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

October 21, 2018

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 doubly corrects the commonly used finite sample correction of Windmeijer (2005) and is very simple to calculate. A nice property of our double correction is that robustness to misspecification is obtained without affecting the order of finite sample correction under correct specification. That is, the proposed variance estimator provides more accurate approximation to the finite sample variance under correct specification and is consistent regardless of misspecification.

1 Introduction

The generalized method of moments (GMM) estimators (L. Hansen, 1982) are widely used in modern economics. In particular, the efficient GMM has the smallest asymptotic variance in the class of GMM estimators. However, researchers have recognized that the two-step procedure which is required to obtain efficiency may give rise to bias in the point estimate and standard error. L. Hansen, Heaton, and Yaron (1996) focused on the finite sample bias of the two-step GMM point estimate and suggested alternative estimators. Windmeijer (2005) proposed a finite sample corrected standard error formula for the two-step linear GMM that accounts for the added variability due to the two-step procedure. His correction (the Windmeijer correction, hereinafter) has been routinely used in practice.[1] Similar to Windmeijer (2005), our focus is the bias of the GMM variance estimator.

We propose an alternative finite sample correction for the variance of the linear one-step, two-step, and iterated GMM estimators. It improves upon the Windmeijer correction by considering

*jungbin.hwang@uconn.edu
†b.kang1@lancaster.ac.uk
‡jay.lee@unsw.edu.au

[1] More than 4,200 citations according to Google Scholar on September 27, 2018.

finite sample bias not only from the two-step procedure but also from over-identification of the moment equation model. Thus, we doubly correct the finite sample bias of the linear GMM variance estimator. Under correct specification, the order of our double correction is the same as the Windmeijer correction. More importantly, our doubly corrected variance estimator is consistent even when the moment equation model is misspecified. Thus, robustness to misspecification is achieved at no cost. The proposed variance estimator is not new though, because they are proposed as the misspecification-robust variance estimator in Lee (2014) and B. Hansen and Lee (2018a). What is new in this paper is to show that the two seemingly unrelated formulas, the finite sample doubly corrected and the misspecification-robust, are in fact equivalent.

Intuition behind our finding is as follows. When the moment equation model is over-identified the sample mean of the moment equation is not equal to zero in finite sample regardless of whether the model is correctly specified or not. Both ours and the Windmeijer correction take into account for additional variations due to the non-zero sample moment, leading to more accurate inference under correct specification. Since our double correction includes terms that are higher-order under correct specification (thus ignored in the Windmeijer correction) but become the first-order under misspecification, our doubly corrected variance estimator is consistent even under misspecification.

Inference with misspecified moment equation models has gained considerable attention in the literature. Maasoumi and Phillips (1982) investigate the limiting distribution of inconsistent instrumental variable (IV) estimators under over-identification. Hall and Inoue (2003) derive the asymptotic distribution of GMM estimators under misspecification. Schennach (2007) proposes the exponentially tilted empirical likelihood estimator exhibiting the nice properties of the empirical likelihood (EL) and the exponential tilting (ET) estimators under misspecification. Otsu (2011) analyses moderate deviation behaviors of GMM. Guggenburger (2012) studies the behavior of the weak instrument robust tests under local violation of the exogeneity assumption. Kitamura, Otsu, and Evdokimov (2013) propose an alternative minimum-Hellinger distance estimator under local deviation from the data generating process. Lee (2014, 2016) propose a nonparametric bootstrap procedure for GMM and generalized empirical likelihood (GEL) estimators that achieves asymptotic refinements regardless of misspecification. Hansen and Lee (2018a) establish the existence and uniqueness of the iterated GMM under misspecification and clustering. Rotemberg (1983) and Andrews (forthcoming) characterize the estimands of the linear overidentified GMM estimators under misspecification.

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 (forthcoming) study the finite sample properties of the one-step and two-step GMM estimators for dependent observations.

Our doubly corrected and misspecification-robust standard errors for the one-step, two-step, and iterated GMM estimators are different than the standard errors of IV and GEL estimators based on many instrument and many weak instrument asymptotics under correct specification, e.g., Bekker

(1994), Han and Phillips (2006), and Newey and Windmeijer (2009). See also Evdokimov and Kolesár (2018) for the valid standard errors in the presence of heterogeneous treatment effects. We point out that our doubly corrected variance formula is not rooted in any alternative asymptotics with many (weak) instruments, but is designed to capture smaller order terms in the standard asymptotics and to focus on finite sample variation due to misspecification in over-identified moment conditions. Thus, this paper can be regarded as a complement to the literature of robust GMM variance estimates in the papers mentioned above.

2 Finite Sample Correction of Windmeijer (2005)

Suppose that we observe a sequence of 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. The moment equation model is correctly specified if

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

for a unique θ_0. We assume that standard regularity conditions for consistency and asymptotic normality of GMM hold, such as the conditions listed in Theorems 2.6 and 3.4 of Newey and McFadden (1994) for i.i.d. observations or Theorems 14 and 15 of Hansen and Lee (2018b) for clustered observations (e.g. panel data). Under correct specification (1), we note that

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

which will be used in determining the order of higher-order terms in Sections 2 and 3.

The one-step GMM estimator is defined as

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

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

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

Define $\Omega = \Omega(\theta_0)$ where

$$\Omega(\theta) = E[g(X_i, \theta)g(X_i, \theta)']. \tag{5}$$

Under regularity conditions, $\Omega_n(\hat{\theta}_1) \xrightarrow{p} \Omega$ and this leads to the efficient GMM. We also define an infeasible two-step GMM estimator $\tilde{\theta}_2$ using $[\Omega_n(\theta_0)]^{-1}$ as the weight matrix:

$$\tilde{\theta}_2 = \arg\min_{\theta \in \Theta} \; g_n(\theta)'[\Omega_n(\theta_0)]^{-1} g_n(\theta). \tag{6}$$

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 $G_i = \partial g(X_i, \theta)/\partial \theta'$. Note that it does not depend on θ due to linearity. Define $G_n = n^{-1} \sum_{i=1}^{n} G_i$ and $G = E[G_i]$. By the first-order Taylor expansion, the 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

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) = -\left\{ G_n'[\Omega_n(\hat{\theta}_1)]^{-1} G_n \right\}^{-1} G_n'[\Omega_n(\hat{\theta}_1)]^{-1} \sqrt{n} g_n(\theta_0). \tag{7}$$

Using a similar expansion, we can get

$$\sqrt{n}(\tilde{\theta}_2 - \theta_0) = -\left\{ G_n'[\Omega_n(\theta_0)]^{-1} G_n \right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1} \sqrt{n} g_n(\theta_0), \tag{8}$$

for the infeasible two-step GMM and

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = -(G_n' W_n^{-1} G_n)^{-1} G_n' W_n^{-1} \sqrt{n} g_n(\theta_0) \tag{9}$$

for the one-step GMM.

