Statistica Sinica Preprint No: SS-2022-0331	
Title	A Linear Errors-in-Variables Model with Unknown Heteroscedastic Measurement Errors
Manuscript ID	SS-2022-0331
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0331
Complete List of Authors	Linh H. Nghiem and Cornelis J. Potgieter
Corresponding Authors	Linh H. Nghiem
E-mails	linh.nghiem@sydney.edu.au
Notice: Accepted version subject to English editing.	

Statistica Sinica

A Linear Errors-in-Variables Model with Unknown Heteroscedastic Measurement Errors

Linh H. Nghiem[*] and Cornelis J. Potgieter[**]

[*]*University of Sydney*

[**]*Texas Christian University & University of Johannesburg*

Abstract: In the classic measurement error framework, covariates are contaminated by independent additive noise. This paper considers parameter estimation in such a linear errors-in-variables model where the unknown measurement error distribution is heteroscedastic across observations. We propose a new generalized method of moment (GMM) estimator that combines a moment correction approach and a phase function-based approach. The former requires distributions to have four finite moments, while the latter relies on covariates having asymmetric distributions. The new estimator is shown to be consistent and asymptotically normal under appropriate regularity conditions. The asymptotic covariance of the estimator is derived, and the estimated standard error is computed using a fast bootstrap procedure. The GMM estimator is demonstrated to have strong finite sample performance in numerical studies, especially when the measurement errors follow non-Gaussian distributions.

Key words and phrases: Asymmetric Distributions; Bootstrap; Generalized Method of Moments; Nutrition; Phase Function; Variance Heterogeneity.

1. Introduction

The errors-in-variables linear model arises when certain covariates suffer from measurement error contamination. This can stem from sources like instrumentation and self-reporting errors, as well as the inadequate use of short-term measurements as proxies for long-term variables. Ignoring measurement error can result in biased estimators, see Carroll et al. (2006) regarding the importance of measurement error correction in understanding the effects of the covariates on the outcome. This paper considers a heteroscedastic measurement error setting, allowing the measurement error covariance to vary across observations. This observation-specific measurement error variance structure, treated as unknown, requires estimation from replicate data. We adopt the classic additive measurement error model wherein the contaminated covariates, i.e the surrogates, are treated as the sum of the true covariates and independent measurement errors, so surrogate variances exceed true covariate variances.

One of the first papers to address the problem of a predictor variable contaminated by measurement error is Wald (1940). Since then, many parametric methods have been proposed, such as the maximum likelihood approach of Higdon and Schafer (2001). The conditional scores approach (Stefanski and Carroll, 1987) and the conditional quasi-likelihood approach (Hanfelt and Liang, 1997) require the conditional distributions of the outcomes and the contaminated covariates to be specified, both in terms of the true covariates. Regression calibration (Carroll and Stefanski, 1990) estimates the true covariates from the contaminated covariates in a validation sample. Simulation-extrapolation (SIMEX) (Stefanski and Cook, 1995) is a computationally-intensive method that adds additional measurement error to estimate model parameters and then extrapolates to the error-free case. All the methods listed in this paragraph require parametric specifications for some distributional components of the model.

Our paper proposes an efficient distribution-free estimator for a linear errors-in-variables model with heteroscedastic measurement errors. Our estimator combines two existing methods: a moment correction approach and a phase-function based estimator. The moment correction approach dates back to Reiersøl (1941) and has also been considered by Gillard (2014) and Erickson et al. (2014). This method requires the existence of model moments up to order $2M$, where $M \geq 2$ is the number of moments that are used to derive moment equations. In contrast, the phase function-based estimator, proposed by Nghiem et al. (2020), does not impose moment conditions on the underlying random variables, but requires the true covariates to have asymmetric distributions. These authors define minimum distance estimators based on the difference of two empirical phase functions. Our paper proposes a new method combining the moment-based and phase function-based approaches within a generalized method of moments (GMM) framework.

We assume that the measurement errors have a joint symmetric distribution, but both the distribution type and the observation-specific covariance matrix are unknown. Our estimator relaxes the asymmetry condition of Nghiem et al. (2020) for a common scenario in practice where all observations have at least two replicates (Carroll et al., 2006). We propose two different weighting schemes to construct weighted empirical phase functions that account for measurement error heteroscedasticity. Furthermore, we present a computationally efficient bootstrap technique to estimate the covariance matrix, serving both GMM estimator computation and standard error estimation. Simulation studies and a data application demonstrate that our combined GMM estimator have strong finite sample performance.

The remaining sections of the paper are organized as follows. Section 2 reviews moment-corrected and phase function estimators. Section 3 introduces the GMM estimator along with a bootstrap approach for covariance matrix estimation. Section 4 conducts simulation studies to compare estimation methods and assess standard error recovery. Section 5 presents an illustrative NHANES dataset analysis, and Section 6 concludes the study.

2. Moment-corrected and phase function estimators

2.1 Heteroscedastic errors-in-variables model formulation

Let $\mathcal{X} = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$ denote an observed sample following a linear errors-in-variables (EIV) model. Here, $\mathcal{D}_j = (\mathbf{W}_j^{(n_j)}, \tilde{\mathbf{Z}}_j, y_j)$ is a random sample following EIV model structure that

$$y_j = \mathbf{X}_j^\top \boldsymbol{\beta}_0 + \mathbf{Z}_j^\top \boldsymbol{\gamma}_0 + \varepsilon_j \quad \text{and} \quad \mathbf{W}_{jk} = \mathbf{X}_j + \mathbf{U}_{jk}, \tag{2.1}$$

for $k = 1, \ldots, n_j$ and $j = 1, \ldots, n$. In this model, we let $y_j \in \mathbb{R}$ denote the outcome of interest, $\mathbf{X}_j = (X_{j1}, \ldots, X_{jp})^\top \in \mathbb{R}^p$ denote the true values of the error-prone covariates, $\mathbf{Z}_j = \left(1, \tilde{\mathbf{Z}}_j^\top\right)^\top$ with $\tilde{\mathbf{Z}}_j = (Z_{j1}, \ldots, Z_{jq})^\top \in \mathbb{R}^q$ denote the error-free covariates, and $\mathbf{W}_{jk} = (W_{jk,1}, \ldots, W_{jk,p})^\top$ denote a contaminated version of $\mathbf{X}_j$ subject to independent measurement error $\mathbf{U}_{jk} = (U_{jk,1}, \ldots, U_{jk,p})^\top$. Furthermore, we let $\mathbf{W}_j^{(n_j)} = (\mathbf{W}_{j1}, \ldots, \mathbf{W}_{jn_j})$ and $\mathbf{U}_j^{(n_j)} = (\mathbf{U}_{j1}, \ldots, \mathbf{U}_{jn_j})$ denote, respectively, the collections of contaminated replicates and measurement errors associated with the $n_j \geq 2$ replicates of the jth observation, and ε_j denote the usual regression error. Also, $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0p})^\top \in \mathbb{R}^p$ and $\boldsymbol{\gamma}_0 = (\gamma_{00}, \gamma_{01}, \ldots, \gamma_{0q})^\top \in \mathbb{R}^{q+1}$ denote, respectively, the coefficients vectors associated with $\mathbf{X}_j$ and $\mathbf{Z}_j$. We assume the measurement errors $\mathbf{U}_{jk}$ have mean zero and covariance $\boldsymbol{\Sigma}_j$ that is potentially distinct for all observations. Consequently, we have $\text{Var}[\mathbf{W}_{jk}] = \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_j$ where $\boldsymbol{\Sigma}_x = \text{Var}[\mathbf{X}_j]$ for $k = 1, \ldots, n_j$

and $j = 1, \ldots, n$. The regression errors ε_j's are assumed to be independent and identically distributed (iid) with mean zero and variance σ_ε^2.

In subsequent sections of this paper, several conditions will be important when considering the estimation methods for the linear EIV model (2.1). These conditions are now presented and discussed. To this end, let operator $\perp$ denote the independence of random quantities, let $\|\boldsymbol{A}\|_{\max} = \max\{|a_{jk}|\}$ denote the element-wise infinity norm of an arbitrary matrix $\boldsymbol{A}$, and let $i = \sqrt{-1}$ denote the imaginary unit.

Condition C1. *For the jth observation, $\mathbf{U}_{jk} \perp \mathbf{U}_{jk'}$ for $k \neq k'$, $k, k' \in \{1, \ldots, n_j\}$. Furthermore, $\mathbf{U}_j^{(n_j)} \perp (\mathbf{X}_j, \mathbf{Z}_j, \varepsilon_j)$. Finally, the random quantities $(\mathbf{X}_j, \mathbf{Z}_j, \varepsilon_j)$, $j = 1, \ldots, n$ are iid copies of random variables $(\mathbf{X}, \mathbf{Z}, \varepsilon)$.*

In addition to requiring independent observations, Condition C1 requires that the measurement error components across the replicates associated within a given observation are mutually independent. Furthermore, the measurement errors, true covariates, and regression errors are required to be mutually independent as well. An example of a well-studied scenario that would violate this assumption is the Berkson error model wherein the observed predictor $\mathbf{W}_j$ has a smaller variance than the true predictor $\mathbf{X}_j$, see Song (2021) for an overview. Many time-series error models would also violate Condition C1. Nevertheless, these settings are outside the scope of the current paper.

