Multiple Testing for Different Structures of Spatial Dynamic Panel Data Models

Francesco Giordano, Massimo Pacella, and Maria Lucia Parrella

Abstract In the econometric field, spatio-temporal data is often modeled by spatial dynamic panel data models (SDPD). In the last decade, several versions of the SDPD model have been proposed, based on different assumptions on the spatial parameters and different properties of the estimators. In particular, the classic version of the model assumes that the spatial parameters are homogeneous over location. Another version, proposed recently and called *generalized SDPD*, assumes that the spatial parameters are adaptive over location. In this work we propose a strategy for testing the particular structure of the spatial dynamic panel data model, by means of a multiple testing procedure that allows to choose between the generalized version of the model and some specific versions derived from the general one by imposing particular constraints on the parameters. The multiple test is made using the Bonferroni technique and the distribution of the multiple test statistic is derived by a residual bootstrap resampling procedure.

Keywords Spatio-temporal models · Model testing · Bootstrap

1 The SDPD Models

Consider a multivariate stationary process $\{\mathbf{y}_t\}$ of order p generating the observations at time t from p different locations. The following model

$$\mathbf{y}_t = D(\mathbf{l}_0)\mathbf{W}\mathbf{y}_t + D(\mathbf{l}_1)\mathbf{y}_{t-1} + D(\mathbf{l}_2)\mathbf{W}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (1)$$

has been proposed by [1] as a generalized version of the spatial dynamic panel data model of [2]. The errors $\boldsymbol{\varepsilon}_t$ are serially uncorrelated, they have zero mean value and may show cross-sectional correlation and heteroscedasticity, which means that $\boldsymbol{\varepsilon}_t$

F. Giordano · M. Pacella (✉) · M. L. Parrella
University of Salerno, Fisciano, Italy
e-mail: giordano@unisa.it; mpacella@unisa.it; mparrella@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_69

have a full variance/covariance matrix $\mathbf{\Sigma}_\varepsilon$; the *spatial matrix* $\mathbf{W}$ is a weight matrix with zero main diagonal; the matrices $D(\mathbf{l}_j)$ are diagonal, for $j = 0, 1, 2$, with main diagonal equal to vectors $\mathbf{l}_j = (\lambda_{j1}, \ldots, \lambda_{jp})$, respectively. Model (1) guarantees adaptivity by means of its $3p$ parameters λ_{ji}, $i = 1, \ldots, p$ and $j = 0, 1, 2$, and it is characterized by the sum of three terms: the *spatial component*, driven by matrix $\mathbf{W}$ and the vector parameter $\mathbf{l}_0$; the *dynamic component*, driven by $\mathbf{l}_1$; and the *spatial–dynamic component*, driven by $\mathbf{W}$ and $\mathbf{l}_2$.

Starting from the general model in (1), denoted as *generalized SDPD* model, we derive different models as special cases by considering some constraints on the parameters. The most used among these is the classic *SDPD* of [2], with only three parameters, where the spatial coefficients are constant among locations

$$\mathbf{y}_t = \lambda_0 \mathbf{W}\mathbf{y}_t + \lambda_1 \mathbf{y}_{t-1} + \lambda_2 \mathbf{W}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}'_t, \tag{2}$$

and the errors are homoscedastic and uncorrelated. We call this model *constant SDPD*. Other special cases of the model can be derived from the *generalized SDPD* by testing the significance of specific λ_{ji} coefficients.

2 A Strategy for Testing the Particular Structure of SDPD Models

In the sequel, we assume that $\mathbf{y}_1, \cdots, \mathbf{y}_T$ are T observations from a stationary process defined by (1) or (2). We assume that the process has mean zero and denote with $\mathbf{\Sigma}_j = Cov(\mathbf{y}_t, \mathbf{y}_{t-j}) = E(\mathbf{y}_t \mathbf{y}'_{t-j})$ the autocovariance matrix of the process at lag j, where the prime subscript denotes the transpose operator.

