Analysis of disclosure determinants: a local-relation approach

Abstract

Purpose
In literature on determinants of disclosure scholars generally tend to investigate the existence of relations in "global" terms, by considering the whole range of observed values pertaining to the dependent and independent variables involved in the descriptive model. Despite the different methodologies employed, coherently to this approach, a hypothesis can be only accepted or rejected entirely. This paper intends to contribute to literature by proposing a data-driven method based on smooth curves, which allow scholars to detect the existence of local relations, significant in a limited interval of the dependent variable.

Design/methodology/approach
The employment of smooth curves is simplified by conducting a study on goodwill disclosure. The model deriving by the adoption of the LOWESS smoothed curves may provide an accurate description about complex relations between the extent of disclosure and its expected determinants, whose shape is not completely captured by traditional statistic techniques.

Findings
The model based on LOWESS curves provided a comprehensive description about the complexities, which characterize the relations between disclosure and its determinants. The results show that in some cases, the extent of disclosure is influenced by multi-faceted local relations.

Practical implications
The exemplificative study provide evidences useful for standard setters to improve their comprehension about the inclination of companies in disclosing information on goodwill impairment.

Originality/value
The adoption of smooth curves is coherent with an inductive research approach, where empirical evidence is generalized and evolves into theoretical explanations The method proposed is replicable in all the field of studies, when extant studies come to unclear and contradicting results as consequence of the complex relations investigated.

Keywords

Determinants of corporate disclosure; Local regression; LOWESS smooth curve; Generalized Additive Model

1. Introduction

Since the 1970s, the literature on the determinants of disclosure has been abundant and progressively growing. Recent studies have considered several sets of explicative variables, which are firm-specific and country-specific, but the results obtained are not unequivocal. This is due to a multiplicity of reasons concerning as example: the methods used to measure the

quantity and quality of disclosure, the independent variables, or the descriptive models adopted to empirically test the hypothesis (Chavent et al., 2006; Khlifi and Bouri, 2010). In addition, improper assumptions about the existence of a linear relationship between disclosure and its determinants or multicollinearity issues may produce contradicting results.

Several scholars, trying to find alternative solutions which could improve the representativeness of the relationships between disclosure and its determinants, come to the following common consideration: it is essential to analyse the structure of data and its underlying relations in order to avoid misrepresentations (Cooke, 1998; Parviainen, et al., 2001).

The aim of the paper is to describe a data-driven research approach inspired by the general statements of artificial intelligence techniques that can provide researchers in the comprehension of the shape and complexity of a relation between dependent and independent variables. In particular, we assume that the relation between disclosure and its determinants is not necessarily a general one, but it may change its shape between different ranges of the explanatory variables. In other words, we assume the existence of meaningful local relations, whose robustness is limited within particular intervals of the investigated explanatory variables. The local relations might be an explanation for some of the contradicting results obtained in previous studies.

The inductive approach proposed in the paper is based on LOcally WEighted Scatter plot Smoothing curves (LOWESS) which are essentially drawn directly by the data, used in combination with Generalized Additive Model (GAM). This methodology is intended as a meaningful complementary research approach since it may help scholars to gain insights about how a set of variables influence disclosure. Such an approach may be suitable at least in two cases:

- when the assumptions about the existence of a general linear relation between disclosure and a set of determinants are not supported in the whole range of the explanatory variables, the LOWESS curve may provide evidence about the existence of significant relations at local level, which are valid only within specific sub-intervals of the independent variables;
- when the assumed relation between disclosure and its determinants is supported at a general level, the LOWESS curve may provide more accurate insights about the shape of the relation between the variables, which may be significant in case of nonlinear and complex relations.

The results obtained might contribute to literature by providing evidence that studies adopting only one single theoretical framework may come to an oversimplification of the

investigated phenomenon since the managers may adopt multiple strategies and behaviours in disclosing information, which vary between one range of the dependent variables and another.

The methodology proposed is illustrated through a study concerning the disclosure of goodwill impairment, and the results obtained show that when contradictions emerge in literature about if and how variables may influence disclosure, an approach based on data-exploration may significantly improve the comprehension and the capacity to describe the investigated phenomenon.

The paper is structured as follows. In the next section, we summarize the literature concerning the methodologies generally used to investigate the determinants of disclosure and the principal contradictions that may emerge by their adoption. In the third section, the LOWESS curves and GAM are presented and their potentialities are described; in the fourth section, the application of the LOWESS curves and GAM is described by developing research hypotheses concerning the determinants of disclosure about goodwill impairment. In the fifth section, the data collection and the dependent and independent variables are described; in the sixth section, the hypotheses are tested through a comparison of general and local regressions, and the main findings are discussed. Final considerations and further research conclude the paper.

2. Methods for investigating the determinants of disclosure: the state of the art

Several studies in the field of disclosure determinants show contradictory findings in a wide range of explanatory variables, including company size (Ng and Koh, 1994; Malone et al., 1993); ownership concentration (Donnelly and Mulcahy, 2008; Glaum et al., 2013; Khlifi and Bouri, 2010; Lang, Lins and Miller, 2004); level of debt, profitability (Armitage and Marston, 2008), and so forth.

Chavent et al. (2006) attribute the contradictions to the following methodological shortcomings:

1) there is not a common practice for measuring the extent of disclosure. The disclosure indexes generally employed are determined by totalling several different items, which are typically not weighted and focus on the quantity, rather than the quality, of disclosure;
2) a wide number of explanatory variables are generally included in the descriptive models, (especially in multi-country analysis), which may produce problems of multicollinearity;
3) naive assumptions are often made about the existence of a linear relation between the level of disclosure and its determinants.

Additionally, the contradictions may be also due to the samples employed, which are not always comparable in terms of size and investigated population. (Ernst & Young, 2006; Glaum et al., 2013).

Khlifi and Bouri (2010), argued for the development of "non-conventional" methods of statistical analysis, which should allow researchers to describe the relations with greater accuracy, overcoming the uncertainties about their shape.

Cooke (1998) describes three alternative data-transformation methods (rank transformation, normal scores approach and log of the odds ration) and compares the results obtained by adopting data-transformed ranked regressions and multiple linear regressions (MLR). As results of the comparison Cooke finds that no one of the methods is overwhelming in improving the fitting of data in all the cases, and consequently examination of both the structure of data and the relations between dependent and independent variable becomes of extreme importance in order to avoid errors in interpretation.

