Supplementary Materials:
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

S1 Proof of Lemma 1

In this section, we provide proof of Lemma 1 from Section 3.1 in the main paper. For clarity, the lemma is restated here. To this end, let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ and recall that $\mathbf{S}(\boldsymbol{\theta}) = [\mathbf{S}_L^\top(\boldsymbol{\theta}), \mathbf{S}_{\tilde{D}}^\top(\boldsymbol{\theta})]^\top$ denotes the vector of gradient equations with $\mathbf{S}_L(\boldsymbol{\theta})$ the gradient vector of the corrected L_2 norm $L(\boldsymbol{\theta})$ as defined in Section 2.2, and with $\mathbf{S}_{\tilde{D}}(\boldsymbol{\theta})$ the gradient of the phase function-based statistic $\tilde{D}(\boldsymbol{\theta})$ as defined in Section 2.3 of the main paper. Let $\mathbf{S}_0(\boldsymbol{\theta}) = \lim_{n\to\infty} \mathrm{E}[\mathbf{S}(\boldsymbol{\theta})]$ denote the limiting expectation of the gradient equations. Subsequently, define

$$Q_0(\boldsymbol{\theta}) = \mathbf{S}_0^\top(\boldsymbol{\theta})\boldsymbol{\Omega}_S^{-1}\mathbf{S}_0(\boldsymbol{\theta}).$$

Lemma 1 now follows.

Lemma 1. *Assume that all variables in the model have at least two finite moments. For $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+q+1}$, the function $Q(\boldsymbol{\theta}) \xrightarrow{p} Q_0(\boldsymbol{\theta})$ uniformly.*

The proof of Lemma 1 relies on establishing the conditions established in Lemma 2.9 of Newey and McFadden (1994). Specifically, the proof first shows that function Q converges in probability to Q_0 for a fixed value of $\boldsymbol{\theta}$ and is continuously differentiable for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, it establishes a Lipschitz condition that bounds the difference between Q at two arbitrary parameter values. This Lipschitz condition is crucial to show that Q converges uniformly in probability to Q_0. Finally, the proof shows that $\mathbf{S}(\boldsymbol{\theta})$ and $\nabla\mathbf{S}(\boldsymbol{\theta}) = \partial\mathbf{S}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ converge uniformly to their limiting expected values. Thus, the result from Newey & McFadden applies and the required uniform convergence of Q to Q_0 follows.

Proof of Lemma 1. Consider the function $Q(\boldsymbol{\theta}) = \mathbf{S}^\top(\boldsymbol{\theta})\boldsymbol{\Omega}_S^{-1}\mathbf{S}(\boldsymbol{\theta})$ and a fixed value $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, $Q(\boldsymbol{\theta}) \xrightarrow{P} Q_0(\boldsymbol{\theta})$ by Slutsky's theorem and the continuous mapping theorem. Moreover, $Q(\boldsymbol{\theta})$ is continuously differentiable for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Thus, by the mean value theorem, for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, we have

$$Q(\boldsymbol{\theta}_1) - Q(\boldsymbol{\theta}_2) = \nabla Q(\boldsymbol{\theta}_z)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2),$$

where $\nabla Q(\boldsymbol{\theta}) = \partial Q/\partial\boldsymbol{\theta}$ and $\boldsymbol{\theta}_z$ is a linear interpolant of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Applying the Cauchy-Schwarz inequality, we obtain

$$|Q(\boldsymbol{\theta}_1) - Q(\boldsymbol{\theta}_2)| \leq \|\nabla Q(\boldsymbol{\theta}_z)\|\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

Subsequently, uniform convergence in probability will follow by by Lemma 2.9 of Newey and McFadden (1994) if we can establish the the Lipschitz condition that for some constants $\alpha > 0$ and a sequence $B_n = O_p(1)$ such that for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, we have

$$|Q(\boldsymbol{\theta}_1) - Q(\boldsymbol{\theta}_2)| \leq B_n\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\alpha. \tag{S1.1}$$

Observe that equation (S1.1) will hold for $\alpha = 1$ if $\|\nabla Q(\boldsymbol{\theta}_z)\| = O_p(1)$. Now, by definition $\nabla Q(\boldsymbol{\theta}) = 2[\nabla\mathbf{S}(\boldsymbol{\theta})]^\top\boldsymbol{\Omega}_S^{-1}\mathbf{S}(\boldsymbol{\theta})$ with $\nabla\mathbf{S}(\boldsymbol{\theta}) = \partial\mathbf{S}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ a $(p+q+1)\times(p+q+1)$ matrix of partial derivatives. By another application of the Cauchy-Schwarz inequality, we obtain

$$\|\nabla Q(\boldsymbol{\theta}_z)\| \leq c_S \|\nabla \mathbf{S}(\boldsymbol{\theta}_z)\| \|\mathbf{S}(\boldsymbol{\theta}_z)\| \leq c_S \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\nabla \mathbf{S}(\boldsymbol{\theta})\| \cdot \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\mathbf{S}(\boldsymbol{\theta})\|$$

where c_S is a constant depending only on $\boldsymbol{\Omega}_S$ and not on the arguments $\boldsymbol{\theta}$. Thus, it remains only to be shown that $\mathbf{S}(\boldsymbol{\theta})$ and $\nabla \mathbf{S}(\boldsymbol{\theta})$ converge uniformly to $\mathrm{E}[\mathbf{S}(\boldsymbol{\theta})]$ and $\mathrm{E}[\nabla \mathbf{S}(\boldsymbol{\theta})]$.

