Journal of Computational and Applied Mathematics 255 (2014) 605–615

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Heuristics and metaheuristics for accelerating the computation of simultaneous equations models through a Steiner tree

Jose J. López-Espín [a,*], Antonio M. Vidal [b], Domingo Giménez [c]

[a] Centro de Investigación Operativa, Universidad Miguel Hernández, 03202 Elche, Spain
[b] Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, 46022 Valencia, Spain
[c] Departamento de Informática y Sistemas, Universidad de Murcia, 30071 Murcia, Spain

ARTICLE INFO

Article history:
Received 26 July 2012
Received in revised form 25 November 2012

Keywords:
Simultaneous equations models
Econometrics
QR-decomposition
Genetic algorithms
Steiner tree

ABSTRACT

Simultaneous equations models can be solved with a variety of algorithms. Some methods use the QR-decomposition of several matrices associated to the equations in the system. To accelerate the computation of those QR-decompositions and consequently the solution of simultaneous equations models, the QR-decomposition of a matrix associated to an equation can be obtained from that of another matrix associated to another equation which contains the variables of the first equation. A Steiner tree can be used, with nodes representing the equations in the model and with edges whose associated weight is the cost of computing the QR-decomposition of an equation from that of another equation. The Steiner tree of the graph associated to the simultaneous equations model gives the order of computation of the QR-decompositions of lowest computational cost. But the number of nodes in the graph is very large, and exact methods to obtain the Steiner tree are not applicable. In this paper, the application of heuristics and metaheuristics to approach the Steiner tree of the graph associated to a simultaneous equations model is considered. A heuristic and a genetic algorithm are presented and analyzed. The quality of the tree obtained and its usability in an algorithm to solve simultaneous equations models efficiently is experimentally studied.

1. Introduction

Simultaneous equations models (SEM) is a statistical technique used to represent the interdependence of some variables. When a system can be represented with a simultaneous equations model, the values of some endogenous variables are estimated from those of some exogenous variables. The equations in the system are formed by linear combinations of the exogenous and endogenous variables. Many applications can be modeled with simultaneous equations models. These have traditionally been used in econometrics [1,2] and in economic analysis [3], but today they are used in a wide variety of fields such as public health [4], psychology [5], or sociology studies (the study of the divorce rate [6] and the demand management policy at an airport [7]), and they can be useful in problems where there is simultaneous influence between variables, such as coevolutionary games [8].

This work is a continuation of [9], where efficient algorithms for the Two-Stage Least Squares (2SLS) method using QR-decomposition are developed and studied, and the common columns of matrices associated to the equations in the system

* Corresponding author. Tel.: +34 966658725.
E-mail addresses: jlopez@umh.es, josejmat@hotmail.com (J.J. López-Espín), avidal@dsic.upv.es (A.M. Vidal), domingo@um.es (D. Giménez).

0377-0427/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cam.2013.06.013

are exploited. The QR-decomposition of a matrix associated to an equation can be obtained from that of another matrix associated to another equation which contains common variables with the first by updating the QR-decomposition. This leads to a reduction in the execution time of the 2SLS method, although this was not considered in the previous work.

A similar problem was tackled in [10], where the decomposition of a set of matrices with common columns was investigated. The authors consider a weighted directed graph where each matrix is represented by a node, and the associated weight of the edges is the cost of computing the QR-decomposition of a matrix from that of another matrix. The Minimum Spanning Tree (MST) provides the order of computation for the QR-decompositions which gives the lowest execution time. A heuristics is developed in which artificial nodes are included in the graph, so the computation of the QR-decomposition of the artificial nodes helps to reduce the cost of the QR-decompositions.

These ideas are adapted and applied in this work to accelerate the solution of SEM by 2SLS with QR-decomposition. Here, the Steiner Tree (ST) is used to provide the order for the QR-decompositions which gives the lowest execution time. The nodes represent the equations in the simultaneous equations model. There are edges from a node to other nodes which correspond to equations whose variables are included in the set of variables in the equation associated to the source node, and the weight associated to an edge is the cost of computing the QR-decomposition of the matrix associated to the equation corresponding to the target node from that of the matrix associated to the equation of the source node.