Parviainen et al. (2001) compares the adoption of the Generalized Method of Moments (GMM), which is largely used in econometric literature, with MLR. The GMM is a computational method that can provide robust estimators for a statistical nonlinear model (Hall, 2005). The results obtained provide empirical evidence for several advantages (Parviainen et al., 2001): nonlinear relations are detected through the GMM, and consequently the level of significance is improved; the variables that are significant in a linear model maintain or even improve their significance by adopting the GMM, since a nonlinear relation can mask itself as a linear one. As a main drawback, the GMM can produce conflicting results since the elaboration method is based on a set of assumptions that should require the researcher to know ex ante what kind of relation is underlying the data.

When the correct functional form about a phenomenon is unknown, or when complex phenomena may not be represented accurately by any of the general functions fitted by the traditional techniques, the simplifications used by researchers may lead to misrepresentations of the structure within the data (Chambers et al., 1983, Jacoby, 2000). In such cases, a direct visualization of data may let the data "to speak for themselves" (Jacoby, 2000) and scholars may be provided with a representation of the relationships between variables, without needing to meet the a priori assumptions required by traditional correlations.

In line with this, we propose the use of LOcally WEighted Scatter plot Smoothing curves (LOWESS), which are drawn directly by the data collected. As main advantage, the LOWESS can fit a wide range of data distributions without requiring any preliminary assumption about the shape of the underlying relationships.

In additions, the LOWESS curves are suitable to detect the existence of non-monotonic relations, composed by multiple different local relations, which are only robust within particular intervals of the investigated explanatory variables.

The LOWESS curves allow for the adoption of a Generalized Additive Model (GAM) that expresses the multivariate structure of the data and provides evidence about the "net effect" of a given explanatory variable after the effects of remaining independent variables have been removed, similarly to a MLR.

We assume that the LOWESS could be helpful in explaining some of the contradicting results obtained in previous studies. In the next section, a detailed description of the LOWESS curves in combination with GAM is provided.

3. LOWESS smooth curves and Generalized Additive Model

A smooth curve allows showing how the mean values of a Y variable change in different ranges of an X variable. When the smooth curve has the shape of a horizontal line, this means that the investigated variables are not correlated throughout the entire plotting area. Conversely, if a relation between variables exists, the smooth curve should reflect the shape of the structural patterns in the data set.

As with linear regression, the smooth curve is drawn by minimizing the variance of the residuals or prediction errors, at local level. The LOWESS is one of the most frequently used bivariate smoother based on non-parametric regressions (Cleveland, 1979). In particular, the LOWESS method is motivated by the assumption that in a range of x-values, neighbouring values of the independent variable are the best indicators of the dependent variable. Indeed, it is perhaps unreasonable to expect a single functional relationship between Y and X throughout the range of X (Cleveland, 1994). The basic idea is that, for a predictor x, the regression function $g(x)$ can be locally approximated by the value of a function belonging to a specified parametric class. Such a local approximation is obtained by fitting a regression surface to the data points within a chosen neighbourhood of the x.

The weighted least squares are used to fit linear or quadratic functions of the predictors at the centres of neighbourhoods: the method uses only the closest neighbour observations in relation to each data-point in order to minimize the weighted residual sum of squares. In practice, the decision about the degree of the polynomial (λ) – that the LOWESS procedure fits the data – is often based upon a visual inspection of the scatter plot, where λ is set equal to 1 or 2 when a local linear polynomial or locally quadratic fitting is suitable to fit the data, respectively.

The width of each neighbourhood is defined by the smoothing parameter (α) which is a fraction, ranging from 0 to 1. As (α) increases, the fitted curve becomes smoother, but it may fail in passing through the centre of the entire point cloud. Conversely, as (α) decreases, the local regression becomes more sensitive to outliers and the resulting curve may be affected by noise. The rule of thumb is to choose the smallest parameter that provides a smooth fit, able to capture the overall structure in the data. An analysis of the residual plots may be helpful either to assess the goodness of fit and the correct understanding of the underlying structure of data..

Data points in a given local neighbourhood are weighted in an inverse proportion to their distance from the centre of the neighbourhood, consequently higher weights are assigned to observations that are closest to the centre. The most commonly used function is the *tricube* weight function. The LOWESS procedure performs iterative reweighting to provide robust fitting and ensure that outliers do not affect excessively the smooth curve. The robustness step excludes *de facto* cases in which large residuals on local regression occur, with small probability of being observed; as such, the occurrence of the smooth curve 'chasing the outliers' in the scatter plot is avoided.

Although LOWESS is often used as a descriptive tool, it allows to perform an inferential analysis, in order to generalize the results about the structure of the population which the sample belongs to (Cleveland and Devlin 1988, Efron and Tibshirani 1993, Beck and Jackman,1998 Fox 1999, Jacoby 2000). The inferential analysis is useful to assess the reliability of the fitted curve in representing the precise functional form of the relationship between variables. When the assumptions required (i.e. the residuals between fitted and observed values should be normally distributed) are met, an F-test can be used to measure the improvement in fit provided by the LOWESS curve, when moving from simpler to more complex models (Beck and Jackman,1998; Fox 1999; Jacoby, 2000).

Statistical inference allows answering the question whether the LOWESS curve measures "anything more than noise due to sampling variability" (Jacoby, 2000). In this case, the F-test enables to evaluate the fitting performance of the curve against the null hypothesis of no functional dependence between variables. The formula is:

$$F = \frac{(TSS_y - RSS_{lowess})/(df_{lowess} - 1)}{(RSS_{lowess})/(n - df_{lowess})}$$

where TSS_y is the total sum of squared residuals between the observed values in Y and their mean; RSS_{lowess} is the sum of squared LOWESS residuals, df_{lowess} is the degree of freedom

associated with the fitted curve (or the equivalent number of parameters), and *n* is the total number of observations. If the obtained value of F-statistic is higher than the corresponding critical value, and therefore its p-value is lower than 0.05, the null hypothesis cannot be rejected and consequently the LOWESS curve is able to capture structural pattern within data.

The next step of an inferential analysis might be to evaluate whether the LOWESS curve provides any improvement in fit, compared to the simpler, linear regression curve (Fox 1999; Jacoby, 2000). The formula of F-statistic will be:

$$F = \frac{\left(RSS_{linregr} - RSS_{lowess}\right)/(df_{lowess} - 2)}{(RSS_{lowess})/(n - df_{lowess})}$$

where $RSS_{linregr}$ is the sum of squared residuals for linear regression; and 2 are the corresponding degrees of freedom (being this regression a bivariate equation).

When the p-value of F- statistic is lower than 0.05 it is possible to affirm that the linear fit provides a misrepresentation of the relationship between variables, and that nonlinearity is not dependent by the "noise" deriving from sampling error, while the LOWESS curve enables to discern such curvilinear relationship and to explicate its form.

