ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# A doubly corrected robust variance estimator for linear GMM☆

Jungbin Hwang [a,1], Byunghoon Kang [b,2], Seojeong Lee [c,*,3]

[a] University of Connecticut, United States of America
[b] Lancaster University, United Kingdom
[c] University of New South Wales, Australia

ARTICLE INFO

*Article history:*
Received 21 August 2019
Received in revised form 29 May 2020
Accepted 8 September 2020
Available online xxxx

*JEL classification:*
C12
C13
C23
C26

*Keywords:*
Generalized method of moments
Variance correction
Panel data
Model misspecification

ABSTRACT

We propose a new finite sample corrected variance estimator for the linear generalized method of moments (GMM) including the one-step, two-step, and iterated estimators. Our formula also corrects the over-identification bias in variance estimation on top of the commonly used finite sample correction of Windmeijer (2005), which corrects the bias from estimating the efficient weight matrix, so is *doubly* corrected. An important feature of the proposed double correction is that it automatically provides robustness to misspecification of the moment condition. In contrast, the conventional variance estimator and the Windmeijer correction are inconsistent under misspecification. That is, the double correction formula proposed in this paper provides a convenient way to obtain improved inference under correct specification and robustness against misspecification at the same time.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

The generalized method of moments (GMM) estimators (Hansen, 1982) are widely used in economics. Among the class of GMM estimators, the efficient GMM has the smallest asymptotic variance, which can be obtained via a two-step procedure. However, researchers have found that the standard error of the two-step efficient GMM is often severely downward biased. To solve this problem, Windmeijer (2005) proposed a finite sample bias-corrected standard error formula for the two-step linear GMM. Specifically, his formula corrects the bias arising from using the efficient weight matrix being evaluated at an estimate, instead of the true value. The correction formula (the Windmeijer correction, hereafter) has been routinely used in practice.[4]

☆ We thank the Co-Editor Xiaohong Chen, an anonymous Associate Editor, and an anonymous referee for valuable comments. We also thank Frank Windmeijer, Atsushi Inoue, Bruce Hansen, Frank Kleibergen, Whitney Newey, Richard Smith, Yixiao Sun and participants at ANZESG2019, ESAM2019, SETA2019, NASMES2019, IPDC2019, KER 2019, ESEM 2019 and CFE 2019 for helpful comments and suggestions.

* Corresponding author.
*E-mail addresses:* jungbin.hwang@uconn.edu (J. Hwang), b.kang1@lancaster.ac.uk (B. Kang), jay.lee@unsw.edu.au (S. Lee).

1 Hwang acknowledges the research support from Department of Economics at University of Connecticut, United States.

2 Kang acknowledges the research support from Lancaster University Management School, United Kingdom (Research Pump-Prime Grants ECA6396).

3 Lee acknowledges that this research was supported under the Australian Research Council Discovery Early Career Research Award (DECRA) funding scheme (project number DE170100787).

4 More than 5200 citations according to Google Scholar on May 26, 2020.

https://doi.org/10.1016/j.jeconom.2020.09.010
0304-4076/© 2021 Elsevier B.V. All rights reserved.

Please cite this article as: J. Hwang, B. Kang and S. Lee, A doubly corrected robust variance estimator for linear GMM. Journal of Econometrics (2021), https://doi.org/10.1016/j.jeconom.2020.09.010.

However, the Windmeijer correction does not take into account the *over-identification bias*, which is another important source of bias in the GMM standard error. The over-identification bias arises from the fact that the over-identified sample moment condition is nonzero in general while it converges in probability to zero under correct specification.

We propose a new finite sample correction, which takes into account the over-identification bias for the variance of the linear one-step, two-step, and iterated GMM estimators. For one-step GMM, such as the two-stage least squares (2SLS) estimators, the proposed finite sample correction is new as the Windmeijer correction does not cover the one-step GMM. For two-step and iterated GMM, the proposed correction improves upon the Windmeijer correction by additionally correcting for the over-identification bias. Thus, we *doubly* correct the finite sample bias of the linear GMM variance estimator.

The order of our double correction terms is equal to the order of the sample moment condition. Under correct specification or local misspecification (where the population moment condition is modeled as a drifting sequence within a $n^{-1/2}$-neighborhood of zero), these terms are $O_p(n^{-1/2})$ so that the double correction becomes a finite-sample correction for the variance. Readers are also referred to a working paper version of this paper (Hwang et al., 2020), which provides a stochastic expansion of the GMM estimators under local misspecification and shows that the double correction terms estimate the (co)variance of higher-order terms that increase with the over-identification bias.

Under (global) misspecification, however, the double correction terms no longer degenerate to zero asymptotically because the stochastic order of the sample moment condition becomes $O_p(1)$. The conventional variance estimator and the Windmeijer correction omit these $O_p(1)$ terms under misspecification. This implies that the conventional variance estimator and the Windmeijer correction are inconsistent, while our doubly corrected variance estimator is consistent regardless of whether the moment condition model is (locally or globally) misspecified or not.

Since the doubly corrected variance estimators are robust to misspecification, it is not surprising that the formulas coincide with the misspecification-robust variance estimator in Lee (2014) for one-step and two-step GMM and Hansen and Lee (2020) for iterated GMM. Indeed the simulation results reported in those papers show that the misspecification-robust variance estimator often performs better than the conventional sandwich variance estimator under correct specification. This paper provides an answer to this seemingly puzzling result by taking an alternative path to obtain the misspecification-robust variance estimator formula. Our approach provides new insight into the misspecification-robust formula as a finite-sample correction closely related to the well-known Windmeijer (2005) correction. We show that the misspecification-robust variance estimators of Lee (2014) and Hansen and Lee (2020) provide the same order of finite-sample correction with the Windmeijer (2005) correction under correct specification and local misspecification. To the best of our knowledge, this paper is the first to show the equivalence between the finite-sample corrected variance formula and the robust variance formula in misspecified GMM.

From a practical point of view, our paper's finding implies that accurate inference under correct specification and robust inference under misspecification can be achieved simultaneously, without knowing whether the moment condition is correctly specified. Moreover, practitioners can easily implement bootstrapping the t statistic standardized by the doubly corrected variance estimator and obtain more accurate t-tests and confidence intervals with smaller errors in the size and the coverage. Lee (2014) shows that this bootstrap procedure is robust to misspecification and does not require an ad hoc correction in the bootstrap sample called recentering.

The finite sample correction of the proposed formula and the Windmeijer formula work for linear models. For nonlinear models, the order of the remainder term is the same as the correction terms, so that the corrections do not necessarily provide improvements under correct specification.

Robust inference with possibly misspecified moment condition models has gained considerable attention in the literature. For linear instrumental variable (IV) models, Maasoumi and Phillips (1982) investigate the limiting distribution of inconsistent IV estimators. Guggenberger (2012) studies the behavior of the weak instrument robust tests under local misspecification. Kang (2018) derives higher-order expansions of IV estimators allowing for local violations of the instrument validity condition. Lee (2018) shows that the moment condition is misspecified under treatment effect heterogeneity and proposes a robust variance estimator for 2SLS.

More generally, Hall and Inoue (2003) derive the asymptotic distribution of GMM in misspecified moment condition models. Schennach (2007) proposes an alternative GEL-type estimator robust to global misspecification. Ai and Chen (2007) investigate the asymptotic properties of the sieve minimum distance estimator under misspecified conditional moment restrictions model. Otsu (2011) analyzes moderate deviation behaviors of GMM. Kitamura et al. (2013) propose an estimator that achieves optimal minimax robust properties under local misspecification. Lee (2014, 2016) propose a robust nonparametric bootstrap procedure for GMM and GEL estimators. Hansen and Lee (2020) provide robust inference theory for iterated GMM. Rotemberg (1983) and Andrews (2019) characterize the estimands of the linear GMM under misspecification. Andrews et al. (2017) propose to measure the effect of model misspecification on the sensitivity of parameter estimates for the minimum distance estimators. Bonhomme and Weidner (2018) and Armstrong and Kolesár (2019) consider minimax and GMM inference under possible misspecification, respectively.

Finite sample properties of GMM estimators, including the iterated and the continuously updating (CU) GMM, are investigated by Hansen et al. (1996). Bond and Windmeijer (2005) provide simulation evidence on the finite sample performance of the asymptotic and bootstrap tests based on GMM estimators. Hwang (2020) develops fixed-cluster asymptotics and finite-sample corrected variance formula for cross-sectionally dependent data. Hwang and Sun (2018) employ fixed-smoothing asymptotics to provide a more accurate comparison between the one-step and two-step GMM procedures for time-series observations.

Our doubly corrected robust variance estimators are generally different from the many instruments and many weak instruments robust variance estimators for IV, GMM, and GEL, e.g., Bekker (1994), Han and Phillips (2006), Newey and Windmeijer (2009), and Evdokimov and Kolesár (2018). Since our double correction formula does not use the many (weak) instruments asymptotics, it is not robust under such sequences.

The remainder of the paper is organized as follows. Section 2 reviews the Windmeijer correction. Section 3 proposes the doubly corrected variance estimator. Section 4 shows that the doubly corrected variance estimator is misspecification-robust. Section 5 discusses the iterated GMM and the CU GMM estimators. Section 6 derives the double correction formula for cross-sectional IV and panel data with the difference GMM. Section 7 provides extensive simulation results comparing the double correction and other variance estimators. Finally, Section 8 concludes. All the proofs are collected in Appendix.

## 2. Finite sample correction of Windmeijer (2005)

Suppose that we observe a sequence of i.i.d. random vectors $X_i \in \mathbb{R}^{d_x}$ for $i = 1, \ldots, n$. Let $g(X_i, \theta)$ be a $q \times 1$ moment function where $\theta$ is a $k \times 1$ parameter vector. We assume $q > k$ so that the model is over-identified and $g(X_i, \theta)$ is linear in parameter. When the model is just-identified ($q = k$) the correction terms are zeros and the analysis becomes trivial. The moment condition model is correctly specified if

$$E[g(X_i, \theta_0)] = 0 \tag{1}$$

for the true value $\theta_0$. Assume $E[\|g(X_i, \theta_0)\|^2] < \infty$ so that

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i, \theta_0) \equiv g_n(\theta_0) = O_p(n^{-1/2}).$$

Thus, the sample moment condition converges in probability to zero at the rate of $n^{-1/2}$ under correct specification (1). This will be used in determining the order of higher-order terms in Sections 2 and 3.

The one-step GMM estimator is defined as

$$\hat{\theta}_1 = \arg\min_{\theta \in \Theta} g_n(\theta)' W_n^{-1} g_n(\theta),$$

where $W_n$ is a $q \times q$ positive definite weight matrix which takes the form of $n^{-1}\sum_{i=1}^{n} W(X_i)$ and $W(X_i)$ does not depend on any unknown parameter. Common choices of $W(X_i)$ are the identity matrix and $Z_i Z_i'$, where $Z_i$ is the instrument vector in IV regressions. Let $W = EW_n$, a positive definite matrix of constants.

The two-step efficient GMM estimator using $\hat{\theta}_1$ as a preliminary (initial) estimator is defined as

$$\hat{\theta}_2 = \arg\min_{\theta \in \Theta} g_n(\theta)' [\Omega_n(\hat{\theta}_1)]^{-1} g_n(\theta),$$

where

$$\Omega_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(X_i, \theta) g(X_i, \theta)'.$$

Define $\Omega = \Omega(\theta_0)$ where $\Omega(\theta) = E\Omega_n(\theta)$. Since $\Omega_n(\hat{\theta}_1)$ is consistent for the asymptotic variance of the moment function, the two-step GMM estimator is efficient.

We also define an infeasible two-step GMM estimator $\tilde{\theta}_2$ using $[\Omega_n(\theta_0)]^{-1}$ as the weight matrix:

$$\tilde{\theta}_2 = \arg\min_{\theta \in \Theta} g_n(\theta)' [\Omega_n(\theta_0)]^{-1} g_n(\theta).$$

Investigating the limiting behavior of $\sqrt{n}(\tilde{\theta}_2 - \theta_0)$ will help us understand the higher-order behavior of the feasible two-step estimator $\sqrt{n}(\hat{\theta}_2 - \theta_0)$.

Let $G(X_i) = \partial g(X_i, \theta)/\partial \theta'$. Note that it does not depend on $\theta$ due to the linearity. Define $G_n = n^{-1}\sum_{i=1}^{n} G(X_i)$ and $G = EG_n$, which is assumed full column-rank. By the first-order Taylor expansion, the first-order condition (FOC) of the (feasible) two-step GMM estimator can be written as

$$0 = G_n'[\Omega_n(\hat{\theta}_1)]^{-1} g_n(\hat{\theta}_2) = G_n'[\Omega_n(\hat{\theta}_1)]^{-1}\left[g_n(\theta_0) + G_n(\hat{\theta}_2 - \theta_0)\right], \tag{2}$$

and so we have

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) = -\left\{G_n'[\Omega_n(\hat{\theta}_1)]^{-1} G_n\right\}^{-1} G_n'[\Omega_n(\hat{\theta}_1)]^{-1}\sqrt{n} g_n(\theta_0). \tag{3}$$

Using a similar expansion, we can get

$$\sqrt{n}(\tilde{\theta}_2 - \theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1} G_n\right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n} g_n(\theta_0) \tag{4}$$

for the infeasible two-step GMM estimator, and

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = -(G_n' W_n^{-1} G_n)^{-1} G_n' W_n^{-1} \sqrt{n} g_n(\theta_0) \tag{5}$$

for the one-step GMM estimator.

Asymptotically (3) and (4) have the same limiting distribution so that using $\Omega_n(\hat{\theta}_1)$ instead of $\Omega_n(\theta_0)$ does not affect the first-order asymptotic analysis. However, by expanding $\Omega_n(\hat{\theta}_1)$ around $\theta_0$ and using (5), Windmeijer (2005) shows that the extra finite sample variation caused by higher-order terms can be estimated and one can improve the accuracy of the variance estimator for linear moment conditions.

To see this, we use the first-order Taylor expansion of $\Omega_n(\hat{\theta}_1)$ on the right-hand side (RHS) of (3) around $\theta_0$:

$$\begin{aligned}\sqrt{n}(\hat{\theta}_2 - \theta_0) &= -\left\{G_n'[\Omega_n(\theta_0)]^{-1} G_n\right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1} \sqrt{n} g_n(\theta_0) + D_n \sqrt{n}(\hat{\theta}_1 - \theta_0) + R_n \\ &= \sqrt{n}(\tilde{\theta}_2 - \theta_0) + \underbrace{D_n \sqrt{n}(\hat{\theta}_1 - \theta_0)}_{=O_p(n^{-1/2})} + R_n,\end{aligned} \tag{6}$$

where

$$\begin{aligned} D_n &= F_{1n} + F_{2n}, \\ F_{1n} &= -\left.\frac{\partial \left\{G_n'[\Omega_n(\theta)]^{-1} G_n\right\}^{-1}}{\partial \theta'}\right|_{\theta=\theta_0} G_n'[\Omega_n(\theta_0)]^{-1} g_n(\theta_0), \\ F_{2n} &= -\left\{G_n'[\Omega_n(\theta_0)]^{-1} G_n\right\}^{-1} \left.\frac{\partial \{G_n'[\Omega_n(\theta)]^{-1} g_n(\theta_0)\}}{\partial \theta'}\right|_{\theta=\theta_0} \end{aligned}$$

are $k \times k$ matrices and $R_n$ is the remainder term. Since $g_n(\theta_0) = O_p(n^{-1/2})$ both $F_{1n}$ and $F_{2n}$ are $O_p(n^{-1/2})$. Thus, the second term on the RHS of (6) is of order $O_p(n^{-1/2})$ assuming that $\sqrt{n}(\hat{\theta}_1 - \theta_0) = O_p(1)$. The remainder term $R_n$ is of order $O_p(n^{-1})$ because of the linearity of the moment function, provided that the higher moments of $g(X_i, \theta_0)$ and $G(X_i)$ exist (the formal justification is given in the proof of Theorem 1). Thus, by taking into account for the variation caused by the $O_p(n^{-1/2})$ term, the finite sample variance of $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ can be more accurately approximated. Note that the expansion (6) only holds for linear moment condition models.

