

# A Fast Estimator for Binary Choice Models with Spatial, Temporal, and Spatio-Temporal Interdependence

Julian Wucherpfennig<sup>1</sup>, Aya Kachi<sup>2</sup>,  
Nils-Christian Bormann<sup>3</sup> and Philipp Hunziker<sup>4\*</sup>

<sup>1</sup> Centre for International Security, Hertie School, Friedrichstraße 180, 10117 Berlin, Germany

<sup>2</sup> Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4052 Basel, Switzerland.

Email: [aya.kachi@unibas.ch](mailto:aya.kachi@unibas.ch)

<sup>3</sup> Department of Philosophy, Politics & Economics, Witten/Herdecke University, Alfred-Herrhausen-Str. 50, 58448 Witten, Germany

<sup>4</sup> Network Science Institute, Northeastern University, 177 Huntington Ave, Boston, MA 02115, USA

## Abstract

Binary outcome models are frequently used in the social sciences and economics. However, such models are difficult to estimate with interdependent data structures, including spatial, temporal, and spatio-temporal autocorrelation because jointly determined error terms in the reduced-form specification are generally analytically intractable. To deal with this problem, simulation-based approaches have been proposed. However, these approaches (i) are computationally intensive and impractical for sizable datasets commonly used in contemporary research, and (ii) rarely address temporal interdependence. As a way forward, we demonstrate how to reduce the computational burden significantly by (i) introducing analytically-tractable pseudo maximum likelihood estimators for latent binary choice models that exhibit interdependence across space *and* time and by (ii) proposing an implementation strategy that increases computational efficiency considerably. Monte Carlo experiments show that our estimators recover the parameter values as good as commonly used estimation alternatives and require only a fraction of the computational cost.

*Keywords:* spatial autocorrelation, temporal autocorrelation, simultaneity, discrete choice models, pseudo maximum likelihood

## 1 Introduction

Modeling binary outcomes—such as war, regime transitions, or policy adaption—poses considerable methodological challenges in the presence of spatial and/or temporal autocorrelation resulting from interdependent outcomes across and within units. The methodological difficulty stems from the likelihood function that involves an analytically intractable  $NT$ -dimensional integral.<sup>1</sup> Simulation-based estimation strategies including Gibbs sampling (LeSage 2000) and recursive importance sampling (RIS) (Beron and Vijverberg 2004) have been proposed to overcome this challenge. While these techniques promise to provide reliable estimates of spatial, and more recently spatio-temporal interdependence (Franzese, Hays, and Cook 2016), they are computationally burdensome (see Calabrese and Elkind 2014).<sup>2</sup> As social scientists implement research designs at increasing resolutions (e.g., at the grid-cell level) and with increasingly large datasets, simulation-based approaches quickly become infeasible.

- 1 Spatial probit is a special case where the marginal probability has a closed form but the likelihood function requires an evaluation of a multivariate normal distribution, which again cannot be computed exactly (Beron and Vijverberg 2004).
- 2 Franzese, Hays, and Cook (2016) and Calabrese and Elkind (2014) provide extensive reviews of the spatial probit literature, and useful comparisons among simulation-based estimation methods. Similarly, Beck *et al.* (2001) discuss a Bayesian estimation strategy for the binary temporal autoregressive model.

*Political Analysis* (2021)  
vol. 29: 570–576  
DOI: 10.1017/pan.2020.54

**Published**  
24 March 2021

**Corresponding author**  
Aya Kachi

**Edited by**  
Jeff Gill

© The Author(s), 2021. Published by Cambridge University Press on behalf of the Society for Political Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

\*Current affiliation: Google.

We provide a new estimator to address spatial, temporal, and spatio-temporal forms of interdependence embedded in binary outcome data. We build on a pseudo maximum likelihood estimator (PMLE) for binary spatially autoregressive models proposed by Smirnov (2010), and extend it to cases of temporal and spatio-temporal interdependence. So far, only Franzese, Hays, and Cook’s (2016) RIS estimator has addressed the spatio-temporal case. In addition, we reduce the estimation costs by proposing an implementation strategy that avoids direct matrix inversion (of large “interdependence multipliers”), and instead relies on a combination of iterative gradient procedures and approximations that yield an estimation algorithm with almost (only) linear complexity in  $N$ .