Condition C2. *For the jth observation, random quantities* $\tilde{\mathbf{X}}_j = [\mathbf{X}_j^\top, \tilde{\mathbf{Z}}_j^\top]^\top$ *and* ε_j *satisfy* $\mathrm{E}\left(\|\tilde{\mathbf{X}}_j \tilde{\mathbf{X}}_j^\top\|_{\max}^4\right) < \infty$ *and* $\mathrm{E}(\varepsilon_j^4) < \infty$*. Also, for the replicates associated with observation j,* $\mathrm{E}\left(\|\mathbf{U}_{jk}\,\mathbf{U}_{jk}^\top\|_{\max}^4\right) < \infty$*,* $j = 1, \ldots, n$*.*

While this paper imposes no parametric distributional assumption on the underlying variables in the model, we do require that the covariates, regression errors, and measurement errors have distributions with at least four finite moments. Examples of situations where Condition C2 would be violated would be if the covariates and/or measurement error terms followed a multivariate t distribution with fewer than 4 degrees of freedom, or a multivariate stable law with index parameter $\alpha < 2$. Condition C2 is central to the moment-corrected approach of Section 2.2 having an asymptotic normal distribution. While the phase function-based approach of Section 2.3 is less concerned with the higher-order moments, the following two conditions are central to it.

Condition C3. *The replicate measurement error vectors* $\mathbf{U}_{jk}$ *have a distribution symmetric about zero with strictly positive characteristic function,* $\phi_{\mathbf{U}_j}(\boldsymbol{t}) = \mathrm{E}[\exp(i\boldsymbol{t}^\top \mathbf{U}_{jk})] > 0,\ k = 1, \ldots, n_j$*, for all* $\boldsymbol{t} = (t_1, \ldots, t_p)^\top \in \mathbb{R}^p$ *and possibly distinct for each* $j = 1, \ldots, n$*. Similarly, the regression errors* ε_j *have a distribution that is symmetric around zero with strictly positive characteristic function* $\phi_\varepsilon(t) = \mathrm{E}[\exp(it\varepsilon_j)] > 0$ *for all* $t \in \mathbb{R}$ *and common to all* $j = 1, \ldots, n$*.*

Many distributions commonly encountered in the measurement error literature satisfy Condition C3, including the Gaussian, the Laplace, and Student's t distributions. Excluded by this condition are distributions only taking values on a bounded interval, for example, the (multivariate) uniform distribution, because the corresponding characteristic functions are negative for some $\boldsymbol{t}$.

In order to state the next condition, the phase function of a random variable has to be defined. For an arbitrary random variable V, let $\phi_V(t)$ denote the characteristic function of V. The phase function of V is then defined as $\rho_V(t) = \phi_V(t)/|\phi_V(t)|$ with $|\phi_V(t)|^2 = \phi_V(t)\bar{\phi}_V(t)$ being the squared complex norm and $\bar{\phi}_V(t)$ the complex conjugate of $\phi_V(t)$. For a more in-depth discussion of the phase function and its properties, consult Delaigle and Hall (2016) and Nghiem et al. (2020).

Condition C4. *Let $V(\boldsymbol{\beta},\boldsymbol{\gamma}) = \mathbf{X}^\top\boldsymbol{\beta} + \mathbf{Z}^\top\boldsymbol{\gamma}$ and let $\rho_V(t|\,\boldsymbol{\beta},\boldsymbol{\gamma})$ denote the corresponding phase function of V. Note that this phase function depends on parameters $(\boldsymbol{\beta},\boldsymbol{\gamma})$. Then, $\rho_V(t|\,\boldsymbol{\beta},\boldsymbol{\gamma})$ is continuously differentiable with respect to all elements of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Furthermore, $\partial\rho_V(t|\,\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\beta_k \neq 0$ for $k = 1,\ldots,p$ and $\partial\rho_V(t|\,\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\gamma_k \neq 0$ for $k = 0,\ldots,q$.*

Condition C4 may appear esoteric. In essence, this condition imposes a joint skewness requirement on the true covariates, as any symmetric variable independent of the other variables will not contribute to the phase function in any way. A related sufficient condition used by Nghiem et al. (2020) is that for covariates $\tilde{\mathbf{X}} = (\mathbf{X}^\top, \tilde{\mathbf{Z}}^\top)^\top = (\tilde{X}_1, \ldots, \tilde{X}_{p+q})^\top$ and true parameter values $\tilde{\boldsymbol{\beta}}_0 = (\tilde{\beta}_{01}, \ldots, \tilde{\beta}_{0,p+q}) = (\beta_{01}, \ldots, \beta_{0p}, \gamma_{01}, \ldots, \gamma_{0q})^\top$, there exists no subset of variables $\mathcal{P} \subseteq \{1, \ldots, p+q\}$ such that $\sum_{k \in \mathcal{P}} \tilde{\beta}_{0k} \tilde{X}_k$ has a symmetric distribution.

Condition C5. *The true parameter* $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\gamma}_0^\top)^\top$ *is an interior point of a compact and convex parameter space* $\boldsymbol{\Theta} \subseteq \mathbb{R}^{p+q+1}$.

Condition C5 is a regularity condition imposed on the parameter space; here it is satisfied when all parameter values are finite.

2.2 Moment correction

Moment correction is a well-established approach to estimate the parameters of the heteroscedastic EIV model as per equation (2.1). As moment correction is also central to the new estimation method proposed in Section 3, a brief overview is provided here. The interested reader can consult Buonaccorsi (2010, Section 5.4) for more details.

Let $\boldsymbol{W}_j = n_j^{-1} \sum_{k=1}^{n_j} \boldsymbol{W}_{jk} = \boldsymbol{X}_j + \boldsymbol{U}_j$ denote the averaged contaminated replicates of the jth observation. Here, $\boldsymbol{U}_j = n_j^{-1} \sum_{k=1}^{n_j} \boldsymbol{U}_{jk}$, whose variance is $\mathrm{Var}\,(\boldsymbol{U}_j) = n_j^{-1} \boldsymbol{\Sigma}_j$. Moment correction relies on the corrected L_2 norm $L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = n^{-1} \sum_{j=1}^{n} \left(y_j - \mathbf{W}_j^\top \boldsymbol{\beta} - \mathbf{Z}_j^\top \boldsymbol{\gamma}\right)^2 - n^{-1} \sum_{j=1}^{n} n_j^{-1} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_j \boldsymbol{\beta}$, which satisfies $\mathrm{E}[L(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)] = \sigma_\varepsilon^2$ when Condition C1 holds. The gradients corresponding to this function are

$$\mathbf{S}_{L,\beta} = \frac{\partial L}{\partial \boldsymbol{\beta}} = -\frac{2}{n}\sum_{j=1}^{n} \mathbf{W}_j \left(y_j - \mathbf{W}_j^\top \boldsymbol{\beta} - \mathbf{Z}_j^\top \boldsymbol{\gamma}\right) - \frac{2}{n}\sum_{j=1}^{n} \frac{1}{n_j} \boldsymbol{\Sigma}_j \boldsymbol{\beta},$$
$$\mathbf{S}_{L,\gamma} = \frac{\partial L}{\partial \boldsymbol{\gamma}} = -\frac{2}{n}\sum_{j=1}^{n} \mathbf{Z}_j \left(y_j - \mathbf{W}_j^\top \boldsymbol{\beta} - \mathbf{Z}_j^\top \boldsymbol{\gamma}\right).$$

The moment corrected estimator is subsequently defined to be the solution of the $p+q+1$ estimating equations $\mathbf{S}_L = [\mathbf{S}_{L,\beta}^\top, \mathbf{S}_{L,\gamma}^\top]^\top = \mathbf{0}$. These estimating equations are well-defined in a statistical sense whenever the underlying variables have at least two finite moments. However, to establish that the estimators are asymptotically normally distributed, Conditions C2 and C5 are required to ensure the variances of the gradients $\mathbf{S}_{L,\beta}$ and $\mathbf{S}_{L,\gamma}$ are finite.

