On Nonparametric Estimation of the Schumpeterian Link between Innovation and Firm Size: Evidence from Belgium, France, and Germany

Irene Bertschek*)[1]   Horst Entorf**)

*) CORE and Institut de Statistique, Louvain-la-Neuve
**) University of Mannheim

April 1995

Correspondence to:
Irene Bertschek
Institut de Statistique
Voie du Roman Pays 34,
B-1348 Louvain-la-Neuve
e-mail:bertschek@stat.ucl.ac.be

[1]We are grateful to W. Pohlmeier and to the Ifo-Institute for providing us with the German data, to B. Crepon, J. Mairesse and to INSEE for the French data set, and to J. Bughin and F. Fecher for the Belgian data set. The analyses based on the French data were performed when both authors were visiting CREST-INSEE. Thanks also go to Ph. Barbe, W. Härdle, N. Rhomari and to participants of a seminar at the Humboldt-Universität zu Berlin for helpful comments. The first author acknowledges financial support by the Human Capital and Mobility Program, ERBCHBICT 941032, and by the Deutsche Forschungsgemeinschaft, SFB373 at the Humboldt-Universität zu Berlin.

Abstract

This paper analyses the Schumpeterian link between innovative activity and firm size by means of the nonparametric Nadaraya-Watson estimator and of the partially linear approach by Speckman. Four data sets referring to the manufacturing industries of three European countries are available for the analysis. We demonstrate how nonparametric methods can produce more reliable conclusions than conventional methods. For this purpose, the roles of bandwidth choice, wild bootstrap, density estimation and trimming are studied. For the German data set of 1984 and for the French data set we find that small firms and large firms are more innovative than firms of intermediate size while the relation is rather hump-shaped for Germany 1989 and decreasing for Belgium. Including an additional parametric component into the estimations based on the French data contributes considerably to the explanation of innovative activity without affecting the U-shaped link betweeen innovation and firm size.

KEYWORDS: innovative activity, Nadaraya-Watson estimator, partially linear model, cross-validation, wild bootstrap

JEL-classification: C14, L11.

1 Introduction

Since Schumpeter's (1942) seminal work, claiming that large firms are more innovative than small firms[2] , the link between innovative activity and firm size is one of the most discussed topics in industrial economics, a fact which is not least due to the controversal empirical results.

This study presents an international comparison of the innovation – firm size relation based on four data sets of three European countries. A nonparametric estimation method allows a more cautious analysis of the data since it relies more on the data than do the commonly used parametric approaches, and thus it should give new and more reliable insights.

The purpose of the paper is twofold: First, it attempts to investigate the feasibility and practicability of estimating nonparametric and semiparametric models by kernel methods in the context of an example related to industrial economics. Second, it analyses the still widely discussed relationship between innovation and firm size by methods which were up until now rarely applied for economic analyses.

A comprehensive and critical survey of the empirical literature on the relationship between innovation and firm size is given by Cohen and Levin (1989). The majority of earlier empirical literature finds innovative activity increasing with firm size, either monotonically (e.g. Link, 1980) or more than proportionately, at least up to a certain threshold (e.g. Scherer, 1965a; Scherer, 1965b). Recent empirical studies show a rather nonlinear relationship in the sense that small and large firms are more innovative than firms of intermediate size, for instance Acs and Audretsch (1990) and Bound, Cummins, Griliches, Hall and Jaffe (1984) for the U.S., Cremer and Sirbu (1978) for France, and Entorf (1988) for Germany.

Up to now, only parametric frameworks have been used for the analysis of the innovation — firm size relationship. However, if this is nonlinear, a nonparametric approach could be more appropriate, since it does not impose a certain functional form but relies on the data and can therefore take account of eventual nonlinearities. In this analysis the Nadaraya-Watson estimator based on kernel density estimation is applied. Basing the statistical inference on estimated densities is particularly

[2]It has to be mentioned that most articles on the link between innovation and firm size investigate the hypothesis that innovative activity increases more than proportionately with firm size. However, referring to Cohen and Levin (1989, p. 1071), Schumpeter just claimed that large firms disposed of R&D laboratories which "are the source of most innovation".

important for the validity of Schumpeter's hypothesis, since the main conclusions are drawn from the evidence of very large and very small firms, i.e. from the performance of the firms belonging to the extreme tails of the firm size density function.

A further model is considered including a nonparametric component and a parametric component. This semiparametric model allows us to take into account variables which are supposed to have a linear impact on the dependent variable, here innovative activity. The model is estimated by the method proposed by Speckman (1988). It is based on partial residuals, using the Nadaraya-Watson weights for estimating the nonparametric component and OLS for estimation of the parametric component.

For our analysis, four different data sets are available: First, a data set containing 1088 firms of the German manufacturing industry for the year 1984. Second, a corresponding data set for the year 1989 consisting of 962 firms. Third, a data set with 366 firms of the French manufacturing industry averaged over the years 1984 to 1988 and fourth, a data set with 277 firms of the Belgian manufacturing industry averaged over the years 1977 to 1983. In the first two cases a variable of innovative output is used — the percentage share of sales due to products in the innovation phase, while in the French and in the Belgian data sets measures of innovative input (R&D-capital and R&D-expenditures, respectively) are available.

For the German data set of 1984 and for the French data set we find that small firms and large firms are more innovative than firms of intermediate size while the relation is rather hump-shaped for Germany 1989 and decreasing for Belgium. Including an additional parametric component into the estimations based on the French data contributes significantly to the explanation of innovative activity without affecting the U-shaped link betweeen innovation and firm size.

The proceeding of the paper is as follows: Section (2) presents the econometric tools. Section (3) gives a description of the data. Section (4) shows the estimation results, and section (5) contains some concluding remarks.

2 The Econometric Tools

The advantage of nonparametric estimation is that no assumption has to be made with respect to the functional link m between dependent and independent variable as well as with respect to the distribution of the independent variable or of the error term. In case of randomly drawn pairs of observations $\{Y_i \in I\!R, X_i \in I\!R\}_{i=1}^n$ a nonparametric regression model can be formed as

$$Y_i = m(X_i) + \varepsilon_i \tag{1}$$

where $\{\varepsilon_i\}_{i=1}^n$ are random variables with mean zero and variance σ^2, X_i and ε_i are supposed to be independent.

2.1 The Nadaraya-Watson Estimator

Departing from model (1) the conditional expectation of Y

$$m(x) = E(Y|X = x) \tag{2}$$

can be calculated as

$$m(x) = \frac{\int y f(x, y) dy}{f(x)} \tag{3}$$

where $f(x)$ is the marginal density $\int f(x, y)dy$ of x. Expression (3) can be estimated by the *Nadaraya-Watson* method which is based on kernel density estimation:

$$\hat{m}_h(x) = \frac{n^{-1}h^{-1}\sum_{i=1}^n K(\frac{x-X_i}{h})Y_i}{\hat{f}(x)}. \tag{4}$$

The estimator is a weighted average of the response variable Y,

$$\hat{m}_h(x) = n^{-1}\sum_{i=1}^n W_{hi}(x)Y_i, \tag{5}$$

the weights $W_{hi}(x)$ depend on the distance between x and X_i, the smoothing parameter h and the kernel function K[3]:

$$W_{hi}(x) = \frac{h^{-1}\sum_{i=1}^n K(\frac{x-X_i}{h})}{\hat{f}(x)} \tag{6}$$

The denominator of $\hat{m}(x)$ represents the *Rosenblatt-Parzen* kernel density estimator (Rosenblatt, 1956; Parzen, 1962) with respect to the explanatory variable x:

$$\hat{f}(x) = n^{-1}h^{-1}\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \tag{7}$$

[3] A kernel is usually chosen as a continuous, bounded and symmetric real function which integrates to one.

In our analysis we always use the *quartic kernel*

$$
\begin{aligned}
K(u) &= \frac{15}{16}(1-u^2)^2 \qquad \text{if} \qquad |u|<1 \\
&= 0 \qquad\qquad\qquad\quad \text{otherwise}
\end{aligned}
\tag{8}
$$

but of course other kernels can be used as well. However, it is well known that the choice of the kernel is not as important as the choice of the smoothing parameter (cf. for instance Silverman, 1986, p. 43).

Proposition 3.1.1 in Härdle (1990a, p. 29) shows under which assumptions the Nadaraya-Watson estimator is consistent. Among others, the following condition must hold:

$$
n \to \infty, \quad h_n \to 0, \quad nh_n \to \infty \tag{9}
$$

i.e., the bandwidth should tend towards zero as the number of observations increases, but not too fast. This qualitative guideline, however, does not help much in finding the "optimal" degree of smoothing in practice. The larger the bandwidth h, the smoother the estimated curve, and the smaller the bandwidth h, the higher the variance of the estimated curve. Regression curves with high variance can make an interpretation of the estimation difficult whereas curves which are too smooth might hide important structural peculiarities.

Although the Nadaraya-Watson estimator has already been proposed in 1964, by Nadaraya and Watson separately, its application to economic problems is still scarce which is probably due to the fact that nonparametric methods are quite time-consuming and computer-intensive. For an example of an application in economics see Härdle (1990a, p. 6). Here, he estimates Engle curves relating expenditures for potatoes to net income. Applications of nonparametric techniques to biometry are more common, for example, in order to estimate the height growth of children in dependence of their age (Gasser and Müller, 1984).

2.2 On the Problem of Bandwidth Choice

In parametric regression a "good" estimator is defined by criteria such as unbiasedness, efficiency, rate of convergence. For instance, according to the Gauss-Markov theorem, which is a small sample result, the simple ordinary least squares estimator (OLS) is known to be BLU - best, linear, unbiased. In addition, OLS can be shown to be $\sqrt{n}$-consistent, a common property for parametric estimators.

