CHAPTER 2 - STATE OF THE ART

"Innovation is vital. ...However, much remains to be done"

European Commission, Green Paper on Innovation, 1995

2.1. Introduction

It is generally agreed that R&D investments contribute to the growth of productivity and output. Indeed, an impressive number of empirical studies, at all aggregation levels - i.e., business unit, firm, industry, regional, and country levels -, come to the conclusion that R&D is a major source of economic growth. However, one recurring problem is that it is virtually impossible to compare these investigations. This assertion raises two fundamental questions. First, why are the myriad of existing studies not comparable? Second, since practically all studies converge towards the same ultimate conclusion, why would this disharmony be a bone of contention?

The incompatibility ensues from the fact that there are noticeable differences in the empirical models used. They differ on the econometric specification, data sources, number of economic units, measurement techniques of R&D and economic performances, and studied period. The subjacent conundrum, in our view, is that one cannot determine whether the different results reflect actual differences in the efficiency of R&D or whether these differences are the outcome of various empirical practices. Any attempt of international comparison from these studies is, as a matter of fact, challenging, if not inconceivable.

The objective of the present chapter is twofold. First we want to provide a critical survey of the existing literature. Second, we want to set the empirical methodology that will be used throughout the following chapters. In this respect, the key issue is to assert whether various methodological and measurement issues are likely to affect the estimated impact of R&D on economic performances.

The remainder of this chapter is structured as follows. Section 2.2 presents the two main theoretical frameworks for productivity measurement - i.e., the primal and the dual approaches - and their main underlying theoretical hypotheses. The basic model that is adopted for our empirical purposes is derived from the primal approach. It consists basically in examining the link between a total factor productivity growth measure and an R&D intensity measure. Although this particular empirical framework is subject to several criticisms, it is still extensively performed.

Then, section 2.3 provides a detailed survey of the existing industry-level studies. This analysis of the literature clarifies various issues regarding the sensitiveness of the empirical results to *i)* the chosen theoretical framework, *ii)* the nature of the estimated parameters (output elasticities *vs* marginal impacts), *iii)* the structure of the data set (cross-section *vs* panel data), *iii)* the output proxy (value added *vs* total sales), *iv)* the inclusion of other variables and different time periods, and *v)* the country considered.

Section 2.4 tackles the main technical and conceptual issues related to the evaluation of the rate of return to R&D. More precisely, it discusses the implications of potential measurement errors related to the computation of a total factor productivity growth index (the likely bias concern the hypotheses of constant returns to scale and perfect competition, the measurement of labor cost shares, the double counting of R&D expenditures, and the utilization rates of production capacities) and to the computation of R&D capital stocks (the potential measurement errors concern the assumed R&D depreciation rate, lag, benchmark year, and R&D deflator). Finally, some econometric issues are put forward with respect to the endogeneity of R&D, to the inclusion of industry and time dummies, and to the treatment of highly influential observations. Section 2.5 concludes.

2.2. Economic growth, total factor productivity and the role of R&D

The traditional framework for productivity growth measurement departs generally from the assumption that producers face either a production function $Y = Y(X, t)$ - the so called primal approach - , or equivalently a cost function $C = C(p, Y, t)$ - the so called dual approach -, where Y is total output, X is a vector of k inputs (such as labor, capital, and intermediate inputs) with their corresponding prices p, and t is a time trend which reflects technological progress. Total factor productivity growth (*TFPG*) measures are generally derived from the elasticities of these functions with respect to time (the inputs being stable):

$$\partial \ln Y / \partial \ln t \equiv \tau_{Yt} \qquad (2.1)$$

$$\partial \ln C / \partial \ln t \equiv \tau_{Ct} \qquad (2.2)$$

These elasticities (τ) characterize the residuals of total output (cost) growth less the contribution of the inputs. Under the neoclassical model's assumptions of instantaneous adjustment, constant returns to scale, and perfect competition, Ohta (1975) shows that $\tau_{Yt} = -\tau_{Ct}$. The two elasticities isolate technical change or the well known Solow residual (see Solow (1958)), also called Hicks-neutral technical progress. The primal measure of *TFPG* may be interpreted as the potential increase in output from t to $t+1$ for a fixed amount of inputs. In order to get a numerically 'computable' index of the 'primal' residual we take the differential of the production function

$$d \ln Y / dt = \left[\frac{1}{Y}\right]\left[\frac{\partial Y}{\partial t} - \sum_k \frac{\partial Y}{\partial X_k}\frac{d \ln X_k}{dt}\right] \qquad (2.3)$$

and, with the hypotheses of profit maximization and perfect competition, marginal costs can be measured by the market price of output

$$p_Y \left[\frac{\partial Y}{\partial X_k}\right] = p_k \ ,$$

and it comes from equations (2.1) and (2.3) that the computation of the 'primal' residual requires the following transformation:

$$\frac{d \ln Y}{d t} - \frac{d \ln X}{d t} = \frac{\dot{Y}}{Y} - \sum_k s_k \left[\frac{\dot{X}_k}{X_k}\right] = \tau_{Yt} \ ,$$

where s_k is the share of input k in the value of total output ($s_k = p_k X_k / p_Y Y$). Under the neoclassical assumptions, the observed share of labor (and other inputs) is an exact measure of the output elasticity of labor (and other inputs).

A similar development allows to compute the 'dual' residual as follows:

$$\tau_{Ct} = \frac{d \ln c}{d\,t} - \frac{d \ln p}{d\,t} = \frac{\dot{c}}{c} - \sum_k w_k \frac{\dot{p}_k}{p_k} ,$$

where c are the unit costs and C the total costs:

$$c\,(p,t) = \frac{C(p,t,Y)}{Y},$$

and w_k is the share of input k in total costs

$$w_k = \frac{X_k\; p_k}{C} \;.$$

Under the standard hypotheses of constant returns to scale, instantaneous adjustment and perfect competition, the dual cost share are equal to the primal input share ($s_k = w_k$) since $p_Y\, Y = C$. The primal measure of *TFPG* is equivalent to the dual measure because *(i)* instantaneous adjustment guarantees full utilization of inputs and the value of marginal products of inputs covers their hired costs; *(ii)* perfect competition ensures no returns to market power; and *(iii)* constant returns to scale impose the equality between marginal cost and average cost of each input and annihilate potential profits ensuing supply characteristics (e.g., scale economies). This *TFPG* measurement (the residual) is commonly perceived as an approximation of technical change.

The studies which analyze the role of R&D investment extend Solow's growth-accounting framework to include an accumulated stock of R&D as an additional input. The standard methodology is firstly to apply the growth-accounting framework to the inputs of capital and labor (and sometimes to intermediate inputs and energy consumption) in order to calculate the *TFPG*, generally through the primal approach. This observed *TFPG* is then considered as being a function of R&D investments (the own efforts to improve the level of knowledge), exogenous technical progress, and a random influence. The following system of equation is generally referred to in order to evaluate the contribution of research to output growth:

$$Y = TFP\ F(L, K)$$
$$TFP = G(R, O)$$
$$R_t = \sum w_h\ I^R_{t-h}$$

Where Y is output, L and K are measures of labor and capital inputs, respectively, TFP is the current state of technology (total factor productivity), R is the measure of accumulated research capital (as a proxy for the knowledge stock), O stands for the other forces affecting productivity (among which disembodied technical change), I^R measures the real gross R&D investment in period t, and w_h connects the level of past research to the current state of knowledge. Technology is implicitly assumed to be Hicks neutral (or output augmenting) instead of Harrod neutral (or labor augmenting).

2.2.1. The primal approach

The economy is composed of n industries. For estimation purposes, the explicit structure of an industry i's production function is generally of the Cobb-Douglas type, which has a useful log additive form, and O is approximated by an exponential trend.

$$Y_i = \exp\left[\,\lambda_i\, t + u_i\right] L_i^{\alpha_i}\ K_i^{\beta_i}\ R_i^{\gamma_i} \qquad i = 1, ..., \mathrm{n} \tag{2.4}$$

where u is a random term, λ is the rate of disembodied technical change and α, β, and γ are the output elasticities with respect to labor, capital and R&D capital stock, respectively. The estimation of these parameters may be done by taking either the natural logarithm of equation (2.4):

$$\ln Y_i = \lambda_i\, t + \alpha_i \ln L_i + \beta_i \ln K_i + \gamma_i \ln R_i + u_i^1 \tag{2.5}$$

or the rate of growth (first difference) of equation (2.5) as follows:

$$\frac{\dot{Y}_i}{Y_i} = \lambda_i + \alpha_i \frac{\dot{L}_i}{L_i} + \beta_i \frac{\dot{K}_i}{K_i} + \gamma_i \frac{\dot{R}_i}{R_i} + u_i^2\ , \tag{2.6}$$

where $\dfrac{\dot{Y}}{Y} = \dfrac{\partial \ln Y_t}{\partial t} = \dfrac{1}{Y_t}\cdot\dfrac{\partial Y_t}{\partial t} = \ln Y_t - \ln Y_{t-1}$.

Under this specification, the assumption of constant returns to scale requires that $\beta = (1-\alpha)$. It further implicitly assumes that there is no interaction between labor and capital. We will come back later on these particular topics. It is worth noticing that in the 'level' specification (2.5) exogenous technical change is estimated through a time trend, whereas in the growth dimension it becomes a constant. A second difference between equations (2.5) and (2.6) is that the latter does not relate differences in productivity levels to differences in R&D capital stock. The advantage of the growth rate specification is that the estimates are expected not to be potentially biased by (correlated) industry-specific fixed effects.[1] On the other hand, it ignores the cross-sectional variability of the data.

Irrespective of the function is in its level or first difference form, some authors rely on a labor productivity specification by dividing output and the right-hand side variables by labor. In equations (2.5) and (2.6), we have deliberately allowed the output elasticities with respect to labor, fixed capital, and R&D capital to vary across industries. However since the available time span is most often very short, all the investigations at the industry level constrain these elasticities to be equal across industries (e.g., $\alpha_i = \alpha$; $\beta_i = \beta$; and $\gamma_i = \gamma$, for all i).

The studies which rely on the growth-accounting framework conventionally derive an index of total factor productivity growth (*TFPG*) from equation (2.6) as follows:

$$TFPG_i \equiv \frac{\dot{Y}_i}{Y_i} - \hat{\alpha}_i \frac{\dot{L}_i}{L_i} - (1-\hat{\alpha}_i)\frac{\dot{K}_i}{K_i} = \lambda_i + \gamma_i \frac{\dot{R}_i}{R_i} + u_i^3 \qquad (2.7)$$

As shown in the above discussion, the computation of this total factor productivity growth index requires the assumption of constant returns to scale with respect to labor and capital and fair payment of these traditional inputs (i.e., a perfect competition environment). In other words, the output elasticities with respect to labor (capital) are assumed to be equal to the labor (capital) costs share in total output:

$$\hat{\alpha}_i = s_{L,i} = p_{L,i}\, L_i \,/\, p_{Y,i}\, Y_i \ .$$

Because of these assumptions, the total factor productivity growth equation (2.7) is on the one hand more restricted than equation (2.6). On the other hand, it has the empirical advantage of allowing the output elasticities with respect to labor and capital to vary across industries. In this respect, equation (2.7) is less restrictive than equation (2.6). In the latter, the industries have the same production function coefficients, which is unlikely to be the case in actual fact. In the former, each industry has its own, a priori

[1] See Griliches and Mairesse (1983b, 1995) for a discussion on these differences.

imposed, labor and capital coefficients. A second advantage of equation (2.7) over equation (2.6) arises in the time series approach: there is no risk of collinearity bias between the fixed capital stock or labor and the R&D capital stock. A third advantage of the *TFPG* approach over the growth rate specification is that it does not suffer from any potential simultaneity bias between output on the one hand, and capital and labor on the other. In the latter equation a simultaneous relationship between output and labor would challenge the exogeneity assumption required for the econometric estimations.[2]

By definition, the output elasticity of R&D capital stock is given by:

$$\gamma = \frac{\partial Y}{\partial R}\frac{R}{Y} = \rho \cdot \frac{R}{Y}, \qquad (2.8)$$

that is, the marginal product of R&D (or the rate of return to R&D) multiplied by the ratio of research capital to production. Equation (2.7) forces the output elasticity of R&D capital stock γ to be equal across industries, causing the rates of return to R&D to be different. However, Griliches (1979, 1986) and Griliches and Lichtenberg (1984a) argue that the rates of return to R&D should be equal across industries in capital market equilibrium. The assumption of a common output elasticity of R&D is bothersome when estimated across industries which are characterized by long-term differences in R&D intensities. Consequently, equation (2.7) may be generalized as follows:

$$TFPG_i = \lambda_i + \rho_i \frac{\Delta R_i}{Y_i} + u_i^4 \qquad (2.9)$$

ΔR indicating the increment of the R&D capital stock, or net investments in R&D. If it is assumed that the depreciation rate of R&D investments is close to zero, the net R&D investment ΔR corresponds to the gross R&D investment I^R. There is a main difference between equations (2.7) and (2.9). The former assumes a constant output elasticity with respect to R&D (γ) and, consequently, the rates of return to R&D (ρ) are inversely correlated to the R&D intensities. Conversely, in the latter the rate of return to R&D is constant across industries and the elasticities are proportional to the R&D intensities.