To implement the above moment correction, knowledge of the variance-covariance matrices $\boldsymbol{\Sigma}_j$ is required. When these are unknown, they can be consistently estimated using

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{n_j(n_j-1)} \sum_{k=1}^{n_j-1} \sum_{k'>k}^{n_j} (\mathbf{W}_{jk} - \mathbf{W}_{jk'})(\mathbf{W}_{jk} - \mathbf{W}_{jk'})^\top. \tag{2.2}$$

The covariance estimator in (2.2) follows upon noting that

$$\mathrm{E}\left[(\mathbf{W}_{jk} - \mathbf{W}_{jk'})(\mathbf{W}_{jk} - \mathbf{W}_{jk'})^\top\right] = \mathrm{E}\left[(\mathbf{U}_{jk} - \mathbf{U}_{jk'})(\mathbf{U}_{jk} - \mathbf{U}_{jk'})^\top\right] = 2\,\boldsymbol{\Sigma}_j$$

for any pair of replicates $(\mathbf{W}_{jk}, \mathbf{W}_{jk'})$ with $k \neq k'$. The second equality follows from the independence of measurement error terms $(\mathbf{U}_{jk}, \mathbf{U}_{jk'})$, and the assumed mean and covariance structure of the $\mathbf{U}_{jk}$. Subsequently, the estimator $\hat{\boldsymbol{\Sigma}}_j$ is defined by averaging over the squared differences of the $n_j(n_j-1)/2$ such pairs associated with the jth observation. Therefore, in practice, the gradient $\mathbf{S}_{L,\beta}$ is replaced by an approximation $\hat{\mathbf{S}}_{L,\beta}$ which substitutes unknown $\boldsymbol{\Sigma}_j$ by $\hat{\boldsymbol{\Sigma}}_j$ for $j = 1, \ldots, n$. The moment-corrected estimators are then calculated as the solution of the estimating equation

$$\hat{\mathbf{S}}_L = [\hat{\mathbf{S}}_{L,\beta}^\top, \mathbf{S}_{L,\gamma}^\top]^\top = \mathbf{0}. \tag{2.3}$$

These estimators $\hat{\boldsymbol{\Sigma}}_j$ and the gradient $\hat{\mathbf{S}}_L$ will be further used in Section 3.

2.3 Phase function-based estimation

Recently, in the context of a homoscedastic EIV model without replicate data, Nghiem et al. (2020) proposed a phase function-based estimator. Their approach, unlike the moment-corrected estimator, does not require estimation of the measurement error covariances, but still leads to a consistent estimator even when the underlying random variables do not have finite variances. We propose here a new variation of the phase function method that adjusts for heteroscedasticity of the measurement error, which is made possible by the availability of replicates. Furthermore, the asymmetric linear combination assumption of Nghiem et al. (2020) is replaced by Condition C4.

For the model defined in (2.1), define $V_{0j} = \mathbf{X}_j^\top \boldsymbol{\beta}_0 + \mathbf{Z}_j^\top \boldsymbol{\gamma}_0$ so that $y_j = V_{0j} + \varepsilon_j$, $j = 1, \dots, n$. Since $(\mathbf{X}_j, \mathbf{Z}_j, \varepsilon_j)$ are independent copies of $(\mathbf{X}, \mathbf{Z})$ by Condition C1, the random variable V_{0j} is an independent copy of $V_0 := V(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) = \mathbf{X}^\top \boldsymbol{\beta}_0 + \mathbf{Z}^\top \boldsymbol{\gamma}_0$ and y_j is an independent copy of $Y = V_0 + \varepsilon$. Recall that $\phi_Y(t)$ and $\rho_Y(t)$ denote, respectively, the characteristic and phase functions of Y, and the same holds for $\phi_{V_0}(t)$ and $\rho_{V_0}(t)$ in terms of V_0. From Condition C3, the phase function $\rho_Y(t)$ is given by

$$\rho_Y(t) = \frac{\phi_Y(t)}{|\phi_Y(t)|} \overset{(i)}{=} \frac{\phi_{V_0}(t)\phi_\varepsilon(t)}{|\phi_{V_0}(t)\phi_\varepsilon(t)|} \overset{(ii)}{=} \frac{\phi_{V_0}(t)}{|\phi_{V_0}(t)|} = \rho_{V_0}(t),$$

where (i) follows from the independence of the regression error as per Condition C1, and (ii) follows from Condition C3. Moreover, recalling that $\mathbf{W}_j$ and $\mathbf{U}_j$ are the averaged replicates and measurement error terms, respectively. Since the term $\mathbf{U}_j^\top \boldsymbol{\beta}$ also has a symmetric distribution around zero, a similar argument shows that the phase function of arbitrary linear combination $\tilde{V}_j(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma} = \mathbf{X}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma} + \mathbf{U}_j^\top \boldsymbol{\beta}$ is the same as the phase function of $V(\boldsymbol{\beta}, \boldsymbol{\gamma})$, i.e $\rho_{\tilde{V}_j}(t|\,\boldsymbol{\beta}, \boldsymbol{\gamma}) = \rho_V(t|\,\boldsymbol{\beta}, \boldsymbol{\gamma})$ for all $j = 1, \dots, n$.

Our method now proceed by equating two different empirical phase functions. Firstly, based on the outcomes y_j, define empirical phase function

$$\hat{\rho}_Y(t) = \frac{\sum_{j=1}^n \exp(ity_j)}{\left[\sum_{j=1}^n \sum_{k=1}^n \exp\left\{it\left(y_j - y_k\right)\right\}\right]^{1/2}},$$

and based on covariates $(\mathbf{W}_j, \mathbf{Z}_j)$, we define the *weighted* empirical phase function (WEPF),

$$\hat{\rho}_V(t|\,\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{\sum_{j=1}^n q_j \exp\left\{it(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma})\right\}}{\left(\sum_{j=1}^n \sum_{k=1}^n q_j q_k \exp\left[it\left\{(\mathbf{W}_j - \mathbf{W}_k)^\top \boldsymbol{\beta} + (\mathbf{Z}_j - \mathbf{Z}_k)^\top \boldsymbol{\gamma}\right\}\right]\right)^{1/2}}. \tag{2.4}$$

where the weights $\{q_j\}_{j=1}^n$ satisfy $q_j \geq 0$ and $\sum_{j=1}^n q_j = 1$. The phase function-based estimator is motivated by noting that the population level equivalents $\rho_Y(t)$ and $\rho_V(t|\,\boldsymbol{\beta}, \boldsymbol{\gamma})$, under the asymmetry imposed on the covariates by Condition C4, are equal if and only if $(\boldsymbol{\beta}, \boldsymbol{\gamma}) = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$. Thus, the estimator for $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ is defined to be the minimizer of the discrepancy

$$D(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \int_{-\infty}^{\infty} |\hat{\rho}_Y(t) - \hat{\rho}_V(t|\,\boldsymbol{\beta}, \boldsymbol{\gamma})|^2 \omega(t) dt, \tag{2.5}$$

where $\omega(t)$ is a weighting function to ensure the integral is finite. Direct minimization of (2.5) is computationally expensive. Following the example of Nghiem et al. (2020), we consider the alternative minimization problem

$$\tilde{D}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \int_0^{t^*}\left[C_y(t)\sum_{j=1}^n q_j \sin\left\{t\left(\mathbf{W}_j^\top\boldsymbol{\beta}+\mathbf{Z}_j^\top\boldsymbol{\gamma}\right)\right\} - S_y(t)\sum_{j=1}^n q_j \cos\left\{t\left(\mathbf{W}_j^\top\boldsymbol{\beta}+\mathbf{Z}_j^\top\boldsymbol{\gamma}\right)\right\}\right]^2 K_{t^*}(t)dt \qquad (2.6)$$

where $C_y(t) = n^{-1}\sum_{j=1}^n \cos(t\,y_j)$, $S_y(t) = n^{-1}\sum_{j=1}^n \sin(t\,y_j)$, and $K_{t^*}(t)$ is a kernel function that is only non-zero on the interval $[0,t^*]$. Note that the two forms (2.5) and (2.6) are equivalent for appropriate choices of $\omega(t)$ and $K_{t^*}(t)$. Following Delaigle and Hall (2016), we use $K_{t^*}(t) = (1-t/t^*)^2$ for numerical implementation with t^* being the largest t such that $|\hat{\phi}_y(t)| \leq n^{-1/2}$. For any fixed $(\boldsymbol{\beta},\boldsymbol{\gamma})$, by a similar argument to Nghiem and Potgieter (2018), the WEPF in (2.4) is a consistent estimator of $\rho_V(\boldsymbol{\beta},\boldsymbol{\gamma})$ for any set of weights having $\max_j q_j = O(n^{-1})$. Therefore, our intended goal is to adjust for measurement error heteroscedasticity through an appropriate choice of weights; we will elaborate on this in Section 3.2.

3. A GMM estimator

3.1 Generalized method of moments

The moment-corrected and phase function-based estimators from Section 2 rely on different yet complementary sets of assumptions. On the one hand, moment correction is a least squares approach with a variance correction term, thus using information from the first two underlying model moments. On the other hand, as discussed in Nghiem et al. (2020), the phase function method essentially uses all *odd* moments of the underlying model, provided these moments exist. Hence, a method combining these two approaches will generally make use of more model information, allowing for the possibility of a more efficient estimator. In this paper, we describe how to combine the two methods using a generalized method of moments (GMM) approach, which is widely used to estimate parameters in over-identified systems (Hansen, 1982).

