JOURNAL OF APPLIED ECONOMETRICS, VOL. 11, 649–667 (1996)

COMPARING FOURIER AND TRANSLOG SPECIFICATIONS OF MULTIPRODUCT TECHNOLOGY: EVIDENCE FROM AN INCOMPLETE PANEL OF FRENCH FARMERS

MARC IVALDI

GREMAQ, Ecole des Hautes Etudes en Sciences Sociales, Place Anatole France, F–31042 Toulouse, France. Email: ivaldi@cix.cict.fr

NORBERT LADOUX

GREMAQ and Commissariat à l'Energie Atomique, 31–33 Rue de la Fédération, F–75752 Paris, France

HERVÉ OSSARD AND MICHEL SIMIONI

ESR, Institut National de Recherche Agronomique, BP 27, F–31326 Castanet-Tolosan, France

SUMMARY

Selecting a functional form for a cost or profit function in applied production analysis is a crucial step in assessing the characteristics of a technology. The present study highlights differences in the description of a technology which are induced by fitting a Fourier or a translog cost functions. On average, both forms provide similar information on the technology. However, estimation results and statistical tests tend to favour the Fourier specification. This is mainly because many trigonometric terms appear to be significant in our application. Accordingly, our results show that the Fourier cost function is able to represent a broader range of technological structures than the translog. The paper presents an application to the case of an incomplete panel of French farmers. In the process, three technical issues are addressed: the missing data problem, the choice of the order of approximation and the conditions ensuring asymptotic normality of Fourier parameters' estimates.

1. INTRODUCTION

When evaluating economic policies, micro data are useful not only because they allow us to measure their average effects but also because they can provide us with insights into their differentiated impacts, given the broad heterogeneity of individual responses. Although necessary, micro data are not sufficient to measure the dispersion of policy effects. With regard to this objective, the choice of approximations which are used to parameterize technological constraints or individual preferences is crucial. We re-examine this long-debated issue by fitting a Fourier and a translog cost function on individual data. Our results show on average, that these two forms provide equivalent descriptions of the technology. The Fourier form is, however, able to represent a broader range of cost structures, which is desirable for calibrating economic policies. Therefore our study sheds light on the application of seminonparametric methods in production analysis, which are often advocated on theoretical grounds but more seldom implemented.

Selecting a functional form for the cost or profit function in applied production analysis is an important step in assessing the characteristics of a technology. Until recently, most studies have

CCC 0883–7252/96/060649–19

Received 22 August 1991
Revised 26 March 1996

favoured functional forms achieving second-order approximations of any arbitrary, twice-differentiable cost function at a given point; this is the so-called Diewert flexibility. (See Diewert, 1971, 1974.) The translog form, which can be interpreted as a Taylor series expansion and which was introduced by Christensen, Jorgenson and Lau (1973), is by far the most popular of the Diewert-flexible forms. Nonetheless, the literature on applied production analysis and on estimation of misspecified models has raised some qualifications regarding the translog specification. First, Guilkey, Lovell and Sickles (1983), when comparing it with two other Diewert-flexible forms (the generalized Leontief and the generalized Cobb–Douglas), have shown that the translog outperforms these other forms. However, the former provides a reliable representation only if the technology is not too complex. (See also Lau, 1986, on this discussion.) Second, White (1980) has shown that, while second-order approximations allow us to attain any arbitrary function at a given point, nothing implies that the true function is consistently estimated at this point. (See also Florens, Ivaldi and Larribeau-Nori, 1991.)

The Fourier form introduced by Gallant (1981, 1982, 1984) is an answer to White's criticism. Indeed, the Fourier form possesses a global flexibility property: It can asymptotically approximate any continuous function, i.e. it achieves the so-called Sobolev-flexibility, which is a distinguishing feature. The Fourier specification allows us to approximate not only an arbitrary cost function but also any of its derivatives (up to a fixed order). Moreover, inferences drawn from estimates of the Fourier are not affected by specification errors. Since it is a combination of polynomial and trigonometric expansions, the Fourier form is indeed endowed with the property that the order of approximation can increase with the sample size. Thus, it provides an infinite dimensional parametrization, which Diewert-flexible forms cannot generate. Finally, by fixing the order of approximation in the expansion, estimation of the resulting Fourier may be performed through the usual parametric estimation methods. In this sense the methodology involved with the implementation of a Sobolev-flexible form is seminonparametric, according to the terminology introduced by Elbadawi, Gallant and Souza (1983).

The present study highlights the differences in the description of a technology that are induced by fitting a Fourier or a translog cost function. First the cost-cum-share systems derived from these forms are estimated on an incomplete panel of French farms over the period 1984–86. Second, discussion on the results focuses on three aspects of the technology: concavity, curvature of isoquants (as measured by Morishima elasticities of substitution), and cost complementarity. In practice we look at the dissimilarities in the empirical distributions of these technological characteristics, computed over the sample from estimates of the two forms. Since we do not know the true technology, which is the usual case in applied econometrics, our purpose is to show the losses or gains in using one form instead of another.

The results are as expected with regard to the literature on approximations we discussed earlier. On average, both forms provide similar information on the technology. However, estimation results and statistical tests tend to favour the Fourier specification. This is mainly because many trigonometric terms appear to be significant in our application. Accordingly, our results show that the Fourier cost function is able to represent a broader range of technological structures than the translog.

To estimate these cost functions, three technical issues must be addressed. First, one must take care of the missing data problem inherent in an incomplete panel by using an adequate estimation method. Second, the order of approximation for the Fourier must be selected. This choice is in part driven by the dimensionality of the specific application we consider, in which the technology is described by means of two outputs and three inputs. Third, to assess the significance of Fourier parameters, we propose to perform a bootstrap method, which should be regularly applied in similar circumstances. (See Efron and Tibshirani, 1986; Veall, 1992.)

To compare the Fourier and the translog cost functions, we could invoke recent results on the asymptotic properties of series estimators. (See Andrews, 1991; Gallant and Souza, 1991.) These properties link the choice of the order of approximation and the sample size. Our application indicates that, for a given order of approximation, the sample size required to achieve asymptotic normality is so large that it is seldom met in practice. Hence the two forms are compared in terms of their empirical economic implications, given that the quality of parameter estimates is judged by means of the bootstrap method.

The paper is organized as follows. Section 2 provides simplified representations of the Fourier and translog flexible forms which are useful for practical purposes. We focus on the link between the identification of the Fourier parameters and the choice of the order of approximation given the characteristics of the technology. Then it is shown that a translog form is nested into a Fourier form. Section 3 is devoted to the estimation procedure and its results. The estimation procedure combines techniques for incomplete panels without time effects (see Baltagi, 1985; Wansbeek and Kapteyn, 1989; Verbeek and Nijman, 1992, among others) with Zellner's (1962) seemingly unrelated regression framework. Lastly, we discuss the accuracy of parameter estimates from results of a bootstrap technique. Section 4 compares the distributions of technological characteristics derived from the Fourier and translog estimates. Section 5 offers some conclusions.

2. THE FOURIER AND TRANSLOG FLEXIBLE FORMS

First, we present the Fourier flexible form introduced by Gallant (1981,1982) in a way suitable for empirical implementation. Then we show how the translog is nested in a Fourier and how a certain order of approximation can be chosen in practice.

