EXECUTIVE PAY AND CORPORATE FINANCIAL PERFORMANCE.
AN EXPLORATIVE DATA ANALYSIS
Ulrike Graßhoff, Joachim Schwalbach and Stefan Sperlich*

Abstract
The relationship between executive pay and corporate financial performance continues to attract wide academic, media and policy attention. The very high salaries enjoyed by senior executives in corporations in the US are often contrasted with the relatively low pay received by executives in Europe and Asia. Empirical research on executive pay has mainly concentrated on the pay-performance relationship. Although the adopted data sets were very different within and across countries, the results are very similar and show very low pay-for-performance elasticities. Despite the similar results, several methodological issues are still uncovered. Almost all studies assume linear or semi-log linear pay functions without applying a test of the adequate functional form. Most models do not allow for variations across corporations, industries, countries and time. It is assumed that pay functions are homogeneous across corporations, variations are captured by the fixed effects in the constants and assumption about the errors. The purpose of the paper is to circumvent these possible misspecifications by adopting an explorative data analysis using nonparametric methods which impose rather weak restrictions on the model. We start with the most general model but use methods that allow for a stepwise closer look by specifying the various objectives of investigation or the model we deduce from the previous results. In particular, we study heterogeneity between various industry groups. The results show quite clearly that all this methodological issues matter empirically, e.g. industry effects are important, assumptions of additivity crucial and nonlinearities strong and leads to underestimations of the elasticities in a standard parametric model. In sum, the results might have far reaching implications for further empirical studies on executive pay. At least, it weakens the concern expressed by many in that field that strong pay-for-performance incentives for executives are missing.

* Graßhoff, Institute of Management, Humboldt-University Berlin; Schwalbach, Institute of Management, Humboldt-University Berlin; Sperlich, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, e-mail: stefan@est-econ.uc3m.es. This research was financially supported by Sonderforschungsbereich 373, Deutsche Forschungsgemeinschaft, at Humboldt-Universität zu Berlin and the Spanish "Dirección General de Enseñanza Superior" (DGES), reference number PB98-0025. We thank Stefan Profit for helpful comments.
1 Introduction

The relationship between executive pay and corporate financial performance continues to attract wide academic, media and policy attention. The very high salaries enjoyed by senior executives in corporations primarily in the Northern America are often contrasted with the relatively low pay received by executives in Europe and Asia. The stark differences in executive pay across some economies becomes an issue particularly in transatlantic mergers, like in the case of Daimler-Benz/Chrysler, BP/Amoco and Deutsche Bank/Bankers Trust. Despite the fascination with executive pay issues there is little academic research comparing executive pay across economies which can be used to design efficient transnational pay structures. As an exception see the international comparison by Conyon and Schwalbach (1999).

Standard academic research is based on agency theoretical assumptions that incentive contracts are characterized by asymmetric information between owners (and their representatives) and executives. The owners of the corporation have incomplete information about the effort of the executives to maximize the long-term value of the corporation. To align owners' with executives' interest, agency theory suggested to design a contract in which executive pay is tied to observed corporate performance. In standard contracts, executives receive a fixed income and performance-related pay. The proportion of fixed and variable pay depends on the executives' degree of risk and effort aversion as well as on firm and industry specific factors.

Empirical research on executive pay has mainly concentrated on the pay-performance relationship. Although the adopted data sets were very different within and across countries, the results are very similar and show very low pay-for-performance elasticities. Estimates vary between 0.1 and 0.15 for the United States, about 0.01 for Japan and about 0.06 for Germany, see the surveys by Murphy (1998), Schwalbach and Graßhoff (1997) or Graßhoff and Schwalbach (1997). This low estimates led some researchers to believe that the lack of strong pay-for-performance relationship reveal that most executive contracts are not incentive compatible, see for instance Jensen and Murphy (1990) and for the opposite view Graßhoff and Schwalbach (1999).

In contrast to the low pay-for-performance elasticities, empirical results show anonymously that executive pay is strongly influenced by firm size and industry effects. The pay-for-firm size elasticities varied within the magnitude of about 0.18 to 0.3 whereas the industry effects can lead to pay levels in some industry which can be 300 percent higher than in the industry with the lowest pay levels.3

Despite the similar results across very different data sets, several methodological issues are still uncovered. Almost all studies assume linear or semi-log linear pay functions without applying a test of the adequate functional form. Furthermore, most models do not allow for variations across corporations, industries, countries and time. It it assumed that pay func-

---

2For a most recent survey of the literature, see Murphy (1998).
3See for instance Murphy (1998) and Graßhoff and Schwalbach (1997).
tions are homogeneous across corporations and variations are captured by the fixed effects in the constants and by the assumption about the errors. The methodological deficiencies in the executive pay studies might also be a cause for the surprisingly similar results across very different data sets.

