2012

# Estimation in Autoregressive Population Models

Thomas Suesse
*University of Wollongong*, tsuesse@uow.edu.au

# Estimation in Autoregressive Population Models


## Abstract

Autoregressive (AR) models for spatial and social interaction have been proposed by many authors. A sample of units is obtained and the model is applied to this sample. Estimation methods such as the maximum likelihood method (ML) have been employed and investigated in the literature. The main assumption is that a response depends on other responses, when these units interact. Some of those units will be in the sample and some in the non-sample. Therefore the model should apply to the whole population rather to the sample only. Under such a population model, the marginal model for the responses of the sample is generally not of the same form and depends on covariates and interactions of non-sample units. Standard estimation methods using the sample information only are inappropriate. In this paper we investigate the performance of the standard ML method and a modified ML version that is based on the population model. Due to the population size, we also consider an approximate ML method. The results show that the standard ML method yields biased estimates and the modified ML version along with the approximate method perform far better.



## Keywords

Autoregressive models, spatial models, social networks, sample surveys

## Publication Details

Suesse, Thomas, Estimation in Autoregressive Population Models, Proceedings of the Fifth Annual ASEARC Conference - Looking to the future - Programme and Proceedings, 2 - 3 February 2012, University of Wollongong.

# Estimation in Autoregressive Population Models

Thomas Suesse

Centre for Statistical and Survey Methodology, University of Wollongong,
Wollongong, NSW, 2522, AUSTRALIA
*tusesse@uow.edu.au*

**Abstract**

Autoregressive (AR) models for spatial and social interaction have been proposed by many authors. A sample of units is obtained and the model is applied to this sample. Estimation methods such as the maximum likelihood method (ML) have been employed and investigated in the literature. The main assumption is that a response depends on other responses, when these units interact. Some of those units will be in the sample and some in the non-sample. Therefore the model should apply to the whole population rather to the sample only. Under such a population model, the marginal model for the responses of the sample is generally not of the same form and depends on covariates and interactions of non-sample units. Standard estimation methods using the sample information only are inappropriate. In this paper we investigate the performance of the standard ML method and a modified ML version that is based on the population model. Due to the population size, we also consider an approximate ML method. The results show that the standard ML method yields biased estimates and the modified ML version along with the approximate method perform far better.

*Keywords:* Autoregressive models, spatial models, social networks, sample surveys

## 1. Introduction

Spatial autoregressive models are popular for accounting for the spatial dependence between responses; see for example Ord (1975), Dorein (1981), Anselin (1988). The spatial distance of units *i* and *j* is accounted for by some corresponding non-negative weight, denoted by $W_{ij}$ , representing the possible degree of interaction of two units. By convention $W_{ii}$=0. The smaller the distance between two units, the larger is the spatial dependence, reflected by a high weight. The autoregressive (AR) model for the *i*th response $Y_i$, the *i*th covariate vector $\mathbf{X}_i$ and the *p*-dimensional vector of regression coefficients $\boldsymbol{\beta}$ has the form

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \rho\sum_j W_{ij}Y_j + \varepsilon_i, i = 1,...,n$$

with the common assumption that the errors are normally distributed, i.e. $\varepsilon_i \sim N(\mu = 0, V = \sigma^2)$. The parameter $\rho$ is a measure of the spatial dependence. If it is zero then the responses are independent. The model can also be written in a more compact matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{Y} + \boldsymbol{\varepsilon} \quad \text{with}$$
$$\boldsymbol{\varepsilon} \sim N(\boldsymbol{\mu} = 0, \Sigma = \mathbf{I}_n\sigma^2), \qquad (1)$$

where $\mathbf{I}_n$ is the n-by-n identity matrix. As an alternative, the AR model may be applied to the errors (ARerr), often also called disturbance model, and has the following form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \rho\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{v} \quad \text{with}$$
$$\mathbf{v} \sim N(\boldsymbol{\mu} = 0, \boldsymbol{\Sigma} = \mathbf{I}_n\sigma^2). \qquad (2)$$

Matrix **W** is often row or column normalised, for example row-normalised is often defined as all rows sum to 1. This also implies that the infinity norm is one, i.e. $\|W\| := \max_i(\sum_j |W_{ij}|) = 1$. This restricts the parameter space of $\rho$ to the open interval (-1,1), which will be assumed here.