To this end, consider the components $\mathbf{S}_L(\boldsymbol{\theta})$ and $\nabla \mathbf{S}_L(\boldsymbol{\theta})$. These functions are continuous for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Furthermore, provided the random variables $(\mathbf{X}_j, \mathbf{Z}_j)$, $\mathbf{U}_j$, and ε_j used to define Q have finite variances, there exists dominating functions $d_{L,1}(\boldsymbol{\theta})$ and $d_{L,2}(\boldsymbol{\theta})$ such that $\|\mathbf{S}_L(\boldsymbol{\theta})\| \leq d_{L,1}(\boldsymbol{\theta})$ and $\|\nabla \mathbf{S}_L(\boldsymbol{\theta})\| \leq d_{L,2}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ by a uniform law of large numbers as in Hoadley (1971) or Pötscher and Prucha (1989).

Next, consider the component $\mathbf{S}_{\tilde{D}}(\boldsymbol{\theta})$ and $\nabla \mathbf{S}_{\tilde{D}}(\boldsymbol{\theta})$. Again, these functions are continuous for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. While slightly tedious, one can also verify that a dominating function $d_{\tilde{D},1}(\boldsymbol{\theta})$ exists for $\mathbf{S}_{\tilde{D}}(\boldsymbol{\theta})$ provided $(\mathbf{X}_j, \mathbf{Z}_j)$ and $\mathbf{U}_j$ have finite first moments. Similarly, a dominating function $d_{\tilde{D},2}(\boldsymbol{\theta})$ exists for $\nabla \mathbf{S}_{\tilde{D}}(\boldsymbol{\theta})$ provided $(\mathbf{X}_j, \mathbf{Z}_j)$ and $\mathbf{U}_j$ have finite second moments. Again, by the same uniform law of large numbers, uniform convergence is achieved. Consequently, we have $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\mathbf{S}(\boldsymbol{\theta})\| = O_p(1)$ and $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\nabla \mathbf{S}(\boldsymbol{\theta})\| = O_p(1)$, the conditions of Newey and McFadden (1994) are satisfied, and the required uniform convergence of $Q(\boldsymbol{\theta})$ to $Q_0(\boldsymbol{\theta})$ follows. □

S2 Proof of Theorem 1

This section presents the proof of Theorem 1 from Section 3.1 in the main paper. For clarity, the theorem is restated here.

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

As $Q_0(\boldsymbol{\theta})$ is a positive definite quadratic form in terms of $\mathbf{S}_0(\boldsymbol{\theta})$, the global minimum occurs at a point $\boldsymbol{\theta}^*$ if and only if $\mathbf{S}_0(\boldsymbol{\theta}^*) = \mathbf{0}$ for a unique value of $\boldsymbol{\theta}^*$. The proof of Theorem 1 thus relies on showing that $Q_0(\boldsymbol{\theta})$, the uniform-in-probability limit of $Q(\boldsymbol{\theta})$, has a unique global minimum at the true parameter values $\boldsymbol{\theta}_0$ due to $\mathbf{S}_0(\boldsymbol{\theta}) = \mathbf{0}$ only when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Throughout the proof, for any random variable V, we let $\phi_V(t)$ denote its characteristic function. For any complex number z, let $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ be its real and imaginary parts, respectively.

Proof of Theorem 1. By Lemma 1, the GMM objective function $Q(\boldsymbol{\theta})$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$. For consistency of the estimators obtained by minimizing $Q(\boldsymbol{\theta})$, it suffices to establish that limiting function $Q_0(\boldsymbol{\theta})$ has a unique global minimum at the true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$. This proof will separately consider the L_2 norm and phase function components contributing to $Q(\boldsymbol{\theta})$. Particularly, we will prove the two following statements:

Statement 1: The first $p+q+1$ elements of $\boldsymbol{S}_0(\boldsymbol{\theta})$, which correspond to the estimating equations from the corrected L_2 norm, has a *unique* solution at $\boldsymbol{\theta}_0$.

Statement 2: $\boldsymbol{\theta}_0$ is always a solution to the last $p+q+1$ elements of $\boldsymbol{S}_0(\boldsymbol{\theta})$, which correspond to the estimating equations from the phase function distance $\tilde{D}$.

Given the two previous statements, Theorem 1 will follow immediately. Indeed, these two statements imply that $\boldsymbol{\theta}_0$ is the unique solution for $\boldsymbol{S}_0(\boldsymbol{\theta}) = 0$ as a whole. This also establishes that $Q_0(\boldsymbol{\theta}_0) = 0$, meaning $Q_0(\boldsymbol{\theta})$ has a unique global minimum of zero at $\boldsymbol{\theta}_0$. As a result, the estimator $\hat{\boldsymbol{\theta}}$ that minimizes $Q(\boldsymbol{\theta})$ is consistent for $\boldsymbol{\theta}_0$.

Hence it remains to prove the two statements, and we will do it in two separate subsections.

