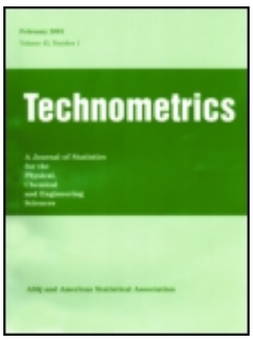

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

A Bayesian Approach for Interpreting Mean Shifts in Multivariate Quality Control

Matthias H. Y. Tan & Jianjun Shi

To cite this article: Matthias H. Y. Tan & Jianjun Shi (2012) A Bayesian Approach for Interpreting Mean Shifts in Multivariate Quality Control, Technometrics, 54:3, 294-307, DOI: 10.1080/00401706.2012.694789

To link to this article: https://doi.org/10.1080/00401706.2012.694789

Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

A Bayesian Approach for Interpreting Mean Shifts in Multivariate Quality Control

Matthias H. Y. TAN and Jianjun SHI

H. Milton Stewart School of Industrial
and Systems Engineering, College of
Engineering, Georgia Institute of Technology
Atlanta, GA 30332
(*mtan6@gatech.edu*; *jianjun.shi@isye.gatech.edu*)

Multivariate quality characteristics are often monitored using a single statistic or a few statistics. However, it is difficult to determine the causes of an out-of-control signal based on a few summary statistics. Therefore, if a control chart for the mean detects a change in the mean, the quality engineer needs to determine which means shifted and the directions of the shifts to facilitate identification of root causes. We propose a Bayesian approach that gives a direct answer to this question. For each mean, an indicator variable that indicates whether the mean shifted upward, shifted downward, or remained unchanged is introduced. Prior distributions for the means and indicators capture prior knowledge about mean shifts and allow for asymmetry in upward and downward shifts. The mode of the posterior distribution of the vector of indicators or the mode of the marginal posterior distribution of each indicator gives the most likely scenario for each mean. Evaluation of the posterior probabilities of all possible values of the indicators is avoided by employing Gibbs sampling. This renders the computational cost more affordable for high-dimensional problems. This article has supplementary materials online.

KEY WORDS: Fault isolation; Gibbs sampling; Hierarchical Bayes; Interpretation of out-of-control signal; Multivariate statistical process control; Variable selection.

1. INTRODUCTION

The performance or quality of a process is often characterized by multiple variables. Thus, effective control of a process can only be achieved by jointly monitoring all relevant variables. Yeh, Lin, and McGrath (2006) and Bersimis, Psarakis, and Panaretos (2007) reviewed the literature on control charts for the covariance matrix. Control charts for the mean include the well-known Hotelling's T^2 charts and multivariate cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) charts (Bersimis, Psarakis, and Panaretos 2007). Monitoring multivariate quality characteristics using a small number of summary statistics is a common practice. However, these statistics does not effectively support the diagnosis task, which is to determine the cause of the out-of-control signal.

Identification of mean shifts among a large number of quality characteristics often provides important information for the diagnosis task. For example, the quality control of fruit juice can be performed by comparing randomly selected samples against base samples using a multivariate control chart for amino acids and other constituents. If the control chart indicates differences between samples, information on the identity of the variables whose mean shifted and the shift directions can help to determine the source of adulteration of the juice. For instance, Zhang et al. (2009) gave criteria for identifying the various types of adulterations of pomegranate juice (e.g., amino acid proline >25 mg/L is indicative of added grape products).

This article proposes a Bayesian method for identifying the means that shifted and the directions of the shifts. The method provides a new diagnostic tool for phase II monitoring. For p quality characteristics, there are 3^p possible scenarios for the means since each mean can increase, decrease, or remain in control. In the Bayesian approach, each scenario corresponds to a value of an indicator vector and the most probable of the 3^p scenarios is found by sampling from the posterior distribution of the indicator via Gibbs sampling.

The problem of interpreting an out-of-control signal from a T^2 control chart has been widely studied in the literature. Many of the proposed approaches attempt to identify a subset of variables that has the most significant contribution to the large observed T^2 value by decomposing T^2. This is the basic idea underpinning Murphy (1987); Doganaksoy, Faltin, and Tucker (1991); Runger, Alt, and Montgomery (1996); and Mason, Tracy, and Young (1995, 1997). Mason, Tracy, and Young (1995, 1997) proposed decomposing the T^2 statistic into independent components. However, there are several problems with the T^2 decomposition method. First, for p variables, there are $p!$ possible decompositions; this makes the T^2 decomposition impractical for high-dimensional problems. Second, there are no clear-cut rules for jointly interpreting the components of the decomposition. Third, results are sensitive to the significance levels used. Li, Jin, and Shi (2008) introduced a solution to the first two problems, which is to use a Bayesian causal network that describes the causal relationship between variables.

Hawkins (1991) proposed a procedure that is based on the likelihood-ratio tests of a shift in each mean. A comparison of the approaches proposed by Murphy (1987); Doganaksoy,

TECHNOMETRICS, AUGUST 2012, VOL. 54, NO. 3
DOI: 10.1080/00401706.2012.694789

Faltin, and Tucker (1991); Hawkins (1991); and Mason, Tracy, and Young (1995) is given by Das and Prakash (2008).

Recently, Wang and Jiang (2009) proposed a penalized likelihood variable selection method to identify variables with shifted means. Zou, Jiang, and Tsung (2011) proposed a heuristic Bayesian information criterion (BIC) for shift detection; the search for a model that minimizes the BIC criterion is restricted to those models that are optimal with respect to an adaptive-LASSO-type penalized likelihood. Their method can be applied to diagnose both covariance and mean shifts. Hereafter, we abbreviate the methods proposed by Wang and Jiang (2009) and Zou, Jiang, and Tsung (2011) as WJPLM and LEB, respectively. Capizzi and Masarotto (2011) and Zou and Qiu (2009) proposed EWMA-type control charts that employ test statistics based on the least angle regression algorithm and the adaptive-LASSO-type penalized likelihood function, respectively. Diagnostic information is a by-product of the charting statistics.

Unlike most of the available techniques in the literature, the Bayesian approach described in this article gives more specific and direct information about shifts in the mean, that is, it gives the means that shifted upward, those that shifted downward, and those that remained in control. It allows prior knowledge to be incorporated in a statistical framework. We believe that this is an advantage rather than a disadvantage since in all cases, engineering knowledge must be used to identify assignable causes whenever a control chart signals.

Our approach is inspired by George and McCulloch's (1993) Bayesian stochastic search variable selection (SSVS) approach. However, it is *different* from SSVS. Our approach is developed for solving the problem of comparing the means of two different populations, that is, phase I and phase II process means. In contrast, SSVS is developed for solving the problem of variable selection in regression. SSVS is limited to independent and identically distributed residual error settings, and there are only two decisions per variable (whether the coefficient is zero or nonzero). On the other hand, our approach is for general multivariate normal variables and there are three decisions per variable. Furthermore, the priors we use can capture prior information such as most likely shift magnitudes that have values different from zero, and different ranges and probabilities of upward and downward shifts. These features cannot be modeled by SSVS. Finally, we also develop an empirical Bayes' (EB) method for specifying some of the prior parameters.

Unlike our approach, WJPLM and LEB can only incorporate prior information via assumptions (such as assumptions on the number of variables that shifted and allowable shift directions), which have strong effects on results. Moreover, these approaches ignore uncertainty incurred in estimating the in-control means and covariance matrix (the phase I estimates are assumed to be equal to the population parameters), whereas our approach takes this uncertainty into account. Ignoring estimation uncertainty may not be justifiable when the phase I sample size is small. However, the proposed approach incurs a high computation cost, except when compared with the best subset variant of WJPLM. Note that our approach cannot be compared with the methods proposed by Capizzi and Masarotto (2011) and Zou and Qiu (2009) as those methods are for simultaneous monitoring and diagnosis. Hence, in this article, we shall only compare our approach with WJPLM and LEB.

The article is organized as follows. In Section 2, we briefly describe the mean diagnostic problem, state our assumptions, and review the T^2 chart and WJPLM. Section 3 gives our proposed Bayesian hierarchical model. Guidelines for specifying prior distributions are given in Section 4. Section 5 gives a Gibbs sampling procedure for sampling from the posterior distribution of the parameters, and decision rules for mean shifts. Section 6 presents three examples and Section 7 concludes the article.

2. MULTIVARIATE QUALITY CONTROL

This section describes the diagnostic problem addressed by this article and states the assumptions and notations used throughout the article. We also review the Hotelling's T^2 control chart, the t-test, and WJPLM. While the proposed approach can be used independently of any control chart, an interesting application of the approach is the diagnosis of an out-of-control signal given by a T^2 chart. Discussion of the T^2 chart also provides a nice context for introducing the mean shift diagnosis problem.

Throughout this article, we assume that the variables of interest $\mathbf{x} = (x_1, \ldots, x_p)^T$ have a multivariate normal distribution with in-control mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}$, that is, $\mathbf{x} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$. We let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ denote the N in-control phase I observations, and n denote the sample size of the phase II sample *suspected* to have been drawn from a common normal distribution different from $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$. All samples are assumed independent. Furthermore, in this article, we assume that the covariance matrix remains in control so that $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in phase II. The objective of the mean shift diagnosis problem is to identify which components of $\boldsymbol{\mu}$ are different from $\boldsymbol{\mu}_0$. We let $\bar{\mathbf{x}} = \sum_{i=1}^{N} \mathbf{x}_i/N$ and $\mathbf{S} = \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$ denote the sample mean and the sample dispersion matrix of the phase I sample respectively. The sample covariance matrix of the phase I sample is denoted by $\widehat{\boldsymbol{\Sigma}} = \mathbf{S}/(N-1)$. Similarly, we let $\bar{\mathbf{x}}_{\mathrm{f}}$ and $\mathbf{S}_{\mathrm{f}}$ denote the phase II sample mean and dispersion matrix respectively (the subscript f stands for future).

The standard control chart for monitoring $\mathbf{x}$ is the T^2 chart with statistic $T^2 = n(\bar{\mathbf{x}}_{\mathrm{f}} - \bar{\mathbf{x}})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\bar{\mathbf{x}}_{\mathrm{f}} - \bar{\mathbf{x}})$ plotted with an upper control limit (UCL). It can be shown that if each phase I observation is independently distributed as $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, and $\bar{\mathbf{x}}_{\mathrm{f}} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/n)$, then $T^2 \sim \frac{(N+n)(N-1)p}{N(N-p)} F(p, N-p)$, where $F(p, N-p)$ is the F-distribution with p and $N-p$ degrees of freedom. Thus, to control the Type I error at α, the UCL should be determined using the upper 100α percentile of $F(p, N-p)$.

When $T^2 >$ UCL, the control chart indicates that either the mean has shifted, that is, $\boldsymbol{\mu} = E(\bar{\mathbf{x}}_{\mathrm{f}}) \neq \boldsymbol{\mu}_0$, or the covariance matrix has changed, or both. This article assumes that the covariance matrix remains in control. Whether the assumption of in-control covariance matrix is reasonable can be decided through the use of control charts for the covariance matrix (see Yeh, Lin, and McGrath 2006) or through tests of the hypothesis $H_0 : \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1$, where $\boldsymbol{\Sigma}_0$ is the covariance matrix for the phase I samples and $\boldsymbol{\Sigma}_1$ is the covariance matrix for the phase II sample that triggered the out-of-control signal. A standard test for H_0 is the likelihood-ratio test (Timm 2002).

If it is determined that the mean shifted, supplementary information can be provided by computing t statistics for testing each of the hypotheses H_0: $\mu_{0i} = \mu_i$, where μ_{0i} is the ith component of $\boldsymbol{\mu}_0$, and μ_i is the ith component of $\boldsymbol{\mu}$. The t statistic for testing H_0: $\mu_{0i} = \mu_i$ is given by

$$t_i = \frac{\bar{x}_{fi} - \bar{x}_i}{\sqrt{\tilde{\sigma}_i^2\left(\frac{1}{n} + \frac{1}{N}\right)}}, \tag{1}$$

where $\tilde{\sigma}_i^2$ is the *pooled* sample variance for the ith variable (the sample variance of the combined phase I and phase II samples), and $\bar{x}_{fi}$ and $\bar{x}_i$ are the ith components of $\bar{\mathbf{x}}_f$ and $\bar{\mathbf{x}}$, respectively. We interpret the t_i's using the following decision rule:

$$\begin{aligned} &\text{If } t_i < -\gamma, \text{ then } \mu_i < \mu_{0i}; \text{if } t_i > \gamma, \text{ then } \mu_i > \mu_{0i};\\ &\text{if } -\gamma \le t_i \le \gamma, \text{ then } \mu_i = \mu_{0i}. \end{aligned} \tag{2}$$

The positive scalar γ is some suitably chosen cutoff.