The purpose of the paper is to circumvent these difficulties of possible misspecification by adopting an explorative data analysis that uses nonparametric methods which impose rather weak restrictions on the model. Additionally, since they estimate the functions locally, the estimated general functional form or curvature does not get affected by outliers in the horizontal direction. Applying smoothing methods we can start with the most general model to get first an idea of what the data tell us. Our methods allow then for a stepwise closer look by specifying the various objectives of investigation or the model we deduce from the previous results.

Finally, nonparametric regression analysis will be applied to detect possible heterogeneity between sample subgroups. In particular, we study the pay function for various industry groups and expect that pay functions vary across industries and do not support the homogeneity assumption by parametric regression analysis.

The paper is organized as follows: In the next section the standard analysis of executive pay studies is reviewed. In section 3, the data base will be described. Section 4 presents results from parametric regressions as a benchmark for the nonparametric regression analysis which will be explained in section 5.

2 The Standard Empirical Model

In analyzing executive pay the standard empirical model contains corporate size and financial performance as determinants of pay. Corporate size is a measure of managerial discretion and financial performance is an indicator for managerial incentive compatibility. Both hypotheses are derived from agency theory as explained above. Basically, two types of regression equations are assumed:

\[ C_{it} = \alpha_i + \beta P_{it-1} + \gamma S_{it-1} + \epsilon_{it} \]  
\[ \ln C_{it} = \alpha_i + \beta P_{it} + \gamma \ln S_{it} + \epsilon_{it} \]  

where \( C_{it} \) stands for executive pay, \( P_{it} \) reflects measures of financial performance and \( S_{it} \) represents size for firm \( i \) at time \( t \). The terms \( \epsilon_{it} \) are the stochastic error terms whereas the parameters \( \alpha_i \) are mostly modeled as firm-specific fixed effects. In the first equation linearity is assumed and pay is determined by ex-post firm-specific measures with a time-lag of one year. The second equation is semi-log linear with no time-lag. In the previous empirical studies there is no rigorous statistical test to legitimize the linearity or semi-log linearity of the equations, see Rosen (1992). Most studies are based on the second equation and therefore we will concentrate our analysis on model (2).
Empirical models based on an equation of the form given in (2) usually employ panel data sets consisting of observations for the relevant variables by a sample of \( i = 1, \ldots, N \) firms over a time horizon of \( t = 1, \ldots, T \) years. In most data sets the number of available \( N \) firms is considerably larger than the length of the times series which may cause two problems. First, if for example one is interested in dealing with different firm-specific behavior, estimation for individual parameters by time series methods could be inefficient. Supposing that the slopes are homogeneous across firms leads to the belief that all individual variation is captured by the fixed effects and the assumptions about the errors. Second, for a large number \( N \)-firms one would expect that there might be groups of firms which may show significant different behavior. Therefore, the technique of pooling could be too restrictive since the estimates of the sensitivity parameters are usually given in some form of weighted means over individual effects and possible group effects are smoothed and therefore undetected.

In this paper we impose variation across industry groups in equation (2) which allows to detect possible heterogeneity across groups in terms of different functional forms and interactions between group dependent exogenous variables. Furthermore, we deal with the problem of outliers in the data for the independent variables, which could lead to biased estimates for pay sensitivities. We think that all these aspects are essential for empirical investigation in this context. Since we found them often neglected in empirical research in executive pay, it might weaken the seemingly robust but in many respects puzzling results.

