A General Spatial ARMA Model: Theory and Application

René van der Kruk

Tinbergen Institute and Vrije Universiteit Amsterdam, department of Spatial Economics

http://www.renevdkruk.com

Abstract: In this paper the theoretical framework of a general spatial auto-regressive moving-average model is presented together with a hedonic pricing application. There is a close relationship between univerate auto-regressive and moving average (ARMA) time series models on the one hand and univerate spatial auto-regressive and moving average (SARMA) models on the other hand. The SARMA model is the spatial analogue of the well-known class of ARMA models that is developed to model time-series processes. While ARMA models relate observations in time using so-called time link matrices, in SARMA models observations in space are related using spatial link matrices. The idea behind both types of link matrices is that a certain structure is imposed on the data before the actual parameters are estimated. The advantage of the general SARMA model is that the scale of spatial processes can be modeled explicitly: a separate spatial link parameter can be estimated for each (higher order) spatial link.

1 Introduction

In this paper the theoretical framework of a general spatial auto-regressive moving-average model is presented together with a hedonic pricing application. There is a close relationship between univerate auto-regressive and moving average (ARMA) time series models on the one hand and univerate spatial auto-regressive and moving average (SARMA) models on the other hand. The SARMA model is the spatial analogue of the well-known class of ARMA models that is developed to model time-series processes. While ARMA models relate observations in time using so-called time link matrices, in SARMA models observations in space are related using spatial link matrices. The idea behind both types of link matrices is that a certain structure is imposed on the data before the actual parameters are estimated. The advantage of the general SARMA model is that the scale of spatial processes can be modeled explicitly: a separate spatial link parameter can be estimated for each (higher order) spatial link.

This paper is organized as follows. In section 2 the univariate ARMA time series model is formulated in terms of time link matrices. In section 3 the univariate spatial SARMA model is formulated using spatial link matrices. In section 4 an example is used to clarify some issues related to the use of spatial link matrices. In section 5 the maximum likelihood estimation procedure for the SARMA model is treated. Most theoretical results on the SARMA model are collected in the appendix of this paper. In section 6 the spatial autocorrelation function is introduced. In section 7 the likelihood ratio test for (residual) spatial autocorrelation in the general SARMA model is described. In section. In section 8 the data are described that are used in section 9 to apply the theoretical results presented in this paper. Section 10 gives a short summary of the main results and findings.

2 The univariate ARMA time series model

The general auto-regressive and moving average (ARMA) time series model is used to describe a process that evolves over time. In this time series context values of a variable at a certain point in time are related to past values. Consider the most commonly used (linear) functional form of the ARMA(p,q) model:

(1)
$$\begin{aligned} y_t &= \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q} = \\ &= \sum_{i=1}^{p} \theta_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i} \end{aligned}$$

In the equation above p is the order of the auto-regressive part and q denotes the order of the moving average part of the time series.

The time structure of the ARMA model in equation (1) can easily be revealed:

(2)
$$\mathbf{y} = \sum_{i=1}^{p} \theta_i T_i \mathbf{y} + \boldsymbol{\varepsilon} + \sum_{i=1}^{q} \alpha_i T_i \boldsymbol{\varepsilon}$$

In this equation:

(3) $$\mathbf{y} = \left(y_T \quad \cdots \quad y_t \quad \cdots \quad y_0 \right)'$$

(4) $$T_i = \{t_{i,jk}\}$$

(5) $$t_{i,jk} = \begin{cases} 1 & \text{for } j = 1,\ldots,T-i \text{ and } k = j+1 \\ 0 & \text{elsewhere} \end{cases}$$

(6) $$\boldsymbol{\varepsilon} = \left(\varepsilon_T \quad \cdots \quad \varepsilon_t \quad \cdots \quad \varepsilon_0 \right)'$$

The matrix T_i represents the relations between all time series observations in the sample and can therefore be called a time link matrix. The key role of the time link matrices is to impose some structure on the time units that is necessary to estimate the model. For example, the first order time link matrix T_1 relates time units that are closest in terms of time. In general, the p-th order time link matrix T_p relates time units that are p steps away in time space. Note that $T_i = (T_1)^i$ and that T_i equals the zero matrix if $i > T$, where $T + 1$ is the total number of time series observations. This is the reason why the condition $\max\{p,q\} < T$ is usually imposed.

In order to illuminate the notation used above a special case of the ARMA model is presented here as an example.

Example: the univariate AR(p) time series model

The AR(p) time series model is usually expressed as:

(7) $$y_t = \sum_{i=1}^{p} \theta_i y_{t-i} + \varepsilon_t = \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \varepsilon_t$$

This equation can also be expressed in terms of time link matrices (assuming $T > 2p$):

$$\text{(8)} \quad \begin{pmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-p+1} \\ y_{T-p} \\ y_{T-p-1} \\ \vdots \\ y_1 \\ y_0 \end{pmatrix} = \begin{bmatrix} 0 & \theta_1 & \cdots & \theta_{p-1} & \theta_p & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \theta_{p-2} & \theta_{p-1} & \theta_p & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \theta_1 & \theta_2 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \theta_1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \theta_1 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \cdot \begin{pmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-p+1} \\ y_{T-p} \\ y_{T-p-1} \\ \vdots \\ y_1 \\ y_0 \end{pmatrix} + \begin{pmatrix} \varepsilon_T \\ \varepsilon_{T-1} \\ \vdots \\ \varepsilon_{T-p+1} \\ \varepsilon_{T-p} \\ \varepsilon_{T-p-1} \\ \vdots \\ \varepsilon_1 \\ \varepsilon_0 \end{pmatrix} =$$

$$= \theta_1 \cdot \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \cdot \begin{pmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-p+1} \\ y_{T-p} \\ y_{T-p-1} \\ \vdots \\ y_1 \\ y_0 \end{pmatrix} + \cdots + \theta_p \cdot \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \cdot \begin{pmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-p+1} \\ y_{T-p} \\ y_{T-p-1} \\ \vdots \\ y_1 \\ y_0 \end{pmatrix} + \begin{pmatrix} \varepsilon_T \\ \varepsilon_{T-1} \\ \vdots \\ \varepsilon_{T-p+1} \\ \varepsilon_{T-p} \\ \varepsilon_{T-p-1} \\ \vdots \\ \varepsilon_1 \\ \varepsilon_0 \end{pmatrix}$$

