X + U^{(1)}}{\sqrt{1 + \sigma_U^2}}\right) \\ &= \frac{\beta_X \text{Var}(X) + \text{Cov}(\varepsilon, X) + \text{Cov}(\varepsilon, U^{(1)})}{\sqrt{1 + \sigma_U^2}} \\ &= \frac{\beta_X}{\sqrt{1 + \sigma_U^2}}, \end{aligned}$$

where the second equality uses $\text{Var}(\hat{X}^{(1)}) = 1$ and final equality uses $\text{Var}(X) = 1$, $\text{Cov}(\varepsilon, X) = 0$, and independence of $U^{(1)}$ from (X, ε) . Similarly,

$$\rho \equiv \text{Cor}(\tilde{X}^{(1)}, \tilde{X}^{(2)}) = \frac{\overbrace{\text{Cov}(X + U^{(1)}, X + U^{(2)})}^{\substack{=\text{Var}(X)=1, \\ \text{cross terms again vanish}}}}{\sqrt{\text{Var}(X + U^{(1)}) \text{Var}(X + U^{(2)})}} = \frac{1}{1 + \sigma_U^2}.$$

Equivalently, ρ is the probability limit of the sample correlation $\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})$ of the standardized split scores. Therefore,

$$\text{plim} \frac{\hat{\beta}_{\hat{X}^{(1)}}}{\sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}} = \frac{\beta_X / \sqrt{1 + \sigma_U^2}}{\sqrt{1/(1 + \sigma_U^2)}} = \beta_X,$$

By demonstrating that the probability of the adjusted slope estimate equals the true parameter β_X , the equation establishes that this correlation-corrected approach yields a statistically consistent estimator.

Now, now, another common method for correcting for measurement error in a predictor is the use of instrumental variables (IV). Although its use today is often motivated by causal identification, IV can also, under certain assumptions, produce consistent estimates of slope coefficients when a predictor is measured with error.¹⁰

In the standard observable predictor setting, if X is measured with error following the setup above and we seek to estimate the slope coefficient on X in a linear regression predicting some other variable Y , we can correct for the resulting attenuation bias by leveraging an instrument Z with the following properties: (i) $\text{Cov}(X, Z) \neq 0$; (ii) $\text{Cov}(Z, U) = 0$; (iii) $\text{Cov}(Z, \varepsilon) = 0$. In other words, a valid instrument in this setting must be correlated with the true predictor X and be uncorrelated with both the measurement error in X and the error term in the regression predicting Y with X .

Following Angrist and Pischke (2009), the two-stage least squares (2SLS) IV estimator of the slope if we were able to observe $\tilde{X}$ could be written as (in this observable [non-latent] predictor set up):

$$\begin{aligned} \hat{\beta}_{\text{IV}, \tilde{X}} &= \frac{\hat{\beta}_{Y \sim Z}}{\hat{\beta}_{\tilde{X} \sim Z}} = \frac{\text{Cov}(Y, Z) / \text{Var}(Z)}{\text{Cov}(\tilde{X}, Z) / \text{Var}(Z)} = \frac{\text{Cov}(X\beta_X + \varepsilon, Z)}{\text{Cov}(X + U, Z)} \\ &= \frac{\beta_X \text{Cov}(X, Z) + \text{Cov}(\varepsilon, Z)}{\underbrace{\text{Cov}(X, Z) + \text{Cov}(U, Z)}_{\substack{\text{under valid-IV orthogonality} \\ \text{Cov}(\varepsilon, Z) = \text{Cov}(U, Z) = 0}}} = \frac{\beta_X \text{Cov}(X, Z)}{\text{Cov}(X, Z)}, \end{aligned} \quad (\text{A.I.1})$$

since by assumption $\text{Cov}(Z, \varepsilon) = 0$ and $\text{Cov}(Z, U) = 0$. Hence it follows that in this observable predictor setup

$$\text{plim} \hat{\beta}_{\text{IV}, \tilde{X}} = \beta_X. \quad (\text{A.I.2})$$

¹⁰Although we focus on measurement error here, note that in a causal framework, IV can be used to address measurement error as well as simultaneity and omitted variable bias under suitable assumptions (see Pancost and Schaller, 2022).

Conversely, in the latent predictor case, we can obtain a suitable instrumental variable by splitting the indicators used to estimate the latent variable into two sets and estimating the latent trait separately using each of these two sets of indicators, calling these estimates $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ as above. We can then use $\hat{X}^{(1)}$ as the regressor and $\hat{X}^{(2)}$ as the instrument, or vice-versa.

Because these are both measures of the same latent trait, they should be highly correlated with each other (which we can also check empirically) and also highly correlated with the true latent trait X (which we cannot check empirically).¹¹

We assume that the split-specific measurement errors are orthogonal across splits on the unstandardized latent scale.¹² For the direction with $\hat{X}^{(1)}$ as the predictor and $\hat{X}^{(2)}$ as the instrument, the IV conditions require

$$\text{Cov}(\hat{X}^{(2)}, U^{(1)}) = 0, \quad \text{Cov}(\hat{X}^{(2)}, \varepsilon) = 0, \quad \text{Cov}(\hat{X}^{(2)}, \hat{X}^{(1)}) \neq 0.$$

The reverse direction requires the analogous conditions $\text{Cov}(\hat{X}^{(1)}, U^{(2)}) = 0$, $\text{Cov}(\hat{X}^{(1)}, \varepsilon) = 0$, and relevance of $\hat{X}^{(1)}$ for $\hat{X}^{(2)}$. With $\tilde{X}^{(1)} = X + U^{(1)}$ and $\tilde{X}^{(2)} = X + U^{(2)}$, these conditions follow from the stronger maintained model in which $\text{Cov}(X, \varepsilon) = 0$ and $(U^{(1)}, U^{(2)})$ are mean-zero, mutually independent, and independent of (X, ε) . Here, $U^{(1)}$ is the error in the unstandardized true-scale measure $\tilde{X}^{(1)}$, not $X - \hat{X}^{(1)}$, because $\hat{X}^{(1)}$ has been rescaled by the identifying normalization.

In contrast to the IV setup with an observable predictor, our estimates $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ are not on the same scale as X itself. If we had $\tilde{X}^{(1)} = X + U^{(1)}$ to use as our predictor, then the IV results in Equations A.I.1 and A.I.2 would hold. In our case, however, we use as our predictor $\hat{X}^{(1)} = \frac{\tilde{X}^{(1)}}{\sqrt{1 + \sigma_U^2}}$, i.e. a rescaled version of the unstandardized split measure since it is standardized by dividing by $\sigma_{\tilde{X}^{(1)}} = \sqrt{1 + \sigma_U^2}$. This implies that our IV estimate using $\hat{X}^{(1)}$ will have the following relationship with the IV estimate using $\tilde{X}^{(1)}$:

$$\begin{aligned} \hat{\beta}_{\text{IV}, \hat{X}^{(1)} | \hat{X}^{(2)}} &= \frac{\text{Cov}(Y, \hat{X}^{(2)})}{\text{Cov}(\hat{X}^{(1)}, \hat{X}^{(2)})} \\ &= \hat{\beta}_{\text{IV}, \tilde{X}^{(1)} | \tilde{X}^{(2)}} \underbrace{\frac{1/\sqrt{1 + \sigma_U^2}}{1/(1 + \sigma_U^2)}}_{=\sqrt{1 + \sigma_U^2}} \\ &= \hat{\beta}_{\text{IV}, \tilde{X}^{(1)} | \tilde{X}^{(2)}} \sqrt{1 + \sigma_U^2}. \end{aligned} \tag{A.I.3}$$

and hence

$$\text{plim } \hat{\beta}_{\text{IV}, \hat{X}^{(1)} | \hat{X}^{(2)}} = \beta_X \sqrt{1 + \sigma_U^2}. \tag{A.I.4}$$

In this way, using IV to correct for measurement error in this latent predictor setting actually *over*-corrects for the attenuation bias in standard linear regression, resulting in coefficient estimates that are inflated in magnitude on average since $\sqrt{1 + \sigma_U^2} > 1$ whenever $\sigma_U^2 > 0$, i.e., whenever there is

¹¹In the latent predictor case, different indicators of the same latent trait, and especially scales constructed based on multiple indicators, are typically highly correlated with each other (see e.g. Ansolabehere et al., 2008), suggesting that the weak instruments problem should rarely be severe in practice.

¹²This is analogous to the conditional independence assumption used in many latent traits models, i.e., that conditional on X , the indicators are independent of each other.

any measurement error in the predictor. Note that here, we are using the notation $A | B$ to indicate that A is our predictor and B is our instrumental variable.

This implies a further correction that can be accomplished by multiplying the IV estimate by the reciprocal of this factor, so that

$$\hat{\beta}_{IV, \hat{X}^{(1)} | \hat{X}^{(2)}}^* = \frac{\hat{\beta}_{IV, \hat{X}^{(1)} | \hat{X}^{(2)}}}{\sqrt{1 + \sigma_U^2}} = \hat{\beta}_{IV, \hat{X}^{(1)} | \hat{X}^{(2)}} \sqrt{\rho}. \quad (\text{A.I.5})$$

With this correction in place, the same asymptotic logic as earlier implies

$$\text{plim } \hat{\beta}_{IV, \hat{X}^{(1)} | \hat{X}^{(2)}}^* = \beta_X. \quad (\text{A.I.6})$$

Hence, by adjusting the split-half IV estimate to account for the identifying restriction used for the estimated latent traits (which itself is based on the measurement error variance σ_U^2), we can obtain a consistent estimate of the linear regression slope coefficient on the true latent X predicting Y .

Of course, the correction used in $\hat{\beta}_{IV, \hat{X}^{(1)} | \hat{X}^{(2)}}^*$ is based on the variance of the measurement error (on the true latent variable's scale) σ_U^2 which we do not typically know. As discussed in the previous section, however, we use these replicate measures, $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$, to learn about σ_U^2 and, consequently, about the split-half reliability ρ . Since the sample correlation between these replicate measurements consistently estimates ρ , we can calculate our corrected estimator as the product of the IV slope estimate and the square root of the correlation:

$$\hat{\beta}_{IV, \hat{X}^{(1)} | \hat{X}^{(2)}}^* = \hat{\beta}_{IV, \hat{X}^{(1)} | \hat{X}^{(2)}} \sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}. \quad (\text{A.I.7})$$

Another way of understanding the corrected IV estimator is to note that if we transformed our measures $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ from having variance 1 to having variance $1 + \sigma_U^2$, which would effectively mean we are in the standard measurement error setup, then the standard IV estimator would provide consistent estimates.¹³ A convenient way to do this is to estimate σ_U^2 with $[1/\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)}) - 1]$, because $\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})$ consistently estimates $\frac{1}{1 + \sigma_U^2}$. Then, note that since we want the variance of our predictor to be $1 + \sigma_U^2$ we can multiply it by $1/\sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}$ and use this transformed predictor in a standard IV analysis.¹⁴ Obviously, multiplying the predictor by $1/\sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}$ has the result of dividing the estimated slope coefficient by the same factor, which is equivalent to multiplying it by $\sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}$ as in Equation A.I.7 above.

Also as above, we may want to calculate our corrected IV estimator over all equally sized partitions of the indicators for estimating $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$. For each such partition, take the average of the corrected estimate with $\hat{X}^{(1)}$ as the regressor and $\hat{X}^{(2)}$ as the instrument and the corrected estimate with $\hat{X}^{(2)}$ as the regressor and $\hat{X}^{(1)}$ as the instrument, then choose as our estimator the median of all such partition averages. In fact, this estimator, which we call $\hat{\beta}_{IV}^*$, is equivalent to $\hat{\beta}_{OLS}^*$

¹³We actually would not need to apply this transformation to whichever one is used as the instrument, since any affine transformation of the instrument will cancel out in the expression for the IV estimator.

¹⁴Since the ratio of the uncorrected OLS and uncorrected IV estimators provides information about the split-half reliability ρ in this standardized split model (see Pancost and Schaller, 2022), yet another equivalent estimator can be constructed based on the square root of the product of the baseline OLS and uncorrected IV estimators using the partitioned $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$. Since this approach is less intuitive, we do not discuss it here.

(see below). In other words, the correlation-adjusted OLS estimator and the correlation-adjusted IV estimator are equivalent and below we refer to them together as the corrected estimator.

Despite the equivalence shown between $\hat{\beta}_{OLS}^*$ and $\hat{\beta}_{IV}^*$ in the bivariate case, one advantage of the IV approach is the ease with which it can be extended to multivariate regression (discussed more later). If one predictor is measured with error, we may either have other “control” variables or the predictor measured with error may itself be a control variable in which case adjusting for this may be important to prevent bias in the other coefficient(s) of interest.¹⁵ We may even have multiple predictors measured with error, and believe that we have instruments that are suitable to address this. In both the OLS and IV cases, we employ the bootstrap to assess uncertainty, as this row-level resampling allows us to incorporate uncertainty not only in the coefficient estimate but also in the correction factor itself in finite samples.

Finally, the framework here assumes classical (mean-zero, independent) measurement error. When non-classical error is a concern, identification can still sometimes be recovered under additional conditions; see, for example, Hu and Schennach (2008). In settings where external instruments are unavailable, instruments can sometimes be constructed from heteroskedasticity (Lewbel, 2012); these ideas complement our split-indicator strategy.

To see that $\hat{\beta}_{IV}^*$ is equivalent to $\hat{\beta}_{OLS}^*$ note that $\hat{\beta}_{OLS, \hat{X}^{(1)} \sim \hat{X}^{(2)}}$, the linear regression slope coefficient estimate when predicting $\hat{X}^{(1)}$ with $\hat{X}^{(2)}$, is equal to

$$\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)}) \sqrt{\frac{\text{Var}(\hat{X}^{(1)})}{\text{Var}(\hat{X}^{(2)})}} = \widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)}),$$

since the variances of both $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ are equal to 1 given that each is standardized. This also implies that $\hat{\beta}_{OLS, \hat{X}^{(2)} \sim \hat{X}^{(1)}} = \widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})$ since $\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)}) = \widehat{\text{Cor}}(\hat{X}^{(2)}, \hat{X}^{(1)})$. Therefore, we can write the uncorrected IV slope coefficient using $\hat{X}^{(1)}$ as our predictor and $\hat{X}^{(2)}$ as our instrument as

$$\begin{aligned} \hat{\beta}_{IV, \hat{X}^{(1)} | \hat{X}^{(2)}} &= \frac{\hat{\beta}_{OLS, Y \sim \hat{X}^{(2)}}}{\underbrace{\hat{\beta}_{OLS, \hat{X}^{(1)} \sim \hat{X}^{(2)}}}_{=\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}} \\ &\quad \text{because both split scores have variance 1} \\ &= \frac{\hat{\beta}_{OLS, Y \sim \hat{X}^{(2)}}}{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}. \end{aligned}$$

¹⁵Note in the latter case we may not care about appropriately adjusting the IV coefficient on the predictor that is measured with error.

and hence we have

$$\begin{aligned}
\hat{\beta}_{\text{IV}, \hat{X}^{(1)} | \hat{X}^{(2)}}^* &= \hat{\beta}_{\text{IV}, \hat{X}^{(1)} | \hat{X}^{(2)}} \sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})} \\
&= \frac{\hat{\beta}_{\text{OLS}, Y \sim \hat{X}^{(2)}}}{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})} \sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})} \\
&= \frac{\hat{\beta}_{\text{OLS}, Y \sim \hat{X}^{(2)}}}{\sqrt{\widehat{\text{Cor}}(\hat{X}^{(1)}, \hat{X}^{(2)})}} \\
&= \hat{\beta}_{\text{OLS}, \hat{X}^{(2)}}^*
\end{aligned} \tag{A.I.8}$$

For a single IV direction, the corrected IV estimator with $\hat{X}^{(1)}$ as the regressor and $\hat{X}^{(2)}$ as the instrument equals the corrected OLS estimator using $\hat{X}^{(2)}$ as the predictor. Therefore, the symmetrized within-partition IV average equals the symmetrized within-partition OLS average.