A more modern approach to variable selection is WJPLM. This approach applies the forward selection method or the best subset method to the variable selection problem, where the response is given by the column vector $\mathbf{R}(\bar{\mathbf{x}}_f - \bar{\mathbf{x}})$, the regressors are given by the columns of $\mathbf{R}$, and $\mathbf{R}$ is defined by the Cholesky decomposition $(\widehat{\boldsymbol{\Sigma}}/n)^{-1} = \mathbf{R}^T\mathbf{R}$. The forward selection algorithm is terminated when the model size is L and the best subset method chooses the best model of size L, where L is specified by the engineer.

3. BAYESIAN HIERARCHICAL MODEL

In this section, we shall develop a Bayesian hierarchical model that can be used to determine the means that shifted and the directions of the shifts when given a suspected out-of-control phase II sample of size n. Prior information is obtained from phase I data, which is then combined with the likelihood for phase II to yield the desired posterior inference.

It can be shown that $\bar{\mathbf{x}} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/N)$, $\mathbf{S} \sim W(\boldsymbol{\Sigma}, N-1)$, where $W(\boldsymbol{\Sigma}, N-1)$ denotes a Wishart distribution with scale matrix $\boldsymbol{\Sigma}$ and $N-1$ degrees of freedom, $\bar{\mathbf{x}}$ and $\mathbf{S}$ are independent, and $(\bar{\mathbf{x}}, \mathbf{S})$ is a sufficient statistic for $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$. Thus, if we use noninformative priors for $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ given by $p(\boldsymbol{\mu}_0) \propto 1$, $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{p+1}{2}}$, the resulting posterior distributions for $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}^{-1}$ are (e.g., see Box and Tiao 1973, chap. 8)

$$p(\boldsymbol{\mu}_0|\bar{\mathbf{x}}, \mathbf{S}) \propto |1 + N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)|^{-\frac{N}{2}}, \tag{3}$$

$$p(\boldsymbol{\Sigma}^{-1}|\bar{\mathbf{x}}, \mathbf{S}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{N-p-2}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}. \tag{4}$$

For a phase II sample with sample mean $\bar{\mathbf{x}}_f \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$ and sample dispersion matrix $\mathbf{S}_f \sim W(\boldsymbol{\Sigma}, n-1)$, the likelihood is

$$\begin{aligned} l\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}|\bar{\mathbf{x}}_f, \mathbf{S}_f\right) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}n(\bar{\mathbf{x}}_f - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}_f - \boldsymbol{\mu})\right]\\ \times |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} \exp\left[-\frac{1}{2}\text{trace}\left(\boldsymbol{\Sigma}^{-1}\mathbf{S}_f\right)\right]. \end{aligned} \tag{5}$$

Note that if $n = 1$, we simply set $\mathbf{S}_f = \mathbf{0}$. It is of interest to determine which components of $\boldsymbol{\mu} - \boldsymbol{\mu}_0$ are nonzero and the signs of the nonzero components. A similar problem arises in the area of Bayesian variable selection (George and McCulloch 1993, 1997) in which the objective is to determine the variables with nonzero coefficients in a linear model. The key idea in Bayesian variable selection is to model each regression coefficient as a mixture of two distributions, where coefficients drawn from one distribution tend to be close to zero, and coefficients drawn from the other distribution tend to be large. Indicator variables are introduced to indicate the distribution from which a coefficient is drawn.

In a similar vein, we introduce indicator variables $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^T$ so that $\delta_i = -1$ indicates that μ_i has decreased, $\delta_i = 0$ indicates that μ_i has remained unchanged, and $\delta_i = 1$ indicates that μ_i has increased. The reason that we let δ_i take on three levels rather than two (one indicating that the ith mean shifted and the other indicating that it remained in control) is that this would allow us to handle cases where the most likely shift magnitudes are known to be large (instead of close to zero) and cases where we have asymmetrical prior information about upward and downward shifts.

The parameters of the Bayesian model *for phase II* are $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\Sigma}^{-1}$, and we need to specify the prior distribution $p(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1})$. Note that it is not justifiable to use the posterior of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$ from phase I as the prior for phase II. The phase II sample mean is expected to be different, and the proposed method is intended to unravel the differences. We assume $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1}$ are independent and $p(\boldsymbol{\mu}|\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\boldsymbol{\delta})$ so that $p(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\boldsymbol{\delta})p(\boldsymbol{\delta})p(\boldsymbol{\Sigma}^{-1})$. The assumption that $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1}$ are independent is reasonable because in many cases, prior knowledge suggests that mean shifts do not depend on the inverse covariance matrix. The dependence structure captured by the covariance matrix is due to *common cause* variation; on the other hand, mean shifts are due to *special cause* variation. For the sake of mathematical tractability, we further assume that $\boldsymbol{\mu}|\boldsymbol{\delta}$ has a multivariate normal distribution, that is,

$$p(\boldsymbol{\mu}|\boldsymbol{\delta}) \propto |\boldsymbol{\psi}_\delta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\theta}_\delta)^T\boldsymbol{\psi}_\delta^{-1}(\boldsymbol{\mu} - \boldsymbol{\theta}_\delta)\right\}, \tag{6}$$

where $\boldsymbol{\theta}_\delta$ is the mean and $\boldsymbol{\psi}_\delta$ is the covariance matrix. The subscript $\boldsymbol{\delta}$ of $\boldsymbol{\theta}_\delta$ and $\boldsymbol{\psi}_\delta$ indicates that the mean and the covariance matrix depend on $\boldsymbol{\delta}$.

Because we assume that $p(\boldsymbol{\mu}|\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ is given by (6), the joint distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$, given $\boldsymbol{\delta} = \mathbf{0}$, does not match the phase I posterior for $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1})$ (see the remark at the end of Section 4.2). However, we adopt (6) because it provides a prior that is easy to interpret and tune to capture prior information. Moreover, the prior distribution $p(\boldsymbol{\mu}|\boldsymbol{\delta} = \mathbf{0})$, with some reasonable values for $\boldsymbol{\theta}_\mathbf{0}$ and $\boldsymbol{\psi}_\mathbf{0}$ (which shall be given in Section 4.2), works very well.

Detailed specification of the parameters of $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and the prior for $\boldsymbol{\delta}$, $p(\boldsymbol{\delta})$ shall be discussed in the next section. Here, we shall discuss the specification of the prior $p(\boldsymbol{\Sigma}^{-1})$ for $\boldsymbol{\Sigma}^{-1}$ since this is a simpler problem. It is obvious that we should set $p(\boldsymbol{\Sigma}^{-1})$ equal to the distribution given by (4) if the covariance matrix is assumed to remain unchanged. Thus,

$$p(\boldsymbol{\Sigma}^{-1}) \propto \left|\boldsymbol{\Sigma}^{-1}\right|^t \exp\left[-\frac{1}{2}\text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right], \tag{7}$$

where $t = (N - p - 2)/2$ as indicated in (4). Note that $N \ge p + 1$ is sufficient for the Wishart distribution (7) to be nondegenerate. This condition is almost always met in practice.

Our proposed Bayesian hierarchical model consists of $l(\bar{\mathbf{x}}_f, \mathbf{S}_f|\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$, $p(\boldsymbol{\mu}|\boldsymbol{\delta})$, $p(\boldsymbol{\Sigma}^{-1})$, and $p(\boldsymbol{\delta})$. The quantities of primary interest are the posterior probabilities of $\boldsymbol{\delta}$, that is, $p(\boldsymbol{\delta}|\bar{\mathbf{x}}_f, \mathbf{S}_f)$. The posterior probability of $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^T$ is the probability that the state of the ith mean is given by δ_i, $i = 1, \ldots, p$ in light of prior knowledge, phase I data, and the phase II data $(\bar{\mathbf{x}}_f, \mathbf{S}_f)$. A comparison of the posterior probabilities of all possible values of $\boldsymbol{\delta}$ would yield information about which combination of out-of-control means is more likely. In addition, the marginal posterior distribution of δ_i can also be useful for deciding whether δ_i shifted upward, downward, or remained in control. However, a direct computation of the posterior distribution of $\boldsymbol{\delta}$ is clearly infeasible for practical problems. Section 5 provides a Gibbs sampling procedure that solves this computational problem. Before describing that procedure, we discuss the specification of prior distributions for $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ in the next section.

4. SPECIFICATION OF PRIOR DISTRIBUTIONS

Our suggestions on the specification of the prior distributions $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and $p(\boldsymbol{\delta})$ are given in this section. Prior distributions for these two parameters play a crucial role in mean shift detection.

4.1 Prior Distribution for Indicator Variables

A straightforward choice for $p(\boldsymbol{\delta})$ is the independence prior

$$p(\boldsymbol{\delta}) = \prod_{i=1}^{p} p_{1i}^{I(\delta_i=-1)} p_{2i}^{I(\delta_i=0)} p_{3i}^{I(\delta_i=1)}, \tag{8}$$

where p_{1i}, p_{2i}, and $p_{3i} = 1 - p_{1i} - p_{2i}$ are the prior probabilities that the ith mean shifted downward, remained in control, and shifted upward, respectively. Adopting the independence prior for $p(\boldsymbol{\delta})$ is equivalent to assuming that the δ_i's are a priori independent, which is justified in cases where there is no prior knowledge about the causal relationships between the variables. We propose that the *default choice* of $\mathbf{p}_i = (p_{1i}, p_{2i}, p_{3i})$ be taken as

$$\mathbf{p}_i = (0.25, 0.5, 0.25)\,. \tag{9}$$

This choice is justified if the engineer is ignorant about whether the ith mean changed or not, and if it changed, whether it increased or decreased. If we specify $\mathbf{p}_i$ as given in (9) for all $i = 1, \ldots, p$, the expected number of shifted means is $p/2$; this quantity provides one way to check whether it is reasonable to specify $\mathbf{p}_i$ according to (9). Other choices of $\mathbf{p}_i$ might be considered if there is some information about shifts in the ith mean. For instance, if the ith variable is a smaller-the-better quality characteristic, a downward shift in the ith mean may be unlikely and p_{1i} should be small.