S2.1 Proof of Statement 1

Consider first the corrected L_2 norm function with estimating equations $\mathbf{S}_L(\boldsymbol{\theta}) = [\mathbf{S}_{L,\boldsymbol{\beta}}(\boldsymbol{\theta})^\top, \mathbf{S}_{L,\boldsymbol{\gamma}}(\boldsymbol{\theta})^\top]^\top$ as defined in equation (2.3) of the main paper with

$$
\begin{aligned}
\mathbf{S}_{L,\boldsymbol{\beta}}(\boldsymbol{\theta}) &= -\frac{2}{n}\sum_{j=1}^{n} \mathbf{W}_j\big(y_j - \mathbf{W}_j^\top\boldsymbol{\beta} - \mathbf{Z}_j^\top\boldsymbol{\gamma}\big) - \frac{2}{n}\sum_{j=1}^{n}\frac{1}{n_j}\boldsymbol{\Sigma}_j\boldsymbol{\beta},\\
\mathbf{S}_{L,\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= -\frac{2}{n}\sum_{j=1}^{n} \mathbf{Z}_j\big(y_j - \mathbf{W}_j^\top\boldsymbol{\beta} - \mathbf{Z}_j^\top\boldsymbol{\gamma}\big).
\end{aligned}
$$

The expected values $\mathrm{E}[\mathbf{S}_{L,\boldsymbol{\beta}}(\boldsymbol{\theta})]$ and $\mathrm{E}[\mathbf{S}_{L,\boldsymbol{\gamma}}(\boldsymbol{\theta})]$ are found by evaluating the conditional expectations of the component summands. For $\mathbf{S}_{L,\boldsymbol{\beta}}(\boldsymbol{\theta})$, we have

$$\begin{aligned}
&\mathrm{E}\Big[\mathbf{W}_j\big(y_j - \mathbf{W}_j^\top\boldsymbol{\beta} - \mathbf{Z}_j^\top\gamma\big)\Big|\mathbf{X}_j,\mathbf{Z}_j\Big] \\
&= \mathrm{E}\Big[\big(\mathbf{X}_j + \mathbf{U}_j\big)\big(\mathbf{X}_j^\top\boldsymbol{\beta}_0 + \mathbf{Z}_j^\top\boldsymbol{\gamma}_0 + \varepsilon_j - \mathbf{X}_j^\top\boldsymbol{\beta} - \mathbf{Z}_j^\top\boldsymbol{\gamma} - \mathbf{U}_j^\top\boldsymbol{\beta}\big)\Big|\mathbf{X}_j,\mathbf{Z}_j\Big] \\
&\overset{(i)}{=} \mathbf{X}_j\mathbf{X}_j^\top\big(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\big) + \mathbf{X}_j\mathbf{Z}_j^\top\big(\gamma_0 - \boldsymbol{\gamma}\big) - n_j^{-1}\boldsymbol{\Sigma}_j\boldsymbol{\beta},
\end{aligned}$$

where step (i) follows from the independence of $\mathbf{U}_j$ and $(\mathbf{X}_j,\mathbf{Z}_j)$, as well as noting that $\mathrm{E}\big[\mathbf{U}_j\mathbf{U}_j^\top\big] = n_j^{-1}\boldsymbol{\Sigma}_j$. Similarly, for $\mathbf{S}_{L,\boldsymbol{\gamma}}(\boldsymbol{\theta})$, we have

$$\mathrm{E}\Big[\mathbf{Z}_j\big(y_j - \mathbf{W}_j^\top\boldsymbol{\beta} - \mathbf{Z}_j^\top\gamma\big)\Big|\mathbf{X}_j,\mathbf{Z}_j\Big] = \mathbf{Z}_j\mathbf{X}_j^\top\big(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\big) + \mathbf{Z}_j\mathbf{Z}_j^\top\big(\gamma_0 - \gamma\big).$$

Letting $(\mathbf{X},\mathbf{Z})$ denote an independent copy of $(\mathbf{X}_j,\mathbf{Z}_j)$, we have

$$\begin{aligned}
\mathrm{E}[\mathbf{S}_{L,\boldsymbol{\beta}}(\boldsymbol{\theta})] &= -2\,\mathrm{E}\big[\mathbf{X}\mathbf{X}^\top\big]\big(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\big) - 2\,\mathrm{E}\big[\mathbf{X}\mathbf{Z}^\top\big]\big(\gamma_0 - \gamma\big) \\
\mathrm{E}[\mathbf{S}_{L,\boldsymbol{\gamma}}(\boldsymbol{\theta})] &= -2\,\mathrm{E}\big[\mathbf{Z}\mathbf{X}^\top\big]\big(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\big) - 2\,\mathrm{E}\big[\mathbf{Z}\mathbf{Z}^\top\big]\big(\gamma_0 - \gamma\big).
\end{aligned}$$

As a consequence of Lemma 1, $\mathbf{S}_{L,\boldsymbol{\beta}}(\boldsymbol{\theta})$ and $\mathbf{S}_{L,\boldsymbol{\gamma}}(\boldsymbol{\theta})$ converge uniformly to $\mathrm{E}[\mathbf{S}_{L,\boldsymbol{\beta}}(\boldsymbol{\theta})]$ and $\mathrm{E}[\mathbf{S}_{L,\boldsymbol{\gamma}}(\boldsymbol{\theta})]$, the first $p+q+1$ components of $\mathbf{S}_0(\boldsymbol{\theta})$. It is straightforward to see that the corresponding system of equations

$$\mathrm{E}[\mathbf{S}_{L,\boldsymbol{\beta}}(\boldsymbol{\theta})] = \mathbf{0} \qquad \text{and} \qquad \mathrm{E}[\mathbf{S}_{L,\boldsymbol{\gamma}}(\boldsymbol{\theta})] = \mathbf{0}$$

has a unique solution at $\boldsymbol{\theta} = \boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \gamma_0)$.