[2] Griliches and Mairesse (1995) recall the critical point made by Marshak and Andrews (1944) and others in reaction to the estimation of the Cobb Douglas production function. In brief, they state that one cannot really treat the right-hand side variables as 'independent' variables and proceed with estimation by ordinary least squares as still done by most applied practitioners. In order to circumvent the potential simultaneity bias between output and labor, Griliches and Mairesse (1984) put forward a system of two semireduced form equations in order to estimate simultaneously output and labor (the left-hand side variables) on fixed capital stock and the R&D capital stock (the right-hand side variables).

Equation (2.9) also differs from equation (2.7) by establishing a direct relation between the productivity *growth* and the *level* of the net R&D intensity (instead of the *growth* of the R&D capital stock as in equation (2.7)).

The parameter ρ is often interpreted as being the rate of return to R&D. In fact, the exact meaning of this parameter depends mainly on how R&D intensity and output are measured. It may be qualified as 'net' or 'gross', 'excess', and, as far as industry-level data are used, 'social' or 'internal' rate of return to R&D. If one takes into account a non-zero depreciation rate of R&D capital it is a 'net' rate of return, whereas if a zero depreciation rate is assumed, it is a 'gross' rate of return. In most studies the estimated rate of return to R&D is called 'excess' because the data on labor and fixed capital are not corrected for double counting (i.e., the employment and the fixed capital devoted to research activities). Finally, it is a social return at the industry level because it also includes the potential intra-industry R&D externalities between the firms belonging to a particular industry. It is computed from the aggregation of output and R&D expenditures across firms. Note that the term 'social' is not fully appropriate because it does not include the potential inter industry R&D spillovers. For this reason the term 'internal' (to the industry) might be a more appropriate qualification at the industry level.

Another attractive feature of equation (2.9) is that, under the hypothesis of no depreciation, it requires only R&D intensity to be computed from R&D investments. That is, it does not require to build an R&D capital stock. This explains why a substantial number of papers have adopted equation (2.9), with a zero depreciation rate, instead of equation (2.7).

2.2.2. The dual approach

The studies which follow the dual approach are much more recent and less numerous than the ones relying on the primal approach. The former firstly appeared in the eighties with Bernstein (1988) at the micro level, Bernstein and Nadiri (1988) at the industry level, and Mohnen and Nadiri (1985) at the macro level. The latter approach has been extensively used since the sixties, the pioneers being Mansfield (1965) at the firm level, Leonard (1971) at the industry level, and Griliches (1964) at the regional level. The specifications related to the dual approach are more diversified than the ones related to the primal approach (see the survey by Mairesse and Mohnen (1995)). First, there are various *a priori* assumptions about which inputs are variable and which ones are quasi-fixed, and which inputs are considered as endogenous. Second, a choice has to be made between temporary equilibrium and a dynamic model. Third, autonomous (or disembodied) technical progress is generally excluded from the dual specification in order to avoid collinearity between the time trend and other variables.

We have seen at the beginning of this subsection that, under the standard neoclassical assumptions, the residual computed from the primal approach was similar to the one computed through the dual approach. All the studies relying on the dual approach relax most of these assumptions and allow for imperfect competition, adjustment costs, and interaction between the classical inputs. That is, they adopt much more flexible functional forms than the studies relying on the primal approach. The main drawbacks of these studies are *(i)* the structure of the data set is generally composed by a limited number of industries over a long time period; *(ii)* it requires reliable data on factor prices which are barely available at the industry level; *(iii)* the computation of cost based shares. For instance, the computation of the cost share related to fixed capital stock requires the construction of an appropriate rental price for capital which is difficult to obtain for countries other than the US (cf. Neusser (1993)).

2.3. Review of the literature

The investigations which assess econometrically the relationship between R&D and productivity can be classified according to *(i)* the underlying theoretical approach (primal or dual), *(ii)* the aggregation level of their data (firms, industries, or countries), and *(iii)* the structure of the data set (time series, cross-section, or panel data). This subsection is devoted to a review of the literature and attempt to provide some answer towards the five following questions.

1- What are the main specifications that have been used and how much do they affect the results (primal *vs* dual; *TFP* *vs* labor productivity; elasticity *vs* rate of return)?
2- What are the implications of using a cross-section data set instead of panel data?
3- Are the results robust to the output proxy (sales *vs* value added)?

4- Are the results sensitive to the inclusion of other factors and to the time period considered?
5- Does the rate of return to R&D vary across countries?

Since we intend to use data at the industry level, we shall concentrate on the works at the same level instead of summarizing the entire literature. This focus on a more limited number of studies allows to extend the scope of analysis of existing surveys in the sense that specifications, data structure and the ensuing hypotheses of each analysis are to be summarized. Nevertheless, our analysis is to be complemented by some informative results obtained through micro and macro studies. The former are much more numerous than industry-level studies, whereas the latter are best suited for international comparisons. Recent surveys of the literature at all aggregation levels are presented in Capron (1992) and Mairesse and Mohnen (1990, 1995). Mairesse and Sassenou (1991) provide a survey on the investigations dealing exclusively with firm data.

Table 2.1 summarizes the results of the empirical studies at the industry level. The first column present the authors, the second and the third ones describe the data set (e.g., country of origin, number of industries, time periods). The fourth column shows other explanatory variables that are sometimes added in the production function with their impact (when available) on output growth between parentheses. A substantial number of papers, characterized by an 'S', also estimate the rate of return to external R&D which measure the spillover effects of the R&D implemented in other industries. These externalities are the subject of Chapter 4 and are not analyzed in this chapter.

The columns denominated 'Q' and 'Str' indicate respectively the proxy used for output growth - total sales or value added (VA) - and the structure of the data set - cross-section (C.S.), time series (T.S.) or panel data (Panel) -. The column entitled 'Model' characterizes the specification underlying the estimated rates of return to R&D. It indicates the left-hand side variables - e.g., total factor productivity (*TFP*), labor productivity (*LP*), or output (*Q*) - and the main right-hand side variable - e.g., R&D intensity (*IR*) for the estimation of rates of return, or R&D capital stock (*R*) for the estimation of output elasticities -. The last two columns present the depreciation rates used to compute the R&D capital stocks and the estimated rates of return to R&D. When authors estimate directly elasticities and do not convert them into rates of return (cf. equation (2.8)), elasticities are reported in the last column. Finally, among the many estimates that are sometimes reported in particular studies, only the most representative ones are reported in Table 2.1. The studies relying on a cross-sectional data set are presented chronologically in the first part of the table, and the panel data studies are reported in the second part.

2.3.1. Five starting choices

(i) Specifications

All the studies presented in Table 2.1 rely on the primal approach with the Cobb Douglas production function. Most of them assume constant returns to scale with respect to labor and fixed capital. The estimated rates of return to R&D range from zero to 150%.[3] It seems that the choice between a *TFP* specification and a labor productivity specification does not affect significantly the results. Odagiri (1985) using a cross-section of Japanese industries and Griliches (1980) using a panel of US industries show that the estimated impact of R&D on output growth does not vary significantly from one specification to the other. The investigations favoring the dual approach at the industry level are reported in the appendix (Table A2.1). The estimated rates of return vary from 9% to 56% and are therefore compatible with the range obtained through the more numerous works based on the primal approach.

(ii) Elasticity *vs* rate of return

Another important choice has to be made regarding the estimation of elasticities or marginal impacts (e.g., rates of return). As previously mentioned, the choice depends on which of the two is likely to be stable across industries or time. Recall also that, under a zero depreciation rate hypothesis (which is arguably dubious), the rate of return specification does not require the computation of R&D capital stocks. The measurement of these stocks needs long time series on R&D investments, which are not always available. This is probably why most studies presented in Table 2.1 use the constant rate of return specification.

[3] However, if we exclude Mansfield (1980, 1988), who differentiates between basic and applied R&D, the range goes from zero to 64%.

Table 2.1.
Internal rate of return to R&D at the industry level- survey of existing estimates- Primal

Authors	Database	period	Other [1]	Q	Str.	Model [2]	δ^3	ρ (%)
1. Cross-sectional studies								
Leonard (1971)	USA, 16 ind.	1957-68	Educ. level=(0)	Sales	C.S.	Δ Q, IR	0%	9
Globerman (1972)	Canada, 11 ind.	1960-68	C4=(0), FA=(-)	Sales	C.S.	Δ Q, IR	0%	36
Griliches (1973)	USA, 85 ind.	1958-63	DHT=(-)	V.A.	C.S.	Δ TFP, IR	0%	40
Terleckyj (1974, 1980a,b)	USA, 20 ind.	1948-66	S, GOV=(-), U=(-)	V.A. Sales	C.S.	Δ TFP, IR	0%	20-37 0
Mansfield (1980)	USA, 20 ind.	1948-66	S, U=(-), B/A	V.A.	C.S.	Δ TFP, IR	0%	150
Sveikauskas (1981)	USA, 144 ind.	1959-69	S, Di	Sales	C.S.	Δ TFP, IR	0%	17-39
Scherer (1982a)	USA, 87 ind.	1964-69 1973-78	S	Sales	C.S.	Δ LP, IR	0%	13 29
Griliches and Mairesse (1983a)	USA, 15 ind. France, 15 ind.	1967-78	Two subperiods, Dt	Sales	CS	Δ TFP, IR	0%	23 33
Griliches and Lichtenberg (1984a)	USA, 27 ind.	1969-76	none	Sales	C.S.	Δ TFP, IR	0%	5
Griliches and Lichtenberg (1984b)	USA, 193 ind.	1959-68 1964-73 1969-78	S	Sales	C.S.	Δ TFP, IR	0%	29 11 31
Odagiri (1985)	Japan, 15 ind.	1973-77 1973-77	S, U=(0)	V.A.	C.S.	Δ TFP, IR Δ LP, IR	0% 0%	2 3
Mansfield (1988)	Japan, 17 ind.	1960-79	none B/A	V.A.	C.S.	Δ TFP, IR	0%	42 60
Hanel (1988)	Québec, 11 ind.	1971-82	S, 2 subperiods	V.A.	C.S.	Δ LP, IR	0%	64
Sterlacchini (1989)	U.K., 15 ind.	1945-83	S	Sales	C.S.	Δ TFP, IR	0%	12-20
Goto and Suzuki (1989)	Japan, 50 ind.	1978-83	S	V.A.	C.S.	Δ TFP, IR	0%	26-29
Griliches (1994a)	USA, 143 ind.	1973-89	Ut	Sales	C.S.	Δ TFP, IR	0%	36
Sakurai *et al.* (1997)	10 countries,13 ind	1970-80 1980-90	S, Dt, 2 subperiods	Sales	C.S.	ΔTFP, IR	0%	13-17 15-17
van Meijl (1997)	France, 30 ind.	1978-92	S, Ut	V.A. Sales	C.S.	ΔTFP, IR	0%	0-19 0-13
Vuory (1997)	Finland, 20 ind.	1981-93	S, Dt, 3 subperiods	V.A.	C.S.	ΔTFP, IR	weig.	0-18
Wolff (1997)	USA, 48 ind.	1958-87	S, Dt, 3 subperiods	Sales	C.S.	ΔTFP, IR	0%	10-13
Verspagen (1997)	9 countries, 22 ind.	1979-89	S	V.A.	C.S.	Δ TFP, IR	0%	0

1. C4 = index of industry concentration; FA = percentage of assets held by foreign corporations in each industry; DHT = dummy variable for R&D intensity >15%; S = inter industry or international R&D spillover variables; GOV = percentage of sales to government; U = union members as a percentage of workers; Dt = time dummies for subperiods; Di = industry dummies. **2.** Indicates the empirical model used to estimate the rate of return to R&D; all variables are in logarithm; Δ indicates growth rates; TFP = total factor productivity; LP = labor productivity; Q = output; R = R&D capital stock; IR = R&D intensity. **3.** δ indicates the assumed R&D depreciation rate.

Table 2.1. (continued)
Internal rate of return to R&D at the industry level- survey of existing estimates- Primal

Authors	Database	period	Other [1]	Q	Str.	Model [2]	δ^3	ρ (%)
2. Panel data studies								
Griliches (1980)	USA, 39 ind.	1959-77	Dt, Di, Kage, elasticity	Sales	Panel	TFP, R LP, R	0%	3-6 4
Nadiri (1980)	USA, 11 ind. 5 durables ind. 6 non durables ind.	1958-75	none	V.A.	Panel	Q, R	10%	20 12 86
Griliches and Lichtenberg (1984a)	USA, 27 ind.	1959-76	Dt, Di, Kage, Ut elasticity	Sales	Panel	TFP, R	0%	0
Suzuki (1985)	Japan, 5 ind.	1970-82	Di, Tr=(+)	V.A.	Panel	LP, RF*	10%	45
Adams (1990)	USA, 18 ind.	1953-80	S, Di	Sales	Panel	Δ TFP, IR	13%	.002
Yamada *et al.* (1991)	Japan, 45 ind.	1975-82	S, Py, Pk elasticity	V.A.	Panel	LP, R	0%	-17 - 44
Mohnen and Ducharme (1996)	Canada, 25 ind.	1967-83	S	Sales	Panel	Δ TFP, IR	0%	17-39
Levy and Terleckyj (1989)	USA, Telecom.	1958-85	P/G, elasticity	Sales	T.S.	Q, R	0%	36

1. Py - index of output prices; Pk - index of capital prices; S = inter industry or international R&D spillover variables; Dt = time dummies; Di = industry dummies; Tr = trend; Kage = age of fixed capital ; P/G = the author(s) differentiate(s) between government R&D and private R&D, the rate of return to the latter is reported. * Suzuki uses as independent variable the sum of R&D investments and foreign technology payments. **2.** Indicates the empirical model used to estimate the rate of return to R&D; all variables are in logarithm; Δ indicates growth rates; TFP = total factor productivity; LP = labor productivity; Q = output; R = R&D capital stock; IR = R&D intensity. **3.** δ indicates the assumed R&D depreciation rate.

