Economics Bulletin

Volume 40, Issue 4

The equivalence of two-step first difference and forward orthogonal deviations GMM

Robert F. Phillips
George Washington University

Abstract

Generalized method of moments (GMM) estimation of a dynamic panel data model often relies on transforming the data. This note provides a necessary and sufficient condition on the instruments for two-step GMM based on differencing to be equivalent to two-step GMM based on forward orthogonal deviations. The same condition is necessary and sufficient for system GMM, based on differencing, to be equal to system GMM using forward orthogonal deviations.

I am grateful to an anonymous referee for suggestions that improved the paper.
Citation: Robert F. Phillips, (2020) "The equivalence of two-step first difference and forward orthogonal deviations GMM", *Economics Bulletin*, Volume 40, Issue 4, pages 2865-2871
Contact: Robert F. Phillips - rphil@gwu.edu.
Submitted: August 10, 2020. **Published:** October 23, 2020.

1 Introduction

Schmidt et al. (1992) and Arellano and Bover (1995) showed that, under certain conditions, generalized method of moments (GMM) estimation of a panel data model is invariant to transformation. Specifically, the same GMM estimator can be obtained using two different transformations of the data, both of which remove fixed effects. They identified using all available instruments as sufficient for invariance to transformation conclusions.

Phillips (2019), however, showed that using all available instruments is unnecessary and provided a sufficient *and necessary* instruments condition assuring the same GMM estimate can be calculated two ways. The importance of having a necessary and sufficient condition is that it tells us not just when two different transformations lead to the same GMM estimator, it also tells us when two different transformations cannot lead to the same estimator. If two different transformations do not lead to the same GMM estimator, it becomes worthwhile to ask which transformation produces the estimator with better sampling properties (see, e.g., Hayakawa, 2009; Phillips, 2020). Moreover, Phillips (2020) demonstrated that even when the instruments condition is satisfied and, therefore, two different transformations lead to the same estimate, the transformations are not necessarily the same in terms of computational efficiency. One transformation can produce a computational algorithm that is orders of magnitude faster than the algorithm based on the other transformation (Phillips, 2020).

However, Phillips (2019) did not cover GMM estimation when optimal weighting is used in the presence of conditional heteroskedasticity. This note extends the numerical equivalency result in Phillips (2019) to GMM when optimal weighting is used. It also extends the equivalency result to system GMM estimation of dynamic panel data models. The results provided here and in Phillips (2019) indicate that the necessary and sufficient instruments condition for the numerical equivalency of different transformations applies to a broad range of cases.

2 Numerically equivalent transformations

Consider the model

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\iota}_T\eta_i + \boldsymbol{v}_i, \qquad i = 1,\ldots,n. \tag{1}$$

Here $\boldsymbol{X}_i$ is a $T\times K$ matrix of observations on the explanatory variables for the ith individual, $\boldsymbol{v}_i$ is a vector of errors, η_i is a scalar unobserved fixed effect, and $\boldsymbol{\iota}_T$ is a $T\times 1$ vector of ones.

When lags of the dependent variable appear in $\boldsymbol{X}_i$, the parameters in $\boldsymbol{\beta}$ are often estimated by applying two-step GMM after the data are transformed. The transformation removes the fixed effect η_i. Let $\boldsymbol{K}$ be the $R\times T$ transformation matrix that satisfies $\boldsymbol{K}\boldsymbol{\iota}_T = \boldsymbol{0}$.

