The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–14

dcxtab: Doubly-Corrected and Misspecification-Robust (DCMR) Variance Estimator for Linear GMM in Stata

First author
First affiliation
City, State/Country
Email address

Second author
Second affiliation
City, State/Country
Email address

Third author
Third affiliation
City, State/Country
Email address

Fourth author
Fourth affiliation
City, State/Country
Email address

Abstract. This paper introduces *Stata* commands [R] **dcxtab** which implement the difference and system GMM in linear dynamic panel data with a new finite sample corrected and misspecification-robust variance estimator. The command implements doubly-corrected standard errors of one-step (2SLS), two-step, and iterated-GMM in addition to the well-known Windmeijer (2005, *Journal of Econometrics*) corrected standard errors. Furthermore, the corrected variance estimator is consistent even under the misspecified moment condition model, while the conventional heteroskedasticity-robust variance estimator and the Windmeijer correction are inconsistent. We describe how to use [R] **dcxtab** command in economic applications and Monte Carlo experiment.

Keywords: st0001, generalized method of moments, dynamic panel, Windmeijer corrected standard error, misspecification-robust standard error.

1 Introduction

[**To be added: Motivation/Literature**]

Stata code [R] **dcxtab** implements doubly-corrected misspecification-robust (DCMR) standard errors for the GMM estimators in dynamic panel data models. DCMR standard error provides finite-sample correction on top of the commonly used Windmeijer (2005) corrected standard error. Further, it also provides consistent variance estimator when the moment condition model is misspecified (e.g., invalid instruments, misspecified lag specifications, heterogeneous effects, etc) while conventional heteroskedasticity-robust standard error and Windmeijer (2005) standard error are inconsistent. Thus, DCMR standard errors provides a convenient way to obtain improved inference under correctly specified model and robustness against misspecification at the same time.

[R] **dcxtab** can be used for Arellano-Bond (Arellano and Bond 1991) estimator, "difference GMM", as well as Arellano-Bover/Blundell-Bond (Arellano and Bover 1995; Blundell and Bond 1998) estimator, "system GMM" in linear dynamic panel model

© *yyyy* StataCorp LLC

with "small-T, large-N" setup. Similar to popular STATA commands [R] **xtabond** and [R] **xtabond2**, [R] **dcxtab** can be used for the dynamic panel model with additional strictly exogenous, predetermined, or endogenous variables as well by stacking and modifying additional moment conditions into the instrument sets.

[R] **dcxtab** implements DCMR standard errors for one-step, two-step, and iterated-GMM estimators. To our knowledge, the DCMR standard errors have not been available at any STATA commands including popular commands [R] **xtabond**, [R] **xtdpd**, and [R] **xtabond2**. Recently, Kripfganz (2018) develops STATA module `xtdpdgmm` which implements iterated-GMM estimators as well as nonlinear moment conditions in dynamic panel data models, but not for doubly-corrected misspecification-robust (DCMR) standard errors.

2 Linear Dynamic Panel GMM

We begin by introducing a linear dynamic panel regression:

$$\begin{aligned} y_{it} &= \alpha_0 y_{it-1} + x_{it}'\beta_0 + \epsilon_{it}, \\ \epsilon_{it} &= \eta_i + v_{it}, \end{aligned} \tag{1}$$

for $i = 1, ..., N$, $t = 1, ..., T$, where $x_{it} = (x_{it}^1, ..., x_{it}^{k-1})' \in R^{k-1}$. The unknown true parameter vector is $\theta_0 = (\alpha_0, \beta_0')' \in R^k$. We assume that the vector of regressors $w_{it} = (y_{it-1}, x_{it}')'$ is correlated with the unobserved individual effect η_i and is predetermined with respect to u_{it}, i.e., $E(w_{it}u_{it+s}) = 0$ for $s = 0, ..., T-t$. Assuming T grows slower than N grows to infinity, it is well known that popular fixed-effects or within-group estimator suffer from the Nickel bias (Nickel, 1981). To fix this issue, Anderson and Hsiao (1981) consider the first-differenced equation