Consider a logarithmic transformation of a cost function

$$\ln c = g(w, y) \tag{1}$$

where w is a vector of logarithms of factor prices and y the vector of logarithms of output quantities. There are I inputs and J outputs. So the vector $z = (w', y')$ has dimension $P = I + J$. The cost function satisfies positive linear homogeneity, monotonicity and concavity in input prices. By Shepard's lemma the gradient of g with respect to w, denoted by

$$s = (s_1, \ldots, s_i, \ldots, s_I)' = \nabla_w g(w, y) = \nabla_w g \tag{2}$$

gives the system of factor cost shares. From the usual definition of price elasticities given in numerous papers (see Blackorby and Russell, 1989, for instance) it is easy to prove that own- and cross-price elasticities can be written

$$\varepsilon_{ii} = \frac{H_{ii} + s_i(s_i - 1)}{s_i} \qquad i = 1, \ldots, I \tag{3}$$

and

$$\varepsilon_{ij} = \frac{H_{ij} + s_i s_j}{s_i} \qquad \forall i \neq j \tag{4}$$

respectively, where $H = [H_{ij}] = \nabla^2_w g$ is the Hessian associated to the function g with respect to w. As pointed out by Blackorby and Russell (1989), the Allen elasticity does not provide a meaningful measure of curvature of isoquants. In this respect one can use the Morishima elasticity between inputs i and j which can be computed as

$$m_{ij} = \varepsilon_{ji} - \varepsilon_{ii} \tag{5}$$

2.1. The Fourier Cost Function

The Fourier cost function is defined as[1]

$$g_K(z/\theta) = a_0 + b'z + \tfrac{1}{2} z'Cz + \sum_{a=1}^{A} \{u_{0a} + [u_a \cos(k'_a z) - v_a \sin(k'_a z)]\} \tag{6}$$

where θ is the vector of parameters, a_0, $b' = (b'_w, b'_y)$, $\mathrm{vec}(C)'$, $(u_{0a}, u_a, v_a)_{a=1}^{A}$a = 1.[2] Equation (6) is an approximation to the order K of the true cost function when the sum of the absolute values of the components of each vector k_a is less than or equal to K. The so called multi-index k_a is a P-vector with integer components satisfying the condition that the first non-zero element must be strictly positive. The number of such multi-indexes is denoted by A, which is clearly a function of K For instance, consider the case with $K = 3$ and $P = 2$; then it is straightforward to check that $A = 12$ since the set of admissible multi-indexes is

$$\{(0,1)',(0,2)',(0,3)',(1,0)',(2,0)',(3,0)',(1,1)',(1,2)',(1,-1)',(1,-2)',(2,1)',(2,-1)'\}$$

As shown by Gallant (1982, theorem 2, p. 297), the Fourier approximation is flexible in terms of the Sobolev norm (which explicitly considers approximation errors on the true function and its derivatives as well)when the matrix C in equation (6) is determined by the system

$$C \equiv -\sum_{a=1}^{A} u_{0a} k_a k'_a \tag{7}$$

Using this last expression and equation (2), the parametric form of the factor cost shares is

$$s = b_w - \sum_{a=1}^{A} \{u_{0a} k'_a z + [y_a \sin(k'_a z) + v_a \cos(k'_a z)]\} k_a \tag{8}$$

while, accordingly, the P by P Hessian matrix of the cost function is

$$H = \sum_{a=1}^{A} \{u_{0a} + [u_a \cos(k'_a z) - v_a \sin(k'_a z)]\} k_a k'_a \tag{9}$$

which is symmetric as C is symmetric by definition.

Finally, for $g_K(z/\theta)$ to be a cost function, one must impose homogeneity with respect to the I first components of vector z, i.e. the factor prices. This condition holds when

$$\sum_{i=1}^{I} b_{w,i} = 1 \qquad \text{and} \qquad u_{0a} = u_a = v_a = 0 \qquad \text{for any } a \text{ such that} \sum_{i=1}^{I} k_{ai} \neq 0 \tag{10}$$

[1] When estimating the Fourier form, the vector z must be scaled by a factor in order to normalize data which must fall on the interval $]0,2\pi[$, according to a procedure proposed by Gallant (1981).

[2] $\mathrm{vec}(C)$ puts the non-redundant elements of a symmetric matrix C in a column vector. Note also that b is a column vector of dimension I, which is decomposed in two components of dimensions I and J corresponding to inputs and outputs, respectively.

Then imposing homogeneity reduces the number of admissible multi-indexes. In the example above for $K = 3$ and $P = 2$, the only remaining index is $(1, -1)'$.[3]

2.2. The Translog Cost Function

From the Fourier cost function (6), the translog is obtained by deleting all u_a and v_a since in this case, equation (6) is just a quadratic form in the logarithms of input prices and output quantities Its expression is

$$g_K(z/\theta) = b_0 + b'z + \tfrac{1}{2}z'Cz \tag{11}$$

where the constant term is now $b_0 = a_0 + \sum_{a=1}^{A} u_{0a}$. The matrix C, defined as above, is here the Hessian matrix and it must contain $(P(P-1)/2)$ distinct elements when homogeneity is imposed, for the translog to be flexible. (See Fuss, McFadden and Mundlak, 1978.)

2.3. The Choice of the Order of Approximation

When the order of approximation is large enough, system (7) is overidentified. For instance, when $K = 4$ and $P = 3$ (three factors and no output), then $A = 9$ (i.e. the number of parameters u_{0a}) while there are only three *free* parameters to be estimated (since C contains six distinct components).[4] Then one must set some of the parameters u_{0a} to be equal to zero. Notice that, in practice, any u_{0a} could be chosen as far as the remaining ones correspond to the independent columns of the $(P(P+1)/2)$ by A matrix Δ in the following system:

$$\operatorname{vec}(C) = \Delta u \tag{12}$$

where $u = (u_{01}, \ldots, u_{0A})'$.

Performing this task is equivalent to retaining only the parameters u_{0a} that do correspond to multi-indexes which can be used to build up a set of matrices of the form $k_a k_a'$, spanning the vector subspace of P by P symmetric matrices written as

$$H_A = \sum_{a=1}^{A} \mu_a k_a k_a' \tag{13}$$

Given equation (9), this basis of matrices allows us to generate the Hessian matrix of the translog cost function if one sets

$$\mu_a = -\mu_{0a}$$

and the Hessian of the Fourier form if

$$\mu_a = -\{u_{0a} + [u_a \cos(k_a'z) - v_a \sin(k_a'z)]\}$$

Therefore the approximation of a cost function requires a number of multi-indexes at least equal to the number of distinct elements of the Hessian matrix. This rule is trivial for the translog case. With the Fourier form, however, the rule can be violated while the number of parameters exceeds the number of distinct elements of the Hessian matrix. Consider a case with three inputs and two outputs, i.e. $P = 5$. Taking a second-order approximation for the Fourier, one obtains nine admissible multi-indexes under the homogeneity condition.[5] Then it is not

[3] When $P = 2$ one considers a unit cost function where output quantities do not appear. i.e. when returns-to-scale are constant.

[4] Under the homogeneity condition,the multi-indexes are $(1,0,-1)'$, $(1,-1,0)'$, $(0,1,-1)'$, $(1,1,-2)'$, $(2,-1,-1)'$, $(1,-2,1)$, $(2,0,-2)'$, $(2,-2,0)'$, and $(0,2,-2)'$.

enough to identify the parameters of the second-order terms of the corresponding translog cost function, i.e., the ten distinct elements of matrix C. Moreover, while the second-order Fourier approximation contains 27 second-order parameters (nine u_{0a}, nine u_a and nine v_a) and six first-order parameters, the number of muti-indexes is not sufficient to generate a full rank Hessian matrix, i.e., a flexible cost function. Therefore a second-order approximation is not sufficient and we need at least a third-order approximation.[6] Hence given equation (11), the translog is nested in a third-order approximation for the Fourier. In this case, the number of multi-indexes is $A = 27$.[7] Thus the translog can be viewed as a third-order *constrained* approximation of the true cost function in terms of the Sobolev norm.