S2.2 Proof of Statement 2

Consider next the phase function-based criterion $\tilde{D}(\boldsymbol{\theta})$ directly. From Lemma 1, we have the uniform convergence of $\tilde{D}(\boldsymbol{\theta})$ to a limiting function $D_0(\boldsymbol{\theta})$.

We will now evaluate this limiting function. Recall that

$$\tilde{D}(\boldsymbol{\theta}) = \int_0^{t^*} \left(C_y(t) \left[\sum_{j=1}^{n} q_j \sin\left\{ t \left(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma} \right) \right\} \right] \right.$$
$$\left. - S_y(t) \left[\sum_{j=1}^{n} q_j \cos\left\{ t \left(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma} \right) \right\} \right] \right)^2 K_{t^*}(t) dt. \quad \text{(S2.2)}$$

Let $V_0 = \mathbf{X}^\top \boldsymbol{\beta}_0 + \mathbf{Z}^\top \boldsymbol{\gamma}_0$ and $Y = V_0 + \varepsilon$. For arbitrary t, by the weak law of large numbers,

$$C_y(t) = \frac{1}{n} \sum_{j=1}^{n} \cos(y_j t) \xrightarrow{p} \mathrm{E}\left[\cos(Yt)\right] = \mathrm{Re}[\phi_Y(t)] = \mathrm{Re}[\phi_{V_0}(t)] \phi_\varepsilon(t),$$

where the last equality follows upon noting that ε has a real-valued characteristic function. Similarly, $S_y(t) \xrightarrow{p} \mathrm{Im}[\phi_{V_0}(t)]\phi_\varepsilon(t)$. Furthermore, noting that for any $\boldsymbol{\theta}$, the random variables $\sin\left[t(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma})\right]$ and $\cos\left[t(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma})\right]$ are bounded, and subsequently have finite variances. By a generalized weak law of large numbers,

$$\sum_{j=1}^{n} q_j \sin\left\{ t \left(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma} \right) \right\} \xrightarrow{p} \mathrm{E}\left[\sum_{j=1}^{n} q_j \sin\left\{ t \left(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma} \right) \right\} \right].$$

Letting $V_j(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{X}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma}$, we have

$$\mathrm{E}\left[\sin\left\{ t \left(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma} \right) \right\} \right] = \mathrm{Im}\left\{ \phi_{V_j(\boldsymbol{\beta},\boldsymbol{\gamma})}(t) \right\} \phi_{\mathbf{U}_j^\top \boldsymbol{\beta}}(t),$$

where we make use of the fact that $\mathbf{U}_j^\top \boldsymbol{\beta}$ has a symmetric distribution about zero and hence a real-valued characteristic function. Since $(\mathbf{X}_j, \mathbf{Z}_j)$ are *iid* by Condition C1, the random variables $V_j(\boldsymbol{\beta}, \boldsymbol{\gamma}),\ j = 1, \ldots, n$ are also *iid* and have a common characteristic function $\phi_{V(\boldsymbol{\beta},\boldsymbol{\gamma})}(t) = \phi_{V_j(\boldsymbol{\beta},\boldsymbol{\gamma})}(t)$ for $j = 1, \ldots, n$. As a consequence, we have

$$\mathrm{E}\left[\sum_{j=1}^{n} q_j \sin\left\{ t \left(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma} \right) \right\} \right] = \mathrm{Im}\left\{ \phi_{V(\boldsymbol{\beta},\boldsymbol{\gamma})}(t) \right\} \left(\sum_{j=1}^{n} q_j \phi_{\mathbf{U}_j^\top \boldsymbol{\beta}}(t) \right).$$

Similarly,

$$\mathrm{E}\left[\sum_{j=1}^{n} q_j \cos\left\{t\left(\mathbf{W}_j^\top \boldsymbol{\beta} + \mathbf{Z}_j^\top \boldsymbol{\gamma}\right)\right\}\right] = \mathrm{Re}\left\{\phi_{V(\boldsymbol{\beta},\boldsymbol{\gamma})}(t)\right\}\left(\sum_{j=1}^{n} q_j \phi_{\mathbf{U}_j^\top \boldsymbol{\beta}}(t)\right).$$

