Computing all roots of the likelihood equations of seemingly unrelated regressions

Mathias Drton*

Department of Statistics, University of Chicago, Chicago, IL 60637, USA

Abstract

Seemingly unrelated regressions are statistical regression models based on the Gaussian distribution. They are popular in econometrics but also arise in graphical modeling of multivariate dependencies. In maximum likelihood estimation, the parameters of the model are estimated by maximizing the likelihood function, which maps the parameters to the likelihood of observing the given data. By transforming this optimization problem into a polynomial optimization problem, it was recently shown that the likelihood function of a simple bivariate seemingly unrelated regressions model may have several stationary points. Thus local maxima may complicate maximum likelihood estimation. In this paper, we study several more complicated seemingly unrelated regression models, and show how all stationary points of the likelihood function can be computed using algebraic geometry.
© 2005 Elsevier Ltd. All rights reserved.

* Corresponding address: University of Chicago, Department of Statistics, 5734 S. University Ave, Eckhart Hall 116, Chicago, IL, 60637, USA. Tel.: +1 773 702 8336; fax: +1 773 702 9810.
E-mail address: drton@galton.uchicago.edu.

1. Introduction

Seemingly unrelated regressions (SUR) are multivariate regression models with correlated response (or dependent) variables that follow a joint Gaussian distribution. Usually different regressions contain different covariates (or independent variables) and seem "unrelated". However, due to the correlated response variables the regressions are only "seemingly unrelated" and contain valuable information about each other (Zellner, 1962). SUR play "a central role in contemporary econometrics" (Goldberger, 1991, p. 323) but also appear in other contexts (Rochon, 1996a,b; Verbyla and Venables, 1988). Moreover, SUR arise in the context of Gaussian graphical models (Andersson et al., 2001, Section 5; Richardson and Spirtes, 2002, Section 8.5).

The parameters of a SUR model can be estimated efficiently, i.e. with small variance, by maximizing the likelihood function, which maps the parameters to the likelihood of observing the given data. Oberhofer and Kmenta (1974) and Telser (1964) give two popular algorithms for this maximization. In general, however, these algorithms will not globally maximize the likelihood function, which indeed may be multimodal; a fact neglected in the literature (Drton and Richardson, 2004, Section 6). Drton and Richardson (2004) demonstrated the possibility of multimodality in a study of a bivariate SUR model that may have a likelihood function with five stationary points. In this paper, we use algebraic geometry to apply the approach of Drton and Richardson (2004) to more general SUR models. In Sections 2 and 3 we give an introduction to SUR and show how maximum likelihood estimation can be performed by solving a polynomial optimization problem, opening the door for tools from algebraic geometry. With these tools, we first revisit the work by Drton and Richardson (2004), see Section 4, and then obtain new results on more general SUR models (Section 5). In particular, we identify examples of SUR models, for which all stationary points of the likelihood function can be computed.

2. Seemingly unrelated regressions

In SUR a family of response variables, indexed by a finite set R, is stochastically modeled using a family of covariates, indexed by a finite set C. All response variables and all covariates are observed on a finite set of subjects N. We denote the cardinalities of the three sets also by R, C and N, respectively. The observations can be represented by two matrices X and Y. The matrix $Y = (Y_{rm}) \in \mathbb{R}^{R \times N}$ has the (r, m)-entry equal to the observation of response variable $r \in R$ on subject $m \in N$, and the matrix $X = (X_{cm}) \in \mathbb{R}^{C \times N}$ has the (c, m)-entry equal to the observation of covariate $c \in C$ on subject $m \in N$. For $c \in C$ and $r \in R$, $X_c \in \mathbb{R}^N$ and $Y_r \in \mathbb{R}^N$ denote the c-th and r-th rows of X and Y, respectively. Similarly, X^m and Y^m, $m \in N$, denote the m-th columns of X and Y, respectively. Clearly, X_c and Y_r comprise all observations of the c-th covariate and the r-th response variable; X^m and Y^m comprise all covariate and response variable observations on the m-th subject.