$$\Delta y_{it} = \Delta w_{it}'\theta_0 + \Delta v_{it}, \ t = 2, ..., T, \tag{2}$$

and propose a consistent IV estimator that employs the lagged w_{it} as the instrument. Building upon this, a seminar work by Arellano and Bond (1991) examines the problem of estimating θ in the context of GMM framework by finding $q = kT(T-1)/2$ sequential instruments:

$$\underset{(T-1)\times k(T-1)T/2}{Z_i} = diag(z_{i2}', ..., z_{iT}'), \tag{3}$$

with all possible lagged instruments $z_{it} = (y_{i0}, ..., y_{it-2}, x_{i1}', ..., x_{it-1}')', 2 \leq t \leq T$ and $\Delta v_i = (\Delta v_{i2}, \cdots, \Delta v_{iT})'$. The total number of observations is $n = N(T-1)$. The moment conditions are then given by

$$E\left(Z_i'\Delta v_i\right) = E[Z_i'(\Delta y_i - \Delta w_i'\theta_0)] = 0. \tag{4}$$

Panel-GMM methods proposed by Arellano and Bond (1991) and Blundell and Bond (1998) as well as finite-sample corrected inference by Windmeijer (2005) are routinely used by applied economists through a convenient official STATA command [R] **xtabond**, [R] **xtdpdsys**, and a user-written extension [R] **xtabond2** by Roodman (2009).

When the moment condition in (4) is overidentified, $q > k$ iff $T \geq 2$, the key underlying assumption to validate aforementioned theories and STATA commands is that the moment condition is correctly specified in the sense that

$$E[Z_i'(\Delta y_i - \Delta w_i'\theta)] = 0, \tag{5}$$

for a unique value $\theta_0 \in R^k$. Often, this may be too restrictive in reality as long as one takes a perspective the moment condition (4) is a summarization of the approximated true moment process. The followings are two empirically plausible scenarios studied in Hwang, Kang and Lee (2019) and Hansen and Lee (2019) which imply the moment conditions (4) are misspecified, i.e. $E[Z_i'(\Delta y_i - \Delta w_i'\theta)] \neq 0$ for all $\theta \in R^k$:

- At least one of the regression coefficients is heterogenous across individuals, and the underlying probability distributions of the heterogeneous parameters depend on individual fixed effects.
- The dynamic specifications deviate from data generating process, e.g. the AR(1) lag lengths in (1) are misspecified.

It turns out that the moment misspecification causes bias in conventional robust standard error estimation, and thus a correct GMM inference ends up with a significant adjustment in covariance matrix calculation, e.g. Hall and Inoue (2003), Lee (2014), Hansen and Lee (2019). The STATA commands `dcxtab` estimates doubly-corrected misspecification-robust (DCMR) standard error that correct for this standard error bias.

A recent work by Hwang, Kang, and Lee (2019) point out that the sample moment condition does not hold for any finite sample size n almost surely whenever the model is over-identified, i.e., $g_N(\theta) \neq 0$ where $g_N(\theta) = N^{-1}\sum_{i=1}^{N} Z_i'(\Delta y_i - \Delta w_i'\theta)$. They call this the over-identification bias, which is non-zero for any N in general for over-identified models. The over-identification bias can be an important source of bias in the GMM standard error even under correctly specified moment conditions. Hwang, Kang, and Lee (2019) show that DCMR standard error actually provides finite-sample correction on top of the popularly used Windmeijer (2005) corrected standard error.