Letting $h(t,\boldsymbol{\beta}) = \phi_\varepsilon(t)^2\left[\sum_{j=1}^{n} q_j \phi_{\mathbf{U}_j^\top \boldsymbol{\beta}}(t)\right]^2$, and recalling the established uniform convergence for $(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \boldsymbol{\Theta}$, the statistic $\tilde{D}(\boldsymbol{\theta})$ converges uniformly to

$$\begin{aligned} D_0(\boldsymbol{\theta}) = \int_0^{t^*} & \Big[\mathrm{Re}\left\{\phi_{V_0}(t)\right\}\mathrm{Im}\left\{\phi_{V(\boldsymbol{\beta},\boldsymbol{\gamma})}(t)\right\} \\ & - \mathrm{Im}\left\{\phi_{V_0}(t)\right\}\mathrm{Re}\left\{\phi_{V(\boldsymbol{\beta},\boldsymbol{\gamma})}(t)\right\}\Big]^2 h(t,\boldsymbol{\beta}) K_{t^*}(t)\,dt \\ = & \int_0^{t^*}\left[\mathrm{Im}\{\phi_{V_0 - V(\boldsymbol{\beta},\boldsymbol{\gamma})}(t)\}\right]^2 h(t,\boldsymbol{\beta}) K_{t^*}\,dt, \end{aligned}$$

where random variables V_0 and $V(\boldsymbol{\beta},\boldsymbol{\gamma})$ are independent. Note that

$$\mathrm{Im}\left\{\phi_{V_0 - V(\boldsymbol{\beta},\boldsymbol{\gamma})}(t)\right\} = 0 \text{ for all } t \in \mathbb{R}$$

if and only if the distribution of $V_0 - V(\boldsymbol{\beta},\boldsymbol{\gamma})$ is symmetric about 0. When Condition C4 holds, this is only true for $\boldsymbol{\theta} = \boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$. On the other hand, when Condition C4 does not hold, $D_0(\boldsymbol{\theta})$ may have infinitely many global minima, but one of those minima still occurs at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

We conclude by noting that $\tilde{D}(\boldsymbol{\theta})$ is continuous, as is the gradient vector $\nabla\tilde{D}(\boldsymbol{\theta}) = \partial\tilde{D}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. From Lemma 1, it subsequently follows that $\nabla\tilde{D}(\boldsymbol{\theta})$ also converges uniformly to $\nabla D_0(\boldsymbol{\theta})$. Thus, $\nabla D_0(\boldsymbol{\theta}_0) = \mathbf{0}$, even though this is not a unique solution to this system of equations. Note also that $\nabla D_0(\boldsymbol{\theta}_0)$ represents the last $p+q+1$ elements of $\boldsymbol{S}_0(\boldsymbol{\theta})$. The proof is now complete.

S3 Calculating the Quasi-Likelihood Weights

The quasi-likelihood weights are defined in Section 3.2 of the main paper to be the minimizer of the L_2 discrepancy

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

where $\hat{\boldsymbol{\mu}}_q = \sum_{j=1}^{n} q_j \mathbf{W}_j$, subject to $q_j \geq 0$, $j = 1, \ldots, n$ and $\sum_{j=1}^{n} q_j = 1$. To find the minimizer, let $\boldsymbol{\Omega}_j = \boldsymbol{\Sigma}_x + n_j^{-1}\boldsymbol{\Sigma}_j$ for $j = 1, \ldots, n$. Some algebraic manipulation gives

$$\begin{aligned} L(\boldsymbol{q}) = &\sum_{j=1}^{n} \mathbf{W}_j^\top \boldsymbol{\Omega}_j^{-1} \mathbf{W}_j - \hat{\boldsymbol{\mu}}_q^\top \left(\sum_{j=1}^{n} \boldsymbol{\Omega}_j^{-1}\mathbf{W}_j\right) \\ &- \left(\sum_{j=1}^{n} \mathbf{W}_j^\top \boldsymbol{\Omega}_j^{-1}\right) \hat{\boldsymbol{\mu}}_q + \hat{\boldsymbol{\mu}}_q^\top \left(\sum_{j=1}^{n} \boldsymbol{\Omega}_j^{-1}\right) \hat{\boldsymbol{\mu}}_q. \end{aligned}$$

Note that $L(\boldsymbol{q})$ is a function of $\boldsymbol{q}$ only through $\hat{\boldsymbol{\mu}}_q$. To calculate the weights, we define the function $D(\boldsymbol{q})$ to be only the terms in $L(\boldsymbol{q})$ involving $\hat{\boldsymbol{\mu}}_q$, and also introducing two Lagrange-multiplier type terms. The first of these ensures the weights q_j sum to 1, while the second ensures a numerically stable solution by constraining the squared differences between the q_j. The resulting function to be minimized is

$$\begin{aligned} D(\boldsymbol{q}, \lambda) \;=\; & -\hat{\boldsymbol{\mu}}_q^\top \left(\sum_{j=1}^{n} \boldsymbol{\Omega}_j^{-1}\mathbf{W}_j\right) - \left(\sum_{j=1}^{n} \mathbf{W}_j^\top \boldsymbol{\Omega}_j^{-1}\right) \hat{\boldsymbol{\mu}}_q + \hat{\boldsymbol{\mu}}_q^\top \left(\sum_{j=1}^{n} \boldsymbol{\Omega}_j^{-1}\right) \hat{\boldsymbol{\mu}}_q \\ & + 2\lambda \left(\sum_{j=1}^{n} q_j - 1\right) + \gamma \sum_{j,k} (q_j - q_k)^2. \end{aligned}$$