Things are different for nonparametric estimators since first, they are generally biased, second, they are only locally linear, and third, they converge at a speed slower than parametric estimators. However, this is the price to pay for a higher flexibility. Due to the bias, it is not enough just to look at the variance of an estimator in order to draw conclusions about its efficiency. Therefore, criterion functions like the pointwise *mean squared error* (MSE)

$$d_M(x,h) = E(\hat{m}_h(x) - m(x))^2 \tag{10}$$

have to be used to measure the precision of the estimator in terms of variance and squared bias:

$$d_M(x,h) = [E(\hat{m}_h(x)^2) - (E\hat{m}_h(x))^2] + (E\hat{m}_h(x) - m(x))^2. \tag{11}$$

Global measures like the *mean integrated squared error* (MISE)

$$d_M(x,h) = E\int(\hat{m}_h(x) - m(x))^2 f(x)w(x)dx, \tag{12}$$

where $w(x)$ represents a weight function, can be considered to analyze the speed of convergence. This speed is usually of order $n^{-r}, 0 < r < 1/2$ and depends on the smoothness of the true curve m, the dimension d of the explanatory variables X, the kth derivative of m, which has to be estimated, and on the type of the estimator (Härdle, 1990a, p. 91). Under several regularity conditions, kernel estimators are asymptotically optimal, i.e. they achieve the optimal (where optimal means the lower and achievable) rate of convergence, by a suitable choice of bandwidth sequence and kernel function (Härdle, 1990a; Stone, 1982). In order to derive the theoretically optimal bandwidth one can calculate the *approximated MSE*, A-MSE, by linearizing the Nadaraya-Watson estimator and by means of a Taylor series expansion which gives the following terms for the variance and the squared bias:

$$variance \quad = \quad \frac{\sigma^2(x)}{nhf(x)}c_K \tag{13}$$

$$bias^2 \quad = \quad h^4\frac{(m''f + 2m'f')(x)^2}{4f(x)^2}d_K^2 \tag{14}$$

with the two kernel constants:

$$c_K = \int K^2(u)du \quad d_K = \int u^2K(u)du. \tag{15}$$

Hence, the A-MSE of the Nadaraya-Watson is:

$$\frac{\sigma^2(x)}{nhf(x)}c_K + \frac{h^4}{4}\left\{\frac{m''(x)f(x)+2m'(x)f'(x)}{f(x)}\right\}^2 d_K^2. \tag{16}$$

Increasing the bandwidth h reduces the variance but increases the bias of the estimate and vice versa. The "optimal" bandwidth is proportional to $n^{-1/5}$ in the one-dimensional case, and it can be shown to be proportional to $n^{-1/(4d+1)}$ for dimension d (Härdle, 1990a, pp. 46, 99–100). It should be clear that first, such results are based on approximations, and second, that the optimal bandwidth depends on factors which are unknown in practice, such as the second derivative $m''(x)$ of the true regression curve $m(x)$. One way out of this dilemma is to base the bandwidth choice on the data which are to be analysed.

Automatic bandwidth selectors are based on the minimization of quadratic error measures of the regression curve such as the MSE or the *average squared error* (ASE) by balancing the trade-off between squared bias and variance. For instance, the ASE

$$d_A = n^{-1}\sum_{i=1}^{n}(m(X_i) - \hat{m}_h(X_i))^2, \tag{17}$$

could be estimated by the *prediction error*

$$p(h) = n^{-1}\sum_{i=1}^{n}(Y_i - \hat{m}_h(X_i))^2. \tag{18}$$

A closer look at $p(h)$ reveals that it would give a biased estimate of d_A since observation Y_i is used for estimating $\hat{m}(X_i)$. The smallest bandwidth would be chosen as the most favorable since observations Y_i can best be predicted by themselves, and hence, the best regression curve would be an interpolation of the observations Y_i.

One way to obtain an unbiased estimate of d_A is *cross-validation*:

$$CV(h) = n^{-1}\sum_{i=1}^{n}[Y_i - \hat{m}_{h,i}(X_i)]^2 \tag{19}$$

where $\hat{m}_{h,i}(X_i)$ represents the *leave-one-out* estimator excluding observation Y_i from the estimate $\hat{m}_h(X_i)$.

Another possibility is the concept of *penalizing functions* where the prediction error is multiplied by a correction function which penalizes values of h which are too small and thus asymptotically cancels the bias of $p(h)$ as an estimate of d_A. Several penalizing functions are proposed in the literature which are all asymptotically eqivalent. But, as Härdle (1990a, p. 165) points out, the estimated bandwidth tends to the optimal bandwidth at an extremely slow speed.

Probably the most popular concept of penalizing functions is *generalized cross-validation* (GCV) proposed by Craven and Wahba (1979). The cross-validation function can be expressed as the sum of the ordinary residuals divided by the factor $1-W_{hi}(X_i-X_i)=1-W_{hi}(0)$, where W_{hi} represents the weights used for averaging the observations Y_i (compare equation (6)):

$$CV(h)=n^{-1}\sum_{i=1}^{n}\left(\frac{Y_i-\hat{m}(X_i)}{1-W_{hi}(0)}\right)^2 \tag{20}$$

GCV replaces the factors $1-W_{hi}(0)$ by their average $1-n^{-1}tr(W_h(X_i))$ since the diagonal elements of the weight matrix W_h are all the same. Hence, the GCV function is written as:

$$GCV(h)=n^{-1}\frac{\sum_{i=1}^{n}(Y_i-\hat{m}(X_i))^2}{1-n^{-1}trW_h(X_i))^2}. \tag{21}$$

where the numerator is the prediction error $p(h)$ and the denominator denotes the penalizing function (see for instance Green and Silverman, 1994, pp. 30–36). For the Nadaraya-Watson weights the penalizing function becomes:

$$\Xi(n,h,f(x))=(1-n^{-1}h^{-1}K(0)/\hat{f}(x))^{-2}. \tag{22}$$

The aim is to find the bandwidth which minimizes the CV or GCV-function. Therefore, the function has to be calculated for a grid of different values of h which makes this kind of bandwidth choice a computer-intensive task. Calculation time can be reduced by applying discretizing methods like WARPing (weighted averaging of rounded points) as proposed by Härdle and Scott (1992) which leads asymptotically to the same results (see also Härdle, 1990b, ch. 6 for details). More details on automatic bandwidth selectors can be found in Härdle (1990a, ch. 5).

2.3 Estimating Pointwise Confidence Intervals by the Wild Bootstrap Method

It can be shown that the suitably normalized term $(\hat{m}_h(x)-m(x))$ is asymptotically normally distributed with the mean vector corresponding to the asymptotic bias and an identity covariance matrix (Härdle, 1990a, theorem 4.2.1 and pp. 99–100). These results can be used to define confidence intervals for the curve estimate, while neglecting the bias term. One possibility taking account of the bias term is undersmoothing, i.e. choosing the bandwidth proportional to $n^{-1/5}$ times a sequence which tends slowly to zero. But how is this sequence determined in practice?

Another possibility incorporating the bias is given by the *wild bootstrap* method (Härdle and Mammen, 1991) which, moreover, allows for heteroscedasticity of the error terms and corrects for local skewness. The idea is to resample from each single residual instead of resampling from the whole set of residuals. The latter is used for instance by Härdle and Bowman (1988) for estimating pointwise confidence intervals.

In the case of wild bootstrap, a bootstrapped residual ε_i^* is drawn from each *single* estimated residual $\hat{\varepsilon}_i = Y_i - \hat{m}_h(X_i)$ according to a two-point distribution, for instance

$$P(\varepsilon_i^* = \hat{\varepsilon}_i a) = \gamma \tag{23}$$

$$P(\varepsilon_i^* = \hat{\varepsilon}_i b) = 1 - \gamma \tag{24}$$

where $a = (1-\sqrt{5})/2$, $b = (1+\sqrt{5})/2$, and $\gamma = (5+\sqrt{5})/10$. This choice of parameters guarantees that $E\varepsilon^* = 0$, $E\varepsilon^{*2} = \hat{\varepsilon}_i^2$ and $E\varepsilon^{*3} = \hat{\varepsilon}_i^3$ (Härdle, 1990a, p. 107).

To calculate new observations Y_i^* a second regression has to be run with a bandwidth g larger than h (for details see Härdle, 1990a, pp. 106–108).

$$Y_i^* = \hat{m}_g(X_i) + \varepsilon_i^* \tag{25}$$

According to Härdle (1990a, pp. 129–138) the estimate $\hat{m}_g(x)$ should be oversmoothed with bandwidth g proportional to $n^{-1/9}$ (in contrast to $n^{-1/5}$ for h). A new estimate $\hat{m}^*(x)$ is then obtained using the data $\{Y_i^*, X_i\}$. Theorem 4.2.2 in Härdle (1990a, p. 107) shows that the distribution of $\hat{m}_h^*(x) - \hat{m}_g(x)$ approximates the distribution of $\hat{m}_h(x) - m(x)$. The $\alpha/2$ and the $1-\alpha/2$ empirical quantiles of the replications of $\hat{m}_h^*(x)$ then represent the lower und upper pointwise confidence intervals. The results improve with the number of bootstraps.

Härdle and Marron (1991) use the wild bootstrap method for constructing simultaneous error bars. Härdle, Huet and Jolivet (1994) achieve a better coverage probability by explicit estimation of the bias term than by undersmoothing. In their approach the coverage accuracy is improved by wild bootstrap. However, the advantage of explicit estimation of the bias over undersmoothing might play a considerable role for huge data sets only (Härdle et al., 1994, p. 2).

Neumann (1994) uses the approach of wild bootstrap for the construction of confidence *bands*. His analysis is based on undersmoothing the local linear estimator proposed by Fan (1992).

2.4 Partially Linear Models

In some cases it might be important to incorporate parametric components into the econometric model and to decompose the specification into a parametric and a nonparametric part. This is, for instance, useful when dummy variables are to be included into the model which imply a parallel shift of an estimated curve, e.g. firms' innovative behavior might differ according to the industry they belong to. Thus, it is useful to include sectoral dummies in order to control for such effects.

A partially linear model can be written as follows:

$$Y_i = Z_i\beta + m(X_i) + \varepsilon_i. \tag{26}$$

The disturbance terms are assumed to be uncorrelated with mean zero and variance σ^2. There is a linear impact from Z and a nonlinear one from X. In the following, X is assumed to be one-dimensional while Z is one- or more-dimensional. Partially linear models belong to the group of semiparametric models, since there is one parametric and one nonparametric component.

In the following, the approach by Speckman (1988) is analysed more deeply. There are also other approaches for estimating a partially linear model, for instance the one proposed by Green, Jennison and Seheult (1985). Both methods are reviewed in the monograph by Green and Silverman (1994).