4.2 Prior Distribution for Mean

We now discuss the specification of $p(\boldsymbol{\mu}|\boldsymbol{\delta})$, which we have assumed to be a normal distribution with mean $\boldsymbol{\theta}_\delta$ and covariance matrix $\boldsymbol{\psi}_\delta$. We set

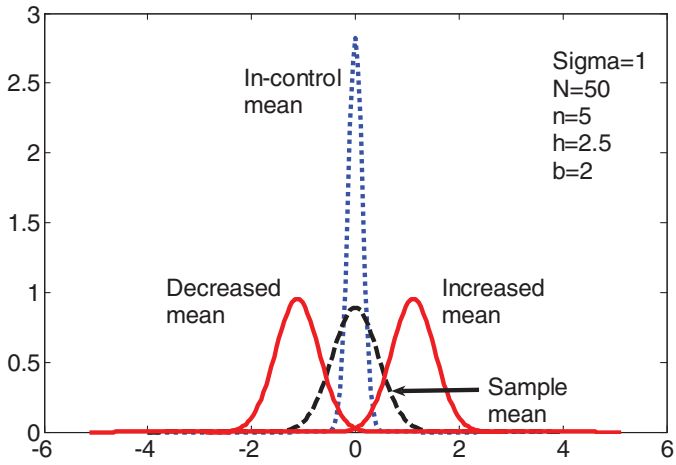

Figure 1. Density functions of $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|(\delta_i = -1)$ (standardized prior for decreased mean), $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|(\delta_i = 0)$ (standardized prior for in-control mean), $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|(\delta_i = 1)$ (standardized prior for increased mean), and $(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i|(\mu_i = \bar{x}_i, \sigma_i = \hat{\sigma}_i)$ (distribution of standardized phase II sample mean). The online version of this figure is in color.

$$\begin{aligned}\boldsymbol{\theta}_\delta = (&\bar{x}_1 - I(\delta_1 = -1)c_{1d} + I(\delta_1 = 1)c_{1u}, \ldots, \bar{x}_p \\ &- I(\delta_p = -1)c_{pd} + I(\delta_p = 1)c_{pu})^T\end{aligned} \tag{10}$$

$$\begin{aligned}\boldsymbol{\psi}_\delta = \operatorname{diag}\Bigg\{&\frac{[a_{1d}^{2I(\delta_1=-1)} a_{1u}^{2I(\delta_1=1)}]\widehat{\sigma}_1^2}{N}, \ldots, \\ &\frac{[a_{pd}^{2I(\delta_p=-1)} a_{pu}^{2I(\delta_p=1)}]\widehat{\sigma}_p^2}{N}\Bigg\},\end{aligned} \tag{11}$$

where $\hat{\sigma}_i$ is the sample standard deviation of the phase I data for the ith variable, that is, the square root of the ith diagonal element of $\widehat{\boldsymbol{\Sigma}}$. It follows from (10) and (11) that $\mu_i|\boldsymbol{\delta} = \mu_i|\delta_i$ and $\mu_1|\delta_1, \ldots, \mu_p|\delta_p$ are independently distributed. We also obtain

$$\begin{aligned}\mu_i|(\delta_i = 0) &\sim N\left(\bar{x}_i, \frac{\hat{\sigma}_i^2}{N}\right), \\ \mu_i|(\delta_i = -1) &\sim N\left(\bar{x}_i - c_{id}, \frac{a_{id}^2\hat{\sigma}_i^2}{N}\right), \\ \mu_i|(\delta_i = +1) &\sim N\left(\bar{x}_i + c_{iu}, \frac{a_{iu}^2\hat{\sigma}_i^2}{N}\right).\end{aligned} \tag{12}$$

Figure 1 illustrates a standardized version of these priors, which will be discussed later in Section 4.2.2. Note that $\mu_i|(\delta_i = 0) \sim N(\bar{x}_i, \hat{\sigma}_i^2/N)$ is entirely determined from phase I data and it is a good approximation of the posterior distribution of μ_{0i} derived from (3) when N is large. On the other hand, the distributions of $\mu_i|(\delta_i = -1)$ and $\mu_i|(\delta_i = +1)$ are each controlled by two parameters. The choices of these parameters are crucial; thus, this subject shall be discussed in the remainder of the section.

Note that $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and $p(\boldsymbol{\delta})$ should be required to jointly satisfy certain restrictions. Suppose that $[\bar{x}_i + \varepsilon_i, \bar{x}_i + \zeta_i]$ and $[\bar{x}_i - \zeta_i, \bar{x}_i - \varepsilon_i]$, where ε_i is small and ζ_i is large, contain all possible upward and downward shifts, respectively. Then, it is logical to require that $p(\delta_i = 1|\mu_i)/p(\delta_i = -1|\mu_i)$ be larger

than one in the interval $[\bar{x}_i + \varepsilon_i, \bar{x}_i + \zeta_i]$ and smaller than one in $[\bar{x}_i - \zeta_i, \bar{x}_i - \varepsilon_i]$. This is equivalent to requiring that

$$o(\mu_i) = \log\left[\frac{p(\mu_i|\,\delta_i = 1)p(\delta_i = 1)}{p(\mu_i|\,\delta_i = -1)p(\delta_i = -1)}\right] \tag{13}$$

satisfies

$$\begin{aligned} &o(\mu_i) > 0\ \forall \mu_i \in [\bar{x}_i + \varepsilon_i, \bar{x}_i + \zeta_i] \text{ and} \\ &o(\mu_i) < 0\ \forall \mu_i \in [\bar{x}_i - \zeta_i, \bar{x}_i - \varepsilon_i]. \end{aligned} \tag{14}$$

Equation (14) will be satisfied whenever $p(\delta_i = 1) = p(\delta_i = -1)$ ($p_{1i} = p_{3i}$ when $p(\boldsymbol{\delta})$ is given by (8)), $c_{id} = c_{iu}$, and $a_{id} = a_{iu}$. It will also be satisfied when $a_{id} = a_{iu}$ and $p(\delta_i = 1)/p(\delta_i = -1) = \exp[(c_{iu}^2 - c_{id}^2)/(2a_i^2\hat{\sigma}_i^2/N)]$. Otherwise, (14) should be checked.

In Sections 4.2.1 and 4.2.2, we shall discuss the specification of $p(\boldsymbol{\mu}|\,\boldsymbol{\delta})$ when prior information about mean shifts is available and when little prior information is available.

Remark. Due to the assumption that $p(\boldsymbol{\mu}|\,\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\,\boldsymbol{\delta})$, $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}|\boldsymbol{\delta} = \mathbf{0})$ is not the same as $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1}|\bar{\mathbf{x}}, \mathbf{S})$, where $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1}|\bar{\mathbf{x}}, \mathbf{S})$ is the posterior distribution $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1})$, given phase I data. To satisfy the requirement that $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}|\boldsymbol{\delta} = \mathbf{0})$ equals $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1}\big|\,\bar{\mathbf{x}}, \mathbf{S})$, we can set

$$\begin{aligned} &p(\boldsymbol{\mu}|\,\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) \\ &\propto \left|\mathbf{D}_{\boldsymbol{\delta}}\boldsymbol{\Sigma}^{-1}\mathbf{D}_{\boldsymbol{\delta}}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}N(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})^T\mathbf{D}_{\boldsymbol{\delta}}\boldsymbol{\Sigma}^{-1}\mathbf{D}_{\boldsymbol{\delta}}(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})\right\}. \end{aligned} \tag{15}$$

In this case, we can have $p(\boldsymbol{\mu}|\boldsymbol{\delta} = \mathbf{0}, \boldsymbol{\Sigma}^{-1}) = N(\bar{\mathbf{x}}, \boldsymbol{\Sigma}/N)$ by setting $\boldsymbol{\theta}_{\mathbf{0}} = \bar{\mathbf{x}}$ and $\mathbf{D}_{\mathbf{0}} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Together with $p(\boldsymbol{\Sigma}^{-1})$ given in (7), $p(\boldsymbol{\mu}|\,\boldsymbol{\delta} = \mathbf{0}, \boldsymbol{\Sigma}^{-1})p(\boldsymbol{\Sigma}^{-1})$ would now be the same as $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1}|\bar{\mathbf{x}}, \mathbf{S})$. However, this prior has the disadvantage that the represented prior knowledge is hard to understand because the distribution has a complicated form. For interpretability, we may want $\mathbf{D}_{\boldsymbol{\delta}}$ to be diagonal so that the variances of $\mu_i|\,(\delta_i = -1)$ and $\mu_i|\,(\delta_i = 1)$ depend only on the ith diagonal element of $\boldsymbol{\Sigma}$ and not on the other elements. However, the existence of correlations between the μ_i's given $(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta})$ for $\boldsymbol{\delta} \neq \mathbf{0}$ is hard to interpret. For example, why would μ_1 and μ_2 be correlated when μ_1 shifts upward and μ_2 shifts downward, and why would they have the same correlation as x_1 and x_2? Certainly, we can let the entire covariance matrix of $p(\boldsymbol{\mu}|\,\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1})$ depend on $\boldsymbol{\delta}$ and set all the correlations involving μ_i to zero whenever $\delta_i \neq 0$. But it would then be difficult to sample $\boldsymbol{\Sigma}^{-1}$ from its full conditional distribution as the distribution would not be a Wishart distribution. Compared with (15), a prior for the mean given by (12) is arguably easier to interpret. For this reason, we prefer (12) over (15) despite the fact that (15) gives the correct prior for the case where $\boldsymbol{\delta} = \mathbf{0}$.

4.2.1 Case 1: Prior Information About Mean Shifts Is Available. In cases where the quality engineer has a good idea about mean shifts, the specification of the prior parameters in (10) and (11) is a rather straightforward exercise. The prior information that needs to be elicited for upward and downward shifts in each mean is the *most likely value and range*.

Clearly, c_{id} and c_{iu} should be set equal to most likely magnitudes of downward and upward shifts of the ith mean, respectively. Now, if the range of upward shifts in the ith mean is $[\varepsilon_{iu}, \zeta_{iu}]$, then we should set a_{iu} so that $c_{iu} + 2a_{iu}\hat{\sigma}_i/\sqrt{N} \geq \zeta_{iu}$ and $c_{iu} - 2a_{iu}\hat{\sigma}_i/\sqrt{N} \leq \varepsilon_{iu}$. If the range of magnitudes of downward shifts is $[\varepsilon_{id}, \zeta_{id}]$, then we should set a_{id} so that $-c_{id} - 2a_{id}\hat{\sigma}_i/\sqrt{N} \leq -\zeta_{id}$ and $-c_{id} + 2a_{id}\hat{\sigma}_i/\sqrt{N} \geq -\varepsilon_{id}$. Note that ± 2 constants are used because a normal random variable has a high (roughly 95%) probability of being within two standard deviations of its mean. These considerations give

$$\begin{aligned} a_{iu} &= \frac{\sqrt{N}}{2\hat{\sigma}_i}\max\left\{\zeta_{iu} - c_{iu}, c_{iu} - \varepsilon_{iu}\right\}, \\ a_{id} &= \frac{\sqrt{N}}{2\hat{\sigma}_i}\max\left\{\zeta_{id} - c_{id}, c_{id} - \varepsilon_{id}\right\}. \end{aligned} \tag{16}$$

Three forms of prior knowledge can be captured by the proposed priors $p(\boldsymbol{\mu}|\,\boldsymbol{\delta})$ and $p(\boldsymbol{\delta})$. The first is the most likely magnitudes of a shift, which is often large. Assignable causes often produce large shifts; for instance, in thermocompression processes, a marked decrease in the strength of gold–gold bonds is observed when surface contamination is present (Jellison 1975). Moreover, large shifts are more critical and more easily detected than small shifts. Second, it is often the case that information about mean shifts is asymmetrical, that is, an increase in a mean is likely to be within a certain range and a decrease is likely to be in another range, and the most likely upward and downward shifts are different. For instance, if x_i is a smaller-the-better quantity (Wu and Hamada 2009, p. 268), we should set $c_{iu} > c_{id}$, whereas for the larger-the-better quantity, we should set $c_{iu} < c_{id}$. This is because process improvements are often smaller than process deteriorations. Third, based on knowledge of the process, the engineer may have reason to believe that a particular mean is more likely to have increased than decreased and vice versa. These three forms of prior knowledge cannot be incorporated by modeling each mean as a mixture of two normal distributions (one for the in-control mean and one for the out-of-control mean).

4.2.2 Case 2: Little or No Prior Information About Mean Shifts Is Available. In cases where prior information about mean shifts is hard to obtain, we consider using a symmetric prior for the mean with *only two parameters*, that is, $c_{id} = c_{iu} = c_i = h\hat{\sigma}_i/\sqrt{n}$ and $a_{id} = a_{iu} = a$. This implies that $\mu_i|\delta_i \sim N(\bar{x}_i + \delta_i h\hat{\sigma}_i/\sqrt{n}, a^{2|\delta_i|}\hat{\sigma}_i^2/N)$, which gives

$$(\mu_i - \bar{x}_i)/\hat{\sigma}_i|\,\delta_i \sim N\left(\delta_i h/\sqrt{n}, a^{2|\delta_i|}/N\right). \tag{17}$$

Figure 1 illustrates this prior. It can be observed that the large reduction in the number of parameters is achieved through *standardization* of all variables (variable i is standardized by $\bar{x}_i$ and $\hat{\sigma}_i$), and assuming that the magnitudes of upward and downward shifts of the standardized variables can be approximately modeled by the same normal distribution $N(h/\sqrt{n}, a^2/N)$. Our recommendations for specifying the prior distributions for the mean and indicators are summarized in Figure 2. Detailed discussions about these choices are given below.

