Solution of Simultaneous Equations Models by High Performance Methods

José Juan López Espín
Universidad Miguel Hernández (Elche, Spain)
Domingo Giménez Cánovas
Universidad de Murcia (Murcia, Spain)

This work has been funded in part by Spanish MCYT and FEDER under Grant TIC2003-08238-C02-02
Contents

- Introduction
- Simultaneous equations models
- The identification problem
- Methods and algorithms
  - OLS
  - ILS
  - 2SLS
- Experimental results
- Conclusions and future works
Introduction

- The solution of a S.E.M. in high performance parallel systems is studied.
- The methods analyzed here are ILS and 2SLS.
- Parallel algorithms for shared and distributed memory have been developed.
- The methods have been analyzed in different parallel systems.
Simultaneous Equations Models

The scheme of a system with \( M \) equations, \( M \) endogenous variables and \( k \) predetermined variables is (structural form)

\[
Y_{1t} = \beta_{12} Y_{2t} + \beta_{13} Y_{3t} + \ldots + \beta_{1M} Y_{Mt} + \gamma_{11} X_{1t} + \ldots + \gamma_{1k} X_{kt} + u_{1t}
\]
\[
Y_{2t} = \beta_{21} Y_{1t} + \beta_{23} Y_{3t} + \ldots + \beta_{2M} Y_{Mt} + \gamma_{21} X_{1t} + \ldots + \gamma_{2k} X_{kt} + u_{2t}
\]
\[
\vdots
\]
\[
Y_{Mt} = \beta_{M1} Y_{1t} + \beta_{M2} Y_{2t} + \beta_{M3} Y_{3t} + \ldots + \beta_{MM-1} Y_{M-1t} + \gamma_{M1} X_{1t} + \ldots + \gamma_{Mk} X_{kt} + u_{Mt}
\]

These equations can be represented in matrix form

\[
BY_t + GX_t + u_t = 0
\]
Simultaneous Equations Models

The structural form can be expressed in reduced form

\[ Y_t = \mathbf{P} \mathbf{X}_t + \nu_t \]

with \( \mathbf{P} = - B^{-1} \mathbf{G} \) and \( \nu_t = - B^{-1} \mathbf{u}_t \)

\[ Y_{1t} = p_{11} X_{1t} + \ldots + p_{1k} X_{kt} + \nu_{1t} \]

... 

\[ Y_{Mt} = p_{M1} X_{1t} + \ldots + p_{Mk} X_{kt} + \nu_{Mt} \]
Contents

- Introduction
- Simultaneous equations models
- The identification problem
- Methods and algorithms
  - OLS
  - ILS
  - 2SLS
- Experimental results
- Conclusions and future works
The identification problem

- Underidentified >> not solve
- Overidentified >> 2SLS
- Just-identified >> ILS (also 2SLS)

- Order conditions (necessary)
- Range conditions (necessary and sufficient)
Contents

- Introduction
- Simultaneous equations models
- The identification problem
- Methods and algorithms
  - OLS
  - ILS
  - 2SLS
- Experimental results
- Conclusions and future works
OLS (Method)

OLS (Ordinary Least Square) can be used to solve a regression model

\[ Y_t = a_1 X_{1t} + ... + a_n X_{nt} + u_t \]

In matrix form

\[ Y = b X + u \]

The expression of the estimator is

\[ \hat{b} = (X'X)^{-1} X' Y \]

The determination coefficient is also calculated
ILS (Method)

The technique ILS needs the equation to be **exactly identified**, which means the values of structural coefficients can be univocally obtained from the reduced form

\[-B_i \Pi = \Gamma_i\]

- The models in **reduced** form are calculated using **OLS**
- The structural form is calculated from reduced models
2SLS (Method)

- OLS can not be used in structural form because random variable and endogenous variables are correlated.
- Endogenous variables are replaced for approximations (proxys variables).
- The proxy of Y is calculated using OLS with Y and the exogenous in the system.
- When the endogenous have been replaced, we use OLS with the variables of this equation (special OLS).
Parallel Algorithm

