

Article

Estimation of Multilevel Simultaneous Equation Models through Genetic Algorithms

Rocío Hernández-Sanjaime * , Martín González  and Jose J. López-Espín 

Center of Operations Research, Miguel Hernández University, 03202 Elche, Spain;
martin.gonzaleze@umh.es (M.G.); jlopez@umh.es (J.J.L.-E.)

* Correspondence: rocio.hernandezs@umh.es

Received: 21 October 2020; Accepted: 18 November 2020; Published: 24 November 2020



Abstract: Problems in estimating simultaneous equation models when error terms are not intertemporally uncorrelated has motivated the introduction of a new multivariate model referred to as Multilevel Simultaneous Equation Model (MSEM). The maximum likelihood estimation of the parameters of an MSEM has been set forth. Because of the difficulties associated with the solution of the system of likelihood equations, the maximum likelihood estimator cannot be obtained through exhaustive search procedures. A hybrid metaheuristic that combines a genetic algorithm and an optimization method has been developed to overcome both technical and analytical limitations in the general case when the covariance structure is unknown. The behaviour of the hybrid metaheuristic has been discussed by varying different tuning parameters. A simulation study has been included to evaluate the adequacy of this estimator when error terms are not serially independent. Finally, the performance of this estimation approach has been compared with regard to other alternatives.

Keywords: multilevel simultaneous equation model; maximum likelihood estimation; genetic algorithms; optimization

1. Introduction

Simultaneous equation models (SEM) [1] and multilevel models [2] have been widely discussed in the statistical literature and many applications arise in econometrics [3,4], medicine [5,6] or social sciences [7,8]. Simultaneous equation models contemplate jointly dependent or endogenous variables in a system of regression equations whereas multilevel models allow dealing with hierarchical or clustered data. However, clustered data may be problematic in the simultaneous equation models framework yielding to misleading results.

Indeed, grouped correlated observations may be precisely one of the causes that accounts for unobserved heterogeneity in simultaneous equation models. Typically, these models assume serially independent data, but it is well-known that ignoring the intraclass correlation across observations induced by groupings can lead to significant bias in the parameter estimates [9]. Under the motivation of overcoming limitations of traditional statistical models, it is necessary to think of new methodologies.

Multilevel simultaneous equation models combining both simultaneity and hierarchically structured data emerge as a valuable solution in this sort of situation. However, the literature exploring the confluence of these two factors in a single statistical models is scarce and it is worth investigating. To the best of our knowledge, theory and practice of multilevel simultaneous equation modelling are basically confined to recursive systems. So far, the existing methodology considers multilevel models in which the endogeneity of some of the variables has been adjusted including additional equations that create a recursive simultaneous equation model. Applications of these types of systems are essentially found in studies analysing resources allocation in the educational system or factors influencing women fertility [10,11].

Alternatively, starting out from a different point of view, a novel modelling framework denominated Multilevel Simultaneous Equation Model (MSEM) has been first introduced in Hernández-Sanjaime et al. [12]. This proposed approach is not restricted to recursive systems and could be gainfully used for estimating statistical models when faced with endogenous variables and temporal correlated observations. Potential applications would comprehend extensions of simultaneous equations models to problems in which error terms are temporally correlated. For instance, economic problems with temporal data, social science studies with hierarchical data (e.g., students in schools, people in districts...) or longitudinal research such as clinical trials, as long as these applications also entail jointly dependent variables.

This latter approach introduces a matrix distribution with a double variance structure in a SEM in which the assumption of intertemporally uncorrelated error terms is violated. The incorporation of an among-row and an among-column covariance matrix pattern aims to solve the misspecification of the model when this circumstance occurs. The maximum likelihood estimation of the parameters of an MSEM has been already established theoretically and the adequacy of the model has been assessed empirically when the double covariance structure is known [12].

Previously, under the assumption of known variance-covariance matrices, solutions to parameter estimates have been explored using a general-purpose optimization solver that selects 2SLS estimates of the coefficient matrices as starting values. The present paper addresses the estimation of the parameters of a multilevel simultaneous equation model in the general case, that is, when the double covariance structure is unknown. Although wanting to keep the same line of work in the general case, we realised that using only an optimization procedure would not be sufficient. Because of the absence of an analytical solution of the system of likelihood equations and the increasing complexity of the search space, a heuristic is required. Namely, a genetic algorithm that creates multiple sets of random parameters is considered.

Genetic algorithms are metaheuristics commonly used to generate high-quality solutions to optimization problems as an alternative to exhaustive search procedures [13,14]. In particular, genetic algorithms have been applied for estimating simultaneous equation models [15] and in the problem of finding the best SEM from a set of variables [16]. Likewise, the use of other metaheuristics or hybridations of several methods has been examined in the SEM context. A unified shared-memory metaheuristic scheme and parallel versions of different metaheuristics (GRASP, genetic algorithms, scatter search and their combinations) for obtaining a SEM from a dataset of variables have been presented and compared in Almeida et al. [17].

In this paper, a hybrid metaheuristic is developed to solve the problem of estimating an MSEM. The metaheuristic consists of a standard genetic algorithm that includes a call to the *nlm* general-purpose optimization function comprised in the statistical software R [18]. The obtained solution is efficient and its computational cost is considerably lower than finding a solution through exhaustive search procedures.