To define the GMM estimator, recall that the phase function-based estimator is found by minimizing function $\tilde{D}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ in (2.6). The minimizer can be expressed as the solution of the $p+q+1$ estimating equations $\mathbf{S}_{\tilde{D}} = [\mathbf{S}_{\tilde{D},\boldsymbol{\beta}}^\top, \mathbf{S}_{\tilde{D},\boldsymbol{\gamma}}^\top]^\top = \mathbf{0}$ where

$$\mathbf{S}_{\tilde{D},\boldsymbol{\beta}} = \frac{\partial \tilde{D}}{\partial \boldsymbol{\beta}} \text{ and } \mathbf{S}_{\tilde{D},\boldsymbol{\gamma}} = \frac{\partial \tilde{D}}{\partial \boldsymbol{\gamma}}. \quad (3.7)$$

Thus, if we simultaneously consider the estimating equations from moment correction, $\hat{\mathbf{S}}_L = \mathbf{0}$ in (2.3), and the phase-function based estimators, $\mathbf{S}_{\tilde{D}} = \mathbf{0}$ defined above in (3.7), we have a system of $2(p+q+1)$ estimating equations in terms of the $p+q+1$ model parameters. More formally, let

$$\mathbf{S} := \mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = [\hat{\mathbf{S}}_L^\top, \mathbf{S}_D^\top]^\top = [\hat{\mathbf{S}}_{L,\boldsymbol{\beta}}^\top, \mathbf{S}_{L,\boldsymbol{\gamma}}^\top, \mathbf{S}_{\tilde{D},\boldsymbol{\beta}}^\top, \mathbf{S}_{\tilde{D},\boldsymbol{\gamma}}^\top]^\top \quad (3.8)$$

denote the vector of $2(p+q+1)$ gradient equations. The system $\mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{0}$ is generally over-identified and does not have an exact solution. Suppressing the dependence of $\mathbf{S}$ on $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, we define a quadratic form in $\mathbf{S}$,

$$Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{S}^\top \boldsymbol{\Omega}_S^{-1} \mathbf{S}, \tag{3.9}$$

where $\boldsymbol{\Omega}_S = \big[\mathrm{Var}(\mathbf{S})\big]_{(\boldsymbol{\beta}, \boldsymbol{\gamma})=(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)}$ is the $2(p+q+1) \times 2(p+q+1)$ covariance matrix corresponding of the gradient equations $\mathbf{S}$ evaluated at the true parameter values $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$. The GMM estimator is then defined to be the minimizer of $Q(\boldsymbol{\beta}, \boldsymbol{\gamma})$. Minimizing the GMM discrepancy function $Q(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is equivalent to projecting $\mathbf{S}$ onto a $(p+q+1)$-dimensional subspace and solving the resulting equations, so the GMM estimator can be thought of as the solution to an optimal linear combination of the estimating equations in $\mathbf{S}$.

We next study the asymptotic properties of the proposed GMM estimator. Both the consistency and asymptotic normality of our proposed estimator follow from the properties of the GMM approach under suitable regularity conditions. First, we establish the uniform convergence of $Q(\boldsymbol{\beta}, \boldsymbol{\gamma})$.

Lemma 1. *Assume that all random variables in the model* (2.1) *have at least two finite moments. Then, for* $(\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \boldsymbol{\Theta}$ *as per Condition C5, the function* $Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) \xrightarrow{p} Q_0(\boldsymbol{\beta}, \boldsymbol{\gamma})$ *uniformly, where* $Q_0(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{S}_0^\top \boldsymbol{\Omega}_S^{-1} \mathbf{S}_0$ *and* $\mathbf{S}_0 :=$

$\mathbf{S}_0(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \lim_{n\to\infty} \mathrm{E}[\mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\gamma})]$ *is the limiting expectation of the gradient equations.*

The proof of Lemma 1 is presented in Section S1 of the Supplementary Materials. The proof relies of verifying sufficient conditions for uniform convergence as per Lemma 2.9 of Newey and McFadden (1994). The uniform convergence of $Q(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is an essential step in establishing the consistency of the GMM estimator in our next theorem.

Theorem 1. *Consider the heteroscedastic linear EIV model defined in* (2.1)*. Assume Conditions C1, C3 and C5 hold. Also assume that all variables in the model have at least two finite moments. Finally, assume the weights* q_j *used for constructing the empirical phase function in* (2.4) *satisfy* $\max_j q_j = O(n^{-1})$*. Then, the estimator obtained by minimizing* $Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{S}^\top \boldsymbol{\Omega}_S^{-1} \mathbf{S}$ *is consistent for true value* $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$*.*

Theorem 1 follows from Theorem 2.1 of Newey and McFadden (1994) combining the uniform convergence in Lemma 1 along with establishing that $Q_0(\boldsymbol{\beta}, \boldsymbol{\gamma})$ has a global minimum at the true parameters $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$. The proof is presented in Section S2 of the Supplementary Materials. We further note that for the GMM estimator to be consistent, Condition C4 (covariate asymmetry) is not a requirement. When it does not hold, some of the components of $\mathbf{S}$ converge to 0 for all values of the underlying parameters. However, the limiting quadratic form still has a unique minimum at the true parameter values, because the estimating equations originating with the moment-corrected approach do not rely on asymmetry. With this in mind, we focus the GMM method on a situation where $\partial\rho_V(t|\,\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\beta_k \neq 0$ for *some* $k = 1,\ldots,p$ and/or $\partial\rho(t|\,\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\gamma_k \neq 0$ for *some* $k = 1,\ldots,q$. When some of these partial derivatives are non-zero, then evaluation of the empirical phase function contributes information to the estimation procedure and it is possible that the efficiency of the estimator is improved. Finally, we establish the asymptotic normality of the proposed estimator.

Theorem 2. *Consider the heteroscedastic linear EIV model defined in* (2.1). *Assume Conditions C1, C2, C3, and C5 hold. Furthermore, assume the weights* q_j *used for constructing the empirical phase function in* (2.4) *satisfy* $\max_j q_j = O(n^{-1})$. *Then, the estimator* $(\hat{\boldsymbol{\beta}}_{\text{gmm}}, \hat{\gamma}_{\text{gmm}})$ *obtained by minimizing* $Q(\boldsymbol{\beta},\boldsymbol{\gamma}) = \mathbf{S}^\top\boldsymbol{\Omega}_S^{-1}\mathbf{S}$ *satisfies*

$$n^{1/2}\left\{\left(\hat{\boldsymbol{\beta}}_{\text{gmm}}^\top\ \hat{\gamma}_{\text{gmm}}^\top\right)^\top - \left(\boldsymbol{\beta}_0^\top\ \ \boldsymbol{\gamma}_0^\top\right)^\top\right\} \sim N\left(\mathbf{0}, \left(\boldsymbol{P}_1\boldsymbol{\Omega}_S^{-1}\boldsymbol{P}_1^\top\right)^{-1}\right)$$

where

$$\boldsymbol{P}_1 = \mathrm{E}\left[\left(\frac{\partial\,\mathbf{S}}{\partial\,\boldsymbol{\beta}}^\top, \frac{\partial\,\mathbf{S}}{\partial\,\boldsymbol{\gamma}}^\top\right)\right]$$

with the expectations in $\boldsymbol{P}_1$ *evaluated at the true parameter values* $(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)$.

Theorem 2 follows from Theorem 3.4 of Newey and McFadden (1994). Compared to the required conditions for the consistency established in Theorem 1, Theorem 2 requires stronger moment conditions that all the variables in the model having four finite moments as per Condition C2. We furthermore note that versions of Theorems 1 and 2 still hold if we replace the covariance matrix $\mathbf{\Omega}_S$ by any positive definite matrix $\mathbf{\Omega}_*$. Nevertheless, the choice of $\mathbf{\Omega}_S$ leads to the most asymptotically efficient estimator, as discussed in Section 5.2 of Newey and McFadden (1994).

In practice, the covariance matrix $\mathbf{\Omega}_S$ is unknown and needs to be replaced by a suitable estimator $\hat{\mathbf{\Omega}}_S$. As per Section 4 of Newey and McFadden (1994), using a consistent estimator of $\mathbf{\Omega}_S$ leaves the asymptotic distribution of the estimators unchanged. We therefore propose a bootstrap resampling algorithm in Section 3.3 to obtain a suitable estimator $\hat{\mathbf{\Omega}}_S$. First, however, we explore the calculation of weights q_j, $j = 1, \ldots, n$ in the WEPF to adjust for measurement error heteroscedasticity.

3.2 Choice of phase function weights

The weighted phase function in (2.4) requires the specification of weights $\{q_j\}_{j=1}^n$ with $q_j \geq 0$ and $\sum_{j=1}^n q_j = 1$. If the underlying distributions were known, it would be possible to directly minimize an asymptotic variance metric directly related to this estimated phase function – see Nghiem and Potgieter (2018) for a derivation of the pointwise asymptotic variance of such a WEPF. However, since the underlying distributions are assumed unknown, the direct approach is unavailable to us. We therefore present here two weighting schemes that are easily implemented in practice.

