QQ Models: Joint Modeling for Quantitative and Qualitative Quality Responses in Manufacturing Systems

Xinwei Deng and Ran Jin*

Department of Statistics and Grado Department of Industrial and Systems Engineering

Virginia Tech

Abstract

A manufacturing system with both quantitative and qualitative (QQ) quality responses (as a QQ system) is widely encountered in many cases. For example, in a lapping process of the semiconductor manufacturing, the quality of wafer's geometrical characteristics is often measured by the total thickness variation as a quantitative response and the conformity of site total indicator reading as a binary qualitative response. The QQ responses are closely associated with each other in a QQ system, but current methodologies often model the two types of quality responses separately. This paper presents a novel modeling approach, called "QQ models", to jointly model the QQ responses through a constrained likelihood estimation. The QQ models can jointly select significant predictors by incorporating inherent features of QQ systems, leading to accurate variable selection and prediction. Both simulation studies and a case study in a lapping process are used to evaluate the performance of the proposed method.

Keywords: Data fusion; Model selection; Nonnegative garrote; Quantitative and qualitative quality control.

1   Introduction

In manufacturing systems, both quantitative and qualitative (QQ) quality responses are often used for quality control. We call such a system "a QQ system". QQ systems are widely

───────────

*Address for correspondence: Ran Jin, Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061 (E-mail: jran5@vt.edu).

encountered in various manufacturing processes. For example, in a lapping process of a semi-conductor manufacturing, the wafers are lapped to improve the thickness uniformities and surface finish as shown in Figure 1 (Ning et al., 2012). The total thickness variation (TTV) is a quantitative quality response to characterize the range of the wafer thickness. The conformity of site total indicator readings (STIR) is a qualitative response with binary output, which is used to indicate whether the STIR is larger than the tolerance or not. The STIR is the span of deviation readings of the front surface in a pre-defined site, which represents the flatness of wafers. The detailed definitions of TTV and STIR can be found from the semiconductor manufacturing industry (O'Mara et al., 1990). The TTV and the conformity of STIR are two key measurements for the quality of wafer geometrical characteristics. Both of them are affected by the same set of process variables in the lapping manufacturing, such as pressure, rotation speed and lapping time (Marinescu et al., 2007). Common root causes of QQ responses need to be identified for variation reduction and quality improvement. However, the association between the two variables is not clear from engineering domain knowledge.

Therefore, joint models of QQ responses are needed for quality-process modeling in the lapping process. For another example, in a solar cell lamination process, solar cells are grouped and laminated to panels. The panels may have both power loss and solar cell cracks (Pilla et al., 2002; Paggi et al., 2013). The quality of the process often considers the power loss as a quantitative response and uses a binary response to indicate if the panel has cracked solar cells after the lamination. For quality improvement of solar cells, it is natural to consider a joint modeling of the power loss and the crack of the solar cells.

In manufacturing systems, the quantitative quality variables are widely used for quality control. Meanwhile, the qualitative quality variables also exist because of heuristic judgment or the limitation of sensor measurements. For example, the quantitative quality variables cannot be accurately measured in real time due to sensor limitations. Instead, some qualitative quality response variables are relatively easy to collect for facilitating the real time data collection. For the lapping process in Figure 1, the sensor is only capable to indicate whether the STIR is within the tolerance or not, which results in a qualitative quality measurement. With the qualitative quality measurement available, modeling such responses with respect to
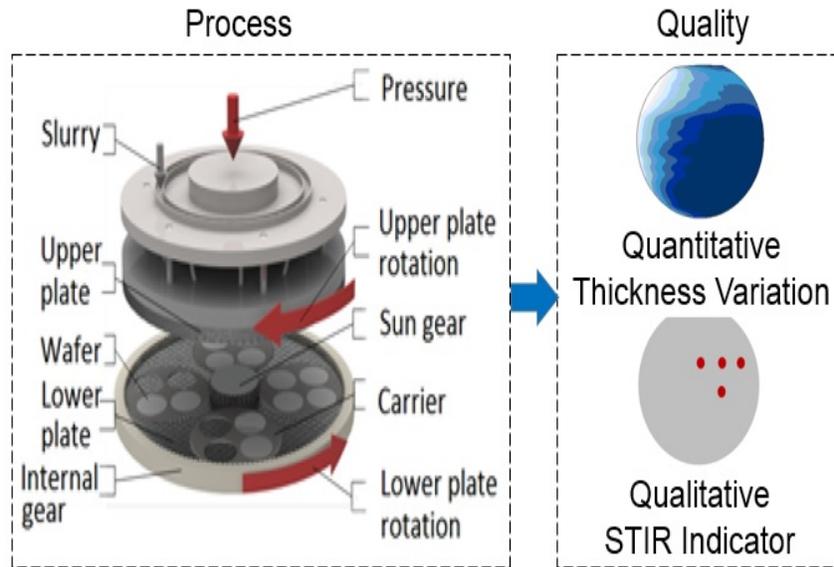
Figure 1: An Illustration of A Lapping Process with QQ Responses (Redrawn from Ning et al., 2012, and Zhao et al., 2011 with Authors' Permission).

the corresponding predictor variables will be very useful because it can clearly provide useful information on whether the manufacturing is conforming or not (e.g., a binary indicator for a failed manufacturing is clearly non-conforming). Such information will be reflected in the variable selection in our proposed methods, which will be discussed later.

Although both QQ quality responses exist in a manufacturing system, current methods often focus on developing quality-process models separately for the two types of quality responses. For quantitative quality variables, statistical models and engineering models are widely used to model the quality-process relationship based on observational data or design of experimental data (Fong and Lawless, 1998; Shi, 2006; Wu and Hamada, 2009; Jin and Shi, 2012). Quality control are performed based on these models, such as process monitoring (Hawkins, 1991, 1993; Woodall, 2000, 2007; Montgomery, 2001; Qiu, 2014), diagnosis (Apley and Shi, 1998; Zhou et al., 2003; Shi, 2006), and control (Joseph, 2003; Jin and Ding, 2004; Shi, 2006; Jin and Shi, 2012). Modeling and quality control methods for qualitative quality variables are also proposed, such as in the area of design of experiments, statistical process control and Run-to-Run control for categorical variables (May, et al., 1991; Spanos and Chen, 1997; Wang and Tsung, 2007; Lin and Wang, 2011). Qiu (2014) provides a com-

prehensive summary on the recent development in statistical process control. In general, the above modeling and quality control methods do not account for the association between the two types of quality responses. In the literature of biometrical study, Fitzmaurice and Laird (1997) investigated a joint modeling for the association of QQ responses. In their work, a conditional model of quantitative response is considered conditioned on the qualitative response, leading to marginal regression models. However, the marginal regression models may not be useful for quality control in the manufacturing systems, since they may not provide an accurate prediction of the quality responses. In the literature of process monitoring, Qiu (2008) suggested monitoring the mixed responses based on categorizing the quantitative responses. Thus the monitoring can be carried out based on the joint distribution of the categorized responses. It is possible to apply similar strategies to model the QQ responses without imposing the normality assumption. In manufacturing systems, it often requires to model the quantitative responses directly for quality control purpose, as the lapping process discussed earlier. Recently, mixed graphical models (Chen et al., 2014; Yang et al., 2014) have been proposed to study the general association of QQ responses. They mainly focus on the correlation or partial correlation of the responses, rather than the exact dependency of the QQ responses.

In this paper, we propose a new joint modeling framework, called "QQ models", for both QQ quality responses in a QQ system. We focus on the qualitative response with binary output, though it is likely to accommodate the multi-level qualitative response as discussed in Section 6. We adopt a logistic regression model for the qualitative response. For the quantitative response, we consider the linear regression models conditioned on the qualitative response. To address the association of the QQ responses, we consider that the conditional linear regression models are different based on different values of the binary output. The proposed method enables a more direct and accurate prediction of the quality responses, rather than a prediction of the expectation of the quantitative response. Note that the prediction for the qualitative response is usually very informative in the manufacturing system. We will first construct the prediction of qualitative response, then predict the quantitative response conditioned on the estimated qualitative response. Recall the lapping

process example, the QQ responses of TTV and STIR indicator are jointly determined by the lapping process variables and the quality covariates before lapping. When the STIR indicator is zero (i.e., STIR satisfies the tolerance), the lapping process is likely to be conforming. Thus, the lapping process is effective to reduce TTV, and the predictors related to process conditions may be more important to affect the TTV than the quality covariates before lapping. If the STIR indicator is one (i.e., STIR is too large to satisfy the tolerance), then the lapping operations may not be conforming. Thus, the quality covariates before lapping become important to affect the TTV. Therefore, the coefficients in the conditional regression models for TTV may vary depending on the values of the STIR indicator. It is beneficial to consider different conditional regression models for quantitative response under different values of the qualitative response, reflecting the intrinsic heterogeneity of the underlying models in the QQ system. This engineering perception is successfully validated by the results of the case study in Section 5.