Asymptotic Normality and Bootstrap Validity. Although we have discussed consistency, we can also now consider asymptotic Normality. Fix a partition p . More generally, if the estimator is reported for a finite collection of partitions $\mathcal{P} = \{p_1, \dots, p_P\}$, treat $\mathcal{P}$ as fixed and not growing with n ; if partitions are randomly generated in implementation, the formal argument is conditional on the realized set of partitions, which is then held fixed. Let O_i denote the full row-level data for unit i : the outcome, any observed covariates, and the complete indicator vector from which the split-specific latent scores are constructed. Let $\hat{\zeta}_p$ denote the finite-dimensional vector of sample summaries for partition p (e.g., the split-half correlation and uncorrected moments) with population limit ζ_{0p} , and let $\hat{\beta}_p^*$ denote the resulting correlation-corrected estimator targeting the true slope β_X .

Now, assume: (i) $O_1, \dots, O_n$ are i.i.d.; (ii) the low-dimensional first- and second-stage summaries entering the corrected estimator have finite $2 + \delta$ moments for some $\delta > 0$ (e.g., finite fourth moments suffice); (iii) for the implemented latent-score procedure on split p , after imposing the normalization and sign conventions, the finite-dimensional vector $\hat{\zeta}_p$ of partition-specific quantities entering the correction and second-stage estimator admits a joint asymptotically linear expansion: specifically:

$$\sqrt{n}(\hat{\zeta}_p - \zeta_{0p}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\zeta,p}(O_i) + o_p(1), \quad \mathbb{E}[\psi_{\zeta,p}(O_i)] = 0, \quad \mathbb{E}[\|\psi_{\zeta,p}(O_i)\|^2] < \infty$$

for influence function $\psi_{\zeta,p}(\cdot)$; (iv) every population split-half correlation entering the correction is positive and bounded away from zero; and (v) the relevant population regression/IV moment matrices are full rank, with sample analogues nonsingular with probability approaching one.

Here, $\psi_{\zeta,p}(\cdot)$ represents the influence function, which isolates and quantifies the individual contribution of a single observation O_i to the overall estimation error. We can justify applying this linear representation because the components of $\hat{\zeta}_p$ —specifically, the sample variances, covariances, and first-stage regression moments—are standard Z-estimators (or smooth functions of sample moments). Under the regularity conditions specified earlier (such as finite fourth moments), the theory of M-estimation guarantees that such estimators are regular and asymptotically linear.

Consequently, via a stacked estimating-equation expansion, the complex, non-linear error of $\hat{\zeta}_p$ can be mathematically linearized into a simple sum of independent, mean-zero individual compo-

nents (ψ). This linearization is crucial: by expressing the error as a sum of independent variables with finite variance, we satisfy the conditions for the Central Limit Theorem, guaranteeing that the intermediate ingredients $\hat{\zeta}_p$ converge to a joint multivariate Normal distribution. Because the final corrected estimator $\hat{\beta}_p^*$ is a smooth, continuously differentiable function of these Normally distributed ingredients, we can subsequently invoke the Delta Method to establish its asymptotic Normality.

Notably, these high-level conditions does not require the full first-stage latent-score object itself to be finite-dimensional; unit-specific latent scores or other nuisance components may have dimension growing with N , provided the low-dimensional summaries entering the corrected estimator are root- N regular. In standard finite-dimensional M/Z-estimation settings, this condition follows from a stacked estimating-equation argument, but we use that only as a sufficient special case rather than as a generic characterization of all latent-score procedures.

Under this set of conditions, the generated-regressor first stage affects the corrected estimator only through $\hat{\zeta}_p$. The partition-specific corrected estimator can then be written as

$$\hat{\beta}_p^* = \phi_p(\hat{\zeta}_p),$$

where ϕ_p is continuously differentiable in a neighborhood of ζ_{0p} , the population limit of the sample summaries for partition p , because ϕ_p is a smooth algebraic function of sample moments and the population split-half correlation is strictly positive (bounded away from zero). In the bivariate OLS case, ϕ_p is a smooth function of sample variances and covariances; in the IV case, it is a smooth function of the corresponding first-stage, reduced-form, and covariance moments; and in either case the square-root split-half correction is smooth because the population split-half correlation is bounded away from zero. Therefore,

$$\sqrt{n}(\hat{\beta}_p^* - \beta_X) = \nabla \phi_p(\zeta_{0p})^\top \sqrt{n}(\hat{\zeta}_p - \zeta_{0p}) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \sigma_p^*),$$

where $(\sigma_p^*)^2$ denotes the asymptotic variance of the corrected estimator for partition p , defined as $\nabla \phi_p(\zeta_{0p})^\top \mathbb{E}[\psi_{\zeta,p}(O_i)\psi_{\zeta,p}(O_i)^\top] \nabla \phi_p(\zeta_{0p})$ (Stefanski and Boos, 2002). Hence, the fixed-partition corrected estimator is root- N consistent and asymptotically Normal. The same argument applies to vector-valued latent-predictor coefficients by replacing derivatives with Jacobians.

Now, let $\hat{\beta}_p^{*,b}$ denote the row-bootstrap analog computed from a bootstrap sample $O_1^b, \dots, O_n^b$ drawn with replacement from the empirical distribution of the rows. For the bootstrap to target the sampling distribution of the implemented estimator, each bootstrap replicate must rerun the full procedure using the same fixed partition or fixed finite set of partitions: re-estimate the split-specific latent scores from the bootstrap sample, re-center/re-scale and sign-align them using the same convention as in the original sample, recompute the split-half correlation(s), recompute the corrected regression or IV estimate(s), and then apply the same partition aggregation rule. Holding the latent scores, the split-half correlation, or the correction factor fixed across bootstrap samples would condition on part of the estimator and generally understate uncertainty. If, in addition, the ordinary row bootstrap consistently reproduces the joint law of $\sqrt{n}(\hat{\zeta}_p - \zeta_{0p})$ for the implemented latent-score procedure under full re-estimation, the same normalization/sign-alignment convention, and the same realized fixed partition set, then the smooth-functional argument yields bootstrap

consistency for any fixed partition:

$$\sup_t \left| \Pr^* \left\{ \sqrt{n}(\hat{\beta}_p^{*,b} - \hat{\beta}_p^*) \leq t \right\} - \Pr \left\{ \sqrt{n}(\hat{\beta}_p^* - \beta_X) \leq t \right\} \right| \xrightarrow{P} 0,$$

where $\Pr^*(\cdot)$ denotes the bootstrap law conditional on the observed sample. We state bootstrap validity for the fixed-partition estimator as an estimator-specific high-level result, rather than as a generic theorem covering all possible latent-score estimators. Under the same positive-correlation assumption, the event that a valid partition is dropped because its estimated split-half correlation is nonpositive has probability $o(1)$, so such screening does not affect the first-order argument. If $\mathcal{P} = \{p_1, \dots, p_P\}$ is a fixed finite collection of partitions, stacking the corresponding low-dimensional summaries yields joint asymptotic Normality of $(\hat{\beta}_{p_1}^*, \dots, \hat{\beta}_{p_P}^*)$. Hence any aggregation rule $g(\hat{\beta}_{p_1}^*, \dots, \hat{\beta}_{p_P}^*)$ that is continuously differentiable at $(\beta_X, \dots, \beta_X)$ —for example, a fixed weighted average over pre-specified partitions—is also asymptotically Normal, and the same ordinary row bootstrap is valid for that smooth aggregation under the same high-level bootstrap condition.

The finite-partition median aggregation rule requires a caveat. Let

$$\hat{\beta}_{\text{med}} = \text{med}\{\hat{\beta}_{p_1}^*, \dots, \hat{\beta}_{p_P}^*\},$$

with P fixed. When all valid partitions share the same population limit β_X , the median map is evaluated at the non-smooth point $(\beta_X, \dots, \beta_X)$. Accordingly, the preceding smooth delta-method/bootstrap argument does not by itself justify the ordinary n -out-of- n row bootstrap for $\hat{\beta}_{\text{med}}$. If

$$\sqrt{n}(\hat{\beta}_{p_1}^* - \beta_X, \dots, \hat{\beta}_{p_P}^* - \beta_X) \xrightarrow{d} (Z_1, \dots, Z_P),$$

with joint Gaussian limit, then

$$\sqrt{n}(\hat{\beta}_{\text{med}} - \beta_X) \xrightarrow{d} \text{med}(Z_1, \dots, Z_P),$$

up to the convention used when P is even. This limit is generally non-Gaussian, although the estimator remains root- N consistent. We therefore use the ordinary row bootstrap for the median-aggregated estimator as a practical resampling approximation to the full implemented algorithm, not as a generic theorem implied by the smooth-bootstrap result above. A fully formal generic route is either to replace the median by a smooth aggregation over a fixed finite set of partitions, for which the ordinary row bootstrap applies directly, or to use subsampling or an m -out-of- n bootstrap for the finite-partition median under the additional regularity conditions required for nonsmooth functionals (Politis and Romano, 1994). For further discussion, see [A.III.3](#).

A.I.2 Multivariate Corrected IV

We now give the multivariate version of the split-indicator IV correction. All variables are taken to be centered and an intercept assumed to be present. Let $\beta_X \in \mathbb{R}^q$ and consider

$$Y_i = X_i^\top \beta_X + C_i^\top \gamma + \varepsilon_i,$$

where $X_i = (X_{i1}, \dots, X_{iq})^\top$ is a vector of identified latent predictors and C_i is a vector of p observed background covariates. The components of X_i need not be mutually independent, and X_i may be correlated with C_i . The only scale normalization imposed here is componentwise: $\text{Var}(X_{ik}) = 1$ for $k = 1, \dots, q$, with directions aligned to the target latent directions.

For split $j \in \{1, 2\}$, write the unstandardized split measure as

$$\tilde{X}_i^{(j)} = X_i + U_i^{(j)},$$

and the corresponding identified split score (componentwise standardized to unit variance, but not yet corrected) as

$$\hat{X}_i^{(j)} = D^{-1} \tilde{X}_i^{(j)} = D^{-1}(X_i + U_i^{(j)}), \quad D = \text{diag}(d_1, \dots, d_q), \quad d_k = \sqrt{1 + \sigma_{U,k}^2}.$$

Here $\sigma_{U,k}^2 = \text{Var}(U_{ik}^{(j)})$, assumed equal across $j = 1, 2$ for a given component k . Thus, each split score is standardized component by component to have variance of one.

Then, let

$$R_i = \begin{pmatrix} \hat{X}_i^{(1)} \\ C_i \end{pmatrix}, \quad Z_i = \begin{pmatrix} \hat{X}_i^{(2)} \\ C_i \end{pmatrix}.$$

The split score $\hat{X}_i^{(1)}$ is treated as endogenous, while C_i is included as an exogenous regressor and therefore also appears in the instrument vector. Assume finite second moments; from the population IV conditions:

$$\begin{aligned} \mathbb{E}[Z_i \varepsilon_i] &= 0 \quad (\text{exogeneity}), & \mathbb{E}[Z_i \{U_i^{(1)}\}^\top] &= 0 \quad (\text{no cross-split contamination}), \\ Q_{ZR} &\equiv \mathbb{E}[Z_i R_i^\top] \text{ is nonsingular} \quad (\text{relevance condition}). \end{aligned}$$

These conditions follow, for example, if the split errors are mean-zero, mutually independent across splits, independent of X_i , C_i , and ε_i , and the usual first-stage rank condition holds.

Following the standard IV setup, the population IV coefficient vector, $\kappa^{IV} = (\delta_X^{IV\top}, \gamma_C^{IV\top})^\top$, is the unique solution to the moment condition equation:

$$\mathbb{E}[Z_i \{Y_i - R_i^\top \kappa\}] = \mathbf{0}_{(q+p) \times 1}. \quad (\text{A.I.9})$$

To identify the population IV estimand, evaluate the moment condition at the parameter vector implied by the latent-score rescaling:

$$\kappa_0 = \begin{pmatrix} D\beta_X \\ \gamma \end{pmatrix}.$$

When we evaluate the IV residual at $\kappa_0 = ((D\beta_X)^\top, \gamma^\top)^\top$, so that $R_i^\top \kappa_0 = \{\hat{X}_i^{(1)}\}^\top D\beta_X + C_i^\top \gamma$,

$$\begin{aligned}
Y_i - R_i^\top \kappa_0 &= X_i^\top \beta_X + \underbrace{C_i^\top \gamma - C_i^\top \gamma}_{=0} + \varepsilon_i - \{\hat{X}_i^{(1)}\}^\top D\beta_X \\
&= X_i^\top \beta_X + \varepsilon_i - \underbrace{\{D^{-1}(X_i + U_i^{(1)})\}^\top D\beta_X}_{\substack{=(X_i + U_i^{(1)})^\top D^{-1}D\beta_X = (X_i + U_i^{(1)})^\top \beta_X \\ \text{since } D \text{ is diagonal so } (D^{-1})^\top = D^{-1} \text{ and } D^{-1}D = I}} \\
&= \underbrace{X_i^\top \beta_X - (X_i + U_i^{(1)})^\top \beta_X}_{= -\{U_i^{(1)}\}^\top \beta_X} + \varepsilon_i \\
&= \varepsilon_i - \{U_i^{(1)}\}^\top \beta_X.
\end{aligned}$$

Substituting this residual into the IV moment function and using the exogeneity and no-cross-split-contamination assumptions,

$$\begin{aligned}
\mathbb{E}[Z_i\{Y_i - R_i^\top \kappa_0\}] &= \mathbb{E}[Z_i\{\varepsilon_i - \{U_i^{(1)}\}^\top \beta_X\}] \\
&= \underbrace{\mathbb{E}[Z_i\varepsilon_i] - \mathbb{E}[Z_i\{U_i^{(1)}\}^\top \beta_X]}_{\substack{= 0 \text{ by exogeneity} \\ \text{and no cross-split contamination}}} \\
&= 0.
\end{aligned}$$

Thus, κ_0 satisfies the population IV moment condition in Equation A.I.9. Since Q_{ZR} , the population instrument–regressor cross-moment matrix, is non-singular under the assumptions outlined above, the population IV solution is unique, so

$$\delta_X^{IV} = D\beta_X \quad \text{and} \quad \gamma_C^{IV} = \gamma.$$

Uncorrected IV therefore recovers the correct coefficient on the observed controls here in the multivariate case, but it recovers $D\beta_X$, not β_X , for the standardized latent predictors. The reason is the scale effect: standard IV would recover β_X if the unstandardized latent-scale regressor $\tilde{X}_i^{(1)} = X_i + U_i^{(1)}$ were used, but the observed latent score is $D^{-1}\tilde{X}_i^{(1)}$, so its IV coefficient is inflated componentwise by D .

