Irina PEAUCELLE
CEPREMAP

September 1988

ON SOME DIFFICULTIES IN EMPIRICAL TESTING OF ECONOMIC THEORIES

Paper prepared for the conference on International Perspectives on Profitability and Accumulation.
New York University, September 16-18, 1988.

I - INTRODUCTION

The aim of economic theories is the analyse and explanation of phenomenon such as production, distribution and consumption. They have the form of logical structures or models relating a certain number of variables, specifying the causality connection and describing the conditions of application. That logical structures are to be compatible in general with some observed phenomenon (1). Still the necessity to dispose of model easy to handle constrains the initial selected representation to be oversimplified. It can be obtain by a decrease of the number of pertinent variables, the other be supposed fixed -the usual condition of "ceteris paribus" or by the choice of a given level of aggregation, for instance, the production may be analysed at the firm, sector or country level, ..., or, finally, by various restrictive assumptions concerning the dynamic specification : the behavioral relations can be assumed time independent, the trends assumed to be constant...

These various simplifications have the undeniable advantage to give a model which is both tractable and operational for simulation purposes, but they can move the representation far away from the real phenomenon, which was at the basis of the theoretical ex post construction. It is important then to compare, the theoretical formulation to the realms of possibility corresponding to the period of interest. This empirical check has to be performed cautiously, since in the absence of controlled experiment one can not hope that the available datas meet exactly all underlying assumptions of the model.

a) Definition of variables

The concepts and the variables have often different definitions in Economics and in Statistics : for example an economic agent (consumer,

----

(1) A systematic introduction to the problem of verification of social and economic assumptions was given by N. KONDRATIEV in 1926.

firm...) is defined, from an economic point of view, as a decision unit, but, from statistical point of view, it is a sample survey unit. The same difficulties appear for the variables, and they can even be more important, because some variables have no correspondent statistical definition, or are not observable at all. In particular it is the case of variables such as expectations [permanent income], or notions such as Marxian plus-value.

b) Disaggregation level

Obviously it would be better to retain a level of disaggregation in empirical study close to the level used in economic theory. The economic theories are often developed for some level, but are applied without modifications at a more aggregated level. Implicitly such a practice may only be justified if there exists some representative individual that is if the conditions for perfect aggregation are satisfied. These conditions, frequently replaced by the stronger one of homogeneity of individuals, is likely not satisfied in practice. Then the results of empirical testing have to be interpreted taking into account the possible aggregation biases.

c) Dynamic aspect

The theories are often tested from historical datas. These datas have their own evolution which may be partially inconsistent with the model. Generally the variables used as control variables for economic policy are not necessary the same as the ones chosen as exogenous in the model, or this set of variables may vary during the period of observation and this may induce a modification in the causality directions. Moreover the economic theories introduce very seldom a description of the evolutions of agents behaviour. Therefore it is not surprising that theories elaborated at one period of time may be unadequate when applied to another period.

Below we propose to test the adequation of one of the simplest specifications for the determination of profit rate ; the period of analysis is 1960-1983 and the tests are performed for the seven OECD countries. Several models are considered depending on the kind of heterogeneity between countries. First we write and estimate the model as if homogeneity exists. In

this case we have likely errors and the estimated coefficients have to be interpreted taking into account these errors. We know that for linear models the heterogeneity biases may be important, but that it is often possible to interprete the coefficients in a different way in order to correct from this bias. Another solution which is usually followed is to modify the initial specification in order to take into account the heterogeneity ; the new specification is more complicated than the previous one, however the interpretation can be done directly without bias and moreover such a formulation may allow an evaluation of the importance of individual heterogeneity.

It is clear that if heterogeneity between countries exists, if the theoretical postulat is fundamentally correct, but if the model is estimated and interpreted under the homogeneity assumption, we might erroeously reject this theory.

II - HETEROGENEITY IN EMPIRICAL STUDY (AN EXAMPLE)

We use in this paper formulations of trends in the rate of profit in manufacturing close to those of T. WEISSKOPF (1985). He investigates the idea that the rate of profit trend depends on the workers' incentive. The incentive is firstly the level of real wage, but it changes also when the worker expects that it might become unemployed. WEISSKOPF estimated the regressions with wage and unemployment as explanatory variables for eight countries. The expected signs are positive for the unemployment variable and negative for the wage variable. The estimations confirm these ideas only in some cases, essentially for the following countries : United-Kingdom, Italy and Canada.