In general, one may infer that the estimates of the rates of return to R&D are consistent with those derived from elasticities (the fourth column in Table 2.1 and Table A2.1 determine whether the parameters are elasticities).

Micro-level studies are more informative about this issue. Griliches and Mairesse (1983a) do not find evidence of a statistically significant relationship between *TFP* growth and the growth in the R&D capital stock across US and French firms. However, when the *TFP* growth is regressed on R&D intensity, the relationship is positive and significant. In other words, it seems that significant parameters are obtained only under the assumption of constant rate of return. Crépon and Mairesse (1993) find out that a great heterogeneity in the output elasticities of R&D (computed at the firm level) across industries may appear, and they tend to increase with the R&D intensity. This heterogeneity across industries is all the more confirmed by the papers relying on the dual approach at the industry level. For instance, Odagiri and Kinukawa (1997) estimate elasticities that range from negative to positive values across industries. This high variability across industries leads Mairesse and Mohnen (1995) to suggest that the assumption of constant rates of return to R&D is perhaps not unreasonable. In any case, the authors posit that the two alternative approaches are, on average, compatible. Further, Hall and Mairesse (1994) present some results suggesting that it is particularly the case with time series.

(iii) Cross section *vs* time series

An additional source of heterogeneity in the estimation of rates of return to R&D relates to the cross-sectional or temporal structure of the data set. The first part of Table 2.1 presents the studies favoring the cross-section and the second part those exploiting the time series. Again, most authors rely on the former dimension, probably to avoid the construction of R&D capital stocks and also because of the lack of data concerning long term R&D investments series. Does the use of time series affect the estimated impact of R&D on output growth?

An important point is that the specifications used by panel data studies are much more heterogeneous than the ones used in cross-section studies. The former take more often the variables in level than in growth rates, compute sometimes R&D capital stocks with positive depreciation rates, and are more inclined to estimate elasticities instead of rates of return. All cross-sectional estimates rely on the zero depreciation rate hypothesis which allows to use gross - instead of net - R&D investments and therefore yields gross excess internal rates of return to R&D.

Our concern here is to see whether cross-section investigations yield higher or lower rates of return to R&D than panel data studies. The estimates in the cross-sectional studies range from 2% to 64% and in the panel data studies from 0% to 45%. Further, if Suzuki (1985)'s estimate is withdrawn from the latter because of the foreign technology payments that are added to R&D investments (which means that the parameter partly includes the returns from foreign R&D), the range goes from 0% to 39%. It is also worth noticing that out of the six remaining panel data studies, three are associated to a return to R&D lower than 5%. We may therefore infer that estimates from time series data are generally lower than estimates from cross-section data.

In his survey of micro-level studies, Capron (1992) observes that estimates in the time dimension provide more controversial results than cross-section studies. With panel data, a large number of coefficients are non significantly different from zero. The author evaluates, across 46 parameters taken in micro-level studies, an average output elasticity of R&D estimated through cross-section superior to 5% than those obtained from time series analyses. Recent micro-level studies by Hall and Mairesse (1994) with French firms and Harhoff (1994) with German firms confirm that cross-sectional estimates are generally higher than temporal estimates, which are sometimes non significantly different from zero. Finally, our own estimates, similar to the ones of Capron (1992), have been run with the information provided in Table 2.1. The results, which are presented in Table A2.3 of the appendix, show that the empirical studies based on panel data obtain an average rate of return of 13%, whereas cross section studies yield an average rate of return of 35%.

Time series estimates might be relatively lower for several potential reasons: *(i)* wrong lags may be associated to the R&D capital stock; *(ii)* high collinearity may appear between the R&D variable and the time trend reflecting exogenous technical change; *(iii)* there may be measurement errors in variables or important variables missing, such as the rate of utilization of fixed capital; *(iv)* there may be some simultaneity bias between output growth and R&D; and *(v)* the temporal variability of R&D investments and of R&D capital stock is relatively low as compared to the variability across firms or industries (with R&D capital stocks, the temporal variability is also reduced by construction). However it is worth noticing that an important advantage of temporal studies over cross sectional studies is that the former are not affected by the potential bias due to omitted variables such as the characteristics of firms or industries, provided industry and/or firm fixed effects are taken into account. In this respect, Nelson (1988) criticizes the interpretation of the empirical results contained in cross-section studies. He argues that R&D intensity is not exogenous; in fact, it is determined by technological opportunities and appropriability conditions specific to each industries. The inclusion of industry dummies allow to control for these factors in panel data investigations.

(iv) Total sales *vs* value added

About half of the investigations in Table 2.1 use total sales to proxy output and the other half value added. In general the second proxy is used when data on intermediate inputs is not available. At the firm level, Griliches and Mairesse (1984) demonstrate that the omission of intermediate inputs in the sales regression yields an upward bias in the R&D elasticity because intermediate inputs are correlated with R&D. It seems that even if one controls for materials, the choice may have a substantial impact on the estimated rates of return.

Capron (1992) estimates that across 17 estimates of 9 cross-section studies at the industry level, the average rate of return to industrial R&D is equal to 38% when total sales are used and to 28% otherwise. Capron presents similar figures for micro-level works. This tends to show that the studies based on value added as a proxy for output yield somewhat smaller estimated rates of return to R&D.

However, the existing estimates presented in Table 2.1 do not really confirm Capron's findings. This is supported by three observations. First, with value added, the rates of return to R&D vary from 0% to 86%, whereas with total sales, they vary from 0% to 39%.[4] Second, The results that are presented in Table A2.3 clearly indicate that the average return to R&D is about 35%, and does not change significantly with the output proxy used.

The third observation is that a few authors evaluate the consequences of using total sales instead of value added, within a given empirical frame. Terleckyj (1980) shows that with the two-factor (i.e., with value added) *TFP* growth, the estimated return to R&D is significant and equal to 27%, whereas with the three-factor *TFP* growth, the return falls to 20%, but the difference is insignificantly different from zero. The results provided by Wolff and Nadiri (1993) suggest that the *TFP* growth estimated with sales (i.e., taking into account intermediate products) is about one half of the *TFP* growth computed with value added. van Meijl (1997), using a cross-section of French industries, obtains a rate of return to R&D equal to 13% with total sales and to 19% with value added. That is, in the light of these three papers and of the results presented in Table 2.1, we are more inclined to support the idea that the use of value added in the computation of productivity yields identical or somewhat higher estimated rates of return to R&D than using total sales.

(v) Other variables and different time periods

The inclusion of other variables into the empirical models generally contributes to decrease the order of magnitude of the estimated impact of R&D on output growth. For instance, inter industry R&D spillovers variables, the utilization rate of fixed capital, the share of union members in labor forces, time dummies (or trend) and industry dummies in panel data studies, have most often a downward effect on the rate of return (or elasticity) to R&D when they are introduced into the empirical specification.

[4] Including Mansfield (1980), this range goes from zero to 150%.

Concerning the interindustry spillover effects, Table A2.3 in the appendix suggests that the inclusion of external R&D into the empirical model reduces the estimated return to R&D from 35% to 25%.

There does not seem to be a noticeable change in the estimates with respect to the period considered (see Table A2.3). However, Griliches and Lichtenberg (1984a, 1984b) and Scherer (1982a), among others, all to amply demonstrate that the rate of return to R&D vary with the time periods considered. The general commitment is that the R&D productivity relationship has changed over time, especially during the seventies where it has sharply fallen. Since then, there is some evidence supporting the idea that the relationship has recovered. Lichtenberg and Siegel (1991), Scherer (1993), and Hall (1993) find that the R&D coefficients come close to disappearing in the late 70's and the early 80's but recover prominently in the mid- to the late eighties.

Notwithstanding the existing empirical evidence about the flagrant changing relationship between R&D and productivity growth over the years, Griliches (1994) suggests another story. He argues, with some strong empirical evidence for US industries, that *TFP* growth might be wrongly measured due to unreliable price indexes in high-tech industries. When the price indexes of three industries out of 143 are adjusted, Griliches obtains no change at all in the estimated return to R&D.

2.3.2. International comparability

The major implication ensuing from the five preceding issues is that the estimated impacts of R&D on productivity growth are barely comparable across studies, even for a country of reference. The different specifications, output proxies, structures of database and time periods prevent any serious international comparison. The only way to get a valuable insight into potential international differences in the rates of return to R&D is to refer to the studies which evaluate, within an homogenous empirical framework, the impact of R&D activities for different countries. Griliches and Mairesse (1983a) is the only study at the sectorial level that allows for an international comparison. Their results, estimated across 15 industries in France and in the US, suggest that the rate of return to R&D is substantially higher, though non significantly different, in France (33%) than in the US (23%). There is obviously a need for more homogenous international comparison at the industry level.

The papers that allow for international comparisons are much more numerous at the macro level and to a somewhat lesser extent at the firm level. They are summarized in Table 2.2. Griliches and Mairesse also (1983a) present French-US estimates at the micro level. Similarly to their estimates at the industry level, French firms seem to have a higher rate of return to R&D (31%) than their American counterparts (19%). However, these results have to be taken cautiously because their French sample is much less numerous and more research intensive than the US sample. The analysis by Hall and Mairesse (1995) is perhaps more robust with respect to a potential 'selection' bias; they rely on panel data over the period 1981-1989 comprising more than one thousand of firms in both France and the US. Their results based on either within estimates (with firm dummies) or first differences estimates (growth rates) show that the output elasticity of R&D is positive and significant for the US firms (9% to 17%), but non significant for the French firms.

Griliches and Mairesse (1990) compare the rate of return to R&D for US and Japanese firms. Without industry dummies and with constant returns to scale imposed their results show a slightly higher rate of return to R&D among the Japanese firms (56%) than among the US firms (41%). However, when industry dummies and/or free returns to scale are included, the estimated rate of return to R&D becomes insignificant in Japan whereas it is about 25% for the US firms. Capron and Cincera (1997) obtain similar output elasticities with respect to R&D for US and Japanese firms (28%), but a relatively weaker impact for European firms (22%).

Macro level studies have the advantage of being 'free' of any selection bias. They are presented in the second part of Table 2.2. Recall that the estimated returns to R&D in these studies are 'social' in the sense that they incorporate both intra- and inter industry spillovers. In this respect, the macro-level estimates should be higher than the micro- or industry-level estimates. All these works converge towards the recognition that Japan has the highest rate of return to R&D and is closely followed by Germany. The US stands in an intermediate position and France and the UK are both characterized by the lowest impact of R&D.[5]

[5] Nadiri and Kim (1996) is the only study, out of the seven macro-level investigations, that presents a slightly different picture. The rates of return to R&D are very similar across countries; 16% in Italy; 15% in Japan, Germany, France, and Canada; and 14% in the UK and the US.

Table 2.2.
International comparison studies- micro and macro level[1]

Authors	Database	period	Other [1]	Q	δ	ρ (%)
Firm level						
Griliches and Mairesse (1983a)	185 French firms	1973-78	Capital, Dc, C.S.	Primal, ΔLP, total sales,	0%	31
	343 US firms		no materials	no interm. inputs, IR		19
Griliches and Mairesse (1990)	1032 Japanese firms	1973-80	Capital, C.S.	Primal, ΔLP, total sales,	0%	56
	968 US firms			no interm. inputs, IR		41
Hall and Mairesse (1995)	1232 French firms	1981-89	Labor, capital, Df,	Primal, Q, total sales,	15%	0
	1073 US firms		Panel, no materials	no interm. inputs, R, elasticities		17
	French firms		First difference			0
	US firms					9
Capron and Cincera (1997)	378 US firms	1987-94	Capital, Labor, S,	Primal, Q, total sales ,	15%	12-28
	133 Japanese firms		panel	R, elasticities		7-28
	101 European firms					0-22
Country level						
Soete and Patel (1985)	USA	1956-82	Tr, Dc, no capital	Primal, LP, R, elasticity	15%	10
	Japan		Time series			33
	France					6
	Germany					27
	UK					6
Mohnen and Nadiri (1985)	USA	1965-78	K and R fixed,	Dual, Cost function	10%	16
	Japan		Time series	elasticities,		23
	Germany					20
	France					13
Mohnen, Nadiri, and Prucha (1986)	USA	1965-78	K and R fixed	Dual, Cost function	10%	11
	Japan		Time series			15
	Germany					13
Guellec (1991)	USA	1960-87	Time series	Primal, total factor	10%	15
	Japan			productivity		26
	Germany					28
	France					16
	UK					9
Mohnen (1992)	USA	1964-85	S, K and R fixed	Dual, Cost function	10%	15
	Japan		Time series			19
	France					1
	Germany					14
	UK					6
Bernstein and Mohnen (1994)	USA	1962-88	S, K and R fixed	Dual, Cost function	10%	13
	Japan		Time series			14
Nadiri and Kim (1996)	USA	1964-91	K and R fixed, S,	Dual, Cost function,	10%	14
	Japan		time series	rates of return		15
	Germany					15
	France					15
	UK					14
	Italy					16
	Canada					15

1. cf. Table 2.1 for the definition of the abbreviations.