In a nutshell, our suggestion concerning the choices of the parameters h and a is based on the rationale that $\mu_i|(\delta_i = 1)$ should be centered at the upper tail of $\bar{x}_{fi}|(\mu_i = \bar{x}_i, \sigma_i = \hat{\sigma}_i)$ (this determines h), and the densities of $\mu_i|(\delta_i = 0)$ and $\mu_i|(\delta_i = 1)$ should have at least a slight overlap (this determines a). Figure 1

Simplified Priors for the Mean and Indicators

1. ith element of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$: $\bar{x}_i + \delta_i c_i = \bar{x}_i + \delta_i h\hat{\sigma}_i/\sqrt{n}$.
2. $\boldsymbol{\psi}_{\boldsymbol{\delta}}$: Diagonal matrix with ith diagonal element $\left(a^{2|\delta_i|}\right)\hat{\sigma}_i^2/N$.
3. Prior for indicators: $p(\boldsymbol{\delta}) = \prod_{i=1}^{p} p_{1i}^{I(\delta_i=-1)} p_{2i}^{I(\delta_i=0)}(1-p_{1i}-p_{2i})^{I(\delta_i=1)}$.
   Default choice for p_{1i} and p_{2i}: $p_{1i} = 0.25$, $p_{2i} = 0.5$.

Choice of Hyperparameters h and a:

Suggestion 1: Compare results obtained from several reasonable choices.

Set $a = max\left\{\frac{h}{b}\sqrt{\frac{N}{n}} - 1,1\right\}$. Try several values of $h \in [2,8]$ and $b \in [0.5,2]$. Make sure that $\left[-h/\sqrt{n} - 2a/\sqrt{N}, h/\sqrt{n} + 2a/\sqrt{N}\right]$ contains all $\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i$.

Suggestion 2: Estimate prior parameters via the empirical Bayes method.

Find $I = \left\{i: \left|\bar{x}_{fi} - \bar{x}_i\right|/\hat{\sigma}_i > 2\right\}$. Set $h = h^{EB} = \begin{cases} \sqrt{n}\sum_{i\in I}\frac{|\bar{x}_{fi}-\bar{x}_i|}{\hat{\sigma}_i}, & if\ I\ is\ nonempty \\ 2, & if\ I\ is\ empty \end{cases}$ and

$$a = a^{EB} = \begin{cases} max\left\{\sqrt{N}stdev\left\{\frac{|\bar{x}_{fi}-\bar{x}_i|}{\hat{\sigma}_i}, i \in I\right\}, \frac{h^{EB}}{2}\sqrt{\frac{N}{n}} - 1,1\right\}, & if\ |I| \geq 2 \\ max\left\{\frac{h^{EB}}{2}\sqrt{\frac{N}{n}} - 1,1\right\}, & if\ |I| \leq 1 \end{cases}.$$

Alternatively, set $h = h^{EB}$, $a = max\left\{\frac{h^{EB}}{b}\sqrt{\frac{N}{n}} - 1,1\right\}$, and try several values of b. We call this the partially empirical Bayes method.

Figure 2. Summary of recommended prior parameter choices when little or no prior information about mean shifts is available.

illustrates this idea. Note that if we plot $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|\delta_i$, the same plot would be obtained for all i. Thus, we need to use *only one* plot to check whether the prior specification is reasonable.

In the design of control charts, n is chosen so that $\hat{\sigma}_i/\sqrt{n}$ is small compared with C_i to get good power for detecting a shift of size C_i (Montgomery 2009, p. 247). Similarly, for the proposed approach, the relative sizes of $\hat{\sigma}_i/\sqrt{n}$ and c_i determine the Type I and Type II error rates, where the Type I error rate is defined as the average number of in-control means declared out-of-control and the Type II error rate is defined as the average number of out-of-control means whose shift directions are misidentified. Thus, it is convenient to measure c_i in units of $\hat{\sigma}_i/\sqrt{n}$, that is, $c_i = h\hat{\sigma}_i/\sqrt{n}$. We recommend that h be at least 2 to achieve reasonable Type I error. The reason for this choice is as follows. When the ith mean is in control, $\bar{x}_{fi}|(\mu_i = \mu_{0i}, \sigma_i) \sim N(\mu_{0i}, \sigma_i^2/n)$ and so, $\bar{x}_{fi}$ falls within the interval $(\bar{x}_i - 2\hat{\sigma}_i/\sqrt{n}, \bar{x}_i + 2\hat{\sigma}_i/\sqrt{n})$ most of the time (see Figure 1). Thus, if we specify h to be less than 2, the Type I error rate would be high. On the other hand, if h is too large $(h\hat{\sigma}_i/\sqrt{n} > 3a\hat{\sigma}_i/\sqrt{N})$, the Type II error rate would be high for small shifts. This is because for fixed a, a large h implies that large shifts are likely and small shifts are unlikely. The problem can be mitigated somewhat by choosing a large a. However, prior information is diluted because the priors for $\mu_i|(\delta_i = 1)$ and $\mu_i|(\delta_i = -1)$ would overlap somewhat and also allow very large shifts.

We set

$$a = \max\left\{\frac{c_i}{b\hat{\sigma}_i/\sqrt{N}} - 1, 1\right\} = \max\left\{\frac{h}{b}\sqrt{\frac{N}{n}} - 1, 1\right\}, \quad (18)$$

where $b > 0$. This arises from the following consideration. For reasons explained in the next paragraph, we would like to choose the constant a so that there is a reasonable overlap between the priors for the increased mean, in-control mean, and decreased mean. Thus, we set $\bar{x}_i + c_i - ba\hat{\sigma}_i/\sqrt{N} = \bar{x}_i + b\hat{\sigma}_i/\sqrt{N}$. This means that b standard deviation units above the center of $\mu_i|(\delta_i = 0)$ should be b standard deviation units below the center of $\mu_i|(\delta_i = 1)$. By symmetry, setting $\bar{x}_i - c_i + ba\hat{\sigma}_i/\sqrt{N} = \bar{x}_i - b\hat{\sigma}_i/\sqrt{N}$ yields the same result. Since there is typically less information about shifted means than in-control means, we should set a to be 1 or larger. These considerations lead to (18).

If $b \leq 2$, there is at least a slight overlap between the density functions of $\mu_i|(\delta_i = 0)$ and $\mu_i|(\delta_i = 1)$, and between the density functions of $\mu_i|(\delta_i = 0)$ and $\mu_i|(\delta_i = -1)$. This ensures that the Gibbs sampler does not get stuck in one of the conditional distributions $\mu_i|\delta_i$. Moreover, too large values for b imply strong prior knowledge on the mean shifts. This can give rise to large Type I error rates because $\delta_i = 1$ and $\delta_j = 0$ for all $j \neq i$ cannot explain the data well if $\bar{x}_{fi} \gg \bar{x}_i + c_i + 2a\hat{\sigma}_i/\sqrt{N}$ and $\bar{x}_{fj} = \bar{x}_j$ for all $j \neq i$. The data may be better explained by $(\delta_i, \delta_l) = (1, 1)$ and $\delta_j = 0$ for all $j \notin \{i, l\}$. The smaller the value of b, the larger is the variances of $\mu_i|(\delta_i = -1)$ and $\mu_i|(\delta_i = 1)$, and so, the prior contains less information about the mean shifts. This reduces shift detection power since for $\mu_i > \bar{x}_i$, $p(\delta_i = 1|\mu_i)/p(\delta_i = 0|\mu_i)$ is reduced. Thus, we recommend that b be at least 0.5.

Note that a small b (which gives a large a) produces considerable overlap between the densities of $\mu_i|(\delta_i = 1)$ and $\mu_i|(\delta_i = -1)$, which may seem unreasonable because if a mean shifted downward, it cannot be much larger than $\bar{x}_i$. However, because $o(\mu_i)$ is a straight line with positive slope through $\bar{x}_i$ regardless of the value of b, the prior distribution makes sense regardless of the value of b. Nonetheless, the slope of $o(\mu_i)$ is $2Nc_i/a^2\hat{\sigma}_i^2$, which decreases with an increase in a. As a consequence, we may observe that if the posterior probability that

$\delta_i = 1$ is largest, the posterior probability that $\delta_i = -1$ can be a close second and vice versa. However, this does not affect decisions about the mean shifts as the decisions are based on the value of δ_i that gives the largest posterior probability (see Section 5).

For the analysis of a real dataset, we suggest that several values of h and b be tried and the resulting decisions be compared. The values of h and b should be chosen so that the interval $[-h/\sqrt{n} - 2a/\sqrt{N}, h/\sqrt{n} + 2a/\sqrt{N}]$ contains all $(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i$ (e.g., see Figure 8). Otherwise, if $(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i > h/\sqrt{n} + 2a/\sqrt{N}$, $\delta_i = 1$ alone cannot explain this large deviation well and if $(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i < h/\sqrt{n} - 2a/\sqrt{N}$, $\delta_i = -1$ alone cannot explain this large deviation well. We have found that decisions based on decision rule 1 or decision rule 2 (see Section 5) are often *robust* to changes in values of b and h. However, trying several values give confidence to the decisions obtained from the proposed approach.

We can also choose h and b using an *empirical Bayesian (EB) approach*. Because $(\mu_i - \bar{x}_i)/\hat{\sigma}_i | (\delta_i = -1) \sim N(-h/\sqrt{n}, a^2/N)$, and $(\mu_i - \bar{x}_i)/\hat{\sigma}_i | (\delta_i = 1) \sim N(h/\sqrt{n}, a^2/N)$, the quantity $(|\mu_i - \bar{x}_i|/\hat{\sigma}_i) | (\delta_i \neq 0)$ would be approximately $N(h/\sqrt{n}, a^2/N)$ if $h/\sqrt{n}$ is large compared with $a/\sqrt{N}$. Since $\mu_1 | \delta_1, \ldots, \mu_p | \delta_p$ are independently distributed, $(|\mu_1 - \bar{x}_1|/\hat{\sigma}_1) | (\delta_1 \neq 0), \ldots, (|\mu_p - \bar{x}_p|/\hat{\sigma}_p) | (\delta_p \neq 0)$ are independent and identically distributed with approximate common distribution $N(h/\sqrt{n}, a^2/N)$. Based on this observation, we can estimate h and a as follows. Let $I = \{i : |\bar{x}_{fi} - \bar{x}_i|/\hat{\sigma}_i > 2\}$, that is, I is the set of indices associated with large discrepancy between phase I and phase II means. Then, we can estimate h by

$$h^{EB} = \begin{cases} \sqrt{n} \sum_{i \in I} \dfrac{|\bar{x}_{fi} - \bar{x}_i|}{\hat{\sigma}_i}, & \text{if } I \text{ is nonempty,} \\ 2, & \text{if } I \text{ is empty.} \end{cases} \tag{19}$$

We propose two ways to choose a. The first is to simply use (18) and try various values of b; the second is to estimate a by

$$a^{EB} = \begin{cases} \max\left\{ \sqrt{N}\,\text{stdev}\left\{ \dfrac{|\bar{x}_{fi} - \bar{x}_i|}{\hat{\sigma}_i}, i \in I \right\}, \right. \\ \qquad \left. \dfrac{h^{EB}}{2}\sqrt{\dfrac{N}{n}} - 1, 1 \right\}, & \text{if } |I| \geq 2, \\ \max\left\{ \dfrac{h^{EB}}{2}\sqrt{\dfrac{N}{n}} - 1, 1 \right\}, & \text{if } |I| \leq 1, \end{cases} \tag{20}$$

where $\text{stdev}\{|\bar{x}_{fi} - \bar{x}_i|/\hat{\sigma}_i, i \in I\}$ denotes the standard deviation of the set of values $\{|\bar{x}_{fi} - \bar{x}_i|/\hat{\sigma}_i, i \in I\}$ and $|I|$ denotes the number of elements in I. We call the approach of setting $h = h^{EB}$ and trying several values of b the partially empirical Bayes' (PEB) method and we call the approach of setting $h = h^{EB}$ and $a = a^{EB}$ the EB method.

The reason we do not set a^{EB} equal to $\sqrt{N}\,\text{stdev}\{|\bar{x}_{fi} - \bar{x}_i|/\hat{\sigma}_i, i \in I\}$ is that we want to ensure that a^{EB} is not smaller than (18) with $b = 2$. Too small a value for a^{EB} is undesirable for reasons discussed previously. In addition, we want to ensure that the prior does not rule out shifts in variables whose indices are not in I.