In the same spirit PEAUCELLE-PETIT (1988) proposed a specification based on the assumption of representative individual and expected signs are obtained.

How to explain these differences between the two approaches ? A possible answer is the omission of heterogeneity phenomenon in the second approach and thus it seems useful to have a precise analysis of the discrepancies between the results country by country and the average result.

a) <u>Variability of profit rate</u>

The variability of datas in cross countries analysis is well known. It both appears in longitudinal and in cross sections. The total variance of a variable Y_{it} (with $i=1, \ldots, N$; $t=1, \ldots, T$) where i denotes the country, and t denotes the year can be decomposed into the individual variability and the temporal variability :

$$\underbrace{\frac{1}{NT}\sum_i \sum_t (Y_{it} - Y_{..})^2}_{\text{TOTAL}} = \underbrace{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{T} \sum_t (Y_{it} - Y_{i.})^2}_{\text{TEMPORAL}} + \underbrace{\frac{1}{N}\sum_{i=1}^{N} (Y_{i.} - Y_{..})^2}_{\text{INDIVIDUAL}}$$

In this paper we analyze seven OECD countries : Canada, France, Italy, Japan, United Kingdom, United States and West Germany between 1960 and 1983. The variability of logarighm of profit rate is due for 48 % to temporal variability and for 52 % to individual variability.

b) <u>Time trends of variables</u>

Another way of evaluating the two components of heterogeneity (temporal and individual) consists in using a descriptive model in which these two effects are clearly distinguished. For instance, if we assume that the logarithm of profit rate, denoted by y, has a linear trend, it is possible to write an approximate formulation such as :

$$Y_{it} = \alpha_i t + \delta_i .$$

If this formulation is compatible with the available data, we have the temporal variability which is equal to :

$$\frac{1}{N}\sum_{i=1}^{N} \frac{1}{T} \sum_t \left(y_{it} - y_{i.}\right)^2 = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{T} \sum_t \left(\alpha_{it} + \delta_i - \alpha_i \frac{T+1}{2} - \delta_i\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} \alpha_i^2 \frac{(T-1)^2}{12}$$

$$= \frac{(T-1)^2}{12} \left[\frac{1}{N} \Sigma (\alpha_i - \alpha_.)^2 + \alpha_.^2\right].$$

It is a simple function of time and of the average and variability of trend coefficients α_i.

In a similar way the individual variability is approximately equal to :

$$\frac{1}{N}\sum_{i=1}^{N} \left(y_{i.} - y_{..}\right)^2 = \frac{1}{N}\sum_{i=1}^{N} \left(\alpha_i \frac{T+1}{2} \delta_i - \alpha_. \frac{T+1}{2} - \delta_.\right)^2$$

$$= \frac{(T+1)^2}{4} \frac{1}{N}\sum_{i=1}^{N} \left(\alpha_i - \alpha_.\right)^2 + (T+1) \frac{1}{N}\sum_{i=1}^{N} \left(\alpha_i - \alpha_.\right)\left(\delta_i - \delta_.\right) +$$

$$+ \frac{1}{N}\sum_{i=1}^{N} \left(\delta_i - \delta_.\right)^2$$

The values of the individual coefficients α_i and δ_i are the following ones :

COUNTRY	CA	FR	IT	JAP	UK	US	WG
α_i	-.013	-.022	-.043	-.029	-.024	-.032	-.03
δ_i	2.3	2.6	2.8	3.3	2.2	2.9	2.9

We deduce the statistical summaries such as average variability and covariability :

$\alpha_{.}$	-.028
$\frac{1}{N}\Sigma\ (\alpha_i - \alpha_{.})^2$	0.74×10^{-4}
$\frac{1}{N}\Sigma\ (\delta_i - \delta_{.})^2$	0.12
$\frac{1}{N}\Sigma\ (\alpha_i - \alpha_{.})(\delta_i - \delta_{.})$	$-.17 \times 10^{-2}$

The decomposition of variability deduced from the trend model is then the following :

TOTAL VARIABILITY	TEMPORAL VARIABILITY	INDIVIDUAL VARIABILITY	RESIDUAL VARIABILITY
0.16	0.03 (19%)	0.09 (56%)	0.04 (25%)

We see that the individual variability remains the most important one and the temporal variability from the previous analysis is made up of temporal and residual variability.

