Estimation of simultaneous equation models with measurement error

Kees Van Montfort
Free University, Department of Econometrics,
De Boelelaan 1105, 1081 HV Amsterdam, The Netherland

Jan De Leeuw
UCLA, Department of Mathematics,
405 Hilgard Avenue, Los Angeles, CA 90024-1555, USA

Abstract: We consider a simultaneous equation model with measurement error. Using characteristic functions and spline approximation we derive a class of asymptotically unbiased estimators of the unknown model parameters. The estimators are functions of the empirical characteristic functions of the observed random variables. Asymptotically, they are normally distributed and the values of their variances reach the Cramer-Rao lower bound. Finally some numerical examples will be discussed.
Keywords: simultaneous equations, measurement error, characteristic functions, spline approximation, BGLS-estimators.

1. Introduction

Simultaneous equation models with measurement error are well-known in econometrics (see for instance Geraci (1976), Hsiao (1976,1983), Aigner et al. (1984), Merckens and Bekker (1993)). In this paper we will deal with the reduced form, where the matrix of parameters of the endogenous variables is equal to the identity matrix:

$$\underline{Y} = A\,\underline{X} + \underline{u} \qquad (1a)$$

$$\underline{y} = \underline{Y} + \underline{e} \qquad (1b)$$

$$\underline{x} = \underline{X} + \underline{d} \qquad (1c)$$

, where
$\underline{y}$ and $\underline{Y}$ are respectively (gx1)-vectors of observed and unobserved endogenous random variables;
$\underline{x}$ and $\underline{X}$ are respectively (kx1)-vectors of observed and unobserved exogenous random variables;
A is a (gxk)-matrix of parameters;
$\underline{u}$ and $\underline{e}$ are (gx1)-vectors of unobserved disturbances with mean zero, and
$\underline{d}$ is a (kx1)-vector of unobserved disturbances with mean zero.

Instead of the latent variables $\underline{Y}$ and $\underline{X}$, we observe $\underline{y}$ and $\underline{x}$, respectively. Assumptions about the distributions of the random variables will not be made. We only assume, that all the disturbances have mean zero and are independent of each other, and independent of $\underline{Y}$ and $\underline{X}$.

Hsiao (1976) has developed a full information maximum likelihood method to estimate the unknown model parameters for this model from the covariance matrix of the observed variables. In addition, he assumes that all the random variables are normally distributed and that the covariance matrices of the disturbances are known. Of course such assumptions are quite restrictive and may not be satisfied in most applications. Omitting these additional assumptions, the Ordinary Least Squares (OLS) estimators of our model parameters, which use first and second order moments, are usually biased (see Van Montfort et al. (1987)). The differences between the expected values of the OLS-estimators and the real values of A are linear functions of the variances and covariances of the disturbances.

In this paper we use characteristic functions to estimate A asymptotically consistent. Van Montfort et al. (1989) used already characteristic functions for estimating a linear regression model with measurement errors, which is a special case of our model (1).

The characteristic function of a random variable is a transformation of the corresponding density function. As a consequence, using complete characteristic functions no information of the distributions of the variables is lost. Using the mean and the variance of a variable we only deal with a limited amount of information about the characteristic function in the point zero. The mean of a random variable is the first order derivative of the corresponding characteristic function in the point zero, multiplied by i. The second order moment of a random variable is minus the second order derivative of the characteristic function in the point zero. Furthermore, every distribution has a corresponding characteristic function, even if the higher order moments do not exist. For instance, the Cauchy distribution and the Student distribution with two degrees of freedom have no second and higher order moments, while the characteristic functions exist. Finally, a third advantage of using characteristic functions is the fact that empirical characteristic functions have small variances, even for small sample sizes.

We can express the logarithms of the characteristic functions of the observed variables $\underline{x}$ and $\underline{y}$ in terms of the logarithms of the characteristic functions of $\underline{X}$, $\underline{d}$ and $\underline{u}+\underline{e}$:

$$\ln c_{x,y}(v_1,...,v_k,w_1,...,w_g) = \ln c_X(v_1 + \Sigma_{i=1}^{g} w_iA_{i,1},.........,v_k + \Sigma_{i=1}^{g} w_iA_{i,k}) +$$

$$\Sigma_{i=1}^{k} \ln c_d^{(i)}(v_i) + \Sigma_{i=1}^{g} \ln c_{u+e}^{(i)}(w_i) \qquad (2)$$

, where $c_d^{(i)}$ corresponds with component i of $\underline{d}$, $c_{u+e}^{(i)}$ corresponds with component i of $\underline{u}+\underline{e}$ and $v_1,...,v_k,w_1,...,w_g$ arbitrary points on the real line.