What this example shows is that time link matrices impose a structure on the time series observations. This structure does not seem to be entirely arbitrary. For example, it seems reasonable to assume that y_4 is influenced by y_{4-i} and not the other way around. Moreover, one presupposes that the impact of y_4 on y_{4-i} is the same as the effect of y_{15} on y_{15-i}, namely θ_i. In other words, in the standard ARMA model the parameters are assumed to be constant over time. This assumption arises from the fact that, by definition, the time period between $t = 14$ and $t = 14 - i$ equals the time lag between $t = 15$ and $t = 15 - i$.

In the next section the link matrix concept introduced in this section is generalized in order to be able to construct a new class of ARMA models, which can be used to describe spatial processes.

3 The univariate spatial ARMA model

The spatial analogue of the time link matrix introduced in the previous section is the spatial link matrix. The general univariate spatial auto-regressive moving average, SARMA(p,q), model can be formulated as follows:

$$(9) \qquad \begin{cases} \mathbf{y} = \sum_{i=1}^{p} \theta_i S_i \mathbf{y} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} = \sum_{i=1}^{q} \alpha_i S_i \boldsymbol{\varepsilon} + \boldsymbol{\mu} \end{cases}$$

Note that the disturbance structure of the spatial ARMA model is different from the one used in time series ARMA models. In equation (9):

$$(10) \qquad S_i = \{s_{i,jk}\}$$

$$(11) \qquad s_{i,jk} = \begin{cases} 1 & \text{if spatial units } j \text{ and } k \text{ are } i-\text{th order neighbors} \\ 0 & \text{elsewhere} \end{cases}$$

$$(12) \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_T & \cdots & \mu_t & \cdots & \mu_0 \end{pmatrix}'$$

The matrix S_i encloses the relations between all spatial observations in the sample. S_1 is the first order spatial link matrix that is commonly used in spatial literature. It defines the first order spatial relation between adjacent spatial units. S_D is the highest order spatial link matrix, i.e. it represents the highest order spatial contiguity. Note that it is usually not the case that $S_i = (S_1)^i$ and that S_i equals the zero matrix if $i > D$, where D is the diameter of the graph depicting all spatial relations (see figure 1). It seems useful to choose p and q such that that max$\{p,q\} \leq D$.

The advantage of the SARMA model is that the order of contiguity of spatial units can be modeled explicitly: a separate spatial link parameter can be estimated for each (higher order) spatial link. The following section gives an idea of how to construct a SARMA model using spatial data. It also treats some differences between time link and spatial link matrices.

4 Spatial link matrices

Anselin and Smirnov (1996) use a simple example of spatial dependence, which can be represented by the following graph.

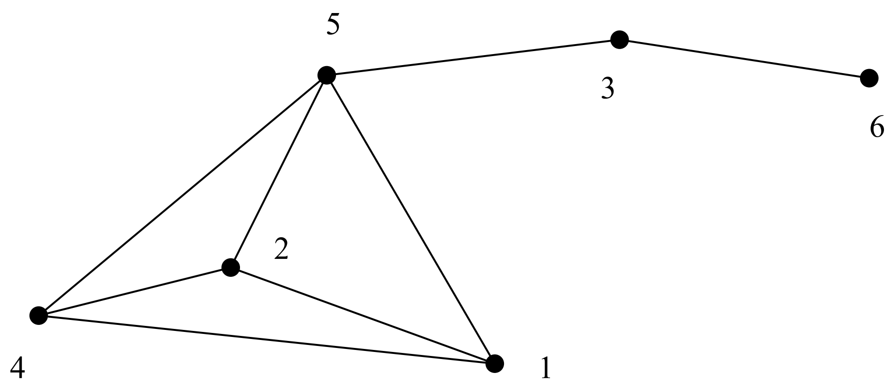

Figure 1: graph representing the dependence between spatial units.

Some of the spatial units in the graph have a direct connection in space, while other spatial units are only related indirectly. For example, one could think of cities that can be linked by means of infrastructure or countries that may or may not have certain socio-cultural relations.

Equations (10) and (11) can now be used to construct the spatial link matrices corresponding to the spatial relations depicted in figure 1:

$$(13) \qquad S_1 = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} S_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} S_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

This example shows some of the key differences between time link matrices and spatial link matrices.

First, while every row in a time link matrix relates a single observation in time to only one past observation, this one-to-one relation is not necessarily present when it comes to spatial link matrices. For example, the first order spatial link matrix relates spatial unit 1 to spatial units 2, 4, and 5. The third order spatial link matrix, however, does not relate spatial unit 1 to any other spatial unit.