- Try to parallelize at the uppest level
- Two type of parallel systems:
  - Shared memory
  - Distributed memory
- One equation
- A complete SEM
OLS (Shared Memory)

- Create X and Y
- Calculate X’X
- Calculate \((X'X)^{-1}\)
- Calculate \((X'X)^{-1}X'\)
- Calculate \((X'X)^{-1}X'Y\)
- Parallel in the multiplications
- Inverse: **Lapack** is used
- \((X'X)^{-1}X'\) is used in successive OLS
ILS for 1 equation
(Shared Memory)

- ILS is
  - A loop to find Pi Matrix
  - SolutionILS (create and solve a system)

- Make the first iteration (All the threads work together)

- Each thread make several iterations \((N/np)\), and all of them use \((X'X)^{-1}X'\) in OLS (it was calculated in the first iteration)

- Parallel at low level in SolutionILS

- Use OLS

  - Not important time when there is a big sample size
ILS for a system
(Shared Memory)

- ILS in different equations can shared the Pi matrix
- ILS needs more time to solve the first equations (it doesn’t have Pi matrix).
- In a system (with a lot of just-identified equations)
  - All the threads work to solve the first equation
  - The rest of the equations are distributed between the threads
2SLS for 1 equation (Shared Memory)

- 2SLS is
  - A loop to find proxies variables
  - Special OLS

- Calculate the first proxy (All threads working together)

- Each thread calculates several proxies, and all of them use \((X'X)^{-1}X'\) in OLS (it was calculated in the first proxy).

- Parallel at low level in special OLS
Distributed Memory

- Process 0 takes the system data and sends them to the other processes.
- Each process (including process 0) solves \( \frac{N}{np} \) equations.
- Each process sends the solutions to process 0.
Contents

- Introduction
- Simultaneous equations models
- The identification problem
- Methods and algorithms
  - OLS
  - ILS
  - 2SLS
- Experimental results
- Conclusions and future works
Computer System

- HPC160: 32GFlops, 4 nodes each node with four processors (1Ghz) (OnpenMP algorithms have been tested here).
- A cluster of ten biprocessors Intel Xeon 2 with SCI connection (MPI algorithms have been tested here).
- The MPI algorithms have also been tested in Marenostrum. Marenostrum compromises 2282 JS20 compute nodes and 42 p615 servers. Each node has two processors at 2.2 Ghz running Linux operating system with 4 GB of memory RAM and 40 GB local disk storage.
## ILS time (sequential)