3 Data Description

The regression analysis will be performed for \( j = 1, \ldots, 4 \) industry groups and equation (2) is extended to:

\[
\ln C_{jt} = \alpha_j t + \beta_j t P_{jt-1} + \gamma_j t \ln S_{jt-1} + \varepsilon_{jt}
\]  

(3)

Our data base is drawn from varies annual executive pay reports by "Kienbaum Vergütungsberatung". The data contain average annual total pay (fixed and variable) by the top executives of German stock companies (Vorstand of Aktiengesellschaften, in short AG’s) and "companies of limited liabilities" (Geschäftsführer of the Gesellschaft mit beschränkter Haftung, in short GmbH's). In total, we use data of up to 339 manufacturing firms for the period of 1988 to 1994. Company size is measured by the number of employees and corporate financial performance by the rate of return on sales (ROS). Companies are grouped into the following four distinct industry groups: (1) Basis industries, (2) Capital goods, (3) Consumer goods and (4) Food, drinks and tobacco. These industry groups contain the following industries in which the sample companies are operating (number of companies in parenthesis for the year 1990):

- Standard descriptive statistics about the data can be found in Tables 3 to 5 in the appendix. Table 2 reveal that the data set consists of relatively few firms in group 3 (consumer goods)

4The measures for independent variables predate the compensation measure to recognize that compensation contracts are written after observing values for the independent variables.
Table 1: Industry Groups.

and group 4 (food, drinks and tobacco) which we have to take into account when discussing the efficiency of regression results. Furthermore, the data set consists of an uneven distribution of small (less than 500 employees) and large (more then 50.000 employees) firms across industry groups within the period of 1988 to 1994.

Table 4 provides means and medians for the executive compensation and the firm size measures for each group over the six years. One can see that the level of executive compensation is slightly higher in groups 1 and 2 which might be explained by the relatively high proportion of large firms in that groups. Comparing values of means and medians one can observe significant size differences among large firms. We solve this heterogeneity problem by taking the logarithm of the firm size variable (see equation (3)).

Table 5 shows that the range of the observed financial performance of the firms in the sample vary considerable within and across industry groups. Therefore, diagnostic analysis is needed to decide if there exist singular observations, which have high influence on regression estimation results.

4 Parametric Regression Results

The regression results for the cross-sectional analysis of equation (3) are summarized in Table 2. The estimated sensitivity parameter for the firm size variable can directly be interpreted as the estimated size elasticity in each case. The parameter for the financial performance variable multiplied with the arithmetic mean in each case will represent an estimation for the performance elasticity at the mean. We will use the notation $\delta s_{zt}$ and $\delta p_{zt}$ for the size and
performance elasticity in group \( j \) at time point \( t \), so we have:

\[
\varepsilon_{S,t}^{j} = \gamma_{S,t}^{j} \quad \varepsilon_{P,t}^{j} = \beta_{P,t}^{j} \cdot \text{mean}(P_{jt}^{t-1})
\]

meaning that one percent increase in firm size will result roughly in an \( \varepsilon_{S,t}^{j} \) increase in executive compensation by holding the performance variable fixed, and respectively, one percent increase of the performance measure at its mean will cause an \( \varepsilon_{P,t}^{j} \) percent increase in compensation.

If one looks at the results for group 1 (basic industries) we can see that both size and performance influences are estimated by positive values which are significantly different from zero for each of the six years. The estimated size elasticities are around 0.25 for the years 1989-1991 then decrease in the year 1992. The estimated performance elasticities evaluated at the means are around 0.12 for the years 1989-1991, then decrease slightly to 0.09 for the next two years whereas the value for 1994 is considerably low with 0.03.

For industry group 2 (capital goods) we obtain significant positive values for the size elasticities in each year, but concerning the estimated parameters for the performance variable we are faced with an irregular scheme over the years. For the years 1989 and 1991 we get positive significant estimations resulting in estimated elasticities about 0.05. For the remaining years we obtain insignificant values which are negative for the last two years. Looking at the industry groups 3 (consumer goods) and 4 (food, drinks and tobacco) we can find some similarity meaning that some estimates are significant some are not with estimated values that spread considerably.

Remembering the wide range of the values of the performance variable (ROS), it is not surprising to see inconsistency in the regression results. Therefore, we formed further diagnostics due to outlying data points in the independent variables. Let us demonstrate the procedure and the consequences for the results by some examples:

Looking at the data for group 2 in year 1990 there is a firm with a ROS value of 0.7 in combination with a value of 1,667 for employees and a compensation value of 412,700. We compute the centered leverage point values (diagonal values of the "hat matrix"⁵, see Belsley, Kuh and Welsch (1980)), as a measure to identify cases with unusual combinations of values for the independent variables and cases which may have a large impact on the regression model. For this firm we obtain the value 0.65. Since values which are considerably larger than the ratio of the number of independent and the number of observations are considered as crucial, we decide to classify this value as far away from the bulk of the data. We suspect that its influence on the regression result is over-weighted meaning that the correspondent value for the dependent variable, the value for compensation, determines the estimation of the regression line essentially. Deleting this data point of the \( N_j = 127 \) data points for this group and year and running a new regression we compute a positive significant value