2.3.3. In a nutshell

The Most important findings which sort out of this survey are:

(i) The estimated internal gross excess rates of return at the industry level range from 0% to 60%, most of them being significantly different from zero, and their average is 35%.

(ii) The various specifications (e.g., primal or dual approach, labor productivity or total factor productivity) yield comparable estimates of the rates of return to R&D.

(iii) The estimates based on panel data (e.g., with a temporal dimension) are lower than those based on cross-section data. However, the former have the advantage of being much less subject to potential bias ensuing from important omitted variables such as technological opportunity and appropriability conditions.

(iv) The rates of return to R&D derived from the estimation of output elasticities of R&D are likely to be lower than the rates of return directly estimated. However, both types of estimates are fully compatible and this compatibility improves within panel data frameworks.

(v) Using value added instead of total sales as a proxy for output yields slightly higher estimates of the rate of return to R&D. In any case, if data on intermediate inputs is not available, value added has to be used.

(vi) The estimated rate of return to R&D decreases when other variables are included in the model (such as time dummies, time trend, the utilization rate of production capacities, allowing for returns to scale and including external R&D) and is sensitive to the period considered.

(vii) International comparisons are hindered by these various sources of heterogeneity across studies. However, the studies which adopt a similar empirical framework for different countries, both at the macro and the micro level, converge towards the observation that Japan is associated to the highest rate of return to R&D and is closely followed by Germany. The US stands in an intermediate position and the UK and France are both characterized by a lower return.

2.4. Technical and conceptual issues in measuring the rate of return to R&D

Once a choice has been made concerning the structure of the data set, the output proxy, and the empirical specification, there are still myriad of measurement issues concerning the operational construction of the various variables. These issues are discussed in the present subsection.

The main empirical model that is to be estimated in the next chapter is the one characterized by equation (2.9). Through this model, the objective is to estimate directly the rate of return to R&D by regressing a total factor productivity growth index (i.e., the 'residual') on the R&D intensity across 22 industries from 1980 to 1990 (computed from net R&D investments). In what follows we first discuss five issues related to the measurement of total factor productivity growth: the hypothesis of constant returns to scale, the hypothesis of perfect competition, the measurement of input shares, the double counting and expensing of capital and labor devoted to research activities, and the rate of utilization of production capacities. Then, four issues are tackled concerning the measurement of the main indicator of knowledge: the computation of R&D capital stocks (depreciation rates, lags, and benchmark values), and the prices deflator of R&D investment series. Finally, some econometric issues are tackled. They concern the inclusion of time and industry dummies, the potential presence of simultaneity biases between output and R&D, and the potential presence of influential observations.

2.4.1. The measurement of total factor productivity growth

(i) TFP - Returns to scale

The hypothesis of constant returns to scale is one of the most controversial issues associated to the estimation of the contribution of R&D to output growth. This controversy,[6] which originally appeared in the early sixties, is due to the fact that it is empirically difficult to determine whether an improvement of productivity is the result of increasing returns to scale or whether it is the result of technical change. Further, this drawback is worsened by the recognition that one of the forces which allow for economies of scale is technological progress itself. Levy and Terleckyj (1989)'s algebraic representation gives a good picture of this issue. Let us assume that we have data on output (Y), labor (L), and capital (K). The production function, technical change, and economies of scale have to be estimated from Y, L, and K. If technical change is neutral (output augmenting) from time 1 to time 2, then at time 1 and at time 2 we would observe:

Time 1: $Y = f(L, K)$ Time 2: $A\,Y = A\,f(L, K), \quad A > 1.$

The same inputs in the second period would produce more output than in the first period. Now, if we consider a situation with returns to scale, and if the production function is homogenous of degree s then, as in the preceding example and if inputs are increased in proportion c at time 2:

Time 1: $Y = f(L, K)$ Time 2: $c^s\,Y = f(cL, cK).$

There are economies of scale if $s > 1$, diseconomies of scale if $s < 1$, and constant returns to scale if $s = 1$. Assume that output rises from Y to AY from time 1 to time 2, and L and K increase to cL and cK. This observation could be interpreted as pure economies of scale of degree s, where s would be found by solving the following equation: $c^s = A$. However, keeping these notations, the factor s could also be expressed as the product of constant returns to scale, c^1, and technical progress amounting to c^{s-1}:

$$c^1\,c^{s-1}\,Y = AY\,.$$

And alternatively, it could also be interpreted as the product of any scale effect x, c^x, and technical progress of amount c^{s-x}:

$$c^x\,c^{s-x}\,Y = AY\,.$$

[6] It is also called the Solow-Stigler controversy; see Solow (1961) and Stigler (1961).

This simplified example illustrates fairly well that with an homothetic production function, productivity growth may be due either to increasing returns to scale, or technical progress, or both. Further, if increasing returns to scale are partly the result of technical progress, it may be empirically difficult to clearly distinguish technical progress, economies of scale due to technical progress, and economies of scale due to an increase in inputs. In any case, it seems that not allowing for 'free' returns to scale might yield a biased evaluation of technical change and of the extent to which R&D contributes to it. Griliches and Ringstad (1971) argue that, even with free returns to scale, a Cobb-Douglas type production function might still be used to evaluate the impact of R&D on output growth.

Levy and Terleckyj (1989) find that one cannot reject the hypothesis of constant returns to scale for the US telecommunication industry.[7] This is an interesting result since we would *a priori* expect increasing returns to scale in that particular industry. However, the estimated rates of return to R&D turn out to be very sensitive to whether Levy and Terleckyj impose the constant returns to scale assumption or not.

The analyses which estimate the impact of R&D on a total factor productivity index do not, in general, test for increasing or decreasing returns to scale. Those which rely on a Cobb-Douglas production function and estimate the output elasticities with respect to labor and fixed capital make sometimes this test. Among all the investigations at the industry level presented in Table 2.1 only Nadiri (1980), Griliches (1980), and van Meijl (1997) test whether relaxing the hypothesis of constant returns to scale has an important effect on the estimated rates of return to direct R&D.[8] Nadiri (1980), with eleven US manufacturing industries, cannot reject the hypothesis of constant returns to scale with respect to labor and capital, and the estimated output elasticity of R&D is stable with respect to the allowance for free returns to scale. This stability of the estimates is not confirmed by Griliches (1980), who notices that the estimated output elasticity of R&D falls when free returns to scale are allowed by including total man-hours among the right-hand side variables. van Meijl (1997) obtains the opposite results. When the author suppress the constant returns to scale constraint - which are in fact decreasing returns to scale -, the estimated rate of return to R&D is higher than with the constant returns to scale.

---

[7] This industry is particularly well suited for that kind of analysis, partly because the output of this industry is rather homogenous and partly because a considerable amount of data on outputs and inputs is available and of good quality.

[8] Yamada *et al.* (1991) also allow for free returns to scale but do not compare their results with a constant returns to scale assumption.

At the micro level, Griliches and Mairesse (1990) estimate sharp diminishing returns to scale across Japanese firms and, to a lesser extent, across US firms. Furthermore, in both countries, the allowance for free returns to scale implies much lower estimated rates of return to R&D, which are even non significantly different from zero for Japanese firms. Hall and Mairesse (1995), using a panel data of French firms, also display decreasing returns to scale (though not always significantly different from zero) and a noticeable sensitivity of the estimated impact of R&D on output growth with respect to the adoption of the constant returns to scale hypothesis or not.

The general procedure of estimating returns to scale is to assume that they are stable across industries or firms. However, the evaluation for particular industries, over a sufficiently large time span, should most probably give a better approximation of the returns to scale prevailing in each industry. In fact, such empirical evidence on economies of scale is rather scarce. Levy and Terleckyj (1989) make the test for the US telecommunication industry, which is characterized by constant returns to scale. Beason and Weinstein (1996) evaluate a scale parameter for 10 Japanese industries over the period 1955-1990. They can not reject the hypothesis of constant returns to scale for any sectors, even for the few ones which exhibit apparently diminishing returns to scale. Similarly, Nakamura (1993) estimates constant scale parameters using a cost function approach for four medium to high-tech Japanese industries.

What can we conclude about this 'returns to scale' issue? The algebraic development demonstrates that it might be difficult to distinguish empirically the effect of increasing returns to scale from the effects of technical progress. And this distinction might become even more puzzling if increasing returns to scale are considered to be partly due to technical change.[9] Empirical works at the firm or industry levels generally reach the conclusion that there are decreasing returns to scale and that relaxing the hypothesis of constant returns to scale reduces the estimated impact of R&D (or the output elasticity with respect to R&D). They confirm in some respects that there is a close link between the two forces at work. Yet, some single industry studies do not reject the constant returns to scale hypothesis, and therefore validate the majority of studies at the firm and industry level that explicitly rely on this apparently not too restrictive hypothesis.

[9] Intuitively, increasing returns to scale may be due also to technobgical characteristics of the production process that would make firms or industries more productive with larger output levels, independently of technical change.

In any case, we would like to test whether increasing or decreasing returns to scale are at work and whether they affect our estimated rates of return to R&D. In the presence of non constant returns to scale (the scale parameter being $(1+\varepsilon^{sc})$, with ε^{sc} different from zero), the 'true' total factor productivity index (*TFP**) would be different from the 'observed' *TFP*:

$$TFP^* = \frac{Y}{L^{\alpha+\varepsilon^{sc}}\ K^{(1-\alpha)+\varepsilon^{sc}}} = \frac{Y}{L^{\alpha}\ K^{1-\alpha}} \cdot \left[L\ K\right]^{-\varepsilon^{sc}} = TFP \cdot \left[L\ K\right]^{-\varepsilon^{sc}}$$

Therefore, in terms of growth rates, and from equation (2.7):

$$TFPG^* = \frac{\dot{Y}}{Y} - \hat{\alpha}\,\frac{\dot{L}}{L} - (1-\hat{\alpha})\,\frac{\dot{K}}{K} - \varepsilon^{sc}\left[\frac{\dot{L}}{L} + \frac{\dot{K}}{K}\right] = TFPG - \varepsilon^{sc}\left[\frac{\dot{L}}{L} + \frac{\dot{K}}{K}\right]$$

In other words, depending on the scale parameters, the observed *TFPG* might be biased downward or upward. Since we have no information about the likely sign and amplitude of ε, it has to be estimated empirically. Equation (2.9) has to include the sum of the rates of growth of labor and capital among the right-hand side variables in order to test for the potential presence of increasing or decreasing returns to scale:

$$TFPG_i = \lambda_i + \rho\,\frac{\Delta R_i}{Y_i} + \varepsilon^{sc}\left[\frac{\dot{L}_i}{L_i} + \frac{\dot{K}_i}{K_i}\right] + u_i \qquad (2.10)$$ [5]

(ii) TFP - Perfect competition

Another source of mismeasurement of *TFPG* may be at work if market power exists. This issue is closely linked to the previous one because the hypothesis of perfect competition is inconsistent with increasing returns to scale. With market power, the output price (p_Y), or marginal revenue, would be above the marginal cost (*MC*). Hall (1988) and Morrison (1992) show that the markup ratio between p_Y and *MC* is determined by the demand elasticity (taking the annotations of equations (2.1) to (2.3)):

$$\frac{p_Y}{MC} = \frac{p_Y}{p_Y + Y\cdot(\partial p_Y / \partial Y)} = \frac{p_Y}{p_Y\cdot(1+\partial \ln p_Y / \partial \ln Y)} = \frac{1}{1+\tau_{PY}}\ ,$$

where τ_{PY} is the inverse demand elasticity of prices facing the firms. In other words, in the case of imperfect competition, total cost (*C*) is lower than total revenue (p_Y . *Y*) :

$$\left(1 + \tau_{PY}\right) = \frac{C}{p_Y\ Y}\ ,$$

and the input share adjustment in the value of total output (s_k) becomes a function of the input share in total costs (w_k) and of the demand elasticity of prices:

$$s_k = w_k\ \left(1 + \tau_{PY}\right).$$

The markup correction of the primal measure of total factor productivity growth (see equation 2.3) is

$$\tau_{Yt}^{pc} = \frac{\dot{Y}}{Y} - \sum_k w_k\ (1 + \tau_{PY}) \left[\frac{\dot{X}_k}{X_k}\right] = \tau_{Yt} - \tau_{PY} \cdot \sum_k w_k \left[\frac{\dot{X}_k}{X_k}\right]$$

Therefore, the 'true' total factor productivity *TFPG**, derived from the primal approach, may be expressed as the observed *TFPG* measure corrected for the 'markup adaptation':

$$TFPG^* = TFPG - \varepsilon^{pc} \cdot \left[\hat{\alpha}\ \frac{\dot{L}}{L} + (1 - \hat{\alpha})\ \frac{\dot{K}}{K}\right];$$

where ε^{pc} ($= \tau_{PY}$) reflects the markup adjustment factor, which may be estimated by transforming equation (2.9) as follows:

$$TFPG_i = \lambda_i + \rho\, \frac{\Delta\ R_i}{Y_i} + \varepsilon^{pc} \cdot \left[\hat{\alpha}_i\ \frac{\dot{L}_i}{L_i} + (1 - \hat{\alpha}_i)\ \frac{\dot{K}_i}{K_i}\right] + u_i^{6}\ . \qquad (2.11)$$

Recall that $\hat{\alpha}_i$ in equation (2.11) is the evaluation of *s(L)*, the share of labor compensation in value added.

