Implementing Groupwise-Heteroskedasticity-Robust Variance-Covariance Estimators in Fixed-Effects Panel Data Regression with Stata

Giovanni S. F. Bruno[1]

[1]giovanni.bruno@unibocconi.it - Bocconi University

XVIII Conferenza Italiana degli Utenti di Stata - Firenze, 9-10 maggio 2024

Introduction

- The **heteroskedasticity-robust variance-covariance estimator** (HR) (White, 1980), consistent for the OLS var-cov matrix in OLS regression under arbitrary heteroskedasticity, is implemented in Stata by the `regress` option `vce(robust)`.
- Stock and Watson (2008) (SW) show that HR

$$\hat{V}^{HR} = \frac{N}{N-n}\left(\sum_i^n\sum_t^T \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)^{-1}\left(\sum_i^n\sum_t^T \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\hat{u}_{it}^2\right)\left(\sum_i^n\sum_t^T \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)^{-1}$$

is inconsistent in fixed-effect panel-data regression,

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\beta + u_{it}, \qquad i = 1, \ldots, n, \; t = 1, \ldots, T,$$

under arbitrary heteroskedasticity and *fixed* T, $n \to \infty$ asymptotics.

==> HR is not available for panel data `[xt]` procedures in Stata. There, `vce(robust)` implements the cluster-robust estimator (Liang and Zeger 1986; Arellano, 1987), which is consistent for arbitrary heteroskedasticity and correlation within individual clusters (Hansen 2007).

Groupwise Heteroskedasticity

- **Groupwise heteroskedasticity** (GH): $E\left(u_{it}^2|X_i\right) = \mu_i^2$ (the conditional variance of the idiosyncratic error is time-invariant, but can vary across individuals.)
- I demonstrate (Bruno, 2024) that HR is (*fixed* T, $n \rightarrow \infty$) consistent in fixed-effect panel data regression under GH.
- I also prove that the statistic used by SW to bias-adjust HR turns out to be another (*fixed* T, $n \rightarrow \infty$) consistent estimator under GH. It uses the group-means of the squared fixed-effect residuals, $\overline{\hat{u}_i^2}$, and can be referred to as GH-robust (GHR)

$$\hat{V}^{GHR} = \frac{N}{N-n}\left(\sum_i^n \sum_t^T \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)^{-1}\left(\sum_i^n \overline{\hat{u}_i^2} \sum_t^T \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)\left(\sum_i^n \sum_t^T \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'\right)^{-1}$$

Monte Carlo

Monte Carlo experiments in Bruno (2024) demonstrate that both HR and, especially, GHR have better finite-sample properties, in terms of bias and MSE, than the SW bias-adjusted estimator (FE) and, especially, the cluster-robust estimator (CLUS) under GH.

- Next, the Monte Carlo results (100000 replications) for the relative bias of HR, FE, GHR and CLUS and for sample sizes n=100, 500, 1000 and T=5, 10, 20.

Table: Relative bias under GH

T	n	Relative BIAS			
		$\hat{\sum}^{HR}$	$\hat{\sum}^{FE}$	$\hat{\sum}^{GHR}$	$\hat{\sum}^{clus}$
5	100	-0.0444	-0.0517	-0.0227	-0.0934
10	100	-0.0177	-0.0191	-0.0068	-0.0750
20	100	-0.0092	-0.0095	-0.0036	-0.0731
5	500	-0.0098	-0.0114	-0.0047	-0.0217
10	500	-0.0040	-0.0044	-0.0015	-0.0153
20	500	-0.0017	-0.0018	-0.0004	-0.0151
5	1000	-0.0056	-0.0064	-0.0031	-0.0110
10	1000	-0.0025	-0.0027	-0.0016	-0.0086
20	1000	-0.0004	-0.0004	-0.0002	-0.0070

Stata implementation of HR

The command `areg` can be used to implement the fixed-effect estimator for β with var-cov estimator HR.