Z_i and X_i are supposed to be related in a nonlinear functional form

$$Z_i = \xi(X_i) + \eta_i \tag{27}$$

Green and Silverman (1994, p. 86) point out that it is not important if the relation between Z and X is plausible or not. Thus, both Y and Z can be expressed as a function of X. This suggests the use of partial residuals in order to estimate β. Using the smoother matrix (or weight matrix) $W_{hn}(x) = n^{-1}W_h(x)$ where $W_h(x)$ is defined by equation (6) we get the residuals $\varepsilon_Y = Y - W_{hn}(x)Y$ and $\varepsilon_Z = Z - W_{hn}(x)Z$. Regressing ε_Y onto ε_Z by OLS we obtain the parametric part of our semiparametric model:

$$\hat{\beta}_S = (\varepsilon_Z'\varepsilon_Z)^{-1}\varepsilon_Z'\varepsilon_Y. \tag{28}$$

With $Y_i - Z_i\beta = m(X_i) + \varepsilon$ we are then able to estimate the nonparametric term

$$\hat{m}_S = W_{hn}(x)(Y - Z\beta). \tag{29}$$

Using an appropriately chosen bandwidth and if several other regularity conditions hold, the Speckman-estimator is $\sqrt{N}$-consistent and asymptotically normal (Speckman, 1988, theorem 4).

In addition, it can be shown that (in case of symmetric smoother matrix W_{hn}) the variance of the estimator for β is $O(\frac{1}{n})$ which is the typical parametric rate (Speckman, 1988, theorems 1 and 2). $\hat{\beta}_S$ can have substantially smaller bias than the estimator proposed by Green et al. (1985) ($o(h^{2p})$ versus $O(h^p)$) in case of correlation between X and Z. Hence, if the bandwidth h is chosen such that the best possible rate of consistency of $\hat{g}_S$ is achieved, the squared bias of $\hat{\beta}_S$ is of a smaller order than its variance, and the MSE tends to zero at rate $O(\frac{1}{n})$ (Green and Silverman, 1994, p. 87). As Green and Silverman (1994, p. 88) point out, more practical experience is needed to draw conclusions on the performance of these estimators when applied to real data.

For Speckman's method, there is again the problem of bandwidth choice. An important result is that generalized cross-validation as an automatic bandwidth selector for kernel regression can also be used in the partially linear framework. The generalized cross-validation function turns out to be:

$$GCV(h) = \frac{n^{-1}\|(I - A(h))Y\|^2}{[1 - n^{-1}tr(A(h))]^2} \tag{30}$$

where the numerator denotes the sum of squared residuals. $A(h)$ denotes the matrix transforming Y into $\hat{Y}$, $\hat{Y} = A(h)Y$. For the Speckman estimator, it looks as follows:

$$A_S(h) = W_{hn} + P_Z(I - W_{hn}) \tag{31}$$

$$\tag{32}$$

where P_Z is the projection matrix

$$P_Z \quad = \quad \varepsilon_Z(\varepsilon_Z'\varepsilon_Z)^{-1}\varepsilon_Z'. \tag{33}$$

The asymptotic normality of the estimator makes the calculation of F-statistics possible (Speckman, 1988, p. 426). Therefore, the regression sum of squares (SS) has to be divided by the residual sum of squares (RSS) and adjusted by the corresponding degrees of freedom. From the estimate of $\hat{\beta}_S$ the fitted values of Y explained by the parametric component $Z\beta$ are calculated:

$$\hat{\varepsilon}_Y = \varepsilon_Z\hat{\beta}_S \tag{34}$$

Thus the corresponding regression sum of squares SS referring to the fit of the parametric part calculated in the first step is given by

$$SS = \hat{\varepsilon}_Y'\hat{\varepsilon}_Z. \tag{35}$$

The RSS[4] is

$$RSS = (Y - Z\hat{\beta}_S - \hat{m}_S(X))'(Y - Z\hat{\beta}_S - \hat{m}_S(X)). \tag{36}$$

In addition, the degrees of freedom have to be determined. Speckman (1988) suggests the following term called *equivalent degrees of freedom* by Green and Silverman (1994):

$$EDF^* = tr(I - A_S(h))'(I - A_S(h)). \tag{37}$$

According to Green and Silverman (1994, p. 37) the equivalent degrees of freedom can be approximated by

$$EDF = tr(I - A(h)) \tag{38}$$

which is much easier and faster to calculate. As Buja, Hastie and Tibshirani (1989) have shown,

$$(n - EDF^*) \approx 1.25(n - EDF) - 0.5. \tag{39}$$

The F-statistic is then obtained by $SS/(RSS/EDF)$. Accordingly, a possible estimator of the variance σ^2 is

$$\hat{\sigma}_h^2 = \frac{RSS}{EDF}. \tag{40}$$

[4] According to the theorem of Frisch-Waugh-Lovell the residual $(Y - Z\hat{\beta}_S - \hat{m}_S(X))$ is equal to the residual $\varepsilon_Y - \varepsilon_Z\hat{\beta}_S$, see for instance Davidson and MacKinnon (1993, p. 20).

In fact, Robinson (1988) develops the same estimator as Speckman (1988) but uses *trimming* in order to avoid erratic behavior due to the random denominator of the Nadaraya-Watson estimator. Trimming means that those observations with smallest density are cut off. In this case one has to optimize over the bandwidth as well as over the trimming parameter. Robinson (1988) also notes that it is impossible to include an intercept into the regression without further restrictions since by redefinition of $m(X_i)$ the intercept cannot be identified. If a constant term is included,

$$Y = (\alpha + Z\beta) + \{m(X) - \alpha\}. \tag{41}$$

For the same reason, Z must not be perfectly predictable by X.

3 The Data

For the empirical analysis two cross-sections are available for 1984 and for 1989 containing 1088 and 962, resp., firms of the West German manufacturing industry. The data have been collected by the Ifo-Institute, Munich. The questionnaires refer to business units[5] of the companies and not to the entire company which takes into account the different behavior of different product groups. Once a year the Ifo-business survey (Ifo-Konjunktur-Test, see in particular Nerlove, 1983, for the importance of the Ifo-business survey for economic research) refers to the firms' innovative activity; the firms are asked for the decomposition of sales according to the four product cycle phases innovation phase, growth phase, stagnation phase and shrinking phase. In this context the innovation phase is used as an indicator of innovative activity. Thus, considerations with respect to the German data only refer to product innovation, not to process innovation. Firm size is measured as number of persons employed in the business unit.[6]

The French data set contains 366 firms of the manufacturing industry averaged over the period 1984 to 1987. The data have been collected by the Ministry of Research and Technology ("Enquêtes Annuelles sur les Moyens Consacrés à la Recherche et au Developpement dans les Entreprises"), they refer, as the German data, to business units and are generally answered by all business units with at least

[5]In the following the expressions business unit and firm are used as synonyms.

[6]The condition for a firm to be in the sample is that the percentages of products in the four different product cycle phases sum up to 100.

one researcher working full-time. Here, the dependent variable is a proxy of R&D-capital (a weighted sum of R&D expenditures) is used which is calculated as follows: The initial R&D-capital is supposed to be five times the R&D-expenditures of the first year data are available (initial years are between 1980 and 1983). In addition, a constant depreciation rate of 15% per year is assumed. Firm size is measured by the employment average over the years 1984 to 1987, where employees in the domain R&D are excluded. For details on the French data, see Crepon and Mairesse (1993), as well as Hall and Mairesse (1995) for a theoretical justification of the proceeding to calculate the R&D-capital (the procedure is also discussed by Griliches, (1979))[7].

Finally, the fourth data set contains 277 firms of 14 sectors of the Belgian manufacturing industry. There are two data sources; the "Centrale des Bilans de la Banque Nationale de Belgique (BNB)" and the "Office National de Sécurité Sociale (ONSS)" (cf. Fecher and Perelman, 1989, for details on the data). The variables are averaged over the period 1977 to 1983. We included all firms for which information on R&D-expenditures (in prices of 1980) and employment is available for at least two of the seven years.

The variables of innovative input (R&D-capital and R&D-expenditures)are taken per employee (per head). All variables except those of innovative output (percentage share of sales due to products in the innovation phase) are taken in natural logarithms[8].

4 The Empirical Results

4.1 Results of the Nonparametric Approach

The following figures show the results of the Nadaraya-Watson estimation. Figures 1 to 4 refer to the German data set with the innovation phase as an indicator of innovative activity. Figures 5 and 6 are based on the French data, and Figures 7 and 8, finally, on the Belgian data set.

Figures 1, 3, 5 and 7 show the nonparametric regression function calculated with

[7]the data were originally prepared to estimate production functions and to measure the productivity of R&D.

[8]Besides employment, sales might also be considered as adequate measure of firm size, in particular for firms working with capital-intensive production-techniques. However, such a variable is not available in these data sets.

the bandwidth obtained by generalized cross-validation together with the scatter plot of observations. The density of the explanatory variable ln(employment) calculated with the bandwidth chosen according to the so-called *rule of thumb* using the quartic kernel (for details see Silverman, 1986, pp. 40–45). In Figures 2, 4, 6 and 8 the regression curve is presented together with the bootstrapped 90% pointwise confidence intervals where the bandwidth g is calculated such that the relation between g and h is like $n^{-1/9}$ to $n^{-1/5}$ (see section 2.3). In addition, for each data set three OLS-regressions were run, one with firm size (ln(employment)) as independent variable, one with firm size and squared firm size, and one with firm size, squared firm size and cubic firm size as explanatory variables. The specification which gives the best results with respect to the F-test is depicted in the figures. All estimation results are presented in Table 1, where bold letters indicate the estimations for the OLS-fits presented in the graphics. The bandwidths chosen for the kernel estimations are given in Table 2. The OLS-curve is depicted with 90% confidence intervals of the nonparametric regression model in order to test if the two models deviate significantly from each other. Finally, the figures show the OLS-curve and the nonparametric regression curve when 5% of observations with smallest density are cut off.

The results are contradictory for the German data sets of 1984 and 1989. The estimated curve for 1984 (Figure 2a) shows a U-shape whereas the data of 1989 reveal a hump-shaped relationship (Figure 4a). Note, however, that the OLS-estimation for 1984 indicates insignificant results. The OLS-estimation for 1989 shows significant t-values but a comparison of the nonparametric fit and the OLS-fit reveals steeper slopes in case of OLS such that it deviates from the 90% confidence interval of the Nadaraya-Watson estimate.