In order to write the formula for a two-step GMM estimator, let $\widehat{\boldsymbol{\beta}}$ denote an initial estimator of $\boldsymbol{\beta}$ and set $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}$. Moreover, let $\tilde{\boldsymbol{y}}_i = \boldsymbol{K}\boldsymbol{y}_i$, $\tilde{\boldsymbol{X}}_i = \boldsymbol{K}\boldsymbol{X}_i$, and $\tilde{\boldsymbol{e}}_i = \boldsymbol{K}\boldsymbol{e}_i$ $(i = 1,\ldots,n)$. Next, let $\boldsymbol{z}_{it}$ be a $k_t\times 1$ vector of instruments $(t = 1,\ldots,R)$, and let $\boldsymbol{Z}_i$ be a

$R \times \sum_{t=1}^{R} k_t$ block-diagonal matrix given by

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{z}_{i1}' & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{z}_{i2}' & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{z}_{iR}' \end{pmatrix}. \tag{2}$$

Then the two-step GMM estimator of $\boldsymbol{\beta}$, based on the transformation matrix $\boldsymbol{K}$, is

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_K &= \left[\sum_i \tilde{\boldsymbol{X}}_i' \boldsymbol{Z}_i \left(\sum_i \boldsymbol{Z}_i' \tilde{\boldsymbol{e}}_i \tilde{\boldsymbol{e}}_i' \boldsymbol{Z}_i \right)^{-1} \sum_i \boldsymbol{Z}_i' \tilde{\boldsymbol{X}}_i \right]^{-1} \\ &\quad \times \sum_i \tilde{\boldsymbol{X}}_i' \boldsymbol{Z}_i \left(\sum_i \boldsymbol{Z}_i' \tilde{\boldsymbol{e}}_i \tilde{\boldsymbol{e}}_i' \boldsymbol{Z}_i \right)^{-1} \sum_i \boldsymbol{Z}_i' \tilde{\boldsymbol{y}}_i. \end{aligned} \tag{3}$$

Given a suitable restriction on the instruments, the estimator $\widehat{\boldsymbol{\beta}}_K$ can be written another way provided $\boldsymbol{K}$ is such that $\boldsymbol{K}\boldsymbol{K}'$ is positive definite. Specifically, set $\boldsymbol{F} = \boldsymbol{U}\boldsymbol{K}$, where $\boldsymbol{U}$ is the upper-triangular Cholesky factorization of $(\boldsymbol{K}\boldsymbol{K}')^{-1}$. Next, let $\ddot{\boldsymbol{y}}_i = \boldsymbol{F}\boldsymbol{y}_i$, $\ddot{\boldsymbol{X}}_i = \boldsymbol{F}\boldsymbol{X}_i$, and $\ddot{\boldsymbol{e}}_i = \boldsymbol{F}\boldsymbol{e}_i$ $(i = 1, \ldots, n)$. Finally, define

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_F &= \left[\sum_i \ddot{\boldsymbol{X}}_i' \boldsymbol{Z}_i \left(\sum_i \boldsymbol{Z}_i' \ddot{\boldsymbol{e}}_i \ddot{\boldsymbol{e}}_i' \boldsymbol{Z}_i \right)^{-1} \sum_i \boldsymbol{Z}_i' \ddot{\boldsymbol{X}}_i \right]^{-1} \\ &\quad \times \sum_i \ddot{\boldsymbol{X}}_i' \boldsymbol{Z}_i \left(\sum_i \boldsymbol{Z}_i' \ddot{\boldsymbol{e}}_i \ddot{\boldsymbol{e}}_i' \boldsymbol{Z}_i \right)^{-1} \sum_i \boldsymbol{Z}_i' \ddot{\boldsymbol{y}}_i. \end{aligned} \tag{4}$$