Equation (2) represents a simple additive relation between the observed and the unobserved random variables. We can eliminate the logarithms of the characteristic functions of the unobserved errors $\underline{d}$ and $\underline{u}+\underline{e}$ in (2):

$$\ln c_{x,y}(v_1,...,v_k,w_1,...,w_g) -$$

$$\ln c_{x,y}(v_1,0,...,0) - .... - \ln c_{x,y}(0,...,0,w_g)$$

$$= \ln c_X(v_1 + \Sigma_{i=1}^{g} w_iA_{i,1},.........,v_k + \Sigma_{i=1}^{g} w_iA_{i,k}) -$$

$$\ln c_X(v_1,0,...,0) - .... - \ln c_X(w_gA_{g,1},.........,w_gA_{g,k}) \qquad (3)$$

The means of the disturbances are assumed to be zero. Therefore the first order derivatives of "$\ln c_x(.,...,.) - \ln c_X(.,...,.)$" and "$\ln c_y(.,...,.) - \ln_Y(.,...,.)$" in the point zero have to be zero. The appendix shows that under this condition the identification of the matrix A in (1) also implies the identification of A by the set of equations (3). In this paper we approximate the logarithm of the characteristic functions of the unobserved random variable in system (3) by a tensor-product B-spline, which is the product of several univariate B-splines. Next, the resulting system will be used to derive estimators of the matrix A.

Section two discusses the approximation of an arbitrary function by B-splines. In section three we approximate the logarithm of the characteristic function of the random vector $\underline{X}_t$ in equation (3) by a tensor-product B-spline. In section four we derive a Best Generalized Least Squares (BGLS) estimator of the matrix A with the help of system (3) and the results of section three. In section five using the Method of Sieves (Geman and Hwang (1982)) we deal with some asymptotic properties of our BGLS estimator of section four. Our BGLS estimators of the elements of A are asymptotically normally distributed and their variances reach the Cramer-Rao lower bound asymptotically. Finally we discuss some numerical examples.

2. Approximation by splines

Spline functions have lately found great use as approximating functions. Ordinary polynomials are inadequate in many situations. Wold (1974) mentioned that in many situations the behavior in one region may be totally uncorrelated to the behavior in another region. Polynomials along with most other mathematical functions have the property, that their behavior in a small region determines their behavior everywhere. Splines do not suffer this handicap since they are defined piecewise. Furthermore De Leeuw and Van Rijckevorsel (1981) notice that the use of splines takes away the disadvantage of the crudeness or discontinuity of the step functions at the break-points, while maintaining the advantage of not approximating strictly linear functions only.

Lets define a given, fixed partition of (n-1) subsets of an interval I with n knots. Spline-functions, defined on an interval I, are piecewise polynomial functions, because they consist of (n-1) s^{th} -order polynomials, one defined on each subinterval of I. In each of the interior knots the polynomials are connected, and (s-1) times continous differentiable (see De Boor (1978), Powell (1981), Schumaker (1981) and Chui (1988)). The location of the interior knots is relatively unimportant if the curve to be fitted is reasonably smooth, although it is helpful to have more knots in regions where the non linearity is most severe (Winsberg and Ramsay (1980)). This insensitivity to knot choice implies that in practice the knots may be chosen a priori and held fixed during the analysis. If the curve to be fitted is not smooth and we know the function to be approximated well enough knot optimization is mainly useful as a free parameter. Wold (1974) and Van Rijckevorsel (1987) give some indications for knot placement optimization. However, Van Rijckevorsel concludes that even if we would accept a greater dimensionality, an uniform knot placement with a generous number of knots would do just as well. Furthermore he warns for the realistic danger of overfitting.

The set of spline functions on an interval I forms a linear space of dimensionality (n+s). This means that each spline can be uniquely expressed as a linear combination of (n+s) "basis splines" of order s. Often B-splines (see Powell (1981), page 229) are used as "basis splines". B-splines are a special kind of splines. They are non-zero on (s+1) consecutive intervals, and zero outside that interval.

Defining a spline function in terms of polynomials is convenient once the polynomial coefficients are known. In the computational process of least squares curve fitting, however, it is simpler to define a spline function in terms of B-splines. In this paper we approximate a multivariate function by a linear function of products of several univariate B-splines (= a tensor-product B-spline).