Before we end this section, we point out that the EB method determines the prior for the mean entirely from data. Hence, it is very convenient in practice.

5. GIBBS SAMPLING AND DECISION RULES FOR MEAN SHIFTS

This section gives a Gibbs sampling procedure for sampling from the posterior distribution of $\boldsymbol{\delta}$ and decision rules for identifying mean shifts.

Since the sample space of $\boldsymbol{\delta}$ consists of 3^p points, a direct calculation of all 3^p probabilities is infeasible even when p is moderately large. To give an idea of the growth of 3^p with p, note that $3^{10} = 59{,}049$ and $3^{15} = 14{,}348{,}907$. However, not all values of $\boldsymbol{\delta}$ are equally of interest. We are primarily interested in those values of $\boldsymbol{\delta}$ with large probabilities, especially the posterior mode. This is adequate information for making inference about the mean shifts and the directions of the shifts. Gibbs sampling is a tool that allows us to discover the most probable values of $\boldsymbol{\delta}$ (Gelfand and Smith 1990; George and McCulloch 1993, 1997).

From the joint posterior distribution,

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta} | \bar{\mathbf{x}}_f, \mathbf{S}_f) \propto l(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1} | \bar{\mathbf{x}}_f, \mathbf{S}_f) p(\boldsymbol{\mu} | \boldsymbol{\delta}) p(\boldsymbol{\Sigma}^{-1}) p(\boldsymbol{\delta}), \tag{21}$$

we can easily obtain the full conditional distributions

$$p\left(\boldsymbol{\Sigma}^{-1} \middle| \bar{\mathbf{x}}_f, \mathbf{S}_f, \boldsymbol{\delta}, \boldsymbol{\mu}\right) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{n}{2}+t} \times \exp\left\{ -\frac{1}{2} \text{trace}\left[\boldsymbol{\Sigma}^{-1} (\mathbf{S}_f + \mathbf{S} + n(\bar{\mathbf{x}}_f - \boldsymbol{\mu})(\bar{\mathbf{x}}_f - \boldsymbol{\mu})^T) \right] \right\}, \tag{22}$$

$$p\left(\boldsymbol{\mu} | \bar{\mathbf{x}}_f, \mathbf{S}_f, \boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}\right) \propto |\mathbf{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}|^{\frac{1}{2}} \times \exp\left\{ -\frac{1}{2} (\boldsymbol{\mu} - \mathbf{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}})^T \mathbf{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\mu} - \mathbf{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}) \right\}, \tag{23}$$

$$p\left(\boldsymbol{\delta} | \bar{\mathbf{x}}_f, \mathbf{S}_f, \boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}\right) \propto |\boldsymbol{\psi}_{\boldsymbol{\delta}}|^{-\frac{1}{2}} \times \exp\left\{ -\frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})^T \boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1} (\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}}) \right\} p(\boldsymbol{\delta}), \tag{24}$$

where $\mathbf{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}} = (\boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1} + n\boldsymbol{\Sigma}^{-1})^{-1} (\boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1} \boldsymbol{\theta}_{\boldsymbol{\delta}} + n\boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}}_f)$ and $\mathbf{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}} = (\boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1} + n\boldsymbol{\Sigma}^{-1})^{-1}$. The conditional distributions (22)–(24) enable us to use a Gibbs sampling algorithm (Figure 3) for sampling from the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta})$. The algorithm given in Figure 3 holds for general $p(\boldsymbol{\mu} | \boldsymbol{\delta})$ given by (6). For the simplified prior given by (12), Step 4 of the Gibbs sampling algorithm given in Figure 3 simplifies to Step 4′ given in Figure 4. The Gibbs sampling algorithm creates a Markov chain

$$(\boldsymbol{\Sigma}^{-1})^1, \boldsymbol{\mu}^1, \delta_1^1, \ldots, \delta_p^1, (\boldsymbol{\Sigma}^{-1})^2, \boldsymbol{\mu}^2, \delta_1^2, \ldots, \delta_p^2, \ldots, (\boldsymbol{\Sigma}^{-1})^i, \boldsymbol{\mu}^i, \delta_1^i, \ldots, \delta_p^i, \ldots, \tag{25}$$

which has the property that the values of $\boldsymbol{\delta}^i = (\delta_1^i, \ldots, \delta_p^i)^T$ that appear in the simulation would be a sample from $p(\boldsymbol{\delta} | \bar{\mathbf{x}}_f, \mathbf{S}_f)$ for i large enough. Here, the reason for working directly with $\boldsymbol{\Sigma}^{-1}$ instead of $\boldsymbol{\Sigma}$ becomes evident. Direct sampling of $\boldsymbol{\Sigma}^{-1}$ avoids the need to invert $\boldsymbol{\Sigma}$ in the computation of $\mathbf{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}$ and $\mathbf{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}$. In addition, $\boldsymbol{\Sigma}$ has an inverse Wishart distribution and we are not aware of a direct method for generating random matrices from this distribution; the most common method is

1. Start with $\boldsymbol{\mu}^0 = \bar{\mathbf{x}}$ and $\boldsymbol{\delta}^0 = (\delta_1^0, \ldots, \delta_p^0)^T$. Set $i = 1$.

2. Sample $(\boldsymbol{\Sigma}^{-1})^i$ from the Wishart distribution with scale matrix $\left(\mathbf{S}_f + \mathbf{S} + n(\bar{\mathbf{x}}_f - \boldsymbol{\mu}^{i-1})(\bar{\mathbf{x}}_f - \boldsymbol{\mu}^{i-1})^T\right)^{-1}$ and $v = n + 2t + p + 1 = N + n - 1$ degrees of freedom.

3. Sample $\boldsymbol{\mu}^i$ from a normal distribution with mean $\mathbf{g}_{\boldsymbol{\delta}^{i-1},\boldsymbol{\Sigma}^i} = \left(\boldsymbol{\psi}_{\boldsymbol{\delta}^{i-1}}^{-1} + n(\boldsymbol{\Sigma}^{-1})^i\right)^{-1}\left(\boldsymbol{\psi}_{\boldsymbol{\delta}^{i-1}}^{-1}\boldsymbol{\theta}_{\boldsymbol{\delta}^{i-1}} + n(\boldsymbol{\Sigma}^{-1})^i\bar{\mathbf{x}}_f\right)$, and covariance matrix $\mathbf{V}_{\boldsymbol{\delta}^{i-1},\boldsymbol{\Sigma}^i} = \left(\boldsymbol{\psi}_{\boldsymbol{\delta}^{i-1}}^{-1} + n(\boldsymbol{\Sigma}^{-1})^i\right)^{-1}$.

4. Define $\boldsymbol{\delta}_j^i(r) = (\delta_1^i, \ldots, \delta_{j-1}^i, r, \delta_{j+1}^{i-1}, \ldots, \delta_p^{i-1})^T$. For $j = 1, \ldots, p$, sample δ_j^i from the discrete distribution:

$$p(r) = \frac{\left|\boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(r)}\right|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(r)}\right)^T \boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(r)}^{-1}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(r)}\right)\right\} p(\boldsymbol{\delta}_j^i(r))}{\sum_{q=-1}^{1}\left|\boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(q)}\right|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(q)}\right)^T \boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(q)}^{-1}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(q)}\right)\right\} p\left(\boldsymbol{\delta}_j^i(q)\right)}, r \in \{-1,0,1\}$$

Figure 3. Gibbs sampling algorithm.

to take the inverse of a draw from the corresponding Wishart distribution.

Mean shift directions would be identified from $\boldsymbol{\delta}$ values that have a high posterior probability or δ_i values that have a high marginal posterior probability. We give two decision rules in Figure 5 to determine mean shift directions. The idea underlying the first decision rule is that of making a decision based on the mode of the marginal posterior distribution of each δ_j, while the idea underlying the second decision rule is that of making a decision based on the posterior mode of $\boldsymbol{\delta}$. We estimate these modes using the close-to-steady-state portion ($i > \tau$) of (25), where τ is the burn-in period. This gives decision rules 1 and 2.

Before we end this section, we point out that while the focus in this article is on shift directions, the shift magnitudes can be easily estimated from the components of $E(\boldsymbol{\mu}|\bar{\mathbf{x}}_f, \mathbf{S}_f) - \bar{\mathbf{x}}$. This latter quantity can be estimated from the samples generated by (25).

6. EXAMPLES

We shall give three examples to demonstrate the effectiveness of our proposed methodology. In the first example, we show that the proposed approach is superior to the t-test and is as good as WJPLM over wide ranges of recommended prior parameter values. Note that Wang and Jiang (2009) proposed the use of the forward selection algorithm for model selection. In contrast, we shall use the best subset method. The latter performs better than the former, but it incurs a high computation cost. Even with the use of fast algorithms such as those proposed by Furnival and Wilson (1974), the computation can be too costly for high-dimensional problems (e.g., $p > 30$). This is primarily due to the exponential increase in the number of possible models with the dimension of the problem. In comparison, the most computationally intensive parts of the proposed approach are Steps 2 and 3 in Figure 3 (assuming Step 4′ in Figure 4 is used), which require $O(p^3)$ operations. We conclude that the proposed approach requires $O(N_{\text{iter}}p^3)$ operations, where N_{iter} is the number of iterations of the Markov chain (25). This means that the proposed approach can be more affordable than the best subset variant of WJPLM when p is large.

In the second example, we compare our approach with the LEB diagnostic procedure proposed by Zou, Jiang, and Tsung (2011). The example demonstrates that our proposed method combined with the likelihood-ratio test for equality of covariance matrices yields a powerful method for diagnosing shifts. The example also includes an analysis of a real dataset; for this dataset, the proposed method yields the same shift decision reached by LEB for wide ranges of prior parameters. The third example analyzes an interesting set of process monitoring data for a fruit juice process, originally given by Fuchs and Kenett (1998) and reproduced by Beltran (2006). Finally, in Appendix A (available in the online supplementary materials), we give an example that shows that proper choices of h and b can yield performance superior to WJPLM with correct L, and that asymmetric choices of $p(\mu_i)$ and $p(\delta_i)$ that correctly reflect the true state of nature can give even better results. MATLAB

4'. Define $\boldsymbol{\delta}_j^i(r) = (\delta_1^i, \ldots, \delta_{j-1}^i, r, \delta_{j+1}^{i-1}, \ldots, \delta_p^{i-1})^T$. For $j = 1, \ldots, p$, sample δ_j^i from the discrete distribution:

$$p(r) = \frac{\left(\phi_j(r)\right)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left[\mu_j^i - \theta_j(r)\right]^2\left(\phi_j(r)\hat{\sigma}_j^2/N\right)^{-1}\right\} p(\boldsymbol{\delta}_j^i(r))}{\sum_{q=-1}^{1}\left(\phi_j(q)\right)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left[\mu_j^i - \theta_j(q)\right]^2\left(\phi_j(q)\hat{\sigma}_j^2/N\right)^{-1}\right\} p\left(\boldsymbol{\delta}_j^i(q)\right)}, r \in \{-1,0,1\},$$

where $\theta_j(r) = \left(\bar{x}_j - I(r=-1)c_{jd} + I(r=1)c_{ju}\right)$ and $\phi_j(r) = a_{jd}^{2I(r=-1)}a_{ju}^{2I(r=1)}$.

Figure 4. Simplification of Step 4 of Gibbs sampling algorithm given in Figure 3.

Decision Rule 1: For each j, choose the value d_j of δ_j that appears most frequently in (25) for $i > \tau$ and make decision $(d_1, \ldots, d_p)$.
Decision Rule 2: Choose the value $\mathbf{d}$ of $\boldsymbol{\delta}$ that appears most frequently in (25) for $i > \tau$ and make decision $\mathbf{d}$.

Figure 5. Decision rules for identifying mean shifts using steady-state Gibbs sampler output $\{\boldsymbol{\delta}^i : i > \tau\}$.

code for implementing the EB version of the proposed approach is given in Appendix E (available in the online supplementary materials).