⁵With \( X \) the matrix of observations of the independent variables, the matrix \( X(X'X)^{-1}X' \) is called the Hat matrix. The diagonal values of this matrix lie between 0 and 1, and the larger the value is, the greater is the distance measure which shows how far away the data point is away from the middle of the data distribution.
Table 2: Parametric regression results.

different from zero. Table 2 demonstrates this issue by forming two columns for the year 1990 for the groups 2 to 4 one giving the results for the original data, one by deleting one data point, who gives the highest value for the leverage criteria. By this we can state that outlying data points are in some cases responsible for strange regression results. The following non-parametric analysis will give us the tools to have a closer look on the relations by estimation methods which are not touched by outlying data points concerning the independent variables. By this we will focus on possible heterogeneous behavior over groups and time due to interaction effects between the independent variables on one side and due to possible different nonlinear functional forms on the other side.
5 Nonparametric analysis

In this section we provide a brief overview to the nonparametric estimation methods we have used for the explorative analysis presented in the previous section. A more detailed description of the methods can be found in the Appendix 8.2. In this paper we are not interested in the statistical properties of these estimators, so we only briefly discuss the difference in their interpretation in this section. For a detailed comparison of these estimators, see Sperlich, Linton and Härdle (1999).

We start with a general multidimensional regression estimator, in particular the two dimensional Nadaraya -Watson estimator and a local polynomial estimator. Then we turn to the problem of estimation in additive models. Therefore we first present the backfitting algorithm and afterwards introduce the more recently developed marginal integration estimator. The two dimensional Nadaraya -Watson estimator will give us a first visual impression of the relations. We can realize whether there is a homogeneous behavior across groups and years concerning functional forms. Further we can get hints about possible interactions between the independent variables and by this we can decide for each group and each time point whether the restriction on an additive model will be reasonable.

The backfitting algorithm is first projecting the multidimensional regression problem into the space of additive models. In this special subspace it is calculating the optimal regression fit for the underlying data, regardless whether the true model is additive or not.

The Marginal Integration Estimator is always estimating the marginal effects of various inputs. This is done by integrating out the other direction (≈ marginal integration). It does not necessarily look for the optimal fit.

In case of non-additivity thus the backfitting is doing a kind of ANOVA decomposition and still provides a good regression fit, whereas the M.I.E. yields consistent estimates for the marginal effects but given non-additivity, their sum is not an estimate for the regression function. Consequently, estimation results of backfitting are closer to the corresponding estimates of parametric additive models.

There exist several proposals how to test for additivity. But most of them are restricted on fixed alternatives, e.g. they allow for a multiplicative interaction of two explanatory variables. More general approaches have been suggested by Gozalo, Linton (1999) and Sperlich, Tjøstheim, Yang (1999). In both articles they use nonparametric estimation, testing and bootstrap methods and a reasonable performance can hardly be expected with (much) less than only hundred observations. So we have to restrict here on arguing e.g. for or against additivity based on certain estimation results but can not prove 95% significance.

However, in case of only two input variables we can plot a three dimensional regression fit to get visually an idea of the structure. Furthermore, taking into account the above mentioned differences between backfitting and M.I.E., a comparison of their estimation results is also providing information about separability. If the estimates differ a lot, i.e. the variance explaining additive components (in backfitting) do not coincide with the marginal effects (given by M.I.E.), then obviously some interaction is present!
Figure 1: The 2-dimensional Nadaraya-Watson estimation for 1989/90. Plotted are the expected executive pay vs size (left axes) and ROS (right axes). First row: group 1 and 2, second row: group 3 and 4.

6 Nonparametric Regression Results

6.1 The 2-dimensional Nadaraya-Watson estimator

To get a primary visual impression of the possible functional forms we first applied the multidimensional, in our case two dimensional, NW estimator. Therefore we used the quartic kernel with bandwidth \( h = 2.5 \times \text{std}(x) \), where \( \text{std}(x) \) represents the standard deviation of the independent variables.

Please notice that since our estimator is a local adaptive one, our results are not touched by possible outliers in the \( x \)-direction. For better presentation we show the plots over trimmed ranges.