(iii) TFP - Measurement of input shares

A third potential problem is that the evaluation of the input shares in value added might be slightly biased, even in the case of constant returns to scale and perfect competition. In the case of such statistical imperfections, the 'true' *TFP** would be slightly different from the observed *TFP*:

$$TFP^* = \frac{Y}{L^{\hat{\alpha}+\varepsilon^{is}}\ K^{(1-\hat{\alpha}-\varepsilon^{is})}} = \frac{Y}{L^{\hat{\alpha}}\ K^{1-\hat{\alpha}}} \cdot \left[\frac{L}{K}\right]^{-\varepsilon^{is}} = TFP \cdot \left[\frac{L}{K}\right]^{-\varepsilon^{is}}$$

Therefore, in terms of growth rates, and from equation (2.7):

$$TFPG^* = \frac{\dot{Y}}{Y} - \hat{\alpha}\,\frac{\dot{L}}{L} - (1-\hat{\alpha})\,\frac{\dot{K}}{K} - \varepsilon^{is}\left[\frac{\dot{L}}{L} - \frac{\dot{K}}{K}\right] = TFPG - \varepsilon^{is}\left[\frac{\dot{L}}{L} - \frac{\dot{K}}{K}\right]$$

where ε^{is} can be estimated and reflects the degree to which the input shares evaluation are biased, holding for constant returns to scale. Testing for the presence of such a bias, equation (2.9) has to be transformed as follows:

$$TFPG_i = \lambda_i + \rho\,\frac{\Delta R_i}{Y_i} + \varepsilon^{is}\left[\frac{\dot{L}_i}{L_i} - \frac{\dot{K}_i}{K_i}\right] + u_i \qquad (2.12)$$ [7]

(iv) TFP - Double counting and expensing of R&D expenditures

The double counting of R&D personnel and fixed capital devoted to research activities is a problem that may stay with us throughout the next chapters. Ideally, we would like to distinguish between capital and labor used to produce current 'output', and capital and labor used in research activities (e.g., the production of future knowledge and the maintenance of the current stock). In fact it is usually impossible to observe these different input components and investigators are therefore forced to use aggregate values for capital and labor in their empirical evaluation.

Relying on variables which are not corrected for double counting of R&D means that the estimated rate of return to R&D is an '*excess*' return to R&D. If the marginal productivity of capital and R&D capital were similar, we would expect an estimation of ρ equal to zero. By the same token, if the marginal productivity of R&D capital is 5 % higher than the marginal productivity of production capital, we would expect the estimated ρ to be close to 5%. In other words the excess rate of return to R&D reflect the extent to which R&D inputs are characterized by an *'above and beyond normal remuneration'* (cf. Griliches 1979) to traditional inputs. The conventional view is to interpret the computed excess rate of return to R&D as either a risk premium or a supra-normal rate of profit on R&D investments.

Schankerman (1981) provides the first analysis of the potential mismeasurement caused by the double-counting of R&D. The author also examines the bias that arises by treating R&D as an intermediate expense rather than as a capital asset. In this case, if a value added measure is used for output, the observed value added suffers from the fact that R&D investments are treated as intermediate inputs and are expensed out (because of special fiscal rules in favor of R&D spending). According to Cunéo and Mairesse (1984), this is particularly true for materials used in R&D activities in France and for all R&D expenditures in the US.

In the following, we discuss the effect of crediting double counting and expensing on the growth accounting context. The true total factor productivity growth (*TFPG**) should be computed from labor and capital which are not included in R&D (L^Y and K^Y); from equation (2.7) we have:

$$TFPG^* \equiv \frac{\dot{Y}^Y}{Y^Y} - \hat{\alpha}^Y \frac{\dot{L}^Y}{L^Y} - \hat{\beta}^Y \frac{\dot{K}^Y}{K^Y} \tag{2.13}$$

and the traditional (or observed) measures of fixed capital and labor (K and L) include research inputs (K^R and L^R) as well as traditional inputs (K^Y and L^Y):

$$\begin{aligned} L &= L^Y\,(1 + L^R / L^Y) \equiv L^Y\,(1 + h_L) \\ K &= K^Y\,(1 + K^R / K^Y) \equiv K^Y\,(1 + h_K) \end{aligned} \tag{2.14}$$

where h_L is the ratio of R&D labor to that of traditional labor, and h_K is similarly defined for fixed capital. In growth rate form, (2.14) becomes

$$\begin{aligned} \frac{\dot{L}}{L} &= \left(\frac{1}{1+h_L}\right)\frac{\dot{L}^Y}{L^Y} + \left(\frac{h_L}{1+h_L}\right)\frac{\dot{L}^R}{L^R}\ , \\ \frac{\dot{K}}{K} &= \left(\frac{1}{1+h_K}\right)\frac{K^Y}{K^Y} + \left(\frac{h_K}{1+h_K}\right)\frac{\dot{K}^R}{K^R}\ . \end{aligned} \tag{2.15}$$

The measured factor shares are also biased by double counting

$$\hat{\alpha} = \hat{\alpha}^Y\,(1 + h_L) \tag{2.16}$$

$$\hat{\beta} = \hat{\beta}^Y \, (1 + h_K)$$

where constant returns to scale imply that $\hat{\beta} = 1 - \hat{\alpha}$. Using (2.13) and (2.15) and assuming that value added is correctly measured, the observed residual defined in (2.7) turns to be

$$TFPG = TFPG^* - \left[\hat{\alpha}^Y h_L \frac{\dot{L}^R}{L^R} + \hat{\beta}^Y h_K \frac{\dot{K}^R}{K^R} \right]. \qquad (2.17)$$

The bracketed term represents the extent to which the observed residual is biased downward when not corrected for the double counting of labor and capital used for research activities. This excess return bias exists provided R&D inputs do not remain constant over time. If the growth rate of research input is very small, one might expect the real rate of return to R&D to be close to the estimated 'excess' rate of return to R&D.

Schankerman also suspects that value added (Y) might be improperly measured. He argues that current R&D is typically expensed - e.g., subtracted from gross output as an intermediate input, inducing that the measured value added is too small by the amount of R&D investments: $Y = Y^Y - I^R$. In growth rate terms, this equality becomes:

$$\frac{\dot{Y}}{Y} = (1 + \theta) \frac{\dot{Y}^Y}{Y^Y} - \theta \frac{\dot{I}^R}{I^R} \, , \qquad (2.18)$$

where $\theta = I^R/Y$ is the observed R&D intensity. From equations (2.17) and (2.18), the observed residual is

$$TFPG = TFPG^* - \left[\hat{\alpha}^Y h_L \frac{\dot{L}^R}{L^R} + \hat{\beta}^Y h_K \frac{\dot{K}^R}{K^R} \right] + \theta \left[\frac{\dot{Y}^Y}{Y^Y} - \frac{\dot{I}^R}{I^R} \right]. \qquad (2.19)$$

The last term characterizes the 'expensing bias', which can be positive or negative and may therefore either countervail or intensify the bias due to double counting. Gross value added is used in equation (2.19), which is defined to include depreciation on traditional fixed capital and R&D capital stock. In the case of net value added - e.g., net of depreciation (δ) on R&D capital (R) - the measured value added is too high by the depreciated R&D: $Y = Y^Y - I^R + \delta R$. In this case, the expensing bias becomes

$$TFPG = TFPG* - \left[\hat{\alpha}^Y h_L \frac{\dot{L}^R}{L^R} + \hat{\beta}^Y h_K \frac{\dot{K}^R}{K^R}\right] + \theta\left[\frac{\dot{Y}^Y}{Y^Y} - \frac{\dot{I}^R}{I^R}\right] - \delta\phi\left[\frac{\dot{Y}^Y}{Y^Y} - \frac{\dot{R}}{R}\right], \quad (2.20)$$

where $\phi = R/Y$ is the R&D capital stock intensity. The main implication of equations (2.19) and (2.20) is that the 'excess return' interpretation is conceptually incorrect as far as one allows for double expensing. Depending on the growth rates of R&D investment, R&D capital stock, and corrected value added, the sign and the amplitude of the total bias is not *a priori* predictable. Schankerman (1981)'s estimates implemented at the US micro level do, however, validate the excess return interpretation. He finds that the expensing bias is downward and reinforces the double counting bias. The downward bias seems to be sensitive to the technological intensity of the firm's industry and the degree of government interventionism. Across the firms in the Chemicals, Motor vehicles, and Petroleum refineries industries, the bias is equal to about 50%, whereas it raises to about 800% in the industries of Electronic equipment and Aircraft.

Cunéo and Mairesse (1984) argue that the double counting bias is more likely to be present in the cross-sectional dimension. In the temporal dimension - e.g., within or first difference estimates - it should vanish if the ratios of research labor to traditional labor (L^R/L^Y) and of research fixed capital to traditional fixed capital stock (K^R/K^Y) are stable over time. Their econometric results, based on a database of French firms, confirm it: the total estimates are much more biased than the within estimates.

Schankerman (1981) does not use a growth rate specification, such as in equations (2.21), (2.23), and (2.24), but rather a cross-sectional framework with the variables expressed in 'levels'. One may therefore wonder if the excess return interpretation is still valid under a specification in growth rates. Mairesse and Hall (1995), using a panel of French firms, find out that correcting value added and labor for double counting raises the R&D capital elasticity by about .07 in the cross section dimension, .03 in the within dimension, but not at all when using first differences (or growth rates). Hall and Mairesse (1995), also relying on a data set of French firms, obtain slightly different results. Since both total and within estimates are significantly biased by the double counting of labor and fixed capital, they do not corroborate Cunéo and Mairesse's (1984) results. However, they find that the bias due to double counting disappears in a first difference formulation.

Summing up, the different ways to test for the potential effects of double counting (2C), double expensing (EX1), and double expensing of net R&D capital stock (EX2) are characterized by

$$TFPG_i^{2C} \equiv TFPG_i + \left[\hat{\alpha}_i^Y \; h_{L,i} \frac{\dot{L}_i^R}{L_i^R} + \hat{\beta}_i^Y \; h_{K,i} \frac{\dot{K}_i^R}{K_i^R} \right] = \lambda_i + \rho^{2C} \frac{\Delta R_i}{Y_i} + u_i^8 \;, \quad (2.21)$$

$$TFPG_i^{EX1} \equiv TFPG_i^{2C} - \theta_i \left[\frac{\dot{Y}_i^Y}{Y_i^Y} - \frac{\dot{I}_i^R}{I_i^R} \right] = \lambda_i + \rho^{EX1} \frac{\Delta R_i}{Y_i} + u_i^9 \;, \quad (2.22)$$

$$TFPG_i^{EX2} \equiv TFPG_i^{EX1} + \delta_i \; \phi_i \left[\frac{\dot{Y}_i^Y}{Y_i^Y} - \frac{\dot{R}_i}{R_i} \right] = \lambda_i + \rho^{EX2} \frac{\Delta R_i}{Y_i} + u_i^{10} \;, \quad (2.23)$$

respectively.

(v) TFP - The rate of utilization of production capacities and labor hoarding

The measure of total factor productivity growth that is to be used is generally considered (cf. Griliches (1979)) to be affected by short-term fluctuations in capacity utilization. Some adjustments should be made for it in order to get a correct measure of the shift in technical change. Yet, such adjustment is most of the time impossible to implement, due to the lack of reliable indicators at the industry (or firm) level. Although utilization rates commonly relate to the use of fixed capital and equipment, a similar phenomenon appear with the utilization rate of employees, or labor hoarding. Maddison (1987) suggests that in normal circumstances one would not expect labor hoarding to be significant in advanced capitalist economies because market forces would cause workers to be laid off if they ceased to be productive.[10] The only country in which labor hoarding might have an influence on the productivity measures is Japan. This is perhaps an additional explanation of the diminishing returns to scale estimated by Griliches and Mairesse (1990) with a cross-section of Japanese firms. The authors attribute it partly to their inability to properly account for the problem of varying capacity utilization.

[10] According to Maddison (1987), only Japan may have been subject to a substantial labor hoarding because *(i)* a relatively high ratio of employed persons are self employed or family workers, and *(ii)* a significant portion of wage and salary earners have lifetime job security. The first reason has been substantially reduced over the past forty years. In 1950 this proportion in Japan was 60 percent and in 1984 it was 26 percent; whereas in the US it was respectively 20 and 9 percent.

It may be that within a panel data framework, unstable utilization capacities are more likely to yield abnormal measures of total factor productivity. Much of the short-run deviations from the long-run trend of productivity may be primarily attributable to the fluctuations in the level of capacity utilization. However, the few attempts in the empirical literature to correct for such fluctuations do not observe any noticeable change in the estimated impact of R&D on output growth.