Asymptotically (7) and (8) 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 (9), 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 equation models.

To see this, we use the first-order Taylor expansion of $\Omega_n(\hat{\theta}_1)$ in the RHS of (7) around θ_0:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_2 - \theta_0) &= -\left\{ G_n'[\Omega_n(\theta_0)]^{-1} G_n \right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1} \sqrt{n} g_n(\theta_0) + D_n \sqrt{n}(\hat{\theta}_1 - \theta_0) + O_p(n^{-1}) \\ &= \sqrt{n}(\tilde{\theta}_2 - \theta_0) + \underbrace{D_n \sqrt{n}(\hat{\theta}_1 - \theta_0)}_{= O_p(n^{-1/2})} + O_p(n^{-1}), \end{aligned} \tag{10}$$

where

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

are $k \times k$ matrices. 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 (10) 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 (10) only holds for linear moment equation models.

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

$$
\widehat{V}_c(\hat{\theta}_2) = \widetilde{V}(\hat{\theta}_2) + \widehat{D}_n \widetilde{V}(\hat{\theta}_2) + \widetilde{V}(\hat{\theta}_2) \widehat{D}_n' + \widehat{D}_n \widetilde{V}(\hat{\theta}_1) \widehat{D}_n', \tag{11}
$$

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

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

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 $\sqrt{\widehat{V}_c(\hat{\theta}_2)/n}$.

3 Double Correction

The Windmeijer correction takes into account for the extra variability due to using the 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 $g_n(\theta_0) \neq 0$. This property holds in general for over-identified models.

We show that non-zero $g_n(\theta_0)$ causes additional finite sample variability in (10). These additional terms are not considered in the Windmeijer correction (11). We propose alternative variance estimators that fully incorporate the additional variations induced by non-zero $g_n(\theta_0)$. These variance estimators will replace $\widetilde{V}(\hat{\theta}_2)$ and $\widetilde{V}(\hat{\theta}_1)$ in (11) without affecting the order of finite sample corrections, leading to the doubly corrected variance estimator.

Assume that

$$\begin{aligned} G_n - G =& O_p(n^{-1/2}), & (12) \\ \text{vec}(\Omega_n(\theta_0) - \Omega) =& O_p(n^{-1/2}), & (13) \\ \text{vec}(W_n - W) =& O_p(n^{-1/2}), & (14) \end{aligned}$$

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

$$[\Omega_n(\theta_0)]^{-1} - \Omega^{-1} = -\Omega^{-1}\left(\Omega_n(\theta_0) - \Omega\right)[\Omega_n(\theta_0)]^{-1}, \tag{15}$$

we can write

$$\begin{aligned} & G_n'[\Omega_n(\theta_0)]^{-1}g_n(\theta_0) \\ &= \underbrace{G'\Omega^{-1}g_n(\theta_0)}_{=O_p(n^{-1/2})} + \underbrace{(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\left(\Omega_n(\theta_0) - \Omega\right)\Omega^{-1}g_n(\theta_0)}_{=O_p(n^{-1})} + O_p(n^{-3/2}). \end{aligned} \tag{16}$$

Using (16), (8) can be written as

$$\begin{aligned} \sqrt{n}(\tilde{\theta}_2 - \theta_0) =& -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}\left[G'\Omega^{-1}\sqrt{n}g_n(\theta_0)\right. & (17) \\ & \left.+\sqrt{n}(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}\left(\Omega_n(\theta_0) - \Omega\right)\Omega^{-1}g_n(\theta_0)\right] & (18) \\ & + O_p(n^{-1}). & (19) \end{aligned}$$

Similarly,

$$\begin{aligned} \sqrt{n}(\hat{\theta}_1 - \theta_0) =& -\left\{G_n'W_n^{-1}G_n\right\}^{-1}\left[G'W^{-1}\sqrt{n}g_n(\theta_0)\right. & (20) \\ & \left.+\sqrt{n}(G_n - G)'W^{-1}g_n(\theta_0) - G'W^{-1}\sqrt{n}\left(W_n - W\right)W^{-1}g_n(\theta_0)\right] & (21) \\ & + O_p(n^{-1}), & (22) \end{aligned}$$

which simplifies to

$$\sqrt{n}(\hat{\theta}_1 - \theta_0) = -\left\{G_n'G_n\right\}^{-1}\left[G'\sqrt{n}g_n(\theta_0) + \sqrt{n}(G_n - G)'g_n(\theta_0)\right] \tag{23}$$

when $W_n = I$. From the above expansions it is clear that (i) to allow for non-zero $g_n(\theta_0)$ we also 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 (10) is not changed.

Using the expansions (17)-(19) an (20)-(22), the expansion of the two-step GMM can be written as

$$\sqrt{n}(\hat{\theta}_2-\theta_0) = -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}\left[G'\Omega^{-1}\sqrt{n}g_n(\theta_0)\right. \tag{24}$$

$$\left.+\sqrt{n}(G_n-G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}\left(\Omega_n(\theta_0)-\Omega\right)\Omega^{-1}g_n(\theta_0)\right] \tag{25}$$

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

$$\left.+\sqrt{n}(G_n-G)'W^{-1}g_n(\theta_0) - G'W^{-1}\sqrt{n}\left(W_n-W\right)W^{-1}g_n(\theta_0)\right] \tag{27}$$

$$+ O_p(n^{-1}). \tag{28}$$

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 (25) and (27). 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 (27) are $O_p(n^{-1})$ multiplied by D_n, which is the same order with the remainder term. Thus, considering the extra terms in estimating the variance of the one-step GMM does not necessarily provide finite sample corrections. However, including these terms are critical to get robustness to misspecification, which is shown in the next Section.

Note that this expansion is not a higher-order stochastic (or Edgeworth) expansion.