For the first approach, note that the variable $\sum_{j=1}^{n} q_j \, \mathbf{W}_j^\top \boldsymbol{\beta}_0$ is an unbiased estimator of $\mathrm{E}(\mathbf{X}^\top \boldsymbol{\beta}_0)$. Therefore, one can consider finding a set of weights that minimize the variance of the above mean estimator. Specifically, we have $\mathrm{Var}\left(\sum_j q_j \, \mathbf{W}_j^\top \, \boldsymbol{\beta}_0\right) = \sum_j q_j^2 \, \boldsymbol{\beta}_0^\top (\boldsymbol{\Sigma}_x + n_j^{-1} \, \boldsymbol{\Sigma}_j) \, \boldsymbol{\beta}_0$, which is minimized by weights $q_j = a_j^{-1} / \sum_{k=1}^{n} a_k^{-1}$, where $a_j = \boldsymbol{\beta}_0^\top (\boldsymbol{\Sigma}_x + n_j^{-1} \, \boldsymbol{\Sigma}_j) \, \boldsymbol{\beta}_0$. Unfortunately, these weights depend on the unknown true $\boldsymbol{\beta}_0$ and are therefore impossible to calculate. We propose the following proxy estimator based on a "minimax" argument. Let λ_j denote the largest eigenvalue of $\boldsymbol{\Sigma}_x + n_j^{-1} \, \boldsymbol{\Sigma}_j$. We then have $a_j \leq \lambda_j \| \, \boldsymbol{\beta}_0 \, \|^2$ with $\| \, \boldsymbol{\beta}_0 \, \|^2 = \sum_{k=1}^{p} \beta_{0k}^2$ denoting the squared L_2 norm. Therefore, we calculate weights based on replacing the a_j by the corresponding upper bounds. Note that using the upper bounds has the potential to *underweight* observations with the large measurement error i.e. their influence is even further mitigated. The proposed minimax weights are given by

$$\hat{q}_j^{\mathrm{mm}} = \frac{\hat{\lambda}_j^{-1}}{\sum_{k=1}^{n} \hat{\lambda}_k^{-1}}, j = 1, \ldots, n, \tag{3.10}$$

where $\hat{\lambda}_j$ is the largest eigenvalue of $\hat{\mathbf{\Sigma}}_x + n_j^{-1}\hat{\mathbf{\Sigma}}_j$ with $\hat{\mathbf{\Sigma}}_j$ given in (2.2) and with $\hat{\mathbf{\Sigma}}_x = (n-1)^{-1}\sum_{i=1}^n(\mathbf{W}_j - \overline{\mathbf{W}})(\mathbf{W}_j - \overline{\mathbf{W}})^\top - n^{-1}\sum_{j=1}^n n_j^{-1}\hat{\mathbf{\Sigma}}_j$, with $\overline{\mathbf{W}} = n^{-1}\sum_{j=1}^n \mathbf{W}_j$.

For the second approach, note that for all $j = 1, \ldots, n$, we have $\mathrm{E}(\mathbf{W}_j) = \mathrm{E}(\mathbf{X})$. Therefore, the quantity $\hat{\boldsymbol{\mu}}_q = \sum_{j=1}^n q_j \mathbf{W}_j$ is an unbiased estimator of $\mathrm{E}(\mathbf{X})$. Recalling that $\mathrm{Var}(\mathbf{W}_j) = \mathbf{\Sigma}_x + n_j^{-1}\mathbf{\Sigma}_j$, we define the L_2 discrepancy

$$L(\boldsymbol{q}) = \sum_{j=1}^n (\mathbf{W}_j - \hat{\boldsymbol{\mu}}_q)^\top \left(\mathbf{\Sigma}_x + n_j^{-1}\mathbf{\Sigma}_j\right)^{-1} (\mathbf{W}_j - \hat{\boldsymbol{\mu}}_q)^\top, \tag{3.11}$$

with $\boldsymbol{q} = (q_1, \ldots, q_n)$. Our second weighting scheme proposes minimizing $L(\boldsymbol{q})$ in terms of the weights $\boldsymbol{q}$ subject to $q_j \geq 0$ and $\sum_{j=1}^n q_j = 1$. We note that minimizing $L(\boldsymbol{q})$ is equivalent to maximizing the log-likelihood of the $\{\mathbf{W}_j\}_{j=1}^n$ in terms of $\boldsymbol{q}$ assuming each $\mathbf{W}_j$ follows a multivariate normal distribution. We therefore refer to this approach as the quasi-likelihood weighting scheme. In Section S3 of the Supplementary Materials, we show that the quasi-likelihood weights are found by solving a system of linear equations. Replacing $\mathbf{\Sigma}_x$ and the $\mathbf{\Sigma}_j$ in (3.11) with their corresponding estimator $\hat{\mathbf{\Sigma}}_x$ and $\hat{\mathbf{\Sigma}}_j$, we denote the resulting minimizing weights by $\hat{q}_j^{\mathrm{ql}}$ for $j = 1, \ldots, n$. The performance of the minimax and quasi-likelihood weights are further considered in the simulation studies of Section 4.

3.3 GMM covariance matrix and standard error estimation

Implementation of the proposed GMM estimator requires a suitable estimator for $\boldsymbol{\Omega}_S$ in (3.9), where $\boldsymbol{\Omega}_S$ is the covariance matrix of the $2(p+q+1)$ gradient equations $\mathbf{S}$ at the true parameter values $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$. We propose a computationally–efficient strategy based on the estimating function bootstrap approach of Hu and Kalbfleisch (1997); this strategy is also used in Nghiem et al. (2020) for estimating the variances of the phase function-based estimators.

Recall that $\mathcal{X} = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$ denotes the random sample from the linear EIV model with $\mathcal{D}_j = (\mathbf{W}_j^{(n_j)}, \tilde{\mathbf{Z}}_j, y_j)$, $j = 1, \ldots, n$. The bth bootstrap sample, $\mathcal{X}_b^* = \{\mathcal{D}_{b1}^*, \ldots, \mathcal{D}_{bn}^*\}$, $b = 1, \ldots, B$, is obtained by sampling n times with replacement from $\mathcal{X}$. No re-sampling is done at the replicate level. For the bth bootstrap sample, we compute $\hat{\boldsymbol{\Sigma}}_{b,x}^*$, $\hat{\boldsymbol{\Sigma}}_{b,j}^*$, and $\hat{q}_{b,j}^*$ which correspond to the observation-level measurement error covariance matrices, and the weights for the phase function estimator calculated using $\mathcal{X}_b$ for $j = 1, \ldots, n$. We subsequently use these quantities and our bootstrap sample to evaluate $\mathbf{S}_b^*(\hat{\boldsymbol{\beta}}_{\text{in}}, \hat{\boldsymbol{\gamma}}_{\text{in}}) = [\hat{\mathbf{S}}_{b,L}^*(\hat{\boldsymbol{\beta}}_{\text{in}}, \hat{\boldsymbol{\gamma}}_{\text{in}})^\top \ \mathbf{S}_{b,\tilde{D}}^*(\hat{\boldsymbol{\beta}}_{\text{in}}, \hat{\boldsymbol{\gamma}}_{\text{in}})^\top]^\top$ as defined in equation (3.8). Here, $(\hat{\boldsymbol{\beta}}_{\text{in}}, \hat{\boldsymbol{\gamma}}_{\text{in}})$ denote consistent initial estimators of $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$; in our implementation, the moment-corrected estimators $(\hat{\boldsymbol{\beta}}_{\text{mc}}, \hat{\boldsymbol{\gamma}}_{\text{mc}})$ are used as initial estimators. Finally, using the B bootstrap samples, we estimate $\boldsymbol{\Omega}_S$ by $\hat{\boldsymbol{\Omega}}_S^* = B^{-1}\sum_{b=1}^{B} \mathbf{S}_b^*(\hat{\boldsymbol{\beta}}_{\text{in}}, \hat{\boldsymbol{\gamma}}_{\text{in}})\, \mathbf{S}_b^*(\hat{\boldsymbol{\beta}}_{\text{in}}, \hat{\boldsymbol{\gamma}}_{\text{in}})^\top$. The method is fast and can be implemented without the need to minimize bootstrap versions of the discrepancy function $Q(\boldsymbol{\beta}, \boldsymbol{\gamma})$; implementation requires only evaluation of the gradient vector for each bootstrap sample prior to calculating $\hat{\mathbf{\Omega}}^*_S$. We have found that $B = 100$ bootstrap samples lead to a good performance. The GMM estimator is finally computed by minimizing $\hat{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{S}^\top \left(\hat{\mathbf{\Omega}}^*_S\right)^{-1} \mathbf{S}$.

As pointed out by a referee, our approach is similar to the traditional two–step feasible GMM approach, in which we first obtain a consistent estimator $(\hat{\boldsymbol{\beta}}_{\mathrm{mc}}, \hat{\boldsymbol{\gamma}}_{\mathrm{mc}})$ and then use it to estimate the covariance matrix $\mathbf{\Omega}_S$ in the GMM discrepancy function. Note that an iterated GMM approach can be pursued, whereby this estimator and a new round of bootstrap samples are used to update the GMM covariance matrix and then minimize the resulting statistic. Asymptotically, this iterated GMM estimator is equivalent to the two-step estimator, although in finite samples, the relative performance between two–step and iterative GMM estimators have been reported to be mixed, see Hansen et al. (1996).