The estimators of the parameters for the *generalized SDPD* model (1) have been proposed and analyzed by Dou et al. [1]. Denote such estimators with $(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i})'$, where the index $i = 1, \ldots, p$ indicates the specific location. For the sake of brevity, we do not report the details of such estimators here.

In order to test the structure of the *SDPD* model, we define the test statistics

$$\hat{D}_{ji} = \sqrt{n}\left(\hat{\lambda}_{ji} - \bar{\lambda}_j\right), \quad i = 1, \ldots, p, \text{ and } j = 0, 1, 2. \tag{3}$$

In the (3), we are comparing the estimator under the generalized model, $\hat{\lambda}_{ji}$, with the estimator under the standard model with constant coefficients, which is evaluated by $\bar{\lambda}_j = \frac{1}{p}\sum_{k=1}^{p} \hat{\lambda}_{jk}$, the mean value of the estimates over different locations, for $j = 0, 1, 2$. Note that large values of the statistics in the (3) denote a preference for the *generalized SDPD* model. Instead, when the true model has constant parameters, as in the *SDPD* model of [2], the statistics in (3) are expected to be around zero. In order to give an empirical evidence of this, Fig. 1 shows the estimated density (based on $N = 250$ replications of the model) of the statistic $\hat{D}_{ji} = \sqrt{n}\left(\hat{\lambda}_{ji} - \bar{\lambda}_j\right)$, for

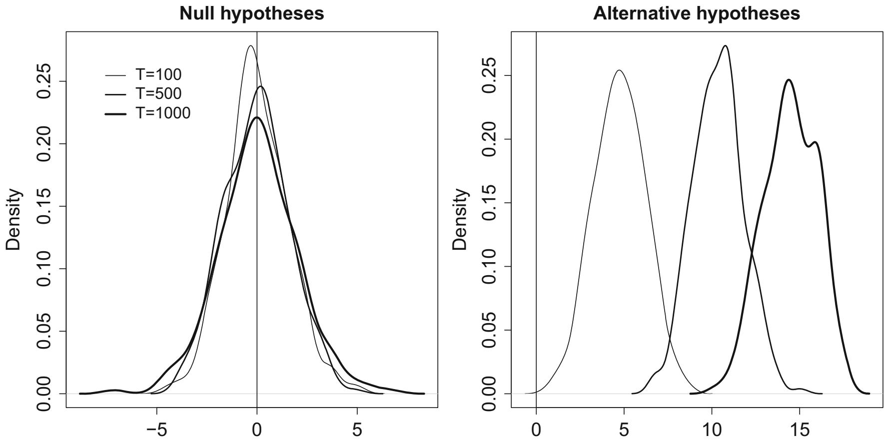

Fig. 1 Estimated densities (based on $N = 250$ replications of the model) of the statistic in (3), for $j = 2$, $i = 1$ and dimension $p = 50$, with different time series lengths (denoted by the line width, as indicated in the legend). The left side refers to the case generated under the Null hypothesis of true *constant SDPD* model. The right side refers to the case generated under the alternative hypothesis of true *generalized SDPD* model

$j = 2$, $i = 1$ and dimension $p = 50$, with different time series lengths (going from $T = 100$ to $T = 1000$ and denoted by the line width, as indicated in the legend). The left side of the figure refers to a case where the true model is the *constant SDPD* model, with constant parameters, therefore this is a case generated under the Null hypothesis. In such a case, as expected, the distribution of the statistic is centered around zero. The right side of the figure refers to a case where the true model is a *generalized SDPD*, with non-constant parameters, therefore this is a case generated under the alternative hypothesis. In the last case, as expected, the statistic $\hat{D}_{ji}$ is far away from zero. Moreover, as required for consistency, the value of the statistic increases for increasing time series length. Similar results can be shown for other values of i, j and p.

