A SIMPLE SIMULATION STUDY TO ASSESS THE ABILITY OF THE MACML ESTIMATOR TO RECOVER PARAMETERS

Supplemental note to

"Incorporating Spatial Dynamics and Temporal Dependency in Land Use Change Models"

by Raghuprasad Sidharthan and Chandra Bhat

We undertake a simple simulation exercise to examine the ability of the MACML estimation approach to recover the parameters. All the notations in this note are based on those defined in the main paper.

A four-alternative choice situation ($I = 4$) with four time periods ($T = 4$) is considered for the simulation exercise (this scenario matches with the dimensions of the empirical study in this paper). A total of $Q = 200$ observation units are assumed (the observation units correspond to parcels in the case of the empirical application in the current paper). Three independent variables are used in the utility equation and each of them are generated from a standard univariate normal distribution (these are the elements of the x_{qti} vector). A random coefficient (across observation units) is assumed on the first variable, while fixed coefficients are assumed on the other two variables. Observation-specific random effects are also introduced, with the mean effect for the first alternative normalized to zero (equivalent to setting $\tilde{a}_1 = 0$ in Section 2.1). A diagonal specification is considered for the covariance matrix of $\tilde{\boldsymbol{\Lambda}}$, with the assumption that there is no random effect for the first alternative (note, however, that, as discussed in Section 2.1, a more general covariance specification is estimable for $\tilde{\boldsymbol{\Lambda}}$). In particular, using the normalization procedure in Section 2.1., $\tilde{\boldsymbol{\Lambda}}$ is specified as:

$$\tilde{\boldsymbol{\Lambda}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

The covariance matrix $\tilde{\boldsymbol{\Psi}}$ for the error term vector $\tilde{\boldsymbol{\eta}}_{qt}$ is also diagonal, but now with the fixed value of 0.5 along the diagonal for each alternative. Such a matrix is again a restrictive case of

1

the more general $\tilde{\boldsymbol{\Psi}}$ covariance matrix discussed in Section 2.1. Note that such a structure simplifies the simulation, since the elements of $\tilde{\boldsymbol{\Psi}}$ are not estimated. The reason for such restrictions on the $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\boldsymbol{\Psi}}$ matrices in the simulation design is to restrict the number of parameters to be estimated (given the moderate size sample used in the experiments) and to focus on the spatial and temporal dependency patterns.

To generate the spatial lag dependency, the 200 observation points are located on a rectangular grid of size 3,800 meters (2.375 miles) by 1800 meters (1.125 miles). Each observation point is 200 meters away from its closest neighbor. The spatial weight matrix W (of size 200×200) is created using the inverse of the square of distance on the coordinate plane between observational units. Finally, the first-order AR(1) temporal dependency parameters ρ is specified to be 0.6, and the spatial lag parameter δ is specified to be 0.5.

In total, the simulation design includes 13 parameters: three mean coefficients on the exogenous variables (corresponding to the **b** coefficient vector), one random coefficient element (corresponding to the Cholesky matrix in $\tilde{\boldsymbol{\Omega}} = \mathbf{LL}'$, which in our simulation design is the standard deviation of $\breve{\beta}_1$), three mean coefficients for the random effects (corresponding to the $\tilde{\mathbf{A}}$ vector), three parameters for the random effects (corresponding to the Cholesky decomposition of the $\tilde{\boldsymbol{\Lambda}}$ matrix, which in our simulation design are the standard deviations of $\breve{\alpha}_2, \breve{\alpha}_3, and \ \breve{\alpha}_3$), and the ρ and δ parameters. The simulation experiments entail assuming underlying "true" values for these parameters and generating data sets for estimation. Specifically, using the pre-specified parameters, the utility vector $\mathbf{U}$ of Equation (3) in the main paper is generated by drawing realizations of $\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\alpha}}$, and $\tilde{\boldsymbol{\eta}}$ from their underlying distributions. Then, for each observation unit and choice occasion, the alternative with the highest utility is designated as the chosen alternative. This variable constitutes the discrete dependent variable. The above procedure is repeated 50 times with different realizations of the $\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\alpha}}$, and $\tilde{\boldsymbol{\eta}}$ to generate 50 different data sets. For each dataset, the 13 model parameters are estimated using the MACML method, considering all pairs of observations in the CML function.

The results from the estimations are translated to measures of performance by comparing the estimated parameters with the "true" parameter values. To evaluate the ability of the MACML procedure to recover the parameters accurately, we compute an absolute percentage

bias (APB) measure for each parameter, which is the deviation of the mean estimate for the parameter.

Table 1 presents the results. The MACML inference approach does quite well in recovering parameters for all distance bands, with the mean APB value being around 6% (see last row of the APB column). More extensive theoretical and simulation studies are needed to better understand and characterize the ability of the MACML estimator to recover parameters under alternative spatial-temporal dependency scenarios, and also to investigate estimator efficiency considerations. But, the simple simulation exercise undertaken here suggests that the MACML method is able to recover the true parameters remarkably well for the spatial lag unordered response model with temporal autocorrelation.

Table 1. MACML Estimation Results of 50 Simulated Datasets with 200 Individuals and 4 Time Periods

Parameter		True	MACML estimate		
Notation in Paper	Components	True	Mean est.	Abs. Bias	Abs. %age bias (APB)
B	b1	0.500	0.521	0.021	4.2%
	b2	0.700	0.710	0.010	1.5%
	b3	-0.600	-0.627	0.027	4.5%
Chol($\widetilde{\mathbf{\Omega}}$)	$SD - \breve{\beta}_1$	1.000	1.033	0.033	3.3%
$\widetilde{\mathbf{A}}$	a2	0.300	0.334	0.034	11.3%
	a3	-0.400	-0.385	0.015	3.7%
	a4	0.500	0.569	0.069	13.8%
Chol($\widetilde{\mathbf{\Lambda}}$)	$SD - \breve{\alpha}_2$	1.000	1.047	0.047	4.7%
	$SD - \breve{\alpha}_3$	1.000	1.009	0.009	0.9%
	$SD - \breve{\alpha}_4$	1.000	1.119	0.119	11.9%
ρ	rho	0.600	0.576	0.024	4.0%
δ	del	0.500	0.465	0.035	6.9%
Mean value across parameters				**0.037**	**5.9%**

4