The estimators $\widehat{\boldsymbol{\beta}}_K$ and $\widehat{\boldsymbol{\beta}}_F$ are defined if the inverses in expressions (3) and (4) exist. This fact imposes a restriction on the number of instruments. Specifically, we must have $\sum_{t=1}^{R} k_t \leq n$. To see this, consider the $\sum_{t=1}^{R} k_t \times \sum_{t=1}^{R} k_t$ matrix $\sum_i \boldsymbol{Z}_i' \ddot{\boldsymbol{e}}_i \ddot{\boldsymbol{e}}_i' \boldsymbol{Z}_i$ in (4). We have $\sum_i \boldsymbol{Z}_i' \ddot{\boldsymbol{e}}_i \ddot{\boldsymbol{e}}_i' \boldsymbol{Z}_i = \boldsymbol{W}'\boldsymbol{W}$ for $\boldsymbol{W}' = \begin{pmatrix} \boldsymbol{Z}_1' \ddot{\boldsymbol{e}}_1 & \boldsymbol{Z}_2' \ddot{\boldsymbol{e}}_2 & \ldots & \boldsymbol{Z}_n' \ddot{\boldsymbol{e}}_n \end{pmatrix}$. Moreover, $\text{rank}(\boldsymbol{W}'\boldsymbol{W}) = \text{rank}(\boldsymbol{W})$ (see, e.g., Greene, 2012, p. 986). The dimension of the matrix $\boldsymbol{W}$ is $n \times \sum_{t=1}^{R} k_t$. Therefore, $n \geq \sum_{t=1}^{R} k_t$ is necessary for the inverse $\left(\sum_i \boldsymbol{Z}_i' \ddot{\boldsymbol{e}}_i \ddot{\boldsymbol{e}}_i' \boldsymbol{Z}_i\right)^{-1}$ to exist.

Assuming the estimators $\widehat{\boldsymbol{\beta}}_K$ and $\widehat{\boldsymbol{\beta}}_F$ are defined, Theorem 1 provides a condition on the instruments that is necessary and sufficient for $\widehat{\boldsymbol{\beta}}_F = \widehat{\boldsymbol{\beta}}_K$.

Theorem 1. *A necessary and sufficient condition for* $\widehat{\boldsymbol{\beta}}_F = \widehat{\boldsymbol{\beta}}_K$ *is*

C1: every entry in $\boldsymbol{z}_{is}$ *is a linear combination of entries in* $\boldsymbol{z}_{it}$ $(s = 1, \ldots, t,\ t = 1, \ldots, R)$.

Proofs are provided in the appendix.

The estimator given by (3) uses an asymptotically optimal weighing matrix in the sense that the GMM estimator based on this weighting matrix is efficient among all those estimators that exploit the moment restrictions $E\left(\boldsymbol{Z}_i' \tilde{\boldsymbol{e}}_i\right) = \boldsymbol{0}$ (see Arellano, 2003, pp. 190–191).

Similarly, the estimator defined by (4) is asymptotically efficient among those GMM estimators that rely on the moment restrictions $E\left(\boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\right)=\boldsymbol{0}$. Moreover, the two sets of moment restrictions — $E\left(\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\right)=\boldsymbol{0}$ and $E\left(\boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\right)=\boldsymbol{0}$ — imply one another if Condition C1 is satisfied.[1]

An important special case of Theorem 1 is a first-differenced panel data model. In this case, $\boldsymbol{K}=\boldsymbol{D}$, where

$$\boldsymbol{D}=\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix}. \tag{5}$$

Premultiplying through (1) by $\boldsymbol{D}$ gives the first-difference model

$$\tilde{\boldsymbol{y}}_i=\tilde{\boldsymbol{X}}_i\boldsymbol{\beta}+\tilde{\boldsymbol{v}}_i, \qquad i=1,\ldots,n, \tag{6}$$

where the tth entry in $\tilde{\boldsymbol{y}}_i$ is the first difference $y_{i,t+1}-y_{it}$ $(t=1,\ldots,T-1)$, and the entries in $\tilde{\boldsymbol{X}}_i$ and $\tilde{\boldsymbol{v}}_i$ are similarly differenced.