The analysis of trends may also be performed for the other variables of interest, i.e. logarithm of wages (denoted W_{it}) and logarithm of unemployment (denoted by U_{it}). The models are :

$$y_{it} = \alpha_i t + \delta_i,$$
$$U_{it} = \beta_i t + \varepsilon_i,$$
$$W_{it} = \gamma_i t + \sigma_i.$$

For every country the coefficients α_i are negative and the coefficients β_i, γ_i are positive. On average they are equal to :

$\alpha_{.} = -.03$ $\beta_{.} = .03$ and $\gamma_{.} = .07$

However for some countries the assumption of a linear trend is perhaps inadequate and it may be better to distinguish some subperiods. For instance we give below the evolution of the three variables for Japan during the period 1960-1970. We observe the slowdown of profit rate, but the unemployment stay stable and the wages even drop until 1965. After 1970 the profit rate stabilizes and wages and unemployment rise rapidly (cf. figures in appendix). Another special case concerns the U.S. economy where the profit rate evolution is cyclical and the rate of unemployment drop till 1969.

c) Links between the variables

The previous analysis of variability is essentially a descriptive one. Now we are interested in the links between the profit rate and the two possible explanatory variables : wage and unemployment.

In the theory, the profit rate has a slow-down trend due to rise of organical structure of capital. K. MARX has enumerated the factors which influence in both ways the speed of its slowing down. We first will focus on one of the most important among them : the wages.

The profits and the wages are the components of a new created value and their respective volume can increase if at a unit of time the labour creates a bigger value. Obviously, one of two components may rise to the detriment of the other one whatever be the value created. We get that the wages play an ambiguous role in determination of profit volume. On the one hand, its increase can cause the decrease of profits in absolute value, on the other hand, one hope that its increase comes with increase of labour intensity which is equivalent to bigger value creation.

The slowing-down of profit rate is the relative decrease of profit in comparison with capital. Naturally, if profits rise in absolute value, it produces the braking of profit rate slowing-down.

One another factor of labour intensity increasing may be the unemployment (the fear to loose a job). The employed workers creat a bigger value in this situation and at the same time the share wages/profits may be in favour of profits. But we do not have to forget that unemployment can mean also the fall in absolute volume of labour. That is another part of the story of its influence on profit rate, the opposite one.

If the rise of unemployment (or wages) implies the braking of slowing-down of profit rate, then the coefficients b_i in the following relations should be positive and the coefficients c_i should be negative :

$$Y_{it} = b_i x_{it} + c_i t + a_i,$$

where x_{it} denote u_{it}or w_{it}. Since the coefficients are different, this model is disaggregated by country. We observe that the expected signs are the right ones only in 43 % of cases with unemployment as explanatory variable, and more less (29%) with real wage variable.

COUNTRY	EXPLANATORY VARIABLES			
	UNEMPLOYMENT		REAL WAGE	
	b_i	c_i	b_i	c_i
CANADA	-	+	-	+
FRANCE	+	-	+	-
ITALY	+	-	-	+
JAPAN	-	+	-	-
U.K.	+	-	-	+
U.S.A.	-	-	+	-
WEST GERMANY	-	-	-	-

Therefore there exist different influencis as concerning the impact of the two chosen factors on the profit rate and these influences depend on the country. But can we have some general idea of that phenomenon ?

d) <u>Approximate formulations</u>

In order to consider all individual differences we have chosen previously the coefficients of the relations depending on individual and we have retained the disaggregated model :

$$Y_{it} = A_i + x_{it} \; b_i + C_i t + V_{it} \qquad E\, V_{it} = 0 \; ; \; i=1, \dots N \; ; \; t = 1, \dots T.$$

We suppose here the disaggregated model well specified. The coefficients a_i and b_i are time independent. It would be possible to also introduce time heterogeneity, but in this note we focus on the individual component of heterogeneity.

Frequently the approches by "representative individual" or by "aggregated" model are used. In the first case the coefficients are assumed to be constant for all individuals (countries). This formulation can be seen as a submodel for the disaggregated model, obtained by assuming that the homogeneity condition is satisfied. In the second case the relations are written directly between the aggregated variables.

