SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN

WORKING PAPER SERIES

Nonparametric Decomposition of Time Series Data with Inputs

by

Edward P. Santos[1] and Erniel B. Barrios[2]

UPSS Working Paper No. 2010-11
September 2010

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

Abstract

[1] Instructor, School of Statistics, University of the Philippines Diliman
[2] Professor, School of Statistics, University of the Philippines Diliman

The backfitting algorithm commonly used in estimating additive models is used to decompose the component shares explained by a set of predictors on a dependent variable in the presence of linear dependencies (multicollinearity) among the predictors. Multicollinearity of independent variables affects the consistency and efficiency of ordinary least squares estimates of the parameters. We propose an estimation procedure that address this problem by estimating shares of each predictor one at a time, the sequence depends on the initial guess of the relative importance of the variables in the model.

Simulated data show that backfitting the ordinary least squares procedure and additive smoothing splines are comparable and superior over the ordinary least squares in estimating share of the contribution of the different predictors to the dependent variable. The superiority of the predictive ability of the method is also more apparent as multicollinearity worsens. The method is used in modeling sales data with predictors including marketing activation indicators, competition measures, distribution indicators, economic and weather indicators. The sales data (time series) that is characterized by severe multicollinearity and inadequate linear fit illustrates the advantage of additive model estimated through backfitting with spline smoother over ordinary least squares and the backfitted linear smoother in terms of predictive ability (MAPE) and interpretability of the estimated shares of the predictors to the dependent variable.

Keywords: multicollinearity, ordinary least Squares, additive Models, backfitting, nonparametric regression

1. Introduction

The utility of a model depends on how realistic it captures the dynamics in which the independent variables influence the dependent variable. It is also important that the predictors whom the modeler has direct control of are included so that it can be used in optimal control. In business, it is important for a model to capture various internal factors like indicators of marketing efforts and external factors like economic indicators, such model can be used in determining the optimal level of internal factors conditional on the levels of external factors. This simulation can be facilitated with a decomposition of how much each of the internal and external factors contributes to the magnitude of the dependent variable is available.

In modeling of continuous dependent variables (e.g., sales), the least squares procedure is commonly used in the construction of a mathematical in which the dependent variables interacts with the predictors. Ordinary least squares theory (OLS) provides estimates of the regression coefficients for each predictor X_i in the model $\underline{Y} = \underline{X\beta} + \underline{\varepsilon}$ by minimizing the error sum of squares $\left\| \underline{Y} - \underline{X\beta} \right\|^2$. Subsequently, the fitted model can assist in the computation of the relative share of each variable to the magnitude of the dependent variable. The OLS estimate of the regression parameter $\underline{\beta}$ is given by $\underline{\hat{\beta}} = \left(\underline{X'X}\right)^{-1} \underline{X'Y}$. This requires the inversion of the inner product of the design matrix X, whose stability relies on whether the design matrix is orthogonal or not. As the design matrix becomes non-orthogonal, ill-conditioning on $\underline{X'X}$ is magnified resulting to excessively large elements of $\left(\underline{X'X}\right)^{-1}$. OLS suffers severely with increased levels of association and departure from orthogonality of the predictors. The problem of multicollinearity distorts the size and direction of the model parameters and affects prediction as well. Multicollinearity occurs very often in business models and other applications due to the inherent association among natural predictors of the dependent variable. Having two predictors with values that move together would automatically result in a near-linear dependency, even if it is only due to the common effect of a latent variable (that may or may not be included in the set of predictors). Estimation of the effects and component shares of these regressors to the dependent variable becomes complicated.

The simplest and most commonly used remedy for multicollinearity is variable deletion. The modeler first identifies the source of the linear dependency and deletes (one or more) independent variables to remove the association, alleviating the possible non-invertibility of $\underline{X'X}$. However, this forfeits the presence of a modeling framework that ought to explain the theoretical basis of certain relationships and assumes that the effects of the excluded predictors are negligible.

Ridge-type estimators and principal component regression (PCR) are also used as a modeling strategy when $\underline{X'X}$ becomes ill-conditioned. Ridge regression adjusts $\underline{X'X}$ into $\underline{X'X} + k\underline{I}$ (with k a parameters and $\underline{I}$ an identity matrix) to avoid perfect linear dependencies of the columns of $\underline{X'X}$. This technique however shows mixed results against OLS under different scenarios, partly due to the method of selection of k. PCR meanwhile produces indices that are hard to interpret and are far away in form from the theoretical expectations on the component estimates as they are given in the original model. Also, PCR doesn't provide leeway for better estimation of predictor-shares that are deemed to be more important than other predictors. Management and policy focus are generally not equally divided among all the determinants, as some prognostic factors are often given more weight in decision making than others. Both ridge regression and principal component regression produces estimates that are biased. The amount of bias in ridge regression is proportional to k, unlike in principal component regression, where the bias is dependent on the last few components excluded as regressors.

Suppose that sales as the response variable and exogenous variables (maybe internally controllable or not) are available. An ideal marketing mix is desirable, i.e., which among those controllable factors like availability of the product, exposure time on television, etc., should be prioritized. This will require estimation of the contribution share of the different determinants to aggregate sales.