The rest of the paper is organised as follows. Section 2 reviews the main characteristics of simultaneous equation models. In Section 3, the multilevel simultaneous equation models is defined and its estimation via the maximum likelihood method is established. Section 4 describes the basis of the hybrid metaheuristic proposed for obtaining the parameters of an MSEM. Section 5 discusses different options for model estimation and reports the simulation results. Finally, the main conclusions, methodology limitations and future research are outlined in Section 6.

2. Simultaneous Equation Models

In the general linear regression model, one deals with a situation in which a dependent variable is expressed as a function of a set of explanatory variables that are either non-stochastic or at least independent of the error terms. Frequently, in some problems this assumption cannot hold. Many times we are interested in a model in which the dependent variable in one equation can also determine some of the explanatory variables by entering another equation (or equations), that is, a model in which a

number of variables are mutually and simultaneously determined. It is these types of situations that are considered in the theory of simultaneous equation models.

The structural form of a simultaneous equation model for a system with m equations, m endogenous variables (which influence and are influenced by other variables) and k predetermined variables (which influence but are not influenced by the system) is

$$Y = YA + XB + E \tag{1}$$

where $Y = [y^1, \dots, y^m]$ is a $N \times m$ matrix of N observations of m endogenous variables, $X = [x^1, \dots, x^k]$ is a $N \times k$ matrix of N observations of k non-random predetermined variables which contains both exogenous and lagged endogenous variables, and $E = [e^1, \dots, e^m]$ is a $N \times m$ matrix of the structural disturbances of the system. The matrices A ($m \times m$) and B ($k \times m$) are the endogenous and exogenous unknown coefficient matrices, respectively (by convention, $a_{ii} = 0, i = 1, 2, \dots, m$). The rows of E , denoted e_t , have the properties

$$e_t^T \sim N(0, \Sigma) \quad E(e_t^T, e_{t'}^T) = \delta_{tt'} \Sigma \quad t, t' = 1, 2, \dots, N \tag{2}$$

$\delta_{tt'}$ being the Kronecker delta and Σ a positive definite matrix.

Thus, the error terms e_t . ($t = 1, \dots, N$) are assumed to be serially independent random vectors normally distributed with 0 mean vector and covariance matrix Σ . Additionally, we make the customary assumptions that error terms are uncorrelated with the predetermined variables of the system and there is no linear dependence among the predetermined variables

$$E(X^T E) = 0 \quad \text{and} \quad \text{rank}(X) = k \tag{3}$$

Finally, assuming that $(I - A)$ is non-singular, the endogenous variables can be uniquely determined in terms of the predetermined and random variables of the system and the reduced form of the model becomes

$$Y = X\Pi + V \quad \text{where} \quad \Pi = B(I - A)^{-1} \quad \text{and} \quad V = E(I - A)^{-1} \tag{4}$$

Before considering the estimation of the model, we shall examine the identification problem. If the order condition $m_i - 1 \leq k - k_i$ holds, where m_i and k_i are the number of endogenous and exogenous variables in the i -th equation ($i = 1, \dots, m$), an equation is identified. It is only in such case that parameters can be calculated and two-stage least squares (2SLS), indirect least squares (ILS), three-stage least squares (3SLS) or maximum likelihood (ML) are the most commonly estimation methods [19].

The preceding distributional assumptions about the structural disturbances involve that error terms may be contemporaneously correlated but are intertemporally uncorrelated. Nevertheless, the property that characterizes clustered data is the intraclass correlation and in the presence of that type of data structure, we cannot assert that the random terms of the system are intertemporally uncorrelated. Thus, we can hardly expect to obtain parameter estimates having desirable properties. In such case, the estimation techniques fail and it is necessary to establish alternative methods that provide at least consistent, unbiased and possibly efficient estimators.

3. Multilevel Simultaneous Equation Model

3.1. Definition of the Model

Consider a general simultaneous equation model specified as in (1) with observed data clustered into l independent groups

$$Y_j = Y_j A + X_j B + E_j \quad j = 1, \dots, l \quad \text{independent groups} \tag{5}$$

where Y_j, X_j denote the observations of endogenous and predetermined variables in group j and E_j the matrix of structural disturbances of the system in group j . Note that A, B are the same for all groups assuming that the observed data are generated by a single set of parameter values representing a unique structural model.

As it is well known, if groupings are ignored during the estimation process, one runs the risk of drawing erroneous statistical conclusions. Typically, multilevel models would be used to allow for data organized into groups, but these models are not designed to deal with endogenous variables. One option is to modify simultaneous equation model distributional assumptions so that error terms can be serially dependent, thereby taking into account data correlation within groups. This point and the fact that A and B are the same across groups will differentiate this proposal from previous research approaches when handling heterogeneity in simultaneous equation models framework.

The matrix normal distribution [20] incorporates an among-row and among-column covariance matrix structure that has motivated the development of a novel multivariate approach referred as to Multilevel Simultaneous Equation Model (MSEM) [12]. This separable variance-covariance patterned matrix appears to be adequate for the double objective targeted in this work and will be considered in the error terms distribution.