Summarizing, the LOWESS method is particularly suitable when the point-cloud is affected by noise and sparse data can hinder to visually identify the structural patterns or to evaluate the functional form of dependency between two variables (Beck and Jackman,1998; Jacoby, 2000). In these cases, the traditional statistic modelling techniques are not able to reveal the existence of complex relationships, and therefore to provide any guidance about the exact nature of these patterns. The LOWESS procedure may be a viable alternative also to increase understanding of the data set by putting in evidence some intervals where the predictor produces a significant influence over the dependent variable (Miller and Hall, 2010).

The adoption of LOWESS does not require relevant complexities when the curves are drawn following a bivariate analysis, even if the resulting spectrum of research is limited since the effects produced by conditioning influences of dependent variables are excluded. To this purpose, a Generalized Additive Model (GAM) may allow to combine the benefit of LOWESS curves with the wide spectrum of a multivariate model of analysis. A GAM explains, in additive form, the (non-linear) relationship between the dependent and independent variables (Hastie and Tibshirani, 1990) using non-parametric functions, such as LOWESS.

The basic form of a GAM that describes the relationship between the dependent variable Y and a series of predictors X_k can be expressed as follows:

$$\eta(Y_i) = \alpha + \sum_{j=1}^{k} f(X_{ij}) + \varepsilon_i$$

The first part of Equation indicates that dependent variable is connected to the predictors by η, a so-called *link function*, that is a transformation describing how to obtain the predicted Y values. Among the several types of link functions, we have an *additive regression model* or *additive model* when an identity function is used for which η(Y) = Y. On the second part of Equation, each f_j corresponds to the non-parametric function describing the relationship between the transformed Y and the *i*th predictor, while α indicates the intercept and ε the disturbance term.

Therefore, an additive model can be considered as a generalization of a MLR model which is able to expresses the *multivariate* structure of the data, by a summation in which the linear regression coefficient for each variable (the additive term) is replaced with a smooth function, the LOWESS, to maximize the quality of prediction of the dependent variable. The additive nature of the model ensures that the impact of each *Xj* on *Y* can be interpreted as the "net effect" of the predictor, after the effects of other *k*-1 independent variables have been removed, just as in multiple linear regression.

As a main result, the GAM provides a set of scatterplots, showing the predictor values against the partial residuals, that is, transformations of the Y values that remove the effect of the remaining predictors (Hastie and Tibshirani, 1990). In other words, the values on the vertical axis in each plot correspond to adjusted Y values that allow to understand the actual nature of the relationship between the predictor and the dependent variable, controlling (or holding constant) all the other independent variables in the model.

An iterative estimation procedure, usually called *backfitting* is used to isolate the net effect of each independent variable X_j, after which it is possible to use *bivariate* LOWESS fitting to describe the functional dependence of Y on X_j, taking into account the *multivariate* structure within data (Jacoby, 2000).

These graphs can reveal effects that are not immediately observable in the bivariate scatterplots drawn on the raw values. Indeed the effect of the predictor X_j could be obscured by the strong effect of another variable. The scatterplots obtained from a GAM allow to overcome this obstacle, showing clearly how the response variable depends upon a predictor, *given* other independent variables.

GAMs can be considered a sensible compromise between simple linear and more complex models, combining flexibility and ease of interpretation, especially in case of nonlinear relationships and significant noise in the data (Hastie and Tibshirani, 1990).

GAMs flexibility is due to their non-parametric local properties, while alternative nonlinear descriptive models, based on polynomials or logarithmic transformations, generally allow to detect only global relations characterized by a unique functional specification. Furthermore, complex models, based on parametric transformations, often present multicollinearity problems and are not effective in case of anomalous behaviors at the ends of data range. Compared to these functional forms, GAMs are simpler and allow to obtain a better prediction of the dependent variable values (Beck and Jackman,1998).

4. Applying the LOWESS curves. Research design and description of hypotheses

In the current section and in the one following, we describe the adoption of the LOWESS smooth curves and the GAM trough an exemplificative study dealing with determinants of goodwill disclosure. The results obtained allows us to discuss the usefulness of these methods in complementing the hypothesis testing approaches generally adopted in the field of studies of financial disclosure.

The research framework is summarized in Figure 1.

Insert Figure 1 approximately here

Figure 1. Research framework

Goodwill disclosure is a crucial issue in accounting studies, although literature is incomplete and largely heterogeneous even in well-structured and developed markets (Carvalho et al., 2016). In table 1 some of the contradictions emerging by comparing studies in which similar assumptions are formulated and similar statistical methods are used to test the hypotheses are summarized.

Insert Table 1 approximately here

Table 1. Empirical studies on the determinants of goodwill impairment disclosure in annual reports

The amount of recognized goodwill and impairment losses (column A), considered either in absolute values or as a percentage of total assets, are assumed to trigger a deeper disclosure (Verriest, Gaeremynck, 2009; Casey, O'Mahoney, 2011; Glaum et al., 2013; Bepari et al.,

2014), even if, in Tsalavoutas et al. (2014), the results reveal the existence of a negative relation; meanwhile, in Devalle and Rizzato (2013), the same hypothesis is not supported.

Size is generally assumed to positively influence the level of goodwill disclosure (D'Alauro, 2014, Jenkins, Pevzner, 2015), but this hypothesis is not supported in all cases (column B). As previously argued by Khlifi and Bouri (2010) contradicting results in this case may be a consequence of the variable used to measure company size.

Concerning profitability (column C), the ambiguity remarkably increases because the general assumption that better performances drive deeper disclosure is often not supported. Additionally, when the results show the existence of a significant relation between profitability and disclosure, the direction of the relation is often uncertain (Prencipe, 2004).

In several articles, the determinants associated to corporate governance are expected to significantly influence the extent of disclosure, in particular when accounting standards allow managers to exert a significant degree of subjectivity, as in the case of goodwill impairment. The association between corporate governance and disclosure is tested trough several explanatory variables such as the percentage of independent directors, the segregation of duties between the chief executive officer and the president of the board of directors, the presence of an audit committee, or the number of meetings yearly held by the audit committee, and so forth. Reasonably, independence, structured and frequent control mechanisms and effective corporate governance practice may positively influence the extent of disclosure, even if these associations are far to be clear in literature (Allegrini and Greco, 2013).

Coherently with the above described theoretical premises, we formulate the research hypotheses summarized in Table 2.

Insert Table 2 approximately here

Table 2. Hypothesis of the empirical research about determinants of goodwill disclosure impairment

It is worth emphasizing that our particular research interest in the present study is to investigate the existence of local relations, which we expect to occur particularly when the existence of a unique monotonic relation is not supported.