This bootstrap quantity $\hat{\mathbf{\Omega}}^*_S$ can also subsequently be used to estimate standard errors of the GMM estimators. Based on the GMM covariance matrix from Theorem 2, the asymptotic covariance matrix of $(\hat{\boldsymbol{\beta}}_{\mathrm{gmm}}, \hat{\boldsymbol{\gamma}}_{\mathrm{gmm}})$ is estimated by $(\hat{\boldsymbol{P}}_1 \hat{\mathbf{\Omega}}^*_S \hat{\boldsymbol{P}}_1^\top)^{-1}$. Here, $\hat{\boldsymbol{P}}_1$ an empirical counterpart of the expected gradient $\boldsymbol{P}_1$ and is evaluated at the GMM estimator $(\hat{\boldsymbol{\beta}}_{\mathrm{gmm}}, \hat{\boldsymbol{\gamma}}_{\mathrm{gmm}})$. Specifically, the matrix

$\hat{\boldsymbol{P}}_1$ is given by

$$\hat{\boldsymbol{P}}_1 = \begin{bmatrix} 2n^{-1}\sum_{j=1}^{n}\mathbf{W}_j\,\mathbf{W}_j^\top - 2n^{-1}\sum_{j=1}^{n}\hat{\boldsymbol{\Sigma}}_j & 2n^{-1}\sum_{j=1}^{n}\mathbf{W}_j\,\mathbf{Z}_j^\top \\ 2n^{-1}\sum_{j=1}^{n}\mathbf{Z}_j\,\mathbf{W}_j^\top & 2n^{-1}\sum_{j=1}^{n}\mathbf{Z}_j\,\mathbf{Z}_j^\top \\ \left.\dfrac{\partial\hat{\mathbf{S}}_{\tilde{D},\boldsymbol{\beta}}}{\partial\boldsymbol{\beta}}\right|_{(\hat{\boldsymbol{\beta}}_{\rm gmm},\hat{\boldsymbol{\gamma}}_{\rm gmm})} & \left.\dfrac{\partial\hat{\mathbf{S}}_{\tilde{D},\boldsymbol{\beta}}}{\partial\boldsymbol{\gamma}}\right|_{(\hat{\boldsymbol{\beta}}_{\rm gmm},\hat{\boldsymbol{\gamma}}_{\rm gmm})} \\ \left.\dfrac{\partial\hat{\mathbf{S}}_{\tilde{D},\boldsymbol{\gamma}}}{\partial\boldsymbol{\beta}}\right|_{(\hat{\boldsymbol{\beta}}_{\rm gmm},\hat{\boldsymbol{\gamma}}_{\rm gmm})} & \left.\dfrac{\partial\hat{\mathbf{S}}_{\tilde{D},\boldsymbol{\gamma}}}{\partial\boldsymbol{\gamma}}\right|_{(\hat{\boldsymbol{\beta}}_{\rm gmm},\hat{\boldsymbol{\gamma}}_{\rm gmm})} \end{bmatrix}.$$

For implementation, the partial derivatives corresponding to the phase function criterion are calculated numerically. The estimated standard errors are given by the square root of the diagonal elements of $(\hat{\boldsymbol{P}}_1\hat{\boldsymbol{\Omega}}_S^*\hat{\boldsymbol{P}}_1^\top)^{-1}$. In Section 4.3, we consider the performance of these estimated standard errors.

4. Simulation studies

4.1 Simulation description

The finite-sample performance of the GMM estimators was evaluated using a simulation study. We considered three different settings for mutiple linear EIV models (2.1) where each observation has the same number of replicates, $n_j = n_{\rm rep}$ for $j = 1, \ldots, n$, with $n_{\rm rep} \in \{2, 3\}$. We chose sample sizes $n \in \{250, 500, 1000\}$ and generate $M = 500$ samples for each simulation configuration. A univariate linear EIV model was also considered in Section S4.1 of the Supplementary Materials. Details of the simulation settings are given as follows:

(I) A model with $p = 2$ error-prone and $q = 0$ error-free covariates. The true covariates were generated from a Gaussian copula (Xue-Kun Song, 2000), where the two covariates have scaled half-normal marginals with variance 1, i.e. $X_{jk} \overset{iid}{\sim} (1 - 2/\pi)^{-1/2} \left|N(0,1)\right|$ for $k = 1, \ldots, n_{\text{rep}}$, and correlation 0.5. Three distributions were considered for the measurement error vectors $\mathbf{U}_{jk}$, namely a bivariate normal, a bivariate t with 2.5 degrees of freedom, and a contaminated bivariate normal $\mathbf{U}_{jk} \sim 0.9\, N(\mathbf{0}, \mathbf{\Sigma}_j) + 0.1\, N(\mathbf{0}, 10^2\mathbf{\Sigma}_j)$. These distributions were all scaled to have covariance matrices $\mathbf{\Sigma}_j$ for the replicates associated with the jth observation, where $\mathbf{\Sigma}_j = \mathbf{D}_j \boldsymbol{R} \mathbf{D}_j$ with $\mathbf{D}_j$ the marginal standard deviations and $\boldsymbol{R}$ the correlation matrix. We considered two choices of $\boldsymbol{R}$, namely the identity matrix and a model with common correlation $\rho = 0.5$ between all measurement error components. The diagonal elements of the matrix $\boldsymbol{D}_j$ were independently generated from the uniform distribution $\sqrt{n_j} \times U(\sqrt{0.2}, \sqrt{1.5})$ by which the marginal signal-to-noise ratios $\text{Var}(X_{jk})/\text{Var}(U_{jk})$ range from 2/3 (fairly weak signal) to 5 (fairly strong signal). The true model coefficients were set to $\boldsymbol{\beta}_0 = (1, 0.5)^\top$ and intercept $\gamma_0 = 2$. The regression error ε_j was generated to match the distribution of the measurement error in each scenario and with constant variance $\sigma_\varepsilon^2 = 0.25$ for $j = 1, \ldots, n$.

(II) A model with $p = 2$ error-prone and $q = 2$ error-free covariates, differing from setting I due to the inclusion of two error-free covariates. The true covariates $(\mathbf{X}_j, \mathbf{Z}_j)$ were generated to have scaled half-normal marginals with the joint structure specified by a Gaussian copula with correlation 0.5 between each pair of predictors. The true model coefficients were $\boldsymbol{\beta}_0 = (1, 0.5)^\top$ and $\boldsymbol{\gamma}_0 = (2, 1, 0.5)^\top$. All other configurations are equivalent to the specifications in setting I.

(III) The model as in setting II, but the two error-free covariates were generated to be symmetric having standard normal marginals. The predictor correlation structure was still specified by a Gaussian copula with correlation parameter 0.5 between each pair of predictor variables.

Among the three error distributions considering, the $t_{2.5}$ does not have four finite moments. Consequently, asymptotic normality as per Theorem 2 does not hold in this case. However, the consistency requirements are still satisfied. We also note that setting III departs from Condition C4, allowing us to explore the effect of symmetric predictors when using the GMM method.

For each generated sample, we computed several different estimators. Firstly, we compute the true ordinary least squares (OLS) estimator which regresses y_j on the true uncontaminated data $(\mathbf{X}_j, \mathbf{Z}_j)$, as well as a naive OLS estimator which regresses y_j on $(\mathbf{W}_j, \mathbf{Z}_j)$ with $\mathbf{W}_j$ denoting the averaged replicates for the jth observations. Next, we computed the moment-corrected estimator with plug-in covariance matrices $\hat{\boldsymbol{\Sigma}}_j$ as per Section 2.2. Finally, we computed three versions of the proposed GMM estimator, corresponding to an equal weighting scheme, the minimax weights, and the quasi-likelihood weights as per Section 3.2. For the GMM estimators, covariance matrix $\hat{\boldsymbol{\Omega}}^*_S$ was estimated using $B = 100$ bootstrap samples.

A robust performance metric was adopted to evaluate estimator performance. Let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\gamma}_0^\top)^\top$ be the true coefficients, and $\hat{\boldsymbol{\theta}}^{(m)} = (\hat{\boldsymbol{\beta}}^{(m)\top}, \hat{\boldsymbol{\gamma}}^{(m)\top})^\top$ denote the estimators obtained using one of the outlined approaches in the mth generated sample, $m = 1, \ldots, M$. To remove outliers, we constructed the $M \times (p + q + 1)$ error matrix $\boldsymbol{A}$ with rows $\boldsymbol{A}_m = (\hat{\boldsymbol{\theta}}^{(m)} - \boldsymbol{\theta})^\top$. Next, we formed a vector of medians $\boldsymbol{A}_{\text{med}}$ by taking the median of each column of $\boldsymbol{A}$. Then, we computed the Mahalanobis distance between each row of $\boldsymbol{A}$ and $\boldsymbol{A}_{\text{med}}$ using the robust minimum covariance determinant estimator of Rousseeuw and Driessen (1999). Finally, we removed the rows with Mahalanobis distances larger than the 90th percentile and calculate the robust mean square error matrix $\text{MSE}_{\text{rob}} = \tilde{\boldsymbol{A}}^\top \tilde{\boldsymbol{A}}/\tilde{M}$ where $\tilde{\boldsymbol{A}}$ denotes the matrix $\boldsymbol{A}$ with the outlier rows removed and $\tilde{M}$ is the number of rows in $\tilde{\boldsymbol{A}}$. The quantity MSE_{rob} is a robust estimator of $\text{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]$, and $\det(1000 \times \text{MSE}_{\text{rob}})$ is the reported performance metric. We reported the determinant rather than the trace as it better accounts for unequal variances for and correlations between estimated parameters. Smaller (larger) values of $\det(\text{MSE}_{\text{rob}})$ indicate better (worse) performance of estimators in a square error loss sense.