Moreover, the proposed QQ models adopt a joint likelihood approach for parameter estimation and use the nonnegative garrote approach (Breiman, 1995) for efficient variable selection. The nonnegative garrote approach accommodates the association of significant predictors among QQ models through flexible constraints. The proposed QQ models intend to encourage that the significant variables in the logistic regression model are kept as significant in at least one of the conditional linear regression models. It is because that the significant variables from modeling the qualitative response are expected to contribute an important role for modeling other quantitative quality responses. For example, in manufacturing scale-up, modeling the binary response whether the manufacturing is conforming or not will provide useful information. In this model, a significant variable should be important for quality control in general, and thus should also be kept as a significant variable for other quantitative quality responses. To address the computational consideration of parameter estimation, we also develop a fast computation algorithm by iteratively approximating the objective function using quadratic approximations.

The rest part of the paper is organized as follows. In Section 2, we detail the proposed QQ models for jointly modeling both types of responses. In Section 3, we develop an efficient

algorithm for fast computation. In Section 4, simulation studies are conducted to examine the effectiveness in prediction and variable selection of the proposed QQ models. In Section 5, the proposed QQ models are applied to the lapping process as a case study. Finally, we draw the conclusion and discuss the future work in Section 6.

2   The Proposed QQ Models

We start with one quantitative response y and one qualitative response z with binary output, though it is possible to extend the models for multiple responses, as discussed in Section 6. Let us denote the observed data are $(\boldsymbol{x}_i, y_i, z_i), i = 1, \ldots, n$, where $y_i \in \mathcal{R}$ and $z_i \in \{0, 1\}$. Here the predictor vector $\boldsymbol{x} = (x_1, \ldots, x_p)'$ contains p predictors, which can be process variables or initial quality covariates. To jointly model the QQ responses y and z given $\boldsymbol{x}$, we follow a joint probability density function $f(y, z|\boldsymbol{x}) = f(y|z, \boldsymbol{x})f(z|\boldsymbol{x})$, where $f(\cdot)$ denotes a general density function. The conditional model on $y|z$ is considered to be linear regressions, while the model of z follows a logistic regression. Specifically, we propose a joint modeling of y and z as

$$z|\boldsymbol{x} = \begin{cases} 1, & w.p. \ p(\boldsymbol{x}) \\ 0, & w.p. \ 1 - p(\boldsymbol{x}) \end{cases} \quad \text{with } p(\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}'\boldsymbol{\eta})}{1 + \exp(\boldsymbol{x}'\boldsymbol{\eta})}, \tag{2.1}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)'$ is a vector of parameter coefficients, and

$$y|z, \boldsymbol{x} \sim N(z\boldsymbol{x}'\boldsymbol{\beta}^{(1)} + (1 - z)\boldsymbol{x}'\boldsymbol{\beta}^{(2)}, \sigma^2), \tag{2.2}$$

where $\boldsymbol{\beta}^{(m)} = (\beta_1^{(m)}, \ldots, \beta_p^{(m)})', m = 1, 2$. The above proposed model indicates that $y|z = 1, \boldsymbol{x} \sim N(\boldsymbol{x}'\boldsymbol{\beta}^{(1)}, \sigma^2)$ and $y|z = 0, \boldsymbol{x} \sim N(\boldsymbol{x}'\boldsymbol{\beta}^{(2)}, \sigma^2)$. We assume the same variances for the two conditional distributions of $y|z = 1, \boldsymbol{x}$ and $y|z = 0, \boldsymbol{x}$. It means that the difference between two conditional distributions are mainly driven by their mean functions. If the two conditional distributions have different variances, one can easily have two different variance parameters, which will not change the nature of the model formulation. Recall that the relationship between the quantitative response y and its predictors depends on the output of qualitative response z. To accommodate this consideration, the proposed method incorporates two conditional linear regression models for $y|z = 1, \boldsymbol{x}$ and $y|z = 0, \boldsymbol{x}$. If these two

linear models are the same, i.e., $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$, then the quantitative response y and the qualitative response z are independent. In this situation, one can model the quantitative response y in regardless of the qualitative response z. Alternatively, it is important to take accounts of the effects of the qualitative response z when modeling the quantitative response y. Note that the proposed models in (2.1) and (2.2) provide a joint modeling of $y|z, \boldsymbol{x}$ and $z|\boldsymbol{x}$. Now we can also derive the conditional model on $z|y, \boldsymbol{x}$. The probability of z conditioned on y can be expressed as $P(z = 1|y, \boldsymbol{x}) = f(y|z = 1, \boldsymbol{x})P(z = 1, \boldsymbol{x})/f(y, \boldsymbol{x})$. Thus, we have

$$P(z = 1|y, \boldsymbol{x}) = \frac{\exp(\frac{(y-\boldsymbol{x}'\boldsymbol{\beta}^{(1)})^2}{-2\sigma^2}) \exp(\boldsymbol{x}'\boldsymbol{\eta})}{\exp(\frac{(y-\boldsymbol{x}'\boldsymbol{\beta}^{(1)})^2}{-2\sigma^2}) \exp(\boldsymbol{x}'\boldsymbol{\eta}) + \exp(\frac{(y-\boldsymbol{x}'\boldsymbol{\beta}^{(2)})^2}{-2\sigma^2})}. \tag{2.3}$$

It implies that the distribution $z|y$ becomes

$$z|y, \boldsymbol{x} = \begin{cases} 1, & w.p. \ p_y(\boldsymbol{x}) \\ 0, & w.p. \ 1 - p_y(\boldsymbol{x}) \end{cases} \quad \text{with } p_y(\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}'\boldsymbol{\eta})}{h(y, \boldsymbol{x}) + \exp(\boldsymbol{x}'\boldsymbol{\eta})}, \tag{2.4}$$

where $h(y, \boldsymbol{x}) = \exp((y - \boldsymbol{x}'\boldsymbol{\beta}^{(2)})^2)/\exp((y - \boldsymbol{x}'\boldsymbol{\beta}^{(1)})^2)$. Therefore, the proposed method can provide both conditional models of $y|z, \boldsymbol{x}$ and $z|y, \boldsymbol{x}$, upon which is more flexible for the manufacturing system.

To estimate the parameters in the proposed models, we consider the joint likelihood estimation approach, which enables the parameter estimation to borrow strength from two models of QQ responses. The log-likelihood function is

$$l(\boldsymbol{\eta}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}) = \log\left\{\prod_{i=1}^n f(z_i)f(y_i|z_i)\right\} = \log\left\{\prod_{i=1}^n f(z_i)\right\} + \log\left\{\prod_{i=1}^n f(y_i|z_i)\right\}$$

$$= \log\left\{\prod_{i=1}^n \left[p(\boldsymbol{x}_i)^{z_i}(1 - p(\boldsymbol{x}_i))^{1-z_i}\right]\right\} \tag{2.5}$$

$$+ \log\left\{\prod_{i=1}^n \left[(\sigma^2)^{-1/2}\exp(\frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}^{(1)})^2}{-2\sigma^2})\right]^{z_i}\left[(\sigma^2)^{-1/2}\exp(\frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}^{(2)})^2}{-2\sigma^2})\right]^{1-z_i}\right\}$$

up to some constant independent of parameters. Now we can estimate the parameters by minimizing the negative log-likelihood function upon some constraints for pursing sparse model structures. A sparse model structure only contains a few significant variables in the model, which is often called as variable selection (Miller, 2002; Hastie et al., 2009).

It aims to achieve a parsimonious model with meaningful interpretation. Specifically, we consider the nonnegative garrote approach to pursue a sparse model for both models of QQ responses. The nonnegative garrote approach is originally introduced by Breiman (1995) for linear models. Several researchers (Yuan and Lin, 2007; Xiong, 2012) have further developed the theoretical properties of nonnegative garrote approach. This approach is also used in other statistical models, such as logistic regressions (Makalic and Schmidt, 2011) and additive models (Cantoni et al., 2011). In the nonnegative garrote approach, the key idea is to re-parameterize the parameter coefficients such that flexible constraints can be imposed to pursue a parsimonious model. Let $\beta_k^{(1)} = \theta_k^{(1)} \tilde{\beta}_k^{(1)}$, $\beta_k^{(2)} = \theta_k^{(2)} \tilde{\beta}_k^{(2)}$, and $\eta_k = \tau_k \tilde{\eta}_k$, where $\tilde{\beta}_k^{(1)}$, $\tilde{\beta}_k^{(2)}$, and $\tilde{\eta}_k$ are some initial estimations of the model parameters, such as least squares estimation or marginal maximum likelihood estimation (MLE). The constraints of this optimization problem should encourage the sparsity of the QQ models, and reflect the engineering perception. Thus, we consider estimating the parameters through a constrained optimization problem as follows.

$$
\begin{aligned}
\min \quad & -2l(\boldsymbol{\eta}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}) \\
s.t. \quad & \beta_k^{(1)} = \theta_k^{(1)} \tilde{\beta}_k^{(1)}, \beta_k^{(2)} = \theta_k^{(2)} \tilde{\beta}_k^{(2)}, \eta_k = \tau_k \tilde{\eta}_k; \\
& \theta_k^{(1)} + \theta_k^{(2)} + \tau_k \leq t_k, k = 1, \ldots, p; \\
& \tau_k \leq \theta_k^{(1)} + \theta_k^{(2)}, k = 1, \ldots, p; \\
& \theta_k^{(1)} \geq 0, \theta_k^{(2)} \geq 0, \tau_k \geq 0, k = 1, \ldots, p; \\
& \sum_{k=1}^{p} t_k \leq M, t_k \geq 0.
\end{aligned}
\tag{2.6}
$$

Here the parameters for optimization includes $\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \ldots, \theta_p^{(1)})'$, $\boldsymbol{\theta}^{(2)} = (\theta_1^{(2)}, \ldots, \theta_p^{(2)})'$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)'$, $\boldsymbol{t} = (t_1, \ldots, t_p)'$, and σ^2. The parameters t_k's are nuisance parameters to encourage the strong association of the significant variables in the regression models for $y|z$ and the logistic regression model for z. Specifically, if a $t_k = 0$, the values of $\theta_k^{(1)}$, $\theta_k^{(2)}$ and τ_k will be forced to be zeros simultaneously. It implies that the kth predictor variable is not significant in the QQ system. Thus, this variable is not selected by the QQ models, which encourages the sparsity of the QQ models. The constraints $\tau_k \leq \theta_k^{(1)} + \theta_k^{(2)}$ is to encourage a variable to be selected in at least one of linear regression models, if this variable appears

8

to be significant in the logistic regression model. This constraint reflects the engineering perceptions that the predictors for modeling qualitative response are usually informative for quality control, and they are also expected to be significant for modeling the quantitative response in the manufacturing system.