This discussion clearly indicates that the choice of order of approximation is first driven by identification purposes. We return on this point later.

3. ESTIMATION PROCEDURE AND RESULTS

Annual accounting data on French farms are regularly collected in a survey called *Réseau d'Information Comptable Agricole* (RICA).[8] From the three consecutive surveys of the years 1984 to 1986 we extract an unbalanced panel of fruit producers on the basis of two criteria: for each farm the production of apples must be different from zero[9] and the productive acreage of the orchard must be greater than 5 acres. The missing data problem is addressed below.

Our objective is to recover the production structure of these farms by means of a Fourier or translog cost function. The technology combines three aggregated inputs, namely capital (including land), labour and materials to produce two outputs, the production of apples and an aggregate of other productions. Given the selection criteria of our sample defined above, the quantity of apples produced by each farmer is always positive. The level of production of the second output, which combines mainly various types of fruit and some secondary products, is also strictly positive for all individuals in the sample. This is not due to the chosen selection criteria. Given that the survey covers a large proportion of existing French farms producing apples, one must simply assume that there are no French farms specializing in apple production exclusively. Two reasons may explain that there is no zero productions. On the one, there is an agricultural constraint on rotating the types of production on the same land. On the other hand, diversification is considered as one of the methods to manage uncertainty in agriculture.

Based on some implicit separability assumptions, measures of input prices and output quantities are obtained by aggregation using the technique of Törnqvist indexes whenever needed. Indeed, the surveys provide very disaggregated data on all types of inputs and outputs. (See Appendix 2 for data description and variable definitions.) Cost and share variables are defined according to this aggregation. Finally, price and quantity indexes are rescaled according to a procedure proposed by Gallant (1982) for estimating a Fourier cost function.

[5] These nine multi-indexes are $(0,0,0,1,0)'$, $(0,0,0,0,1)'$, $(0,1,-1,0,0)'$, $(1,0,-1,0,0)'$, $(1,-1,0,0,0)'$, $(0,0,0,1,-1)$, $(0,0,0,1,1)'$, $(0,0,0,2,0)'$, and $(0,0,0,0,2)'$. Notice that the first three components of these indexes refer to the three input prices.

[6] Note that, if the true cost function is assumed to be homothetic, a second-order approximation generates flexible forms. Indeed it is the only type of cost Function that can be produced with the list of multi-indexes given in footnote 3. This can be seen by applying the rule for identifying system (12) and expanding the expression of the Hessian matrix. To save space, we do not report the formal proof of this statement.

[7] The list of the 27 multi-indexes is given in Appendix 1.

[8] Note that similar databases exist for many countries of the European Union.

[9] This criterion is used to select producers of apples in the population of fruit producers. Note that there is no farmer producing only apples. Obviously selecting producers of apples prevents us extending the results of the analysis to other farmers.

The Fourier cost function combines second-order polynomial terms and trigonometric functions. Eubank and Speckman (1990,1991) show that polynomial-trigonometric regressions have attractive features compared to non-parametric methods. In addition to permitting the use of standard regression methodologies, this type of approximation alleviates boundary problems and their associated estimators are able to reach optimal rates of convergence for any sufficiently smooth regression function. The estimation of the Fourier cost function presents similar advantages that we now discuss.

3.1. Estimation Procedure

As usual in applied production analysis, after having added disturbance terms to the cost function (1) and each of the factor cost shares defined by (2), the econometric task consists of estimating the cost-cum-share system obtained by stacking the cost function and (I-1) factor shares. One share equation is omitted because, otherwise, the covariance matrix of the errors on the share system would be singular due to the adding-up constraint. Using standard notation in econometrics, the model to be estimated is the seemingly unrelated regression system

$$y_{mit} = x_{mit}'\beta m + \mu_{mi} + \varepsilon_{mit} \tag{14}$$

where y and x denote endogenous and exogenous variables respectively, and where m, i, and t index equations ($m = 1, \ldots, M$), individuals ($i = 1, \ldots, N$) and time periods ($t = 1, \ldots, T_i$), respectively. Given that we will be dealing with incomplete panel data, the time length T_i is specific to individual i since it gives the number of consecutive periods for which data on this individual are not missing.[10]

Observations across individuals are assumed to be independent. The individual effects μ_{mi} are time-invariant random variables distributed independent across individuals. Moreover,

$$\mathrm{E}(\mu_{mi}) = 0 = \mathrm{E}(\varepsilon_{mit}) \qquad \text{for any } m,i,t \tag{15}$$

and

$$\mathrm{E}\left(\begin{bmatrix}\mu_{gi}\\ \varepsilon_{git}\end{bmatrix}[\mu_{hj} \;\; \varepsilon_{hjs}]\right) = \begin{bmatrix}\sigma^2_{gh}(\mu)\delta(i=j) & 0\\ 0 & \sigma^2_{gh}(\varepsilon)\delta(i=j)\delta(t=s)\end{bmatrix} \tag{16}$$

for any g, h, i, j, t, s, where $\delta(i=j) = 1$ if $i = j$ and 0 otherwise. Finally, regressors and disturbances are orthogonal.

Given these stochastic assumptions, the usual estimation procedure follows Chamberlain's (1984) methodology which consists of estimating a multivariate system obtained as follows. Stacking observations by individual (with the time index t going fast and the equation index m going slow), the model becomes

$$y_i = X_i\beta + \mu_i \otimes e_i + \varepsilon_i \tag{17}$$

where μ_i, is the $M \times 1$ vector of individual effects, e_i is a $T \times 1$ vector of ones, the matrix X_i contains the observations for individual i on all the non-redundant exogenous variables entering in the M equations, β is the corresponding vector of parameters, and the symbol $\otimes$ designates

[10] Our procedure does not treat the cases where an individual quits the survey then re-enters after several periods. A crude treatment of such cases is to consider an individual before quitting and after re-entry as two separate observations. Since such cases happen only twice in the application below, we believe that this treatment will not have a significant impact on the estimation results.

the Kronecker product. Given the definition of X_i and β, this setup allows us to easily impose cross-equation restrictions when it is necessary as in our application.

Let $u_i = \mu_i \otimes e_i + \varepsilon_i$ *and* Σ_ε (resp. Σ_μ) be the $M \times M$ covariance matrix with generic element $\sigma^2_{gh}(\varepsilon)$ (resp. $\sigma^2_{gh}(\mu)$. The covariance matrix of u_i is

$$\Omega_i = \Sigma_\varepsilon \otimes Q_i + \Sigma_i \otimes (J_i/T_i) \tag{18}$$

where $Q_i = I_i - (J_i/T_i)$, with $J_i = e_i e_i'$ and I_i being the $T_i \times T_i$ identity matrix, and where Σ_i is a matrix of generic term $\sigma^2_{gh}(\varepsilon) + T_i\sigma^2_{gh}(\mu)$. As is well known, equation (18) can be viewed as the main step in the decomposition of Ω_i with respect to its eigenvalues. (See Nerlove, 1971.) From a result by Baltagi (1980), powers of matrix Ω_i are given by

$$\Omega_i^n = \Sigma_\varepsilon^n \otimes Q_i + \Sigma_i^n \otimes (J_i/T_i) \tag{19}$$

The Generalized Least Squares estimator of β is defined as

$$\beta_{\text{GLS}} = (\Sigma_{i=1}^n X_i'\Omega_i^{-1}X_i)^{-1}\Sigma_{i=1}^N X_i'\Omega_i^{-1}y_i \tag{20}$$

The practical implementation of this estimator is not described here.[11]

Concerning the properties of this estimator, which are well known, the classical presentation we have followed should not be misleading. What we stress here is that the Fourier flexible form allows us to apply standard regression estimators. However, the main advantage of the Fourier is that it has a variable number of parameters (as a function of the approximation order) depending on the sample size. Then the asymptotic properties of the GLS estimator should be related to the choice of the approximation order of the Fourier form.