Pending the full definition of the dcxtab syntax in Section 3, the next subsections illustrate the command `dcxtab` by copying an influential paper, Acemoglu, Johnson, Robinson, and Yared (2008, AJRY hereinafter). AJRY investigate a causal effect of income on democracy by considering the dynamic panel regression with $k = 2$ and additional time-effect term μ_t in (1). The pre-determined regressor x_{it} is replaced with its one-period lag x_{it-1}. y_{it} is a measure of democracy (fhpolrigaug), x_{it} is log income per capita (L.lrgdpch). The parameter of interest β_0 which is the effect of income on democracy. Their panel data set is based on a sample of 127 countries annually from 1960 to 2000 with observations at either 5-year or 10-year frequencies. Readers are referred to Hansen and Lee (2019) for comprehensive derivations and theories on the empirical applications with AJRY data.

2.1 One-step GMM estimation

We first show how dynamic panel model DCMR standard errors correct misspecification and over-identification biases in the one-step GMM estimation methods. Let $Z = (Z_1', ..., Z_N')'$ be the instrument matrix, and $\Delta Y = (\Delta y_1', ..., \Delta y_N')'$, $\Delta w = (\Delta w_1', ..., \Delta w_N')'$, $\Delta y_i = (\Delta y_{i2}, ..., \Delta y_{iT})'$, and $\Delta w_i = (\Delta w_{i2}, ..., \Delta w_{iT})'$ be the data. The command `dcxtab` uses the initial GMM weighting matrix $\widehat{W} = n^{-1}\sum_{i=1}^{N} Z_i' H Z_i$. By default, H is a matrix with 2's on the main diagonal, -1's on the first off-diagonals and zero elsewhere. Other choices of $q \times q$ H matrix are described in Section 3. The one-step GMM estimator is given by

$$\widehat{\theta}_1 = (\Delta w' Z \widehat{W}^{-1} Z' \Delta w)^{-1} \Delta w' Z \widehat{W}^{-1} Z' \Delta Y. \tag{6}$$

The doubly corrected variance estimator of $\widehat{\theta}_1$ is given by

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

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

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

$$\widetilde{V}(\widehat{\theta}_1) = n^2 \left(\Delta w' Z \widehat{W}^{-1} Z' \Delta w\right)^{-1} \Delta w' Z \widehat{W}^{-1} \widehat{\Omega}_1 \widehat{W}^{-1} Z' \Delta w \left(\Delta w' Z \widehat{W}^{-1} Z' \Delta w\right)^{-1}$$

where

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

The syntax of `dcxtab` is similar to that of `xtabond2` which estimates $\widehat{\theta}_1$ and its conventional standard errors by typing:

```
. use "data_Acemoglu_et_al.dta"
. drop yr1 yr2
. xtabond2 fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv(yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant
```

(*output omitted*)

The command `dcxtab` below estimates the same $\widehat{\theta}_1$, but provides its bias-corrected DCMR.

```
. dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv(yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant dc
Dynamic panel-data estimation:
Group variable: code_numeric                    Number of obs      =       838
Time variable : year_numeric                    Number of groups   =       127
Number of instruments = 55                      Obs per group: min =         0
```

```
Wald chi2(11) =    160.61                                  avg =      6.60
Prob > chi2   =     0.000                                  max =         9
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    L2.lrgdpch
    D.(yr3 yr4 yr5 yr6 yr7 yr8 yr9 yr10 yr11)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(1/10).L.fhpolrigaug
------------------------------------------------------------------------------
                            Doubly-Corrected Robust Standard Errors for One-step GMM
 fhpolrigaug |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 fhpolrigaug |
        L1.  |   .4888286   .0950592     5.14   0.000      .302516    .6751412
             |
     lrgdpch |
        L1.  |  -.1291144   .0882869    -1.46   0.144    -.3021536    .0439248
             |
         yr3 |    .055474   .0303024     1.83   0.067    -.0039176    .1148656
         yr4 |   .0439763   .0382014     1.15   0.250     -.030897    .1188496
         yr5 |  -.0397544   .0541803    -0.73   0.463    -.1459459     .066437
         yr6 |   .0198622   .0585803     0.34   0.735     -.094953    .1346775
         yr7 |   .1103302   .0654022     1.69   0.092    -.0178558    .2385161
         yr8 |   .1111113   .0742949     1.50   0.135     -.034504    .2567265
         yr9 |   .1458195     .07378     1.98   0.048     .0012133    .2904257
        yr10 |   .1897608   .0869634     2.18   0.029     .0193157    .3602059
        yr11 |   .1927664   .0879794     2.19   0.028        .02033    .3652028
------------------------------------------------------------------------------
Choice of Inital GMM estimator: Arellano Bond (1991)
```