Second, in contrast with time series processes it is usually the case that distances between neighboring spatial units are not constant over space. This can be a source of serious modelling problems if spatial units are clustered, i.e. if some spatial units are close to each other and if others are more remote. A simple alternative would be to replace equation (11) by the following equation:

$$(14) \qquad s_{i,jk} = \begin{cases} f(d_{j,k}) & \text{if spatial units } j \text{ and } k \text{ are } i-\text{th order neighbors} \\ 0 & \text{elsewhere} \end{cases}$$

where $f(d_{j,k})$ denotes a function $f(.)$ of $d_{j,k}$, the (Euclidian) distance between spatial units j and k. For example, Pace and Gilley (1997) use normalized weights based on the distance between spatial units. However, the problem with this *ad hoc* solution is that estimation results will crucially depend on the specification of the distance function $f(.)$.

In the application that is treated later on in this paper the spatial link matrices are constructed as follows: in the first stage the Euclidian distances between all spatial units are calculated and sorted ascending. In the second step the calculation of all spatial link matrices is based on the following rule: $s_{i,jk} = 1$ if rank($d_{j,k} = i$) and $s_{i,jk} = 0$ in all other cases. In other words, the i-th order spatial link matrix relates each spatial unit to the spatial unit that is number i on the sorted list of all distances between the spatial unit under consideration and each other spatial unit.

In the next section the maximum likelihood estimation procedure of the spatial ARMA model is described.

5 The maximum likelihood estimation procedure for the SARMA model

In the appendix of this paper the general spatial ARMA model is described together with a derivation of the elements in the information matrix. In this section a special case of the SARMA model described in equation (A.1) is considerd. It is assumed that $\Omega = \sigma^2 I_n$. Under this restriction the log likelihood function in equation (A.6) becomes:

$$\begin{aligned} L(\mathbf{y}|\boldsymbol{\theta}) = &-\frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}(A\mathbf{y} - X\boldsymbol{\beta})' B' B (A\mathbf{y} - X\boldsymbol{\beta}) - \\ &-\frac{n}{2}\ln(\sigma^2) + \ln(|B|) + \ln(|A|) \end{aligned} \tag{15}$$

Maximization of the log likelihood function in equation (15) yields the maximum likelihood estimates for β and σ^2:

(16) $$\hat{\boldsymbol{\beta}} = [(XB)'BX]^{-1}(XB)'BA\mathbf{y} = [B'X'BX]^{-1}B'X'BA\mathbf{y}$$

(17) $$\hat{\boldsymbol{\mu}} = B(A\mathbf{y} - X\hat{\boldsymbol{\beta}})$$

(18) $$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\mu}}}{n} = \frac{1}{n}(A\mathbf{y} - X\hat{\boldsymbol{\beta}})'B'B(A\mathbf{y} - X\hat{\boldsymbol{\beta}})$$

Substitution of the estimates for β and σ^2 into the log likelihood function (15) results in a concentrated likelihood of the following form:

(19) $$L(\mathbf{y}|\boldsymbol{\theta}) = C - \frac{n}{2}\ln(\hat{\sigma}^2) + \ln(|B|) + \ln(|A|)$$

where $$C = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}$$

The estimation process of the SARMA model can proceed according to the following algorithm:

1. choose some starting values for $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$;
2. estimate $\boldsymbol{\beta}$ using equation (16) : yields $\hat{\boldsymbol{\beta}}$;
3. estimate σ^2 using equations (17) and (18) : yields $\hat{\sigma}^2$;
4. find $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ that maximize $L(\mathbf{y} \mid \boldsymbol{\theta})$ defined in equation (15) : yields $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\delta}}$;
5. proceed with step 2 unless a certain convergence criterion is met;
6. given $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\delta}}$, compute $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$.

Note that step 4 in this estimation algorithm necessitates the use of a nonlinear optimization routine.

In many applications there are no *a priori* reasons to choose a particular specification of the SARMA model. Consequently, to a large extent the data will determine which model is appropriate. In the next section the spatial analogue of the time series sample autocorrelation coefficient is introduced.

6 The spatial autocorrelation function

In time series analysis, before estimating any model it is common to estimate (partial) autocorrelation coefficients. Often this gives some idea about which model specification might be correct. The time series sample autocorrelation function describes the estimated correlation between y_t and its lag y_{t-l} as a function of l. The l-th order sample autocorrelation coefficient is defined as:

$$\hat{\rho}_l = \frac{T+1}{T+1-l} \cdot \frac{\sum_{j=l}^{T} (y_j - \bar{y})(y_{j-l} - \bar{y})}{\sum_{j=0}^{T} (y_j - \bar{y})^2} = \frac{T+1}{T+1-l} \cdot \frac{\tilde{\mathbf{y}}' T_1^{l} \tilde{\mathbf{y}}}{\tilde{\mathbf{y}}' \tilde{\mathbf{y}}} \qquad (20)$$

where $\tilde{\mathbf{y}} = (I_{T+1} - \frac{1}{T+1} \iota_{T+1} \iota_{T+1}')\mathbf{y}$

The l-th order sample spatial autocorrelation coefficient can be defined as:

$$\hat{\rho}_l = \frac{T+1}{T+1-l} \cdot \frac{\sum_{j=0}^{T} \sum_{k=0}^{T} (y_j - \bar{y})(y_k - \bar{y}) s_{l,jk}}{\sum_{j=0}^{T} (y_j - \bar{y})^2} = \frac{T+1}{T+1-l} \cdot \frac{\tilde{\mathbf{y}}' S_l \tilde{\mathbf{y}}}{\tilde{\mathbf{y}}' \tilde{\mathbf{y}}} \qquad (21)$$

The sample spatial autocorrelation coefficients can be used to construct sample partial spatial autocorrelation coefficients using a recursive formula due to Durbin (1959). This coefficient measures the strength of the l-th order spatial autocorrelation among pairs of spatial units while accounting for (i.e., removing the effects of) all spatial autocorrelations below spatial order l.