6.1 Example 1: Performance Comparison I

We consider a problem where $p = 12$, $N = 90$, and $n = 6$ (12 variables, 90 phase I samples, 6 out-of-control phase II samples). The in-control mean is $\boldsymbol{\mu}_0 = \mathbf{0}$ and the population covariance matrix $\boldsymbol{\Sigma}_0$ (given in Appendix B available in the online supplementary materials) is generated randomly from the inverse Wishart distribution with scale matrix equal to the identity matrix $\mathbf{I}$, and $p + 1$ degrees of freedom. This distribution yields a random correlation matrix that has marginally uniformly distributed correlations (Barnard, McCulloch, and Meng 2000). The shifted mean is $\boldsymbol{\mu}_1 = (\Delta\sigma_1, -\Delta\sigma_2, \Delta\sigma_3, -\Delta\sigma_4, 0, \ldots, 0)$. The prior distributions $p(\boldsymbol{\mu}|\, \boldsymbol{\delta})$ and $p(\boldsymbol{\delta})$ are specified according to Figure 2. We change Δ, h, and b according to the experimental design given in Table 1 (the EB method is denoted by EB), and we replicate each run in the design 100 times. For each replicate, we simulate $\bar{\mathbf{x}}$, $\mathbf{S}$, $\bar{\mathbf{x}}_f$, and $\mathbf{S}_f$ independently from $N(\mathbf{0}, \boldsymbol{\Sigma}_0/N)$, $W(\boldsymbol{\Sigma}_0, N-1)$, $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_0/n)$, and $W(\boldsymbol{\Sigma}_0, n-1)$, respectively. Given $\bar{\mathbf{x}}$, $\mathbf{S}$, $\bar{\mathbf{x}}_f$, and $\mathbf{S}_f$, we obtain samples from $p(\boldsymbol{\delta}|\, \bar{\mathbf{x}}_f, \mathbf{S}_f)$ using the Gibbs sampling algorithm in Figure 3, with Step 4 replaced by Step 4′ given in Figure 4, and we apply the decision rules given in Figure 5. Two performance measures are calculated:

1. Type I error rate = (number of in-control means incorrectly identified as out-of-control)/(total number of in-control means).
2. Type II error rate = (number of out-of-control means incorrectly identified as in-control or whose shift direction is incorrectly identified)/(total number of out-of-control means).

Adding a constant to both $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ will not change the performance of our procedure. It has the desirable *invariance property* that shifting or rescaling all observations by the same amount does not change the posterior distribution of $\boldsymbol{\delta}$ (if the priors are specified according to Figure 2).

For all simulation runs, we terminate sampling from the Markov chain (25) after $N_{\text{iter}} = 3000$ iterations and we set the burn-in period at $\tau = 1000$. These choices are found to work well in many trial runs. Table 1 presents the results of the simulation. The error rates for decision rule 1 are given in the columns labeled D1 and the error rates for decision rule 2 are given in the columns labeled D2.

It can be observed from Table 1 that except for a few cases, decision rules 1 and 2 have a comparable performance. Table 1 also suggests that a smaller b tends to give smaller Type I error rates and a larger h ($h/\sqrt{n} > \Delta$) tends to give larger Type II error rates. Note that for $\Delta = 1.4$, $b = 0.5$, and $b = 2$ give comparable Type II error rates, but $b = 0.5$ gives smaller Type I error rates. Thus, for larger shifts, a smaller b may be better. For all shift sizes, the EB method gives good performance; its superior performance when the shift size is small, that is, $\Delta = 0.6$, is noteworthy.

We plot the Type II error rate versus the Type I error rate for the five different priors (results for decision rule 1 are used), the t-test, and WJPLM in Figure 6. Symbols A, B, C, and D denote the priors given by $(h, b) = (\sqrt{6}, 0.5), (1.8\sqrt{6}, 0.5), (\sqrt{6}, 2), (1.8\sqrt{6}, 2)$, respectively, and symbol E denotes the EB method. The continuous curve for the t-test is obtained by changing γ in (2) over small steps (the Type I and Type II error rates are both functions of γ). WJPLM with model size (specified number of mean shifts) L is plotted as $+L$ in the figure. Figure 6 clearly indicates that our approach and WJPLM are superior to the t-test. When $\Delta = 0.6$, our approach performs similarly to WJPLM with $L = 2$ and $L = 3$ (specified number of mean shifts less than four), while when $\Delta = 1$ and $\Delta = 1.4$, our approach has performance comparable with WJPLM with correctly specified number of mean shifts, that is, $L = 4$. In particular, for $\Delta = 0.6$, the EB method performs like WJPLM with $L = 3$, and for $\Delta = 1$ and $\Delta = 1.4$, the EB method performs like WJPLM with $L = 4$. Note that selecting the model of correct size with minimum residual sum of squares, which is WJPLM with $L = 4$, is probably the best performance we can expect from a frequentist method. Since, in practice, the actual

Table 1. Factor settings and results for simulation experiment

			Type II error rate		Type I error rate	
Shift Δ	h	b	D1	D2	D1	D2
0.6	$\sqrt{6}$	0.5	0.445	0.433	0.011	0.030
	$1.8\sqrt{6}$	0.5	0.525	0.528	0.009	0.016
	$\sqrt{6}$	2	0.340	0.373	0.054	0.083
	$1.8\sqrt{6}$	2	0.505	0.503	0.014	0.031
	EB		0.330	0.335	0.053	0.086
1	$\sqrt{6}$	0.5	0.080	0.068	0.014	0.011
	$1.8\sqrt{6}$	0.5	0.098	0.110	0.009	0.013
	$\sqrt{6}$	2	0.028	0.035	0.040	0.045
	$1.8\sqrt{6}$	2	0.073	0.068	0.016	0.019
	EB		0.043	0.053	0.035	0.044
1.4	$\sqrt{6}$	0.5	0.000	0.000	0.006	0.005
	$1.8\sqrt{6}$	0.5	0.008	0.008	0.005	0.010
	$\sqrt{6}$	2	0.000	0.000	0.025	0.026
	$1.8\sqrt{6}$	2	0.008	0.008	0.015	0.018
	EB		0.003	0.000	0.014	0.013

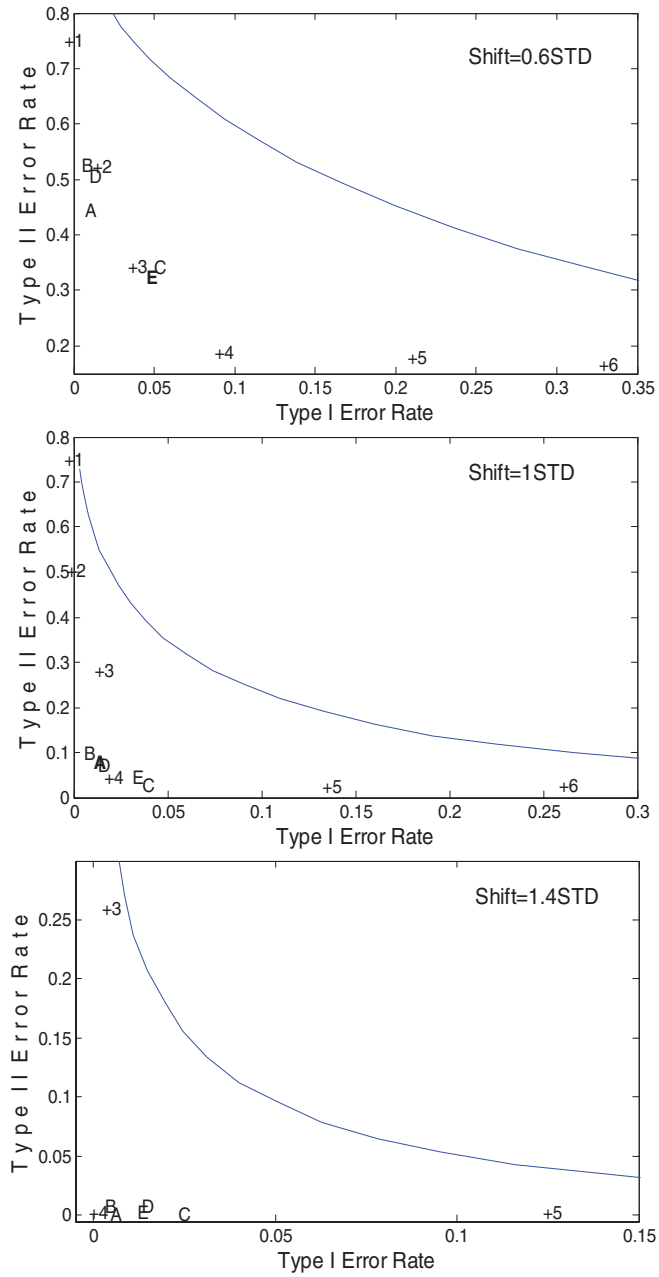

Figure 6. Type II error rate versus Type I error rate for the proposed approach, WJPLM, and t-test. Top: $\Delta = 0.6$, middle: $\Delta = 1$, bottom: $\Delta = 1.4$. The solid line is for the t-test. The symbol $+L$ is for WJPLM with model size L. Symbols A–D are for prior distributions in Table 1. Symbol E is for the EB approach. The online version of this figure is in color.

number of means that shifted cannot be known, the performance of the EB method is impressive.

It can also be concluded from an examination of Table 1 and Figure 6 that the proposed approach is *robust*. The Type I and Type II error rates for the proposed approach change only slightly (compared with changes for WJPLM with different L), even though h and b are changed over wide ranges.

We have also investigated the effect of n on performance. When $\Delta = 0.6$ and n is increased to 12, the Type II and Type I error rates of the EB method decrease to 0.098 and 0.024, respectively, for decision rule 1, and 0.075 and 0.023, respectively, for decision rule 2 (compare with Table 1). For $\Delta = 1$ and $n = 12$, the estimated Type II error rates are 0.005 (for both decision rules) and the estimated Type I error rates are 0.0175 and 0.02. Thus, huge improvements can be attained with an increase in n.

Table 2. Type I and Type II error rates for poor prior distribution for mean

			Type II error rate		Type I error rate	
Shift	h	b	D1	D2	D1	D2
1	$0.6\sqrt{6}$	2	0.000	0.003	0.173	0.191
1.4	$0.6\sqrt{6}$	2	0.000	0.000	0.375	0.394

We shall now illustrate a case of poor prior specification. The prior given by $h = 0.6\sqrt{6}$ and $b = 2$ is a poor choice for mean shifts $\Delta = 1, 1.4$. This is illustrated in Figure 7 (left), which plots the distribution of $(\mu_i - \bar{x}_i)/\hat{\sigma}_i^2 | \delta_i$ together with the estimated in-control distribution of the standardized sample mean $(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i | (\mu_i = \bar{x}_i, \sigma_i = \hat{\sigma}_i)$. It can be observed that shifts of sizes $\hat{\sigma}_i$ and $1.4\hat{\sigma}_i$ (vertical lines) are unlikely under this prior since the two vertical lines fall far out in the tail of the prior for the increased mean. Hence, the performance of these choices is poor when $\Delta = 1$ or $\Delta = 1.4$, as shown in Table 2. In particular, we observe that the Type I error rates are very large compared with the values in Table 1. An example of a good prior specification ($h = \sqrt{6}$ and $b = 0.5$) is given on the right side of Figure 7. In this case, shifts of sizes $\Delta = 1, 1.4$ are in regions of concentration of the prior density for the increased/decreased mean.

6.2 Example 2: Comparison With LEB

In this example, we compare our method with the LASSO-based diagnostic procedure LEB introduced by Zou, Jiang, and Tsung (2011). In section 4.1 of their article, they considered a few simulated cases that involve only mean shifts. In these cases, $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = (1, 1, 0, \ldots, 0)^T$ and $\boldsymbol{\Sigma}_0 = (0.5^{|i-j|})$. We shall compare our approach with LEB for four cases of (N, n, p): (50, 25, 4), (1000, 25, 4), (100, 50, 6), and (1000, 50, 6).

