Lancaster University Management School

Economics Working Paper Series

2019/016

A Doubly Corrected Robust Variance Estimator for Linear GMM

Jungbin Hwang, Byunghoon Kang and Seojeong Lee

LUMS home page: http://www.lancaster.ac.uk/lums/

A Doubly Corrected Robust Variance Estimator for Linear GMM

Jungbin Hwang[*]   Byunghoon Kang[†]
University of Connecticut   Lancaster University

Seojeong Lee[‡]
University of New South Wales

August 12, 2019[§]

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 additionally corrects for the over-identification bias in variance estimation on top of the commonly used finite sample correction of Windmeijer (2005) which corrects for the bias from estimating the efficient weight matrix, so is *doubly* corrected. Formal stochastic expansions are derived to show the proposed double correction estimates the variance of some higher-order terms in the expansion. In addition, the proposed double correction 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 proposed double correction formula provides a convenient way to obtain improved inference under correct specification and robustness against misspecification at the same time.

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 for the bias arising from using the efficient weight matrix being evaluated at an estimate, rather than the true value. The correction formula (the Windmeijer correction, hereinafter) has been routinely used in practice.[1]

However, the Windmeijer correction does not take into account for 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 while it converges in probability to zero under correct specification.

We propose a new finite sample correction which takes into account for 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.

We provide an additional formal justification of the proposed double correction by deriving the stochastic expansions of the GMM estimators under local misspecification where the moment condition is modeled as a local drifting sequence around zero. The result shows that the double correction estimates the (co)variance of some higher-order terms which increase with the over-identification bias. Rothenberg (1984) provides a general treatment of the stochastic expansion and Newey and Smith (2004) derive the stochastic expansion of GMM and the generalized empirical likelihood (GEL) estimators under correct specification.

The order of our double correction equals the order of the sample moment condition. Under correct specification or local misspecification, these terms are $O_p(n^{-1/2})$ so that the double correction is a finite-sample correction for the variance. However, the double correction terms become $O_p(1)$ when the moment condition is (globally) misspecified because the sample moment condition is $O_p(1)$. 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 coincides with the misspecification-robust variance estimator in Lee (2014) for the one-step and two-step GMM and B. Hansen and Lee (2019) for the iterated GMM. Indeed the simulation results reported in those papers show that the misspecification-robust variance estimator often performs better than the conventional variance estimator under correct specification. This paper provides an answer to this seemingly puzzling result by taking an alternative path to obtain the identical variance estimator formula.

From a practical point of view, the equivalence between the misspecification-robust formula and the doubly corrected formula implies that accurate inference under correct specification and

---

[1]More than 4,700 citations according to Google Scholar on Aug 5, 2019.

robust inference under misspecification can be achieved simultaneously, without knowing whether the moment condition is correctly specified or not. Moreover, it can be easily implemented to obtain more accurate t tests and confidence intervals (smaller errors in the size and the coverage) by bootstrapping the t statistic studentized by the doubly corrected variance estimator. 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 works 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. Even for nonlinear models, however, the doubly corrected variance estimator is still consistent for the asymptotic variance of the GMM estimator under misspecification. So the robustness property holds for both linear and nonlinear models.

Robust inference with possibly misspecified moment condition models has gained considerable attention in the literature. Maasoumi and Phillips (1982) investigate the limiting distribution of inconsistent instrumental variable (IV) estimators. Guggenberger (2012) studies the behavior of the weak instrument robust tests under local misspecification. Lee (2018) shows that the moment condition is misspecified under treatment effect heterogeneity and proposes a robust variance estimator for 2SLS. Hall and Inoue (2003) derive the asymptotic distribution of GMM under misspecification. Schennach (2007) proposes an alternative GEL-type estimator robust to global misspecification. Otsu (2011) analyses moderate deviation behaviors of GMM. Kitamura, Otsu, and Evdokimov (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 (2019) provide robust inference theory for the iterated GMM. Rotemberg (1983) and Andrews (2019) characterize the estimands of the linear GMM under misspecification. Andrews, Gentzkow, and Shapiro (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, Heaton, and Yaron (1996). Bond and Windmeijer (2005) provide simulation evidence on the finite sample performance of the asymptotic and bootstrap tests based on GMM estimators. Hwang and Sun (2018) develop alternative asymptotics and compare the finite sample properties of the one-step and two-step GMM estimators for dependent observations.

Our doubly corrected robust variance estimators are generally different than the many instrument and many weak instrument robust variance estimators for IV, GMM, and GEL estimators proposed by 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 derives the stochastic expansion of the one-step and two-step GMM estimators under local misspecification and shows that the double correction estimates the variance of some higher-order terms in the expansion. Section 5 shows that the doubly corrected variance estimator is misspecification-robust. Section 6 discusses the iterated GMM and the CU GMM. Section 7 derives the double correction formula for cross-sectional IV and the difference GMM. Finally Section 8 provides extensive simulation results comparing the double correction and other variance estimators. All the proofs are collected in the 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 $i = 1, ..., n$. Let $g(X_i, \theta)$ be a $q \times 1$ moment function where θ 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 zero and the analysis becomes trivial. The moment condition model is correctly specified if

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

for a unique θ_0. Assume $E[|g(X_i, \theta_0)|^2] < \infty$. Then by the Lindeberg-Lévy central limit theorem

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

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. We also assume standard regularity conditions for consistency and asymptotic normality of the GMM estimators, e.g., the conditions of Theorems 2.6 and 3.4 of Newey and McFadden (1994).

The one-step GMM estimator is defined as

$$\hat{\theta}_1 = \arg \min_{\theta \in \Theta} \ g_n(\theta)' W_n^{-1} g_n(\theta), \tag{3}$$

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 $W = E W_n$, a positive definite matrix of constants.

Taking $\hat{\theta}_1$ as a preliminary (initial) estimator, the two-step efficient GMM estimator is obtained by minimizing

$$\hat{\theta}_2 = \arg \min_{\theta \in \Theta} \ g_n(\theta)'[\Omega_n(\hat{\theta}_1)]^{-1} g_n(\theta), \tag{4}$$

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 is efficient.

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

$$\tilde{\theta}_2 = \arg\min_{\theta\in\Theta} \ g_n(\theta)'[\Omega_n(\theta_0)]^{-1}g_n(\theta). \tag{5}$$

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 θ due to linearity. Define $G_n = n^{-1}\sum_{i=1}^{n}G(X_i)$ and $G = EG_n$. By the first-order Taylor expansion, the first-order condition (FOC) of the (feasible) two-step GMM can be written as

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

and so we have

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) = -\left\{G_n'[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right\}^{-1}G_n'[\Omega_n(\hat{\theta}_1)]^{-1}\sqrt{n}g_n(\theta_0). \tag{6}$$

Using a similar expansion, we can get

$$\sqrt{n}(\tilde{\theta}_2 - \theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n}g_n(\theta_0), \tag{7}$$

for the infeasible two-step GMM and

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = -(G_n'W_n^{-1}G_n)^{-1}G_n'W_n^{-1}\sqrt{n}g_n(\theta_0) \tag{8}$$

for the one-step GMM.

Asymptotically (6) and (7) 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 θ_0 and using (8), Windmeijer (2005) shows that the extra finite sample variations caused by higher-order terms can be estimated and the accuracy of the variance estimate can be improved for linear moment condition models.

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

(6) around θ_0:

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) = -\left\{G'_n[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}G'_n[\Omega_n(\theta_0)]^{-1}\sqrt{n}g_n(\theta_0) + D_n\sqrt{n}(\hat{\theta}_1 - \theta_0) + O_p(n^{-1})$$

$$= \sqrt{n}(\tilde{\theta}_2 - \theta_0) + \underbrace{D_n\sqrt{n}(\hat{\theta}_1 - \theta_0)}_{=O_p(n^{-1/2})} + O_p(n^{-1}), \tag{9}$$

where

$$D_n = F_{1n} + F_{2n},$$

$$F_{1n} = -\left.\frac{\partial \left\{G'_n[\Omega_n(\theta)]^{-1}G_n\right\}^{-1}}{\partial\theta'}\right|_{\theta=\theta_0} G'_n[\Omega_n(\theta_0)]^{-1}g_n(\theta_0),$$

$$F_{2n} = -\left\{G'_n[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}\left.\frac{\partial G'_n[\Omega_n(\theta)]^{-1}g_n(\theta_0)}{\partial\theta'}\right|_{\theta=\theta_0}$$

are $k \times k$ matrices. Since $g_n(\theta_0) = O_p(n^{-1/2})$ both F_{1n} and F_{2n} are $O_p(n^{-1/2})$. Combining these with the assumption that $\sqrt{n}(\hat{\theta}_1 - \theta_0) = O_p(1)$, the second term in the above expansion (9) has the order of $O_p(n^{-1/2})$, which is of a lower order than the last term. 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 (9) only holds for linear moment condition models.

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

$$\widehat{V}_w(\hat{\theta}_2) = \widetilde{V}(\hat{\theta}_2) + \widehat{D}_n\widetilde{V}(\hat{\theta}_2) + \widetilde{V}(\hat{\theta}_2)\widehat{D}'_n + \widehat{D}_n\widetilde{V}(\hat{\theta}_1)\widehat{D}'_n, \tag{10}$$

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

$$\widetilde{V}(\hat{\theta}_1) = \left(G'_nW_n^{-1}G_n\right)^{-1}\left(G'_nW_n^{-1}\Omega_n(\hat{\theta}_1)W_n^{-1}G_n\right)\left(G'_nW_n^{-1}G_n\right)^{-1},$$

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

$$\widehat{D}_n[.,j] = \left\{G'_n[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right\}^{-1}G'_n\left\{[\Omega_n(\hat{\theta}_1)]^{-1}\left.\frac{\partial\Omega_n(\theta)}{\partial\theta_{[j]}}\right|_{\theta=\hat{\theta}_1}[\Omega_n(\hat{\theta}_1)]^{-1}\right\}g_n(\hat{\theta}_2),$$

$$\frac{\partial\Omega_n(\theta)}{\partial\theta_{[j]}} = \Upsilon_j(\theta) + \Upsilon'_j(\theta),$$

$$\Upsilon_j(\theta) = \frac{1}{n}\sum_{i=1}^{n}g_i(X_i, \theta)\frac{\partial g_i(X_i, \theta)'}{\partial\theta_{[j]}}.$$

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

6

3 Double Correction

The Windmeijer correction takes into account 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 θ, 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 (9). These additional terms are not considered in the Windmeijer correction (10). We propose alternative variance estimators that fully incorporate the additional variations induced by the over-identification bias. These variance estimators will replace $\widetilde{V}(\hat{\theta}_2)$ and $\widetilde{V}(\hat{\theta}_1)$ in (10) without affecting the order of finite sample corrections, leading to our doubly corrected variance estimator.