The following sequence of command lines compare conventional standard errors and DCMR standard errors.

```
. qui dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv(yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant dc

. estimates store dcxtab_one_dc

. qui dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv(yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant conv

. estimates store dcxtab_one_conv

. estimates table dcxtab_one_conv dcxtab_one_dc, se b(%7.4f) se(%7.4f) ///
> modelwidth(18) title("dcxtab: One-step GMM estimations ") style(columns)

dcxtab: One-step GMM estimations
```

Variable	dcxtab_one_conv	dcxtab_one_dc
fhpolrigaug		
L1.	0.4888	0.4888
	0.0850	0.0951
lrgdpch		
L1.	-0.1291	-0.1291
	0.0759	0.0883
yr3	0.0555	0.0555
	0.0302	0.0303
yr4	0.0440	0.0440

	0.0342	0.0382
yr5	-0.0398	-0.0398
	0.0490	0.0542
yr6	0.0199	0.0199
	0.0544	0.0586
yr7	0.1103	0.1103
	0.0599	0.0654
yr8	0.1111	0.1111
	0.0662	0.0743
yr9	0.1458	0.1458
	0.0659	0.0738
yr10	0.1898	0.1898
	0.0783	0.0870
yr11	0.1928	0.1928
	0.0780	0.0880

legend: b/se

2.2 Two-step GMM estimation

Next, consider the two-step efficient GMM estimator

$$\widehat{\theta}_2 = (\Delta w' Z \widehat{\Omega}_1^{-1} Z' \Delta w)^{-1} \Delta w' Z \widehat{\Omega}_1^{-1} Z' \Delta Y, \tag{7}$$

which can be implemented by typing:

```
dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv( yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant two dc
```

(*output omitted*)

To save the space, we omit the output result.

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

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

where

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

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

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

$$\widehat{D}_n = \left(\Delta w' Z \widehat{\Omega}_1^{-1} Z' \Delta w \right)^{-1} \Delta w' Z \widehat{\Omega}_1^{-1}$$

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

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

where

$$\widetilde{V}(\widehat{\theta}_2) = n^2\left(\Delta w' Z\widehat{\Omega}_1^{-1}Z'\Delta w\right)^{-1}. \tag{8}$$

The following sequence of command lines implement the two-step GMM estimator, and caculate and compare corresponding conventional standard errors, Windmeijer's corrected standard errors, and DCMR standard errors.

```
. qui dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv( yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant two dc
. estimates store dcxtab_two_dc
. qui dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv( yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant two conv
. estimates store dcxtab_two_conv
. qui dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv( yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant two wind
. estimates store dcxtab_two_wind
. estimates table dcxtab_two_conv dcxtab_two_wind dcxtab_two_dc, se b(%7.4f) ///
> se(%7.4f) modelwidth(18) title("dcxtab: Two-step GMM estimations ") style(columns)
dcxtab: Two-step GMM estimations
```

Variable	dcxtab_two_conv	dcxtab_two_wind	dcxtab_two_dc
fhpolrigaug			
L1.	0.5279	0.5279	0.5279
	0.0472	0.1047	0.0968
lrgdpch			
L1.	-0.0121	-0.0121	-0.0121
	0.0476	0.0648	0.0487
yr3	0.0044	0.0044	0.0044
	0.0129	0.0218	0.0176
yr4	-0.0139	-0.0139	-0.0139
	0.0179	0.0284	0.0234
yr5	-0.1086	-0.1086	-0.1086
	0.0325	0.0469	0.0369
yr6	-0.0441	-0.0441	-0.0441
	0.0297	0.0441	0.0323
yr7	0.0079	0.0079	0.0079
	0.0350	0.0514	0.0392
yr8	0.0172	0.0172	0.0172
	0.0387	0.0570	0.0435
yr9	0.0352	0.0352	0.0352
	0.0415	0.0609	0.0452
yr10	0.0395	0.0395	0.0395
	0.0491	0.0729	0.0537
yr11	0.0517	0.0517	0.0517
	0.0515	0.0756	0.0572