Note that the quadratic form on the right hand side of equation (21) is just a multiplication of a constant and the basic form of Moran's (1948) I test statistic for spatial autocorrelation that specifies the variation of $\boldsymbol{y}$ around its mean.

It is possible to simulate the theoretical spatial autocorrelation coefficients of a simple SARMA model using the following equation:

$$\begin{cases} A\mathbf{y} = \boldsymbol{\varepsilon} \\ B\boldsymbol{\varepsilon} = \boldsymbol{\mu} \end{cases} \Leftrightarrow \mathbf{y} = A^{-1}B^{-1}\boldsymbol{\mu} \tag{22}$$

It is easy to simulate spatial data with equation (22) by simply generating a vector μ of independent standard normally distributed disturbances for a given set of spatial link matrices and parameter vectors α and δ. The simulated vector $\boldsymbol{y}$ can in turn be inserted into equation (21) to calculate the theoretical spatial autocorrelation coefficients.

7 The likelihood ratio test for residual spatial autocorrelation

It is possible to test for the presence of residual spatial autocorrelation in the SARMA model described in section 5. In this section the likelihood ratio test is described. Starting from the log likelihood function of the model

$$\begin{aligned} L^1(\mathbf{y}|\boldsymbol{\theta}) = &-\frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}(A\mathbf{y} - X\boldsymbol{\beta})'B'B(A\mathbf{y} - X\boldsymbol{\beta}) - \\ &-\frac{n}{2}\ln(\sigma^2) + \ln(|B|) + \ln(|A|) \end{aligned} \tag{23}$$

one can impose restrictions under the null hypothesis.

For example, the null hypothesis H_0: $\delta = \boldsymbol{0}$ can be tested against the alternative hypothesis H_1: $\delta \neq \boldsymbol{0}$, i.e. $B = I_n$, which yields:

$$L^0(\mathbf{y}|\boldsymbol{\theta}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}(A\mathbf{y} - X\boldsymbol{\beta})'(A\mathbf{y} - X\boldsymbol{\beta}) - \frac{n}{2}\ln(\sigma^2) + \ln(|A|) \tag{24}$$

The likelihood ratio test for residual spatial autocorrelation is based on the difference between equations (23) and (24). The coefficients in the log likelihood functions can be replaced by their maximum likelihood estimates, yielding the following likelihood ratio test statistic:

$$(25) \qquad LR^{(\delta)} = 2[L^1(\mathbf{y}|\boldsymbol{\theta}) - L^0(\mathbf{y}|\boldsymbol{\theta})] = n[\ln(\hat{\sigma}_0^2) - \ln(\hat{\sigma}_1^2)] + 2\ln(|B|) \sim \chi^2(d_3)$$

Analogue, it is possible to test the null hypothesis H_0: $\alpha = \boldsymbol{0}$ against the alternative hypothesis H_1: $\alpha \neq \boldsymbol{0}$, i.e. $A = I_n$, which yields the following test statistic:

$$(26) \qquad LR^{(\alpha)} = n[\ln(\hat{\sigma}_0^2) - \ln(\hat{\sigma}_1^2)] + 2\ln(|A|) \sim \chi^2(d_1)$$

8 Data description

Harrison and Rubinfeld (1978) apply ordinary least squares estimation to analyze the demand for clean air using housing market data from the Boston Standard Metropolitan Statistical Area in 1970. The dependent variable in the hedonic equation is the median value of owner-occupied houses in each of the 506 census tracts. There are 14 non-constant independent variables.

Pace and Gilley (1997) add the location of each tract in latitude and longitude out of the 1970 census to this data set and use a two-dimensional grid search to estimate a SARMA(0,1) model, i.e. a spatial moving average model of order one.

In the next section a class of more general SARMA models is estimated using the Harrison and Rubinfeld data.[1]

[1] The SARMA models are estimated using a slightly modified version of the Econometrics Toolbox developed by J. P. LeSage (which is downloadable free of change at internet page http://www.spatial-econometrics.com). You can download the program files and data necessary to estimate some of the models in this paper at http://www.renevdkruk.com.

9 Application: A hedonic SARMA model

In this section the theoretical SARMA model introduced in section 5 will be estimated using the data described in the previous section. However, first some attention is paid to the spatial autocorrelation coefficients in order to see which model specification might be correct.

Figure 2 depicts the sample (partial) spatial autocorrelation coefficients, which are calculated up to and including order ten using equation (21).