Griliches (1980) notices that allowing for differential energy intensity (often used as a proxy for utilization rates) or adding capacity utilization measures, leads to no appreciable improvement or modification of the estimated output elasticity with respect to R&D capital stock. Similarly, the output elasticity of R&D capital stock, estimated at the US industry level by Griliches and Lichtenberg (1984a), is invariant to the inclusion of capacity utilization proxies,[11] and these variables do not contribute to improve the overall fit of the regression results. A few other empirical works take into account an indicator of utilization capacity but do not report the estimation results when the indicator is not included among the regressors. Dollar and Wolff (1993) use capacity utilization indices to adjust their macroeconomic *TFP* measures. They notice that the corrected *TFP* measure is almost identical to unadjusted *TFP*.

With time series cross sections of industries, a very partial solution adopted to reduce the productivity swings induced by fluctuating utilization capacities is to compute a moving average of the traditional measure of productivity. For instance, Adams (1990) with a panel composed of 18 US industries over a 28 years period, and Mohnen and Ducharme (1995) with a panel of 25 Canadian industries over a 17 years period, use a five years moving average of total factor productivity growth as dependent variable.

2.4.2. Measurement of R&D capital stock

There are four noteworthy points to be drawn underlying the computation procedure of R&D capital stock. First, a certain length of time is needed between the time that the investment in research is made and the time that the products embodying the new technology is sold and contribute to improve the productivity of an industry. Second,

[11] When capacity utilization indexes are not available, the most widely used proxies are the age of capital, the average annual hours of work, and the energy intensity (or consumption). The first proxy is also used as an indicator of returns to scale (cf. Griliches (1980)).

investments in R&D may depreciate and become obsolete. Third, the benchmark year has to be evaluated. Fourth, an appropriate R&D price index has to be used in order to deflate the R&D investment series.

The main implication of the presence of lags and depreciation rates is that the net impact of research activities is smaller than would appear at first sight. Despite the fact that research activities indubitably suffer from depreciation and lags, a significant number of empirical studies have largely ignored these factors because of the difficulties in determining them. The rationale usually given for ignoring depreciation is that it should be relatively small in a context of high growth rates of R&D expenditures.

The following equation shows how a knowledge capital stock, R, can be 'formally' obtained by allowing for both depreciation of research and a gestation lag:

$$R_t = \sum w_h \, I_{t-h}^R \ ,$$

where I^R is the gross investment in R&D and w_h reflects the assumed R&D lags and obsolescence schemes. Since there is no information to specify the lag structure and obsolescence scheme, it is commonly assumed that the lag structure is constant over time and that the obsolescence scheme is linear. Therefore, the construction of the R&D capital stock may be implemented through the perpetual inventory method:

$$R_t = I_{t-\theta}^R + (1-\delta)\, R_{t-1} \ , \qquad (2.24)$$

where θ is the average lag operator, and δ the assumed depreciation rate. Assuming that the growth rate of I^R, g, has been stable along a sufficiently large period, the benchmark R&D capital stock R_0 can be obtained as follows:[12]

$$R_0 = I_{0-\theta}^R \ \frac{1+g}{g+\delta} \qquad (2.25)$$

These lags, depreciation rates, and benchmark issues are largely ignored in the whole empirical literature. Most investigators prefer to rely on a specification identical to equation (2.9), with the no depreciation rate hypothesis and/or no or few lags, avoiding the apparent difficulties associated to the construction of an unambiguous measure of R.

[12] This method is described in appendix A2.2.

The next three points enter more closely into these three issues with respect to the existing literature, and to their empirical implications.

(i) The depreciation rate.

We have seen that the most common practice in empirical studies is to impose the zero depreciation rate hypothesis. This is particularly true with the cross sections, as can be seen in Table 2.1. The zero-depreciation procedure is a first-best practice when there is no available R&D investment figures over a sufficiently long period. Once the temporal dimension is investigated, the zero depreciation is still used but apparently less often than the arbitrarily chosen 10 percent depreciation rate, constant across industries (cf. the depreciation rates used by the empirical studies surveyed in Table 2.1, Table 2.2, and Table A2.1).

As noticed by Mairesse and Sassenou (1991, footnotes 19 and 20) and Capron (1992, p. 67-68), the estimates of R&D elasticities (γ in equation 2.7) may be considered as *a priori* much more robust than the estimates of rates of return (ρ in equation 2.9) with respect to the rates of obsolescence of R&D. This can be demonstrated as follows. Assuming a constant rate of growth of R&D investments, the R&D capital stock in each year could be expressed as in the benchmark equation (2.25):

$$R_t = I_t^R \cdot \frac{1+g}{g+\delta} \qquad (2.26)$$

Since the second term in the right-hand side of equation (2.26) is constant, estimating the output elasticity with respect to R&D investments or R&D capital stock would yield similar results. This is due to the fact that the constant term would be assimilated by the intercept when one takes the logarithmic model or it would disappear in the growth rates model characterized by equation (2.7). As long as the growth rate of R&D investment is stable over time, one can use indifferently the R&D capital stock or the gross R&D investments in order to estimate the output elasticity of R&D. The story is different when the focus is on the direct estimation of the rate of return to R&D. Here the assumption about the rate of obsolescence is crucial because, according to equation (2.9), we need to compute the net R&D intensity, which has as numerator

$$\Delta R_t = R_t - R_{t-1} = I_t^R - \delta R_{t-1}.$$

With the same assumption of a stable growth rate of R&D investments, this equation becomes:

$$\Delta R_t = I_t^R \frac{g}{g+\delta} \qquad (2.27)$$

Therefore, the use of gross instead of net R&D investment leads to the underestimation of the return to R&D as far as the last term of equation (2.27) is smaller than one, which is the case when the rate of obsolescence is positive. This result is interesting in the sense that it does not corroborate the conventional expectation that the estimated net rate of return to R&D should be smaller than the gross return, and that this difference should be roughly equal to the assumed depreciation rate. For instance, if the growth rate of R&D investments and the rate of obsolescence of knowledge are equal to 5% and 15%, respectively, then the net R&D is equal to one quarter of the gross R&D. Therefore, one would expect the estimated net rate of return to R&D to be higher (not lower) than the estimated gross return, by a factor of 4. On the other hand, if the growth rate of R&D investment is negative, and if the depreciation rate is smaller than the absolute value of g, using gross instead of net R&D investment would yield an overestimate of the return to R&D.

Equation (2.9) clearly indicates that the relevant concept for ΔR is the net R&D expenditures, not the gross. And the above discussion shows that using gross R&D investments instead of net might yield biased estimates of the net return to R&D. However, if net R&D investments are to be used, one have to make an explicit choice about the depreciation of R&D capital. One may therefore wonder if the obsolescence rate is really different from zero and, in the affirmative, what is its likely amplitude.

R&D capital stock depreciates because the created knowledge is forgotten and therefore useless without continued efforts at exercising, retrieving and transmitting it. In addition, some of the newer inventions render obsolete a significant share of previously acquired knowledge. Another force that fosters the rate of depreciation of research is related to the appropriability conditions. The more knowledge diffuse towards competing firms, the weaker the appropriability of the invention and the higher its depreciation rate.

Pakes and Schankerman (1984), referring to the different causes that may render knowledge obsolete, argue that the depreciation rate of R&D should be higher than that of fixed capital. From an econometric analysis of patent renewal data of five European countries, they estimate a 25% average rate of obsolescence of R&D which may vary from 18% to 36%. Focusing on French patent data, Pakes and Schankerman (1985) find a 28% depreciation rate. Still with patent renewal data, Bosworth (1978) evaluates a 10% to 15% range for rate of obsolescence of R&D capital in the UK. Schankerman and Pakes (1986) extend their analysis geographically and find a low depreciation rate in France (10%), closely followed by Germany (11%). The UK is associated to a higher rate of obsolescence equal to 17%. Schankerman (1991) provides further evidence from patent granted in France that the depreciation rates of knowledge vary across industries. The author evaluates a 3% depreciation rate for the Drugs industry, 4% in the Chemicals, 10% in Machinery, and 15% in Electronics. Although these results confirm that R&D capital stock depreciates, it should be noticed that a number of problems exist with the patent renewal data approach.

First, it is assumed that firms try to maximize the net profit of each patent. This kind of maximization exercise is in fact costly (both in terms of time and information required for it) and may be difficult to implement in the large corporations that hold many patents. Second, some patent holders might be highly risk averse, choosing to renew most of their patents. Third, the rate of obsolescence derived from patent renewal data may underestimate the real rate if patents are renewed even though they have no more economic value. Fourth, Goto and Suzuki (1989) encounter another problem with patent renewal data ensuing from the nature of the patent holder. Individual patent holders tend to abandon their patents more frequently than corporations. This difference in the composition of patent holders between industries (individuals *vs* corporations) is reflected in the rate of obsolescence of R&D capital across industries. Fifth, there are other ways of appropriating the benefit of research activities, such as head start and secrecy, which means that patent data cover only a share of industrial innovations.

Goto and Suzuki (1989) opt for the analysis of the Japanese Science and Technology Agency's survey of the 'life span' of technology in order to derive depreciation rates. This survey-based questionnaire gives the average length of time during which patents generate royalty revenues, and/or during which the product embodied by the technology generates profits. Taking the inverse of this average 'life span', Goto and Suzuki obtain an estimation of the rate of obsolescence of R&D capital. It turns out from their procedure that the yearly depreciation rate of R&D varies substantially across industries, from 6% for the food industry to 25% for precision machinery. The depreciation is generally higher for the industries in which technology is advancing rapidly.

To what extent does the choice of an R&D depreciation rate affect the estimated rate of return to R&D in empirical studies? Trying to give an answer to this question may be an alternative way to estimate the magnitude of the depreciation of R&D. Several authors choose their R&D capital stock through a sensitivity analysis of their econometric results with respect to variations in the depreciation rate. It should be noticed, however, that their conclusions diverge depending on the parameter of interest - e.g., output elasticities or rates of return to R&D -. This can be anticipated from equations (2.26) and (2.27).

Using a cross-section of US industries, Griliches and Lichtenberg (1984a) test for different R&D depreciation rates and their data favor the hypothesis of no depreciation of R&D capital. Under the 0% depreciation rate hypothesis their econometric results provide both the highest R^2 and *t*-statistic on private R&D intensity. Both of these statistics decline monotonically as the assumed depreciation rises. Yet, the estimated rates of return to private R&D sharply increase as the assumed depreciation rate goes from 0% to 10%, 20% and 30%, for the three subperiods studied by the authors.

In his panel data study of US industries, Griliches (1980) adopts the zero percent depreciation hypothesis. However, as opposed to Griliches and Lichtenberg (1984a), his experiments with the depreciation rates of 10, 20, and 30 percent lead largely to the same results. He notices that the data can hardly distinguish between them, showing a slight preference for 0 or 10% (cf. Griliches (1980), footnote 4, p. 345). This weak sensitiveness to the depreciation rate is most probably due to the fact the author estimates output elasticities of R&D instead of rates of return. Also using a panel data of US industries, Adams (1990) obtains a best fit for a 13% rate of obsolescence.

What can be learned from micro-level investigations? Hall and Mairesse (1995) conclude from their analysis of a panel of French firms that a 25% depreciation rate might perform slightly better than the 15% depreciation rate, and the higher depreciation rate yield a slightly lower output elasticity with respect to R&D (for the within and first-differenced estimates). Nonetheless, the differences are not significant enough to induce a definitive conclusion. They also estimate the rates of return to R&D and observe that there is no differences between the estimated gross and net rates of return to R&D. Cunéo and Mairesse (1984) and Griliches and Mairesse (1984), with panels of French and US firms, respectively, produce stable estimates of the output elasticity with respect to R&D capital, whatever the depreciation rate (0%, 15%, or 30%). These works confirm the idea suggested by equations (2.26) and (2.27) that the estimates of elasticities are more robust - e.g., less sensitive to the hypothesized depreciation rate - than the estimates of rates of return.

Goto and Suzuki (1989) obtain large differences in the estimated rates of return, depending on whether net or gross R&D intensities are used. These differences cannot always be explained by the depreciation rate of R&D. For instance, the firms in the sector of Organic chemicals are associated to a net rate of return to R&D of 81%, whereas their gross rate of return if equal to 56%. The reverse is true in other industries, such as Drugs and medicines, where the gross rate of return to R&D (42%) is above the net rate of return (23%). Therefore, as suggested by equation (2.27), net rates of return to R&D can not be deduced from their gross counterpart, and the assumed depreciation rate substantially affects the estimated parameter. In other words, disregarding R&D depreciation in the measurement of R&D intensity may have serious implications on the estimated returns to R&D.

(ii) The lags.

Several lag issues are involved in the R&D-productivity nexus. They reflect the time span between R&D investments and invention, between invention and development, and between development and commercialization or production. These three lags are sometimes generalized to two main components: the ‘gestation lag’ - e.g., the mean lag between project inception and completion - and the ‘application lag’ - e.g., the mean lag from project completion to commercial application -. Depending on the nature of the research implemented by firms, the whole lag may vary substantially. Griliches (1973) suggests that for basic research the average lag appears to be around 5 to 8 years, whereas for the bulk of industrial research (applied and development) the lag is much shorter, about two to three years, but still significant.