Note that there is a tuning parameter M in (2.6), which needs to be specified based on the data. An appropriate selection of M can balance the trade-off between the model fitting and model parsimoniousness. If the value of M is set to be zero in (2.6), all values of the estimated parameters will be equal to zeros, i.e., none of predictor variables will be selected in the model. If the value of M is set to be sufficiently large, the proposed method will select all predictors in the model. The estimation of parameters will be the same as the MLE approach. The common methods to select tuning parameters include cross-validation and information criterion approaches, such as Akaike information criterion (AIC), Bayesian information criterion (BIC), and C_p criteria (Burnham and Anderson, 2002). In this work, we use the BIC to find an optimal value of the tuning parameter M. The BIC score for the proposed models can be written as

$$BIC(M) = -2l(\boldsymbol{\tau}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) + q \log(n), \tag{2.7}$$

where q is the number of nonzero estimates of parameters in the models. Specifically, we can generate a grid for M, such as $M \in \mathcal{C} = \{m_1, \ldots, m_t\}$. For each grid point m_j in $\mathcal{C}$, we evaluate the corresponding BIC scores, and find the optimal choice of M with the minimal BIC score among all grid points in $\mathcal{C}$.

3   Computational Algorithm

Solving the proposed optimization problem in (2.6) is non-trivial because of the non-convexity of the objective function. It is also likely that there are a large number of parameters to optimize. To address this challenge, we propose a quadratic approximation for the original objective function, thus to convert the optimization in (2.6) to a constrained quadratic optimization problem. The key idea is based on a sequence of local quadratic approximation. It is well known that the constrained sequential quadratic optimization can guarantee a global

9

optimum with fast computation (Boyd and Vandenberghe, 2004).

Denote $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ to be the design matrix, $\boldsymbol{z} = (z_1, \ldots, z_n)^T$ to be the binary response vector. Without loss of the generality, we assume that the first n_1 observations with binary response $z = 1$, and the remaining n_2 observations with response $z = 0$, where $n = n_1 + n_2$. Then, we can define $\boldsymbol{y}_{(1)} = (y_1, \ldots, y_{n_1})^T$ and $\boldsymbol{y}_{(2)} = (y_{n_1+1}, \ldots, y_n)^T$. Similarly, we can correspondingly define $\boldsymbol{X}_{(1)}$ and $\boldsymbol{X}_{(2)}$, respectively. Given an estimate of $\boldsymbol{\eta}$, we first apply the quadratic approximation for the log-likelihood function $l(\boldsymbol{\eta}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)})$ in (2.6). The detail of the quadratic approximation for $l(\boldsymbol{\eta}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)})$ is described in Part A of Supplemental Materials. Specifically, we approximate the original objective function in (2.6) by

$$
(\tilde{\boldsymbol{z}} - \boldsymbol{X}\boldsymbol{\eta})^T \boldsymbol{W} (\tilde{\boldsymbol{z}} - \boldsymbol{X}\boldsymbol{\eta}) + n \log(\sigma^2)
$$
$$
+ \frac{1}{\sigma^2} \left[(\boldsymbol{y}_{(1)} - \boldsymbol{X}_{(1)}\boldsymbol{\beta}^{(1)})^T (\boldsymbol{y}_{(1)} - \boldsymbol{X}_{(1)}\boldsymbol{\beta}^{(1)}) + (\boldsymbol{y}_{(2)} - \boldsymbol{X}_{(2)}\boldsymbol{\beta}^{(2)})^T (\boldsymbol{y}_{(2)} - \boldsymbol{X}_{(2)}\boldsymbol{\beta}^{(2)}) \right] \quad (3.1)
$$

Here $\tilde{\boldsymbol{z}} = \boldsymbol{X}\boldsymbol{\eta}_c + \boldsymbol{W}^{-1}(\boldsymbol{z} - \boldsymbol{p})$ with $\boldsymbol{W} = diag(p(\boldsymbol{x}_1, \boldsymbol{\eta}_c)(1 - p(\boldsymbol{x}_1, \boldsymbol{\eta}_c)), \ldots, p(\boldsymbol{x}_n, \boldsymbol{\eta}_c)(1 - p(\boldsymbol{x}_n, \boldsymbol{\eta}_c)))$, and $\boldsymbol{p} = (p(\boldsymbol{x}_1, \boldsymbol{\eta}_c), \ldots, p(\boldsymbol{x}_n, \boldsymbol{\eta}_c))'$ provided by an current estimate $\boldsymbol{\eta}_c$. In this case, the above objective function of the approximate quadratic problem is clearly a convex function with respect to unknown parameters $\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \ldots, \theta_p^{(1)})'$, $\boldsymbol{\theta}^{(2)} = (\theta_1^{(2)}, \ldots, \theta_p^{(2)})'$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)'$, and $\boldsymbol{t} = (t_1, \ldots, t_p)'$. Since the parameter σ^2 is not involved in any constraint, by taking the derivative of the objective function with respect to σ^2, we can obtain the estimate of σ^2 explicitly as follows,

$$
\hat{\sigma}^2 = \frac{1}{n} \left[(\boldsymbol{y}_{(1)} - \boldsymbol{X}_{(1)}\boldsymbol{\beta}^{(1)})^T (\boldsymbol{y}_{(1)} - \boldsymbol{X}_{(1)}\boldsymbol{\beta}^{(1)}) + (\boldsymbol{y}_{(2)} - \boldsymbol{X}_{(2)}\boldsymbol{\beta}^{(2)})^T (\boldsymbol{y}_{(2)} - \boldsymbol{X}_{(2)}\boldsymbol{\beta}^{(2)}) \right]. \quad (3.2)
$$

Note that the objective function in (3.1) is provided by an initial estimate $\boldsymbol{\eta}_c = \boldsymbol{\eta}_0$. We can choose a proper value of $\boldsymbol{\eta}_0$ and update the estimations in an iterative procedure. The iterative algorithm is summarized as follows:

Algorithm 1.

Step 1: *Set an initial estimate $\sigma^2 = \sigma_0^2 > 0$, also an initial estimate of $\boldsymbol{\eta}_c = \boldsymbol{\eta}_0$.*

Step 2: *Update the weight $\boldsymbol{W} = diag(p(\boldsymbol{x}_1, \boldsymbol{\eta}_c)(1 - p(\boldsymbol{x}_1, \boldsymbol{\eta}_c)), \ldots, p(\boldsymbol{x}_n, \boldsymbol{\eta}_c)(1 - p(\boldsymbol{x}_n, \boldsymbol{\eta}_c)))$, and the adjusted response $\tilde{\boldsymbol{z}} = \boldsymbol{X}\boldsymbol{\eta}_c + \boldsymbol{W}^{-1}(\boldsymbol{z} - \boldsymbol{p})$, where $\boldsymbol{p} = (p(\boldsymbol{x}_1, \boldsymbol{\eta}_c), \ldots, p(\boldsymbol{x}_n, \boldsymbol{\eta}_c))'$.*

Step 3: *Obtain the estimate $\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\beta}}^{(2)}$ by solving the optimization in (3.1).*

Step 4: *Update the estimate $\hat{\sigma}^2$ by plugging $\hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\beta}}^{(2)}$ obtained in Step 2 into (3.2).*

Step 5: *Check if convergence, i.e., the maximum absolute differences of each element of $\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\beta}}^{(2)}$ and the estimates in the last iteration are smaller than tolerance level $\delta = 0.001$. Otherwise, set $\sigma^2 = \hat{\sigma}^2$, $\boldsymbol{\eta}_c = \hat{\boldsymbol{\eta}}$, and go back to **Step 2**.*

For the initial estimation of σ^2 and $\boldsymbol{\eta}_c$, we simply choose σ_0^2 as the residual variance by fitting a linear regression model for the quantitative response, and $\boldsymbol{\eta}_0$ as the marginal logistic regression model parameters. For the initial estimation in the nonnegative approach for re-parameterization, we use the least squares estimates for $\tilde{\beta}_k^{(1)}$, $\tilde{\beta}_k^{(2)}$, and $\tilde{\eta}_k$. When the least squares estimates is not available, we would choose a ridge regression estimator (Hoerl and Kennard, 1970) for the linear regression models and the logistic regression model.