We propose to use a class of additive models (Hastie and Tibshirani, 1990) to produce estimates of the components explained by a set of predictor variables on a single continuous response with muticollinearity present among the independent variables. In particular, the backfitting algorithm as a common method for estimating additive models (Hastie and Tibshirani, 1990) is used in estimating the component shares of the different predictors.

When the effects of each predictor are analyzed and set aside one-at-a-time, the problem of attribution of common shares by two or more predictors can be avoided. The decomposition method for the effects incorporates the option of using a flexible nonparametric form for the component effects, increasing the optimality of estimation. This is true since forecasts are not bound by the restrictive form of the traditional linear model. It is also possible to incorporate the relative importance of the regressors to each other in explaining the response Y into the procedure through their priority of entry into the backfitting algorithm. In this manner, estimated shares of indices that are deemed to be of more importance in policy-making can be obtained with more accuracy and confidence relative to other predictors of Y. In the procedure, the possible constraint of additivity of effects can be easily avoided by a flexible form (nonparametric) of each component explained by the X_j's; i.e., the function f_j that links the j^{th} regressor to Y.

2. Some Modeling Issues and Strategies

The presence of linear relationships among predictors of a dependent variable masks the nature of the regression coefficients in OLS, often, parameter estimates contradict theory. The repressive proportional-to-β form of the components binds the gains brought about by assuming a linear model which is supposed to induce simplicity in the interpretation of share estimates. The assumption of additivity of effects would only be a small price to pay for poor model specification if the component shares of the independent variables are estimated as close as possible to their true values. Lauridsen and Mur (2006) summarizes the consequences of multicollinearity on forecasting and inference (unpredictable estimation results, high standard errors, coefficients in the incorrect direction, and unfeasible sizes of parameters). Also as cited in Lauridsen and Mur (2006), several works illustrate the asymmetry of multicollinearity's effects: Chatterjee and Hadi (1988) and Belsley et.al. (1980) show that sharp linear dependencies affect the $\hat{\beta}$'s in OLS but not necessarily the predicted values and residuals. There are also instances where the number of regressors and the degree of multicollinearity itself directly affect the estimates, e.g., Farrell (2002) observed that estimates reacted directly to the number of predictors and the level of linear dependencies.

There are many solutions proposed in the literature to address the multicollinearity problem. Ridge regression is among the early solutions as proposed by Hoerl and Kennard (1970), where the OLS estimate is modified as $\underline{\hat{\beta}}(k) = (\underline{X'X} + k\underline{I})^{-1}\underline{X'Y}$ where $k \geq 0$. While the estimator incurs bias, the gain is in terms of mitigating the ill-conditioning of X'X that can lead to unstable OLS estimates. Several modifications of ridge regression were proposed, e.g., Yalian and Hu (2008) introduced a random component into the ridge specification leading to a stochastic mixed ridge estimator. The mixed specification is advantageous for large values of the ridge parameter, but the reverse happens as k decreases. Similarly, Lipovetsky (2006) introduced a revised version of a two-parameter ridge model and defined the net effect of a predictor X_j to Y as $\beta_j(X'Y)_j$ and showed that these quantities can be interpreted well in the ridge model unlike in OLS. The net effect quantity measures the share of a regressor X_i on the coefficient of determination or total variability explained (R^2). Defining the component share or effect in this manner though shows that OLS generally will have better forecasting power, and translates negative component effects as being uninterpretable (Lipovetsky, 2006). This indicated the need for another quantity to summarize the effects of a regressor that incorporates the model parameter β_j at the same time.

Baye and Parker (1984) meanwhile developed the division of $r-k$ of estimators that include OLS, ridge regression, and PCR as special cases and compared their respective performances by their MSE's theoretically. The functional similarities in form and performance of ridge regression and PCR are also noted by Lipovetsky and Conklin (2001). Later, Sarkar (2006) derived the theoretical bases for the increased efficiency of this general type of estimator over OLS using a mean squared error matrix. However, these theoretical results have yet to be validated in practice.

Kosfeld and Lauridsen (2008) compared the performance of factor analysis regression (FAR) and PCR relative to OLS in the presence of variables with varying strengths of multicollinearity by Monte Carlo simulations. FAR and PCR provided favorable estimates at strong and extreme multicollinearity, with FAR superior at moderate levels. However, this is true only for cases where the effects of the independent variables are near identical to each other. Furthermore, FAR and PCR variations tend to equalize the effect of predictors that belong to the same factor group (Kosfeld and Lauridsen, 2008), and doesn't distinguish between component shares with high specificity. Similarly, in analyzing the use of sparse principal components in nonstationary time series, Lansangan and Barrios (2009) presented this same equalizing effect of PC regression on the component loadings of different variables. This is disadvantageous in cases where some indices that affect the response have distinctively large contributions to Y in the true decomposition, notwithstanding the importance of a predictor relative to the other independent variables in the model. These varied results show that multicollinearity cannot be solved extensively through various methodologies.