It is known that when the coefficients a_i, b_i, c_i can be considered as independently drawn from the same distribution of mean A, B, C and if they are independent from the values of the exogenous variables x_{it}, than the disaggregated model is both compatible with the model for "representative individual" and with the "aggregated" model. But when heterogeneity is present the two simplified models are not compatible with the disaggregated model. In particular, the residual of the simplified model is correlated with the exogenous variable and this implies biases for the O.L.S. estimators. In order to avoid estimation biases it is sometimes possible to introduce partial heterogeneity. For instance we can estimate a model in differences :

$$Y_{it} - Y_{i,t-1} = (x_{i,t} - x_{i,t-1})\, b^{**} + C + V_{it}^{**}$$

or consider the within relation,

$$Y_{it} - Y_{i.} = (x_{it} - x_{i.})\, b^{***} + V_{it}^{***}$$

where $Y_{i.}$ is the temporal mean of values Y_{it}. For instance, the specification in differences is compatible with a model :

$$Y_{it} = x_{it} b^{**} + a_i + Ct + V_{it}$$

where heterogeneity exists on the constant term.

The study of the O.L.S. estimators for these different kinds of model is presented in GOURIEROUX-PEAUCELLE (1988). The biases of estimation for each model can be studied explicitly. We have to compare the estimated value with a descriptive summary of the individual coefficients b_i. This summary will be chosen as an average of individual coefficients :

$$B = \frac{1}{n} \sum_j b_j$$

In the sequel we present the estimations for the following regressions :

1) Disaggregated models : $Y_{it} = a_i + b_i \ X_{it} + C_i t + V_{it}$

2) Representative individual model: $Y_{it} = a + b \ X_{it} + Ct + V^*_{it}$

3) Model in differences : $\Delta Y_{it} = b^{**} \ \Delta \ X_{it} + C + V^{**}_{it}$

4) Within model : $Y_{it} - Y_{i.} = (X_{it} - x_{i.}) b^{***} + V^{***}_{it}$

Our estimations are based on the time series data on whole period for the time explanatory variables. When the exogenous variable is unemployed the estimations have also been performed on subperiod 1970-1983. By distinguishing this subperiod we will try to eliminate the countertendencies observed for Japan.

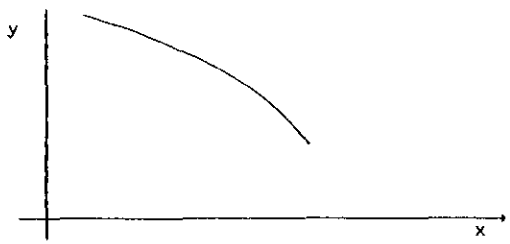
We might expect that the estimator in the equation of representative individual, which can contain the effect of neglected heterogeneity (specific ignorance represented by a_i) will perceptibly differ from the other ones. We see also that the difference between all other estimators is very large in the equations with unemployment. The slope coefficient b can have even different signs. Clearly, direct conclusions based on one of this estimated model may lead to accept or reject a given theory. For instance if we consider the model with heterogeneity and the estimated average coefficient, and if the sign is positive, we do accept the role of unemployment on reduction of speed

of profit rate fall.

In fact the significant difference in full data series estimator might be explained by cross-section heterogeneity and also by the fact that the same theory likely not apply for all countries for the whole period. But when we choose the time series data from 1971 up 1983, the estimators are more similar and in particular they have the same sign. The differences between them indicate the heterogeneity which subsists into individuals ; however the results are compatible with a same theory.

COUNTRY	EXOGENOUS VARIABLES		
	UNEMPLOYMENT		REAL WAGE
	1960-1983	1971-1983	1960-1983
$B = \frac{1}{N} \sum b_i$	+ 0.06	- 0.35	- 0.84
b repres. ind.	- 0.19	- 0.31	- 0.3
b^{**} difference	- 0.345	- 0.31	- 0.92
b^{***} within	- 0.21	- 0.17	- 0.69

