ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume LXI 239 Number 7, 2013

http://dx.doi.org/10.11118/actaun201361072151

HETEROSKEDASTICITY, TEMPORAL AND SPATIAL CORRELATION MATTER

Ladislava Grochová, Luboš Střelec

Received: August 26, 2013

Abstract

GROCHOVÁ LADISLAVA, STŘELEC LUBOŠ: *Heteroskedasticity, temporal and spatial correlation matter.* Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2013, LXI, No. 7, pp. 2151–2155

As economic time series or cross sectional data are typically affected by serial correlation and/or heteroskedasticity of unknown form, panel data typically contains some form of heteroskedasticity, serial correlation and/or spatial correlation. Therefore, robust inference in the presence of heteroskedasticity and spatial dependence is an important problem in spatial data analysis. In this paper we study the standard errors based on the HAC of cross-section averages that follows Vogelsang's (2012) fixed-b asymptotic theory, i.e. we continue with Driscoll and Kraay approach (1998). The Monte Carlo simulations are used to investigate the finite sample properties of commonly used estimators both not accounting and accounting for heteroskedasticity and spatiotemporal dependence (OLS, GLS) in comparison to brand new estimator based on Vogelsang's (2012) fixed-b asymptotic theory in the presence of cross-sectional heteroskedasticity and serial and spatial correlation in panel data with fixed effects. Our Monte Carlo experiment shows that the OLS exhibits an important downward bias in all of the cases and almost always has the worst performance when compared to the other estimators. The GLS corrected for HACSC performs well if time dimension is greater than cross-sectional dimension. The best performance can be attributed to the Vogelsang's estimator with fixed-b version of Driscoll-Kraay standard errors.

heteroskedasticity, serial correlation, spatial correlation, Monte Carlo simulation, panel data, HAC estimator

As it is generally known, economic data arises from time series or cross sectional studies which typically exhibit serial correlation and/or heteroskedasticity of unknown form. As well as panel data described by econometrics models typically contains some form of heteroskedasticity, serial correlation and/or spatial correlation[1]. Therefore, robust inference in the presence of heteroskedasticity and spatial dependence is an important problem in spatial data analysis (Kim and Sun, 2011). For statistical inference in such models it is essential to use some covariance matrix estimators that can consistently estimate the covariance of the model parameters. The use of heteroskedasticity consistent (HC) covariance matrix estimators in cross sectional data and of heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators in time series contexts is common in applied econometrics. Note that the popularity of these robust covariance matrix estimators is due to their consistency under weak functional form assumptions (Hansen, 2007).

As Kim and Sun (2013) state, there are several robust covariance estimators with correlated panel data. Recently, a large number of works have focused on spatial HAC estimation of covariance matrices of parameter estimation, e.g. Conley (1996, 1999), Conley and Molinari (2007), Robinson (2007), Kelejian and Prucha (2007), Hansen (2007), Kim and

1 Among studies dealing with the issue of robust estimates of panel data burdened with heteroskedasticity and/or serial correlation can be listed e.g. Arellano (1987), Dama (2013), Grochová and Střelec (2013) etc.

Sun (2011, 2013), Vogelsang (2012), Moscone and Tosetti (2012) etc.

First approach dealing with HAC estimator is based on the clustered covariance estimator by extending the White standard error (White, 1980) to account for serial correlation (see Arellano, 1987). Consequently, Driscoll and Kraay (1998) suggest a different approach that uses a time series HAC estimator (e.g. Newey and West, 1987) applied to cross sectional averages of moment condition. Using the Monte Carlo simulations they provide the evidence that their estimator performs better than the ordinary least squares (OLS) and seemingly unrelated regression (SUR) estimators overcoming also the "large cross-sectional dimension (*N*) compared to time dimension (*T*)" problem. For inference with dependent data using cluster covariance estimators see e.g. Bester *et al.* (2011a).