3. Approximating the logarithms of the characteristic functions of the unobserved exogenous random variables by splines

Consider the simultaneous equation model with measurement error of section one. In this section we approximate the logarithms of the characteristic functions of the unobserved exogenous random variables $\underline{X}$ in equation (3) by splines. Section four will derive estimators of matrix A with the help of the results of this section.

We notice that the logarithm of the characteristic function of $\underline{X}$ is a multiple valued function. The single valued interpretation of it is only true for the real part, which corresponds with the logarithm of the distance between the complex value of the characteristic function and the origin in the complex space (see Epps (1993)):
$\text{Re}\,[\ln c_X(v_1,\ldots,v_k)] = \ln \{(\text{Re}\,c_X(v_1,\ldots,v_k))^2 + (\text{Im}\,c_X(v_1,\ldots,v_k))^2\}^{1/2}$.
For the imaginary part, which is equal to the argument of the characteristic function in the complex space, we have equality modulo 2π. Thus we have
$\text{Im}\,[\ln c_X(v_1,\ldots\ldots,v_k)] = \text{Im}\,[\ln c_X(v_1,\ldots\ldots,v_k)] + 2\pi m$, with m an arbitrary integer. This is a bit inconvenient to work with numerically. We can ignore this numerical problem if the arguments of the characteristic functions have to lie in the interval $[-\pi,\pi]$. Of course, in the case $\underline{X}$ is symmetrically distributed with mean zero, the imaginary part of $\ln c_X(.,\ldots,.)$ is equal to zero.

Now, subdivide the interval $[-\pi,\pi]$ in an arbitrary fixed number of disjunct subintervals and define two approximation tensor-product B-splines of order three on the space $[-\pi,\pi]\times\ldots\times[-\pi,\pi]$ (see, for instance, Powell (1981), page 229). Approximating the real and the imaginary parts of $\ln c_X(.,\ldots,.)$ separately by these two multivariate splines (call them B and C) we can reformulate equation (3):

$$\ln c_{x,y}(v_1,\ldots,v_k,w_1,\ldots,w_g) -$$

$$\ln c_{x,y}(v_1,0,\ldots,0) - \ldots\ldots\ldots\ldots - \ln c_{x,y}(0,\ldots,0,w_g)$$

$$\approx B(v_1 + \textstyle\sum_{i=1}^{g} w_iA_{i,1},\ldots\ldots\ldots,v_k + \sum_{i=1}^{g} w_iA_{i,k}) -$$

$$B(v_1,0,\ldots,0) - \ldots\ldots\ldots\ldots - B(w_gA_{g,1},\ldots\ldots\ldots,w_gA_{g,k})$$

$$i * C(v_1 + \textstyle\sum_{i=1}^{g} w_iA_{i,1},\ldots\ldots\ldots,v_k + \sum_{i=1}^{g} w_iA_{i,k}) -$$

$$i * C(v_1,0,...,0) - .... - i * C(w_g A_{g,1},........., w_g A_{g,k}) \qquad (4)$$

The left hand side and the right hand side of (4) are functions of $(v_1,...,v_k,w_1,...,w_g)$. For every different value of the combination of $(v_1,...,v_k,w_1,...,w_g)$ relation (4) contains a different equation. So, relation (4) is a system of an infinite number of equations. From Winsberg and Ramsey (1980) it follows that for fixed values of the model parameters the unknown spline coefficients of the tensor-product B-spline representations are determinated uniquely by (4).

In the following of this paper we ignore the imaginary part of (4) for reasons of simplicity. However, if we keep this part the principles of our derivations remain the same.

4. Derivation of a BGLS estimator of the matrix A

With the help of the characteristic function relation (4) and the Best Generalized Least Squares (BGLS) method we derive estimators of the coefficients of the matrix A. Therefore we define the following Best General Least Squares function BGLS:

$$BGLS = Z^T V^{-1} Z \, , \qquad (5)$$

, with Z a column vector;
, $Z_i = \underline{L}_X(grid_i) - R_X(grid_i)$
, where $\underline{L}_X(.)$ is the empirical function of the left hand side of equation (4)
and R_X is the right hand side of equation (4)
and $grid_i$ is an arbitrary point $(v_1,...,v_k,w_1,...,w_g)$ in the (k+g)-dimensional space $[-\pi,\pi]\times......\times[-\pi,\pi]$
, and V the asymptotic dispersion matrix of the vector Z.