With respect to this issue, theorem 5 of Gallant and Souza (1991, pp. 345–346) provides a necessary and sufficient condition for ensuring asymptotic normality of any linear combination of Fourier parameter estimates. However, when this condition is checked on our data set, we found that several thousands of observations are needed if we consider a Fourier at a third-order approximation. As the Gallant and Souza condition has been derived under the assumption of homoscedastic data, in principle it applies only to panel data where observations are assumed to be independent across individuals (which is the maintained hypothesis of our study). However, we conjecture that the extension of this condition to panel data under more general stochastic assumptions should not rule out a similar restrictive condition. Now, data sets with a sample size large enough for the condition to hold are rare in applied production analysis.

To bypass this difficulty we propose first to use bootstrap techniques in order to assess the distribution of GLS estimates of the Fourier and translog cost functions, and then to compare the two forms on their empirical economic implications.

3.2. Estimation Results

Depending on the choice of cost function (Fourier or translog), a computer program (using a GAUSS code) allows us (1) to construct the matrix of elementary multi-indexes and to select the subset of indexes for the homogeneity condition to hold; (2) to eliminate some multi-indexes in order to identify the matrix C (see the discussion in Section 2.3); (3) to scale the data for the variables to take values on $]0,2\pi[$; and (4) to construct the SUR system and to implement the iterated GLS estimation.

In the two outputs–three inputs case, the Fourier form involves 69 parameters if a third-order approximation (i.e. $K = 3$) is chosen.[12] The corresponding nested translog form is described by

[11] Computer programs are available from the authors.

[12] Recall that here $A = 27$. Then there are 27 u_a, 27 v_a, 10 u_{0a} (since the matrix C has 10 distinct elements), 5 first-order terms and the constant. The list of the 27 multi-indexes is given in Appendix 1.

15 parameters. For a fourth-order approximation the number of parameters for the Fourier form is as high as 146. We deemed this number to be too high, given the structure of our panel data set (see Table 1). Accordingly, we do not report the results when $K = 4$. Eastwood and Gallant (1991) suggest a procedure for choosing statistically the order of approximation. However, our application shows that this choice is very limited in practice given the need to identify the Fourier cost function parameters (see our discussion in Section 2.3) and given the number of observations available.

According to Table I, only 92 farmers report their financial data for each year. In our context, addressing the missing data problem is certainly a source of efficiency gain. Comparing the estimators applied to unbalanced panels with the estimators applied to balanced panels offers a way to detect the possibility of 'unbalancedness' bias. Verbeek and Nijman (1992) propose such a procedure based on a version of the Hausman test. Applied to our case, we cannot reject the null hypothesis of no bias (the Chi-square statistics are 25·7 and 871·7 for the translog and the Fourier case, respectively). Therefore we are required to deal with the missing data problem. As suggested by an anonymous referee, this result may also indicate the presence of attrition bias. However, because of the surveying process, the reasons explaining why a farmer is not surveyed a certain year are in general not related to the economic conditions of the farmers production activity. Indeed the RICA ensures that the survey is random. Therefore we adopt the estimation procedure presented above in order to use all the potential statistical information.

Estimation results are presented in Table II. Given the very large difference in the number of parameters between the two forms we are not surprised that a Wald test leads to a strong rejection of the translog.[13] However, the asymptotic properties of Wald tests are not known here since they depend on the order of approximation of the Fourier (see the above discussion of the asymptotic normality of GLS estimates). This is in part why we compare later the two forms on the basis of their abilities to adequately represent the technology.

First, one must assess the quality of the estimates of the Fourier and translog cost functions. On the basis of standard errors (computed in the usual way), nineteen parameters of the Fourier form are significant in a two-tailed test at the 5% level, and many more at the 10% level. In particular, many trigonometric terms have a significant coefficient. For the translog form 11 (resp. 13) parameters are significant at the 5% (resp. 10%) level.

In order to confirm or reject this conclusion we construct confidence intervals for the estimated t-ratios associated to the Fourier parameters by performing a bootstrap. Following Eakin, McMillen and Buono (1990) we generate 2500 samples by random draws with replacement in our data set. In order to maintain our assumption of independent observations across individuals, we define a data point as the observed values (including missing ones) for a given individual over three periods and we resample data points so defined. By this procedure we retain the temporal pattern of individual behaviours and the structure of missing data. Hence, our stochastic assumptions are not modified by the chosen sampling.

Table I. Structure of the panel of farmers (sample sizes)

Year	Unbalanced panel	Missing cases	Balanced panel
1984	130	54	92
1985	135	49	92
1986	140	44	92

[13] The value of the Wald test statistic is not reported but is available upon request.