This function is minimized over $(\boldsymbol{q}, \lambda)$, while γ is a user-specified constant ensuring a numerically stable solution. Now, defining

$$\mathbf{A}_1 = \sum_{j=1}^{n} \boldsymbol{\Omega}_j^{-1}\mathbf{W}_j \quad \text{and} \quad \mathbf{A}_2 = \sum_{j=1}^{n} \boldsymbol{\Omega}_j^{-1}$$

the target function can be written in the convenient form

$$D(\boldsymbol{q}, \lambda) = -2\hat{\boldsymbol{\mu}}_q^\top \mathbf{A}_1 + \hat{\boldsymbol{\mu}}_q^\top \mathbf{A}_2 \hat{\boldsymbol{\mu}}_q + 2\lambda \left(\sum_{j=1}^{n} q_j - 1 \right) + \gamma \sum_{i,j} (q_i - q_j)^2.$$

This function is quadratic in $(\boldsymbol{q}, \lambda)$ with a global minimum that can be found by solving a linear system of equations. Taking partial derivatives of with respect to the q_k, for $k = 1, \ldots, n$, we have estimating equations

$$\frac{\partial D}{\partial q_k} = -2\mathbf{V}_k^\top \mathbf{A}_1 + 2\mathbf{V}_k^\top \mathbf{A}_2 \hat{\boldsymbol{\mu}}_{\boldsymbol{q}} + 2\lambda + 2\gamma \sum_{j \neq k} (q_k - q_j) = 0.$$

Here, $\mathbf{V}_k$ denotes a $n \times 1$ column vector of zeros, with the kth entry equal to 1. Writing these estimating equations explicitly in terms of the weights q gives

$$-\mathbf{V}_k^\top \mathbf{A}_1 + \sum_{j=1}^{n} q_j \left(\mathbf{V}_k^\top \mathbf{A}_2 \mathbf{V}_j \right) + \lambda + \gamma \sum_{j \neq k} (q_k - q_j) = 0, \; k = 1, \ldots, n.$$

The minimizer of $D(\boldsymbol{q}, \lambda)$ is found by solving the linear system

$$\begin{pmatrix} \mathbf{V}_1^\top \mathbf{A}_2 \mathbf{V}_1 + (n-1)\gamma & \mathbf{V}_1^\top \mathbf{A}_2 \mathbf{V}_2 - \gamma & \cdots & \mathbf{V}_1^\top \mathbf{A}_2 \mathbf{V}_n - \gamma & 1 \\ \mathbf{V}_2^\top \mathbf{A}_2 \mathbf{V}_1 - \gamma & \mathbf{V}_2^\top \mathbf{A}_2 \mathbf{V}_2 + (n-1)\gamma & \cdots & \mathbf{V}_2^\top \mathbf{A}_2 \mathbf{V}_n - \gamma & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{V}_n^\top \mathbf{A}_2 \mathbf{V}_1 - \gamma & \mathbf{V}_n^\top \mathbf{A}_2 X_2 - \gamma & \cdots & \mathbf{V}_n^\top \mathbf{A}_2 \mathbf{V}_n + (n-1)\gamma & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{V}_1^\top \mathbf{A}_1 \\ \mathbf{V}_2^\top \mathbf{A}_1 \\ \vdots \\ \mathbf{V}_n^\top \mathbf{A}_1 \\ 1 \end{pmatrix}.$$

Numerical exploration suggests the solution $\hat{\boldsymbol{q}}$ of this system, which is easily obtained numerically, is fairly robust with regards to the specific choice of γ. In a simulation study not reported here, the choices $\gamma = 1/n$ and $\gamma = \log(n)$ resulted in nearly identical estimators of the observation-specific weights. Furthermore, any value of $\gamma > 0$ resulted in a numerically stable system to solve. We therefore use $\gamma = 1/n$ in the remainder of the paper.

S4 Additional Simulation Results

S4.1 Simulation for simple linear EIV models

In this section, we report a set of simulation results for the simple errors-in-variables model $y_j = \gamma + \beta X_j + \epsilon_j,\ W_{jk} = X_j + U_{jk}$ for $j = 1, \ldots, n$ and $k = 1, \ldots, n_{\text{rep}}$. In this setting, the true covariates X_j are generated from the half-normal distribution scaled to have variance 1, i.e. $X_j \overset{iid}{\sim} (1 - 2/\pi)^{-1/2} \left|N(0, 1)\right|$. Three scenarios are considered for the distribution of the measurement error terms, namely normal, $U_{jk} \sim N(0, \sigma_j^2)$, Student's t with 2.5 degrees of freedom, $U_{jk} \sim \sigma_j/\sqrt{5}\, t_{2.5}$, and a contaminated normal, $U_{jk} \sim \sigma_j/\sqrt{10.9}\{0.9N(0, 1) + 0.1N(0, 10^2)\}$. In each scenario, the U_{jk} are generated independently, have mean 0, and are scaled to have $\text{Var}(U_{jk}) = \sigma_j^2$ for $k = 1, \ldots, n_{\text{rep}}$. In all the three scenarios, the measurement error variances σ_j^2 are generated from the uniform distribution $n_{\text{rep}} \times U(0.2, 1.5)$, so the signal-to-noise ratio $\text{Var}(X_j)/\text{Var}(U_j)$ for the averaged replicate measurement error U_j ranges from 2/3 (fairly weak signal) to 5 (fairly strong signal). The true intercept and slope are set to be $\gamma_{00} = 2$ and $\beta = 1$, respectively. The regression error ϵ_j's are generated to match the distribution of the measurement error in each scenario and with constant variance $\sigma_\varepsilon^2 = 0.25$.