In this regression setting, the matrix X is assumed to be deterministic and fixed but the matrix Y is modeled to follow a multivariate normal distribution, where the mean vector of Y_r, $r \in R$, is a linear combination of some X_c, $c \in C_r \subseteq C$,

$$\mathrm{E}[Y_r] = \sum_{c \in C_r} \beta_{rc} X_c \in \mathbb{R}^N, \quad r \in R. \tag{1}$$

Here $(C_r \mid r \in R)$ is a fixed family of subsets of C indexing the covariates involved in each one of the R regressions. The weights β_{rc} in (1) are called *regression coefficients*. Setting $\beta_{rc} = 0$ if $c \notin C_r$, we can define a matrix of regression coefficients $B = (\beta_{rc}) \in \mathbb{R}^{R \times C}$. The random vectors Y^m, $m \in N$, are assumed to be independent with common positive definite covariance matrix

$$\mathrm{Var}[Y^m] = \Sigma \in \mathbb{R}^{R \times R}, \quad m \in N. \tag{2}$$

Letting

$$\mathcal{C_R} = \cup(\{r\} \times C_r \mid r \in R) \subseteq R \times C, \tag{3}$$

the *seemingly unrelated regressions model* is the family of normal distributions

$$\mathbf{N}(\mathcal{C_R}) = \big(\mathcal{N}(BX, \Sigma \otimes I_N) \,\big|\, (B, \Sigma) \in \mathbb{B}(\mathcal{C_R}) \times \mathbb{P}\big). \tag{4}$$

Here $\mathcal{N}$ is the multivariate normal distribution on $\mathbb{R}^{R \times N}$; I_N is the $N \times N$ identity matrix; $\otimes$ is the Kronecker product; B and Σ are the mean and the variance parameters; and *the parameter space* $\mathbb{B}(\mathcal{C_R}) \times \mathbb{P}$ is the Cartesian product of the linear space

$$\mathbb{B}(\mathcal{C_R}) = \big\{ B \in \mathbb{R}^{R \times C} \,\big|\, B = (\beta_{rc}), \; \beta_{rc} = 0 \; \forall (r, c) \notin \mathcal{C_R} \big\} \tag{5}$$

and the cone $\mathbb{P}$ of all positive definite real $R \times R$ matrices. The response matrix Y is then an observation from some (unknown) distribution in the model,

$$Y \sim \mathcal{N}(BX, \Sigma \otimes I_N), \quad (B, \Sigma) \in \mathbb{B}(\mathcal{C_R}) \times \mathbb{P}.$$

If $N \geq R + C$ and X is a matrix of full rank, then with probability one the $(R + C) \times N$ matrix obtained by stacking X and Y has full rank,

$$\mathrm{rank} \begin{pmatrix} Y \\ X \end{pmatrix} = R + C. \tag{6}$$

We assume (6) to hold throughout the paper.

3. Maximum likelihood estimation by polynomial optimization

The *probability density function* $f_{(B,\Sigma)} : \mathbb{R}^{R \times N} \to (0, \infty)$ of the distribution $\mathcal{N}(BX, \Sigma \otimes I_n)$ can be written as

$$f_{(B,\Sigma)}(Y) = \frac{1}{\sqrt{(2\pi)^{RN} |\Sigma|^N}} \exp\left\{ -\frac{1}{2} \mathrm{tr}\big[\Sigma^{-1} (Y - BX)(Y - BX)' \big] \right\}.$$

For data Y, the *likelihood function* $L : \mathbb{B}(\mathcal{C_R}) \times \mathbb{P} \to (0, \infty)$ of the model $\mathbf{N}(\mathcal{C_R})$ is defined as

$$L(B, \Sigma) = f_{(B,\Sigma)}(Y).$$

In maximum likelihood estimation the parameters (B, Σ) are estimated by

$$(\hat{B}, \hat{\Sigma}) = \arg\max\{L(B, \Sigma) \mid (B, \Sigma) \in \mathbb{B}(\mathcal{C}_\mathcal{R}) \times \mathbb{P}\}. \tag{7}$$

It follows from (6) that the maximum of the likelihood function exists.