Table II. Estimation results

	Fourier estimates				Fourier estimates			Translog estimates				Fourier estimates	
Coeff.	Value	*t*-ratio	*ks* test	Prob.	Value	*t*-ratio	Coeff.	Value	*t*-ratio	*ks* test	Prob.	Value	*t*-ratio
a_0	6·301	1·62	0·016	0·16	12·290	354·15[a]	u_{12}	−0·361	−2·97[a]	0·014	0·25	—	—
b_k	0·090	6·13[a]	0·009	0·90	0·203	27·38[a]	v_{12}	−0·108	−1·74	0·023	0·00	—	—
b_1	0·321	10·39[a]	0·010	0·87	0·697	52·61[a]	u_{013}	−0·008	−2·39[a]	0·011	0·63	—	—
b_{y1}	−1·180	−1·78	0·014	0·33	0·107	9·84[a]	u_{13}	−0·180	−4·83[a]	0·016	0·12	—	—
b_{y2}	4·670	1·83	0·014	0·28	0·063	5·42[a]	v_{13}	−0·005	1·16	0·016	0·14	—	—
u_{01}	−0·443	−2·36[a]	0·025	0·00	0·014	1·86	u_{14}	−0·042	−0·64	0·013	0·36	—	—
u_1	−0·184	−0·50	0·017	0·10	—	—	v_{14}	0·059	1·07	0·016	0·15	—	—
v_1	−1·222	−2·64[a]	0·015	0·23	—	—	u_{015}	−0·001	−0·28	0·012	0·52	0·002	0·97
u_{02}	1·352	1·70	0·015	0·23	−0·011	−1·04	u_{15}	0·007	2·39[a]	0·013	0·39	—	—
u_2	3·782	1·90	0·019	0·04	—	—	v_{15}	−0·004	−1·25	0·024	0·00	—	—
v_2	−2·368	−2·94[a]	0·021	0·01	—	—	u_{16}	−0·005	−1·03	0·023	0·01	—	—
u_{03}	−0·056	−0·62	0·017	0·10	0·141	2·70[a]	v_{16}	−0·001	−0·23	0·016	0·12	—	—
u_3	0·979	2·68[a]	0·014	0·32	—	—	u_{17}	0·004	1·43	0·019	0·03	—	—
v_3	−0·118	−0·70	0·014	0·31	—	—	v_{17}	−0·006	−1·57	0·011	0·65	—	—
u_{04}	−0·005	−0·41	0·013	0·40	0·088	12·63[a]	u_{18}	−0·003	−1·20	0·014	0·25	—	—
u_4	−0·007	0·66	0·011	0·65	—	—	v_{18}	0·002	0·86	0·018	0·05	—	—
v_4	0·004	0·29	0·010	0·84	—	—	u_{19}	−0·004	0·96	0·013	0·37	—	—
u_{05}	−0·008	−1·90	0·011	0·61	0·034	6·71[a]	v_{19}	0·005	1·19	0·010	0·84	—	—
u_5	0·001	0·16	0·014	0·30	—	—	u_{20}	−0·003	−0·71	0·015	0·23	—	—
v_5	−0·008	−1·02	0·012	0·57	—	—	v_{20}	0·002	0·57	0·013	0·41	—	—
u_{06}	0·007	−3·48[a]	0·020	0·02	0·018	5·48[a]	u_{21}	−0·003	−0·99	0·015	0·21	—	—
u_6	−0·007	−1·19	0·014	0·32	—	—	v_{21}	−0·006	−2·15[a]	0·009	0·90	—	—
v_6	−0·005	0·80	0·010	0·82	—	—	u_{22}	0·308	3·14[a]	0·022	0·01	—	—
u_7	−0·332	−1·37	0·017	0·08	—	—	v_{22}	0·304	3·10[a]	0·010	0·81	—	—
v_7	−0·891	−2·97[a]	0·017	0·07	—	—	u_{23}	0·057	0·98	0·015	0·22	—	—
u_{08}	0·000	0·09	0·019	0·04	0·001	0·54	v_{23}	0·065	1·34	0·017	0·10	—	—
u_8	−0·012	−2·52[a]	0·011	0·61	—	—	u_{24}	−0·227	−2·63[a]	0·016	0·13	—	—
v_8	−0·004	−0·92	0·016	0·16	—	—	v_{24}	0·128	1·11	0·011	0·70	—	—
u_{09}	−0·015	−2·89[a]	0·021	0·01	0·007	2·89[a]	u_{25}	0·152	0·46	0·019	0·04	—	—
u_9	−0·001	−0·18	0·017	0·10	—	—	v_{25}	−1·281	−2·74[a]	0·023	0·00	—	—
v_9	0·002	0·43	0·010	0·86	—	—	v_{25}	−1·281	−2·74[a]	0·023	0·00	—	—
u_{10}	0·002	0·37	0·018	0·07	—	—	u_{26}	−0·050	−1·88	0·017	0·08	—	—
v_{10}	−0·004	−0·97	0·017	0·10	—	—	v_{26}	0·021	0·86	0·011	0·59	—	—
u_{11}	−0·006	−0·96	0·014	0·25	—	—	v_{27}	−0·075	-1·47	0·020	0·02	—	—
v_{11}	0·008	1·31	0·033	0·00	—	—	v_{27}	−0·161	-1·51	0·015	0·18	—	—

[a] A parameter significant at the 5% level.

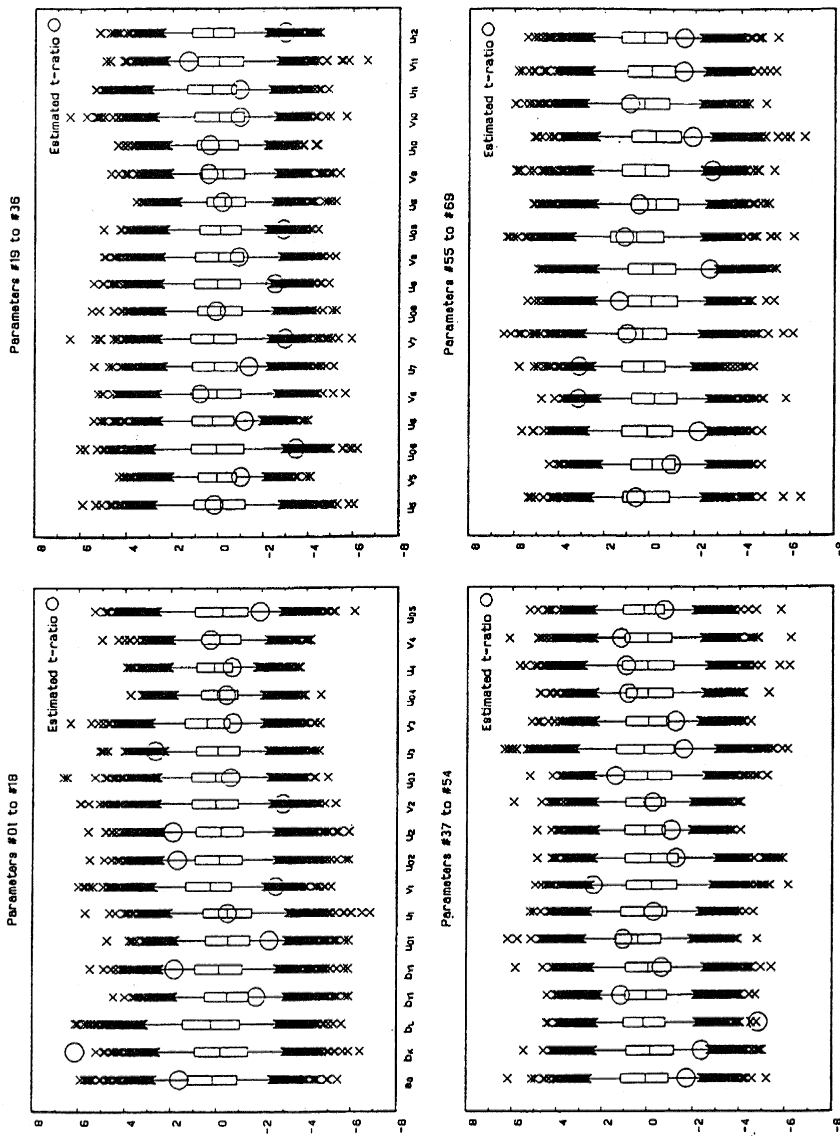

Figure 1. Box plots of bootstrap t-ratios. (see footnote 14 for a definition of box plots)

For each sample we perform the GLS estimation and compute the t-ratios for all the parameters. From the empirical distributions of these bootstrap t-ratios we can then assess the validity of the estimated t-ratios: a parameter will be significant if its estimated t-ratio does not belong to the 95% confidence interval of the empirical distribution of the corresponding bootstrap t-ratios. Figure 1 reports box plots where the estimated t-ratios of parameters are represented by a circle while the bootstrap t-ratios belonging to the significance region are represented by crosses.[14] Amongst the 19 parameters that appeared to be significant from the GLS estimation, significance is confirmed for 14; 4 parameters (namely, u_{013}, u_{15}, v_{21}, and u_{24}) are on the edge of the significance region; one parameter, u_{01}, is no longer significant. For the remaining parameters rejection is confirmed.

For the sake of completeness, the Kolmogorov–Smirnov statistics for testing the normality of the distribution of boostrap standardized coefficients, (i.e. t-ratios) are given in Table 2. The upper-tail probabilities computed according to a formula proposed by Lilliefors (1967) are also provided. The normality is almost never rejected. This implies that t-ratios have asymptotically a normal distribution, which justifies inferences with these statistics.

