ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 55-56 – 1999

Capital-Labor Substitution Heterogeneity with Endogenous Switching Regression

Nicolas COTE-COLISSON, François LEGENDRE *

ABSTRACT. – The aim of this paper is to show that it would be easier, for firms, to substitute capital for low skilled labor than to substitute capital for high skilled labor. To test this conjecture, we use panel data spanned from 1980 to 1987 on nearly 800 French manufacturing firms. We first propose a semi-reduced form obtained conditionally to a specific price setting rule as, at the micro level, the selling price is not observed. We next argue for estimating a system of two equations as the production function alone is not the best way to reveal the elasticity of substitution in production. Our system can be interpreted as the following recursive firms' behavior. On the one hand, firms choose the optimal capital per worker level; on the other hand, the optimal production and price levels. We also can capture the between-firms heterogeneity by means of individualizing a deep production function parameter. Finally, we propose to trace heterogeneity in elasticity of substitution by considering the presence of two groups of firms. The first group includes the low elasticity of substitution firms; the second group includes the high elasticity of substitution firms. We develop a switching regression model to carry out firms classification, by endogenous selection. Eventually, we find that the average skills of the work force is a factor which contributes to accounting for the probability that a firm belongs in either one group or another. This paper then gives a (soft) evidence to support the French current policy to subsidize the "low wages jobs".

Hétérogénéité de la substitution capital-travail et régression à changement de régime endogène

RÉSUMÉ. – Le but de ce travail est de montrer qu'il serait plus facile, pour les entreprises, de remplacer du travail peu qualifié par du capital que de remplacer du travail qualifié par du capital. Pour tester cette conjecture, nous avons recours à des données de panel de près de 800 entreprises industrielles françaises suivies de 1980 à 1987. Tout d'abord, nous développons une forme semi-réduite obtenue conditionnellement à une règle particulière de fixation des prix de vente pour pallier le fait que ces prix ne sont pas, au niveau micro-économique, observés. Ensuite, nous plaidons pour estimer un système de deux équations pour évaluer plus précisément l'élasticité de substitution capital-travail. Ce système retrace le comportement récursif suivant des entreprises. En premier lieu, elles choisissent le niveau optimal du capital par tête ; en second lieu, elles fixent les niveaux du produit et du prix de vente. L'hétérogénéité inter-entreprise est spécifiée sous deux formes : d'un côté, par l'individualisation d'un paramètre structurel de la fonction de production ; de l'autre côté, en supposant la présence de deux groupes d'entreprises. Le premier groupe est constitué des entreprises à faible élasticité de substitution ; le second à forte élasticité de substitution. La ventilation des entreprises est basée sur un modèle de « *switching regression* », avec une règle endogène de sélection. Nous trouvons, finalement, que le degré moyen de qualification de la main-d'œuvre est un facteur qui explique la probabilité pour une entreprise d'appartenir à l'un ou à l'autre des deux groupes. Ces résultats contribuent ainsi à justifier la politique d'allégement de charges sur les bas salaires conduite en France depuis 1993.

* N. COTE-COLISSON: Université Paris-I, Centre d'Économie Mathématique et d'Économétrie; F. LEGENDRE: EPS-Adeps (Université Nancy-II et CNRS) et Conseil supérieur de l'emploi, des revenus et des coûts.

We are grateful to the *Direction de la prévision* of French *Ministère de l'Économie et des Finances* for financial research support. The data used are from the *Centrale des bilans* of the *Direction de la prévision, Bureau des études fiscales.* Anne GENTY gave us a decisive help for the English translation. We thank the two anonymous referees for useful comments. Any errors are, of course, ours.

1 Introduction

Unemployment, in Europe, first affects the least skilled workers. Although this fact could be only a "*stylized fact*" which can be challenged, [1] we believe that a better explanation of the causes of unemployment could be obtained by distinguishing the various skills of the labor force.

The aim of this paper is to show that it would be easier, for firms, to substitute capital for low skilled labor than to substitute capital for high skilled labor. [2] Current unemployment would then be "classical unemployment" related to a labor cost too high for the least skilled workers. This view supports the macroeconomics policy recommendation to subsidize this jobs through a reduction of the employer's social contribution on low wages (along others, Commissariat général du Plan [1993], DRÈZE and MALINVAUD [1994], Commissariat général du Plan [1994] or Conseil supérieur de l'emploi, des revenus et des coûts [1996]).

To test this conjecture, we use panel data spanned from 1980 to 1987 on nearly 800 manufacturing firms. We should obtain more accurate results with such data than with macroeconomic data. However, the panel does not always incorporate very substantial information as the data come from firms' records. Skill structures are not available and we use wages per employee as proxy for the average skills of the labor force. Nevertheless, the distinction, at the micro level, between quantity and price is not available. The standard way to avoid this difficulty is to use sectorial deflators. But this is not very satisfactory as the latter do not include useful information. Finally, panel data econometrics emphasize the potential bias entailed by not specifying the between-firms heterogeneity. In the copious literature devoted to labor demand, it is stressed the biased results obtained when the various categories of workers are aggregated as long as specific elasticities to each component of labor exists (along others, BRESSON, KRAMARZ and SEVESTRE [1992], HAMERMERSH [1993] or DORMONT [1996]).

Then, we first develop, in section 2, a model with imperfect competition to reckon how the value added price is set. Thus we can use only observed data in the estimates but our results depend on our price setting hypothesis. The model is useful to detect the sources of between-firms heterogeneity that we can incorporate in a "*deep structural*" manner. We address heterogeneities in two ways. On one hand, we consider a firm-specific effect on a parameter of the production function. On the other hand, we postulate the presence of two groups of firms, the first characterized by a low elasticity of capital-labor substitution; the second by a high elasticity. We shall exploit a switching regression model to classify, by endogenous selection, each firm of the sample. We then explore a particular route for "*freeing up*" the elasticity of

1. Because, for example, this situation comes from spillover effects where the high skilled workers supersede low skilled workers or because firms, in order to hoard labor, prefer to first fire the least skilled employees (who not have specific human capital).
2. Our investigations are then complementary to SNEESSENS works (for example, MAILLARD and SNEESSENS [1993] and SNEESSENS [1994]) on biased technical progress.

substitution as an alternative to most usual ways (for example, a random coefficient model *à la* Swamy).

Section 3 reports the results. Heterogeneities are strong because the elasticity of substitution in the first group is about .6 when it is about .9 in the second group. So, the paper proposes a methodology for estimating some structural parameters with incomplete administrative data. We do not try to parametrize the elasticity of capital-labor substitution as a function of some firm-specific variables. We prefer use a 2-step procedure to eventually relate the probability of being in each regime to firm-specific variables. We find that the average skills of the labor force (approximated by the wages per employee) is a significant factor: a firm is more likely to be in the low elasticity group when the skills of its labor force are high.

2 The Theoretical and the Estimated Models

In a first stage, we quickly set the theoretical model. This model is very general. In a second stage, we then explain how the latter can be improved in order to trace a first source of between-firms heterogeneity. Finally, we present the estimated model, a switching regression model which entails us to trace a second source of heterogeneity about elasticity of capital-labor substitution.

2.1. A Simple Model of Imperfect Competition

In this first subsection, we expose a simple model of imperfect competition. The model is especially useful to identify the evolution of the firm's profits. We can elude, in this way, the fact that prices, at the micro level, are not observable.

Let us start by a production function with constant elasticity of substitution (noted σ)

$$Q = A(\delta K^{(\sigma-1)/\sigma} + (1-\delta) L^{(\sigma-1)/\sigma})^{v\sigma/(\sigma-1)} \tag{1}$$

where Q is the level of production (in fact the value added); K and L, the factors of production stocks (resp. capital and labor); A, a dimensional term (also the total factor productivity); δ and $1-\delta$, terms which measure in first approximation the relative contribution of capital and labor to the production; and v, the returns to scale parameter.

Let us denote by C the user cost of capital and by W the total wages per employee and let us suppose that the firm is "*price taker*" (as competition is perfect on the factors of production markets or as those costs are predetermined by negotiations). Then the cost function, for the firm, can be deduced for the following program

$$\min_{K \text{ and } L} CK + WL$$

subject to the production function constraint. The cost function is then

$$C(Q) = \left(\frac{Q}{A}\right)^{1/v} (\delta^{\sigma} C^{1-\sigma} + (1-\delta)^{\sigma} W^{1-\sigma})^{1/(1-\sigma)}$$

The duality theory can help to interpret this formula. The latter is an inverse CES specification: when σ goes to 0 (*i.e.* the factors of production are complementary – the CES operator goes to the min operator), the cost function is directly the addition of the factors costs.

This program also leads to the following capital per worker ratio:

(2) $$\frac{K}{L} = \left(\frac{\delta}{1-\delta}\right)^{\sigma} \left(\frac{C}{W}\right)^{-\sigma}$$

We see that the elasticity of this ratio with respect to the relative cost of the factors of production is equal to σ according to the CES specification of the production function.

It is not very interesting to exactly specify the imperfect competition environment of the firm. Let us suppose that the equilibrium leads to the following price setting rule

(3) $$P = \mu\, C'(Q)$$

where P is the selling price, μ a parameter like a mark-up rate, and $C'(Q)$ the marginal cost of production. This rule could be justified in different ways. It is first possible to invoke competition *à la* COURNOT; alternatively competition *à la* BERTRAND. In both cases, the marginal cost of production is still a reference as the selling price is a mark-up on this marginal cost (see BENASSY [1989] for recent development on this subject).