As subsequent step, we tested the research hypotheses by mean of a MLR model (see Figure 1) and in parallel with this we drawn the LOWESS smooth curves for all the variables included in our analysis. By comparing the results obtained by MLR and LOWESS, we selected the determinants expected to produce linear and nonlinear effects on disclosure, which have

been considered in a Generalized Mixed Model, in order to include conditioning influences. The process followed is detailed in section 6.

5. Sample, data collection and description of variables

The sample investigated to provide empirical testing to the research hypotheses included 260 non-financial companies listed on the four major European stock exchanges with capitalization higher than 1.000 USD/Millions (measured at 30th November 2016): LSE Group, Euronext, Deutsche Boerse, Nasdaq Nordic Exchanges. The four financial markets represented the 87% of the overall European market capitalization. The sample was stratified by market cap and observations were randomly selected. The sample size has been calculated according to a confidence interval of 90% and a sampling error of 5%.

The sample is heterogeneous in terms of turnover (mean 5,764 €/1,000; std. dev. 23,295) and total assets (mean 10,429 €/1,000; std. dev. 45,306 €/1,000). On average, the incidence of goodwill on total assets is rather relevant (19%).

In table 3 the explanatory variables selected according to the research hypotheses are summarized.

Insert Table 3 approximately here

Table 3. Summary of independent variables

For what concern size, we tested H2 by using multiple explanatory variables in order to collect evidences about how the measure used to test such a hypothesis may influence the significance of results.

As dependent variable a disclosure index (GWDScore) is employed, which is a self-constructed measure of goodwill disclosure. It has been created by following Botosan, (1997) and Prencipe (2004) as a sum of substantially unweighted items, which express the Ias 36 disclosure requirements. The list of disclosure items included in the GWDScore allowed a maximum score of 20. Details about the disclosure index can be provided by the authors upon request. Internal consistency of the GWDScore has been assessed by mean of the Cronbach's alpha coefficient and considering the coherence of the correlations obtained between GWDScore and the traditional disclosure determinants. Data have been partially manual-collected, sourcing by the sample companies' annual reports (descriptive notes to consolidated financial statements), and extracted by the Thompson Reuters Datastream where available. The investigated period was 2015.

6. LOWESS and GAM in place. Description of the method and discussion of results.

6.1 Hypotheses testing through MLR

Following the research framework described in section 4, we estimated a MLR model to test the hypotheses about the existence of a linear relationship between the GWDScore and the expected determinants.

A ranking transformation was applied to all the variables (Lang and Lundholm, 1993), which was suitable since the dependent variable is measured through a score, while the nature of independent variables is mixed (some of them are continuous and infinite variables, some other are continuous but finite, some other are discrete). In such a case the analysis of raw data is likely to produce significant biases, which could explain some of the contradictions emerging in previous studies.

In order to check for multicollinearity, the Variance Inflation Factors (VIF) test was used. The values of SALES and ASSETS indicate that these variables are influencing each other, therefore ASSETS was removed from the model because it is less correlated to the dependent variable. The standardized regression coefficients obtained for all predictors are shown in Table 4.

Insert Table 4 approximately here

Table 4. Standardized regression coefficients

The results showed that SALES is positively associated to goodwill disclosure, in line with general literature on corporate disclosure. Given the large consensus in literature about the influence produced by company size on disclosure, SALES may be considered as control variables in this study. Nevertheless, the LOWESS curves might provide interesting insights also in this case. Conversely the expected association between VALUE and disclosure was not significant, confirming that in some cases the robustness of results may vary according to the explanatory variable adopted to test a research hypothesis.

Also INCGW and INCIMPLOSS, the determinants representing the relevance of recognized goodwill and impairment losses, resulted significantly associated with disclosure; conversely VARGW did not resulted to significantly impact on disclosure.

All the remaining hypotheses about profitability (ROI) and corporate governance (INDDIRECT and TOTBOARD), were not supported.

According to the MLR, the extent of goodwill disclosure is influenced by the following three variables: SALES, INCGW, INCIMPLOSS. As additional check for robustness, the AIC

stepwise method was also used to select the best subset of variables (Yamashita et al., 2007). The resulting model, which is composed by the same three variables, satisfies the basic assumptions of regression analysis, but the relatively low values of goodness-of-fit indices suggest that it fails to incorporate all of the structure in the data; consequently, the linear relationship might represent an oversimplification.

6.2 LOWESS curves and Generalized Additive Model

In parallel with MLR a set of scatterplots has been used for visually examining the bivariate relationship between each predictor and the response variable. A LOWESS curve has been fitted in each plot in order to check for the existence of nonlinearities within the data (i.e. patterns that may not conform to any simple structure).

As mentioned above, the principles and tools of statistical inference (such as ANOVA or F-test) are suitable to test if a LOWESS curve is capturing a significant relationship between X and Y, or conversely if it is not adding any additional insight to the hypothesis of no functional dependence (Cleveland and Devlin 1988, Efron and Tibshirani 1993, Beck and Jackman,1998 Fox 1999, Jacoby 2000). In line with this the F-test was used to determine the appropriate fitting parameters, comparing two nested LOWESS curves and testing if the more complicated one provides a significant improvement in fit over the simpler one (Jacoby, 2000). Once the "best" LOWESS curve for each variable has been obtained, the following two hypotheses were tested (Fox, 1999):

1) the fitted curve describes a significant relationship between the explanatory variable and the dependent variable, against the null hypothesis of no functional dependence;

2) the fitted curve improves the representation of the data with respect to the linear regression model.

Insert Table 5 approximately here

Table 5. F-Tests for bivariate LOWESS curves

The LOWESS curves fitted to INCGW and VARGW are the only for which both the hypotheses were supported, meaning that those variables are expected to produce nonlinear effects on the response variable. It is worth of mention notwithstanding the significance of the results obtained by the MLR, the LOWESS curve for INCGW provided a more accurate representation of the relations, consequently the variable is assumed to produce nonlinear effects on disclosure.

It is worth nothing that VARGW did not result to significantly impact on goodwill disclosure according to the results obtained with MLR and this provide meaning to our study showing that traditional methods may not be effective in detecting complex nonlinear relations.

As a difference, the MLR showed that SALES and INCIMPLOSS produce linear effects on goodwill disclosure, while for the LOWESS, the two conditions for significance were not supported. For all the remaining variables no functional dependence emerged, neither of linear nor nonlinear type.