3.3. Comments

Recall that the estimated equation for the Fourier cost function is, for a typical observation,

$$g_k(z_{it}/\theta) = a_0 + b'z_{it} + \tfrac{1}{2} z'_{it}Cz_{it} + \sum_{a=1}^{A} \{u_{0a}\cos(k'_a z_{it}) - v_a \sin(k'_a z_{it})]\} + \mu_i + \varepsilon_{it} \qquad (21)$$

Our results show that some trigonometric terms are significant. The Fourier can be viewed as a translog with an heteroscedastic term composed of trigonometric expansions on input prices, an individual effect and the usual disturbance term representing measurement error on the dependent variable. It is worth noting that in applied production analysis using panel data, and in particular in the literature on frontier models (see Cornwell, Schmidt and Sickles,1990, for instance), it is common to posit that the error term is correlated with input prices and output quantities.

4. EMPIRICAL ANALYSIS OF THE COST STRUCTURE

Although the above discussion seems to favour the Fourier over the translog from a statistical point of view, it is of interest to compare the descriptions of the technology they reveal. The true technology is of course unknown, so that the purpose of this comparison is simply to illustrate the different policy perspectives that we can draw from the two estimated cost functions. Our comparison focuses on three aspects of the technology: Concavity, substitution effects and cost complementarity.

4.1. Concavity

Since the Fourier is more flexible than the translog, one could expect to reject more often concavity of the cost function with Fourier than with translog. Previous empirical applications have already noticed this point. (See Devezeaux, Ivaldi and Ladoux, 1990; Chalfant and

[14] A horizontal line is drawn through the box at the median of the estimated elasticities, the upper and lower ends of the box are the upper and lower quartiles, and vertical lines go up and down from the box to the extremes of the estimated elasticities. The elasticities that are very extreme are plotted by themselves.

Wallace, 1991.) In our study, concavity is only violated for two observations corresponding to two different individuals at two different periods, for the Fourier case. Hence, concavity is not a relevant criterion for choosing between the two forms, since these two exceptions are acceptable given the heterogeneity of the sample.

4.2. Substitution Effects

Using expressions (4) and (5), the Morishima cross-elasticities of substitution and the own-price elasticities are computed from our estimates for each individual at each period. Their standard errors are also obtained by means of an approximation theorem on non-linear combinations of normal distributions due to Serfling (1980). This theorem can be applied here since the bootstrap has shown that distributions of parameters are well approximated by normal distributions. These standard errors are very small and therefore almost all elasticities are highly significant. They are not reported and instead, Table 3 gives the empirical means and standard deviations of all elasticities.

For each elasticity the empirical means obtained from the Fourier and the translog are very close. However, the standard deviations are two to four times larger for the Fourier case than for the translog. So the Fourier estimates are able to represent a larger range of technological and/or behavioural characteristics than the translog. Recall that the Fourier elasticities have widespread empirical distributions because they vary not only with the factor shares as in the translog case but also with the curvature of the cost function due to the presence of the trigonometric terms. (This can be observed by inspecting equations (7) and (9).) This result illustrates the fact that the second derivatives for the translog cost function are constant. This property cannot be justified on an *a priori* basis, even though it is sufficient for ensuring Diewert flexibility, i.e. to ensure a good local approximation. As correct measures of the curvature of the cost function,

Table III. Empirical means and standard errors of elasticities[a]

Year	1984		1985		1986	
Form	Translog	Fourier	Translog	Fourier	Translog	Fourier
Capital	−0·493	−0·496	−0·489	−0·482	−0·501	−0·507
	0·050	0·109	0·052	0·114	0·058	0·120
Labour	−0·214	−0·203	−0·199	−0·195	−0·213	−0·225
	0·022	0·070	0·027	0·068	0·025	0·075
Material	−0·287	−0·286	−0·278	−0·284	−0·283	−0·297
	0·016	0·081	0·031	0·114	0·022	0·100
Capital–labour	0·591	0·569	0·580	0·572	0·603	0·621
	0·078	0·166	0·093	0·171	0·083	0·176
Labour–capital	0·486	0·482	0·492	0·503	0·495	0·536
	0·052	0·189	0·060	0·185	0·058	0·195
Capital–material	0·610	0·584	0·590	0·574	0·620	0·624
	0·079	0·137	0·093	0·143	0·083	0·144
Material–capital	0·510	0·504	0·473	0·475	0·498	0·504
	0·062	0·114	0·074	0·128	0·083	0·129
Labour–material	0·386	0·403	0·375	0·404	0·373	0·416
	0·034	0·156	0·046	0·165	0·048	0·166
Material–labour	0·405	0·418	0·386	0·407	0·390	0·419
	0·045	0·142	0·059	0·157	0·065	0·157

[a] In each cell, the figure on the first line is the mean, the second one being the standard error.

the Morishima elasticities computed from the Fourier indicate that the 'ease of substitution' can vary on a larger range than with the translog.

From Table 3, capital, labour and material appear to be substitutes on average, whatever the forms chosen, since all Morishima elasticities are positive.[15] Note that, from equation (5), the cross-price elasticity is positive (resp. negative) if the Morishima elasticity is greater (resp. smaller) in absolute values than the own-price elasticity.[16] Given the average values reported in Table 3, the cross-price elasticities among the three inputs are all positive.

The results are useful to design an environmental policy, for instance. Based on the values of elasticities between capital and materials, a policy which aims at reducing the use of fertilizers or herbicides in the sector of fruit production by increasing taxes on materials would have the drawback of increasing production costs.[17] Since materials and capital are shown to be substitutes, this negative effect could be counterbalanced by favouring a decrease in interest rates to lower capital price.

Since wage rates adjust very slowly in France, and since the values of the own-price elasticities are small, the one for capital being the highest, a general decrease in interest rates appears to be an appropriate way to reduce marginal costs. If, at the same time, fallows are encouraged, one may expect a decrease in costs without an increase in production as land belongs to capital. (As a well-known attitude of farmers is to idle their least productive land, one may be doubtful of the efficiency of this policy without efficient monitoring, for instance by scrutinizing the land use from satellite observations. The revenue of the farmer could be maintained even if product prices are no longer subsidized. This would be in the interest of consumers. But such an analysis requires the knowledge of the demand side, which is beyond the scope of this paper.

4.3. Cost Complementarity

Table 4 displays some statistics on the distributions of changes in marginal cost of the main product (i.e. apples) with respect to the quantity of the other products. The difference between Fourier and translog is again striking. First, the cost complementarity estimates with the Fourier show a larger range of changes than with the translog. Second, while the marginal cost of apples is decreasing with the quantity of other products for all farmers and all periods under the translog case,[18] it is decreasing for about half of the sample in the Fourier case and increasing for the other half in each period. Hence, on the one hand, the translog approximation indicates that the technology favours diversification of the production; on the other, with the Fourier, there is a non-negligible subset of the sample for which specialization in the production of apples should be an optimal behaviour.

These results seem to suggest that, on average, the degree of specialization is not high enough in the sector of apple production. But this is questionable if one takes into account uncertainty in the cost analysis of agricultural structures. While full specialization often appears to be an optimal strategy *ex post* when output prices are known, farmers must choose their technology

[15] Note also that, by pairwise comparison and computing the usual statistics for comparing means, the Morishima elasticities are not symmetric.

[16] For a very few cases, the cross-price elasticities are negative. But this is negligible.

[17] With respect to environmental policies, governments of OECD countries have been always very reluctant to impose taxes on polluting inputs. Recently however, this type of policy has received attention from the European Commission. (See Abler and Shortle, 1993.)

[18] Changes in marginal cost under the translog case are close to zero but always negative. Recall that only the signs of these changes matter for the analysis of cost complementarity.