It remains to estimate the scale correction D^{-1} . For each component k , define the population split-half correlation

$$\rho_k \equiv \text{Cor}(\hat{X}_{ik}^{(1)}, \hat{X}_{ik}^{(2)}),$$

where the correlation is taken across observations i . Because correlations are unchanged by componentwise affine standardization,

$$\begin{aligned}
\rho_k &= \frac{\text{Cov}(X_{ik} + U_{ik}^{(1)}, X_{ik} + U_{ik}^{(2)})}{\sqrt{\text{Var}(X_{ik} + U_{ik}^{(1)}) \text{Var}(X_{ik} + U_{ik}^{(2)})}} \\
&= \frac{1}{1 + \sigma_{U,k}^2} = d_k^{-2},
\end{aligned}$$

where the second equality uses orthogonality of X_{ik} , $U_{ik}^{(1)}$, and $U_{ik}^{(2)}$, together with $\text{Var}(X_{ik}) = 1$ and $\text{Var}(U_{ik}^{(1)}) = \text{Var}(U_{ik}^{(2)}) = \sigma_{U,k}^2$. Thus, the reciprocal scale factor for component k is

$$d_k^{-1} = \sqrt{\rho_k}.$$

Equivalently, the full correction matrix is

$$D^{-1} = \text{diag}(d_1^{-1}, \dots, d_q^{-1}) = \text{diag}(\sqrt{\rho_1}, \dots, \sqrt{\rho_q}).$$

These population identities translate directly into the sample estimator. The first step is ordinary 2SLS using the observed, standardized split scores: regressors are $(\hat{X}^{(1)}, C)$, instruments are $(\hat{X}^{(2)}, C)$, and the observed covariates C are included in both because they are treated as exogenous. The second step is a post-estimation rescaling of only the coefficients on the latent predictors, using the estimated split-half correlations. Formally, let

$$\hat{\kappa}^{IV} = \begin{pmatrix} \hat{\beta}_{\hat{X}^{(1)}|\hat{X}^{(2)}}^{IV} \\ \hat{\gamma}_C^{IV} \end{pmatrix}$$

denote the sample IV/2SLS estimator using regressors $(\hat{X}^{(1)}, C)$ and instruments $(\hat{X}^{(2)}, C)$.

In the exactly identified case, there are as many instruments as regressors, so the IV estimator is the solution to the sample analog of the population moment condition,

$$\mathbb{P}_n[Z_i\{Y_i - R_i^\top \kappa\}] = 0,$$

where $\mathbb{P}_n$ denotes the empirical averaging operator. Equivalently,

$$\mathbb{P}_n[Z_i Y_i] - \mathbb{P}_n[Z_i R_i^\top] \kappa = 0.$$

Provided $\mathbb{P}_n Z_i R_i^\top$ is nonsingular, solving these equations gives

$$\hat{\kappa}^{IV} = (\mathbb{P}_n[Z_i R_i^\top])^{-1} \mathbb{P}_n[Z_i Y_i], \quad \mathbb{P}_n[f_i] \equiv \frac{1}{n} \sum_{i=1}^n f_i.$$

This is the usual 2SLS estimator written in its identified moment-equation form. Define

$$\hat{D}^{-1} = \text{diag} \left(\sqrt{\widehat{\text{Cor}}(\hat{X}_{\cdot 1}^{(1)}, \hat{X}_{\cdot 1}^{(2)})}, \dots, \sqrt{\widehat{\text{Cor}}(\hat{X}_{\cdot q}^{(1)}, \hat{X}_{\cdot q}^{(2)})} \right),$$

where $\hat{X}_{\cdot k}^{(j)} = (\hat{X}_{1k}^{(j)}, \dots, \hat{X}_{nk}^{(j)})^\top$ denotes the sample vector of split- j scores for component k . Set

$$\hat{\beta}_X^* = \hat{D}^{-1} \hat{\beta}_{\hat{X}^{(1)}|\hat{X}^{(2)}}^{IV}.$$

If the split-half correlations are positive and bounded away from zero, the law of large numbers gives $\hat{D}^{-1} \xrightarrow{p} D^{-1}$, while standard IV consistency gives $\hat{\beta}_{\hat{X}^{(1)}|\hat{X}^{(2)}}^{IV} \xrightarrow{p} D\beta_X$. Therefore, by Slutsky's theorem,

$$\text{plim} \hat{\beta}_X^* = D^{-1} D\beta_X = \beta_X.$$

The observed-control coefficient is already correctly scaled:

$$\text{plim } \hat{\gamma}_C^{IV} = \gamma.$$

In practice, the signs of the split-specific latent dimensions should be aligned to the target latent directions before computing the correlations, and very small split-half correlations should be treated as a sign of weak-instrument or weak-measurement dynamics, where caution is warranted.¹⁶

A.I.3 A Stylized Comparison with the Method of Composition

In this stylized comparison, suppose the point estimate can be written as

$$\hat{X} = \frac{X + U_{\text{MOC}}}{\sqrt{1 + \sigma_{U,\text{MOC}}^2}}.$$

Write the corresponding pre-standardized MOC draw as

$$\tilde{X}_{\text{MOC}}^{(t)} = X + U_{\text{MOC}} + \xi^{(t)},$$

and the standardized MOC regressor as

$$\hat{X}_{\text{MOC}}^{(t)} = \frac{X + U_{\text{MOC}} + \xi^{(t)}}{\sqrt{1 + \sigma_{U,\text{MOC}}^2 + \sigma_{\xi}^2}},$$

where $\xi^{(t)}$ is mean-zero draw noise independent of X , U_{MOC} , and ε . Then

$$\text{plim } \hat{\beta}_{\text{MOC}}^{(t)} = \frac{\beta_X}{\sqrt{1 + \sigma_{U,\text{MOC}}^2 + \sigma_{\xi}^2}},$$

whereas the regression using the point estimate is centered at

$$\frac{\beta_X}{\sqrt{1 + \sigma_{U,\text{MOC}}^2}}.$$

¹⁶Note that the diagonal correction above is appropriate for componentwise unit-variance identification and should not be interpreted as a full whitening or rotation correction. If a measurement procedure applies a general nonsingular non-diagonal transformation $\hat{X} = A(X + U)$, the population IV coefficient δ_X^{IV} satisfies

$$A^\top \delta_X^{IV} = \beta_X, \quad \delta_X^{IV} = A^{-\top} \beta_X, \quad \beta_X = A^\top \delta_X^{IV}.$$

Componentwise split-half correlations do not identify A ; estimating a full rotation or shear correction requires additional information or restrictions.

Because $\sigma_\xi^2 > 0$ implies that

$$\frac{1}{\sqrt{1 + \sigma_{U,\text{MOC}}^2 + \sigma_\xi^2}} < \frac{1}{\sqrt{1 + \sigma_{U,\text{MOC}}^2}},$$

the MOC draw-level regressions are more attenuated in this stylized setting. This result is not a claim about all Bayesian models. It describes the common two-stage MOC implementation in which X is sampled from $p(X | W)$ rather than from $p(X | W, Y)$. Posterior means or posterior draws from a fully specified Bayesian model need not decompose into a true score plus independent classical noise, so the calculation is a stylized additive-noise comparison (not a general MOC theorem).

A.I.4 Alternative Identification Restrictions and Scale Transformations

Here, we derive the attenuation factor for bivariate linear regression under identifying restrictions other than the mean 0, standard deviation 1 assumption for both the true latent trait X and the latent-trait estimates $\hat{X}$ that is assumed in the paper.

We do this by noting that if we have (possibly different) sets of identifying restrictions for X and $\hat{X}$ then they can be represented by affine transformations of the paper’s standardized measures. Formally, we can write the true latent trait under its new identifying restrictions as X' and the estimated latent trait under its new identifying restrictions as $\hat{X}'$. Then for some constants a, c and positive scale constants b, d after aligning directions, we have:

$$X' = \frac{X - a}{b} \tag{A.I.10}$$

$$\hat{X}' = \frac{(\hat{X} - c)}{d} \tag{A.I.11}$$

which allows us to derive the attenuation factor under any identification restrictions. Note that it may seem odd to assume different identifying restrictions for the true latent traits and for the estimates of them, but we derive our results here under the broadest conditions, which include cases where the same identifying restriction is used on the true and estimated latent traits.

We derive the attenuation factor by noting that the subtraction terms in the numerators of Equations A.I.10 and A.I.11 do not affect the true definition of β_X or its estimate $\hat{\beta}_{\hat{X}}$. Dividing X by b would multiply β_X by b and dividing $\hat{X}$ by d would multiply $\hat{\beta}_{\hat{X}}$ by d . Together, this means that the new attenuation factor is d/b times the attenuation factor in the paper, so that

$$\text{plim } \hat{\beta}_{\hat{X}'} = \frac{d}{b\sqrt{1 + \sigma_U^2}} \beta_{X'}. \tag{A.I.12}$$

For example, if one were to use the identifying restriction that two different observations’ true latent trait values were -1 and 1, respectively (sometimes referred to as the “Kennedy-Helms” identification restriction used to fix the ideology a very liberal and very conservative legislator at these two values), one could simply take the values estimated for the two legislators and find the affine transformation (values of c and d in Equation A.I.11) that would place the two legislators at

these positions.

However, if we define the true latent trait X using the same identifying restrictions we apply to its estimate, we recreate the dynamic discussed in the main text. In standard classical measurement error, the observed predictor has a larger variance than the true predictor. In the latent case, applying identical restrictions forces both scales to match. Also, this argument treats a , b , c , and d as fixed constants. If anchors or scale constants are estimated from the data, especially from selected observations, these constants are random and can affect uncertainty calculations and asymptotic approximations.

Appendix II: Evaluating the Sensitivity of the Corrected Estimator to Violations of $\sigma_{U,1}^2 = \sigma_{U,2}^2$

In our main derivations in the paper, particularly in showing how the correlation between the two sets of latent-trait estimates can be used to estimate the measurement error variance, we assumed that the measurement error variance was the same for both of our replicate measures $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ which were obtained by partitioning our indicators into two equally sized sets and estimating the latent trait separately based on each of the two sets. Although this may be roughly true given that each set of estimates is based on the same number of indicators (or nearly the same number in the case of an odd number of indicators), it is unlikely that these variances will be exactly the same in practice.

Note, for example, that different indicators may contain different amounts of information about the latent trait. To use an example from educational testing, a question that every single student answers incorrectly does not give us *any* information about the relative intelligence of the test takers. For example, even if half of test takers answer a question correctly, their responses may not correlate with overall performance—correct answers might reflect regional familiarity with question content or other non-intelligence factors. And, some highly informative indicators may discriminate very strongly between those with low and high levels of the latent trait.

To the extent that the values of $\sigma_{U,1}^2$ and $\sigma_{U,2}^2$ differ, our estimate of ρ based on $\text{Cor}(\hat{X}^{(1)}, \hat{X}^{(2)})$ and hence the implied correction factor applied to the standard OLS or IV estimators will be incorrect which will cause our corrected estimators to be biased. But how large is this bias likely to be in practice if such assumption violations arise? In other words, how different must $\sigma_{U,1}^2$ and $\sigma_{U,2}^2$ be to meaningfully affect the ultimate estimates of β_X that our corrected estimators produce? ...and how much are these possible biases mitigated by our strategy of averaging over slope estimates from each of the different partitions of indicators into two sets? In this section, we explore these questions by deriving an expression for the degree to which estimates are affected by given values of the two variances and discussing the implications for our estimators in practice.

Recall that in the main text, we showed that if we estimate a standard linear regression with our observable estimates $\hat{X}$, then the population slope, equivalently the probability limit of the OLS slope estimator, satisfies

$$\text{plim } \hat{\beta}_{\hat{X}} = \left(\text{plim } \hat{\beta}_{\hat{X}} \right) \sqrt{1 + \sigma_U^2} = \frac{1}{1 + \sigma_U^2} \beta_X \sqrt{1 + \sigma_U^2} = \frac{1}{\sqrt{1 + \sigma_U^2}} \beta_X.$$

Since as we show, $\text{Cor}(\hat{X}^{(1)}, \hat{X}^{(2)}) = \frac{1}{1+\sigma_U^2}$ we can obtain a corrected estimator by dividing $\hat{\beta}_{\hat{X}}$ by $\sqrt{\text{Cor}(\hat{X}^{(1)}, \hat{X}^{(2)})}$.

But for the correlation between the two latent-trait estimates $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ to appropriately estimate the correction factor in this way, we need to assume that the measurement error variance for both of these estimates of the latent trait are the same, i.e., that $\sigma_{U,1}^2 = \sigma_{U,2}^2$, which is why we subscript the variance above with U rather than $U, 1$ or $U, 2$.

If the variance of the measurement error for $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ is not the same, then the correlation estimates something different. Specifically, we have

$$\begin{aligned}
\text{Cor}(\hat{X}^{(1)}, \hat{X}^{(2)}) &= \text{Cor}(\tilde{X}^{(1)}, \tilde{X}^{(2)}) = \frac{\text{Cov}(\tilde{X}^{(1)}, \tilde{X}^{(2)})}{\sqrt{\sigma_{\tilde{X}^{(1)}}^2 \sigma_{\tilde{X}^{(2)}}^2}} \\
&= \frac{\text{Cov}(X + U^{(1)}, X + U^{(2)})}{\sqrt{(\sigma_X^2 + \sigma_{U,1}^2)(\sigma_X^2 + \sigma_{U,2}^2)}} \\
&= \frac{\text{Cov}(X, X) + \text{Cov}(X, U^{(2)}) + \text{Cov}(U^{(1)}, X) + \text{Cov}(U^{(1)}, U^{(2)})}{\sqrt{(\sigma_X^2 + \sigma_{U,1}^2)(\sigma_X^2 + \sigma_{U,2}^2)}} \\
&= \frac{\sigma_X^2 + 0 + 0 + 0}{\sqrt{(\sigma_X^2 + \sigma_{U,1}^2)(\sigma_X^2 + \sigma_{U,2}^2)}} \\
&= \frac{\sigma_X^2}{\sqrt{(\sigma_X^2 + \sigma_{U,1}^2)(\sigma_X^2 + \sigma_{U,2}^2)}} \\
&= \frac{1}{\sqrt{(1 + \sigma_{U,1}^2)(1 + \sigma_{U,2}^2)}},
\end{aligned} \tag{A.II.1}$$

since as above we assume by identification that $\sigma_X^2 = 1$.