The Windmeijer correction of the variance of $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ is obtained by

$$\widehat{V}_w(\hat{\theta}_2) = \widetilde{V}(\hat{\theta}_2) + \widehat{D}_n \widetilde{V}(\hat{\theta}_2) + \widetilde{V}(\hat{\theta}_2) \widehat{D}_n' + \widehat{D}_n \widetilde{V}(\hat{\theta}_1) \widehat{D}_n', \tag{7}$$

where $D[., j]$ denotes the $j$th column of $D$, $\theta_{[j]}$ denotes the $j$th element of $\theta$, and

$$\begin{aligned} \widetilde{V}(\hat{\theta}_1) &= \left(G_n' W_n^{-1} G_n\right)^{-1} \left(G_n' W_n^{-1} \Omega_n(\hat{\theta}_1) W_n^{-1} G_n\right) \left(G_n' W_n^{-1} G_n\right)^{-1}, \\ \widetilde{V}(\hat{\theta}_2) &= \left\{G_n'[\Omega_n(\hat{\theta}_1)]^{-1} G_n\right\}^{-1}, \\ \widehat{D}_n[., j] &= \left\{G_n'[\Omega_n(\hat{\theta}_1)]^{-1} G_n\right\}^{-1} G_n' \left\{[\Omega_n(\hat{\theta}_1)]^{-1} \left.\frac{\partial \Omega_n(\theta)}{\partial \theta_{[j]}}\right|_{\theta=\hat{\theta}_1} [\Omega_n(\hat{\theta}_1)]^{-1}\right\} g_n(\hat{\theta}_2), \\ \frac{\partial \Omega_n(\theta)}{\partial \theta_{[j]}} &= \Upsilon_j(\theta) + \Upsilon_j'(\theta), \\ \Upsilon_j(\theta) &= \frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta) \frac{\partial g(X_i, \theta)'}{\partial \theta_{[j]}}. \end{aligned}$$

Since the estimate of $F_{1n}$ is equal to zero because of the FOC, $0 = G_n'[\Omega_n(\hat{\theta}_1)]^{-1} g_n(\hat{\theta}_2)$, it does not appear in the variance estimator formula. The standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_w(\hat{\theta}_2)/n}$.

## 3. Double correction

The Windmeijer correction accounts for the extra variability due to using an estimated parameter in the weight matrix. This correction is effective because $\widehat{D}_n \neq 0$, which is due to $g_n(\hat{\theta}_2) \neq 0$ in finite sample. In fact, $g_n(\theta) \neq 0$ for all $\theta$ almost surely if at least one of the moments is continuously distributed, which trivially implies $g_n(\theta_0) \neq 0$. We call this the *over-identification bias*, which is non-zero for any $n$ in general for over-identified models.

We show that the over-identification bias causes additional finite sample variability in (6). These additional terms are not considered in the Windmeijer correction (7). We propose alternative variance estimators that fully incorporate the additional variation induced by the over-identification bias. These variance estimators will replace $\widetilde{V}(\hat{\theta}_2)$ and $\widetilde{V}(\hat{\theta}_1)$ in (7) without affecting the order of finite sample corrections, leading to our doubly corrected variance estimator.

Define $g = g(\theta_0)$ where $g(\theta) = E[g(X_i, \theta)]$. Assume that

$$G_n - G = O_p(n^{-1/2}), \tag{8}$$
$$\text{vec}(W_n - W) = O_p(n^{-1/2}), \tag{9}$$
$$\text{vec}(\Omega_n(\theta_0) - \Omega) = O_p(n^{-1/2}), \tag{10}$$

which hold under appropriate regularity conditions. Since $G'\Omega^{-1}g = 0$ by the population FOC and

$$[\Omega_n(\theta_0)]^{-1} - \Omega^{-1} = -\Omega^{-1}(\Omega_n(\theta_0) - \Omega)[\Omega_n(\theta_0)]^{-1},$$

we can write

$$\begin{aligned} &G_n'[\Omega_n(\theta_0)]^{-1}g_n(\theta_0) \\ &= \underbrace{G'\Omega^{-1}g_n(\theta_0)}_{=O_p(n^{-1/2})} + \underbrace{(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}(\Omega_n(\theta_0) - \Omega)\Omega^{-1}g_n(\theta_0)}_{=O_p(n^{-1})} + O_p(n^{-3/2}). \end{aligned} \tag{11}$$

Using (11), the expansion of the infeasible two-step GMM estimator (4) can be written as

$$\sqrt{n}(\tilde{\theta}_2 - \theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}\left[G'\Omega^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{12}$$
$$\left. + \sqrt{n}(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}(\Omega_n(\theta_0) - \Omega)\Omega^{-1}g_n(\theta_0)\right] \tag{13}$$
$$+ O_p(n^{-1}). \tag{14}$$

Similarly,

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = -\left\{G_n'W_n^{-1}G_n\right\}^{-1}\left[G'W^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{15}$$
$$\left. + \sqrt{n}(G_n - G)'W^{-1}g_n(\theta_0) - G'W^{-1}\sqrt{n}(W_n - W)W^{-1}g_n(\theta_0)\right] \tag{16}$$
$$+ O_p(n^{-1}), \tag{17}$$

which simplifies to

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = -\left\{G_n'G_n\right\}^{-1}\left[G'\sqrt{n}g_n(\theta_0) + \sqrt{n}(G_n - G)'g_n(\theta_0)\right]$$

when $W_n = I$. From the above expansions, we learn the followings. First, we need to consider the extra variation from $\sqrt{n}(G_n - G)$ and $\sqrt{n}(\Omega_n(\theta) - \Omega)$ (or $\sqrt{n}(W_n - W)$ for the one-step GMM) to account for the over-identification bias. Second, the order of the remainder term of the original expansion (6) is not changed.

Using the expansions (12)–(14) an (15)–(17), the expansion of the two-step GMM can be written as

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}\left[G'\Omega^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{18}$$
$$\left. + \sqrt{n}(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}(\Omega_n(\theta_0) - \Omega)\Omega^{-1}g_n(\theta_0)\right] \tag{19}$$
$$- D_n\left\{G_n'W_n^{-1}G_n\right\}^{-1}\left[G'W^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{20}$$
$$\left. + \sqrt{n}(G_n - G)'W^{-1}g_n(\theta_0) - G'W^{-1}\sqrt{n}(W_n - W)W^{-1}g_n(\theta_0)\right] \tag{21}$$
$$+ O_p(n^{-1}). \tag{22}$$

Note that (22) is the sum of (14), (17), and $R_n$ in (6). The first two terms are $O_p(n^{-1})$ under (8)–(10).

In finite sample, $g_n(\theta_0) \neq 0$ because $g_n(\theta) \neq 0$ for all $\theta$ almost surely, and this causes extra variation through the terms in (19) and (21). Similar to the Windmeijer correction, by taking into account for these (asymptotically negligible) terms in estimating the variance we can make more accurate inference.

Since $D_n = O_p(n^{-1/2})$, the terms in (21) multiplied by $D_n$ are $O_p(n^{-1})$, which is the same order as the remainder term. Thus, considering those terms in (21) does not necessarily provide finite sample corrections. However, including these terms are critical to getting robustness to misspecification, which is shown in Section 4.

The expansion for the one-step GMM is (15)–(17), and those terms in (16) are $O_p(n^{-1/2})$. Thus, considering the finite sample variation caused by these terms provides a more accurate variance estimator formula. This correction for the one-step GMM is not considered in Windmeijer (2005) and is new.

The doubly corrected variance estimator of $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ is

$$\widehat{V}_{dc}(\hat{\theta}_2) = \widehat{V}(\hat{\theta}_2) + \widehat{D}_n\widehat{C}(\hat{\theta}_1, \hat{\theta}_2) + \widehat{C}(\hat{\theta}_1, \hat{\theta}_2)'\widehat{D}_n' + \widehat{D}_n\widehat{V}_{dc}(\hat{\theta}_1)\widehat{D}_n',$$

where

$$\widehat{V}(\hat{\theta}_2) = \left(G_n'[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1}\Sigma_n(\hat{\theta}_2, \Omega_n(\hat{\theta}_1))\left(G_n'[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1},$$

$$\widehat{V}_{dc}(\hat{\theta}_1) = \left(G_n' W_n^{-1} G_n\right)^{-1} \Sigma_n(\hat{\theta}_1, W_n) \left(G_n' W_n^{-1} G_n\right)^{-1},$$
$$\widehat{C}(\hat{\theta}_1, \hat{\theta}_2) = \left(G_n' W_n^{-1} G_n\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} m_i(\hat{\theta}_1, W_n) m_i(\hat{\theta}_2, \Omega_n(\hat{\theta}_1))' \left(G_n' [\Omega_n(\hat{\theta}_1)]^{-1} G_n\right)^{-1},$$

and

$$\begin{aligned} \Sigma_n(\theta, \Xi_n(\phi)) =& \frac{1}{n} \sum_{i=1}^{n} m_i(\theta, \Xi_n(\phi)) m_i(\theta, \Xi_n(\phi))', \\ m_i(\theta, \Xi_n(\phi)) =& G_n' [\Xi_n(\phi)]^{-1} g(X_i, \theta) + G(X_i)' [\Xi_n(\phi)]^{-1} g_n(\theta) \\ & - G_n' [\Xi_n(\phi)]^{-1} \Xi(X_i, \phi) [\Xi_n(\phi)]^{-1} g_n(\theta), \\ \Xi_n(\phi) =& \frac{1}{n} \sum_{i=1}^{n} \Xi(X_i, \phi). \end{aligned} \tag{23}$$

When $\Xi_n(\phi) = \Xi(X_i, \phi) = I$, the last term of $m(\theta, \Xi_n(\phi))$ drops. Note that $G(X_i)$ and $\Xi(X_i, \phi)$ in $m_i(\theta, \Xi_n(\phi))$ are not centered because the FOCs hold evaluated at $(\hat{\theta}_2, \Omega_n(\hat{\theta}_1))$ and $(\hat{\theta}_1, W_n)$, respectively.

The doubly corrected variance estimator for the two-step GMM, $\widehat{V}_{dc}(\hat{\theta}_2)$, provides the same order of finite sample correction as the Windmeijer correction, $\widehat{V}_w(\hat{\theta}_2)$. The standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{dc}(\hat{\theta}_2)/n}$.

The doubly corrected variance estimator for the one-step GMM, $\widehat{V}_{dc}(\hat{\theta}_1)$, accounts for the variations up to the order of $O_p(n^{-1/2})$ in the expansion (15)–(17). This correction is not considered in Windmeijer (2005). The standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{dc}(\hat{\theta}_1)/n}$.

## 4. Robustness to misspecification

The variance estimators considered so far, the doubly corrected, the Windmeijer corrected, and the conventional, are consistent for the asymptotic variance of $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ under correct specification, $E[g(X_i, \theta_0)] = 0$. In words, the correct specification means that an over-identified model exactly holds at the true value $\theta_0$, but this may be too restrictive in reality. Indeed, the sample moment condition does not hold for any finite sample size $n$ almost surely if the model is over-identified, i.e., $g_n(\hat{\theta}) \neq 0$, provided that at least one of the moments is continuously distributed. Thus, it is reasonable to view the assumed moment condition model as the best-approximating model and to allow for possible misspecification.

Under (global) misspecification, which is defined as

$$E[g(X_i, \theta)] = \delta(\theta) \neq 0, \ \forall \theta \in \Theta,$$

where $\delta(\theta)$ is a vector of constants and $\Theta$ is the parameter space, the GMM estimator is consistent for the pseudo-true value which is defined as the unique minimizer of the population GMM criterion given the weight matrix (Hall and Inoue, 2003). In addition, the asymptotic variance has more terms that are assumed away under correct specification. Thus, the conventional variance estimators are no longer consistent under misspecification. Lee (2014) proposes variance estimators for one-step and two-step GMM under misspecification. Hansen and Lee (2020) propose a similar robust variance estimator for iterated GMM. These variance estimators are shown to be consistent regardless of misspecification and they are referred to as the misspecification-robust variance estimator, hereafter.

Since the misspecification-robust variance estimators contain additional terms that are not present in the conventional variance estimator, one may conjecture that it is less accurate than the conventional formula under correct specification. We show that this conjecture is not true by proving that the doubly corrected variance estimator $\widehat{V}_{dc}(\hat{\theta}_2)$ is the misspecification-robust variance estimator.

To formally show that $\widehat{V}_{dc}(\hat{\theta}_2)$ is consistent for the asymptotic variance under misspecification, we introduce additional definitions. Define the one-step and two-step GMM (pseudo-) true values as

$$\theta_1 = \arg\min_{\theta \in \Theta} E[g(X_i, \theta)]' W^{-1} E[g(X_i, \theta)], \tag{24}$$
$$\theta_2 = \arg\min_{\theta \in \Theta} E[g(X_i, \theta)]' [\Omega(\theta_1)]^{-1} E[g(X_i, \theta)]. \tag{25}$$

In general $\theta_1 \neq \theta_2$, but $\theta_1 = \theta_2 = \theta_0$ under correct specification. Write $g_j = E[g(X_i, \theta_j)]$ and $\Omega_j = \Omega(\theta_j)$ for $j = 1, 2$. In the presence of global misspecification, $g_n(\theta_j) = O_p(1)$ for $j = 1, 2$.

The robustness of $\widehat{V}_{dc}(\hat{\theta}_2)$ holds for the following reasons. Recall that the formulas for $\widehat{V}_{dc}(\hat{\theta}_2)$ and $\widehat{V}_w(\hat{\theta}_2)$ are given by

$$\widehat{V}_{dc}(\hat{\theta}_2) = \widehat{V}(\hat{\theta}_2) + \widehat{D}_n \widehat{C}(\hat{\theta}_1, \hat{\theta}_2) + \widehat{C}(\hat{\theta}_1, \hat{\theta}_2)' \widehat{D}_n' + \widehat{D}_n \widehat{V}_{dc}(\hat{\theta}_1) \widehat{D}_n',$$
$$\widehat{V}_w(\hat{\theta}_2) = \widetilde{V}(\hat{\theta}_2) + \widehat{D}_n \widetilde{V}(\hat{\theta}_2) + \widetilde{V}(\hat{\theta}_2) \widehat{D}_n' + \widehat{D}_n \widetilde{V}(\hat{\theta}_1) \widehat{D}_n'.$$

The correction term $\widehat{D}_n$ corrects the bias in the variance due to using the weight matrix $[\Omega_n(\hat{\theta}_1)]^{-1}$ instead of $[\Omega_n(\theta_1)]^{-1}$. Since both $\widehat{V}_{dc}(\hat{\theta}_2)$ and $\widehat{V}_w(\hat{\theta}_2)$ have $\widehat{D}_n$, this bias is corrected in both variance estimators. What is not accounted for in the Windmeijer corrected variance estimator is the additional variation in the sample Jacobian $G_n$, and the sample weight matrices $W_n$ and $\Omega_n(\theta_1)$. These variations are asymptotically negligible under correct specification but become the first-order under misspecification. Our doubly corrected variance estimator $\widehat{V}_{dc}(\hat{\theta}_2)$ accounts for these variations.

The expansion of the GMM estimators in misspecified models is quite similar to those in correctly specified models, except that we need to allow different pseudo-true values for one-step and two-step GMM, and the moment condition evaluated at the pseudo-true value is not equal to zero.

Consider the FOC of the two-step GMM estimator (2). By expanding $g_n(\hat{\theta}_2)$ around $\theta_2$, and $\Omega_n(\hat{\theta}_1)$ around $\theta_1$, we can write

$$\begin{aligned}\sqrt{n}(\hat{\theta}_2-\theta_2) &= -\left\{G_n'[\Omega_n(\theta_1)]^{-1}G_n\right\}^{-1}G_n'[\Omega_n(\theta_1)]^{-1}\sqrt{n}g_n(\theta_2)+D_n^*\sqrt{n}(\hat{\theta}_1-\theta_1)+R_n^*\\ &= \sqrt{n}(\tilde{\theta}_2^*-\theta_2)+D_n^*\sqrt{n}(\hat{\theta}_1-\theta_1)+R_n^*,\end{aligned} \tag{26}$$

where $\tilde{\theta}_2^*$ is defined as

$$\tilde{\theta}_2^* = \arg\min_{\theta\in\Theta}\ g_n(\theta)'[\Omega_n(\theta_1)]^{-1}g_n(\theta), \tag{27}$$

and

$$\begin{aligned}D_n^* &= F_{1n}^*+F_{2n}^*,\\ F_{1n}^* &= -\left.\frac{\partial\left\{G_n'[\Omega_n(\theta)]^{-1}G_n\right\}^{-1}}{\partial\theta'}\right|_{\theta=\theta_1}G_n'[\Omega_n(\theta_1)]^{-1}g_n(\theta_2),\\ F_{2n}^* &= -\left\{G_n'[\Omega_n(\theta_1)]^{-1}G_n\right\}^{-1}\left.\frac{\partial G_n'[\Omega_n(\theta)]^{-1}g_n(\theta_2)}{\partial\theta'}\right|_{\theta=\theta_1},\end{aligned}$$

and $R_n^*$ is the remainder term of order $O_p(n^{-1/2}\|g_n(\theta_2)\|)$ (which is formally justified with the order of other terms in the proof of Theorem 1). Since $D_n^* = O_p(\|g_n(\theta_2)\|)$, the order of finite sample correction depends on the degree of misspecification, from being $O_p(n^{-1/2})$ under correct specification to $O_p(1)$ under misspecification. Note that both $\sqrt{n}(\tilde{\theta}_2^*-\theta_2)$ and $D_n^*\sqrt{n}(\hat{\theta}_1-\theta_1)$ in (26) are $O_p(1)$ under misspecification and this will alter the first-order asymptotic variance.