Given $\boldsymbol{K}=\boldsymbol{D}$, the appropriate $\boldsymbol{F}$ is the forward orthogonal deviations transformation matrix

$$\begin{aligned}\boldsymbol{F} &= \operatorname{diag}\left(c_1,c_2,\ldots,c_{T-1}\right)\\ &\times\begin{pmatrix} 1 & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix},\end{aligned} \tag{7}$$

with $c_t^2=(T-t)/(T-t+1)$ $(t=1,\ldots,T-1)$ (see Arellano, 2003, p. 17). Premultiplying through (1) by $\boldsymbol{F}$ gives

$$\ddot{\boldsymbol{y}}_i=\ddot{\boldsymbol{X}}_i\boldsymbol{\beta}+\ddot{\boldsymbol{v}}_i, \qquad i=1,\ldots,n. \tag{8}$$

The tth entry of $\ddot{\boldsymbol{y}}_i$ is

$$\ddot{y}_{it}=c_t\left[y_{it}-\frac{1}{T-t}\left(y_{i,t+1}+\cdots y_{iT}\right)\right], \qquad t=1,\ldots,T-1.$$

The entries in $\ddot{\boldsymbol{X}}_i$ and $\ddot{\boldsymbol{v}}_i$ are similarly transformed. Thus, the forward orthogonal deviations transformation removes from each variable the within average of future values of the variable.

Arellano (2003, p. 153) indicates that two-step GMM applied to (8) is equivalent to

[1] To see this, note that there exists a nonsingular matrix $\boldsymbol{C}$ satisfying $\boldsymbol{C}\boldsymbol{Z}_i'=\boldsymbol{Z}_i'\boldsymbol{U}$ if, and only if, Condition C1 is satisfied (see Lemma A.1 in the appendix). Moreover, $E\left(\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\right)=\boldsymbol{0}$ implies $\boldsymbol{C}E\left(\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\right)=\boldsymbol{0}$. But $\boldsymbol{C}E\left(\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\right)=E\left(\boldsymbol{Z}_i'\boldsymbol{U}\boldsymbol{K}\boldsymbol{e}_i\right)=E\left(\boldsymbol{Z}_i'\boldsymbol{F}\boldsymbol{e}_i\right)=E\left(\boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\right)$. Hence, $E\left(\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\right)=\boldsymbol{0}$ implies $E\left(\boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\right)=\boldsymbol{0}$. Conversely, $E\left(\boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\right)=\boldsymbol{0}$ implies $\boldsymbol{C}^{-1}E\left(\boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\right)=\boldsymbol{0}$. But $E\left(\boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\right)=\boldsymbol{C}E\left(\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\right)$. Hence, $E\left(\boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\right)=\boldsymbol{0}$ implies $\boldsymbol{C}^{-1}\boldsymbol{C}E\left(\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\right)=E\left(\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\right)=\boldsymbol{0}$.

two-step GMM applied to (6) provided all available instrumental variables are used. If the instrumental variables are predetermined variables and all available instruments are used, this choice of instrumental variables satisfies the instruments condition in C1. This case is important, but using all available instrumental variables is not necessary for condition C1 to be satisfied. For example, only some of the predetermined variables can be used provided the instrumental variables used each period are also used in later periods. Or, suppose some of the regressors are strictly exogenous, and only strictly exogenous variables are used as instrumental variables. Using only strictly exogenous variables as instrumental variables will satisfy condition C1 provided the instrumental variables used in period s are also used in period t, for each $t \geq s$. In both of these examples the instrumental variables used each period is a subset — possibly a proper subset but not necessarily a proper subset — of the instrumental variables used in every later period. Phillips (2019) shows that cases like these satisfy C1.