For model (5), imposing the condition that each group has the same number of units, n , we assume that error terms follow a matrix normal distribution

$$E_j \sim N_{n,m}(0, U, \Sigma) \quad j = 1, \dots, l \tag{6}$$

where $0 (n \times m)$ is the mean of the distribution and $U (n \times n)$ and $\Sigma (m \times m)$ are symmetric positive definite matrices specifying the covariance between units of the same group and the covariance between variables, i.e., the temporal and contemporaneous correlation, respectively (note that the condition of equal number of units in all groups let us suppose that the U covariance matrix is common across groups).

Applying some basic properties to express the values of the endogenous variables in the reduced form

$$Y_j = X_j B(I - A)^{-1} + E_j(I - A)^{-1} \quad \text{and} \quad W_j = E_j(I - A)^{-1} \tag{7}$$

we have that,

$$W_j \sim N(0, U, ((I - A)^{-1})^T \Sigma (I - A)^{-1}) \quad j = 1, \dots, l \tag{8}$$

Therefore, replacing W_j by the observable quantities $Y_j - X_j B(I - A)^{-1}$, the matrix normal density function is

$$f_j(W_j) = c^{-1} \exp \left[-\frac{1}{2} \text{tr} \left(U^{-1} (Y_j(I - A) - X_j B) \Sigma^{-1} (Y_j(I - A) - X_j B)^T \right) \right] \tag{9}$$

with $c = (2\pi)^{nm/2} |U|^{m/2} |((I - A)^{-1})^T \Sigma (I - A)^{-1}|^{n/2}$.

3.2. Maximum Likelihood Estimation

Given the above distributional assumptions, the log-likelihood function for a random sample of l independent and identically distributed (i.i.d.) groups is

$$L(A, B, U, \Sigma) = -\frac{nm}{2} \ln(2\pi) - \frac{ml}{2} \ln|U| - \frac{nl}{2} \ln|((I - A)^{-1})^T \Sigma (I - A)^{-1}| - \frac{1}{2} \sum_{j=1}^l \text{tr} \left(U^{-1} (Y_j(I - A) - X_j B) \Sigma^{-1} (Y_j(I - A) - X_j B)^T \right) \tag{10}$$

The problem is to maximize L with respect to the parameters A, B, U and Σ given the sample data $(X_j, Y_j) (j = 1, \dots, l)$ and a fixed number of groups l .

Applying matrix derivatives [21–23], we obtain the system of likelihood equations:

$$\frac{\partial L(\cdot)}{\partial U} = -mlU^{-1} + \frac{ml}{2}diag(U^{-1}) + \sum_{j=1}^l \left(U^{-1}(Y_j(I - A) - X_jB)\Sigma^{-1}(Y_j(I - A) - X_jB)^T U^{-1} \right) - \frac{1}{2} \sum_{j=1}^l diag \left(U^{-1}(Y_j(I - A) - X_jB)\Sigma^{-1}(Y_j(I - A) - X_jB)^T U^{-1} \right) = 0 \quad (11)$$

$$\frac{\partial L(\cdot)}{\partial \Sigma} = -nl\Sigma^{-1} + \frac{nl}{2}diag(\Sigma^{-1}) + \sum_{j=1}^l \left(\Sigma^{-1}(Y_j(I - A) - X_jB)^T U^{-1}(Y_j(I - A) - X_jB)\Sigma^{-1} \right) - \frac{1}{2} \sum_{j=1}^l diag \left(\Sigma^{-1}(Y_j(I - A) - X_jB)^T U^{-1}(Y_j(I - A) - X_jB)\Sigma^{-1} \right) = 0 \quad (12)$$

$$\frac{\partial L(\cdot)}{\partial B} = \sum_{j=1}^l (X_j^T U^{-1} Y_j)(I - A)\Sigma^{-1} - \sum_{j=1}^l (X_j^T U^{-1} X_j)B\Sigma^{-1} = 0 \quad (13)$$

$$\frac{\partial L(\cdot)}{\partial (I - A)} = nl((I - A)^{-1})^T - \sum_{j=1}^l (Y_j^T U^{-1} Y_j)(I - A)\Sigma^{-1} - Y_j^T U^{-1} X_j B \Sigma^{-1} = 0 \quad (14)$$

The maximum likelihood estimator (MLE) is defined as the global maximum of the (log)-likelihood function. The usual method to calculate this estimator involves solving the system of likelihood equations described above by setting each derivative equal to zero. Unfortunately, in this case the system has not a closed solution and it is not possible to isolate all the unknown parameters implicated.

4. A Genetic Algorithm for MSEM Estimation

The maximization of the log-likelihood function requires the resolution of a nonlinear system of equations that implies cumbersome calculations. In the absence of a closed analytical solution, the computational cost of completely exploring the space of solutions using exhaustive search procedures makes this option operationally unfeasible. As an alternative, the method here suggested for finding the MLE is to use a hybrid metaheuristic technique, which consists of an optimized genetic algorithm. This implementation is able to improve the solution provided by the genetic algorithm by computing the R general-purpose optimization function *nlm* from the stats package.

This section outlines the genetic algorithm programmed for estimating the multilevel simultaneous equation model (MSEM), which is described as follows. A population of chromosomes composed of multiple sets of parameters *A*, *B*, *U* and Σ is explored. Each chromosome determines an initial solution for the MSEM and represents a candidate for the MLE. The population is updated in the crossover and it can be enhanced by using an optimization solver in two different points of the algorithm. First, it can be randomly improved after the mutation function and secondly, once the genetic algorithm has concluded, a predefined set of chromosomes from among the fittest ones will be also optimized. The fitness of a chromosome is calculated evaluating the maximum likelihood function. The general scheme of a genetic algorithm is used (Algorithm 1) and its functions and parameters are described in detail hereunder.