Note that Zou, Jiang, and Tsung (2011) applied their method to diagnose changes in the elements of the covariance matrix as well as the means. Thus, to ensure a fair comparison, our method is combined with the likelihood-ratio test of equality of two covariance matrices (Timm 2002). We assume that if the null hypothesis is rejected, then the means are declared in control but the covariance matrix is declared out of control. For comparison with the results in Zou, Jiang, and Tsung (2011), we measure the performance of the *combined method* with two quantities: the relative frequency with which the combined method yields perfectly correct decisions on both means and covariance matrix (C), and the expected number of errors in mean shift decisions (ENEM). Note that Zou, Jiang, and Tsung (2011) employed the expected number of errors in the decisions on all parameters ENE as a performance measure instead of ENEM. However, this performance measure cannot be adopted here because we would need a diagnostic method for identifying the elements in the covariance matrix

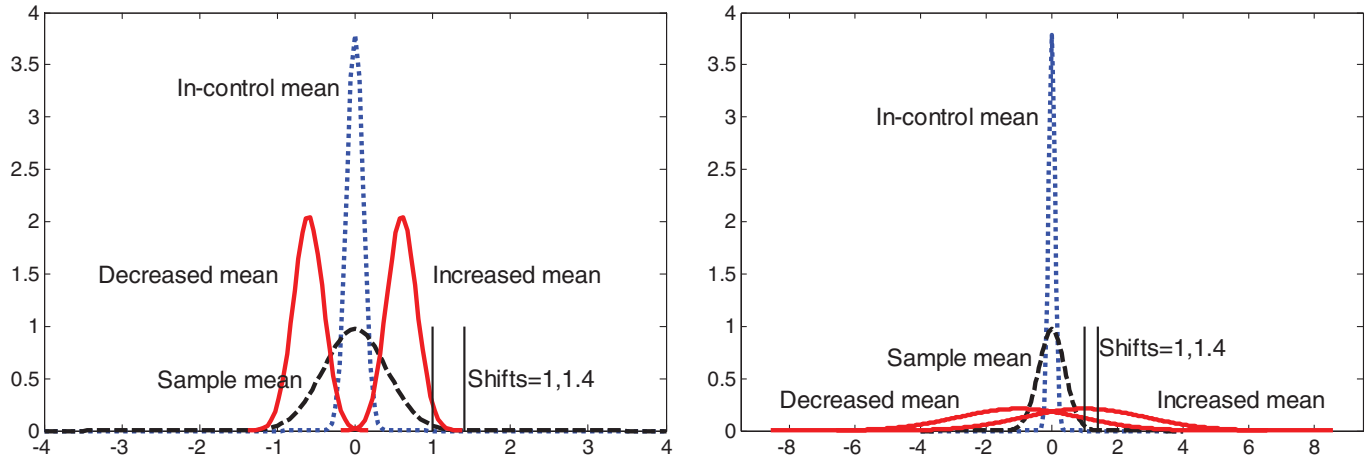

Figure 7. Poor prior specification (left), and good prior specification (right). The online version of this figure is in color.

that shifted. Nevertheless, if the combined method is enhanced to include a diagnostic procedure for the covariance matrix, we will have ENE $\in$ [ENEM, ENEM $+ \alpha p(p+1)/2$], where α is the significance level of the test for the covariance matrix and $p(p+1)/2$ is the number of elements in the covariance matrix.

Table 3 summarizes the results for four different choices of (h, b), that is, (3, 0.5),(3, 1),(6, 0.5), and (6, 1), and the PEB method with $b = 0.5$ and $b = 1$. The table also reproduces the performance estimates for LEB given by Zou, Jiang, and Tsung (2011). The number of simulations for each prior is 200 and we set $N_{\text{iter}} = 3000$ and $\tau = 1000$. For $n = 25$, the shift sizes are $5\sigma_i/\sqrt{n}, i = 1, 2$, and for $n = 50$, the shift sizes are $7.07\sigma_i/\sqrt{n}, i = 1, 2$. Thus, $h = 3$ and $h = 6$ can be thought of as poor and moderately good guesses of the mean shifts, respectively. The significance level of the covariance matrix test is fixed at 0.05 and the cutoff point is obtained from a chi-squared approximation. Simulation suggests that this approximation is accurate. Thus, for our approach, ENE $\in$ [ENEM, ENEM + 0.5] for $p = 4$ and ENE $\in$ [ENEM, ENEM + 1.05] for $p = 6$. For the purpose of comparison, we set ENE = ENEM + 0.25 for $p = 4$ and ENE = ENEM + 0.525 for $p = 6$, which are perhaps worst-case estimates.

We observe that except for the combination of $(N, n, p) = (100, 50, 6)$ and $(h, b) = (3, 1)$, the performances of the six priors are significantly better than LEB (larger values of C and smaller values of ENE are preferred). Some of the best results are obtained with the PEB method. The marked decrease in performance when $(N, n, p) = (100, 50, 6)$ and $(h, b) = (3, 1)$ is because $(h, b) = (3, 1)$ yields a somewhat informative prior centered far from the true shift, and $N = 100$ is not large enough for accurately estimating the in-control parameters. It can be concluded from the results in Table 3 that for processes that frequently suffer from mean shifts, it is better to use the combined approach than to use the LEB procedure to perform a simultaneous check on all parameters for a shift.

We now apply the proposed approach (without the covariance matrix test) to the wine quality control example discussed by Zou, Jiang, and Tsung (2011), which is based on a real dataset. This problem involves 11 variables that are measurements from various physicochemical tests. Zou, Jiang, and Tsung (2011) showed that the correct change point (which is known) can be found with a change-point method. We take all observations after the change point as the out-of-control sample. Thus, we have $N = 870$ phase I samples, and $n = 11$ out-of-control samples. We compute $\lambda_i = \sqrt{n}(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i$ for $i = 1, \ldots, 11$, and found that the four largest λ_i are for $i = 4, 5, 8, 11$ with values 1.95, 3.32, 3.88, and -4.38, respectively. The fifth largest λ_i only has magnitude 1.45. Eight priors given by $(h, b) \in \{2, 3, 5, 7\} \times \{1, 1.5\}$ and the EB method are tried. For the EB method, we set $I = \{4, 5, 8, 11\}$. All nine prior distributions

Table 3. Estimates of performance measures C, ENEM, and ENE for six prior distribution choices and LEB

		(N, n, p)							
		(50,25,4)		(1000,25,4)		(100,50,6)		(1000,50,6)	
h	b	C	ENEM (ENE)	C	ENEM (ENE)	C	ENEM (ENE)	C	ENEM (ENE)
3	0.5	0.81	0.28 (0.53)	0.91	0.15 (0.40)	0.88	0.16 (0.69)	0.84	0.25 (0.78)
3	1	0.80	0.30 (0.55)	0.86	0.20 (0.45)	0.69	0.42 (0.94)	0.80	0.25 (0.78)
6	0.5	0.86	0.22 (0.47)	0.90	0.19 (0.44)	0.90	0.17 (0.69)	0.95	0.10 (0.63)
6	1	0.88	0.19 (0.44)	0.89	0.18 (0.43)	0.83	0.27 (0.79)	0.90	0.15 (0.67)
PEB	0.5	0.88	0.17 (0.42)	0.92	0.13 (0.38)	0.92	0.16 (0.68)	0.93	0.12 (0.65)
PEB	1	0.82	0.24 (0.49)	0.87	0.18 (0.43)	0.87	0.21 (0.73)	0.90	0.16 (0.68)
		C	ENE	C	ENE	C	ENE	C	ENE
LEB		0.37	1.04	0.36	1.26	0.51	0.85	0.50	0.94

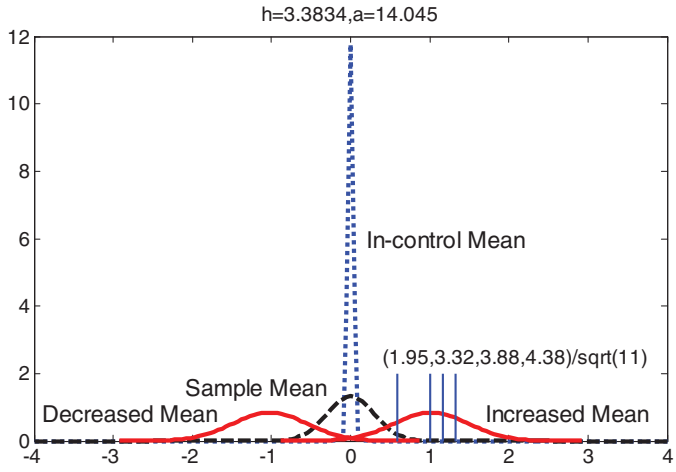

Figure 8. Plot of prior distributions obtained via the EB method and density of $(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i | (\mu_i = \bar{x}_i, \sigma_i = \hat{\sigma}_i)$. The online version of this figure is in color.

result in appreciable probability for $(\mu_i - \bar{x}_i)/\hat{\sigma}_i | (\delta_i = 1)$ over the interval $[1.95/\sqrt{11}, 4.38/\sqrt{11}]$ (see Figure 8 for a plot of the priors obtained via the EB method). The results obtained for all priors are similar. All nine posterior modes of $\boldsymbol{\delta}$ are identical, with a 1 in the fifth and eighth positions, a -1 in the eleventh position, and zeros elsewhere. The results are obtained with $N_{\text{iter}} = 20{,}000$ and $\tau = 10{,}000$. For the EB method, the marginal posterior distributions of $\delta_1, \ldots, \delta_{11}$ are plotted in Figure 9. The results clearly indicate that the fifth and eighth means shifted upward and the eleventh mean shifted downward. This agrees with the conclusions reached by Zou, Jiang, and Tsung (2011) using LEB.

Finally, note that although we set $I = \{4, 5, 8, 11\}$, the EB method identifies a shift in μ_5, μ_8, and μ_{11} but not in μ_4. Thus, the choice of I does not determine the means that will be identified as shifted.

6.3 Example 3: Fruit Juice Data

We analyze process monitoring data collected from a fruit juice process given in table 4.3a of Beltran (2006). The data, which is in Appendix C (available in the online supplementary materials) of this article, consist of the concentrations in microgram per standard volume of $p = 11$ amino acids (names of the amino acids are given in Appendix C, available online). We take the phase I observations as the first 25 observations and the phase II observations as the remaining $n = 11$ observations. However, there are some phase I observations that plot outside the 95% control limit of a T^2 chart for the phase I observations (see Bersimis, Psarakis, and Panaretos 2007 for a formula for the control limit). We removed the four observations (rows 16, 20, 22, and 25 of the data in Appendix C) that fall above the control limit, reconstructed the control chart, and further removed two observations (rows 11 and 12) that fall above the revised control limit. The remaining observations all appear to be in control. Thus, we have $N = 19$ in-control phase I samples. A T^2 chart for the fruit juice process (without the T^2 value for observations that were removed) is plotted in Figure 10. We observe that four of the phase II samples plot above the 95% UCL.

Ten of the values of $\lambda_i = \sqrt{n}(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i$, for $i = 1, \ldots, 11$, exceed two in absolute value; only $|\lambda_3| < 2$. Taking $I = \{1, 2, 4, \ldots, 10\}$, we find that $h^{\text{EB}} = 3.58$ and $a^{\text{EB}} = 2.01 = \sqrt{N}\,\text{stdev}\{\frac{(\bar{x}_{fi} - \bar{x}_i)}{\hat{\sigma}_i}, i \in I\} > \frac{h^{\text{EB}}}{2}\sqrt{\frac{N}{n}} - 1 = 1.35$. The posterior mode of $\boldsymbol{\delta}$ ($N_{\text{iter}} = 20{,}000$ and $\tau = 10{,}000$) is $(0, 0, 0, 1, 0, 0, 0, 1, -1, -1, 0)$ and it has a probability of about 0.3. The marginal posterior distributions of the δ_i's, which are plotted in Figure 11, give the same information. This indicates that means 4 and 8 shifted upward, whereas means 9 and 10 shifted downward. Although $\lambda_3 = 0.141$ is the smallest in absolute value, the posterior probability that $\delta_3 = -1$ is quite high (about 0.3). This can be explained as follows. Variable 3 has positive correlations with variables 4 and 8 but negative correlations with variables 9 and 10. Thus, if it is assumed that one or more of means 4, 8, 9, and 10 did not shift so that the unusually large or small values of $\bar{x}_{fi}, i = 4, 8, 9, 10$, are explained by natural variation, we would expect λ_3 to be quite large and positive. Since λ_3 is small, mean 3 could have shifted downward. We also tried $(h, b) = (2.5, 0.8), (5, 1)$. For these

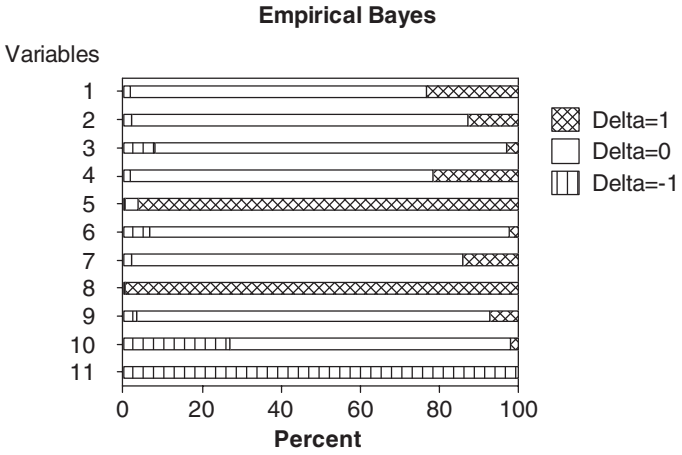

Figure 9. Marginal posterior distribution of each indicator for the EB method—wine data.