Additive models are extensions of the classical linear model specification $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon$ and are previously viewed as preliminary and descriptive tools. Hastie and Tibshirani (1990) argued that the concept of additive models serving as an initial procedure to locate the patterns and behavior of the predictors relative to the response, suggesting a possible parametric form for which to model Y at a later stage. An additive model can be specified as $Y = \alpha + \sum_{j=1}^{p} f_j\left(X_j\right) + \varepsilon$, where the individual functions f_j representing the link between the j^{th} predictor and Y. Note that each f_j can be distinct from the others, with each regressor linked to the response in a different way. This removes the constraint present in the linear model, where each X_j component to be decomposed is proportional to $\beta_j X_j$. Also, the representative functions do not have to be specified fully and can be nonparametric. This adds flexibility in estimating the relationship between Y and X's. A proper and effective design for the f_j's would bring out an increased gain in estimation of the decomposed predictor effects and would more than compensate for the additive approximation to the true regression surface $f\left(X_1, X_2, \cdots, X_p\right)$.

To estimate and additive model, backfitting is commonly used. The Backfitting Algorithm's iterative procedure proceeds as follows (Hastie and Tibshirani, 1990):

(i) Initialize: $\alpha = \text{average}(y_i), f_j = f_j^0, j = 1, \ldots, p$

(ii) Cycle: $j = 1, \ldots, p, 1, \ldots, p, \ldots$ (1)

$$f_j = S_j\left(\mathbf{y} - \alpha - \sum_{k \neq j} f_k \mid x\right)$$

(iii) Continue (ii) until the individual functions stabilize.

where the f_j's are the true functions that represent the component or share that is being explained by the j^{th} regressor. The individual estimators S_j are smoothers used to quantify the effect of the j^{th} regressor. Note that in the representation, S_j can have any functional form, e.g., $X_j \beta_j$ in OLS, or a histogram smoother in nonparametric regression. Also, the exact form of the S_j's need not be specified (possibility nonparametric). The class of cubic smoothing splines for the S_j's provide solution to the penalized least squares (Hastie and Tibshirani, 1990) criterion:

$$\sum_{i=1}^{n}\left\{y_i - f(x_i)\right\}^2 + \lambda \int_a^b \left\{f''(t)\right\}^2 dt \,. \tag{2}$$

Then the function $f(\cdot)$ that minimizes this criteria is the cubic spline estimate of the true link between Y and X_i. The solution $f(\cdot)$ is a balance between the smoothness of the estimates and model fit. The choice of λ is determined via the generalized cross validation criterion (GCV), (Hastie and Tibshirani, 1990). The general form of the GCV is

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{y_i - f_\lambda(x_i)}{1 - tr(S_\lambda)/n}\right\}^2 . \tag{3}$$

This entails creating a jackknifed fit at each value x_i of X and computing for prediction and squared errors, leaving out one observation at a time (Hastie and Tibshirani, 1990). $f_\lambda(x_i)$ represents the smooth estimate at x_i with S_λ representing the smoothing matrix (Hastie and Tibshirani, 1990). The common choices for the spline knot locations are the data points (x_i, y_i) themselves. For backfitting OLS, the partial residuals are modeled at each iteration using X_j as the lone independent variable by least squares. This is different from the classical OLS procedure since the components of each variable are not estimated simultaneously, but iteratively using the partial residuals (excluding the effect of X_j). In the backfitting algorithm, step (iii) rarely presents cases of non-convergence of the effects $\hat{f}_j$ (Hastie and Tibshirani, 1990).

3. Decomposition Method

Given a regression model

$$\underline{Y} = \underline{X\beta} + \underline{\varepsilon} \quad (4)$$

Let $\underline{X} = (\underline{X}_1, \underline{X}_2)$ and $\underline{\beta} = (\underline{\beta}_1, \underline{\beta}_2)$. If $\underline{X}_2$ is ignored and $\underline{Y}$ is regressed on $\underline{X}_1$, the OLS estimate of $\underline{\beta}_1$ is biased if $\underline{X}_1' \underline{X}_2 \neq 0$. This implies that our understanding on the contribution of each predictor is contaminated once they possess some overlaps on the proportion of $\underline{Y}$ they ought to explain. The contribution of the predictors $X_1, X_2, \ldots, X_p$ on the dependent variable Y is an important information that help facilitate optimal control and decision-making. Suppose Y is sales and X_1 is expenditure on television advertisements, while X_2 is expenditure on mall activation. These indicators are usually time series realization of a stochastic process. A business strategist would be interested to know which strategy sales is most sensitive to. This will be an important input in formulating an optimum marketing mix, i.e., prioritization between television advertisements versus mall activation. The standardized estimates of the regression parameters suffice to produce the information required in the formulation of the marketing mix. In many business and economic settings where indicators are monitored over time, there is a tendency for these indicators to drift simultaneously as they progress over time, see for example, Lansangan and Barrios (2009). In this case, multicollinearity problem yields undesirable effects on the least squares estimates, e.g., unstable or inefficient estimates of the regression coefficients. Consequently, the estimated share of each predictor is also affected.