4.1. Initialization and End Conditions

The initial population is randomly created and the population size (*PopSize*) is stated at the beginning. The generational process in the metaheuristic is repeated until it accomplishes the end condition. Namely, the algorithm stops when it reaches the maximum number of iterations (*MaxIter*), which will be studied for different predefined values.

4.2. Evaluating a Chromosome and Selecting the Best Ranking

The fitness value of each chromosome is calculated using Equation (10). For the selection of the fittest individuals, a benchmark set of chromosomes (*BenchSet*) composed of the individuals with higher fitness score is included. In each generation, a comparison of the evaluations of all chromosomes in the population is made and the benchmark set is updated removing those individuals with lower fitness value.

4.3. Crossover

In each iteration a proportion of the existing population is selected to breed a new generation and only a preset number of the individuals in the best ranking of the benchmark set will be chosen for reproduction (*RepSize*).

The individuals selected from the set of chromosomes to produce offspring are randomly paired in *CrossSize* couples, so that the same chromosome can be picked more than once and can be matched with multiple couples. Each pair of ascendent chromosomes (*ascendent1* and *ascendent2*) will produce a new solution or *descendent*. The number of chromosomes that are created in each generation (*CrossSize*) is also a prespecified parameter.

Different methods can be used to combine the ascendants. In this work, the new individual inherits each of the matrices of parameters A , B , U or Σ completely from one of the ascendants. Therefore, for each new chromosome, a binary code randomly decides if a matrix is inherited from *ascendent1* encoded with 1 or the *ascendent2* encoded with 0.

4.4. Mutation

In each iteration, a small probability of mutation denoted by P_{mut} is considered. The mutation will only affect the elements of the diagonal of the covariance matrices U or Σ . If a chromosome from the new offspring is randomly chosen for mutation, all elements in the diagonal of any of these matrices will be subjected to mutation jointly by changing their numerical values.

4.5. Chromosomes Improvement

After mutation, a percentage of the descendent chromosomes is randomly chosen to be improved with a probability called P_{imp} . An optimization solver is included in the algorithm so that the new chromosomes can increment their quality to survive. The solver uses the selected chromosome as initial solution for maximizing the likelihood function (10). If it finds a better set of parameters for the model being estimated, the descendent chromosome is updated. If not, the chromosome remains equal.

Finally, once the descendents have been subjected to mutation and improvement, they can become part of the benchmark set. Each descendent is evaluated and the benchmark set is reordered considering these newly chromosomes. If the fitness score of a descendent is higher than any of the individuals in the benchmark set, the descendent will be included in the benchmark set in the corresponding position determined by its fitness value and the chromosome on the bottom of the ranking will be eliminated.

4.6. Optimization

Once the genetic algorithm has finished, the programme includes again a call to the optimization solver. In this step, a fixed number of the best ranked chromosomes (*OptSize*) is selected to be optimized. As described before, the chosen chromosomes are used as initial solutions by the solver and the fitness values of the resulting optimized chromosomes are calculated. Lastly, these scores are compared and the final solution shown by the metaheuristic corresponds to the parameters stored in the fittest individual. This step guarantees the search of a better final solution starting from a valid good one.

Algorithm 1: Parameterized schema for the hybrid metaheuristic

```

Initialize(ParamIni) → S_ini
ComputeFitness(S_ini, ParamIni)
Select(S_ini, ParamSelIni) → S_ref
while not EndCondition(S_ref, ParamEndCon) do
    Select(S_ref, ParamSel) → S_sel
    Combine(S_sel, ParamCom) → S_com
    Mutate(S_com, ParamMut) → S_mut
    Improve(S_com, S_mut, ParamImp) → S_imp_com, S_imp_mut
    ComputeFitness(S_com, S_mut, S_imp_com, S_imp_mut, ParamCom)
    Include(S_com, S_mut, S_imp_com, S_imp_mut, S_ref, ParamInc) → S_ref
end while
Improve(S_ref, ParamImpPost) → S_opt
Select(S_opt, ParamSelBest)
    
```

5. Experiment Results

Experiments have been executed in a DELL PowerEdge R730 node with 2 Intel(R) Xeon(R) CPU E5-2650 v3 @ 2.30 GHz, with 20 cores (40 threads) at 2.4 GHz and 25 MB SmartCache memory. Tests were carried out in C code importing the *nlm* optimization function from statistical software R (GNU R version 3.5.2).

In the following simulations, some parameters have been fixed to predefined values. The initial population, *PopSize*, is integrated by 300 chromosomes and in each generation, a benchmark set, *BenchSet*, of 100 chromosomes is considered. The number of individuals selected from the benchmark set to be used in the crossover is *RepSize* = 20 and in each generation, *CrossSize* = 25 new chromosomes are created with a probability of mutation of 25 percent.