For the French data set a variable of innovative input, R&D-capital, is analysed (Figures 5, 6). The relationship betweeen ln(R&D-capital p.h.) and ln(employment) is U-shaped, i.e. R&D-capital per employee is larger for small and large firms than for those of intermediate size. The range of the fitted values varies between 3.2 and 4.5, implying a range between 26.7 and 92.9 thousand French francs per employee (measured in prices of 1980). The nonparametric regression curve is close to the OLS-curve in the middle part, but it deviates at the two ends (Figure 6b).

Figures 7 and 8, based on the Belgian data, seem to confirm the results obtained by the French data set and the German data set of 1984, at least at first glance. The estimated density, Figure 7b, is much more skewed to the right than in case of the other data sets. From Figures 8a and 8b it can be seen that ln(R&D-expenditures per head) are first decreasing, then increasing with firm size. According to the estimates small and large firms deviate quite considerably from medium-sized firms: The estimated range of R&D-capital per head varies between 53.4 and 565.9 thousand Belgian Francs per head. The Nadaraya-Watson estimate turns out to be quite wiggly in the middle part such that for the purpose of interpretation one might feel inclined to take a bandwidth larger than the one obtained by cross-validation. The curves of the confidence intervals deviate from the regression curve slightly more at the right end than in the left and in the middle part of the curves.

The OLS-estimates seem to differ considerably from the Nadaraya-Watson estimates when they are plotted together with the pointwise confidence intervals. Only for Germany 1984 the OLS-estimate fits quite well into the confidence intervals of the nonparametric regression curve, but it crosses the upper interval curve at the right end.

Since it is hard to draw reliable conclusions where estimation is based on a sparse number of observations, in a further step, 5% of observations with smallest density are cut off (according to the densities estimated with the bandwidth obtained by the rule of thumb). Note, that cutting off only makes sense *after* estimating the density or the regression function (otherwise you are left with new outliers at the boundaries and you have to restart).

In the case of the German data of 1984 (Figure 2d) the downward movements on the lower and the upper bounds of the nonparametric regression curve disappear whereas for 1989 they are less profound (Figure 4d). The regression curves for the French and the Belgian data (Figures 6d and 8d) are affected by cutting off observations with large employment numbers. For Belgium, the nonparametric curve estimate now has a rather decreasing tendency. The deviation in R&D-capital and R&D-expenditure per employee reduces to 26.7 to 59.7 thousand French Francs and to 55.10 to 276.6 thousand Belgian Francs, respectively.

To summarize the results: they differ with respect to the two cross-sections of the German data; for 1984, innovative activity is first decreasing and then increasing with firm size, while for 1989 the corresponding relationship is inversed.

This result shows that innovative activity does not only depend on firm size but is also influenced by other factors such as industry conditions, market structure, institutional framework, etc. (Cohen and Levin, 1989, pp. 1070–74). Since the analysis of the French data set gives qualitatively the same results as that for the German data in 1984, one might alternatively suspect that the result for 1989 is rather an exception.

In the case of the Belgian data conclusions have to be drawn more cautiously. Without a cut off of outliers the estimations strongly confirm the results obtained for the German cross-section of 1984 and those obtained with the French data set. However, taking into account regions with sparse density, R&D-expenditures per head are decreasing with firm size. Here, OLS would give a too stylized picture of the U-shaped relationship (Figure 8d).

4.2 Results of the Semiparametric Approach

For reasons of space the application of Speckman's estimator is demonstrated only for the French data set. Two kinds of variables are considered for the parametric component of the model, the value added per employee and three sector dummies referring to the sectors of intermediate goods, G, investment goods, I, and consumer goods, C. Including value added per head, VA, aims at controlling the R&D – firm size relationship for potential productivity differentials among firms. More productive firms are expected to be more innovative than less productive firms. As before, the nonparametric part is the link between innovative activity and firm size.

Different specifications are estimated by the Speckman approach (Table 3) as well as by OLS (Table 4) where the latter includes the simple and the quadratic term of ln(employment) since this gives best results with respect to the firm size. In addition, the OLS-estimations generally include a constant term. For the partially linear models the significance of the estimated coefficients is measured by F-tests as proposed by Speckman (see section 2.4).

In fact, for specifications 1 and 2, the OLS results do not differ much from those of the Speckman approach with respect to the size and the significance of the estimated coefficients as well as with respect to the residual sum of squares. Including all three sector dummies, however, leads to complications in the partially linear model. In case of specifications 4 and 5 the cross-validation function cannot be calculated for the whole grid of bandwidths since the matrix $\varepsilon_Z' \varepsilon_Z$ sometimes turns out to be singular [9]. The reason for this behavior is that three dummies summing up to a constant have the same effect as including a constant term which leads, as already mentioned, to identification problems. Using, instead, the bandwidths obtained by cross-validation for specifications 2 and 3 leads to the results given in Table 3. Specification 5 is quite similar to the OLS-specification 4 with the coefficients of the dummies slightly smaller in absolute value.

Some results of estimating the partially linear model according to specification 2 are depicted in Figure 9. The decomposition of the estimate in its parametric and nonparametric component is shown together with the fit including both components and the residual sum of squares in Figure 9a. When the parametric fit is added to the nonparametric fit it becomes U-shaped (note the large dimension on the vertical axis when compared to Figure 6 and Figure 9b). However, the depiction of the residual sum of squares implies that there still remains a considerable part of the variance unexplained, a result which is, however, not surprising when using cross-sectional data.

Figure 9b includes the regression curves obtained by the purely nonparametric model together with the OLS-curve and the estimate of the semiparametric model with specification 2. In order to make the structure of the curves comparable the estimate of the nonparametric component of the partially linear model is parallely shifted by the mean of the parametric component. Accordingly, the OLS-estimate of the simple and the quadratic term of the firm size variable is shifted by the mean of the remaining term of the regression (otherwise the curves would lie too far apart from each other and the structure would be invisible).

In fact, the curve of the partially linear model turns out to be a little flatter than the curve of the purely nonparametric model which is partly due to the fact that a larger bandwidth is chosen in case of the semiparametric model. The OLS-estimate is close to the Speckman-estimate in the middle part but deviating more at the ends. However, the interpretation that small and large firms are more innovative than middle-sized firms can also be maintained when taking into account additional variables which are supposed to have a linear impact on innovative activity.

[9]Including only the three sector dummies and dropping the value added per head leads to a cross-validation function with several local minima and with sharp declines and increases.

5 Conclusions

The purpose of this analysis is to study the feasibility of nonparametric regression techniques in the context of a problem related to industrial economics. The Nadaraya-Watson estimator, based on kernel density estimation, is applied in order to analyse the relationship between innovation and firm size. This approach is well settled in biometry but still less established in economics. Particular attention is paid to the bandwidth choice, a sensitive point for the hesitance of economists to apply nonparametric methods.

Regression fits of parametric OLS estimation lead to similar general conclusions, but are sometimes deviating at the very ends of the range of the explaning variable, i.e. for very small and very large firms. Since these regions are the most important ones for the validity of Schumpeter's hypothesis, they are analysed more deeply. Here, bootstrapped confidence intervals and trimming outliers enables us to draw more reliable conclusions than with conventional methods.

The analysis of the Schumpeterian link between innovation and firm size confirms previous findings by Acs and Audretsch (1990), Bound et al. (1984), Cremer and Sirbu (1978), and Entorf (1988) that small and large firms are more innovative than firms of intermediate size in the case of the German data set of 1984 as well as for the French data set. However, the estimations show a rather hump-shaped relationship for the German data of 1989 and a decreasing one for the Belgian data. These results reflect the fact that innovative activity does not only depend on firm size but also on cyclical effects, market structure, etc.

Further investigating the French R&D peformance by including additional variables into the regression model which are supposed to have linear impacts on the firms' innovative activity and by applying Speckman's (1988) semiparametric approach based on partial residuals contributes significantly to the explanation of the variance in ln(R&D-capital per head). Nevertheless, the partial relationship between ln(R&D-capital p.h.) and ln(employment) remains U-shaped, implying that small and large firms are more innovative than middle-sized firms.

Given the experience of our empirical study, is it really worth doing nonparametric regression in economic research when it takes more computation time and is more difficult to interpret? We would respond to this question affirmatively.

Parametric methods are only one way of investigating economic relationships. Nonparametric methods enforce empirical researchers to have a closer look at their data. The (unsolved) problem, however, remains that the bandwidth choice is unpopular with economists, who prefer to estimate structural parameters derived from first economic principals. However, it should be noted, that bandwidth choice is not artificial but based on *statistical* theory (for a detailed discussion, see Härdle, 1990a) and the method of cross-validation as an automatic bandwidth selector is already quite well-established and practicable.

Moreover, parametric regressions are also subject to many sensitive assumptions, in particular when they are used for explorative purposes. For instance, exploring the relationship between firm size and innovative activity using OLS suggests a particular functional form with additive components.

An important advantage of the Nadaraya-Watson estimator over OLS is that it takes into account the density of the explanatory variable in the sense that observations in areas of sparse data have smaller weights than those of high density. However, the bandwidth is taken to be fixed over the whole range of observations. An improvement could be made by letting the bandwidth vary with the density, i.e. using smaller bandwidths in areas of high density and larger bandwidths in regions whith thinly scattered data (cf. Silverman, 1986, pp. 21–23). A cut off of observations with smallest density attempts to compensate for this shortcoming since it concentrates on those observations with high densities.

Partially linear models attempt to combine the extreme types of modelling a relationship, i.e. either purely parametrically or purely nonparametrically. The estimator proposed by Speckman (1988) applying OLS to the partial residuals is easily implemented, and with generalized cross-validation as a method of automatic bandwidth slection it gives reasonable results. F-tests can be calculated as in the case of OLS.