Suppose that an outcome indicator y_t is to be decomposed according to the share of the different predictors $x_{1t}, x_{2t}, \ldots, x_{pt}$. We would like to decompose y_t according to the additive model

$$y_t = f_1(x_{1t}) + f_2(x_{2t}) + \ldots + f_p(x_{pt}) + \varepsilon_t \quad (5)$$

The nonparametric nature of (f_j) will mitigate the ill-condition that a linear model usually suffer as a consequence of linear dependencies among the dependent variables. The methodology facilitates the use of a class of additive models (Hastie and Tibshirani, 1990) in decomposing the effects of predictors (possibly collinear) on y_t. The steps in partitioning the components are as follows:

1. Determine the priority of inclusion of the independent variables for entry to the backfitting algorithm. Pairwise correlation between the dependent variable and the predictors can be used (predictor with higher correlation with the dependent variable comes first). It is also possible to consider the degree of importance of the variable in optimally controlling for the variation of the dependent variable, e.g., marketing expenditure is more important than distribution of competition, thus, it goes first in the iteration of the backfitting algorithm.
2. Estimate the component explained by X_j (f_j) in the additive model (5) based on the smoother S_j, i.e., determine $\hat{f}_j$. In each iteration of the backfitting algorithm, S_j smoothes the partial residuals $\left(\underline{Y} - \hat{\alpha} - \sum_{k \neq j} \hat{f}_k\right)$, excluding the variable X_j in the computation.

3. Continue cycling through steps (2) and (3) of the algorithm until the estimates of the shares stabilize. Convergence of estimates is assumed if the value of $\hat{f}_j$ do not change more than δ% from the last iteration.
4. Compute for the proportion of estimated share of each predictor on the dependent variable. The estimated share of X_j is defined as:

$$\text{Estimated Share}(X_j) = \frac{\sum_{i=1}^{n}\left(\frac{\hat{\beta}_j x_{ji}}{Y_i}\right)}{n} \tag{6}$$

from the linear model in (4), or

$$\text{Estimated Share}(X_j) = \frac{\sum_{i=1}^{n}\left(\frac{\hat{f}_{ji}}{Y_i}\right)}{n} \tag{6a}$$

from the nonparametric model in (5).

The definition of shares above is advantageous since the component explained by a variable is the average of the j^{th} additive effect over all observations. It incorporates both the range of X_j and the parameter β_j (if the true model is linear). It is interpretable regardless of the size and magnitude of the independent variables as well, even under multicollinearity.

Casals, et al (2010), decomposed a state-space model into the contributions of the different inputs. The result is used in estimating the return on investments (ROI) of advertising.

4. Simulation Study

A simulation study is conducted under the assumption that there are groups of correlated indices that affect the response variable in the model. The basic specification has the form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + k\varepsilon_i \tag{7}$$

where

$$X_1, X_2, X_3 \sim Uniform(a,b)$$

and

$$X_{4i} = \beta_1^{'} X_{1i} + k_1 \varepsilon_{1i}, \varepsilon_1 \sim N(\mu_1, \sigma_1^2) \tag{8}$$

$$X_{5i} = \beta_1^{''} X_{2i} + \beta_2^{''} X_{3i} + k_2 \varepsilon_{2i}, \varepsilon_2 \sim N(\mu_2, \sigma_2^2) \tag{9}$$

There are three primary predictor variables are considered: X_1, X_2, and X_3. The pairwise correlations among X_1, X_2, and X_3 are minimized (smaller than 0.3) to simulate a set of distinct and weakly dependent regressors. Two groups of multicollinearity are induced, one in X_4 and the other in X_5. X_4 is assumed to depend on X_1 alone while X_5 depends on X_2 and X_3. This captures events where the dependencies between the independent variables are present over several groups. Without loss of generality, the error terms ε_1, ε_2, and ε are assumed to have a $N(0,1)$ distribution. k is used to control model fit of (7) while k_1, and k_2 are used to control the extent of multicollinearity among the predictors. Table 1 summarizes the simulation parameters for both the cross-section and well as time series data.

Table 1. Simulation Parameters

	Cross Sectional Data	Time Series Data
K	1 (good model fit) 10000 (moderately good model fit) 1000000(poor model fit)	1 (good model fit) 1000 (moderately good model fit) 1000000 (poor model fit)
k_1	1 (strong multicollinearity) 100 (moderate multicollinearity) 500(weak/no multicollinearity)	1 (strong multicollinearity) 100 (moderate multicollinearity) 500(weak/no multicollinearity)
k_2	1 (strong multicollinearity) 100 (moderate multicollinearity) 500(weak/no multicollinearity)	1 (strong multicollinearity) 100 (moderate multicollinearity) 500(weak/no multicollinearity)
size/length of sample	20, 100	40, 400

The range of X_1, X_2, and X_3, they are set to take values from -100 to 100. Different parameters are controlled to assess the performance of the additive modeling procedure. Of extreme importance are k_1 and k_2 respectively. These control the level of multicollinearity present (1 for strong, 100 for moderately strong, and 500 for weak multicollinearity), as they indicate the size of the error terms in the dependencies of X_4 on X_1 and of X_5 on X_2 and X_3. It is also of interest to assess the performance of additive modeling and backfitting under small and large sample sizes (20/40 as small/short and 100/200 as large/long, respectively). In addition, the inclusion of a constant *k* as a multiplier to the model error measures model fit (also, possible departures from additivity).

