HETEROSKEDASTICITY, TEMPORAL AND SPATIAL CORRELATION MATTER

Ladislava Grochová, Luboš Střelec

Received: August 26, 2013

Abstract

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. 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.
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:

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

where

- $y_{it}$ is the dependent variable,
- $x_{it}$ is the explanatory variable,
- $\epsilon_{it}$ is the error term.

Without loss of generality, we set $\beta = 1$ and $\alpha_i \sim N(0, 0.5)$. For explanatory variable $x_{it}$ we assume that

$$x_{it} = 0.5 \alpha_i + v_{it},$$

where

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

For error term $\epsilon_{it}$ we assume that

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

where $\delta_i \sim \text{Unif}(0.3, 0.6)$ 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 $\epsilon_{it}$ by a term that includes panel unit specific independent variable $x_{it}$, i.e.

$$w_{it} = \rho w_{it-1} + \sqrt{0.5 \cdot 0.5 x_{it} u_{it}},$$

where

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

Following Kelejian and Prucha (2007) we assess all units to be located at locations $(a_i, b_i)$, for $a_i, b_i = 1, \ldots, V / N$. 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 $v_{it}$, i.e. $\text{Corr}(x_{it}, v_{it}) = 0$.

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.

### 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

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

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


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.


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