Table 1: Results of the OLS-estimations*)

explan. variab.	innov.phase (%) Germany 1984	innov. phase (%) Germany 1989	ln(R&D-capital p.h.), France	ln(R&D-expend. p.h.), Belgium
	dependent variables			
const.	0.14	0.12	4.04	5.40
	(11.74)	(10.49)	(10.03)	(11.52)
ln(emp)	-0.001	0.001	-0.07	-0.15
	(-0.48)	(0.53)	(-1.29)	(-1.67)
R^2	0.00	0.00	0.01	0.01
F-test	0.23	0.28	1.66	2.77
significance	(0.63)	(0.59)	(0.19)	(0.20)
const.	0.15	**0.07**	**9.15**	**9.68**
	(5.20)	**(2.11)**	**(7.12)**	**(6.93)**
ln(emp)	-0.01	**0.02**	**-1.54**	**-1.82**
	(-1.02)	**(2.06)**	**(-4.33)**	**(-3.48)**
$(\ln(emp))^2$	0.001	**-0.002**	**0.10**	**0.15**
	(0.95)	**(-1.99)**	**(4.18)**	**(3.24)**
R^2	0.001	0.004	0.05	0.05
F-test	0.57	2.13	9.59	6.7
significance	(0.57)	(0.12)	(0.00)	(0.001)
const.	**0.10**	-0.001	5.61	8.50
	(1.43)	(-0.02)	(1.73)	(2.03)
ln(emp)	**0.03**	0.06	0.10	-1.12
	(0.71)	(1.69)	(0.07)	(-0.46)
$(\ln(emp))^2$	**-0.01**	-0.01	-0.14	0.02
	(-0.91)	(-1.41)	(-0.68)	(0.05)
$(\ln(emp))^3$	**0.0004**	0.0004	0.01	0.01
	(1.05)	(1.13)	(1.18)	(0.30)
R^2	0.002	0.01	0.05	0.05
F-test	0.75	1.85	6.87	4.48
significance	(0.52)	(0.14)	(0.00)	(0.004)
# obs.	1088	962	366	277

*)t-values in brackets. The bold letters indicate the estimations chosen for the OLS-fits in the graphics.

Table 2: Results of bandwidth choice for density estimation and regression curve estimation*)

data set	h for density estimation	h for regression curve estimation	h for oversmooth in confid. intervals
German data '84	0.89	2.24	4.17
German data '89	0.96	3.33	6.14
French data	0.68	1.17	1.98
Belgian data	1.08	0.86	1.41

*)The bandwidth for density estimation is chosen by the "rule of thumb"; the bandwidth for regression curve estimation is based on cross-validation where a grid of 20 bandwidths is given.

Table 3: The parametric part of the semiparametric Speckman-approach*)

	dependent variable: ln(R&D-capital p.h.)				
expl. vars.	1	2	3	4**)	5**)
G	–	-0.01	-0.24	-0.32	-0.40
		(0.01)	(3.46)	(7.93)	(13.08)
		(0.91)	(0.06)	(0.005)	(0.0)
I	–	0.92	0.35	0.53	0.42
		(67.88)	(7.69)	(22.02)	(14.18)
		(0.0)	(0.01)	(0.0)	(0.0)
C	–	–	–	-0.03	-0.39
				(0.09)	(12.17)
				(0.77)	(0.0)
ln(VA)	1.09	1.51	–	1.49	1.51
	(57.06)	(126.2)		(119.6)	(126.0)
	(0.0)	(0.0)		(0.0)	(0.0)
h	1.529	1.529	1.17	1.17	1.529
RSS	396.98	340.63	435.32	346.39	341.52
F-test‡)	57.06	41.80	8.16	30.20	29.58
EDF	360.18	358.19	357.64	357.06	358.60

*) The numbers in the first brackets below the estimated coefficents give a single F-test as proposed by Speckman; in the second brackets the corresponding significance levels are given.
**) The bandwidth is given as resulting from generalized cross-validation for specifications 1 and 2. ‡) All F-tests imply significance.

Table 4: Results of the OLS-estimations of the partially linear model*)

	dependent variable: ln(R&D-capital p.h.)			
expl. vars.	1	2	3	4
constant	2.56	-0.45	8.82	–
	(1.75)	(-0.31)	(6.97)	
ln(empl)	-1.34	-1.10	-1.43	-1.09
	(-4.05)	(-3.54)	(-4.07)	(-3.54)
ln(empl)^2	0.09	0.07	0.09	0.07
	(-4.06)	(3.31)	(3.82)	(3.31)
G	–	-0.02	-0.27	-0.47
		(-0.19)	(-1.85)	(-0.33)
I	–	0.90	0.30	0.45
		(6.42)	(2.05)	(0.33)
C	–	–	–	-0.45
				(-0.31)
ln(VA)	1.11	1.52	–	1.52
	(7.73)	(10.37)		(10.37)
RSS	398.76	343.57	446.19	343.57
R^2	0.19	0.30	0.09	0.30
F-test†)	27.37	30.52	8.70	25.43

*)t-values in brackets. †) All F-tests imply significance.

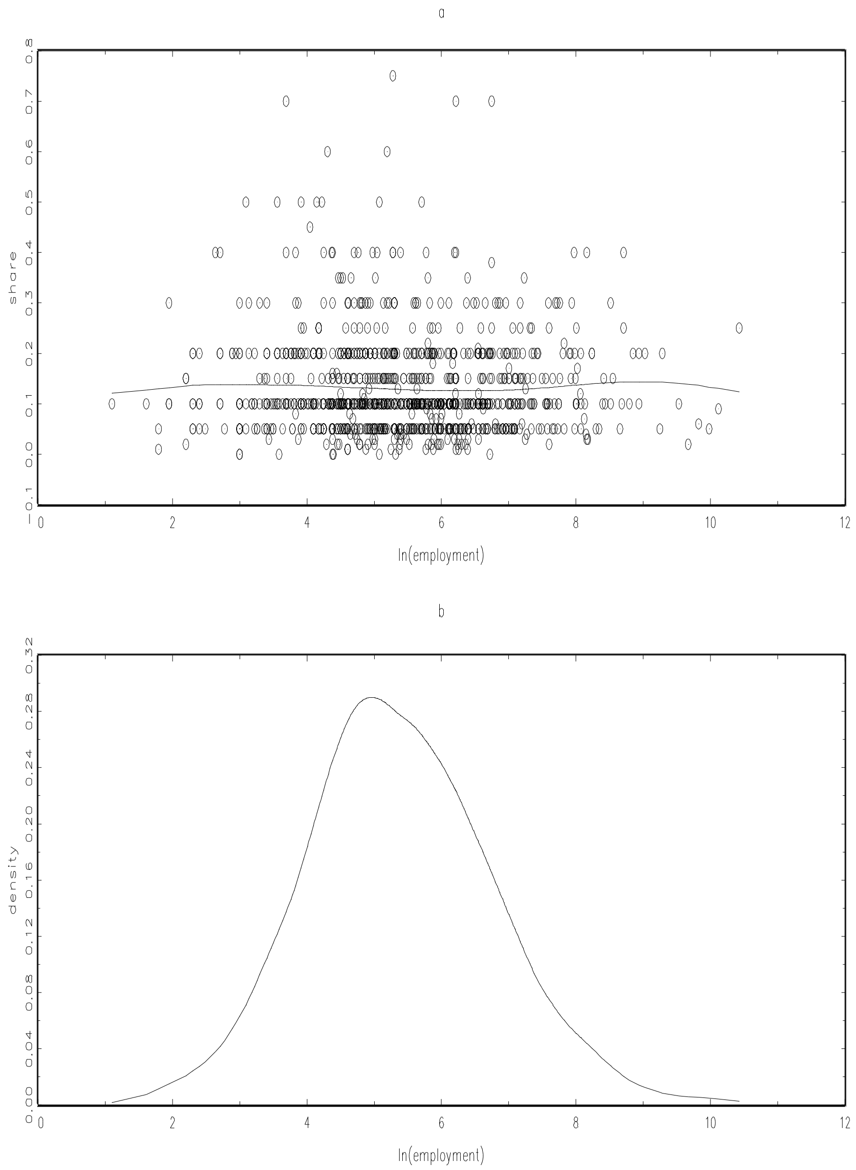


Figure 1: Firm size and percentage share of sales due to products in the innovation phase, Germany 1984. a) scatter plot with Nadaraya-Watson estimate; b) density of ln(employment).

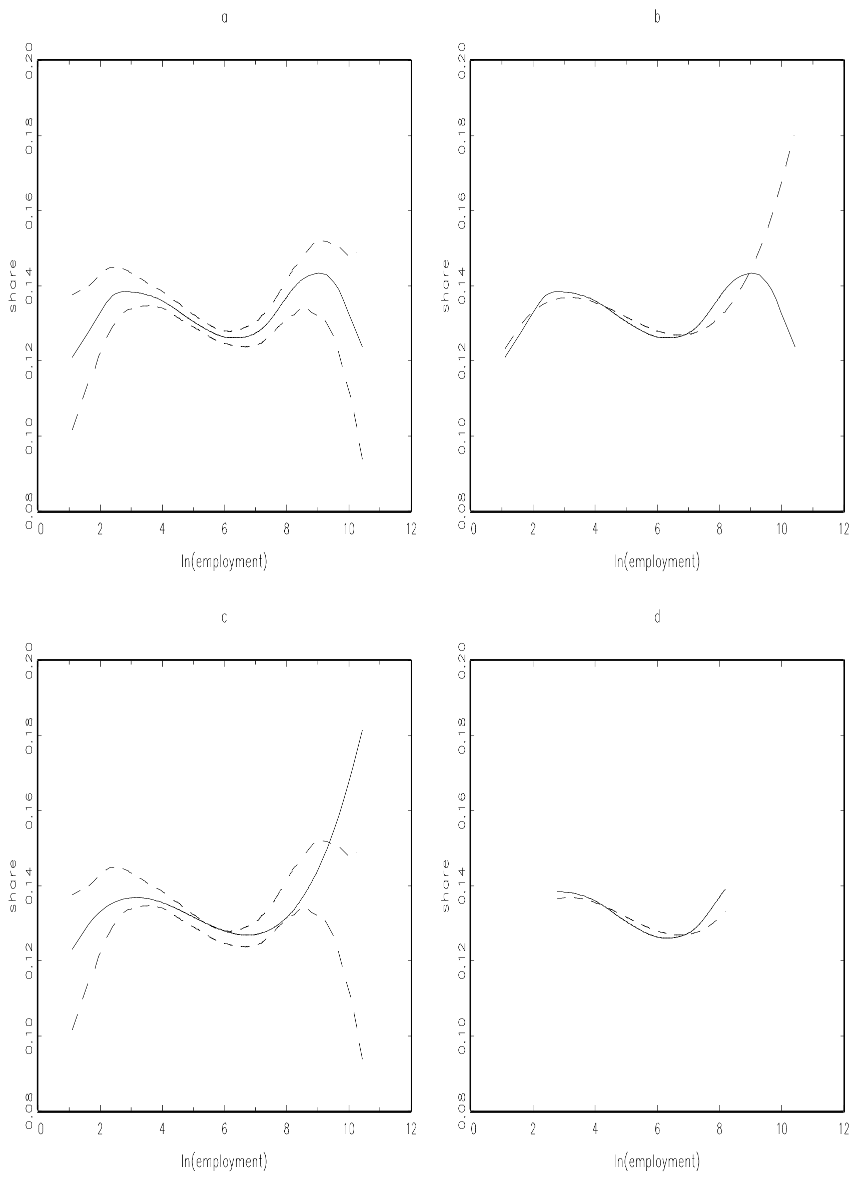


Figure 2: Firm size and percentage share of sales due to products in the innovation phase, Germany 1984. a) curve estimate and 90% pointwise confidence intervals; b) nonparametric curve estimate (solid), OLS-estimate (dashed); c) OLS-estimate and 90% pointwise confidence intervals of Nadaraya-Watson estimate; d) Nadaraya-Watson estimate (solid) and OLS-estimate (dashed) when 5% of observations are trimmed.