Second approach is based on extension of the spatial HAC estimator applied to time series averages of moment conditions. For this purpose when a distance measure is available in spatial setting, the robust standard errors can be obtained using the approaches of Conley (1996, 1999), Kelejian and Prucha (2007), Bester *et al.* (2011b) or Kim and Sun (2011, 2013) etc. which are extensions of nonparametric kernel HAC robust standard errors to the spatial context (see Vogelsang, 2012). Hence, Conley (1996, 1999) proposes a spatial HAC estimator based on the assumption that each observation is a realization of a random process, which is stationary and mixing, at a point in a two-dimensional Euclidean space. Later, Conley and Molinari (2007) examine the performance of this estimator using Monte Carlo simulation.

Kim and Sun (2011) study spatial HAC estimation of covariance matrices of parameter estimators. They generalize the spatial HAC estimator proposed by Kelejian and Prucha (2007) to be applicable to general linear and nonlinear spatial models and establish its asymptotic properties. They provide the consistency conditions, the rate of convergence and the asymptotic truncated MSE of the estimator. They also determine the optimal bandwidth parameter which minimizes the asymptotic truncated MSE. Consequently, they extend the spatial HAC estimator applied to time series averages of moment conditions in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms (Kim and Sun, 2013).

Moscone and Tosetti (2012) propose a HAC estimator for the covariance matrix of the fixed effects estimator in a panel data model with unobserved fixed effects and errors that are both serially and spatially correlated. They suggest a HAC covariance matrix estimator in the context of a panel data model with unobserved fixed effects, where errors are allowed to be both spatially and serially correlated. Their Monte Carlo simulations show that mentioned approach is quite robust to various forms of serial and cross sectional dependence.

The next approach combines the previous two approaches presented e.g. by Born and Breitung (2010) who focus on serial correlation and both cross-sectional and time dependent heteroskedasticity in the error terms in fixed panel data models. They propose a test robust to the presence of spatiotemporal HAC disturbances.

Recently, Vogelsang (2012) develops an asymptotic theory for test statistics in linear panel models that are robust to heteroskedasticity, autocorrelation and/or spatial correlation. He analyzed two classes of standard errors, which are based on nonparametric HAC matrix estimators. The first class is based on averages of HAC estimators across individuals in the cross-section. The second class is based on the HAC of cross-section averages and was proposed by Driscoll and Kraay (1998).

In this contribution we continue with Driscoll and Kraay approach, i.e. we study the second noted class of standard errors based on the HAC of cross-section averages that follows Vogelsang's fixed-b asymptotic theory.

This paper is organized as follows. In the first part, we reviewed the existing literature on HAC and HACSC estimators. In the second part, we describe the methods, resources and simulation setup. In the third part, we present and discuss some interesting results. The last sections are Conclusions and Summary.

METHODS AND RESOURCES

Monte Carlo simulations are used to investigate the finite sample properties of classical widely used estimators (OLS, GLS) in comparison to a new estimator based on Vogelsang's (2012) fixed-b asymptotic theory in the presence of cross-sectional heteroskedasticity, autocorrelation and spatial correlation (HACSC) in panel data with fixed effects. Fixed-b version of the estimator with Driscoll-Kraay standard errors produces Driscoll and Kraay (1998) standard errors of coefficients estimated by fixed-effects regression. The error structure is assumed to be HACSC. We use this nonparametric technique as the size of the cross-sectional dimension in finite samples does not constitute a constraint on feasibility, even in the case that *N* is greater than *T* (Vogelsang, 2012).

First, we focus on consequences in case that covariance matrix estimates are not corrected for HACSC. Consequently, we use the GLS estimator, originally proposed by Parks (1967) as an early solution to heteroskedasticity and spatiotemporal dependence in the residuals. As this estimator is not designed for medium- and large-scale panels since this method is infeasible if *T* is smaller than *N* and it tends to underestimate standard errors of parameters, we employ the fixed-b Vogelsang's estimator, the advantage of which is that it computes test statistics that correspond to the chosen kernel and bandwidth (Vogelsang, 2012). Finally we compare finite sample properties of these estimators under three scenarios: the ratio of N / T is equal to, greater and less than 1.