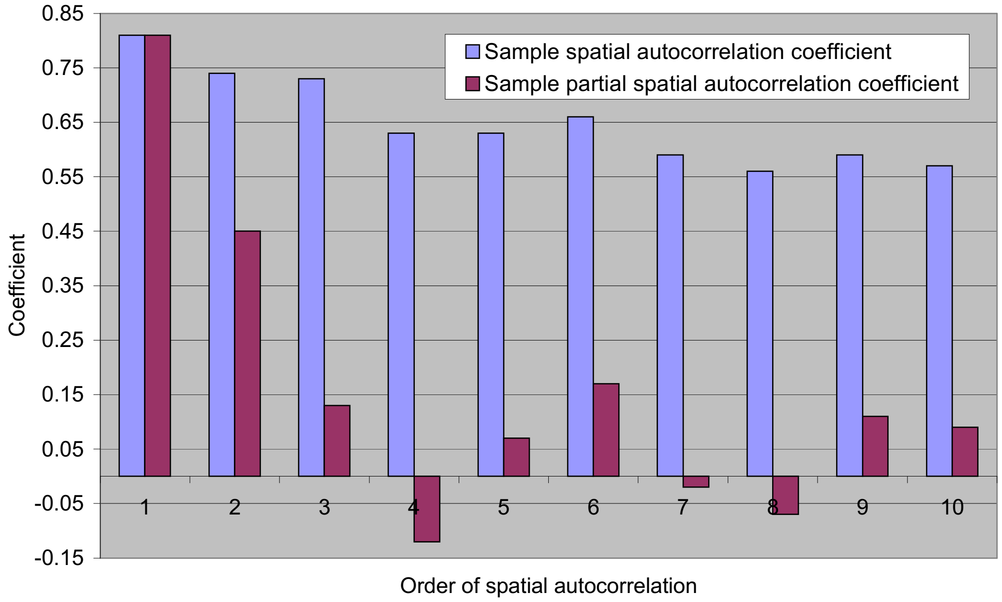

Figure 2: Some sample (partial) spatial autocorrelation coefficients of the log of the median housing prices in the Boston census tracts.

The figure above shows that the first order sample spatial autocorrelation coefficient is quite high. The higher order spatial autocorrelation coefficients decrease gradually. The first order sample partial spatial autocorrelation coefficient by definition equals the sample spatial autocorrelation coefficient. The higher order sample partial spatial autocorrelation coefficients converge rather fast towards zero for successively higher spatial lags.

It is possible to simulate the theoretical spatial autocorrelation coefficients of SARMA(10,10) models using equation (22). Table 1 contains the parameter vectors of a selection of these models. The first model shows gradually decreasing parameter values, whereas the third model is characterized by less spatial dependence. The fourth and fifth model both contain a first order spatial link parameter that is close to one, i.e. the spatial unit root case. Figure 3 presents all the theoretical spatial autocorrelation coefficients for each model based on one and the same random draw of μ.

Table 1: The parameter vectors of 5 theoretical SARMA(10,10) models.

	model 1		model 2		model 3		model 4		model 5	
order	α	ρ	α	ρ	α	ρ	α	ρ	α	ρ
1	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.995	0.995	0.95
2	0.85	0.85	0.75	0.75	0.55	0.55	0.85	0.85	0.85	0.85
3	0.75	0.75	0.55	0.55	0.15	0.15	0.75	0.75	0.75	0.75
4	0.65	0.65	0.35	0.35	0	0	0.65	0.65	0.65	0.65
5	0.55	0.55	0.15	0.15	0	0	0.55	0.55	0.55	0.55
6	0.45	0.45	0	0	0	0	0.45	0.45	0.45	0.45
7	0.35	0.35	0	0	0	0	0.35	0.35	0.35	0.35
8	0.25	0.25	0	0	0	0	0.25	0.25	0.25	0.25
9	0.15	0.15	0	0	0	0	0.15	0.15	0.15	0.15
10	0.05	0.05	0	0	0	0	0.05	0.05	0.05	0.05

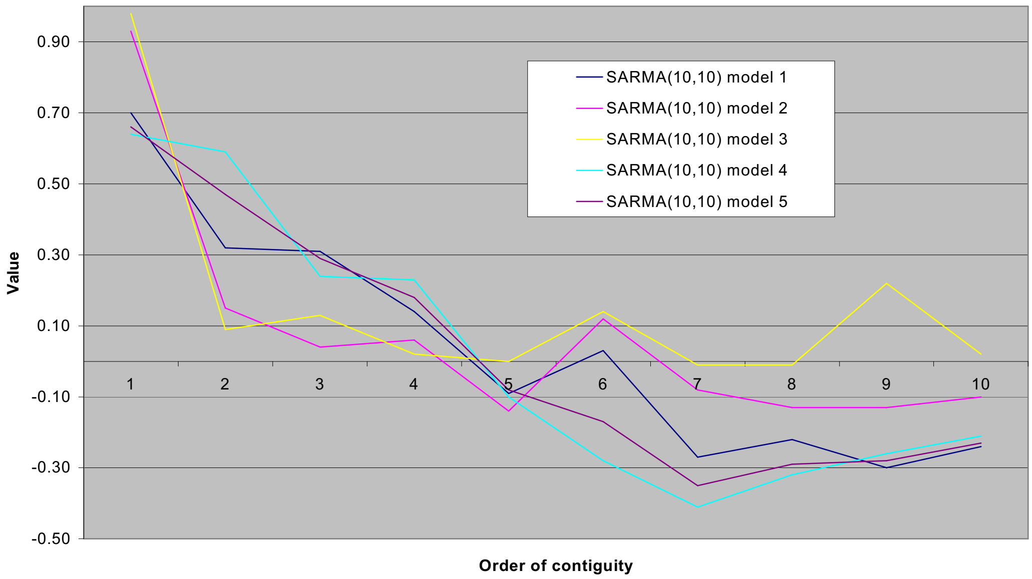

Figure 3: The theoretical spatial autocorrelation coefficients for the five SARMA(10,10) models described in table 1.

Figure 3 shows that it is difficult to choose between different types of SARMA models when comparing both the sample and the theoretical spatial autocorrelation coefficients. Moreover, it seems difficult to detect spatial unit roots. One way out is to estimate different kinds of SARMA models.

Table 2 shows the estimation results of four types of SARMA models, as described in section 5, along with the estimates of the classical linear regression model.

Table 2: The estimates of 4 SARMA models and the classical linear regression model.