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

$$\widehat{V}_{dc}(\hat{\theta}_2) = \widehat{V}(\hat{\theta}_2) + \widehat{D}_n\widehat{C}(\hat{\theta}_1,\hat{\theta}_2) + \widehat{C}(\hat{\theta}_1,\hat{\theta}_2)'\widehat{D}_n' + \widehat{D}_n\widehat{V}_{dc}(\hat{\theta}_1)\widehat{D}_n', \tag{29}$$

where

$$\widehat{V}(\hat{\theta}_2) = \left(G_n'[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1}\Psi_n(\hat{\theta}_2,\Omega_n(\hat{\theta}_1))\left(G_n'[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1}, \tag{30}$$

$$\widehat{V}_{dc}(\hat{\theta}_1) = \left(G_n'W_n^{-1}G_n\right)^{-1}\Psi_n(\hat{\theta}_1,W_n)\left(G_n'W_n^{-1}G_n\right)^{-1}, \tag{31}$$

$$\widehat{C}(\hat{\theta}_1,\hat{\theta}_2) = \left(G_n'W_n^{-1}G_n\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\psi(\hat{\theta}_1,W_n)\psi(\hat{\theta}_2,\Omega_n(\hat{\theta}_1))'\left(G_n'[\Omega_n(\hat{\theta}_1)]^{-1}G_n\right)^{-1}, \tag{32}$$

and

$$\begin{aligned}\Psi_n(\theta,\Xi_n(\phi)) =& \frac{1}{n}\sum_{i=1}^{n}\psi(\theta,\Xi_n(\phi))\psi(\theta,\Xi_n(\phi))', \\ \psi(\theta,\Xi_n(\phi)) =& G_n'[\Xi_n(\phi)]^{-1}g(X_i,\theta) + G(X_i)'[\Xi_n(\phi)]^{-1}g_n(\theta) \\ & - G_n'[\Xi_n(\phi)]^{-1}\Xi(X_i,\phi)[\Xi_n(\phi)]^{-1}g_n(\theta), \\ \Xi_n(\phi) =& \frac{1}{n}\sum_{i=1}^{n}\Xi(X_i,\phi).\end{aligned} \tag{33}$$

When $\Xi_n(\phi) = \Xi(X_i,\phi) = I$, the last term of $\psi(\theta,\Xi_n(\phi))$ needs to be dropped. Note that

$\psi(\theta, \Xi_n(\phi))$ 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 corrected one, $\widehat{V}_c(\hat{\theta}_2)$. 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 (20)-(22). The standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{dc}(\hat{\theta}_1)/n}$.

4 Robustness to Misspecification

Both the Windmeijer corrected and the conventional variance estimators, $\widehat{V}_c(\hat{\theta}_2)$ and $\widetilde{V}(\hat{\theta}_2)$, are consistent for the asymptotic variance of $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ under correct model specification, $E[g(X_i, \theta_0)] = 0$. In words, this 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 equation 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 equation model as the best approximating model and allow for possible misspecification. Under misspecification, all the conventional GMM variance estimators, either finite sample corrected or not, are no longer consistent.

Based on the result of Hall and Inoue (2003), Lee (2014) proposes variance estimators for the one- and two-step GMM which are consistent regardless of misspecification (the misspecification-robust variance estimator, hereinafter). However, its finite sample behavior under correct specification has not been investigated, though it has been generally viewed less accurate than the conventional variance estimator because it includes additional terms that are presumed to be zero under correct specification.

We show that this intuition is not true: The doubly corrected variance estimator $\widehat{V}_{dc}(\hat{\theta}_2)$ is the misspecification-robust variance estimator. The double correction procedure in the previous Section is in fact equivalent to the one achieving robustness against misspecification.

We first consider a globally misspecified moment equation model which is defined as

$$E[g(X_i, \theta)] = \delta(\theta) \neq 0, \ \forall \theta \in \Theta, \tag{34}$$

where $\delta(\theta)$ is constant for n and Θ is a compact parameter space (Hall and Inoue, 2003). In this case, 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. This implies that 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. We may alternatively view the model as locally misspecified model, e.g., Otsu (2011) and Guggenberger (2012). Since the first-order asymptotic variance and the true value are not affected in this case, the analysis become trivial. Thus, we assume global misspecification in this Section to show the equivalence of the doubly corrected and the misspecification-robust formula.

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. Second, the additional correction that accounts for the effect of non-zero $g_n(\theta_0)$ corrects for the variability arising from non-zero $g_n(\theta_0)$ in the limit. Overall, the double correction makes the expansions (24)-(28) robust to misspecification.

To formally show that $\widehat{V}_{dc}(\hat{\theta}_2)$ is consistent for the asymptotic variance under misspecification, we first list some definitions and sufficient conditions. Let θ_1 and θ_2 be the pseudo-true values that correspond to $\hat{\theta}_1$ and $\hat{\theta}_2$. It may be that $\theta_1 \neq \theta_2$ but $\theta_1 = \theta_2 = \theta_0$ under correct specification. Define $g_j = E[g(X_i, \theta_j)]$ and $\Omega_j = \Omega(\theta_j)$ for $j = 1, 2$. Assume that for $j = 1, 2$,

$$
\begin{aligned}
\hat{\theta}_j - \theta_j =& O_p(n^{-1/2}), \\
G_n - G =& O_p(n^{-1/2}), \\
\text{vec}(\Omega_n(\theta_j) - \Omega_j) =& O_p(n^{-1/2}).
\end{aligned}
$$

Hall and Inoue (2003) provide a complete list of relevant definitions and sufficient conditions.

Take the FOC of the two-step GMM. We can use the expansions (7)-(9) by letting $\theta_1 \neq \theta_2$. Then (10) can be written as

$$
\begin{aligned}
\sqrt{n}(\hat{\theta}_2 - \theta_2) =& - \left\{ G_n'[\Omega_n(\theta_1)]^{-1} G_n \right\}^{-1} G_n'[\Omega_n(\theta_1)]^{-1} \sqrt{n} g_n(\theta_2) \\
&+ D_n^* \sqrt{n}(\hat{\theta}_1 - \theta_1) + O_p(n^{-1/2} \|g_n(\theta_2)\|) \\
=& \sqrt{n}(\tilde{\theta}_2^* - \theta_2) + D_n^* \sqrt{n}(\hat{\theta}_1 - \theta_1) + O_p(n^{-1/2} \|g_n(\theta_2)\|), \qquad (35)
\end{aligned}
$$

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

$$
\tilde{\theta}_2^* = \arg\min_{\theta \in \Theta} \; g_n(\theta)'[\Omega_n(\theta_1)]^{-1} g_n(\theta), \qquad (36)
$$

and

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

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 (35) 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 (36) can be expanded as

$$\sqrt{n}(\tilde{\theta}_2^* - \theta_2) = -\left\{G_n'[\Omega_n(\theta_1)]^{-1}G_n\right\}^{-1}\left\{G_n'[\Omega_n(\theta_1)]^{-1}\sqrt{n}\left(g_n(\theta_2) - g_2\right)\right.$$
$$\left.+\sqrt{n}\left(G_n - G\right)'\left[\Omega_n(\theta_1)\right]^{-1}g_2 - G'\Omega_1^{-1}\sqrt{n}\left(\Omega_n(\theta_1) - \Omega_1\right)\left[\Omega_n(\theta_1)\right]^{-1}g_2\right\}. \qquad (37)$$