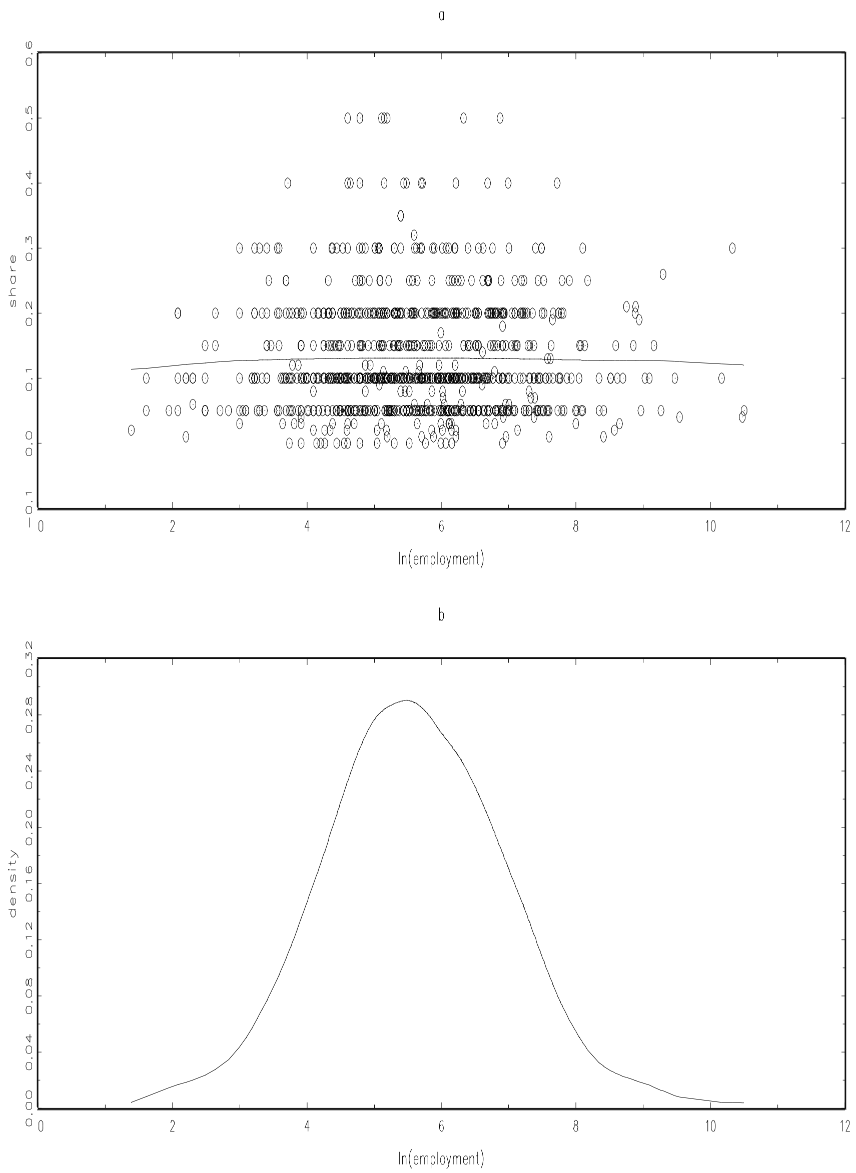


Figure 3: Firm size and percentage share of sales due to products in the innovation phase, Germany 1989. a) density of ln(employment); b) scatter plot with Nadaraya-Watson estimate.

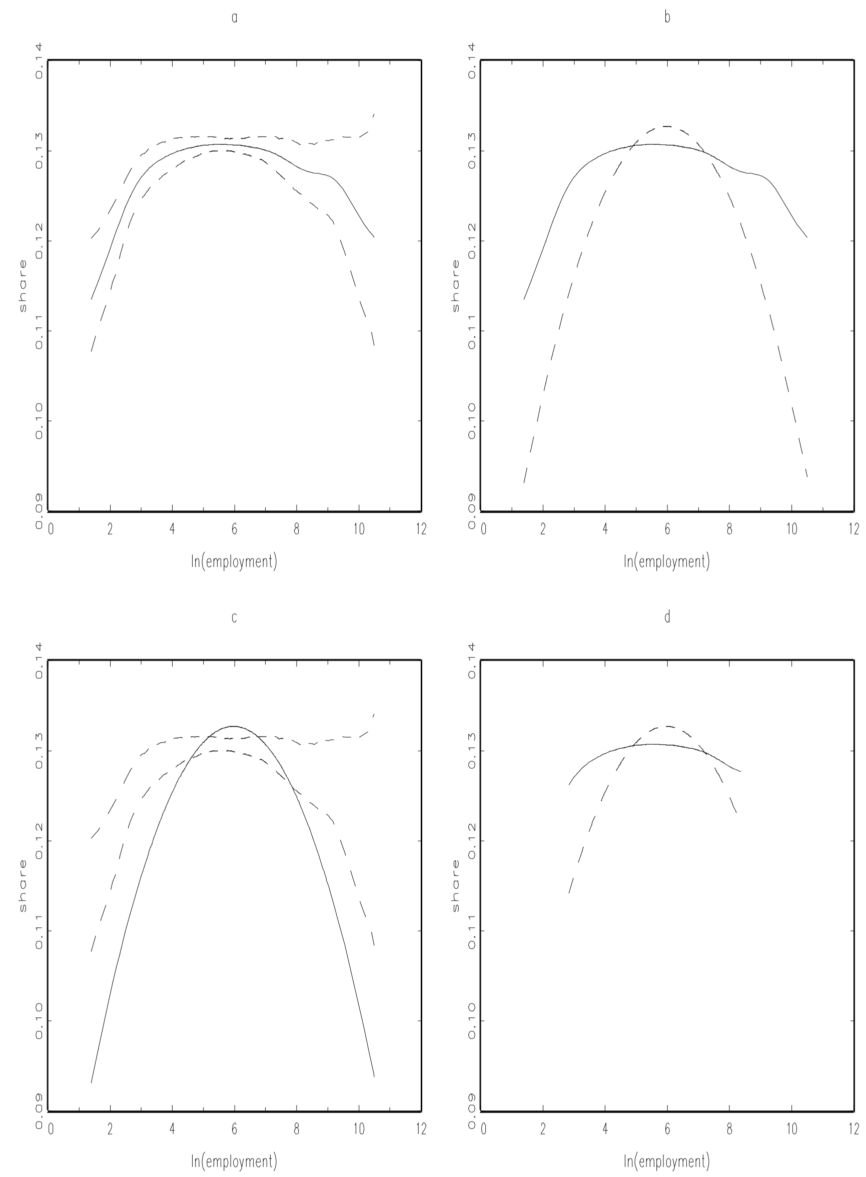


Figure 4: Firm size and percentage share of sales due to products in the innovation phase, Germany 1989. a) curve estimate and 90% pointwise confidence intervals; b) nonparametric curve estimate (solid), OLS-estimate (dashed); c) OLS-estimate and 90% pointwise confidence intervals of Nadaraya-Watson estimate; d) Nadaraya-Watson estimate (solid) and OLS-estimate (dashed) when 5% of observations are trimmed.

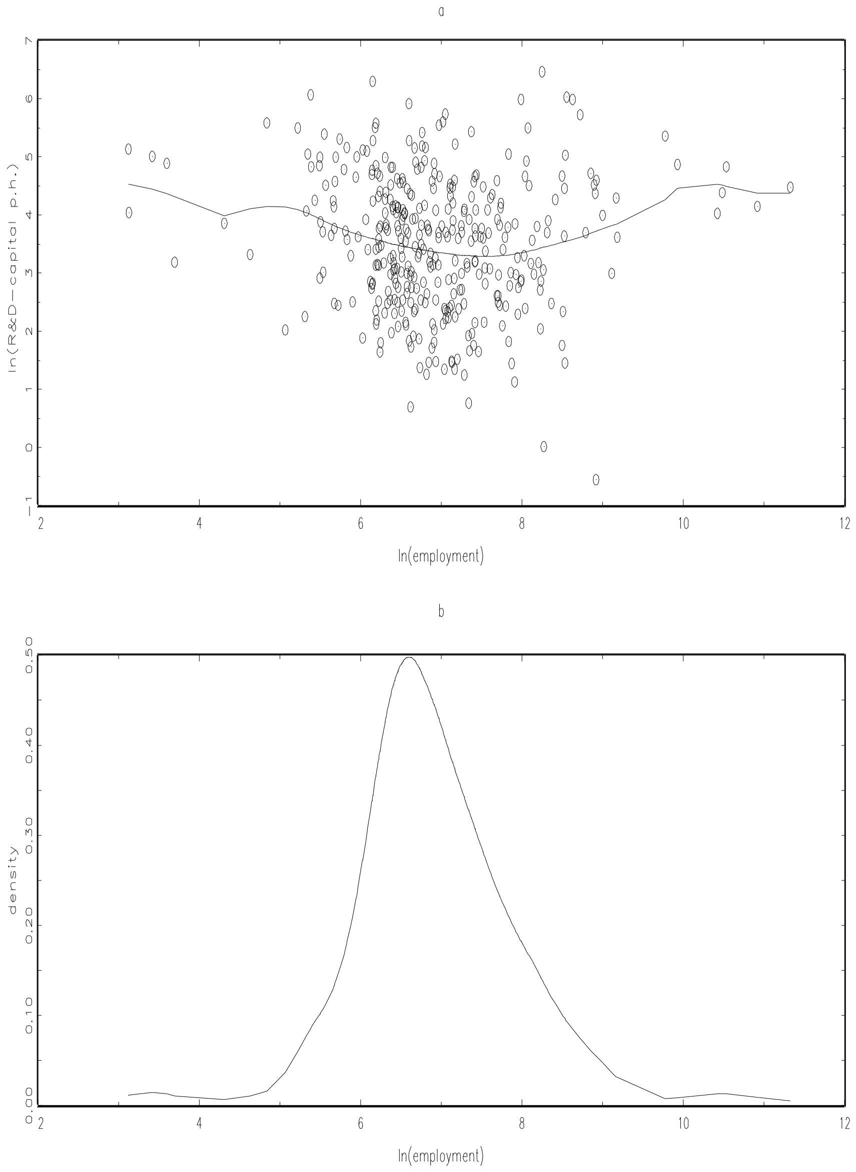


Figure 5: Firm size and R&D-capital per head, France, 1984–1987. a) density of ln(employment); b) scatter plot with Nadaraya-Watson estimate.