Assume that

$$G_n - G = O_p(n^{-1/2}), \tag{11}$$

$$\text{vec}(W_n - W) = O_p(n^{-1/2}), \tag{12}$$

$$\text{vec}(\Omega_n(\theta_0) - \Omega) = O_p(n^{-1/2}), \tag{13}$$

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

$$[\Omega_n(\theta_0)]^{-1} - \Omega^{-1} = -\Omega^{-1}(\Omega_n(\theta_0) - \Omega)[\Omega_n(\theta_0)]^{-1}, \tag{14}$$

we can write

$$G_n'[\Omega_n(\theta_0)]^{-1}g_n(\theta_0)$$
$$= \underbrace{G'\Omega^{-1}g_n(\theta_0)}_{=O_p(n^{-1/2})} + \underbrace{(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}(\Omega_n(\theta_0) - \Omega)\Omega^{-1}g_n(\theta_0)}_{=O_p(n^{-1})} + O_p(n^{-3/2}). \tag{15}$$

Using (15), the expansion of the infeasible two-step GMM (7) can be written as

$$\sqrt{n}(\tilde{\theta}_2 - \theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}\left[G'\Omega^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{16}$$
$$+\sqrt{n}(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}(\Omega_n(\theta_0) - \Omega)\Omega^{-1}g_n(\theta_0)\Big] \tag{17}$$
$$+ O_p(n^{-1}). \tag{18}$$

Similarly,

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = -\left\{G_n'W_n^{-1}G_n\right\}^{-1}\left[G'W^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{19}$$
$$+\sqrt{n}(G_n - G)'W^{-1}g_n(\theta_0) - G'W^{-1}\sqrt{n}(W_n - W)W^{-1}g_n(\theta_0)\Big] \tag{20}$$
$$+ O_p(n^{-1}), \tag{21}$$

7

which simplifies to

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = -\left\{G'_n G_n\right\}^{-1} \left[G'\sqrt{n}g_n(\theta_0) + \sqrt{n}(G_n - G)'g_n(\theta_0)\right] \tag{22}$$

when $W_n = I$. From the above expansions it is clear that (i) to take into account for the over-identification bias we need to consider the extra variations from $\sqrt{n}(G_n - G)$ and $\sqrt{n}(\Omega_n(\theta) - \Omega)$ (or $\sqrt{n}(W_n - W)$ for the one-step GMM), and (ii) the order of the remainder term of the original expansion (9) is not changed.

Using the expansions (16)-(18) an (19)-(21), the expansion of the two-step GMM can be written as

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) = -\left\{G'_n[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1} \left[G'\Omega^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{23}$$

$$+\sqrt{n}(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}\left(\Omega_n(\theta_0) - \Omega\right)\Omega^{-1}g_n(\theta_0)\right] \tag{24}$$

$$- D_n\left\{G'_n W_n^{-1}G_n\right\}^{-1}\left[G'W^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{25}$$

$$+\sqrt{n}(G_n - G)'W^{-1}g_n(\theta_0) - G'W^{-1}\sqrt{n}\left(W_n - W\right)W^{-1}g_n(\theta_0)\right] \tag{26}$$

$$+ O_p(n^{-1}). \tag{27}$$

In finite sample, $g_n(\theta_0) \neq 0$ because $g_n(\theta) \neq 0$ for all θ, and this causes extra variations through the terms in (24) and (26). 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 (26) are $O_p(n^{-1})$ multiplied by D_n, which is the same order with the remainder term. Thus, considering those terms in (26) does not necessarily provide finite sample corrections. However, including these terms are critical to get robustness to misspecification, which is shown in Section 5.

The expansion for the one-step GMM is (19)-(21) and those terms in (20) are $O_p(n^{-1/2})$. Thus, by considering the finite sample variation caused by these terms provide more accurate inference. This correction for the one-step GMM is not considered in Windmeijer (2005) and is new.

The relationship between the expansions in (19)-(21) and (23)-(27), and their respective higher-order stochastic expansions will be explained in the next Section.

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

$$\widehat{V}_{dc}(\hat{\theta}_2) = \widehat{V}(\hat{\theta}_2) + \widehat{D}_n\widehat{C}(\hat{\theta}_1, \hat{\theta}_2) + \widehat{C}(\hat{\theta}_1, \hat{\theta}_2)'\widehat{D}'_n + \widehat{D}_n\widehat{V}_{dc}(\hat{\theta}_1)\widehat{D}'_n, \tag{28}$$

where

$$\widehat{V}(\hat{\theta}_2) = \left(G'_n[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1} \Sigma_n(\hat{\theta}_2, \Omega_n(\hat{\theta}_1)) \left(G'_n[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1}, \tag{29}$$

$$\widehat{V}_{dc}(\hat{\theta}_1) = \left(G'_n W_n^{-1}G_n\right)^{-1} \Sigma_n(\hat{\theta}_1, W_n) \left(G'_n W_n^{-1}G_n\right)^{-1}, \tag{30}$$

$$\widehat{C}(\hat{\theta}_1, \hat{\theta}_2) = \left(G'_n W_n^{-1}G_n\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} m_i(\hat{\theta}_1, W_n)m_i(\hat{\theta}_2, \Omega_n(\hat{\theta}_1))' \left(G'_n[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1}, \tag{31}$$

and

$$\Sigma_n(\theta, \Xi_n(\phi)) = \frac{1}{n} \sum_{i=1}^n m_i(\theta, \Xi_n(\phi)) m_i(\theta, \Xi_n(\phi))', \tag{32}$$

$$m_i(\theta, \Xi_n(\phi)) = G_n'[\Xi_n(\phi)]^{-1} g(X_i, \theta) + G(X_i)'[\Xi_n(\phi)]^{-1} g_n(\theta)$$
$$- G_n'[\Xi_n(\phi)]^{-1} \Xi(X_i, \phi)[\Xi_n(\phi)]^{-1} g_n(\theta),$$

$$\Xi_n(\phi) = \frac{1}{n} \sum_{i=1}^n \Xi(X_i, \phi).$$

When $\Xi_n(\phi) = \Xi(X_i, \phi) = I$, the last term of $m(\theta, \Xi_n(\phi))$ drops. Note that $m_i(\theta, \Xi_n(\phi))$ does not have to include the centered processes for $G(X_i)$ and $\Xi(X_i, \phi)$ 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, $\widehat{V}_{dc}(\hat{\theta}_1)$, takes into account for the variations up to the order of $O_p(n^{-1/2})$ in the expansion (19)-(21). 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   Stochastic Expansion of GMM Estimators

In this section, we provide formal stochastic expansions of GMM estimators assuming that the moment condition evaluated at the true value is a local drifting sequence around zero:

$$E[g(X_{in}, \theta_0)] = \frac{\delta}{\sqrt{n}} \tag{33}$$

for some nonzero $\delta \in \mathbb{R}^q$ that depends on θ_0. Note that the observations now form a triangular array $\{X_{in} : i = 1, ..., n, n \in \mathbb{N}\}$ because (33) changes with n. Under (33) and regularity conditions, $\hat{\theta}_1$ and $\hat{\theta}_2$ are both consistent for θ_0 and by the central limit theorem

$$\sqrt{n}(g_n(\theta_0) - E[g(X_{in}, \theta_0)]) = O_p(1). \tag{34}$$

This setup is referred to as local misspecification in the literature, e.g., Newey (1985), Otsu (2011), Guggenberger (2012), Conley, Hansen, and Rossi (2012), Andrews, Gentzkow, and Shapiro (2017), Bonhomme and Weidner (2018), and Armstrong and Kolesár (2019). Since $g_n(\theta_0)$ is still $O_p(n^{-1/2})$, local misspecification and over-identification bias are essentially equivalent. This setup is useful in distinguishing the effect of the over-identification bias (local misspecification) on the point estimate and on the variance. That is, the local-to-zero device provides a better approximation of the finite sample distribution of the estimator but it should not be taken as the true data-generating process.

Under correct specification (1), Newey and Smith (2004) provide a thorough analysis on the higher-order bias and variance of the stochastic expansions of GMM and GEL estimators up to the order of $O_p(n^{-1})$ in the form of

$$\sqrt{n}(\hat{\theta} - \theta_0) = \psi_0 + \psi_1/\sqrt{n} + \psi_2/n + R_n \tag{35}$$

where ψ_0, ψ_1, ψ_2 are all $O_p(1)$ and $R_n = O_p(n^{-3/2})$. Our analysis reveals additional higher-order terms of the order $O_p(n^{-1/2})$ in the RHS of (35) as well as the first order asymptotic bias in the point estimate. Specifically, we derive the exact expressions of ψ_0 and ψ_1 for the one-step and the two-step GMM to show that our double correction approximates the (co)variance of some higher-order terms. Deriving the full expression of ψ_2 is not attempted because the derivation is quite tedious and is not required to show the relationship between the double correction and the higher-order expansion.

Let $g(\theta) = E[g(X_{in}, \theta)]$ and write $g = g(\theta_0)$. Define

$$\widetilde{g}(\theta) = \sum_{i=1}^{n} \left(g(X_{in}, \theta) - g(\theta) \right)/\sqrt{n},$$

$$\widetilde{G} = \sum_{i=1}^{n} (G(X_{in}) - G)/\sqrt{n},$$

$$\widetilde{W} = \sum_{i=1}^{n} (W(X_{in}) - W)/\sqrt{n},$$

$$\widetilde{\Omega}(\theta) = \sum_{i=1}^{n} (g(X_{in}, \theta)g(X_{in}, \theta)' - \Omega(\theta))/\sqrt{n},$$

and write $\widetilde{g} = \widetilde{g}(\theta_0)$ and $\widetilde{\Omega} = \widetilde{\Omega}(\theta_0)$. First we consider the one-step GMM estimator.