This implies that when $\sigma_{U,1}^2 \neq \sigma_{U,2}^2$ our proposed correction to the OLS slope estimator will be affected. To see this, note that the naive OLS estimator using $\hat{X}^{(1)}$ as the predictor has probability limit $\beta_X / \sqrt{1 + \sigma_{U,1}^2}$. If we divide by $\sqrt{\text{Cor}(\hat{X}^{(1)}, \hat{X}^{(2)})}$ to correct the estimator, then we will not cancel out the attenuation factor and instead the corrected estimator has probability limit

$$\begin{aligned}
\text{plim } \hat{\beta}_{\hat{X}^{(1)}}^* &= \frac{\beta_X}{\sqrt{1 + \sigma_{U,1}^2}} \underbrace{\left[(1 + \sigma_{U,1}^2)(1 + \sigma_{U,2}^2) \right]^{\frac{1}{4}}}_{\text{inverse square-root split correlation}} \\
&= \beta_X \frac{(1 + \sigma_{U,2}^2)^{\frac{1}{4}}}{(1 + \sigma_{U,1}^2)^{\frac{1}{4}}}.
\end{aligned} \tag{A.II.2}$$

The larger the difference between $\sigma_{U,1}^2$ and $\sigma_{U,2}^2$, the farther this fraction will be from 1, and hence the farther the probability limit of the corrected slope will be from β_X .

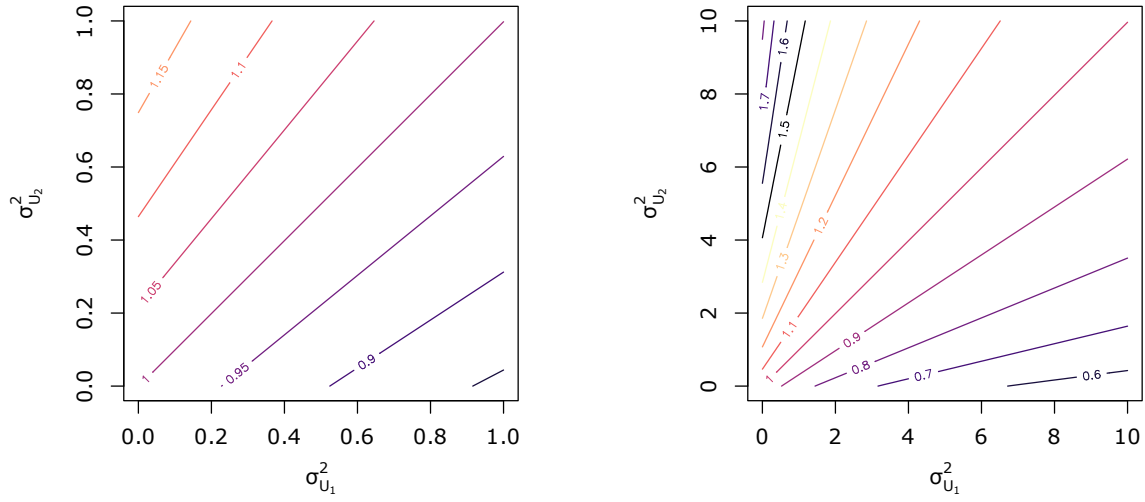


Figure A.II.1: This figure depicts the value of the factor $\frac{(1+\sigma_{U,2}^2)^{\frac{1}{4}}}{(1+\sigma_{U,1}^2)^{\frac{1}{4}}}$ from Equation A.II.2 as a function of $\sigma_{U,1}^2$ and $\sigma_{U,2}^2$. The left pane shows the values of this factor when each of these variances ranges between 0 and 1, while the right pane shows values as the variances range between 0 and 10.

But how large does the difference between $\sigma_{U,1}^2$ and $\sigma_{U,2}^2$ have to be in order to produce a meaningfully large skew from using this correlation-based correction? In order to assess this Figure A.II.1 plots the factor multiplying β_X in Equation A.II.2 as a function of both measurement error variances. The left pane shows the factor multiplying β_X on the right side of Equation A.II.2 as the measurement error variances for the two estimates of the latent trait, each of which ranges between 0 and 1. For most applications, we would expect *ex ante* that these parameters would fall into this range, i.e., that the measurement error variance for each of the two sets of estimates would not be larger than the sample variance of the true latent trait itself. Within this range, we observe that the factor is small but not always negligible, reaching values below 0.85 (a 15 percent downward skew in the slope estimate) and above 1.15 (a 15 percent upward skew in the slope estimate) for some instances. But note that these values are ones in which one of the measures has measurement error variance very close to 0, and the other has measurement error variance very close to 1. If we were to partition a set of indicators evenly into two sets, it would seem unlikely that we would be in such a situation. But skewing the slope estimate by smaller amounts, e.g., by 5 or 10 percent in either direction, may still be problematic.

The right pane shows the impact of this assumption violation for a much wider (perhaps implausibly so) range of values for the two variances, with each going from 0 to 10. We see that this factor reaches values much farther from 1. Note, however, that for many of these values much farther from 1, we again would be dealing with dramatically differing values for $\sigma_{U,1}^2$ and $\sigma_{U,2}^2$, e.g., 1 and 5, which may be unlikely to occur in most settings.

More importantly, recall that our proposed corrected estimator $\hat{\beta}_X^*$ averages over the two ways to use a given partition, in the case of the IV formulation, with $\hat{X}^{(1)}$ as the regressor and $\hat{X}^{(2)}$ as

the instrument, and also with $\hat{X}^{(2)}$ as the regressor and $\hat{X}^{(1)}$ as the instrument. As discussed in the main text, this would be expected to mitigate the skews that result from differing measurement error variances. This is because when these two measurement errors are not equal, the correlation-based correction factor will be too high when using one of the two sets of estimates as the predictor and too low when using the other one. Specifically, when averaging in this way, we would obtain a factor of

$$\overbrace{\frac{1}{2} \left[\frac{(1 + \sigma_{U,2}^2)^{\frac{1}{4}}}{(1 + \sigma_{U,1}^2)^{\frac{1}{4}}} + \frac{(1 + \sigma_{U,1}^2)^{\frac{1}{4}}}{(1 + \sigma_{U,2}^2)^{\frac{1}{4}}} \right]}^{\text{average of the two reciprocal directional factors}} = \frac{\underbrace{r + r^{-1}}_{\text{Letting } r \equiv \left(\frac{1 + \sigma_{U,2}^2}{1 + \sigma_{U,1}^2} \right)^{1/4} > 0}}{2} \geq \sqrt{r \cdot r^{-1}} = 1. \quad (\text{A.II.3})$$

The inequality is the arithmetic mean–geometric mean inequality applied to the positive pair r and r^{-1} . Equality requires $r = r^{-1}$, or equivalently $r = 1$, which occurs exactly when $\sigma_{U,1}^2 = \sigma_{U,2}^2$. Thus, averaging the two directions removes the first-order imbalance between the reciprocal factors, but it does not eliminate the remaining second-order inflation when the two measurement-error variances differ.

Figure A.II.2 shows how this new averaged factor varies over values of each variance for a correlation-corrected slope estimator that is averaged over using $\hat{X}^{(1)}$ and using $\hat{X}^{(2)}$ as the predictor. We see that the factor is much closer to 1 for most plausible parameter values, although it is not exactly 1 when the two variances differ. In the left pane, the factor is between 1 and 1.01 for nearly all possible variances, and is between 1 and 1.002 for a wide range of variances. Even when we let the variances range between 0 and 10, we see that only highly unequal variances produce factors above 1.02. As discussed above, we may be skeptical that measurement error variances would be this unequal in practice when using the same number of indicators to obtain estimates. This suggests that violations of the assumption that $\sigma_{U,1}^2 = \sigma_{U,2}^2$ are unlikely to be severe enough in practice to have a significant impact on our corrected estimates.

Furthermore, simple and straightforward checks could be used to reassure researchers that a particular partition of indicators into two sets does not result in a large imbalance of information about the latent trait between these two sets of indicators. For example, if using a factor analytic or ideal point model to estimate the latent trait, one could inspect the factor loadings or discrimination parameters from estimating the model using the full set of indicators in order to make sure that one of the two partitions does not contain many that are close to zero and with other containing many that are far from zero. Researchers could also inspect standard errors, posterior standard deviations, or other measures of uncertainty that their chosen measurement model may produce. Even when using simpler strategies, such as additive or averaging, to produce latent-trait estimates, inspecting the correlation matrix of the indicators and the distributions of each individual indicator may be enough to prevent the unexpected use of sets of indicators that each contain dramatically different amounts of information about the latent trait. Many of these things are good practice in any application of measurement models, not just for our specific aims of addressing measurement error in latent predictors.

Also note that the expression in Equation A.II.3 holds for a particular partition of the indicators, where the estimates are averaged using one or the other as the predictor. Our proposed corrected estimator $\hat{\beta}^*$ takes the median over all such within-partitions averages of the corrected

slope estimate, which can further reduce sensitivity to unstable partitions but does not remove the componentwise imbalance factor for any given partition.

Overall, then, we can expect that our correlation-corrected estimators (recall that the corrected OLS and IV estimators are equivalent for the bivariate regression case) should perform well in a wide range of situations and that in practical applications with balanced splits, the assumption that these two measurement error variances are identical is unlikely to meaningfully affect our proposed estimators.

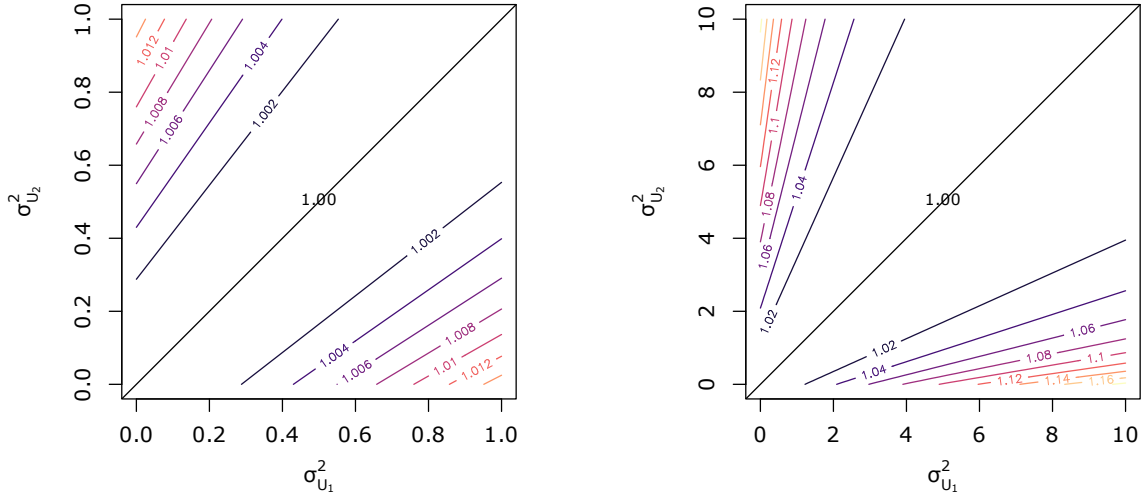


Figure A.II.2: This plot displays the value of the factor $\left[\frac{(1+\sigma_{U,2}^2)^{\frac{1}{4}}}{(1+\sigma_{U,1}^2)^{\frac{1}{4}}} + \frac{(1+\sigma_{U,1}^2)^{\frac{1}{4}}}{(1+\sigma_{U,2}^2)^{\frac{1}{4}}} \right] / 2$ from Equation A.II.3 as a function of $\sigma_{U,1}^2$ and $\sigma_{U,2}^2$. The left pane shows the values of this factor when each of these variances ranges between 0 and 1, while the right pane shows values as the variances range between 0 and 10.

Appendix III: Additional Simulation Details & Results

A.III.1 Simulation Design Details

We assume a two-parameter probit-link IRT model for the indicators, as well as a standard linear regression of the observed dependent variable Y on the latent trait X . This setup is common in political science and other literatures, with estimation, often via Gibbs sampling with data augmentation or other methods, being well studied (see e.g. Clinton et al., 2004; Martin et al., 2011).

Formally, we assume that the indicators are generated according to:

$$\Pr(W_{ij} = 1) = \Phi(X_i \eta_j + \alpha_j). \quad (\text{A.III.1})$$

with independence across individuals i and items j . Further, we assume a standard linear regression

model for an observed dependent variable Y where the latent trait X is the predictor:

$$Y_i = \beta_0 + X_i\beta_X + \varepsilon_i. \quad (\text{A.III.2})$$

with $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Standard independent conjugate priors are used on all parameters: $X_i \sim N(0, 1)$, $\eta_j, \alpha_j, \beta_0, \beta_X \sim N(0, 25)$, $\sigma_\varepsilon^2 \sim \text{Inverse-Gamma}(.01, .01)$.

We adopt the same data-generating process as above; for each simulation iteration we:

1. Generate simulated matrix of indicators W^{Sim} by randomly drawing from the predictive distribution of the ideal point model in Equation A.III.1.
2. Estimate respondent ideal points $\hat{X}^{\text{Sim}}$ from the simulated indicator matrix W^{Sim} , rescaling these estimated latent scores to have mean 0 and variance 1 with higher values corresponding to more conservative positions.
3. Estimate the slope coefficient relating X with Y as specified in Equation A.III.2 using simple OLS with the true latent predictor, with the estimated predictor, and with the split-indicator IV approach (with uncorrected and corrected estimators).

A.III.2 Ideology Estimation

Before discussing a larger-scale simulation that varies the number of observations and indicators used in estimation, we first provide a simple illustration of the performance of the estimators discussed above through a set of simulations using a single fixed set of parameter values. To choose these “true” data-generating parameter values, we look at data from the 2016 Cooperative Congressional Election Survey (CCES). This survey includes a large number of policy questions on which respondents are asked their positions (e.g., whether respondents support background checks for all gun sales or whether they think women should always be allowed to have an abortion as a matter of choice). We use 30 of these policy questions as our indicators from which we estimate respondents’ ideologies (ideal points) and the item parameters based on the Clinton, Jackman, and Rivers (CJR) ideal point model.¹⁷ We then use the estimated ideal points to predict 2016 presidential vote (1=Trump, 0=Clinton).

We then set the true values for 1,000 respondents’ ideal points to those point estimates for X_i (after imposing the identifying restriction that these ideal points have mean 0, variance 1, and that higher values correspond to more conservative positions) and set the true values of the discrimination/loading parameters (η_j) and item intercept/difficulty-style parameters (α_j) to those estimated from these indicators in our CCES sample. We also set the true values for the intercept β_0 , the slope β_X , and the error variance $\text{Var}(\varepsilon_i)$ to roughly correspond to the values estimated in the linear regression predicting presidential vote with the estimated ideal points using the baseline OLS approach.¹⁸

¹⁷We use the `binIRT` function in the `emIRT` package. We drop several policy items that contain a large proportion of missing values. We also drop any respondents having missing values on any of these indicators or on the presidential vote question or who stated they did not vote or voted for someone other than Trump or Clinton, and then, we randomly select a subset of respondents to use.

¹⁸The estimated slope using our corrected estimates is slightly under 0.4, but for the purposes of the simulations, we use the value of 0.4 since a round number is easier to use as a baseline for comparison in figures and elsewhere. Note

Although these simulations rely on a single set of true parameters, these values are based on real data of the sort that is often used in latent predictor applications. Furthermore, these simple simulations will illustrate some basic characteristics of the various estimators and their relative performance, at least in this specific example. Figure A.III.1 shows the distributions of simulated slope estimates (β_X) based on 100,000 simulation draws for each estimator.