Similarly to the zero-depreciation hypothesis, the shortness of the available time series concerning R&D investments often forces investigators to ignore the lag issue when estimating the impact of research activities on output growth. The most common hypothesis is one of small or no lag. Different types of analyses provide evidence that the lag between R&D expenditures and productivity growth may be of several years. Some studies rely on ‘questionnaire’ to firms, while others prefer to rely on a best fit approximation of the lag structure.

Pakes and Schankerman (1984) summarize the mean R&D lag calculated by Rapoport (1971) and Wagner (1968). They obtain lags which vary between one and three years, according to the industry considered. More particularly, the average lags evaluated by Rapoport is equal to 1.7 in the Chemical industry, 2.4 in Machinery, and 1.2 in Electronics. The corresponding figures presented by Wagner are 2.6 for the Durables industries and 2.2 for the Nondurables industries. Goto and Suzuki (1989) rely on the survey of R&D activities of major firms implemented by the Japanese Economic Planning Agency. According to them, the length of R&D lag is two years for Electrical machinery, Electronics and components, and Metal products; five years for Drugs and medicines; and three years for the remaining industries. Except for the Drugs industry, these figures are broadly similar to those derived from Rapoport (1971) and Wagner (1968).

The studies which rely on a best fitting strategy - e.g., minimization of the sum of squared residuals or maximization of the R-squared - generally find higher lags. Branch (1974) finds out that the lags between the number of patents granted to US firms (with an application lag of 4 years) and sales growth vary from 1 to 3 years, depending on the industries. Ravenscraft and Scherer (1982) use data not at the firm level but rather at the level of individual 'business', from 42 units selling a distinct set of products to an identifiable group of customers and in competition with a well-defined set of competitors. They find strong evidence that the lag structure is roughly bell-shaped, with a mean lag of four to six years before R&D activities begin to generate profits.

Adams (1990) with eighteen US industries finds that the best fitting lags on own knowledge capital stock and on knowledge spillovers are, respectively, 20 and 30 years. Yet these lags are associated mainly to the basic research emanating from universities, the stock of knowledge is derived from articles count data of scientific papers. These lags are much longer than the lags estimated with R&D expenditures at the industry level. This is due to a strong orientation of industrial R&D towards applied research, which is known to be associated to much shorter lags. Mohnen and Ducharme (1996) find an optimal lag of three to six years for the R&D capital stock at the industry level in Canada.

In cross section studies, productivity growth rate is averaged over a period of several years, and the R&D intensity is computed generally for the beginning of the period or for one of the first years. Since inter industry differences in R&D intensity are relatively stable over time for both industries and firms, the assumed lag has no great importance in these investigations. For instance, as observed by Mansfield (1980), the coefficients of correlation between an industry's ratio of R&D expenditures to value added in 1953 and its ratio in 1975 is above 85%.

Among the three methods adopted in the literature - i.e., patent renewal data, questionnaire studies, and best fitting strategy - the first and the second ones seem to be more appropriate for our industry level analysis. Indeed, the best fitting strategy might not be suitable for such an aggregation level. Griliches (1979) recommends assuming a particular form *a priori*, as we should " *...not expect the data to answer such fine question*. ". Nevertheless, we shall refer in the next chapter both to the industry-specific lags provided by the Japanese Economic Planning Agency, and to a best fitting strategy in order to get a clear insight into this lag issue.

(iii) The benchmark stock.

Benchmark figures for R&D capital stocks are not available either at the macro level or the industry level. Some authors take the first observation on R&D investments as the benchmark; whereas others multiply it by a constant across firms or industries (for instance, the assumption of a 15% depreciation rate and a 5% annual rate of growth of R&D investments would imply the multiplication of the R&D expenditures in the first year by a factor of 4 in order to get the benchmark stock). A more accurate approach is to rely more formally on formula (2.14) and to try to have a good approximation of the annual growth rate of R&D investments during a sufficiently long period preceding the benchmark year.

Since this kind of data is generally scarce, authors approximate the R&D growth rate with either the ex-post growth rates of R&D, or the growth rate of fixed capital (or output) prior to the studied period, or the output growth rate. Few studies test for the sensitiveness of their results to the procedure used to compute the benchmark stock. Using a panel of French firms Hall and Mairesse (1995) obtain a slightly higher output elasticity when the knowledge capital is constructed from a longer history of R&D.

(iv) R&D and output - Price indexes

The effect of R&D on output growth might be biased downward due to errors in the producer prices used to deflate value added or total sales. According to Griliches (1979), product quality in high-technology industries rises faster than elsewhere and such quality improvements may be swept up by improper prices indexes. In fact, R&D intensive goods are likely to be underestimated and their prices overestimated. Further, Adams (1990) argues that output prices may be an incorrect weight when value added is used as output proxy in the production function. Griliches (1994a) presents a striking illustration of the implications ensuing the use of more reliable deflators. When adjusting the output price indexes for three out of 143 industries the estimated rate of return to R&D increases from 30 to 46% for the period 1978-89.[13] By withdrawing the computer industry from this sample, the price adjustment for two industries increases the estimated return to R&D from 12 to 35%.

Using panel data, Mairesse and Mohnen (1995) suggest that quality improvements may be captured by time and industry dummies, especially in the absence of reliable - i.e., adjusted for quality - price indexes. However, when time dummies are introduced into the empirical model, one generally gets lower estimates of the rate of return to R&D, not higher, as could be expected when adjusting prices for quality improvements. This explains why Griliches (1995) states that the results of the production function estimation should be viewed essentially as lower bound estimates of the contribution of R&D to the social output.

The rate of inflation in R&D is also very difficult to measure. For the lack of anything better, the most widely used R&D deflator in empirical studies is the GDP deflator. However, a proper R&D deflator for R&D outlays would have to reflect the prices of the different components of R&D (intermediate inputs, labor compensation, and fixed capital). Little is known about the extent to which real R&D deflators in various industries would differ from the GDP deflator. Mansfield, Romeo and Switzer (1981) estimate industry specific R&D deflators from a detailed data set obtained for about one-ninth of all company-financed R&D in the US. In their analysis, the industry-specific R&D price index is a Laspeyres index of price components of R&D. Their result show that the GDP deflator underestimates the rate of price increase for R&D inputs during the period 1969-1979 in most of the eight industrial sectors studied.

Jankowski (1991), using a similar methodology, evaluates the R&D deflators of 12 manufacturing industries over the period 1969-1988. From data on five different

[13] The *TFP* growth in the computer industry is downwardly adjusted by deflating material (mainly other computer components and semiconductors) purchases in this industry by the same output price index. The *TFP* growth in the semiconductors (electronic components) is upwardly adjusted with a similar transformation. The growth of *TFP* in the pharmaceutical industry is upwardly adjusted to reflect the exclusion of price declines due to the introduction of generics in the measurement procedures.

research inputs, Jankowski provides some evidence that, at the aggregate level, the GDP deflators are reasonable approximations of inflationary changes in R&D input costs. Therefore, an index specific to R&D does not appear absolutely essential for deriving acceptable estimates. However, on an industry-specific basis, broad-based GDP deflators may be much less appropriate for calculating real R&D expenditures.

The issue that arises when estimating the return to R&D at the industry- or company-levels, is to determine whether the estimates are sensitive to the use of an industry-specific R&D deflator. Harhoff (1994), using a panel composed by German firms, reports that his results are weakly sensitive to the adoption of sector-specific *vs* aggregate R&D investment deflator. Since there are no published R&D deflators at the industry level to our knowledge, the use of GDP deflators, which are highly similar to different types of aggregate R&D deflators, seem to be reliable to deflate the R&D investment series. Due to the lack of data for other countries, we shall test only for Japan whether using an aggregate R&D deflator affects the estimated rate of return to R&D at the industry-level.

2.4.3. Trade and TFP growth

Another factor that may foster the rate of technological change, and hence productivity growth, is related to export performance. This idea, known as the export-led growth hypothesis, suggests that countries or industries which do well in their export performance are likely to do well in their productivity performance. The question that arises is to determine the nature of this particular link. A strong correlation between the two aggregates might reflect either a growth accounting identity - exports as component of output - or a real causal link. The key issue here would be to determine whether the estimated rates of return to R&D would be affected if exports were taken into account in the empirical model.

In the first case, since exports is part of output, a sharp increase might allow for increasing returns to scale internal to the industry, which would be incorporated in the observed total factor productivity index. For instance, Maddison (1987) assumes that foreign trade produced ten percent of economies of scale over the period 1973-84 at the aggregate level of six large economies. The second case is more interesting since it sees export performance as having a stimulating influence on the industry considered. This influence may be characterized through technological spillovers as well as other externalities. Entering into international competition calls for an improved efficiency, provides incentives for innovations, and spurs specialization which, in turn, gives rise to external economies of scale. The recent theoretical literature on international trade has particularly focused on this relationship between productivity and exports.

Marin's (1992) survey on this literature shows that the link between these two aggregates is quite ambiguous. In the 'new' trade theory technical efficiency and trade seem to be the central link. However the likely effect of trade on technological growth is still not conclusive, depending mostly on the hypotheses of departure. The trade effect appears to depend on the type of competition assumed in the domestic economy, on whether entry and exit are frictionless, and on whether the market structure is sensitive to trade disturbances. Marin posits that exports are likely to affect productivity growth, but the sign of the causal impact is ambiguous. He also argues that whether exports would cause productivity gains or losses can, in the end, only be decided empirically, by letting the data 'speak' by itself, and without imposing too many theory-induced restrictions.

In order to explore the comovement of exports and productivity in the US, the UK, Germany and Japan, Marin (1992) uses quarterly observations over the period 1960-1987. His causality tests confirm that exports cause labor productivity in all four countries, inducing that past information on exports can improve the forecast of productivity. However, he recognizes that the cumulated quantitative impact of lagged exports seems to be negligible and range from -0.13 and 0.03 with positive signs for Germany and negative signs for the three other countries. Further, exports lagged six to eight quarters have still an impact on actual productivity measure, meaning that exports may affect productivity growth with a two years lag. In the light of Marin's results, we shall test whether two years lagged exports affect the estimated returns to R&D.

2.4.4. Econometric issues

Three econometric issues are to be discussed. The first one relates to the endogeneity of R&D, the second one to the inclusion of industry and time dummies into the model, and the third one to the potential presence of highly influential observations.

(i) The endogeneity of R&D.

A problem that may appear in the estimation procedures is the 'simultaneity' problem. If R&D is chosen on the basis of economic incentives, it is likely to be influenced by the forces affecting output, fixed investments and labor.[14] In practice, the bias associated to simultaneity are more likely to appear when the knowledge variable is proxied by contemporaneous R&D investments or contemporaneous R&D capital stock. As far as a substantial number of lags are taken into account, the simultaneity between R&D capital stock and output should vanish. In other words, assuming some lags between the time of the R&D investment and the time of its potential effect on output growth suffices to rule out the effect of output on R&D investments.

(ii) Industry and time dummies.

In general, the productivity growth of industries may differ for reasons not incorporated in R&D variables (such as market structure, appropriability conditions, industry size, technological opportunity, government intervention). These differences among industries, provided they are stable over time, may be endogenized by industry dummies. If a model similar to equations (2.7) and (2.9) has to be estimated, industry dummies are deemed to capture disembodied technical change specific to each industry. One can see in Table 2.1 that three out of the seven panel data studies at the industry level do not allow for industry-specific intercepts. We shall test in the next chapter whether the presence of industry dummies into the regression equation affects the estimated rates of return to R&D.

Mohnen and Ducharme (1996) observe that when they introduce industry dummies among the right-hand side variables in order to take into account the potential differences between industries in the impact of disembodied technical change, the estimated rate of return to R&D gets smaller and less significant. According to the authors, this reduction of the estimated impact of R&D might be an indication that the technological opportunities specific to each industry are difficult to be distinguished from their rate of return to R&D.

[14] Recall that one of the reasons underlying our choice of the total factor productivity approach instead of the traditional production function is related to potential simultaneity bias among output, capital, and labor in the latter specification. Hall and Mairesse (1995) use input measures from the beginning of the year- i.e., lagged one year - to mitigate the effect of simultaneity between factor choice and output, but recognize that simultaneity can still be a problem.

The introduction of time dummies for each year in the estimation are deemed to take into account changes over time in the rate of productivity growth which are common to all industries. Although these time dummies are generally used to pick exogenous shocks, their presence may also have a downward effect on the estimated rate of return to R&D. It is worth noticing that the introduction of time dummies is a well established practice in micro-level works but is much less frequent among industry-level investigations. Out of seven panel data analyses presented in Table 2.1, only two include time dummies among the right-hand side variables (namely, Griliches (1980) and Griliches and Lichtenberg (1984a)) and Suzuki (1985) includes a time trend. Using time dummies (instead of a trend) allows the time effects to vary over time and reduces potential multicollinearity problems with the other right-hand side variables.

(iii) Outliers and influential data points.

An additional important issue when measuring the rate of return to R&D is related to the definition of the sample from which the estimates are derived. In both industry- and firm-level investigations, the presence or the absence of outliers and influential data points might drastically modify the estimated parameters. Nevertheless, very few studies on the relationship between R&D activities and output growth, at the firm- and industry-levels, test for the presence of such 'dirty' or aberrant observations.[15]

Micro-level studies generally adopt an arbitrary data-cleaning procedure before turning to the estimation of parameters. The observations - i.e., the firms - for which some variables have suspicious values according to some criteria are simply erased from the sample.[16] To our knowledge, Lichtenberg and Siegel (1989) is the only firm-level paper which attempts to test the sensitivity of the estimates to the presence of potential outliers and influential data points. The authors test the sensitivity of their results using two methods: estimation by least absolute deviation and deletion of influential outliers.