<table>
<thead>
<tr>
<th>N</th>
<th>K</th>
<th>n</th>
<th>k</th>
<th>d</th>
<th>ILS</th>
<th>OLS</th>
<th>%</th>
<th>Loop</th>
<th>%</th>
<th>solution ils</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>200</td>
<td>201</td>
<td>0</td>
<td>100</td>
<td>0,120115</td>
<td>0,044921</td>
<td>37,40%</td>
<td>0,055663</td>
<td>46,34%</td>
<td>0,01953</td>
<td>16,26%</td>
</tr>
<tr>
<td>500</td>
<td>200</td>
<td>201</td>
<td>0</td>
<td>500</td>
<td>0,415031</td>
<td>0,141599</td>
<td>34,12%</td>
<td>0,252925</td>
<td>60,94%</td>
<td>0,01953</td>
<td>4,71%</td>
</tr>
<tr>
<td>500</td>
<td>200</td>
<td>201</td>
<td>0</td>
<td>1000</td>
<td>1,42966</td>
<td>0,896467</td>
<td>62,70%</td>
<td>0,512685</td>
<td>35,86%</td>
<td>0,01953</td>
<td>1,37%</td>
</tr>
<tr>
<td>1000</td>
<td>400</td>
<td>401</td>
<td>0</td>
<td>100</td>
<td>0,586903</td>
<td>0,226558</td>
<td>38,60%</td>
<td>0,218746</td>
<td>37,27%</td>
<td>0,14062</td>
<td>23,96%</td>
</tr>
<tr>
<td>1000</td>
<td>400</td>
<td>401</td>
<td>0</td>
<td>500</td>
<td>3,225525</td>
<td>1,981408</td>
<td>61,43%</td>
<td>1,099588</td>
<td>34,09%</td>
<td>0,14258</td>
<td>4,42%</td>
</tr>
<tr>
<td>1000</td>
<td>400</td>
<td>401</td>
<td>0</td>
<td>1000</td>
<td>6,720575</td>
<td>4,204021</td>
<td>62,55%</td>
<td>2,372025</td>
<td>35,29%</td>
<td>0,14258</td>
<td>2,12%</td>
</tr>
<tr>
<td>1500</td>
<td>600</td>
<td>601</td>
<td>0</td>
<td>100</td>
<td>1,627899</td>
<td>0,629871</td>
<td>38,69%</td>
<td>0,537099</td>
<td>32,99%</td>
<td>0,458</td>
<td>28,13%</td>
</tr>
<tr>
<td>1500</td>
<td>600</td>
<td>601</td>
<td>0</td>
<td>500</td>
<td>8,538899</td>
<td>5,655166</td>
<td>66,23%</td>
<td>2,420852</td>
<td>28,35%</td>
<td>0,45605</td>
<td>5,34%</td>
</tr>
<tr>
<td>1500</td>
<td>600</td>
<td>601</td>
<td>0</td>
<td>1000</td>
<td>20,277933</td>
<td>12,77222</td>
<td>62,99%</td>
<td>7,042835</td>
<td>34,73%</td>
<td>0,45898</td>
<td>2,26%</td>
</tr>
<tr>
<td>2000</td>
<td>800</td>
<td>801</td>
<td>0</td>
<td>100</td>
<td>3,345639</td>
<td>1,345677</td>
<td>40,22%</td>
<td>0,918928</td>
<td>27,47%</td>
<td>1,07225</td>
<td>32,05%</td>
</tr>
<tr>
<td>2000</td>
<td>800</td>
<td>801</td>
<td>0</td>
<td>500</td>
<td>21,808178</td>
<td>15,07881</td>
<td>69,14%</td>
<td>5,652236</td>
<td>25,92%</td>
<td>1,06932</td>
<td>4,90%</td>
</tr>
<tr>
<td>2000</td>
<td>800</td>
<td>801</td>
<td>0</td>
<td>1000</td>
<td>41,598812</td>
<td>29,4799</td>
<td>70,87%</td>
<td>11,04081</td>
<td>26,54%</td>
<td>1,06834</td>
<td>2,57%</td>
</tr>
</tbody>
</table>
## ILS parallel

<table>
<thead>
<tr>
<th>N</th>
<th>K</th>
<th>n</th>
<th>1h</th>
<th>2 th</th>
<th>%</th>
<th>speedup</th>
<th>3 th</th>
<th>%</th>
<th>speedup</th>
<th>4 th</th>
<th>%</th>
<th>speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>400</td>
<td>401</td>
<td>678991</td>
<td>4095625</td>
<td>60.32%</td>
<td>1657845</td>
<td>30761</td>
<td>45.30%</td>
<td>22073</td>
<td>26689</td>
<td>39.31%</td>
<td>25441</td>
</tr>
<tr>
<td>2000</td>
<td>800</td>
<td>801</td>
<td>41445496</td>
<td>2623973</td>
<td>63.31%</td>
<td>1579494</td>
<td>20206</td>
<td>48.75%</td>
<td>205118</td>
<td>17455</td>
<td>42.11%</td>
<td>23745</td>
</tr>
<tr>
<td>3000</td>
<td>1200</td>
<td>1201</td>
<td>12563448</td>
<td>7720849</td>
<td>61.45%</td>
<td>1627211</td>
<td>60203</td>
<td>47.92%</td>
<td>208685</td>
<td>52839</td>
<td>42.06%</td>
<td>23777</td>
</tr>
<tr>
<td>5000</td>
<td>2000</td>
<td>2001</td>
<td>43063632</td>
<td>2651268</td>
<td>61.57%</td>
<td>1624266</td>
<td>202049</td>
<td>46.92%</td>
<td>213135</td>
<td>178243</td>
<td>41.39%</td>
<td>2416</td>
</tr>
<tr>
<td>7000</td>
<td>2800</td>
<td>2801</td>
<td>97565448</td>
<td>6124561</td>
<td>62.77%</td>
<td>159302</td>
<td>46708</td>
<td>47.87%</td>
<td>208884</td>
<td>414346</td>
<td>42.47%</td>
<td>23547</td>
</tr>
<tr>
<td>10000</td>
<td>4000</td>
<td>4001</td>
<td>20246293</td>
<td>1268925</td>
<td>62.67%</td>
<td>1595547</td>
<td>961.9</td>
<td>47.51%</td>
<td>210482</td>
<td>845408</td>
<td>41.76%</td>
<td>23949</td>
</tr>
</tbody>
</table>