Theorem 1. *Suppose that (33) holds, $\widetilde{g} = O_p(1)$, $\widetilde{G} = O_p(1)$, $\widetilde{W} = O_p(1)$, the FOC of the one-step GMM holds with probability approaching one (w.p.a. 1), G is full column rank, and $W > 0$. In addition, suppose that the second moment of $g(X_{in}, \theta_0)$, $G(X_{in})$, and $W(X_{in})$ exist and are finite. Then, the one-step GMM estimator has the following expansion w.p.a.1.*

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = \eta_W + \widetilde{\psi}_{W,0} + \frac{1}{\sqrt{n}} \left(\widetilde{\psi}_{W,1} + \widetilde{q}_W + \widetilde{B}_W \left(\eta_W + \widetilde{\psi}_{W,0} \right) \right) + O_p\left(\frac{1}{n} \right) \tag{36}$$

where

$$\eta_W = -(G'W^{-1}G)^{-1}G'W^{-1}\delta, \tag{37}$$

$$\widetilde{\psi}_{W,0} = -(G'W^{-1}G)^{-1}G'W^{-1}\widetilde{g}, \tag{38}$$

$$\widetilde{\psi}_{W,1} = -(G'W^{-1}G)^{-1}\left(\widetilde{G}'W^{-1}\delta - G'W^{-1}\widetilde{W}W^{-1}\delta\right), \tag{39}$$

$$\widetilde{q}_W = -(G'W^{-1}G)^{-1}\left(\widetilde{G}W^{-1}\widetilde{g} - G'W^{-1}\widetilde{W}W^{-1}\widetilde{g}\right) \tag{40}$$

$$\widetilde{B}_W = -(G'W^{-1}G)^{-1}\left(\widetilde{G}'W^{-1}G - G'W^{-1}\widetilde{W}W^{-1}G + G'W^{-1}\widetilde{G}\right). \tag{41}$$

The first term in the expansion of Theorem 1, η_W, is the constant bias due to local misspecification. It shifts the mean of the distribution but does not alter the variance. This bias cannot be consistently estimated from the data in general. Conley, Hansen, and Rossi (2012) impose a prior distribution on δ and Armstrong and Kolesár (2019) set a pre-specified bound on δ to construct bias-corrected confidence intervals.

The conventional variance estimator estimates the variance of $\widetilde{\psi}_{W,0}$. The doubly corrected variance estimator (30) estimates the variance of $\widetilde{\psi}_{W,0} + \widetilde{\psi}_{W,1}/\sqrt{n}$. Since the variance of $\widetilde{\psi}_{W,1}$ increases with δ, the double correction would be more effective with a larger δ. This does not necessarily imply potentially larger bias in the point estimate, however. The magnitude of change in η_W and in the variance due to δ depends on the data matrices. For example, Andrews, Gentzkow, and Shapiro (2017) use η_W with a given δ to check how sensitive the estimator is. So if the estimator is not sensitive (to local violation of the moment condition), then η_W does not change much while the variance may be significantly affected. Table 1 shows an example that the bias in the point estimate does not increase with δ while the standard deviation of the estimator increases substantially ($n = 50$).

On the other hand, the double correction omits some terms of the same order, $(\widetilde{q}_W + \widetilde{B}_W(\eta_W + \widetilde{\psi}_{W,0}))/\sqrt{n}$. We discuss these terms one by one. First, the variance of $(\widetilde{q}_W + \widetilde{B}_W\widetilde{\psi}_{W,0})/\sqrt{n}$ does not increase with δ because $\widetilde{g}$ is a centered process. Second, $\mathrm{Var}(\widetilde{B}_W\eta_W/\sqrt{n})$ increases with δ but its sample analogue is zero because the sample analogue of η_W is zero (FOC). So the (co)variance of this term cannot be estimated.

An estimator of the higher-order variance up to $O_p(n^{-1})$ of Newey and Smith (2004) would include the variance of $(\widetilde{q}_W + \widetilde{B}_W(\eta_W + \widetilde{\psi}_{W,0}))/\sqrt{n}$ and the covariance between $\widetilde{\psi}_{W,0}$ and $(\widetilde{q}_W + \widetilde{B}_W(\eta_W + \widetilde{\psi}_{W,0}))/\sqrt{n} + \widetilde{\psi}_{W,2}/n$ where $\widetilde{\psi}_{W,2}$ is the higher-order term corresponding ψ_2 in (35). Since these terms are not taken into account, the doubly corrected variance estimator is not an estimator of the higher-order variance of Newey and Smith (2004). Although the higher-order variance up to $O_p(n^{-1})$ could be estimated by further expanding (36) up to the remainder term of the order $O_p(n^{-3/2})$ the resulting estimator would include many terms to be estimated which would not be practical. Thus, the doubly corrected variance estimator can be viewed as a convenient alternative to the higher-order variance estimator.

Next we consider the two-step GMM estimator.

11

Theorem 2. *Suppose that the assumptions of Theorem 1 hold. In addition, $\widetilde{\Omega} = O_p(1)$, the FOC of the two-step GMM holds w.p.a.1, $\Omega > 0$, and the fourth moment of $g(X_{in}, \theta_0)$ exists and is finite. Then the two-step GMM estimator has the following expansion w.p.a.1.*

$$
\sqrt{n}(\hat{\theta}_2 - \theta_0)
$$

$$
= \left[\eta_\Omega + \frac{1}{\sqrt{n}} \left((D + H_{\eta_\Omega}) \, \eta_W \right) \right]
$$

$$
+ \widetilde{\psi}_{\Omega,0} + \frac{1}{\sqrt{n}} \left(\widetilde{\psi}_{\Omega,1} + \left(D + \widetilde{C} + H_{\eta_\Omega} + H_{\widetilde{\psi}_{\Omega,0}} \right) \widetilde{\psi}_{W,0} + \widetilde{q}_\Omega + \widetilde{B}_\Omega \left(\eta_\Omega + \widetilde{\psi}_{\Omega,0} \right) + \left(\widetilde{C} + H_{\widetilde{\psi}_{\Omega,0}} \right) \eta_W \right)
$$

$$
+ \frac{1}{n} D \widetilde{\psi}_{W,1} + O_p \left(\frac{1}{n} \right)
$$

where

$$
\eta_\Omega = -(G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\delta, \tag{42}
$$

$$
\widetilde{\psi}_{\Omega,0} = -(G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\widetilde{g}, \tag{43}
$$

$$
\widetilde{\psi}_{\Omega,1} = -(G'\Omega^{-1}G)^{-1} \left(\widetilde{G}'\Omega^{-1}\delta - G'\Omega^{-1}\widetilde{\Omega}\Omega^{-1}\delta \right), \tag{44}
$$

$$
\widetilde{q}_\Omega = -(G'\Omega^{-1}G)^{-1} \left(\widetilde{G}'\Omega^{-1}\widetilde{g} - G'\Omega^{-1}\widetilde{\Omega}\Omega^{-1}\widetilde{g} \right), \tag{45}
$$

$$
\widetilde{B}_\Omega = -(G'\Omega^{-1}G)^{-1} \left(\widetilde{G}'\Omega^{-1}G - G'\Omega^{-1}\widetilde{\Omega}\Omega^{-1}G + G'\Omega^{-1}\widetilde{G} \right), \tag{46}
$$

$$
D[.,j] = (G'\Omega^{-1}G)^{-1}G'\Omega^{-1} \left. \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \right|_{\theta=\theta_0} \Omega^{-1}\delta, \tag{47}
$$

$$
\widetilde{C}[.,j] = (G'\Omega^{-1}G)^{-1}G'\Omega^{-1} \left. \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \right|_{\theta=\theta_0} \Omega^{-1}\widetilde{g}, \tag{48}
$$

$$
H_v[.,j] = (G'\Omega^{-1}G)^{-1}G'\Omega^{-1} \left. \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \right|_{\theta=\theta_0} \Omega^{-1}Gv. \tag{49}
$$

The terms in the square brackets in the expansion, $\eta_\Omega + (D + H_{\eta_\Omega})\eta_W/\sqrt{n}$ is the constant bias due to local misspecification which cannot be consistently estimated in general. It only shifts the mean of the distribution but does not affect the variance.

The conventional, the Windmeijer correction, and the doubly corrected variance estimators estimate the variance of

$$
\text{conventional: } \widetilde{\psi}_{\Omega,0},
$$

$$
\text{Windmeijer: } \widetilde{\psi}_{\Omega,0} + \frac{D\widetilde{\psi}_{W,0}}{\sqrt{n}},
$$

$$
\text{double correction: } \widetilde{\psi}_{\Omega,0} + \frac{\widetilde{\psi}_{\Omega,1} + D\widetilde{\psi}_{W,0}}{\sqrt{n}} + \frac{D\widetilde{\psi}_{W,1}}{n},
$$

respectively. Similar to the one-step GMM expansion, the double correction estimates the variance of the higher-order terms that increase with δ: D, $\widetilde{\psi}_{\Omega,1}$, and $\widetilde{\psi}_{W,1}$. In contrast, some (or all) of

these terms are omitted in the Windmeijer correction (or the conventional variance estimator). This implies that those variance estimators tend to differ more with a larger δ, which is also supported by the simulation experiment in Section 8. Note that a larger δ does not necessarily imply larger bias in the point estimate via η_Ω and η_W. For example, Table 1 shows that the bias in the two-step GMM point estimate decreases with δ but the three variance estimators differ substantially for a large δ.

Including $(\widetilde{\psi}_{\Omega,1} + D\widetilde{\psi}_{W,0})/\sqrt{n} + D\widetilde{\psi}_{W,1}/n$ in the double correction formula is also important to get robustness to global (fixed) misspecification, where $\delta = O(\sqrt{n})$, because these terms become $O_p(1)$. In this case, the expansion should be centered around the pseudo-true value which is defined as the unique minimizer of the GMM criterion and η_W and η_Ω become zero because of the population FOC. We show that the doubly corrected variance estimator is misspecification-robust in the next section.

By a similar argument with the one-step GMM, both the doubly corrected and the Windmeijer corrected variance estimators are not the higher-order variance estimator of Newey and Smith (2004) because some higher-order terms that constitute the higher-order variance up to $O_p(n^{-1})$ are omitted in both corrections.

Remark (Edgeworth expansion) The improved approximation to the finite sample variance of the GMM estimators by the double correction is different from the higher-order refinement via the Edgeworth expansion which expands the finite sample distribution function of the (standardized) test statistic. Using analytical expansions (Rothenberg, 1984; Hansen, 2006; Kundhi and Rilstone, 2013) or the bootstrap (Hall, 1992; Hall and Horowitz, 1996; Andrews, 2002; Lee, 2014, 2016), a critical value can be obtained which gives a smaller error in the size of the test or in the coverage probability of the confidence interval.