In the context of time series data, the model in (7) is still maintained, but this time with the behavior of X_1, X_2, and X_3 governed by an AR(1) process where

$$(1-\phi B)X_t = a_t \qquad (10)$$

and $a_t \sim N(0,1)$ with B as the backshift lag operator (Wei, 2006). The value of ϕ influences stationarity of the series $\{X_t\}$, i.e., (10) is stationary if $|\phi|<1$. In the simulation study, ϕ is assigned the value of 0.5. Stationarity is targeted in the simulation to control the present dependencies in the model to appear only through X_4 and X_5, since nonstationary processes exhibit joint movement, see Lansangan and Barrios (2009). The empirical behavior and summary values of the estimated components and percent differences are analyzed and compared between those estimated using ordinary least squares and the proposed additive modeling scheme.

The data-generating model in (7) is simulated from

$$Y_i = 12 + 20X_{1i} + 3X_{2i} - 14X_{3i} - 17X_{4i} + 10X_{5i} + k\varepsilon_i \quad (11)$$

with $X_{4i} = 9X_{1i} + k_1\varepsilon_{1i}$ and $X_{5i} = 4X_{2i} + 6X_{3i} + k_2\varepsilon_{2i}$. The values reported are averages over several cases and scenarios or averaged over the respective replicate datasets. The performance of least squares regression (OLS), backfitted OLS (BF), and the nonparametric smoothing spline additive model (SS) are assessed by the percentage differences of their estimated component shares from the true component shares. These are examined both when the independent variables are classified as one of the uncorrelated regressors (X_1, X_2, and X_3) or in the correlated set (X_4 and X_5) or with respect to the priority of entry of the independent variables into the model measured by the size of their true shares on Y ($\text{share}(X_j)$). When analyzing accuracy that are compiled according to the importance of the predictors X_1 to X_5, it is possible for the independent variables to change priorities depending on the case replicate for that specific scenario.

Note that due to the presence of multicollinearity, the estimated shares of the predictors observably are underestimates of their true values in general with negative component shares. This is due to the bias induced during estimation since the regressors are non-orthogonal.

5. Results and Discussions

The percentage differences of the estimated shares using OLS, BF, and SS are shown on Table 2 as averaged over all cross-sectional data by priorities of the estimation of shares of the independent variables. Considering that the data generating model is linear in form, the backfitting procedure using the parametric form of least squares as a smoother produces estimated shares that are closer to their true values for the first three most important predictors of Y. However, OLS yields closer share estimates for the fourth priority variable. In this situation though, the nonparametric SS performs at par with OLS (44.36% difference). On the last priority variable however, BF tremendously outperforms OLS and SS with a share percent difference of -308.15. This is more than a threefold increase in accuracy compared to the two other estimators.

Given that the true model is linear, backfitting indeed produced accurate estimates of component shares of the different predictors, the more accurate are those of shares estimated ahead in the iterative process.

Table 2. Average Share Size Differences from the True Share Values by Priority of *X* entering the Model (Cross Sectional Data)

Priority in Model Entry	Method		
	OLS	BF	SS
First	-149.87	-101.68	-158.24
Second	-165.81	-98.93	-167.65
Third	-155.38	-97.77	-152.15
Fourth	41.26	-104.38	-44.36
Fifth	-1900.69	-308.15	-1051.52

Table 3 summarizes the difference between estimated shares and the true share values for time series data with AR(1) regressors. Again, because of the linear form of the data generating model, backfitted linear smoothers yield estimates of shares that are closer to the true values. The additive modeling procedure performs (approximately) at par with OLS in share estimation.

Table 3. Average Share Size Differences from the True Share Values by Priority of *X* entering the Model (Time Series Data)

Priority in Model Entry	Method		
	OLS	BF	SS
First	-137.58	-115.19	-133.89
Second	-119.89	-114.46	-128.12
Third	278.03	55.87	252.20
Fourth	103.77	47.70	205.83
Fifth	-828.96	-627.62	-1598.71

5.1 Effect of Linear Dependencies of the Regressors

Table 4 summarizes the average percent difference of the share estimators according to the category of the predictors (set of uncorrelated, set of multicollinear independent variables). For the group of uncorrelated regressors, BF produces share estimates that are closer to their true values for X_1, X_2, and X_3. BF estimates the share of X_2 most optimally with a fifteen-fold increase in accuracy in the cross-section case. SS performs relatively at par with OLS, with a more accurate shares for X_2 and X_3 and a worse component for X_1. In the correlated set, OLS outperforms both BF and SS in the share of X_4. However, BF produces slightly better shares than OLS in X_5.

Table 4. Average Share Size Differences from the True Share Values by Nature of *X* at Moderate and Weak levels of Multicollinearity (Cross-Sectional Data)

Nature of Predictors	Predictor	Method		
		OLS	BF	SS
Uncorrelated Predictors	X_1	-376.75	-99.42	668.42
	X_2	-1715.86	-112.83	-1598.05
	X_3	-956.85	-94.62	-578.33
Correlated Predictors	X_4	-4.11	-99.44	-112.07
	X_5	-784.03	-522.25	-774.07

The average share differences between the estimated share and the true values for time series data are summarized in Table 5. Estimates obtained by backfitting the least squares estimator are closer to their actual values in the uncorrelated predictor set compared to their regular OLS counterparts. In the collection of correlated predictors though, there is an apparent reversing of pattern from the results in cross-section data. BF outperforms OLS in the variable with less collinear parts (X_4).