To eliminate Q (the value added by quantity which is not observed) from the former expression, it is possible, using in the one hand the production function (1) and in the other hand the price rule (3), to get the quasi-reduced form

(4) $$Y = \frac{\mu}{v}(\delta K^{(\sigma-1)/\sigma} + (1-\delta) L^{(\sigma-1)/\sigma})^{\sigma/(\sigma-1)} (\delta^{\sigma} C^{1-\sigma} + (1-\delta)^{\sigma} W^{1-\sigma})^{1/(1-\sigma)}$$

where Y is the value added in nominal terms — $Y = P\,Q$. This formula can be interpreted as a production function in nominal terms where the levels of the factors of production stocks and the levels of the factors of production costs are related to the value added in nominal terms.

The A term, the total factor productivity, vanishes from the latter expression because it is entirely passed on to the selling price. There are two alternatives when one tries to estimate at the micro level production functions. One may use sectorial deflators and implicitly suppose that all firms, in the industry, have increase their price in the same proportion. Conversely, one may not want to use such deflators and suppose that the productivity gains are passed on to price. We prefer to use the second route: it seems to us to be closer to industrial realities.

It is now necessary to put emphasis on the effective specification used for our panel data.

2.2. The First Source of Individual Heterogeneity

The exposed model needs to be amended as it is fitted to panel data. The production function in nominal terms we try to fit is as follows (in logarithm to avoid heteroskedasticity)

$$\log Y_{it} = \log \frac{\mu}{v} + \frac{\sigma}{\sigma - 1} \log(\delta K_{it}^{(\sigma-1)/\sigma} + (1-\delta) L_{it}^{(\sigma-1)/\sigma}) + \frac{1}{1-\sigma} \log(\delta^{\sigma} C_{it}^{1-\sigma} + (1-\delta)^{\sigma} W_{it}^{1-\sigma}) \tag{5}$$

The subscripts $i = 1, \ldots, N$ and $t = 1, \ldots, T$ denote the double data dimension. The first is about firm (N is the number of overall firms); the second, year (T is the number of overall years).

It is necessary to be specific about between-firms heterogeneity. The standard way is to put in the model an individual fixed effect (indeed a stochastic effect). By introducing dummy specific firm variables in the regression, the average differences between firms are suppressed.

However, this specification is *ad hoc*; the individual effect would correspond to individualizing the μ coefficient (the mark-up rate) or the v coefficient (the returns to scale parameter). It is hard to believe that the main source of between-firm diversity comes from those two coefficients. We believe that it is more natural to first look at total factor productivity (the A parameter). Some firms are more efficient, in a permanent way, than others for a lot of reasons that we do not need to enumerate. But our model tells us that high productivity is passed on to low prices: the more efficient firms by this way can increase their economic activity. It is then not justified to put an individual fixed effect in the regression.

We can, moreover, capture other important specific effects. First, it seems necessary to explain the capital-intensitivity of the occupation. Some industries are more capital-intensive than others. Individualizing the δ coefficient is the formal way to capture such heterogeneity. This specification allows us not to put an excessive emphasis on the relative cost of the factors of production in explaining the capital per worker ratio. This ratio is now specified as follows

$$\frac{K_{it}}{L_{it}} = \left(\frac{\delta_i}{1-\delta_i} \right)^{\sigma} \left(\frac{C_{it}}{W_{it}} \right)^{-\sigma}$$

On the one hand, the $\delta_i/(1-\delta_i)$ term captures the specific of the capital-intensive level of the occupation. On the other hand, the relative cost of the factors of production captures the substitution between factors of production from the cost minimization.

Eventually, it is better to exploit all the hypotheses used. The elasticity of substitution (our parameter of interest) is not well revealed by the production function. It is for the capital per worker ratio that this parameter is much more meaningful as equation (2) tell us

$$\frac{K}{L} = \left(\frac{\delta}{1-\delta}\right)^{\sigma} \left(\frac{C}{W}\right)^{-\sigma}$$

It seems then hard not to exploit this information. We propose then to estimate the following system

$$\text{(6)} \qquad \log \frac{K_{it}}{L_{it}} = \sigma \log \frac{\delta_i}{1-\delta_i} - \sigma \log \frac{C_{it}}{W_{it}} + g_{1\,t} + u_{1\,it}$$

$$\log Y_{it} = \log \frac{\mu}{v} + \frac{\sigma}{\sigma-1} \log \left(\delta_i K_{it}^{(\sigma-1)/\sigma} + (1-\delta_i) L_{it}^{(\sigma-1)/\sigma}\right) +$$

$$\text{(7)} \qquad \frac{1}{1-\sigma} \log(\delta_i^{\sigma} C_{it}^{1-\sigma} + (1-\delta_i)^{\sigma} W_{it}^{1-\sigma}) + g_{2\,t} + u_{2\,it}$$

To simplify things, we do not identify two technical progress terms (the first which augments the capital, the second which augments the labor). We only put in the regression temporal dummies (see the $g_{1\,t}$ and $g_{2\,t}$ terms) with, for purpose of identification, the constrains $\sum_t g_{1\,t} = 0$ and $\sum_t g_{2\,t} = 0$. Those dummies also capture short term shocks (if they have the same impact on all firms). Finally, $u_{1\,it}$ and $u_{2\,it}$ are the two error terms for each equation. On first order, the system includes $N + 2(T-1) + 2$ parameters to estimate; the δ_i (by quantity N), the $g_{1\,t}$ and $g_{2\,t}$ (by quantity $2(T-1)$ only because the two constraints), the $\log(\mu/v)$ term, and the elasticity of substitution σ.

This system of equations would depict a recursive firm's behavior as following. In a first stage, the firm chooses the optimal capital per worker ratio (first equation). In a second stage, it chooses the product level and the sell price level (second equation) (see, in a different context, HÉNIN and POUCHAIN [1980] for a much more elaborated justification). This framework lets us retain the absence of instantaneous correlation between u_1 and u_2. This hypothesis, only acceptable in a first approximation, is retained as the estimator takes a simpler form (note also, if there is instantaneous correlation, the estimator is still unbiased).

2.3. **The Second Source of Individual Heterogeneity**

We now shall address heterogeneity related to the elasticity of capital-labor substitution. We capture such heterogeneity in the following way. We suppose that our firms' population is a mix of two sub-populations. The first group would be the low elasticity of substitution group; the second the high elasticity group. We consequently need a specific estimation method. On the one hand, we can use a model with exogenous selection (the population is separated into two groups on the grounds of an exogenous variable) or a model with endogenous selection (the estimation method must, in addition, classify each firm). We opt for the second solution as we do not need to introduce *a priori* information about the way the firms fall into one or another group. With panel data, we can retain, to ensure identification, that each firm stays in the same group all along the period.

In appendix B, the model used is explained in detail. It is a switching regression model which distinguishes, for each observation, two alternative explanations. Moreover, an auxiliary equation models the group classification. On cross section, the model is as follows (where, for sake of simplicity, we retain only one explanatory variable for each equation)

$$(8) \qquad y_i = \begin{cases} a^a x_i^a + u_i^a & \text{if} \quad bz_i + v_i < 0 \\ a^b x_i^b + u_i^b & \text{if} \quad bz_i + v_i \geqslant 0 \end{cases} \qquad i = 1,\ldots,N.$$

The superscripts a and b are related to each regime. Thus the model includes one dependent variable (the $N \times 1$ vector $\underline{y}$), three explanatory variables (the $N \times 1$ vectors $\underline{x}^a$, $\underline{x}^b$, and $\underline{z}$), three unknown coefficients to estimate (a^a and a^b which are the regression coefficient in each regime, and b which is the regression coefficient of the auxiliary equation), and three error terms (the $N \times 1$ vectors $\underline{u}^a$, $\underline{u}^b$, and $\underline{v}$). On panel data, the model is now, with the subscripts i and t

$$(9) \quad y_{it} = \begin{cases} a^a x_{it}^a + u_{it}^a & \text{if} \quad bz_{it} + v_{it} < 0 \\ a^b x_{it}^b + u_{it}^b & \text{if} \quad bz_{it} + v_{it} \geqslant 0 \end{cases} \qquad i = 1,\ldots,N; \quad t = 1,\ldots,T.$$

We assume that each firm is always in the same regime. The auxiliary equation reads therefore as follows

$$(10) \qquad b_i + v_{it}$$

As a result, we introduced a firm-specific fixed effect in the auxiliary equation. This effect captures all the causes which can explain that the firm i is in the first or in the second group. To be more precise, the estimated model is of the following form:

if $\quad b_i + v_{it} < 0$

$$\begin{cases} \log \dfrac{K_{it}}{L_{it}} = & \sigma^a \log \dfrac{\delta_i^a}{1-\delta_i^a} - \sigma^a \log \dfrac{C_{it}}{W_{it}} + g_{1t}^a + u_{1it}^a \\ \log Y_{it} = & \log \dfrac{\mu^a}{v^a} + \dfrac{\sigma^a}{\sigma^a - 1} \log\Big(\delta_i^a K_{it}^{(\sigma^a-1)/\sigma^a} + (1-\delta_i^a) L_{it}^{(\sigma^a-1)/\sigma^a}\Big) + \\ & \dfrac{1}{1-\sigma^a} \log\Big(\delta_i^{a\,\sigma^a} C_{it}^{1-\sigma^a} + (1-\delta_i^a)^{\sigma^a} W_{it}^{1-\sigma^a}\Big) + g_{2t}^a + u_{2it}^a \end{cases}$$