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

$$\sqrt{n}(\hat{\theta}_1 - \theta_1) = -\left\{G_n'W_n^{-1}G_n\right\}^{-1}\left[G_n'W_n^{-1}\sqrt{n}\left(g_n(\theta_1) - g_1\right)\right.$$
$$\left.+\sqrt{n}(G_n - G)'W_n^{-1}g_1 - G'W^{-1}\sqrt{n}\left(W_n - W\right)W_n^{-1}g_1\right]. \qquad (38)$$

The expansions (37) and (38) are misspecification-robust versions of (8) and (9), being specific about the probability limits of the one-step and two-step GMM estimators and taking into account for non-zero $g_n(\theta_j)$ for $j = 1, 2$.

Noting that $G_n'[\Omega_n(\theta_1)]^{-1}g_n(\hat{\theta}_2) = G_n'W_n^{-1}g_n(\hat{\theta}_1) = 0$, it is straightforward to see that the estimators $\widehat{V}(\hat{\theta}_2)$, $\widehat{V}_{dc}(\hat{\theta}_1)$, and $\widehat{C}(\hat{\theta}_1, \hat{\theta}_2)$ given in (30)-(32) are consistent for the asymptotic (co)variances of (37) and (38). Finally $\widehat{D}_n - D_n^* = o_p(1)$ because $F_{1n}^* \xrightarrow{p} 0$. This proves our claim that $\widehat{V}_{dc}(\hat{\theta}_2)$ is the misspecification-robust variance estimator for the two-step GMM.

We next consider a locally misspecified moment condition with a sequence of local alternatives departed from (1) as

$$E[g(X_{in}, \theta_0)] = \frac{\delta_0}{\sqrt{n}} \qquad (39)$$

for some nonzero $\delta_0 \in \mathbb{R}^q$ that depends on θ_0. Note that the observations now form a triangular array $\{X_{in} : i = 1, ..., n,\ n \in \mathbb{N}\}$. Under the locally misspecified moment condition in (39) and some regular conditions, one can show that $\hat{\theta}_1$ and $\hat{\theta}_2$ are both $\sqrt{n}$-consistent estimators for the true parameter θ_0. This is different from the globally misspecified moment condition in (39) where the probability limit of GMM estimators differ by the choice of weighting matrix. We assume that

$$\sqrt{n}\bar{g}_n(\theta_0) := \sqrt{n}(g_n(\theta_0) - E[g(X_{in}, \theta_0)]) = O_p(1),$$

and thus $g_n(\theta_0)$ is $O_p(n^{-1/2})$. Then, the stochastic orders of expansions calculated under the correctly specified moment conditions in Section 3 does not change. This implies that under the locally misspecified moments our doubly corrected variance formula still provides the same order of finite sample correction upto $O_p(n^{-1/2})$. However, it is important to point out that the doubly corrected variance formula does not explictly identify whether the finite sample variation is caused by the local misspecification in (39) or the over-identified the sample moment process $g_n(\theta)$. To be more specific, let

$$\eta_0 = -\left\{G'\Omega^{-1}G\right\}^{-1}G'\Omega^{-1}\delta_0,$$

be the asymptotic bias parameter of $\hat{\theta}_2$ which is induced by the locallly misspecified moment conditions in (39), e.g. Hall (2005) and Andrews, Gentzkow, and Shapiro (2017). From the results in Hall (2005), Conley, Hansen, and Rossi (2012),

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) \xrightarrow{d} N(\eta_0, (G'\Omega^{-1}G)^{-1}).$$

Thus, it is natural to re-expand (24)–(28) considering the effect of local (psuedo) true value η_0 on the finite sample distribution of $\hat{\theta}_2$. Let $\tilde{\eta}_n$ be

$$\tilde{\eta}_n = -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1}\delta_0$$

an infeasible estimator of η_n. Then,

$$\sqrt{n}(\hat{\theta}_2 - \theta_0) + \eta_0 = (\eta_0 - \tilde{\eta}_n) - \left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}\left[G'\Omega^{-1}\sqrt{n}\bar{g}(\theta_0)\right. \tag{40}$$

$$\left. +\sqrt{n}(G_n - G)'\Omega^{-1}\bar{g}(\theta_0) - G'\Omega^{-1}\sqrt{n}\left(\Omega_n(\theta_0) - \Omega\right)\Omega^{-1}\bar{g}(\theta_0)\right] \tag{41}$$

$$+ D_n\left\{G_n'W_n^{-1}G_n\right\}^{-1}\left[G'W^{-1}\sqrt{n}\bar{g}(\theta_0)\right. \tag{42}$$

$$\left. +\sqrt{n}(G_n - G)'W^{-1}\bar{g}(\theta_0) - G'W^{-1}\sqrt{n}\left(W_n - W\right)W^{-1}\bar{g}(\theta_0)\right] \tag{43}$$

$$+ O_p(n^{-1}). \tag{44}$$

The asymptotic bias term $\tilde{\eta}_n$ in the rightside of (40) deviates from the population bias term η_0, and the difference between two quantities can cause an extra source of variations in $\sqrt{n}(\hat{\theta}_2 - \theta_0)$. The stochastic order of this difference is $O_p(n^{-1/2})$ which is up to the same as the order of the second term in (40)-(44). However, the local misspecification parameter δ_0 in η_0 cannot be consistently estimated in any types of GMM models. Therefore, both practically and theoretically, it is not possible to identify the exact nature of variations of GMM estimators in the locally misspecified moment conition. Neverthelss, our doubely corrected variance formula corrects the extra variations of $\hat{\theta}_2$ from the local misspecifiction as well as the finite sample moment process.

All the results in the previous Sections including the current one still hold if we replace $\Omega_n(\theta)$ with the centered weight matrix

$$\Omega_n^c(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(g(X_i,\theta) - g_n(\theta)\right)\left(g(X_i,\theta) - g_n(\theta)\right)'. \tag{45}$$

The matrices D_n and $\widehat{D}_n$ would need to be modified accordingly. The centered weight matrix is consistent for the covariance matrix of the moment equation under misspecification (so we would need to define Ω_j as the covariance matrix rather than the second moment). B. Hansen (2018) recommends using the centered weight matrix for this reason. Hall (2000) shows that the GMM over-identification test statistic with a centered heteroskedasticity-and-autocorrelation-consistent (HAC) weight matrix leads to more powerful tests in the time series setting.