5   Robustness to Misspecification

Both the Windmeijer corrected and the conventional variance estimators, $\widehat{V}_w(\hat{\theta}_2)$ and $\widetilde{V}(\hat{\theta}_2)$, 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, correct specification means that an over-identified model exactly holds at a unique parameter value θ_0, but this may be too restrictive in reality. Indeed, the sample moment condition model does not hold for any finite sample size n almost surely if the model is over-identified, i.e., $g_n(\hat{\theta}) \neq 0$. 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

$$E[g(X_i, \theta)] = \delta(\theta) \neq 0, \ \forall \theta \in \Theta, \tag{50}$$

where $\delta(\theta)$ is a vector of constants and Θ is a compact 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). Thus, the pseudo-true values of the one-step and two-step GMM may differ from each other. 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 the one-step and two-step GMM under misspecification. Hansen and Lee (2019) propose a similar robust variance estimator for the iterated GMM. These variance estimators are shown to be consistent regardless of misspecification and they are referred to as the misspecification-robust variance estimator, hereinafter.

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

The equivalence holds for the following reasons. First, the Windmeijer correction that accounts for the effect of $\hat{\theta}_1$ in the weight matrix corrects for the pseudo-true value of the one-step GMM being different from that of the two-step GMM. Our double correction corrects for this bias too. Second, the double correction corrects for the over-identification bias which becomes asymptotically non-negligible misspecification bias. Overall, the double correction makes the expansions (23)-(27) robust to misspecification.

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

$$\theta_1 = \arg\min_{\theta} E[g(X_i, \theta)]'W^{-1}E[g(X_i, \theta)], \tag{51}$$

$$\theta_2 = \arg\min_{\theta} E[g(X_i, \theta)]'[\Omega(\theta_1)]^{-1}E[g(X_i, \theta)]. \tag{52}$$

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 $j = 1, 2$. (Global) misspecification implies that $g_n(\theta_j) = O_p(1)$ for $j = 1, 2$. Sufficient conditions for consistency and asymptotic normality under misspecification for the one-step and two-step GMM is given in Hall and Inoue (2003) for stationary data and for the one-step and iterated GMM is given in B. Hansen and Lee (2019) for clustered data.

Take the FOC of the two-step GMM. We can use the expansions (6)-(8) by letting $\theta_1 \neq \theta_2$. Then (9) can be written as

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_2 - \theta_2) = & -\left\{G_n'[\Omega_n(\theta_1)]^{-1}G_n\right\}^{-1}G_n'[\Omega_n(\theta_1)]^{-1}\sqrt{n}g_n(\theta_2) \\
& + D_n^*\sqrt{n}(\hat{\theta}_1 - \theta_1) + O_p(n^{-1/2}\|g_n(\theta_2)\|) \\
= & \sqrt{n}(\tilde{\theta}_2^* - \theta_2) + D_n^*\sqrt{n}(\hat{\theta}_1 - \theta_1) + O_p(n^{-1/2}\|g_n(\theta_2)\|),
\end{aligned} \tag{53}$$

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

$$\tilde{\theta}_2^* = \arg\min_{\theta \in \Theta} \ g_n(\theta)'[\Omega_n(\theta_1)]^{-1}g_n(\theta), \tag{54}$$

14

and

$$D_n^* = F_{1n}^* + F_{2n}^*,$$

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

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

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 (global) misspecification. Note that both $\sqrt{n}(\tilde{\theta}_2^* - \theta_2)$ and $D_n^* \sqrt{n}(\hat{\theta}_1 - \theta_1)$ in (53) are $O_p(1)$ under misspecification and this will alter the first-order asymptotic variance.

Using the fact that $G'\Omega_j^{-1}g_j = 0$ for $j = 1, 2$, the FOC of the infeasible two-step GMM (54) can be expanded as

$$\sqrt{n}(\tilde{\theta}_2^* - \theta_2) = - \left\{ G_n'[\Omega_n(\theta_1)]^{-1}G_n \right\}^{-1} \left\{ G_n'[\Omega_n(\theta_1)]^{-1}\sqrt{n}\left(g_n(\theta_2) - g_2\right) \right.$$
$$\left. + \sqrt{n}\left(G_n - G\right)'[\Omega_n(\theta_1)]^{-1}g_2 - G'\Omega_1^{-1}\sqrt{n}\left(\Omega_n(\theta_1) - \Omega_1\right)[\Omega_n(\theta_1)]^{-1}g_2 \right\}. \qquad (55)$$

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

$$\sqrt{n}(\hat{\theta}_1 - \theta_1) = - \left\{ G_n'W_n^{-1}G_n \right\}^{-1} \left[G_n'W_n^{-1}\sqrt{n}\left(g_n(\theta_1) - g_1\right) \right.$$
$$\left. + \sqrt{n}(G_n - G)'W_n^{-1}g_1 - G'W^{-1}\sqrt{n}\left(W_n - W\right)W_n^{-1}g_1 \right]. \qquad (56)$$

The expansions (55) and (56) are misspecification-robust versions of (7) and (8), allowing for different probability limits of the one-step and two-step GMM estimators and taking into account for the misspecification (over-identification) bias.

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

Theorem 3. *Suppose that* (50) *holds,* $G_n - G = O_p(n^{-1/2})$, $W_n - W = O_p(n^{-1/2})$, $\Omega_n(\theta_j) - \Omega_j = O_p(n^{-1/2})$, $\hat{\theta}_j - \theta_j = O_p(n^{-1/2})$ *for* $j = 1, 2$, W *and* Ω_1 *are nonsingular, and* G *is full column rank. Then, as* $n \to \infty$, *for* $j = 1, 2$,

$$\widehat{V}_{dc}(\hat{\theta}_j)^{-1/2}\sqrt{n}(\hat{\theta}_j - \theta_j) \xrightarrow{d} N(0, I_k). \qquad (57)$$

Theorem 3 also implies that the 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, the bootstrap t test and CIs based on the GMM t statistic studentized with the conven-

tional or the Windmeijer standard error requires an additional recentering procedure in resampling to correct for the over-identification bias to achieve higher-order refinements. Furthermore, the conventional bootstrap procedure with recentering 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 (Weight matrix) All the results so far hold if we replace $\Omega_n(\theta)$ with the centered weight matrix

$$\Omega_n^c(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left(g(X_i, \theta) - g_n(\theta) \right) \left(g(X_i, \theta) - g_n(\theta) \right)'. \tag{58}$$

The matrices D_n and $\widehat{D}_n$ would need to be modified accordingly. The centered weight matrix is consistent for the asymptotic variance matrix of the moment equation under misspecification. B. Hansen (2019) 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.

6 Iterated GMM and Continuously Updating GMM

Both the Windmeijer and our double correction correct for 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 of B. Hansen and Lee (2019) and the continuously-updating (CU) GMM of L. Hansen, Heaton, and Yaron (1996). We show that the answer is yes for the iterated GMM and also show that the double correction formula is the same as the misspecification-robust formula. For the CU GMM, the answer is negative.

Assume 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

$$0 = G_n'[\Omega_n(\hat{\theta})]^{-1} g_n(\hat{\theta}) \tag{59}$$

where $\hat{\theta}$ is the iterated GMM. 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 (2019). By applying the first-order Taylor expansion around θ_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

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left\{ G_n'[\Omega_n(\theta_0)]^{-1} G_n (I_k - D_n) \right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1} \sqrt{n} g_n(\theta_0) + O_p(n^{-1}). \tag{60}$$

Windmeijer (2000) proposes a finite sample corrected variance estimator based on the expansion (60). We proceed one additional step. By further expanding to take into account for the over-identification bias, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left\{ G_n'[\Omega_n(\theta_0)]^{-1}G_n(I_k - D_n) \right\}^{-1} \left[G'\Omega^{-1}\sqrt{n}g_n(\theta_0) \right. \tag{61}$$

$$\left. + \sqrt{n}(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}\left(\Omega_n(\theta_0) - \Omega \right)\Omega^{-1}g_n(\theta_0) \right] + O_p(n^{-1}). \tag{62}$$

Since the remainder term is $O_p(n^{-1})$, by estimating the variance of the terms in (61)-(62) 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 the iterated GMM is

$$\widehat{V}_{dc}(\hat{\theta}) = \{G_n'[\Omega_n(\hat{\theta})]^{-1}G_n(I_k - \widehat{D}_n)\}^{-1}\Sigma_n(\hat{\theta}, \Omega_n(\hat{\theta}))\{G_n'[\Omega_n(\hat{\theta})]^{-1}G_n(I_k - \widehat{D}_n)\}^{-1\prime}, \tag{63}$$

where $\Sigma_n(\hat{\theta}, \Omega_n(\hat{\theta}))$ is defined in (32) and $\widehat{D}_n$ is evaluated at $\hat{\theta}$. Not surprisingly, this formula is identical to the misspecification-robust variance estimator for the iterated GMM of Hansen and Lee (2019). The finite sample corrected formula suggested by Windmeijer (2000) is

$$\widehat{V}_w(\hat{\theta}) = (I_k - \widehat{D}_n)^{-1} \left(G_n'[\Omega_n(\hat{\theta})]^{-1}G_n \right)^{-1} (I_k - \widehat{D}_n)^{-1\prime}. \tag{64}$$

On the other hand, a similar finite sample correction may not be obtained for the 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 θ is scalar. The FOC is

$$0 = \left(G_n - \frac{1}{2}g_n(\hat{\theta})'[\Omega_n(\hat{\theta})]^{-1} \left. \frac{\partial \Omega_n(\theta)}{\partial \theta} \right|_{\theta = \hat{\theta}} \right)' [\Omega_n(\hat{\theta})]^{-1}g_n(\hat{\theta}). \tag{65}$$

This shows that even when the moment function is linear, the effective derivative that forms 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.