For each generated sample, we compute the same set of estimators and used the same criterion $\det(\text{MSE}_{\text{rob}})$ as described in Section 4.2 of the main paper. Note that in the simple linear EIV model, the two heteroscedastic weighting schemes (minimax and quasi-likelihood) result in the same estimates. Table S1 summarizes the results.

It comes as no surprise that the naive estimator has the worst performance among the estimators considered. The large values observed

Table S1: Simple Linear EIV model performance naive OLS (Naive), moment-corrected (MC), and GMM estimators with equal (Equal), minimax (MM) and quasi-likelihood (QL) weights with $n_{\text{rep}} \in \{2, 3\}$ as measured by $\det(1000 \times \text{MSE}_{\text{rob}})$.

U	n	$n_{\text{rep}} = 2$				$n_{\text{rep}} = 3$			
		Naive	MC	GMM		Naive	MC	GMM	
				Equal	QL			Equal	QL
Normal	250	381.98	29.58	26.80	25.80	397.80	28.47	25.08	25.24
	500	181.67	9.44	6.21	6.65	160.34	8.13	6.52	6.10
	1000	88.78	2.13	1.63	1.56	90.83	1.80	1.40	1.40
$t_{2.5}$	250	263.14	20.89	10.56	8.74	322.03	17.25	9.32	7.99
	500	128.75	5.66	2.29	1.90	143.20	3.75	2.05	1.60
	1000	87.26	1.44	0.52	0.42	72.70	1.27	0.49	0.36
Cont.Normal	250	19.96	22.19	2.18	1.97	20.05	12.25	1.94	1.26
	500	44.30	15.28	0.98	0.66	46.25	11.49	0.90	0.48
	1000	98.12	14.27	0.75	0.48	90.87	12.58	0.66	0.37

demonstrate the consequences of ignoring measurement errors when they are present. When considering the various corrected estimators, the GEE estimators outperform the moment-corrected approach regardless of the weighting scheme used. The improvement of GMM estimators over moment correction is especially pronounced in contaminated normal error scenario, with the $t_{2.5}$ error scenario also showing some large decreases in estimation error. These cases illustrate the value of combining moment correction with the phase function-based approach. When comparing the two GMM weighting schemes, the estimator with quasi-likelihood weights also outperforms the equal weights estimator in almost all cases. The only exceptions occur under the normal measurement error model, where equal weighting in two instances performs better than quasi-likelihood weighting.

S4.2 Additional results for multiple linear EIV models

In this section, we summarize simulation results equivalent to those in Section 4.2 of the main paper. Tables S2 through S4 are analogous to Tables 1 through 3, but with $n_{\text{rep}} = 3$ whereas the main paper presents results for $n_{\text{rep}} = 2$. For greater specificity regarding the simulation configurations, see the descriptions in Section 4.2 of the main paper. We note here here that similar conclusions can be drawn for the case with $n_{\text{rep}} = 3$ replicates. Specifically, when the measurement error follows a $t_{2.5}$ or contaminated normal distribution, the GMM estimators outperform moment correction. On the other hand, when the measurement error follows a normal distribution, there isn't a clear preference for either the moment corrected or GMM estimators, each in turn outperforming the other. Finally, Table S5 is analogous to Table 4 of the main paper, showing the accuracy with which the standard error is estimated for the model parameters.

Table S2: 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 $n_{\mathrm{rep}} = 3$ as measured by $\det(1000 \times \mathrm{MSE}_{\mathrm{rob}})$.

R	**U**	n	True	Naive	MC	GMM		
						Equal	MM	QL
$\rho = 0$	Normal	250	0.010	14.237	1.565	2.082	2.023	2.087
		500	0.001	3.385	0.188	0.182	0.180	0.180
		1000	0.000	0.944	0.028	0.029	0.031	0.030
	$t_{2.5}$	250	0.012	20.132	1.827	0.967	0.805	0.802
		500	0.001	4.778	0.108	0.075	0.072	0.071
		1000	0.000	1.364	0.018	0.014	0.012	0.012
	Cont.Normal	250	0.009	13.100	1.009	0.853	0.650	0.676
		500	0.001	3.744	0.128	0.100	0.073	0.080
		1000	0.000	1.036	0.017	0.015	0.012	0.013
$\rho = 0.5$	Normal	250	0.009	34.094	2.511	3.097	2.892	2.771
		500	0.001	6.323	0.286	0.252	0.274	0.253
		1000	0.000	1.715	0.042	0.042	0.044	0.042
	$t_{2.5}$	250	0.008	39.974	2.152	1.134	1.004	1.129
		500	0.001	8.377	0.166	0.105	0.108	0.111
		1000	0.000	3.272	0.027	0.020	0.017	0.019
	Cont.Normal	250	0.009	41.149	2.271	1.761	1.400	1.475
		500	0.001	10.698	0.311	0.241	0.157	0.191
		1000	0.000	2.413	0.035	0.027	0.019	0.023

Table S3: 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 $n_{\text{rep}} = 3$ as measured by $\det(1000 \times \text{MSE}_{\text{rob}})$.