(11) if $\quad b_i + v_{it} \geqslant 0$

$$\begin{cases} \log \dfrac{K_{it}}{L_{it}} = & \sigma^b \log \dfrac{\delta_i^b}{1-\delta_i^b} - \sigma^b \log \dfrac{C_{it}}{W_{it}} + g_{1t}^b + u_{1it}^b \\ \log Y_{it} = & \log \dfrac{\mu^b}{v^b} + \dfrac{\sigma^b}{\sigma^b - 1} \log\Big(\delta_i^b K_{it}^{(\sigma^b-1)/\sigma^b} + (1-\delta_i^b) L_{it}^{(\sigma^b-1)/\sigma^b}\Big) + \\ & \dfrac{1}{1-\sigma^b} \log\Big(\delta_i^{b\,\sigma^b} C_{it}^{1-\sigma^b} + (1-\delta_i^b)^{\sigma^b} W_{it}^{1-\sigma^b}\Big) + g_{2t}^b + u_{2it}^b \end{cases}$$

Each system of two equations to estimate is a little complicated. It is not linear with respect to σ^j and to the δ_i^j, $j = a,b$. Likewise, the two equations are interrelated by σ^j and the δ_i^j. However, when σ^j and the δ_i^j are given, the system is much simpler: it can be estimated by a standard method, equation after equation. In fact, the system is linear with the other coefficients (the $\log(\mu^j/v^j)$, $g_{1\,t}^j$ and $g_{2\,t}^j$ terms; $j = a,b$) and the two equations are not interrelated (as we retain a non instantaneous correlation hypothesis). Then, it takes the following form (when σ^j is given set at $\overline{\sigma}^j$ and when each δ_i^j is given set at $\overline{\delta}_i^j$)

$$\log\frac{K_{it}}{L_{it}} - \overline{\sigma}^j \log\frac{\overline{\delta}_i^j}{1-\overline{\delta}_i^j} + \overline{\sigma}^j \log\frac{C_{it}}{W_{it}} = g_{1\,t}^j + u_{1\,it}^j \quad j = a,b$$

$$\log Y_{it} - \frac{\overline{\sigma}^j}{\overline{\sigma}^j-1}\log\left(\overline{\delta}_i^j K_{it}^{(\overline{\sigma}^j-1)/\overline{\sigma}^j} + (1-\overline{\delta}_i^j)L_{it}^{(\overline{\sigma}^j-1)/\overline{\sigma}}\right)$$

$$-\frac{1}{1-\overline{\sigma}^j}\log\left(\overline{\delta}_i^{j\,\overline{\sigma}^j} C_{it}^{1-\overline{\sigma}^j} + (1-\overline{\delta}_i^j)^{\overline{\sigma}} W_{it}^{1-\overline{\sigma}^j}\right)$$

$$= \log\frac{\mu^j}{v^j} + g_{2\,t}^j + u_{2\,it}^j \qquad j = a,b$$

Furthermore, for σ^j, $\log(\mu^j/v^j)$, $g_{1\,t}^j$, and $g_{2\,t}^j$ given, the system is like a non linear system of interrelated equations. It takes the following form (when σ^j is given set at $\overline{\sigma}^j$, $\log(\mu^j/v^j)$ given set at $\log(\overline{\mu}^j/\overline{v}^j)$, each $g_{1\,t}^j$ given set at $\overline{g}_{1\,t}^j$, and each $g_{2\,t}^j$ given set at $\overline{g}_{2\,t}^j$)

$$\log\frac{K_{it}}{L_{it}} + \overline{\sigma}^j \log\frac{C_{it}}{W_{it}} - \overline{g}_{1\,t}^j = \overline{\sigma}^j \log\frac{\delta_i^j}{1-\delta_i^j} + u_{1\,it}^j \quad j = a,b$$

$$\log Y_{it} - \log\frac{\overline{\mu}^j}{\overline{v}^j} - \overline{g}_{2\,t}^j = \frac{\overline{\sigma}^j}{\overline{\sigma}^j-1}\log\left(\delta_i^j K_{it}^{(\overline{\sigma}^j-1)/\overline{\sigma}^j} + (1-\delta_i^j)L_{it}^{(\overline{\sigma}^j-1)/\overline{\sigma}^j}\right)$$

$$+\frac{1}{1-\overline{\sigma}^j}\log\left(\delta_i^{j\,\overline{\sigma}^j} C_{it}^{1-\overline{\sigma}^j} + (1-\delta_i^j)^{\overline{\sigma}^j} W_{it}^{1-\overline{\sigma}^j}\right) + u_{2\,it}^j \quad j = a,b$$

The two equations are then interrelated. To estimate the δ_i^j, it is necessary to take account of the heteroskedasticity between the equations; but it is possible to proceed by individual and the computational burden is then substantially light.

In the appendix B, we explain in greater detail the exact form of the ML estimator and the computational procedure. The numerical resolving can be summarized by the three following items.

- The overall procedure is conducted conditional to a value for σ^a and a value for σ^b, given before; the procedure is then duplicated for different trial values of σ^a and σ^b.
- A "zigzag" procedure, iterative one, is carry out. Now, the "macro" coefficients ($\log(\mu^j/v^j)$, the $g_{1\,t}^j$, the $g_{2\,t}^j$ and the variances of u_1^j and u_2^j) are given for $j = a,b$ and the "micro" (the δ_i^j) are estimated; now, the "micro" coefficients are given and the "macro" coefficients are estimated.

TABLE 1
The Asked Questions and their Resolution

Question	Answer
1 How not to use sectorial deflators which do not include pertinent information as selling prices at the micro level are not observed?	Develop an imperfect competition model and estimate a quasi reduced form conditional by the price setting rule.
2 How to capture the data heterogeneity as the diversity of capital-intensitivity is not only a matter of relative cost of the factors of production?	Individualize δ and $1-\delta$ terms of the production function which are the relative contribution of capital and labor to the production.
3 How to estimate accurately the elasticity of substitution while this parameter does not enter as main effect in the production function?	Estimate a system of two equations, the first which captures the optimal behavior of capital per worker setting, the second which derives from the optimal setting behavior of the level of production and of selling price.
4 How to address elasticity of substitution heterogeneity as some firms can more easily substitute capital to labor than others?	Let us suppose two groups of firms with respect to the elasticity value and use an endogenous switching regression model.

We show, in appendix B, that this computational procedure is convergent and it coincides with ML estimator.

- The δ_i^j are estimated by individuals. In fact, we show that the procedure is separable: it is not necessary to optimize on $\delta_1^j, \delta_2^j, \ldots, \delta_N^j$; it is sufficient to maximize on δ_1^j, then on δ_2^j, etc. Each δ_i^j is between 0 and 1; we thus proceed by grid with 101 trials for each firm (a .01 step was chosen).

We use GAUSS (a specialized computational language) and IBM-PC like computer to carry out the computations. The programs are available upon request.

In this section, we have introduced a lot of complications, each with a view to answering a precise question. Table 1, as abstract, lists the asked questions and their resolution.

3 The Results

The results of the estimated model are first displayed. Sub-section 3.2 looks for explanatory factors to classify firms in each group.

3.1. **Two Groups of Firms**

The data used are about nearly 800 industrial firms, from 1981 to 1987. See appendix A for a more precise data presentation (see also LEGENDRE [1990]). The user cost of capital is a much more elaborated measure than the financial charges over debt ratio. It is evaluated as a weighted average cost in the different ways to finance a firm (see, among others, KING [1974], AUERBACH [1983], and LEGENDRE [1990] for the development of a such user cost measure).

Table 2 illustrates the computational procedure used. The two competing groups differ only in terms of the value of the elasticity of capital-labor substitution. The first regime, denoted by *a*, includes firms with low elasticity of substitution; the second regime, denoted by *b*, includes firms with high elasticity of substitution. For different values of σ^a (then this coefficient is given, we start by a first grid from .35 to .75 by .1 step) and different values of σ^b (a grid from .65 to 1.25 by .1 step is used), the table reports two figures: the log-likelihood value and the number of firms in group *a*. We take care to keep $\sigma^b - \sigma^a$ above .3 so that the two regimes remain identifiable. One can see, on this table, that a maximum likelihood is achieved about $\sigma^a = .55$ and about $\sigma^b = .95$. With this maximum, the numbers of firms in each group are almost equal (41 % in group *a*, the low elasticity group).

Table 3 displays the result for the second grid. The maximum likelihood is obtained for $\sigma^a = .57$ and for $\sigma^b = .93$. The detailed results for each group are reported in table 4.