7 Examples

7.1 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

$$\hat{\theta}_1 = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y. \tag{66}$$

where $Y = [y_1, \cdots, y_n]'$, $X = [X_1, \cdots, X_n]'$, and $Z = [Z_1, \cdots, Z_n]'$ 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

$$\hat{\theta}_2 = (X'Z\widehat{\Omega}_1^{-1}Z'X)^{-1}X'Z\widehat{\Omega}_1^{-1}Z'Y \tag{67}$$

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

$$\widehat{V}_{dc}(\hat{\theta}_1) = \left(\frac{1}{n}X'Z(Z'Z)^{-1}Z'X\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{m}_{1i}\hat{m}_{1i}'\left(\frac{1}{n}X'Z(Z'Z)^{-1}Z'X\right)^{-1}, \tag{68}$$

$$\widehat{V}_{dc}(\hat{\theta}_2) = \widehat{V}(\hat{\theta}_2) + \widehat{D}_n\widehat{C}(\hat{\theta}_1, \hat{\theta}_2) + \widehat{C}(\hat{\theta}_1, \hat{\theta}_2)'\widehat{D}_n' + \widehat{D}_n\widehat{V}_{dc}(\hat{\theta}_1)\widehat{D}_n', \tag{69}$$

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

$$\hat{\theta}_s = (X'Z\widehat{\Omega}_{s-1}^{-1}Z'X)^{-1}X'Z\widehat{\Omega}_{s-1}^{-1}Z'Y, \tag{70}$$

where

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

We iterate the s-step GMM estimator until convergence given a preset tolerance, e.g. $\|\hat{\theta}_s - \hat{\theta}_{s-1}\| < 10^{-5}$ 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

$$\widehat{V}_{dc}(\hat{\theta}) = \widehat{H}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{m}_i \hat{m}_i'\right)\widehat{H}^{-1'}, \tag{71}$$

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

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

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

$$\widehat{V}_w(\hat{\theta}) = \widehat{H}^{-1}\left(\frac{1}{n^2}X'Z\widehat{\Omega}^{-1}Z'X\right)\widehat{H}^{-1'}, \tag{72}$$

$$\widetilde{V}(\hat{\theta}) = \left(\frac{1}{n^2}X'Z\widehat{\Omega}^{-1}Z'X\right)^{-1}. \tag{73}$$

7.2 A Panel Data Model

Consider a panel data model with a scalar regressor

$$y_{it} = x_{it}\beta + \eta_i + v_{it}, \tag{74}$$

for $i = 1, ..., N$ and $t = 1, ..., T$ where η_i is the unobserved individual effects, the unknown parameter of interest is β, 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,

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

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

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

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

Our doubly corrected variance estimator can be used for the model (74) 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 $\widehat{W} = n^{-1}\sum_{i=1}^{N} Z_i'HZ_i$, where H is a matrix with 2's on the main diagonal, -1's on the first off-diagonals and zero 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 $Z = (Z_1', ..., Z_N')'$ is the instrument matrix, $\Delta Y = (\Delta y_1', ..., \Delta y_N')'$, $\Delta X = (\Delta x_1', ..., \Delta x_N')'$, $\Delta y_i = (\Delta y_{i2}, ..., \Delta y_{iT})'$, and $\Delta x_i = (\Delta x_{i2}, ..., \Delta x_{iT})'$. Note that scaling the weight matrix does not affect the estimator. The doubly corrected variance estimator of $\hat{\beta}_1$ is given by

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

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

where $\Delta\hat{v}_{1i} = \Delta y_i - \Delta x_i\hat{\beta}_1$ and $\Delta\hat{v}_1 = (\Delta\hat{v}_{11}', ..., \Delta\hat{v}_{1N}')'$. 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

$$\widehat{\Omega}_1 = \frac{1}{n}\sum_{i=1}^{N} Z_i'\Delta\hat{v}_{1i}\Delta\hat{v}_{1i}'Z_i. \tag{75}$$

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 $\Delta\hat{v}_2 = (\Delta\hat{v}_{21}', ..., \Delta\hat{v}_{2N}')'$. 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

$$\widetilde{V}(\hat{\beta}_2) = n^2 \left(\Delta X'Z\widehat{\Omega}_1^{-1}Z'\Delta X\right)^{-1}. \tag{76}$$

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

$$\hat{\beta}_s = (\Delta X'Z\widehat{\Omega}_{s-1}^{-1}Z'\Delta X)^{-1}\Delta X'Z\widehat{\Omega}_{s-1}^{-1}Z'\Delta Y, \tag{77}$$

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, e.g. $\|\hat{\beta}_s - \hat{\beta}_{s-1}\| < 10^{-5}$ 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 $\Delta\hat{v} = (\Delta\hat{v}_1', ..., \Delta\hat{v}_N')'$ be the $n \times 1$ residual vector.

The doubly corrected variance estimator for the 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

$$
\widehat{V}_w(\hat{\beta}) = \widehat{H}^{-1}\left(\frac{1}{n^2}\Delta X'Z\widehat{\Omega}^{-1}Z'\Delta X\right)\widehat{H}^{-1'}, \tag{78}
$$

$$
\widetilde{V}(\hat{\beta}) = \left(\frac{1}{n^2}\Delta X'Z\widehat{\Omega}^{-1}Z'\Delta X\right)^{-1}. \tag{79}
$$

8 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 (58). 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 compared with those based on the uncentered one (reported) they are not reported.

8.1 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

$$
y_i = x_i\beta_0 + e_i
$$

$$
E\left(z_ie_i\right) = 0 \tag{80}
$$

where x_i and β_0 are scalar and $z_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i})'$ is a vector of instrumental variables. We estimate β_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

$$y_i = x_i \beta_0 + \frac{\alpha_0}{\sqrt{n}} (z_{1i} - z_{2i} + z_{3i} - z_{4i}) + e_i, \tag{81}$$

$$x_i = \pi_0 (z_{1i} + z_{2i} + z_{3i} + z_{4i}) + u_i,$$

$$e_i = 0.5 u_i + \sqrt{1 - 0.5^2} v_i,$$

$$z_i \sim N(0, I_4), \begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & z_{1i}^2 \end{bmatrix} \right).$$

We set $\beta_0 = 1$, vary α_0 from 0 to 1 in steps of 0.2, and set the first-stage coefficient π_0 so that the first-stage $R^2 = 0.2$. We set the number of observations as $n = 50, 100, 500$.

The parameter α_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 (80) fails to hold in finite samples, but it holds asymptotically.

Means and standard deviations of one-step (2SLS), two-step, and iterated GMM estimators 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 remains accurate regardless of misspecification, including the correct specification case ($\alpha_0 = 0$); the means of corrected standard errors are very close to the standard deviations for all values of α_0, especially for the two-step and the 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 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 α_0. As is well known, the conventional standard error for the two-step is severely downward biased when $\alpha_0 = 0$, and this bias also increases with α_0. The Windmeijer corrected standard error works well under correct specification, but does not fully account for additional variations due to non-zero α_0. The result is similar for the iterated GMM.

8.2 Linear Dynamic Panel Model

8.2.1 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 $i = 1, ..., N$ and $t = 1, ..., T$,

$$y_{it} \ = \ \rho_0 y_{i,t-1} + \eta_i + \nu_{it}, \tag{82}$$

where η_i is an unobserved individual-specific effect and ν_{it} is an error term. The parameter of interest ρ_0 is estimated by the difference GMM based on a set of moment conditions:

$$E[y_{i,t-s}(\Delta y_{it} - \rho_0 \Delta y_{i,t-1})] = 0, \ \ t = 3, ...T, \ \text{and} \ s \geq 2, \tag{83}$$

The moment conditions are derived from taking differences of (82), 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 (83). A sufficient condition for this to hold is that the model (82) coincides with the true DGP, but this is unlikely to be true. A reasonable deviation from the assumed model (82) is heterogeneity in ρ_0 across i. We assume the following DGP. For $i = 1, ..., N$ and $t = 1, ...T$,

$$y_{it} = \rho_i y_{i,t-1} + \eta_i + \nu_{it},$$
$$\eta_i \sim N(0,1); \ \rho_i \sim \Phi(\alpha_0 \eta_i); \ \nu_{it} \sim N(0, 0.5^2),$$
$$y_{i1} = \frac{\eta_i}{1 - \rho_i} + u_{i1}; \ u_{i1} \sim N\left(0, \frac{1}{1 - \rho_i^2}\right),$$

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

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

where γ_j is the jth autocovariance. The last equation becomes zero at $\rho = E[\rho_i]$ if ρ_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 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 ρ for $s = 2, 3, ..., t - 1$. In the DGP, η_i and ρ_i are dependent through α_0 and a larger α_0 leads to larger heterogeneity. We vary α_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 α_0 under global misspecification. However, by varying α_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 the two-step and iterated GMM estimators, the doubly corrected standard errors are as accurate as the Windmeijer correction for small values of α_0 (including correct specification $\alpha_0 = 0$) but dominate the other in accuracy for larger values of α_0. The doubly corrected standard error for the one-step GMM is slightly upward biased for small values of α_0, but this bias decreases with a larger sample size $N = 500$.

8.2.2 Misspecified Lag Length

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

$$y_{it} = \beta_0 x_{it} + \eta_i + v_{it}, \tag{84}$$

for $i = 1, ..., N$ and $t = 1, ..., T$. The unknown parameter of interest is β_0, and the regressor x_{it} is predetermined with respect to v_{it}, i.e., $E(x_{it} v_{it+s}) = 0$ for $s = 0, ..., T - t$. We use the first differenced GMM estimator and the number of moment conditions is $T(T-1)/2$ as in Section 7.2.

The DGP is

$$
\begin{aligned}
y_{it} &= \beta_0 x_{it} + \alpha_0 x_{it-1} + \eta_i + v_{it}, \tag{85} \\
x_{it} &= 0.5 x_{it-1} + \eta_i + 0.5 v_{it-1} + \epsilon_{it}, \\
\eta &\sim N(0,1) \text{ and } \epsilon_{it} \sim N(0,1), \\
v_{it} &= \delta_i \tau_t \omega_{it} \text{ and } \omega_{it} \sim \chi_1^2 - 1.
\end{aligned}
$$

We generate initial 50 time periods with $\tau_t = 0.5$ for $t = -49, \ldots, 0$ and $x_i \sim N(\eta_i/0.5, 1/0.75)$ same as Windmeijer (2005). The parameter α_0 in (85) governs the degree of misspecification. When $\alpha_0 = 0$, the model (84) is correctly specified which reduces to that of Windmeijer (2005).

Tables 4 and 5 report estimation results for $\beta_0 = 1$, $N = 100, 500$ and $T = 4, 6$. The degree of misspecification α_0 is varied across $\{0, 0.025, 0.05, 0.1\}$. The first column ($\alpha_0 = 0$) in Table 4 replicates Monte Carlo studies in Windmeijer (2005, Table 1).

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 simulation experiments, the Windmeijer correction works best under correct specification, but becomes downward biased as α_0 increases. Note that deviation from the 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 $N = 500$.