Moreover, if, and only if, Condition C1 is satisfied, there is more than one way to calculate a system GMM estimator. To see this, let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$, and let $\boldsymbol{X}_i$ denote a $T \times K$ matrix with $(y_{i,t-1}, \boldsymbol{x}'_{it})$ in its tth row ($t = 1, \ldots, T$). Next set $\boldsymbol{y}_i^+ = (\boldsymbol{y}'_i, (\boldsymbol{I}^* \boldsymbol{y}_i)')'$ and $\boldsymbol{X}_i^+ = (\boldsymbol{X}'_i, (\boldsymbol{I}^* \boldsymbol{X}_i)')'$, where $\boldsymbol{I}^*$ is a $(T-1) \times T$ matrix obtained by deleting the first row of a T-dimensional identity matrix. The well-known system GMM estimator studied by Arellano and Bover (1995) and Blundell and Bond (1998) relies on differencing the observations in the first T rows in $\boldsymbol{y}_i^+$ and $\boldsymbol{X}_i^+$. Specifically, it uses the transformed data $\tilde{\boldsymbol{y}}_i^+ = \boldsymbol{K}^+ \boldsymbol{y}_i^+$ and $\tilde{\boldsymbol{X}}_i^+ = \boldsymbol{K}^+ \boldsymbol{X}_i^+$ ($i = 1, \ldots, n$), where

$$\boldsymbol{K}^+ = \begin{pmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix},$$

and $\boldsymbol{I}$ is a $T-1$ dimensional identity matrix. Moreover, set

$$\boldsymbol{Z}_i^+ = \begin{pmatrix} \boldsymbol{Z}_{1i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_{2i} \end{pmatrix}$$

($i = 1, \ldots, n$). The system GMM estimator is

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{K^+} &= \left[\sum_i \tilde{\boldsymbol{X}}_i^{+\prime} \boldsymbol{Z}_i^+ \left(\sum_i \boldsymbol{Z}_i^{+\prime} \tilde{\boldsymbol{e}}_i^+ \tilde{\boldsymbol{e}}_i^{+\prime} \boldsymbol{Z}_i^+ \right)^{-1} \sum_i \boldsymbol{Z}_i^{+\prime} \tilde{\boldsymbol{X}}_i^+ \right]^{-1} \\ &\quad \times \sum_i \tilde{\boldsymbol{X}}_i^{+\prime} \boldsymbol{Z}_i^+ \left(\sum_i \boldsymbol{Z}_i^{+\prime} \tilde{\boldsymbol{e}}_i^+ \tilde{\boldsymbol{e}}_i^{+\prime} \boldsymbol{Z}_i^+ \right)^{-1} \sum_i \boldsymbol{Z}_i^{+\prime} \tilde{\boldsymbol{y}}_i^+, \end{aligned}$$

where $\tilde{\boldsymbol{e}}_i^+ = \tilde{\boldsymbol{y}}_i^+ - \tilde{\boldsymbol{X}}_i^+ \widehat{\boldsymbol{\beta}}$ ($i = 1, \ldots, n$) and $\widehat{\boldsymbol{\beta}}$ is an initial estimator of $\boldsymbol{\beta}$.

The same system GMM estimator can be constructed using forward orthogonal deviations if, and only if, C1 is satisfied. To see this, let

$$\boldsymbol{F}^+ = \begin{pmatrix} \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix},$$

where $\boldsymbol{F}$ is given by Eq. (7). Also, let $\ddot{\boldsymbol{y}}_i^+ = \boldsymbol{F}^+\boldsymbol{y}_i^+$ and $\ddot{\boldsymbol{X}}_i^+ = \boldsymbol{F}^+\boldsymbol{X}_i^+$ $(i = 1, \ldots, n)$. Then define