Using the population FOC $G'\Omega_j^{-1}g_j = 0$ for $j = 1, 2$, the FOC of the infeasible two-step GMM estimator (27) can be expanded as

$$\begin{aligned}\sqrt{n}(\tilde{\theta}_2^*-\theta_2) = -\left\{G_n'[\Omega_n(\theta_1)]^{-1}G_n\right\}^{-1}\Big\{&G_n'[\Omega_n(\theta_1)]^{-1}\sqrt{n}\,(g_n(\theta_2)-g_2)\\ &+\sqrt{n}\,(G_n-G)'\,[\Omega_n(\theta_1)]^{-1}g_2 - G'\Omega_1^{-1}\sqrt{n}\,(\Omega_n(\theta_1)-\Omega_1)\,[\Omega_n(\theta_1)]^{-1}g_2\Big\}.\end{aligned} \tag{28}$$

The FOC of the one-step GMM can be expanded similarly:

$$\begin{aligned}\sqrt{n}(\hat{\theta}_1-\theta_1) = -\left\{G_n'W_n^{-1}G_n\right\}^{-1}\Big[&G_n'W_n^{-1}\sqrt{n}\,(g_n(\theta_1)-g_1)\\ &+\ \sqrt{n}(G_n-G)'W_n^{-1}g_1 - G'W^{-1}\sqrt{n}\,(W_n-W)\,W_n^{-1}g_1\Big].\end{aligned} \tag{29}$$

The expansions (28) and (29) are misspecification-robust versions of (4) and (5), allowing for different probability limits of the one-step and the two-step GMM estimators, and taking into account for the misspecification (over-identification) bias. Under correct specification, $g_1 = g_2 = 0$ and (28) and (29) coincide with (4) and (5).

Now we list assumptions for the main result.

**Assumption 1.**

(i) $\theta_j$ is unique and is in the interior of the parameter space $\Theta$ for $j = 1, 2$
(ii) $X_1, \ldots, X_n$ are i.i.d.
(iii) $W$ and $\Omega(\theta_1)$ are nonsingular
(iv) $G$ is full column rank
(v) $E[\|g(X_i,\theta_1)\|^4] < \infty$, $E[\|g(X_i,\theta_2)\|^2] < \infty$, $E[\|G(X_i)\|^4] < \infty$, $E[\|W(X_i)\|^2] < \infty$

Assumption 1 is mild regularity conditions for consistency and asymptotic normality of the one-step and the two-step linear GMM estimators allowing for global misspecification. For misspecified models, Hall and Inoue (2003) provide a list of conditions for one-step and two-step GMM in the time series context. Hansen and Lee (2020) provide a list of conditions for one-step and iterated GMM under the independent but not necessarily identically distributed (i.n.i.d.) sampling.

The following theorem shows that the doubly corrected variance estimators of one-step and two-step linear GMM are consistent for the asymptotic variance matrices under misspecification. The proof is given in Appendix.

**Theorem 1.** *Suppose that Assumption 1 holds. As $n \to \infty$, for $j = 1, 2$,*

$$\sqrt{n}(\hat{\theta}_j - \theta_j) \xrightarrow{d} N(0, V_j)$$

*and*

$$\widehat{V}_{dc}(\hat{\theta}_j) \xrightarrow{p} V_j.$$

Theorem 1 holds regardless of whether the model is correctly specified or not. Thus, $\widehat{V}_{dc}(\hat{\theta}_j)$ for $j = 1, 2$, provides a finite sample correction under correct specification and it remains consistent under misspecification. In contrast, the Windmeijer corrected variance estimator, $\widehat{V}_w(\hat{\theta}_2)$, is not first-order consistent under misspecification. Assuming the linearity of moment condition, the proofs of Lemma 2 and Theorem 1 in Appendix can be used to rigorously justify the stochastic orders of the higher-order terms mentioned in the previous sections.

For nonlinear models, $\widehat{V}_{dc}(\hat{\theta}_j)$ formula can be adjusted by considering the second derivative of the moment function, which coincides with the misspecification-robust formula of Lee (2014) and Hansen and Lee (2020). The same consistency result with Theorem 1 is shown in those papers, but stronger assumptions on the moment/Jacobian processes and the compact parameter space are required to use the uniform law of large numbers.

Theorem 1 also implies that the nonparametric i.i.d. bootstrap $t$ test and confidence intervals (CIs) based on the GMM $t$ statistic studentized with the doubly corrected standard error automatically achieve higher-order refinements over the asymptotic $t$ test and CIs regardless of misspecification (Lee, 2014). In contrast, those bootstrap $t$ test and CIs based on the GMM $t$ statistic studentized with the conventional or the Windmeijer standard error require an additional recentering procedure in resampling to correct for the over-identification bias to achieve higher-order refinements, see Hall and Horowitz (1996) and Andrews (2002). Furthermore, the conventional nonparametric i.i.d. bootstrap procedure for GMM is not valid under misspecification. Thus, our doubly corrected variance estimator formula provides a very convenient way to get more accurate but also robust bootstrap tests and CIs.

**Remark 1** (*Stochastic Expansion Under Local Misspecification*)**.** From the way it is constructed, we can directly see that the doubly corrected variance formulas provide a finite-sample variance correction with the same argument as Windmeijer (2005). It is expected to work the best when the sample moment condition is large. Since a nonzero sample moment condition can also be due to a locally misspecified moment condition, one can seek an additional justification of the double correction by deriving the stochastic expansions of the GMM estimators under such a sequence. In a working paper version of this paper (Hwang et al., 2020), we provide formal stochastic expansions of the one-step and the two-step GMM estimators by allowing the population moment condition evaluated at the true value is $E[g(X_{in}, \theta_0)] = \delta/\sqrt{n}$ for some $\delta \neq 0$. We also discuss how the first-order terms $D_n^*$, (28), and (29) in the double correction are related to the higher-order terms under local misspecification. Interestingly, the analysis reveals that the double correction effectively estimates the (co)variances of higher-order terms that depend on $\delta$ and thus is fully robust to additional variations due to local misspecification. In contrast, the Windmeijer corrected variance estimator only partially considers the higher-order terms that depend on $\delta$, making it only partly robust to local misspecification. Finally, the stochastic expansions show that both the doubly corrected and the Windmeijer corrected variance estimators are not higher-order variance estimators which would estimate additional terms up to $O(n^{-1})$. For the general treatment of the stochastic expansion, see Rothenberg (1984). Newey and Smith (2004) derive the stochastic expansion of GMM and the generalized empirical likelihood (GEL) estimators under correct specification.

**Remark 2** (*Weight Matrix*)**.** The main results hold if we replace $\Omega_n(\theta)$ with the centered weight matrix

$$\Omega_n^c(\theta) = \frac{1}{n}\sum_{i=1}^{n} (g(X_i, \theta) - g_n(\theta))\,(g(X_i, \theta) - g_n(\theta))', \tag{30}$$

with some specifics need to be modified accordingly. Specifically, the two-step GMM pseudo-true value $\theta_2$ defined in (25) is now defined with $\Omega_n^c(\theta_1)$. In addition, the derivative of the centered weight matrix is

$$\Upsilon_j^c(\theta) = \frac{1}{n}\sum_{i=1}^{n}(g(X_i, \theta) - g_n(\theta))\left(\frac{\partial g(X_i, \theta)}{\partial \theta_{[j]}} - \frac{\partial g_n(\theta)}{\partial \theta_{[j]}}\right)'$$

so that

$$\frac{\partial \Omega_n^c(\theta)}{\partial \theta_{[j]}} = \Upsilon_j^c(\theta) + \Upsilon_j^{c\prime}(\theta).$$

The centered weight matrix is consistent for the asymptotic variance matrix of the moment equation under misspecification. Hansen (2020) recommends using the centered weight matrix for this reason. Hall (2000) shows that the GMM over-identification test statistic with a centered heteroskedasticity-and-autocorrelation-consistent (HAC) weight matrix leads to more powerful tests in the time series setting.

## 5. Iterated GMM and continuously updating GMM

Both the Windmeijer and our double correction correct the extra variation due to the weight matrix being evaluated at an estimate rather than the true value. A natural question is whether similar finite sample corrections can be obtained for other GMM estimators, namely the iterated GMM estimator and the continuously-updating (CU) GMM estimator of Hansen et al. (1996). We show that the answer is yes for the iterated GMM estimator, and the double correction formula is the same as the misspecification-robust formula. For the CU GMM estimator, the answer is negative.

Assume the correct specification. The iterated GMM estimator is obtained by iterating the two-step efficient GMM estimator until convergence. By iteration the dependence of the final estimator on the previous step estimators disappears. The FOC is given by

$$0 = G_n'[\Omega_n(\hat{\theta})]^{-1} g_n(\hat{\theta}),$$

where $\hat{\theta}$ is the iterated GMM estimator. Assume that $g_n(\theta_0) = O_p(n^{-1/2})$ and $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ whose sufficient conditions are provided in Hansen and Lee (2020). By applying the first-order Taylor expansion around $\theta_0$ to $g_n(\hat{\theta})$ and $\Omega_n(\hat{\theta})$ sequentially

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1} G_n\right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n} g_n(\theta_0) + D_n\sqrt{n}(\hat{\theta} - \theta_0) + O_p(n^{-1}),$$

and thus

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1} G_n(I_k - D_n)\right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n} g_n(\theta_0) + O_p(n^{-1}). \tag{31}$$

Windmeijer (2000) proposes a finite sample corrected variance estimator based on the expansion (31). We proceed one additional step. By further expanding to take into account for the over-identification bias, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1} G_n(I_k - D_n)\right\}^{-1} \left[G'\Omega^{-1}\sqrt{n} g_n(\theta_0)\right. \tag{32}$$

$$\left. + \sqrt{n}(G_n - G)'\Omega^{-1} g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}\,(\Omega_n(\theta_0) - \Omega)\,\Omega^{-1} g_n(\theta_0)\right] + O_p(n^{-1}). \tag{33}$$

Since the remainder term is $O_p(n^{-1})$, by estimating the variance of the terms in (32)–(33) up to $O_p(n^{-1/2})$ we can get the same order of finite sample correction with the doubly corrected two-step GMM variance estimator.

The doubly corrected variance estimator for iterated GMM is

$$\widehat{V}_{dc}(\hat{\theta}) = \{G_n'[\Omega_n(\hat{\theta})]^{-1} G_n(I_k - \widehat{D}_n)\}^{-1}\,\Sigma_n(\hat{\theta}, \Omega_n(\hat{\theta}))\{G_n'[\Omega_n(\hat{\theta})]^{-1} G_n(I_k - \widehat{D}_n)\}^{-1\prime},$$

where $\Sigma_n(\hat{\theta}, \Omega_n(\hat{\theta}))$ is defined in (23) and $\widehat{D}_n$ is evaluated at $\hat{\theta}$. Not surprisingly, this formula is identical to Hansen and Lee (2020)'s misspecification-robust variance estimator for iterated GMM. The finite sample corrected formula suggested by Windmeijer (2000) is

$$\widehat{V}_w(\hat{\theta}) = (I_k - \widehat{D}_n)^{-1}\left(G_n'[\Omega_n(\hat{\theta})]^{-1} G_n\right)^{-1}(I_k - \widehat{D}_n)^{-1\prime}.$$

On the other hand, a similar finite sample correction may not be obtained for CU GMM. Windmeijer (2005) showed that if the derivative of the moment function is a function of the parameter, then the proposed formula would not necessarily give finite sample corrections. The same argument applied to CU GMM. Let $\hat{\theta}$ be the CU GMM estimator. For simplicity, let $k = 1$ so that $\theta$ is scalar. The FOC is

$$0 = \left(G_n - \frac{1}{2} g_n(\hat{\theta})'[\Omega_n(\hat{\theta})]^{-1}\left.\frac{\partial \Omega_n(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}}\right)' [\Omega_n(\hat{\theta})]^{-1} g_n(\hat{\theta}).$$

This shows that even when the moment function is linear, the effective Jacobian term in the FOC still depends on the parameter. Thus, the misspecification-robust variance formula for CU GMM does not necessarily provide a finite sample correction under correct specification. Since GEL estimators have similar non-linear FOC even with linear moment functions, we expect similar conclusions.

## 6. Examples

### 6.1. Cross-sectional IV

Consider the linear IV model $y_i = X_i'\theta + e_i$ with the moment conditions $E[Z_i e_i] = 0$. The two-stage least squares (2SLS) estimator is given by

$$\hat{\theta}_1 = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y,$$

where $Y = [y_1, \ldots, y_n]'$, $X = [X_1, \ldots, X_n]'$, and $Z = [Z_1, \ldots, Z_n]'$ are $n \times 1$, $n \times k$, and $n \times q$ data matrices. Using the 2SLS as the preliminary estimator, the two-step efficient GMM estimator is given by

$$\hat{\theta}_2 = (X'Z\widehat{\Omega}_1^{-1}Z'X)^{-1}X'Z\widehat{\Omega}_1^{-1}Z'Y,$$

where $\widehat{\Omega}_1 = \frac{1}{n}\sum_{i=1}^n Z_i Z_i' \hat{e}_{1i}^2$, $\hat{e}_{1i} = y_i - X_i'\hat{\theta}_1$. Also define $\hat{e}_{2i} = y_i - X_i'\hat{\theta}_2$ and the $n \times 1$ residual vector $\hat{e}_j = Y - X\hat{\theta}_j$ for $j = 1, 2$.

The doubly corrected variance estimators of the 2SLS and two-step GMM are

$$\widehat{V}_{dc}(\hat{\theta}_1) = \left(\frac{1}{n}X'Z(Z'Z)^{-1}Z'X\right)^{-1}\frac{1}{n}\sum_{i=1}^n \hat{m}_{1i}\hat{m}_{1i}'\left(\frac{1}{n}X'Z(Z'Z)^{-1}Z'X\right)^{-1},$$

$$\widehat{V}_{dc}(\hat{\theta}_2) = \widehat{V}(\hat{\theta}_2) + \widehat{D}_n\widehat{C}(\hat{\theta}_1, \hat{\theta}_2) + \widehat{C}(\hat{\theta}_1, \hat{\theta}_2)'\widehat{D}_n' + \widehat{D}_n\widehat{V}_{dc}(\hat{\theta}_1)\widehat{D}_n',$$

where

$$\widehat{V}(\hat{\theta}_2) = \left(\frac{1}{n^2}X'Z\widehat{\Omega}_1^{-1}Z'X\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \hat{m}_{2i}\hat{m}_{2i}'\right)\left(\frac{1}{n^2}X'Z\widehat{\Omega}_1^{-1}Z'X\right)^{-1},$$

$$\widehat{C}(\hat{\theta}_1, \hat{\theta}_2) = \left(\frac{1}{n}X'Z(Z'Z)^{-1}Z'X\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \hat{m}_{1i}\hat{m}_{2i}'\right)\left(\frac{1}{n^2}X'Z\widehat{\Omega}_1^{-1}Z'X\right)^{-1},$$

$$\widehat{D}_n = \frac{2}{n}\left(X'Z\widehat{\Omega}_1^{-1}Z'X\right)^{-1}X'Z\widehat{\Omega}_1^{-1}\sum_{i=1}^n Z_i\left(\hat{e}_{1i}Z_i'\widehat{\Omega}_1^{-1}Z'\hat{e}_2\right)X_i',$$

$$\hat{m}_{1i} = X'Z(Z'Z)^{-1}Z_i\hat{e}_{1i} + X_iZ_i'(Z'Z)^{-1}Z'\hat{e}_1 - X'Z(Z'Z)^{-1}Z_iZ_i'(Z'Z)^{-1}Z'\hat{e}_1,$$

$$\hat{m}_{2i} = \frac{1}{n}X'Z\widehat{\Omega}_1^{-1}Z_i\hat{e}_{2i} + \frac{1}{n}X_iZ_i'\widehat{\Omega}_1^{-1}Z'\hat{e}_2 - \frac{1}{n^2}X'Z\widehat{\Omega}_1^{-1}Z_iZ_i'\hat{e}_{1i}^2\widehat{\Omega}_1^{-1}Z'\hat{e}_2.$$

It is worth observing that the doubly corrected variance estimator $\widehat{V}_{dc}(\hat{\theta}_2)$ reduces to the Windmeijer corrected one $\widehat{V}_w(\hat{\theta}_2)$ if (i) the last two terms in $\hat{m}_{2i}$ and $\hat{m}_{1i}$ are ignored and (ii) $\hat{e}_{1i}$ replaces $\hat{e}_{2i}$ in $\hat{m}_{2i}$. By (i) and (ii), the variance estimators $\widehat{V}(\hat{\theta}_2)$ and $\widehat{V}_{dc}(\hat{\theta}_1)$ reduce to conventional ones $\widetilde{V}(\hat{\theta}_2)$ and $\widetilde{V}(\hat{\theta}_1)$, and $\widehat{C}(\hat{\theta}_1, \hat{\theta}_2)$ becomes $\widetilde{V}(\hat{\theta}_2)$. In general, however, $\widehat{V}_{dc}(\hat{\theta}_2) \neq \widehat{V}_w(\hat{\theta}_2)$ because $Z'\hat{e}_j \neq 0$ for $j = 1, 2$, so the last two terms of $\hat{m}_{ji}$ are non-zero. Furthermore, it is critical (and reasonable) to use $\hat{e}_{2i}$ in $\hat{m}_{2i}$ to get robustness under misspecification.