Table 5. Average Share Size Differences from the True Share Values by Nature of *X* at Moderate and Weak levels of Multicollinearity (Time Series Data)

Nature of Predictors	Predictor	Method		
		OLS	BF	SS
Uncorrelated Predictors	X_1	1259.66	484.93	1121.14
	X_2	-5.45	-76.63	124.15
	X_3	-2781.19	-1792.28	-3642.02
Correlated Predictors	X_4	388.31	108.09	555.16
	X_5	-27.89	-67.99	-50.16

5.2 Effect of Sample Size on the Decomposition

The performance of the three estimators did not exhibit consistent pattern according to the different levels of the sample size. For small sample sizes (n = 20), backfitting outperforms OLS and SS in the cross-section case, with OLS and SS comparable. This pattern though does not hold in the time series context, since having a short series (40 data points) shows that OLS, BF, and SS perform comparatively similar. Also in the cross-section case, BF outperforms both OLS and SS, again with OLS and SS performing comparably with a large sample size. The same holds true for the time series case with n = 200.

Table 6. Average Share Size Differences from the True Value for the First Priority Variable by Sample Size

Cross-Section Data			
Sample Size	Method		
	OLS	BF	SS
20	-160.91	-102.15	-178.04
100	-138.83	-101.21	-138.43
Predictor	Method		
	OLS	BF	SS
40	-103.36	-103.88	-98.19
200	-171.81	-126.49	-169.59

5.3 Effect of Model Fit in the Decomposition of Components

Varying model fit was simulated using the constant k, multiplied in the error term. The model fit deteriorates as this contamination constant increases. A large value of the constant k is tantamount to a model that is not sufficiently linear in form. Hence, under this scenario, nonparametric methods are known to exhibit relative advantage over other procedures.

In the cross section case, as k increases from 1 (near-perfect fit) to 10000 (moderate fit) to 1000000 (poor model fit), the percent difference of the share of the first priority variable remains relatively constant (around -100%), while the percent differences for OLS and SS increase accordingly. This shows the robustness of the backfitting procedure even if there is increasing departure from model fit. There are no recognizable differences in the accuracy of OLS and SS in the cross-section case under varying model fit scenarios. For the time series cases, the average percent difference for the first priority variable using OLS, BF, and SS don't deviate from each other substantially for near-perfect and moderate model fit. The BF method shows increased robustness and accuracy in the time series data with severe departure from model fit as shown by its -153.98% deviation, clearly more accurate than component predictions from OLS (-223.66%) and SS (-232.43%).

Table 7. Average Share Size Difference from the True Value for the First Priority Variable by Model Fit

Cross-Section Data			
Contamination Constant (k)	**Method**		
	OLS	**BF**	**SS**
1	-102.38	-103.10	-102.38
10,000	-138.83	-101.21	-138.43
1,000,000	-217.52	-100.65	-235.72
Contamination Constant (k)	**Method**		
	OLS	**BF**	**SS**
1	-106.47	-105.40	-106.48
1,000	-82.62	-86.19	-62.77
1,000,000	-223.66	-153.98	-232.43

5.4 Effect of Severity of Multicollinearity on the Decomposition of Components

The two multicollinear variables X_4 and X_5 represent the presence of linear dependencies in the model. The level and severity of the multicollinearity is controlled by k_1 and k_2 respectively, modifying the relationship of X_4 on X_1 and of X_5 on X_2 and X_3 respectively. Table 8 summarizes the performance of the component estimators for cross-section and time-series data at varying numbers and severity of multicollinearity.

The backfitted estimates increases in prediction accuracy and robustness of share estimates for the primary predictor as the severity of multicollinearity increases. BF performs only comparably to OLS and SS when there is moderate/weak multicollinearity on both variables. The better performance of BF though appears when there is at least one strong multicollinearity present in X_4 or X_5, since the percentage differences for OLS and SS increase rapidly to -172% and -179.20% respectively (BF still remains at around 101%). At extreme multicollinearity though, the parametric backfitting method starkly outperforms both OLS and SS, with SS performing poorer than the other two.

For the time series case, the three estimators perform at-par with each other at weaker multicollinearity. For example, if there is only a weak or moderate multicollinearity on both X_4 and X_5, the accuracy of OLS, BF, and SS are comparable. If there is strong multicollinearity on only one of the variables, there are only fractional differences in the accuracy of the three methods. The more optimal performance of additive modeling is illustrated in the case of having both sets of multicollinearity in X_4 and X_5 at an extreme level (both k_1 = 1 and k_2=1). Note that the optimality and accuracy of backfitting is slightly more persistent in the cross-section data compared to the time series data. Also, the comparable performance of OLS and the additive model methodology in situations of weak or moderate multicollinearity is consistent with theory. Due to the weak linear dependencies, the estimates of component estimates from OLS do not suffer as much bias as one would expect to arise in cases of severe multicollinearity. Thus, BF and SS don't present much gain in cases with weakly collinear independent variables. The added optimality of backfitting arises only in scenarios of harsh multicollinearity among the predictors.