legend: b/se

2.3 Iterated GMM estimation

`dcxtab` implements the iterated GMM estimator by typing:

```
. dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv( yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant iter dc
Dynamic panel-data estimation:
Group variable: code_numeric                    Number of obs      =       838
Time variable : year_numeric                    Number of groups   =       127
Number of instruments = 55                      Obs per group: min =         0
Wald chi2(11) =    158.95                                      avg =      6.60
Prob > chi2   =     0.000                                      max =         9
------------------------------------------------------------------------------

Instruments for first differences equation
  Standard
    L2.lrgdpch
    D.(yr3 yr4 yr5 yr6 yr7 yr8 yr9 yr10 yr11)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(1/10).L.fhpolrigaug
Number of Iterations: 27
```

fhpolrigaug	Doubly-Corrected Robust Standard Errors for Iterated GMM Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fhpolrigaug						
L1.	.7438099	.1280335	5.81	0.000	.4928689	.9947509
lrgdpch						
L1.	-.0092376	.0390891	-0.24	0.813	-.085851	.0673757
yr3	-.025657	.0112029	-2.29	0.022	-.0476143	-.0036996
yr4	-.0380719	.0174569	-2.18	0.029	-.0722867	-.0038571
yr5	-.1118897	.0315965	-3.54	0.000	-.1738177	-.0499616
yr6	-.0283989	.0294001	-0.97	0.334	-.0860219	.0292242
yr7	.0047411	.0351174	0.14	0.893	-.0640876	.0735699
yr8	.0135705	.0352385	0.39	0.700	-.0554957	.0826368
yr9	.0151596	.0379762	0.40	0.690	-.0592724	.0895915
yr10	-.0037889	.0440525	-0.09	0.931	-.0901303	.0825525
yr11	.02055	.0476351	0.43	0.666	-.0728131	.1139131

```
Choice of Inital GMM estimator: Arellano Bond (1991)
```

The following sequence of command lines implement the iterated GMM estimator, and compare corresponding conventional standard errors, Windmeijer's corrected standard errors, and DCMR standard errors.

```
. qui dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv( yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant iter dc
. estimates store dcxtab_iter_dc
. qui dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv( yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant iter conv
. estimates store dcxtab_iter_conv
. qui dcxtab fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, ///
> gmm(L.(fhpolrigaug)) iv( yr*) iv(L2.lrgdpch, passthru) noleveleq noconstant iter wind
. estimates store dcxtab_iter_wind
. estimates table dcxtab_iter_conv dcxtab_iter_wind dcxtab_iter_dc, se b(%7.4f) ///
```

```
> se(%7.4f)  modelwidth(18) title("dcxtab: Two-step GMM estimations ") style(columns)
dcxtab: Two-step GMM estimations
```

Variable	dcxtab_iter_conv	dcxtab_iter_wind	dcxtab_iter_dc
fhpolrigaug			
L1.	0.7438	0.7438	0.7438
	0.0434	0.1495	0.1280
lrgdpch			
L1.	-0.0092	-0.0092	-0.0092
	0.0395	0.0536	0.0391
yr3	-0.0257	-0.0257	-0.0257
	0.0086	0.0132	0.0112
yr4	-0.0381	-0.0381	-0.0381
	0.0158	0.0234	0.0175
yr5	-0.1119	-0.1119	-0.1119
	0.0288	0.0393	0.0316
yr6	-0.0284	-0.0284	-0.0284
	0.0247	0.0403	0.0294
yr7	0.0047	0.0047	0.0047
	0.0306	0.0465	0.0351
yr8	0.0136	0.0136	0.0136
	0.0322	0.0497	0.0352
yr9	0.0152	0.0152	0.0152
	0.0350	0.0529	0.0380
yr10	-0.0038	-0.0038	-0.0038
	0.0412	0.0621	0.0441
yr11	0.0205	0.0205	0.0205
	0.0430	0.0659	0.0476