The rest of the parameters of the metaheuristic are studied experimentally. Table 1 reports the fitness provided by the algorithm for different values of the tuning parameters when estimating a medium size MSEM problem. Specifically, *Alg.Prev* denotes the fitness returned by the metaheuristic before entering in the optimization step 4.6, while *Alg.End* designates the fitness of a solution after the optimization is accomplished, if applicable. The cost of the algorithm measured in seconds is also presented. Likewise, *Prev* express the time spent by the procedure before entering in the optimization step and *Post* the time after this step is fulfilled.

Table 1. Average fitness and run time (over 10 simulation runs) for different parameters configuration of the hybrid metaheuristic.

Mode			Fitness		Time	
<i>MaxIter</i>	<i>OptSize</i>	<i>P_{imp}</i>	<i>Alg.Prev</i>	<i>Alg.End</i>	<i>Prev</i>	<i>Post</i>
10	0	0	-16,726.56	-16,726.56	0.32	0.32
50	0	0	-12,005.68	-12,005.68	0.90	0.90
100	0	0	-12,028.44	-12,028.44	1.60	1.60
10	10	0	-17,927.11	-10,568.38	0.32	1751.52
50	10	0	-12,098.04	-10,346.21	0.90	1732.36
100	10	0	-11,753.68	-10,319.02	1.60	1739.39
10	0	10	-9944.31	-9944.31	3663.19	3663.19
50	0	10	-8949.24	-8949.24	17,441.91	17,441.91
100	0	10	-8488.46	-8488.46	33,893.78	33,893.78
10	10	5	-10,364.71	-9992.68	2024.01	3776.39
50	10	5	-8814.78	-8756.06	8983.55	10,519.69
100	10	5	-8953.06	-8838.36	17,051.52	18,454.82

From the results, the rest of the parameters of the hybrid metaheuristic are henceforth established as follows. The probability of improvement, P_{imp} , is fixed to 5% in each iteration. The number of chromosomes to be optimized at the end of the genetic algorithm from among those ranked in the top of the benchmark set, $OptSize$, is fixed to 10. Note that the combination $P_{imp} = 0$ and $OptSize = 0$ corresponds to a standard genetic algorithm. The generational process is repeated until the end condition is reached, $MaxIter = 10$. Figure 1 summarises the iterative process and the parameters values eventually selected in each step of the hybrid metaheuristic which will be used to obtain the MLE.

The analysis in Table 1 indicates that including the call to the nlm function after mutation ($P_{imp} \neq 0$) or at the end of the process ($OptSize \neq 0$) outperforms in either case the standard genetic algorithm fitness value, as can be seen in column $Alg.End$. The results also suggest that inserting the optimization after the mutation function ($P_{imp} = 10$ and $OptSize = 0$) has a higher positive impact than just optimizing the fittest individuals at the end of the metaheuristic ($P_{imp} = 0$ and $OptSize = 10$). In fact, this last option does not provide a substantial improvement with regard to a basic genetic algorithm. Moreover, the simulation study illustrates that there is no significant difference between implementing the solver only after mutation and applying both optimizations ($P_{imp} = 5$ and $OptSize = 10$). However, we consider this last configuration a more complete choice since it would allow for at least a minimal enhancement involving a negligible cost in time. On the other hand, for any of the versions of the optimized genetic algorithm, it appears that the different values for the maximum number of iterations provide similar fitness scores. In contrast, the computer time speeds up as the number of iterations increases and leads us to choose $MaxIter = 10$ as end condition.