We can parameterize $\mathbb{B}(\mathcal{C}_\mathcal{R})$ by mapping a vector

$$\beta = (\beta_{rc} \mid (r, c) \in \mathcal{C}_\mathcal{R}) \in \mathbb{R}^{\mathcal{C}_\mathcal{R}},$$

to the matrix $B(\beta) \in \mathbb{B}(\mathcal{C}_\mathcal{R})$ with entry $B(\beta)_{rc} = \beta_{rc}$ if $(r, c) \in \mathcal{C}_\mathcal{R}$ and $B(\beta)_{rc} = 0$ otherwise. Define $\ell : \mathbb{R}^{\mathcal{C}_\mathcal{R}} \times \mathbb{P} \to \mathbb{R}$ by

$$\begin{aligned} \ell(\beta, \Sigma) &= \log L(B(\beta), \Sigma) \\ &\propto -\tfrac{N}{2} \log |\Sigma| - \tfrac{1}{2}\mathrm{tr}\big[\Sigma^{-1}\big(Y - B(\beta)X\big)\big(Y - B(\beta)X\big)'\big]. \end{aligned} \tag{8}$$

Clearly we can solve (7) by finding

$$(\hat{\beta}, \hat{\Sigma}) = \arg\max\{\ell(\beta, \Sigma) \mid (\beta, \Sigma) \in \mathbb{R}^{\mathcal{C}_\mathcal{R}} \times \mathbb{P}\}, \tag{9}$$

and setting $\hat{B} = B(\hat{\beta})$. The standard approach for solving (9) is to solve the *likelihood equations*

$$\left(\frac{\partial \ell(\beta, \Sigma)}{\partial \beta}, \frac{\partial \ell(\beta, \Sigma)}{\partial \Sigma}\right) = 0. \tag{10}$$

It can be shown that (10) holds if and only if

$$\Sigma = \frac{1}{N}\big(Y - B(\beta)X\big)\big(Y - B(\beta)X\big)' \tag{11}$$

and

$$\beta = \Big[A'(XX' \otimes \Sigma^{-1})A\Big]^{-1} A'\mathrm{vec}(\Sigma^{-1}YX'), \tag{12}$$

where A is a matrix of zeros and ones that satisfies $\mathrm{vec}(B(\beta)) = A\beta$. In fact, each column of A has precisely one entry equal to one and the remaining entries equal to zero. Oberhofer and Kmenta (1974) show how one solution to the likelihood equations can be obtained by alternating between solving (11) for fixed β and solving (12) for fixed Σ. Here, we take a different approach that, for certain SUR models, allows us to compute all solutions to the likelihood equations.

From (6) and (8), it follows that for fixed $\beta \in \mathbb{R}^{\mathcal{C}_\mathcal{R}}$ the function $\ell_\beta : \Sigma \mapsto \ell(\beta, \Sigma)$ is strictly concave with maximizer (11). Thus the *profile log-likelihood function* $\ell_{\mathrm{prof}} : \mathbb{R}^{\mathcal{C}_\mathcal{R}} \to \mathbb{R}$ defined as

$$\ell_{\mathrm{prof}}(\beta) = \max\{\ell(\beta, \Sigma) \mid \Sigma \in \mathbb{P}\} \tag{13}$$

takes on the form

$$\ell_{\mathrm{prof}}(\beta) \propto -\frac{N}{2} \log \left| \frac{1}{N}(Y - B(\beta)X)(Y - B(\beta)X)' \right| - \frac{RN}{2}. \tag{14}$$