4 Simulation

To evaluate the performance of the proposed method, we consider several simulation settings for generating the data with the underlying models in (2.1) and (2.2). Let I_1 be the index set of significant predictor variables in $\boldsymbol{\beta}^{(1)}$ and I_2 be the index set of significant variables in $\boldsymbol{\beta}^{(2)}$. Denote $\bar{\boldsymbol{\beta}}_1 = \{\beta_k^{(1)} : k \in I_1\}$ and $\bar{\boldsymbol{\beta}}_2 = \{\beta_k^{(2)} : k \in I_2\}$. Similarly, we denote by I_l be the index set of significant variables in $\boldsymbol{\eta}$. Denote $\bar{\boldsymbol{\eta}} = \{\eta_k : k \in I_l\}$. Specifically, we consider four examples as follows.

Example 1. I_1 and I_2 are the same, and the values of $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ are similar.

Example 2. I_1 and I_2 are different (i.e., the significant variables are different), but the values of $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ are similar.

Example 3. I_1 and I_2 are the same, but the values of $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ are different.

Example 4. I_1 and I_2 are different, and the values of $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ are also different.

In each example, the index sets I_l, I_1 and I_2 are generated randomly but following the proportion of sparsity s. Here the value of s represents the proportion of nonzero entries in a parameter vector. The entry values of parameter vector $\bar{\boldsymbol{\eta}}$ are generated from uniform distribution $U(-2, 2)$. In Examples 1 and 2, we first generate the parameter vector $\bar{\boldsymbol{\beta}}_1$ with each entry value from normal distribution $N(2, 1)$. Then the entry values of $\bar{\boldsymbol{\beta}}_2$ are

obtained by adding a small perturbation from $N(0, 0.01)$ onto the entry values of generated $\bar{\boldsymbol{\beta}}_1$. In Examples 3 and 4, the entry values of parameter vectors $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ are generated independently from $N(2, 1)$, respectively.

For each example, we generate a training set and a test set based on the models in (2.1) and (2.2). The n data points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in the training set are independent and identically distributed (i.i.d.) sample generated from $N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (\sigma_{ij})_{p \times p}$ with (i, j)th entry $\sigma_{ij} = \rho^{|i-j|}$ and ρ being a correlation parameter. The n data points of the test data set are i.i.d. sample generated from $U(-2, 2)$. The sample sizes for the training data set and the test data set are $n = 100$. The σ^2 in the model (2.2) is chosen to be 1.

To systematically investigate the performance of the proposed method, we consider different scenarios of generating the simulation data by varying the predictor dimensionality p, correlation parameter ρ and proportion of sparsity s. We choose two levels of p with the values $p = 20$ and $p = 50$, two levels of ρ with the values $\rho = 0$ and $\rho = 0.5$, and two levels of s with the values $s = 20\%$ and $s = 50\%$. For every setting of each example, we conduct 50 simulation replicates.

To evaluate the accuracy of the estimated QQ models, we compare the proposed method with two benchmark models: separating modeling using BIC (SMBIC) and modeling with additional predictors using BIC (MABIC). The SMBIC approach ignores the association between the two types of responses. It fits a linear regression model for the quantitative response y and a logistic regression model for the qualitative response z separately, of which both use BIC for variable selection. The MABIC approach considers adding one of QQ responses as predictor in modeling the other response. Specifically, the MABIC is to fit a linear regression model for the quantitative response y by adding the qualitative response z as an additional predictor, and fit a logistic regression model for the qualitative response z by adding the quantitative response y as an additional predictor. The BIC is also used for the variable selection. For all three methods in comparison, the models estimated from the training set are used to compute the prediction errors based on the test set. Specifically, the prediction errors for the quantitative response y is measured by the root mean squared prediction error as $RMSPE = \frac{1}{n} \sum_{k=1}^{2} \sum_{i=1}^{n_k} \sqrt{(y_i^{(k)} - \boldsymbol{x}_i^{(k)'} \hat{\boldsymbol{\beta}}^{(k)})^2}$, and the prediction error

for the qualitative response z with binary output is measured by the misclassification error as $ME = \frac{1}{n}\sum_{i=1}^{n} I(z_i \neq \hat{z}_i)$, where $\hat{z}_i \in \{0, 1\}$ is the prediction of binary response z_i based on the logistic regression in (2.1), and $I(\cdot)$ is an indicator function. Furthermore, we also calculate the accuracy of variable selection for the estimated models. Here we use the total number of falsely selected variables, denoted by γ, as the performance measure. The γ is defined by $\gamma = FP + FN$, where FP represents the number of false positives and FN represents the number of false negatives in variable selection.

Table 1 reports the averages of RMSPEs (or MEs) and standard errors in parenthesis based on 50 simulation replicates. By using the efficient computational algorithm in Section 3, the average of computing time for the proposed method is 3.95 seconds for each simulation replicate (based on a workstation with CPU Xeon Processor E5-2687W, 3.10GHz, 64GB RAM). We denote the estimated linear regression models by QQ_{lm} in (2.2) and the estimated logistic regression model by QQ_{logit} in (2.1) from the proposed QQ models. Similarly, we denote BIC_{lm} and BIC_{logit} as the corresponding estimated models from the SMBIC approach. We denote Add_{lm} and Add_{logit} as the corresponding estimated models from the MABIC approach. For the results of Example 1, the performance of the proposed QQ models is comparable to the SMBIC approach and the MABIC approach. Note that Example 1 considers $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ being similar. It implies that the quantitative response y does not depend much on the qualitative response z. Since there is no hidden association between the QQ responses, separate modeling QQ responses such as the SMBIC and the MABIC approaches would give accurate estimation of parameters. In contrast, Examples 2-4 consider the situations of $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ being different with respect to the significant variables and their values. It means that the two conditional models $y|z = 1, \boldsymbol{x}$ and $y|z = 0, \boldsymbol{x}$ are different, reflecting the dependency between the QQ responses. In these situations, the proposed QQ models generally outperform the SMBIC approach and the MABIC approach. In particular, the RMSPEs from the QQ_{lm} are much smaller than those from the BIC_{lm} and the Add_{lm}. These findings can be explained by the fact that the proposed QQ models consider the dependency between the quantitative response y and qualitative response z through the joint probability $p(y|z, \boldsymbol{x})p(z|\boldsymbol{x})$. While the SMBIC approach considers the probability indepen-

Table 1: Averages and Standard Errors of RMSPEs (or MEs) from 50 Simulation Runs.