Those results do not seem to invalidate our conjecture: the population is a mix of two groups, the low elasticity of substitution in production group and the high elasticity of substitution in production group. For the group *b*, the μ^b/v^b coefficient is estimated below 1. This result comes either from a low mark-up rate (μ^b not very great) either from increasing returns to scale (with $v^b > \mu^b$). The theoretical model is yet invalidate as it supposes $\mu/v > 1$ so the firm revenues are positive. This result however relies on the temporal

TABLE 2
Two Groups of Firms (First Grid)

The first figure is the log-likelihood value
The second figure is the number of firms in group *a* (by %)

		σ^b						
		.65	.75	.85	.95	1.05	1.15	1.25
σ^a	.75					2297 85	2187 89	2203 90
	.65				2537 63	2458 70	2401 77	2306 88
	.55			2713 35	2858 41	2513 38	1817 61	1956 83
	.45		2703 22	2280 21	2425 27	2163 28	2055 32	1721 35
	.35	2071 11	2554 13	2149 13	2208 14	1944 16	1738 16	1422 19

Number of iterations achieved: 100.

dummies normalization so it is not especially informative. The $\widehat{\delta}$ average is about .8 for the group *a* and about .3 for the group *b* but it is only when the elasticity of substitution is equal to 1 that one can interpret the δ parameter as relative contribution of capital to the production. We then get, for the group *a* where the elasticity is very different from 1, a strong deformation for the estimated δ_i distribution. On the other hand, for the group *b*, we can maintain the usual interpretation of the δ parameter.

We now need to better accounting for such a split in the population. This is the aim of the next sub-section.

TABLE 3
Two Groups of Firms (Second Grid)

The first figure is the log-likelihood value
The second figure is the number of firms in group *a* (by %)

		σ^b						
		.92	.93	.94	.95	.96	.97	.98
σ^a	.58	2968 44	2917 44	2816 42	2892 45	2887 45	2920 44	2860 45
	.57	2923 42	3048 43	2800 42	2918 44	2826 42	2946 43	2727 43
	.56	2933 42	2917 42	2851 41	2862 42	2907 43	2878 42	2768 41
	.55	2779 38	2788 40	2655 39	2873 42	2769 42	2710 41	2697 40
	.54	2809 37	2752 37	2743 38	2812 39	2806 41	2683 38	2611 38
	.53	2795 35	2737 36	2707 37	2840 38	2720 38	2606 36	2652 38
	.52	2765 34	2767 32	2693 33	2496 31	2729 34	2649 33	2414 31

Number of iterations achieved: 150.

TABLE 4
The Results of the Two Regimes Model
Estimation Method: Maximum Likelihood

Quantity	Group *a*	Group *b*
σ	.57	.93
$\log(\mu/v)$	.026	– .015
$V(u_1)$	.027	.0089
$V(u_2)$	.081	.013
$\widehat{\delta}$ average	.81	.31
$\widehat{\delta}$ median	.88	.28
Number of firms in group	325	442
Log-likelihood	3048	

Results for temporal dummies are not reported.

TABLE 5
The Factors used of the Descriptive Analysis

Variable	Formula	Description
y	$\log(\overline{Y}_i)$	Size of the firm (in log, average over the period of the value added)
dy	$\log(Y_{iT}) - \log(Y_{i1})$	Value added in nominal terms rate of growth (average over the period)
ksl	$\log(\overline{K_i/L_i})$	Capital per worker (in log, average over the period)
dksl	$\log(K_{iT}/L_{iT}) - \log(K_{i1}/L_{i1})$	Capital per worker rate of growth
w	$\log(\overline{W}_i)$	Average skills of the labor force (in log, we use as proxy the average wages per employee)

3.2. **The Factors of the Elasticity of Capital-Labor Substitution**

We try in this section to identify the factors which can explain the probability that a firm is in one or another group. More precisely, we use two models. In one hand, we employ OLS to explain the quantity $1 - F(-\widehat{b_i})$ where $\widehat{b_i}$ is the estimated coefficient for firm *i* from previous analysis and where $F(\cdot)$ is the c.d.f. of a zero mean, unit variance normally distributed variable – so we want to account for the probability that the firm belongs to group *b*. On the other hand, we employ Probit model to explain the *ex-post* classification of each firm.

The factors of these both statistical analysis are reported in table 5. We retain, on the one hand, some factors which capture structural contrasts between firms (as size or capital per worker); on the other hand, some factors which are more transitory (as the value added rate of growth). However, in the panel, we have the sector of the main activity of each firm (the nomenclature used distinguishes 18 sectors in manufacturing); this variable is introduced through 17 dummy variables. Finally, we try to relate the probability that a firm belongs to one or another group to the skills of the labor force by putting in the regression, as proxy, the average wages per employee.

The results are reported in table 6. In the models (1), the average wages per employee is not statistically different from zero. In the models (2), this factor is put twice in the regression in a non linear fashion (the gross explanatory variable and the square of the explanatory variable). When we try to fit less specific models (with much more variables), we rapidly face multicollinearity problems. In spite of the high number of observations, the estimated coeffi-

TABLE 6

The Factors to Belong in Group b (High Elasticity of Substitution)

Explanatory variable	OLS on $1 - F(-\widehat{b})$				ML Probit			
	Model (1)		Model (2)		Model (1)		Model (2)	
	Estimate	*CP* *	Estimate	*CP*	Estimate	*CP*	Estimate	*CP*
y	.025	.03	.022	.10	.074	1.1	.064	2.8
dy	.011	65	.0069	77	.065	52	.041	69
ksl	– .014	33	– .015	28	– .044	47	– .059	33
dksl	– .21	.01	– .21	.01	– .68	.01	– .72	.01
w	– .053	28	2.2	.12	– .28	18	13.8	.11
w^2	–	–	– .23	.09	–	–	– 1.46	.09

* *CP* stands for critical probability (by %).

Results for the 17 dummy variables (to capture the sector of the firm) are not reported.

cients are not stable and the critical probabilities are high. We then leave it at the models reported in table 6.

Thus, all things being equal, the elasticity of capital-labor substitution seems to be all the higher as, first, the firms are larger, second, as the firms have reduced their capital per worker, and lastly, as the firms have less skilled work force. The “size effect” we obtain may be interpreted as follows. The bigger firms hold a more diversified portfolio of activities. Then the capital-labor substitution is also achieved by structural effects. On figure 1, one can see the contribution pattern of the wages per employee variables from the OLS regression and from the Probit model. A no monotone relationship is emphasized between the skills and the probability of belonging to group *b*. This probability is high for wages within the 25th percentile and the 50th percentile (denoted resp. by Q_1 and Q_2 on the graphic); it then decreases with respect to the higher values of the average wages. Mind to note that the contribution of those variables is limited. From the 90th percentile to the 10th percentile, all things being equal, this probability increases by only 4 % (from .54 to .58) in the OLS regression and by 7 % (from .53 to .60) in the Probit model.

4 Conclusion

In this paper, we do not (but only indirectly) invalidate the following conjecture: the elasticity of capital-labor substitution is oppositely related to the skills of the work force. We give support, as numerous studies (see HAMERMESH [1993]), to the hypothesis that skilled labor is more complementary with capital than unskilled labor. In order to accurately estimate the elasticity of substitution in production from a panel data of near 800 manufacturing firms, we have specified the following micro-based model.

FIGURE 1
Probability for a Firm to be in Group b with Respect to the Labor Force Skills

Plain line: from the OLS model on $1 - F(-\widehat{b})$
Dotted line: from the ML Probit model

First, at the micro level, the selling price is not observed. To avoid using sectorial deflators, we propose a semi-reduced form obtained conditionally to a specific price setting rule. Next, we argue for estimating a system of two equations as the production function alone is not the best way to reveal the elasticity of substitution in production. Our system can be interpreted as the following recursive firms behavior. On the one hand, firms choose the optimal capital per worker level; on the other hand, the optimal production and price levels. Finally, we do not use to capture the between-individual heterogeneity an *ad hoc* specification; we trace this heterogeneity by means of individualizing a deep production function parameter.

Our results show the presence of two groups of firms, constituted on the grounds of a endogenous selection model. The first group includes the low elasticity of substitution firms; this elasticity is about .6. By contrast, the second group includes the high elasticity of substitution firms; this elasticity is about .9. Eventually, we find that the average skills of the work force is a factor which contributes to accounting for the probability that a firm belongs in either one group or another. This paper then gives a (soft) evidence to support the French current policy to subsidize the "*low wages jobs*".

• References

ARTUS P., BISMUTH C. (1980). – "Substitution et coût des facteurs : un lien existe-t-il ?" *Économie et Statistique,* 127.

AUERBACH A. J. (1979). – "Wealth Maximisation and the Cost of Capital", *Quarterly Journal of Economic,* 93.

AUERBACH A. J. (1983). – "Taxation, Corporate Financial Policy and the Cost of Capital", *Journal of Economic Literature,* 21.

BENASSY J.-P. (1989). – "Market Size and Substitutability in Imperfect Competition : a Bertrand-Edgeworth Chamberlin Model", *Review of Economics Studies,* 56.

BERNARD A. (1977). – “Le coût d'usage du capital productif : une ou plusieurs mesures ?”, *Annales de l'INSEE*, 28.

BUA M., GIRARD Ph., LEGENDRE F., REDONDO Ph. (1990). – “Les effets favorables d'une baisse de la fiscalité des entreprises : une évaluation à partir de données individuelles”, *Économie et Statistique*, 229.

BUA M., GIRARD Ph., LEGENDRE F., REDONDO Ph. (1991). – “Financement, fiscalité et croissance des entreprises industrielles”, *Économie et Prévision*, 98.

Commissariat général du Plan (1993). – *L'économie française en perspective*, La Découverte, Paris.

Commissariat général du Plan (1994). – *Coût du travail et emploi : une nouvelle donne*, La Documentation française, Paris.

Conseil supérieur de l'emploi, des revenus et des coûts (1996). – *L'allègement des charges sociales sur les bas salaires*, Rapport au Premier ministre, La Documentation française, Paris.

