ESTIMATION OF A U.S. DAIRY SECTOR MODEL BY MAXIMUM SIMULATED LIKELIHOOD

Carlos Arias and Thomas L. Cox

Abstract:
This paper estimates a multivariate tobit system of monthly wholesale dairy prices where 4 prices are lower censored by the dairy price support program. Using Maximum Simulated Likelihood (MSL) we test/correct for the effects of simulation noise and discuss the relevance of estimating multivariate versus the single tobit equations.


1 Carlos Arias is a lecturer at the Department of Economics, University of Oviedo, Spain. Thomas L. Cox is a professor at the Department of Agricultural and Applied Economics, University of Wisconsin-Madison.

Copyright 1998 by Carlos Arias and Thomas L. Cox. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Estimation of a U.S. Dairy Sector Model by Maximum Simulated Likelihood

1. Introduction

There are many economic examples of systems of equations with censored endogenous variables. Among these examples are systems of demand equations in which some consumers report zero consumption of a given commodity (Wales and Woodland, 1983; Lee and Pitt, 1986). In the case of a system of equations with multiple censored variables, it is necessary to take account both of the censoring and the simultaneity. If the endogenous variables in the system are generated simultaneously, it is unlikely that the disturbances of the respective equations are independent. Therefore, there is some efficiency gain from estimating the equations as a system. Of course, the argument in favor of the estimation of the equations as a system is stronger if the theory predicts cross equation restrictions for the parameters.

The estimation of systems of equations with censored endogenous variables has been analytically treated in the literature. Maddala (1983) and Pudney (1989) present an accessible and thorough discussion of the matter and the problem that remains is basically computational. The likelihood function of the system contains several definite integrals of dimensions that range from one to the number of censored variables in the system. The evaluation of such integrals is not a trivial task for the case of more than three censored variables under the common assumption that the disturbances of the model are distributed multivariate normal.

The literature on estimation of models with multiple censored variables has searched for alternatives to the evaluation of multiple definite integrals without a closed form solution.
Despite many improvements of the initial idea, quadrature methods are still very expensive in terms of computer time. As an alternative, the probability simulation methods seem to be able to evaluate multidimensional integrals keeping a good balance between computation costs and accuracy. The estimation by maximum likelihood where probabilities are simulated rather than calculated is known as Maximum Simulated Likelihood (MSL). One of these probability simulation methods, the Geweke-Hajivassiliou-Keane (GHK) simulator, is applied in the empirical example estimated in the present paper.

The empirical part of this paper estimates a US dairy sector model. Historically, the government, through the price support program, has determined a floor price for farm level milk by buying any quantity of butter, cheese and nonfat dry milk (NFDM) at fixed prices. Therefore, the prices of these three dairy products and farm level milk prices are lower censored endogenous variables. The econometric censoring problems described earlier appear in this model and therefore provide a good example in which to apply the probability simulation technique mentioned earlier.

The structure of the paper is as follows: section 2 presents a model that describes price formation of dairy commodities at the wholesale level in the U.S., section 3 presents a probability simulation method, section 4 presents the results of the estimations and section 5 contains some conclusions.

2.- A Model of Wholesale Dairy Commodities Prices

The wholesale prices of dairy commodities display the following important features: the prices are functions of a set of exogenous variables, the wholesale dairy prices of several commodities are lower censored and the dairy prices are simultaneously determined. The price equations for each commodity can be represented by:
\begin{equation}
    p^*_i = x_i \beta + u_i \quad i = 1, \ldots, m
\end{equation}

where \( p^*_i \) is a latent variable, \( x_i \) is a vector of exogenous variables and \( u_i \) is a random disturbance with mean zero and variance \( \sigma_i^2 \). The relationship between the latent variable \( p^*_i \) and the price of the commodity can be represented by:

\begin{equation}
    p_i = \begin{cases} 
    p^*_i & \text{if } p^*_i > \bar{p}_i \\
    \bar{p}_i & \text{otherwise}
    \end{cases}
\end{equation}

where, \( p_i \) is the price of commodity \( i \) and \( \bar{p}_i \) represents the price support level for commodity \( i \).