The two couples of determinants that have linear (SALES, INCIMPLOSS) and nonlinear (INCGW, VARGW) effects respectively, have been then considered for the construction of a Mixed Generalized Additive Model (Beck and Jackman,1998), which is structured as follows:

$$y_i = \alpha + \sum_{l=1}^{m} \beta_l \, \mathrm{Z}_{i,l} + \sum_{j=1}^{k} f\left(X_{ij}\right) + \varepsilon_i$$

where the first and the second part of the equation are used to model the linear and nonlinear effects respectively.

Multiple checks for robustness were performed using standard F-test, which are suitable when two nested models are compared (Fox, 1999; Jacoby 2000).

As first we compared the mixed model with a more complex model including LOWESS for all the determinants, in order to test whether any nonlinearities, neglected in the bivariate context, emerge in a multivariate case. The results showed that the model including only LOWESS curves did not produce any improvements in fit (see Table 6).

Insert Table 6 approximately here

Table 6. F-test for comparing nested models

Afterwards, we used the F-test to check whether the mixed model improves the representation of the data, over the simpler multiple regression model that can be considered nested within the former.

The p-value associated with the F test statistic is less than 0.05, thus the null hypothesis of no improvement in fit can be rejected. Stated somewhat differently, the additive model is preferable to the multiple regression model, providing a better understanding of the relationship between the response variable and the predictors.

Therefore, the scatterplots obtained from this model can be used to describe the relationships between predictors and dependent variable effectively.

Insert Figure 2 approximately here

Figure 2. GAM Scatterplot. VARGW *net effect*

The LOWESS about VARGW (see figure 2) showed the existence of at least three different local relations whose rank intervals are shown in table 7.

Insert Table 7 approximately here

Table 7. VARGW rank-transformed and raw observations

The first relation involves values ranging from high negative variations of goodwill to null variation, showing that the disclosure decreases as the intensity of negative variations decreases. The second local relation show that disclosure gradually increases as small positive variation of goodwill occur. We may interpret these two tendencies as coherent with the prescriptions of the signalling theory, even if sign of the relation is inverted, since if we look at the % variations in absolute values we may generalize that disclosure increases as the variation increases. In contrast with that, the third local relation shows that the extent of disclosure decreases as increasing positive variations of goodwill are recognized suggesting that managers are less inclined to disclose information about goodwill when the amounts recognized increase substantially after a given threshold.

The analysis of the scatterplot for INCGW shows how the method in discussion may provide valuable insights also for also the existence of significant linear relations was supported (see figure 3).

Insert Figure 3 approximately here

Figure 3. INCGW LOWESS curve

The net effect for INCGW shows that the relation is not completely linear. In particular, a positive relation between INCGW and the extent of disclosure is assumed to exist in a first range of the dependent variable that corresponds approximately to the first half of the rank transformed data (see Table 8).

Insert Table 8 approximately here

Table 8. INCGW rank-transformed and raw observations

Within the second interval of observations the association between the extent of goodwill disclosure and the incidence of goodwill is weaker (the shape of the curve is substantially flat). This shows that the significance of the regression in mainly conditioned by the first range of values. The results contribute to theory providing evidence that managers are more inclined to improve goodwill disclosure, as percentage of goodwill on total assets increases, but this is true only within a given threshold. When the relevance is over that threshold greater relevance of goodwill seem not to impact on the inclination of managers to disclose additional information.

Generally, the 95% confidence bands (shown by dotted lines in figures 2 and 3) are plotted along with the LOWESS curve, providing a visual guide as to whether the nonlinearity of the curve is simply due to sampling fluctuations, or the fitted function is really different from a linear fit (Beck and Jackman, 1998). In figure 2 the confidence bands for VARGW are closer in the center of the figure, meaning that a low variability of estimations occurs. In figure 3 the distance between the dotted lines is constant all along the curve, and meaning that variability of the estimation remains constant.

Summarizing, for both VARGW and INCGW the results obtained by the GAM provide evidence that the relations between such variables and goodwill disclosure are quite complex and multifaceted, like in the case of VARGW, whose relation of dependence has not been detected by the MLR. Furthermore, it's worth noting that such a curve has been obtained without making any *a priori* assumption about the functional form. Consequently, the results may contribute to literature showing that financial disclosure is conditioned by local relations. In those cases, a descriptive model built coherently with the assumption of a unique theoretical framework may lead to an oversimplification of the investigated phenomenon.

7. Conclusions

Extant studies on corporate disclosure and its determinants show uncertain and contradictory results that according to several authors are due to methodology shortcomings, which in some cases fail in providing an effective description of the relations between disclosure and its expected determinants.

Traditional methodologies generally assume the existence of a global relation (whose shape is supposed to be linear or non-linear) between dependent and independent variables.

Coherently with this approach, the research hypotheses are confirmed when the empirical evidence shows that the relation is significant for the whole range of observations.

In our methodological contribution, we start by the assumption that in some cases the extent of disclosure is influenced by relations of multiple kind (i.e. local relations), which are relevant within specific subsets of the independent variables. Consequently, the traditional empirical methods may be usefully complemented by data-driven inductive approaches which allow to detect complex relations of any kind, without requiring preliminary assumptions about the structure of data and the shape of the relation lying beneath the data.

In particular, we propose the use of LOWESS smooth curves in combination with the Generalized Additive Model (GAM). The former are derived directly from a rank-transformed set of data and allow scholars to check for the existence of meaningful drivers, which are associated with disclosure by mean of nonlinear relations. The GAM allows to put together the expected impact produced by all the drivers both the linear- and nonlinear-related with disclosure into a unique model. As result by the GAM a set of scatterplots is obtained, representing the net effect produced by a variable on disclosure after that all other impact produced by the remaining drivers have been removed.

The methodology is suitable in financial disclosure studies for several reasons. First of all, it is easy to implement, since it is not necessary to develop a general descriptive function a priori. Additionally, the methodology proposed is easy to interpret, thus making the method applicable to a broad range of problems.

The adoption of LOWESS and GAM is coherent with an inductive research approach, where empirical evidence is generalized and evolves into theoretical explanations. This approach may be particularly suitable when complex phenomena are investigated, which may result oversimplified if modelled assuming the existence of a global relation.

To demonstrate the methodology, we performed an analysis on determinants of goodwill disclosure, by assuming that a local regression approach is likely to improve the description of the investigated phenomenon.

The results supported our expectation and showed that in some cases a variable may influence the extent of disclosure in a complex and multi-faceted relation, as in the cases of the % variation of recognized goodwill and the incidence of goodwill on total assets.

It is worth to mention that for what concern the % variation of recognized goodwill, the proposed methodology was able to detect the existence of a curvilinear relation, which was completely neglected by the MLR. As a difference the relation described for the incidence of

goodwill on total assets (INCGW) resulted more accurate than what the MLR allowed to represent.