DORMONT B. (1983). – “Substitution et coûts de facteurs. Une approche en termes de modèles à erreurs sur les variables”, *Annales de l'INSEE*, 50.

DORMONT B. (1989). – “Introduction à l'économétrie des données de panel”, *Monographies d'Économétrie*, ADRES-CNRS, Paris.

DORMONT B. (1989). – “Petite apologie des données de panel”, *Économie et Prévision*, 87.

DRÈZE J., MALINVAUD E. (1994). – “Croissance et emploi. l'ambition d'une initiative européenne”, *Observations et diagnostics économiques*, (49).

GAGEY F. (1989). – Dégradation et redressement de la situation des entreprises françaises entre 1973 et 1986 : un essai de synthèse, *Economie et Prévision*, 88-89.

HARTLEY M. J. (December 1978). – Estimating Mixtures of Normal Distributions : Comment”, *Journal of American Statistical Association*.

HÉNIN P-Y., POUCHAIN M. (1980). – “Les comportements d'entreprises en déséquilibre comme processus d'adaptation : le modèle RAMAGE”, In P-Y. Hénin, editor, *Etudes sur l'économie en déséquilibre*, Economica, Paris.

HOUTHAKKER H. S. (1956). – The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis, *Review of Economic Studies*, 23(1).

KIEFER N. M. (1978). – “Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model”, *Econometrica*, 46.

KING M. A. (1974). – “Taxation and the Cost of Capital”, *Review of Economic Studies*, 41.

LANGOT F., PUCCI M. (1994). – “Qualifications et subventions à l'emploi différenciées dans un modèle intertemporel d'équilibre général”, Document de travail, Université de Paris I, MAD.

LEGENDRE F. (1992). – “La distribution des rendements d'échelles dans l'industrie : une illustration à partir d'un panel de plus de 700 entreprises françaises”, *Revue Économique*, 43(1).

MAILLARD B., SNEESSENS H. (1993). – “Caractéristiques de l'emploi et du chômage par PCS : France, 1962-1988”, *Document de travail*, Université de Louvain la neuve, IRES.

MAIRESSE J., GRILICHES Z. (1988). – “Hétérogénéité et panels : y a-t-il des fonctions de production stables”, In *Mélanges économiques, Essais en l'honneur d'E. Malinvaud*, *Economica*, Paris.

OBERHOFER W., KMENTA J. (1974). – “A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models”, *Econometrica*, 42(3), pp. 579-590.

ROCHDI F. (1996). – “Technologies multiples et choix de formes fonctionnelles : application au secteur textile tunisien”, *Thèse de doctorat en sciences économiques*, Université de Bourgogne, France.

SNEESSENS H. (1993). – “Courbe de beveridge et demande de qualifications”, *Document de travail*, Université de Louvain la neuve, IRES.

VILLA P., MUET P.-A., BOUTILLIER M. (1980). – “Une estimation simultanée des demandes d'investissement et de travail”, *Annales de l'INSEE*, pp. 38-39.

APPENDIX A

The Data

The data are from the books of French corporations performing in the manufacturing. They are more precisely described in BUA *et al* [1989] or in LEGENDRE [1990]. The following table collects some statistics about the data we used.

Statistic	log Y	log K	log L	log C	log W
Mean	9.4	8.7	4.3	– .95	4.8
Standard Deviation	1.9	2.2	1.8	.33	.35
Q_1 (percentile 25)	7.8	7.0	2.8	– 1.1	4.6
Q_3 (percentile 75)	11.0	1.5	5.8	– .77	5.0
Between variance in total variance	.98	.99	.99	.70	.66

Y is the value added in nominal terms; K the capital stock; L the employed manpower; C the user cost of the capital and; W the average wages per employee.

APPENDIX B

Econometrics

The aim of this appendix is to specified all the econometrics used in this paper. First, we introduce a convenient method to estimate a model with related equations. Second, in the same way, we propose a computational method for specific non linear models. Finally, we expose the two regimes switching model.

B.1 Econometrics of Related Equations

We introduce, in this subsection, a well-known numerical method to estimate a set of two related equations. The system of two equations is as follows

$$\underline{y}_1 = X_1 \underline{a} + \underline{u}_1 \qquad (12)$$

$$\underline{y}_2 = X_2 \underline{a} + \underline{u}_2 \qquad (13)$$

where $\underline{y}_1$ and $\underline{y}_2$ are the two $N \times 1$ vectors corresponding to the dependent variables (where N is the number of observations by equation), X_1 and X_2 the two $N \times p$ matrixes corresponding to the explanatory variables (where p is the number of unknown coefficients), $\underline{a}$ the $p \times 1$ vector of the coefficients, and $\underline{u}_1$ and $\underline{u}_2$ the two $N \times 1$ vectors of the disturbances. The constraints relating equations are taken into account *via* the vector $\underline{a}$, the same for the two equations. This system could be stacked as follows

$$\underline{y} = X\underline{a} + \underline{u} \quad \text{with} \quad \underline{y} = \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix} X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \text{and} \quad \underline{u} = \begin{pmatrix} \underline{u}_1 \\ \underline{u}_2 \end{pmatrix} \qquad (14)$$

Under homocedasticity and non autocorrelation hypotheses, the covariance matrix of the two stochastic vectors $\underline{u}_1$ and $\underline{u}_2$ are

$$\mathrm{V}(\underline{u}_1) = \sigma_1^2 I_N \quad \text{and} \quad \mathrm{V}(\underline{u}_2) = \sigma_2^2 I_N$$

What's more, if one sets a non correlation hypothesis between the disturbances of the two equations, then the covariance matrix of stochastic vector $\underline{u}$ reads as follows

$$\mathrm{V}(\underline{u}) = \Omega = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \otimes I_N \qquad (15)$$

This model also could be interpreted as an univariate model with a special heteroskedasticity pattern: the variance of the first N disturbances is σ_1^2; of the last N, σ_2^2.

In the normality hypothesis, the log-likelihood is then (irrelevant constants have been removed)

(16) $$\ell(\underline{a},\sigma_1^2,\sigma_2^2|\underline{y}) = -\frac{1}{2}\log|\Omega| - \frac{1}{2}(\underline{y} - X\underline{a})'\Omega^{-1}(\underline{y} - X\underline{a}) =$$

$$-\frac{N}{2}(\log\sigma_1^2 + \log\sigma_2^2)$$

$$-\frac{1}{2}\left(\frac{(\underline{y}_1 - X_1\underline{a})'(\underline{y}_1 - X_1\underline{a})}{\sigma_1^2} + \frac{(\underline{y}_2 - X_2\underline{a})'(\underline{y}_2 - X_2\underline{a})}{\sigma_2^2}\right)$$

The first order conditions can be written as

(17) $$a = \left(\frac{1}{\sigma_1^2}X_1'X_1 + \frac{1}{\sigma_2^2}X_2'X_2\right)^{-1}\left(\frac{1}{\sigma_1^2}X_1'\underline{y}_1 + \frac{1}{\sigma_2^2}X_2'\underline{y}_2\right)$$

(18) $$\sigma_1^2 = \frac{(\underline{y}_1 - X_1\underline{a})'(\underline{y}_1 - X_1\underline{a})}{N}$$

(19) $$\sigma_2^2 = \frac{(\underline{y}_2 - X_2\underline{a})'(\underline{y}_2 - X_2\underline{a})}{N}$$

This system is not linear. One can notice that $\underline{a}$ do not relate separately on σ_1^2 and on σ_2^2 but only on σ_1^2/σ_2^2, so we could write the condition 17 in a simpler form; but it is not much enlightening. By contrast, this pattern made us think of using the following numerical procedure to solve the system. First, the second order terms are given (for example, are set at 1 – so $\sigma_{1\,0}^2 = 1$ and $\sigma_{2\,0}^2 = 1$ where the second subscript denotes the iteration number). Then, the following system is repeatedly evaluated

(20) $$\underline{a}_k = \left(\frac{1}{\sigma_{1\,k-1}^2}X_1'X_1 + \frac{1}{\sigma_{2\,k-1}^2}X_2'X_2\right)^{-1}$$

$$\left(\frac{1}{\sigma_{1\,k-1}^2}X_1'\underline{y}_1 + \frac{1}{\sigma_{2\,k-1}^2}X_2'\underline{y}_2\right) \quad k = 1,2,3,\ldots$$

(21) $$\sigma_{1\,k}^2 = \frac{(\underline{y}_1 - X_1\underline{a}_k)'(\underline{y}_1 - X_1\underline{a}_k)}{N} \quad k = 1,2,3,\ldots$$

(22) $$\sigma_{2\,k}^2 = \frac{(\underline{y}_2 - X_2\underline{a}_k)'(\underline{y}_2 - X_2\underline{a}_k)}{N} \quad k = 1,2,3,\ldots$$

$\underline{a}_1$ is the OLS estimator of the stacked model; this estimator is unbiased and consistent but it is not efficient. $\underline{a}_2$ is known as "*estimated generalized least squares*" estimator. In a first stage, the parameters of the covariance matrix are estimated by a consistent method; in a second stage, the generalized least squares rule is used where the unknown parameters are replaced by the estimator of the first stage.