Appendix: Proofs

In the proof of Theorems 1 and 2, we use following lemma on geometric expansion of a matrix. Lemma 1 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

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

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

by the assumptions of the lemma. $\qquad\qquad\square$

Proof of Theorem 1: In what following the statements hold with probability approaching one. First note that

$$G_n'W_n^{-1}G_n = G'W^{-1}G + (G_n - G)'W^{-1}G + G'W^{-1}(G_n - G) + G'(W_n^{-1} - W^{-1})G + O_p(n^{-1})$$

$$= G'W^{-1}G + O_p(n^{-1/2}).$$

By Lemma 1 with $q = 1$,

$$(G_n'W_n^{-1}G_n)^{-1} = \left(G'W^{-1}G + \frac{1}{\sqrt{n}}\sqrt{n}(G_n'W_n^{-1}G_n - G'W^{-1}G)\right)^{-1}$$

$$= (G'W^{-1}G)^{-1} - \frac{1}{\sqrt{n}}(G'W^{-1}G)^{-1}\sqrt{n}(G_n'W_n^{-1}G_n - G'W^{-1}G)(G'W^{-1}G)^{-1}$$

$$+ O_p(n^{-1})$$

$$= (G'W^{-1}G)^{-1} + \frac{1}{\sqrt{n}}\widetilde{B}_W(G'W^{-1}G)^{-1} + O_p(n^{-1}).$$

In addition,

$$
\begin{aligned}
G_n' W_n^{-1} \sqrt{n} g_n(\theta_0) =& G_n' W_n^{-1} \sqrt{n}(g_n(\theta_0) - g) + \sqrt{n}(G_n - G)' W_n^{-1} g + G' \sqrt{n}(W_n^{-1} - W^{-1}) g + G' W^{-1} \delta \\
=& G' W^{-1} \widetilde{g} + \widetilde{G}' W^{-1} g - G' W^{-1} \widetilde{W} W^{-1} g + \widetilde{G}' W^{-1} \widetilde{g}/\sqrt{n} - G' W^{-1} \widetilde{W} W^{-1} \widetilde{g}/\sqrt{n} \\
& + G' W^{-1} \delta + O_p(n^{-1}).
\end{aligned}
$$

Since the FOC of the one-step GMM holds regardless of misspecification of the moment condition, we use the above expansions to obtain

$$
\begin{aligned}
& \sqrt{n}(\hat{\theta}_1 - \theta_0) \\
=& (G_n W_n^{-1} G_n)^{-1} G_n W_n^{-1} \sqrt{n} g_n(\theta_0) \\
=& - \left[(G' W^{-1} G)^{-1} + \widetilde{B}_W (G' W^{-1} G)^{-1}/\sqrt{n} \right] \times \left[G' W^{-1} \widetilde{g} + \widetilde{G}' W^{-1} g - G' W^{-1} \widetilde{W} W^{-1} g \right. \\
& \left. + \widetilde{G}' W^{-1} \widetilde{g}/\sqrt{n} - G' W^{-1} \widetilde{W} W^{-1} \widetilde{g}/\sqrt{n} + G' W^{-1} \delta \right] + O_p(n^{-1}) \\
=& \eta_W + \widetilde{B}_W \eta_W/\sqrt{n} - (G' W^{-1} G)^{-1} G' W^{-1} \widetilde{g} - (G' W^{-1} G)^{-1} \left(\widetilde{G}' W^{-1} g - G' W^{-1} \widetilde{W} W^{-1} g \right) \\
& - (G' W^{-1} G)^{-1} \left(\widetilde{G} W^{-1} \widetilde{g} - G' W^{-1} \widetilde{W} W^{-1} \widetilde{g} \right)/\sqrt{n} - \widetilde{B}_W (G' W^{-1} G)^{-1} G' W^{-1} \widetilde{g}/\sqrt{n} + O_p(n^{-1}) \\
=& \eta_W + \widetilde{\psi}_{W,0} + (\widetilde{\psi}_{W,1} + \widetilde{q}_W + \widetilde{B}_W(\widetilde{\psi}_{W,0} + \eta_W))/\sqrt{n} + O_p(n^{-1}).
\end{aligned}
$$

$\square$

Proof of Theorem 2: In what following the statements hold with probability approaching one. We first prove some useful expansions. Note that from Theorem 1, $\sqrt{n}(\hat{\theta}_1 - \theta_0) = O_p(1)$. By the first-order Taylor expansion and (13),

$$
\begin{aligned}
\Omega_n(\hat{\theta}_1) =& \Omega_n(\theta_0) + \frac{1}{\sqrt{n}} \sum_j \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \sqrt{n}(\hat{\theta}_{1[j]} - \theta_{0[j]}) \quad\quad (86) \\
=& \Omega_n(\theta_0) - \Omega + \Omega + \frac{1}{\sqrt{n}} \sum_j \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \sqrt{n}(\hat{\theta}_{1[j]} - \theta_{0[j]}) \\
=& \Omega + O_p(n^{-1/2}).
\end{aligned}
$$

By Lemma 1 with $q = 0$,

$$
[\Omega_n(\hat{\theta}_1)]^{-1} = \Omega^{-1} + O_p(n^{-1/2}), \quad\quad (87)
$$

$$
[\Omega_n(\theta_0)]^{-1} = \Omega^{-1} + O_p(n^{-1/2}). \qu\quad (88)
$$

Using (88) and applying Lemma 1 with $q = 1$ to (86),

$$[\Omega_n(\hat{\theta}_1)]^{-1} = [\Omega_n(\theta_0)]^{-1} - \frac{1}{\sqrt{n}}[\Omega_n(\theta_0)]^{-1} \sum_j \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \sqrt{n}(\hat{\theta}_{1[j]} - \theta_{0[j]})[\Omega_n(\theta_0)]^{-1} + O_p(n^{-1}) \quad (89)$$

$$= [\Omega_n(\theta_0)]^{-1} - \frac{1}{\sqrt{n}}\Omega^{-1} \sum_j \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \sqrt{n}(\hat{\theta}_{1[j]} - \theta_{0[j]})\Omega^{-1} + O_p(n^{-1}) \quad (90)$$

Using a similar argument with the proof of Theorem 1 and Lemma 1 with $q = 1$, we obtain

$$(G_n'[\Omega_n(\theta_0)]^{-1}G_n)^{-1} = (G'\Omega^{-1}G)^{-1} + \frac{1}{\sqrt{n}}B_\Omega(G'\Omega^{-1}G)^{-1} + O_p(n^{-1}), \quad (91)$$

$$G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n}g_n(\theta_0) = G'\Omega^{-1}\delta + G'\Omega^{-1}\widetilde{g} + \widetilde{G}'\Omega^{-1}g - G'\Omega^{-1}\widetilde{\Omega}\Omega^{-1}g \quad (92)$$

$$+ \widetilde{G}\Omega^{-1}\widetilde{g}/\sqrt{n} - G'\Omega^{-1}\widetilde{\Omega}\Omega^{-1}\widetilde{g}/\sqrt{n} + O_p(n^{-1}). \quad (93)$$

Note that the assumption of the theorem implies that $\hat{\theta}_2 - \theta_0 = O_p(n^{-1/2})$. Using (90), the first-order Taylor expansion of the FOC of $\hat{\theta}_2$ around θ_0 can be written as

$$0 = G_n'[\Omega_n(\hat{\theta}_1)]^{-1}g_n(\theta_0) + G_n'[\Omega_n(\hat{\theta}_1)]^{-1}G_n(\hat{\theta}_2 - \theta_0) \quad (94)$$

$$= G_n'[\Omega_n(\theta_0)]^{-1}g_n(\theta_0) + G_n'[\Omega_n(\theta_0)]^{-1}G_n(\hat{\theta}_2 - \theta_0) \quad (95)$$

$$+ G_n'\left([\Omega_n(\hat{\theta}_1)]^{-1} - [\Omega_n(\hat{\theta}_0)]^{-1}\right)\left(g_n(\theta_0) + G_n(\hat{\theta}_2 - \theta_0)\right) \quad (96)$$

$$= G_n'[\Omega_n(\theta_0)]^{-1}g_n(\theta_0) + G_n'[\Omega_n(\theta_0)]^{-1}G_n(\hat{\theta}_2 - \theta_0) \quad (97)$$

$$- \frac{1}{\sqrt{n}}G'\Omega^{-1} \sum_j \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \sqrt{n}(\hat{\theta}_{1[j]} - \theta_{0[j]})\Omega^{-1}\left(g_n(\theta_0) + G(\hat{\theta}_2 - \theta_0)\right) + O_p(n^{-3/2}). \quad (98)$$

By arranging terms, multiplying $\sqrt{n}$, using (91)-(92),

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

$$+ \left(G_n'[\Omega_n(\theta_0)]^{-1}G_n\right)^{-1}$$

$$\times \frac{1}{\sqrt{n}}G'\Omega^{-1} \sum_j \frac{\partial \Omega(\theta)}{\partial \theta_{[j]}} \sqrt{n}(\hat{\theta}_{1[j]} - \theta_{0[j]})\Omega^{-1}\left(\sqrt{n}g_n(\theta_0) + G\sqrt{n}(\hat{\theta}_2 - \theta_0)\right) + O_p(n^{-1}).$$

Note that the second term in the RHS is $O_p(n^{-1/2})$. First,

$$- \left(G_n'[\Omega_n(\theta_0)]^{-1}G_n\right)^{-1}G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n}g_n(\theta_0) \quad (99)$$

$$= -\left((G'\Omega^{-1}G)^{-1} + \frac{1}{\sqrt{n}}\widetilde{B}_\Omega(G'\Omega^{-1}G)^{-1}\right) \quad (100)$$

$$\times \left(G'\Omega^{-1}\delta + G'\Omega^{-1}\widetilde{g} + \widetilde{G}'\Omega^{-1}g - G'\Omega^{-1}\widetilde{\Omega}\Omega^{-1}g + \widetilde{G}\Omega^{-1}\widetilde{g}/\sqrt{n} - G'\Omega^{-1}\widetilde{\Omega}\Omega^{-1}\widetilde{g}/\sqrt{n}\right) \quad (101)$$

$$+ O_p(n^{-1}) \quad (102)$$

$$= \eta_\Omega + \frac{1}{\sqrt{n}}\widetilde{B}_\Omega\eta_\Omega + \widetilde{\psi}_{\Omega,0} + \frac{1}{\sqrt{n}}\left(\widetilde{\psi}_{\Omega,1} + \widetilde{q}_\Omega + \widetilde{B}_\Omega\widetilde{\psi}_{\Omega,0}\right) + O_p(n^{-1}). \quad (103)$$