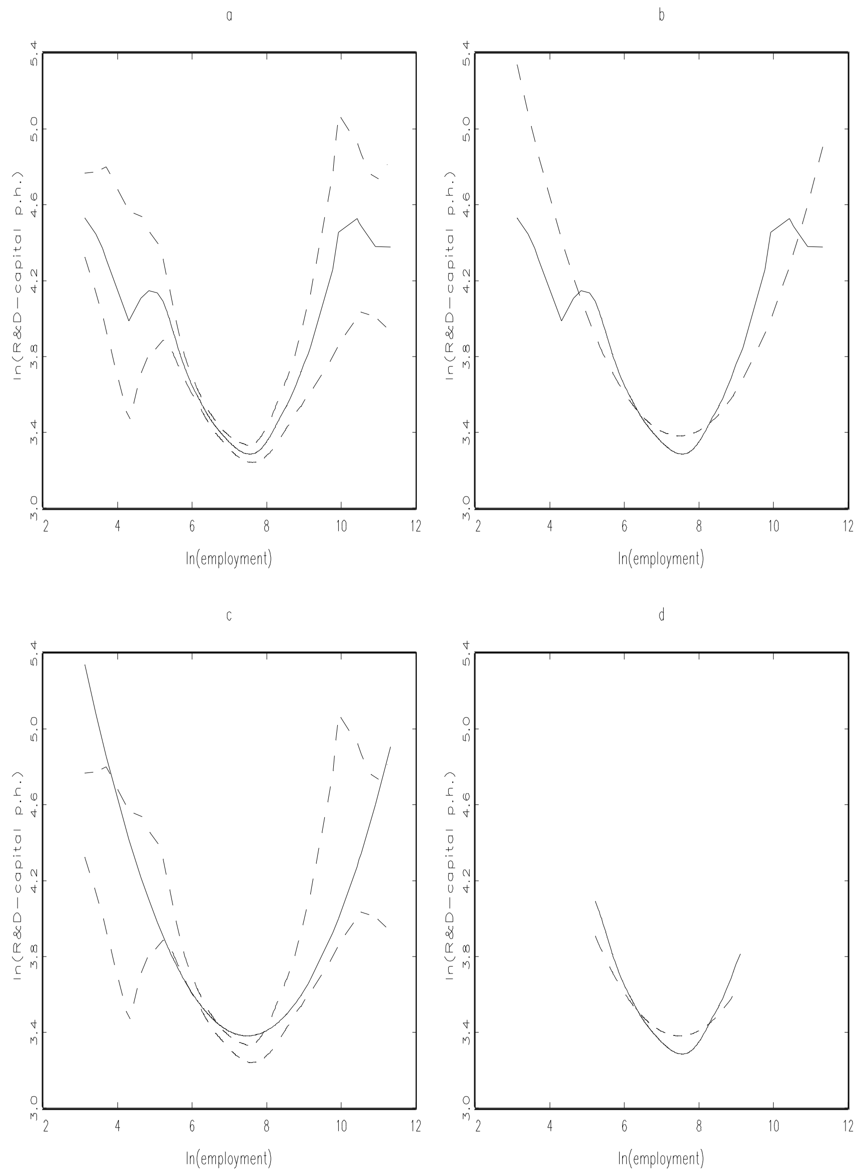
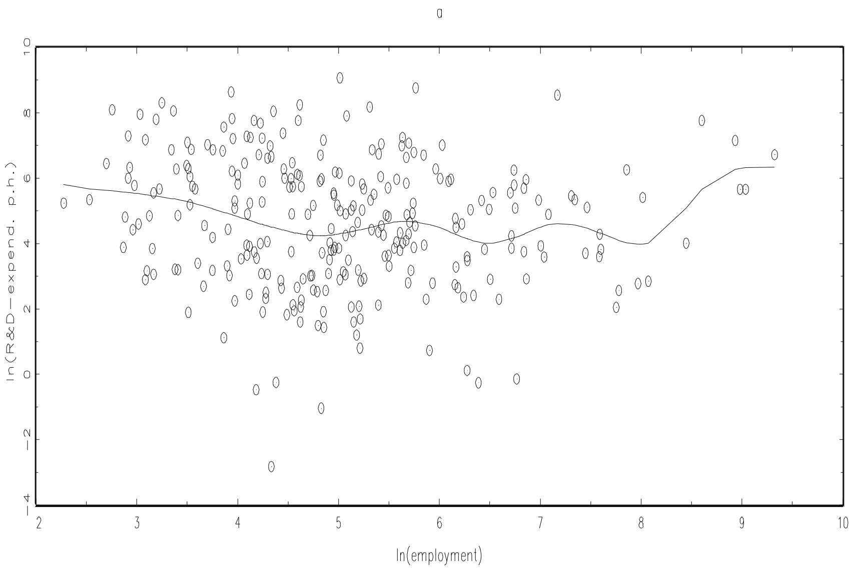


Figure 6: Firm size and R&D-capital per head, France, 1984–1987. a) curve estimate and 90% pointwise confidence intervals; b) nonparametric curve estimate (solid), OLS-estimate (dashed); c) OLS-estimate and 90% pointwise confidence intervals of Nadaraya-Watson estimate; d) Nadaraya-Watson estimate (solid) and OLS-estimate (dashed) when 5% of observations are trimmed.



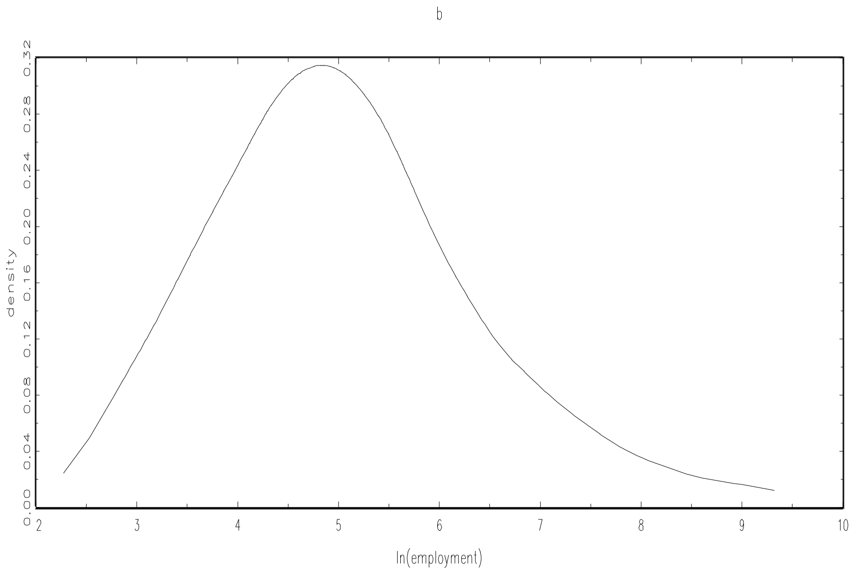


Figure 7: Firm size and R&D-expenditure per head, Belgium,1977–1983. a) density of ln(employment); b) scatter plot with Nadaraya-Watson estimate.

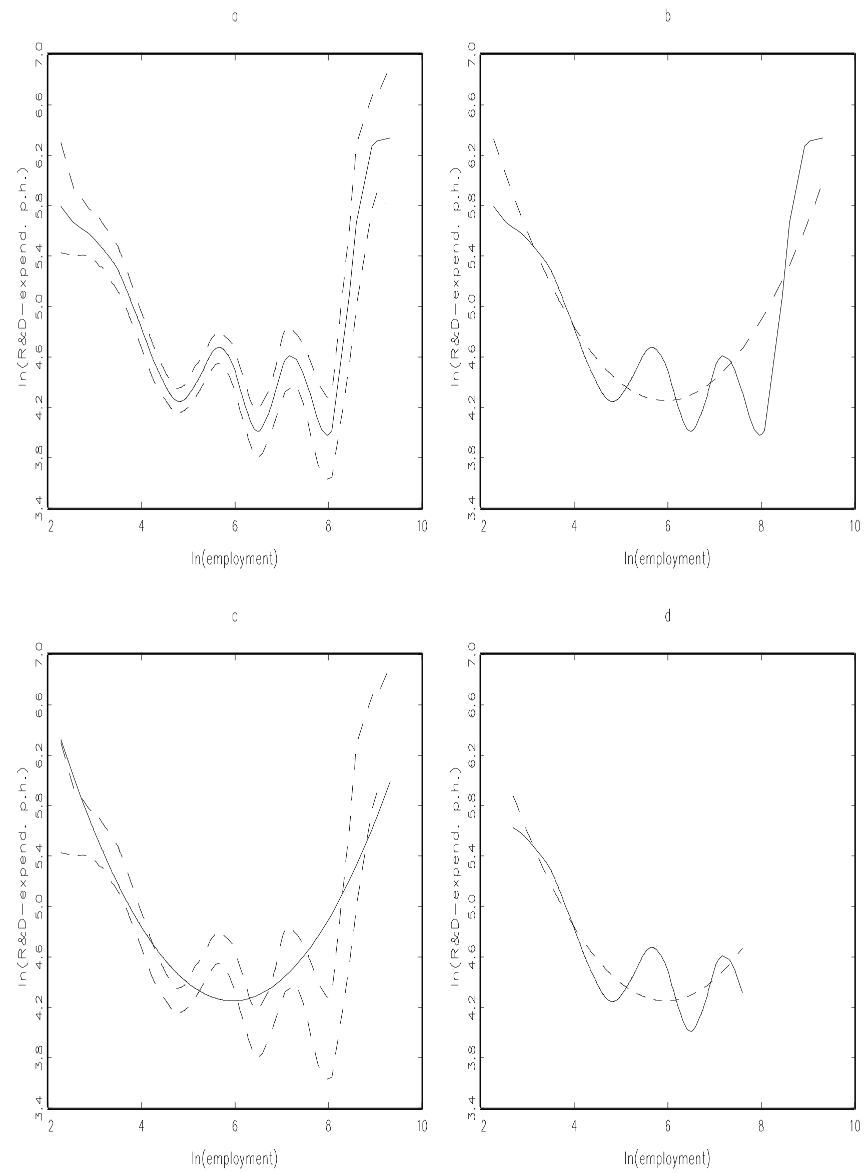


Figure 8: Firm size and R&D-expenditure per head, Belgium, 1977–1983. a) curve estimate and 90% pointwise confidence intervals; b) nonparametric curve estimate (solid), OLS-estimate (dashed); c) OLS-estimate and 90% pointwise confidence intervals of Nadaraya-Watson estimate; d) Nadaraya-Watson estimate (solid) and OLS-estimate (dashed) when 5% of observations are trimmed.