BGLS is a function of the unknown spline coefficients and the unknown model parameters A_{ij} $(i=1,...,g;j=1,...,k)$. We will find BGLS estimates of the model parameters by minimizing the function BGLS. However, the number of components of Z has to be equal or greater than the number of independent spline parameters (="$(s+nknots)^k-1$", where s is the degree of the piecewise polynomials and nknots is the number of knots on the approximation interval (see Chui, 1988)). Otherwise the spline coefficients are not determinated uniquely by the minimizing procedure.

Suppose the logarithms of the characteristic functions of the unobserved exogenous random variables can be exactly approximated by our splines. Assume furthermore that the unknown model parameters are identified by the system of equations $Z^T=(0,...,0)$ and that the asymptotic dispersion matrix of Z is invertable. In that case Feuerverger and Mc Dunnough (1981a,1981b) prove the estimates from (5) converge almost surely to their real values. In addition, the estimators are asymptotically normally distributed with asymptotic variance $\partial Z\ V^{-1}\ \partial Z$ (∂Z is the derivative from Z with respect to the unknown parameters). Feuerverger and McDunnough (1981a, 1981b) also show that by using a sufficiently fine and extensive grid $\{(v_1^{(i)},...,v_k^{(i)},w_1^{(i)},...,w_g^{(i)}),\ i=1,...,\text{dimension } Z\}$ the asymptotic variances of the estimators can be made arbitrarily close to the Cramer-Rao bound. So, our procedure can attain optimal asymptotic efficiency.

An alternative of minimizing BGLS is minimizing some linearized form bgls of BGLS:

$$\text{bgls} = U^T\ V^{-1}\ U$$

, with U a column vector with the same dimension as Z

and $U_i = Z_i(\theta^*) - D^T(\theta^*)*(\theta - \theta^*)$

θ^* is an asymptotically consistent estimator of the vector θ which contains all the model parameters and all the unknown spline coefficients. $D(\theta^*)$ is the matrix containing the first order derivatives of the vector Z to all the model parameters and all the unknown spline coefficients in the point θ^*.

This alternative has been proposed by Ferguson (1958), who showed that the estimates have the same asymptotic properties as estimates found from minimizing BGLS. De Leeuw (1983) and Bentler (1983) call this method a two stage ADF method (the 2SDAF method): first find an asymptotically consistent estimator of θ, say θ^*, next compute $A(\theta^*)$ and find θ^{**} by minimizing bgls.

In this paper θ^* will be the estimator which can be obtained by minimizing (5) with V equal to the identity matrix. So, in step one of the above procedure we have to minimize the function $\lambda = Z^TZ$. However, we will minimize λ by minimizing it over the unknown spline coefficients and the model parameters alternatingly. At one time we fix the set of the spline coefficients and minimize over the model parameters, the other time we fix the model parameters and minimize over the spline coefficients. We continue to do so until no further improvement can be made. This principle has worked very well in many related data analysis contexts (see for instance, Young (1981) and Wold (1982)).

Minimizing over the unknown spline coefficients for fixed model parameters is a quadratic optimization problem with some linear constraints, because the means of the disturbances (= the first order derivatives of "$\ln c_x(.,...,.) - \ln c_X(.,...,.)$" and "$\ln c_y(.,...,.) - \ln_Y(.,...,.)$" in the point zero) have to be zero. Furthermore the equation $\mathrm{Re}\,[\ln c_X(0,......,0)] = 0$ has to be involved. On the other hand minimizing over the model parameters is a nonlinear optimization problem without additional restrictions. For instance, the Newton-Raphson method can be used (see Zangwill (1969)).

5. Some asymptotic properties of the BGLS estimator

Until now we partitioned the interval $[-\pi,\pi]$ into a fixed number of subintervals. In this section the number of knots for our approximation splines is not fixed but depends on the sample-length sl. We will define a relation between the sample-length sl and the maximum number of knots nk_{sl} on the interval $[-\pi,\pi]$. Using this relation we will prove some asymptotic properties of our BGLS estimators.