legend: b/se

2.4 System GMM

In an attempt to imporve the efficiecy and alleviate the problem of weak instrument problem, Blundell and Bond (1998) develop an extension of AB approach in the previous subsection, providing extra moment conditions under $\text{cov}(w_{it}\eta_i) = 0$. In this case, observations on w_{it} up to and including $t = s$ are valid instruments for the levels equation at $t = s$. That is,

$$E[\Delta w_{it}\epsilon_{it}] = 0 \text{ for } t \leq T.$$

To combine the moment conditions for both the first-differenced equations and levels equations, we stack the equations in diferences and levels as below:

$$y_i^+ = \begin{pmatrix} \Delta y_i \\ y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \; w_i^+ = \begin{pmatrix} \Delta w_i \\ w_{i1} \\ \vdots \\ w_{iT} \end{pmatrix},$$

and

$$Z_i^+ = \begin{pmatrix} Z_i & 0 & 0 & \cdots & 0 \\ 0 & w_{i1} & 0 & \cdots & 0 \\ 0 & & w_{i2} & \cdots & \vdots \\ \vdots & & \cdots & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & w_{iT} \end{pmatrix}$$

and define $\Delta Y^+ = (\Delta y_1^{+\prime}, ..., \Delta y_N^{+\prime})'$, $\Delta w^+ = (\Delta w_1^{+\prime}, ..., \Delta w_N^{+\prime})'$. After we replace ΔY , Δw, and Z with ΔY^+, Δw^+, Z^+, repectively, at all of our formule in the previous subsections, we obtain the doubly corrected variance estimators in system GMM.

We illustrate how to use `dcxtab` command using the data set in Arellano and Bond (1991) for System GMM estimator. To save space, we compute doubly corrected robust standard error only for two-step estimator.

```
. use http://www.stata-press.com/data/r7/abdata.dta
. dcxtab n l(1/2).n l(0/1).(w ys) k yr*, gmm(l.n) iv(yr* l(0/1).(w ys) k) ///
> noconstant h(2) two dc
Dynamic panel-data estimation:
Group variable: id                              Number of obs      =       751
Time variable : year                            Number of groups   =       140
Number of instruments = 46                      Obs per group: min =         5
Wald chi2(16) =  65431.88                                      avg =      5.36
Prob > chi2   =     0.000                                      max =         7
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    D.(yr1976 yr1977 yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984 w L.w ys
    L.ys k)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(1/8).L.n
Instruments for levels equation
  Standard
    yr1976 yr1977 yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984 w L.w ys
    L.ys k
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    D.L.n
------------------------------------------------------------------------------
                    Doubly-Corrected Robust Standard Errors for Two-step GMM
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .9084088   .2303481     3.94   0.000     .4569348    1.359883
         L2. |  -.1003836   .0843651    -1.19   0.234    -.2657361    .0649689
             |
           w |
         --. |  -.4859112   .1493596    -3.25   0.001    -.7786507   -.1931717
         L1. |   .3803484   .1711857     2.22   0.026     .0448307    .7158662
             |
          ys |
         --. |   .4476686   .2296163     1.95   0.051    -.0023711    .8977084
         L1. |  -.4384282   .2315624    -1.89   0.058    -.8922822    .0154258
             |
           k |    .162688   .1218592     1.34   0.182    -.0761517    .4015277
      yr1976 |          0  (omitted)
      yr1977 |          0  (omitted)
```

```
      yr1978 |   .5679325   .6835861      0.83   0.406    -.7718717    1.907737
      yr1979 |    .578845   .6851103      0.84   0.398    -.7639464    1.921636
      yr1980 |   .5675788   .6722251      0.84   0.398    -.7499582    1.885116
      yr1981 |   .5252205   .6708533      0.78   0.434    -.7896277    1.840069
      yr1982 |   .5236351   .6742151      0.78   0.437    -.7978023    1.845072
      yr1983 |   .5510275   .6678618      0.83   0.409    -.7579576    1.860013
      yr1984 |   .5208755   .6788509      0.77   0.443    -.8096478    1.851399
------------------------------------------------------------------------------
Choice of Inital GMM estimator: Blundell and Bond (1998)
```