The iterated GMM estimator is obtained as follows. Let $\hat{\theta}_0$ be any initial value. The $s$-step GMM estimator for $s \geq 1$ is given by

$$\hat{\theta}_s = (X'Z\widehat{\Omega}_{s-1}^{-1}Z'X)^{-1}X'Z\widehat{\Omega}_{s-1}^{-1}Z'Y,$$

where

$$\widehat{\Omega}_{s-1} = \frac{1}{n}\sum_{i=1}^n Z_iZ_i'(y_i - X_i'\hat{\theta}_{s-1})^2.$$

We iterate the $s$-step GMM estimator until convergence given a preset tolerance $\epsilon$, i.e. $\|\hat{\theta}_s - \hat{\theta}_{s-1}\| < \epsilon$ to obtain the iterated GMM estimator $\hat{\theta}$. The residuals are $\hat{e}_i = y_i - X_i'\hat{\theta}$. Also let $\hat{e} = Y - X\hat{\theta}$ be the $n \times 1$ residual vector.

The doubly corrected variance estimator is

$$\widehat{V}_{dc}(\hat{\theta}) = \widehat{H}^{-1}\left(\frac{1}{n}\sum_{i=1}^n \hat{m}_i\hat{m}_i'\right)\widehat{H}^{-1\prime},$$

$$\widehat{H} = \frac{1}{n^2}X'Z\widehat{\Omega}^{-1}Z'X - \frac{2}{n^3}X'Z\widehat{\Omega}^{-1}\sum_{i=1}^n Z_i\left(\hat{e}_iZ_i'\widehat{\Omega}^{-1}Z'\hat{e}\right)X_i',$$

$$\hat{m}_i = \frac{1}{n}X'Z\widehat{\Omega}^{-1}Z_i\hat{e}_i + \frac{1}{n}X_iZ_i'\widehat{\Omega}^{-1}Z'\hat{e} - \frac{1}{n^2}X'Z\widehat{\Omega}^{-1}Z_iZ_i'\hat{e}_i^2\widehat{\Omega}^{-1}Z'\hat{e}.$$

In comparison, the Windmeijer corrected and the conventional variance estimators are

$$\widehat{V}_w(\hat{\theta}) = \widehat{H}^{-1}\left(\frac{1}{n^2}X'Z\widehat{\Omega}^{-1}Z'X\right)\widehat{H}^{-1\prime},$$

$$\widetilde{V}(\hat{\theta}) = \left(\frac{1}{n^2}X'Z\widehat{\Omega}^{-1}Z'X\right)^{-1}.$$

### 6.2. A panel data model

Consider a panel data model with a scalar regressor

$$y_{it} = x_{it}\beta + \eta_i + v_{it}, \tag{34}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ where $\eta_i$ is the unobserved individual effects, the unknown parameter of interest is $\beta$, and the single regressor $x_{it}$ is predetermined with respect to $v_{it}$ (possibly including lags of the dependent variable), i.e., $E(x_{it} v_{is}) = 0$ for all $s \geq t$. After first-differencing,

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta v_{it}, \quad t = 2, \ldots, T,$$

the standard approach to estimate $\beta$ is first differenced GMM (Arellano and Bond (1991) estimator) with the moment conditions $E(Z_i' \Delta v_i) = 0$ where $Z_i$ is the $(T-1)$ by $T(T-1)/2$ instrument matrix

$$Z_i = diag(z_{i2}', \ldots, z_{iT}')$$

with all possible lagged instruments $z_{it} = (x_{i1}, \ldots, x_{it-1})'$ for $2 \leq t \leq T$ and $\Delta v_i = (\Delta v_{i2}, \ldots, \Delta v_{iT})'$. The total number of observations is $n = N(T-1)$.

Our doubly corrected variance estimator can be used for the model (34) with additional strictly exogenous, predetermined, or endogenous variables as well as the system GMM estimator (Arellano and Bover (1995) and Blundell and Bond (1998)) by stacking and modifying additional moment conditions into the instrument sets $Z_i$. If the panel is unbalanced, the instrument matrix can be constructed as described in Arellano and Bond (1991).

Using the initial weight matrix $\widehat{W} = n^{-1} \sum_{i=1}^{N} Z_i' H Z_i$, where $H$ is a matrix with 2's on the main diagonal, $-1$'s on the first off-diagonals and zeros elsewhere, the one-step GMM estimator is given by

$$\hat{\beta}_1 = (\Delta X' Z \widehat{W}^{-1} Z' \Delta X)^{-1} \Delta X' Z \widehat{W}^{-1} Z' \Delta Y,$$

where $Z = (Z_1', \ldots, Z_N')'$ is the instrument matrix, $\Delta Y = (\Delta y_1', \ldots, \Delta y_N')'$, $\Delta X = (\Delta x_1', \ldots, \Delta x_N')'$, $\Delta y_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})'$, and $\Delta x_i = (\Delta x_{i2}, \ldots, \Delta x_{iT})'$. Note that scaling the weight matrix does not affect the estimator. The doubly corrected variance estimator of $\hat{\beta}_1$ is given by

$$\widehat{V}_{dc}(\hat{\beta}_1) = n^2 \left(\Delta X' Z \widehat{W}^{-1} Z' \Delta X\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{N} \hat{m}_{1i} \hat{m}_{1i}'\right) \left(\Delta X' Z \widehat{W}^{-1} Z' \Delta X\right)^{-1},$$

$$\hat{m}_{1i} = \Delta X' Z \widehat{W}^{-1} Z_i' \Delta \hat{v}_{1i} + \Delta x_i' Z_i \widehat{W}^{-1} Z' \Delta \hat{v}_1 - \frac{1}{n} \Delta X' Z \widehat{W}^{-1} Z_i' H Z_i \widehat{W}^{-1} Z' \Delta \hat{v}_1,$$

where $\Delta \hat{v}_{1i} = \Delta y_i - \Delta x_i \hat{\beta}_1$ and $\Delta \hat{v}_1 = (\Delta \hat{v}_{11}', \ldots, \Delta \hat{v}_{1N}')'$. The doubly corrected standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{dc}(\hat{\beta}_1)/n}$. In comparison, the conventional variance estimator is given by

$$\widetilde{V}(\hat{\beta}_1) = n^2 \left(\Delta X' Z \widehat{W}^{-1} Z' \Delta X\right)^{-1} \Delta X' Z \widehat{W}^{-1} \widehat{\Omega}_1 \widehat{W}^{-1} Z' \Delta X \left(\Delta X' Z \widehat{W}^{-1} Z' \Delta X\right)^{-1},$$

where

$$\widehat{\Omega}_1 = \frac{1}{n} \sum_{i=1}^{N} Z_i' \Delta \hat{v}_{1i} \Delta \hat{v}_{1i}' Z_i.$$

Next, consider the two-step efficient GMM estimator

$$\hat{\beta}_2 = (\Delta X' Z \widehat{\Omega}_1^{-1} Z' \Delta X)^{-1} \Delta X' Z \widehat{\Omega}_1^{-1} Z' \Delta Y.$$

Let $\Delta \hat{v}_{2i} = \Delta y_i - \Delta x_i \hat{\beta}_2$ and $\Delta \hat{v}_2 = (\Delta \hat{v}_{21}', \ldots, \Delta \hat{v}_{2N}')'$. The doubly corrected variance estimator of $\hat{\beta}_2$ is given by

$$\widehat{V}_{dc}(\hat{\beta}_2) = \widehat{V}(\hat{\beta}_2) + \widehat{D}_n \widehat{C}(\hat{\beta}_1, \hat{\beta}_2) + \widehat{C}(\hat{\beta}_1, \hat{\beta}_2)' \widehat{D}_n' + \widehat{D}_n \widehat{V}_{dc}(\hat{\beta}_1) \widehat{D}_n',$$

where

$$\widehat{V}(\hat{\beta}_2) = n^2 \left(\Delta X' Z \widehat{\Omega}_1^{-1} Z' \Delta X\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{N} \hat{m}_{2i} \hat{m}_{2i}'\right) \left(\Delta X' Z \widehat{\Omega}_1^{-1} Z' \Delta X\right)^{-1},$$

$$\widehat{C}(\hat{\beta}_1, \hat{\beta}_2) = n^2 \left(\Delta X' Z \widehat{W}^{-1} Z' \Delta X\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{N} \hat{m}_{1i} \hat{m}_{2i}'\right) \left(\Delta X' Z \widehat{\Omega}_1^{-1} Z' \Delta X\right)^{-1},$$

$$\hat{m}_{2i} = \Delta X' Z \widehat{\Omega}_1^{-1} Z_i' \Delta \hat{v}_{2i} + \Delta x_i' Z_i \widehat{\Omega}_1^{-1} Z' \Delta \hat{v}_2 - \frac{1}{n} \Delta X' Z \widehat{\Omega}_1^{-1} Z_i' \Delta \hat{v}_{1i} \Delta \hat{v}_{1i}' Z_i \widehat{\Omega}_1^{-1} Z' \Delta \hat{v}_2,$$

$$\widehat{D}_n = \left(\Delta X' Z \widehat{\Omega}_1^{-1} Z' \Delta X\right)^{-1} \Delta X' Z \widehat{\Omega}_1^{-1} \times \frac{1}{n} \sum_{i=1}^{N} \left(Z_i' \Delta x_i \left(\Delta \hat{v}_2' Z \widehat{\Omega}_1^{-1} Z_i' \Delta \hat{v}_{1i}\right) + \left(Z_i' \Delta \hat{v}_{1i}\right) \left(\Delta \hat{v}_2' Z \widehat{\Omega}_1^{-1} Z_i' \Delta x_i\right)\right).$$

The doubly corrected standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{dc}(\hat{\beta}_2)/n}$. Note that the Windmeijer corrected variance estimator is

$$\widehat{V}_w(\hat{\beta}_2) = \widetilde{V}(\hat{\beta}_2) + \widehat{D}_n\widetilde{V}(\hat{\beta}_2) + \widetilde{V}(\hat{\beta}_2)\widehat{D}_n' + \widehat{D}_n\widetilde{V}(\hat{\beta}_1)\widehat{D}_n',$$

where

$$\widetilde{V}(\hat{\beta}_2) = n^2\left(\Delta X'Z\widehat{\Omega}_1^{-1}Z'\Delta X\right)^{-1}.$$

Finally, the iterated GMM estimator is given as follows. Let $\hat{\beta}_0$ be any initial value. The $s$-step GMM estimator for $s \geq 1$ is given by

$$\hat{\beta}_s = (\Delta X'Z\widehat{\Omega}_{s-1}^{-1}Z'\Delta X)^{-1}\Delta X'Z\widehat{\Omega}_{s-1}^{-1}Z'\Delta Y,$$

where

$$\widehat{\Omega}_{s-1} = \frac{1}{n}\sum_{i=1}^{N} Z_i'(\Delta y_i - \Delta x_i\hat{\beta}_{s-1})(\Delta y_i - \Delta x_i\hat{\beta}_{s-1})'Z_i.$$

We iterate the $s$-step GMM estimator until convergence given a preset tolerance $\epsilon$, i.e. $\|\hat{\beta}_s - \hat{\beta}_{s-1}\| < \epsilon$ to obtain the iterated GMM estimator $\hat{\beta}$. The residuals are $\Delta\hat{v}_i = \Delta y_i - \Delta x_i\hat{\beta}$. Also let $\Delta\hat{v} = (\Delta\hat{v}_1', \ldots, \Delta\hat{v}_N')'$ be the $n \times 1$ residual vector.

The doubly corrected variance estimator for iterated GMM is given by

$$\begin{aligned}
\widehat{V}_{dc}(\hat{\beta}) =& \widehat{H}^{-1}\left(\frac{1}{n}\sum_{i=1}^{N}\hat{m}_i\hat{m}_i'\right)\widehat{H}^{-1'},\\
\widehat{H} =& \frac{1}{n^2}\Delta X'Z\widehat{\Omega}^{-1}Z'\Delta X\\
&- \frac{1}{n^3}\Delta X'Z\widehat{\Omega}^{-1}\left(\sum_{i=1}^{N}\left(Z_i'\Delta\hat{v}_i\right)\left(\Delta\hat{v}'Z\widehat{\Omega}^{-1}Z_i'\Delta x_i\right) + Z_i'\Delta x_i\left(\Delta\hat{v}'Z\widehat{\Omega}^{-1}Z_i'\Delta\hat{v}_i\right)\right),\\
\hat{m}_i =& \frac{1}{n}\Delta X'Z\widehat{\Omega}^{-1}Z_i'\Delta\hat{v}_i + \frac{1}{n}\Delta X_i'Z_i\widehat{\Omega}^{-1}Z'\Delta\hat{v} - \frac{1}{n^2}\Delta X'Z\widehat{\Omega}^{-1}Z_i'\Delta\hat{v}_i\Delta\hat{v}_i'Z_i\widehat{\Omega}^{-1}Z'\Delta\hat{v},
\end{aligned}$$

and the doubly corrected standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{dc}(\hat{\beta})/n}$.

In comparison, the Windmeijer corrected and the conventional variance estimators are

$$\begin{aligned}
\widehat{V}_w(\hat{\beta}) &= \widehat{H}^{-1}\left(\frac{1}{n^2}\Delta X'Z\widehat{\Omega}^{-1}Z'\Delta X\right)\widehat{H}^{-1'},\\
\widetilde{V}(\hat{\beta}) &= \left(\frac{1}{n^2}\Delta X'Z\widehat{\Omega}^{-1}Z'\Delta X\right)^{-1}.
\end{aligned}$$

## 7. Simulation

We investigate the finite sample performance of the doubly corrected standard errors proposed in this paper and provide a thorough comparison with the conventional and the Windmeijer corrected ones under correct specification and misspecification. We consider three different setups: (i) a cross-sectional linear IV model with potentially invalid instruments; (ii) a linear dynamic panel model with a random coefficient; (iii) a linear dynamic panel model with possibly misspecified lag specifications. The number of Monte Carlo simulation is 100,000.

In an unreported simulation, we also investigate the performance of the estimators with the centered weight matrix (30). Since the results are similar and there is no obvious pattern of better performance of the point and variance estimators based on the centered weight matrix, they are not reported.

### 7.1. Cross-sectional IV

We use the following simulation design which is a simple linear instrumental variable regression with a single endogenous regressor. The model to be estimated is

$$\begin{aligned}
y_i &= x_i\beta_0 + e_i,\\
E(z_ie_i) &= 0
\end{aligned} \tag{35}$$

where $x_i$ and $\beta_0$ are scalar and $z_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i})'$ is a vector of instrumental variables. We estimate $\beta_0$ by 2SLS (one-step), two-step, and iterated GMM, and calculate the conventional, the Windmeijer corrected, and the doubly corrected standard errors. Our data-generating process (DGP) is

$$\begin{aligned} y_i &= x_i\beta_0 + e_i, \\ x_i &= \pi_0 (z_{1i} + z_{2i} + z_{3i} + z_{4i}) + u_i, \\ e_i &= \frac{\alpha_0}{\sqrt{n}} (z_{1i} - z_{2i} + z_{3i} - z_{4i}) + 0.5u_i + \sqrt{1 - 0.5^2} v_i, \\ z_i &\sim N(0, I_4), \begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & z_{1i}^2 \end{bmatrix} \right). \end{aligned}$$

We set $\beta_0 = 1$, vary $\alpha_0$ from 0 to 1 in steps of 0.2, and set the first-stage coefficient $\pi_0$ so that the first-stage $R^2 = 0.2$. We set the number of observations as $n = 50, 100, 500$.

The parameter $\alpha_0$ is the extent that the exclusion condition is locally violated. At $\alpha_0 = 0$, the model is correctly specified. For $\alpha_0 \neq 0$, we find $E(z_i e_i) = (\alpha_0, -\alpha_0, \alpha_0, -\alpha_0)'/\sqrt{n} \neq 0$, so the moment condition (35) fails to hold in finite samples, but it holds asymptotically.