The top pane of Figure A.III.1 provides a baseline, showing the simulated sampling distribution for the slope coefficient if we actually knew each respondent's ideal point with certainty (i.e. if there was no measurement error in the predictor) and we estimated a simple bivariate linear regression predicting presidential vote with the ideal points X . As expected, the sampling distribution for the slope estimator is centered at the true slope of 0.4. The simulated RMSE is relatively small (0.01). Well over 60 percent of the simulations produced estimates within 0.01 of the true slope, and well over 90 percent produced estimates within 0.02 of the true slope. In a world where the true predictor is measured without error, we can obtain unbiased and relatively precise estimates of the slope coefficient of interest. Of course, with latent predictors, we will never know these true values.

The second pane in Figure A.III.1 plots the simulated distribution of the uncorrected OLS estimator in the case of measurement error. These simulations estimate the slope coefficient on X predicting the presidential vote, assuming we do not actually observe the true latent trait X but instead observe the set of 30 survey questions (indicators) from which we estimate X . In other words, we use *estimated* ideal points $\hat{X}$ as our predictor rather than the true ideal points without any correction or adjustment. As expected, the coefficient estimates are generally smaller in magnitude, with the average estimate falling just below 0.38. The RMSE is more than double that of the known X case.

The third pane of Figure A.III.1 shows estimates from the standard split-indicator IV approach without any corrections for the latent variable case. In these simulations, after simulating the matrix of indicators, we split the 30 indicators into the first 15 and the second 15, estimating ideal points for all respondents separately using each of these two sets. This produces $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$, two separate estimated split scores. We then use 2SLS with $\hat{X}^{(1)}$ as the regressor and $\hat{X}^{(2)}$ as the instrument, and symmetrically with $\hat{X}^{(2)}$ as the regressor and $\hat{X}^{(1)}$ as the instrument where reported. In the standard case of measurement error in an observable (as opposed to latent) predictor, this IV setup can yield consistent estimates of the slope coefficient. But as shown above, when applied to the latent case, it will *over*-correct for the attenuation bias, producing amplification bias. This is illustrated in the figure, where the average estimate is 0.43; the RMSE is nearly 4 times as large as in the known X case (nearly double the value with uncorrected OLS and estimated X).

Finally, the bottom pane of Figure A.III.1 shows split-indicator IV estimates corrected using the correlation between $\hat{X}^{(1)}$ and $\hat{X}^{(2)}$ as described in Equation A.I.7. These estimates are quite similar to those using the true X , but with a very slightly higher bias and RMSE. This suggests the feasibility of estimating β_X with the corrected IV approach in the more realistic case where we cannot use the true X to estimate the measurement error variance of the split-half estimates.

In the last two applications (both corrected IV estimates), this simulation suggests favorable bias dynamics. (Appendix I discusses the possibility of a very enduring small amount of downward

that this value for the slope is likely to be attenuated, i.e., smaller in magnitude than the true slope coefficient, since it was estimated from data that itself includes measurement error in the predictor, but we leave it as is for the true value in our simulations.

finite sample bias.)

This example illustrates the attenuation bias induced in estimates of a bivariate linear regression slope by using a noisy measure of a predictor. In our application, the measure is based on a relatively large number (30) of indicators, resulting in a relatively good signal-to-noise ratio. In cases where fewer indicators or less informative indicators are used, the attenuation bias will be larger, as seen below. It is also shown that while the standard IV approach to dealing with measurement error produces *over*-correction, it is possible to correct for this and obtain slope estimates that are nearly unbiased, with variance barely higher than if the true predictor were known without error.

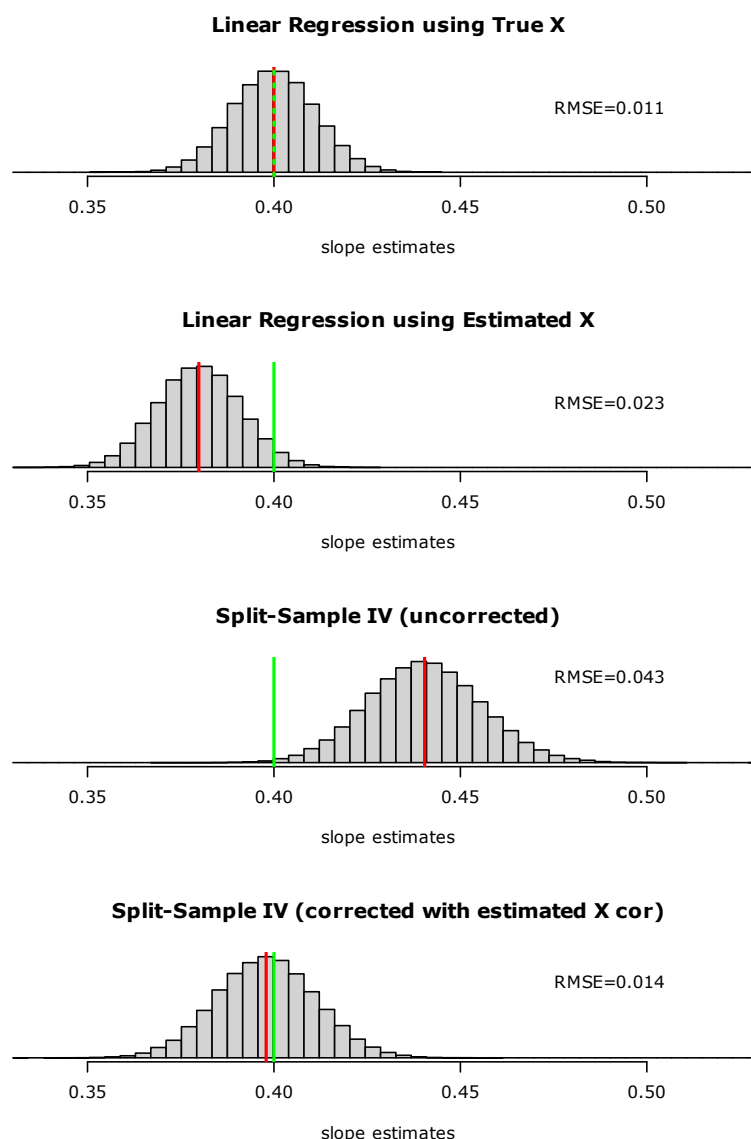


Figure A.III.1: Illustrative simulation results. The true coefficient is highlighted in green (0.40). The mean of the estimates is highlighted in red across the various approaches.

A.III.3 Coverage Analysis

Figure A.III.2 shows that, across most simulated settings, and especially for moderate-to-large M , our corrected procedure retains coverage close to nominal while confidence intervals from the uncorrected OLS and IV estimators frequently fail to contain the true parameter value (0.4) even at moderate sample sizes. With the corrected approach, as the number of indicators (M) grows, measurement error in the latent predictor diminishes, allowing the corrected estimator's coverage to stay near nominal levels; in the smallest- M settings, coverage can be mildly reduced. In contrast, the uncorrected methods exhibit persistently lower coverage, particularly in smaller M regimes. In this way, statistical inference is improved with the corrections described here.

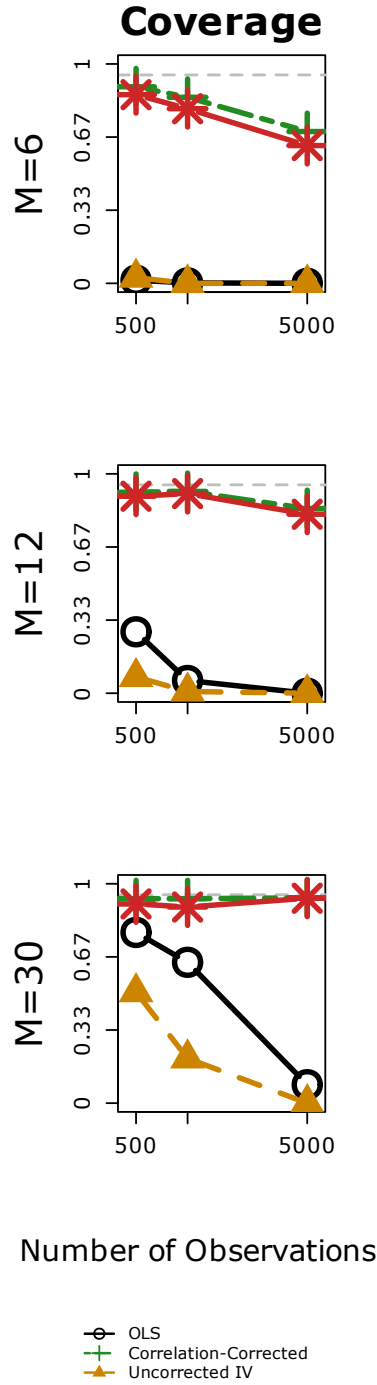


Figure A.III.2: Coverage results varying N and M . Points show empirical 95% interval coverage for each estimator across the simulation designs described in Appendix A.III.1; the horizontal reference line marks nominal coverage.

A closer look at the small- M regime reveals that even the corrected approach can show mildly reduced coverage there, reflecting finite-sample bias when estimating the latent factor from very little data. One possibility is that, in this low-information setting, random variation in the latent

variable estimates via the bootstrap is insufficient to reflect true sampling uncertainty. Another possibility is the presence of small but lingering finite sample bias in this low-information case. Nonetheless, once N and M become sufficiently large, this small- M artifact subsides; the corrected procedure once again not only outperforms the uncorrected OLS or IV, but also yields inferences approaching nominal levels.

In further analyses, we find that subsampling based-bootstrap variants (Politis and Romano, 1994) do not markedly improve coverage of the median bootstrap; windorized mean coverage is slightly higher on average. That said, the main contribution to the undercoverage of the corrected estimator seems to be residual finite indicator bias. Pseudo-target coverage, where the target is the cell mean of the estimator (rather than the true parameter value), is much better for the corrected estimator, rising from approximately 0.70 to 0.84 in the $M = 6$ case and from 0.936 to 0.944 in the $M = 30$ case.

A.III.4 Multivariate Simulation Results

We also evaluate the multivariate corrected IV extension derived in Appendix A.I.2. This simulation focuses on the comparison between uncorrected OLS and the corrected IV estimator because the aim is to assess whether the split-indicator correction carries over when multiple latent predictors and observed covariates enter the same outcome model.

For each Monte Carlo draw, we generate $K \in \{1, 2, 4, 8\}$ standardized latent predictors with an autoregressive correlation structure and pair them with 10 observed covariates that are correlated with the latent predictors. The outcome is generated as a linear function of the latent predictors, the observed covariates, and independent Gaussian noise. For each latent predictor, we then generate eight indicators and estimate split scores. The primary measurement design uses binary indicators generated from a two-parameter logistic model and estimated by EM. As robustness checks, we also consider continuous equal-loading indicators estimated by averaging and continuous congeneric indicators estimated by PCA. In all designs, we vary $N \in \{500, 1000, 5000\}$, use eight split partitions with median aggregation, and apply the correction only to partitions with split-half correlations above 0.01.

Figure A.III.3 reports absolute bias, estimator standard deviation, and RMSE for both the latent-predictor coefficients and the observed-covariate coefficients in the primary binary 2PL design. The multivariate results mirror the bivariate simulation patterns. Uncorrected OLS has persistent bias because measurement error in the latent predictors affects not only their own coefficients but also coefficients on observed covariates correlated with them. The corrected IV estimator substantially reduces this bias. In the largest primary design, with $N = 5000$ and $K = 8$, latent-coefficient RMSE falls from 0.109 under uncorrected OLS to 0.027 under corrected IV, a roughly 4.092-fold improvement. For the observed-covariate coefficients, RMSE falls from 0.102 to 0.021, a roughly 4.767-fold improvement.

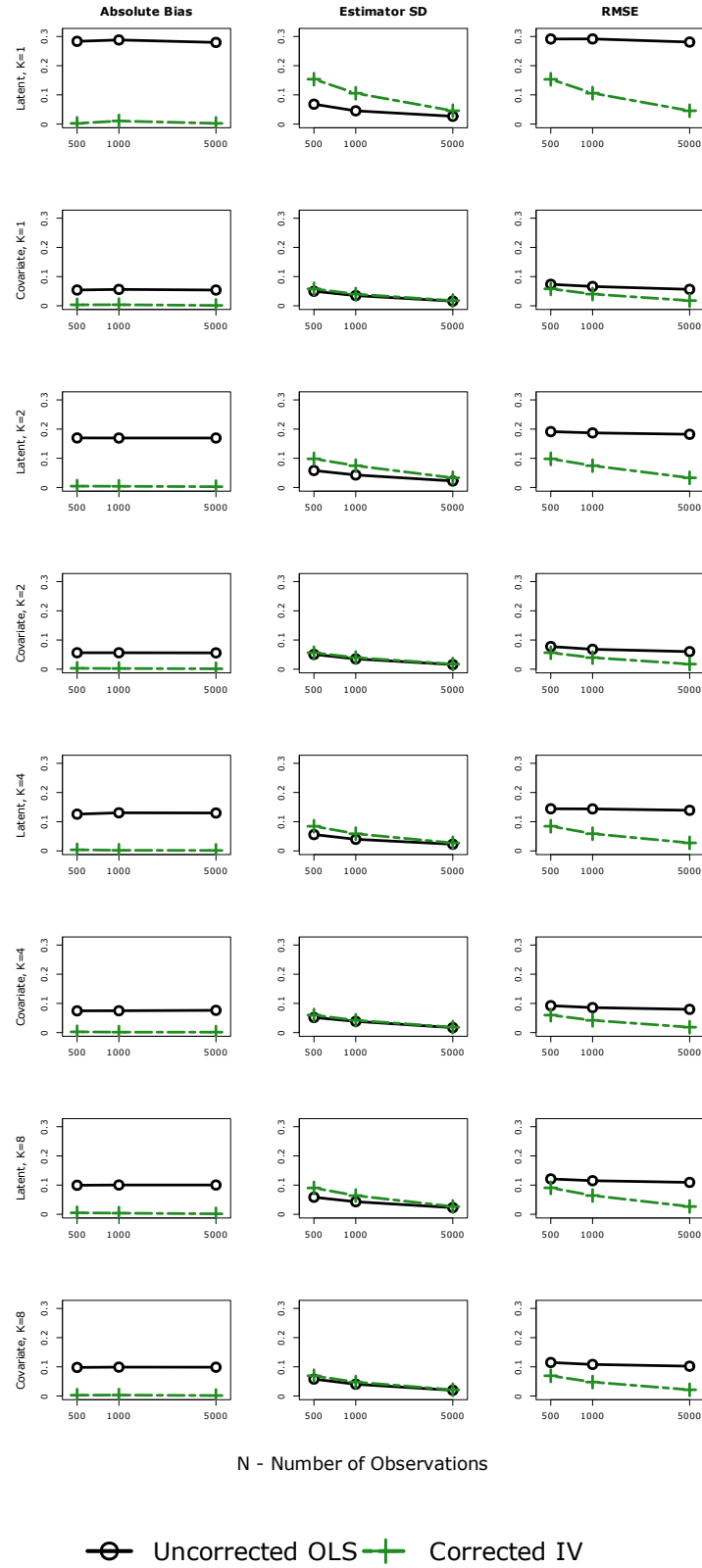


Figure A.III.3: Multivariate simulation results for the primary binary 2PL measurement design. Panels report absolute bias, estimator standard deviation, and RMSE for uncorrected OLS and the multivariate corrected IV estimator, separately for latent-predictor and observed-covariate coefficients.