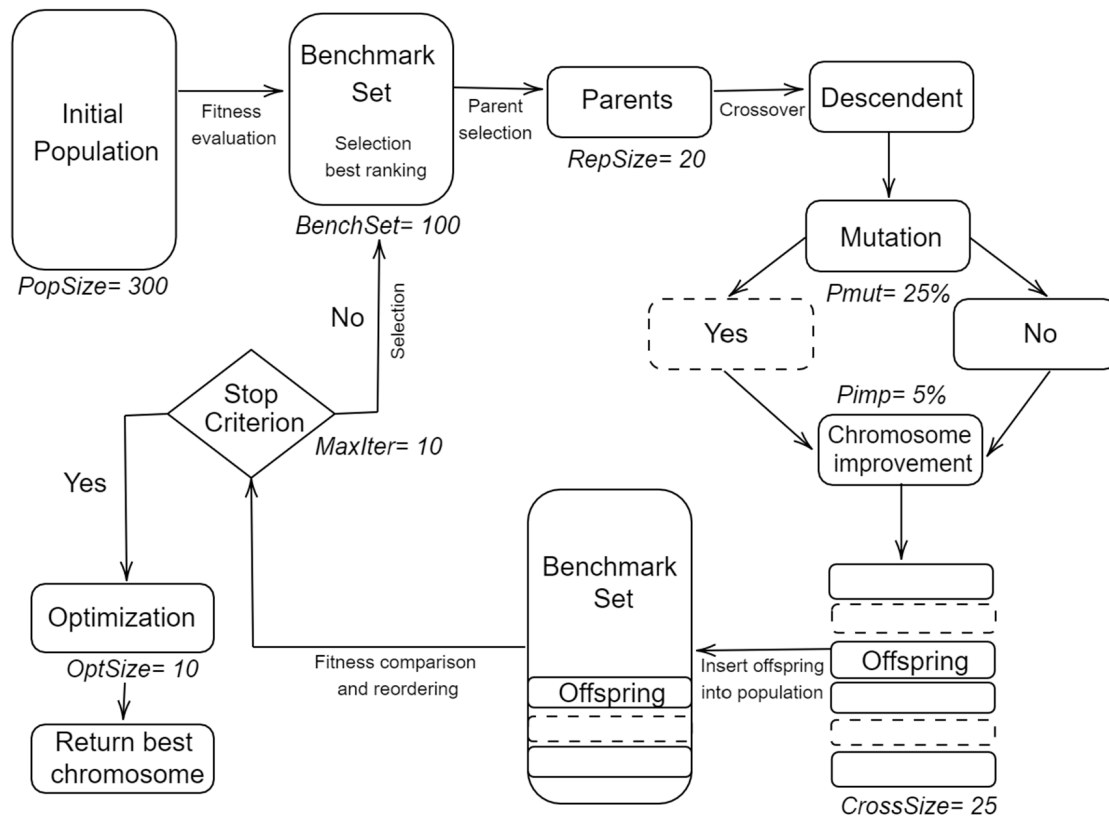


Figure 1. Hybrid genetic algorithm scheme.

Once parameters have been established according to evidences in Table 1, Table 2 shows the experiment outcomes for the hybrid metaheuristic developed in this work. In all experiments, the MSEM was codified so as to satisfy the order condition, thereby ensuring the estimation of the model. Henceforth, we shall assume that the MSEM is identified. The values of the chromosomes

which constitute the entry of the algorithm for the coefficient matrices A and B are, in half of the cases, 2SLS estimates of A and B with a variability from 0% to 0.75%. In the other half, A and B initial entries are selected randomly from the solution space. Although 2SLS method is not designed to deal with temporally correlated observations, 2SLS estimates offer a convenient starting solution for the search algorithm while the variability allows to expand the exploring in the solution space. In this way, we seek to reduce the algorithm dependence on good initial values and speed up the convergence. The starting values for U and Σ are generated randomly in such a way that symmetric positive definite matrices are guaranteed.

Finally, the performance of the hybrid metaheuristic has been examined regarding other estimation methods. The results are compared in terms of fitness, time and accuracy with two-stage least squares (2SLS) and a standard genetic algorithm (GA) with a maximum number of iterations of 10,000 which does not include a call to the nlm optimization function. The former method is the one commonly used in simultaneous equation models, which do not allow for intertemporally correlated error terms. Thus, using 2SLS in the MSEM case would involve an estimation bias. It is worth investigating the evolution of 2SLS estimates as the temporal correlation becomes significant. On the other hand, genetic algorithms are customary heuristics widely used in optimization and search problems when exact solutions are necessarily computationally expensive. Heuristics can provide an approximate solution individually or be used as a good baseline to be supplemented with an optimization procedure. This last case is the one considered in the hybrid metaheuristic. Hence, it is interesting to study whether or not the proposed optimized genetic algorithm produces better estimates than a standard genetic algorithm individually (even if a large number of iterations are permitted in the GA when used individually).

Three different values for endogenous and exogenous variables have been considered. Whatever the MSEM size, the number of groups is $l = 5$ and the number of observations in each group is $n = 30$. In addition, the covariance matrix U has been adjusted for five different ranges of values by dividing the generation interval by $\lambda = 100, 10, 1, 0.1$ and 0.01 . Note that dividing the generation interval by $\lambda = 1$ returns the original range for the entries of U . For each model configuration of parameters size ($m \ k \ l \ n$) and λ , we have simulated five different models and each of them has been estimated in two different ways using whether the hybrid version or the simple genetic algorithm mentioned above. The exogenous variables have been simulated following a matrix normal distribution. The endogenous variables have been calculated using Equation (7), with the error disturbances E_j generated by using the property stated in Arnold [20]:

If $Z_j = (z_{st})$ denotes an $n \times m$ random matrix with $z_{st} \ (s = 1, \dots, n; t = 1, \dots, m)$ i.i.d. $N(0, 1)$, then

$$E_j = U^{1/2} Z_j \Sigma^{1/2} \sim N_{n,m}(0, U, \Sigma) \quad j = 1, \dots, l$$

On the basis of the experimental results, the fitness value of the likelihood function obtained either by the hybrid metaheuristic (F_{HM}) or by the standard genetic algorithm (F_{GA}) always outperforms 2SLS fitness score (F_{2SLS}). In addition, the fitness value of the likelihood function obtained by combining the metaheuristic with an optimization procedure clearly enhances the likelihood value achieved by using only the genetic algorithm. Therefore, the solving method proposed in the hybrid version improves the fitness over any of the other alternative procedures.

To assess the accuracy of the different estimation methods, the Frobenius norm between observed data Y and fitted values \hat{Y} is calculated. Interestingly, 2SLS shows better results for small values of U , but as the range of values of U augments, the hybrid metaheuristic improves 2SLS whatever the MSEM size is. Nevertheless, this finding does not occur when using a standard genetic algorithm. In this case, 2SLS always produces better estimates than the genetic algorithm, with just one exception (size 15 20 5 30 and $\lambda = 0.1$).

Table 2. Average results for fitness, run time and accuracy (over 5 simulation runs) of different estimation methods.