Table 8. Average Share Size Difference from the True Value for the First Priority Variable by the Severity of Multicollinearity

Cross Section Data			
	Method		
	OLS	BF	SS
Strong in 2 Variables	-251.40	-99.99	-300.48
Strong in 1 Variable and Moderate/Weak on the other	-172.20	-101.07	-179.20
Moderate/Weak on Both Variables	-102.15	-102.71	-101.72
Time Series Data			
	Method		
	OLS	BF	SS
Strong in 2 Variables	-455.63	-252.35	-423.74
Strong in 1 Variable and Moderate/Weak on the other	-95.26	-95.66	-95.22
Moderate/Weak on Both Variables	-100.39	-100.42	-100.09

*Strong - k1 (equivalently, k2) is 1; Moderate - k1 (equivalently, k2) is 100; Weak - k1 (equivalently, k2) is 500.

6. Example: Decomposing Sales Data

We illustrate the application of the proposed method for some sales data. The predictors used included price, price of competitor 1, price of competitor 2, availability in retail outlets, availability in retail outlets of competitor 1, availability in retail outlets of competitor 2, out-of-stock in outlets, competitor 1 out-of-stock in outlets, competitor 2 out-of-stock in outlets, marketing expenditures, peso-dollar exchange rate, negative consumer perception of economic conditions, positive consumer perception of economic conditions, growth of gross domestic product, inflation rate, unemployment rate, and weather conditions. The data available is time series observations monitored monthly for 4 years (48 time points).

The full model yields a coefficient of determination of 47%, indicating that a linear model may not be adequate to explain sales behavior. Severe multicollinearity is evident from the data with variance inflation factors for 8 predictors exceeding 10, the highest is at 127. Eight eigenvalues are smaller than 0.01, the smallest is barely zero at 0.0000254. The condition index is very high at 637,195, indicating the presence of severe multicollinearity. While one predictor is significant [competitor 2 out-of-stock ($p<0.0282$)], the overall analysis of variance in not ($p<0.1529$). There are also some indicators with reversed signs from what is theoretically expected from that variable.

After backward elimination was implemented in variable selection, only three indicators were left (availability, competitor 2 is out-of-stock, and weather indicator). While these indicators are important as well, the development of optimal marketing mix can benefit significantly if indicators like marketing expenditures, prices, etc., are included in the final model.

Table 9 summarizes the result of decomposition based on OLS estimation, backfitted linear smoother (BF) and the additive model with spline smoother (SS). The OLS estimator yield a mean absolute prediction error (MAPE) of 8.09%, backfitted linear smoother with 154%, while the spline smoother have the lowest MAPE of 3.67%. The nonpametric (spline smoother) method is superior at this point since the predictors indicate so much multicollinearity and the linear model do not adequately fit the data.

From SS, own price and price of competitor are among the most important predictors, their share in sales are way higher compared to other predictors. Considering that the product belongs to the fast moving consumer good (FMCG) category, pricing is indeed a very important driver of sales. OLS and BF were not able to manifest this behavior. Furthermore, weather indicator that is retained in backward elimination procedure is among the lowest in terms of percentage share contribution on sales.

Table 9. Summary of Estimated Component Shares of Sales

Predictor of Sales	OLS		BF		SS	
	Model Coefficient	Component Share	Component Share	Backfitting Priority	Component Share	Backfitting Priority
Compet_1_Price	5753732	2.9389	3.0183	7	7.9559	7
In_Stock	534047	2.2040	1.0533	1	0.8011	1
Price	-3590650	-2.0856	-2.7629	4	-7.3943	4
Exchange_Rate	-352895	-0.9671	-1.0613	12	-0.7921	12
Compet_1_In_Stock	-561847	-0.9567	-0.6796	2	-1.7544	2
Compet_2_In_Stock	473156	0.9145	0.9631	9	1.7950	9
CCT_Worse	22916547	0.3160	0.2739	17	0.4284	17
Compet_2_Out_Stock	1497234	0.2398	0.2384	6	0.3179	6
Unemploy_Rate	-339495	-0.2095	-0.2630	15	-0.1520	15
CCT_Better	23023400	0.1670	0.0707	11	0.4145	11
Inflation	647601	0.1450	0.1844	3	-0.2353	3
Compet_1_Out_Stock	-199866	-0.0513	-0.0501	10	-0.0675	10
Compet_2_Price	-129998	-0.0471	0.1368	8	-1.1260	8
GDP_Growth	-138738	-0.0316	-0.0529	5	-0.3305	5
Marketing_Expenditure	-6	-0.0145	-0.0210	13	-0.0159	13
Weather	-120064	0.0028	0.0028	16	0.0046	16
Out_Stock	-2734	-0.0006	-0.0404	14	-0.2592	14

7. Conclusions

The additive modeling methodology using the backfitted form of the OLS estimator (BF) and a nonparametric smoothing spline (SS) are used to estimate the explained component shares of a set of predictors on a continuous response variable under multicollinearity. The backfitted estimator (BF) produces more accurate estimates of the shares compared to OLS for the most important variables in the model. In contrast, it is apparent that BF is better in share estimation as the model fit decreases. It also shows robustness in the estimated components. Meanwhile, backfitting yield better results as the number of collinear variables involved increases and the multicollinearity in the data becomes more severe. However, as the level of multicollinearity dissipates, so does the advantage of additive modeling due to adequate estimation under OLS.