Independent variables	**OLS**		**SARMA(1,1)**		**SARMA(4,0)**		**SARMA(0,4)**		**SARMA(4,4)**	
	par	**t-stat**	**par**	**t-stat**	**par**	**t-stat**	**par**	**t-stat**	**par**	**t-stat**
constant	4.56	29.95	3.53	14.35	2.27	0.00	3.64	60.87	3.13	33.46
crime	-0.01	-9.59	-0.01	-8.82	-0.01	-8.03	-0.01	-6.02	-0.01	-6.13
zoning	0.00	0.18	0.00	0.43	0.00	1.11	0.00	1.46	0.00	1.41
amount of industrial land	0.00	0.08	0.00	0.26	0.00	0.69	0.00	0.02	0.00	0.39
contiguity of the Charles river	0.09	2.81	0.04	1.19	0.02	0.91	-0.03	-0.95	-0.02	-0.64
(nitrous oxide concentration)^2	-0.64	-5.71	-0.43	-4.04	-0.28	-3.51	-0.15	-1.04	-0.17	-1.27
(number of rooms)^2	0.01	4.83	0.01	6.10	0.01	7.48	0.01	10.07	0.01	9.17
age	0.00	0.14	0.00	-0.91	0.00	-0.51	0.00	-2.97	0.00	-2.53
distance to the center of Boston	-0.20	-6.01	-0.17	-4.36	-0.15	-6.31	-0.11	-2.61	-0.15	-3.42
accessibility	0.09	4.75	0.08	1.76	0.07	0.00	0.06	2.78	0.06	3.09
tax burden	0.00	-3.46	0.00	-0.80	0.00	0.00	0.00	-3.69	0.00	-2.95
pupils-to-teachers	-0.03	-5.99	-0.02	-5.33	-0.01	-3.20	-0.02	-3.43	-0.02	-3.28
race	0.00	3.55	0.00	3.02	0.00	2.22	0.00	5.85	0.00	5.19
prop. of lower status individuals	-37.49	-15.20	-28.81	-11.72	-23.94	-13.17	-24.93	-11.30	-25.49	-11.69
SARMA Parameters	**OLS**		**SARMA(1,1)**		**SARMA(4,0)**		**SARMA(0,4)**		**SARMA(4,4)**	
	par	**t-stat**	**par**	**t-stat**	**par**	**t-stat**	**par**	**t-stat**	**par**	**t-stat**
α_1	-	-	0.18	6.13	0.18	7.16	-	-	0.07	2.79
α_2	-	-	-	-	0.13	5.05	-	-	0.07	2.63
α_3	-	-	-	-	0.11	4.65	-	-	0.00	0.06
α_4	-	-	-	-	0.05	2.39	-	-	0.04	1.40
δ_1	-	-	0.22	7.86	-	-	0.16	4.32	0.12	2.57
δ_2	-	-	-	-	-	-	0.21	5.13	0.14	2.73
δ_3	-	-	-	-	-	-	0.21	5.24	0.24	4.41
δ_4	-	-	-	-	-	-	0.18	5.12	0.14	2.80
Other statistics	**OLS**		**SARMA(1,1)**		**SARMA(4,0)**		**SARMA(0,4)**		**SARMA(4,4)**	
number of observations	506		506		506		506		506	
number of variables	14		16		18		18		22	
number of iterations	1		22		106		58		481	
maximum log likelihood value	156.96		227.48		254.86		275.21		279.69	
σ^2	0.03		0.02		0.02		0.02		0.02	
R^2	0.81		0.86		0.88		0.89		0.90	
adjusted R^2	0.81		0.86		0.88		0.89		0.89	

From table 2 it appears that the higher order SARMA models perform relatively good: the (adjusted) R^2-values are higher as well as the maximum log likelihood function values. Moreover, most SARMA parameter estimates are significantly different from zero. Remarkably, the estimated NO_x (pollution) parameter becomes less negative if more spatial links are included in the model.

It is possible to test the SARMA(4,4) model against either the SARMA(4,0) model or the SARMA(0,4) model. First, the likelihood ratio test statistic corresponding to the null hypothesis H_0: $\delta = \mathbf{0}$ against the alternative hypothesis H_1: $\delta \neq \mathbf{0}$ equals 49.66 while the critical value of the $\chi^2(4)$ distribution at a critical value of 5% is equal to 9.49. Second, the likelihood ratio test statistic corresponding to the null hypothesis H_0: $\alpha = \mathbf{0}$ against the alternative hypothesis H_1: $\alpha \neq \mathbf{0}$ equals 8.96. This means that the SARMA(4,4) model is not significantly performing better than the SARMA(0,4) model.

10 Conclusion

The advantage of the general SARMA model is that the order of contiguity of spatial units can be modeled explicitly: a separate spatial link parameter can be estimated for each spatial link. In this paper the elements in the information matrix corresponding to the SARMA model are derived and the maximum likelihood estimation procedure for the SARMA model is treated. The spatial autocorrelation function does not seem to be as promising as a tool for the selection of the suitable spatial model in comparison with the likelihood ratio test for (residual) spatial autocorrelation. The hedonic pricing model estimation results show the superiority of the general spatial ARMA model over other (spatial) linear models.

Literature

Anselin, L. 1988. Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht.

Anselin, L and O. Smirnov. 1996. "Efficient algorithms for constructing proper higher order spatial lag operators." *Journal of Regional Science* 36:67-89.

Durbin, J. 1959. "The fitting of time-series models." *Revue de l'Institut International de Statistique* 28:229-249.

Harrison, D., and D. L. Rubinfeld. 1978. "Hedonic housing prices and the demand for clean air." *Journal of Environmental Economics and Management* 5:81-102.

Moran, P. A. P. 1948. "The interpretation of statistical maps." *Journal of the Royal Statistical Society B* 10:243-251.

Pace, R. Kelley, and O.W. Gilley. 1997. "Using the spatial configuration of the data to improve estimation." *Journal of the Real Estate Finance and Economics* 14(3): 333-340.