Size	λ	F_{HM}	F_{GA}	F_{2SLS}	S.prev	S.post	S_{GA}	$\ Y - \hat{Y}_{HM}\ $	$\ Y - \hat{Y}_{GA}\ $	$\ Y - \hat{Y}_{2SLS}\ $
8 12 5 30	100	-3163.36	-3635.03	-411,983.91	434.99	922.77	57.25	537.39	116.48	89.19
	10	-3811.79	-4440.59	-57,824.25	584.70	1087.31	59.51	559.08	302.24	280.04
	1	-4094.60	-5748.64	-34,293.66	504.84	976.59	59.57	1123.97	948.76	899.87
	0.1	-7248.25	-10,943.99	-48,767,714.85	522.52	1030.72	59.67	2264.33	2658.97	2596.77
	0.01	-10,069.02	-61,501.99	-881,112.47	507.95	1008.34	59.51	7696.10	8519.65	8368.07
15 20 5 30	100	-7341.87	-8668.81	-141,301.20	1894.89	3680.11	142.04	1270.48	408.67	186.20
	10	-8411.78	-10,104.22	-7,104,505.59	1781.75	3547.47	141.76	1464.06	841.10	608.62
	1	-11,014.00	-12,694.56	-1,283,916.42	2310.64	4140.55	136.27	2563.91	2433.26	2076.20
	0.1	-12,097.40	-22,999.82	-327,573.88	2188.60	3900.75	127.63	5406.90	6179.77	6306.63
	0.01	-47455.24	-170,551.28	-21,252,787.68	1346.84	3189.38	130.23	14,815.85	20,587.63	18,881.66
22 28 5 30	100	-14,003.91	-16,217.86	-3,125,727.39	3851.21	8308.45	270.85	3922.10	2720.13	344.21
	10	-14,556.42	-17,178.54	-2,558,133.93	5222.54	9589.12	270.27	2274.98	2417.81	917.13
	1	-20,391.21	-28,325.12	-6,585,294.85	5454.31	9287.51	271.69	4343.76	5204.45	3823.36
	0.1	-22,702.70	-33,346.31	-18,947,400.33	3153.00	7526.67	270.77	8860.07	10,557.65	9396.87
	0.01	-48,689.37	-327,171.61	-15,177,293.12	5968.90	10,372.43	271.31	31,943.36	42,114.08	36,997.67

If one compares the norm of both algorithms, for small and medium MSEM problems (i.e., sizes 8 12 5 30 and 15 20 5 30), the hybrid metaheuristic performs better than the standard genetic algorithm for large values of U , in particular, $\lambda = 0.1$ and 0.01 . However, for a big MSEM size i.e., size 22 28 5 30), our algorithm becomes a better estimation method than a standard genetic algorithm even for small values of U . This comparison also illustrates that whenever the hybrid technique improves 2SLS estimates, it also enhances the standard genetic algorithm. Furthermore, it is worth highlighting that when the proposed approach is presented as the best estimation option, an end condition of $MaxIter = 10$ combined with a relatively small percentage of optimization is enough to outperform a standard genetic algorithm with reasonably large number of iterations. This is remarkable since it reinforces the choice of $MaxIter$ tuning parameter from Table 1 and will contribute to alleviate the computational cost.

In summary, these experiments lead to an important conclusion. Overall, 2SLS is a more appropriate estimation method when the U values are small, that is to say, when the error terms are serially uncorrelated (or at least the serial dependence is not very strong). However, as we had postulated, our simulation experiments point out that this technique does not provide satisfactory results when this assumption is violated. As the values of U become greater, the hybrid metaheuristic arises as a preferable option to 2SLS. In that case, our algorithm works best even for a modest sample size and reasonably large models.

Expectedly, as the size of the problem increases the execution time rises and the algorithm can be computationally demanding. The cost of estimating a model is the sum of applying a genetic algorithm and an optimization method. In particular, the last optimization step described in Section 4 originates an evident time impact, as one can compare in the tables above. The time the algorithm needs before entering in the final optimization (S_{prev}) and the total running time once the whole procedure has finished (S_{post}) measured in seconds are displayed. In contrast, the time spent by the standard genetic algorithm (S_{GA}) is also indicated. The hybrid metaheuristic shows the highest execution times but it provides the best outcomes. In addition, the time that the algorithm needs depends heavily on the numerical procedure implemented in nlm for optimizing the likelihood of the system and on the complexity of the fitness function.

6. Conclusions and Future Work

This paper proposes a metaheuristic strategy to estimate a multilevel simultaneous equation model (MSEM) when the double covariance structure is unknown. A genetic algorithm is combined with an optimization procedure in order to obtain the maximum likelihood estimator. Different parameters have been experimentally tuned to reduce the complexity of the search space to a more manageable one. However, further simulation studies are necessary to better inspect the space of parameters and avoid falling in local optima.

The simulation experiments suggest that this hybrid metaheuristic produces better outcomes than other estimation methods such as 2SLS when error terms are not intertemporally uncorrelated. According to the results, the algorithm provides satisfactory solutions when the system of likelihood equations cannot be analytically solved and standard optimal procedures are not feasible to implement. Although the hybrid metaheuristic leads to a more efficient inference, some aspects require further analysis. Moreover, as future work extensions to non-normal error distributional forms (e.g., matrix Student's T distribution) need to be investigated.