By the strict concavity of ℓ_β, (β, Σ) is a stationary point of $\ell(\beta, \Sigma)$ if and only if β is a stationary point of $\ell_{\text{prof}}(\beta)$ and Σ satisfies (11); compare Drton and Richardson (2004, Lemma 1). The same holds for

$$G(\beta) = |(Y - B(\beta)X)(Y - B(\beta)X)'|, \tag{15}$$

which conveniently is a polynomial in β. Thus we can solve (9) by using (11) and solving the unconstrained polynomial program

$$\hat{\beta} = \arg\min\{G(\beta) \mid \beta \in \mathbb{R}^{\mathcal{C}_\mathcal{R}}\}. \tag{16}$$

We try to solve (16) by computing the stationary points of G, i.e. by solving the equations

$$g_{rc} = \frac{\partial G(\beta)}{\partial \beta_{rc}} = 0, \quad (r, c) \in \mathcal{C}_\mathcal{R}. \tag{17}$$

In practice the observations Y and X are available only in finite accuracy and the partial derivatives g_{rc}, $(r, c) \in \mathcal{C}_\mathcal{R}$, are elements of the ring $\mathbb{Q}[\beta]$ of polynomials in β with rational coefficients. In an algebraic approach to solving polynomial equations (Cox et al., 1997, 1998; Sturmfels, 2002) we allow the indeterminants in the polynomial equation system (17) to be complex, i.e. $\beta \in \mathbb{C}^{\mathcal{C}_\mathcal{R}}$, where $\mathbb{C}$ is the field of complex numbers. We define the *maximum likelihood ideal* I_G to be the ideal of $\mathbb{Q}[\beta]$ that is generated by the partial derivatives g_{rc}, $(r, c) \in \mathcal{C}_\mathcal{R}$, i.e.

$$I_G = \langle g_{rc} \mid (r, c) \in \mathcal{C}_\mathcal{R} \rangle; \tag{18}$$

compare Sturmfels (2002, Section 8.4) where maximum likelihood ideals are defined in a different statistical context. Software like Macaulay 2[1] and Singular (Greuel et al., 2001) permits us to check whether I_G is a zero-dimensional ideal. If $\dim(I_G) = 0$, then the variety $V_{\mathbb{C}}(I_G)$, i.e. the set of common complex zeros of the partial derivatives g_{rc}, is a finite set and all its elements can be computed using, for example, Singular or PHCpack.[2] The real points $V_{\mathbb{R}}(I_G) = V_{\mathbb{C}}(I_G) \cap \mathbb{R}^{\mathcal{C}_\mathcal{R}}$ can then be identified and yield the stationary points of G.

4. Revisiting the multimodal bivariate seemingly unrelated regressions with two covariates

Drton and Richardson (2004) study a SUR model with two response variables and two covariates, in which response variable 1 is regressed only on covariate 1, and response variable 2 only on covariate 2. Hence, $R = \{1, 2\}$, $C = \{1, 2\}$, $C_1 = \{1\}$, and $C_2 = \{2\}$. Therefore, $\mathcal{C}_\mathcal{R} = \{(1, 1), (2, 2)\}$, and $B \in \mathbb{B}(\mathcal{C}_\mathcal{R})$ if B is of the form

$$B = \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix} \in \mathbb{R}^{2 \times 2}.$$

Using Singular and the data in Drton and Richardson (2004, Table 1), we can solve (16) as shown in Table 1.

---

[1] http://www.math.uiuc.edu/Macaulay2/.

[2] http://www.math.uic.edu/~jan/.

Table 1
Singular-session for the model in Drton and Richardson (2004)

```
> ring R=0,(b(1..2)), lp;
> matrix X[2][8] = 188,22,-46,77,-103,74,83,101,
.                  55,-216,116,-30,131,195,-311,-239;
> matrix Y[2][8] = 234,-5,6,182,-193,278,62,-68,
.                  497,-326,266,-3,93,558,-584,-224;
> matrix B[2][2] = b(1),0, 0,b(2);
> poly G = det((Y-B*X)*transpose(Y-B*X));
> ideal IG =jacob(G);
> ideal J = groebner(IG);
> dim(J); vdim(J);
0
5
> LIB "solve.lib"; solve(J,6);
[1]:
   [1]:   0.778796
   [2]:   1.538029
[2]:
   [1]:   1.622609
   [2]:   2.034745
[3]:
   [1]:   (1.480687-i*1.547274)
   [2]:   (2.16845+i*0.765283)
[4]:
   [1]:   (1.480687+i*1.547274)
   [2]:   (2.16845-i*0.765283)
[5]:
   [1]:   2.764418
   [2]:   2.504006
```