This method is very easy to implement. In fact, $\underline{a}_k$ can be computed as the OLS estimator of the model

(23) $$\underset{\sim}{\widetilde{y}}_k = \widetilde{X}_k \underline{a}_k + \underset{\sim}{\widetilde{u}}_k \quad \text{with}$$

$$\underline{\widetilde{y}}_k = \begin{pmatrix} \dfrac{1}{\sigma_{1\,k-1}} & \underline{y}_1 \\ \dfrac{1}{\sigma_{2\,k-1}} & \underline{y}_2 \end{pmatrix} \quad \text{and} \quad \widetilde{X}_k = \begin{pmatrix} \dfrac{1}{\sigma_{1\,k-1}} & X_1 \\ \dfrac{1}{\sigma_{2\,k-1}} & X_2 \end{pmatrix}$$

It then suffices to transform all the observable variables of the model and next to use the standard OLS formula.

This numerical iterative procedure leads to a sequence of estimators: $\underline{a}_0$, $\underline{a}_1$, $\underline{a}_2$, ... Is a fixed point reached? Is this fixed point the ML estimator ? See OBERHOFER and KMENTA [1974] on the hypotheses to ensure the numerical convergence of the procedure and the equivalence between "iterated estimated GLS" and ML. In our model, we do get such equivalence.

B.2. **Econometrics of Additive Non Linear Component Model**

In this subsection, we explain how one can estimate special kind of non linear models – models with additive non linear component – by an iterative procedure. First, we set the method in the linear case; next, in the additive non linear component models; eventually, in the panel data context.

B.2.1. **Zigzag Procedure for a Linear Model**

One can use an iterative procedure to compute the coefficient of a linear model. For the sake of simplicity, the model is only with two explanatory variables

(24) $$y_i = ax_i + bz_i + u_i \quad i = 1, \ldots, N$$

where y_i is the dependent variable, x_i and z_i the two explanatory variables, a and b the two unknown coefficients (to estimate), and u_i the disturbance term. With matrix notations, we write the model as

(25) $$\underline{y} = a\underline{x} + b\underline{z} + \underline{u} = (\underline{x} \mid \underline{z}) \begin{pmatrix} a \\ b \end{pmatrix} + \underline{u}$$

where $\underline{y}$, $\underline{x}$, $\underline{z}$ and $\underline{u}$ are $N \times 1$ vectors.

The OLS estimator of a and b (denoted by $\widehat{a}$ and $\widehat{b}$) follows from the formula

(26) $$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \left\{ (\underline{x} \mid \underline{z})'(\underline{x} \mid \underline{z}) \right\}^{-1} (\underline{x} \mid \underline{z})'\underline{y}$$

It is then necessary to inverse a 2×2 positive definite matrix to get this estimator. To avoid this inversion, one can propose the following iterative procedure. First, b_0 is given (to any value). Then, we rewrite the model as follows

$$(\underline{y} - b_0\underline{z}) = a\underline{x} + \underline{v}$$

a_1 is thus the MCO estimator of the previous model

$$a_1 = \frac{\underline{x}'(\underline{y} - b_0\underline{z})}{\underline{x}'\underline{x}}$$

We just have to divide two numbers (and not to invert a matrix). Next, b_1 is the MCO estimator of the following model

$$(\underline{y} - a_1\underline{x}) = b\underline{z} + \underline{w} \quad \text{or} \quad b_1 = \frac{\underline{z}'(\underline{y} - a_1\underline{x})}{\underline{z}'\underline{z}}$$

More strictly, the procedure is defined by the two sequences (for $k = 1,2,3,\ldots$)

$$a_k = \frac{\underline{x}'(\underline{y} - b_{k-1}\underline{z})}{\underline{x}'\underline{x}} = \frac{\underline{x}'\underline{y}}{\underline{x}'\underline{x}} - \frac{\underline{x}'\underline{z}}{\underline{x}'\underline{x}}b_{k-1} \quad \text{and}$$

$$b_k = \frac{\underline{z}'(\underline{y} - a_k\underline{x})}{\underline{z}'\underline{z}} = \frac{\underline{z}'\underline{y}}{\underline{z}'\underline{z}} - \frac{\underline{x}'\underline{z}}{\underline{z}'\underline{z}}a_k$$

We now have to study the behavior of those two sequences. After a few substitutions, we get for a

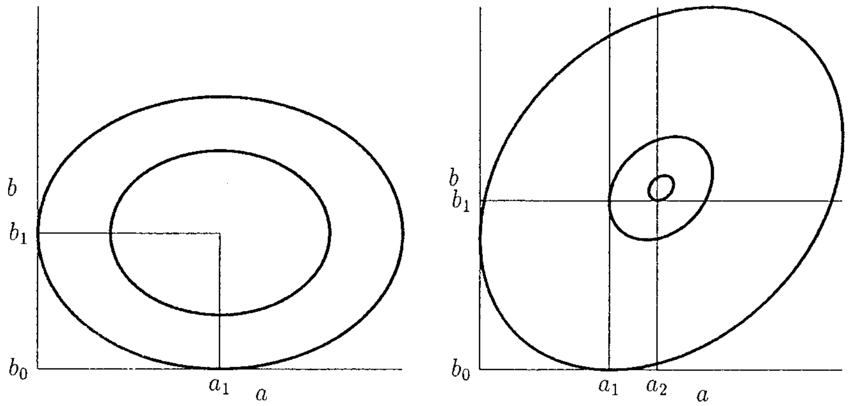
$$a_k = \frac{(\underline{x}'\underline{z})^2}{\underline{x}'\underline{x}\ \underline{z}'\underline{z}}a_{k-1} + \frac{\underline{x}'\underline{y}}{\underline{x}'\underline{x}} - \frac{\underline{x}'\underline{z}\ \underline{z}'\underline{y}}{\underline{x}'\underline{x}\ \underline{z}'\underline{z}}$$

$(\underline{x}'\underline{z})^2/(\underline{x}'\underline{x}\ \underline{z}'\underline{z})$ is the square cosinus of the angle between $\underline{x}$ and $\underline{z}$. It is also the square gross correlation coefficient between $\underline{x}$ and $\underline{z}$. This quantity is then in absolute value less than 1 (it is equal to 1 when $\underline{x}$ and $\underline{z}$ are colinear but the coefficients a and b are not yet identifiable). One can see that the sequence converges and reaches the limit. This limit is really the OLS estimator. The favorable case is when $(\underline{x}'\underline{z})^2/(\underline{x}'\underline{x}\ \underline{z}'\underline{z}) = 0$ (*i.e.* $\underline{x}$ and $\underline{z}$ are orthogonal), then the convergence is reached in one iteration only. This is because the $(\underline{x} \mid \underline{z})'(\underline{x} \mid \underline{z})$ matrix is diagonal and its inverse is

$$\{(\underline{x} \mid \underline{z})'(\underline{x} \mid \underline{z})\}^{-1} = \begin{pmatrix} (\underline{x}'\underline{x})^{-1} & 0 \\ 0 & (\underline{z}'\underline{z})^{-1} \end{pmatrix}$$

The figure 2 gives a suggestive illustration of the zigzag procedure. On this figure, we attempted to point out the criterion (the iso-SSR – sum of squared residuals). We chose to take b_0 equated to 0 and to distinguish the favorable case and the usual case (where $\underline{x}$ and $\underline{z}$ are correlated). As the criterion is a quadratic form, the iso-SSR curves are elliptical. In the favorable case, the axes of the ellipses are parallel to the axes of the graphic. In the usual case, on the graphic, a_1 is the point where the straight line $b = 0$ and an iso-SSR curve are tangential; then, b_1 is the point where the straight line $a = a_1$ and an

FIGURE 2
Zigzag Procedure (Iso Sum of Squared Residuals Curves)



Favorable case : $\underline{x}$ and $\underline{z}$ are orthogonal   Usual case : $\underline{x}$ and $\underline{z}$ are correlated

iso-SSR curve are tangential; etc. This is why we call this numerical resolution procedure the "*zigzag procedure*".

B.2.2. **Zigzag Procedure of Additive Non Linear Component Model**

The latter procedure should be adapted to estimate a non linear model. Let us assume that this model is of the following kind

$$y_i = ax_i + bf_i(b) + u_i \quad i = 1, \ldots, N \tag{27}$$

The symbols are as before: notably, a and b are the two unknown coefficients to estimate. The non linear component $bf_i(b)$ is additive just as the disturbances. We suppose the $f_i(\cdot)$ function to be continuous and derivable $\forall i$.

This model is rather general. For example, the following non linear model

$$y_i = ax_i + (z_i)^b + u_i$$

can be rewritten as

$$y_i = ax_i + b\,\frac{(z_i)^b}{b} + u_i \quad \text{with} \quad f_i(b) = \frac{(z_i)^b}{b}$$

Under usual hypotheses, the ML estimator of the a and b coefficients are the solution to the program

$$\min_{a \text{ and } b} \sum_{i=1}^{N} \{y_i - (ax_i + bf_i(b))\}^2 \tag{28}$$

We assume that, for a given (set at $\overline{a}$), it is fairly easy to get b as the solution to the program

$$\min_{b} \sum_{i=1}^{N} \{y_i - (\overline{a}x_i + bf_i(b))\}^2$$

as, for example like in this paper, b is in a predetermined set. It then suffices to try different values to approximate the solution.