Expression (1) describes the fact that the price of each commodity depends on a set of exogenous variables while expression (2) models the censoring process due to the existence of a price support program. Each price in the model has a "tobit" structure and the estimation of this model is straightforward. However, since the prices are determined simultaneously, the random shocks to the model are likely to be correlated. If that is the case, there are efficiency gains derived of estimating the equations in (2) as a system.

There are not many papers in the literature that estimate a system of tobit equations. A few examples are Hajivassiliou (1993), Cornick et al. (1994) and Feenberg and Skinner (1994). The likelihood function of the system of equations in the case in which all prices are above censoring level is given by:

\begin{equation}
    L = f(u_1, \ldots, u_m),
\end{equation}

where \( f \) is the probability density function of a multivariate normal function with mean zero and variance \( \Omega \). The likelihood function for an observation in which the \( n \) first prices out of \( m \) are censored is:
Expression (4) represents a portion of the likelihood function with an n-dimensional definite integral. Under the common assumption of multivariate normality of the disturbances of the system this integral does not have a closed form solution. Therefore, estimating the system of prices by maximum likelihood requires an efficient method for evaluating the high dimensional definite integrals. In the next section we analyze available methods for evaluation of these integrals.

3.- Probability simulation methods

The numerical methods used to approximate the value of a definite integral are known as quadrature methods. Besides some results of numerical analysis that reduce the number of evaluations needed for a given level of accuracy (Judd, 1996) the computing costs of quadrature methods increases very fast with the dimension of the problem.

For high dimensional problems, the probability simulation methods are an alternative to the costly quadrature methods. These methods are based on the fact that the integral of interest represents the probability of an event in a population. Lerman and Manski (1981) propose generating a pseudo-random sample of observations from the relevant population and using the relative frequency of the event in the sample to approximate the integral of interest. This simulation method is called a "crude frequency simulator" and it was improved in several subsequent papers. Stern (1992) explains the importance of smoothness in a probability simulator and proposes an smooth alternative to the "crude frequency simulator". Geweke (1989) and Borsh-Saupan and Hajivassiliou (1993) proposed the GHK simulator. Hajivassiliou et al. (1996) find that the GHK probability simulator outperforms all other methods by keeping
a good balance between accuracy and computational costs.

The GHK simulator computes the value of the integral:

$$\Pr(a < u < b) = \int_a^b g(u)du$$  \hspace{1cm} (5)$$

where, \(u\) is a random vector distributed multivariate normal with mean \(0\) and variance \(\Omega\) and \(g\) is the density function of the random vector \(u\). The starting point is that:

$$\Pr(a < u < b) = \Pr(a < Le < b)$$  \hspace{1cm} (6)$$

where, \(L\) is the lower triangular Cholesky factor of \(\Omega\), such that \(LL' = \Omega\), and \(e\) is a random vector of independent standard normal variables. The right hand side of expression (6) is easier to simulate than the probability in the left hand side due to the triangular structure of the constrains defined by \(Le\). The intervals defining the event in the right hand side of expression (6) can be written as:

$$\begin{align*}
a_1 & < l_{11} e_1 < b_1 \\
a_2 & < l_{12} e_1 + l_{22} e_2 < b_2 \\
\vdots \\
a_n & < l_{1n} e_1 + \cdots + l_{nn} e_n < b_n
\end{align*}$$  \hspace{1cm} (7)$$

where \(l_{ij}, a_i\) and \(b_i\) are the corresponding elements of \(L\), \(a\) and \(b\). Arranging terms in (7) the event in (6) can be decomposed into the following events:

$$\begin{align*}
A_1 &= \left\{ \frac{a_1}{l_{11}} < e_1 < \frac{b_1}{l_{11}} \right\} \\
A_2 &= \left\{ \frac{a_2 - l_{12} e_1}{l_{22}} < e_2 < \frac{b_2 - l_{12} e_1}{l_{22}} \right\} \\
\vdots \\
A_n &= \left\{ \frac{a_n - l_{1n} e_1 - \cdots - l_{n-1n} e_{n-1}}{l_{nn}} < e_n < \frac{b_n - l_{1n} e_1 - \cdots - l_{n-1n} e_{n-1}}{l_{nn}} \right\}
\end{align*}$$  \hspace{1cm} (8)$$
Expression (8) shows the recursive nature of the constraints that affect the random vector \( e \). As a result, the probability of interest can be written as:

\[
\Pr(a < L_e < b) = \Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_1, A_2)\ldots\Pr(A_n|A_1, \ldots, A_{n-1})
\]  