Figure A.III.4 reports the corresponding split-half and first-stage diagnostics. The median split-half correlation in the largest primary design is 0.465, and the median first-stage F -statistic is 964.128. Across the simulated cells, first-stage diagnostics are well above the conventional weak-instrument threshold of 10, consistent with the stable behavior of the corrected estimator in Figure A.III.3.

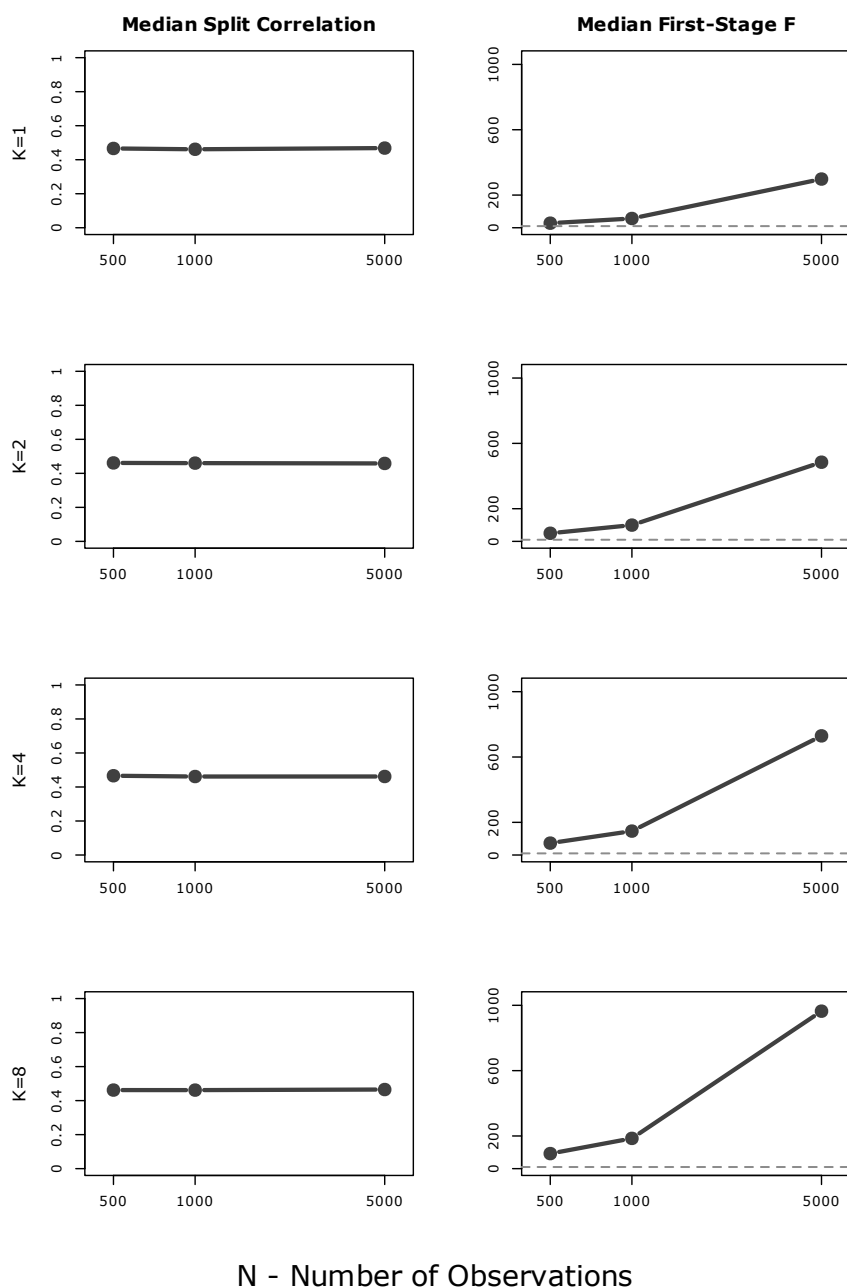


Figure A.III.4: Diagnostics for the primary multivariate binary 2PL simulation. Panels report median split-half correlations and median first-stage F -statistics across values of N and K ; the horizontal reference line marks $F = 10$.

The same qualitative pattern appears in the continuous-measurement robustness checks. Figure A.III.5 shows that, under both the equal-loading averaging design and the congeneric PCA design, the corrected IV estimator again reduces bias and RMSE relative to uncorrected OLS for the latent-predictor coefficients and for the observed-covariate coefficients affected by latent-predictor measurement error.

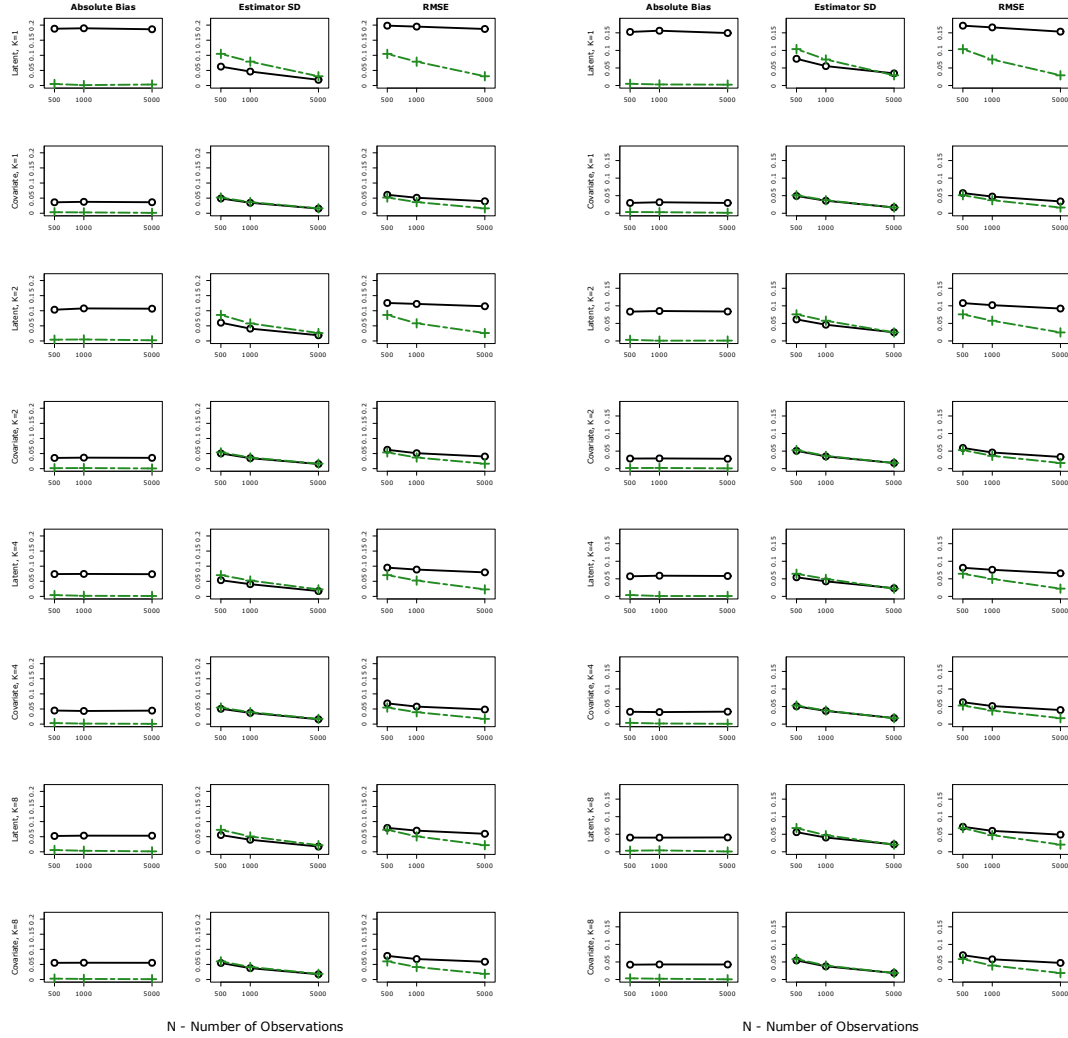


Figure A.III.5: Multivariate simulation robustness checks using continuous measurement designs. The left panel uses equal-loading indicators with averaging; the right panel uses congeneric indicators with PCA. Each point represents an values averaged across latent dimensions. The estimator legend is the same as in Figure A.III.3.

A.III.5 Comparison with Additional Estimation Strategies

Researchers have increasingly turned to Bayesian strategies for inference regarding latent traits. We examine three such approaches: (1) method of composition, (2) multiple over-imputation, and (3) joint factor and outcome models. We compare these strategies using the same data-generating process outlined in the simulation section (see Section 5). The first two approaches share an initial

measurement-model fit but differ in how the outcome enters the subsequent analysis.

Method of composition. The method of composition approach draws from the posterior distribution of the latent trait estimates (Treier and Jackman, 2008; Tanner, 2012). Let $\hat{X}_{\text{MOC}}^{(1)}, \dots, \hat{X}_{\text{MOC}}^{(T)}$ denote pre-standardized posterior draws from $p(X | W)$, where W is the indicator matrix. For each draw, we first place the latent scores on a comparable identified scale, denoting the standardized draw by $\hat{X}_{\text{MOC}}^{(t)}$. Each draw is standardized separately. We then fit the outcome regression

$$Y_i = \beta_0^{(t)} + \beta_{\text{MOC}}^{(t)} \hat{X}_{\text{MOC},i}^{(t)} + \varepsilon_i$$

for each t . In the implementation used for the simulations, the regression is fit by OLS, the covariance matrix is estimated using HC3, and the coefficient vector is optionally drawn from a Normal approximation centered at the OLS coefficient vector with that covariance matrix. The reported MOC slope and uncertainty summarize these coefficient draws across posterior latent-trait draws.

Multiple over-imputation. Multiple over-imputation, described in Rubin (1988) and refined in Blackwell et al. (2017) for outcome and measurement-error settings, treats an error-free variable or latent score as unobserved and imputes it using both measurement-error information and other analysis variables. In our implementation, it starts by fitting the measurement model to the observed indicators. From this first stage, we form a unit-level approximation to each marginal measurement-stage distribution, summarized by the posterior mean and standard deviation,

$$\mu_i = E(X_i | W), \quad \sigma_i = \text{SD}(X_i | W).$$

We then construct an imputation data set containing the observed outcome Y and a latent-score column initialized at $\mu = (\mu_1, \dots, \mu_n)$. The latent-score column is over-imputed for all units, rather than treated as fixed at μ , using an Amelia-style routine; Y remains in the imputation model as an observed variable, while (μ_i, σ_i) are supplied as unit-specific prior/uncertainty information for the latent-score column. This produces completed latent-score vectors $X_{\text{imp}}^{(1)}, \dots, X_{\text{imp}}^{(M_{\text{imp}})}$; in our implementation, $M_{\text{imp}} = 5$. Each completed vector is then standardized and used in the completed-data regression

$$Y_i = \beta_0^{(r)} + \beta_X^{(r)} X_{\text{imp},i}^{(r)} + \varepsilon_i.$$

The over-imputation summary pools the resulting completed-data slope estimates across $r = 1, \dots, M_{\text{imp}}$. Thus, unlike MOC, the outcome enters the imputation step. Unlike full joint estimation, however, the original measurement model is not re-estimated jointly with the outcome equation; instead, the first-stage measurement distribution is approximated through unit-specific summaries supplied to the imputation model. Consequently, these completed-data regressions should generally be interpreted as estimates from a staged approximation to $p(X | W, Y)$, not as draws from the full joint posterior $p(\beta_0, \beta_X, \vartheta, X | W, Y)$, except under strong specification and compatibility conditions.

In the joint modeling strategy, we fit a Bayesian model that simultaneously uses both outcome and observed factor information to generate estimates of the latent quantities of interest. This approach generates a Bayesian posterior over the coefficient estimates. Again, draws of the latent trait, X , are rescaled so that the posterior mean takes on a standard deviation of one.

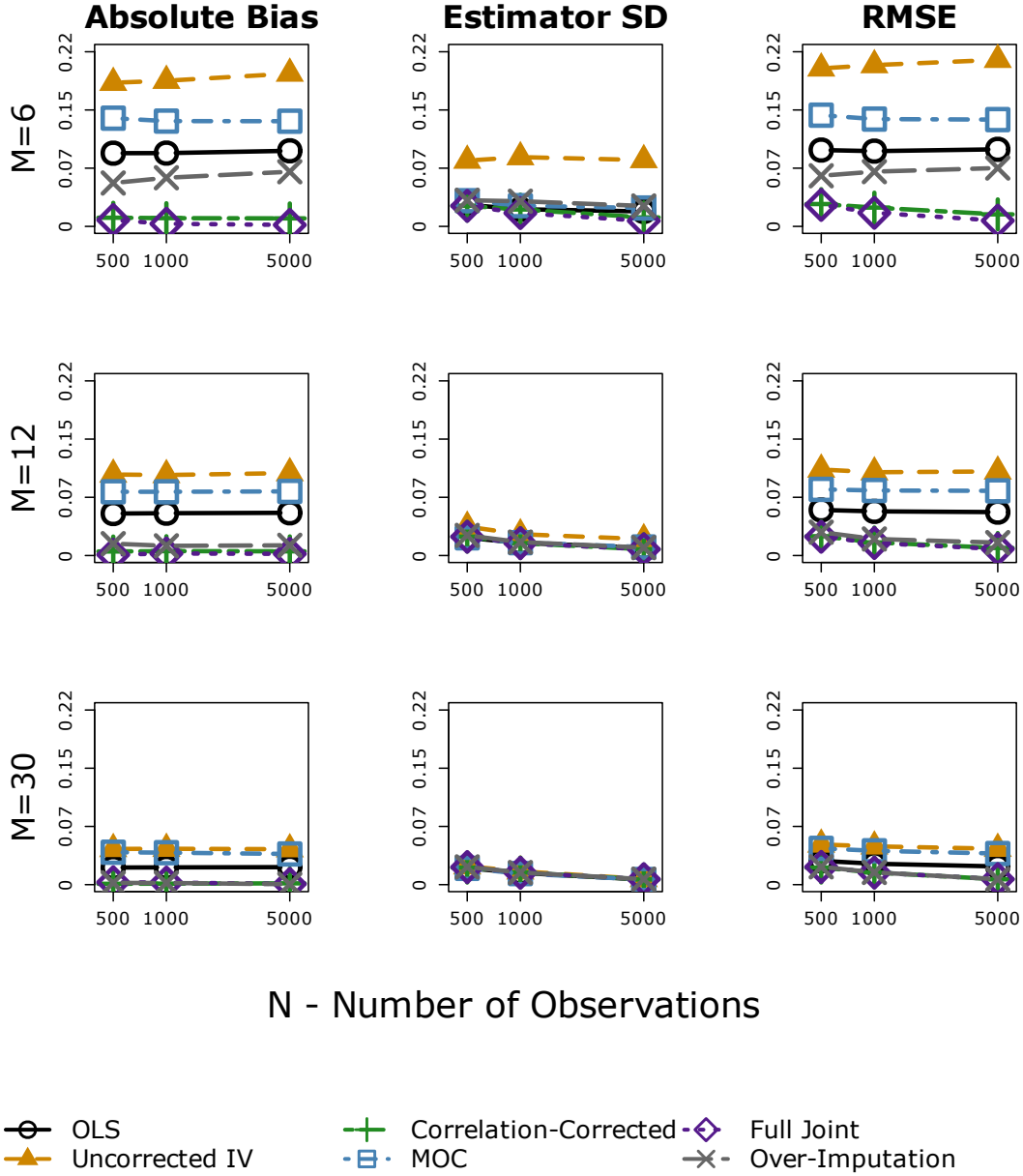


Figure A.III.6: *Additional estimation strategies compared with the main simulation estimators. Results add uncorrected IV to the comparison of OLS, MOC, over-imputation, full joint estimation, and the split-indicator correlation-corrected estimator.*