As computed by `dim` and `vdim`, the maximum likelihood ideal $I_G = $ `IG` is zero-dimensional and of degree five. The five points in the variety $V_{\mathbb{C}}(I_G)$ are computed by `solve`, which lists $\beta_{11} = $ `b(1)` as the first component and $\beta_{22} = $ `b(2)` as the second component. There are three real points in $V_{\mathbb{R}}(I_G)$, which yield the stationary points of the likelihood function of the model $\mathbf{N}(\mathcal{C}_{\mathcal{R}})$. Note that we confirm the values stated in Drton and Richardson (2004, Table 2 with $\beta_{11} = \beta_1$ and $\beta_{22} = \beta_2$). The Gröbner basis computed by the command `groebner(IG)` has two elements that are (i) a quintic in $\beta_{22} = $ `b(2)` and (ii) a sum of a linear function in $\beta_{11} = $ `b(1)` and a quartic in $\beta_{22} = $ `b(2)`. Thus it follows immediately that the stationary points of G can be found from solving a quintic (cf. Drton and Richardson, 2004, Thm. 2).

5. Dimensions and degrees of maximum likelihood ideals

5.1. Seemingly unrelated regressions

The algebraic approach can also be applied to more general models. Here we focus on SUR models $\mathbf{N}(\mathcal{C}_{\mathcal{R}})$ for which $(C_r \mid r \in R)$ consists of disjoint sets; in other models

Table 2
Dimension and degree of maximum likelihood ideals

$\mathcal{C}_{\mathcal{R}}$	$\mathbb{B}(\mathcal{C}_{\mathcal{R}})$	$\dim(I_G)$	$\text{degree}(I_G)$
$\{(1, 1), (2, 2)\}$	$\begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix}$	0	5
$\{(1, 1), (1, 2), (2, 3)\}$	$\begin{pmatrix} \beta_{11} & \beta_{12} & 0 \\ 0 & 0 & \beta_{23} \end{pmatrix}$	0	9
$\{(1, 1), (2, 2), (3, 3)\}$	$\begin{pmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{pmatrix}$	0	29
$\{(1, 1), (1, 2), (1, 3), (2, 4)\}$	$\begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} & 0 \\ 0 & 0 & 0 & \beta_{24} \end{pmatrix}$	1	4
$\{(1, 1), (1, 2), (2, 3), (2, 4)\}$	$\begin{pmatrix} \beta_{11} & \beta_{12} & 0 & 0 \\ 0 & 0 & \beta_{23} & \beta_{24} \end{pmatrix}$	1	8
$\{(1, 1), (2, 2), (3, 3), (4, 4)\}$	$\begin{pmatrix} \beta_{11} & 0 & 0 & 0 \\ 0 & \beta_{22} & 0 & 0 \\ 0 & 0 & \beta_{33} & 0 \\ 0 & 0 & 0 & \beta_{44} \end{pmatrix}$	1	32
$\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\}$	$\begin{pmatrix} \beta_{11} & 0 & 0 & 0 & 0 \\ 0 & \beta_{22} & 0 & 0 & 0 \\ 0 & 0 & \beta_{33} & 0 & 0 \\ 0 & 0 & 0 & \beta_{44} & 0 \\ 0 & 0 & 0 & 0 & \beta_{55} \end{pmatrix}$	2	80

inclusion relations among the sets C_r may be exploited (cf. Andersson and Perlman, 1994). More precisely, we consider models $\mathbf{N}(\mathcal{C}_{\mathcal{R}})$ in which $r_1 < r_2$, $r_1, r_2 \in R$, implies that $c_1 < c_2$ for all $c_1 \in C_{r_1}$ and $c_2 \in C_{r_2}$. Then $\mathbb{B}(\mathcal{C}_{\mathcal{R}})$ is a linear space of block-diagonal matrices.