(9)

The idea behind the GHK simulator is that expression (9) can be difficult to calculate but can be simulated instead. Therefore, the GHK simulator can be written as:

\[
\tilde{\Pr}(a < L_e < b) = \frac{1}{R} \sum_{r=1}^{R} \Pr(A_1)\Pr(A_2|e_{ir})\Pr(A_3|e_{ir}, e_{2r})\ldots\Pr(A_n|e_{ir}, \ldots, e_{n-1r})
\]

(10)

where the \( e_{ir} \)'s are drawn sequentially from independent standard normal distributions truncated by expression (8) and \( R \) is the number of simulations. The truncated random variables \( e_{ir} \) can be generated smoothly using the integral transform theorem (Ross, 1988). Once the \( e_{ir} \)'s are drawn, the terms in the product are calculated as:

\[
\Pr(A_i|e_{ir}, e_{2r}, \ldots, e_{n-1r}) = \Phi \left( \frac{b_i - l_{i1} e_i - \ldots - l_{i,i1} e_{i1}}{l_{ii}} \right) - \Phi \left( \frac{a_i - l_{i1} e_i - \ldots - l_{i,i1} e_{i1}}{l_{ii}} \right)
\]

(11)

where \( \Phi \) is the cumulative distribution function of a standard normal distribution function.

Borsch-Saupan and Hajivassiliou (1993) proved that the probability simulator in (10) is an unbiased estimator of the true probability. However, the logarithm of the simulator is not an unbiased estimator of the logarithm of the true probability. As a consequence, the estimates of the parameters are biased due to simulation noise.

Hajivassiliou (1997) proposes a test for bias generated by simulation noise in MSL estimation. The null hypothesis of the test is:
\( H_0: E \left[ \frac{\partial \ln L(y, \tilde{\theta})}{\partial \theta} \right] = 0 \)

Against the alternative:

\( H_1: E \left[ \frac{\partial \ln L(y, \tilde{\theta})}{\partial \theta} \right] \neq 0 \)

where, \( L \) is a simulated likelihood function, \( y \) is a vector of observations and \( \tilde{\theta} \) is the MSL estimate of the parameter of interest \( \theta \). The rejection of the null hypothesis is interpreted as evidence of simulation noise bias. The test relies on a simulation of the data generating process \( y(\tilde{\theta}) \) to calculate the empirical mean \( (m) \) and variance \( (v) \) of the score variable. Then, the test can be written as:

\[ w = NSm'v^{-1}m \]  

(12)

where \( N \) is the number of observation in the sample and \( S \) the number of simulations per observation. Under the null hypothesis, \( w \) distributes chi-square with degrees of freedom equal to the number of parameters.

The bias created by simulation noise decrease with the number of simulations used for calculation of the GHK simulator \( (R) \). If the proposed test permits to reject the hypothesis of negligible bias due to simulation noise, we can increase the number of simulations in the GHK simulator \( (R) \) until we get an acceptable value of \( w \). Then, we can re-estimate the model with the number of simulation that gets the result of negligible simulation noise bias.

**4.- Estimation and Results**
The “Tobit” equations for the prices of wholesale dairy products in (1) and (2) are estimated by maximum likelihood using the likelihood function in (3) and (4). Since there are observations with up to four censored endogenous variables, these observations require that we simulate rather than calculate the likelihood variable as was discussed in section 3. Table 1 and 2 present the variables used in the estimation of the system.