Monte Carlo experiments demonstrate that the PMLE recovers the parameter values as good as commonly used estimation alternatives—including Bayes, GMM, RIS, and naive probit—in a fraction of the time that simulation-based methods require. Yet, our analyses also accentuate important methodological issues that await solutions in improving the two existing estimators for spatio-temporal models (RIS and PMLE). First, both estimators generate seemingly biased standard errors. Second, we find that the performance of the RIS estimator is sensitive to the choice of data generating process (DGP). The conclusion section elaborates on these points.

## 2 Binary Choice Models with Spatio-Temporal Interdependence

This section specifies a binary choice model with spatio-temporal interdependence, for which we then develop a pseudo maximum likelihood estimators. We focus on the spatio-temporal case, which is applicable to cross-sectional-time-series data, noting that a purely spatial model for cross-sectional data and a purely temporal model for time-series data are nested herein. Full derivations are given in the [Online Appendix](#).

Our analytical point of departure is a discrete-choice spatio-temporal autoregressive (STAR) model as the conventional latent variable formulation:

$$y_{it}^* = \rho \sum_{j=1}^N w_{ij,t} y_{jt}^* + \gamma y_{i,t-1}^* + \mathbf{x}_{it} \boldsymbol{\beta} + u_{it} \tag{1}$$

or in matrix form:

$$\mathbf{y}_{(N \times T)}^* = \rho \mathbf{W} \mathbf{y}^* + \gamma \mathbf{T} \mathbf{y}^* + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}. \tag{2}$$

Here,  $\mathbf{y}^*$  is our latent outcome variable, for which we observe realizations  $\mathbf{y}$ , such that  $y_{it} = 1$  if  $y_{it}^* > 0$  and  $y_{it} = 0$  otherwise. The spatial connectivity matrix  $\mathbf{W}$  captures dependency between units across space,<sup>3</sup>  $\mathbf{T}$  is a temporal connectivity matrix that links the unit’s current outcomes to past realizations thereof,  $\mathbf{X} \boldsymbol{\beta}$  is a vector of covariates with corresponding parameters, and  $\mathbf{u}$  is an error term (zero-mean, iid, on the individual unit level). More specifically,  $\mathbf{W}_{NT \times NT}$  is block-diagonal with blocks of  $\mathbf{W}_{N \times N}^*$  for each time period  $t$ , while  $\mathbf{T}$  is an  $NT \times NT$  matrix full of zeros except for the identity matrices of size  $N$  on the lower first-minor (block) diagonal. The reduced form is:

$$\mathbf{y}^* = (\mathbf{I} - \rho \mathbf{W} - \gamma \mathbf{T})^{-1} \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W} - \gamma \mathbf{T})^{-1} \mathbf{u}. \tag{3}$$

Deriving the likelihood function requires the computation of  $P(y|\boldsymbol{\beta}, \rho, \gamma, \mathbf{X})$ , the joint probability for the observed random variable  $Y$  given the model parameters and regressors, which then requires the marginal CDF of the reduced-form error term,  $(\mathbf{I} - \rho \mathbf{W} - \gamma \mathbf{T})^{-1} \mathbf{u}$ . This computation is analytically intractable (as long as  $\rho \neq 0$ ) due to the *interdependence multiplier*,  $(\mathbf{I} - \rho \mathbf{W} - \gamma \mathbf{T})^{-1}$  (Anselin 2002).<sup>4</sup>

3 As is convention, we row-standardize  $\mathbf{W}$  to ensure stationarity for  $|\rho| < 1$  (Kelejian and Prucha 2010).

4 Spatial probit is a special case where the marginal probability has a closed form but the likelihood function requires an evaluation of a multivariate normal distribution, which again cannot be computed exactly (Beron and Vijverberg 2004).

### 3 A PMLE for Binary Spatio-Temporal Autoregressive (STAR) Models

To circumvent this problem, we turn to a pseudo maximum likelihood method. We build on Smirnov’s (2010) spatial PMLE and extend it to cases of temporal and spatio-temporal interdependence. The remainder of the section illustrates the gist of this derivation. Here, we maintain general mathematical expressions without assuming a specific marginal distribution (e.g., logistic vs. normal). In fact, the PMLE’s feasibility regardless of the error-term distribution is one of the strengths of this approach.<sup>5</sup>