We have selected the results for two representative years, see Figures 1 and 2. Considering the plots over the years we can realize strong functional similarities between the industry groups 1 and 2 while the results for the other groups seem not to be homogeneous at all. Regardless the outliers we see a strong positive relation for compensation to firm size at
least for group 1 and 2, and a weaker one to the performance measure varying over years and groups.

Further we can recognize some interaction of the independent variables especially in group 3 and 4. This can visually be detected as follows. Imagine you cut slices parallel to the $x$-axes. If these slices indicate different functional forms within one direction separability of the inputs is not justified. Regarding this procedure we state additivity for group 1 and 2.

6.2 Estimation results under additivity restrictions

As mentioned above the backfitting procedure is at first projecting the data into the space of additive models and looking there for the optimal fit. For this it makes sense to apply this estimator even if separability is not given as stated for group 3 and 4. We only have to be aware of the problem of interpretation and the deterioration of the regression.
Figure 3: Backfit and Parametric estimates for 1989/90, and for industry groups 1-4 from top to bottom. For the parametric case we have results for untrimmed (dashed) and trimmed (solid) data. For untrimmed data (dashed line) the influence is insignificant.

6.3 Backfitting and Parametric Results

Now we have a closer look to the additive components using the backfitting estimator, see appendix for explanations. For the smoothing we use a local linear kernel smoother with
quartic kernel and bandwidth \( h = (0.5, 0.6)^T \star \text{std}(x) \).

In Figures 3 and 4 we show the nonparametric results in comparison with the parametric results for trimmed and untrimmed data. Again note that the results change only for the parametric model. In all industry groups we see clear nonlinearities for between compensation and financial performance. The low values in the parametric model only describe the
linear influence. But obviously this low values are due to functional misspecification.
Considering the firm size influence we also detect a significant bump in the middle of the
range in 1991/92 for groups 3 and 4. Unfortunately, we do not find an economic explanation
for this feature but on the other hand this can be caused by interactions realized in the
3-dimensional plots in Figures 1 and 2.

6.4 Marginal Integration Results

Finally, we estimate the pure marginal effects of the independent variables. We use again
the local linear kernel smoother with quartic kernel with bandwidths \( h = 1.5 \times \text{std}(x) \),
\( g = 2.5 \times \text{std}(x) \). We present the estimation results together with 'confidence intervals' in
forms of 2\( \sigma \)-bands 6.

As main result we can postulate that these estimation results are consistent with the finding
above. First the nonlinearities of the financial performance influence are strengthened espe­
cially for groups 1 and 2. Second since the above mentioned bumps in the firm size relation
are not existent now, we can conclude that indeed interactions are responsible. In general,
the backfitting results differ from the estimated marginal effects substantially in groups 3 and
4. This again supports that interaction effects definitively have to be taken into account to
get reasonable regression results. We admit that this makes economic interpretation rather
difficult.

7 Conclusions

The paper provides an explorative study for executive pay determination employing different
nonparametric methods. In comparison with parametric methods applied in almost all stud­
ies on executive pay, the advantage of this approach is that the analysis is not restricted to
functional forms, it does not require additivity assumptions and it is not effected by outlier
in the data. The results show quite clearly that all this effects matter. Mainly we found:
(1) Industry effects are important and are detected by introducing industry groups. In our
investigation, maybe, only group 1 and 2 could be considered as one group. (2) The assump­
tion of additivity is crucial (in particular for groups 3 and 4). For these groups, interaction
effects between indepedent variables exist. (3) Nonlinearities are strong between exec­
tutive compensation and financial performance (especially for groups 3 and 4) which lead
to underestimations of the elasticities in a standard parametric model. In sum, the results
might have far reaching implications for further empirical studies on executive pay. At least,
it weakens the concern expressed by many in that field and most prominently phrased by
Jensen and Murphy (1990) that "... the lack of strong pay-for-performance incentives for
CEOs ... is puzzling." (p.262).

6The smoothness in these plots, see Figure 5, is caused by the choice of large bandwidths.