We start with the data generating process for the Monte Carlo simulation that takes HACSC into account. Consider bivariate regression model specification:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}, \text{ for } i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \quad (1)$$

where
y_{it} .... dependent variable,
x_{it} .... explanatory variable,
x_{it} .... error term the next parameters.

Without loss of generality, we set $\beta = 1$ and $\alpha_i \sim N(0, 0.5)$.

For explanatory variable x_{it} we assume that

$$x_{it} = 0.5\alpha_i + v_{it}, \quad (2)$$

where

$$v_{it} \sim N(0, 0.75). \quad (3)$$

For error term ε_{it} we assume that

$$\varepsilon_{it} = \delta_i \sum_{j=i+1}^{N} s_{ij}\varepsilon_{jt} + w_{it}, \quad (4)$$

where $\delta_i \sim Unif(0.3, 0.6)$ and where s_{ij} are elements of a $N \times N$ spatial weights matrix S that is row-normalized. Heteroskedasticity is simulated multiplying the error term ε_{it} by a term that includes panel unit specific independent variable x_{it}, i.e.

$$w_{it} = \rho w_{it-1} + \sqrt{0.5 \quad 0.5 x_{it} u_{it}}\ , \quad (5)$$

where

$$u_{it} \sim N(0, 1 - \rho^2). \quad (6)$$

Following Kelejian and Prucha (2007) we assess all units to be located at locations (a, b), for $a, b = 1, \ldots, \sqrt{N}$. To introduce serial correlation the error term contains autoregressive parameter and is constructed as AR(1) process.

It is assumed that the explanatory variable is uncorrelated with the remainder term ε_{it}, i.e. $Corr(x_{it}, \varepsilon_{it}) = 0$.

The Monte Carlo simulations proceed as follows: $N \in \{5, 50, 100\}$, $T \in \{20, 50, 100\}$ and $\rho \in \{0.0, 0.2, 0.4, 0.6, 0{,}8\}$. To eliminate the time-series process, 200 pre-observations were simulated and dropped for each panel unit. For mentioned parameters the number of replications $M = 5000$ was generated, and parameters and standard errors estimated. Consequently, the average ratio of estimated to true standard errors was computed.

RESULTS

The results of our Monte Carlo simulations are summarized in Tab. I and discussed in this section.

Tab. I describes how three discussed estimators perform in presence of heteroskedasticity and spatiotemporal dependence of error term. The ratio of estimated to true standard error averages from 5000 replications is reported in the table. If the performance is unbiased then the ratio equals to 1. If the estimator is downward/upward biased the value is less/greater than 1.

DISCUSSION AND CONCLUSIONS

The Monte Carlo experiment shows that the OLS exhibits an important downward bias in all of the cases and always has the worst performance when compared to the other estimators. The downward bias worsens as serial correlation and/or cross-sectional dimension increase. The value of the ratio lies in the interval ranging from 0.50 to 0.94. It never reaches the optimal value even if there is no temporal dependence as in our simulation setup heteroskedasticity and spatial dependence are always present. The GLS corrected for HACSC performs better than the OLS and is unbiased or negligibly downward biased in $T > N$ cases as ρ augments. The estimated to true standard errors ratio lies in the interval from 0.78 to 1.05. A considerable biased can be observed in $N > T$ cases in which the ratio drops to values around 0.70. The only estimator that performs well under all scenarios demonstrated by the values of the ratio around 1 is the Vogelsang's estimator with fixed-b version of Driscoll-Kraay standard errors. In contrast to Vogelsang (2012), we have to claim that his estimator undervalues standars errors when N is considerably greater than T that stems from an asymptotic theory. This shortcut, however, rapidly improves as N/T ratio decreases.