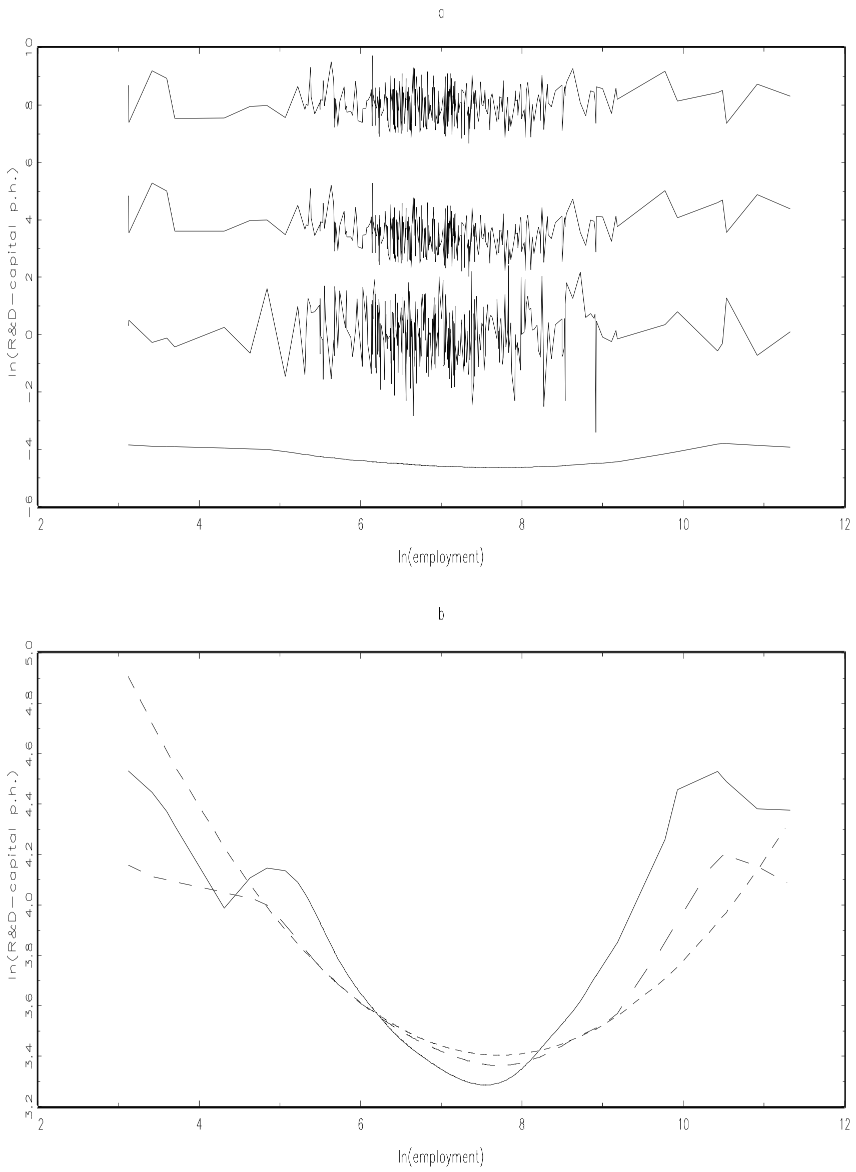


Figure 9: Firm size and R&D-capital per head, France, 1984–1987. a) from top to bottom: fit of the parametric component, fit of the parametric and nonparametric component, residual sum of squares, nonparametric component; b) estimates of the nonparametric component: purely nonparametric model (solid), semiparametric model by Speckman (long dashes) and OLS (short dashes), the last two are shifted by the mean of remaining component.

References

Acs, Z. J. and Audretsch, D. B. (1990). *Innovation and Technological Change, An International Comparison*, Harvester Wheatsheaf, New York.

Bound, J., Cummins, C., Griliches, Z., Hall, B. H. and Jaffe, A. (1984). Who does R&D and who patents?, *in* Z. Griliches (ed.), *R&D, Patents, and Productivity*, University of Chicago, Press for the National Bureau of Economic Research, Chicago.

Buja, A., Hastie, T. J. and Tibshirani, R. J. (1989). Linear smoothers and additive models (with discussion), *Annals of Statistics* **17**(2): 453–555.

Cohen, W. M. and Levin, R. C. (1989). Empirical studies of innovation and market structure, *in* R. Schmalensee and R. D. Willig (eds), *Handbook of Industrial Organization*, Vol. II, Elsevier Science Publishers B.V, New York.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions, *Numerische Mathematik* **31**: 377–403.

Cremer, J. and Sirbu, M. (1978). Une analyse econométrique de l' effort de recherche et developpement de l'industrie française, *Revue Economique* **29**: 940–954.

Crepon, B. and Mairesse, J. (1993). Productivité, recherche-developpement, et qualifications, paper presented at the tenth Meeting of Applied Microeconomics, 4th to 5th June 1993.

Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

Entorf, H. (1988). Die endogene Innovation: Eine mikro-empirische Analyse von Produktphasen als Innovationsindikatoren (endogenous innovations: A micro-empirical analysis of product cycle phases as indicators of innovative activity), *Jahrbücher für Nationalökonomie und Statistik* **204**(2): 175–189.

Fan, J. (1992). Design-adaptive nonparametric regression, *Journal of the American Statistical Association* **87**: 998–1004.

Fecher, F. and Perelman, S. (1989). Productivité, progrès technique et efficacitè: une étude comparative de 14 secteurs industriels belges, *Annales d'Economie et de Statistique* **3**: 93–118.

Gasser, T. and Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method, *Scandinavian Journal of Statistics* **11**: 171–185.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*, Vol. 58 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, London.

Green, P., Jennison, C. and Seheult, A. (1985). Analysis of field experiments by least squares smoothing, *Journal of the Royal Statistical Society* **47**(2): 299–315. Series B.

Griliches, Z. (1979). Issues in assessing the contribution of R&D to productivity growth, *The Bell Journal of Economics* **10**: 92–116.

Hall, B. H. and Mairesse, J. (1995). Exploring the relationship between R&D and productivity in french manufacturing firms, *Journal of Econometrics* **65**(2): 263–293.

Härdle, W. (1990a). *Applied Nonparametric Regression*, number 19 in *Econometrics Society Monographs*, Cambridge University Press.

Härdle, W. (1990b). *Smoothing Techniques with Implementation in S*, Springer Series in Statistics, Springer, New York.

Härdle, W. and Bowman, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands, *Journal of the American Statistical Association* **38**(401): 102–110. Theory and Methods.

Härdle, W., Huet, S. and Jolivet, E. (1994). Better bootstrap confidence intervals for regression curve estimation. SFB-discussion paper, No. 27, Humboldt-University of Berlin.

Härdle, W. and Mammen, E. (1991). Bootstrap methods in nonparametric regression, *in* G. Roussas (ed.), *Nonparametric Functional Estimation and Related Topics*, Vol. 335, Kluwer Academic Publishers, Netherlands, pp. 111–124.

Härdle, W. and Marron, S. (1991). Bootstrap simultaneous error bars for nonparametric regression, *The Annals of Statistics* **19**(2): 778–796.

Härdle, W. and Scott, D. W. (1992). Smoothing by weighted averaging of rounded points, *Computational Statistics* **7**: 97–128.

Link, A. N. (1980). Firm size and efficient entrepreneurial activity: A reformulation of the schumpeter hypothesis, *Journal of Political Economy* **88**: 771–782.

Nadaraya, E. A. (1964). On estimating regression, *Theory of Probability and its Applications* **10**: 186–190.

Nerlove, M. (1983). Expectations, plans, and realizations in theory and practice, *Econometrica* **51**: 1251–1279.

Neumann, M. H. (1994). Bootstrap confidence bands in nonparametric regression, unpublished manuscript.

Parzen, E. (1962). On estimation of a probability density function and mode, *Annals of Mathematical Statistics* **33**: 1065–1076.

Robinson, P. (1988). Root-n-consistent semiparametric regression, *Econometrica* **56**(4): 931–954.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27**: 832–837.

Scherer, F. M. (1965a). Firm size, market structure, opportunity, and the output of patented inventions, *American Economic Review* **55**: 1097–1125.

Scherer, F. M. (1965b). Size of firm, oligopoly, and research: A comment, *Canadian Journal of Economics and Political Science* **31**: 256–266.

Schumpeter, J. A. (1942). *Capitalism, Socialism, and Democracy*, Harper & Row, New York.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Vol. 26 of *Monographs on Statistics and Applied Probability*, Chapman&Hall, London.

Speckman, P. (1988). Kernel smoothing in partial linear models, *Journal of the Royal Statistical Society* **50**(3): 413–436.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression, *Annals of Statistics* **10**: 1040–53.

Watson, G. S. (1964). Smooth regression analysis, *Sankhyā* **26**: 359–372.