13
Figure 5: Marginal Integration estimates for 1991/92 with 2σ bands for industry groups 1-4 from top to bottom.
8 Appendix

8.1 Tables for descriptive statistics and parametric regression results

<table>
<thead>
<tr>
<th>Group</th>
<th>N min-max</th>
<th>Small firms min-max (% of sample)</th>
<th>Large firms min-max (% of sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>81-109</td>
<td>19-25</td>
<td>28-31</td>
</tr>
<tr>
<td>2</td>
<td>100-145</td>
<td>10-16</td>
<td>29-34</td>
</tr>
<tr>
<td>3</td>
<td>39-51</td>
<td>21-37</td>
<td>5-19</td>
</tr>
<tr>
<td>4</td>
<td>22-40</td>
<td>41-61</td>
<td>3-15</td>
</tr>
</tbody>
</table>

Table 3: Sample size and distribution of small and large firms, where small firm means number of employees lower than 500, and large firm a number of employees greater than 50000.

<table>
<thead>
<tr>
<th>Group</th>
<th>Comp.</th>
<th>Empl.</th>
<th>Comp.</th>
<th>Empl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean</td>
<td>556</td>
<td>Mean</td>
<td>538</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>495</td>
<td>Median</td>
<td>467</td>
</tr>
<tr>
<td>2</td>
<td>Mean</td>
<td>12255</td>
<td>Mean</td>
<td>12334</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>2045</td>
<td>Median</td>
<td>2104</td>
</tr>
</tbody>
</table>

Table 4: Descriptive statistics for executive compensation (in thousand DM) and firm size.

<table>
<thead>
<tr>
<th>Group</th>
<th>Return on sales in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>-15 -36 -40 -35</td>
</tr>
<tr>
<td>max</td>
<td>15 70 13 50</td>
</tr>
<tr>
<td>mean</td>
<td>1.90 0.94 1.37 1.20</td>
</tr>
<tr>
<td>median</td>
<td>1.73 1.56 1.22 0.95</td>
</tr>
</tbody>
</table>

Table 5: Descriptive statistics for financial performance of sample firms.
8.2 Nonparametric Regression Methods

This section is devoted to a brief introduction to the nonparametric estimation methods we have used for the explorative analysis presented in the sections 5 and 5.

We start with a general multidimensional regression estimator, in particular the two-dimensional Nadaraya-Watson estimator and a local polynomial estimator. Then we turn to the problem of estimation in additive models. Therefore we first present the backfitting algorithm and afterwards introduce the more recently developed marginal integration estimator.

8.2.1 Nadaraya-Watson and local polynomial estimator

Consider the regression problem of estimating the functional relation between a response variable $Y \in \mathbb{R}$ and its possibly multidimensional explanatory variable $X \in \mathbb{R}^d$, i.e. estimating the conditional expectation $m(x) = E[Y|X = x]$, where $x$ is the realization of $X$.

The underlying model is

$$Y = m(X) + \sigma(X)\varepsilon$$

with $E(\varepsilon) = 0$, $Var(\varepsilon) = 1$, $\varepsilon$ independent of $X$ and $\sigma(\cdot)$ a bounded variance function. Given $\varphi(y|x)$ is the conditional density of $y$ given $x$, $\varphi(x, y)$ the joint density and $\varphi_X(x)$ the marginal one of $X$, we have

$$E[Y|X = x] = \int y \varphi(y|x) dy = \int y \frac{\varphi(x, y)}{\varphi_X(x)} dy. \quad (4)$$

A nonparametric Kernel Density Estimator for $\varphi(x, y)$ with bandwidths $h, g$ and kernel function $K_h(\cdot) = h^{-1}K(h^{-1})$ is

$$\hat{\varphi}_{h,g}(x, y) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)K_g(y - Y_i).$$

Thus, for the denominator of (4) we get

$$\int y \hat{\varphi}_{h,g}(x, y) dy = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)Y_i$$

and the resulting estimator for $E[Y|X = x]$ is

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^{n} W_{hi}(x)Y_i$$

with weights $W_{hi}(x) = \frac{K_h(x - X_i)}{\hat{\varphi}_h(x)}$. Here, $\hat{\varphi}_h(x)$ can be replaced by $\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)$.