Summarizing, this method can be useful to explicate the proper functional specification of relationships between variables particularly when the theory suggests the prevalence of nonlinearities (Jacoby, 2000), or during the exploratory stages of a research, when there is consensus about the variables relevant to explain a relation, while is not clear the kind of pattern lying behind data.

It is a major weakness that the LOWESS curve requires the analyst to make partially arbitrary decisions about the fitting parameters, whose values are determined by a process that is based upon visual inspection of the scatter plots (basic or residual plots).

Nevertheless, the application of the methodology to different data sets might offer scholars additional insights about the existence of local relations. This could be particularly significant in financial disclosure studies to help scholar explain some of the contradictions emerging from previous studies. Nevertheless, the adoption of LOWESS and GAM is generalizable all the field of studies where contradictions emerging by extant studies may require the adoption of tools able to detect complex and non-monotonic relations.

References

Allegrini, M. and Greco, G. (2013). "Corporate boards, audit committees and voluntary disclosure: evidence from Italian Listed Companies". *Journal of Management and Governance*, 17(1), 13, 187-216.

Armitage, S. and Marston, C. (2008). "Corporate disclosure, cost of capital and reputation: evidence from finance directors". *The British Accounting Review*, 40(4), 314-336.

Beck, N. and Jackman, S. (1998). "Beyond linearity by default: generalized additive models". *American Journal of Political Science*, 42(2), 596–627.

Bepari, K., Rahman, S.F. and Mollik, A.T. (2014). "Firms' compliance with the disclosure requirements of IFRS for goodwill impairment testing: Effect of the global financial crisis and other firm characteristics". *Journal of Accounting & Organizational Change*, 10(1), 116-149.

Botosan, C.A. (1997). "Disclosure Level and the Cost of Equity Capital". *The Accounting Review*, 72, 323-349.

Carvalho, C., Rodrigues, A.M. and Ferreira, C. (2016). "Goodwill and Mandatory Disclosure Compliance: A Critical Review of the Literature". *Australian Accounting Review*, 26(4), 376-389.

Casey, J. and O'Mahoney, O. (2011). "An Analysis of the Determinants of Voluntary Impairment Disclosure Post the 2008 Credit Crisis". *Irish Accounting and Finance Association Annual Conference, at http://www.repository.wit.ie/2643/1/Determinants of Voluntary Impairment Disclosure.*

Chambers, J.M., Cleveland, W.S., Kleiner, B. and Tukey, P.A., (1983). *Graphical methods for data analysis*. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA, USA.

Chavent, M., Ding, Y., Fu, L., Stolowy, H. and Wang, H. (2006). "Disclosure and Determinants Studies: An Extension Using the Divisive Clustering Method (DIV)", *European Accounting Review*, 15(2), 181–218.

Cleveland, W.S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots". *Journal of American Statistical Association*, 74(368), 829-836.

Cleveland, W.S. (1993). *Visualizing Data*. Hobart Press.

Cleveland, W.S. (1994). "Coplots, nonparametric regression, and conditionally parameter fits". *Lecture Notes-Monograph Series*, 21-36.

Cleveland W.S., and Devlin S.J. (1988) "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting" Journal of the American Statistical Association, 83, 403, 596-610

Cooke, T.E. (1998). "Regression Analysis in Accounting Disclosure Studies". *Accounting and Business Research*, 28(3), 209-224.

D'Alauro, G. (2014). "The impact of IAS 36 on goodwill disclosure: Evidence of the write-offs and performance effects". *Intangible capital*. 9(3), 754-799.

Devalle, A. and Rizzato, F. (2013). "The determinants of mandatory disclosure of goodwill". An empirical Analysis". *3rd Annual International Conference of Accounting and Finance, Singapore.*

Donnelly, R. and Mulcahy, M. (2008). "Board structure, ownership, and voluntary disclosure in Ireland". *Corporate Governance: An International Review*, 16(5), 416-429.

Efron, B., Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Chapman and Hall, New York.

Ernst & Young (2006). *IFRS: observations on the implementation of IFRS*. London: Ernst & Young.

Fox J. (1999). *Nonparametric regression analysis*. Typescript, McMaster University.

Glaum, M., Schmidt, P., Street, D.L. and Vogel, S., (2013). "Compliance with IFRS 3- and IAS 36-required disclosures across 17 European countries: company- and country-level determinants". *Accounting and Business Research*. 43(3), 163-204.

Hall, A.R. (2005). *Generalized method of moments*. Oxford University press.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized additive models*. Chapman and Hall, New York.

Jacoby, W.G., (2000). "Loess: a nonparametric, graphical tool for depicting relationships between variables". *Electoral Studies*, 19(4), pp. 577-613.

Jenkins N.T. and Pevzner M. (2015). "Discretionary disclosure of goodwill slack: determinants and consequences", *working paper, University of Kentucky. http://www.researchgate.net/publication/280528020*

Khlifi, F. and Bouri, A. (2010). "Corporate Disclosure and Firm Characteristics: A Puzzling Relationship". *Journal of Accounting, Business & Management*, 17(1), 62-89.

Lang, M. and Lundholm, R., (1993). "Cross-Sectional Determinants of Analyst Ratings of Corporate Disclosures". *Journal of Accounting Research*, 31, 246–271.

Lang, M., Lins, K. and Miller, D., (2004). "Concentrated Control, Analyst Following and Valuation: Do Analysts Matter Most When Investors Are Protected Least?" *Journal of Accounting Research*, 42(3), 589–623.

Malone, D., Fries, C. and Jones, T. (1993). "An empirical investigation of the extent of corporate financial disclosure in the oil and gas industry". *Journal of Accounting, Auditing and Finance*, 8(3), 249-273.

Miller, H. and Hall, P. (2010). "Local polynomial regression and variable selection". In: Berger, J.O., Cai, T.T., Johnston, I.M. (Eds), *Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown*, pp. 216-233. Institute of Mathematical Statistics.

Ng, E. and Koh, H. (1994). "An agency theory and probit analytic approach to corporate non-mandatory disclosure compliance". *Asia-Pacific Journal of Accounting*, 1(1), 29-44.

Parviainen A., Hannu J. and Dallas R., (2001). "On the non-linear relationship between disclosure and its determinants". *Applied Economics Letters*, 8(11), 747-750, 2001.

Prencipe, A. (2004). "Proprietary Costs and Determinants of Voluntary Segment Disclosure: Evidence from Italian Listed Companies". *European Accounting Review*, 13(2), 319–340.

Tsalavoutas I., Andrè P. and Dionysiou D., (2014). "Worldwide application of IFRS 3, IAS 38 and IAS 36, related disclosures, and determinants of non-compliance", ACCA Research Report 134.