Example	p	Method	$\rho = 0$		$\rho = 0.5$	
			$s = 20\%$	$s = 50\%$	$s = 20\%$	$s = 50\%$
		QQ_{lm}	0.49 (0.01)	0.50 (0.01)	0.51 (0.01)	0.50 (0.01)
		BIC_{lm}	0.57 (0.01)	0.59 (0.01)	0.57 (0.01)	0.59 (0.01)
1	20	Add_{lm}	0.57 (0.01)	0.60 (0.01)	0.57 (0.01)	0.60 (0.01)
		QQ_{logit}	0.10 (0.00)	0.12 (0.01)	0.21 (0.01)	0.12 (0.01)
		BIC_{logit}	0.15 (0.01)	0.13 (0.01)	0.14 (0.01)	0.12 (0.01)
		Add_{logit}	0.15 (0.01)	0.12 (0.01)	0.15 (0.01)	0.12 (0.01)
		QQ_{lm}	0.72 (0.04)	5.36 (0.64)	0.66 (0.02)	1.69 (0.29)
		BIC_{lm}	0.66 (0.01)	0.80 (0.01)	0.65 (0.01)	0.87 (0.02)
1	50	Add_{lm}	0.66 (0.01)	0.80 (0.01)	0.65 (0.01)	0.88 (0.02)
		QQ_{logit}	0.20 (0.01)	0.25 (0.01)	0.18 (0.01)	0.21 (0.01)
		BIC_{logit}	0.17 (0.01)	0.26 (0.01)	0.18 (0.01)	0.24 (0.01)
		Add_{logit}	0.18 (0.01)	0.24 (0.01)	0.17 (0.01)	0.20 (0.01)
		QQ_{lm}	0.70 (0.02)	0.71 (0.01)	0.57 (0.01)	0.69 (0.02)
		BIC_{lm}	3.77 (0.04)	5.37 (0.06)	2.57 (0.03)	3.62 (0.04)
2	20	Add_{lm}	3.75 (0.04)	5.31 (0.06)	2.56 (0.03)	3.60 (0.04)
		QQ_{logit}	0.19 (0.01)	0.16 (0.01)	0.10 (0.00)	0.14 (0.01)
		BIC_{logit}	0.15 (0.01)	0.12 (0.00)	0.13 (0.01)	0.12 (0.01)
		Add_{logit}	0.15 (0.01)	0.13 (0.01)	0.13 (0.01)	0.12 (0.01)
		QQ_{lm}	3.02 (0.25)	10.04 (0.22)	3.21 (0.17)	8.20 (0.66)
		BIC_{lm}	9.05 (0.13)	8.37 (0.12)	8.43 (0.13)	8.48 (0.16)
2	50	Add_{lm}	8.66 (0.13)	8.44 (0.15)	8.46 (0.12)	8.43 (0.15)
		QQ_{logit}	0.21 (0.01)	0.26 (0.01)	0.25 (0.01)	0.23 (0.01)
		BIC_{logit}	0.17 (0.01)	0.25 (0.01)	0.18 (0.01)	0.23 (0.01)
		Add_{logit}	0.18 (0.01)	0.25 (0.01)	0.20 (0.01)	0.23 (0.01)
		QQ_{lm}	0.59 (0.01)	0.69 (0.01)	0.57 (0.01)	0.76 (0.04)
		BIC_{lm}	2.53 (0.03)	3.35 (0.04)	1.96 (0.02)	3.05 (0.05)
3	20	Add_{lm}	2.54 (0.03)	3.35 (0.04)	1.96 (0.02)	3.00 (0.04)
		QQ_{logit}	0.11 (0.01)	0.13 (0.01)	0.09 (0.00)	0.14 (0.01)
		BIC_{logit}	0.15 (0.01)	0.13 (0.01)	0.14 (0.01)	0.11 (0.01)
		Add_{logit}	0.15 (0.01)	0.12 (0.01)	0.14 (0.01)	0.11 (0.01)
		QQ_{lm}	1.53 (0.10)	9.99 (0.33)	2.21 (0.10)	7.78 (0.39)
		BIC_{lm}	3.52 (0.05)	8.11 (0.15)	4.27 (0.07)	7.93 (0.17)
3	50	Add_{lm}	3.56 (0.06)	8.05 (0.14)	4.29 (0.07)	7.88 (0.15)
		QQ_{logit}	0.16 (0.01)	0.24 (0.01)	0.16 (0.01)	0.19 (0.01)
		BIC_{logit}	0.19 (0.01)	0.27 (0.01)	0.19 (0.01)	0.24 (0.01)
		Add_{logit}	0.17 (0.01)	0.26 (0.01)	0.18 (0.01)	0.23 (0.01)
		QQ_{lm}	0.62 (0.01)	1.44 (0.16)	0.59 (0.01)	0.71 (0.03)
		BIC_{lm}	4.28 (0.06)	5.19 (0.06)	2.78 (0.04)	3.49 (0.04)
4	20	Add_{lm}	4.26 (0.06)	5.18 (0.06)	2.78 (0.04)	3.49 (0.04)
		QQ_{logit}	0.11 (0.01)	0.18 (0.01)	0.11 (0.01)	0.14 (0.01)
		BIC_{logit}	0.14 (0.01)	0.12 (0.01)	0.11 (0.01)	0.12 (0.01)
		Add_{logit}	0.14 (0.01)	0.13 (0.01)	0.12 (0.01)	0.11 (0.00)
		QQ_{lm}	2.90 (0.17)	10.64 (0.22)	3.09 (0.18)	8.86 (0.42)
		BIC_{lm}	7.79 (0.11)	12.11 (0.18)	7.82 (0.10)	11.16 (0.15)
4	50	Add_{lm}	7.64 (0.12)	11.77 (0.17)	7.73 (0.10)	10.89 (0.17)
		QQ_{logit}	0.27 (0.01)	0.25 (0.01)	0.25 (0.01)	0.20 (0.01)
		BIC_{logit}	0.17 (0.01)	0.27 (0.01)	0.17 (0.01)	0.24 (0.01)
		Add_{logit}	0.17 (0.01)	0.26 (0.01)	0.17 (0.01)	0.23 (0.01)

dently as $p(y|\boldsymbol{x})p(z|\boldsymbol{x})$ and the MABIC approach only considers a particular dependency of y and z by treating of one of them as predictor in the modeling. Therefore, the model structure for QQ_{lm} based on $p(y|z, \boldsymbol{x})$ is more favorable to obtain an accurate model in prediction. We also note that the MEs from the QQ_{logit} are comparable to that from the BIC_{logit} and the Add_{logit}. Because the QQ_{logit} is obtained based on marginal distribution $p(z)$ for modeling the binary response. It is expected that QQ_{logit} have comparable prediction performance to the BIC_{logit} and the Add_{logit}, both of which are also based on marginal distribution $p(z|\boldsymbol{x})$ for modeling the binary response.

Under different design matrices and dimensionality of parameters, we also observe that the proposed QQ models give better prediction performance than the SMBIC approach and the MABIC approach for Examples 2-4. It is because that the proposed QQ models take the advantage of the association between the QQ responses regardless of the structure of the design matrix $\boldsymbol{X}$. Under two levels of sparsity s with the same correlation parameter ρ and dimensionality p, we can see that the QQ_{lm} has smaller RMSPEs than the BIC_{lm} and the Add_{lm} in both sparsity levels for $p = 20$. When the dimensionality becomes larger in $p = 50$, the QQ_{lm} has much better performance than the BIC_{lm} and the Add_{lm} at sparse level $s = 20\%$, while the QQ_{lm} has slightly better performance than the BIC_{lm} and the Add_{lm} at sparse level $s = 50\%$. One possible explanation is that when $p = 50$, the number of parameters reaches 150, which is more than the sample size $n = 100$. Such a situation would be in favor of QQ models with more sparse levels to gain better prediction accuracy.

Furthermore, Table 2 examines the performance of variable selection in terms of the number of false selection for the four examples. Here the number of false selection for the linear models, i.e., QQ_{lm}, BIC_{lm} and Add_{lm}, is calculated by the average of the number of false selections with respect to the two conditional linear regression models on $y|z = 1$ and $y|z = 0$. The results in Table 2 show that the proposed QQ models generally have better variable selection accuracy than the SMBIC approach and the MABIC approach. In Examples 1 and 3, note that the linear models of $y|z = 1, \boldsymbol{x}$ and $y|z = 0, \boldsymbol{x}$ have the same significant variables. In these situations, it appears that the QQ_{logit} has better selection accuracy than the BIC_{logit} and the Add_{logit}, while the variable selection accuracy of QQ_{lm} is

Table 2: Averages and Standard Errors of Number of False Selections from 50 Simulation Runs.

Example	p	Method	$\rho = 0$		$\rho = 0.5$	
			$s = 20\%$	$s = 50\%$	$s = 20\%$	$s = 50\%$
1	20	QQ_{lm}	1.74 (0.15)	2.44 (0.13)	5.31 (0.24)	1.92 (0.16)
		BIC_{lm}	0.84 (0.16)	0.32 (0.07)	0.60 (0.14)	0.24 (0.07)
		Add_{lm}	0.82 (0.16)	0.38 (0.07)	0.60 (0.14)	0.28 (0.07)
		QQ_{logit}	0.50 (0.10)	1.52 (0.14)	1.84 (0.16)	3.74 (0.22)
		BIC_{logit}	5.86 (0.52)	4.80 (0.30)	5.56 (0.49)	7.04 (0.29)
		Add_{logit}	6.50 (0.55)	5.58 (0.43)	6.40 (0.56)	8.22 (0.37)
1	50	QQ_{lm}	2.88 (0.48)	20.79 (0.57)	11.34 (0.64)	7.28 (0.44)
		BIC_{lm}	3.08 (0.40)	1.72 (0.25)	2.86 (0.34)	3.26 (0.35)
		Add_{lm}	3.04 (0.39)	1.72 (0.25)	2.90 (0.35)	3.44 (0.36)
		QQ_{logit}	8.04 (0.17)	16.72 (0.35)	5.74 (0.26)	20.50 (0.36)
		BIC_{logit}	9.12 (0.45)	23.12 (0.45)	11.04 (0.44)	26.80 (0.40)
		Add_{logit}	9.46 (0.52)	25.48 (0.60)	12.56 (0.60)	27.98 (0.55)
2	20	QQ_{lm}	1.95 (0.15)	1.76 (0.19)	0.82 (0.09)	1.60 (0.10)
		BIC_{lm}	4.40 (0.21)	6.38 (0.27)	4.56 (0.23)	6.80 (0.33)
		Add_{lm}	4.38 (0.20)	6.42 (0.28)	4.54 (0.23)	6.90 (0.31)
		QQ_{logit}	3.54 (0.25)	5.18 (0.29)	3.60 (0.26)	7.02 (0.31)
		BIC_{logit}	5.58 (0.55)	4.96 (0.33)	6.12 (0.51)	6.78 (0.32)
		Add_{logit}	6.44 (0.60)	5.70 (0.36)	6.66 (0.54)	7.62 (0.40)
2	50	QQ_{lm}	1.12 (0.29)	24.24 (0.47)	2.28 (0.21)	6.46 (0.52)
		BIC_{lm}	14.02 (0.73)	16.80 (0.73)	14.14 (0.60)	17.96 (0.65)
		Add_{lm}	14.10 (0.75)	17.38 (0.75)	14.36 (0.61)	18.16 (0.68)
		QQ_{logit}	10.56 (0.32)	20.96 (0.43)	10.50 (0.34)	22.48 (0.45)
		BIC_{logit}	9.16 (0.41)	22.40 (0.47)	11.08 (0.41)	26.66 (0.40)
		Add_{logit}	9.28 (0.40)	23.94 (0.61)	11.76 (0.42)	27.34 (0.53)
3	20	QQ_{lm}	0.16 (0.09)	0.81 (0.05)	0.98 (0.17)	4.11 (0.09)
		BIC_{lm}	0.72 (0.11)	1.00 (0.17)	1.14 (0.21)	3.02 (0.18)
		Add_{lm}	0.84 (0.14)	1.00 (0.17)	1.10 (0.21)	2.84 (0.20)
		QQ_{logit}	1.78 (0.15)	7.14 (0.24)	1.70 (0.12)	7.22 (0.17)
		BIC_{logit}	5.54 (0.52)	4.82 (0.34)	5.86 (0.56)	6.24 (0.28)
		Add_{logit}	6.54 (0.59)	5.98 (0.45)	6.28 (0.55)	6.74 (0.34)
3	50	QQ_{lm}	2.41 (0.47)	23.82 (0.47)	2.78 (0.32)	17.06 (0.40)
		BIC_{lm}	3.28 (0.38)	8.88 (0.64)	5.08 (0.58)	11.08 (0.51)
		Add_{lm}	3.36 (0.42)	8.92 (0.67)	5.44 (0.60)	11.20 (0.52)
		QQ_{logit}	6.82 (0.30)	17.58 (0.36)	7.20 (0.22)	16.00 (0.34)
		BIC_{logit}	9.88 (0.44)	23.54 (0.39)	11.24 (0.48)	25.92 (0.38)
		Add_{logit}	10.34 (0.53)	24.86 (0.55)	11.88 (0.58)	28.04 (0.61)
4	20	QQ_{lm}	1.96 (0.16)	0.35 (0.09)	0.09 (0.06)	2.23 (0.16)
		BIC_{lm}	4.54 (0.25)	5.70 (0.32)	4.56 (0.23)	5.40 (0.27)
		Add_{lm}	4.54 (0.25)	5.96 (0.33)	4.54 (0.23)	5.40 (0.27)
		QQ_{logit}	2.38 (0.26)	6.36 (0.33)	3.50 (0.17)	5.00 (0.34)
		BIC_{logit}	5.38 (0.55)	4.98 (0.32)	5.30 (0.45)	7.32 (0.29)
		Add_{logit}	6.12 (0.59)	6.20 (0.41)	5.92 (0.52)	8.24 (0.35)
4	50	QQ_{lm}	2.69 (0.40)	19.48 (0.44)	4.97 (0.21)	11.55 (0.49)
		BIC_{lm}	14.20 (0.70)	21.04 (0.72)	15.42 (0.70)	21.66 (0.62)
		Add_{lm}	14.60 (0.63)	21.98 (0.65)	15.34 (0.64)	21.88 (0.61)
		QQ_{logit}	10.58 (0.33)	19.42 (0.47)	10.96 (0.41)	20.34 (0.41)
		BIC_{logit}	8.70 (0.43)	23.48 (0.41)	10.34 (0.48)	25.98 (0.43)
		Add_{logit}	10.02 (0.48)	24.84 (0.54)	10.98 (0.51)	27.14 (0.53)