Table 2 states the dimension and degree of the maximum likelihood ideal for seven examples including the one from Section 4. For the models with zero-dimensional maximum likelihood ideal I_G, we can find all stationary points of the likelihood function by computations analogous to the ones demonstrated in Table 1. The likelihood functions of these models may be multimodal and it would be interesting to find, for each model, reference data for which the cardinality of $V_{\mathbb{R}}(I_G)$ is large. For example, let $\mathcal{C}_{\mathcal{R}} = \{(1, 1), (2, 2)\}$ and choose

$$X = \begin{pmatrix} -0.65 & -0.80 & 1.34 & -1.03 & -1.08 \\ -0.04 & -1.18 & 1.98 & -2.42 & -3.75 \end{pmatrix},$$
$$Y = \begin{pmatrix} 0.14 & -0.73 & 1.40 & -2.29 & -3.30 \\ 0.52 & -1.93 & 3.02 & -6.67 & -9.94 \end{pmatrix}; \tag{19}$$

then the variety of the maximum likelihood ideal of $\mathbf{N}(\mathcal{C}_{\mathcal{R}})$ is purely real, i.e. $V_{\mathbb{R}}(I_G) = V_{\mathbb{C}}(I_G) = 5$. Fig. 1 shows a three-dimensional plot and a contour plot of the profile log-likelihood function for these observations. We conjecture that data with $V_{\mathbb{R}}(I_G) = V_{\mathbb{C}}(I_G)$ exist for all three models in Table 2 that have zero-dimensional maximum likelihood ideal. For models with maximum likelihood ideal of dimension one or higher, it is not clear
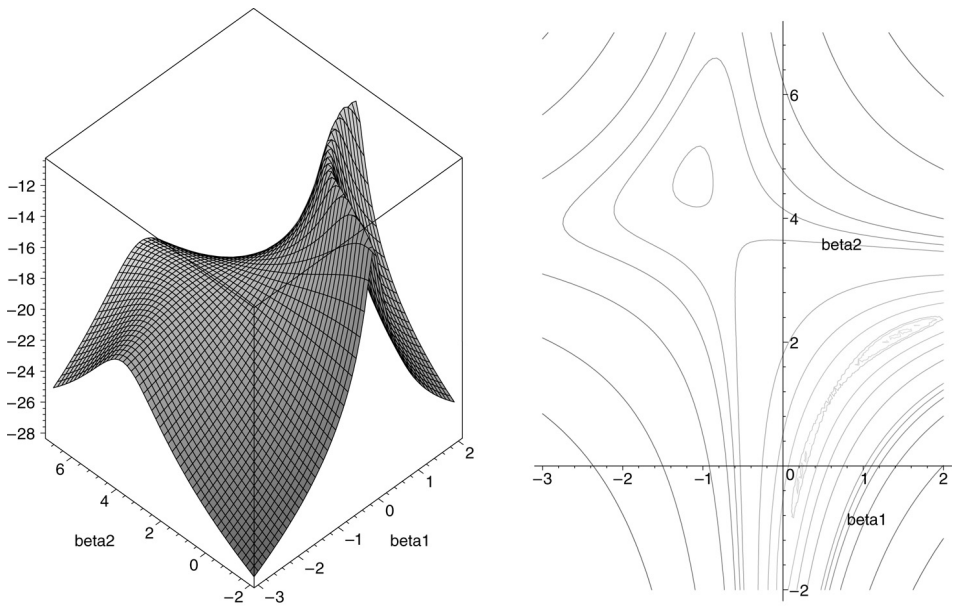
Fig. 1. Three-dimensional plot and contour plot of profile log-likelihood function.

whether $V_{\mathbb{R}}(I_G) = \infty$, i.e. a likelihood function with an infinite number of stationary points, can occur with non-zero probability.