28

Thus, we have $\sqrt{n}(\hat{\theta}_2 - \theta_0) = \eta_\Omega + \widetilde{\psi}_{\Omega,0} + O_p(n^{-1/2})$. Next, using $\sqrt{n}(\hat{\theta}_1 - \theta_0) = \eta_W + \widetilde{\psi}_{W,0} + O_p(n^{-1/2})$,

$$\left(G'_n[\Omega_n(\theta_0)]^{-1}G_n\right)^{-1} \frac{1}{\sqrt{n}} G'\Omega^{-1} \sum_j \frac{\partial\Omega(\theta)}{\partial\theta_{[j]}} \sqrt{n}(\hat{\theta}_{1[j]} - \theta_{0[j]})\Omega^{-1}\left(\sqrt{n}g_n(\theta_0) + G\sqrt{n}(\hat{\theta}_2 - \theta_0)\right)$$

(104)

$$= (G'\Omega^{-1}G)^{-1} \frac{1}{\sqrt{n}} G'\Omega^{-1} \sum_j \frac{\partial\Omega(\theta)}{\partial\theta_{[j]}}(\eta_{W[j]} + \widetilde{\psi}_{W,0[j]})\Omega^{-1}\left(\widetilde{g} + \delta + G(\eta_\Omega + \widetilde{\psi}_{\Omega,0})\right) + O_p(n^{-1})$$

(105)

$$= (G'\Omega^{-1}G)^{-1} \frac{1}{\sqrt{n}} G'\Omega^{-1} \sum_j \frac{\partial\Omega(\theta)}{\partial\theta_{[j]}}\eta_{W[j]}\Omega^{-1}\left(\widetilde{g} + \delta + G(\eta_\Omega + \widetilde{\psi}_{\Omega,0})\right)$$

(106)

$$+ (G'\Omega^{-1}G)^{-1} \frac{1}{\sqrt{n}} G'\Omega^{-1} \sum_j \frac{\partial\Omega(\theta)}{\partial\theta_{[j]}}\widetilde{\psi}_{W,0[j]}\Omega^{-1}\left(\widetilde{g} + \delta + G(\eta_\Omega + \widetilde{\psi}_{\Omega,0})\right) + O_p(n^{-1})$$

(107)

$$= \frac{1}{\sqrt{n}}\left\{(D + H_{\eta_\Omega})\,\eta_W + \left(\widetilde{C} + H_{\widetilde{\psi}_{\Omega,0}}\right)\left(\eta_W + \widetilde{\psi}_{W,0}\right) + (D + H_{\eta_\Omega})\,\widetilde{\psi}_{W,0}\right\} + O_p(n^{-1}).$$

(108)

Combining this with (103),

$$\sqrt{n}(\hat{\theta}_2 - \theta_0)$$

(109)

$$= \eta_\Omega + \frac{1}{\sqrt{n}}\left((D + H_{\eta_\Omega})\,\eta_W\right) + \widetilde{\psi}_{\Omega,0}$$

(110)

$$+ \frac{1}{\sqrt{n}}\left(\widetilde{\psi}_{\Omega,1} + \left(D + \widetilde{C} + H_{\eta_\Omega} + H_{\widetilde{\psi}_{\Omega,0}}\right)\widetilde{\psi}_{W,0} + \widetilde{q}_\Omega + \widetilde{B}_\Omega\left(\eta_\Omega + \widetilde{\psi}_{\Omega,0}\right) + \left(\widetilde{C} + H_{\widetilde{\psi}_{\Omega,0}}\right)\eta_W\right)$$

(111)

$$+ O_p(n^{-1}).$$

(112)

Notice that $D\widetilde{\psi}_{W,1}/n$ in the $O_p(n^{-1})$ remainder term can be obtained if we plug $\sqrt{n}(\hat{\theta}_1 - \theta_0) = \eta_W + \widetilde{\psi}_{W,0} + \widetilde{\psi}_{W,1}/\sqrt{n} + O_p(n^{-1/2})$ into (104). This proves the theorem. $\qquad\square$

Proof of Theorem 3: We first derive the asymptotic variance matrices of the GMM estimators under misspecification and then show that the doubly corrected variance estimators are consistent. Expanding the FOC of the one-step GMM around the pseudo-true value θ_1 we can write

$$\sqrt{n}(\hat{\theta}_1 - \theta_1) = -(G'W^{-1}G)^{-1}\left(GW^{-1}\sqrt{n}(g_n(\theta_1) - g_1) + \sqrt{n}(G_n - G)'W^{-1}g_1\right.$$
$$\left. - G'W^{-1}\sqrt{n}(W_n - W)W^{-1}g_1\right) + O_p(n^{-1/2})$$
$$= -(G'W^{-1}G)^{-1}\left(GW^{-1}\sqrt{n}g_n(\theta_1) + \sqrt{n}G'_nW^{-1}g_1 - G'W^{-1}\sqrt{n}W_nW^{-1}g_1\right)$$
$$+ O_p(n^{-1/2}).$$

The second equality holds by the population FOC $G'W^{-1}g_1 = 0$. Similarly we expand the FOC of

the two-step GMM as

$$\sqrt{n}(\hat{\theta}_2 - \theta_2) = -\left(G'_n[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1}G_n[\Omega_n(\hat{\theta}_1)]^{-1}\sqrt{n}g_n(\theta_2)$$
$$= -\left(G'_n[\Omega_n(\theta_1)]^{-1}G_n\right)^{-1}G_n[\Omega_n(\theta_1)]^{-1}\sqrt{n}g_n(\theta_2) + D^*_n\sqrt{n}(\hat{\theta}_1 - \theta_1) + O_p(n^{-1})$$

where

$$D^*_n = F^*_{1n} + F^*_{2n},$$

$$F^*_{1n} = -\left.\frac{\partial\left\{G'_n[\Omega_n(\theta)]^{-1}G_n\right\}^{-1}}{\partial\theta'}\right|_{\theta=\theta_1} G'_n[\Omega_n(\theta_1)]^{-1}g_n(\theta_2),$$

$$F^*_{2n} = -\left\{G'_n[\Omega_n(\theta_1)]^{-1}G_n\right\}^{-1}\left.\frac{\partial G'_n[\Omega_n(\theta)]^{-1}g_n(\theta_2)}{\partial\theta'}\right|_{\theta=\theta_1}.$$

Note that $F^*_{1n} = O_p(n^{-1/2})$ because $G'_n[\Omega_n(\theta_1)]^{-1}g_n(\theta_2) = O_p(n^{-1/2})$ but $F^*_2 = O_p(\|g_n(\theta_2)\|) = O_p(1)$ under misspecification. To further expand the FOC of the two-step GMM we observe that $D^*_n = D^* + O_p(n^{-1/2})$ where

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

$$\frac{\partial\Omega(\theta)}{\partial\theta_{[j]}} = E\left[g(X_i,\theta)\frac{\partial g(X_i,\theta)'}{\partial\theta_{[j]}}\right] + E\left[\frac{\partial g(X_i,\theta)}{\partial\theta_{[j]}}g(X_i,\theta)'\right].$$

Thus,

$$\sqrt{n}(\hat{\theta}_2 - \theta_2) = -\left(G'\Omega_1^{-1}G\right)^{-1}\left(G'\Omega_1^{-1}\sqrt{n}g_n(\theta_2) + \sqrt{n}G'_n\Omega_1^{-1}g_2 - G'\Omega_1^{-1}\sqrt{n}\Omega_n(\theta_1)\Omega_1^{-1}g_2\right)$$
$$+ D^*\sqrt{n}(\hat{\theta}_1 - \theta_1) + O_p(n^{-1/2})$$
$$= -\left(G'\Omega_1^{-1}G\right)^{-1}\left(G'\Omega_1^{-1}\sqrt{n}g_n(\theta_2) + \sqrt{n}G'_n\Omega_1^{-1}g_2 - G'\Omega_1^{-1}\sqrt{n}\Omega_n(\theta_1)\Omega_1^{-1}g_2\right)$$
$$- D^*(G'W^{-1}G)^{-1}\left(G'W^{-1}\sqrt{n}g_n(\theta_1) + \sqrt{n}G'_nW^{-1}g_1 - G'W^{-1}\sqrt{n}W_nW^{-1}g_1\right)$$
$$+ O_p(n^{-1/2}).$$

Let

$$m_{1i} = G'W^{-1}g(X_i,\theta_1) + G(X_i)'W^{-1}g_1 - G'W^{-1}W_iW^{-1}g_1, \tag{113}$$

$$m_{2i} = G'\Omega_1^{-1}g(X_i,\theta_2) + G(X_i)'\Omega_1^{-1}g_2 - G'\Omega_1^{-1}\Omega_i(\theta_1)\Omega_1^{-1}g_2 \tag{114}$$

where $W_n = n^{-1}\sum_{i=1}^n W_i$ and $\Omega_i(\theta) = g(X_i,\theta)g(X_i,\theta)'$. Since $E[m_{1i}] = E[m_{2i}] = 0$, by the multivariate Lindeberg-Lévy CLT,

$$\sqrt{n}(\hat{\theta}_2 - \theta_2) \xrightarrow{d} N(0, V_2) \tag{115}$$

where

$$V_2 = (G'\Omega_1^{-1}G)^{-1}E[m_{2i}m_{2i}'](G'\Omega_1^{-1}G)^{-1} + D^*(G'W^{-1}G)^{-1}E[m_{1i}m_{2i}'](G'\Omega_1^{-1}G)^{-1}$$
$$+ (G'\Omega_1^{-1}G)^{-1}E[m_{2i}m_{1i}'](G'W^{-1}G)^{-1}D^{*'} + D^*(G'W^{-1}G)^{-1}E[m_{1i}m_{1i}'](G'W^{-1}G)^{-1}D^{*'}.$$

By replacing the population quantities (expectations) of V with the sample means, we directly obtain $\widehat{V}_{dc}(\hat{\theta}_2)$. By the LLN, $\widehat{V}_{dc}(\hat{\theta}_2) \xrightarrow{p} V_2$.

Similarly we have

$$\sqrt{n}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} N(0, V_1) \tag{116}$$

where

$$V_1 = (G'W^{-1}G)^{-1}E[m_{1i}m_{1i}'](G'W^{-1}G)^{-1}. \tag{117}$$

By the LLN, $\widehat{V}_{dc}(\hat{\theta}_1) \xrightarrow{p} V_1$. This completes the proof. $\qquad\square$

References

[1] Andrews, D. W. K. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica*, 70, 119-162.

[2] Andrews, I. (2019). On the structure of IV estimands. *Journal of Econometrics*, 211(1), 294-307.

[3] Andrews, I., Gentzkow, M., Shapiro, J.M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *Quarterly Journal of Economics*, 132(4), 1553-1592.

[4] Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58, 277-297.

[5] Arellano, M., and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68, 29-51.