Some of these estimators are biased and the sign and values of biases might give information about the correlations between heterogeneity and explanatory variables (see the Appendix). For instance the within estimator (b^{***}) differs from the average B of the estimated b_i if some correlation exists between the time variability of exogenous and the coefficients b_i. In our case, the difference has the same sign as cov (Vx_i, b_i). This covariance is negative and corresponds to a negative link between the variability of the exogenous Vx_i and the slopes b_i. In this case the "within" estimator under-value the mean B. The bias due to heterogeneity can give an idea on the function linking x to y. Indeed since the covariance cov (Vx, b) is negative, b appears to be (in average) a decreasing functions of Vx. Moreover since the variability and the average of x are positively correlated, we deduce that $b = \frac{\partial y}{\partial x}$ is (in average) a decreasing function of x, or equivalently that y is a concave function of x. Therefore the bias may also be seen as a consequence of a misspecification of the functional form : this form has been constrained to be a linear one, while a strictly concave function is perhaps more adapted.



This interpretation of the bias means that the acceleration of the profit rate fall with increase of unemployment (or wages). That causal relation for wages is known in economic theory as "profit squeeze". It can mean that the trend of profit rate decrease is not suficiently reduced by the work intensification, that it is amplified by increase of wages and/or the fall in volume of labour.

III - CONCLUDING REMARKS

This paper refers to some methodological problems met by economists in the use of econometrics as a vehicle to prove the rules, postulates, assumptions ... Among various obstacles to get over for obtaining a satisfactory result, we fix our attention on the way from the established facts on individuals to more general consideration. It is the search for minimum information loss in modelisation. The well known problem of aggregation is of this kind. Our topic is the treatment of heterogeneity.

We discuss the biases of estimations when the heterogeneity is not considered. We show that the simple statistical methods with linear models can be used to analyse the heterogeneity problems in some conditions. For sake of simplicity, we use only models with one explanatory variable. It appears that the analysis of biases may help to overshoot some limits of linear model and to give indications for improvement of the initial models, e.g. indications on concavity.

This methodological study is complementary to series of empirical work on the profit rate trends in some OECD countries and on the role of wages and of work precariousness increase in its fall.

The empirical results confirm once more the heterogeneity of countries and invite to be cautious in presenting general economic rules of capitalist product system. It is clear, nevertheless, that the relations remain conflictual among wages and profits. But some implicit consensus may exist for decreasing of unemployment, because the unemployment does not appear to brake the fall of profit rate.

As since the aim of the paper is to invite to be cautious in interpreting economic results as proof of assumptions, our own empirical results have to be also taken cautiously at least because of the extreme simplicity of chosen models.

APPENDIX 1

a) Profit rate in private corporate sector,

b) Unit labor cost in manufacturing (rate of change in real values),

c) Unemployment rate for total labor force.

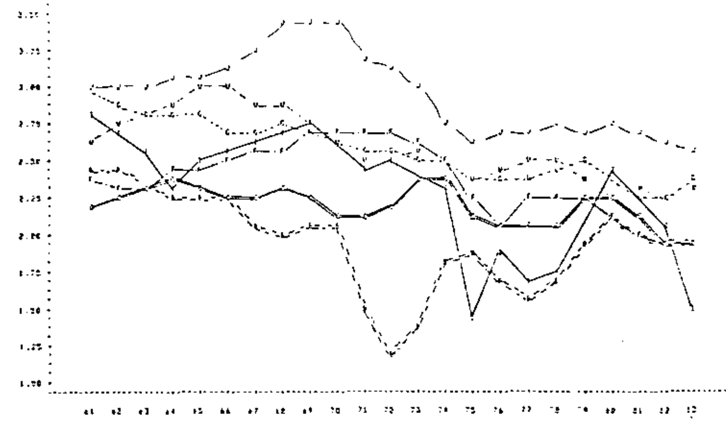
The time interval is 1960-1983.

Symbols used are :

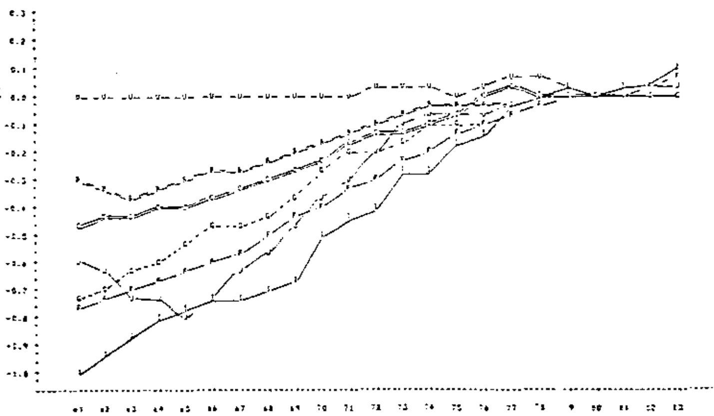
C - Canada
B - United Kingdom
F - France
G - West-Germany
I - Italy
J - Japan
U - USA.