**Table 1. Endogenous variables**

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>WPB</td>
<td>Wholesale Price of Butter</td>
</tr>
<tr>
<td>WPCH</td>
<td>Wholesale Price of Cheese</td>
</tr>
<tr>
<td>WPNFDM</td>
<td>Wholesale Price of Nonfat Dry Milk (NFDM)</td>
</tr>
<tr>
<td>WPICE</td>
<td>Wholesale Price of Ice Cream</td>
</tr>
<tr>
<td>WPOTH</td>
<td>Wholesale Price of Other Products</td>
</tr>
<tr>
<td>MW</td>
<td>Wholesale Price of Manufacturing Milk</td>
</tr>
</tbody>
</table>

**Table 2. Exogenous variables**

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADV</td>
<td>Advertising Expenditures</td>
</tr>
<tr>
<td>MP</td>
<td>Milk Production</td>
</tr>
<tr>
<td>PCE</td>
<td>Personal Consumption Expenditures</td>
</tr>
<tr>
<td>WAGE</td>
<td>Wage Rate in the Food and Kindred Industries</td>
</tr>
<tr>
<td>POPU</td>
<td>Population</td>
</tr>
<tr>
<td>R</td>
<td>Nominal Interest Rate</td>
</tr>
<tr>
<td>PPI</td>
<td>Producer Price Index</td>
</tr>
<tr>
<td>D1</td>
<td>First Quarter of the Year Dummy Variable</td>
</tr>
</tbody>
</table>

The selection of exogenous variables follows Cornick and Cox (1994). The advertising expenditures, personal consumption expenditures and population variables are demand shifters; the coefficients on these variables are expected to be positive. The wages are a measure of producer costs and are expected to have positive coefficients. The role of the producer price index can be seen as proxy for production costs. As a consequence the coefficient of this variable can be expected to be positive. However, this variable can measure the relative behavior of the dairy product price relatively to other prices in the economy. In this
case the sign can be undetermined. An increase in the quantity of milk produced can increase the supply of dairy products, therefore the coefficient on this variable is expected to be negative. The interest rate is included in the regression as a measure of the carrying costs of storage. The coefficient on this variable is expected to be negative. A higher interest rate reduces the demand for storage purposes and decreases the price of the dairy product. Each price equation also contains several lagged variables that take account of the dynamic structure of dairy product prices.

The model is estimated using monthly data for the period January-85 to December-94. The estimates of the parameters of the system are presented in table 3. The standard errors of the parameter estimates are in parenthesis. An "L" before a previously defined variable means that the variable is in logarithms in the model.

From the parameters estimates we find that, other things being equal: 1) The population variable, with the exception of the price of NFDM equation, seems to capture a long term trend of dairy prices decreasing in real terms; 2) all equations show that prices fall in the first quarter of the year; 3) all dairy prices exhibit a negative relationship with milk produced; 4) the personal consumption expenditures have a positive effect on all prices; 5) all the producer price coefficients are positive and significantly different from zero except those for cheese price and manufacturing milk price; and 6) the coefficient on interest rate is not significantly different from zero in all equations. his variable proxies storage carrying charges. 7) The dynamic structure of prices is very important in explaining changes in current price.