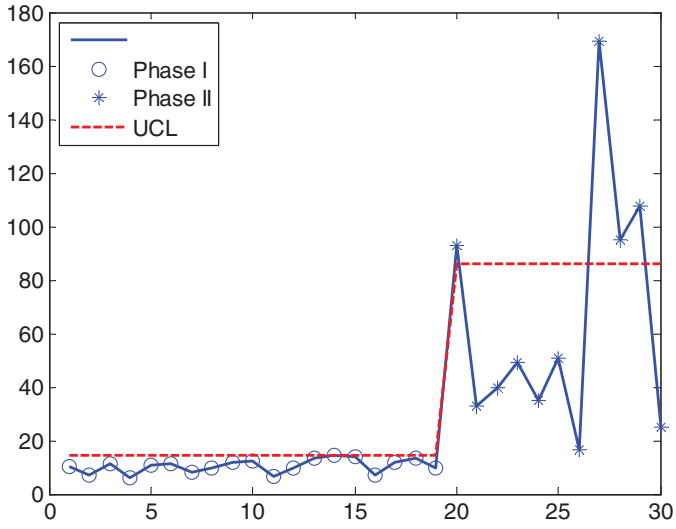

Figure 10. T^2 chart for fruit juice data. The online version of this figure is in color.

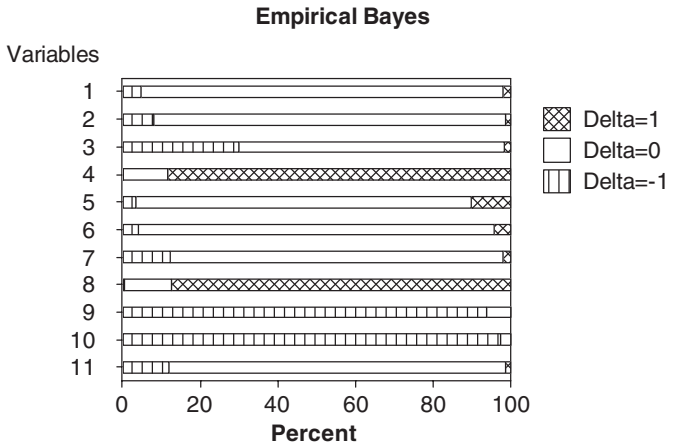

Figure 11. Marginal posterior distribution of each indicator for the EB method—fruit juice data.

priors, the shift decisions given by decision rules 1 and 2 are identical to those obtained with the EB method.

Appendix D (available in the online supplementary materials) gives a plot of the observations for each variable. The figure gives some indication that the means of variables 4, 9, and 10 shifted. However, it seems hard to tell by a visual inspection of the figure whether the mean of variable 8 shifted.

7. CONCLUSIONS

We have proposed a Bayesian approach to identify the means that shifted and the direction of the shifts when a control chart for the mean of normal variables signals, but the data indicate that the covariance matrix remain in control. We introduce an indicator variable for each mean whose values $-1, 0, 1$ indicate whether the mean shifted downward, remained in control, or shifted upward. The prior for each mean conditioned on its indicator captures prior information about the in-control and shifted states of the mean. The Bayesian hierarchical model is specified by prior distributions for the shifts, the indicators of the shift directions, and the covariance matrix. The prior distributions for the in-control mean and the inverse covariance matrix are derived from phase I data. Assumptions are made to simplify the prior distributions and guidelines are given to choose the prior parameters effectively. A Gibbs sampling algorithm for sampling from the posterior distribution of the vector of indicators is given.

We propose two decision rules to identify the most probable state of each mean. The first is to pick the value of each indicator with the highest marginal posterior probability. The second is to pick the posterior mode of the vector of indicators. These quantities can easily be estimated by their sample counterparts. Monte Carlo simulation shows that the proposed Bayesian approach always performs better than the t-test and can have performance comparable with the best subset variant of WJPLM with correctly specified number of mean shifts. Moreover, it can also outperform the LEB approach when shifts in the mean are more common than shifts in the covariance matrix. In all examples, the EB or PEB method for specifying the prior for the mean gives good results. The EB and PEB methods are attractive because the former is entirely data driven and the latter only requires the user to specify one prior parameter. For these reasons, we recommend the EB and PEB methods for practical use.

We have considered modeling the mean with a mixture of three normal distributions. However, it is straightforward to extend our work to include modeling the increased and decreased means with truncated normal distributions. Truncated prior distributions may yield better results. If we let $p(\boldsymbol{\mu}|\boldsymbol{\delta}) \propto |\boldsymbol{\psi}_\delta|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\theta}_\delta)^T \boldsymbol{\psi}_\delta^{-1}(\boldsymbol{\mu} - \boldsymbol{\theta}_\delta)\} I_{\boldsymbol{\Xi}_\delta}(\boldsymbol{\mu})$, where $I_{\boldsymbol{\Xi}_\delta}(\boldsymbol{\mu}) = 1$ if $\boldsymbol{\mu} \in \boldsymbol{\Xi}_\delta$ and $I_{\boldsymbol{\Xi}_\delta}(\boldsymbol{\mu}) = 0$ otherwise, and $\boldsymbol{\Xi}_\delta$ is a hyperrectangular set that depends on $\boldsymbol{\delta}$, then the full conditional distribution of each μ_i is truncated normal. Hence, we can sample from the posterior distribution of $\boldsymbol{\delta}$ using the Gibbs sampling algorithm given in Figure 3 with a modified Step 3. The modified Step 3 is to sample from $\mu_1, \ldots, \mu_p$ one at a time, where each μ_i is sampled from a truncated normal distribution.

Finally, we mention that we are researching the topic of incorporating prior knowledge about causal relationships between variables through the prior $p(\boldsymbol{\delta})$.

SUPPLEMENTARY MATERIALS

A document (pdf) with five appendices providing the following details:

Appendix A: Additional simulation results: an example that compares Type I and Type II error rates for various priors, including asymmetric ones, with WJPLM and t-test.

Appendix B: Covariance matrices used in the simulations reported in this article.

Appendix C: Data for Example 3.

Appendix D: Time series plots for the data analyzed in Example 3.

Appendix E: MATLAB code for implementing the EB approach.

ACKNOWLEDGMENTS

Tan's research was supported by NSF grants DMS0705261 and DMS1007574, and Shi's research was partially supported by NSF grant CMMI-0927574. We thank the editor and two anonymous referees for many valuable comments that helped improve the article significantly.

[Received February 2011. Revised April 2012.]

REFERENCES

Barnard, J., McCulloch, R., and Meng, X. (2000), "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, With Application to Shrinkage," *Statistica Sinica*, 10, 1281–1311. [302]

Beltran, L. A. (2006), "Nonparametric Multivariate Statistical Process Control Using Principal Component Analysis and Simplicial Depth," Ph.D. dissertation, Department of Industrial and Management Systems, University of Central Florida. [301,305]

Bersimis, S., Psarakis, S., and Panaretos, J. (2007), "Multivariate Statistical Process Control Charts: An Overview," *Quality and Reliability Engineering International*, 23, 517–543. [294,305]

Box, G. E. P., and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley. [296]

Capizzi, G., and Masarotto, G. (2011), "A Least Angle Regression Control Chart for Multidimensional Data," *Technometrics*, 53, 285–296. [295]

Das, N., and Prakash, V. (2008), "Interpreting the Out-of-Control Signal in Multivariate Control Chart—A Comparative Study," *International Journal of Advanced Manufacturing Technology*, 37, 966–979. [294]

Doganaksoy, N., Faltin, F. W., and Tucker, W. T. (1991), "Identification of Out-of-Control Quality Characteristics in a Multivariate Manufacturing Environment," *Communications in Statistics—Theory and Methods*, 20, 2775–2790. [294]

Fuchs, C., and Kenett, R. (1998), *Multivariate Quality Control*, New York: Marcel Dekker. [301]

Furnival, G. M., and Wilson, R. W. (1974), "Regressions by Leaps and Bounds," *Technometrics*, 16, 499–511. [301]

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409. [300]

George, E. I., and McCulloch, R. E. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889. [295,296,300]

——— (1997), "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–373. [296,300]

Hawkins, D. M. (1991), "Multivariate Quality Control Based on Regression-Adjusted Variables," *Technometrics*, 33, 61–75. [294]

Jellison, J. L. (1975), "Effect of Surface Contamination on the Thermocompression Bondability of Gold," *IEEE Transactions on Parts, Hybrids, and Packaging*, 11, 206–211. [298]

Li, J., Jin, J., and Shi, J. (2008), "Causation-Based T^2 Decomposition for Multivariate Process Monitoring and Diagnosis," *Journal of Quality Technology*, 40, 46–58. [294]

Mason, R. L., Tracy, N. D., and Young, J. C. (1995), "Decomposition of T^2 for Multivariate Control Chart Interpretation," *Journal of Quality Technology*, 27, 99–108. [294]

——— (1997), "A Practical Approach for Interpreting Multivariate T^2 Control Chart Signals," *Journal of Quality Technology*, 29(1), 396–406. [294]

Montgomery, D. C. (2009), *Introduction to Statistical Quality Control* (6th ed.), New York: Wiley. [299]

Murphy, B. J. (1987), "Selecting Out of Control Variables With the T^2 Multivariate Quality Control Procedure," *The Statistician*, 36, 571–581. [294]

Runger, G. C., Alt, F. B., and Montgomery, D. C. (1996), "Contributors to a Multivariate Statistical Process Control Chart Signal," *Communications in Statistics—Theory and Methods*, 25, 2203–2213. [294]

Timm, N. H. (2002), *Applied Multivariate Analysis*, New York: Springer-Verlag. [295,303]

Wang, K., and Jiang, W. (2009), "High-Dimensional Process Monitoring and Fault Isolation via Variable Selection," *Journal of Quality Technology*, 41, 247–258. [295,301]

Wu, C. F. J., and Hamada, M. S. (2009), *Experiments: Planning, Analysis, and Optimization* (2nd ed.), New York: Wiley. [298]

Yeh, A. B., Lin, D. K. J., and McGrath, R. N. (2006), "Multivariate Control Charts for Monitoring Covariance Matrix: A Review," *Quality Technology and Quantitative Management*, 3, 415–436. [294,295]

Zhang, Y., Krueger, D., Durst, R., Lee, R., Wang, D., Seeram, N., and Heber, D. (2009), "International Multidimensional Authenticity Specification (IMAS) Algorithm for Detection of Commercial Pomegranate Juice Adulteration," *Journal of Agricultural and Food Chemistry*, 57, 2550–2557. [294]

Zou, C., Jiang, W., and Tsung, F. (2011), "A LASSO-Based Diagnostic Framework for Multivariate Statistical Process Control," *Technometrics*, 53, 297–309. [295,301,303,304]

Zou, C., and Qiu, P. (2009), "Multivariate Statistical Process Control Using LASSO," *Journal of the American Statistical Association*, 104, 1586–1596. [295]