Means and standard deviations of one-step (2SLS), two-step, and iterated GMM are computed in Table 1. For all GMM estimators, we report means of the conventional standard errors (se $\hat{\beta}$), the Windmeijer corrected standard errors (se$_w$ $\hat{\beta}$), and the doubly corrected standard errors (se$_{dc}$ $\hat{\beta}$).

Table 1 shows that our doubly corrected standard errors remain accurate regardless of misspecification, including correct specification case ($\alpha_0 = 0$); the means of corrected standard errors are very close to the standard deviations for all values of $\alpha_0$, especially for two-step and iterated GMM. Simulation evidence reassures our theory that the doubly corrected standard errors not only take into account variation in the estimation of the weight matrix but also extra variation due to the non-zero sample moments in the over-identified model even under correct specification. Furthermore, our doubly corrected standard errors are the only valid one under misspecification.

The conventional standard error for the one-step GMM (2SLS) estimator is downward biased under correct specification ($\alpha_0 = 0$), and this bias increases with $\alpha_0$. As is well known, the conventional standard error for the two-step GMM estimator is severely downward biased when $\alpha_0 = 0$, and this bias also increases with $\alpha_0$. The Windmeijer corrected standard error works well under correct specification but does not fully account for additional variations when $\alpha_0$ is non-zero. The result is similar for iterated GMM.

It is worth noting that the two-step and iterated GMM point estimates are sensitive to the local violation parameter $\alpha_0$. Interestingly, the point estimate becomes similar to the true value as the degree of misspecification $\alpha_0$ increases. This is because, in our DGP, the local misspecification bias (which depends on $\alpha_0$), and the higher-order asymptotic bias (which does not depend on $\alpha_0$) have opposite signs so that it happens to cancel out each other as $\alpha_0$ increases. In contrast, the one-step GMM point estimate varies little across $\alpha_0$ because the local misspecification bias is zero.[5] This is specific to this DGP and cannot be generalized. The size of bias in the point estimate decreases as the sample size increases.

### 7.2. Linear dynamic panel model

#### 7.2.1. Random coefficient

We next explore the finite sample performance of the doubly corrected standard error in the presence of heterogeneous effects (random coefficient) in dynamic panel model. We consider the AR(1) dynamic panel model of Blundell and Bond (1998). For $i = 1, \ldots, N$ and $t = 1, \ldots, T$,

$$y_{it} = \rho_0 y_{i,t-1} + \eta_i + \nu_{it}, \tag{36}$$

where $\eta_i$ is an unobserved individual-specific effect and $\nu_{it}$ is an error term. The parameter of interest $\rho_0$ is estimated by the difference GMM based on a set of moment conditions:

$$E[y_{i,t-s}(\Delta y_{it} - \rho_0 \Delta y_{i,t-1})] = 0, \quad t = 3, \ldots, T, \text{ and } s \geq 2, \tag{37}$$

The moment conditions are derived from taking differences of (36), and uses the lagged values of $y_{it}$ as instruments. The number of moment conditions is $(T-1)(T-2)/2$.

The moment conditions are correctly specified if there is a unique parameter that satisfies (37). A sufficient condition for this to hold is that the model (36) coincides with the true DGP, but this is unlikely to be true. A reasonable deviation from the assumed model (36) is heterogeneity in $\rho_0$ across $i$. We assume the following DGP. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$,

$$\begin{aligned} y_{it} &= \rho_i y_{i,t-1} + \eta_i + \nu_{it}, \\ \eta_i &\sim N(0, 1); \ \rho_i \sim \Phi(\alpha_0 \eta_i); \ \nu_{it} \sim N(0, 0.5^2), \end{aligned}$$

[5] The local misspecification bias and the higher-order asymptotic bias can be calculated using the formula in Theorem 2 and 3 in Hwang et al. (2020).

**Table 1**
Monte Carlo results for linear IV: $n = 50, 100, 500$.

| | $\alpha_0$ | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 |
|---|---|---|---|---|---|---|---|
| $n = 50$ | $\hat{\beta}_1$ | 1.0833 | 1.0859 | 1.0823 | 1.0833 | 1.0840 | 1.0811 |
| | sd $\hat{\beta}_1$ | 0.3229 | 0.3226 | 0.3295 | 0.3417 | 0.3503 | 0.3746 |
| | se $\hat{\beta}_1$ | 0.2962 | 0.2963 | 0.2985 | 0.3022 | 0.3047 | 0.3113 |
| | $\text{se}_{dc}$ $\hat{\beta}_1$ | 0.3346 | 0.3363 | 0.3417 | 0.3516 | 0.3619 | 0.3794 |
| | $\hat{\beta}_2$ | 1.0736 | 1.0656 | 1.0517 | 1.0421 | 1.0344 | 1.0243 |
| | sd $\hat{\beta}_2$ | 0.3029 | 0.3035 | 0.3142 | 0.3293 | 0.3470 | 0.3753 |
| | se $\hat{\beta}_2$ | 0.2544 | 0.2549 | 0.2575 | 0.2616 | 0.2646 | 0.2720 |
| | $\text{se}_w$ $\hat{\beta}_2$ | 0.2889 | 0.2900 | 0.2945 | 0.3039 | 0.3133 | 0.3281 |
| | $\text{se}_{dc}$ $\hat{\beta}_2$ | 0.3101 | 0.3144 | 0.3231 | 0.3398 | 0.3570 | 0.3813 |
| | $\hat{\beta}_{iter}$ | 1.0778 | 1.0672 | 1.0519 | 1.0390 | 1.0290 | 1.0154 |
| | sd $\hat{\beta}_{iter}$ | 0.3026 | 0.3028 | 0.3147 | 0.3312 | 0.3517 | 0.3827 |
| | se $\hat{\beta}_{iter}$ | 0.2513 | 0.2531 | 0.2573 | 0.2631 | 0.2680 | 0.2766 |
| | $\text{se}_w$ $\hat{\beta}_{iter}$ | 0.2850 | 0.2859 | 0.2919 | 0.3028 | 0.3140 | 0.3316 |
| | $\text{se}_{dc}$ $\hat{\beta}_{iter}$ | 0.3069 | 0.3086 | 0.3176 | 0.3340 | 0.3500 | 0.3742 |
| $n = 100$ | $\hat{\beta}_1$ | 1.0411 | 1.0408 | 1.0413 | 1.0411 | 1.0402 | 1.0412 |
| | sd $\hat{\beta}_1$ | 0.2326 | 0.2315 | 0.2337 | 0.2373 | 0.2420 | 0.2477 |
| | se $\hat{\beta}_1$ | 0.2212 | 0.2212 | 0.2218 | 0.2229 | 0.2240 | 0.2259 |
| | $\text{se}_{dc}$ $\hat{\beta}_1$ | 0.2354 | 0.2359 | 0.2380 | 0.2414 | 0.2458 | 0.2519 |
| | $\hat{\beta}_2$ | 1.0353 | 1.0238 | 1.0133 | 1.0041 | 0.9940 | 0.9860 |
| | sd $\hat{\beta}_2$ | 0.2153 | 0.2138 | 0.2187 | 0.2239 | 0.2316 | 0.2400 |
| | se $\hat{\beta}_2$ | 0.1956 | 0.1957 | 0.1964 | 0.1977 | 0.1991 | 0.2010 |
| | $\text{se}_w$ $\hat{\beta}_2$ | 0.2089 | 0.2087 | 0.2099 | 0.2130 | 0.2169 | 0.2221 |
| | $\text{se}_{dc}$ $\hat{\beta}_2$ | 0.2135 | 0.2143 | 0.2179 | 0.2239 | 0.2315 | 0.2408 |
| | $\hat{\beta}_{iter}$ | 1.0386 | 1.0260 | 1.0145 | 1.0044 | 0.9931 | 0.9836 |
| | sd $\hat{\beta}_{iter}$ | 0.2143 | 0.2126 | 0.2175 | 0.2228 | 0.2311 | 0.2398 |
| | se $\hat{\beta}_{iter}$ | 0.1946 | 0.1958 | 0.1978 | 0.2000 | 0.2024 | 0.2053 |
| | $\text{se}_w$ $\hat{\beta}_{iter}$ | 0.2073 | 0.2079 | 0.2101 | 0.2140 | 0.2187 | 0.2248 |
| | $\text{se}_{dc}$ $\hat{\beta}_{iter}$ | 0.2123 | 0.2129 | 0.2164 | 0.2226 | 0.2298 | 0.2392 |
| $n = 500$ | $\hat{\beta}_1$ | 1.0081 | 1.0080 | 1.0085 | 1.0080 | 1.0082 | 1.0085 |
| | sd $\hat{\beta}_1$ | 0.1044 | 0.1048 | 0.1047 | 0.1050 | 0.1056 | 0.1061 |
| | se $\hat{\beta}_1$ | 0.1035 | 0.1036 | 0.1036 | 0.1038 | 0.1037 | 0.1038 |
| | $\text{se}_{dc}$ $\hat{\beta}_1$ | 0.1048 | 0.1049 | 0.1050 | 0.1055 | 0.1057 | 0.1062 |
| | $\hat{\beta}_2$ | 1.0066 | 1.0005 | 0.9949 | 0.9885 | 0.9828 | 0.9778 |
| | sd $\hat{\beta}_2$ | 0.0962 | 0.0966 | 0.0969 | 0.0970 | 0.0981 | 0.0989 |
| | se $\hat{\beta}_2$ | 0.0946 | 0.0946 | 0.0946 | 0.0948 | 0.0948 | 0.0949 |
| | $\text{se}_w$ $\hat{\beta}_2$ | 0.0958 | 0.0957 | 0.0956 | 0.0958 | 0.0958 | 0.0962 |
| | $\text{se}_{dc}$ $\hat{\beta}_2$ | 0.0955 | 0.0956 | 0.0958 | 0.0964 | 0.0970 | 0.0979 |
| | $\hat{\beta}_{iter}$ | 1.0074 | 1.0012 | 0.9955 | 0.9891 | 0.9833 | 0.9782 |
| | sd $\hat{\beta}_{iter}$ | 0.0960 | 0.0964 | 0.0966 | 0.0968 | 0.0977 | 0.0985 |
| | se $\hat{\beta}_{iter}$ | 0.0945 | 0.0949 | 0.0951 | 0.0956 | 0.0959 | 0.0962 |
| | $\text{se}_w$ $\hat{\beta}_{iter}$ | 0.0957 | 0.0958 | 0.0959 | 0.0964 | 0.0966 | 0.0972 |
| | $\text{se}_{dc}$ $\hat{\beta}_{iter}$ | 0.0954 | 0.0954 | 0.0956 | 0.0962 | 0.0967 | 0.0975 |

$$y_{i1} = \frac{\eta_i}{1-\rho_i} + u_{i1};\ u_{i1} \sim N\left(0, \frac{1}{1-\rho_i^2}\right),$$

where $\Phi(z)$ is the standard normal cdf. At $\alpha_0 = 0$, the model is correctly specified and $\rho_i = \rho_0 = 0.5$. For $\alpha_0 \neq 0$, the effective moment condition model can be written as

$$\begin{aligned} E[y_{i,t-s}(\Delta y_{it} - \rho \Delta y_{i,t-1})] &= E[y_{i,t-s}(\Delta v_{it} + (\rho_i - \rho)\Delta y_{i,t-1})] \\ &= E[\rho_i y_{i,t-s}\Delta y_{i,t-1}] - \rho(\gamma_{s-1} - \gamma_{s-2}), \end{aligned}$$

where $\gamma_j$ is the $j$th autocovariance. The last equation becomes zero at $\rho = E[\rho_i]$ if $\rho_i$ is independent of the $\{y_{it}\}$ process. If this is the case, then the moment condition model is correctly specified and the estimand is $E[\rho_i]$. Otherwise in general, the moment condition model fails to hold at a single unique parameter value because each of the moment condition

**Table 2**
Monte Carlo results for linear dynamic panel: $N = 100$ and $T = 4, 6$.

| | $\alpha_0$ | 0 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 |
|---|---|---|---|---|---|---|---|---|
| $N = 100$ | $\hat{\rho}_1$ | 0.4256 | 0.4356 | 0.4571 | 0.4925 | 0.5397 | 0.5926 | 0.6473 |
| $T = 4$ | sd $\hat{\rho}_1$ | 0.3324 | 0.3376 | 0.3268 | 0.3173 | 0.3107 | 0.3154 | 0.2763 |
| | se $\hat{\rho}_1$ | 0.3211 | 0.3215 | 0.3130 | 0.3023 | 0.2899 | 0.2775 | 0.2544 |
| | $\text{se}_{dc}$ $\hat{\rho}_1$ | 0.3515 | 0.3520 | 0.3431 | 0.3327 | 0.3196 | 0.3081 | 0.2815 |
| | $\hat{\rho}_2$ | 0.4256 | 0.4350 | 0.4554 | 0.4894 | 0.5351 | 0.5876 | 0.6420 |
| | sd $\hat{\rho}_2$ | 0.3502 | 0.3552 | 0.3443 | 0.3362 | 0.3327 | 0.3259 | 0.2966 |
| | se $\hat{\rho}_2$ | 0.3113 | 0.3115 | 0.3032 | 0.2929 | 0.2806 | 0.2680 | 0.2466 |
| | $\text{se}_w$ $\hat{\rho}_2$ | 0.3376 | 0.3377 | 0.3300 | 0.3208 | 0.3087 | 0.2961 | 0.2734 |
| | $\text{se}_{dc}$ $\hat{\rho}_2$ | 0.3684 | 0.3673 | 0.3594 | 0.3521 | 0.3375 | 0.3258 | 0.3027 |
| | $\hat{\rho}$ | 0.4182 | 0.4276 | 0.4467 | 0.4790 | 0.5234 | 0.5736 | 0.6267 |
| | sd $\hat{\rho}$ | 0.3656 | 0.3702 | 0.3619 | 0.3574 | 0.3568 | 0.3551 | 0.3324 |
| | se $\hat{\rho}$ | 0.3123 | 0.3122 | 0.3038 | 0.2938 | 0.2819 | 0.2688 | 0.2482 |
| | $\text{se}_w$ $\hat{\rho}$ | 0.3483 | 0.3489 | 0.3427 | 0.3350 | 0.3243 | 0.3133 | 0.2909 |
| | $\text{se}_{dc}$ $\hat{\rho}$ | 0.3756 | 0.3773 | 0.3695 | 0.3619 | 0.3495 | 0.3398 | 0.3122 |
| $N = 100$ | $\hat{\rho}_1$ | 0.4234 | 0.4272 | 0.4398 | 0.4626 | 0.4992 | 0.5460 | 0.6008 |
| $T = 6$ | sd $\hat{\rho}_1$ | 0.1469 | 0.1471 | 0.1468 | 0.1480 | 0.1493 | 0.1503 | 0.1477 |
| | se $\hat{\rho}_1$ | 0.1458 | 0.1455 | 0.1441 | 0.1418 | 0.1377 | 0.1308 | 0.1221 |
| | $\text{se}_{dc}$ $\hat{\rho}_1$ | 0.1537 | 0.1540 | 0.1542 | 0.1546 | 0.1545 | 0.1509 | 0.1441 |
| | $\hat{\rho}_2$ | 0.4217 | 0.4249 | 0.4363 | 0.4570 | 0.4909 | 0.5351 | 0.5891 |
| | sd $\hat{\rho}_2$ | 0.1630 | 0.1640 | 0.1650 | 0.1667 | 0.1704 | 0.1727 | 0.1708 |
| | se $\hat{\rho}_2$ | 0.1327 | 0.1324 | 0.1310 | 0.1284 | 0.1242 | 0.1175 | 0.1094 |
| | $\text{se}_w$ $\hat{\rho}_2$ | 0.1635 | 0.1634 | 0.1631 | 0.1626 | 0.1611 | 0.1567 | 0.1493 |
| | $\text{se}_{dc}$ $\hat{\rho}_2$ | 0.1628 | 0.1634 | 0.1646 | 0.1668 | 0.1693 | 0.1687 | 0.1643 |
| | $\hat{\rho}$ | 0.4167 | 0.4194 | 0.4294 | 0.4476 | 0.4762 | 0.5137 | 0.5606 |
| | sd $\hat{\rho}$ | 0.1782 | 0.1799 | 0.1831 | 0.1872 | 0.1972 | 0.2065 | 0.2127 |
| | se $\hat{\rho}$ | 0.1328 | 0.1325 | 0.1312 | 0.1289 | 0.1250 | 0.1191 | 0.1119 |
| | $\text{se}_w$ $\hat{\rho}$ | 0.1778 | 0.1783 | 0.1794 | 0.1822 | 0.1858 | 0.1887 | 0.1887 |
| | $\text{se}_{dc}$ $\hat{\rho}$ | 0.1773 | 0.1784 | 0.1806 | 0.1855 | 0.1916 | 0.1965 | 0.1975 |

imposes a restriction

$$\rho = \frac{E[\rho_i y_{i,t-s} \Delta y_{i,t-1}]}{\gamma_{s-1} - \gamma_{s-2}},$$

but there is no reason that this should hold at a unique $\rho$ for $s = 2, 3, \ldots, t-1$. In the DGP, $\eta_i$ and $\rho_i$ are dependent through $\alpha_0$ and a larger $\alpha_0$ leads to larger heterogeneity. We vary $\alpha_0$ from 0 to 0.3 in steps of 0.05. The pseudo-true value would depend on the instrument set and the value of $\alpha_0$ under global misspecification. However, by varying $\alpha_0$ by a small amount we try to capture local behavior of the standard errors when the pseudo-true value is close to the true value. The sample sizes are $N = 100, 500$ and $T = 4, 6$.