Now, let the multivariate function $\ln c_X(.,...,.)$ be 4 times continuous differentiable on the space $[-\pi,\pi]\times...\times[-\pi,\pi]$. Suppose also that the tensor product spline $B(.,...,.)$ has the knots $-\pi=kn_1<...<kn_N=\pi$ in the interval $[-\pi,\pi]$ and satisfies the property $\ln c_X(kn_{j1},...,kn_{jk})=B(kn_{j1},...,kn_{jk})$ for every $j1,...,jk=1,...,N$.
Then for every $v_{j1},...,v_{jk}$ in the interval $[-\pi,\pi]$ (see Schumaker (1981)):

$$|\ln c_X^{(q)}(v_{j1},...,v_{jk})-B^{(q)}(v_{j1},...,v_{jk})| \leq O(d^{(4-q)}) \qquad (6)$$

with d the maximal distance between two successive knots and $\ln c_X^{(q)}(.,...,.)$ the q^{th}-order derivative of the function $\ln c_X(.,...,.)$.

System (4) contains the logarithm of the characteristic function of $\underline{x}$ and $\underline{y}$, and the tensor-product B-spline approximation of the logarithm of the characteristic function of $\underline{X}$. It is not necessary to make the approximation errors of the abovementioned approximations smaller than the variances of the logarithms of the empirical characteristic functions of $\underline{x}$ and $\underline{y}$ in different points $(v_1,...,v_k,w_1,...,w_g)$. Using this condition and expression (6), we can define the number of knots nk_{sl} for a given sample-length sl in the following way:

nk_{sl} is the largest integer-valued number for which the following constrains are satisfied:

$$d^4 \geq \max_{v1,...,vk,w1,...,wg} \{\mathrm{var}\ \ln \underline{c}_{x,y}(v_1,...,v_k,w_1,...,w_g) \mid v_1,...,v_k,w_1,...,w_g \text{ in } [-\pi,\pi]\} \qquad (7)$$

The left hand side of (7) represents the upperbound for the approximation of the logarithm of the characteristic funtion of $\underline{X}$. The right hand side of the above constraint represents the maximum of the variances of the logarithms of the characteristic function of $\underline{x}$ and $\underline{y}$ in several points in the (k+g)-dimensional space $[-\pi,\pi]\times......\times[-\pi,\pi]$. Dealing with an uniform knot placement the left hand side and the right hand side of (7) are of order $O((nk_{sl})^{-4})$ and $O((sl)^{-1})$. Because we assume $O((nk_{sl})^{-4}) \geq O((sl)^{-1})$, nk_{sl} is of order $O((sl)^{1/4})$.

From (7) we can easily see that nk_{sl} is an increasing function of sl. If sl --> ∞ then nk_{sl} --> ∞. So for sl --> ∞ the logarithms of the characteristic functions of the unobserved exogenous random variables will be approximated exactly. Now from Feuerverger and McDunnough (1981a,1981b) follows that our BGLS estimators of the model parameters are asymptotically normally distributed and their values for the variances reach the Cramer-Rao lower bound.

In this section we used the principle of the nonparametric estimation Method of Sieves (Geman and Hwang (1982)):

> Suppose we have to estimate a parameter which takes values in an infinite dimensional space. Then we can estimate this parameter if we first estimate in a constrained subspace of the parameter space and next relax the constraint as the sample size grows.

Finally we remark that the dimension of Z in expression (5), which is equal to the number of different values for the combinations $(v_1,...,v_k,w_1,...,w_g)$, depends on the number of spline knots. The dimension of Z has to be larger than the number of knots nk on the intervals $[-\pi,\pi]$. If sl --> ∞ and nk_{sl} --> ∞ then dimension Z --> ∞.

6. An example with generated data

Using an example with simulated data we can illustrate the influence of the number of different values for the combinations $(v_1,...,w_g)$ in (5), which is equal to the number of components of the random vector Z. We deal with model (1) and only one endogenous (g=1) and two exogenous random variables (k=2). In that case we have a linear regression model with measurement errors.

Scores on the unobserved disturbances $\underline{u}$, $\underline{e}$ and $\underline{d}=(\underline{d}_1,\underline{d}_2)^T$ are drawn at random from normal distributions with the variances equal to the arbitrary values (0.4;0.2;0.2;0.2), (0.2;0.4;0.2;0.2), (0.2;0.2;0.4;0.2) or (0.2;0.2;0.2;0.4) respectively. Scores on each component of the unobserved exogenous random variable $\underline{X}$ are generated at random by a chi-square distribution with one, two or three degrees of freedom, and by a Student distribution with two degrees of freedom. The matrix of model parameters A is set on the values (1,1), (3,1) and (1,3) successively. The values of the observed random variables $\underline{x}$ and $\underline{y}$ are calculated next on the basis of the above scores and the chosen parameter values.