5 Iterated GMM and Continuously Updating GMM

Both the Windmeijer and the double correction correct for the extra variation due to the weight matrix being evaluated at a preliminary estimate. A natural question is whether similar finite sample corrections can be obtained if the weight matrix is evaluated at the estimator itself, rather than a preliminary estimate. There are two existing GMM estimators that have this property: the iterated GMM (B. Hansen and Lee, 2018a) and the continuously-updating (CU) GMM (L. Hansen, Heaton, and Yaron, 1996). We show that the answer is yes for the iterated GMM and provide a doubly corrected (and misspecification-robust) formula. For the CU GMM, however, it seems difficult to obtain a simple finite sample correction due to the additional $O_p(n^{-1})$ term in the FOC which is difficult to estimate.

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

$$0 = G_n'[\Omega_n(\hat{\theta})]^{-1}g_n(\hat{\theta}) \tag{46}$$

where $\hat{\theta}$ is the iterated GMM. Let θ_0 be the true value or the pseudo-true value. Assume that $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ whose sufficient conditions are provided in B. Hansen and Lee (2018a). By applying the first-order Taylor expansion around θ_0 to $g_n(\hat{\theta})$ and $\Omega_n(\hat{\theta})$ sequentially

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\theta_0) =& -\left\{G_n'[\Omega_n(\hat{\theta})]^{-1}G_n\right\}^{-1} G_n'[\Omega_n(\hat{\theta})]^{-1}\sqrt{n}g_n(\theta_0)\\ =& -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n}g_n(\theta_0)\\ &+ D_n\sqrt{n}(\hat{\theta}-\theta_0) + O_p(n^{-1/2}\|g_n(\theta_0)\|)\end{aligned}$$

and thus

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\theta_0) =& -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n(I_k - D_n)\right\}^{-1} G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n}g_n(\theta_0) & (47)\\ &+ O_p(n^{-1/2}\|g_n(\theta_0)\|). & (48)\end{aligned}$$

Since $g_n(\theta_0) = O_p(n^{-1/2})$, we can get the same order of accuracy with the doubly corrected two-step GMM variance estimator by estimating the variance of the RHS of (47). To account for the variability of non-zero $g_n(\theta_0)$, we further expand (47)-(48) as

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\theta_0) =& -\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n(I_k - D_n)\right\}^{-1} \left[G'\Omega^{-1}\sqrt{n}g_n(\theta_0)\right.\\ &\left.+\sqrt{n}(G_n - G)'\Omega^{-1}g_n(\theta_0) - G'\Omega^{-1}\sqrt{n}\left(\Omega_n(\theta_0)-\Omega\right)\Omega^{-1}g_n(\theta_0)\right]\\ &+ O_p(n^{-1/2}\|g_n(\theta_0)\|).\end{aligned}$$

Similar to the two-step GMM estimator, the order of finite sample corrections depends on the degree of misspecification. Under correct specification, the order is up to $O_p(n^{-1/2})$ while the order is $O_p(1)$ under misspecification so that it corrects the first-order asymptotic variance.

Now the doubly corrected variance estimator is

$$\widehat{V}_{dc}(\hat{\theta}) = \{G_n'[\Omega_n(\hat{\theta})]^{-1}G_n(I_k - \widehat{D}_n)\}^{-1}\Psi_n(\hat{\theta},\Omega_n(\hat{\theta}))\{G_n'[\Omega_n(\hat{\theta})]^{-1}G_n(I_k - \widehat{D}_n)\}^{-1\prime}, \tag{49}$$

where $\Psi_n(\hat{\theta},\Omega_n(\hat{\theta}))$ is defined in (33). Not surprisingly, this formula is identical to the misspecification-robust variance estimator for the iterated GMM of Hansen and Lee (2018a) but they do not discuss finite sample correction of the iterated GMM variance estimator.

Windmeijer (2000) proposes a finite sample corrected variance estimator for the iterated GMM based on a similar argument with the two-step GMM. The formula is

$$\widehat{V}_c(\hat{\theta}) = (I_k - \widehat{D}_n)^{-1}\left(G_n'[\Omega_n(\hat{\theta})]^{-1}G_n\right)^{-1}(I_k - \widehat{D}_n)^{-1\prime}. \tag{50}$$

For simplicity, we also call this variance estimator the Windmeijer corrected one for the iterated GMM.

On the other hand, it is difficult to obtain a similar finite sample correction for the CU GMM. For simplicity, let $k = 1$ so that θ is scalar and assume correct specification. The FOC is

$$0 = G_n'[\Omega_n(\hat{\theta})]^{-1}g_n(\hat{\theta}) - \frac{1}{2}S_n(\hat{\theta})'\left([\Omega_n(\hat{\theta})]^{-1}g_n(\hat{\theta}) \otimes [\Omega_n(\hat{\theta})]^{-1}g_n(\hat{\theta})\right), \tag{51}$$

where

$$S_n(\theta) = \frac{\partial}{\partial\theta'}\text{vec}(\Omega_n(\theta)).$$

The second term of the RHS of (51) is $O_p(n^{-1})$ which is specific to the CU GMM. Let

$$\begin{aligned}
A_n(\theta) &= \left([\Omega_n(\theta)]^{-1}G_n \otimes [\Omega_n(\theta)]^{-1}g_n(\theta_0)\right) + \left([\Omega_n(\theta)]^{-1}g_n(\theta_0) \otimes [\Omega_n(\theta)]^{-1}G_n\right), \\
B_n(\theta) &= \left([\Omega_n(\theta)]^{-1}G_n \otimes [\Omega_n(\theta)]^{-1}G_n\right)(\hat{\theta}-\theta_0)^2 + \left([\Omega_n(\theta)]^{-1}g_n(\theta_0) \otimes [\Omega_n(\theta)]^{-1}g_n(\theta_0)\right).
\end{aligned}$$

By the first-order Taylor expansion of $g_n(\hat{\theta})$ around θ_0 in (51)

$$0 = G_n'[\Omega_n(\hat{\theta})]^{-1}g_n(\theta_0) + G_n'[\Omega_n(\hat{\theta})]^{-1}G_n(\hat{\theta}-\theta_0) - \frac{1}{2}S_n(\hat{\theta})'(A_n(\hat{\theta})(\hat{\theta}-\theta_0) + B_n(\hat{\theta})).$$