Appendix: the SARMA model

In this appendix the general spatial ARMA model is described. Moreover, the elements in the information matrix are derived. The notation used here closely follows Anselin (1988, pp. 61-65 and 76-77) who specifies the SARMA(1,1) model.

Consider the general SARMA model:

$$(27) \qquad \begin{cases} \mathbf{y} = \sum_{i=1}^{d_1} \alpha_i S_i \mathbf{y} + \sum_{i=1}^{d_2} \zeta_i S_i \mathbf{z} + Y\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} = \sum_{i=1}^{d_3} \delta_i S_i \boldsymbol{\varepsilon} + \boldsymbol{\mu} \\ \boldsymbol{\mu} \sim N(0, \Omega) \\ \Omega_{ii} = h_i(z\boldsymbol{\kappa}) \quad h_i > 0 \end{cases}$$

The dimensions of the variables and parameters are as follows:

$\boldsymbol{y}$	$n \times 1$	α_i	1×1	S_i	$n \times n$	d_i	1×1
ζ_i	1×1	$\boldsymbol{z}$	$n \times 1$	Y	$n \times m$	γ	$m \times 1$
ε	$n \times 1$	δ_i	1×1	μ	$n \times 1$	Ω	$n \times n$
$\boldsymbol{\kappa}$	$p \times 1$	h_i	1×1				

The model has $d_1 + d_2 + m + d_3 + 1 + p$ unknown parameters:

$$\theta \;\; = [\ \alpha_1 \dots \alpha_{d1}\ \zeta_1 \dots \zeta_{d2}\ \gamma'\ \delta_1 \dots \delta_{d3}\ \sigma^2\ \boldsymbol{\kappa}']'$$

Now, define:

$$(28) \qquad \begin{cases} A = I - \sum_{i=1}^{d_1} \alpha_i S_i \\ B = I - \sum_{i=1}^{d_3} \delta_i S_i \\ X\beta = (\sum_{i=1}^{d_2} \zeta_i S_i)\mathbf{x} + Y\boldsymbol{\gamma} \end{cases}$$

Where A is $n \times n$, B is $n \times n$, X is $n \times (d_2 + m)$ and β is $(d_2 + m) \times 1$.

The model in equation (A.1) can be rewritten using (A.2) as:

$$(29) \qquad \begin{cases} A\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ B\varepsilon = \boldsymbol{\mu} \end{cases}$$

Since Ω is diagonal, there exists a vector ν of homoskedastic random disturbances:

$$(30) \qquad \begin{aligned} \boldsymbol{v} &= \Omega^{-\frac{1}{2}} B(A\mathbf{y} - X\boldsymbol{\beta}) \\ \left|\frac{\partial \boldsymbol{v}}{\partial \mathbf{y}}\right| &= \left|\Omega^{-\frac{1}{2}} BA\right| = \left|\Omega^{-\frac{1}{2}}\right| |B| |A| \end{aligned}$$

Equation (A.4) gives the following result:

$$(31) \qquad g(\mathbf{y}) = f(\Omega^{-\frac{1}{2}} B(A\mathbf{y} - X\boldsymbol{\beta})) \left|\frac{\partial \boldsymbol{v}}{\partial \mathbf{y}}\right| = (2\pi)^{-\frac{n}{2}} \exp\{-\frac{1}{2}\boldsymbol{v}'\boldsymbol{v}\} \left|\Omega^{-\frac{1}{2}}\right| |B| |A|$$

The likelihood function is thus defined as:

$$(\text{A.6}) \qquad L(\mathbf{y}|\boldsymbol{\theta}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\boldsymbol{v}'\boldsymbol{v} - \frac{1}{2}\ln(|\Omega|) + \ln(|B|) + \ln(|A|)$$

The first order and second order conditions can now be derived. In turn, the elements of the information matrix can be determined..

First order conditions :

$$\frac{\partial L(\mathbf{y}|\theta)}{\partial \alpha_i} = \nu'\Omega^{-\frac{1}{2}} BS_i\mathbf{y} - \mathrm{tr}(A^{-1}S_i)$$

$$\frac{\partial L(\mathbf{y}|\theta)}{\partial \delta_i} = \nu'\Omega^{-\frac{1}{2}} S_i(A\mathbf{y} - X\boldsymbol{\beta}) - \mathrm{tr}(B^{-1}S_i)$$

$$\frac{\partial L(\mathbf{y}|\theta)}{\partial \boldsymbol{\beta}} = \nu'\Omega^{-\frac{1}{2}} BX$$

$$\frac{\partial L(\mathbf{y}|\theta)}{\partial \kappa_p} = \frac{1}{2}\nu'\Omega^{-\frac{3}{2}} H_p B(A\mathbf{y} - X\boldsymbol{\beta}) - \frac{1}{2}\mathrm{tr}(\Omega^{-1}H_p)$$

Second order conditions :