comparable to the BIC_{lm} and the Add_{lm}. An intuitive explanation is that when the linear models of $y|z = 1, \boldsymbol{x}$ and $y|z = 0, \boldsymbol{x}$ have the same significant variables, the conditional model $y|z, \boldsymbol{x}$ in (2.2) only reflects the role of z through the values of estimated coefficients, not on the role of what the significant variables are. It makes the variable selection accuracy of QQ_{lm} comparable to those of BIC_{lm} and Add_{lm}. In Examples 2 and 4, when the linear models of $y|z = 1, \boldsymbol{x}$ and $y|z = 0, \boldsymbol{x}$ become different in significant variables, one can clearly see that the proposed method gains the superiority of variable selection accuracy for both QQ_{lm} and QQ_{logit}. Moreover, as the dimensionality p increases in these examples, the advantages of the proposed method on variable selection become more significant, especially when $s = 20\%$. With the sample size $n = 100$ fixed in this study, the increase of p would make the QQ models having more advantage when the underlying model is sparse.

To check whether the fitted QQ models are over-sparse or under-sparse, we also report the number of false positives and the number of false negatives in Part B of Supplemental Materials. The results show that the number of false positive for the QQ models are small in general, indicating that the fitted QQ models would not be under-sparse. In addition, the number of false positive from the QQ models are much smaller than those from the SMBIC and the MABIC approaches in most cases. For the number of false negatives, we can see that under the dimensionality $s = 20\%$ and $p = 20$, the number of false positives for QQ models are generally smaller than those from the SMBIC and the MABIC approaches. But when the sparsity $s = 50\%$ and $p = 50$, the number of false positives for QQ models become relatively large, indicating that the fitted models can be over-sparse to some extent. One possible explanation is that $p = 50$ means that the number of parameters in QQ model is 150, larger than the sample size $n = 100$. In this situation, the QQ models tends to pursue sparse models to get accuracy on the predication.

In this work, the proposed QQ models assume a linear model for the quantitative response y with normal distributed errors. To evaluate the robustness of the proposed QQ models, we first conduct a set of simulations to check the normality assumption of the linear model residuals $\epsilon = y - z\boldsymbol{x}'\boldsymbol{\beta}^{(1)} + (1 - z)\boldsymbol{x}'\boldsymbol{\beta}^{(2)}$. Specifically, we generate the simulation data following the aforementioned procedure, except changing the distribution of the linear model

residuals from a normal distribution to a Chi-squares distribution (skewed) and a t distribution (heavy tail) with degrees of freedom 5, respectively. We also scale the Chi-square and t distributions to have the same variance $\sigma^2 = 1$ as the normal distribution used in linear model residuals. The prediction and variable selection performance of the QQ models and the SMBIC approach are reported in Part B of Supplemental Materials. We found that the QQ models generally have more accurate prediction and variable selection than the SMBIC approach in Examples 2-4, when the underlying distributions deviate from a normal distribution. The QQ models are robust to the normality assumption of residuals to some extent, since the prediction and variable selection performance of the QQ models for Chi-squares distributed residuals and t distributed residuals are comparable to those of the QQ models for normally distributed residuals. Second, we evaluate the robustness of the linear model assumption for the quantitative response y. Specifically, we generate the simulation data following the aforementioned procedure, but taking cubic of the regression mean in (2.2) as $z(\boldsymbol{x}'\boldsymbol{\beta}^{(1)})^3 + (1 - z)(\boldsymbol{x}'\boldsymbol{\beta}^{(2)})^3$. Thus the underlying model for the quantitative response becomes nonlinear. We fit the data using the QQ models and SMBIC approach with a linear model for the quantitative response. From the results in Part B of Supplemental Materials, both the QQ_{lm} and the BIC_{lm} yield large prediction errors due to the improper assumptions of the model structure. But the QQ_{lm} gives relatively smaller RMSPEs than the SMBIC approach, which may be explained by the flexibility of the QQ models: the quantitative response is conditioned on the different values of z. For the variable selection performance, the QQ_{lm} does not clearly outperform the BIC_{lm} in various cases. It appears that when the linear model structure assumption is not valid, the proposed QQ models may not perform well. Some discussions are provided in Section 6 on how to address this issue.

5 Case Study in the Lapping Process

To further illustrate the merits of the QQ models, we analyze the data from a real case study in the lapping process. Recall the lapping process introduced in Section 1, there are 10 predictors and two quality responses, which are summarized in Part C of Supplemental Materials. In this process, there are four process variables: pressure, rotation speed, lapping
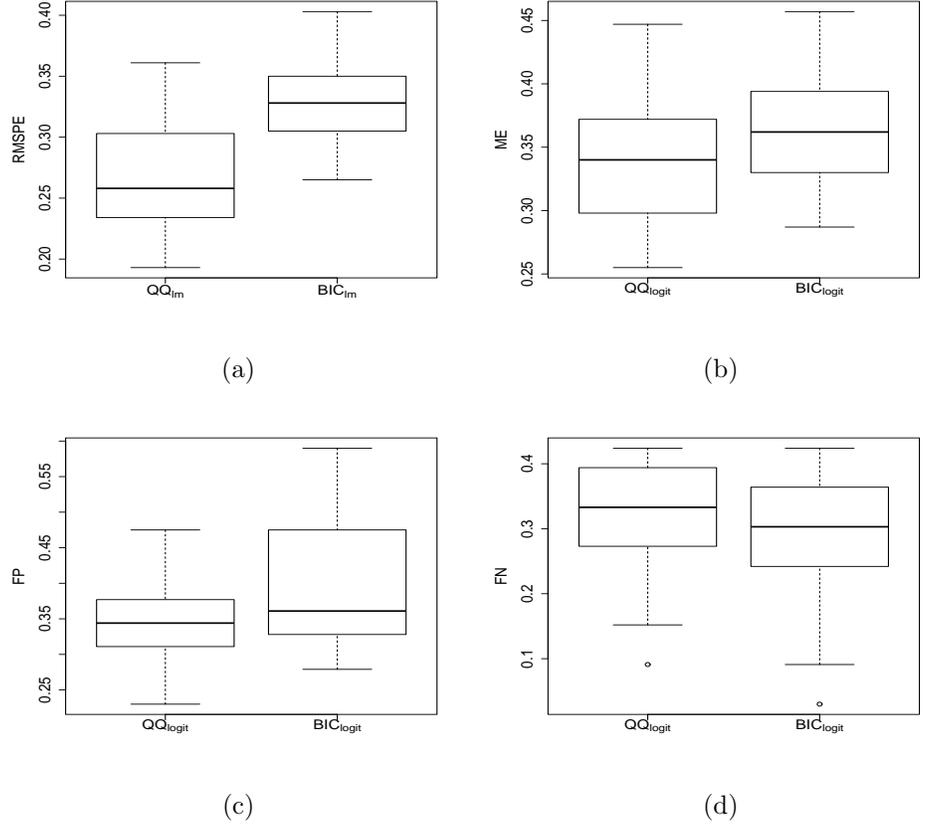
Figure 2: Comparison of Prediction Accuracy of the QQ Models and SMBIC for the Lapping Process: (a) Root mean squared errors (RMSPE); (b) misclassification error (ME); (c) false positive classification error (FP); (d) false negative classification errors (FN).