R	**U**	n	True	Naive	MC	GMM		
						Equal	MM	QL
$\rho = 0$	Normal	250	0.001	34.108	2.509	2.777	2.673	2.403
		500	0.000	1.716	0.096	0.108	0.122	0.108
		1000	0.000	0.089	0.002	0.003	0.003	0.003
	$t_{2.5}$	250	0.002	13.080	1.388	1.133	0.814	0.746
		500	0.000	2.394	0.051	0.042	0.036	0.034
		1000	0.000	0.255	0.002	0.001	0.002	0.001
	Cont.Normal	250	0.001	18.461	1.571	1.483	1.460	1.041
		500	0.000	1.495	0.055	0.055	0.055	0.036
		1000	0.000	0.057	0.001	0.001	0.001	0.001
$\rho = 0.5$	Normal	250	0.001	51.039	9.331	14.770	18.866	14.313
		500	0.000	4.915	0.234	0.266	0.291	0.257
		1000	0.000	0.336	0.009	0.010	0.011	0.011
	$t_{2.5}$	250	0.002	92.843	6.001	3.647	2.989	2.726
		500	0.000	5.910	0.143	0.105	0.104	0.090
		1000	0.000	0.648	0.006	0.004	0.003	0.003
	Cont.Normal	250	0.001	60.581	4.318	3.805	3.496	2.731
		500	0.000	4.936	0.150	0.149	0.128	0.094
		1000	0.000	0.296	0.005	0.004	0.004	0.003

Table S4: 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 $n_{\text{rep}} = 3$ as measured by $\det(1000 \times \text{MSE}_{\text{rob}})$.

R	**U**	n	True	Naive	MC	GMM		
						Equal	MM	QL
$\rho = 0$	Normal	250	0.001	14.862	1.997	1.972	1.964	1.976
		500	0.000	0.789	0.043	0.043	0.043	0.043
		1000	0.000	0.042	0.001	0.001	0.001	0.001
	$t_{2.5}$	250	0.001	13.456	1.132	1.087	0.978	1.056
		500	0.000	0.837	0.022	0.022	0.020	0.020
		1000	0.000	0.118	0.001	0.001	0.001	0.001
	Cont.Normal	250	0.001	21.877	2.346	2.379	2.159	2.282
		500	0.000	1.116	0.053	0.053	0.049	0.052
		1000	0.000	0.055	0.002	0.002	0.001	0.001
$\rho = 0.5$	Normal	250	0.001	51.724	4.557	4.554	4.543	4.555
		500	0.000	3.059	0.145	0.145	0.145	0.145
		1000	0.000	0.180	0.005	0.005	0.005	0.005
	$t_{2.5}$	250	0.001	32.805	2.819	2.777	2.470	2.635
		500	0.000	2.937	0.081	0.076	0.069	0.070
		1000	0.000	0.354	0.003	0.003	0.003	0.003
	Cont.Normal	250	0.001	44.184	4.592	4.594	4.243	4.462
		500	0.000	3.465	0.149	0.149	0.135	0.145
		1000	0.000	0.205	0.004	0.005	0.004	0.004

Table S5: Monte Carlo standard errors (MC-SE) and average of 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 $n_{\mathrm{rep}} = 3$ replicates and $\rho = 0.5$.

Setting	**U**	Coeff	$n = 500$		$n = 1000$	
			MC-SE	Avg-SE	MC-SE	Avg-SE
II	Normal	$\hat{\beta}_1$	0.030	0.029	0.021	0.021
		$\hat{\beta}_2$	0.028	0.029	0.022	0.021
		$\hat{\gamma}_{00}$	0.043	0.044	0.031	0.031
	$t_{2.5}$	$\hat{\beta}_1$	0.023	0.026	0.022	0.019
		$\hat{\beta}_2$	0.027	0.025	0.019	0.019
		$\hat{\gamma}_{00}$	0.035	0.036	0.027	0.025
IV	Normal	$\hat{\beta}_1$	0.036	0.033	0.021	0.022
		$\hat{\beta}_2$	0.032	0.030	0.021	0.022
		$\hat{\gamma}_{00}$	0.061	0.049	0.037	0.035
		$\hat{\gamma}_1$	0.026	0.027	0.020	0.019
		$\hat{\gamma}_2$	0.028	0.028	0.019	0.019
	$t_{2.5}$	$\hat{\beta}_1$	0.029	0.032	0.020	0.021
		$\hat{\beta}_2$	0.027	0.027	0.020	0.020
		$\hat{\gamma}_{00}$	0.043	0.046	0.032	0.030
		$\hat{\gamma}_1$	0.028	0.026	0.022	0.018
		$\hat{\gamma}_2$	0.029	0.025	0.016	0.018

References

Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of Mathematical Statistics*, pages 1977–1991.

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

Pötscher, B. M. and Prucha, I. R. (1989). A uniform law of large numbers for dependent and heterogeneous data processes. *Econometrica: Journal of the Econometric Society*, pages 675–683.