5.2. Submodels of seemingly unrelated regressions

It is obvious that the algebraic approach developed in Section 3 immediately carries over to the submodels of SUR that are of interest in testing equality of regression coefficients. In the model $\mathbf{N}(\mathcal{C}_{\mathcal{R}})$ with $\mathcal{C}_{\mathcal{R}} = \{(1, 1), (2, 2)\}$, for example, we may be interested in testing whether $\beta_{11} = \beta_{22}$. If this is done using a likelihood ratio test, then the likelihood function of the submodel in which $\beta_{11} = \beta_{22}$ is imposed has to be maximized. More precisely, the submodel has the restricted parameter space

$$\{B \in \mathbb{B}(\mathcal{C}_{\mathcal{R}}) \mid \beta_{11} = \beta_{22}\} \times \mathbb{P}. \tag{20}$$

Table 3 lists similarly obtained submodels of the models in Table 2, for which the maximum likelihood ideal is zero-dimensional and the variety $V_{\mathbb{C}}(I_G)$ can be computed.

It should also be noted that submodels of SUR need not inherit unimodal likelihood functions from their parent model. For example, the bivariate SUR model $\mathbf{N}(\mathcal{C}_{\mathcal{R}})$ with $\mathcal{C}_{\mathcal{R}} = \{(1, 1), (2, 1), (2, 2)\}$ is monotone, i.e. the family $(C_r \mid r \in R)$ is totally ordered by inclusion, which guarantees that the likelihood function has precisely one stationary point corresponding to the global maximum (Andersson and Perlman, 1994; Drton et al., 2005). However, the submodel induced by the restriction $\beta_{11} = \beta_{21}$ can be re-expressed in the form of the model studied in Section 4 by means of the linear transformation that changes response Y_2 into $Y_2 - Y_1$. Hence, the submodel does not always have a unimodal likelihood function.

Table 3
Dimension and degree of maximum likelihood ideals of submodels

$\mathcal{C}_{\mathcal{R}}$	Subspace of $\mathbb{B}(\mathcal{C}_{\mathcal{R}})$	$\dim(I_G)$	$\mathrm{degree}(I_G)$
$\{(1, 1), (2, 2)\}$	$\begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{11} \end{pmatrix}$	0	3
$\{(1, 1), (1, 2), (2, 3)\}$	$\begin{pmatrix} \beta_{11} & \beta_{12} & 0 \\ 0 & 0 & \beta_{12} \end{pmatrix}$	0	7
$\{(1, 1), (2, 2), (3, 3)\}$	$\begin{pmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{11} & 0 \\ 0 & 0 & \beta_{33} \end{pmatrix}$	0	11
$\{(1, 1), (1, 2), (1, 3), (2, 4)\}$	$\begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} & 0 \\ 0 & 0 & 0 & \beta_{13} \end{pmatrix}$	0	11
$\{(1, 1), (1, 2), (2, 3), (2, 4)\}$	$\begin{pmatrix} \beta_{11} & \beta_{12} & 0 & 0 \\ 0 & 0 & \beta_{12} & \beta_{24} \end{pmatrix}$	0	23
$\{(1, 1), (2, 2), (3, 3), (4, 4)\}$	$\begin{pmatrix} \beta_{11} & 0 & 0 & 0 \\ 0 & \beta_{11} & 0 & 0 \\ 0 & 0 & \beta_{33} & 0 \\ 0 & 0 & 0 & \beta_{44} \end{pmatrix}$	0	63