[6] Armstrong, T., and Kolesár, M. (2019). Sensitivity analysis using approximate moment condition models. arXiv preprint arXiv:1808.07387.

[7] Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variables estimators. *Econometrica*, 63, 657-681.

[8] Blundell, R., and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87, 115-143.

[9] Bond, S., and Windmeijer, F. (2005). Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews*, 24(1), 1-37.

[10] Bonhomme, S., and Weidner, M. (2018). Minimizing sensitivity to model misspecification. arXiv preprint arXiv:1807.02161.

[11] Conley, T.G., Hansen, C.B. and Rossi, P.E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1), pp.260-272.

[12] Evdokimov, K., and Kolesár, M. (2018). Inference in instrumental variable regression analysis with heterogeneous treatment effects. Working paper.

[13] Guggenberger, P. (2012). On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. *Econometric Theory*, 28(2), 387-421.

[14] Hall, A. R. (2000). Covariance matrix estimation and the power of the overidentifying restrictions test. *Econometrica*, 68(6), 1517-1527.

[15] Hall, A. R., and Inoue, A. (2003). The large sample behavior of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, 114(2), 361-394.

[16] Hall, P. (1992). The Bootstrap and Edgeworth expansion, New York: Springer-Verlag.

[17] Hall, P. and J. L. Horowitz (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* 64, 891-916.

[18] Han, C. and Phillips, P.C. (2006). GMM with many moment conditions. *Econometrica*, 74(1), 147-192.

[19] Hansen, B. E. (2006). Edgeworth expansions for the Wald and GMM statistics for nonlinear restrictions. Econometric theory and practice: frontiers of analysis and applied research, ed. by D. Corbae, SN Durlauf, and BE Hansen, Cambridge, 9-35.

[20] Hansen, B. E. (2019). Econometrics. Unpublished Manuscript. University of Wisconsin.

[21] Hansen, B. E., and Lee, S. (2019). Inference for iterated GMM under misspecification. Working paper.

[22] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029-1054.

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

[24] Hwang, J., and 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. *Journal of Econometrics*, 207(2), 381-405.

[25] Kitamura, Y., Otsu, T., and Evdokimov, K. (2013). Robustness, infinitesimal neighborhoods, and moment restrictions. *Econometrica*, 81(3), 1185-1201.

[26] Kundhi, G., and Rilstone, P. (2013). Edgeworth and saddlepoint expansions for nonlinear estimators. *Econometric Theory*, 29(5), 1057-1078.

[27] Lee, S. (2014). Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators. *Journal of Econometrics*, 178, pp.398-413.

[28] Lee, S. (2016). Asymptotic refinements of a misspecification-robust bootstrap for GEL estimators. *Journal of Econometrics*, 192(1), 86-104.

[29] Lee, S. (2018). A consistent variance estimator for 2sls when instruments identify different LATEs. *Journal of Business & Economic Statistics*, 36(3), 400-410.

[30] Magdalinos, M. A. (1992). Stochastic expansions and asymptotic approximations. *Econometric Theory*, 8(3), 343-367.

[31] Maasoumi, E., and Phillips, P. C. (1982). On the behavior of inconsistent instrumental variable estimators. *Journal of Econometrics*, 19(2), 183-201.

[32] Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics*, 29(3), 229-256.

[33] Newey, W. K., and McFadden, D. (1994). Large sample estimation and hypothesis testing. Handbook of econometrics, 4, 2111-2245.

[34] Newey, W. K., and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1), 219-255.

[35] Newey, W. K., and Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, 77, 687-719.

[36] Otsu, T. (2011). Moderate deviations of generalized method of moments and empirical likelihood estimators. *Journal of Multivariate Analysis*, 102(8), 1203-1216.

[37] Rotemberg, J. J. (1983). Instrumental variable estimation of misspecified models. Working paper, MIT Sloan.

[38] Rothenberg, T. J. (1984). Approximating the distributions of econometric estimators and test statistics. Handbook of econometrics, 2, 881-935.

[39] Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics*, 35(2), 634-672.

[40] 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.

[41] Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators.*Journal of Econometrics*, 126(1), 25-51.

	α_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

Table 1: Monte Carlo Results for Linear IV: $n = 50, 100, 500$

	α_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
	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
	se$_w$ $\hat{\rho}_2$	0.3376	0.3377	0.3300	0.3208	0.3087	0.2961	0.2734
	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
	se$_w$ $\hat{\rho}$	0.3483	0.3489	0.3427	0.3350	0.3243	0.3133	0.2909
	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
	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
	se$_w$ $\hat{\rho}_2$	0.1635	0.1634	0.1631	0.1626	0.1611	0.1567	0.1493
	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
	se$_w$ $\hat{\rho}$	0.1778	0.1783	0.1794	0.1822	0.1858	0.1887	0.1887
	se$_{dc}$ $\hat{\rho}$	0.1773	0.1784	0.1806	0.1855	0.1916	0.1965	0.1975

Table 2: Monte Carlo Results for Linear Dynamic Panel: $N = 100$ and $T = 4, 6$

	α_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

Table 3: Monte Carlo Results for Linear Dynamic Panel: $N = 500$ and $T = 4, 6$

	γ_0	0	0.025	0.05	0.1
$N = 100$	$\hat{\beta}_1$	0.9793	0.9534	0.9268	0.8744
$T = 4$	sd $\hat{\beta}_1$	0.1521	0.1529	0.1538	0.1588
	se $\hat{\beta}_1$	0.1469	0.1468	0.1465	0.1472
	se$_{dc}$ $\hat{\beta}_1$	0.1546	0.1555	0.1563	0.1604
	$\hat{\beta}_2$	0.9849	0.9588	0.9322	0.8773
	sd $\hat{\beta}_2$	0.1404	0.1422	0.1451	0.1552
	se $\hat{\beta}_2$	0.1243	0.1244	0.1242	0.1253
	se$_w$ $\hat{\beta}_2$	0.1390	0.1402	0.1415	0.1473
	se$_{dc}$ $\hat{\beta}_2$	0.1343	0.1366	0.1390	0.1482
	$\hat{\beta}$	0.9858	0.9599	0.9334	0.8781
	sd $\hat{\beta}$	0.1417	0.1439	0.1474	0.1600
	se $\hat{\beta}$	0.1243	0.1244	0.1242	0.1253
	se$_w$ $\hat{\beta}$	0.1393	0.1408	0.1426	0.1507
	se$_{dc}$ $\hat{\beta}$	0.1352	0.1378	0.1406	0.1517
$N = 100$	$\hat{\beta}_1$	0.9755	0.9577	0.9411	0.9060
$T = 6$	sd $\hat{\beta}_1$	0.1027	0.1043	0.1051	0.1083
	se $\hat{\beta}_1$	0.1002	0.1003	0.1004	0.1013
	se$_{dc}$ $\hat{\beta}_1$	0.1056	0.1064	0.1075	0.1107
	$\hat{\beta}_2$	0.9833	0.9649	0.9466	0.9080
	sd $\hat{\beta}_2$	0.0906	0.0929	0.0948	0.1017
	se $\hat{\beta}_2$	0.0716	0.0718	0.0720	0.0731
	se$_w$ $\hat{\beta}_2$	0.0905	0.0914	0.0930	0.0978
	se$_{dc}$ $\hat{\beta}_2$	0.0836	0.0853	0.0876	0.0944
	$\hat{\beta}$	0.9857	0.9671	0.9484	0.9083
	sd $\hat{\beta}$	0.0946	0.0970	0.0998	0.1099
	se $\hat{\beta}$	0.0716	0.0718	0.0720	0.0731
	se$_w$ $\hat{\beta}$	0.0937	0.0952	0.0976	0.1054
	se$_{dc}$ $\hat{\beta}$	0.0866	0.0885	0.0916	0.1006

Table 4: Monte Carlo Results for Linear Panel Model: $N = 100$ and $T = 4, 6$

	γ_0	0	0.025	0.05	0.1
$N = 500$	$\hat{\beta}_1$	0.9958	0.9683	0.9406	0.8853
$T = 4$	sd $\hat{\beta}_1$	0.0685	0.0686	0.0695	0.0716
	se $\hat{\beta}_1$	0.0679	0.0678	0.0677	0.0679
	se$_{dc}$ $\hat{\beta}_1$	0.0686	0.0690	0.0696	0.0716
	$\hat{\beta}_2$	0.9970	0.9710	0.9443	0.8892
	sd $\hat{\beta}_2$	0.0652	0.0657	0.0676	0.0730
	se $\hat{\beta}_2$	0.0632	0.0631	0.0630	0.0634
	se$_w$ $\hat{\beta}_2$	0.0648	0.0653	0.0662	0.0694
	se$_{dc}$ $\hat{\beta}_2$	0.0634	0.0643	0.0658	0.0708
	$\hat{\beta}$	0.9970	0.9712	0.9446	0.8897
	sd $\hat{\beta}$	0.0652	0.0658	0.0678	0.0738
	se $\hat{\beta}$	0.0632	0.0631	0.0630	0.0634
	se$_w$ $\hat{\beta}$	0.0648	0.0652	0.0662	0.0698
	se$_{dc}$ $\hat{\beta}$	0.0634	0.0643	0.0659	0.0715
$N = 500$	$\hat{\beta}_1$	0.9947	0.9758	0.9564	0.9181
$T = 6$	sd $\hat{\beta}_1$	0.0473	0.0476	0.0482	0.0498
	se $\hat{\beta}_1$	0.0469	0.0470	0.0470	0.0473
	se$_{dc}$ $\hat{\beta}_1$	0.0475	0.0479	0.0484	0.0499
	$\hat{\beta}_2$	0.9968	0.9778	0.9582	0.9175
	sd $\hat{\beta}_2$	0.0432	0.0440	0.0454	0.0494
	se $\hat{\beta}_2$	0.0408	0.0408	0.0410	0.0414
	se$_w$ $\hat{\beta}_2$	0.0431	0.0436	0.0445	0.0471
	se$_{dc}$ $\hat{\beta}_2$	0.0413	0.0421	0.0434	0.0470
	$\hat{\beta}$	0.9969	0.9779	0.9584	0.9174
	sd $\hat{\beta}$	0.0433	0.0441	0.0457	0.0504
	se $\hat{\beta}$	0.0408	0.0408	0.0410	0.0414
	se$_w$ $\hat{\beta}$	0.0431	0.0437	0.0447	0.0479
	se$_{dc}$ $\hat{\beta}$	0.0414	0.0423	0.0437	0.0479

Table 5: Monte Carlo Results for Linear Panel Model: $N = 500$ and $T = 4, 6$