Let  $\mathbf{Z}_{NT \times NT} = (\mathbf{I} - \rho\mathbf{W} - \gamma\mathbf{T})^{-1}$  and define  $\mathbf{D}$  as a corresponding diagonal matrix that contains only the diagonal elements of  $\mathbf{Z}$ , with all off-diagonal elements being zeros. This allows us to rewrite the model reduced form as:

$$\mathbf{y}_{(NT \times 1)}^* = \mathbf{Z}\mathbf{X}\boldsymbol{\beta} + \underbrace{(\mathbf{Z} - \mathbf{D})\mathbf{u}}_{\text{higher-order effects}} + \underbrace{\mathbf{D}\mathbf{u}}_{\text{zero-order effects}}. \tag{4}$$

Decomposing the spatial multiplier this way allows us to distinguish between zero-order and higher-order effects of an external shock as it affects observation  $it$ . More specifically, zero-order effects capture those shocks that the unit experiences directly. Higher-order effects are spillovers of external shocks which are transmitted either spatially through other units, or temporally across multiple time-periods. Our decomposition aggregates across both dimensions.

In order to allow for an analytical formulation of a (pseudo) likelihood, we assume that higher-order effects can be “ignored.” Behaviorally, this means that observations may simplify their choice by neglecting aggregate spatial effects of a random shock that are experienced by other (connected) observations. That is, mathematically, we do not expect a systematic effect of a random shock on unit  $i$  that is carried through the off-diagonal elements of the spatial multiplier; that is, it does not affect the choice probability systematically. The assumption is warranted because  $u_{it}$  are i.i.d with mean 0. This dramatically simplifies the stochastic element of the choice probability:

$$P(y_{it} = 1) = P(y_{it}^* \geq 0) = F_u \left( \frac{\sum_s \sum_j \beta z_{ijs} x_{js}}{d_{it}} \right), \tag{5}$$

which is now the cdf of the *univariate* distribution of  $u_{it}$  ( $F_u(\cdot)$ ), such as the standard normal (Probit) or a standard logistic (Logit). This allows us to write down a (pseudo) likelihood function, which is now in closed form. For any binomial link function  $g(\cdot)$ , we have

$$PL(\rho, \gamma, \boldsymbol{\beta} | \mathbf{X}, \mathbf{y}) \propto \prod_{i=1}^N \prod_{t=1}^T \left[ g^{-1} \left( \frac{[\mathbf{Z}\mathbf{X}\boldsymbol{\beta}]_{ij,s}}{d_{it}} \right)^{y_{it}} \left[ 1 - g^{-1} \left( \frac{[\mathbf{Z}\mathbf{X}\boldsymbol{\beta}]_{ij,s}}{d_{it}} \right) \right]^{(1-y_{it})} \right]. \tag{6}$$

Note that this expression requires an estimate for the values for  $\mathbf{y}_{i0}^*$ , that is the values *preceding* the first observed period in order to calculate the first period  $\mathbf{y}_{i1}^*$  (see Equation 1).<sup>6</sup> Assuming mean stationarity, we draw on Kauppi and Saikkonen (2008) and use what can be viewed as the unconditional expectation of  $\mathbf{y}^*$  across all time period (and units):  $E[\mathbf{y}^*] = (\mathbf{I} - \rho\mathbf{W} - \gamma)^{-1} \bar{\mathbf{X}}\boldsymbol{\beta}$ , where  $\bar{\mathbf{X}}$  are the sample means.

5 In our view, this strength goes beyond these distributions. This is useful when one might develop an estimator, for instance, for a hybrid of a binary spatial model and another model from a different model class such as duration and count.  
 6 Because  $\mathbf{y}^*$  is latent, dropping the first period from the likelihood merely shifts the problem to the next period, rather than solving it (cf. Franzese, Hays, and Cook 2016).

## 4 Speeding Up Computation Further

Naive implementations of the proposed PMLE may still be costly to run. We further reduce the estimation costs by proposing an implementation strategy that avoids direct matrix inversion, and instead relies on a combination of iterative gradient procedures and approximations that yield an estimation algorithm with almost linear complexity in  $N$ . (Appendix 3.1 clarifies why and how.)

## 5 Monte Carlo Simulations

### 5.1 Main Results: Estimator Comparisons for the Binary Spatio-Temporal Model

We ran MCs for spatial, temporal, and spatio-temporal PMLE, comparing estimates to those of alternative estimators. Here our discussion focuses on spatio-temporal estimators. Other results are presented in the Online Appendix. The DGP is a spatio-temporal probit as specified in (1), with  $u_i \sim N(0, 1)$ . We ran MCs for probit because most other existing tools are for probit models.<sup>7</sup> Throughout the experiments we set  $\beta_0$  to -0.5 and  $\beta_1$  to 1. The covariate vector  $\mathbf{x}$  is drawn from the standard normal and the spatial weights matrix  $\mathbf{W}$  captures queen-neighborhoods on a square lattice (row-standardized). We repeat the experiments for sample sizes of  $N \times T = \{64 \times 4; 64 \times 16; 256 \times 16\}$  with three combinations of spatial and temporal autocorrelation:  $\{\rho = .25 \ \& \ \gamma = .25; \rho = .25, \ \& \ \gamma = .5; \rho = .5 \ \& \ \gamma = .25\}$ .