SUMMARY

In this paper we studied the standard errors based on the HAC of cross-section averages that follows Vogelsang's (2012) fixed-b asymptotic theory, i.e. we continue with Driscoll and Kraay approach (1998). For this purpose, the Monte Carlo simulations were used to investigate the finite sample properties of classical widely used estimators (OLS, GLS) in comparison to brand new estimator based on Vogelsang's (2012) fixed-b asymptotic theory in the presence of cross-sectional heteroskedasticity and serial and spatial correlation in panel data with fixed effects. The error structure was assumed to be HACSC. We used this nonparametric technique as the size of the cross-sectional dimension in finite samples does not constitute a constraint on feasibility, even in the case that N is greater than T (Vogelsang, 2012). Firstly, we focused on consequences in case that covariance matrix estimates are

I: *Results of the Monte Carlo simulations of estimators' performance in presence of heteroskedasticity and spatiotemporal dependence of error term*

		ρ				
Parameters	**estimator**	**0**	**0.2**	**0.4**	**0.6**	**0.8**
$N = 5$, $T = 20$	OLS	0.929	0.920	0.803	0.750	0.586
	GLS	0.846	0.890	0.825	0.790	0.782
	SCC	0.998	0.972	0.916	0.874	0.859
$N = 5$, $T = 50$	OLS	0.910	0.902	0.822	0.765	0.754
	GLS	1.005	0.978	0.931	0.913	0.893
	SCC	1.061	1.032	1.005	1.001	0.994
$N = 5$, $T = 100$	OLS	0.908	0.892	0.813	0.758	0.697
	GLS	1.048	0.981	0.959	0.935	0.917
	SCC	1.018	1.010	1.008	1.003	1.000
$N = 50$, $T = 20$	OLS	0.937	0.917	0.811	0.763	0.710
	GLS	1.023	0.907	0.722	0.701	0.692
	SCC	0.982	0.944	0.905	0.884	0.803
$N = 50$, $T = 50$	OLS	0.906	0.860	0.816	0.767	0.690
	GLS	1.010	0.933	0.835	0.791	0.777
	SCC	1.033	1.031	1.027	1.007	1.000
$N = 50$, $T = 100$	OLS	0.902	0.835	0.818	0.766	0.639
	GLS	1.027	0.935	0.882	0.832	0.809
	SCC	1.023	1.017	1.013	1.008	1.002
$N = 100$, $T = 20$	OLS	0.924	0.913	0.814	0.769	0.670
	GLS	0.929	0.767	0.726	0.699	0.694
	SCC	1.038	0.905	0.831	0.811	0.797
$N = 100$, $T = 50$	OLS	0.903	0.856	0.820	0.767	0.695
	GLS	0.981	0.924	0.867	0.794	0.779
	SCC	1.004	0.966	0.912	0.875	0.836
$N = 100$, $T = 100$	OLS	0.893	0.835	0.818	0.766	0.496
	GLS	1.003	0.934	0.850	0.803	0.788
	SCC	1.004	1.003	1.002	1.002	1.001

Note: OLS means the standard ordinary least square estimator. GLS stands for the GLS estimator with correction for spatiotemporal dependence and heteroskedasticity in the forcing term. SCC stands for the Vogelsang's estimator with Driscoll-Kraay standard errors corresponding to fixed-b asymptotic theory.

not corrected for spatial HAC. Secondly, we used a number of spatial HAC consistent covariance matrix estimators comparing their finite sample properties under three scenarios: the ratio of N / T is equal to, greater and less than 1.

Our Monte Carlo experiment shows that the OLS exhibits an important downward bias in all of the cases and almost always has the worst performance when compared to the other estimators. The GLS corrected for HACSC performs well if $T > N$ which corresponds to the GLS assumptions. The best performance can be attributed to the Vogelsang's estimator with fixed-b version of Driscoll-Kraay standard errors that exhibits a correct value of the estimated to true standard errors in most of the cases.

Acknowledgement

Results published in the paper are a part of a research project "WWWforEurope" No. 290647 within Seventh Framework Programme supported financially by the European Commission.

REFERENCES

ARELLANO, M., 1987: Computing Robust Standard Errors for Within-groups Estimators. *Oxford Bulletin of Economics and Statistics*, 49, 4: 431–434. ISSN 0305-9049.