In the introduction of the paper we claim that if the endogenous variables in the system are generated simultaneously, it is unlikely that the disturbances of the respective equations are independent and therefore, there is some efficiency gain from estimating the equations as a
<table>
<thead>
<tr>
<th></th>
<th>LWPB</th>
<th>LWPCH</th>
<th>LWPNFDM</th>
<th>LWPICE</th>
<th>LWPOTH</th>
<th>LMW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.197</td>
<td>-0.071</td>
<td>-0.168</td>
<td>0.000</td>
<td>-0.003</td>
<td>-0.077</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.016)</td>
<td>(0.028)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>LADV</td>
<td>-0.004</td>
<td>0.014</td>
<td>-0.039</td>
<td>-0.005</td>
<td>-0.001</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.019)</td>
<td>(0.025)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>LWAGE</td>
<td>-0.786</td>
<td>-0.016</td>
<td>-0.495</td>
<td>-0.013</td>
<td>-0.010</td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.060)</td>
<td>(0.079)</td>
<td>(0.023)</td>
<td>(0.024)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>LMP</td>
<td>-0.117</td>
<td>-0.125</td>
<td>-0.071</td>
<td>0.001</td>
<td>-0.009</td>
<td>-0.167</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.046)</td>
<td>(0.044)</td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>LPCE</td>
<td>0.639</td>
<td>0.5601</td>
<td>0.571</td>
<td>0.167</td>
<td>0.072</td>
<td>0.341</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.029)</td>
<td>(0.044)</td>
<td>(0.020)</td>
<td>(0.024)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>LPOPU</td>
<td>-1.677</td>
<td>-1.391</td>
<td>0.502</td>
<td>-0.420</td>
<td>-0.243</td>
<td>-0.350</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.062)</td>
<td>(0.053)</td>
<td>(0.023)</td>
<td>(0.082)</td>
<td>(0.064)</td>
</tr>
<tr>
<td>LR</td>
<td>-0.013</td>
<td>-0.005</td>
<td>0.002</td>
<td>-0.000</td>
<td>0.000</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.021)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>LPPI</td>
<td>-0.010</td>
<td>0.248</td>
<td>0.675</td>
<td>0.114</td>
<td>0.067</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.032)</td>
<td>(0.113)</td>
<td>(0.020)</td>
<td>(0.019)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>D1</td>
<td>-0.052</td>
<td>-0.024</td>
<td>-0.022</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.014)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>L1</td>
<td>1.148</td>
<td>0.967</td>
<td>0.975</td>
<td>0.884</td>
<td>1.392</td>
<td>1.143</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.068)</td>
<td>(0.084)</td>
<td>(0.041)</td>
<td>(0.025)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>L2</td>
<td>-0.655</td>
<td>-0.353</td>
<td>-0.463</td>
<td>-0.012</td>
<td>-0.450</td>
<td>-0.473</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.061)</td>
<td>(0.064)</td>
<td>(0.041)</td>
<td>(0.025)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>L3</td>
<td>0.288</td>
<td>(0.026)</td>
<td></td>
<td></td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.042)</td>
<td></td>
</tr>
</tbody>
</table>
system. However, the use of probability simulation methods can produce bias in parameter estimates due to simulation noise. Therefore, it is very important to test for the existence of simulation noise bias in the estimation of the parameters. The test presented in expression (16) suggests that for a modest number of simulations in the GHK simulator (R=100) bias due to simulation noise are negligible. This finding is consistent with results presented in Hajivassiliou et al. (1996) that suggest the use of a number of simulations (R) equal to the number of observations in the sample.

The estimation of a system of equations containing several tobit equations is far from trivial and to our best knowledge the computer code is to some extent case specific. On the other hand, the estimation of single tobit equations is an almost trivial task. The natural question to ask is: "What are we really accomplishing when estimating the system of equations containing tobits versus estimating single equations separately?" The question is not well treated in the literature but it is easy to identify tentative answers. First, from a statistical point of view, single equation estimation can be seen as the result of imposing some parametric restrictions in the system. More precisely, the single equation estimation is equivalent to the estimation of a system of equations with a variance-covariance matrix where all the off-diagonal elements are equal to zero. This parametric restriction can be tested following Hajivassiliou (1997). In the present case, we can not reject the restricted model (single equation estimation). However, this test may suffer from the problems that arise in the estimation of the variance-covariance elements (Greene, 1997). Secondly, from an economic point of view, we can check if the unrestricted models provide results that differ from the restricted one. In the present case, there are not important parameter differences between the restricted and unrestricted model.
The importance of the methodology used in this paper should not be obscured by the results that suggest that the single equation tobit equations can do the job in this case. It is important to keep in mind that we obtained this result using MSL. Without this methodology the relevance of the systems of equations would be an open question.

5.- Conclusions

This paper explores the feasibility of using probability simulation methods in maximum likelihood estimation of a system of equations with multiple censored endogenous variables. The GHK probability simulator proves to be useful in estimating models with more than three limited dependent variables.

It is important to stress that the estimation of these models was considered unfeasible until very recently. It is also important to analyze the bias generated by simulation noise when estimating a model by MSL. In the present case, the bias was negligible for a modest number of simulation in the computation of the GHK simulator.

The estimation of a model with several related tobit equations is far from being a trivial task. It is natural to evaluate the relevance of estimating a system of tobit equations versus the estimation of single equation tobits. For the example in this paper, we can not statistically reject the restricted model composed of single tobit equations.
6.- References