Next two are equivalent, providing Windmeijer-corrected two-step system-GMM standard errors.

```
. dcxtab n l(1/2).n l(0/1).(w ys) k yr*, gmm(l.n) iv(yr* l(0/1).(w ys) k) ///
> noconstant h(2) two wind
. xtabond2 n l(1/2).n l(0/1).(w ys) k yr*, gmm(l.n) iv(yr* l(0/1).(w ys) k) ///
> noconstant h(2) two robust
```

3 The dcxtab command

3.1 Syntax

The syntax of the `dcxtab` command is as follows.

`dcxtab` *depvar varlist* [*if*] [*in*] [*weight*] [`,` *gmmtype vcetype*
`ivstyle(`*varlist* [`,` `equation(diff|level|both)` `passthru` `mz`])
`gmmstyle(`*varlist* [`,` `laglimits (a b)` `collapse` `orthogonal`
`equation(diff|level|both)` `passthru` `passthru` `split`])
`level(#)` `h(#)` `noconstant` `noleveleq`]

3.2 Options

gmmtype specifies the type of GMM estimator. *gmmtype* may be one of the following, and the default option is `onestep`:

`onestep` for one-step GMM. This uses the initial weighting matrix that does not depend on model parameters. For difference GMM (with `noleveleq` option below), the default uses inverse of $Z'HZ$ where Z is the instrument matrix and H is a matrix with 2's on the main diagonal, -1's on the first off-diagonals and zero

elsewhere, i.e.

$$\underset{(T-1)\times(T-1)}{H} = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & & 0 & 0 \\ 0 & -1 & 2 & \ddots & 0 & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & \ddots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}. \tag{9}$$

To implement system GMM (Blundell and Bond, 1998), `dcxtab` chooses the default weight matrix is $Z'HZ$ where Z where

$$\underset{(2T-1)\times(2T-1)}{H} = \begin{pmatrix} \underset{(T-1)\times(T-1)}{H} & \underset{(T-1)\times T}{M'} \\ \underset{T\times(T-1)}{M} & \underset{T\times T}{I} \end{pmatrix}, \tag{10}$$

and $M'M = H$, which is same as `xtabond2` by Roodman (2009). See option h(#) for other choices of initial weighitng matrices.

twostep for two-step GMM which uses one-step GMM estimator as an efficient weighting matrix. The one-step GMM used to construct two-step GMM estimator is estimated via the default weight matrices that are described at option onestep. For other types of initial weighitng matrices, see option h(#).

iterated for iterated GMM which is obtained by iterating the two-step efficient GMM estimator until convergence. The initial choice of GMM estimator in the iteration loop is the one-step GMM estimator via the default weight matrices that are described at option onestep. For other types of initial weighitng matrices, see option h(#).

vcetype specifies the type of standard error reported in the estimated regression model. *vcetype* may be one of the following:

dc for doubly-corrected robust-standard errors. This is the default option.

conventional for conventional heteroskedasticity-robust standard errors.

windmeijer for Windmeijer (2005) robust-standard errors. This option cannot be used with onestep.

Following two options `ivstyle` and `gmmstyle` also have suboptions that allow further customization of the instrument matrix and share the same syntax structure with [R] **xtabond2**.