[15] Verspagen (1997) tests for the presence of outliers in the explanatory variables, noticing that the results are quite sensitive to the suppression of these outlying observations.

[16] For instance, the criteria used by Hall and Mairesse (1995) in order to 'clean' their data set are: to remove any observation for which *(i)* value added is zero or negative; *(ii)* the value added by worker is outside of three times the interquartile range above or below the median; *(iii)* the growth rate of value added is less than minus 90% or greater than 300% (and similarly with labor, fixed capital and R&D capital); *(iv)* the inputs devoted to research activities account for more than 50% of the total inputs. The removing criteria used by Capron and Cincera (1997) delete the observations for which *(i)* R&D intensity is less than 0.2% or greater than 50%; *(ii)* and *(iii)* similar to the corresponding criteria of Hall and Mairesse (1995) but with slightly different thresholds. Mairesse and Hall (1995) used only one criterion, which is similar to *(iii)* .

The least absolute deviation estimator is a 'bounded influence' estimator which minimizes the sum of absolute deviations of observations from their means.[17] Since the sum of the *absolute*, instead of *squared*, deviations is minimized, this alternative estimator reduces the impact that influential observations have on the fitted regression function. Outlying observations are detected through the Cook (1977)'s influence statistic that measures change in the estimated parameters vector that results if the *i*th observation is deleted. Using a data set composed by 1092 US firms over three subperiods, they find that with the least absolute deviation estimator the rate of return to total R&D is 29.5% lower than their original least squares estimates. The point estimate of the same parameter declines by 3.8% when outliers are discarded.

Similarly to their micro-level counterparts, industry-level investigations barely test for the presence of influential data points. So far, the only attempt to test for the presence of outliers at the industry-level has been made by Reiss (1990). The author re-examines the analysis of Griliches and Lichtenberg (1984a) by looking for potentially aberrant industries. Using a cross-sectional data set of 27 industries, Reiss identifies four potential outliers. The rates of return to private R&D turns out to be slightly sensitive to the deletion of these four industries, decreasing from 35% in the full sample to 26% in the "clean" sample. For government R&D, the change is more remarkable. In the full sample, the return to R&D subsidies is non significantly different from zero (1%), but it sharply increases to a significant 18% when the outlying industries are discarded.

Griliches (1994a) and Hanel (1994) show that their own cross-section estimates at the industry level are strongly sensitive to the removal of the computer industry. When the outlying computer industry is removed, the estimated rates of return to R&D fall significantly in both studies. This drop is most probably due to the output price indexes used to deflate sales or value added. Indeed, in Griliches (1994a), when the deflator is adjusted for three industries (including computers) out of 143, the estimates are much less sensitive to the withdrawal of the computer industry from the data set. Is it right to treat the computer industry (or any other industries) as an outlier and exclude it from the estimates? Griliches's (1994a) experiment hints that the results derived from the exclusion procedure are different from those derived from the price-adjustment procedure, which are closer to reality. However, although it is difficult in practice to improve the measurement of the outlying observations, the withdrawal of outliers may

[17] As noted by Lichtenberg and Siegel (1991, footnote 32), the standard errors for theparameter estimates through this robust estimator were unknown before the nineties; the statistical properties of the sampling distribution of this estimator were not well defined.

provide a good insight into their influence on the estimated parameters. In any case, looking for outliers allows to determine the industries or the observations for which there are measurement problems or unexpected behaviors.

Panel data studies of industries do not test for the presence of outlying observations. Griliches (1980) is the only empirical analysis among the six panel data investigations presented in Table 2.1 in which there is some sensitiveness analysis regarding outlying industries. Although he does not report any alternative results, Griliches notices that discarding industries whose price deflators are of dubious quality led to no appreciable improvement. This is not surprising since his regression equation includes industry dummies which partially incorporate inter industry differences in the quality of price deflators. That is, in the case of a panel data analysis of industries, and provided that the inter industry differences in quality measurement are stable over the studied period - e.g., provided they may be endogenized by the industry-specific intercepts - one may expect that no industry would appear as an outlier.

However, one may wonder whether some observations, for a given year and a given industry, have a strong influence on the estimated parameters. More generally, are our estimates likely to be biased by some kinds of model failure? In order to provide a proper answer to this question, one has first to define the potential sources of model failure. There are two main categories of model failure that might be serious: gross errors and misspecification. Whereas the former results from keypunch errors, inherently low precision numbers and improperly measured data, the latter derives from the fact that the model itself is not a good approximation. In occurrence, there can be some omitted explanatory variables that could be listed along with R&D capital stock in order to explain total factor productivity growth. In our case, these two sources are closely inter-related since the measured total factor productivity relies on some restrictive hypotheses. For instance, large swings in capacity utilization rates, improper deflators, and strong year-to-year fluctuations in inputs or sales are among the factors that may result in sharp fluctuations of the observed year-to-year total factor productivity growth which would not reflect the actual evolution of technical change.[18] The combined influence of the nine previously discussed technical and conceptual issues suggests that our empirical specification may be subject to some kind of model failure.

---

[18] Wage rigidity combined with fluctuations in value added might result inhighly volatile shares of value added for capital and labor, especially in the short run.

What observations, or what types of data points must be downweighted in order to provide some protection against model failure? Belsley *et al.* (1980) recall that a common practice in applied econometrics is to associate dummies to the largest residuals which correspond to unusual events. This procedure is approximately equivalent to deleting these particular observations; and the resulting distribution of residuals will then appear to be much better behaved. In our case of productivity growth analysis, since it is impractical to model reality in its full complexity, this kind of dummy variable procedure is used in order to prevent contamination of the estimated rates of return to R&D. Formally, the common practice in applied experiments is to:

(i) run an OLS regression,
(ii) examine the observations with large residuals and determine whether they should be treated separately from the bulk of the data, and
(iii) rerun OLS regression with the observations deleted or dummy variables added.

According to Krasker *et al.* (1983) the observations that may drastically change the estimated parameters are not only the ones which have large residuals, but may also be the ones with 'extreme' X (explanatory variables) rows, or both large residuals and extreme X values. They convincingly argue that in the presence of outliers among the data set, the estimates might be totally misleading - i.e., the behavior of the bulk of the data would be misrepresented - and, worse yet, the outlying data points would not necessarily be characterized by large residuals and might therefore go unnoticed. This is due to the fact that by trying to fit all the data well, under the assumption that the model is correct, least squares frequently hide the true nature of the data.

In what follows, we present different 'influence' statistics that allow to test whether the data set contains outlying observations. These various diagnostic tools aim at locating errors or reporting legitimate extreme data that are likely to influence greatly the estimated model. A first possible indicator, *E10*, would rely 'only' on a crude analysis of the error term. The outlying observations would be those whose error terms are situated in the 10% range from the lowest (e.g., negative) and the highest (e.g., positive) error terms.

Another way to characterize outlying observations is to determine whether there are influential X-data.[19] In this case the *leverage* points are determined exclusively from the

[19] See Krasker *et al.* (1983) for more information on this diagnostic test (*HAT*) as well as on the two subsequent tests (*EIS*, and *DIFF*).

analysis of the explanatory variables. The indicator, *HAT*, is derived from the diagonal elements h_i of the *HAT matrix*: $H = X\ (X^T\ X)^{-1}\ X^T$. Since H is a projection matrix, its diagonal elements are included into the range [0, 1] and their sum over all observations (n) is equal to the number of explanatory variables (p). Therefore, a perfectly balanced X-matrix - i.e., with equal leverage for all observations - is one for which $h_i = p/n$. Belsley *et al.* (1980) demonstrate that the observations for which h_i exceeds $3\ p/n$ can be considered as potentially influential.

The *studentized residuals*, *EIS*, correspond to the residuals (e_i) scaled by h_i and $s(i)$, the sample standard deviation excluding the ith row, so that the denominator is stochastically independent of the numerator: $EIS_i = e_i\ /\ s(i)\ (1-h_i)^{0.5}$. The observations for which the studentized residual exceeds 2.5 may be considered potential outliers. Although these residuals are scaled, they might lead to an inadequate diagnostic. In practice, they can be augmented by leverage information derived from the *HAT* matrix.

The last diagnostic tool, *DIFF*, combines the information incorporated in both the studentized residuals and the right-hand side variables. This indicator is computed from the observed influence of a particular observation on regression estimates by comparing ordinary least squares based on the full data set with estimates obtained when one observation at a time has been deleted. The hat matrix diagonal and the studentized residuals are combined as follows: $DIFF_i = (h_i\ /\ (1-h_i))^{0.5}\ EIS_i$. It ensues from the computation of the *DIFF* statistic that the ith observation will have no influence even if *EIS* is large provided h_i is small. Furthermore, large leverage points may be a major source of influence on the fit even when the studentized residual is small. The highly influential data points are those for which *DIFF* is above the value of $3\ (p/n)^{0.5}$.

If highly influential observations are detected, how should they be treated? The appropriate remedies include, but are not restricted to, bounded-influence estimations and more particularly the use of dummy variables according to the three-step procedure described above. Quoting Krasker *et al.* (1983), the *DIFF* diagnostic tool is more appropriate than the *E10* or *EIS* diagnostics. The authors argue that using a dummy variable or withdrawing the observations that are clearly highly influential is appropriate for relatively small samples. No real payoff for using a more formal procedure might be expected. However, if the values of the outlying observations are correct, then their deletion or their corresponding dummy variable might eliminate vital information. Nonetheless, if the main interest is to catch the behavior of the bulk of the observations, downweighting a very small share of the data set is still a reliable practice.

There are two main other types of robust estimations that may be used to cope with very large observed residuals: least absolute deviation (LAD) and generalized least squares. LAD is an estimator that minimizes the sum of the absolute values rather than the square of errors. In practice, it reduces the effect of egregious errors. Referring to the survey of Taylor (1974), this estimator has historically never attracted much attention and its distribution theory has remained a problem. As mentioned by Krasker *et al.* (1983), resistant estimation through LAD has remained peripheral to mainstream econometric work because of computational costs as well as the absence of widely available code designed for this purpose, and overall, because of the lack of convincing theoretical support. Nowadays, the statistical properties are better defined (see Pollard (1990) and Phillips (1991)) but the least absolute deviation estimator is still barely used in the type of studies we survey.

The generalized least squares allows for heteroscedasticity, and when computed by weighted least squares (the weights being derived from the ordinary least squares' residuals), may be contrasted to the bounded-influence estimates related to the analysis of the residuals - e.g., the *E10* and *EIS* diagnostics. The main difference between weighted least squares and LAD, and the bounded-influence statistic *DIFF*, is that 'only' large error variances are downweighted with the former, while highly influential observations are downweighted with the latter. That is, if the data set is actually infected by very influential observations, only the *DIFF* procedure would allow to downweight their effects on the estimated parameters.

2.5. Concluding summary

There are many ways of estimating the rate of return to R&D. The first step we have to bridge is to select an empirical model among the existing possibilities. In the light of most studies in this field, the total factor productivity growth framework is chosen. Although still widely used, this empirical approach is subject to series of criticisms related to the implicit hypotheses that have to be imposed and to numerous measurement issues.

A critical survey of the literature on the relationship between R&D investments and economic performances of manufacturing industries permits to put the empirical analysis of Chapter 3 in context and to clarify some issues related to the structure of the data set and the proxy used for output growth. Five important observations emerge from this analytical survey. They are mainly related to the advantage of using panel data, to the choice of the parameter to be estimated (i.e., elasticity or rate of return), to the output proxy (i.e., total sales or value added), to slight modifications of the empirical model, and to the problems of international comparability.

The broad environment of the empirical analysis being set with respect to these unavoidable preliminary choices, we get to the heart of the matter. There are myriads of technical and conceptual issues that have to be clarified. They are related to the measurement of total factor productivity growth, R&D capital stock, and to the relationship between them. Concerning the total factor productivity indicator, the hypotheses of perfect competition, constant returns to scale, a wrong evaluation of labor cost shares, the double counting of capital and labor devoted to research activities, and the failure to take into account the utilization rate of production capacities might lead to biased estimates of the return to R&D. Concerning the measurement of R&D capital stocks, wrong lag structure, R&D depreciation rates, and R&D deflator might also lead to biased estimates of the impact of R&D. Finally, for estimation purposes, it is important to ensure the exogeneity of the R&D variable, to allow for industry and time dummies and to control for outlying observations.

Several tests are suggested in order to ensure the robustness of the estimated rate of return to R&D. Some of them are trivial but might be crucial, such as including industry and time dummies into the regression equation. Others are less trivial but barely used, such as allowing for free returns to scale, correcting for the double counting of R&D, and relying on a best fitting strategy to determine the most appropriate R&D depreciation rates and lag. Finally, various tests might be seen as pioneering with respect to the present state of the art, such as testing for perfect competition, using industry-specific R&D depreciation rates and lag, and mitigating the influence of outlying observations.

The next chapter is devoted to the empirical analysis and tests the robustness of the estimated rates of return to these various measurement issues.