**Sample size = 1000**
Speedup of ILS (one equation) using OpenMP

- 2 thr
- 3 thr
- 4 thr
Speedup of ILS (complete system) using OpenMP
## 2SLS time (sequential)

<table>
<thead>
<tr>
<th>N</th>
<th>K</th>
<th>n</th>
<th>k</th>
<th>d</th>
<th>2LSL</th>
<th>First proxy</th>
<th>%</th>
<th>Rest of proxys</th>
<th>%</th>
<th>Special OLS</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>200</td>
<td>201</td>
<td>0</td>
<td>100</td>
<td>0,10449</td>
<td>0,044921</td>
<td>42,99%</td>
<td>0,02637</td>
<td>25,23%</td>
<td>0,0332</td>
<td>31,78%</td>
</tr>
<tr>
<td>500</td>
<td>200</td>
<td>201</td>
<td>0</td>
<td>500</td>
<td>0,3496</td>
<td>0,148433</td>
<td>42,46%</td>
<td>0,11035</td>
<td>31,56%</td>
<td>0,09277</td>
<td>26,54%</td>
</tr>
<tr>
<td>500</td>
<td>200</td>
<td>201</td>
<td>0</td>
<td>1000</td>
<td>2,049752</td>
<td>0,969701</td>
<td>47,31%</td>
<td>0,21191</td>
<td>10,34%</td>
<td>0,86131</td>
<td>42,02%</td>
</tr>
<tr>
<td>1000</td>
<td>400</td>
<td>401</td>
<td>0</td>
<td>100</td>
<td>0,519518</td>
<td>0,230463</td>
<td>44,36%</td>
<td>0,10449</td>
<td>20,11%</td>
<td>0,18457</td>
<td>35,53%</td>
</tr>
<tr>
<td>1000</td>
<td>400</td>
<td>401</td>
<td>0</td>
<td>500</td>
<td>4,412018</td>
<td>2,074176</td>
<td>47,01%</td>
<td>0,43357</td>
<td>9,83%</td>
<td>1,85543</td>
<td>42,05%</td>
</tr>
<tr>
<td>1000</td>
<td>400</td>
<td>401</td>
<td>0</td>
<td>1000</td>
<td>9,448</td>
<td>4,417855</td>
<td>46,76%</td>
<td>1,0488</td>
<td>11,10%</td>
<td>3,98037</td>
<td>42,13%</td>
</tr>
<tr>
<td>1500</td>
<td>600</td>
<td>601</td>
<td>0</td>
<td>100</td>
<td>1,449181</td>
<td>0,64842</td>
<td>44,74%</td>
<td>0,25195</td>
<td>17,39%</td>
<td>0,54296</td>
<td>37,47%</td>
</tr>
<tr>
<td>1500</td>
<td>600</td>
<td>601</td>
<td>0</td>
<td>500</td>
<td>12,012446</td>
<td>5,662968</td>
<td>47,14%</td>
<td>1,09666</td>
<td>9,13%</td>
<td>5,27626</td>
<td>43,92%</td>
</tr>
<tr>
<td>1500</td>
<td>600</td>
<td>601</td>
<td>0</td>
<td>1000</td>
<td>27,688742</td>
<td>13,00943</td>
<td>46,98%</td>
<td>2,54486</td>
<td>9,19%</td>
<td>12,1042</td>
<td>43,72%</td>
</tr>
<tr>
<td>2000</td>
<td>800</td>
<td>801</td>
<td>0</td>
<td>100</td>
<td>3,05168</td>
<td>1,374965</td>
<td>45,06%</td>
<td>0,47949</td>
<td>15,71%</td>
<td>1,20798</td>
<td>39,58%</td>
</tr>
<tr>
<td>2000</td>
<td>800</td>
<td>801</td>
<td>0</td>
<td>500</td>
<td>30,612495</td>
<td>14,63634</td>
<td>47,81%</td>
<td>2,44135</td>
<td>7,97%</td>
<td>13,7868</td>
<td>45,04%</td>
</tr>
<tr>
<td>2000</td>
<td>800</td>
<td>801</td>
<td>0</td>
<td>1000</td>
<td>62,037785</td>
<td>29,41932</td>
<td>47,42%</td>
<td>4,88938</td>
<td>7,88%</td>
<td>27,6703</td>
<td>44,60%</td>
</tr>
</tbody>
</table>
2SLS parallel