Data source : OECD estimates.

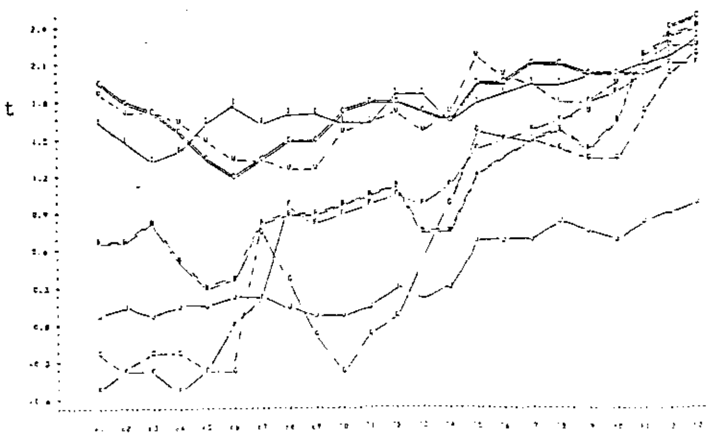
a) log Y_{it}



b) log W_{it}



c) log U_{it}



APPENDIX 2

MODELS WITH SINGLE EXPLANATORY VARIABLE : STUDY OF BIASES

We assume that the parameter of interest is an empirical average of the disaggregated coefficients $B = \frac{1}{n} \sum_{j=1}^{n} b_j$ and, as an example, we consider the two estimators computed from the model with fixed effects.

Than we have :

$$\underset{T}{\text{plim}}\ \overset{=**}{b} - B = \text{Cov}_e^j \left(\frac{VX_j\,(1-\rho_j)}{\sum_i VX_i(1-\rho_i)},\ b_j \right)$$

$$= \frac{1}{\sum_i VX_i\,(1-\rho_i)}\ \text{Cov}_e^j\ [VX_j\,(1-\rho_j),\ b_j],$$

where VX_i is the variability and ρ_j is the first order correlation of X.

$$\underset{T}{\text{plim}}\ \overset{=***}{b} - B = \frac{1}{\sum_i VX_i}\ \text{Cov}_e^j\ [VX_j,\ b_j],$$

$$\underset{T}{\text{plim}}\ \overset{=***}{b} - \underset{T}{\text{plim}}\ \overset{=**}{b} = \text{Cov}_e^j \left(\frac{VX_j}{\sum_i VX_i} - \frac{VX_j\,(1-\rho_j)}{\sum_i VX_i\,(1-\rho_i)},\ b_j \right).$$

By introducing the empirical covariances weighted by $\frac{VX_j}{\sum_i VX_i}$, and denoted $\text{Cov}_{V,e}^i$ we have :

$$\underset{T}{\text{plim}}\ \overset{=***}{b} - \underset{T}{\text{plim}}\ \overset{=**}{b} = \text{Cov}_{V,e}^i \left[1 - (1-\rho_j)\ \frac{\sum_i VX_i}{\sum'_i VX_i(1-\rho_i)},\ b_j \right]$$

$$= \frac{\sum_i VX_i}{\sum_i VX_i\,(1-\rho_i)}\ \text{Cov}_{V,e}^j\ (\rho_j,\ b_j).$$

R E F E R E N C E S

GOURIEROUX, C. and I., PEAUCELLE (1988)
"Hétérogénéité : le Cas Linéaire"
CEPREMAP, N° 8809.

KONDRATIEV, N. (1926)
"The Problem of Prevision"
Economics and Mathematical Methods, Vol XXIV, N°2, 1988, Moscow (in russian).

PEAUCELLE, I. and, P., PETIT (1988)
"Profit et Formes de Motivations Salariales"
Economie Appliquée, tome XLI, N° 1.

WEISSKOPF, T. (1985)
"Worker Security and Capitalist Prosperity : An International Comparative Analysis"
Univ. of Michigan (D.P.).