Models such as (1) and (2) have been extended to a mixture of AR and ARerr models referring to one or more weight matrices and including possibly random effects. In this paper, we restrict our attention to (1) and (2) only. These spatial models are also applied in a social network context, where weights are defined on the social relationship of two people; see for example Leenders (2002) for various options of **W**.

Maximum likelihood estimation for (1) and (2) has been considered by Ord (1975). Recently the generalized method of moments has been investigated as an alternative to account for possibly non-normally distributed errors, see for example Kelejian and Prucha (1999, 2010) and Kapoor, Kelejian and Prucha (2007).

The AR models account for the spatial/social dependence between units *i* and *j*, for example it assumes that the response for unit *i* also depends on the response for unit *j*, if $W_{ij}>0$. The sample contains a set of units. In general, the sample will not contain all units *j* of the population which interact with unit *i*. Therefore the models should be rather applied to the whole population and not only to the sample. Current estimation methods ignore the dependence of sample units with non-sample units. In this paper, we investigate several estimation methods, when the AR models hold for the population. The standard ML method only uses the sample responses, the sample network and sample covariates. A modified ML method for the population model uses the population network, the population covariates and the sample responses.

The latter method is computationally complex due to the population size *N* and an alternative approximate ML method is considered as an alternative.

## 2. Population Autoregressive Models

Assume that models (1) and (2) hold for the population. Marginally model (1) implies

$$\mathbf{Y} \sim N(\boldsymbol{\mu} = \mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} = \sigma^2(\mathbf{A}'\mathbf{A})^{-1}) \quad \text{with}$$

$$\mathbf{A} = \mathbf{I}_n - \rho\mathbf{W} \text{ and} \quad \text{model} \quad (2)$$

$$\mathbf{Y} \sim N(\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} = \sigma^2(\mathbf{A}'\mathbf{A})^{-1}).$$

Let the set of sample units be denoted by *s* and the set of non-sample units by *r* and consider the partition of **Y**, **X** and **W** as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_s \\ \mathbf{Y}_r \end{pmatrix},\ \mathbf{X} = \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{pmatrix} \text{ and } \mathbf{W} = \begin{pmatrix} \mathbf{W}_{ss} & \mathbf{W}_{sr} \\ \mathbf{W}_{rs} & \mathbf{W}_{rr} \end{pmatrix}.$$

Suppose responses are only observed for the sample $\mathbf{Y}_s$ only and that $\mathbf{X}$ and $\mathbf{W}$ are known for the population. A similar situation might also occur by non-response, then $\mathbf{Y}_s$ represents the responses and $\mathbf{Y}_r$ the non-responses. It follows that $\mathbf{Y}_s \sim N(\boldsymbol{\mu}_s = \mathbf{B}_s\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ss} = \sigma^2[(\mathbf{A}'\mathbf{A})^{-1}]_{ss})$ with $\mathbf{B}_s = ((\mathbf{A}^{-1})_{ss}, (\mathbf{A}^{-1})_{sr})$ for the AR model and $\mathbf{Y}_s \sim N(\boldsymbol{\mu}_s = \mathbf{X}_s\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ss} = \sigma^2[(\mathbf{A}'\mathbf{A})^{-1}]_{ss})$ for the ARerr model. Standard estimation approaches (such as ML) are based on $\tilde{\boldsymbol{\Sigma}}_{ss} = \sigma^2(\mathbf{A}_{ss}'\mathbf{A}_{ss})^{-1}$ and $\tilde{\boldsymbol{\mu}}_s = \mathbf{A}_{ss}^{-1}\mathbf{X}_s\boldsymbol{\beta}$ (AR) or $\tilde{\boldsymbol{\mu}}_s = \mathbf{X}_s\boldsymbol{\beta}$ (ARerr). Only when $\mathbf{W}_{sr} = (\mathbf{W}_{rs})' = \mathbf{0}$ (sample does not interact with non-sample), then $\tilde{\boldsymbol{\Sigma}}_{ss} = \boldsymbol{\Sigma}_{ss}$ and $\tilde{\boldsymbol{\mu}}_s = \boldsymbol{\mu}_s$. Therefore in general, the standard estimation methods are based on a mis-specified model. Estimates for $\hat{\boldsymbol{\beta}}$ will be biased for the AR model, but unbiased for the ARerr model. Estimates for $\sigma^2$ and $\rho$ are likely to be biased leading to incorrect standard errors for $\boldsymbol{\beta}$ and possibly mis-interpretation of the spatial/social interaction.