Since $A_n(\theta_0) = O_p(n^{-1/2})$ and $B_n(\theta_0) = O_p(n^{-1})$, the first-order Taylor expansion of $\Omega_n(\hat{\theta})$ around θ_0 gives

$$\begin{aligned}
0 = &\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}G_n'[\Omega_n(\theta_0)]^{-1}g_n(\theta_0) \\
&+ \left(I_k - D_n - \frac{1}{2}\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}S_n(\theta_0)'A_n(\theta_0)\right)(\hat{\theta}-\theta_0) \\
&- \frac{1}{2}\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}S_n(\theta_0)'B_n(\theta_0) + O_p(n^{-3/2}).
\end{aligned}$$

Rearranging terms and multiplying $\sqrt{n}$ on both sides we get

$$\sqrt{n}(\hat{\theta}-\theta_0) = -R_n^{-1}\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}G_n'[\Omega_n(\theta_0)]^{-1}\sqrt{n}g_n(\theta_0) \tag{52}$$

$$+\frac{1}{2}R_n^{-1}\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}S_n(\theta_0)'\sqrt{n}B_n(\theta_0) \tag{53}$$

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

where

$$R_n = I_k - D_n - \frac{1}{2}\left\{G_n'[\Omega_n(\theta_0)]^{-1}G_n\right\}^{-1}S_n(\theta_0)'A_n(\theta_0).$$

Since $B_n(\theta_0) = O_p(n^{-1})$, the term in (53) is $O_p(n^{-1/2})$. So if the extra variation caused by this term can be handled then a finite sample correction to the variance would be possible. However, $B_n(\theta_0)$ may not be accurately estimated because it includes $(\hat{\theta}-\theta_0)^2$ term. Thus, even when the variance of the RHS of (52) is corrected via R_n, it does not provide a finite sample correction compared to the usual first-order asymptotic approximation

$$\sqrt{n}(\hat{\theta}-\theta_0) = \left\{G_n'[\Omega_n(\hat{\theta})]^{-1}G_n\right\}^{-1}G_n'[\Omega_n(\hat{\theta})]^{-1}\sqrt{n}g_n(\theta_0) + O_p(n^{-1/2}). \tag{54}$$

6 Examples

6.1 Cross-sectional IV

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

$$\hat{\theta}_1 = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y. \tag{55}$$

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

$$\hat{\theta}_2 = (X'Z\widehat{\Omega}_1^{-1}Z'X)^{-1}X'Z\widehat{\Omega}_1^{-1}Z'Y \tag{56}$$

where

$$\widehat{\Omega}_1 = \frac{1}{n}\sum_{i=1}^{n} Z_iZ_i'\hat{e}_{1i}^2,$$

$$\hat{e}_{1i} = y_i - X_i'\hat{\theta}_1.$$

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

$$\widehat{V}_{dc}(\hat{\theta}_1) = \left(\frac{1}{n}X'Z(Z'Z)^{-1}Z'X\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_{1i}\hat{\psi}_{1i}'\left(\frac{1}{n}X'Z(Z'Z)^{-1}Z'X\right)^{-1}, \tag{57}$$

$$\widehat{V}_{dc}(\hat{\theta}_2) =\widehat{V}(\hat{\theta}_2) + \widehat{D}_n\widehat{C}(\hat{\theta}_1,\hat{\theta}_2) + \widehat{C}(\hat{\theta}_1,\hat{\theta}_2)'\widehat{D}_n' + \widehat{D}_n\widehat{V}_{dc}(\hat{\theta}_1)\widehat{D}_n', \tag{58}$$

where

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

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

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

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

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

It is worth observing that the doubly corrected variance estimator $\widehat{V}_{dc}(\hat{\theta}_2)$ reduces to the Windmeijer corrected one $\widehat{V}_c(\hat{\theta}_2)$ if (i) the last two terms in $\hat{\psi}_{2i}$ and $\hat{\psi}_{1i}$ are ignored and (ii) $\hat{e}_{1i}$ replaces $\hat{e}_{2i}$ in $\hat{\psi}_{2i}$. 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, $\widehat{V}_{dc}(\hat{\theta}_2) \neq \widehat{V}_c(\hat{\theta}_2)$ because $Z'\hat{e}_j \neq 0$ for $j = 1, 2$, so the last two terms of $\hat{\psi}_{ji}$ are non-zero. Furthermore, it is critical (and reasonable) to use $\hat{e}_{2i}$ in $\hat{\psi}_{2i}$ 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

$$\hat{\theta}_s = (X'Z\widehat{\Omega}_{s-1}^{-1}Z'X)^{-1}X'Z\widehat{\Omega}_{s-1}^{-1}Z'Y, \tag{59}$$

where

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

We iterate the s-step GMM estimator until convergence given a preset tolerance, 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

$$\widehat{V}_{dc}(\hat{\theta}) = \widehat{H}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i\hat{\psi}_i'\right)\widehat{H}^{-1\prime}, \tag{60}$$

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

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

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

$$\widehat{V}_c(\hat{\theta}) = \widehat{H}^{-1}\left(\frac{1}{n^2}X'Z\widehat{\Omega}^{-1}Z'X\right)\widehat{H}^{-1\prime}, \tag{61}$$

$$\widetilde{V}(\hat{\theta}) = \left(\frac{1}{n^2}X'Z\widehat{\Omega}^{-1}Z'X\right)^{-1}. \tag{62}$$

6.2 A Panel Data Model

Consider a panel data model with a scalar regressor

$$y_{it} = x_{it}\beta + \eta_i + v_{it}, \tag{63}$$

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 (63) 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

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

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

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

$$\widehat{\Omega}_1 = \frac{1}{n} \sum_{i=1}^N Z_i' \Delta \hat{v}_{1i} \Delta \hat{v}_{1i}' Z_i. \tag{64}$$

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

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

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

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

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

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

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

where

$$\widetilde{V}(\hat{\beta}_2) = n^2\left(\Delta X'Z\widehat{\Omega}_1^{-1}Z'\Delta X\right)^{-1}. \tag{65}$$

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

$$\hat{\beta}_s = (\Delta X'Z\widehat{\Omega}_{s-1}^{-1}Z'\Delta X)^{-1}\Delta X'Z\widehat{\Omega}_{s-1}^{-1}Z'\Delta Y, \tag{66}$$

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

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

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

$$\widehat{V}_c(\hat{\beta}) = \widehat{H}^{-1}\left(\frac{1}{n^2}\Delta X'Z\widehat{\Omega}^{-1}Z'\Delta X\right)\widehat{H}^{-1\prime}, \tag{67}$$

$$\widetilde{V}(\hat{\beta}) = \left(\frac{1}{n^2}\Delta X'Z\widehat{\Omega}^{-1}Z'\Delta X\right)^{-1}. \tag{68}$$

7 Simulation

We investigate the finite sample performance of the doubly corrected standard errors proposed in this paper and provide a thorough comparison with the conventional and the Windmeijer corrected one 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 (45). 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.