The number of nodes in the graph associated to the model is very large, and exact methods to obtain the Steiner tree are not applicable. In a similar way to [10], we have developed a new heuristic specially conceived for SEM. We have also developed a genetic algorithm with a double purpose: as comparison test and to accelerate the heuristic, where possible. The new heuristic and metaheuristic are used to approximate the ST in a short execution time.

This work considers SEM with a structure of clusters of equations with subsets of equations with a large number of common variables. This type of SEM appears when a big model is built up from a set of lower dimension models. For example, SEM for the economic variables in Europe can be obtained from the models of the different countries by adding some joining equations. Thus, the global model has a large number of variables, but in the equations of a particular country most of the variables are local variables. The world model (managed by the LINK project at the University of Toronto [11]) includes the Spanish model (managed by the CEPREDE or Centro de Predicción Económica [12] and by the Lawrence R. Klein Institute [13] at the University Autónoma de Madrid) and a set of equations to connect the Spanish variables with the global ones. The reduction in the execution time achieved with the method proposed in this paper is more apparent for this type of large structured systems.

The variables used to obtain these SEM continue to provide new data. Thus, the SEM parameters can be recalculated to obtain more accurate estimations. In the Spanish model, for example, the crude oil price (per barrel) is an important variable which provides new data every day and affects many equations. Recalculating the SEM parameters enhances accuracy.

The aim of this work is to reduce the total cost of solving a SEM by using a ST to solve the equations. However, the cost of obtaining the ST often exceeds the reduction obtained. The ST depends on the structure of the SEM and not on the values of the data variables, so it can be reused when the parameters of the SEM are recalculated with new data. Thus, the algorithm proposed in this work is very useful in SEM whose parameters are recalculated many times.

The rest of the paper is organized as follows. Section 2 reviews the background of SEM and 2SLS method based on QR-decomposition. In Section 3, the weighted graph associated to a SEM and the ST problem generated are presented. In Section 4 the heuristic and metaheuristic methods to approach the ST are proposed and studied. In Section 5 an algorithm to solve a SEM through the ST is presented. Section 6 describes the experiments. Finally, in Section 7 conclusions and future works are outlined.

2. Background of SEM and the 2SLS algorithm

This section summarizes the basics of SEM and the 2SLS algorithm and introduces the notation used in the rest of the paper. For a fuller description, see [9].

Consider N interdependent (endogenous) variables which depend on K independent (exogenous) variables. Suppose that each endogenous variable can be expressed as a linear combination of the other endogenous variables, the exogenous variables and the white noise that represents stochastic interference. Then, a SEM is [1]:

$$\begin{aligned} y_1 &= B_{1,2}y_2 + B_{1,3}y_3 + \cdots + B_{1,N}y_N + \Gamma_{1,1}x_1 + \cdots + \Gamma_{1,K}x_K + u_1 \\ y_2 &= B_{2,1}y_1 + B_{2,3}y_3 + \cdots + B_{2,N}y_N + \Gamma_{2,1}x_1 + \cdots + \Gamma_{2,K}x_K + u_2 \\ &\cdots \\ y_N &= B_{N,1}y_1 + \cdots + B_{N,N-1}y_{N-1} + \Gamma_{N,1}x_1 + \cdots + \Gamma_{N,K}x_K + u_N \end{aligned} \tag{1}$$

where $B \in \mathbb{R}^{N\times N}$ and $\Gamma \in \mathbb{R}^{N\times K}$ are matrices of coefficients, and x, y and u are, respectively, exogenous, endogenous and random variables, which are vectors of dimension d, where d is the sample size. Some coefficients of $B_{i,j}$ and $\Gamma_{k,r}$ are zero, and are known a priori. The equation can be represented in matrix form as:

$$YB^T + X\Gamma^T + u = 0 \tag{2}$$

where $Y = (y_1 \dots y_N)\,, X = (x_1 \dots x_K)\,, u = (u_1 \dots u_N)$.