Mallow Model Averaging for Spatial Matrix Selection

Davide Fiaschi*- Lisa Gianmoena†- Angela Parenti‡

February 10, 2017

Long abstract for ERSA 2017

In spatial econometric literature, "the matrix is the fundamental tool used to model the spatial interdependence between regions. More precisely, each region is connected to a set of neighbouring regions by means of a spatial pattern introduced exogenously as a spatial weight matrix W" (Le Gallo et al., 2003, p.110).

The traditional specification of the spatial weights matrix relies on the geographical relation between observations, implying that areal units are neighbours when they share a common border (first-order contiguity), or the distance between their centroids is within a distance cut-off value (distance based contiguity). As pointed out by Anselin and Bera (1998), other specifications of the spatial weights matrix are possible as, for example, weights reflecting whether or not two individuals belong to the same social network, or based on some "economic" distance. Although these specifications are desired, the resulting spatial process must satisfy necessary regularity conditions. "For example, this requires constraints on the extent on the range of interaction and/or the degree of heterogeneity implied by the weights matrices" (Anselin and Bera, 1998, p. 244). Moreover, "in the standard estimation and testing approaches, the weights matrix is taken to be exogenous" (Anselin and Bera, 1998, p. 244). Therefore, the spatial matrix represents the a priori assumption about interaction strength between regions. However, in many cases considerable attention should be given to specifying the spatial matrix to represent as far as possible economic links (see Corrado and Fingleton, 2012).

This paper presents a methodology, called Mallow Model Averaging (MMA), which minimizes the in-sample mean square error (MSE) and of the out-of-sample one-step-ahead mean square forecast error (MSFE) for the **selection** of the spatial weights matrix (Hansen,

*University of Pisa, e-mail: davide.fiaschi@unipi.it
†Università Politecnica delle Marche, e-mail: lisa.gianmonena@univpm.it
‡IMT School for Advanced Studies Lucca, e-mail: angela.parenti@imtlucca.it

2008). As discussed in Hansen (2008) the *model averaging* based on Mallows' criterion has some advantages with respect to other two competing methods, i.e. the simple averaging and the Bayesian averaging. In particular, the simple averaging approach strongly depends on the set of models considered: "if a terrible model is included in the class of forecasting models, simple averaging will pay the penalty" (Hansen, 2008, p. 342). On the other hand, Bayesian model averaging suffers from the arbitrariness of the priors specification. Differently, MMA does not have these problems and optimal MMA weights are asymptotically optimal also with respect to the MSE (Hansen, 2008). In the same vein, Granger and Ramanathan (1984) consider three alternative methods (A, B and C) to obtaining optimal combinations of forecasts. According to Method A the optimal weights are the estimated coefficients of a regression of the observed values on all the competing forecast values with no constant term. In this case, the combined forecast will be unbiased only when the weights sum to one. Method B estimates the optimal weights using a restricted regression, where the sum of squared residuals is minimized under the restriction that the weights add up to one. Finally, Method C estimates the optimal weights from a regression of the observed values on all the competing forecast values and a constant term. The latter method is showed to be the best "because it gives the smallest mean squared error and has an unbiased combined forecast *even if individual forecasts are biased*" (Granger and Ramanathan, 1984, p. 201) .

Our methodology search for the optimal *combination* of different spatial matrices which minimizes a score inspired to *Mallow's* C_p. The optimal model is therefore given by a linear combination of the M candidate models, which only differ for the use of a different spatial matrix, with the vector of weights $\mathbf{w} = \left(w^1, \ldots, w^M\right)$, $\sum_{m=1}^{M} w^m = 1$ and $w^m \geq 0 \; \forall m$, calculated adopting the MMA methodology proposed in Hansen (2008). In short, the methodology operates as follows. Given a panel of dimensions $N \times T$, denote by $\mathbf{y}_T$ the vector of observed variable of interest in the last period T and by $\hat{\mathbf{y}}_T^m$ its predicted value calculated by applying to the initial level $\mathbf{y}_0$ the m-th estimated model; the prediction errors $\hat{\mathrm{e}}^m$ is given by:

$$\hat{\mathbf{e}}^m = \mathbf{y}_T - \hat{\mathbf{y}}_T^m; \tag{1}$$

and the MSFE:

$$\text{MSFE}^m = \left(\hat{\mathbf{e}}^m\right)^\top \hat{\mathbf{e}}^m / N. \tag{2}$$

As the second step, on the base of MSFE^m, we calculate the MMA criterion:

$$C_n(\mathbf{w}) = \mathbf{w}^\top \hat{\mathbf{e}}^\top \hat{\mathbf{e}} \mathbf{w} + 2\mathbf{w}^\top \mathbf{K} s^2 \tag{3}$$

where $\hat{\mathrm{e}} = \left(\hat{\mathrm{e}}^1, ..., \hat{\mathrm{e}}^M\right)$ are the prediction errors of the M models, $\mathbf{K} = (k^1, ..., k^M)$ is the vector of the number of parameters in each estimated model; and s^2 is the estimate of the unknown error variance of the true model. We estimate s^2 by the prediction errors of the model with the minimum MSFE. The optimal MMA weights vector is therefore the vector

$\hat{\mathbf{w}}$ that minimizes $C_n(\mathbf{w})$ under the assumption that $\mathbf{w}$ is a weights vector, i.e.:

$$\hat{\mathbf{w}} = \underset{\mathbf{w}\in[0,1]^M \text{ and } \sum_{m=1}^{M} w^m=1}{\operatorname{argmin}} C_n(\mathbf{w}). \tag{4}$$

We apply the proposed methodology to investigate the optimal combination of different spatial matrices based on several types of proximity in a growth model with spatial dependence inspired by Ertur and Koch (2006) for a sample of 224 EU regions in the period 1991-2014. In particular, we focus on four types of proximity inspired by Boschma (2005), Arbia et al. (2010), and Basile et al. (2012), i.e. i) *geographical proximity*, based on the Euclidean distance between regional centroids; ii) *technological proximity*, based on the distance between regional output composition in 1991 using the same index described in Basile et al. (2012); iii) *social proximity*, based on a synthetic distance between seven indicators of regional social capital in 1991-1993 as described in Basile et al. (2012); and, finally, iv) *institutional proximity*, based on the country membership of regions. We find that all types of proximity should be taken into account in the analysis, with a predominant role of geographical and technological proximities. We check the robustness of these findings with respect to Method C of Granger and Ramanathan (1984).