Table 1 summarizes the results. Overall, the PMLE recovers the parameters accurately and with reasonable precision. However, the PMLE is overconfident in some cases, suggesting that the standard errors (obtained via the Hessian) are too small. Less than one percent of runs do not converge (see Tables A7 and A8 in the Online Appendix).

We contend that our estimator is nevertheless useful for applied researchers. Unlike most existing spatial estimators, it can simultaneously account for both spatial and temporal autocorrelation. So far, only the RIS approach proposed by Franzese, Hays, and Cook (2016) is able to estimate spatio-temporal processes for binary data. However, our PMLE is several orders of magnitude faster: on a standard PC a single run for  $N \times T = 64 \times 16$  takes seven seconds for the PMLE, and nearly 3 h for the RIS (see Figure A3 for a summary of estimation times); for  $N \times T = 256 \times 16$  the PMLE takes nine seconds, while we estimate the RIS to take almost two weeks if estimation time increases linearly (not executed).

More concernedly, our experiments show that the RIS's unbiasedness hinges on the choice of DGP, especially as true values of  $\gamma$  and  $\rho$  increase (c.f. Darmofal 2015, 166). For instance, the biases are not prominent under the DGPs chosen in Franzese, Hays, and Cook (2016) and Calabrese and Elkink (2014) but they are under the DGP we selected. (see Figure 1.)<sup>8</sup>

### 5.2 Other Results: Estimator Comparisons for the Binary Spatial or Temporal Model

For a purely spatial DGP, we compared the Bayesian spatial probit model proposed by LeSage (2000) and implemented by Wilhelm and Godinho de Matos (2013), a linearized spatial GMM (Klier and McMillen 2008), RIS (Franzese, Hays, and Cook 2016), a naive probit with an observed spatial lag, and our PMLE. Our experiments suggest that a Bayesian approach is preferable (see Tables A1, A5, Figure A1, and Calabrese and Elkink (2014)). However, given its speed, PMLE potentially provides useful starting values for the Bayesian approach. Finally, we also examined a setup with just temporal autocorrelation, comparing the RIS and PMLE approach. Again the PMLE outperforms the RIS (see Tables A2, A6 and Figure A2).

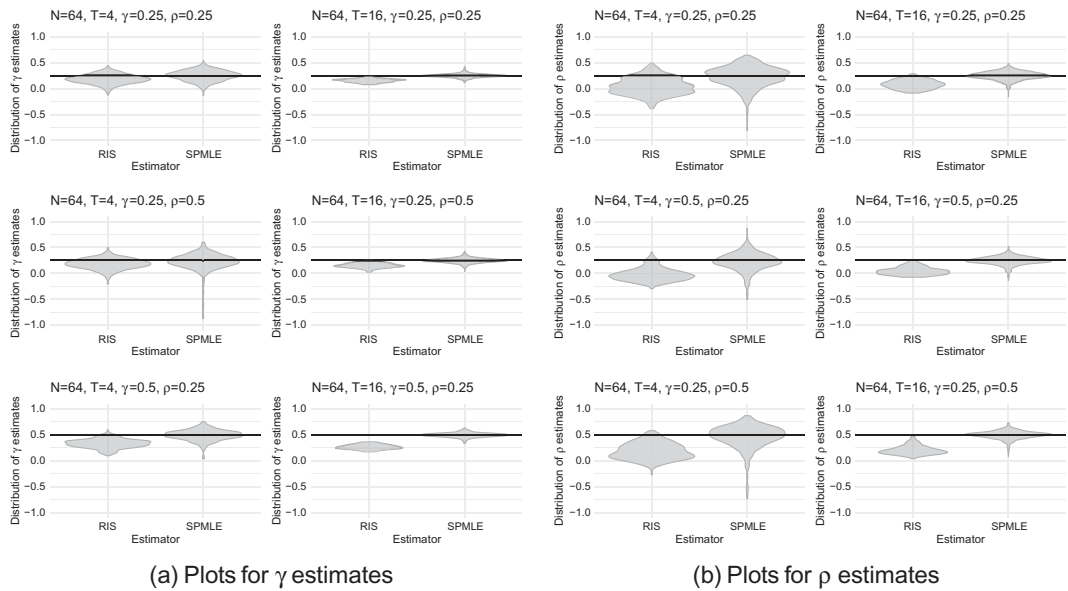
<sup>7</sup> As is in Equation (6), our PMLEs can host other link functions if one wishes.