BESTER, A. C., CONLEY, T. G., and HANSEN, C. B., 2011a: Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, 165, 2: 137–151. ISSN 0304-4076.

BESTER, A. C., CONLEY, T. G., HANSEN, C. B., and VOGELSANG, T. J., 2011b: Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators. *Working Paper, Department of Economics, Michigan State University*. Available online: https://www.msu.edu/~tjv/spatialhac.pdf.

BORN, B., and BREITUNG, J., 2010: Testing for serial correlation in fixed-effects panel data models. *Beiträge zur Jahrestatung des Vereins für Socialpolitik 2010: Ökonomie der Familie – Session: Panel Data Models, No. C15-V2.* Available online: http://www.econstor.eu/bitstream/10419/37346/3/VfS_2010_pid_512.pdf.

CONLEY, T. G., 1996: Econometric modelling of cross sectional dependence. Ph.D. Thesis. University of Chicago, Department of Economics.

CONLEY, T. G., 1999: GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92, 1: 1–45. ISSN 0304-4076.

CONLEY, T., and MOLINARI, F., 2007: Spatial correlation robust inference with errors in location or distance. *Journal of Econometrics*, 140, 1: 76–79. ISSN 0304-4076.

DAMA, M., 2013: Developing a two way error component estimation model with disturbances following a special autoregressive (4) for quarterly data. *Economics Bulletin*, 33, 1: 625–634. ISSN 1545-2921.

DRISCOLL, J., and KRAAY, A. C., 1998: Consistent covariance matrix estimation with spatially dependent data. *Review of Economics and Statistics*, 80, 4: 549–560. ISSN 0034-6535.

GROCHOVÁ, L., and STŘELEC, L., 2013: Performance of the OLS estimator in presence of autocorrelation. In: *11th International Conference of numerical analysis and applied mathematics icnaam 2013: ICNAAM, edited by T. E. Simos et al., AIP Conference Proceedings*, 1558: 1851–1854. Melville, New York: American Institute of Physics. ISBN 978-0-7354-1184-5.

HANSEN, C. B., 2007: Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics*, 141, 2: 597–620. ISSN 0304-4076.

KELEJIAN, H. H., and PRUCHA, I. R., 2007: HAC estimation in a spatial framework. *Journal of Econometrics*, 140, 1: 131–154. ISSN 0304-4076.

KIM, M. S., and SUN, Y., 2013: Heteroskedasticity and spatiotemporal dependence robust inference for linear panel models with fixed effects. *Journal of Econometrics*, 177, 1: 85–108. http://dx.doi.org/10.1016/j.jeconom.2013.07.002. ISSN 0304-4076.

KIM, M. S., and SUN, Y., 2011: Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics*, 160, 2: 349–371. ISSN 0304-4076.

MOSCONE, F., and TOSETTI, E., 2012: HAC estimation in spatial panels. *Economics Letters*, 117, 1: 60–65. ISSN 0165-1765.

NEWEY, W. K., and WEST, K. D., 1987: A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 3: 703–708. ISSN 0012-9682.

PARKS, R., 1967: Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *Journal of the American Statistical Association*, 62, 318: 500–509. ISSN 0162-1459.

ROBINSON, P., 2007: Nonparametric spectrum estimation for spatial data. *Journal of Statistical Planning and Inference*, 137, 3: 1024–1034. ISSN 0378-3758.

VOGELSANG, T. J., 2012: Heteroskedasticity, Autocorrelation, and Spatial Correlation Robust Inference in Linear Panel Models with Fixed-Effects. *Journal of Econometrics*, 166, 2: 303–319. ISSN 0304-4076.

WHITE, H., 1980: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 4: 817–838. ISSN 0012-9682.

Address

Ing. Ladislava Grochová, Ph.D., Ing. Luboš Střelec, Ph.D., Department of Economics, Department of Statistics and Operation Analysis, Mendel University in Brno, 613 00 Brno, Czech Republic, e-mail: ladislava.grochova@mendelu.cz, lubos.strelec@mendelu.cz