7.1 Cross-sectional IV

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

$$\begin{aligned} y_i &= x_i\beta_0 + e_i \\ E\left(z_i e_i\right) &= 0 \end{aligned} \tag{69}$$

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

$$\begin{aligned} y_i &= x_i\beta_0 + \frac{\alpha_0}{\sqrt{n}}\left(z_{1i} - z_{2i} + z_{3i} - z_{4i}\right) + e_i, \qquad (70) \\ x_i &= \pi_0\left(z_{1i} + z_{2i} + z_{3i} + z_{4i}\right) + u_i, \\ e_i &= 0.5u_i + \sqrt{1-0.5^2}v_i, \\ z_i &\sim N\left(0, I_4\right), \begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & z_{1i}^2 \end{bmatrix}\right). \end{aligned}$$

We set $\beta_0 = 1$, vary 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 (69) 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$_c$ $\hat{\beta}$), 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.

7.2 Linear Dynamic Panel Model

7.2.1 Random Coefficient

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

$$y_{it} = \rho_0 y_{i,t-1} + \eta_i + \nu_{it}, \tag{71}$$

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:

$$E[y_{i,t-s}(\Delta y_{it} - \rho_0 \Delta y_{i,t-1})] = 0, \;\; t = 3, ...T, \text{ and } s \geq 2, \tag{72}$$

The moment conditions are derived from taking differences of (71), 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 (72). A sufficient condition for this to hold is that the model (71) coincides with the true DGP, but this is unlikely to be true. A reasonable deviation from the assumed model (71) is heterogeneity in ρ_0 across i. We assume the following DGP. For $i = 1, ..., N$ and $t = 1, ...T$,

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

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 equation model can be written as

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

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 equation model is correctly specified and the estimand is $E[\rho_i]$. Otherwise in general, the moment equation model fails to hold at a single unique parameter value because each of the moment equation 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 $n = 500$.

7.2.2 Misspecified Lag Length

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

$$y_{it} = \beta_0 x_{it} + \eta_i + v_{it}, \tag{73}$$

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 6.2.

The DGP is

$$
\begin{aligned}
y_{it} &= \beta_0 x_{it} + \alpha_0 x_{it-1} + \eta_i + v_{it}, \\
x_{it} &= 0.5 x_{it-1} + \eta_i + 0.5 v_{it-1} + \epsilon_{it}, \\
\eta &\sim N(0,1) \text{ and } \epsilon_{it} \sim N(0,1), \\
v_{it} &= \delta_i \tau_t \omega_{it} \text{ and } \omega_{it} \sim \chi_1^2 - 1.
\end{aligned} \tag{74}
$$

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 (74) governs the degree of misspecification. When $\alpha_0 = 0$, the model (73) 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 $\{0, 0.025, 0.05, 0.1, 0.2, 0.3, 0.4\}$. 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 $N = 500$.

8 Conclusion

We propose doubly corrected standard errors for the one-step, two-step, and iterated GMM estimators in the linear over-identified model. We show that the doubly corrected variance estimators are robust to misspecification. Under correct specification, the double correction provides finite sample correction upon the conventional variance estimator up to the same order with that of Windmeijer (2005). Under misspecification, the doubly corrected variance estimator remains consistent while the conventional and the Windmeijer corrected variance estimators are inconsistent.

References

[1] Andrews, I. (forthcoming). On the structure of IV estimands. *Journal of Econometrics*

[2] 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.

[3] 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.

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

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

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

[7] 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.

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

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

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

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

[12] Hall, A. R. (2005). Generalized method of moments. Oxford university press.

[13] 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.

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

[15] Hansen, B. E. (2018). Econometrics. Unpublished Manuscript. University of Wisconsin.

[16] Hansen, B. E., and Lee, S. (2018a). Inference for iterated GMM under misspecification and clustering. Working paper.

[17] Hansen, B. E., and Lee, S. (2018b). Asymptotic theory for clustered samples. Working paper.

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

[19] 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.

[20] Hwang, J., and Sun, Y. (forthcoming). 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*.

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

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

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

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

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

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

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

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

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

[30] 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.