We report the simulation results in Tables 2 and 3, which are qualitatively similar to the IV setup. Tables 2 and 3 show that the doubly corrected standard errors approximate the standard deviation of the GMM estimators well regardless of misspecification. For two-step and iterated GMM, the doubly corrected standard errors are as accurate as the Windmeijer correction for small values of $\alpha_0$ (including correct specification $\alpha_0 = 0$) but dominate the other in terms of accuracy for larger values of $\alpha_0$. The doubly corrected standard error for one-step GMM is slightly upward biased for small values of $\alpha_0$, but this bias decreases with a larger sample size $N = 500$.

*7.2.2. Misspecified lag length*

We use the baseline linear panel model of Windmeijer (2005) allowing for possible lag length misspecification. The model is

$$y_{it} = \beta_0 x_{it} + \eta_i + v_{it}, \tag{38}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The unknown parameter of interest is $\beta_0$, and the regressor $x_{it}$ is predetermined with respect to $v_{it}$, i.e., $E(x_{it} v_{it+s}) = 0$ for $s = 0, \ldots, T-t$. We use the first differenced GMM estimator. The number of moment conditions is $T(T-1)/2$ as in Section 6.2.

The DGP is

$$\begin{aligned} y_{it} &= \beta_0 x_{it} + \alpha_0 x_{it-1} + \eta_i + v_{it}, \\ x_{it} &= 0.5 x_{it-1} + \eta_i + 0.5 v_{it-1} + \epsilon_{it}, \\ \eta_i &\sim N(0,1), \quad \epsilon_{it} \sim N(0,1), \end{aligned} \tag{39}$$

**Table 3**
Monte Carlo results for linear dynamic panel: $N = 500$ and $T = 4, 6$.

| | $\alpha_0$ | 0 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 |
|---|---|---|---|---|---|---|---|---|
| $N = 500$ | $\hat{\rho}_1$ | 0.4879 | 0.4938 | 0.5170 | 0.5531 | 0.5990 | 0.6502 | 0.7016 |
| $T = 4$ | sd $\hat{\rho}_1$ | 0.1379 | 0.1363 | 0.1330 | 0.1284 | 0.1220 | 0.1131 | 0.1044 |
| | se $\hat{\rho}_1$ | 0.1369 | 0.1356 | 0.1322 | 0.1265 | 0.1192 | 0.1107 | 0.1020 |
| | $se_{dc}$ $\hat{\rho}_1$ | 0.1389 | 0.1377 | 0.1347 | 0.1292 | 0.1222 | 0.1136 | 0.1047 |
| | $\hat{\rho}_2$ | 0.4893 | 0.4950 | 0.5180 | 0.5539 | 0.6001 | 0.6517 | 0.7034 |
| | sd $\hat{\rho}_2$ | 0.1400 | 0.1385 | 0.1357 | 0.1314 | 0.1254 | 0.1164 | 0.1075 |
| | se $\hat{\rho}_2$ | 0.1361 | 0.1348 | 0.1315 | 0.1258 | 0.1186 | 0.1101 | 0.1014 |
| | $se_w$ $\hat{\rho}_2$ | 0.1389 | 0.1377 | 0.1349 | 0.1295 | 0.1225 | 0.1138 | 0.1046 |
| | $se_{dc}$ $\hat{\rho}_2$ | 0.1404 | 0.1393 | 0.1368 | 0.1318 | 0.1251 | 0.1165 | 0.1072 |
| | $\hat{\rho}$ | 0.4891 | 0.4948 | 0.5178 | 0.5537 | 0.5999 | 0.6515 | 0.7032 |
| | sd $\hat{\rho}$ | 0.1403 | 0.1389 | 0.1362 | 0.1319 | 0.1260 | 0.1170 | 0.1080 |
| | se $\hat{\rho}$ | 0.1362 | 0.1349 | 0.1315 | 0.1259 | 0.1187 | 0.1102 | 0.1016 |
| | $se_w$ $\hat{\rho}$ | 0.1393 | 0.1382 | 0.1354 | 0.1301 | 0.1232 | 0.1145 | 0.1052 |
| | $se_{dc}$ $\hat{\rho}$ | 0.1408 | 0.1397 | 0.1373 | 0.1323 | 0.1256 | 0.1170 | 0.1077 |
| $N = 500$ | $\hat{\rho}_1$ | 0.4835 | 0.4869 | 0.4997 | 0.5237 | 0.5621 | 0.6130 | 0.6687 |
| $T = 6$ | sd $\hat{\rho}_1$ | 0.0691 | 0.0691 | 0.0689 | 0.0681 | 0.0676 | 0.0654 | 0.0615 |
| | se $\hat{\rho}_1$ | 0.0690 | 0.0688 | 0.0680 | 0.0664 | 0.0637 | 0.0595 | 0.0544 |
| | $se_{dc}$ $\hat{\rho}_1$ | 0.0698 | 0.0697 | 0.0695 | 0.0691 | 0.0681 | 0.0656 | 0.0611 |
| | $\hat{\rho}_2$ | 0.4842 | 0.4875 | 0.4998 | 0.5227 | 0.5595 | 0.6089 | 0.6639 |
| | sd $\hat{\rho}_2$ | 0.0712 | 0.0714 | 0.0718 | 0.0723 | 0.0735 | 0.0726 | 0.0689 |
| | se $\hat{\rho}_2$ | 0.0677 | 0.0675 | 0.0667 | 0.0651 | 0.0623 | 0.0581 | 0.0530 |
| | $se_w$ $\hat{\rho}_2$ | 0.0711 | 0.0711 | 0.0710 | 0.0708 | 0.0701 | 0.0678 | 0.0634 |
| | $se_{dc}$ $\hat{\rho}_2$ | 0.0708 | 0.0709 | 0.0713 | 0.0722 | 0.0730 | 0.0720 | 0.0682 |
| | $\hat{\rho}$ | 0.4841 | 0.4874 | 0.4996 | 0.5223 | 0.5585 | 0.6066 | 0.6603 |
| | sd $\hat{\rho}$ | 0.0715 | 0.0718 | 0.0723 | 0.0732 | 0.0752 | 0.0757 | 0.0737 |
| | se $\hat{\rho}$ | 0.0677 | 0.0675 | 0.0667 | 0.0651 | 0.0624 | 0.0583 | 0.0534 |
| | $se_w$ $\hat{\rho}$ | 0.0715 | 0.0715 | 0.0715 | 0.0717 | 0.0720 | 0.0713 | 0.0685 |
| | $se_{dc}$ $\hat{\rho}$ | 0.0712 | 0.0713 | 0.0718 | 0.0730 | 0.0746 | 0.0749 | 0.0725 |

$$v_{it} = \delta_i \tau_t \omega_{it}, \quad \omega_{it} \sim \chi^2_1 - 1,$$
$$\delta_i \sim \text{Uniform}[0.5, 1.5], \quad \tau_t = 0.5 + 0.1(t-1).$$

We generate initial 50 time periods with $\tau_t = 0.5$ for $t = -49, \ldots, 0$ and $x_{i,-49} \sim N(\eta_i/0.5, 1/0.75)$ same as Windmeijer (2005). The parameter $\alpha_0$ in (39) governs the degree of misspecification. When $\alpha_0 = 0$, the model (38) is correctly specified which reduces to that of Windmeijer (2005).

The model (38) is misspecified for $\alpha_0 \neq 0$. We discuss the pseudo-true value and its interpretation. Since $0 = E[x_{is}\Delta v_{it}]$, the effective moment condition can be written as

$$0 = E[x_{is}(\Delta y_{it} - \beta^* \Delta x_{it})] = (\beta_0 - \beta^*)E[x_{is}\Delta x_{it}] + \alpha_0 E[x_{is}\Delta x_{it-1}] \tag{40}$$

for all $1 \leq s < t$ and $2 \leq t \leq T$.

First, consider $T = 2$ where there is only one moment condition so that the model is just-identified. By setting $s = 1$ and $t = 2$, the unique solution to (40) is $\beta^* = \beta_0 - \alpha_0$. Since (40) is equal to zero at $\beta^*$, the moment condition is not misspecified, although the lag is misspecified. Since the model is just-identified $\beta^* = \beta_0 - \alpha_0$ is considered as the true value but this is not equal to $\beta_0$ unless $\alpha_0 = 0$. An implication is that model misspecification, and pseudo-true values are not only pertinent to over-identified models.

For $T \geq 3$, the model is overidentified and each of the moment condition holds at

$$\beta^*_{t,s} = \beta_0 + \alpha_0 \frac{E[x_{i,t-1}x_{is}] - E[x_{i,t-2}x_{is}]}{E[x_{it}x_{is}] - E[x_{i,t-1}x_{is}]}.$$

Note that $\beta^*_{t,s}$ can vary across $t$ and $s$ for a nonzero $\alpha_0$. For example, $\beta^*_{3,1} = \beta_0 + 2\alpha_0$ and $\beta^*_{3,2} = \beta_0 - \alpha_0$. The pseudo-true value $\beta^*$ is a weighted average of $\beta^*_{t,s}$'s given $T$ and the weight matrix. One can interpret that $\beta^*$ represents the causal effects of past and present values of $x_{it}$'s on $y_{it}$ in the true DGP.

Tables 4 and 5 report estimation results for $\beta_0 = 1$, $N = 100, 500$ and $T = 4, 6$. The degree of misspecification $\alpha_0$ is varied across $\{0, 0.05, 0.1, 0.2, 0.3\}$. The first column ($\alpha_0 = 0$) in Table 4 replicates Monte Carlo studies in Windmeijer (2005), Table 1). Tables 4 and 5 also report the rejection rates of the $J$ test with the nominal size of 5%. Although the rejection rate increases as the degree of misspecification increases and the sample size increase, the $J$ test tends to not correctly detect the violation of the overidentifying moment restrictions in small samples.

The implication of the results in Tables 4 and 5 are largely unchanged as in two previous simulation experiments; doubly corrected standard errors approximate the standard deviations well regardless of model misspecification. In this

**Table 4**
Monte Carlo results for linear panel model: $N = 100$ and $T = 4, 6$.

| | $\alpha_0$ | 0 | 0.05 | 0.1 | 0.2 | 0.3 |
|---|---|---|---|---|---|---|
| $N = 100$ | $\hat{\beta}_1$ | 0.9793 | 0.9268 | 0.8744 | 0.7702 | 0.6647 |
| $T = 4$ | sd $\hat{\beta}_1$ | 0.1521 | 0.1538 | 0.1588 | 0.1727 | 0.1925 |
| | se $\hat{\beta}_1$ | 0.1469 | 0.1465 | 0.1472 | 0.1513 | 0.1589 |
| | $\text{se}_{dc}$ $\hat{\beta}_1$ | 0.1546 | 0.1563 | 0.1604 | 0.1744 | 0.1944 |
| | $J$ test | 0.0779 | 0.0836 | 0.1127 | 0.2899 | 0.6127 |
| | $\hat{\beta}_2$ | 0.9849 | 0.9322 | 0.8773 | 0.7587 | 0.6238 |
| | sd $\hat{\beta}_2$ | 0.1404 | 0.1451 | 0.1552 | 0.1843 | 0.2207 |
| | se $\hat{\beta}_2$ | 0.1243 | 0.1242 | 0.1253 | 0.1303 | 0.1381 |
| | $\text{se}_{w}$ $\hat{\beta}_2$ | 0.1390 | 0.1415 | 0.1473 | 0.1669 | 0.1919 |
| | $\text{se}_{dc}$ $\hat{\beta}_2$ | 0.1343 | 0.1390 | 0.1482 | 0.1775 | 0.2146 |
| | $J$ test | 0.0301 | 0.0391 | 0.0767 | 0.2605 | 0.5404 |
| | $\hat{\beta}$ | 0.9858 | 0.9334 | 0.8781 | 0.7533 | 0.5977 |
| | sd $\hat{\beta}$ | 0.1417 | 0.1474 | 0.1600 | 0.2000 | 0.2524 |
| | se $\hat{\beta}$ | 0.1243 | 0.1242 | 0.1253 | 0.1303 | 0.1381 |
| | $\text{se}_{w}$ $\hat{\beta}$ | 0.1393 | 0.1426 | 0.1507 | 0.1806 | 0.2230 |
| | $\text{se}_{dc}$ $\hat{\beta}$ | 0.1352 | 0.1406 | 0.1517 | 0.1896 | 0.2391 |
| | $J$ test | 0.0295 | 0.0377 | 0.0727 | 0.2429 | 0.4983 |
| $N = 100$ | $\hat{\beta}_1$ | 0.9755 | 0.9411 | 0.9060 | 0.8368 | 0.7676 |
| $T = 6$ | sd $\hat{\beta}_1$ | 0.1027 | 0.1051 | 0.1083 | 0.1172 | 0.1288 |
| | se $\hat{\beta}_1$ | 0.1002 | 0.1004 | 0.1013 | 0.1037 | 0.1077 |
| | $\text{se}_{dc}$ $\hat{\beta}_1$ | 0.1056 | 0.1075 | 0.1107 | 0.1192 | 0.1306 |
| | $J$ test | 0.3465 | 0.3533 | 0.4157 | 0.6786 | 0.9239 |
| | $\hat{\beta}_2$ | 0.9833 | 0.9466 | 0.9080 | 0.8238 | 0.7318 |
| | sd $\hat{\beta}_2$ | 0.0906 | 0.0948 | 0.1017 | 0.1213 | 0.1431 |
| | se $\hat{\beta}_2$ | 0.0716 | 0.0720 | 0.0731 | 0.0760 | 0.0801 |
| | $\text{se}_{w}$ $\hat{\beta}_2$ | 0.0905 | 0.0930 | 0.0978 | 0.1117 | 0.1285 |
| | $\text{se}_{dc}$ $\hat{\beta}_2$ | 0.0836 | 0.0876 | 0.0944 | 0.1131 | 0.1357 |
| | $J$ test | 0.0205 | 0.0277 | 0.0582 | 0.2762 | 0.6453 |
| | $\hat{\beta}$ | 0.9857 | 0.9484 | 0.9083 | 0.8124 | 0.6885 |
| | sd $\hat{\beta}$ | 0.0946 | 0.0998 | 0.1099 | 0.1427 | 0.1858 |
| | se $\hat{\beta}$ | 0.0716 | 0.0720 | 0.0731 | 0.0760 | 0.0801 |
| | $\text{se}_{w}$ $\hat{\beta}$ | 0.0937 | 0.0976 | 0.1054 | 0.1320 | 0.1722 |
| | $\text{se}_{dc}$ $\hat{\beta}$ | 0.0866 | 0.0916 | 0.1006 | 0.1287 | 0.1675 |
| | $J$ test | 0.0199 | 0.0259 | 0.0551 | 0.2553 | 0.5980 |

simulation experiment, the Windmeijer correction works best under correct specification but becomes downward biased as $\alpha_0$ increases. Note that deviation from correct specification makes the bias of the conventional standard error and the Windmeijer corrected standard error larger, and this bias does not disappear with a larger sample size of $N = 500$.

### 7.3. Size

This section investigates the small sample performance of the $t$ tests with the proposed standard errors. We report the size of the $t$ tests under correct specification for the simulation setups considered in previous sections.

Based on the one-step, two-step, and iterated GMM estimators, Table 6 evaluates the size of the $t$ tests for nominal size 5% using various standard errors. In the column labeled $t$, we report the size of the test based on conventional heteroskedasticity-robust standard errors. $t_w$ is based on the Windmeijer corrected standard errors, while $t_{dc}$ is based on the doubly corrected standard errors in this paper. Further, we report the size of the bootstrap $t$ test using the doubly corrected standard errors ($t_{dc}$-$bs$), which is equivalent to the misspecification-robust (MR) bootstrap of Lee (2014). We calculate the bootstrap critical values based on 1000 bootstrap replications per Monte Carlo replication. Results for larger sample sizes ($n = 500$ in Tables 1, 3 and 5) are not reported here for brevity as the results are similar across different methods.