<sup>8</sup> We were able to retrieve virtually identical RIS estimates to those presented in Franzese, Hays, and Cook (2016) when using their DGP and contiguity matrix (see Table A9). Moreover, this bias is also present in our purely spatial experiments, for which several other estimators are able to retrieve far better, if not unbiased, estimates. This provides confidence that our R implementation of Franzese, Hays, and Cook's (2016) Matlab code (which is a direct translation) is correct.

**Table 1.** Simulation results for **spatio-temporal** PMLE (500 iterations;  $y_0^*$  is estimated by  $E(y_t^*)$ )

	N=64 & T=4				N=64 & T=16				N=256 & T=16			
	$\beta_0 = -0.5$	$\beta_1 = 1$	$\gamma$	$\rho$	$\beta_0 = -0.5$	$\beta_1 = 1$	$\gamma$	$\rho$	$\beta_0 = -0.5$	$\beta_1 = 1$	$\gamma$	$\rho$
Experiment #1: $\gamma = 0.25, \rho = 0.25$												
Mean coefficient estimate	-0.523	1.001	0.249	0.229	-0.484	0.970	0.251	0.249	-0.483	0.964	0.249	0.249
Mean bias	0.171	0.122	0.080	0.154	0.083	0.057	0.036	0.080	0.044	0.039	0.019	0.042
RMSE	0.237	0.155	0.100	0.202	0.105	0.071	0.046	0.099	0.056	0.047	0.024	0.053
Actual SD of estimates	0.236	0.156	0.100	0.201	0.104	0.064	0.046	0.099	0.053	0.031	0.024	0.053
Overconfidence	1.302	1.104	1.086	1.203	1.192	0.971	1.011	1.166	1.161	0.952	1.027	1.167
Experiment #2: $\gamma = 0.25, \rho = 0.5$												
Mean coefficient estimate	-0.614	0.997	0.233	0.459	-0.469	0.930	0.250	0.497	-0.464	0.925	0.247	0.501
Mean bias	0.318	0.162	0.102	0.160	0.136	0.093	0.046	0.075	0.071	0.077	0.023	0.037
RMSE	0.580	0.206	0.141	0.231	0.176	0.113	0.059	0.099	0.087	0.086	0.029	0.047
Actual SD of estimates	0.570	0.206	0.140	0.228	0.174	0.089	0.059	0.099	0.079	0.042	0.029	0.047
Overconfidence	2.027	1.092	1.371	1.474	1.548	1.067	1.210	1.394	1.392	1.011	1.137	1.283
Experiment #3: $\gamma = 0.5, \rho = 0.25$												
Mean coefficient estimate	-0.530	0.937	0.493	0.224	-0.451	0.876	0.496	0.245	-0.438	0.860	0.499	0.246
Mean Bias	0.229	0.169	0.079	0.128	0.143	0.130	0.042	0.079	0.086	0.140	0.022	0.039
RMSE	0.331	0.210	0.104	0.176	0.179	0.149	0.053	0.101	0.103	0.146	0.027	0.049
Actual SD of estimates	0.330	0.200	0.103	0.174	0.173	0.083	0.053	0.101	0.082	0.041	0.027	0.049
Overconfidence	1.464	1.069	1.125	1.354	1.514	1.054	1.142	1.393	1.458	1.058	1.141	1.338

Overconfidence is the standard deviation of the estimated parameter divided by the mean of its estimated standard error.



**Figure 1.** Distribution of  $\gamma$  and  $\rho$  estimates from Monte Carlo simulations for recursive importance sampler and pseudo-maximum-likelihood estimator.

## 6 Conclusion

In this letter, we (i) introduced an analytically tractable pseudo maximum likelihood estimator for binary choice models that exhibit interdependence across space and/or time and (ii) proposed an implementation strategy that increases computational efficiency considerably. Our Monte Carlo experiments demonstrate that the estimators are able to recover the parameters of the DGP, and requires only a fraction of the computational cost of simulation-based methods. For spatio-temporal and temporal models, the PMLE estimator outperforms the only available alternative, the RIS implementation by (Franzese, Hays, and Cook 2016). By contrast, for purely spatial applications the Bayesian approach by Beron and Vijverberg (2004) appears to perform best.