The iterative procedure used is yet defined trough the two sequences

$$b_k = \arg \min_{b} \sum_{i=1}^{N} \{y_i - (a_{k-1}x_i + bf_i(b))\}^2 \text{ and}$$

$$a_k = \frac{\underline{x}'(\underline{y} - b_k \underline{f}_k)}{\underline{x}'\underline{x}} = \frac{\underline{x}'\underline{y}}{\underline{x}'\underline{x}} - \frac{\underline{x}'\underline{f}_k}{\underline{x}'\underline{x}} b_k$$

where $\underline{f}_k$ is the $N \times 1$ vector with the term i equals to $f_i(b_k)$. b_k can be expressed as

$$b_k = \frac{\underline{f}_k'(\underline{y} - a_{k-1}\underline{x})}{\underline{f}_k'\underline{f}_k} = \frac{\underline{f}_k'\underline{y}}{\underline{f}_k'\underline{f}_k} - \frac{\underline{x}'\underline{f}_k}{\underline{f}_k'\underline{f}_k} a_{k-1}$$

So, we can rewrite the sequence for a

$$a_k = \frac{(\underline{x}'\underline{f}_k)^2}{\underline{x}'\underline{x}\ \underline{f}_k'\underline{f}_k} a_{k-1} + \frac{\underline{x}'\underline{y}}{\underline{x}'\underline{x}} - \frac{\underline{x}'\underline{f}_k\ \underline{f}_k'\underline{y}}{\underline{x}'\underline{x}\ \underline{f}_k'\underline{f}_k}$$

or else

$$a_k = \lambda_k a_{k-1} + \mu_k \quad \text{with} \quad \lambda_k = \frac{(\underline{x}'\underline{f}_k)^2}{\underline{x}'\underline{x}\ \underline{f}_k'\underline{f}_k} \quad \text{and} \quad \mu_k = \frac{\underline{x}'\underline{y}}{\underline{x}'\underline{x}} - \frac{\underline{x}'\underline{f}_k\ \underline{f}_k'\underline{y}}{\underline{x}'\underline{x}\ \underline{f}_k'\underline{f}_k}$$

As in the linear case, λ_k is, in absolute value, less than 1. This sequence then converges. The limit is a solution to the maximum likelihood.

B.2.3. **Zigzag Procedure in Panel Data Context**

Two subscripts are necessary to denote panel data. The first subscript (usually denoted by i, $i = 1, \ldots, N$) distinguishes the individual dimension; the second subscript (usually denoted by t, $t = 1, \ldots, T$) distinguishes the temporal dimension. In this paper, in order to capture the differences between individuals, we propose a model including a non linear individual component. Then the latter is

(29) $$y_{it} = ax_{it} + b_i f_{it}(b_i) + u_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T.$$

a is like a macro coefficient, the same for all the individuals in the panel; by contrast, the b_i are micro coefficients.

The criterion to minimize (to maximize the likelihood) is

$$\min_{(a,\,(b_i)_i)} \sum_{i=1}^{N}\sum_{t=1}^{T}\{y_{it} - (ax_{it} + b_i\, f_{it}(b_i))\}^2 \tag{30}$$

For a given (set at $\overline{a}$), the program is a separable one. It is sufficient to find each b_i which is the solution to the programs

$$\min_{b_i} \sum_{t=1}^{T}\{y_{it} - (\overline{a}x_{it} + b_i\, f_{it}(b_i))\}^2 \quad \text{for} \quad i = 1,\ldots,N$$

We suppose that it is not too hard to find the solution (with a trial procedure for example – in the paper, the coefficient is in the]0, 1[interval and we use a grid to try different values).

The iterative procedure is then defined by the following sequences (where the subscript k denotes the iteration number)

$$b_{i\,k} = \arg\min_{b_i} \sum_{t=1}^{T}\{y_{it} - (\overline{a}_{k-1}x_{it} + b_i\, f_{it}(b_i))\}^2 \quad \text{for} \quad i = 1,\ldots,N$$

and

$$a_k = \frac{\underline{x}'(\underline{y} - \underline{z}_k)}{\underline{x}'\underline{x}}$$

where the $NT \times 1$ vector $\underline{z}_k$ is with the term i equals to $z_{it\,k} = b_{i\,k}\, f_{it}(b_{i\,k})$.

As before, the procedure converges. The limit satisfies the first order conditions of the maximum likelihood. The advantages of this procedure are obvious: a drastic decrease of the dimension of the unknown parameters space. Let us denote by M the number of trial values for each b_i; if the program is not separable, one needs to evaluate M^N expressions to find the solution. By contrast, with our method, one needs to only evaluate $N \times M$ expressions.

B.3. Econometrics of Two Regimes Switching Model

In this last subsection, we first expose a very general two regimes switching model. Next, we show how it should be used in a panel data context.

B.3.1. A General Model

We closely follow HARTLEY [1978] but we use slightly different notations. We first assume that the observations can be in either one or another group; *a priori* two different regression models are specified to explain the same dependent variable. For example, one has cross section data on consumption and wants to fit two alternative models: a life cycle model and a liquidity constraint model.

If additional information is available on the way the population splits into the two groups, one just has to run two different regressions on each subpopulation. When such information is not directly available, the following model of endogenous switching can be proposed

$$(31) \qquad y_i = \begin{cases} a^a x_i^a + u_i^a & \text{if} \quad bz_i + v_i < 0 \\ a^b x_i^b + u_i^b & \text{if} \quad bz_i + v_i \geqslant 0 \end{cases} \qquad i = 1,\dots,N.$$

where $\underline{y}$ is the $N \times 1$ vector of dependent variable. $\underline{x}^a$ and $\underline{x}^b$ are the $N \times 1$ vectors of explanatory variables for each regime (for the sake of simplicity, we restrict the presentation to one variable per regime). The $N \times 1$ vector $\underline{z}$ is the explanatory variable of the auxiliary equation which could help to select the regime. On the first order, there are three unknown coefficients to estimate: a^a, a^b, and b. u_i^a, u_i^b and v_i are three random disturbances with zero means and constant variances. They are not correlated (either autocorrelated or between them). Thus, the covariance matrixes are

$$\mathrm{V}(\underline{u}^a) = \sigma^{a2} I_N \quad \mathrm{V}(\underline{u}^b) = \sigma^{b2} I_N \quad \text{and} \quad \mathrm{V}(\underline{v}) = I_N$$

The variance of $v_i \; \forall \, i$ is normalized to 1 so the coefficient b can be identified.

This model is a multi-purpose model; several suggestive cases can be generated. First, one can assume $x_i^a = x_i^b = 1 \; \forall \, i$; then, the sample is a mix of two stochastic laws. (with, resp., a^a and a^b means, and σ^{a2} and σ^{b2} variances). Next, one can assume $z_i = 1 \; \forall \, i$; then no additional information is available to classify the observations. *Ex post*, we can only define a global probability for each regime to occur. Eventually, one can assume $x_i^a = x_i^b \; \forall \, i$; in this case, the model is a model of structural change on the parameters.

Under normality hypothesis, the log-likelihood function for the model can be written as

$$\ell(a^a, a^b, b, \sigma^{a2}, \sigma^{b2} \mid \underline{y}) =$$

$$\sum_{i=1}^{N} \log\{F(-bz_i) f[(y_i - a^a x_i^a)/\sigma^a]/\sigma^a +$$

$$[1 - F(-bz_i)] f[(y_i - a^b x_i^b)/\sigma^b]/\sigma^b\}$$

where $f(\cdot)$ and $F(\cdot)$ denote resp. the p.d.f. and the c.d.f. of a zero mean, unit variance normally distributed variable. The likelihood function is, however, not bounded: regime a, for example, could be estimated as accurately as one wants and the likelihood is then equal to infinity. But the algorithm proposed by HARTLEY [1978] seems to circumvent these difficulties.

The derivative with respect to a^a is equal to

$$(33) \qquad \sum_{i=1}^{N} F(-bz_i) x_i^a (y_i - a^a x_i^a) f[(y_i - a^a x_i^a)/\sigma^a]/\sigma^{a3} \,/\, g_i$$

with

$$g_i = F(-bz_i) f[(y_i - a^a x_i^a)/\sigma^a]/\sigma^a +$$

$$[1 - F(-bz_i)] f[(y_i - a^b x_i^b)/\sigma^b]/\sigma^b \quad \forall \, i = 1,\dots,N$$

The derivative with respect to a^b exhibits the same pattern *mutadis mutandis* – mind to substitute $1-F(-bz_i)$ for $F(-bz_i)$. The derivative with respect to σ^{a2} is a little more complicated

$$(34)\quad \sum_{i=1}^{N} F(-bz_i)\, f[(y_i - a^a x_i^a)/\sigma^a][(y_i - a^a x_i^a)^2/\sigma^{a2} - 1]/\sigma^{a3}/g_i/2$$

Eventually, the derivative with respect to b is

$$(35)\quad \sum_{i=1}^{N} -z_i\, f(-bz_i)\{ f[(y_i - a^a x_i^a)/\sigma^a]/\sigma^a - f[(y_i - a^b x_i^b)/\sigma^b]/\sigma^b\}/g_i$$

Again, the system of the five first order conditions is not linear. To make matters worse, we cannot get an explicit non reduced form for b. However, we can rewrite the latter equation as

$$-b\sum_{i=1}^{N} z_i^2 + \sum_{i=1}^{N} z_i$$

$$\{bz_i - f(-bz_i)[f[(y_i - a^a x_i^a)/\sigma^a]/\sigma^a - f[(y_i - a^b x_i^b)/\sigma^b]/\sigma^b]/g_i\}$$

which leads to an explicit non reduced solution to b.