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{F+} &= \left[\sum_i \ddot{\boldsymbol{X}}_i^{+\prime}\boldsymbol{Z}_i^+ \left(\sum_i \boldsymbol{Z}_i^{+\prime}\ddot{\boldsymbol{e}}_i^+\ddot{\boldsymbol{e}}_i^{+\prime}\boldsymbol{Z}_i^+\right)^{-1} \sum_i \boldsymbol{Z}_i^{+\prime}\ddot{\boldsymbol{X}}_i^+\right]^{-1} \\
&\quad \times \sum_i \ddot{\boldsymbol{X}}_i^{+\prime}\boldsymbol{Z}_i^+ \left(\sum_i \boldsymbol{Z}_i^{+\prime}\ddot{\boldsymbol{e}}_i^+\ddot{\boldsymbol{e}}_i^{+\prime}\boldsymbol{Z}_i^+\right)^{-1} \sum_i \boldsymbol{Z}_i^{+\prime}\ddot{\boldsymbol{y}}_i^+,
\end{aligned}$$

where $\ddot{\boldsymbol{e}}_i^+ = \ddot{\boldsymbol{y}}_i^+ - \ddot{\boldsymbol{X}}_i^+\widehat{\boldsymbol{\beta}}$ $(i = 1, \ldots, n)$.

We can now state Theorem 2.

Theorem 2. *Suppose $\boldsymbol{Z}_{1i}$ is block-diagonal, with $1 \times k_t$ instrument vector $\boldsymbol{z}_{it}'$ in its tth diagonal block $(t = 1, \ldots, T-1)$, and $\widehat{\boldsymbol{\beta}}_{K+}$ and $\widehat{\boldsymbol{\beta}}_{F+}$ use the same initial estimator $\widehat{\boldsymbol{\beta}}$. Then $\widehat{\boldsymbol{\beta}}_{F+} = \widehat{\boldsymbol{\beta}}_{K+}$ if, and only if, C1 is satisfied.*

The result in Theorem 2 holds more generally. In particular, $\boldsymbol{D}$ can be replaced in the definition of $\boldsymbol{K}^+$ with another transformation matrix $\boldsymbol{K}$, provided $\boldsymbol{K}\boldsymbol{K}'$ is positive definite and provided $\boldsymbol{F}$ in $\boldsymbol{F}^+$ is given by $\boldsymbol{F} = \boldsymbol{U}\boldsymbol{K}$, where $\boldsymbol{U}$ is the upper-triangular Cholesky factorization of $(\boldsymbol{K}\boldsymbol{K}')^{-1}$.

Moreover, the conclusion of the theorem does not depend on how $\boldsymbol{Z}_{2i}$ is specified.

3 Concluding remarks

The results provided in this note tell us when two different transformations lead to the same two-step GMM estimators and system estimators. The results also tell us when they do not. The note focused on the equivalence of the first-difference and forward orthogonal deviations transformations. Similar results for one-step GMM based on forward orthogonal deviations and differencing were provided in Phillips (2019); see, for example, Example 1 in Phillips (2019).

There is, however, a difference in the practical import of the results provided in Phillips (2019) and those provided here. Specifically, one-step GMM based on forward orthogonal deviations provides significant computational advantages compared to one-step GMM based on differencing (Phillips, 2020). But we have no reason to expect a similar computational advantage for two-step GMM based on forward orthogonal deviations. This is because the weighting matrix for GMM based on forward orthogonal deviations is block-diagonal when only one step is used, and the block-diagonal weighting matrix is why one-step GMM estimates based on forward orthogonal deviations can often be computed much faster than one-step GMM estimates based on first-differences (see Phillips, 2020). But the weighting matrix is not block-diagonal for two-step GMM based on forward orthogonal deviations.

Appendix: Proofs

The proof of Theorem 1 relies on the following corollary to Theorem 1 in Phillips (2019).

Lemma A.1. *Suppose $\boldsymbol{K}\boldsymbol{K}'$ is positive definite. Let $\boldsymbol{U}$ be the upper-triangular Cholesky factorization of $(\boldsymbol{K}\boldsymbol{K}')^{-1}$, and let $\boldsymbol{Z}_i$ be defined as in (2). Then there is a nonsingular matrix $\boldsymbol{C}$ satisfying $\boldsymbol{C}\boldsymbol{Z}_i' = \boldsymbol{Z}_i'\boldsymbol{U}$ if, and only if, C1 is satisfied.*