`ivstyle(`*varlist*`)` specifies a set of variables to serve as standard instruments. Normally, strictly exogenous regressors are included in `ivstyle` options.

`gmmstyle(`*varlist*`)` specifies a set of variables to be used as bases for "GMM-style" instrument sets described in Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991). By default xtabond2 uses, for each time period, all available lags of the specified variables in levels dated $t-1$ or earlier as instruments for the transformed equation; and uses the contemporaneous first differences as instruments in the levels equation. Normally, predetermined or endogenous regressors are included in `gmmstyle` options.

`noleveleq` specifies that level equation should be excluded from the estimation, yielding difference rather than system GMM.

`h(#)` specifies the form of H, the a priori estimate of the covariance matrix of the idiosyncratic errors at the initial weight matrix $Z'HZ$. For both difference and system GMM estimators, the option `h(1)` chooses H =Identity. For difference GMM, the option `h(2)` (or `h(2)`) chooses the matrix defined in (9). For system GMM, the option `h(2)` chooses the matrix defined in (10), whereas the option `h(3)` selects the following form of matrix H:

$$H = \begin{pmatrix} \underset{(T-1)\times(T-1)}{H} & \underset{(T-1)\times T}{0} \\ \underset{T\times(T-1)}{0} & \underset{T\times T}{I} \end{pmatrix}, \tag{11}$$

as originally suggested by Blundell and Bond (1998).

`noconstant` suppresses the constant term in the levels equation.

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [U] 20.7 Specifying the width of confidence intervals

3.3 Saved Results

The commands [R] **dcxtab** stores the following results in `e()`:

Scalars

`e(N)`	Number of complete observations in untransformed data (system GMM) or transformed data (difference GMM)
`e(N_g)`	Number of included individuals
`e(g_min)`	Lowest number of observations in an included individual
`e(g_max)`	Highest number of observations in an included individual
`e(g_avg)`	Average number of observations per included individual
`e(chi2)`	Wald chi-squared statistic
`e(chi2p)`	p-value of Wald statistic

Macros

`e(iterations)`	Number of iterations after iterated GMM estimation
`e(vcetype)`	`Conventional` for heteroskedasticity-robust Std. Err., `Windmeijer` for Windmeijer (2005) corrected Std. Err., `Doubly Corrected s.e.` for doubly corrected HKL(2019) Std. Err.
`e(ivinsts`i`)`	Variables listed in `ivstyle` group i
`e(gmminsts`i`)`	Variables listed in `gmmstyle` group i

`e(esttype)`	`system` or `difference` GMM	`e(transform)`	`first differences`
`e(depvar)`	Dependent variable	`e(xvars)`	List of regressors
`e(tvar)`	Time variable	`e(ivar)`	Individual (panel) variable
`e(properties)`	`b V`		

Matrices

`e(b)`	coefficient vector	`e(V)`	variance–covariance matrix of the estimators

Functions

`e(sample)`	marks estimation sample

4 Monte Carlo Simulation

[**Can be done briefly. Subset of HKL. Some comparisons using Stata graphs can be done similar to what Hwang, Kang and Lee (2019) and Hansen and Lee (2019) did?**]

5 References

Anderson, T. W. and Hsiao, C Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association 76*: 598-606.

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.

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

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

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.

Hansen, B. E. and Lee, S. 2019. Inference for iterated GMM under misspecification. *Working paper*.

Holtz-Eakin, D., Newey, W., Rosen, H. S. 1988 Estimating Vector Autoregressions with Panel Data. *Econometrica 56*: 1371-1395.

Hwang, J, Kang, B, and Lee, S. 2019. A Doubly Corrected Robust Variance Estimator for Linear GMM. *Working paper*.

Kripfganz, S. 2018. XTDPDGMM: Stata module to perform generalized method of moments estimation of linear dynamic panel data models. *Statistical Software Components S458395, Boston College Department of Economics.*

Nickell, S. 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica 49*: 1417-1426.

Roodman, D. 2009 How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal 9*: 86-136.

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