3 Bootstrap Scheme for the Multiple Testing Procedure

Figure 1 shows that the statistics in (3) can be used as building blocks of a testing procedure in order to identify the specific structure of the spatial dynamic model and to classify it between the two categories of *constant SDPD* and *generalized SDPD*. The hypotheses we need to test are

$$H_i : D_{ji} = 0, \quad \text{vs} \quad H_i' : D_{ji} \neq 0 \quad \text{for } i = 1, \ldots, p, \tag{4}$$

where j denotes the specific spatial parameter, $j = 0, 1, 2$. Test (4) has a multiple testing structure and the problem then becomes how to decide which hypotheses to reject, taking into account the multitude of tests. If many hypotheses are tested jointly, some are bound to appear as significant by chance alone, even if in reality they are not relevant. To prevent us from declaring true null hypotheses to be false, we seek control (at least asymptotically) of the familywise error rate (FWE), which is the probability of making at least one false rejection. The most familiar scheme for controlling the FWE is the well known Bonferroni method: for each null hypothesis H_i , individual p-values p_is are computed and the hypothesis H_i is rejected at global level α if $p_i \leq \alpha/m$.

In order to derive the individual p-values p_is, we use a resampling procedure based on the residual bootstrap approach, to approximate the distribution of the test statistics $\hat{D}_{ji}$. This procedure runs as follows.

1. First obtain the bootstrap errors $\{\boldsymbol{\varepsilon}_t^*\}$ by drawing $B = 999$ replicates independently from the residuals $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$, where $\hat{\mathbf{y}}_t = \bar{\lambda}_0 \mathbf{W}\mathbf{y}_t + \bar{\lambda}_1 \mathbf{y}_{t-1} + \bar{\lambda}_2 \mathbf{W}\mathbf{y}_{t-1}$.
2. Generate the bootstrap series, under the Null hypothesis, as

$$\hat{\mathbf{y}}_t^* = (\mathbf{I}_p - \bar{\lambda}_0 \mathbf{W})^{-1}(\bar{\lambda}_1 \mathbf{I}_p + \bar{\lambda}_2 \mathbf{W})\mathbf{y}_{t-1}^* + \epsilon_t^*.$$

3. Compute the bootstrap statistics $\hat{D}_{ji}^* = \sqrt{n}\left(\hat{\lambda}_{ji}^* - \bar{\lambda}_j^*\right)$, as in (3), with $\hat{\lambda}_{ji}^*$ and $\bar{\lambda}_j^*$ estimated from the bootstrap data $\hat{\mathbf{y}}_t^*$.
4. For a given $i = 1, \ldots, p$, the individual p-value p_i for testing H_i is defined as the probability $P(|D_{ji}^*| > |\hat{D}_{ji}| \Big| \mathbf{y}_1, ..., \mathbf{y}_T)$, which is approximated by the relative frequency of the event $(|D_{ji}^*| > |\hat{D}_{ji}|)$ over the 999 bootstrap replications.

The size of the test (with nominal size $\alpha = 0.1$) and the power have been evaluated empirically for different values of p and T and reported in the following table.

Under the Null	for $j = 0$			for $j = 1$			for $j = 2$		
(=size)	$T = 100$	500	1000	100	500	1000	100	500	1000
$p = 10$	0.124	0.1	0.072	0.184	0.144	0.148	0.136	0.112	0.108
$p = 50$	0.024	0.12	0.092	0.156	0.144	0.172	0.144	0.164	0.24
$p = 100$	0.888	0.2	0.204	0.82	0.216	0.244	0.884	0.128	0.136
Under the Alternative	for $j = 0$			for $j = 1$			for $j = 2$		
(=power)	$T = 100$	500	1000	100	500	1000	100	500	1000
$p = 10$	0.204	1	1	1	1	1	0.988	1	1
$p = 50$	0.056	0.108	0.148	1	1	1	1	1	1
$p = 100$	0.44	0.968	1	1	1	1	1	1	1

References

1. Dou, B., Parrella, M.L., Yao, Q.: Generalized Yule-Walker estimation for spatio-temporal models with unknown diagonal coefficients. J. Econ. **194**(2), 369–382 (2016)
2. Yu, J., de Jong, R., Lee, L.F.: Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. J. Econ. **146**, 118–134 (2008)