From Kapteyn and Wansbeek (1983) and Bekker (1986) it follows that this model is identified. Using the generated $\underline{x}$- and $\underline{y}$-scores and the method of this paper, the model parameters are estimated with sample size 50. We chose splines of order three and an uniform knots placement with an arbitrary number of three knots on the interval $[-\pi,\pi]$. The number of different values for the combination (v_1,v_2,w_1) in (5) had to be equal or greater than the number of independent splines parameters $(=(3+3)^2-4=32)$ and varied from 35 to 60 (35,40,45,50,55,60).

Every specific distribution setting, parameter setting and number of different components of the random vector Z in (5) was replicated 300 times (16 x 3 x 6 x 300 estimation runs). It followed that the number of different values for the combination (v_1,v_2,w_1) in (5) didn't influence the variances and the means of the estimates of the unknown model parameters significantly. However, if this number was "too large" (in this case lager than about 50) there was a realistic danger that the asymptotic dispersion matrix V was not invertable.

The estimation results were reasonable well for each distribution setting, parameter setting and number of different components of Z. The variances of the estimates were maximal about 25 percent of the real model parameter values, while the means of the estimates differed maximal about 10 percent of the real parameter values.

We also used Ordinary Least Squares (OLS) estimates. The simulation runs, where the components of $\underline{X}$ had Student distributions with two degrees of freedom, gave absurd estimates of the model parameters. For the other runs it followed, that their variances had the same size as those of our characteristic function estimates. Additionally, increasing variances of the measurement disturbances caused increasing bias of the means of the OLS estimates. Some parameter and distribution settings had an estimation bias of nearly 30 percent of the real parameter values.

7. An empirical example

We have applied the estimation method of this paper to an empirical example. In this example we deal with two endogenous and two exogenous random variables. Furthermore, we choose the approximation interval $[-\pi,\pi]$, three spline knots on this interval and forty different values for the combination (v_1,v_2,w_1,w_2). If the arguments of the logarithms of the characteristic functions of the endogenous random variables are going beyond the above approximation interval, our method has some complications. In that situation we drop out the values for the combinations (v_1,v_2,w_1,w_2) of which the corresponding arguments are outside the approximation interval. However, using this correction method we are changing the optimizing function.

The data analyzed here are collected by the Netherlands Central Bureau of Statistics (CBS), measuring the average income of one-person families (input variable 1), the average income of multi-persons families (input variable 2), the average basis house-rent per month (output variable 1) and the average sale-worth of houses (output variable 2) for forty separate regions in The Netherlands (1993). All the variables are normalized (means equal to zero and standard deviations equal to one). The regression coefficients, estimated by the method of this paper, are presented in Table 1. The standard deviations of these regression coefficient estimators are computed with the bootstrap method and are drawn between brackets. The corresponding OLS estimates, which were used as starting values for our regression coefficient estimates, are also mentioned.

If the model is correct, the minimizing function BGLS (see (5)) is asymptotically chi-square distributed with degrees of freedom equal to the number of different values for the combination (v_1,v_2,w_1,w_2) (=40), minus the sum of the number of the unknown regression coefficients (=4) and the number of the independent spline coefficients corresponding with the characteristic functions of the two unobserved exogenous random variables ($=(3+3)^2-5=31$). So, the BGLS-function is approximately chi-square distributed with 40-4-31=5 degrees of freedom. In this example the chi-square minimizing function gets the value 0.7. The probability level is 0.98 and we cannot reject the model on the basis of the data. In the first iteration step, where the starting values of the unknown regression coefficients are set equal to the OLS-estimates and the chi-square minimizing function BGLS is optimized by the unknown spline coefficients, the value for this optimizing function is about 7.5 (probability level 0.30).

The regression coefficients $A_{1,1}$ and $A_{1,2}$ for the first output variable are all in the neighbourhood of zero. Therefore we may conclude that per region the average annually basis house-rent does not depent on the average income of families. The second output variable has positive regression coefficients ($A_{2,1}$ and $A_{2,2}$). This can be interpreted that per region the average sale-worth of houses has a positive correlation with the average income of one-person and multi-persons families. For multi-person families is this correlation most positive.

Appendix

Theorem:
Assume that the expectations of the disturbances are zero. Then the identification of the matrix A in (1) also implies the identification of A by the set of equations (3).

Proof:
Suppose that A is identified by (1). Then A also satisfies relation (2) and (3).