The following computational iterative resolution can be suggested. At the beginning of iteration k, one has common values for all unknown parameters. Those values are denoted by a_{k-1}^a, a_{k-1}^b, b_{k-1}, σ_{k-1}^{a2}, and σ_{k-1}^{b2}. Let us introduce the following notation

$$g_{i\,k-1} = F(-b_{k-1}z_i)\, f[(y_i - a_{k-1}^a x_i^a)/\sigma_{k-1}^a]/\sigma_{k-1}^a +$$

$$[1-F(-b_{k-1}z_i)]\, f[(y_i - a_{k-1}^b x_i^b)/\sigma_{k-1}^b]/\sigma_{k-1}^b \quad \forall\, i = 1,\ldots,N$$

and define, first, the diagonal $N\times N$ W_{k-1}^a matrix with the term i as follows

$$W_{i\,k-1}^a = F(-b_{k-1}z_i)\, f[(y_i - a_{k-1}^a x_i^a)/\sigma_{k-1}^a]/\sigma_{k-1}^a/g_{i\,k-1},$$

next, the diagonal W_{k-1}^b matrix defined *mutadis mutandis* as W_{k-1}^a, and, last, the $N\times 1$ $\underline{d}_{k-1}$ vector with the term i as follows

$$d_{i\,k-1} = b_{k-1}z_i - f(-b_{k-1}z_i)\{ f[(y_i - a_{k-1}^a x_i^a)/\sigma_{k-1}^a]/\sigma_{k-1}^a -$$

$$f[(y_i - a_{k-1}^b x_i^b)/\sigma_{k-1}^b]/\sigma_{k-1}^b\}/g_{i\,k-1}$$

It is then possible to define the resolution method by means of the following sequences

$$(36)\quad a_k^a = (\underline{x}^{a\prime} W_{k-1}^a \underline{y})/(\underline{x}^{a\prime} W_{k-1}^a \underline{x}^a) \quad k = 1,2,3,\ldots$$

$$(37)\quad a_k^b = (\underline{x}^{b\prime} W_{k-1}^b \underline{y})/(\underline{x}^{b\prime} W_{k-1}^b \underline{x}^b) \quad k = 1,2,3,\ldots$$

$$(38)\quad b_k = (\underline{z}' \underline{d}_{k-1})/(\underline{z}'\underline{z}) \quad k = 1,2,3,\ldots$$

(39) $\sigma_k^{a2} = (\underline{y} - a_k^a \underline{x}^a)' W_{k-1}^a (\underline{y} - a_k^a \underline{x}^a) / (\sum_{i=1}^{N} W_{i\,k-1}^a) \quad k = 1,2,3,\ldots$

(40) $\sigma_k^{b2} = (\underline{y} - b_k^b \underline{x}^b)' W_{k-1}^b (\underline{y} - b_k^b \underline{x}^b) / (\sum_{i=1}^{N} W_{i\,k-1}^b) \quad k = 1,2,3,\ldots$

It is easy to generalize when there is several explanatory variables in each regression regime; for example, let X^a be the first regime $N \times p^a$ matrix of the explanatory variables (where p^a is the number of the explanatory variables), and $\underline{a}^a$ be the $p^a \times 1$ vector of the unknown parameters. The sequence which defines $\underline{a}^a$ is

$$\underline{a}_k^a = (X^{a\prime} W_{k-1}^a X^a)^{-1} X^{a\prime} W_{k-1}^a \underline{y}$$

In the same vein, one could generalize when there are several explanatory variables in the auxiliary regression equation. Z is the matrix of the explanatory variables and $\underline{b}_k$ is then defined by $(Z'Z)^{-1} Z'\underline{d}_{k-1}$.

In others respects, on can note that the procedure is very easy to implement. The formulas which give a_k^a and a_k^b can be interpreted as weighted OLS. It is then sufficient to adequately transform the variables and then to use the standard OLS computational routines.

We can interpret this iterative procedure as the revision among iterations of the probability that each observation is in either one group or another. One can initiate the procedure with, for a_0^a, a_0^b, σ_0^{a2} and σ_0^{b2}, the OLS estimator on the overall population. One can also choose, for b, zero: the probabilities (to be in the first or the second group) are then equal to .5 and .5. Then, at iteration 0, the residual p.d.f. (denoted by $\alpha_{i\,0}$ and $\beta_{i\,0}$) for each group are

$$\alpha_{i\,0} = f[(y_i - a_0^a x_i^a)/\sigma_0^a]/\sigma_0^a \quad \text{and} \quad \beta_{i\,0} = f[(y_i - a_0^b x_i^b)/\sigma_0^b]/\sigma_0^b$$

The weight for the observation i is

$$W_{i\,0}^a = \frac{\alpha_{i\,0}}{\alpha_{i\,0} + \beta_{i\,0}} \quad \text{and} \quad W_{i\,0}^b = \frac{\beta_{i\,0}}{\alpha_{i\,0} + \beta_{i\,0}} = 1 - W_{i\,0}^a$$

Thus, the weight is related to the accuracy of the fit in each regime. For example, if the estimated residual for the first group is close to zero and if the estimated residual for the second group is away from zero, then $\alpha_{i\,0}$ is very great than $\beta_{i\,0}$ and, in this way, $W_{i\,0}^a$ is close to 1 and $W_{i\,0}^b$ close to 0. The values of the a^a and a^b coefficients are revised *via* a weighted regression, the contribution of each observation being to extend to the relative goodness of the fit in each regime at the previous iteration.

Eventually, b is evaluated, at iteration 1, with OLS in a model where the dependent variable is

$$d_{i\,0} = -f(0)(W_{i\,0}^a - W_{i\,0}^b)$$

The $\underline{z}$ variable must then explain, for each observation, the relative accuracy of the fit for each group.

KIEFER [1978] has showed that a root of the likelihood equations corresponding to a local maximum is consistent, asymptotically normal, and efficient. Let us now quote HARTLEY [1978]:

> Limited Monte Carlo experiments with this algorithm indicate that convergence to a solution of the likelihood equations corresponding to a local maximum of L always obtains, that point estimates are very close to the true parameter values (for moderate sample sizes of 100 observations) [...] If there are multiple roots, while the algorithm converges to a solution of the likelihood equations, there is no guarantee that the "*correct*" root (corresponding to a consistent estimator) will obtain.

In our various experiments, we only check that the likelihood value is growing with each iteration. We never find the following case

$$\ell(a_k^a, a_k^b, b_k, \sigma_k^{a2}, \sigma_k^{b2} | \underline{y}) < \ell(a_{k-1}^a, a_{k-1}^b, b_{k-1}, \sigma_{k-1}^{a2}, \sigma_{k-1}^{b2} | \underline{y})$$

We always improve the criterion among procedure iterations. To speed the process, we slightly alter the definition of the b_k sequence. We interpret the original sequence as giving a direction for movement, denoted by δ_k

$$\delta_k = (\underline{z}' \underline{d}_{k-1})/(\underline{z}'\underline{z}) - b_{k-1}$$

We then try different steps and we retain the step for which the likelihood increase is the highest. Rigorously, this modification can be specified as

$$b_k = b_{k-1} + \lambda_k \delta_k \quad \text{with}$$

$$\lambda_k = \arg \max_{\lambda \in \{.5, 1, 2, 5, 10\}} \ell(a_k^a, a_k^b, b_{k-1} + \lambda \delta_k, \sigma_k^{a2}, \sigma_k^{b2} | \underline{y})$$

For a step equal to 1, one gets $b_k = (\underline{z}' \underline{d}_{k-1})/(\underline{z}'\underline{z})$ as before; for a different step, this alteration speeds the process.

B.3.2. **Two Regimes Switching Model in a Panel Data Context**

In this paper, we use the above model in a special way. Without individual and/or temporal specific coefficients, the model is

$$(41)\quad y_{it} = \begin{cases} a^a x_{it}^a + u_{it}^a & \text{if} \quad bz_{it} + v_{it} < 0 \\ a^b x_{it}^b + u_{it}^b & \text{if} \quad bz_{it} + v_{it} \geqslant 0 \end{cases} \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T.$$

We want to only state, as additional information, that each individual remains in the same regime at all times. An individual is then "type a" or "type b"'. We do not want to give other information but we want to assign *ex post* to each individual a probability to be in one or another group. We then opt for the introduction of a individual specific effect in the auxiliary equation.

Thus, the auxiliary equation is

$$b_i + v_{it}$$

Let us denote by $\underline{\ell}_T$ the $T \times 1$ vector of ones and by $\underline{b}$ the $N \times 1$ vector of b_i term. The auxiliary equation may be written compactly as

$$Z\underline{b} + \underline{v} \quad \text{with} \quad Z = I_N \otimes \underline{\ell}_T$$

where the observations are sorted out by individual and then by time period.

The numerical procedure is still valid but it is then possible to reduce the computational burden. The expression which gives $\underline{b}_k$ is always

$$\underline{b}_k = (Z'Z)^{-1} Z' \underline{d}_{k-1}$$

But as $(Z'Z)^{-1}$ is equal to $\frac{1}{T} I_N$, we can compute each $b_{i\,k}$ by using

$$b_{i\,k} = \frac{\sum_{t=1}^{T} d_{it\,k-1}}{T}$$

It is then not necessary to inverse the huge $N \times N$ $Z'Z$ matrix. This hint is like the transformation in terms of deviations around the mean in the individual dummy variable model.