Let us focus on a modified (correct) ML method leading to consistent estimates and proper standard errors due to correct model specification.

The negative log-likelihood $\tilde{L}$ for the partially observed data expressed in terms of $\mathbf{Y}_s$, $\boldsymbol{\mu}_s$ and $\mathbf{V}_{ss}^{-1} := [(\mathbf{A}'\mathbf{A})^{-1}]_{ss}$ is

$$\tilde{L} = \frac{n}{2}\log(2\pi) + \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log|\mathbf{V}_{ss}| + \frac{1}{2}(\mathbf{Y}_s - \boldsymbol{\mu}_s)'\mathbf{V}_{ss}(\mathbf{Y}_s - \boldsymbol{\mu}_s)$$

Standard optimization procedures require the calculation of the log-likelihood and its 1st and possibly 2nd order derivatives. In each iteration of these algorithms several matrix multiplications and inversions of matrices of order $N \times N$ have to be calculated, all of order $O(N^3)$ using schoolbook matrix multiplication. Improved algorithms as the Strassen algorithm and Coppersmith–Winograd algorithm reduce the complexity to $O(N^{2.807})$ and $O(N^{2.376})$, respectively but are practically only faster for extremely large *N*.

In comparison, standard methods based on the miss-specified model have complexity $O(n^3)$ for schoolbook matrix multiplication. Therefore this increased complexity makes estimation difficult. Alternatively we propose another approximate ML method that is discussed next.

### 3. Approximate Maximum Likelihood Method

If the inverse of a matrix $\mathbf{A}$ exists, then $\mathbf{A}^{-1} = \sum_{j=0}^{\infty} (\mathbf{I}_N - \mathbf{A})^j$. Therefore $\mathbf{A}^{-1} = \mathbf{I}_N + \sum_{k=1}^{\infty} \rho^j \mathbf{W}^j$. A $k$th order Taylor series approximations of $\mathbf{A}^{-1}$ is $\mathbf{A}^{-1} \approx \mathbf{I}_N + \sum_{j=1}^{k} \rho^j \mathbf{W}^j$. Since $|\rho| < 1$ and $\|\mathbf{WW}\| \le \|\mathbf{W}\|\|\mathbf{W}\|$ for the infinity norm, and by previous convention $\|\mathbf{W}\| = 1$, we conclude that the higher order terms $\|\rho^j \mathbf{W}^j\|$ approach zero, because $|\rho| < 1$ and $\|\rho^j \mathbf{W}^j\| \le |\rho| \|\rho^{j-1} \mathbf{W}^{j-1}\|$.

We maximize the same log-likelihood, only replacing $\mathbf{A}^{-1}$ by $\mathbf{I}_N + \sum_{j=1}^{k} \rho^j W^j$, a linear combination of powers of $\mathbf{W}$. However, we do not require the full matrix $\mathbf{A}^{-1}$ but only $(\mathbf{A}^{-1})_{ss}$ and $(\mathbf{A}^{-1})_{sr}$. Fortunately $(\mathbf{I}_N + \sum_{j=1}^{k} \rho^j W^j)_{ss} = \mathbf{I}_n + \sum_{j=1}^{k} \rho^j (W^j)_{ss}$, similarly with *sr* instead of *ss*.