Proof: Let $\boldsymbol{\Phi} = \boldsymbol{K}\boldsymbol{\Omega}\boldsymbol{K}'$, with $\boldsymbol{\Omega}$ a positive definite matrix. For the upper-triangular Cholesky factorization of $\boldsymbol{\Phi}^{-1}$, say $\boldsymbol{U}^*$, there is a nonsingular matrix $\boldsymbol{C}$ satisfying $\boldsymbol{C}\boldsymbol{Z}_i' = \boldsymbol{Z}_i'\boldsymbol{U}^*$ if, and only if, C1 is satisfied (Phillips, 2019, Theorem 1). Set $\boldsymbol{\Omega} = \boldsymbol{I}$. Then $\boldsymbol{U}^* = \boldsymbol{U}$, and the conclusion of Lemma A.1 follows.

A.1. Theorem 1 proof

By Lemma A.1 there is a nonsingular matrix $\boldsymbol{C}$ such that $\boldsymbol{C}\boldsymbol{Z}_i' = \boldsymbol{Z}_i'\boldsymbol{U}$ if, and only if, C1 is met. This fact and $\boldsymbol{F} = \boldsymbol{U}\boldsymbol{K}$ imply

$$\begin{aligned}
\sum_i \ddot{\boldsymbol{X}}_i'\boldsymbol{Z}_i\left(\sum_i \boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\ddot{\boldsymbol{e}}_i'\boldsymbol{Z}_i\right)^{-1}\sum_i \boldsymbol{Z}_i'\ddot{\boldsymbol{X}}_i &= \sum_i \tilde{\boldsymbol{X}}_i'\boldsymbol{U}'\boldsymbol{Z}_i\left(\sum_i \boldsymbol{Z}_i'\boldsymbol{U}\tilde{\boldsymbol{e}}_i\tilde{\boldsymbol{e}}_i'\boldsymbol{U}'\boldsymbol{Z}_i\right)^{-1}\sum_i \boldsymbol{Z}_i'\boldsymbol{U}\tilde{\boldsymbol{X}}_i \\
&= \sum_i \tilde{\boldsymbol{X}}_i'\boldsymbol{Z}_i\boldsymbol{C}'\left(\sum_i \boldsymbol{C}\boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\tilde{\boldsymbol{e}}_i'\boldsymbol{Z}_i\boldsymbol{C}'\right)^{-1}\sum_i \boldsymbol{C}\boldsymbol{Z}_i'\tilde{\boldsymbol{X}}_i \\
&= \sum_i \tilde{\boldsymbol{X}}_i'\boldsymbol{Z}_i\left(\sum_i \boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\tilde{\boldsymbol{e}}_i'\boldsymbol{Z}_i\right)^{-1}\sum_i \boldsymbol{Z}_i'\tilde{\boldsymbol{X}}_i \qquad (9)
\end{aligned}$$

if, and only if, C1 is satisfied. By similar reasoning, we get

$$\sum_i \ddot{\boldsymbol{X}}_i'\boldsymbol{Z}_i\left(\sum_i \boldsymbol{Z}_i'\ddot{\boldsymbol{e}}_i\ddot{\boldsymbol{e}}_i'\boldsymbol{Z}_i\right)^{-1}\sum_i \boldsymbol{Z}_i'\ddot{\boldsymbol{y}}_i = \sum_i \tilde{\boldsymbol{X}}_i'\boldsymbol{Z}_i\left(\sum_i^n \boldsymbol{Z}_i'\tilde{\boldsymbol{e}}_i\tilde{\boldsymbol{e}}_i'\boldsymbol{Z}_i\right)^{-1}\sum_i^n \boldsymbol{Z}_i'\tilde{\boldsymbol{y}}_i, \qquad (10)$$

if, and only if, C1 is satisfied.