Let the matrix B also satify relation (3) and let B differ from A.
In that case follows for every $v_1,\ldots,v_k,w_1,\ldots, w_g$ on the real line:

$$\ln c_X(v_1 + \Sigma_{i=1}^{g} w_i A_{i,1},\ldots\ldots\ldots,v_k + \Sigma_{i=1}^{g} w_i A_{i,k}) -$$

$$\ln c_X(v_1,0,\ldots,0) - \ldots\ldots\ldots\ldots - \ln c_X(w_g A_{g,1},\ldots\ldots\ldots,w_g A_{g,k})$$

$$= \ln c_{X'}(v_1 + \Sigma_{i=1}^{g} w_i B_{i,1},\ldots\ldots\ldots,v_k + \Sigma_{i=1}^{g} w_i B_{i,k}) -$$

$$\ln c_{X'}(v_1,0,\ldots,0) - \ldots\ldots\ldots\ldots - \ln c_{X'}(w_g B_{g,1},\ldots\ldots\ldots,w_g B_{g,k}) \qquad (3')$$

,where $c_{X'}(.,\ldots,.)$ is a characteristic function.

$c_{X'}(.,\ldots,.)$ is not necessarily equal to $c_X(.,\ldots,.)$. Furthermore the first order derivatives of "$\ln c_x(.,\ldots,.) - \ln c_{X'}(.,\ldots,.)$" and "$\ln c_y(.,\ldots,.) - \ln_{Y'}(.,\ldots,.)$" in the point zero have to be zero. This condition follows from the assumption, that the expectations of the disturbances are zero.

Substituting the above relation (3') in (2) gives for the arbitrary points $v_1,\ldots,v_k,w_1,\ldots,w_g$ on the real line the following expression:

$$\ln c_{x,y}(v_1,\ldots,v_k,w_1,\ldots,w_g) =$$

$$\ln c_{X'}(v_1 + \Sigma_{i=1}^{g} w_i B_{i,1},\ldots\ldots\ldots,v_k + \Sigma_{i=1}^{g} w_i B_{i,k}) -$$

$$\ln c_{X'}(v_1,0,\ldots,0) - \ldots\ldots\ldots\ldots - \ln c_{X'}(w_g B_{g,1},\ldots\ldots\ldots,w_g B_{g,k}) +$$

$$\ln c_X(v_1,0,\ldots,0) + \ldots\ldots\ldots\ldots + \ln c_X(w_g A_{g,1},\ldots\ldots\ldots,w_g A_{g,k}) +$$

$$\Sigma_{i=1}^{k} \ln c_d^{(i)}(v_i) + \Sigma_{i=1}^{g} \ln c_{u+e}^{(i)}(w_i)$$

Table 1, Estimates of the regression coefficients.
(corresponding standard deviations between brackets)

$A_{1,1}^{BGLS}$	-0.03334	$A_{1,2}^{BGLS}$	-0.06361
	(0.02618)		(0.00666)
$A_{2,1}^{BGLS}$	0.18111	$A_{2,2}^{BGLS}$	0.78229
	(0.00002)		(0.01102)
df = 5			
$X^2 = 0.7$			
$A_{1,1}^{OLS}$	-0.02825	$A_{1,2}^{OLS}$	-0.04995
	(0.02894)		(0.00884)
$A_{2,1}^{OLS}$	0.25560	$A_{2,2}^{OLS}$	0.82389
	(0.00002)		(0.01024)
df = 5			
$X^2 = 7.5$			

$$= \ln c_{X'}(v_1 + \Sigma_{i=1}^{g} w_i B_{i,1}, \ldots\ldots\ldots, v_k + \Sigma_{i=1}^{g} w_i B_{i,k}) +$$

$$\Sigma_{i=1}^{k} \ln C_d^{(i)}(v_i) + \Sigma_{i=1}^{g} \ln C_{u+e}^{(i)}(w_i) \qquad (2')$$

, where $C_d^{(i)}(.)$ and $C_{u+e}^{(i)}(.)$ are characteristic functions not necessarily equal to $c_d^{(i)}(.)$ and $c_{u+e}^{(i)}(.)$.