Because $(\mathbf{A}'\mathbf{A})^{-1} = \mathbf{A}^{-1}(\mathbf{A}^{-1})'$, a $k$th order Taylor series also leads to a linear combination of powers of $\mathbf{W}$ for this term. If we consider a $k$th order approximation, we need powers of $\mathbf{W}$ of up to order $2k$. In total there are $((k+1)(k+2)-4)/2$ powers of $\mathbf{W}$ needed to approximate $(\mathbf{A}'\mathbf{A})^{-1}$ and $\mathbf{A}^{-1}$ by a $k$th order Taylor series. For $k$=2 four such powers are needed and for $k$=3 eight. The complexity to calculate these powers is $O(N^3)$, but this is only a one-off investment. When the approximate log-likelihood is maximized, the approximisations of $(\mathbf{A}^{-1})_{ss}$, $(\mathbf{A}^{-1})_{sr}$ and $[(\mathbf{A}'\mathbf{A})^{-1}]_{ss}$ are of order $n \times n$ or $N \times n$. In each step of the fitting algorithm it requires matrix multiplication and inversion of at most $n \times n$ matrices implying a complexity of $O(n^3)$. The term $\mathbf{B}_s\mathbf{X}$ consists of powers of $\mathbf{W}$ multiplied by $\mathbf{X}$, which can also be done prior to algorithm, leaving only matrices of order $n \times p$ for the fitting algorithm.

We expect that this approximate ML method is computationally much faster than the exact ML method when $N$ is large compared to $n$.

### Simulation Study

Now we aim to compare the exact ML method (exML) with the approximate ML method of order $k$ (K=$k$) and the standard ML method that only uses the sample (MLss). We consider a simple model with an intercept $\beta_0 = 10$ and a slope $\beta_1 = 20$ for the predictor $X \sim N(0,1)$. The variance parameter is chosen as $\sigma^2 = 1$ and the network/spatial parameter is $\rho = 0.3$. The fitting methods use the standard `R` routine `optim()` to minimise $\tilde{L}$. The method of choice "`L-BFGS-B`" is a box-constraint method applied to a quasi-Newton method (Byrd *et. al.* 1995), only requiring the value of $\tilde{L}$ and its first derivatives. In our experience this method is faster than other methods, as the classical Newton method, which also requires 2nd order derivatives, or non-derivative methods that only require $\tilde{L}$.

**Table 1: Results for estimates of $\beta_0$ and $\beta_1$**

| Estimation Method | Mean estimates | | 100*MSE | | Coverage in % | |
|---|---|---|---|---|---|---|
| | $\beta_0$ | $\beta_1$ | $\beta_0$ | $\beta_1$ | $\beta_0$ | $\beta_1$ |
| exML | -10.001 | 20.001 | 2.36 | 1.20 | 82.4 | 92.9 |
| K=1 | -10.888 | 20.350 | 80.45 | 13.50 | 0.00 | 11.3 |
| K=2 | -10.157 | 19.995 | 4.51 | 1.22 | 62.4 | 92.7 |
| K=3 | -10.070 | 20.010 | 2.66 | 1.21 | 78.1 | 92.9 |
| K=4 | -10.020 | 20.001 | 2.33 | 1.20 | 82.0 | 93.0 |
| K=5 | -10.008 | 20.002 | 2.33 | 1.20 | 82.2 | 93.0 |
| MLss | -13.800 | 20.441 | 1456 | 38.34 | 0.00 | 71.9 |

We consider various $n$=50,100,200,300,400,500 and $N$=200,300,400,500,600,700,800. Tables 1 and 2 show the mean estimates of the 4 parameter for the various methods along with its mean square error (MSE) and the coverage of a 95% Wald type confidence interval for the AR model with $n$=100 and $N$=800. Figure 1 shows the computation time of the methods for the AR model for increasing population size, similarly Figure 2 for increasing sample size but for the ARerr model.