4.2 Simulation results: Parameter recovery

Tables 1 and 2 present results for simulation settings I and II with $n_{\text{rep}} = 2$; results with $n_{\text{rep}} = 3$ and are presented in Section S4.2 of the Supplementary Materials and show similar conclusions. The true OLS estimator stands out with much smaller performance metrics than those of the estimators computed using data with measurement error; note that this true estimator is not available in practice since the true covariates $\mathbf{X}_i$'s are not observed. On the other hand, the naive OLS estimator has the worst performance in all the settings, highlighting the deleterious effect of ignoring measurement error as well as the importance of measurement error correction. Next, we compare all the correction estimators. For the $t_{2.5}$ and contaminated normal errors, all versions of the GMM estimator outperform the moment-corrected estimator, often with significant efficiency gains. For normally distributed errors, the GMM estimators perform competitively compared to moment-correction. That being said, in the normal error scenarios, GMM occasionally performs worse than moment-correction; see Setting II. When comparing GMM weighting schemes, there is no clear supe-

Table 1: Setting I performance of uncontaminated OLS (True), naive OLS (Naive), moment-corrected (MC), and GMM estimators with equal (Equal), minimax (MM) and quasi-likelihood (QL) weights with $n_{\text{rep}} = 2$ as measured by $\det(1000 \times \text{MSE}_{\text{rob}})$.

R	***U***	n	True	Naive	MC	GMM		
						Equal	MM	QL
$\rho = 0$	Normal	250	0.012	9.799	1.394	1.040	1.020	1.102
		500	0.001	2.889	0.240	0.224	0.222	0.234
		1000	0.000	0.668	0.026	0.025	0.027	0.030
	$t_{2.5}$	250	0.018	11.413	0.816	0.576	0.525	0.593
		500	0.001	5.128	0.134	0.100	0.075	0.077
		1000	0.000	1.502	0.021	0.013	0.011	0.010
	Cont.Normal	250	0.010	17.965	1.290	0.693	0.605	0.556
		500	0.001	4.305	0.149	0.105	0.077	0.090
		1000	0.000	1.143	0.022	0.014	0.009	0.011
$\rho = 0.5$	Normal	250	0.008	27.999	3.155	2.698	3.399	2.953
		500	0.001	7.365	0.355	0.362	0.378	0.373
		1000	0.000	1.339	0.037	0.036	0.038	0.039
	$t_{2.5}$	250	0.015	34.374	1.351	0.643	0.661	0.730
		500	0.001	8.390	0.211	0.128	0.112	0.111
		1000	0.000	2.583	0.026	0.017	0.016	0.016
	Cont.Normal	250	0.008	45.002	2.976	1.405	1.048	1.228
		500	0.001	10.191	0.287	0.193	0.120	0.157
		1000	0.000	2.331	0.034	0.024	0.015	0.021

rior weighting scheme, but minimax and quasi-likelihood weights perform better than equal weighting for non-normal errors.

Table 3 presents the results for setting III where the asymmetry condition C4 is violated, again for the case with $n_{\text{rep}} = 2$ replicates. The case with $n_{\text{rep}} = 3$ replicates is summarized in Section S4.2 in the Supplementary Materials. Similar to other settings, there is little difference in the performance between the moment-corrected estimator and the GMM estimator when the measurement

Table 2: Setting II performance of uncontaminated OLS (True), naive OLS (Naive), moment-corrected (MC), and GMM estimators with equal (Equal), minimax (MM) and quasi-likelihood (QL) weights with $n_{\text{rep}} = 2$ as measured by $\det(1000 \times \text{MSE}_{\text{rob}})$.

$\boldsymbol{R}$	$\boldsymbol{U}$	n	True	Naive	MC	GMM		
						Equal	MM	QL
$\rho = 0$	Normal	250	0.001	19.898	3.427	4.418	4.472	3.763
		500	0.000	2.053	0.145	0.161	0.203	0.157
		1000	0.000	0.121	0.003	0.004	0.004	0.004
	$t_{2.5}$	250	0.001	24.749	0.985	0.751	0.594	0.574
		500	0.000	2.359	0.087	0.055	0.052	0.045
		1000	0.000	0.139	0.001	0.001	0.001	0.001
	Cont.Normal	250	0.001	13.924	1.564	1.200	1.048	0.793
		500	0.000	0.985	0.043	0.042	0.038	0.033
		1000	0.000	0.063	0.002	0.002	0.001	0.001
$\rho = 0.5$	Normal	250	0.001	83.078	11.799	11.058	11.721	12.045
		500	0.000	5.037	0.266	0.288	0.332	0.292
		1000	0.000	0.286	0.007	0.008	0.009	0.009
	$t_{2.5}$	250	0.001	81.873	5.663	3.799	3.669	4.338
		500	0.000	4.953	0.130	0.104	0.097	0.086
		1000	0.000	0.329	0.003	0.003	0.003	0.002
	Cont.Normal	250	0.001	71.486	5.833	3.522	2.922	2.260
		500	0.000	3.010	0.150	0.126	0.105	0.088
		1000	0.000	0.230	0.004	0.004	0.003	0.002

error is Gaussian. However, we continue to note that GMM represents a substantive improvement over moment correction for contaminated normal and $t_{2.5}$ error distributions. Also, in the latter two error settings, minimax and quasi-likelihood weights outperform equal weighting; neither of the latter two weighting scheme is clearly preferred.

We also note that across settings I through III, it is observed that the MSE metric for the naive estimators also decreases as sample size increases. This is

Table 3: Setting III performance of uncontaminated OLS (True), naive OLS (Naive), moment-corrected (MC), and GMM estimators with equal (Equal), minimax (MM) and quasi-likelihood (QL) weights with $n_{\text{rep}} = 2$ as measured by $\det(1000 \times \text{MSE}_{\text{rob}})$.

$\boldsymbol{R}$	$\boldsymbol{U}$	n	True	Naive	MC	GMM		
						Equal	MM	QL
$\rho = 0$	Normal	250	0.001	10.963	1.625	1.623	1.615	1.623
		500	0.000	0.926	0.069	0.069	0.068	0.068
		1000	0.000	0.064	0.002	0.002	0.002	0.002
	$t_{2.5}$	250	0.001	11.607	0.944	0.936	0.855	0.890
		500	0.000	0.996	0.041	0.040	0.032	0.036
		1000	0.000	0.062	0.001	0.001	0.001	0.001
	Cont.Normal	250	0.001	12.622	1.665	1.663	1.467	1.572
		500	0.000	1.477	0.068	0.070	0.066	0.067
		1000	0.000	0.072	0.002	0.002	0.001	0.001
$\rho = 0.5$	Normal	250	0.001	54.053	7.288	7.162	7.214	7.338
		500	0.000	2.879	0.147	0.145	0.147	0.148
		1000	0.000	0.169	0.004	0.004	0.004	0.004
	$t_{2.5}$	250	0.001	41.193	2.852	2.782	2.374	2.524
		500	0.000	2.484	0.078	0.074	0.067	0.070
		1000	0.000	0.162	0.002	0.002	0.002	0.002
	Cont.Normal	250	0.001	65.616	5.556	5.552	4.945	5.375
		500	0.000	3.563	0.144	0.143	0.130	0.140
		1000	0.000	0.205	0.004	0.004	0.004	0.004

an artifact of the scaling used. The interested reader can easily verify that the MSE metric for the naive estimator decreases at a much slower rate than the same metric for the MC and GMM estimators.

In conclusion, the GMM estimator has a competitive performance compared to the moment-corrected estimator when the measurement errors are Gaussian. However, the relative decrease in $\det(\text{MSE}_{\text{rob}})$ is often so large for the heavier-tailed error distributions that some would be willing to risk a small loss in efficiency were the error distribution closer to a true normal. Furthermore, the two weighting schemes that account for heteroscedasticity tend to result in better estimators than the equal weighting.

4.3 Simulation results: Standard error estimation

We also performed a simulation study to examine the performance of the asymptotic covariance estimator in estimating the standard errors of the GMM estimators. The data were simulated from Settings I and III with the sample size $n \in \{500, 1000\}$, the measurement error $\mathbf{U}_{jk}$ following either a bivariate normal or bivariate t distribution with 2.5 degrees of freedom, and the correlation matrix $\boldsymbol{R}$ having $\rho = 0.5$. We reported the Monte Carlo estimates of the true standard errors obtained from 500 simulated pairs $(\hat{\boldsymbol{\beta}}_{\text{gmm}}, \hat{\boldsymbol{\gamma}}_{\text{gmm}})$ obtained from independently generated datasets, as well as the average bootstrap plug-in standard errors as defined in Section 3.3 for the GMM estimators with the minimax weighting scheme $\hat{q}_j^{\text{mm}}$ defined in (3.10), while the results for other weighting schemes are similar and hence are omitted. Table 4 presents the results for $n_{\text{rep}} = 2$, while the results for $n_{\text{rep}} = 3$ are presented in Section S4.2 the Supplementary Materials. The average bootstrap plug-in standard errors are similar to the Monte Carlo standard error in all the considered settings, suggesting that the bootstrap procedure in Section 3.3 provides a reliable estimator for the standard errors of the proposed GMM estimators.