Verriest, A. and Gaeremynck, A. (2009). “What determines goodwill impairment”. *Review of Business and Economics*, 54(2), 106-128.

Yamashita, T., Yamashita, K., and Kamimura, R. (2007). “A stepwise AIC method for variable selection in linear regression”. *Communications in Statistics—Theory and Methods*, 36(13), 2395-2403.

Analysis of disclosure determinants: a local-relation approach - Tables

	A. Incidence of goodwill and/or impairment loss	B. Size	C. Profitability	D. Corporate governance	Methodology
Verriest and Gaeremynck, 2009		MK capitalization: **– (S)**	ROA: **+(NS);** Growth MK: **– (S)**	IND/TOT: + (S) Cman-CEO: – (S)	
Casey and O'Mahoney, 2011	Imp. loss: **+ (S)** Imp. loss / TA: **+ (S)**	TA: **+ (NS)**	Profit: **+ (S)**	IND/TOT: – (NS)	
Devalle and Rizzato, 2013	GW/TA: **+ (S)** GW/Equity: **+ (S)** GW: **– (NS)**	Revenues: **+ (S)** Mk capitalization: **– (S)** TA: **+ (NS)**	ROS: **– (NS)**; EBIT: **+ (NS)**; ROE: **+(NS)**		
Glaum et al., 2013	GW/TA: **+ (S)**	AVR (TA, employees, MK capitalization): **+ (NS)**		Audit committee:+ (S)	
Bepari et al., 2014	GW/TA: **+ (S)**	Log (TA): + **(NS)**	ROE: + **(S)**		
D'Alauro, 2014	Imp. loss/TA: **+ (S)**	TA: **+ (S)**	ROE: **+ (S)**; AVG ROE: **+ (S)**	AC Meet: – (NS) AC IND: – (NS)	
Tsalavoutas et al., 2014	Imp. loss: **– (S)**	MK capitalization: **– (S)**	Profit: **+ (NS)**		
Jenkins and Pevzner, 2015		MTBV:**+(NS)** MK capitalization: **– (S)** Fixed assets/TA: **– (NS)**	ROA: **– (S)**		

Table 1. Empirical studies on the determinants of goodwill impairment disclosure in annual reports

(S)/(NS): significant/not significant
GW: goodwill
TA: total asset
IND/TOT: percentage of independent members of the board of directors
Cman-CEO: if the Chairman and CEO are the same person: CEO duality (dummy)
AC Meet: number of meetings of the audit committee held in the year
AC IND: number of independent directors who are members of the audit committee scaled by the total number of directors

	Hypotheses
H1	*The higher the relevance of goodwill in financial statements, the higher the extent of goodwill disclosure*
H1.a	*The higher the % variation of goodwill, the higher the extent of goodwill disclosure*
H1.b	*The higher the incidence of goodwill on total assets, the higher the extent of goodwill disclosure*
H1.c	*The higher the incidence of recognized impairment loss (as a % of the previous year's goodwill), the higher the extent of goodwill disclosure*
H2	*The higher the size, the higher the extent of goodwill disclosure*
H3	*The higher profitability, the higher the extent of goodwill disclosure*
H4	*The higher the quality of corporate governance (in terms of independence and transparency), the higher the extent of goodwill disclosure*
H4.a	*The higher the % of independent directors, the higher the extent of goodwill disclosure*
H4.b	*The higher the number of directors, the higher the extent of goodwill disclosure*

Table 2. Hypothesis of the empirical research about determinants of goodwill disclosure impairment

Hypotheses	Explanatory Variable	Description	Measure adopted/scores assigned
H1.a	VARGW	Goodwill % variation	$(Goodwill_{(y)} – goodwill_{(y-1)})/goodwill_{(y-1)}$
H1.b	INCGW	Goodwill incidence	$Goodwill_{(y)}$/total $assets_{(y)}$
H1.c	INCIMPLOSS	Incidence of impairment loss	Impairment $loss_{(y)}/goodwill_{(y-1)}$
H2	ASSETS	Size	Consolidated total assets
H2	SALES	Size	Consolidated net sales
H2	VALUE	Size	Market Value
H3	ROI	Return on assets	$EBIT_{(y)}$/Total $Assets_{(y)}$
H4.a	INDDIRECT	% of independent directors	Independent Director/Total Nr. Directors
H4.b	TOTBOARD	Nr. of Directors	Number of Directors

Table 3. Summary of independent variables

Hp	Explanatory Variable	Coefficient	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
H1.a	VARGW	0,005	0,067	0,072	0,943	-0,128	0,138
H1.b	INCGW	0,266	0,065	4,096	**< 0,0001**	0,138	0,393
H1.c	INCIMPLOSS	0,182	0,068	2,661	**0,008**	0,047	0,316
H2	SALES	0,402	0,079	5,053	**< 0,0001**	0,245	0,558
H2	VALUE	-0,063	0,069	-0,902	0,368	-0,199	0,074
H3	ROI	-0,026	0,073	-0,358	0,720	-0,170	0,118
H4.a	INDDIRECT	0,030	0,064	0,458	0,647	-0,098	0,157
H4.b	TOTBOARD	-0,079	0,076	-1,042	0,299	-0,230	0,071

Table 4. Standardized regression coefficients (* *p-value < 0.05)*

Hp	Explanatory Variable	F-Test (vs no dependence)	F-Test (vs linear regression)
H1.a	VARGW (λ=2; α=0.5)	0,000*	0,000*
H1.b	INCGW (λ=2; α=0.6)	0,000*	0,000*
H1.c	INCIMPLOSS (λ=1; α=0.5)	0,013*	0,915
H2	SALES (λ=1; α=0.5)	0,000*	0,109
H2	VALUE (λ=2; α=0.5)	0,253	0,161
H3	ROI (λ=2; α=0.5)	0,075	0,040*
H4.a	INDDIRECT (λ=1; α=0.5)	0,867	0,231
H4.b	TOTBOARD (λ=1; α=0.5)	0,191	0,064

Table 5. F-Tests for bivariate LOWESS curves

Comparison	F-test	p-value
Mixed model vs GAM (all LOWESS)	1,063	0,372
MLR vs Mixed model	2,921	0,005*

Table 6. F-test for comparing nested models

VARGW	
Rank interval	**Raw interval**
1st - 115th	-∞%; 0%]
115th - 200th	]0%; +10%]
200th - 260th	]10%; +∞%

Table 7. VARGW rank-transformed and raw observations

INCGW	
Rank interval	**Raw interval**
1st - 130th	0%; 13%]
126th - 260th	]13%; 75%