6. Conclusion

The presented algebraic approach to maximum likelihood estimation in SUR permits us to compute all stationary points of the likelihood function if the maximum likelihood ideal is zero-dimensional. This is the case for three seemingly unrelated regressions models considered in this paper (cf. Table 2): (i) the previously studied model based on $\mathcal{C}_{\mathcal{R}}^{(1)} = \{(1, 1), (2, 2)\}$, (ii) the model with $\mathcal{C}_{\mathcal{R}}^{(2)} = \{(1, 1), (1, 2), (2, 3)\}$, and (iii) the model with $\mathcal{C}_{\mathcal{R}}^{(3)} = \{(1, 1), (2, 2), (3, 3)\}$. Additionally, interesting submodels of SUR may have a zero-dimensional maximum likelihood ideal (cf. Table 3). The computations in Singular that find all stationary points of the likelihood functions of the models with zero-dimensional maximum likelihood ideal are instantaneous for all but the model in Table 3 that has a maximum likelihood ideal of degree 63. Thus we advocate the use of Singular or similarly capable software in statistical data analysis.

In future work it would be interesting to find reference data sets leading to likelihood functions with a large number of stationary points. Moreover, the algebraic approach presented herein could be combined with regression approaches (e.g. Andersson and Perlman, 1994; Drton et al., 2005) in order to identify larger classes of SUR models for which all stationary points of the likelihood function can be computed. Finally, it could be explored whether methods for global minimization of polynomials (Parrilo and Sturmfels, 2003) can be used to find the global maximum of SUR likelihood functions.

Acknowledgements

References

Andersson, S.A., Madigan, D., Perlman, M.D., 2001. Alternative Markov properties for chain graphs. Scand. J. Statist. 28, 33–86.

Andersson, S.A., Perlman, M.D., 1994. Normal linear models with lattice conditional independence restrictions. In: Multivariate Analysis and its Applications. Inst. Math. Statist., vol. 24. Hayward, CA, pp. 97–110.

Cox, D., Little, J., O'Shea, D., 1997. Ideals, Varieties, and Algorithms, 2nd edition. Springer-Verlag, New York.

Cox, D., Little, J., O'Shea, D., 1998. Using Algebraic Geometry. Springer-Verlag, New York.

Drton, M., Andersson, S.A., Perlman, M.D., 2005. Conditional independence models for seemingly unrelated regressions with incomplete data. J. Multivariate Anal., to appear.

Drton, M., Richardson, T.S., 2004. Multimodality of the likelihood in the bivariate seemingly unrelated regressions model. Biometrika 91, 383–392.

Goldberger, A., 1991. A Course in Econometrics. Harvard University Press, Cambridge, MA.

Greuel, G.-M., Pfister, G., Schönemann, H., 2001. SINGULAR 2.0. A Computer Algebra System for Polynomial Computations, Centre for Computer Algebra, University of Kaiserslautern. http://www.singular.uni-kl.de.

Oberhofer, W., Kmenta, J., 1974. A general procedure for obtaining maximum likelihood estimates in generalized regression models. Econometrica 42, 579–590.

Parrilo, P.A., Sturmfels, B., 2003. Minimizing polynomial functions. In: Basu, S., Gonzalez-Vega, L. (Eds.), Algorithmic and Quantitative Real Algebraic Geometry (Piscataway, NJ, 2001). In: DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 60. Amer. Math. Soc., Providence, RI, pp. 83–99.

Richardson, T.S., Spirtes, P., 2002. Ancestral graph Markov models. Ann. Statist. 30, 962–1030.

Rochon, J., 1996a. Accounting for covariates observed post-randomization for discrete and continuous repeated measures data. J. Roy. Statist. Soc. Ser. B 58, 205–219.

Rochon, J., 1996b. Analyzing bivariate repeated measures for discrete and continuous outcome variables. Biometrics 52, 740–750.

Sturmfels, B., 2002. Solving Systems of Polynomial Equations. In: CBMS Regional Conference Series in Mathematics, vol. 97. Conference Board of the Mathematical Sciences, Washington, DC.

Telser, L.G., 1964. Iterative estimation of a set of linear regression equations. J. Amer. Statist. Assoc. 59, 845–862.

Verbyla, A.P., Venables, W.N., 1988. An extension of the growth curve model.. Biometrika 75, 129–138.

Zellner, A., 1962. An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. J. Amer. Statist. Assoc. 57, 348–368.