Table 4: Monte Carlo standard errors (MC-SE) and average of the bootstrap plug-in standard errors (Avg-SE) for the GMM estimators with the minimax weighting scheme in simulation settings I and III with $n_{\text{rep}} = 2$ replicates and $\rho = 0.5$.

Setting	**U**	Coeff	$n = 500$		$n = 1000$	
			MC-SE	Avg-SE	MC-SE	Avg-SE
I	Normal	$\hat{\beta}_1$	0.031	0.030	0.025	0.022
		$\hat{\beta}_2$	0.031	0.029	0.023	0.021
		$\hat{\gamma}_{00}$	0.043	0.044	0.033	0.032
	$t_{2.5}$	$\hat{\beta}_1$	0.029	0.026	0.021	0.018
		$\hat{\beta}_2$	0.029	0.026	0.019	0.019
		$\hat{\gamma}_{00}$	0.038	0.036	0.028	0.025
III	Normal	$\hat{\beta}_1$	0.037	0.033	0.021	0.024
		$\hat{\beta}_2$	0.031	0.031	0.024	0.022
		$\hat{\gamma}_{00}$	0.055	0.050	0.038	0.036
		$\hat{\gamma}_1$	0.028	0.027	0.019	0.020
		$\hat{\gamma}_2$	0.029	0.028	0.019	0.019
	$t_{2.5}$	$\hat{\beta}_1$	0.030	0.028	0.018	0.021
		$\hat{\beta}_2$	0.029	0.028	0.019	0.021
		$\hat{\gamma}_{00}$	0.042	0.043	0.029	0.032
		$\hat{\gamma}_1$	0.024	0.026	0.019	0.018
		$\hat{\gamma}_2$	0.025	0.024	0.016	0.018

5. Analysis of NHANES data

The National Health and Nutrition Examination Survey (NHANES) is a long-running research survey conducted by the National Center for Health Statistics (NCHS). The goal of this longitudinal survey study is to assess the health and nutritional status of both adults and children in the United States, tracking the evolution of this status over time. During the 2009-2010 survey period, participants were interviewed and asked to provide their demographic background as well as information about nutrition habits. Participants also undertook a series of health examinations. To assess the nutritional habits of participants, dietary data were collected using two 24-hour recall interviews wherein the participants self-reported the consumed amount for a series of food items during the 24 hours prior to each interview. Based on these recalls, daily aggregated consumption of water, food energy, and other nutrition components such as total fat and total sugar consumption were computed. We used the 2009-2010 NHANES dietary data to illustrate our new GMM estimation procedure.

In this illustrative analysis, we considered the relationship between participants' BMI (outcome of interest) and their age as well as daily aggregates of energy, protein and fat consumption. As these nutritional variables were calculated based on self-reported data, they are well-known to be subject to measurement error. We restricted our analysis to $n = 1595$ white women and treat the nutritional data from each of the interviews as $n_{\text{rep}} = 2$ independently observed replicates. We fitted the multiple linear EIV model to this data, treating the nutritional quantities energy, protein, and fat consumption as error-prone covariates with age considered an error-free covariate. Furthermore, all of these

Table 5: NHANES study estimated coefficients for naive, moment-corrected (MC) and GMM estimators with equal (Equal), minimax (MM), and quasi-likelihood weights (QL) weights.

	Naive	MC	GMM		
			Equal	MM	QL
Intercept	26.39 (0.18)	26.39 (0.18)	26.63 (0.18)	26.59 (0.18)	26.56 (0.18)
Energy	-0.86 (0.47)	-2.3 (1.05)	-2.7 (0.78)	-2.59 (0.85)	-2.49 (0.83)
Protein	1.29 (0.34)	3.35 (1.03)	3.56 (0.43)	3.68 (0.49)	3.54 (0.47)
Fat	0.54 (0.43)	1.03 (0.99)	0.88 (0.57)	0.89 (0.64)	0.89 (0.59)
Age	3.92 (0.18)	3.71 (0.19)	3.82 (0.16)	3.78 (0.16)	3.79 (0.16)

covariates were standardized in the analysis. We computed the naive, moment-corrected, and GMM estimators, the latter with the three different weighting schemes. We further reported the estimated standard errors for these estimators: The estimated standard errors for the naive estimators were obtained from the linear regression model of the outcome on all the observed covariates, while the standard errors for the moment corrected estimators were calculated from the robust estimator given in Buonaccorsi (2010, Section 5.4). The estimated standard errors for the GMM estimators were calculated in the same manner as in our simulation using the asymptotic covariance and the bootstrap procedure.

Table 5 demonstrates that the consequences of ignoring measurement error are apparent – the naive estimator exhibits dramatic attenuation for all the nutritional variables. On the other hand, the moment-corrected and GMM approaches all appear to correct for the bias as the estimated coefficients for energy, protein and fat intake are much larger in absolute terms. Comparing the different corrected estimators, we note the magnitude of the coefficients corresponding to energy and protein are smaller for moment correction than for GMM. The reverse holds for the coefficient corresponding to fat consumption. The effect of age on BMI is similar across all the estimators, which can be explained by the low correlation between age and each of the nutritional variables (the correlations between age and the averaged energy, protein, and fat intakes are -0.02, 0.10, and 0.03, respectively). Finally, the GMM estimators tend to have lower estimated standard errors than the moment-corrected estimators. This possibly reflects information contributions from the phase function beyond the first two moments.

6. Conclusion

In this paper, we have explored distribution-free solutions for parameter estimation in the linear errors-in-variables model with heteroscedastic measurement errors across cases. The newly proposed solution combines the popular moment-corrected estimator with a phase function-based estimator using a GMM framework. On the one hand, the proposed GMM estimator inherits estimating equations from the moment-corrected estimator and in this way is able to relax a strict asymmetry condition imposed on the covariates as in the phase function-based estimator proposed by Nghiem et al. (2020). On the other hand, the proposed GMM estimator inherits estimating equations from the phase function-based estimator that leverages the skewness of the true covariates if present, and introduces observation-specific weighting to account for the measurement error heteroscedasticity. Our simulation studies show that when the measurement errors are normal, the GMM estimator is competitive with the moment-corrected estimator. Nevertheless, when measurement errors are non-normal, the GMM estimator has superior performance across all simulation settings considered, including ones where some covariates are symmetrically distributed.

We close by noting some future research that could be explored relating to this problem. Firstly, in the multivariate setting, the estimation of observation-level measurement error covariance matrices is a challenge. Some form of covariance regularization could be applied to further improve the estimation efficiency of the regression parameters. Secondly, the idea of a continuously-updating GMM treating the matrix $\boldsymbol{\Omega}_S$ as an explicit function of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ can be explored. This could further improve estimator performance, but does increase the computational complexity of the problem. Finally, the development of tools for exploring and quantifying the skewness of the true covariates subject to symmetric measurement error will help practitioners understand when the GMM approach is beneficial.

Supplementary Materials

The Supplementary Materials contain the proofs of Lemma 1 and Theorem 1 in Sections S1 and S2 respectively. Detailed calculations for the quasi-likelihood weights are in Section S3, and additional simulation results are in Section S4.

References

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*. CRC press.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*, 85(411):652–663.

Delaigle, A. and Hall, P. (2016). Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 231–252.

Erickson, T., Jiang, C. H., and Whited, T. M. (2014). Minimum distance estimation of the errors-in-variables model using linear cumulant equations. *Journal of Econometrics*, 183(2):211–221.

Gillard, J. (2014). Method of moments estimation in linear regression with errors in both variables. *Communications in Statistics-Theory and Methods*, 43(15):3208–3222.

Hanfelt, J. J. and Liang, K.-Y. (1997). Approximate likelihoods for generalized linear errors-in-variables models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):627–637.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280.

Higdon, R. and Schafer, D. W. (2001). Maximum likelihood computations for regression with measurement error. *Computational Statistics & Data Analysis*, 35(3):283–299.

Hu, F. and Kalbfleisch, J. D. (1997). Estimating equations and the bootstrap. *Lecture Notes-Monograph Series*, pages 405–416.

Newey, W. K. and McFadden, D. (1994). Chapter 36: Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

Nghiem, L. and Potgieter, C. J. (2018). Density estimation in the presence of heteroscedastic measurement error of unknown type using phase function deconvolution. *Statistics in Medicine*, 37(25):3679–3692.

Nghiem, L. H., Byrd, M. C., and Potgieter, C. J. (2020). Estimation in linear errors-in-variables models with unknown error distribution. *Biometrika*, 107(4):841–856.

Reiersøl, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica: Journal of the Econometric Society*, pages 1–24.

Rousseeuw, P. J. and Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223.

Song, W. (2021). Nonparametric inference methods for berkson errors. In *Handbook of Measurement Error Models*, pages 271–292. Chapman and Hall/CRC.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, 74(4):703–716.

Stefanski, L. A. and Cook, J. R. (1995). Simulation-extrapolation: the measurement error jackknife. *Journal of the American Statistical Association*, 90(432):1247–1256.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, 11(3):284–300.

Xue-Kun Song, P. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320.