time for low pressure, and lapping time for high pressure. In addition, there are six initial quality covariates before the lapping process. This collected data set contains 254 wafer observations, where 203 wafers have the STIR indicator as 0 (good), and the remaining 51 wafers have the STIR indicator as 1 (bad). To evaluate the performance of the proposed QQ models, we randomly partition the data set into a training data set (50%) and a test data set (50%) with a stratified sampling strategy. Such random partitions repeat for 50 times. For each partition, the proposed QQ models and the SMBIC approach are applied to model the quantitative response TTV and the qualitative response STIR indicator based on the training data set. The prediction errors of the test data set and the number of selections for each predictor variable of the training data set are computed for both methods.

Figure 2 shows the boxplots of RMSPEs (MEs) for the QQ models and the SMBIC approach under the 50 replicates of data partitions. We did not compare the performance of the MABIC approach here because neither of QQ responses will be available and they need to be predicted in real practice. From Figure 2(a), it is clear that the proposed QQ models have much smaller RMSPEs than the SMBIC approach. Figure 2(b) also shows that the misclassification errors of the proposed method are also smaller than that from the SMBIC approach. As shown in Figure 2(c)-(d), the smaller misclassification error of the QQ models is mainly due to the smaller false positive classification error; while the false negative classification errors are comparable for the two methods. The results indicate that the proposed joint modeling of QQ responses enhanced the prediction performance in this case.

Table 3 reports the average numbers of selected variables and their standard errors (in the parenthesis) for both methods. Here the number of selected variables for the linear models, i.e., QQ_{lm} and BIC_{lm}, is calculated by the average of the number of selections with respect to the two conditional linear regression models on $y|z = 1, \boldsymbol{x}$ and $y|z = 0, \boldsymbol{x}$. From the results in Table 3, we can see that the quality covariates $x_5 - x_{10}$ are often selected by QQ_{lm} and QQ_{logit}. While the BIC_{lm} and BIC_{logit} consistently ignore the quality covariates in the models. In fact, the quality covariates are expected to be important in the model if the STIR indicator is 1. From the engineering perception, the lapping process is likely to be nonconforming if the STIR indicator is 1. In this case, the quality covariates representing the initial quality of wafers become important factors for after-lapping wafer quality. The proposed QQ models successfully capture such engineering perception. In contrast, the SMBIC approach considers modeling the two responses TTV and STIR indictor separately, which cannot unveil the significance of quality covariates in this case study. In summary, the proposed QQ models successfully discover this hidden information, which can be further used for quality control and process improvement.

We also evaluate the model assumption for this real-data example. The Q-Q normal plots for the residuals in the model $y|z = 1, \boldsymbol{x}$ and the residuals in the model $y|z = 0, \boldsymbol{x}$ are provided in Part C of Supplemental Materials. From the Q-Q plots, we can see that the distribution

Table 3: Averages and Standard Errors of Selection Percentage from 50 Replicates of Data Partitions for the Lapping Process.

Method	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
QQ_{lm}	0.74	0.62	0.88	0.88	0.65	0.60	0.68	0.59	0.66	0.65
	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)
BIC_{lm}	0.52	0.10	1.00	1.00	0.00	0.02	0.02	0.00	0.00	0.08
	(0.07)	(0.04)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.00)	(0.00)	(0.04)
QQ_{logit}	0.44	0.52	0.64	0.70	0.32	0.18	0.36	0.32	0.22	0.44
	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.05)	(0.07)	(0.07)	(0.06)	(0.07)
BIC_{logit}	0.14	0.04	0.68	0.86	0.00	0.00	0.00	0.04	0.00	0.00
	(0.05)	(0.03)	(0.07)	(0.05)	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)

of the residuals would be close to normal distribution after a linear transformation. It appears that the normality assumption of the residuals may not strictly hold. Note that the prediction performance of the regression models in QQ models is still better than the benchmark models. It appears being reasonable to use the proposed models for prediction. This is also consistent with the conclusion drawn from the simulation, that the proposed QQ models are robust to the normally distributed residual assumption. We also examine the residuals over fitted responses in the model $y|z = 1, \boldsymbol{x}$ and the model $y|z = 0, \boldsymbol{x}$, see Part C of Supplemental Materials. It is clear that for both models the residuals have constant variance, and forms several clusters along fitted response. This is mainly because the lapping data are collected from the design of experiments settings, where the fitted responses are clustered. In addition, the random patterns appeared in the residual plots indicate that the linear model assumption seems to be reasonable for modeling y.

For the two conditional models $y|z = 1, \boldsymbol{x}$ and $y|z = 0, \boldsymbol{x}$, we assume that the two models have the same error variance, i.e., $var(y|z = 1, \boldsymbol{x}) = var(y|z = 0, \boldsymbol{x})$. To check this assumption, we check if the standard errors in the model $y|z = 1, \boldsymbol{x}$ and the model $y|z = 0, \boldsymbol{x}$ are the same by using the F-test. Based on the fitted QQ models, the standard error for the model $y|z = 1, \boldsymbol{x}$ is 0.3135, and the standard error of the error terms for the model $y|z = 0, \boldsymbol{x}$ is 0.2219. The p-value of the F-test is 0.0002, indicating that the standard

errors are different in this case study. Recall the model in (2.2), the proposed method can be easily extended to allow the two conditional models having two different variances without changing the nature of the problem.

To check the logistic model assumption in modeling $z|\boldsymbol{x}$, we performed model diagnostics by using χ^2 test for deviance. For the logistic regression obtained from the QQ models, the null deviance is 433.91, and the model deviance is 398.13. The p-value of the χ^2 test is 1.7×10^{-8}. Thus, it shows that the logistic regression model obtained from the QQ models have a reasonable goodness of fit for the real data.

6   Discussion

A QQ system is widely encountered in manufacturing processes, such as a lapping process, a solar cell lamination process, and nanostructure synthesis. With both QQ responses of quality closely associated, the two types of quality responses often share the same set of potential root causes. Therefore, a joint modeling framework is needed for integrating both types of quality responses. In this paper, we propose a joint likelihood modeling framework called QQ models. The proposed models consider the joint probability of the quantitative and qualitative responses, while the constrained likelihood estimation approach is used for parameter estimation and variable selection. A fast algorithm of parameter estimation is also developed by quadratic approximation to facilitate fast computation. We use both simulation studies and a real case study from the lapping process to demonstrate the effectiveness of the QQ models. Especially in the case study, the proposed QQ models yield better prediction and more meaningful variable selection, which reflect the inherent features of the real manufacturing process.

Note that for complex QQ systems, the quality-process relationship can be highly non-linear. The proposed method can be extended to the nonlinear models. One possibility is to incorporate projection pursuit (Friedman and Stuetzle, 1981) into the proposed QQ models to allow more flexibility on the predictive functions. Specifically, we can extend the models

in (2.1) and (2.2) as

$$z|\boldsymbol{x} = \begin{cases} 1, & w.p. \ p(\boldsymbol{x}) \\ 0, & w.p. \ 1 - p(\boldsymbol{x}) \end{cases} \quad \text{with } p(\boldsymbol{x}) = \frac{\exp(v(\boldsymbol{x}'\boldsymbol{\eta}))}{1 + \exp(v(\boldsymbol{x}'\boldsymbol{\eta}))},$$

$$y|z, \boldsymbol{x} \sim N(zh_1(\boldsymbol{x}'\boldsymbol{\beta}^{(1)}) + (1-z)h_2(\boldsymbol{x}'\boldsymbol{\beta}^{(2)}), \sigma^2),$$

where h_1, h_2 and v are nonparametric functions. The nonparametric forms of h_j and v allow flexible functional structures to be estimated. For the parameter estimation, an efficient algorithm can be developed by iterating the estimation between functions h_1, h_2, v and parameters $\boldsymbol{\eta}$, $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$. Another possibility of extending the proposed QQ models is to adopt the nonparametric methods (Qiu, 2014) to relax the normality assumption, which can be an interesting topic for our future research.