Let $C_d^{(i)}(.)$ and $C_{u+e}^{(i)}(.)$ correspond with the random disturbances $\underline{d}'$ and $(\underline{u}'+\underline{e}')$. Obviously, the first order derivatives of $C_d^{(i)}(.)$ and $C_{u+e}^{(i)}(.)$ in the point zero are equal to zero. This means that each distribution of (y,x), which is generated by the matrix A and the distribution of $(\underline{Y},\underline{X},\underline{u}+\underline{e},\underline{d})$, can also be generated by B and the distribution of $(\underline{Y}',\underline{X}',\underline{u}'+\underline{e}',\underline{d}')$. So, the existence of B implies that A is not identified in (1).

References

Aigner, D.J., C. Hsiao, A. Kapteyn and T.J. Wansbeek (1984). Latent variable models in econometrics, in: Z. Griliches and M.D. Intriligator (eds.), Handbook of Econometrics, vol. 2, North-Holland, Amsterdam.

Bekker, P.A. (1986). Comment on identification in the linear errors in variables model, Econometrica, 54, 215-217.

Bentler, P.M. (1983). Simultaneous equation systems as moment structure models: with an introduction to latent variable models. Journal of Econometrics, 22, 13-42.

De Boor, C. (1978). A practical guide to splines. Springer, Berlin.

Chui, C.K. (1988). Multivariate Splines. NSF Regional Conference Series in Applied Mathematics, nr. 54, Society for Industrial and Applied Mathematics, Philadelphia.

De Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. Journal of Econometrics, 22, 113-137.

De Leeuw, J. and J. van Rijckevorsel (1981). Non linear principal component analysis with B-splines. Research Note RR-81-02. Erasmus University Rotterdam.

Epps, T.W. (1993). Characteristic functions and their empirical counterparts: geometrical interpretations and applications to statistical inference. The American Statistician, 47, 1, 33-38.

Ferguson, T.S. (1958). A method of generating best asymptotically normal estimates with applications to the estimation of bacterial densities. Annals of Mathematical Statistics, 29, 1046-1062.

Feuerverger, A. and P. McDunnough (1981a). On some fourier methods for inference. Journal of the American Statistical Association, 76, 379-387.

Feuerverger, A. and P. McDunnough (1981b). On the efficiency of empirical characteristic function procedures. Journal of the Royal Statistical Society B, 43, 1, 20-27.

Geman, S., and C. Hwang (1982). Nonparametric maximum likelihood estimation by the method of sieves. The Annals of Statistics, 10, 2, 401-414.

Geraci, V.J. (1976). Identification of simultaneous equation models with measurement error. Journal of Econometrics, 4, 263-283.

Hsiao, C. (1976). Identification and estimation of simultaneous equation models with measurement error. International Economic Review, 17, 319-339.

Hsiao, C. (1983). Identification, in: Z. Griliches and M.D. Intriligator (eds.), Handbook of Econometrics, vol. 1, North-Holland, Amsterdam.

Kapteyn, A. and Wansbeek, T. (1983). Identification in the linear errors in variables model, Econometrica, 51, 1847-1849.

Lukacs, E.L. (1970). Characteristic functions, Griffin, London.

Merckens, A. and P.A. Bekker (1993). Identification of simultaneous equation models with measurement error: a computerized evaluation. Statistica Neerlandica, 47, 4, 233-244.

Powell, M. (1981). Approximation theory and methods. Cambridge University Press, Cambridge.

Schumaker, L.L. (1981). Spline functions basic theory. Wiley, New York.

Van Montfort, K., A. Mooijaart en J. De Leeuw (1987). Estimation with errors in variables: estimators based on third order moments, Statistica Neerlandica, 41, 4, 223-239.

Van Montfort, K., A. Mooijaart en J. De Leeuw (1989). Estimation of regression coefficients with the help of characteristic functions, Journal of Econometrics, 28, 267-278.

Van Rijckevorsel, J. (1987). The application of fuzzy coding and horsehoes in multiple correspondence analysis, DSWO Press, Leiden.

Winsberg, S. and J.O. Ramsay (1980). Monotonic transformations to additivity using splines. Biometrika, 67, 3, 669-674.

Wold, S. (1974). Spline functions in data analysis. Technometrics, 16, 1, 1-11.

Wold, H. (1982). Soft modelling: the basic design and some extensions. In: K.G. Joreskog and H. Wold (eds.), System under indirekt observation: Causality, structure, prediction. Amsterdam, North Holland Publishing Co.

Young, F.W. (1981). Quantitative analysis of qualitative data. Psychometrika, 46, 357-388.

Zangwill, W.I. (1969). Nonlinear programming. A unified approach. Englewood Cliffs, Prentice Hall.