For Setups 1 and 2, tests using the conventional standard errors ($t$) are severely oversized. Using the Windmeijer corrected standard errors ($t_w$), and the doubly corrected standard errors ($t_{dc}$) for the two-step and iterated estimator improve the size and perform similarly, although both are moderately oversized. The size distortion decreases when we increase the sample size. Using the MR bootstrap with the doubly corrected standard errors ($t_{dc}$-$bs$) improves the size of the test dramatically. The excellent size property of the MR bootstrap $t$ test is theoretically justified by its asymptotic

**Table 5**
Monte Carlo results for linear panel model: $N = 500$ and $T = 4, 6$.

| | $\alpha_0$ | 0 | 0.05 | 0.1 | 0.2 | 0.3 |
|---|---|---|---|---|---|---|
| $N = 500$ | $\hat{\beta}_1$ | 0.9958 | 0.9406 | 0.8853 | 0.7754 | 0.6650 |
| $T = 4$ | sd $\hat{\beta}_1$ | 0.0685 | 0.0695 | 0.0716 | 0.0787 | 0.0889 |
| | se $\hat{\beta}_1$ | 0.0679 | 0.0677 | 0.0679 | 0.0698 | 0.0734 |
| | $\text{se}_{dc}$ $\hat{\beta}_1$ | 0.0686 | 0.0696 | 0.0716 | 0.0787 | 0.0889 |
| | $J$ test | 0.0732 | 0.1212 | 0.3934 | 0.9524 | 1.0000 |
| | $\hat{\beta}_2$ | 0.9970 | 0.9443 | 0.8892 | 0.7660 | 0.6225 |
| | sd $\hat{\beta}_2$ | 0.0652 | 0.0676 | 0.0730 | 0.0891 | 0.1092 |
| | se $\hat{\beta}_2$ | 0.0632 | 0.0630 | 0.0634 | 0.0657 | 0.0693 |
| | $\text{se}_w$ $\hat{\beta}_2$ | 0.0648 | 0.0662 | 0.0694 | 0.0798 | 0.0930 |
| | $\text{se}_{dc}$ $\hat{\beta}_2$ | 0.0634 | 0.0658 | 0.0708 | 0.0865 | 0.1060 |
| | $J$ test | 0.0447 | 0.1014 | 0.3824 | 0.9426 | 0.9996 |
| | $\hat{\beta}$ | 0.9970 | 0.9446 | 0.8897 | 0.7635 | 0.6014 |
| | sd $\hat{\beta}$ | 0.0652 | 0.0678 | 0.0738 | 0.0946 | 0.1238 |
| | se $\hat{\beta}$ | 0.0632 | 0.0630 | 0.0634 | 0.0657 | 0.0693 |
| | $\text{se}_w$ $\hat{\beta}$ | 0.0648 | 0.0662 | 0.0698 | 0.0839 | 0.1052 |
| | $\text{se}_{dc}$ $\hat{\beta}$ | 0.0634 | 0.0659 | 0.0715 | 0.0911 | 0.1185 |
| | $J$ test | 0.0445 | 0.1004 | 0.3822 | 0.9410 | 0.9995 |
| $N = 500$ | $\hat{\beta}_1$ | 0.9947 | 0.9564 | 0.9181 | 0.8423 | 0.7662 |
| $T = 6$ | sd $\hat{\beta}_1$ | 0.0473 | 0.0482 | 0.0498 | 0.0543 | 0.0600 |
| | se $\hat{\beta}_1$ | 0.0469 | 0.0470 | 0.0473 | 0.0485 | 0.0504 |
| | $\text{se}_{dc}$ $\hat{\beta}_1$ | 0.0475 | 0.0484 | 0.0499 | 0.0543 | 0.0602 |
| | $J$ test | 0.3557 | 0.4796 | 0.7820 | 0.9995 | 1.0000 |
| | $\hat{\beta}_2$ | 0.9968 | 0.9582 | 0.9175 | 0.8276 | 0.7251 |
| | sd $\hat{\beta}_2$ | 0.0432 | 0.0454 | 0.0494 | 0.0602 | 0.0725 |
| | se $\hat{\beta}_2$ | 0.0408 | 0.0410 | 0.0414 | 0.0428 | 0.0448 |
| | $\text{se}_w$ $\hat{\beta}_2$ | 0.0431 | 0.0445 | 0.0471 | 0.0548 | 0.0640 |
| | $\text{se}_{dc}$ $\hat{\beta}_2$ | 0.0413 | 0.0434 | 0.0470 | 0.0573 | 0.0694 |
| | $J$ test | 0.0388 | 0.1139 | 0.4500 | 0.9906 | 1.0000 |
| | $\hat{\beta}$ | 0.9969 | 0.9584 | 0.9174 | 0.8219 | 0.6961 |
| | sd $\hat{\beta}$ | 0.0433 | 0.0457 | 0.0504 | 0.0659 | 0.0882 |
| | se $\hat{\beta}$ | 0.0408 | 0.0410 | 0.0414 | 0.0428 | 0.0448 |
| | $\text{se}_w$ $\hat{\beta}$ | 0.0431 | 0.0447 | 0.0479 | 0.0592 | 0.0770 |
| | $\text{se}_{dc}$ $\hat{\beta}$ | 0.0414 | 0.0437 | 0.0479 | 0.0620 | 0.0823 |
| | $J$ test | 0.0387 | 0.1131 | 0.4489 | 0.9902 | 1.0000 |

refinements, which is formally shown by Lee (2014). Perhaps surprisingly, using the doubly corrected standard errors for the $t$ test and bootstrap $t$ test improves size properties considerably for the one-step estimator where the Windmeijer correction is not available.

For Setup 3, $t$ based on the one-step estimator and $t_w$ with the two-step estimator have good size properties, and the same results can be found in Windmeijer (2005, Figure 1). $t_{dc}$ has similar size properties to $t_w$, but is slightly undersized for the one-step estimator and slightly oversized for the two-step and iterated estimators. The MR bootstrap test is slightly undersized. Windmeijer (2005) and Bond and Windmeijer (2005) report similar results for the nonparametric bootstrap test of Hall and Horowitz (1996) based on the two-step estimator and explain that the performance of the bootstrap deteriorates with an increasing number of moment conditions. Since $t_{dc}$ becomes oversized with the number of moment conditions, this suggests that some components of the doubly corrected variance estimator may be sensitive to the number of moment conditions. This issue deserves further future investigation.

## 8. Conclusion

We propose doubly corrected standard errors for the over-identified one-step, two-step, and iterated GMM estimators in linear moment conditions. Our double correction considers the over-identification bias in variance estimation on top of the commonly used finite-sample correction of Windmeijer (2005), which corrects the bias from estimating the efficient weight matrix.

Monte Carlo simulations show that our doubly corrected formulas have a much more accurate variance estimator and result in size accuracy than conventional ones regardless of the model misspecification. Our formulas do not entail much extra computational cost because their main ingredients are traditional GMM estimators using original data. Thus we

**Table 6**
Finite sample sizes of the $t$-test (Nominal size 5%).

| | One-step | | | Two-step | | | | Iterated | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $t$ | $t_{dc}$ | $t_{dc}$-$bs$ | $t$ | $t_w$ | $t_{dc}$ | $t_{dc}$-$bs$ | $t$ | $t_w$ | $t_{dc}$ | $t_{dc}$-$bs$ |
| Setup 1: Cross-sectional IV (Section 7.1) | | | | | | | | | | | |
| $n = 50$ | 0.109 | 0.084 | 0.072 | 0.142 | 0.112 | 0.110 | 0.077 | 0.151 | 0.123 | 0.122 | 0.071 |
| $n = 100$ | 0.079 | 0.065 | 0.064 | 0.096 | 0.082 | 0.082 | 0.069 | 0.100 | 0.085 | 0.085 | 0.068 |
| Setup 2: Dynamic panel (Section 7.2.1) | | | | | | | | | | | |
| $N = 100, T = 4$ | 0.074 | 0.067 | 0.060 | 0.096 | 0.076 | 0.075 | 0.061 | 0.106 | 0.080 | 0.079 | 0.057 |
| $N = 100, T = 6$ | 0.089 | 0.082 | 0.069 | 0.154 | 0.082 | 0.091 | 0.065 | 0.178 | 0.078 | 0.090 | 0.055 |
| Setup 3: Dynamic panel (Section 7.2.2) | | | | | | | | | | | |
| $N = 100, T = 4$ | 0.047 | 0.039 | 0.030 | 0.073 | 0.047 | 0.056 | 0.027 | 0.078 | 0.049 | 0.058 | 0.025 |
| $N = 100, T = 6$ | 0.060 | 0.046 | 0.041 | 0.123 | 0.055 | 0.074 | 0.034 | 0.137 | 0.057 | 0.076 | 0.025 |

strongly recommend that practitioners use our formulas in linear GMM models, including linear IV and dynamic panel regression.

## Appendix. Proofs

In the proofs, "LLN" refers to the weak law of large numbers, "CLT" refers to the multivariate Lindeberg–Lévy central limit theorem (e.g. Theorem 6.3. of Hansen, 2020) and "CMT" refers to the continuous mapping theorem.

Lemma 1 is on geometric expansion of a matrix. It builds on Corollary 1 of Magdalinos (1992).

**Lemma 1.** *Let $X_n$ and $Y_n$ be square random matrices. If $X_n^{-1}$ and $(X_n + Y_n/\sqrt{n})^{-1}$ exist and $X_n^{-1}$ and $Y_n$ are of order $O_p(1)$, then following holds for any nonnegative integer $q$,*

$$\left(X_n + \frac{1}{\sqrt{n}}Y_n\right)^{-1} = \sum_{j=0}^{q}\left(-\frac{1}{\sqrt{n}}X_n^{-1}Y_n\right)^j X_n^{-1} + O_p(n^{-(q+1)/2}).$$

**Proof of Lemma 1.** Let $S_n = X_n^{-1}Y_n$ and consider the following identity,

$$\sum_{j=0}^{q}\left(-\frac{1}{\sqrt{n}}\right)^j S_n^j\left(I + \frac{1}{\sqrt{n}}S_n\right) = I - \left(-\frac{1}{\sqrt{n}}\right)^{q+1} S_n^{q+1}.$$

Using $I + n^{-1/2}S_n = X_n^{-1}(X_n + n^{-1/2}Y_n)$, $S_n = X_n^{-1}Y_n$ and rearranging terms, we have

$$\begin{aligned}\left(X_n + \frac{1}{\sqrt{n}}Y_n\right)^{-1} &= \sum_{j=0}^{q}\left(-\frac{1}{\sqrt{n}}\right)^j S_n^j X_n^{-1} + \left(-\frac{1}{\sqrt{n}}\right)^{q+1} S_n^{q+1}\left(X_n + \frac{1}{\sqrt{n}}Y_n\right)^{-1} \\ &= \sum_{j=0}^{q}\left(-\frac{1}{\sqrt{n}}\right)^j \left(X_n^{-1}Y_n\right)^j X_n^{-1} + \left(-\frac{1}{\sqrt{n}}\right)^{q+1} O_p(1)\end{aligned}$$

by the assumptions of the lemma. □

Let

$$\frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} = E\left[\frac{\partial \Omega_n(\theta)}{\partial \theta_{[j]}}\right], \quad \frac{\partial^2 \Omega(\theta)}{\partial \theta_{[j]}\partial \theta_{[l]}} = E\left[\frac{\partial^2 \Omega_n(\theta)}{\partial \theta_{[j]}\partial \theta_{[l]}}\right].$$

This is a slight abuse of notation because the LHS does not mean differentiation of the expectation. Lemma 2 establishes some useful convergence results.

**Lemma 2.** *Under Assumption 1, the followings hold:*

$$g_n(\theta_1) - g_1 = O_p(n^{-1/2}), \tag{41}$$
$$g_n(\theta_2) - g_2 = O_p(n^{-1/2}), \tag{42}$$
$$G_n - G = O_p(n^{-1/2}), \tag{43}$$
$$W_n - W = O_p(n^{-1/2}), \tag{44}$$
$$\Omega_n(\theta_1) - \Omega_1 = O_p(n^{-1/2}), \tag{45}$$

$$W_n^{-1} - W^{-1} = O_p(n^{-1/2}), \tag{46}$$
$$[\Omega_n(\theta_1)]^{-1} - [\Omega_1]^{-1} = O_p(n^{-1/2}), \tag{47}$$

*and for each j, lth element of* $\theta$,

$$\frac{\partial \Omega_n(\theta_1)}{\partial \theta_{[j]}} - \frac{\partial \Omega(\theta_1)}{\partial \theta_{[j]}} = O_p(n^{-1/2}), \tag{48}$$
$$\frac{\partial^2 \Omega_n(\theta_1)}{\partial \theta_{[j]}\partial \theta_{[l]}} - \frac{\partial^2 \Omega(\theta_1)}{\partial \theta_{[j]}\partial \theta_{[l]}} = O_p(n^{-1/2}), \tag{49}$$

*In addition, if* $\hat{\theta}_1 - \theta_1 = O_p(n^{-1/2})$, *then*

$$\Omega_n(\hat{\theta}_1) - \Omega_1 = O_p(n^{-1/2}), \tag{50}$$
$$[\Omega_n(\hat{\theta}_1)]^{-1} - [\Omega_1]^{-1} = O_p(n^{-1/2}). \tag{51}$$

**Proof of Lemma 2.** Under Assumption 1 (ii) and (v), (41)–(45) immediately follow by CLT.

Since $g(X_i, \theta)$ is linear in $\theta$,

$$\frac{\partial \Omega_n(\theta_1)}{\partial \theta_{[j]}} = \frac{1}{n}\sum_{i=1}^n g(X_i, \theta_1)\frac{\partial g(X_i, \theta_1)'}{\partial \theta_{[j]}} + \frac{1}{n}\sum_{i=1}^n \frac{\partial g(X_i, \theta_1)}{\partial \theta_{[j]}} g(X_i, \theta_1)',$$
$$\frac{\partial^2 \Omega_n(\theta_1)}{\partial \theta_{[j]}\partial \theta_{[l]}} = \frac{1}{n}\sum_{i=1}^n \frac{\partial g(X_i, \theta_1)}{\partial \theta_{[l]}}\frac{\partial g(X_i, \theta_1)'}{\partial \theta_{[j]}} + \frac{1}{n}\sum_{i=1}^n \frac{\partial g(X_i, \theta_1)}{\partial \theta_{[j]}}\frac{\partial g(X_i, \theta_1)'}{\partial \theta_{[l]}}.$$

Under Assumption 1 (ii) and (v), we apply CLT and the Cauchy–Schwarz inequality to have (48) and (49).

Next we show (50). Observing that the third derivative of $\Omega_n(\theta)$ with respect to $\theta$ equals zero, we use the second-order Taylor expansion to obtain

$$\begin{aligned}\Omega_n(\hat{\theta}_1) =& \Omega_1 + (\Omega_n(\theta_1) - \Omega_1) + \sum_j \frac{\partial \Omega(\theta_1)}{\partial \theta_{[j]}}(\hat{\theta}_{1[j]} - \theta_{1[j]}) + \sum_j \left(\frac{\partial \Omega_n(\theta_1)}{\partial \theta_{[j]}} - \frac{\partial \Omega(\theta_1)}{\partial \theta_{[j]}}\right)(\hat{\theta}_{1[j]} - \theta_{1[j]}) \\ &+ \frac{1}{2}\sum_j\sum_l \frac{\partial^2 \Omega(\theta_1)}{\partial \theta_{[j]}\partial \theta_{[l]}}(\hat{\theta}_{1[j]} - \theta_{1[j]})(\hat{\theta}_{1[l]} - \theta_{1[l]}) \\ &+ \frac{1}{2}\sum_j\sum_l \left(\frac{\partial^2 \Omega_n(\theta_1)}{\partial \theta_{[j]}\partial \theta_{[l]}} - \frac{\partial^2 \Omega(\theta_1)}{\partial \theta_{[j]}\partial \theta_{[l]}}\right)(\hat{\theta}_{1[j]} - \theta_{1[j]})(\hat{\theta}_{1[l]} - \theta_{1[l]}).\end{aligned}$$

By Assumption 1 (ii) and (v), the Cauchy–Schwarz inequality implies

$$\frac{\partial \Omega(\theta_1)}{\partial \theta_{[j]}} < \infty \quad \text{and} \quad \frac{\partial^2 \Omega(\theta_1)}{\partial \theta_{[j]}\partial \theta_{[l]}} < \infty. \tag{52}$$

Provided that $\hat{\theta}_1 - \theta_1 = O_p(n^{-1/2})$, (50) holds by (45), (48), (49), and (52).