Although we focus on one quantitative response and one qualitative response in this work, the proposed method can be generalized to the case of multiple responses. Suppose the multiple quantitative responses are $y_1, \ldots, y_m$ and multiple binary responses are $z_1, \ldots, z_t$. For example, a multi-level qualitative response can be transformed into a set of dummy binary responses. For $m > 1$ and $t = 1$, we can generalize the QQ models by multi-response regression (Breiman and Friedman, 1997) as

$$(y_1, \ldots, y_m)|z, \boldsymbol{x} \sim N(z\boldsymbol{x}'\boldsymbol{B}^{(1)}) + (1-z)\boldsymbol{x}'\boldsymbol{B}^{(2)}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{B}^{(1)}, \boldsymbol{B}^{(2)}$ are coefficient matrices, and $\boldsymbol{\Sigma}$ is a covariance matrix. For $m > 1$ and $t > 1$ with multiple binary responses, considering all 2^t conditional models $(y_1, \ldots, y_m|z_1 = 1, \ldots, z_t = 1), \ldots, (y_1, \ldots, y_m|z_1 = 0, \ldots, z_t = 0)$ may only work for a small t. Alternatively, we can develop a two-stage approach by first using independence screening (Fan and Lv, 2008) to select important qualitative responses. That is to conduct t independent conditional models $(y_1, \ldots, y_m|z_1), \ldots, (y_1, \ldots, y_m|z_t)$ and select $u < t$ qualitative responses z_j's with best fitting. Then the full conditional models will be applied for the selected u qualitative responses. For the joint distribution of u qualitative responses, we can consider multi-logit models (McCullagh, 1980) and Ising models (Ravikumar et al., 2010) as possible solutions.

The proposed method can be extended to accommodate interaction effects of the predictors. The hierarchical and heredity principles (Wu and Hamada, 2009) can also be incorpo-

rated under the nonnegative garrote approach for variable selection. Other types of variable selection and group variable selection techniques, such as lasso (Tibshirani, 1996) and group non-negative garrotes (Yuan and Lin, 2006), will be investigated for the proposed method. A Bayesian modeling framework will also be developed to integrate information for both types of responses.

We would like to point out that there can be other possibilities to model the joint distribution of QQ responses. In the proposed QQ models, we have considered the joint distribution $f(y, z|\boldsymbol{x}) = f(y|z, \boldsymbol{x})f(z|\boldsymbol{x})$ for model construction. Alternatively, one may consider the model construction based on $f(y, z|\boldsymbol{x}) = f(z|y, \boldsymbol{x})f(y|\boldsymbol{x})$. However, it is not clear how to appropriately quantify $f(z|y, \boldsymbol{x})$ in a general formulation, which can be an interesting topic for future research. When the association of QQ responses becomes more complicated than the conditional distribution, it may not be appropriate to use the QQ models. Other techniques such as nonparametric methods can be potentially useful to deal with the related research problems.

Acknowledgement

References

Apley, D. and Shi, J. (1998). Diagnosis of Multiple Fixture Faults in Panel Assembly. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **120**, 793–801.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Breiman, L. (1995). Better Subset Regression using the Nonnegative Garrote, *Technometrics*, **37**, 373–384.

Breiman, L. and Friedman, J. H. (1997) Predicting Multivariate Responses in Multiple Linear Regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **59**, 3–54.

Cantoni, E., Flemming, J. M., and Ronchetti, E. (2011). Variable Selection in Additive Models by Non-negative Garrote, *Statistical Modelling*, **11(3)**, 237–252.

Chen, S., Witten, D., and Shojaie, A. (2014). Selection and Estimation for Mixed Graphical Models, *arXiv:1311.0085*.

Draper, N. R. and John, J. A. (1988). Response-Surface Designs for Quantitative and Qualitative Variables. *Technometrics*, **30**, 423–428.

Fan, J. and Lv, J. (2008) Sure Independence Screening for Ultrahigh Dimensional Feature Space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**, 849–911.

Fitzmaurice, G.M. and Laird, N.M. (1997). Regression Models for Mixed Discrete and Continuous Responses with Potentially Missing Values, *Biometrics*, **53**, 110–122.

Fong, D. and Lawless, J. F. (1998). The Analysis of Process Variation Transmission with Multivariate Measurements, *Statistica Sinica*, **8**, 151–164.

Friedman, J. H. and Stuetzle, W. (1981) Projection Pursuit Regression. *Journal of the American Statistical Association*, **76**, 817–823.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition*, Springer-Verlag.

Hawkins, D.M. (1991). Multivariate Quality Control Based on Regression-Adjusted Variables, *Technometrics*, **33**, 61–75.

Hawkins, D.M. (1993). Regression Adjustment for Variables in Multivariate Quality Control, *Journal of Quality Technology*, **25**, 170–182.

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

Jin, J. and Ding, Y. (2004). Online Automatic Process Control using Observable Noise Factors for Discrete Part Manufacturing. *IIE Transactions*, **36**, 899–911.

Jin, R., and Shi, J. (2012). Reconfigured Piecewise Linear Regression Tree for Multistage Manufacturing Process Control, *IIE Transactions*, **44**, 249–261.

Joseph, R. (2003). Robust Parameter Design with Feed-Forward Control, *Technometrics*, **45**, 284–292.

Lin, J. and Wang, K. (2011). Online Parameter Estimation and Run-to-Run Process Adjustment using Categorical Observations, *International Journal of Production Research*, **49**, 4103–4117.

McCullagh, P. (1980) Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **42**, 109–142.

Makalic E. and Schmidt, D. F. (2011). Logistic Regression with the Nonnegative Garrote. *Lecture Notes in Computer Science*, **7106**, 82–91.

Marinescu, I. D., Uhlmann, E., and Doi, T. K. (2007). *Handbook of Lapping and Polishing*, CRC Press, Boca Raton, FL.

May, G.S., Huang, J., and Spanos, C.J. (1991). Statistical Experimental Design in Plasma Etch Modeling, *IEEE Transactions on Semiconductor Manufacturing*, **4**, 83–98.

Miller, A. J. (2002). *Subset Selection in Regression, 2nd ed.* Chapman and Hall, London.

Montgomery, D.C. (2001). *Introduction to Statistical Quality Control, 4th ed.*, Wiley, New York, NY.

Ning, Y., Bian, Y. and Liu, B. (2012). Improving a Lapping Process using Robust Parameter Design and Run-to-Run Control, *Journal of the Chinese Institute of Industrial Engineers*, **29(2)**, 111–124.

O'Mara, W. C., Herring, R. B., and Hunt, L. P. (1990). *Handbook of Semiconductor Silicon Technology*, Noyes Publications, Park Ridge, NJ.

Paggi, M., Corrado, M., and Rodriguez, M. A. (2013). A Multi-physics and Multi-scale Numerical Approach to Microcracking and Power-loss in Photovoltaic Modules, *Composite Structures*, **95**, 630–638.

Pilla, M., Galmiche, F. and Maldague, X. P. (2002). Thermographic Inspection of Cracked Solar Cells, *Proc. SPIE 4710*, Thermosense XXIV, 699-703.

Qiu, P. (2008). Distribution-free Multivariate Process Control based on Log-linear Modeling. *IIE Transactions*, **40**, 664-677.

Qiu, P. (2014). *Introduction to Statistical Process Control*, Chapman&Hall/CRC, Boca Raton, FL.

Ravikumar, P., Wainwright, M. J., and Lafferty, J. D. (2010) High-dimensional Ising Model Selection using l_1-regularized Logistic Regression.*The Annals of Statistics*, **38**, 1287–1319.

Shi, J. (2006). *Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*, CRC Press, Taylor&Francis Group.

Spanos, C. J. and Chen, R. L. (1997). Using Qualitative Observations for Process Tuning and Control, *IEEE Transactions on Semiconductor Manufacturing*, **10**, 307–316.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.

Wang, K.B. and Tsung, F. (2007). Run-to-Run Process Adjustment Using Categorical Observations, *Journal of Quality Technology*, **39**, 312–325.

Woodall, W. H. (2000). Controversies and Contradictions in Statistical Process Control. *Journal of Quality Technology*, **32**, 341–350.

Wu, C. F. J., and Hamada, M. (2009). *Experiments: Planning, Analysis, and Optimization, 2nd Edition*, Wiley.

Xiong, S. (2012). Some Notes on the Nonnegative Garrote. *Technomatrics*, **52**, 349–361.

Yuan, M, and Lin, Y. (2006). Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68(1)**, 49–67.

Yang, E., Baker, Y., Ravikumar, P., Allen, G. I., and Liu, Z. (2014). Mixed Graphical Models via Exponential Families. In proceedings of *International Conference on Artificial Intelligence and Statistics (AISTATS)*, **33**, 1042–1050.

Yuan, M., and Lin, Y. (2007). On the Nonnegative Garrote Estimator. *Journal of the Royal Statistical Society, Series B*, **69(2)**, 143-161.

Zhao, H., Jin, R., Wu, S., and Shi, J. (2011). PDE-constrained Gaussian Process Model on Material Removal Rate of Wire Saw Slicing Process, *ASME Transactions, Journal of Manufacturing Sciences and Engineering*, **133**, 021012.

Zhou, S., Ding, Y, Chen, Y., and Shi, J. (2003). Diagnosability Study of Multistage Manufacturing Processes Based on Linear Mixed-effects Models, *Technomatrics*, **45**, 312–325.