Because of the computational burden, the simulation results reported above are limited. The execution time grows as the size of the model increases and a parallel version of the algorithm may be a preferable option. The development of a shared memory version may boost the efficiency of processor usage reducing response times in the solution of the problem. Finally, other appealing metaheuristic techniques (Scatter Search, GRASP...) need to be examined and alternative optimization solvers must be explored.

Author Contributions: Conceptualization and Methodology, R.H.-S. and J.J.L.-E.; Software, M.G.; Writing—original draft preparation and Editing, R.H.-S.; Writing—review and Supervision, J.J.L.-E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministerio de Economía y Competitividad of Spain under Grant TIN2016-8056-R and a predoctoral contract from the Generalitat Valenciana and the European Social Fund to R.H.-S. under Grant ACIF/2018/219.

Acknowledgments: The authors gratefully acknowledge the computer resources and assistance provided by the Scientific Computing and Parallel Programming Group of the University of Murcia for the simulation study. The authors would like to thank the anonymous reviewers for their constructive comments and useful suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Hausman, J.A. Specification and estimation of simultaneous equation models. *Handb. Econom.* **1983**, *1*, 391–448.
- Goldstein, H. *Multilevel Statistical Models*; John Wiley & Sons: Chichester, UK, 2011; Volume 922.
- Klein, L.R. Economic fluctuations in the United States, 1921–1941. *Econ. J.* **1950**, *61*, 387–389.
- Dornbusch, R.; Fischer, S. *Macroeconomics*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1984.
- King, T.M. Using simultaneous equation modeling for defining complex phenotypes. *BMC Genet. Biomed. Cent.* **2003**, *4*, S10. [[CrossRef](#)]
- Simone, I.L.; Ceccarelli, A.; Tortorella, C.; Bellacosa, A.; Pellegrini, F.; Plasmati, I.; De Caro, M.F.; Lopez, M.; Girolamo, F.; Livrea, P. Influence of Interferon beta treatment on quality of life in multiple sclerosis patients. *Health Qual. Life Outcomes* **2006**, *4*, 96. [[CrossRef](#)] [[PubMed](#)]
- Ressler, R.W.; Waters, M.S. Female earnings and the divorce rate: A simultaneous equations model. *Appl. Econ.* **2000**, *32*, 1889–1898. [[CrossRef](#)]
- Graham-Bermann, S.A.; Miller-Graff, L. Community-based intervention for women exposed to intimate partner violence: A randomized control trial. *J. Fam. Psychol.* **2015**, *29*, 537. [[CrossRef](#)] [[PubMed](#)]
- Jedidi, K.; Jagpal, H.S.; DeSarbo, W.S. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Mark. Sci.* **1997**, *16*, 39–59. [[CrossRef](#)]
- Steele, F.; Vignoles, A.; Jenkins, A. The effect of school resources on pupil attainment: A multilevel simultaneous equation modelling approach. *J. R. Stat. Soc. Ser. A* **2007**, *170*, 801–824. [[CrossRef](#)]
- Steele, F. Selection effects of source of contraceptive supply in an analysis of discontinuation of contraception: Multilevel modelling when random effects are correlated with an explanatory variable. *J. R. Stat. Soc. Ser. A* **2003**, *166*, 407–423. [[CrossRef](#)]
- Hernández-Sanjaime, R.; González, M.; López-Espín, J.J. Multilevel simultaneous equation model: A novel specification and estimation approach. *J. Comput. Appl. Math.* **2020**, *366*, 112378. [[CrossRef](#)]
- Mitchell, M. *An Introduction to Genetic Algorithms*; MIT Press: Cambridge, MA, USA, 1998.

14. Goldberg, D.E.; Holland, J.H. Genetic algorithms and machine learning. *Mach. Learn.* **1988**, *3*, 95–99. [[CrossRef](#)]
15. López, J.J.; Giménez, D. Genetic algorithms for simultaneous equation models. *International Symposium on Distributed Computing and Artificial Intelligence 2008 (DCAI 2008)*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 215–224.
16. López-Espín, J.J.; Giménez, D. Obtaining simultaneous equation models from a set of variables through genetic algorithms. *Procedia Comput. Sci.* **2010**, *1*, 427–435. [[CrossRef](#)]
17. Almeida, F.; Giménez, D.; Lopez-Espin, J.J. Obtaining Simultaneous Equation Models through a unified shared-memory scheme of metaheuristics. In *Proceedings of the 2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum, Anchorage, AK, USA, 16–20 May 2011*; pp. 1981–1988.
18. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2012; ISBN 3-900051-07-0.
19. Dhrymes, P.J. *Econometrics: Statistical Foundations and Applications*; Springer: New York, NY, USA, 1974.
20. Arnold, S.F. *The Theory of Linear Models and Multivariate Analysis*; Wiley: New York, NY, USA, 1981.
21. Harville, D.A. *Matrix Algebra from a Statistician's Perspective*; Springer: New York, NY, USA, 1997; Volume 1.
22. Dwyer, P.S. Some applications of matrix derivatives in multivariate analysis. *J. Am. Stat. Assoc.* **1967**, *62*, 607–625. [[CrossRef](#)]
23. Petersen, K.B.; Pedersen, M.S. The matrix cookbook. *Tech. Univ. Den.* **2008**, *7*, 510.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).