- `areg` implements the fixed-effect estimator if the name of the panel variable is specified in the option `absorb()`.
- With the option `vce(robust)`, `areg` implements HR (the reason why HR is legitimate here is that `areg` is thought for datasets with many groups, but not a number that grows with the sample size).
- Next, an example with a simulated data-set designed with GH, $\beta = 1$, n=1000 and T=5.

```
. areg y x, absorb(ivar) vce(robust)

Linear regression, absorbing indicators            Number of obs =     5000
                                                   F(  1,  3999) =  1318.94
                                                   Prob > F      =   0.0000
                                                   R-squared     =   0.6030
                                                   Adj R-squared =   0.5037
                                                   Root MSE      =   1.0211

------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.012517   .0278799    36.32   0.000     .9578568    1.067177
       _cons |  -.0196672   .0144444    -1.36   0.173    -.0479862    .0086519
-------------+----------------------------------------------------------------
        ivar |   absorbed                                    (1000 categories)

. mat V0=e(V)

. mat list V0

symmetric V0[2,2]
               x       _cons
    x  .00077729
_cons  .00001076   .00020864
```

The matrix V0 from `areg` is identical to the matrix V_HR computed manually

```
. mat list V_HR

symmetric V_HR[2,2]
                x      _cons
    x  .00077729
_cons  .00001076  .00020864
```

Stata implementation of GHR

This requires a few lines of coding based on standard Stata commands and the community-contributed command avar (Baum and Schaffer, 2013), which simplifies computation of the filling of "sandwich" var-cov estimators

```
. use sim_data

.

. qui xtreg y x, fe

. sca N=e(N)

. sca G=e(N_g)

. sca k=e(rank)

. predict res,e

. gen res2=res^2

. bysort ivar: egen gm_res2=mean(res2)

. gen gm_res05=gm_res2^.5
```

```
. preserve

. xtdata, fe clear

. mat accum XX=x
(obs=5000)

. mat XX=XX/N

. mat XXi=invsym(XX)

. qui avar gm_res05 (x), rob

. mat S_GHR=r(S)

. mat V_GHR=XXi*S_GHR*XXi/N

. mat V_GHR=V_GHR*(N/(N-G-k+1))

. mat list V_GHR

symmetric V_GHR[2,2]
               x      _cons
    x  .00014268
_cons  9.400e-07  .00012037

. restore
```

```
. qui areg y x, absorb(ivar) vce(robust)

. erepost V=V_GHR

. areg

Linear regression, absorbing indicators           Number of obs =      5000
                                                  F(  1,   3999) = 1318.94
                                                  Prob > F      =   0.0000
                                                  R-squared     =   0.6030
                                                  Adj R-squared =   0.5037
                                                  Root MSE      =   1.0211

------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.012517    .011945    84.77   0.000     .9890981    1.035936
       _cons |  -.0196672   .0109713    -1.79   0.073    -.0411771    .0018428
-------------+----------------------------------------------------------------
        ivar |   absorbed                                      (1000 categories)

.
```

erepost (Jann, 2007) replaces the HR matrix currently in memory with the GHR matrix. This is then used by the second areg to compute the GHR std. err.'s, which evidently allow greater precision than HR.

Conclusion

- In the fixed effect panel regression under GH, the var-cov estimators HR and, especially, GHR permit more precise inference than the SW bias-adjusted estimator and, especially, the cluster-robust estimator.
- HR and GHR can be easily implemented in Stata.
- I am currently working on a test for the null of GH against the alternative of arbitrary heteroskedasticity, based on the discrepancy between HR and the bias-adjusted estimator (or equivalently between HR and GHR) under *fixed* T, $n \to \infty$ asymptotics. Preliminary Monte Carlo results on size and power are encouraging.

- Arellano, M., 1987. Computing robust standard errors for within-groups estimators. Oxford Bulletin of Economics and Statistics 49 (4), 431–34.
- Baum, C. F., Schaffer, M. E., 2013. avar: Asymptotic covariance estimation for iid and non-iid data robust to heteroskedasticity, autocorrelation, 1- and 2-way clustering, common cross-panela utocorrelated disturbances, etc., http://ideas.repec.org/c/boc/bocode/XXX.html.
- Bruno, G. S. F., 2024. Groupwise-heteroskedasticity-robust variance-covariance estimators in fixed effects panel data regression.
- Hansen, C. B., 2007. Asymptotic properties of a robust variance matrix estimator for panel data when t is large. Journal of Econometrics 141, 597–620.
- Jann, B., 2007. erepost: Stata module to repost the estimation results., available from http://ideas.repec.org/c/boc/bocode/s456850.html.
- Liang, K.-Y., Zeger, S. L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

- Stock, J. H., Watson, M. W., 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. Econometrica 76, 155–74.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48 (4), 817–838.