<table>
<thead>
<tr>
<th>N</th>
<th>K</th>
<th>n</th>
<th>1 th</th>
<th>2 th</th>
<th>3 th</th>
<th>4 th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>%</td>
<td>Speed up</td>
<td>Speed up</td>
</tr>
<tr>
<td>1000</td>
<td>400</td>
<td>401</td>
<td>9,448</td>
<td>4,890532</td>
<td>51.76%</td>
<td>1,9319</td>
</tr>
<tr>
<td>2000</td>
<td>800</td>
<td>801</td>
<td>62,037785</td>
<td>34,04818</td>
<td>54.88%</td>
<td>1,82206</td>
</tr>
<tr>
<td>3000</td>
<td>1200</td>
<td>1201</td>
<td>193,61516</td>
<td>107,871</td>
<td>55.71%</td>
<td>1,79488</td>
</tr>
<tr>
<td>5000</td>
<td>2000</td>
<td>2001</td>
<td>680,42969</td>
<td>376,9001</td>
<td>55.39%</td>
<td>1,80533</td>
</tr>
<tr>
<td>7000</td>
<td>2800</td>
<td>2801</td>
<td>1503,1566</td>
<td>836,4596</td>
<td>55.65%</td>
<td>1,79705</td>
</tr>
<tr>
<td>10000</td>
<td>4000</td>
<td>4001</td>
<td>3238,9675</td>
<td>1833,813</td>
<td>56.62%</td>
<td>1,76625</td>
</tr>
</tbody>
</table>

Sample size = 1000
Speedup of 2SLS (1 equation) using OpenMP
Speedup of 2SLS (complete system) using OpenMP
## Distributed Memory (ILS complete system using MPI)