[31] 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
	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
	se$_{wc}$ $\hat{\beta}_2$	0.2889	0.2900	0.2945	0.3039	0.3133	0.3281
	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
	se$_{wc}$ $\hat{\beta}_{iter}$	0.2850	0.2859	0.2919	0.3028	0.3140	0.3316
	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
	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
	se$_{wc}$ $\hat{\beta}_2$	0.2089	0.2087	0.2099	0.2130	0.2169	0.2221
	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
	se$_{wc}$ $\hat{\beta}_{iter}$	0.2073	0.2079	0.2101	0.2140	0.2187	0.2248
	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
	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
	se$_{wc}$ $\hat{\beta}_2$	0.0958	0.0957	0.0956	0.0958	0.0958	0.0962
	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
	se$_{wc}$ $\hat{\beta}_{iter}$	0.0957	0.0958	0.0959	0.0964	0.0966	0.0972
	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$_c$ $\hat{\rho}_2$	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$_c$ $\hat{\rho}$	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$_c$ $\hat{\rho}_2$	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$_c$ $\hat{\rho}$	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_c $\hat{\rho}_2$	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_c $\hat{\rho}$	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_c $\hat{\rho}_2$	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_c $\hat{\rho}$	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	0.2	0.3	0.4
$N = 100$	$\hat{\beta}_1$	0.9793	0.9534	0.9268	0.8744	0.7702	0.6647	0.5590
$T = 4$	sd $\hat{\beta}_1$	0.1521	0.1529	0.1538	0.1588	0.1727	0.1925	0.2196
	se $\hat{\beta}_1$	0.1469	0.1468	0.1465	0.1472	0.1513	0.1589	0.1699
	se_{dc} $\hat{\beta}_1$	0.1546	0.1555	0.1563	0.1604	0.1744	0.1944	0.2191
	$\hat{\beta}_2$	0.9849	0.9588	0.9322	0.8773	0.7587	0.6238	0.4795
	sd $\hat{\beta}_2$	0.1404	0.1422	0.1451	0.1552	0.1843	0.2207	0.2622
	se $\hat{\beta}_2$	0.1243	0.1244	0.1242	0.1253	0.1303	0.1381	0.1481
	se_c $\hat{\beta}_2$	0.1390	0.1402	0.1415	0.1473	0.1669	0.1919	0.2192
	se_{dc} $\hat{\beta}_2$	0.1343	0.1366	0.1390	0.1482	0.1775	0.2146	0.2551
	$\hat{\beta}$	0.9858	0.9599	0.9334	0.8781	0.7533	0.5977	0.4147
	sd $\hat{\beta}$	0.1417	0.1439	0.1474	0.1600	0.2000	0.2524	0.3081
	se $\hat{\beta}$	0.1243	0.1244	0.1242	0.1253	0.1303	0.1381	0.1481
	se_c $\hat{\beta}$	0.1393	0.1408	0.1426	0.1507	0.1806	0.2230	0.2695
	se_{dc} $\hat{\beta}$	0.1352	0.1378	0.1406	0.1517	0.1896	0.2391	0.2899
$N = 100$	$\hat{\beta}_1$	0.9755	0.9577	0.9411	0.9060	0.8368	0.7676	0.6993
$T = 6$	sd $\hat{\beta}_1$	0.1027	0.1043	0.1051	0.1083	0.1172	0.1288	0.1427
	se $\hat{\beta}_1$	0.1002	0.1003	0.1004	0.1013	0.1037	0.1077	0.1130
	se_{dc} $\hat{\beta}_1$	0.1056	0.1064	0.1075	0.1107	0.1192	0.1306	0.1442
	$\hat{\beta}_2$	0.9833	0.9649	0.9466	0.9080	0.8238	0.7318	0.6367
	sd $\hat{\beta}_2$	0.0906	0.0929	0.0948	0.1017	0.1213	0.1431	0.1666
	se $\hat{\beta}_2$	0.0716	0.0718	0.0720	0.0731	0.0760	0.0801	0.0849
	se_c $\hat{\beta}_2$	0.0905	0.0914	0.0930	0.0978	0.1117	0.1285	0.1460
	se_{dc} $\hat{\beta}_2$	0.0836	0.0853	0.0876	0.0944	0.1131	0.1357	0.1598
	$\hat{\beta}$	0.9857	0.9671	0.9484	0.9083	0.8124	0.6885	0.5350
	sd $\hat{\beta}$	0.0946	0.0970	0.0998	0.1099	0.1427	0.1858	0.2348
	se $\hat{\beta}$	0.0716	0.0718	0.0720	0.0731	0.0760	0.0801	0.0849
	se_c $\hat{\beta}$	0.0937	0.0952	0.0976	0.1054	0.1320	0.1722	0.2202
	se_{dc} $\hat{\beta}$	0.0866	0.0885	0.0916	0.1006	0.1287	0.1675	0.2107

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

	γ_0	0	0.025	0.05	0.1	0.2	0.3	0.4
$N = 500$	$\hat{\beta}_1$	0.9958	0.9683	0.9406	0.8853	0.7754	0.6650	0.5546
$T = 4$	sd $\hat{\beta}_1$	0.0685	0.0686	0.0695	0.0716	0.0787	0.0889	0.1011
	se $\hat{\beta}_1$	0.0679	0.0678	0.0677	0.0679	0.0698	0.0734	0.0785
	se$_{dc}$ $\hat{\beta}_1$	0.0686	0.0690	0.0696	0.0716	0.0787	0.0889	0.1013
	$\hat{\beta}_2$	0.9970	0.9710	0.9443	0.8892	0.7660	0.6225	0.4643
	sd $\hat{\beta}_2$	0.0652	0.0657	0.0676	0.0730	0.0891	0.1092	0.1291
	se $\hat{\beta}_2$	0.0632	0.0631	0.0630	0.0634	0.0657	0.0693	0.0739
	se$_c$ $\hat{\beta}_2$	0.0648	0.0653	0.0662	0.0694	0.0798	0.0930	0.1069
	se$_{dc}$ $\hat{\beta}_2$	0.0634	0.0643	0.0658	0.0708	0.0865	0.1060	0.1269
	$\hat{\beta}$	0.9970	0.9712	0.9446	0.8897	0.7635	0.6014	0.4007
	sd $\hat{\beta}$	0.0652	0.0658	0.0678	0.0738	0.0946	0.1238	0.1510
	se $\hat{\beta}$	0.0632	0.0631	0.0630	0.0634	0.0657	0.0693	0.0739
	se$_c$ $\hat{\beta}$	0.0648	0.0652	0.0662	0.0698	0.0839	0.1052	0.1279
	se$_{dc}$ $\hat{\beta}$	0.0634	0.0643	0.0659	0.0715	0.0911	0.1185	0.1458
$N = 500$	$\hat{\beta}_1$	0.9947	0.9758	0.9564	0.9181	0.8423	0.7662	0.6895
$T = 6$	sd $\hat{\beta}_1$	0.0473	0.0476	0.0482	0.0498	0.0543	0.0600	0.0673
	se $\hat{\beta}_1$	0.0469	0.0470	0.0470	0.0473	0.0485	0.0504	0.0530
	se$_{dc}$ $\hat{\beta}_1$	0.0475	0.0479	0.0484	0.0499	0.0543	0.0602	0.0672
	$\hat{\beta}_2$	0.9968	0.9778	0.9582	0.9175	0.8276	0.7251	0.6148
	sd $\hat{\beta}_2$	0.0432	0.0440	0.0454	0.0494	0.0602	0.0725	0.0854
	se $\hat{\beta}_2$	0.0408	0.0408	0.0410	0.0414	0.0428	0.0448	0.0471
	se$_c$ $\hat{\beta}_2$	0.0431	0.0436	0.0445	0.0471	0.0548	0.0640	0.0736
	se$_{dc}$ $\hat{\beta}_2$	0.0413	0.0421	0.0434	0.0470	0.0573	0.0694	0.0822
	$\hat{\beta}$	0.9969	0.9779	0.9584	0.9174	0.8219	0.6961	0.5330
	sd $\hat{\beta}$	0.0433	0.0441	0.0457	0.0504	0.0659	0.0882	0.1143
	se $\hat{\beta}$	0.0408	0.0408	0.0410	0.0414	0.0428	0.0448	0.0471
	se$_c$ $\hat{\beta}$	0.0431	0.0437	0.0447	0.0479	0.0592	0.0770	0.0993
	se$_{dc}$ $\hat{\beta}$	0.0414	0.0423	0.0437	0.0479	0.0620	0.0823	0.1061

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