Figure A.III.6 presents the results for these approaches. We observe that in this simulation design, our method performs nearly identically to the most principled method of full joint estimation. In the two-stage MOC implementation, posterior latent-trait draws are noisier than the posterior mean used as a point estimate, which explains the additional attenuation in the stylized calculation in Appendix I. This point does not apply to a correctly specified full joint model in the same way, because the outcome enters the joint posterior for the latent traits and the regression coefficient. Both the joint model and multiple imputations face questions around how to perform

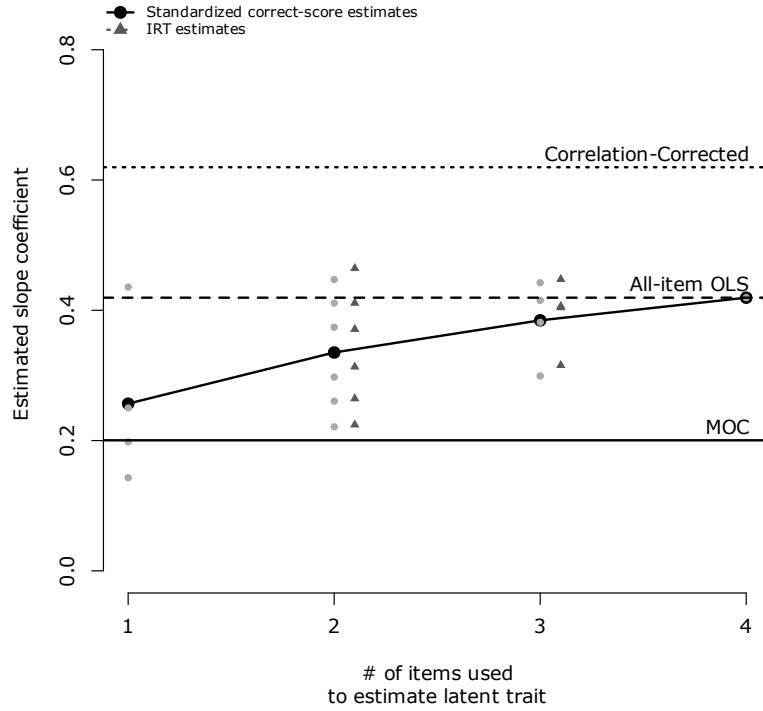


Figure A.IV.1: Grey points are uncorrected OLS slope coefficients from regressing duty-to-vote on estimated knowledge, where knowledge is constructed from each nonempty subset of the four indicators. Black dots are the mean slope within each indicator-count. Horizontal lines show the unadjusted OLS estimate using all four items, the split-indicator correlation-corrected estimate, and the MOC estimate. Additional details appear in Appendix A.IV.3.

standardization, whether that be within or across posterior draws.

In contrast, simple OLS estimates based on sample splitting, adjusted for measurement error as proposed in this paper, provide a compelling alternative. Beyond improvements in RMSE, this approach does not require fitting a larger-scale Bayesian model, thereby avoiding the increased computational costs associated with generating uncorrelated draws from the posterior distribution.

Appendix IV: Additional Empirical Results

A.IV.1 American Politics: Ideology & Presidential Vote

We here consider an application from American politics, in particular, from the literature on the relationship between latent ideology and partisan vote choice—a relationship that has been extensively studied and shown to be quite strong (Jessee, 2009; Tausanovitch and Warshaw, 2013). To what extent might measurement error in the estimation of latent ideology lead to over- or under-estimation in estimated relationships with the various strategies previously described?

Using the same design as outlined in Section A.III.1, we explore the relationship between latent

ideology, as measured by 30 questions on the 2016 CCES regarding respondents' policy positions, and vote choice for president. We again vary the number of observations used and policy indicators considered in estimating latent traits, using random sub-samples and computing average estimates. Latent traits are estimated as described above using the expectation-maximization algorithm.

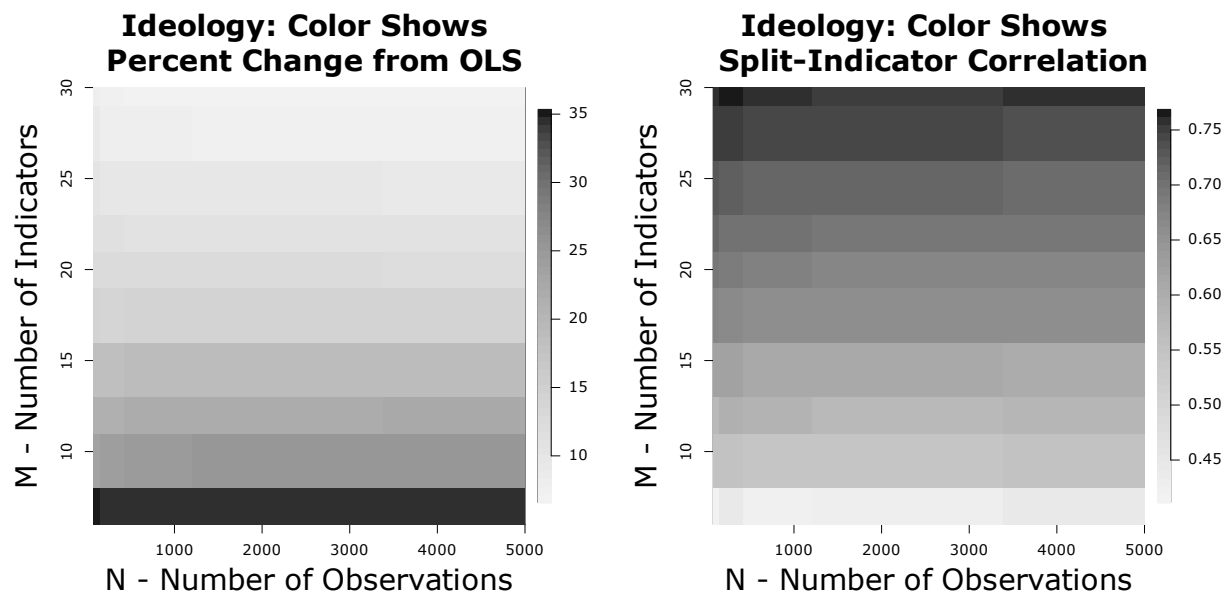


Figure A.IV.2: Measurement-error amplification for partisan-ideal-point predictors. Heat maps vary the number of observations (N , x -axis) and the number of policy indicators used to estimate the latent trait (M , y -axis). Left panel: percentage increase (relative to naïve OLS) in the absolute magnitude of the vote-share coefficient after applying the split-indicator correction; lighter shades indicate larger upward revisions. Right panel: split-indicator correlations between the two halves of the indicator set, with higher values signaling greater measurement stability.

Table A.IV.1: Percentage change in effect estimate of ideology on vote choice, corrected OLS vs. OLS. Rows correspond to observation number; columns to indicators used in estimating latent ideology.

	$M = 6$	10	12	14	18	20	22	24	28	30
$n = 80$	35.2	23.7	21.3	18.6	14.6	12.7	11.7	10.5	9.0	8.3
225	34.4	24.2	21.2	18.2	14.0	12.6	11.3	9.9	8.2	7.4
632	34.5	25.0	22.0	19.0	14.2	12.5	10.9	9.7	7.8	7.0
1778	34.6	25.4	22.2	19.1	14.1	12.4	10.9	9.5	7.7	6.9
5000	34.4	25.4	22.4	19.2	14.2	12.3	10.8	9.5	7.7	6.9

A.IV.2 Comparison: Corrected Estimator and Uncorrected IV in Democracy and Growth Application

We see in Table A.IV.2 that, if using the uncorrected IV strategy, we overestimate the effect of democracy on growth relative to the corrected IV strategy using a split-indicator approach. This overestimation is largest when the number of observations and the number of Freedom House sub-indicators used in estimation are small. The overestimation decreases as the number of observations and the number of sub-indicators increase, although again at a faster rate for the number of sub-indicators than for the number of observations. This again highlights the importance of gathering extensive information around the latent trait of interest.

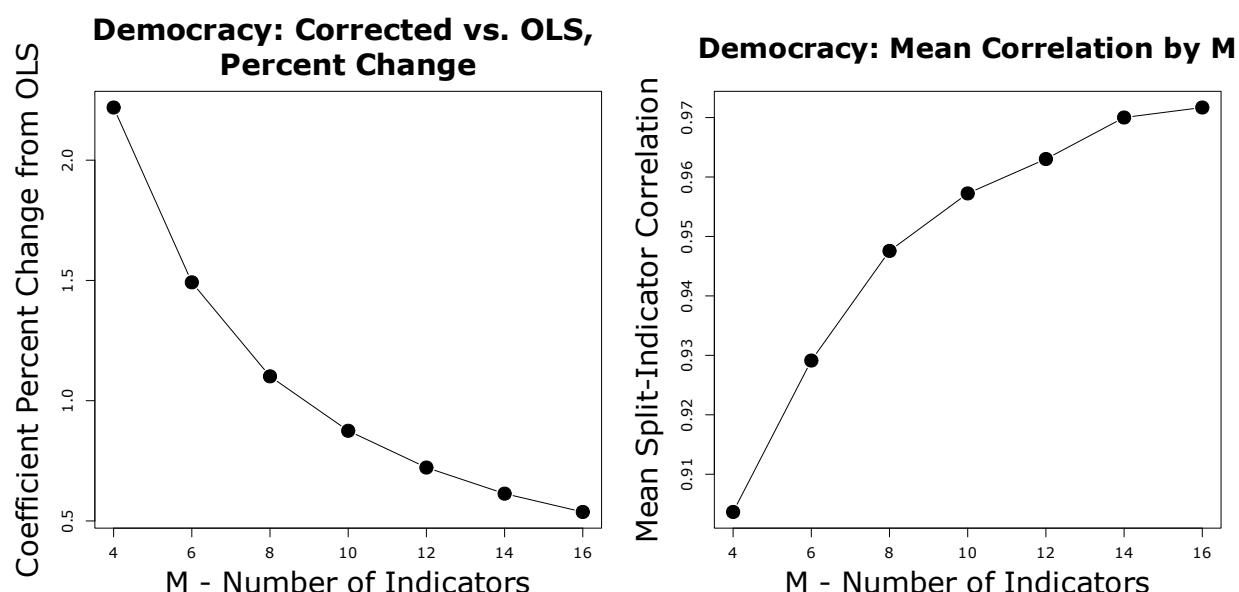


Figure A.IV.3: Democracy-growth example: Sensitivity to the size of the indicator set. Left panel: percentage change in the estimated effect of democracy on long-difference GDP growth after the split-indicator correction, plotted against the number of Freedom-House sub-indicators used (M). Right panel: mean split-indicator correlations for each value of M , showing how measurement error shrinks as more indicators are included.

Table A.IV.2: Percentage change in effect estimate of democracy on growth (long difference), corrected IV vs. uncorrected IV. Rows correspond to observation number; columns to indicators used in estimating latent democratic status.

	$M = 4$	6	8	10	12	14	16
$n = 184$	-4.8	-3.5	-2.6	-2.1	-1.7	-1.5	-1.3

A.IV.3 Effects of Political Knowledge

We here present additional empirical results on the application involving political knowledge.

Table A.IV.3: *Percentage change in effect estimate of knowledge on vote duty, corrected IV vs. OLS. Rows correspond to observation number; columns to indicators used in estimating latent knowledge.*

	$M = 4$
$n = 80$	49.6
225	50.7
632	48.7
1778	49.9
5000	49.8

Table A.IV.4: *Percentage change in effect estimate of knowledge on vote duty, corrected IV vs. uncorrected IV. Rows correspond to observation number; columns to indicators used in estimating latent knowledge.*

	$M = 4$
$n = 80$	-45.0
225	-41.3
632	-41.1
1778	-41.0
5000	-40.8

A.IV.4 Additional Information, Knowledge Example

We here provide the full text of questions from the American National Election Study (ANES) data used for the political knowledge and duty to vote application. These data were from the pre-release of the 2024 Time Series Study (downloaded 3/11/2025).

- V241612 For how many years is a United States Senator elected - that is, how many years are there in one full term of office for a U.S. Senator? [Respondents specify number of years, 6 is correct answer]
- V241613 On which of the following does the U.S. federal government currently spend the least? [Foreign aid (correct), Medicare, National defense, Social Security]
- V241614 Do you happen to know which party currently has the most members in the U.S. House of Representatives in Washington? [Democrats, Republicans (correct)]

V241615 Do you happen to know which party currently has the most members in the U.S. Senate?
[Democrats (correct), Republicans]

V241218x V241218x is a summary variable based on V241215, V241216 and V241217.

V241215 Different people feel differently about voting. For some, voting is a duty - they feel they should vote in every election, regardless of their feelings about the candidates and parties. For others, voting is a choice - they feel free to vote or not to vote, depending on their feelings about the candidates and parties. For you personally, is voting mainly a duty, mainly a choice, or neither a duty nor a choice?[Mainly a duty, Mainly a choice, neither a duty nor a choice]

V241216 How strongly do you feel that voting is a duty? [Very strongly, somewhat strongly, or not strongly / Not strongly, somewhat strongly, or very strongly]?

V241217 How strongly do you feel that voting is a choice? [Very strongly, somewhat strongly, or not strongly / Not strongly, somewhat strongly, or very strongly]?

Appendix V: Measurement Error and Latent Predictors in the Political Science Literature

To assess the state of the political science literature regarding measurement error in latent predictors, we conducted an original survey of all articles published in the three top substantive political science journals (*APSR*, *AJPS*, *JOP*) as well as *Political Analysis* (*PA*) between 2000 and 2023. We observe in Figure A.V.1 that the overall percentage of articles mentioning latent variables and, especially, measurement error is increasing, while the number of published works mentioning both measurement error and latent variables has remained relatively constant throughout the study period. This survey of recent political science articles demonstrates that while many researchers are aware of and concerned about measurement error, the full extent of its impact on analyses involving latent variables is, perhaps, not yet widely recognized or addressed in applied work.