Unfortunately this estimator is biased, and choosing $h \to \infty$, the estimates $\hat{m}(x) \to \frac{1}{n} \sum Y_i$ for all $x$. To get rid of at least of the bias in linear direction, we have to take into account
the first derivative of $m$. This can be done by local polynomial estimation. For the ease of presentation let $x \in \mathbb{R}^1$. Consider the Taylor expansion for sufficiently smooth functions $m(\cdot)$:

$$m(t) \approx m(x) + m'(x)(t - x) + m''(x)(t - x)^2 \frac{1}{2} + \cdots + m(p)(t - x)^p \frac{1}{p}$$

for $t \in U_h(x)$ together with the weighted least squares problem

$$\min_{\beta} \sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1(X_i - x) - \beta_2(X_i - x)^2 - \cdots - \beta_p(X_i - x)^p\}^2 K_h(X_i - x).$$

Obviously the resulting $\beta = (\beta_0, \ldots, \beta_p)^T$ provides us with estimates for $\frac{1}{\nu}m^{(\nu)}(x), \nu = 0, 1, \ldots, p$. Here $m^{(\nu)}$ is the $\nu$'th derivative of $m$.

### 8.2.2 Additive Models

We consider the same regression problem as above with $X \in \mathbb{R}^d, d \geq 2$. The only difference is that for a better interpretability and some wished theoretical properties we assume the underlying model to be separable, i.e.

$$m(x) = E(Y|X) = c + \sum_{j=1}^{d} f_j(X_j).$$

We call the function $f_j(\cdot)$ additive components and set for identifiability reasons

$$E[ f_j(X_j) ] = 0 \quad \text{for all} \quad j = 1, \ldots, d.$$

Please note that this centering of the $f_j$ is no restriction since in a multidimensional regression model we always can only identify the relative influence of each factor $j$.

### 8.2.3 The Backfitting Algorithm

This estimation technique has been introduced by Buja, Hastie and Tibshirani (1989). They consider the following problem:

$$\text{minimize} \quad E[Y - m(X)]^2 \quad \text{s.t.} \quad m(X) = c + \sum_{j=1}^{d} f_j(X_j).$$

From projection theory we know that there exists a unique solution for this problem with:

$$E[\{Y - m(X)\} | X_\alpha] = 0 \iff f_\alpha(X_\alpha) = E \left[ \left\{ Y - \sum_{j \neq \alpha} f_j(X_j) \right\} | X_\alpha \right], \text{ for all } \alpha.$$

A feasible algorithm to estimate these $f_\alpha$ non parametrically is:
1. initialize $\hat{\varphi} = 0, \hat{c} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

2. repeat for $\alpha = 1, \ldots, d$, $r = (r_1, \ldots, r_n)^T$
   $$r_i = y_i - \hat{c} - \sum_{j \neq \alpha} \hat{f}_j(X_{ij})$$
   $$\hat{f}_\alpha(X_{i\alpha}) = S(r|X_{i\alpha}),$$

where $S(\cdot | \cdot)$ is a smoothing operator such as the Nadaraya-Watson or the local polynomial estimator, i.e. we regress $r$ on $X_\alpha = (X_{1\alpha}, \ldots, X_{n\alpha})^T$ at point $X_{i\alpha}$.

3. proceed until convergence is reached, i.e. the estimates $\hat{f}_\alpha$ do not change significantly any more.

Consistency of the backfitting estimator under weak conditions has recently been proved by Mammen, Linton and Nielsen (1999).

8.2.4 The Marginal Integration Estimator

This method is calculating the marginal effects of each input factor. In an additive separable model as (5) the marginal effects correspond to the additive functions $f_j$. The idea of this method is to pre-estimate the multi-dimensional regression surface and to integrate out the directions not of interest. Due to the indentifiability conditions $EX_{X_\alpha}[f_\alpha(X_\alpha)], EY = c$ we have for $X_\alpha = (X_1, \ldots, X_{\alpha-1}, X_{\alpha+1}, \ldots, X_d)$ with the marginal density $\varphi_\alpha$

$$EX_{X_\alpha} = [m(x_\alpha, X_\alpha)] = EX_{X_\alpha} \left[ c + \sum_{j \neq \alpha} f_j(x_j) + f_\alpha(x_\alpha) \right]$$

$$= c + f_\alpha(x_\alpha)$$

Replacing $EX_{X_\alpha}$ by taking the average over the observations and $m(x_\alpha, X_\alpha)$ by an appropriate pre-estimate, we get the so called Marginal Integration Estimator (M.I.E.) $\hat{f}_\alpha(x_\alpha) = \frac{1}{n} \sum_{i=1}^{n} \hat{m}(x_\alpha, X_{i\alpha})$. This method has first been introduced by Tjøstheim and Auestad (1994) and Linton and Nielsen (1995).
References