### Conclusion

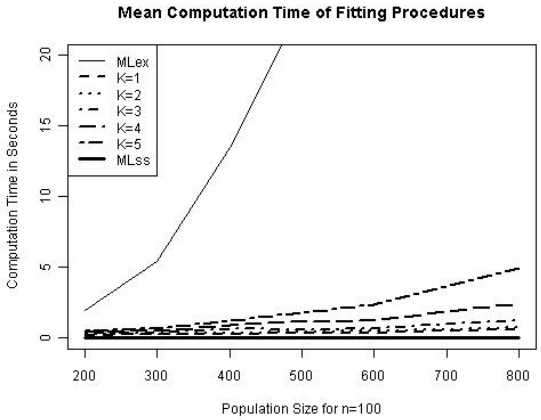
The results show that the standard ML method yields biased estimates of $\boldsymbol{\beta}$ (does not apply for ARerr model) and for $\rho$, $\sigma^2$ (for both models). Mostly the effect of the spatial or social dependence is underestimated. Practically this means that the

spatial/network effect reported in the literature is underestimated and that estimates $\boldsymbol{\beta}$ are biased (AR model only). This might lead to miss-interpretation of results. In small area estimation this might have the effect of biased estimates, because the estimation is based on a model for the sample, but then the estimates are applied to a population model.

**Table 2: Results for estimates of $\rho$ and $\sigma^2$**

| Estimation Method | Mean estimates | | 10,000*MSE for $\rho$ 100*MSE for $\sigma^2$ | | Coverage in % | |
|---|---|---|---|---|---|---|
| | $\rho$ | $\sigma^2$ | $\rho$ | $\sigma^2$ | $\rho$ | $\sigma^2$ |
| exML | 0.300 | 0.967 | 0.945 | 1.927 | 73.0 | 51.1 |
| K=1 | 0.311 | 1.141 | 2.451 | 4.583 | 63.8 | 37.9 |
| K=2 | 0.310 | 0.969 | 2.180 | 1.931 | 56.3 | 51.9 |
| K=3 | 0.301 | 0.969 | 0.977 | 1.924 | 74.5 | 51.7 |
| K=4 | 0.300 | 0.967 | 0.958 | 1.927 | 73.3 | 51.3 |
| K=5 | 0.300 | 0.967 | 0.946 | 1.927 | 73.2 | 51.2 |
| MLss | 0.056 | 9.971 | 600.6 | 8137 | 100 | 0.00 |

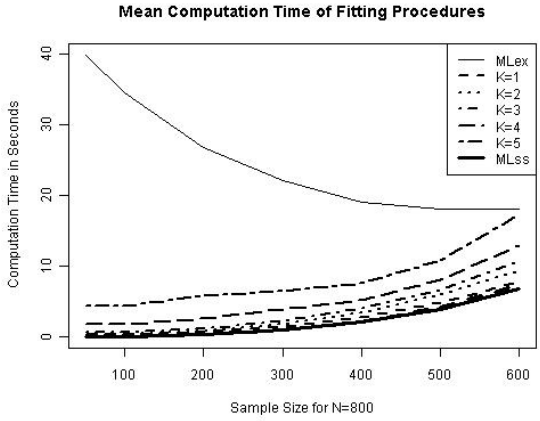
**Figure 1: Computation Time in Seconds for increasing *N* for fixed *n*=100 and the AR model**




The approximate ML method is much faster than the exact one and for $k$=4 it yields results that are practically identically to those of the exact method. A problem remains, for very large populations the reduction of complexity is not sufficient to make proper estimation feasible.

In this paper, we did not focus on conditional AR (CAR) models, a wider class of models that contain the considered models in this paper, also known as simultaneous AR (SAR) models, as a sub-class. Similar results as for SAR models are expected for CAR models that are not SAR models.

**Figure 2: Computation Time in Seconds for increasing *n* for fixed *N*=800 and the ARerr model**




## References


Anselin, L. (1988) Spatial *Econometrics: Methods and Models*, Kluwer Academic Publishers, Boston.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing **16**, 1190–1208.

Ord, J.K. (1975) Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* **70**, 120–126.

Doreian, P. (1980) Linear models with spatially distributed data. *Sociological Methods and Research* **9**, 29–60.

Doreian, P., 1981. Estimating linear models with spatially distributed data. In: Leinhardt, S. (Ed.), *Sociological Methodology*. Jossey-Bass, San Francisco, 359–388.

Kelejian, H.H., Prucha, I.R. (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* **40**, 509-533.

Kapoor, M., Kelejian, H.H., Prucha, I.R. (2007) Panel data models with spatially correlated error components, *Journal of Econometrics* **140**, 97-130.

Kelejian, H.H., Prucha, I.R. (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* **157**, 53-67.