Specific model scenarios are analyzed for the SS method in the cross-section and time series contexts. Under strong multicollinearity in one variable, poor model fit, and a small sample size, the SS method outperforms both BF and OLS in the cross section case. However, in time series data, OLS and BF also perform very well even though SS provides more accurate estimates for the most important model variable. The actual time series data that is characterized by severe multicollinearity and inadequate linear fit illustrates the advantage of SS over OLS and BF in terms of predictive ability (MAPE) and interpretability of the estimated shares of the predictors to the dependent variable.

References

Alvarez, J. and Pardinas, J., 2009, Additive Models in Censored Regression, Computational Statistics and Data Analysis, 53: 3490-3501.

Baye, M., and Parker, D., 1984, Combining Ridge and Principal Component Regression, Communications in Statistics - Theory and Methods, 13(2): 197 – 205.

Belsley, D., Kuh, E., and Welsh, R., 1980, Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Casals, J., Jerez, M., and Sotoca, S., 2010, Decomposition of a State-Space Model with Inputs, Journal of Statistical Computation and Simulation, 80(9): 979 – 992.

Chatterjee, S., and Hadi, A., 1988, Sensitivity Analysis in Linear Regression. New York: Wiley.

Farrell, AM, 2002, Effects of multiple performance measures, multicollinearity, and task structure on individuals' judgment performance. Ph.D. dissertation, Michigan State University, United States – Michigan.

Friedman, J., Hastie, T., and Tibshirani, R., 2000, Additive Logistic Regression: A Statistical View of Boosting, The Annals of Statistics, 28(2): 337-407.

Greenberg, E., and Parks, R., 1997, A Predictive Approach to Model Selection and Multicollinearity, Journal of Applied Econometrics, 12(1): 67-75.

Hastie, T., and Tibshirani, R., 1990, Generalized Additive Models, London: Chapman and Hall.

Hastie, T., and Tibshirani, R., 2000, Bayesian Backfitting, Statistical Science, 15(3): 196-223.

Hoerl, A., and Kennard R., 1970, Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12: 55–67.

Kosfeld, R., and Lauridsen, J., 2008, Factor Analysis Regression. Statistical Papers, 49(4): 653-667.

Landagan, O., and Barrios, E., 2007, An Estimation Procedure for a Spatial-Temporal Model, Statistics and Probability Letters, 77:401-406.

Lansangan, J.R., and Barrios, E., 2009, Principal Components Analysis of Nonstationary Time Series Data, Statistics and Computing, 19(2): 173-187.

Lauridsen, J. and Mur, J., 2006, Multicollinearity in Cross-Sectional Regressions. Journal of Geographical Systems, 8(4): 317-333.

Lipovetsky, A., 2006, Two-Parameter Ridge Regression and its Convergence to the Eventual Pairwise Model, Mathematical and Computer Modelling, 44(3-4): 304-318.

Lipovetsky, S., and Conklin, M., 2001, Multiobjective Regression Modifications for Collinearity, Computers and Operations Research, 28(13): 1333-1345.

Luo, Z., 1998, Backfitting in Smoothing Spline ANOVA, The Annals of Statistcs, 26(5): 1733-1759.

Mammen, E., Linton, O., and Nielsen, J., 1999, The Existence and Asymptotic Properties of a Backfitting Projection Algorithm under Weak Conditions, The Annals of Statistics, 27(5): 1443-1490.

Morlini, I., 2006, On Multicollinearity and Concurvity in Some Nonlinear Multivariate Models, Statistical Methods and Applications, 15(1): 3-26.

Opsomer, J., 2000, Asymptotic Properties of Backfitting Estimators, Journal of Multivariate Analysis, 73:166-179.

Sarkar, N., 1996, Mean Square Error Matrix Comparison of Some Estimators in Linear Regressions with Multicollinearity, Statistics and Probability Letters, 30(2): 133-138.

Smith, M., 1998, Additive Nonparametric Regression with Autocorrelated Errors, Journal of The Royal Statistical Society B, 60(2): 311-331.

Wan, A., 2002, On Generalized Ridge Regression Estimators under Collinearity and Balanced Loss, Applied Mathematics and Computation, 129(2-3): 455-467.

Wang, G., 1996, How to Handle Multicollinearity in Regression Modeling. The Journal of Business Forecasting Methods and Systems, 15(1): 23.

Wei, W., 2006, Time Series Analysis: Univariate and Multivariate Methods, Boston: Pearson Addison Wesley.

Yalian, L., and Hu, Y., 2008, A New Stochastic Mixed Ridge Estimator in Linear Regression Model. Statistical Papers – Online Article Published March 1, 2010.