A.2. Theorem 2 proof

By Lemma A.1, C1 is necessary and sufficient for the existence of a nonsingular matrix $\boldsymbol{C}$ satisfying $\boldsymbol{C}\boldsymbol{Z}_{1i}' = \boldsymbol{Z}_{1i}'\boldsymbol{U}$. Let

$$\boldsymbol{C}^+ = \begin{pmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}$$

and

$$\boldsymbol{U}^+ = \begin{pmatrix} \boldsymbol{U} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}.$$

Then $\boldsymbol{C}^{+}$ is nonsingular and $\boldsymbol{C}^{+}\boldsymbol{Z}_i^{+\prime} = \boldsymbol{Z}_i^{+\prime}\boldsymbol{U}^{+}$ if, and only if, C1 is met. This fact and $\boldsymbol{F}^{+} = \boldsymbol{U}^{+}\boldsymbol{K}^{+}$ gives

$$\sum_i \ddot{\boldsymbol{X}}_i^{+\prime}\boldsymbol{Z}_i^{+}\left(\sum_i \boldsymbol{Z}_i^{+\prime}\ddot{\boldsymbol{e}}_i^{+}\ddot{\boldsymbol{e}}_i^{+\prime}\boldsymbol{Z}_i^{+}\right)^{-1}\sum_i \boldsymbol{Z}_i^{+\prime}\ddot{\boldsymbol{X}}_i^{+}$$
$$= \sum_i \tilde{\boldsymbol{X}}_i^{+\prime}\boldsymbol{Z}_i^{+}\left(\sum_i \boldsymbol{Z}_i^{+\prime}\tilde{\boldsymbol{e}}_i^{+}\tilde{\boldsymbol{e}}_i^{+\prime}\boldsymbol{Z}_i^{+}\right)^{-1}\sum_i \boldsymbol{Z}_i^{+\prime}\tilde{\boldsymbol{X}}_i^{+}$$

and

$$\sum_i \ddot{\boldsymbol{X}}_i^{+\prime}\boldsymbol{Z}_i^{+}\left(\sum_i \boldsymbol{Z}_i^{+\prime}\ddot{\boldsymbol{e}}_i^{+}\ddot{\boldsymbol{e}}_i^{+\prime}\boldsymbol{Z}_i^{+}\right)^{-1}\sum_i \boldsymbol{Z}_i^{+\prime}\ddot{\boldsymbol{y}}_i^{+}$$
$$= \sum_i \tilde{\boldsymbol{X}}_i^{+\prime}\boldsymbol{Z}_i^{+}\left(\sum_i \boldsymbol{Z}_i^{+\prime}\tilde{\boldsymbol{e}}_i^{+}\tilde{\boldsymbol{e}}_i^{+\prime}\boldsymbol{Z}_i^{+}\right)^{-1}\sum_i \boldsymbol{Z}_i^{+\prime}\tilde{\boldsymbol{y}}_i^{+}$$

by derivations similar to those establishing Eq.s (9) and (10).

References

Arellano, M. (2003) *Panel Data Econometrics*, Oxford University Press: Oxford.

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

Blundell, R. and S. Bond (1998) “Initial conditions and moment restrictions in dynamic panel data models” *Journal of Econometrics* **87**, 115–143.

Greene, W. H. (2012) *Econometric Analysis*, 7th ed., Prentice Hall: Upper Saddle River, NJ.

Hayakawa, K. (2009) “First difference or forward orthogonal deviation- which transformation should be used in dynamic panel data models?: A simulation study” *Economics Bulletin* **29**, 2008–2017.

Phillips, R. F. (2019) “A numerical equivalence result for generalized method of moments” *Economics Letters* **179**, 13–15.

Phillips, R. F. (2020) “Quantifying the advantages of forward orthogonal deviations for long time series” *Computational Economics* **55**, 653–672.

Schmidt, P., Ahn, S. C., and D. Wyhowski (1992) “Comment” *Journal of Business & Economic Statistics* **10**, 10–14.