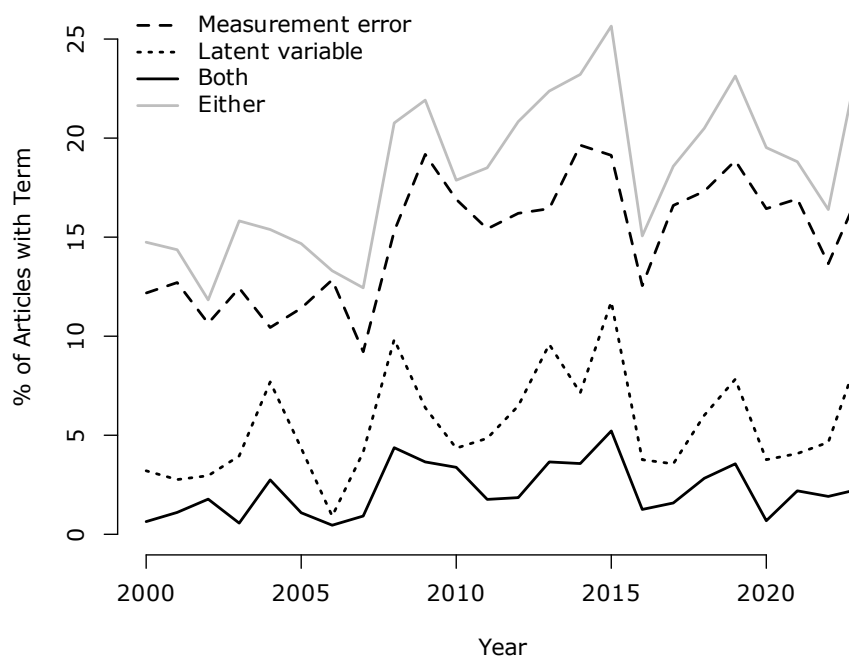


Figure A.V.1: While the overall number of mentions in top political science journals of measurement error and latent variables has been increasing, mentions of the intersection of the two have remained relatively constant in the survey period. The graph shows the number of articles containing each term by year for APSR, AJPS, JOP and PA from 2000 through 2023.

Appendix VI: A `lpmec` Package Tutorial

`lpmec` is an R package for latent-predictor measurement error correction under identification restrictions (i.e., when a key regressor is an estimated latent trait from observed indicators, leading to attenuation bias in outcome regressions). Its core workflow repeatedly splits the indicator matrix into independent halves (and can average over many such partitions), estimates the latent score separately in each half (via EM IRT for binary/ordinal items using `emIRT`, PCA, simple averaging, or a user-supplied estimator), and then applies split-indicator IV and OLS-style corrections that use the cross-half relationship to de-attenuate regression coefficients. The main entry points are `lpmec_onerun()` (single partition) and `lpmec()` (bootstrap + partition aggregation with percentile confidence intervals), with S3 `print/summary/plot` methods and built-in data (e.g., `KnowledgeVoteDuty`); the codebase also includes sensitivity reporting via `sensemakr`'s extreme robustness value.

For Bayesian estimation, `lpmec` supports MCMC either through an R backend (`pscl::ideal`) or an optional Python backend (`numpyro` via `reticulate`), plus an over-imputation variant using `Amelia`. The `mcmc_joint` option implements a joint Bayesian model that couples an IRT measurement model for the indicators with an outcome model (e.g., $Y_i \sim \mathcal{N}(\beta_0 + \beta_X X_i, \sigma)$), so

posterior uncertainty in the latent trait propagates directly into inference on the regression slope. When using the NumPyro/JAX backend, this can take advantage of JAX's JIT/XLA compilation (and, optionally, GPU/TPU acceleration) and supports subsampling-based kernels for larger problems.

We average over `n_partition` split-indicator partitions:

```
# install.packages("devtools")
devtools::install_github(
  "cjerzak/lpmec-software",
  subdir = "lpmec"
)

library(lpmec)
data(KnowledgeVoteDuty)

# Latent error correction method, with partitioning and bootstrap
results <- lpmec::lpmec(
  Y = KnowledgeVoteDuty$voteduty,
  observables = as.matrix(KnowledgeVoteDuty[, 2:5]),
  n_boot = 32L,
  n_partition = 10L,
  estimation_method = "em"
)

print(results)
summary(results)
```

Appendix VII: Code and Data Availability Statement

Additional data and materials will be made available online. For a code implementation, see

[GitHub.com/cjerzak/lpmec-software](https://github.com/cjerzak/lpmec-software)

for open-source software implementing the methods described in the paper. Replication data and scripts will be made available at a Harvard Dataverse Repository.

Appendix VIII: Winsorized-Mean Simulation Results

Figures [A.VIII.1–A.VIII.3](#) repeat the simulation analysis from Appendix [A.III.1](#) using winsorized-mean aggregation over partitions rather than median aggregation. The median-aggregation results remain the baseline results in the main text and earlier appendix figures.

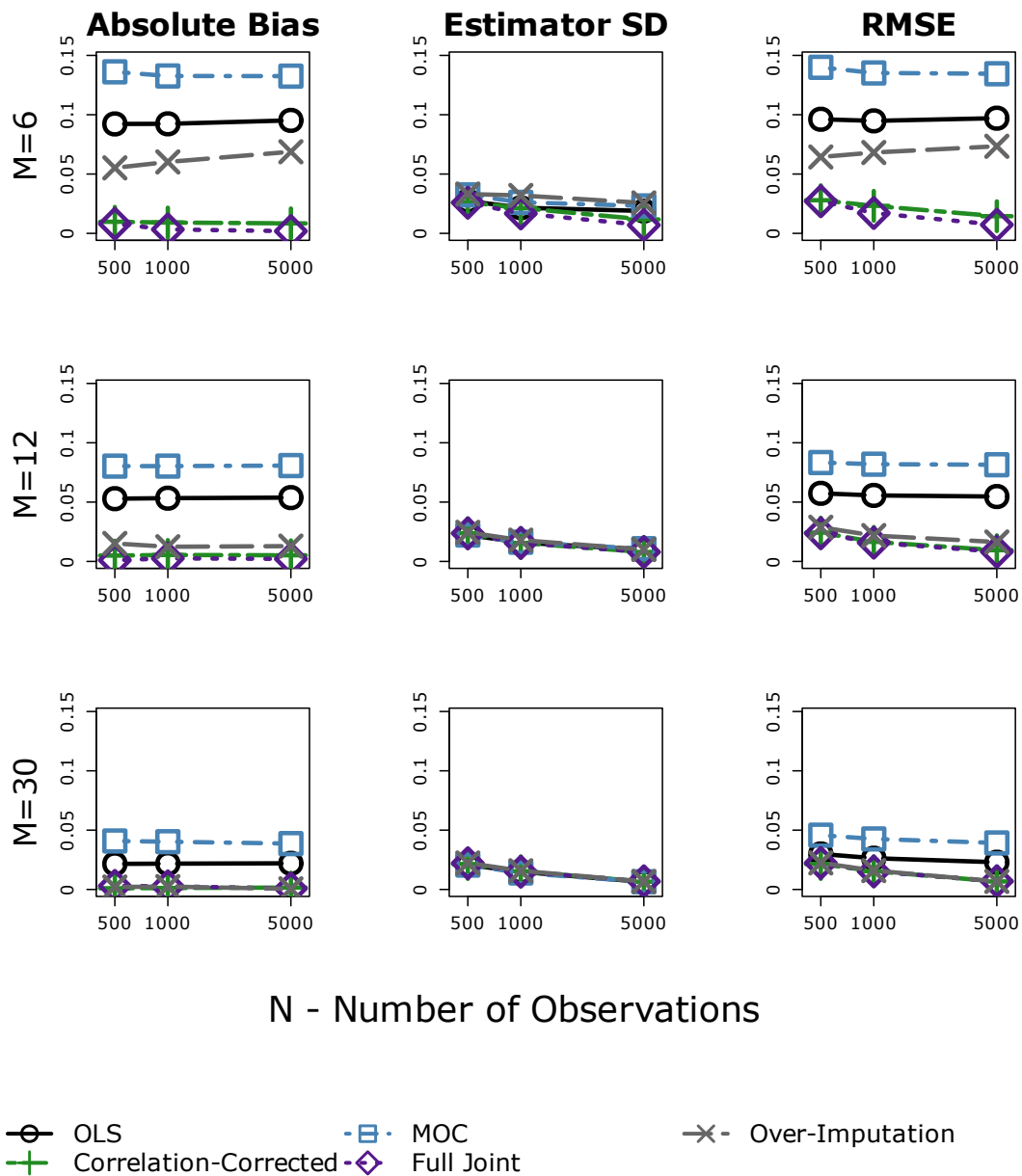


Figure A.VIII.1: Simulation results varying N and M using winsorized-mean aggregation over partitions. Results compare OLS, MOC, over-imputation, full joint estimation, and the split-indicator correlation-corrected estimator; median aggregation remains the baseline specification in the main text.

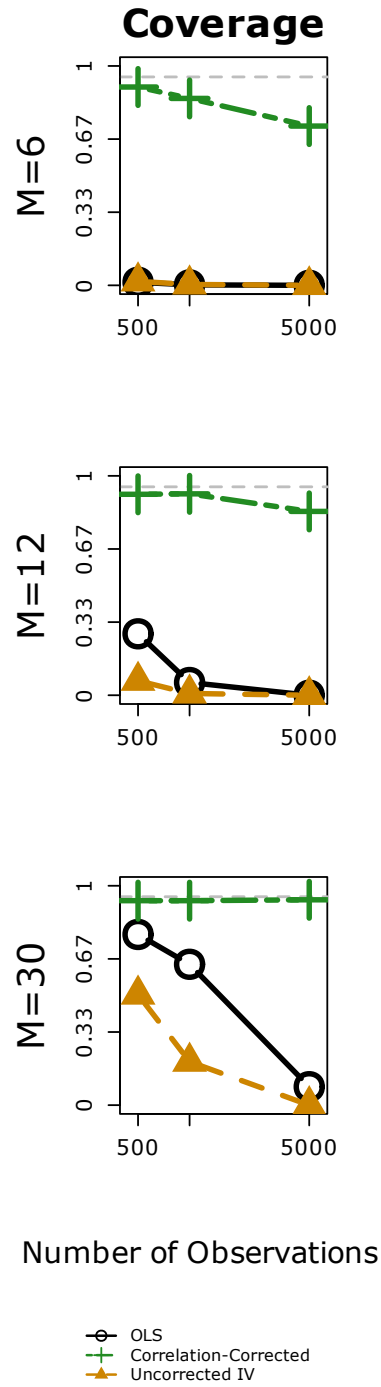


Figure A.VIII.2: Coverage results varying N and M using winsorized-mean aggregation over partitions. Points show empirical 95% interval coverage; the horizontal reference line marks nominal coverage.

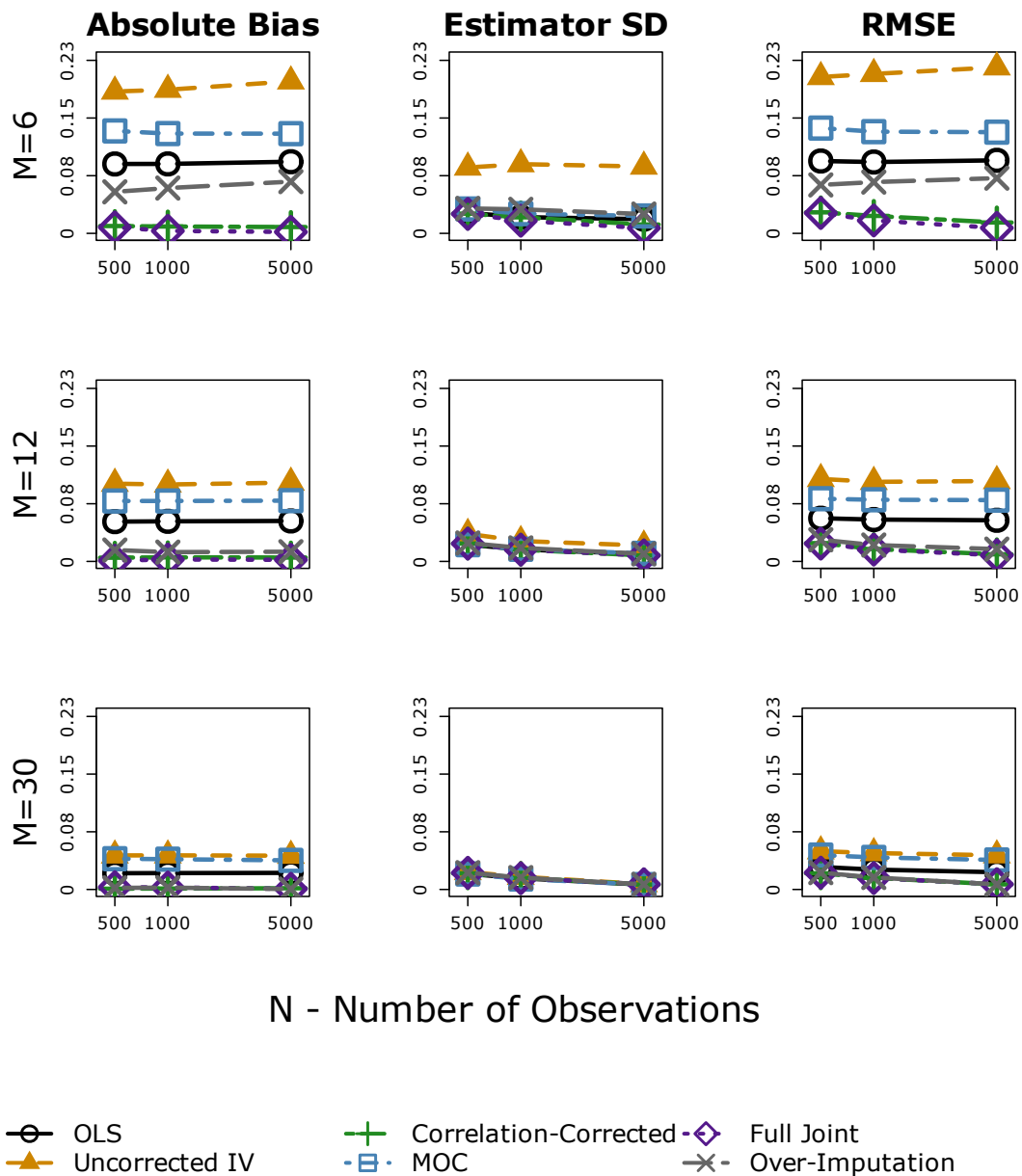


Figure A.VIII.3: Additional estimation strategies using winsorized-mean aggregation over partitions. Results add uncorrected IV to the comparison of OLS, MOC, over-imputation, full joint estimation, and the split-indicator correlation-corrected estimator; median aggregation remains the baseline specification in the main text.