Table IV. Statistics on cost complementarity

Year	1984		1985		1986	
Form	Translog	Fourier	Translog	Fourier	Translog	Fourier
Mean	−0·00032	0·00012	−0·00031	−0·00021	−0·00032	0·00038
Standard deviation	0·00049	0·00340	0·00033	0·00289	0·00048	0·00350
Minimum	−0·00386	−0·01242	−0·00335	−0·01218	−0·00492	−0·01584
Maximum	−0·00011	0·01981	−0·00011	0·00957	−0·00011	0·00965

ex ante, and they usually prefer not to be over-specialized since they are risk adverse. For this reason, results obtained with the Fourier may be of some help in characterizing which farms are too much or not enough specialized and in improving the management of farms. This is of interest given the new path of the European agricultural policy since a better understanding of individual production conditions is required to give to the farmers the correct incentives for producing more efficiently.

5. CONCLUSIONS

The large degree of flexibility of the Fourier form has been proved in several theoretical studies. This suggests that the Fourier is particularly suitable for estimating cost functions on panel data sets which usually contain variables characterized by large variances. For this type of data, traditional flexible forms should not be adequate, as they are known to well approximate the true technology only in a neighbourhood of a given point.

In this paper we compare the Fourier to the translog, which is the most commonly used Diewert-flexible form. The data used for the purpose of estimation are drawn from an incomplete panel of fruit producers, extracted from a survey of the French agriculture. Recall that our estimation procedure allows us to treat missing observations, i e. farmers for whom data are not available for some years of the estimation period.

Some noticeable conclusions arise from our analysis. As expected, the means of empirical distributions of both the Morishima elasticities and the cost complementarity values are very close for the two forms. However, the standard deviations of these distributions are significantly larger with Fourier than with translog. This result confirms that the Fourier, as a global approximation, allows us to capture the heterogeneity of the sample while the translog only reveals the average behaviour of farmers. When one is to choose between various agricultural policies (such as taxation to give incentives to save water or to reduce fertilizer utilization), being able to account for the heterogeneity may help to design relevant policies. It is also an interesting exercise to compare the results obtained through the two cost functions with the knowledge of agricultural experts. For example, the description of the technology made with translog implies that farmers should diversify their production which is at variance with both the results from Fourier and the success of farmers who have specialized in the production of apples, as noted by several agricultural experts.

Our analysis also shows the usefulness of bootstrap techniques in the context of seminon-parametric methods. While the Fourier can be estimated by means of standard regression techniques, it seems that even for a low order of approximation, the sample size required to validate the asymptotic properties of Fourier estimates would be difficult to meet in applied research. In usual circumstances bootstrapping the distributions of parameters appears to be a practicable solution.

As far as large panel data sets exist, our study offers some evidence of gains in choosing Fourier to approximate the true cost function. It remains that the Fourier is not parsimonious in terms of number of parameters, particularly if one wants to have a disagregated description of the technology (for the purpose of studying the effects of different exogenous shocks on the changes in total costs for instance). This may limit the advantages of this approach very rapidly.

APPENDIX 1: THE SET OF MULTI-INDEXES

To represent the third-order approximation of a cost function in the Sobolev norm, for a technology based on three inputs and two outputs, we need the following 27 multi-indexes:

0	0	0	1	1	0	1	0	1	0	0	0	0	1	0	1	1	1	1	1	0	0	0	0	0	0	0
0	0	0	1	−1	0	0	−1	1	−1	1	0	1	0	−1	1	−1	0	0	0	0	0	0	0	0	0	0
0	0	0	−1	0	−1	0	0	−1	0	−1	0	−1	0	0	−1	0	−1	−1	−1	−1	0	0	0	0	0	0
1	0	1	0	0	0	1	0	0	0	0	1	−1	2	−1	1	1	1	0	0	−1	1	2	2	0	3	0
0	1	−1	0	0	0	1	1	1	−1	−1	2	0	−1	0	0	0	0	1	−1	0	−2	1	0	2	0	3

APPENDIX 2: SAMPLE SUMMARY STATISTICS AND VARIABLE DEFINITIONS

Our data set is extracted from three annual surveys of French farms which report accounting information. These data are aggregated to construct two outputs, i.e. production of apples and other productions, and three inputs: capital, labour and materials. Remember that, for estimating a cost function, all we need are the cost shares and prices of inputs. Some details about their computation are now given.

(1) Production of apples: The quantity of apples produced is directly measured from the survey. The revenue generated by this production is also available which permits us to compute an average price.
(2) Other production: The total revenue obtained from these various products are divided by the output price of apples to obtain a quantity index measured in units of apple quantity.
(3) Capital: The cost of capital is measured by the sum of financial expenses, depreciation and land taxes. All these amounts are given by the accounting system as well as the measures of landed capital and fixed assets. Accordingly, the price of capital is the sum of the rental

Table AI. 1984–130 farmers

Variables	*Cost*	S_K	S_L	S_M	y_1	y_2	w_K	w_L	w_M
Mean	509 921	0·17	0·52	0·31	0·000	0·000	0·000	−0·014	1·882
St. dev.	368 672	0·07	0·10	0·09	1·640	1·176	0·655	0·099	0·385
Max	2 274 573	0·35	0·77	0·61	3·153	1·741	2·065	0·391	3·118
Min	132 877	0·04	0·19	0·15	−6·065	−7·306	−1·714	−0·241	0·783

Table AII. 1985–135 farmers

Variables	*Cost*	S_K	S_L	S_M	y_1	y_2	w_K	w_L	w_M
Mean	540 734	0·16	0·54	0·30	−1·076	−0·375	0·005	0·170	1·856
St. dev.	388 363	0·06	0·11	0·10	1·737	1·013	0·661	0·095	0·378
Max	2 438 817	0·33	0·79	0·64	2·314	2·879	1·725	0·572	2·804
Min	168 392	0·04	0·21	0·14	−6·903	−3·534	−1·688	−0·048	0·549

Table AIII. 1986–145 farmers

Variables	*Cost*	S_K	S_L	S_M	y_1	y_2	w_K	w_L	w_M
Mean	541 026	0·19	0·47	0·34	−0·862	−0·204	0·098	−0·011	1·839
St. dev.	451 928	0·07	0·11	0·11	1·803	0·994	0·590	0·195	0·384
Max	3 557 328	0·40	0·76	0·61	2·774	1·580	1·580	0·543	2·754
Min	117 901	0·04	0·26	0·14	−8·263	−4·405	−1·784	−0·707	0·992

price of acquisition (measured by dividing the financial expenses by the sum of long and medium term debt), the rate of depreciation (obtained by dividing the depreciation by the value of capital, i.e. land and fixed assets) and the user cost of capital (measured by dividing the land taxes by the value of land).

(4) Labour: Two types of labour are aggregated, the household's worktime and the hired labour force. The total wage bill and the total number of hours worked by hired workers are available. Then an average wage can be derived for each farm. We assume that the household's worktime must be evaluated at the mean of wages of hired workers computed over the last percentile of the distribution of these wages.[19] A rationale for this choice is the following. Consider a farmer (or his other family) who would have no worker. It means that the farmer's wage as a worker must be higher than the highest wages prevailing on the labour market; indeed, at the market price, he would prefer working on his farm rather than selling his labour force elsewhere. Since the survey offers an estimate of hours spent by the farmer's family for production purposes, we are then able to evaluate the cost of the household's worktime, and to aggregate the two types of labour using the technique of the Törnqvist index.