However, our PMLE approach comes with one drawback: its standard errors are potentially biased (but apparently less so than those given by the RIS). In the broader context of composite maximum likelihood approaches, Varin, Reid, and Firth (2011) provide an extensive review of this property. In short, the standard errors obtained via the Hessian of the PMLE tend to be underestimated for certain parameters. Generally speaking, the literature finds that the bias is greater when  $NT$  is not sufficiently large compared to the variables included in the model. Our own first-cut Monte Carlo simulations (only with a single configuration of parameter values) indicate that this bias can emerge especially in the standard error estimate for the spatial parameter  $\rho$ . One approach would be a parametric bootstrap. Another approach would be an approximation, such as the use of a sandwich estimator. In this realm, for PMLEs in particular, a sandwich estimator using the Godambe information matrix appears promising (Varin, Reid, and Firth 2011).

## Acknowledgments

We would like to thank Frederick Boehmke, Lars-Erik Cederman, Scott Cook, Olga V. Chyzh, Martin Elff, Jos Elkink, Robert J. Franzese Jr., Christian Kleiber, Dennis Quinn, Andrea Ruggeri, Emily Schilling, Martin Steinwand, Oliver Westerwinter, Michael Ward, Christopher Zorn as well as audiences at various conferences for their useful comments.

## Data Availability Statement

The replication materials for this paper can be found at Harvard Dataverse at <https://doi.org/10.7910/DVN/9SKYLD> (Wucherpfennig *et al.* 2020).

## Supplementary Material

To view supplementary material for this article, please visit <https://dx.doi.org/10.1017/pan.2020.54>.

## References

- Anselin, L. 2002. "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models." *Agricultural Economics* 27(3):247–267.
- Beck, N., D. Epstein, S. Jackman, and S. O'Halloran. 2001. "Alternative Models of Dynamics in Binary Time-Series-Cross-Section Models: The Example of State Failure." Paper presented at the Annual Meeting of the Society for Political Methodology, Emory University.
- Beron, K. J., and W. P. M. Vijverberg. 2004. "Probit in a Spatial Context: A Monte Carlo Analysis." In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, edited by L. Anselin, R. J. G. M. Florax, and S. J. Rey, 169–195. Berlin: Springer.
- Calabrese, R., and J. A. Elkink. 2014. "Estimators of Binary Spatial Autoregressive Models: A Monte Carlo Study." *Journal of Regional Science* 54(4):664–687.
- Darmofal, D. 2015. *Spatial Analysis for the Social Sciences*. New York: Cambridge University Press.
- Franzese, R. J., J. C. Hays, and S. J. Cook. 2016. "Spatial-and Spatiotemporal-Autoregressive Probit Models of Interdependent Binary Outcomes." *Political Science Research and Methods* 4(1):151–173.
- Kauppi, H., and P. Saikkonen. 2008. "Predicting US Recessions with Dynamic Binary Response Models." *The Review of Economics and Statistics* 90(4):777–791.
- Kelejian, H. H., and I. R. Prucha. 2010. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances." *Journal of Econometrics* 157(1):53–67.
- Klier, T., and D. P. McMillen. 2008. "Clustering of Auto Supplier Plants in the United States: Generalized Method of Moments Spatial Logit for Large Samples." *Journal of Business & Economic Statistics* 26(4):460–471.
- LeSage, J. P. 2000. "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models." *Geographical Analysis* 32(1):19–35.
- Smirnov, O. A. 2010. "Modeling Spatial Discrete Choice." *Regional Science and Urban Economics* 40(5):292–298.
- Varin, C., N. Reid, and D. Firth. 2011. "An Overview of Composite Likelihood Methods." *Statistica Sinica* 21(1):5–42.
- Wilhelm, S., and M. Godinho de Matos. 2013. "Estimating Spatial Probit Models in R." *The R Journal* 5(1):130–143.
- Wucherpfennig, J., A. Kachi, N.-C. Bormann, and P. Hunziker. 2020. "Replication Data for: A Fast Estimator for Binary Choice Models with Spatial, Temporal, and Spatio-Temporal Interdependence". <https://doi.org/10.7910/DVN/9SKYLD>, Harvard Dataverse, V1.