Finally, applying Lemma 1 with $q = 0$ to (44), (45), and (50), we obtain (46), (47), and (51), respectively. □

**Proof of Theorem 1.** The proof consists of two parts. We first derive the asymptotic variance matrices of the one-step and two-step GMM estimators without assuming correct specification of the moment condition. Then we show that the doubly corrected variance estimators are consistent.

First, Lemma 2, CMT, and the Woodbury matrix identity together imply that

$$\begin{aligned}(G_n' W_n^{-1} G_n)^{-1} =& (G'W^{-1}G)^{-1}(I_k + (G_n' W_n^{-1} G_n - G'W^{-1}G)(G'W^{-1}G)^{-1})^{-1} \\ =& (G'W^{-1}G)^{-1}(I_k + o_p(1)),\end{aligned} \tag{53}$$

and similarly

$$(G_n'[\Omega_n(\theta_1)]^{-1} G_n)^{-1} = (G'\Omega_1^{-1}G)^{-1}(I_k + o_p(1)). \tag{54}$$

The FOC of the one-step GMM estimator is

$$0 = G_n' W_n^{-1} g_n(\hat{\theta}_1), \tag{55}$$

which holds with probability approaching zero by Assumption 1(i). By expanding (55) around the pseudo-true value $\theta_1$ and rearranging, we have

$$\begin{aligned}\hat{\theta}_1 - \theta_1 &= -(G_n' W_n^{-1} G_n)^{-1} G_n' W_n^{-1} g_n(\theta_1) \\ &= -(G' W^{-1} G)^{-1}(I_k + o_p(1)) G_n' W_n^{-1} g_n(\theta_1).\end{aligned}$$

By Lemma 2 and the population FOC $G'W^{-1}g_1 = 0$,

$$G_n' W_n^{-1} g_n(\theta_1) = G' W^{-1} g_n(\theta_1) + G_n' W^{-1} g_1 - G' W^{-1} W_n W^{-1} g_1 + O_p(n^{-1}).$$

Let

$$m_{1i} = G' W^{-1} g(X_i, \theta_1) + G(X_i)' W^{-1} g_1 - G' W^{-1} W(X_i) W^{-1} g_1.$$

Since $Em_{1i} = 0$, by CLT under Assumption 1(v)

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{1i} \xrightarrow{d} N(0, Em_{1i} m_{1i}'), \tag{56}$$

as $n \to \infty$. Now we can write

$$\begin{aligned}\sqrt{n}(\hat{\theta}_1 - \theta_1) &= -(G' W^{-1} G)^{-1}(I_k + o_p(1)) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{1i} + O_p(n^{-1/2}) \right) \\ &= -(G' W^{-1} G)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{1i} + o_p(1) \qquad (57) \\ &\xrightarrow{d} N(0, V_1), \qquad (58)\end{aligned}$$

as $n \to \infty$ where $V_1$ is the asymptotic variance matrix of one-step GMM under misspecification given by

$$V_1 = (G' W^{-1} G)^{-1} Em_{1i} m_{1i}' (G' W^{-1} G)^{-1}.$$

The FOC of the two-step GMM estimator is

$$0 = G_n' [\Omega_n(\hat{\theta}_1)]^{-1} g_n(\hat{\theta}_2).$$

By expanding $g_n(\hat{\theta}_2)$ around $\theta_2$ and then expanding $[\Omega_n(\hat{\theta}_1)]^{-1}$ around $\theta_1$,

$$\begin{aligned}\hat{\theta}_2 - \theta_2 &= -(G_n' [\Omega_n(\hat{\theta}_1)]^{-1} G_n)^{-1} G_n [\Omega_n(\hat{\theta}_1)]^{-1} g_n(\theta_2) \\ &= -(G_n' [\Omega_n(\theta_1)]^{-1} G_n)^{-1} G_n [\Omega_n(\theta_1)]^{-1} g_n(\theta_2) + D_n^*(\hat{\theta}_1 - \theta_1) + R_n^*, \qquad (59)\end{aligned}$$

where $D_n^* = F_{1n}^* + F_{2n}^*$ and $R_n^*$ is the remainder term.

First, $F_{1n}^* = o_p(1)$ by Lemma 2 and the population FOC $G' \Omega_1^{-1} g_2 = 0$. By Lemma 2, $F_{2n}^* = D^* + o_p(1)$, where

$$D^*[., j] = \left(G' \Omega_1^{-1} G\right)^{-1} G' \Omega_1^{-1} \left. \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \right|_{\theta = \theta_1} \Omega_1^{-1} g_2.$$

Hence we can write $D_n^* = D^* + o_p(1)$.

Next we show $R_n^* = O_p(n^{-1})$. Let

$$Q_n(\theta) = \left\{ G_n' [\Omega_n(\theta)]^{-1} G_n \right\}^{-1} G_n' [\Omega_n(\theta)]^{-1} g_n(\theta_2).$$

Since the third derivative of $\Omega_n(\theta)$ w.r.t. $\theta$ equals zero, the fifth derivative of $Q_n(\theta)$ w.r.t. $\theta$ equals zero by the chain rule. Thus,

$$\begin{aligned}R_n = &\frac{1}{2!} \sum_j \sum_l \frac{\partial^2 Q_n(\theta_1)}{\partial \theta_{[j]} \partial \theta_{[l]}} (\hat{\theta}_{1[j]} - \theta_{1[j]})(\hat{\theta}_{1[l]} - \theta_{1[l]}) \\ &+ \frac{1}{3!} \sum_j \sum_l \sum_m \frac{\partial^3 Q_n(\theta_1)}{\partial \theta_{[j]} \partial \theta_{[l]} \partial \theta_{[m]}} (\hat{\theta}_{1[j]} - \theta_{1[j]})(\hat{\theta}_{1[l]} - \theta_{1[l]})(\hat{\theta}_{1[m]} - \theta_{1[m]}) \\ &+ \frac{1}{4!} \sum_j \sum_l \sum_m \sum_s \frac{\partial^4 Q_n(\theta_1)}{\partial \theta_{[j]} \partial \theta_{[l]} \partial \theta_{[m]} \partial \theta_{[s]}} (\hat{\theta}_{1[j]} - \theta_{1[j]})(\hat{\theta}_{1[l]} - \theta_{1[l]})(\hat{\theta}_{1[m]} - \theta_{1[m]})(\hat{\theta}_{1[s]} - \theta_{1[s]}).\end{aligned}$$

By LLN under Assumption 1, $\frac{\partial^2 Q_n(\theta_1)}{\partial \theta_{[j]} \partial \theta_{[l]}} = O_p(1)$, $\frac{\partial^3 Q_n(\theta_1)}{\partial \theta_{[j]} \partial \theta_{[l]} \partial \theta_{[m]}} = O_p(1)$, and $\frac{\partial^4 Q_n(\theta_1)}{\partial \theta_{[j]} \partial \theta_{[l]} \partial \theta_{[m]} \partial \theta_{[s]}} = O_p(1)$. Since we have shown $\hat{\theta}_1 - \theta_1 = O_p(n^{-1/2})$, it follows that $R_n^* = O_p(n^{-1})$. Note that this also justifies the expansion (6) by setting $\theta_1 = \theta_2 = \theta_0$ and acknowledging that $g_n(\theta_0) = O_p(n^{-1/2})$ under correct specification.

By multiplying $\sqrt{n}$ on both sides (59) can now be written as

$$\sqrt{n}(\hat{\theta}_2 - \theta_2) = -(G'\Omega_1^{-1}G)^{-1}(I_k + o_p(1))G_n[\Omega_n(\theta_1)]^{-1}\sqrt{n}g_n(\theta_2) \tag{60}$$

$$+ D^*\sqrt{n}(\hat{\theta}_1 - \theta_1) + o_p(1). \tag{61}$$

Take (60). By Lemma 2 and the population FOC $G'\Omega_1^{-1}g_2 = 0$,

$$G_n[\Omega_n(\theta_1)]^{-1}g_n(\theta_2) = G'\Omega_1^{-1}g_n(\theta_2) + G_n'\Omega_1^{-1}g_2 - G'\Omega_1^{-1}\Omega_n(\theta_1)\Omega_1^{-1}g_2 + O_p(n^{-1}). \tag{62}$$

Let

$$m_{2i} = G'\Omega_1^{-1}g(X_i, \theta_2) + G(X_i)'\Omega_1^{-1}g_2 - G'\Omega_1^{-1}g(X_i, \theta_1)g(X_i, \theta_1)'\Omega_1^{-1}g_2.$$

Since $Em_{2i} = 0$, by CLT under Assumption 1(v)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} m_{2i} \\ m_{1i}\end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} Em_{2i}m_{2i}' & Em_{2i}m_{1i}' \\ Em_{1i}m_{2i}' & Em_{1i}m_{1i}'\end{pmatrix}\right),$$

as $n \to \infty$. Replacing (62) and (57) into (60) and (61), respectively, to obtain

$$\sqrt{n}(\hat{\theta}_2 - \theta_2) = -\begin{bmatrix}(G'\Omega_1^{-1}G)^{-1} & D^*(G'W^{-1}G)^{-1}\end{bmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} m_{2i} \\ m_{1i}\end{pmatrix} + o_p(1)$$

$$\xrightarrow{d} N(0, V_2), \tag{63}$$

as $n \to \infty$ where $V_2$ is the asymptotic variance matrix of two-step GMM under misspecification given by

$$\begin{aligned} V_2 =& (G'\Omega_1^{-1}G)^{-1}Em_{2i}m_{2i}'(G'\Omega_1^{-1}G)^{-1} + D^*(G'W^{-1}G)^{-1}Em_{1i}m_{2i}'(G'\Omega_1^{-1}G)^{-1} \\ &+ (G'\Omega_1^{-1}G)^{-1}Em_{2i}m_{1i}'(G'W^{-1}G)^{-1}D^{*'} + D^*(G'W^{-1}G)^{-1}Em_{1i}m_{1i}'(G'W^{-1}G)^{-1}D^{*'}. \end{aligned}$$

Since we have shown the asymptotic variance matrices of the GMM estimators robust to misspecification, the first part of the proof is complete.

Note that $m_{1i} = G'W^{-1}g(X_i, \theta_0)$, $m_{2i} = G'\Omega_0^{-1}g(X_i, \theta_0)$, and $D^* = 0$ under correct specification and $V_1$ and $V_2$ coincide to the conventional asymptotic variance matrices.

Next we show that $\widehat{V}_{dc}(\hat{\theta}_1) \xrightarrow{p} V_1$ and $\widehat{V}_{dc}(\hat{\theta}_2) \xrightarrow{p} V_2$. Since (53) and (54) hold, by the CMT it suffices to show

$$\frac{1}{n}\sum_{i=1}^{n} m_i(\hat{\theta}_2, \Omega_n(\hat{\theta}_1))m_i(\hat{\theta}_2, \Omega_n(\hat{\theta}_1))' \xrightarrow{p} Em_{2i}m_{2i}', \tag{64}$$

$$\frac{1}{n}\sum_{i=1}^{n} m_i(\hat{\theta}_1, W_n)m_i(\hat{\theta}_2, \Omega_n(\hat{\theta}_1))' \xrightarrow{p} Em_{1i}m_{2i}', \tag{65}$$

$$\frac{1}{n}\sum_{i=1}^{n} m_i(\hat{\theta}_1, W_n)m_i(\hat{\theta}_1, W_n)' \xrightarrow{p} Em_{1i}m_{1i}', \tag{66}$$

$$\widehat{D}_n \xrightarrow{p} D^*. \tag{67}$$

(64)–(66) follow if we show

$$m_i(\hat{\theta}_2, \Omega_n(\hat{\theta}_1)) = m_{2i} + O_p(n^{-1/2}), \tag{68}$$

$$m_i(\hat{\theta}_1, W_n) = m_{1i} + O_p(n^{-1/2}). \tag{69}$$

Since by Lemma 2, (58), and (63), for $j = 1, 2$,

$$\begin{aligned} g_n(\hat{\theta}_j) =& g_j + (g_n(\theta_j) - g_j) + G(\hat{\theta}_j - \theta_j) + (G_n - G)(\hat{\theta}_j - \theta_j) \\ =& g_j + O_p(n^{-1/2}) \end{aligned}$$

and

$$g(X_i, \hat{\theta}_j) = g(X_i, \theta_j) + G(X_i)(\hat{\theta}_j - \theta_j) = g(X_i, \theta_j) + O_p(n^{-1/2}).$$

Thus, (68) and (69) follow. To show (67), we use (42), (43), (48), (51) of Lemma 2. □

## References

Ai, C., Chen, X., 2007. Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. J. Econometrics 141 (1), 5–43.
Andrews, D.W.K., 2002. Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. Econometrica 70, 119–162.
Andrews, I., 2019. On the structure of IV estimands. J. Econometrics 211 (1), 294–307.
Andrews, I., Gentzkow, M., Shapiro, J.M., 2017. Measuring the sensitivity of parameter estimates to estimation moments. Q. J. Econ. 132 (4), 1553–1592.
Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econom. Stud. 58, 277–297.
Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. J. Econometrics 68, 29–51.
Armstrong, T., Kolesár, M., 2019. Sensitivity analysis using approximate moment condition models. arXiv preprint arXiv:1808.07387.
Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variables estimators. Econometrica 63, 657–681.
Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. J. Econometrics 87, 115–143.
Bond, S., Windmeijer, F., 2005. Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. Econometric Rev. 24 (1), 1–37.
Bonhomme, S., Weidner, M., 2018. Minimizing sensitivity to model misspecification. arXiv preprint arXiv:1807.02161.
Evdokimov, K., Kolesár, M., 2018. Inference in Instrumental Variable Regression Analysis with Heterogeneous Treatment Effects. Working Paper.
Guggenberger, P., 2012. On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Econometric Theory 28 (2), 387–421.
Hall, A.R., 2000. Covariance matrix estimation and the power of the overidentifying restrictions test. Econometrica 68 (6), 1517–1527.
Hall, P., Horowitz, J.L., 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64, 891–916.
Hall, A.R., Inoue, A., 2003. The large sample behavior of the generalized method of moments estimator in misspecified models. J. Econometrics 114 (2), 361–394.
Han, C., Phillips, P.C., 2006. GMM with many moment conditions. Econometrica 74 (1), 147–192.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.
Hansen, B.E., 2020. Econometrics. University of Wisconsin, Unpublished Manuscript.
Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. J. Bus. Econom. Statist. 14 (3), 262–280.
Hansen, B.E., Lee, S., 2020. Inference for iterated GMM under misspecification. Econometrica, Forthcoming.
Hwang, J., 2020. Simple and trustworthy cluster-robust GMM inference. J. Econometrics, Forthcoming.
Hwang, J., Kang, B., Lee, S., 2020. A doubly corrected robust variance estimator for linear GMM. arXiv working paper. https://arxiv.org/abs/1908.07821.
Hwang, J., Sun, Y., 2018. Should we go one step further? An accurate comparison of one-step and two-step procedures in a generalized method of moments framework. J. Econometrics 207 (2), 381–405.
Kang, B., 2018. Higher Order Approximation of IV Estimators with Invalid Instruments. Working Paper.
Kitamura, Y., Otsu, T., Evdokimov, K., 2013. Robustness, infinitesimal neighborhoods, and moment restrictions. Econometrica 81 (3), 1185–1201.
Lee, S., 2014. Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators. J. Econometrics 178, 398–413.
Lee, S., 2016. Asymptotic refinements of a misspecification-robust bootstrap for GEL estimators. J. Econometrics 192 (1), 86–104.
Lee, S., 2018. A consistent variance estimator for 2sls when instruments identify different LATEs. J. Bus. Econom. Statist. 36 (3), 400–410.
Maasoumi, E., Phillips, P.C., 1982. On the behavior of inconsistent instrumental variable estimators. J. Econometrics 19 (2), 183–201.
Magdalinos, M.A., 1992. Stochastic expansions and asymptotic approximations. Econometric Theory 8 (3), 343–367.
Newey, W.K., Smith, R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72 (1), 219–255.
Newey, W.K., Windmeijer, F., 2009. Generalized method of moments with many weak moment conditions. Econometrica 77, 687–719.
Otsu, T., 2011. Moderate deviations of generalized method of moments and empirical likelihood estimators. J. Multivariate Anal. 102 (8), 1203–1216.
Rotemberg, J.J., 1983. Instrumental Variable Estimation of Misspecified Models. Working Paper, MIT Sloan.
Rothenberg, T.J., 1984. Approximating the distributions of econometric estimators and test statistics. Handb. Econom. 2, 881–935.
Schennach, S.M., 2007. Point estimation with exponentially tilted empirical likelihood. Ann. Statist. 35 (2), 634–672.
Windmeijer, F., 2000. A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators. IFS Working Papers W00/19, Institute for Fiscal Studies.
Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. J. Econometrics 126 (1), 25–51.