Table 8. INCGW rank-transformed and raw observations

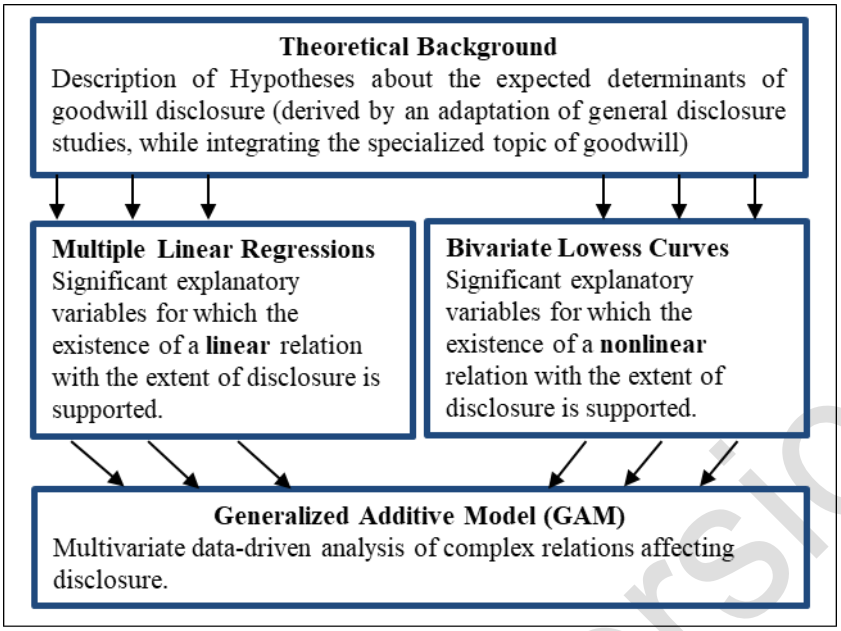

Figure 1. Research framework

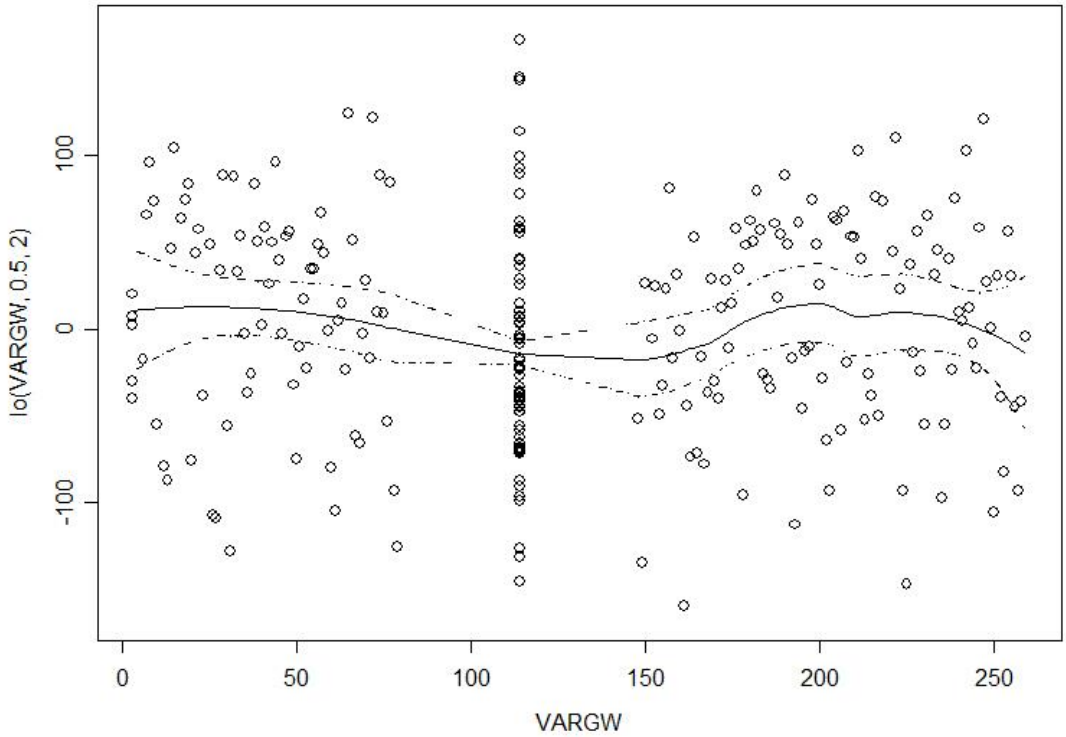

Figure 2. GAM Scatterplot. VARGW *net effect*

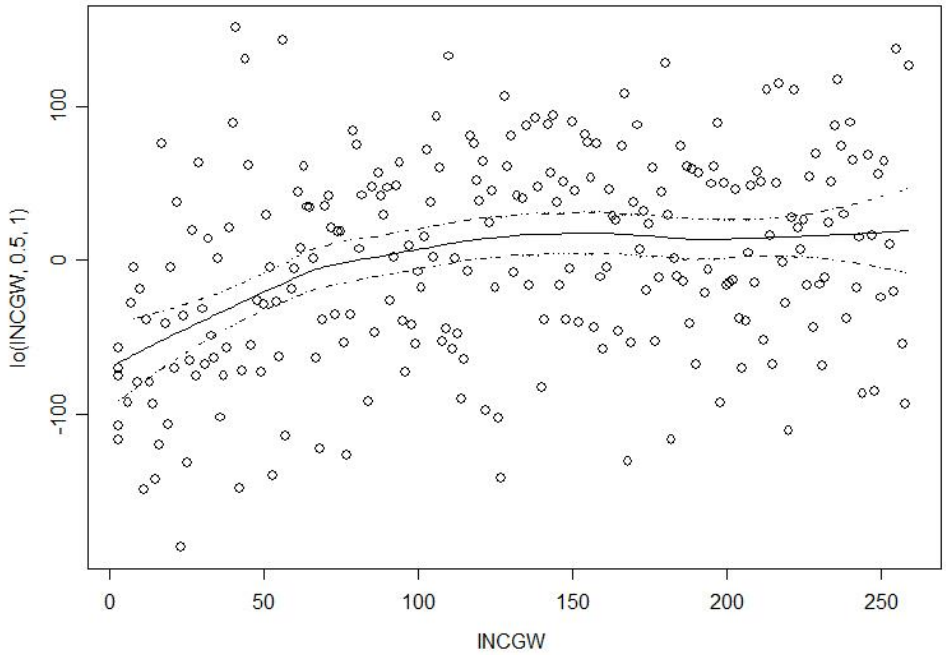

Figure 3. GAM Scatterplot. INCGW net effect