(5) Materials: These cover all items that are completely renewed each year such as chemicals or energy. Quantities of all inputs gathered in this group are documented in the survey, but not all their prices and costs. Non-documented prices are obtained from external sources: Mostly these are indexes published by the French statistical institute INSEE and are identical for all farms. Then the Törnqvist index of price for materials can then be obtained.

Some annual statistics on these indexes are presented in Tables A1–A3. Cost is the total cost in French Francs, S_i is the cost share of input $i (i = K, L, M$, which stands for capital, labour and materials respectively), y is the logarithm of quantity index of product $j (j = 1{,}2$ for apples and other products respectively) and w is the logarithm of price index of input $i (i = K, L, M)$. Note that the means of the two output quantity indexes and the price index of capital are equal to zero

[19] This specific choice should be tested. In any case the distribution of observed wages has a small variance, so that the potential bais introduced by considering only the last percentile of this distribution is relatively small.

in 1984, since this year serves as the base year and the 'within' means for this year (computed over the individuals) define the individual of reference. The means of the two other price indexes differ from zero because computation of these indexes involves the use of prices that do not vary across farms.

ACKNOWLEDGEMENTS

Financial support from the Institut National de la Recherche Agronomique (INRA) is gratefully acknowledged. The authors wish to thank S. Monier-Dilhan for her skilful help on the database. Earlier versions of the paper were presented at the GREMAQ-INRA Seminar, Toulouse, May 1990, the Conference on Panel Data, Paris, June 1990, and the European Meeting of the Econometric Society, Cambridge, September 1991. Insightful remarks and comments by H. Cremer, J. P. Florens, M. J. Lee, Q. Vuong, and four anonymous referees have been invaluable.

REFERENCES

Abler, D. and J. Shortle (1993), 'Environmental policies and induced innovation: the case of agriculture', mimeo, Pennstate University.

Andrews, D. W. K. (1991), 'Asymptotic normality of series estimators for nonparametric and semiparametric regression models', *Econometrica*, **59**, 307–345.

Baltagi, B. H. (1980), 'On seemingly unrelated regression with error components', *Econometrica*, **48**, 1547–1551.

Baltagi, B. H. (1985), 'Pooling cross-sections with unequal time-series length', *Economics Letters*, **18**, 133–136.

Blackorby, C. and R. R. Russel (1989), 'Will the real elasticity of substitution please stand up? (A comparison of the Allen/Uzawa and Morishima elasticities)', *The American Economic Review*, **79**, 882–888.

Chalfant, J. A. and N. E. Wallace (1991), 'Testing the translog specification with the Fourier cost function', Department of Agricultural and Ressource Economics, University of California, Berkeley, WP–585.

Chamberlain, G. (1984), 'Panel data', in Z. Griliches and M.D. Intrilligator (eds), *Handbook of Econometrics*, Vol. 2, North-Holland, Amsterdam.

Christensen, L. R., D. W. Jorgenson and L. J. Lau (1973), 'Transcendental logarithmic production frontiers', *Review of Economics and Statistics*, **55** 28–45.

Cornwell, C., P. Schmidt and R. C. Sickles (1990), 'Production frontiers with cross-sectional and time-series variations in efficiency levels', *Journal of Econometrics*, **4**, 185–200.

Devezeaux de Lavergne, J. G., M. Ivaldi and N. Ladoux (1990), 'La forme flexible de Fourier: Une évaluation sur données macroéconomiques', *Annales d'Economie et de Statistiques*, **17**, 121–144.

Diewert, W. E. (1971), 'An application of the Shepard duality theorem, a generalized Leontief production function', *Journal of Political Economy*, **79**, 481–507.

Diewert, W. E. (1974), 'Applications of duality theory', in M. D. Intriligator and D. A. Kendriks (eds), *Frontiers in Quantitative Economics*, North-Holland, Amsterdam.

Eakin, B. K., D. P. McMillen and M. J. Buono (199()), 'Constructing confidence intervals using the bootstrap: an application to a multiproduct cost functions', *Review of Economics and Statistics*, **72**, 339–345.

Eastwood, B. J. and A. R. Gallant (1991), 'Adaptive truncation rules for semi non parametric estimators that achieve asymptotic normality', *Econometric Theory*, **7**, 307–340.

Efron, B. and R. Tibshirani (1986), 'Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy', *Statistical Science*, **1**, 54–77.

Elbadawi, I., A. R. Gallant and G. Souza (1983), 'An elasticity can be estimated consistently without *a priori* knowledge of the functional form', *Econometrica*, **51**, 1731–1751.

Eubank, R. L. and P. Speckman (1990), 'Curve fitting by polynomial-trigonometric regression', *Biometrika*, **77**, 1–9.

Eubank, R. L. and P. Speckman (1991), 'Convergence rate for trigonometric and polynomial-trigonometric regressions estimators', *Statistics and Probability Letters*, **11**, 119–124.

Florens, J. P., M. Ivaldi and S. Larribeau-Nori (1991), 'Sobolev approximations of unknown regressions', mimeo, GREMAQ, Toulouse.

Fuss, M., D. McFadden and Y. Mundlak (1978), 'A survey of functional forms in the economic analyses of production', in M. Fuss and D. McFadden (eds), *Production Economics: A Dual Approach to Theory and Applications*, North-Holland, Amsterdam.

Gallant, A. R. (1981), 'On the bias in flexible functional forms and an essentially unbiased form: the Fourier flexible form', *Journal of Econometrics*, **15**, 21–245.

Gallant, A. R. (1982), 'Unbiased determination of production technologies', *Journal of Econometrics*, **20**, 285–323.

Gallant, A. R. (1984), 'The Fourier flexible form', *American Journal of Agricultural Economics*, **66**, 204–208.

Gallant, A. R. and G. Souza (1991), 'On the asymptotic normality of Fourier flexible form estimates', *Journal of Econometrics*, **50**, 329–353.

Guilkey, D. K., C. A. K. Lovell and R.C. Sickles (1983), 'A comparison of the performance of three functional forms', *International Economic Review*, **24**, 591–617.

Lau, J. L. (1986), 'Functional forms in econometric model building', in Z. Griliches and M. D. Intriligator (eds), *Handbook of Econometrics*, Vol. 3, North-Holland, Amsterdam.

Lilliefors, H. W. (1967), 'On the Kolmogorov–Smirnov test for normality with mean and variance unknown', *Journal of the American Statistical Association*, **64**, 399–402.

Ministère l'Agriculture (1986), 'Résultats économiques des exploitations agricoles', *RICA*, no 2, Paris.

Ministère l'Agriculture (1987), 'Résultats économiques des exploitations agricoles', *RICA*, no 4, Paris.

Ministère l'Agriculture (1988), 'Résultats économiques des exploitations agricoles', *RICA*, no 7, Paris.

Nerlove, M. (1971), 'A note on error component models', *Econometrica*, **39**, 383–396.

Serfling, T. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley, New York.

Veall, M. R. (1992), 'Bootstrapping the process of model selection: an econometric example', *Journal of Applied Econometrics*, **7**, 93–99.

Verbeek, M. and T. Nijman (1992), 'Incomplete panels and selection bias', in L. Mátyás and P. Sevestre (eds), *The Econometrics of Panel Data*, Kluwer Academic Publishers, Dordrecht.

Wansbeek, T. and A. Kapteyn (1989), 'Estimation of the error-components model with incomplete panels', *Journal of Econometrics*, **41**, 341–361.

White, H. (1980), 'Using least squares to approximate unknown regression functions', *International Economic Review*, **21**, 149–170.

Zellner, A. (1962), 'An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias', *Journal of the American Statistical Association*, **57**, 348–368.