<table>
<thead>
<tr>
<th>N</th>
<th>1 th</th>
<th>2 th</th>
<th>sp</th>
<th>4 th</th>
<th>sp</th>
<th>8 th</th>
<th>sp</th>
<th>16 th</th>
<th>sp</th>
<th>24 th</th>
<th>sp</th>
<th>32 th</th>
<th>sp</th>
<th>64 th</th>
<th>sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>54,4</td>
<td>27,7</td>
<td>2,0</td>
<td>16,2</td>
<td>3,3</td>
<td>8,2</td>
<td>6,6</td>
<td>4,5</td>
<td>12,1</td>
<td>3,2</td>
<td>16,9</td>
<td>2,6</td>
<td>20,7</td>
<td>1,2</td>
<td>45,7</td>
</tr>
<tr>
<td>2000</td>
<td>1473,3</td>
<td>763,7</td>
<td>1,9</td>
<td>432,1</td>
<td>3,4</td>
<td>217,5</td>
<td>6,8</td>
<td>110,6</td>
<td>13,3</td>
<td>74,6</td>
<td>19,8</td>
<td>57,6</td>
<td>25,6</td>
<td>30,7</td>
<td>48,1</td>
</tr>
<tr>
<td>3000</td>
<td>6680,2</td>
<td>3387,1</td>
<td>2,0</td>
<td>1970,5</td>
<td>3,4</td>
<td>996,7</td>
<td>6,7</td>
<td>492,7</td>
<td>13,6</td>
<td>332,2</td>
<td>20,1</td>
<td>248,7</td>
<td>26,9</td>
<td>130,7</td>
<td>51,1</td>
</tr>
<tr>
<td>4000</td>
<td>18117,8</td>
<td>9358,9</td>
<td>1,9</td>
<td>5373,4</td>
<td>3,4</td>
<td>2622,4</td>
<td>6,9</td>
<td>1339,0</td>
<td>13,5</td>
<td>900,3</td>
<td>20,1</td>
<td>675,6</td>
<td>26,8</td>
<td>349,1</td>
<td>51,9</td>
</tr>
<tr>
<td>5000</td>
<td>43289,0</td>
<td>21363,3</td>
<td>2,0</td>
<td>12332,4</td>
<td>3,5</td>
<td>6221,6</td>
<td>7,0</td>
<td>5845,2</td>
<td>7,4</td>
<td>2079,9</td>
<td>20,8</td>
<td>1569,7</td>
<td>27,6</td>
<td>827,7</td>
<td>52,3</td>
</tr>
<tr>
<td>6000</td>
<td>83270,4</td>
<td>41971,2</td>
<td>2,0</td>
<td>23551,9</td>
<td>3,5</td>
<td>11993,3</td>
<td>6,9</td>
<td>5984,5</td>
<td>13,9</td>
<td>4021,2</td>
<td>20,7</td>
<td>3064,2</td>
<td>27,2</td>
<td>1553,6</td>
<td>53,6</td>
</tr>
</tbody>
</table>
Distributed Memory (2SLS complete system using MPI)

<table>
<thead>
<tr>
<th>N</th>
<th>1 th</th>
<th>2 th</th>
<th>sp</th>
<th>4 th</th>
<th>sp</th>
<th>8 th</th>
<th>sp</th>
<th>16 th</th>
<th>sp</th>
<th>24 th</th>
<th>sp</th>
<th>32 th</th>
<th>sp</th>
<th>64 th</th>
<th>sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>377,7</td>
<td>188,7</td>
<td>2,0</td>
<td>106,1</td>
<td>3,6</td>
<td>53,2</td>
<td>7,1</td>
<td>26,5</td>
<td>14,3</td>
<td>17,8</td>
<td>21,2</td>
<td>13,8</td>
<td>27,5</td>
<td>7,0</td>
<td>53,6</td>
</tr>
<tr>
<td>2000</td>
<td>4995,4</td>
<td>2505,7</td>
<td>2,0</td>
<td>1415,6</td>
<td>3,5</td>
<td>726,4</td>
<td>6,9</td>
<td>379,1</td>
<td>13,2</td>
<td>245,2</td>
<td>20,4</td>
<td>180,3</td>
<td>27,7</td>
<td>94,9</td>
<td>52,6</td>
</tr>
<tr>
<td>3000</td>
<td>20037,0</td>
<td>9987,2</td>
<td>2,0</td>
<td>5778,2</td>
<td>3,5</td>
<td>2983,5</td>
<td>6,7</td>
<td>1494,4</td>
<td>13,4</td>
<td>1013,1</td>
<td>19,8</td>
<td>720,2</td>
<td>27,8</td>
<td>377,4</td>
<td>53,1</td>
</tr>
<tr>
<td>4000</td>
<td>58256,5</td>
<td>29324,6</td>
<td>2,0</td>
<td>16844,4</td>
<td>3,5</td>
<td>8289,9</td>
<td>7,0</td>
<td>4109,5</td>
<td>14,2</td>
<td>2811,3</td>
<td>20,7</td>
<td>2091,0</td>
<td>27,9</td>
<td>1098,2</td>
<td>53,0</td>
</tr>
</tbody>
</table>
Contents

- Introduction
- Simultaneous equations models
- The identification problem
- Methods and algorithms
  - OLS
  - ILS
  - 2SLS
- Experimental results
- Conclusions and future works
Conclusions and Future works

- Sometimes a Simultaneous Equations Model needs special software and be solved in High Performance Systems
- Tools will be made freely available to the scientific community
- Optimize the developed methods
- Develop other methods (full information)
- Application to real problems