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = -(BX)'\Omega^{-1}BX$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \alpha_i \partial \alpha_j} = -\mathrm{tr}(A^{-1}S_jA^{-1}S_i) - (BS_j\mathbf{y})'\Omega^{-1}BS_i\mathbf{y}$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \delta_i \partial \delta_j} = -\mathrm{tr}(B^{-1}S_jB^{-1}S_i) - [S_j(A\mathbf{y} - X\boldsymbol{\beta})]'\Omega^{-1}S_i(A\mathbf{y} - X\boldsymbol{\beta})$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \boldsymbol{\beta}\partial \alpha_i} = \frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \alpha_i \partial \boldsymbol{\beta}'} = -(BX)'\Omega^{-1}BS_i\mathbf{y}$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \boldsymbol{\beta}\partial \delta_i} = \frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \delta_i \partial \boldsymbol{\beta}'} = -(BX)'\Omega^{-1}S_i(A\mathbf{y} - X\boldsymbol{\beta}) - \nu'\Omega^{-\frac{1}{2}}S_iX$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \boldsymbol{\beta}\partial \kappa_p} = \frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \kappa_p \partial \boldsymbol{\beta}'} = -\frac{1}{2}(BX)'\Omega^{-2}H_pB(A\mathbf{y} - X\boldsymbol{\beta}) - \frac{1}{2}\nu'\Omega^{-1\frac{1}{2}}H_pBX$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \alpha_i \partial \delta_j} = \frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \delta_j \partial \alpha_i} = -[S_j(A\mathbf{y} - X\boldsymbol{\beta})]'\Omega^{-1}BS_i\mathbf{y} - \nu'\Omega^{-\frac{1}{2}}S_jS_i\mathbf{y}$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \alpha_i \partial \kappa_p} = \frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \kappa_p \partial \alpha_i} = -\frac{1}{2}(BS_i\mathbf{y})'\Omega^{-2}H_pB(A\mathbf{y} - X\boldsymbol{\beta}) - \frac{1}{2}\nu'\Omega^{-1\frac{1}{2}}H_pBS_i\mathbf{y}$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \delta_i \partial \kappa_p} = -\frac{1}{2}[S_i(A\mathbf{y} - X\boldsymbol{\beta})]'\Omega^{-2}H_pB(A\mathbf{y} - X\boldsymbol{\beta}) - \frac{1}{2}\nu'\Omega^{-1\frac{1}{2}}H_pS_i(A\mathbf{y} - X\boldsymbol{\beta})$$

$$\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial \kappa_p \partial \kappa_q} = \frac{1}{2}\mathrm{tr}(\Omega^{-2}H_qH_p) - \frac{1}{2}\mathrm{tr}(\Omega^{-1}H_{pq}) - \frac{1}{4}[B(A\mathbf{y} - X\boldsymbol{\beta})]'\Omega^{-3}H_qH_pB(A\mathbf{y} - X\boldsymbol{\beta}) -$$

$$-\frac{3}{4}\nu'\Omega^{-\frac{5}{2}}H_qH_pB(A\mathbf{y} - X\boldsymbol{\beta}) + \frac{1}{2}\nu'\Omega^{-\frac{3}{2}}H_{pq}B(A\mathbf{y} - X\boldsymbol{\beta})$$

Expected values :

$$\mathrm{E}[\varepsilon] = \mathrm{E}[\mu] = \mathrm{E}[\nu] = 0$$

$$\mathrm{E}[\varepsilon\varepsilon'] = B^{-1}\Omega(B^{-1})'$$

$$\mathrm{E}[\mu\mu'] = \Omega$$

$$\mathrm{E}[\nu\nu'] = I$$

$$\mathrm{E}[\mathbf{y}] = A^{-1}X\beta$$

$$\mathrm{E}[\mathbf{y}\mathbf{y}'] = A^{-1}X\beta(A^{-1}X\beta)' + A^{-1}B^{-1}\Omega(A^{-1}B^{-1})'$$

Elements information matrix :

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right] = (BX)'\Omega^{-1}BX$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\alpha_i\partial\alpha_j}\right] = \mathrm{tr}(A^{-1}S_jA^{-1}S_i) + \mathrm{tr}((BS_jA^{-1}X\boldsymbol{\beta})'\Omega^{-1}BS_iA^{-1}X\boldsymbol{\beta}) + \mathrm{tr}(\Omega(BS_jA^{-1}B^{-1})'\Omega^{-1}BS_iA^{-1}B^{-1})$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\delta_i\partial\delta_j}\right] = \mathrm{tr}(B^{-1}S_jB^{-1}S_i) + \mathrm{tr}(\Omega(S_jB^{-1})'\Omega^{-1}S_iB^{-1})$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\boldsymbol{\beta}\partial\alpha_i}\right] = -\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\alpha_i\partial\boldsymbol{\beta}'}\right] = (BX)'\Omega^{-1}BS_iA^{-1}X\boldsymbol{\beta}$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\boldsymbol{\beta}\partial\delta_i}\right] = -\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\delta_i\partial\boldsymbol{\beta}'}\right] = 0$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\boldsymbol{\beta}\partial\kappa_p}\right] = -\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\kappa_p\partial\boldsymbol{\beta}'}\right] = 0$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\alpha_i\partial\delta_j}\right] = -\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\delta_j\partial\alpha_i}\right] = \mathrm{tr}((S_jB^{-1})'\Omega^{-1}BS_iA^{-1}B^{-1}\Omega) + \mathrm{tr}(S_jS_iA^{-1}B^{-1})$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\alpha_i\partial\kappa_p}\right] = -\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\kappa_p\partial\alpha_i}\right] = \frac{1}{2}\mathrm{tr}(\Omega^{-2}H_p\Omega(BS_iA^{-1}B^{-1})' + \Omega^{-1}H_pBS_iA^{-1}B^{-1})$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\delta_i\partial\kappa_p}\right] = \frac{1}{2}\mathrm{tr}(\Omega^{-2}H_p\Omega(S_iB^{-1})' + \Omega^{-1}H_pS_iB^{-1})$$

$$-\mathrm{E}\left[\frac{\partial^2 L(\mathbf{y}|\theta)}{\partial\kappa_p\partial\kappa_q}\right] = \frac{1}{2}\mathrm{tr}(\Omega^{-2}H_qH_p)$$